Front. Math. China 2017, 12(5): 1139–1162
https://doi.org/10.1007/s11464-017-0631-6

RESEARCH ARTICLE

Convergence of ADMM for multi-block nonconvex separable optimization models

Ke GUO[1], **Deren HAN**[1], **David Z. W. WANG**[2], **Tingting WU**[3]

1 School of Mathematical Sciences and Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, China
2 School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore
3 School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract For solving minimization problems whose objective function is the sum of two functions without coupled variables and the constrained function is linear, the alternating direction method of multipliers (ADMM) has exhibited its efficiency and its convergence is well understood. When either the involved number of separable functions is more than two, or there is a nonconvex function, ADMM or its direct extended version may not converge. In this paper, we consider the multi-block separable optimization problems with linear constraints and absence of convexity of the involved component functions. Under the assumption that the associated function satisfies the Kurdyka-Lojasiewicz inequality, we prove that any cluster point of the iterative sequence generated by ADMM is a critical point, under the mild condition that the penalty parameter is sufficiently large. We also present some sufficient conditions guaranteeing the sublinear and linear rate of convergence of the algorithm.

Keywords Nonconvex optimization, separable structure, alternating direction method of multipliers (ADMM), Kurdyka-Lojasiewicz inequality
MSC 90C26, 65K10, 49J52, 49M27

1 Introduction

In this paper, we consider the following nonconvex optimization problem:

Received July 5, 2016; accepted January 16, 2017
Corresponding author: Deren HAN, E-mail: handeren@njnu.edu.cn

$$
\begin{aligned}
&\min \ \sum_{i=1}^{m} f_i(x_i) \\
&\text{s.t.} \quad A_1x_1 + A_2x_2 + \cdots + A_{m-1}x_{m-1} + x_m = b,
\end{aligned} \tag{1}
$$

where

$$f_1 \colon \mathbb{R}^{n_1} \to \mathbb{R} \cup \{+\infty\}$$

is a proper lower semicontinuous function,

$$f_i \colon \mathbb{R}^{n_i} \to \mathbb{R},\ i = 2, 3, \ldots, m-1, \quad f_m \colon \mathbb{R}^s \to \mathbb{R},$$

are continuous differentiable functions with ∇f_i being Lipschitz continuous with modulus $L_i > 0$, $A_i \in \mathbb{R}^{s\times n_i}$, $i = 1, 2, \ldots, m-1$, is a given matrix, and $b \in \mathbb{R}^s$ is a vector.

The direct extension of the classic alternating direction method of multipliers (ADMM) (initiated from [11–13]) for solving problem (1) reads as

$$
\begin{cases}
x_1^{k+1} \in \arg\min\limits_{x_1} \{\mathscr{L}_\beta(x_1, x_2^k, \ldots, x_m^k, \lambda^k)\}, \\
x_2^{k+1} \in \arg\min\limits_{x_2} \{\mathscr{L}_\beta(x_1^{k+1}, x_2, x_3^k, \ldots, x_m^k, \lambda^k)\}, \\
\cdots, \\
x_m^{k+1} \in \arg\min\limits_{x_m} \{\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, \ldots, x_{m-1}^{k+1}, x_m, \lambda^k)\}, \\
\lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + \cdots + A_{m-1}x_{m-1}^{k+1} + x_m^{k+1} - b),
\end{cases} \tag{2}
$$

which can be viewed as a Gauss-Seidel implementation of the well-known augmented Lagrangian algorithm for linear constraint optimization problems. Here and throughout the paper, $\mathscr{L}_\beta(\cdot)$ denotes the augmented Lagrangian function for (1):

$$
\begin{aligned}
\mathscr{L}_\beta(x_1, x_2, \ldots, x_m, \lambda) := \sum_{i=1}^{m} f_i(x_i) - \Big\langle \lambda, \sum_{i=1}^{m-1} A_ix_i + x_m - b \Big\rangle \\
+ \frac{\beta}{2} \Big\| \sum_{i=1}^{m-1} A_ix_i + x_m - b \Big\|^2,
\end{aligned} \tag{3}
$$

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

When $m = 2$ and the involved component functions f_1 and f_2 are both convex and some very mild conditions are satisfied, ADMM is proved to converge to a solution of (1) globally. Under some further conditions (for special problems where strong convexity or some error bound conditions hold), ADMM can achieve linear convergence [4,16,20,34]. For the case either there are three or more separable blocks in model (1), or there are nonconvex component functions (even for the two-block case), ADMM may not converge [9]. On the other hand, there are many applications that can naturally be modeled or reformulated as a multi-block linearly constrained minimization model whose objective function is the sum of more than two functions without coupled variables, such as phase retrieval [32], nonconvex background/foreground extraction problem [33]. In fact, heuristic applications of ADMM in solving these problems result in very well numerical results. Such a gap between the high efficiency of ADMM in numerical experiments and lack of convergence result attracts the researchers' more and more attentions on it, and there have been a few developments. Here, we summarize the progress from two aspects.

(i) The multi-block case. Han and Yuan [15] first theoretically considered this problem and they proved that when all the objective functions are strongly convex, the direct extension ADMM scheme is globally (linear) convergent, provided that the penalty parameter is smaller than a threshold. Then, this condition was relaxed and only one or more functions in the objective are required to be strongly convex to ensure the convergence [8,25]. On the other hand, some researchers suggested twisting the ADMM scheme slightly. For examples, in [18,19], it was suggested to correct the output of ADMM scheme to generate a new iterate and the resulting prediction-correction schemes are guaranteed to be convergent. Numerically, the original ADMM scheme usually performs better than all the twisted variants with provable convergence (see, e.g., [17]); and it is the most convenient scheme to be implemented compared with its variants. Hong and Luo [21] suggested attaching a shrinkage factor to the Lagrange multiplier updating step and it was shown that the convergence of ADMM is guaranteed when this factor is small enough to satisfy some error bound conditions.

(ii) The case that there is at least one nonconvex component function. For two block nonconvex separable optimization problem, under the assumption that the associated function satisfies the Kurdyka-Lojasiewicz (KL) inequality, Guo et al. [14] proved that any cluster point of the iterative sequence generated by the alternating direction method is a critical point provided that the penalty parameter is greater than $2L$, where L is the Lipschitz constant of the gradient of one of the involving function. Under some further conditions on the problem's data, they also analyzed the rate of convergence of the algorithm. Li and Pong [24] showed that if the penalty parameter in the augmented Lagrangian function associated to the problem is chosen sufficiently large and the sequence generated by the algorithm has a cluster point, then it gives a critical point of the nonconvex problem when one of the component objective functions is twice continuously differentiable with bounded Hessian, and the other one is a proper closed function. Hong et al. [22] analyzed the convergence of the ADMM for solving certain special nonconvex problems, i.e., the consensus and sharing problems. They proved that the sequence generated by ADMM converges to the set of stationary solutions, provided that the penalty parameter in the augmented Lagrangian function is chosen to be sufficiently large.

The purpose of this paper is to prove the convergence of the classic ADMM for multi-block nonconvex optimization problems (1). Using the important KL inequality (see Definition 3 below), we prove that if the augmented Lagrangian function is a KL function, then the sequence generated by ADMM converges to a critical point of the augmented Lagrangian function. If some further conditions on the problem's data hold, we then prove the sublinear and linear rate of convergence of the algorithm. The importance of KL inequality is due to the fact that many functions emerged in the modern application models satisfy this inequality. Especially, when the function belongs to some functional classes, e.g., semi-algebraic, subanalytic, and log-exp (see [2,3,5,6] and references therein). These facts originate in the pioneering and fundamental work of Łojasiewicz [26] and Kurdyka [23]; work which was recently extended to nonsmooth functions is in [5,6].

The rest of this paper is organized as follows. In Section 2, we present some preliminary materials that will be used in our next analysis. In Section 3, we prove the convergence of scheme (2). Then, we establish the convergence rate for scheme (2) in Section 4. Finally, we draw some conclusions.

2 Preliminaries

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

The following notation and definitions are quite standard and can be found in, e.g., [27,29,30]. Let $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a point-to-set mapping. Then its graph is defined by

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

For any subset $S \subseteq \mathbb{R}^n$ and any point $x \in \mathbb{R}^n$, the distance from x to S, denoted by $d(x,S)$, is defined as

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

When $S = \emptyset$, we set $d(x,S) := +\infty$ for all x. Let $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. We denote

$$v := (x,y) \in \mathbb{R}^n \times \mathbb{R}^m, \quad \|v\|^2 := \|x\|^2 + \|y\|^2.$$

Definition 1 Given a function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the effective domain and the epigraph of f are defined by

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}, \quad \operatorname{epi} f := \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \colon f(x) \leqslant \alpha\},$$

respectively. We say that the function f is proper (resp., lower semicontinuous) if $\operatorname{dom} f$ (resp., $\operatorname{epi} f$) is nonempty (resp., closed).

Definition 2 Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.

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

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

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

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

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

Remark 1 In view of Definition 2, the following conclusions hold.

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

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

(iii) If $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous and $h \colon \mathbb{R}^n \to \mathbb{R}$ is continuous differentiable, then

$$\partial(h + f)(x) = \nabla h(x) + \partial f(x), \quad \forall\, x \in \operatorname{dom} f.$$

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

Definition 3 ([2], Kurdyka-Łojasiewicz inequality) Let $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leqslant +\infty$, set

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

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

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

(ii) φ is C^1 on $(0, \eta)$ and continuous at 0;

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

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

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

Definition 4 ([3], Kurdyka-Łojasiewicz function) Denote Φ_η the set of functions which satisfy (i)–(iii) in Definition 3. If f satisfies the KL property at each point of $\operatorname{dom} \partial f$, then f is called a KL function.

Remark 2 One can easily check that the Kurdyka-Łojasiewicz property is automatically satisfied at any noncritical point $x^* \in \operatorname{dom} f$; see, e.g., [2, Lemma 2.1, Remark 3.2 (b)].

Lemma 1 ([7], Uniformized KL property) *Let Ω be a compact set, and let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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 $\varepsilon, \eta > 0$ and $\varphi \in \Phi_\eta$ such that for all $\overline{x} \in \Omega$ and for all x in the intersection*

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

one has

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

Lemma 2 [28] *Let $h\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous differentiable function with gradient ∇h is Lipschitz continuous with the modulus $L_h > 0$. Then, for any $x, y \in \mathbb{R}^n$, we have*

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

3 Convergence

In this section, we prove the convergence of the ADMM procedure (2). However, in the following, we only consider the case $m = 3$ because in the convergence analysis, the proof for $m > 3$ follows the same roadmap as $m = 3$. When $m = 3$, problem (1) reduces to

$$\begin{aligned} &\min\ \ f_1(x_1) + f_2(x_2) + f_3(x_3) \\ &\text{s.t.}\ \ \ A_1x_1 + A_2x_2 + x_3 = b. \end{aligned} \tag{4}$$

The corresponding algorithm (2) becomes

$$\begin{cases} x_1^{k+1} \in \arg\min\limits_{x_1}\{\mathscr{L}_\beta(x_1, x_2^k, x_3^k, \lambda^k)\}, \\ x_2^{k+1} \in \arg\min\limits_{x_2}\{\mathscr{L}_\beta(x_1^{k+1}, x_2, x_3^k, \lambda^k)\}, \\ x_3^{k+1} \in \arg\min\limits_{x_3}\{\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3, \lambda^k)\}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b), \end{cases} \tag{5}$$

where the augmented Lagrangian function (3) reduces to

$$\begin{aligned} \mathscr{L}_\beta(x_1, x_2, x_3, \lambda) := \sum_{i=1}^{3} f_i(x_i) &- \langle \lambda, A_1x_1 + A_2x_2 + x_3 - b\rangle \\ &+ \frac{\beta}{2}\|A_1x_1 + A_2x_2 + x_3 - b\|^2. \end{aligned} \tag{6}$$

First, we make some assumptions.

Assumption 1 Let $f_1 \colon \mathbb{R}^{n_1} \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function, and let $f_2 \colon \mathbb{R}^{n_2} \to \mathbb{R}$ and $f_3 \colon \mathbb{R}^{s} \to \mathbb{R}$ be continuously differentiable functions with ∇f_2 and ∇f_3 being Lipschitz continuous with modulus $L_2 > 0$ and $L_3 > 0$, respectively. Set $L := \max\{L_2, L_3\}$. Furthermore, assume the following holds:

(i) $\beta > \max\{2L, L/\mu\}$;

(ii) $A_1^{\mathrm{T}} A_1 \succeq \mu I$, $A_2^{\mathrm{T}} A_2 \succeq \mu I$ for some $\mu > 0$.

Let

$$\delta := \min\Big\{\frac{\beta - L}{2} - \frac{L^2}{\beta}, \frac{\beta\mu - L}{2}\Big\}. \tag{7}$$

Then it follows from (i) of Assumption 1 that $\delta > 0$.

Definition 5 We say that $(x_1^*, x_2^*, x_3^*, \lambda^*)$ is a critical point of the augmented Lagrangian function $\mathscr{L}_\beta(\cdot)$ in (6), if it satisfies

$$\begin{cases} A_1^{\mathrm{T}} \lambda^* \in \partial f_1(x^*), \\ \nabla f_2(x_2^*) = A_2^{\mathrm{T}} \lambda^*, \\ \nabla f_3(x_3^*) = \lambda^*, \\ A_1 x^* + A_2 x_2^* + x_3^* - b = 0. \end{cases} \tag{8}$$

The set of critical points of $\mathscr{L}_\beta(\cdot)$ is denoted by $\operatorname{crit} \mathscr{L}_\beta$.

Remark 3 Actually, if (x_1^*, x_2^*, x_3^*) is a local minimizer of problem (4), then (x_1^*, x_2^*) is a local minimizer of the following problem:

$$\min_{x_1, x_2} \ f_1(x_1) + f_2(x_2) + f_3(b - A_1 x_1 - A_2 x_2).$$

By [30, Theorem 8.15], it follows from (iii) of Remark 1 that

$$\begin{cases} 0 \in \partial f_1(x_1^*) - A_1^T \nabla f_3(b - A_1 x_1^* - A_2 x_2^*), \\ 0 = \nabla f_2(x_2^*) - A_2^T \nabla f_3(b - A_1 x_1^* - A_2 x_2^*). \end{cases} \tag{9}$$

Since

$$A_1 x_1^* + A_2 x_2^* + x_3^* = b,$$

setting

$$\lambda^* := \nabla f_3(x_3^*),$$

we know that system (8) holds in view of (9). Hence, system (8) is indeed the first-order necessary condition of (4).

Before the proof, let us present the variational characterization of scheme (5). Invoking the optimality condition for (5), we have

$$\begin{cases} 0 \in \partial f_1(x_1^{k+1}) - A_1^{\mathrm{T}}\lambda^k + \beta A_1^{\mathrm{T}}(A_1x_1^{k+1} + A_2x_2^k + x_3^k - b), \\ 0 = \nabla f_2(x_2^{k+1}) - A_2^{\mathrm{T}}\lambda^k + \beta A_2^{\mathrm{T}}(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b), \\ 0 = \nabla f_3(x_3^{k+1}) - \lambda^k + \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b), \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b). \end{cases} \tag{10}$$

Using the last equality and rearranging terms, we obtain

$$\begin{cases} A_1^{\mathrm{T}}\lambda^{k+1} + \beta A_1^{\mathrm{T}}(A_2x_2^{k+1} - A_2x_2^k) + \beta A_1^{\mathrm{T}}(x_3^{k+1} - x_3^k) \in \partial f_1(x_1^{k+1}), \\ \nabla f_2(x_2^{k+1}) = A_2^{\mathrm{T}}\lambda^{k+1} + \beta A_2^{\mathrm{T}}(x_3^{k+1} - x_3^k), \\ \nabla f_3(x_3^{k+1}) = \lambda^{k+1}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b). \end{cases} \tag{11}$$

In the sequel, for convenience, we often use the notation $\{v^k := (x_2^k, x_3^k)\}$. We begin our analysis with the following lemma.

Lemma 3 *Let $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by algorithm* (5). *Then we have*

$$\mathscr{L}_\beta(w^{k+1}) \leqslant \mathscr{L}_\beta(w^k) - \delta\|v^{k+1} - v^k\|^2. \tag{12}$$

Proof From the definition of the augmented Lagrangian function $\mathscr{L}_\beta(\cdot)$ in (6), it follows that

$$\begin{aligned} &\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) \\ &\quad = \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k) + \langle \lambda^k - \lambda^{k+1}, A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b\rangle \\ &\quad = \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k) + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 \end{aligned} \tag{13}$$

and

$$\begin{aligned} &\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k) - \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k) \\ &\quad = f_3(x_3^k) - f_3(x_3^{k+1}) + \langle \lambda^k, x_3^{k+1} - x_3^k\rangle + \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b\|^2 \\ &\qquad -\frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b\|^2. \end{aligned} \tag{14}$$

Since ∇f_3 is Lipschitz continuous with modulus $L_3 \leqslant L$, it follows from Lemma 2 and the third equality of (11) that

$$f_3(x_3^k) - f_3(x_3^{k+1}) \geqslant \langle \lambda^{k+1}, x_3^k - x_3^{k+1}\rangle - \frac{L}{2}\|x_3^k - x_3^{k+1}\|^2. \tag{15}$$

Inserting (15) into (14) yields

$$\begin{aligned}&\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k) - \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k)\\&\quad \geqslant \langle \lambda^{k+1} - \lambda^k, x_3^k - x_3^{k+1}\rangle + \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b\|^2\\&\qquad - \frac{L}{2}\|x_3^k - x_3^{k+1}\|^2 - \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b\|^2.\end{aligned} \tag{16}$$

From the fourth equation of (11), we know

$$A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b = \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) + (x_3^k - x_3^{k+1}).$$

Thus,

$$\begin{aligned}&\langle \lambda^{k+1} - \lambda^k, x_3^k - x_3^{k+1}\rangle + \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b\|^2\\&\quad = \langle \lambda^{k+1} - \lambda^k, x_3^k - x_3^{k+1}\rangle + \frac{\beta}{2}\left\|\frac{1}{\beta}(\lambda^k - \lambda^{k+1}) + (x_3^k - x_3^{k+1})\right\|^2\\&\quad = \frac{\beta}{2}\|x_3^k - x_3^{k+1}\|^2 + \frac{1}{2\beta}\|\lambda^{k+1} - \lambda^k\|^2.\end{aligned} \tag{17}$$

Substituting (17) into (16), we obtain

$$\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k) - \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k) \geqslant \frac{\beta - L}{2}\|x_3^k - x_3^{k+1}\|^2. \tag{18}$$

On the other hand, since $\nabla f_3(z^{k+1}) = \lambda^{k+1}$ and ∇f_3 is Lipschitz continuous, we get

$$\|\lambda^{k+1} - \lambda^k\| \leqslant L\|x_3^{k+1} - x_3^k\|. \tag{19}$$

Consequently, it follows from (13), (18), and (19) that

$$\begin{aligned}&\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})\\&\quad \leqslant \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k) - \Big(\frac{\beta - L}{2} - \frac{L^2}{\beta}\Big)\|x_3^{k+1} - x_3^k\|^2.\end{aligned} \tag{20}$$

Similarly,

$$\begin{aligned}&\mathscr{L}_\beta(x_1^{k+1}, x_2^k, x_3^k, \lambda^k) - \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k)\\&\quad = f_2(x_2^k) - f_2(x_2^{k+1}) + \langle \lambda^k, A_2x_2^{k+1} - A_2x_2^k\rangle\\&\qquad + \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^k + x_3^k - b\|^2 - \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b\|^2.\end{aligned} \tag{21}$$

Since ∇f_2 is Lipschitz continuous with modulus $L_2 \leqslant L$, it follows from Lemma 2 and the second equality of (10) that

$$\begin{aligned}f_2(x_2^k) - f_2(x_2^{k+1}) &\geqslant \langle A_2^{\mathrm{T}}\lambda^k - \beta A_2^{\mathrm{T}}(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b), x_2^k - x_2^{k+1}\rangle\\&\quad - \frac{L}{2}\|x_2^k - x_2^{k+1}\|^2.\end{aligned} \tag{22}$$

Inserting (22) into (21) and by (ii) of Assumption 1 yield

$$\begin{aligned}
&\mathscr{L}_\beta(x_1^{k+1}, x_2^k, x_3^k, \lambda^k) - \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k) \\
&\quad \geqslant -\beta\langle A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b, A_2x_2^k - A_2x_2^{k+1}\rangle \\
&\qquad + \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^k + x_3^k - b\|^2 \\
&\qquad - \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b\|^2 - \frac{L}{2}\|x_2^k - x_2^{k+1}\|^2 \\
&\quad = \frac{\beta}{2}\|A_2x_2^k - A_2x_2^{k+1}\|^2 - \frac{L}{2}\|x_2^k - x_2^{k+1}\|^2 \\
&\quad \geqslant \frac{\beta\mu - L}{2}\|x_2^{k+1} - x_2^k\|^2. \qquad (23)
\end{aligned}$$

Thus, it follows from (20) and (23) that

$$\begin{aligned}
&\mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) \\
&\leqslant \mathscr{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k) - \Big(\frac{\beta - L}{2} - \frac{L^2}{\beta}\Big)\|x_3^{k+1} - x_3^k\|^2 \\
&\leqslant \mathscr{L}_\beta(x_1^{k+1}, x_2^k, x_3^k, \lambda^k) - \Big(\frac{\beta - L}{2} - \frac{L^2}{\beta}\Big)\|x_3^{k+1} - x_3^k\|^2 - \frac{\beta\mu - L}{2}\|x_2^{k+1} - x_2^k\|^2 \\
&\leqslant \mathscr{L}_\beta(x_1^k, x_2^k, x_3^k, \lambda^k) - \delta\|v^{k+1} - v^k\|^2,
\end{aligned}$$

where the third inequality follows from (7) and the fact that x_1^{k+1} is the global minimizer of $\mathscr{L}_\beta(x_1, x_2^k, x_3^k, \lambda^k)$ with respect to variable x_1, i.e.,

$$\mathscr{L}_\beta(x_1^{k+1}, x_2^k, x_3^k, \lambda^k) \leqslant \mathscr{L}_\beta(x_1^k, x_2^k, x_3^k, \lambda^k).$$

The proof is complete. □

Remark 4 Since $\delta > 0$, in view of Lemma 3, we know that $\mathscr{L}_\beta(\cdot)$ is monotonicity nonincreasing.

Remark 5 In fact, if we assume that f_2 is a convex function instead of a smooth function, then we can also prove Lemma 3 holds. In this situation, $L := L_3$ and we assume

(i) $\beta > 2L$, then the corresponding

$$\delta := \min\Big\{\frac{\beta - L}{2} - \frac{L^2}{\beta}, \frac{\beta\mu}{2}\Big\} > 0;$$

(ii) $A_1^{\mathrm{T}}A_1 \succeq \mu I$, $A_2^{\mathrm{T}}A_2 \succeq \mu I$ for some $\mu > 0$.

Since the proof can go in a similar way as Lemma 3 and for the sake of clarity, we move the corresponding proof to Appendix.

Lemma 4 *Let $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by algorithm* (5) *which is assumed to be bounded. Then*

$$\sum_{k=0}^{+\infty}\|w^{k+1} - w^k\|^2 < +\infty. \qquad (24)$$

Proof Since $\{w^k\}$ is bounded, there exists a subsequence $\{w^{k_j}\}$ such that $w^{k_j} \to w^*$. Due to the continuity of f_2 and f_3 and lower semicontinuity of f_1, $\mathscr{L}_\beta(\cdot)$ is lower semicontinuous, and hence,

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

Consequently, $\mathscr{L}_\beta(w^{k_j})$ is bounded from below, which, together with the fact that $\mathscr{L}_\beta(\cdot)$ is nonincreasing, means that $\mathscr{L}_\beta(w^{k_j})$ is convergent. Moreover, we have $\mathscr{L}_\beta(w^k)$ is convergent and $\mathscr{L}_\beta(w^k) \geqslant \mathscr{L}_\beta(w^*)$. Rearranging terms of (12) yields

$$\delta\|v^{k+1} - v^k\|^2 \leqslant \mathscr{L}_\beta(w^k) - \mathscr{L}_\beta(w^{k+1}).$$

Summing up the above inequality for all $k \geqslant 0$, we get

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

Since $\delta > 0$, we have

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

Thus,

$$\sum_{k=0}^{+\infty} \|x_2^{k+1} - x_2^k\|^2 < +\infty, \quad \sum_{k=0}^{+\infty} \|x_3^{k+1} - x_3^k\|^2 < +\infty.$$

Consequently, it follows from (19) that

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

Recall that

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b), \\ \lambda^k = \lambda^{k-1} - \beta(A_1x_1^k + A_2x_2^k + x_3^k - b), \end{cases}$$

and hence,

$$\lambda^{k+1} - \lambda^k = (\lambda^k - \lambda^{k-1}) + \beta(A_1x_1^k - A_1x_1^{k+1}) + \beta(A_2x_2^k - A_2x_2^{k+1}) + \beta(x_3^k - x_3^{k+1}).$$

Then it follows that

$$\begin{aligned} &\|\beta(A_1x_1^k - A_1x_1^{k+1})\|^2 \\ &\quad = \|(\lambda^{k+1} - \lambda^k) - (\lambda^k - \lambda^{k-1}) - \beta(x_3^k - x_3^{k+1}) - \beta(A_2x_2^k - A_2x_2^{k+1})\|^2 \\ &\quad \leqslant 4(\|\lambda^{k+1} - \lambda^k\|^2 + \|\lambda^k - \lambda^{k-1}\|^2 + \beta^2\|x_3^{k+1} - x_3^k\|^2 \\ &\qquad + \beta^2\|A_2\|^2\, \|x_2^{k+1} - x_2^k\|^2). \end{aligned} \tag{25}$$

Using (ii) of Assumption 1, we have

$$\|\beta(A_1x_1^k - A_1x_1^{k+1})\|^2 \geqslant \beta^2\mu\|x_1^{k+1} - x_1^k\|^2. \tag{26}$$

Substituting (26) into (25) implies

$$\sum_{k=1}^{+\infty} \|x_1^{k+1} - x_1^k\|^2 < +\infty.$$

Therefore, we obtain (24). □

Remark 6 If $\mathscr{L}_\beta(\cdot)$ is bounded from below, it is easy to deduce (24) without using the boundedness of $\{w^k\}$.

Lemma 5 *Let $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by algorithm* (5)*. Then there exists $\zeta > 0$ such that*

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

Proof From the definition of the augmented Lagrangian function $\mathscr{L}_\beta(\cdot)$ in (6), it follows that

$$\begin{cases} \partial_{x_1}\mathscr{L}_\beta(w^{k+1}) = \partial f_1(x_1^{k+1}) - A_1^{\mathrm{T}}\lambda^{k+1} + \beta A_1^{\mathrm{T}}(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b), \\ \partial_{x_2}\mathscr{L}_\beta(w^{k+1}) = \nabla f_2(x_2^{k+1}) - A_2^{\mathrm{T}}\lambda^{k+1} + \beta A_2^{\mathrm{T}}(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b), \\ \partial_{x_3}\mathscr{L}_\beta(w^{k+1}) = \nabla f_3(x_3^{k+1}) - \lambda^{k+1} + \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b), \\ \partial_\lambda\mathscr{L}_\beta(w^{k+1}) = -(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b). \end{cases}$$

This, together with (11), yields

$$\begin{cases} A_1^{\mathrm{T}}(\lambda^k - \lambda^{k+1}) + \beta A_1^{\mathrm{T}}(A_2x_2^{k+1} - A_2x_2^k) + \beta A_1^{\mathrm{T}}(x_3^{k+1} - x_3^k) \in \partial_{x_1}\mathscr{L}_\beta(w^{k+1}), \\ A_2^{\mathrm{T}}(\lambda^k - \lambda^{k+1}) + \beta A_2^{\mathrm{T}}(x_3^{k+1} - x_3^k) \in \partial_{x_2}\mathscr{L}_\beta(w^{k+1}), \\ \lambda^k - \lambda^{k+1} \in \partial_{x_3}\mathscr{L}_\beta(w^{k+1}), \\ \dfrac{1}{\beta}(\lambda^{k+1} - \lambda^k) \in \partial_\lambda\mathscr{L}_\beta(w^{k+1}). \end{cases}$$

Define

$$\begin{aligned} &(\hat{x}_1^{k+1}, \hat{x}_2^{k+1}, \hat{x}_3^{k+1}, \hat{\lambda}^{k+1}) \\ &\quad := \Big(A_1^{\mathrm{T}}(\lambda^k - \lambda^{k+1}) + \beta A_1^{\mathrm{T}}(A_2x_2^{k+1} - A_2x_2^k) + \beta A_1^{\mathrm{T}}(x_3^{k+1} - x_3^k), \\ &\qquad\quad A_2^{\mathrm{T}}(\lambda^k - \lambda^{k+1}) + \beta A_2^{\mathrm{T}}(x_3^{k+1} - x_3^k), \lambda^k - \lambda^{k+1}, \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\Big). \end{aligned}$$

Then we have

$$(\hat{x}_1^{k+1}, \hat{x}_2^{k+1}, \hat{x}_3^{k+1}, \hat{\lambda}^{k+1}) \in \partial\mathscr{L}_\beta(w^{k+1}).$$

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

$$\|(\hat{x}_1^{k+1}, \hat{x}_2^{k+1}, \hat{x}_3^{k+1}, \hat{\lambda}^{k+1})\| \leqslant \zeta_1\|x_2^{k+1} - x_2^k\| + \zeta_2\|x_3^{k+1} - x_3^k\| + \zeta_3\|\lambda^{k+1} - \lambda^k\|.$$

By setting $\zeta_4 := \zeta_2 + L\zeta_3$, it follows from (19) that

$$\begin{aligned} d(0, \partial\mathscr{L}_\beta(w^{k+1})) &\leqslant \|(\hat{x}_1^{k+1}, \hat{x}_2^{k+1}, \hat{x}_3^{k+1}, \hat{\lambda}^{k+1})\| \\ &\leqslant \zeta_1\|x_2^{k+1} - x_2^k\| + \zeta_4\|x_3^{k+1} - x_3^k\| \\ &\leqslant \sqrt{\zeta_1^2 + \zeta_4^2} \cdot \|v^{k+1} - v^k\|, \end{aligned}$$

where the third inequality follows from the Cauchy inequality. By setting

$$\zeta := \sqrt{\zeta_1^2 + \zeta_4^2},$$

we complete the proof. □

In the following result, we summarize several properties of the limit point set. Let $\{w^k\}$ be a sequence generated by the ADMM procedure (5) from a starting point w^0. The set of all limit points is denoted by $S(w^0)$, i.e.,

$$S(w^0) := \{w^*\colon \exists \text{ subsequence } \{w^{k_j}\} \text{ of } \{w^k\} \text{ converges to } w^*\}.$$

Lemma 6 *Let $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by algorithm* (5) *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*

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

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

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

$$\inf_{k\in\mathbb{N}} \mathscr{L}_\beta(w^k) = \lim_{k\to+\infty} \mathscr{L}_\beta(w^k). \tag{27}$$

Proof We prove the results item by item.

(i) Obviously, $S(w^0)$ is a nonempty bounded set and $d(w^k, S(w^0)) \to 0$ as $k \to +\infty$. Thus, we only need to show that $S(w^0)$ is a closed set. To see this, let $p^n \in S(w^0)$ and $p^n \to \hat{w}$, we just need to prove $\hat{w} \in S(w^0)$. Indeed, since $p^n \to \hat{w}$, for any fixed $i > 0$, we can find n_i such that

$$\|p^{n_i} - \hat{w}\| \leqslant \frac{1}{2i}. \tag{28}$$

For any n, $p^n \in S(w^0)$, then there exits a subsequence $\{w_{k_j}^n\}$ of $\{w^k\}$ that converges to p^n. That is, for any n,

$$w_{k_j}^n \to p^n, \quad j \to +\infty. \tag{29}$$

Since $p^{n_i} \in S(w^0)$, for fixed n_i, it follows from (29) that there exits j_{n_i} such that

$$\|w^{n_i}_{k_{j_{n_i}}} - p^{n_i}\| \leqslant \frac{1}{2i}. \tag{30}$$

Thus, it follows from (28) and (30) that

$$\|w^{n_i}_{k_{j_{n_i}}} - \hat{w}\| \leqslant \|w^{n_i}_{k_{j_{n_i}}} - p^{n_i}\| + \|p^{n_i} - \hat{w}\| \leqslant \frac{1}{i}.$$

Therefore, $\{w^{n_i}_{k_{j_{n_i}}}\}$ is a subsequence of $\{w^k\}$ that converges to $\hat{w}$. Hence, $\hat{w} \in S(w^0)$.

(ii) Let $w^* := (x_1^*, x_2^*, x_3^*, \lambda^*) \in S(w^0)$. Then there exists a subsequence $(x_1^{k_j}, x_2^{k_j}, x_3^{k_j}, \lambda^{k_j})$ of $(x_1^k, x_2^k, x_3^k, \lambda^k)$ converges to $(x_1^*, x_2^*, x_3^*, \lambda^*)$. Note that Lemma 4 implies

$$\|w^{k+1} - w^k\| \to 0, \quad k \to +\infty. \tag{31}$$

Then we know that $(x_1^{k_j+1}, x_2^{k_j+1}, x_3^{k_j+1}, \lambda^{k_j+1})$ also converges to $(x_1^*, x_2^*, x_3^*, \lambda^*)$. Since x_1^{k+1} is a global minimizer of $\mathscr{L}_\beta(x_1, x_2^k, x_3^k, \lambda^k)$ for the variable x_1, it holds

$$\mathscr{L}_\beta(x_1^{k+1}, x_2^k, x_3^k, \lambda^k) \leqslant \mathscr{L}_\beta(x_1^*, x_2^k, x_3^k, \lambda^k). \tag{32}$$

It follows from (31), (32), and the continuity of $\mathscr{L}_\beta(\cdot)$ with respect to x_2, x_3, and λ, we have

$$\begin{aligned}\limsup_{j\to+\infty} \mathscr{L}_\beta(x_1^{k_j+1}, x_2^{k_j+1}, x_3^{k_j+1}, \lambda^{k_j+1}) &= \limsup_{j\to+\infty} \mathscr{L}_\beta(x_1^{k_j+1}, x_2^{k_j}, x_3^{k_j}, \lambda^{k_j}) \\ &\leqslant \mathscr{L}_\beta(x_1^*, x_2^*, x_3^*, \lambda^*). \end{aligned} \tag{33}$$

On the other hand, from the lower semicontinuity of $\mathscr{L}_\beta(\cdot)$, we have

$$\liminf_{j\to+\infty} \mathscr{L}_\beta(x_1^{k_j+1}, x_2^{k_j+1}, x_3^{k_j+1}, \lambda^{k_j+1}) \geqslant \mathscr{L}_\beta(x_1^*, x_2^*, x_3^*, \lambda^*). \tag{34}$$

The above two relations (33) and (34) show that

$$\lim_{j\to+\infty} f_1(x_1^{k_j+1}) = f_1(x_1^*).$$

Because of the continuity of ∇f_2 and ∇f_3 and the closedness of ∂f_1, taking limit in (11) along the subsequence $(x_1^{k_j+1}, x_2^{k_j+1}, x_3^{k_j+1}, \lambda^{k_j+1})$ and using (31) again, we obtain

$$\begin{cases} A_1^{\mathrm{T}}\lambda^* \in \partial f_1(x_1^*), \\ \nabla f_2(x_2^*) = A_2^{\mathrm{T}}\lambda^*, \\ \nabla f_3(x_3^*) = \lambda^*, \\ A_1x_1^* + A_2x_2^* + x_3^* - b = 0. \end{cases}$$

Then, $(x_1^*, x_2^*, x_3^*, \lambda^*)$ satisfies system (8), and hence, $w^* \in \operatorname{crit} \mathscr{L}_\beta$.

(iii) For any point $(x_1^*, x_2^*, x_3^*, \lambda^*) \in S(w^0)$, there exists a subsequence $(x_1^{k_j}, x_2^{k_j}, x_3^{k_j}, \lambda^{k_j})$ converges to $(x_1^*, x_2^*, x_3^*, \lambda^*)$. By means of (33), (34), and $\{\mathscr{L}_\beta(w^k)\}_{k\in N}$ is nonincreasing, we obtain

$$\lim_{k\to+\infty} \mathscr{L}_\beta(x_1^k, x_2^k, x_3^k, \lambda^k) = \mathscr{L}_\beta(x_1^*, x_2^*, x_3^*, \lambda^*).$$

Therefore, $\mathscr{L}_\beta(\cdot)$ is constant on $S(w^0)$. Moreover, (27) holds. □

Remark 7 Based on [7, Remark 5], we can also show that $S(w^0)$ is a connected set; for more details, see [7].

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

Theorem 1 *Let* $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ *be the sequence generated by algorithm* (5) *which is assumed to be bounded. Suppose that* $\mathscr{L}_\beta(\cdot)$ *is a KL function. Then* $\{w^k\}$ *has finite length, that is,*

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

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

Proof From the proof of Lemma 6, it follows that $\mathscr{L}_\beta(w^k) \to \mathscr{L}_\beta(w^*)$ for all $w^* \in S(w^0)$. We consider two possible cases.

(i) The first case is that there exists an integer k_0 such that

$$\mathscr{L}_\beta(w^{k_0}) = \mathscr{L}_\beta(w^*).$$

Rearranging terms of (12) and by Remark 4, for any $k > k_0$, we have

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

and so, for any $k > k_0$, we have $v^{k+1} = v^k$. Associated with (19), (25), and (26), for any $k > k_0 + 1$, it follows that $w^{k+1} = w^k$ and the assertion holds.

(ii) The second case is that $\mathscr{L}_\beta(w^k) > \mathscr{L}_\beta(w^*)$ for all k. Since $d(w^k, S(w^0))$ $\to 0$, it follows that for any $\varepsilon > 0$, there exists $k_1 > 0$ such that for any $k > k_1$, $d(w^k, S(w^0)) < \varepsilon$. Again, since $\mathscr{L}_\beta(w^k) \to \mathscr{L}_\beta(w^*)$, it follows for all $\eta > 0$, there exists $k_2 > 0$ such that for any $k > k_2$,

$$\mathscr{L}_\beta(w^k) < \mathscr{L}_\beta(w^*) + \eta.$$

Consequently, for all $\varepsilon, \eta > 0$, when $k > \widetilde{k} := \max\{k_1, k_2\}$,

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

Since $S(w^0)$ is a nonempty compact set and $\mathscr{L}_\beta(\cdot)$ is constant on $S(w^0)$, applying Lemma 1 with $\Omega := S(w^0)$, we deduce that for any $k > \widetilde{k}$,

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

Since

$$\mathscr{L}_\beta(w^k) - \mathscr{L}_\beta(w^{k+1}) = (\mathscr{L}_\beta(w^k) - \mathscr{L}_\beta(w^*)) - (\mathscr{L}_\beta(w^{k+1}) - \mathscr{L}_\beta(w^*)),$$

using the concavity of φ, we get

$$\begin{aligned}&\varphi(\mathscr{L}_\beta(w^k) - \mathscr{L}_\beta(w^*)) - \varphi(\mathscr{L}_\beta(w^{k+1}) - \mathscr{L}_\beta(w^*)) \\ &\quad \geqslant \varphi'(\mathscr{L}_\beta(w^k) - \mathscr{L}_\beta(w^*))(\mathscr{L}_\beta(w^k) - \mathscr{L}_\beta(w^{k+1})).\end{aligned}$$

Thus, associating with

$$d(0, \partial\mathscr{L}_\beta(w^k)) \leqslant \zeta\|v^k - v^{k-1}\|, \quad \varphi'(\mathscr{L}_\beta(w^k) - \mathscr{L}_\beta(w^*)) > 0,$$

we know

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

For convenience, we set

$$\Delta_{p,q} := \varphi(\mathscr{L}_\beta(w^p) - \mathscr{L}_\beta(w^*)) - \varphi(\mathscr{L}_\beta(w^q) - \mathscr{L}_\beta(w^*)).$$

Combining Lemma 3 and the above inequality yields that for all $k > \widetilde{k}$,

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

and hence,

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

By using the fact

$$2\sqrt{\alpha\beta} \leqslant \alpha + \beta, \quad \forall\, \alpha, \beta > 0,$$

we obtain

$$2\|v^{k+1} - v^k\| \leqslant \|v^k - v^{k-1}\| + \frac{\zeta}{\delta}\,\Delta_{k,k+1}. \tag{35}$$

Summing up (35) for $k = \widetilde{k} + 1, \widetilde{k} + 2, \ldots, m$ yields

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

Notice that

$$\varphi(\mathscr{L}_\beta(w^{m+1}) - \mathscr{L}_\beta(w^*)) > 0.$$

Rearranging terms and letting $m \to +\infty$ yield

$$\sum_{k=\widetilde{k}+1}^{+\infty} \|v^{k+1} - v^k\| \leqslant \|v^{\widetilde{k}+1} - v^{\widetilde{k}}\| + \frac{\zeta}{\delta}\varphi(\mathscr{L}_\beta(w^{\widetilde{k}+1}) - \mathscr{L}_\beta(w^*)), \tag{36}$$

which means

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

Thus, we can deduce

$$\sum_{k=0}^{+\infty} \|x_2^{k+1} - x_2^k\| < +\infty, \quad \sum_{k=0}^{+\infty} \|x_3^{k+1} - x_3^k\| < +\infty.$$

Moreover, it follows from (19) that

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

On the other hand, it follows from (25) and (26) that

$$\begin{aligned}
&\|x_1^{k+1} - x_1^k\| \\
&\leqslant \sqrt{\frac{4}{\beta^2\mu}}\,(\|\lambda^{k+1} - \lambda^k\|^2 + \|\lambda^k - \lambda^{k-1}\|^2 + \beta^2\|x_3^k - x_3^{k+1}\|^2 \\
&\quad + \beta^2\|A_2\|^2\,\|x_2^{k+1} - x_3^k\|^2)^{1/2} \\
&\leqslant \sqrt{\frac{4}{\beta^2\mu}}\,(\|\lambda^{k+1} - \lambda^k\| + \|\lambda^k - \lambda^{k-1}\| + \beta\|x_3^k - x_3^{k+1}\| + \beta\|A_2\|\,\|x_2^{k+1} - x_2^k\|).
\end{aligned}$$

Hence,

$$\sum_{k=1}^{+\infty} \|x_1^{k+1} - x_1^k\| < +\infty.$$

Moreover, we note that

$$\|w^{k+1} - w^k\| \leqslant \|x_1^{k+1} - x_1^k\| + \|x_2^{k+1} - x_2^k\| + \|x_3^{k+1} - x_3^k\| + \|\lambda^{k+1} - \lambda^k\|.$$

Therefore,

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

$\{w^k\}$ is a Cauchy sequence (see [7, p. 482] for a simple proof), and thus is convergent. The assertion then follows immediately from Lemma 6. □

Remark 8 Actually, [2,3] proved an abstract convergence result for descent methods satisfying a sufficient decrease assumption, and allowing a relative error tolerance. However, as stated in [24], their results cannot be applied directly to our algorithm. In fact, their sufficient descent property in our case reads, there exits $\theta > 0$ such that

$$\mathscr{L}_\beta(w^{k+1}) \leqslant \mathscr{L}_\beta(w^k) - \theta \|w^{k+1} - w^k\|^2, \tag{37}$$

while we only have

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

which is not sufficient for (37) holding.

Next, we give some sufficient conditions to guarantee the sequence $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by the ADMM (5) is bounded.

Lemma 7 *Let $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by algorithm* (5)*. Suppose that*

$$\inf_{x_3} \Big\{ f_3(x_3) - \frac{1}{4L} \|\nabla f_3(x_3)\|^2 \Big\} =: \overline{f}_3 > -\infty.$$

If

$$\liminf_{\|x_1\| \to +\infty} f(x_1) = +\infty, \quad \liminf_{\|x_2\| \to +\infty} f_2(x_2) = +\infty, \tag{38}$$

then $\{w^k\}$ is bounded.

Proof From Lemma 3, we know that

$$\mathscr{L}_\beta(x_1^k, x_2^k, x_3^k, \lambda^k) \leqslant \mathscr{L}_\beta(x_1^1, x_2^1, x_3^1, \lambda^1).$$

Then, combining with $\lambda^k = \nabla f_3(x_3^k)$, we get

$$\begin{aligned}
&\mathscr{L}_\beta(x_1^1, x_2^1, x_3^1, \lambda^1) \\
&\quad \geqslant f_1(x_1^k) + f_2(x_2^k) + f_3(x_3^k) - \langle \lambda^k, A_1 x_1^k + A_2 x_2^k + x_3^k - b \rangle \\
&\qquad + \frac{\beta}{2} \|A_1 x_1^k + A_2 x_2^k + x_3^k - b\|^2 \\
&\quad = f_1(x_1^k) + f_2(x_2^k) + f_3(x_3^k) - \frac{1}{2\beta} \|\lambda^k\|^2 + \frac{\beta}{2} \Big\| A_1 x_1^k + A_2 x_2^k + x_3^k - b - \frac{1}{\beta} \lambda^k \Big\|^2 \\
&\quad = f_1(x_1^k) + f_2(x_2^k) + \Big(f_3(x_3^k) - \frac{1}{4L} \|\nabla f_3(x_3^k)\|^2 \Big) + \Big(\frac{1}{4L} - \frac{1}{2\beta} \Big) \|\lambda^k\|^2 \\
&\qquad + \frac{\beta}{2} \Big\| A_1 x_1^k + A_2 x_2^k + x_3^k - b - \frac{1}{\beta} \lambda^k \Big\|^2 \\
&\quad \geqslant f_1(x_1^k) + f_2(x_2^k) + \overline{f}_3 + \Big(\frac{1}{4L} - \frac{1}{2\beta} \Big) \|\lambda^k\|^2 \\
&\qquad + \frac{\beta}{2} \Big\| A_1 x_1^k + A_2 x_2^k + x_3^k - b - \frac{1}{\beta} \lambda^k \Big\|^2.
\end{aligned}$$

Observe that, (38) implies that

$$\inf_{x_1} f_1(x_1) > -\infty, \quad \inf_{x_2} f_2(x_2) > -\infty.$$

It follows from these and $\beta > 2L$ that

$$\{x_1^k\}, \quad \{x_2^k\}, \quad \{\lambda^k\}, \quad \left\{\frac{\beta}{2}\left\|A_1x_1^k + A_2x_2^k + x_3^k - b - \frac{1}{\beta}\lambda^k\right\|^2\right\},$$

are bounded. Therefore, $\{x_3^k\}$ is bounded, and hence, $\{w^k\}$ is bounded. □

4 Convergence rate

In this section, we establish the convergence rate for the ADMM procedure (2). Similar to the last section, we only consider the case $m = 3$. The main result is summarized in the following theorem.

Theorem 2 *Let $\{w^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by algorithm (5) and converges to $\{w^* := (x_1^*, x_2^*, x_3^*, \lambda^*)\}$. Assume that $\mathscr{L}_\beta(\cdot)$ has the KL property at $(x_1^*, x_2^*, x_3^*, \lambda^*)$ with $\varphi(s) = cs^{1-\theta}$, $\theta \in [0, 1)$, $c > 0$. Then the following estimations hold*:

(i) *if $\theta = 0$, then the sequence $\{w^k\}$ converges in a finite number of steps*;

(ii) *if $\theta \in (0, 1/2]$, then there exist $c > 0$ and $\tau \in [0, 1)$ such that*

$$\|(x_1^k, x_2^k, x_3^k, \lambda^k) - (x_1^*, x_2^*, x_3^*, \lambda^*)\| \leqslant c\tau^k;$$

(iii) *if $\theta \in (1/2, 1)$, then there exists $c > 0$ such that*

$$\|(x_1^k, x_2^k, x_3^k, \lambda^k) - (x_1^*, x_2^*, x_3^*, \lambda^*)\| \leqslant ck^{(\theta-1)/(2\theta-1)}.$$

Proof We first consider the case that $\theta = 0$; then $\varphi(s) = cs$ and $\varphi'(s) = c$. If $\{w^k\}$ does not converge in a finite number of steps, then the KL property at $(x_1^*, x_2^*, x_3^*, \lambda^*)$ yields for any k sufficiently large, $c \cdot d(0, \partial L_\beta(w^k)) \geqslant 1$, a contradiction to Lemma 5.

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

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

The triangle inequality yields $\Delta_k \geqslant \|v^k - v^*\|$, and it is therefore sufficient to estimate Δ_k. With these notations, it follows from (36) that

$$\Delta_{\widetilde{k}+1} \leqslant (\Delta_{\widetilde{k}} - \Delta_{\widetilde{k}+1}) + \frac{\zeta}{\delta}\varphi(\mathscr{L}_\beta(w^{\widetilde{k}+1}) - \mathscr{L}_\beta(w^*)).$$

Again, by the KL property at $(x^*, y^*, z^*, \lambda^*)$, we have

$$\varphi'(\mathscr{L}_\beta(w^{\widetilde{k}+1}) - \mathscr{L}_\beta(w^*))d(0, \partial\mathscr{L}_\beta(w^{\widetilde{k}+1})) \geqslant 1,$$

which is equivalent to

$$(\mathscr{L}_\beta(w^{\widetilde{k}+1}) - \mathscr{L}_\beta(w^*))^\theta \leqslant c \cdot (1-\theta)d(0, \partial\mathscr{L}_\beta(w^{\widetilde{k}+1})). \tag{39}$$

Using Lemma 5, we get

$$d(0, \partial\mathscr{L}_\beta(w^{\widetilde{k}+1})) \leqslant \zeta \cdot \|v^{\widetilde{k}+1} - v^{\widetilde{k}}\| = \zeta(\Delta_{\widetilde{k}} - \Delta_{\widetilde{k}+1}). \tag{40}$$

Combining (39) and (40), we obtain that there exists $\gamma > 0$ such that

$$\varphi(\mathscr{L}_\beta(w^{\widetilde{k}+1}) - \mathscr{L}_\beta(w^*)) = c \cdot (\mathscr{L}_\beta(w^{\widetilde{k}+1}) - \mathscr{L}_\beta(w^*))^{1-\theta} \leqslant \gamma(\Delta_{\widetilde{k}} - \Delta_{\widetilde{k}+1})^{(1-\theta)/\theta},$$

and hence,

$$\Delta_{\widetilde{k}+1} \leqslant (\Delta_{\widetilde{k}} - \Delta_{\widetilde{k}+1}) + \frac{\zeta}{\delta}\gamma(\Delta_{\widetilde{k}} - \Delta_{\widetilde{k}+1})^{(1-\theta)/\theta}.$$

Sequences satisfying such inequalities have been studied by Attouch and Bolte [1]. It follows that

- if $\theta \in (0, 1/2]$, then there exists $c_1 > 0$ and $\tau \in [0, 1)$ such that

$$\|v^k - v^*\| \leqslant c_1\tau^k,$$

and

- if $\theta \in (1/2, 1)$, then there exists $c_2 > 0$ such that

$$\|v^k - v^*\| \leqslant c_2 k^{(\theta-1)/(2\theta-1)}.$$

Thus, we have

- if $\theta \in (0, 1/2]$, then there exists $c_1 > 0$ and $\tau \in [0, 1)$ such that

$$\|x_2^k - x_2^*\| \leqslant c_1\tau^k, \quad \|x_3^k - x_3^*\| \leqslant c_1\tau^k, \tag{41}$$

and

- if $\theta \in (1/2, 1)$, then there exists $c_2 > 0$ such that

$$\|x_2^k - x_2^*\| \leqslant c_2 k^{(\theta-1)/(2\theta-1)}, \quad \|x_3^k - x_3^*\| \leqslant c_2 k^{(\theta-1)/(2\theta-1)}. \tag{42}$$

Recall that ∇f_3 is Lipschitz continuous with modulus $L_3 \leqslant L$. It follows from (8) and (11) that

$$\|\lambda^k - \lambda^*\| = \|\nabla f_3(x_3^k) - \nabla f_3(x_3^*)\| \leqslant L\|x_3^k - x_3^*\|. \tag{43}$$

Furthermore, from the relations

$$\lambda^k = \lambda^{k-1} - \beta(A_1x_1^k + A_2x_2^k + x_3^k - b)$$

and

$$A_1x_1^* + A_2x_2^* + x_3^* = b,$$

it follows that

$$\beta(A_1x_1^k - A_1x_1^*) = \beta(A_2x_2^* - A_2x_2^k) + \beta(x_3^* - x_3^k) + (\lambda^{k-1} - \lambda^*) + (\lambda^* - \lambda^k).$$

Therefore, there exists $\overline{\gamma} > 0$ such that

$$\begin{aligned}\|x_1^k - x_1^*\| &\leqslant \overline{\gamma}\Big(\|A_2\|\,\|x_2^k - x_2^*\| + \|x_3^k - x_3^*\| + \frac{1}{\beta}\,\|\lambda^{k-1} - \lambda^*\| + \frac{1}{\beta}\,\|\lambda^* - \lambda^k\|\Big)\\ &\leqslant \overline{\gamma}\Big[\|A_2\|\,\|x_2^k - x_2^*\| + \Big(1 + \frac{L}{\beta}\Big)\|x_3^k - x_3^*\| + \frac{L}{\beta}\,\|x_3^{k-1} - x_3^*\|\Big], \qquad (44)\end{aligned}$$

where the second inequality follows from (43). Combining (41)–(44), we get the desired inequalities immediately. □

5 Conclusions

In this paper, we analyzed the convergence of alternating direction method of multipliers (ADMM) for solving multi-block linearly constrained nonconvex minimization model without coupled variables where none of the involving functions are convex. Under the assumption that the associated function satisfies the Kurdyka-Łojasiewicz (KL) inequality, we proved that any cluster point of the iterative sequence generated by the algorithm is a critical point, provided that the penalty parameter is sufficiently large. Particularly, when the data functions f_1, f_2, and f_3 are semi-algebraic, the convergence rate of the algorithm was also established.

Furthermore, we prove Lemma 3 holds under the assumption that one of the objective functions is convex. In this case, under the assumption that the associated function satisfies the KL inequality, we can similarly prove that any cluster point of the iterative sequence generated by ADMM is a critical point, provided that the penalty parameter is greater than $2L$, where L is the Lipschitz constant of the gradient of one of the involving function. When the data functions f_1, f_2, and f_3 are semi-algebraic, we can also show the convergence rate of the algorithm.

As we have mentioned in the introduction, the nonconvex separable optimization model (1) finds many interesting application and ADMM exhibits great success in solving the model. One of our future research topic is using the model and algorithm to some other application field such as traffic assignment problem [10,31].

Acknowledgements This work was supported by a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. The research of Ke Guo was also supported by the National Natural Science Foundation of China (Grant No. 11371015); Deren Han by (Grant Nos. 11625105, 11371197, 11431002); Tingting Wu by (Grant No. 11501301), Jiangsu Planned Projects for Postdoctoral Research Funds (Grant No. 1501071B), and the Foundation of Jiangsu Key Lab for NSLSCS (Grant No. 201601).

Appendix Proof of Remark 5

Proof of Remark 5 Similar to the proof of Lemma 3, we can show

$$\begin{aligned}&\mathscr{L}_\beta(x_1^{k+1},x_2^{k+1},x_3^{k+1},\lambda^{k+1})\\&\quad\leqslant \mathscr{L}_\beta(x_1^{k+1},x_2^{k+1},x_3^{k},\lambda^{k})-\Big(\frac{\beta-L}{2}-\frac{L^2}{\beta}\Big)\|x_3^{k+1}-x_3^k\|^2.\end{aligned}\tag{A.1}$$

Recall that

$$\begin{aligned}&\mathscr{L}_\beta(x_1^{k+1},x_2^{k},x_3^{k},\lambda^{k})-\mathscr{L}_\beta(x_1^{k+1},x_2^{k+1},x_3^{k},\lambda^{k})\\&\quad=f_2(x_2^k)-f_2(x_2^{k+1})+\langle\lambda^k,A_2x_2^{k+1}-A_2x_2^k\rangle\\&\qquad+\frac{\beta}{2}\|A_1x_1^{k+1}+A_2x_2^k+x_3^k-b\|^2-\frac{\beta}{2}\|A_1x_1^{k+1}+A_2x_2^{k+1}+x_3^k-b\|^2.\end{aligned}\tag{A.2}$$

Since f_2 is a convex function, it follows from the second equality of (10) that

$$f_2(x_2^k)-f_2(x_2^{k+1})\geqslant\langle A_2^{\mathrm{T}}\lambda^k-\beta A_2^{\mathrm{T}}(A_1x_1^{k+1}+A_2x_2^{k+1}+x_3^k-b),x_2^k-x_2^{k+1}\rangle.\tag{A.3}$$

Inserting (A.3) into (A.2) and by means of (ii) of Remark 5, we have

$$\begin{aligned}&\mathscr{L}_\beta(x_1^{k+1},x_2^{k},x_3^{k},\lambda^{k})-\mathscr{L}_\beta(x_1^{k+1},x_2^{k+1},x_3^{k},\lambda^{k})\\&\quad\geqslant-\beta\langle A_1x_1^{k+1}+A_2x_2^{k+1}+x_3^k-b,A_2x_2^k-A_2x_2^{k+1}\rangle\\&\qquad+\frac{\beta}{2}\|A_1x_1^{k+1}+A_2x_2^k+x_3^k-b\|^2-\frac{\beta}{2}\|A_1x_1^{k+1}+A_2x_2^{k+1}+x_3^k-b\|^2\\&\quad\geqslant\frac{\beta\mu}{2}\|x_2^{k+1}-x_2^k\|^2.\end{aligned}\tag{A.4}$$

Thus, it follows from (A.1) and (A.4) that

$$\begin{aligned}&\mathscr{L}_\beta(x_1^{k+1},x_2^{k+1},x_3^{k+1},\lambda^{k+1})\\&\quad\leqslant\mathscr{L}_\beta(x_1^{k+1},x_2^{k+1},x_3^{k},\lambda^{k})-\Big(\frac{\beta-L}{2}-\frac{L^2}{\beta}\Big)\|x_3^{k+1}-x_3^k\|^2\\&\quad\leqslant\mathscr{L}_\beta(x_1^{k+1},x_2^{k},x_3^{k},\lambda^{k})-\Big(\frac{\beta-L}{2}-\frac{L^2}{\beta}\Big)\|x_3^{k+1}-x_3^k\|^2-\frac{\beta\mu}{2}\|x_2^{k+1}-x_2^k\|^2\\&\quad\leqslant\mathscr{L}_\beta(x_1^{k},x_2^{k},x_3^{k},\lambda^{k})-\delta\|v^{k+1}-v^k\|^2,\end{aligned}$$

where the third inequality follows from (i) of Remark 5 and the fact that x_1^{k+1} is the global minimizer of $\mathscr{L}_\beta(x_1,x_2^k,x_3^k,\lambda^k)$ with respect to variable x_1, i.e.,

$$\mathscr{L}_\beta(x_1^{k+1},x_2^{k},x_3^{k},\lambda^{k})\leqslant\mathscr{L}_\beta(x_1^{k},x_2^{k},x_3^{k},\lambda^{k}).$$

The proof is complete. □

References

1. Attouch H, Bolte J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math Program, 2009, 116: 5–16
2. Attouch H, Bolte J, Redont P, Soubeyran A. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality. Math Oper Res, 2010, 35: 438–457
3. Attouch H, Bolte J, Svaiter B F. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math Program, 2013, 137: 91–129
4. Boley D. Local linear convergence of ADMM on quadratic or linear programs. SIAM J Optim, 2013, 23: 2183–2207
5. Bolte J, Daniilidis A, Lewis A. The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J Optim, 2007, 17: 1205–1223
6. Bolte J, Daniilidis A, Lewis A, Shiota M. Clarke subgradients of stratifiable functions. SIAM J Optim, 2007, 18: 556–572
7. Bolte J, Sabach S, Teboulle M. Proximal alternating linearized minimization for non-convex and nonsmooth problem. Math Program, 2014, 146: 459–494
8. Cai X J, Han D R, Yuan X M. The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex. Comput Optim Appl, 2017, 66: 39–73
9. Chen C H, He B S, Ye Y Y, Yuan X M. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math Program, 2016, 155: 57–79
10. Du B, Wang D Z W. Continuum modeling of park-and-ride services considering travel time reliability and heterogeneous commuters—A linear complementarity system approach. Transportation Research Part E: Logistics and Transportation Review, 2014, 71: 58–81
11. Gabay D. Applications of the method of multipliers to variational inequalities. In: Fortin M, Glowinski R, eds. Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. Amsterdam: North-Holland, 1983, 299–331
12. Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Comput Math Appl, 1976, 2: 17–40
13. Glowinski R, Marrocco A. Approximation par éléments finis d'ordre un et résolution par pénalisation dualité d'une classe de problèmes non linéaires. RAIRO, Analyse numérique, 1975, 9(2): 41–76
14. Guo K, Han D R, Wu T T. Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints. Int J Comput Math, 2016, DOI: 10.1080/00207160.2016.1227432
15. Han D R, Yuan X M. A note on the alternating direction method of multipliers. J Optim Theory Appl, 2012, 155: 227–238
16. Han D R, Yuan X M. Local linear convergence of the alternating direction method of multipliers for quadratic programs. SIAM J Numer Anal, 2013, 51: 3446–3457
17. Han D R, Yuan X M, Zhang W X. An augmented-Lagrangian-based parallel splitting method for separable convex programming with applications to image processing. Math Comp, 2014, 83: 2263–2291
18. He B S, Tao M, Yuan X M. Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J Optim, 2012, 22: 313–340
19. He B S, Tao M, Yuan X M. Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming. Preprint
20. He B S, Yuan X M. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J Numer Anal, 2012, 50: 700–709

21. Hong M, Luo Z Q. On the linear convergence of alternating direction method of multipliers. Math Program, 2016, DOI: 10.1007/s10107-016-1034-2
22. Hong M, Luo Z Q, Razaviyayn M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM J Optim, 2016, 26: 337–364
23. Kurdyka K. On gradients of functions definable in o-minimal structures. Ann Inst Fourier (Grenoble), 1998, 48: 769–783
24. Li G, Pong T K. Global convergence of splitting methods for nonconvex composite optimization. SIAM J Optim, 2015, 25: 2434–2460
25. Li M, Sun D F, Toh K C. A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. Asia-Pac J Oper Res, 2015, 32: 1550024
26. Lojasiewicz S. Une propriété topologique des sous-ensembles analytiques réels. Les équations aux dérivées partielles, 1963, 117: 87–89
27. Mordukhovich B. Variational Analysis and Generalized Differentiation, I. Basic Theory. Grundlehren Math Wiss, Vol 330. Berlin: Springer, 2006
28. Nesterov Y. Introductory Lectures on Convex Optimization: A Basic Course. Boston: Kluwer Academic Publishers, 2004
29. Rockafellar R T. Convex Analysis. Princeton Univ Press, 2015
30. Rockafellar R T, Wets R J B. Variational Analysis. Berlin: Springer, 1998
31. Wang D Z W, Xu L L. Equilibrium trip scheduling in single bottleneck traffic flows considering multi-class travellers and uncertainty—a complementarity formulation. Transportmetrica A: Transport Science, 2016, 12(4): 297–312
32. Wen Z W, Yang C, Liu X, Marchesini S. Alternating direction methods for classical and ptychographic phase retrieval. Inverse Problems, 2012, 28: 115010
33. Yang L, Pong T K, Chen X J. Alternating direction method of multipliers for non-convex background/foreground extraction. 2015, arXiv: 1506.07029
34. Yang W H, Han D R. Linear convergence of alternating direction method of multipliers for a class of convex optimization problems. SIAM J Numer Anal, 2016, 54: 625–640