Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL
Pacific Journal of Optimization
Vol. 14, No. 3, pp. 489–506, 2018

CONVERGENCE OF ADMM FOR OPTIMIZATION PROBLEMS WITH NONSEPARABLE NONCONVEX OBJECTIVE AND LINEAR CONSTRAINTS*

Ke Guo, Deren Han† and Tingting Wu

Abstract: In this paper, we consider the convergence properties of the alternating direction method of multipliers (ADMM) for solving linearly constrained nonconvex minimization model whose objective contains coupled functions. Under the assumption that the associated functions satisfy the Kurdyka-Łojasiewicz inequality, we prove that the iterative sequence generated by the ADMM converges to a critical point of the augmented Lagrangian function, when the penalty parameter in the augmented Lagrangian function is sufficiently large. This result recovers that in [19] when the coupled term in the objective function vanishes, i.e., the range of the penalty parameter that ensure the convergence of ADMM for solving two-block separable nonconvex optimization problem. We also analyze the convergence rate of the algorithm under some suitable conditions on the problem's data. Some extensions are also presented.

Key words: *alternating direction method of multipliers, Kurdyka-Lojasiewicz inequality, nonconvex optimization, coupled objective function, nonsmooth analysis*

Mathematics Subject Classification: *90C26, 65K10, 49J52, 49M27*

1 Introduction

In this paper, we consider the following nonconvex optimization problem

$$\begin{aligned} \min \quad & f_1(x) + f_2(y) + H(x,y) \\ s.t. \quad & Ax + y = b, \end{aligned} \tag{1.1}$$

where $f_1 : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ is a proper lower semicontinuous function, $f_2 : \mathcal{R}^m \to \mathcal{R}$ is a continuously differentiable function whose gradient ∇f_2 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. Problem (1.1) captures a number of important applications arising in various areas, for example, the control of a smart grid system [1, 7, 26], the appliance load model [31], cognitive radio network [33, 36]; just mention a few and the interested readers can refer to [21] for more discussion.

Let $\beta > 0$ be a given parameter. Define the augmented Lagrangian function for problem (1.1) as

$$\mathcal{L}_\beta(x,y,\lambda) := f_1(x) + f_2(y) + H(x,y) - \lambda^T(Ax + y - b) + \frac{\beta}{2}\|Ax + y - b\|^2, \tag{1.2}$$

*This research is supported by the Natural Science Foundation of China (Grant No. 11801455, 11625105, 11371197 and 11431002) and the Fundamental Research Funds of China West Normal University (Grant No. 17E084, 18B031).

where λ is the Lagrangian multiplier associated with the linear constraints. Based on alternately optimizing the augmented Lagrangian function $\mathcal{L}_\beta(\cdot)$ for one variable but with the others fixed, the alternating direction method of multipliers (ADMM), originally proposed in [17], generates the iterative sequence with the following recursion:

$$\begin{cases} x^{k+1} \in \operatorname{argmin}_x\{\mathcal{L}_\beta(x, y^k, \lambda^k)\}, \\ y^{k+1} \in \operatorname{argmin}_y\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k)\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b), \end{cases} \tag{1.3}$$

starting with $(y^0, \lambda^0) \in \operatorname{dom} f_2 \times \mathcal{R}^m$. Here, $\operatorname{dom} f_2$ denotes the domain of f_2.

When f_1, f_2 are proper lower semicontinuous convex functions and $H \equiv 0$, the convergence of the two-block ADMM has long been established, we refer the readers to [10, 11, 12, 14, 15, 17, 20, 34] for a discussion about convergence and convergence rate in the convex setting. As mentioned in [22], when the objective functions are not separable across the variables, the convergence of ADMM for the general problem (1.1) is still an open question, even the objective functions are convex. Recently, Gao and Zhang [16] considered the case where H is a smooth convex function and f_1, f_2 are convex functions. Under the assumptions that ∇H is Lipschitz continuous and f_2 is strongly convex, they proved the sequence generated by the proximal ADMM for problem (1.1) converges to an optimal solution. Chen et al. [8] analyzed the convergence of the two-block ADMM for problem (1.1) with coupled quadratic objective function.

In this paper, we consider the problem (1.1) without assuming the convexity of f_1, f_2, and the coupled term H. A very important technique to prove the convergence for nonconvex optimization problems is assuming the objective function satisfying the Kurdyka-Łojasiewicz (KL) inequality (See Definition 2.4). Using the important KL inequality, we prove that if the augmented Lagrangian function is a KL function, then the sequence generated by the ADMM (1.3) converges to a critical point of the augmented Lagrangian function. We refer the reader to [2, 3, 5, 13, 18, 19, 23, 24, 30] for the recent literature concerning the convergence of nonconvex optimization problem relying on the KL inequality.

The rest of this paper is organized as follows. In section 2, we summarize some preliminary materials and useful results for further analysis. In section 3, we present the convergence of the scheme (1.3), where the convergence rate is also provided under some further assumptions. Some extensions are made in section 4. Finally, we give some concluding remarks.

2 Preliminaries

In this section, we first introduce some notations that will be frequently used in the analysis later.

Let $x \in \mathcal{R}^n$, $y \in \mathcal{R}^m$, we denote $v := (x, y) \in \mathcal{R}^n \times \mathcal{R}^m$. Throughout this paper, we use the convention $0 \cdot \infty = 0$. 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.

We now recall a few definitions concerning subdifferential calculus for nonsmooth functions, see, e.g. [28, 32] for more details.

Given a function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$, we define its epigraph by

$$\text{epi } f := \{(x, \alpha) \in \mathcal{R}^n \times \mathcal{R} : f(x) \leq \alpha\}.$$

We say that the function f is proper (respectively, lower semicontinuous) if the above set is nonempty (respectively, closed). The domain of the function f is defined by

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

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 $x \in \text{dom } f$, written $\hat{\partial} f(x)$, is the set of vectors $x^* \in \mathcal{R}^n$ that satisfy

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

When $x \notin \text{dom} f$, we set $\hat{\partial} f(x) := \emptyset$.
(ii). The limiting-subdifferential, or simply the subdifferential, of f at $x \in \text{dom } f$, 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 proper lower semicontinuous and $g : \mathcal{R}^n \to \mathcal{R}$ is continuous 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. The set of critical points of f is denoted by crit f.

Let us recall an important property of subdifferential calculus.

Lemma 2.3 ([2])**.** *Suppose that $H(x, y, z) := f(x) + g(y) + h(z)$, where $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$, $g : \mathcal{R}^m \to \mathcal{R} \cup \{+\infty\}$ and $h : \mathcal{R}^p \to \mathcal{R} \cup \{+\infty\}$ are proper lower semicontinuous functions. Then for all $(x, y, z) \in$ dom $H =$ dom $f \times$ dom $g \times$ dom h, we have*

$$\partial H(x, y, z) = \partial_x H(x, y, z) \times \partial_y H(x, y, z) \times \partial_z H(x, y, z).$$

The Kurdyka-Łojasiewicz property plays a central role in our analysis. Below, we recall some essential elements.

Definition 2.4 ([2] **Kurdyka-Łojasiewicz inequality**)**.** Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leq +\infty$, set

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

We say the function f has the KL property at $x^* \in \text{dom}\, \partial f$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of x^*, and a continuous concave function $\varphi : [0, \eta) \to \mathcal{R}_+$, such that
(i). $\varphi(0) = 0$;
(ii). φ is C^1 on $(0, \eta)$ and continuous at 0;
(iii). $\varphi'(s) > 0, \forall s \in (0, \eta)$;
(iv). for all x in $U \cap [f(x^*) < f < f(x^*) + \eta]$, the KL inequality holds:

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

Definition 2.5 ([3] **KL function**)**.** Denote Φ_η be the set of functions which satisfy (i), (ii) and (iii). If f satisfies the KL property at each point of dom ∂f, then f is called a KL function.

Remark 2.6. *One can easily check that the KL property is automatically satisfied at any noncritical point $x^* \in$ dom f; e.g., Lemma 2.1 and Remark 3.2 (b) of [2].*

Lemma 2.7 ([5] **Uniformized KL property**)**.** *Let Ω be a compact set and let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper 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,

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

Definition 2.8 ([4])**.** A proper lower semicontinuous function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ is called semiconvex with constant $\omega \geq 0$ if the function

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

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

Remark 2.9. (i). Definition 2.8 is equivalent to

$$f(y) \geq f(x) + \langle p, y - x\rangle - \frac{\omega}{2}\|y - x\|^2, \tag{2.2}$$

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

(ii). 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 [9] and primal-lower-nice functions [27]. Moreover, it is easily seen that every twice continuous differentiable function f with bounded second derivative is semiconvex. For general properties of semiconvex functions, see, e.g. [4, 6].

Definition 2.10. We say (x^*, y^*, λ^*) is a critical point of the augmented Lagrangian function $\mathcal{L}_\beta(\cdot)$ (1.2), if it satisfies:

$$\begin{cases} A^T\lambda^* - \nabla_x H(x^*, y^*) \in \partial f_1(x^*), \\ \lambda^* - \nabla_y H(x^*, y^*) = \nabla f_2(y^*), \\ Ax^* + y^* - b = 0. \end{cases} \tag{2.3}$$

The following lemma is useful in the derivation of the main results.

Lemma 2.11 ([29]). *Let $h : \mathcal{R}^n \to \mathcal{R}$ be a continuously differentiable function whose gradient ∇h is Lipschitz continuous with constant $L > 0$, then for any $x, y \in \mathcal{R}^n$, we have*

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

For any two vectors x and y with the same dimensions, we have

$$\|x + y\|^2 \leq (1 + \xi)\|x\|^2 + (1 + \frac{1}{\xi})\|y\|^2, \ \forall \xi > 0. \tag{2.4}$$

3 Convergence

In this section, we prove the convergence of the ADMM (1.3) under the following assumption.

Assumption 3.1. *Let $f_1 : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a semiconvex function with constant $\omega > 0$, $f_2 : \mathcal{R}^m \to \mathcal{R}$ be a continuously differentiable function whose gradient ∇f_2 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 holds:*

(i) $\inf_{(x,y) \in \mathcal{R}^n \times \mathcal{R}^m} H(x, y) > -\infty, \ \inf_{x \in \mathcal{R}^n} f_1(x) > -\infty, \ \inf_{y \in \mathcal{R}^m} f_2(y) > -\infty;$

(ii) *For any fixed x, the partial gradient $\nabla_y H(x, y)$ is globally Lipschitz with constant $L_2(x)$, that is*

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

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

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

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

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

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

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

(v) *$A^T A \succeq \mu I$ for some $\mu > 0$;*

(vi)

$$\beta > \max\{\frac{(L_3 + \omega) + \sqrt{(L_3 + \omega)^2 + 16\mu M^2}}{2\mu}, \frac{(L_1 + L_2) + \sqrt{(L_1 + L_2)^2 + 16(L_1^2 + M^2)}}{2}\}. \tag{3.1}$$

Note that, if we set

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

we know $\delta > 0$ in view of (vi) of Assumption 3.1.

Remark 3.1. Now, we specify the conditions in Assumption 3.1 to some special cases of f_1, f_2 and H. For example, the SCAD-ℓ_2 model model proposed in [35] can be rewritten in the following equivalent form:

$$\begin{aligned} \min_{x,y} \quad & \|Mx - My - b\|^2 + \sum_{i=1}^{n} g_\lambda(|x_i|) + \|y\|^2, \\ s.t. \quad & x + y = 0. \end{aligned} \tag{3.2}$$

where

$$g_\lambda(|x_i|) := \begin{cases} \lambda|x_1^i|, & |x^i 1| \le \lambda, \\ \frac{-(x_1^i)^2 + 2a\lambda|x_1^i| - \lambda^2}{2(a-1)}, & \lambda < |x_1^i| \le a\lambda, \\ \frac{(a+1)\lambda^2}{2}, & |x_1^i| > a\lambda. \end{cases}$$

If we set $f_1(x) := \sum_{i=1}^n g_\lambda(|x_i|)$, $f_2(y) := \|y\|^2$, $H(x,y) := \|Mx - My - b\|^2$ and $A := I$, then problem (3.2) fits our scenario. Since $\nabla_x H(x,y) = 2M^T(Mx - My - b)$ and $\nabla_y H(x,y) = -2M^T(Mx - My - b)$, it is easy to know the conditions (i)-(v) in Assumption 3.1 hold.

Before the proof, let us present the variational characterization of the scheme (1.3). Invoking the optimality condition for (1.3), we have

$$\begin{cases} 0 \in \partial f_1(x^{k+1}) + \nabla_x H(x^{k+1}, y^k) - A^T\lambda^k + \beta A^T(Ax^{k+1} + y^k - b), \\ 0 = \nabla f_2(y^{k+1}) + \nabla_y H(x^{k+1}, y^{k+1}) - \lambda^k + \beta(Ax^{k+1} + y^{k+1} - b), \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b). \end{cases} \tag{3.3}$$

Using the last equality and rearranging terms, we obtain

$$\begin{cases} A^T\lambda^{k+1} + \beta A^T(y^{k+1} - y^k) - \nabla_x H(x^{k+1}, y^k) \in \partial f_1(x^{k+1}), \\ \lambda^{k+1} - \nabla_y H(x^{k+1}, y^{k+1}) = \nabla f_2(y^{k+1}), \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b). \end{cases} \tag{3.4}$$

In the sequel for convenience, we often use the notation $\{w^k := (x^k, y^k, \lambda^k)\}_{k\in N}$ and $\{v^k := (x^k, y^k)\}_{k\in N}$. We begin our analysis with the following lemma.

Lemma 3.2. *Let $\{w^k\}_{k\in N}$ be the sequence generated by the ADMM (1.3) which is assumed to be bounded, then we have*

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

Proof. From the definition of the augmented Lagrangian function $\mathcal{L}_\beta(\cdot)$, we have

$$\begin{aligned} \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}) &= \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) + \langle \lambda^k - \lambda^{k+1}, Ax^{k+1} + y^{k+1} - b \rangle \\ &= \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2. \end{aligned} \tag{3.6}$$

Note that

$$\begin{aligned} &\mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) \\ &= f_1(x^{k+1}) + f_2(y^k) + H(x^{k+1}, y^k) - \langle \lambda^k, Ax^{k+1} + y^k - b \rangle \\ &\quad + \frac{\beta}{2}\|Ax^{k+1} + y^k - b\|^2 - \{f_1(x^{k+1}) + f_2(y^{k+1}) + H(x^{k+1}, y^{k+1}) \\ &\quad - \langle \lambda^k, Ax^{k+1} + y^{k+1} - b \rangle + \frac{\beta}{2}\|Ax^{k+1} + y^{k+1} - b\|^2\} \end{aligned}$$

$$= f_2(y^k) - f_2(y^{k+1}) + \langle \lambda^k, y^{k+1} - y^k \rangle + \frac{\beta}{2}\|Ax^{k+1} + y^k - b\|^2 - \frac{\beta}{2}\|Ax^{k+1} + y^{k+1} - b\|^2$$
$$+H(x^{k+1}, y^k) - H(x^{k+1}, y^{k+1}). \tag{3.7}$$

By the Lipschitz continuity of ∇f_2, it follows from Lemma 2.11 and the second equality of (3.4) that

$$f_2(y^k) - f_2(y^{k+1}) \geq \langle \lambda^{k+1} - \nabla_y H(x^{k+1}, y^{k+1}), y^k - y^{k+1} \rangle - \frac{L_1}{2}\|y^{k+1} - y^k\|^2. \tag{3.8}$$

By simple manipulations and using $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b)$, we know

$$\frac{\beta}{2}\|Ax^{k+1} + y^k - b\|^2 - \frac{\beta}{2}\|Ax^{k+1} + y^{k+1} - b\|^2$$
$$= \frac{\beta}{2}\|y^{k+1} - y^k\|^2 + \langle \lambda^k - \lambda^{k+1}, y^k - y^{k+1} \rangle. \tag{3.9}$$

Since $\nabla_y H(x^{k+1}, \cdot)$ is Lipschitz with constant $L_2(x^{k+1})$, it follows from Lemma 2.11 that

$$H(x^{k+1}, y^k) - H(x^{k+1}, y^{k+1}) \geq \langle \nabla_y H(x^{k+1}, y^{k+1}), y^k - y^{k+1} \rangle - \frac{L_2(x^{k+1})}{2}\|y^{k+1} - y^k\|^2. \tag{3.10}$$

Substituting (3.8), (3.9) and (3.10) into (3.7) yields

$$\mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) \geq \frac{\beta - L_1 - L_2(x^{k+1})}{2}\|y^{k+1} - y^k\|^2. \tag{3.11}$$

On the other hand,

$$\mathcal{L}_\beta(x^k, y^k, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k)$$
$$= f_1(x^k) + f_2(y^k) + H(x^k, y^k) - \langle \lambda^k, Ax^k + y^k - b \rangle + \frac{\beta}{2}\|Ax^k + y^k - b\|^2$$
$$-\{f_1(x^{k+1}) + f_2(y^k) + H(x^{k+1}, y^k) - \langle \lambda^k, Ax^{k+1} + y^k - b \rangle + \frac{\beta}{2}\|Ax^{k+1} + y^k - b\|^2\}$$
$$= f_1(x^k) - f_1(x^{k+1}) + \langle \lambda^k, Ax^{k+1} - Ax^k \rangle + \frac{\beta}{2}\|Ax^k + y^k - b\|^2 - \frac{\beta}{2}\|Ax^{k+1} + y^k - b\|^2$$
$$+H(x^k, y^k) - H(x^{k+1}, y^k). \tag{3.12}$$

Combining (2.2) and the first relation of (3.4), and using the semiconvexity of f_1, we have

$$f_1(x^k) \geq f_1(x^{k+1}) + \langle A^T \lambda^{k+1} + \beta A^T(y^{k+1} - y^k) - \nabla_x H(x^{k+1}, y^k), x^k - x^{k+1} \rangle - \frac{\omega}{2}\|x^{k+1} - x^k\|^2.$$

Again since $\nabla_x H(\cdot, y^k)$ is Lipschitz with constant $L_3(y^k)$, it follows from Lemma 2.11 that

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

By (v) of Assumption 3.1, it follows that

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

Thus, combining the above three inequalities with (3.12), we obtain

$$\mathcal{L}_\beta(x^k, y^k, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) \geq \frac{\beta\mu - L_3(y^k) - \omega}{2}\|x^{k+1} - x^k\|^2. \tag{3.13}$$

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

$$\begin{aligned}
&\|\lambda^{k+1} - \lambda^k\|^2 \\
&= \|\nabla f_2(y^{k+1}) + \nabla_y H(x^{k+1}, y^{k+1}) - \nabla f_2(y^k) - \nabla_y H(x^k, y^k)\|^2 \\
&\le 2\|\nabla f_2(y^{k+1}) - \nabla f_2(y^k)\|^2 + 2\|\nabla_y H(x^{k+1}, y^{k+1}) - \nabla_y H(x^k, y^k)\|^2 \\
&\le 2L_1^2\|y^{k+1} - y^k\|^2 + 2M^2\|x^{k+1} - x^k\|^2 + 2M^2\|y^{k+1} - y^k\|^2 \\
&= (2L_1^2 + 2M^2)\|y^{k+1} - y^k\|^2 + 2M^2\|x^{k+1} - x^k\|^2, \qquad (3.14)
\end{aligned}$$

which, together with (3.6), (3.11) and (3.13) implies

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

where the second inequality follows from (vi) of the Assumption 3.1. The proof is complete. □

If (3.1) holds, then $\delta > 0$, and it follows Lemma 3.2 that $\{\mathcal{L}_\beta(\omega^k)\}_{k\in N}$ is monotonicity nonincreasing. The range of β in the assumption is not optimal. In fact, the range of β is nothing but $\beta > 2L$ proposed in [19] when $H \equiv 0$. We show this in the following.

First, we recalculate (3.14). By (2.4), for any $\xi > 0$ we have

$$\begin{aligned}
&\|\lambda^{k+1} - \lambda^k\|^2 \\
&= \|\nabla f_2(y^{k+1}) + \nabla_y H(x^{k+1}, y^{k+1}) - \nabla f_2(y^k) - \nabla_y H(x^k, y^k)\|^2 \\
&\le (1+\xi)\|\nabla f_2(y^{k+1}) - \nabla f_2(y^k)\|^2 + (1 + \frac{1}{\xi})\|\nabla_y H(x^{k+1}, y^{k+1}) - \nabla_y H(x^k, y^k)\|^2 \\
&\le 2L_1^2\|y^{k+1} - y^k\|^2 + 2M^2\|x^{k+1} - x^k\|^2 + 2M^2\|y^{k+1} - y^k\|^2 \\
&= ((1+\xi)L_1^2 + (1 + \frac{1}{\xi})M^2)\|y^{k+1} - y^k\|^2 + (1 + \frac{1}{\xi})M^2\|x^{k+1} - x^k\|^2. \qquad (3.15)
\end{aligned}$$

Substituting (3.11), (3.13) and (3.15) into (3.6), we obtain

$$\begin{aligned}
\mathcal{L}_\beta(w^{k+1}) \quad &\le \quad \mathcal{L}_\beta(w^k) + (\frac{L_3(y^{k+1}) + \omega}{2} - \frac{\beta\mu}{2} + \frac{M^2}{\beta}(1 + \frac{1}{\xi}))\|x^{k+1} - x^k\|^2 \\
&\qquad + (\frac{(1+\xi)L_1^2}{\beta} + \frac{(1+\frac{1}{\xi})M^2}{\beta} - \frac{\beta - L_1 - L_2(x^{k+1})}{2})\|y^{k+1} - y^k\|^2.
\end{aligned}$$

By simple calculation, when

$$\beta > \max\{\beta_1, \beta_2\}, \qquad (3.16)$$

where

$$\beta_1 := \frac{(L_1 + L_2) + \sqrt{(L_1 + L_2)^2 + 8[(1+\xi)L_1^2 + (1 + \frac{1}{\xi})M^2]}}{2}$$

and

$$\beta_2 := \frac{L_3 + \omega + \sqrt{(L_3 + \omega)^2 + 8\mu M^2(1 + \frac{1}{\xi})}}{2\mu}.$$

Then, we have

$$\mathcal{L}_\beta(w^{k+1}) \leq \mathcal{L}_\beta(w^k) - \bar{\delta}\|v^{k+1} - v^k\|^2,$$

where $\bar{\delta} := \min\{\frac{L_3+\omega}{2} - \frac{\beta\mu}{2} + \frac{M^2}{\beta}(1+\frac{1}{\xi}), \frac{(1+\xi)L_1^2}{\beta} + \frac{(1+\frac{1}{\xi})M^2}{\beta} - \frac{\beta - L_1 - L_2}{2}\} > 0$.

If $H \equiv 0$, then we have $L_2 = L_3 = M = 0$. Moreover, we have $\xi = 0$. Thus (3.16) reduces to

$$\beta > \max\{2L_1, \frac{\omega}{\mu}\}. \tag{3.17}$$

Furthermore, if f_1 is a convex function, then $\omega = 0$ and (3.17) is the same condition as that proposed in [19] to ensure (3.5) holds. Notice that, in [19], they only assume f_1 is a proper semicontinuous function. The reason is that, to prove (3.5) holds, it does not need to estimate the difference between $\mathcal{L}_\beta(x^k, y^k, \lambda^k)$ and $\mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k)$ in [19]. Consequently, from this point of view, the condition in [19] is a special case of our condition.

Lemma 3.3. *Let $\{w^k\}_{k\in N}$ be the sequence generated by the ADMM (1.3) which is assumed to be bounded. Then the following holds*

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

Proof. Since $\{w^k\}_{k\in N}$ is bounded, then there exists a subsequence $\{w^{k_j}\}_{j\in N}$ such that $w^{k_j} \to w^*$. First, we know $\mathcal{L}_\beta(\cdot)$ is lower semicontinuous due to the continuity of f_2 and H and the closedness of f_1. That is

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

Consequently, $\{\mathcal{L}_\beta(w^{k_j})\}_{j\in N}$ is bounded from below. Note that, Lemma 3.2 implies that $\{\mathcal{L}_\beta(\omega^k)\}_{k\in N}$ is nonincreasing and thus $\{\mathcal{L}_\beta(w^{k_j})\}_{j\in N}$ is convergent. Moreover, we have $\{\mathcal{L}_\beta(w^k)\}_{k\in N}$ is convergent and $\mathcal{L}_\beta(w^k) \geq \mathcal{L}_\beta(w^*)$. Rearranging terms of (3.5) leads to

$$\delta\|v^{k+1} - v^k\|^2 \leq \mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^{k+1}),$$

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

$$\sum_{k=0}^{n} \delta\|v^{k+1} - v^k\|^2 \leq \mathcal{L}_\beta(w^0) - \mathcal{L}_\beta(w^{n+1}) \leq \mathcal{L}_\beta(w^0) - \mathcal{L}_\beta(w^*) < +\infty,$$

Since $\delta > 0$, we have $\sum_{k=0}^{+\infty} \|v^{k+1} - v^k\|^2 < +\infty$. Thus,

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty, \quad \sum_{k=0}^{+\infty} \|y^{k+1} - y^k\|^2 < +\infty. \tag{3.18}$$

Moreover, it follows from (3.14) and (3.18) that $\sum_{k=0}^{+\infty} \|\lambda^{k+1} - \lambda^k\|^2 < +\infty$. Therefore, we obtain $\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\|^2 < +\infty$. □

Lemma 3.4. *Let $\{w^k\}_{k\in N}$ be the sequence generated by the ADMM (1.3) which is assumed to be bounded. Then, there exists $\zeta > 0$ such that*

$$d(0, \partial\mathcal{L}_\beta(w^{k+1})) \leq \zeta\|v^{k+1} - v^k\|.$$

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

$$\begin{cases} \partial_x \mathcal{L}_\beta(w^{k+1}) = \partial f_1(x^{k+1}) + \nabla_x H(x^{k+1}, y^{k+1}) - A^T\lambda^{k+1} + \beta A^T(Ax^{k+1} + y^{k+1} - b), \\ \partial_y \mathcal{L}_\beta(w^{k+1}) = \nabla f_2(y^{k+1}) + \nabla_y H(x^{k+1}, y^{k+1}) - \lambda^{k+1} + \beta(Ax^{k+1} + y^{k+1} - b), \\ \partial_\lambda \mathcal{L}_\beta(w^{k+1}) = -(Ax^{k+1} + y^{k+1} - b). \end{cases} \tag{3.19}$$

Substituting (3.4) into (3.19) produces

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

Set $(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}) := (A^T(\lambda^k - \lambda^{k+1}) + \beta A^T(y^{k+1} - y^k) + \nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^{k+1}, y^k), \lambda^k - \lambda^{k+1}, \frac{1}{\beta}(\lambda^{k+1} - \lambda^k))$, then it follows from Lemma 2.3 that $(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}) \in \partial \mathcal{L}_\beta(w^{k+1})$. Furthermore, there exist $\zeta_1, \zeta_2, \zeta_3 > 0$ such that

$$\begin{aligned} &\|(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1})\| \\ &\leq \zeta_1\|y^{k+1} - y^k\| + \zeta_2\|\lambda^{k+1} - \lambda^k\| + \zeta_3\|\nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^{k+1}, y^k)\|. \end{aligned} \tag{3.20}$$

Again since ∇H is Lipschitz continuous on bounded subsets and $\{(x^k, y^k)\}$ is bounded, it follows

$$\|\nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^{k+1}, y^k)\| \leq M\|y^{k+1} - y^k\|. \tag{3.21}$$

From (3.14) we know

$$\|\lambda^{k+1} - \lambda^k\| \leq \sqrt{2}M\|x^{k+1} - x^k\| + \sqrt{2L_1^2 + 2M^2}\|y^{k+1} - y^k\|. \tag{3.22}$$

By setting $\zeta := \sqrt{(\zeta_1 + \sqrt{2L_1^2 + 2M^2} \cdot \zeta_2 + M\zeta_3)^2 + (\sqrt{2}\zeta_3 M)^2}$, it follows from (3.20), (3.21) and (3.22) that

$$\begin{aligned} &d(0, \partial \mathcal{L}_\beta(w^{k+1})) \\ &\leq \|(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1})\| \\ &\leq (\zeta_1 + \sqrt{2L_1^2 + 2M^2}\zeta_2 + M\zeta_3)\|y^{k+1} - y^k\| + \sqrt{2}\zeta_3 M\|x^{k+1} - x^k\| \\ &\leq \zeta\|v^{k+1} - v^k\|, \end{aligned}$$

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

Let $\{w^k\}_{k\in N}$ be a sequence generated by the ADMM (1.3) from a starting point w^0. The set of all limit points is denoted by $S(w^0)$, i.e.,

$$S(w^0) := \{w^* : \exists \text{ a subsequence } \{w^{k_j}\}_{j\in N} \text{ of } \{w^k\}_{k\in N} \text{ converges to } w^*\}.$$

In the following, we summarize several properties of the limit point set.

Lemma 3.5. *Let $\{w^k\}_{k\in N}$ be the sequence generated by the 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*

$$d(w^k, S(w^0)) \to 0, as \quad k \to +\infty;$$

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

(iii) $\mathcal{L}_\beta(\cdot)$ *is finite and constant on* $S(w^0)$, *equal to* $\inf_{k\in N} \mathcal{L}_\beta(w^k) = \lim_{k\to+\infty} \mathcal{L}_\beta(w^k)$.

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 $(x^*, y^*, \lambda^*) \in S(w^0)$, there exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}_{j\in N}$ of $\{(x^k, y^k, \lambda^k)\}_{k\in N}$ converges to (x^*, y^*, λ^*). Since x^{k+1} is the global minimizer of $\mathcal{L}_\beta(x, y^k, \lambda^k)$ for the variable x, it holds

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

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

$$\limsup_{j\to+\infty} \mathcal{L}_\beta(x^{k_j+1}, y^{k_j}, \lambda^{k_j}) = \limsup_{j\to+\infty} \mathcal{L}_\beta(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}) \le \mathcal{L}_\beta(x^*, y^*, \lambda^*). \tag{3.23}$$

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

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

Then by combining (3.23) and (3.24) together we can get

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

which implies

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

Passing to the limit in (3.4) 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 $\nabla f_2(\cdot), \nabla_x H(\cdot,\cdot), \nabla_y H(\cdot,\cdot)$, it follows that

$$\begin{cases} A^T\lambda^* - \nabla_x H(x^*, y^*) \in \partial f_1(x^*), \\ \lambda^* - \nabla_y H(x^*, y^*) = \nabla f_2(y^*), \\ Ax^* + y^* - b = 0. \end{cases}$$

Then, (x^*, y^*, λ^*) is a critical point of (1.3), i.e., $(x^*, y^*, \lambda^*) \in \text{crit } \mathcal{L}_\beta$.
(iii). For any point $(x^*, y^*, \lambda^*) \in S(w^0)$, there exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}_{j\in N}$ converges to (x^*, y^*, λ^*). Since $\{\mathcal{L}_\beta(w^k)\}_{k\in N}$ is nonincreasing, combining (3.23) and (3.24) lead to

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

Therefore, $\mathcal{L}_\beta(\cdot)$ is constant on $S(w^0)$. Moreover, $\inf_{k\in N} \mathcal{L}_\beta(w^k) = \lim_{k\to+\infty} \mathcal{L}_\beta(w^k)$. □

We are now ready for proving the main result of this paper.

Theorem 3.6. *Let* $\{w^k\}_{k\in N}$ *be the sequence generated by the ADMM (1.3) which is assumed to be bounded. Suppose that* $\mathcal{L}_\beta(\cdot)$ *is a KL function, then* $\{w^k\}_{k\in N}$ *has finite length, that is*

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

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

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

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

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

implying that $v^{k+1} = v^k$ for any $k > k_0$. Associated with (3.14), for any $k > k_0 + 1$, it follows that $w^{k+1} = w^k$ and the assertion holds.

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

$$\delta\|v^{k+1} - v^k\|^2 \leq \zeta\|v^k - v^{k-1}\|\Delta_{k,k+1}, \tag{3.27}$$

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

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

Since $S(w^0)$ is a nonempty compact set and $\mathcal{L}_\beta(\cdot)$ is constant on $S(w^0)$, applying Lemma 2.7 with $\Omega := S(w^0)$, we deduce that for any $k > \tilde{k}$

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

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

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

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

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

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

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

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

$$2\|v^{k+1} - v^k\| \leq \|v^k - v^{k-1}\| + \frac{\zeta}{\delta}\Delta_{k,k+1}. \tag{3.28}$$

Summing the inequality (3.28) over $k = \tilde{k} + 1, \dots, m$ yields

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

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

$$\sum_{k=\tilde{k}+1}^{+\infty}\|v^{k+1}-v^k\|\le\|v^{\tilde{k}+1}-v^{\tilde{k}}\|+\frac{\zeta}{\delta}\varphi(\mathcal{L}_\beta(w^{\tilde{k}+1})-\mathcal{L}_\beta(w^*)),\tag{3.29}$$

which implies

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

Thus, it follows that

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

These, together with (3.22), we obtain

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

Moreover, note that

$$\begin{aligned}&\|w^{k+1}-w^k\|\\&=(\|x^{k+1}-\lambda^k\|^2+\|y^{k+1}-y^k\|^2+\|\lambda^{k+1}-\lambda^k\|^2)^{\frac{1}{2}}\\&\le\|x^{k+1}-\lambda^k\|+\|y^{k+1}-y^k\|+\|\lambda^{k+1}-\lambda^k\|.\end{aligned}\tag{3.30}$$

Therefore,

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

and $\{w^k\}_{k\in N}$ is a Cauchy sequence (see P.482 of [5] for a simple proof), which converges. The assertion then follows immediately from Lemma 3.5. □

We now establish the convergence rate for the ADMM (1.3). The proof of the following result is similar to that in [19] and hence omitted here.

Theorem 3.7. *Let $\{w^k\}_{k\in N}$ be the sequence generated by the ADMM (1.3) and converges to $w^*:=(x^*,y^*,\lambda^*)$. 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 $\{w^k\}_{k\in N}$ 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*

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

4 Extensions

The simple structure of ADMM allows us to extend it to the more general setting involving $p > 2$ blocks for which Theorem 3.1 holds. Thus, we can easily extend our previous analysis to the more general model:

$$\begin{aligned} \min_{x_i} \quad & \sum_{i=1}^{p} f_i(x_i) + H(x_1, \cdots, x_p) \\ s.t. \quad & A_1x_1 + A_2x_2 + \cdots + A_{p-1}x_{p-1} + x_p = b, \end{aligned} \tag{4.1}$$

where $f_1 : \mathcal{R}^{n_1} \to \mathcal{R} \cup \{+\infty\}$, $f_i : \mathcal{R}^{n_i} \to \mathcal{R}, i = 2, \cdots, p-1$, $f_p : \mathcal{R}^m \to \mathcal{R}$ and $H : \mathcal{R}^{n_1} \times \cdots \mathcal{R}^{n_{p-1}} \times \mathcal{R}^m \to \mathcal{R}$ are functions, $A_i \in \mathcal{R}^{m \times n_i}, i = 1, \cdots, p-1$ is a given matrix, $b \in \mathcal{R}^m$ is a vector. Using the Gauss-Seidel idea in solving the associated augmented Lagrangian function, it is natural to directly extend the scheme (1.3) to the problem (4.1) with $p > 2$, and the resulting iterative recursion is:

$$\begin{cases} x_1^{k+1} \in \text{argmin}_{x_1}\{\mathcal{L}_\beta^p(x_1, x_2^k, \cdots, x_m^k, \lambda^k)\}, \\ \cdots \\ x_i^{k+1} \in \text{argmin}_{x_i}\{\mathcal{L}_\beta^p(x_1^{k+1}, \cdots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \cdots, x_p^k, \lambda^k)\}, \\ \cdots \\ x_p^{k+1} \in \text{argmin}_{x_p}\{\mathcal{L}_\beta^p(x_1^{k+1}, \cdots, x_{p-1}^{k+1}, x_p, \lambda^k)\}, \\ \lambda^{k+1} = \lambda^k - \beta(\sum\limits_{i=1}^{p-1} A_i x_i^{k+1} + x_p^{k+1} - b), \end{cases} \tag{4.2}$$

where

$$\mathcal{L}_\beta^p(x_1, \cdots, x_p, \lambda) := \sum_{i=1}^{p} f_i(x_i) + H(x_1, \cdots, x_p) - \langle \lambda, \sum_{i=1}^{p-1} A_i x_i + x_p - b \rangle + \frac{\beta}{2} \| \sum_{i=1}^{p-1} A_i x_i + x_p - b \|^2$$

is the augmented Lagrangian function of problem (4.1), λ is the Lagrangian multiplier associated with the linear constraint, and $\beta > 0$ is the penalty parameter.

Next, we use the notation $\{\hat{w}^k := (x_1^k, \cdots, x_p^k, \lambda^k)\}_{k\in N}$ and $\{\hat{v}^k := (x_1^k, \cdots, x_p^k)\}_{k\in N}$ for succinctness.

Theorem 4.1. *Let $f_1 : \mathcal{R}^{n_1} \to \mathcal{R} \cup \{+\infty\}$ be a semiconvex function with constant $\omega > 0$, $f_i : \mathcal{R}^{n_i} \to \mathcal{R}, i = 2, \cdots, p-1$, $f_p : \mathcal{R}^m \to \mathcal{R}$ and $H : \mathcal{R}^{n_1} \times \cdots \mathcal{R}^{n_{p-1}} \times \mathcal{R}^m \to \mathcal{R}$ be smooth functions with $\inf H(x_1, \cdots, x_p) > -\infty$, $\inf\limits_{x_i \in \mathcal{R}^{n_i}} f_i(x_i) > -\infty$ for any $i = 1, \cdots, p-1$, $\inf\limits_{x_p \in \mathcal{R}^m} f_p(x_p) > -\infty$. Let $\{\hat{w}^k\}_{k\in N}$ be the sequence generated by algorithm (4.2) which is assumed to be bounded, then there exists $\delta_1 > 0$ such that*

$$\mathcal{L}_\beta^p(\hat{w}^{k+1}) \leq \mathcal{L}_\beta^p(\hat{w}^k) - \delta_1 \|\hat{v}^{k+1} - \hat{v}^k\|^2.$$

Moreover, suppose that $\mathcal{L}_\beta^p(\cdot)$ is a KL function, then $\{\hat{w}^k\}_{k\in N}$ has finite length, that is

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

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

Proof. Since the proof is similar to Lemma 3.2 and Theorem 3.6, we omit the details. □

Remark 4.2. If $H \equiv 0$, then problem (4.1) and algorithm (4.2) reduce to those considered in [18].

Remark 4.3. Recently, Li, Sun and Toh [25] present a majorized ADMM with indefinite proximal terms for solving linearly constrained two-block convex composite optimization problems. From a numerical point of view, it is advantageous to pick an indefinite S or T whenever possible. Motivated by this, we can also solve both subproblems by adding indefinite terms, leading to the following scheme: Let S and T be given symmetric, possibly indefinite matrix. The ADMM with indefinite proximal terms for problem (1.1) reads as:

$$\left\{\begin{array}{l} x^{k+1} \in \operatorname{argmin}_x\{\mathcal{L}_\beta(x, y^k, \lambda^k) + \frac{1}{2}\|x - x^k\|_S^2\}, \\ y^{k+1} \in \operatorname{argmin}_y\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_T^2\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b). \end{array}\right. \tag{4.3}$$

Furthermore, we can also extend the algorithm (4.2) with indefinite proximal terms to solve (4.1). We omit these details for succinctness.

5 Conclusions

In this paper, we analyzed the convergence of the alternating direction method of multipliers (ADMM) for solving linearly constrained nonconvex minimization model whose objective contains coupled functions. Under the assumption that the associated functions satisfy the Kurdyka-Łojasiewicz inequality, we proved that the iterative sequence generated by the ADMM 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. Some extensions are also presented.

Acknowledgement

The authors would like to thank Dr. Caihua Chen at Nanjing University for his suggestions on this topic.

References

[1] M. Alizadeh, X. Li, Z. Wang and A. Scaglione, Melton, R, Demand side management in the smart grid: Information processing for the power switch, *IEEE Signal Process. Mag.* 59 (2012) 55–67.

[2] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality, *Math. Oper. Res.* 35 (2010) 438–457.

[3] H. Attouch, J. Bolte and B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program.* 137 (2013) 91–129.

[4] J. Bolte, A. Daniilidis, O. Ley and L. Mazet, Characterization of Łojasiewicz inequalities: subgradient flows, talweg, convexity, *Trans. Amer. Math. Soc.* 362 (2010) 3319–3363.

[5] J. Bolte, S. Sabach and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problem, *Math. Program.* 146 (2014) 459–494.

[6] P. Cannarsa and C. Sinestrari, *Semiconcave functions, Hamilton-Jacobi equations, and optimal control,* Springer Science & Business Media, 2004.

[7] T.H. Chang, M. Alizadeh and A. Scaglione, Coordinated home energy management for real-time power balancing, in *Proc. IEEE PES Gen. Meet*, San Diego, USA, 2012, pp. 22–26.

[8] C.H. Chen, M. Li, X. Liu, Y.Y. Ye, Extended ADMM and BCD for nonseparable convex minimization models with quadratic coupling terms: convergence analysis and insights, *arXiv:1508.00193v3.*

[9] M. Degiovanni, A. Marino and M. Tosques, Evolution equations with lack of convexity, *Nonlin. Anal.* 9 (1985) 1401–1443.

[10] J. Eckstein, Some saddle-function splitting methods for convex programming, *Optim. Methods Softw.* 4 (1994) 75–83.

[11] J. Eckstein and D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.* 55 (1992) 293–318.

[12] M. Fortin and R. Glowinski, On decomposition-coordination methods using an augmented Lagrangian, in *Augmented Lagrangian Methods: Applications to the solution of boundary problem*, M. Fortin and R. Glowinski (ed.), North-Holland, Amsterdam 1983.

[13] P. Frankel, G. Garrigos and J. Peypouquet, Splitting methods with variable metric for Kurdyka- Łojasiewicz functions and general convergence rates, *J. Optim. Theory Appl.* 165 (2015) 874–900.

[14] D. Gabay, Applications of the method of multipliers to variational inequalities, in *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski (eds.), North-Holland, Amsterdam, 1983, pp. 299–331.

[15] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, *Comput. Math. Appl.* 2 (1976) 17–40.

[16] X. Gao and S.Z. Zhang, First-order algorithms for convex optimization with nonseparable objective and coupled constraints, *J. Oper. Res. Soc. China.* DOI 10.1007/s40305-016-0131-5.

[17] 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.* 2 (1975) 41–76.

[18] K. Guo, D.R. Han, D.Z.W.Wang and T.T. Wu, Convergence of ADMM for multi-block nonconvex separable optimization models, *Front. Math. China* 12 (2017) 1139–1162.

[19] K. Guo, D.R. Han and T.T. Wu, Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints, *Int. J. Comput. Math.* 94 (2017) 1653–1669.

[20] D.R. Han and X.M. Yuan, Local linear convergence of alternating direction method of multipliers for quadratic programs, *SIAM J. Numer. Anal.* 51 (2013) 3446–3457.

[21] M. Hong, T. Chang, X. Wang, M. Razaviyayn, S. Ma and Z. Luo, A block coordinate descent method of multipliers: convergence analysis and applications, *IEEE Int. Conf. Acoust. Speech Signal Process.* (2014) 7739–7743.

[22] M. Hong, Z. Luo and M. Razaviyayn, Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems, *SIAM J. Optim.* 26 (2016) 337–364.

[23] G.Y. Li and T.K. Pong, Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems, *Math. Program.* 159 (2016) 371–401.

[24] G.Y. Li and T.K. Pong, Global convergence of splitting methods for nonconvex composite optimization, *SIAM J. Optim.* 25 (2015) 2434–2460.

[25] M. Li, D.F. Sun and K.C. Toh, A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization, *SIAM J. Optim.* 26 (2016) 922–950.

[26] N. Li, L. Chen and S.H. Low, Optimal demand response based on utility maximization in power networks, in *Proc. IEEE PES Gen. Meet*, Detroit, MI, 2011, pp. 1–8.

[27] S. Marcellin and L. Thibault, Evolution problems associated with primal lower nice functions, *J. Convex Anal.* 13 (2006) 385–421.

[28] B. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory,* Grundlehren der Mathematischen Wissenschaften, Springer, Berlin, 2006.

[29] Y. Nesterov, *Introduction Lectures on Convex Optimization : A Basic Course,* Springer, 2004.

[30] P. Ochs, Y. Chen, T. Brox and T. Pock, iPiano: Inertial proximal algorithm for non-convex optimization, *SIAM J. Imaging Sci.* 7 (2014) 1388–1419.

[31] J.V. Paatero and P.D. Lund, A model for generating household electricity load profiles, *Int. J. Energy Res.* 30 (2006) 273–290.

[32] R.T. Rockafellar and R.J.-B Wets, *Variational Analysis,* Springer, Berlin, 1998.

[33] G. Scutari and D.P. Palomar, MIMO cognitive radio: A game theoretical approach, *IEEE Trans. Signal Process.* 58 (2010) 761–780.

[34] W.H. Yang and D.R. Han, Linear convergence of alternating direction method of multipliers for a class of convex optimization problems, *SIAM J. Numer. Anal.* 54 (2016) 625–640.

[35] L. Zeng and J. Xie, Group variable selection via SCAD-ℓ_2, *Statistics.* 48 (2014) 49–66.

[36] Q. Zhao and B.M. Sadler, A survey of dynamic spectrum access, *IEEE Signal Process. Mag.* 5 (2007) 79–89.

Manuscript received 10 January 2016
revised 17 September 2016
accepted for publication 14 March 2017

Ke Guo
School of Mathematics and information
China West Normal University
Nanchong 637002, China.
E-mail address: keguo2014@126.com

Deren Han
School of Mathematics and System Science
Beihang University, Beijing 100191, China
E-mail address: handr@buaa.edu.cn

Tingting Wu
School of Science, Nanjing University of Posts and
Telecommunications, Nanjing 210023, China
E-mail address: wutt@njupt.edu.cn