The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex

Xingju Cai* **Deren Han†** and **Xiaoming Yuan‡**

November 18, 2014

Abstract

The alternating direction method of multipliers (ADMM) is a benchmark for solving a two-block linearly constrained convex minimization model whose objective function is the sum of two functions without coupled variables. Meanwhile, it is known that the convergence is not guaranteed if the ADMM is directly extended to a multiple-block convex minimization model whose objective function has more than two functions. Recently, some authors have actively studied the strong convexity condition on the objective function to sufficiently ensure the convergence of the direct extension of ADMM or the resulting convergence when the original scheme is appropriately twisted. However, these strong convexity conditions still seem too strict to be satisfied by some applications for which the direct extension of ADMM work well; and the twisted schemes are less efficient or convenient to implement than the original scheme of the direct extension of ADMM. We are thus motivated to understand why the original scheme of the direct extension of ADMM works for some applications and under which realistic conditions its convergence can be guaranteed. We answer this question for the three-block case where there are three separable functions in the objective; and show that when one of them is strongly convex, the direct extension of ADMM is convergent. Note that the strong convexity of one function does hold for many applications. We further estimate the worst-case convergence rate measured by the iteration complexity in both the ergodic and nonergodic senses for the direct extension of ADMM, and show that its globally linear convergence in asymptotical sense can be guaranteed under some additional conditions.

Key words: Alternating Direction Method of Multipliers, Convergence Analysis, Convex Programming, Separable Structure.

1 Introduction

Many applications can be modeled or reformulated as convex minimization models with separable objective functions without coupled variables but the variables are coupled by linear

*School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, P.R. China. Email: `caixingju@njnu.edu.cn`. This author was supported by the NSFC grant 11401315, NSF from Jiangsu province BK 20140914.

†School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, P.R. China. Email: `handeren@njnu.edu.cn`. The research is supported by a project funded by PAPD of Jiangsu Higher Education Institutions and the NSFC grants 11371197 and 11431002.

‡Department of Mathematics, Hong Kong Baptist University, Hong Kong, P.R. China. Email: `xmyuan@hkbu.edu.hk`. This author was supported by the General Research Fund from Hong Kong Research Grants Council: HKBU 203613.

constraints. When the objective function is the sum of two convex functions without coupled variables, the alternating direction method of multipliers (ADMM) proposed in [12] has been well studied in the literature. It can be regarded as a splitting version of the classical augmented Lagrangian method (ALM) in [20, 30]; and it has found a variety of applications in many areas, mainly because of its feature of generating easier subproblems in iterations. We refer to, e.g., [2, 10, 13], for some reviews on the ADMM. On the other hand, there are many applications that can be modeled or reformulated as a multiple-block linearly constrained convex minimization model whose objective function is the sum of more than two functions without coupled variables. In fact, many applications can be captured by a three-block model where there are three functions in the objective, see e.g., the image alignment problem in [29], the robust principal component analysis model with noisy and incomplete data in [33], the latent variable Gaussian graphical model selection in [4, 28], the stable principal component pursuit with nonnegative constraint in [34], the quadratic discriminant analysis model in [27] and the quadratic conic programming in [22].

Inspired by the mentioned applications, we focus on the three-block linearly constrained separable convex minimization model:

$$\min\left\{\sum_{i=1}^{3}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{3}A_ix_i=b,\ x_i\in\mathcal{X}_i,\ i=1,2,3\right\},\tag{1.1}$$

where $\theta_i:\mathcal{R}^{n_i}\to\mathcal{R}$ are all closed proper convex functions (could be nonsmooth); $\mathcal{X}_i\subseteq\mathcal{R}^{n_i}$ are nonempty closed convex sets; $A_i\in\mathcal{R}^{l\times n_i}$; $b\in\mathcal{R}^l$; and $\sum_{i=1}^{3}n_i=n$. The solution set of (1.1) is assumed to be nonempty throughout our discussion. In our analysis, we also use the notation $\mathbf{x}:=(x_1,x_2,x_3)$ and $\mathcal{X}:=\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{X}_3$. Let the augmented Lagrangian function of (1.1) be

$$\mathcal{L}_\beta(x_1,x_2,x_3,\lambda):=\sum_{i=1}^{3}\theta_i(x_i)-\lambda^T(\sum_{i=1}^{3}A_ix_i-b)+\frac{\beta}{2}\left\|\sum_{i=1}^{3}A_ix_i-b\right\|^2$$

with $\lambda\in\mathcal{R}^l$ the Lagrange multiplier and $\beta>0$ a penalty parameter. Then, the direct extension of ADMM for (1.1) reads as

$$\begin{cases}
x_1^{k+1}=\arg\min\left\{\mathcal{L}_\beta(x_1,x_2^k,x_3^k,\lambda^k)\mid x_1\in\mathcal{X}_1\right\}, & \text{(1.2a)}\\
x_2^{k+1}=\arg\min\left\{\mathcal{L}_\beta(x_1^{k+1},x_2,x_3^k,\lambda^k)\mid x_2\in\mathcal{X}_2\right\}, & \text{(1.2b)}\\
x_3^{k+1}=\arg\min\left\{\mathcal{L}_\beta(x_1^{k+1},x_2^{k+1},x_3,\lambda^k)\mid x_3\in\mathcal{X}_3\right\}, & \text{(1.2c)}\\
\lambda^{k+1}=\lambda^k-\beta(\sum_{i=1}^{3}A_ix_i^{k+1}-b). & \text{(1.2d)}
\end{cases}$$

Although it is natural to consider the direct extension of ADMM (1.2a)-(1.2d) and this scheme does work very well for many concrete applications of (1.1)(e.g., [16, 29, 33, 34]), it has been shown in [5] that the scheme (1.2a)-(1.2d) is not necessarily convergent if no further conditions are posed on the model (1.1). Some works have been studied in the literature for guaranteeing the convergence of (1.2a)-(1.2d) and they can be classified into two categories. The first category is twisting the scheme (1.2a)-(1.2d) slightly and no more condition is assumed for the original model. For examples, in [17, 18], it was suggested to correct the output of (1.2a)-(1.2d) to generate a new iterate and the resulting prediction-correction schemes are guaranteed to be convergent. In [19], the x_i-subproblems are twisted slightly to render the convergence. In [21], the authors suggested attaching a shrinkage factor to the Lagrangian-multiplier updating step in (1.2d) and it was shown that the convergence of (1.2a)-(1.2d) is guaranteed when this factor is small enough to satisfy some error bound conditions. Note that these just-mentioned works are for the more general case with $m \geq 3$ blocks in its objective function. Most recently, it was studied in [22] that if certain proximal terms are used to regularize the x_i-subproblems, then the convergence of (1.2a)-(1.2d) can be guaranteed only when one of the functions in the objective of (1.1) is strongly convex. Indeed, a relaxation factor in $(0, \frac{1+\sqrt{5}}{2})$ initiated in [11] can be attached to the Lagrangian-multiplier updating step (1.2d) on the cost that the proximal terms and the penalty parameter β should be chosen more judiciously. More specifically, via some careful analysis, we actually can see that if the factor is chosen to be close to $\frac{1+\sqrt{5}}{2}$, then this implies that the penalty parameter β should be very close to 0 or the matrix of one proximal term must be positive definite and its smallest eigenvalue must be big enough. Note that the original scheme (1.2a)-(1.2d) usually performs all the twisted variants with provable convergence numerically, as shown in [16, 33]; and it seems to be the most convenient scheme to be used in practice compared with its variants.

The second category is assuming more conditions on the model (1.1) while retaining the original iterative scheme (1.2a)-(1.2d). Indeed, prior to the work [5] in which it was shown that (1.2a)-(1.2d) is not necessarily convergent, it was proved in [14] that the scheme (1.2a)-(1.2d) is convergent if all the functions θ_i are strongly convex and the penalty parameter β is chosen in a certain interval. Then, in [6, 23], this condition was relaxed and only two or more functions in the objective are required to be strongly convex to ensure the convergence of (1.2a)-(1.2d). However, assuming the strong convexity for two functions still excludes most of the applications that can be efficiently solved by the scheme (1.2a)-(1.2d). Thus, these conditions are of only theoretical interests and they are too strict to be satisfied by the mentioned applications. In other words, there is still a gap between the empirical efficiency of (1.2a)-(1.2d) for a variety of applications and the lack of theoretical conditions that can both ensure the convergence of (1.2a)-(1.2d) and be satisfied by such an application.

We are thus motivated to answer the question of under which realistic condition, that can be easily satisfied by some applications of (1.1), the original scheme (1.2a)-(1.2d) is guaranteed to be convergent. Our main purpose is to answer this question. More specifically, we show that the convergence of (1.2a)-(1.2d) can be ensured whenever one function in the objective of (1.1) is strongly convex, together with some minor restrictions on the coefficient matrices A_2 and A_3 in (1.1) and the penalty parameter β. Note that the strong convexity condition is satisfied by many of the mentioned applications of (1.1). Thus, the efficiency of (1.2a)-(1.2d) which has been long observed empirically, can be theoretically justified. A by-product of our analysis is establishing the convergence rate for the scheme (1.2a)-(1.2d), including the worst-case convergence rate measured by the iteration complexity and the linear convergence rate in asymptotical sense, under some additional assumptions. Note that the convergence rate of (1.2a)-(1.2d) has been studied in [6, 21, 23, 24] under stronger conditions.

The rest of this paper is organized as follows. Some necessary preliminaries for further analysis are provided in Section 2. In Section 3, we prove the convergence of the scheme (1.2a)-(1.2d) under the assumption that one of the three functions in (1.1) is strongly convex. Then, we establish the worst-case convergence rate measured by the iteration complexity in the ergodic and nonergodic senses in Sections 4 and 5, respectively. In Section 6, under some stronger conditions, we show that the scheme (1.2a)-(1.2d) is globally linear convergent. Finally, some concluding remarks are drawn in Section 7.

2 Preliminaries

In this section, we summarize some notation and preliminaries to be used for further analysis.

2.1 Notation

All vectors are column vectors. For two arbitrary vectors $x \in \mathcal{R}^n$ and $y \in \mathcal{R}^m$, we simply use $u = (x, y)$ to denote their adjunction. That is, (x, y) denotes the vector $(x^T, y^T)^T$. For $p \geq 1$, $\|x\|_p$ denotes the p-norm of x. For $p = 2$, we simply denote it as $\|x\|$.

Let M be a positive definite matrix; we use $\|x\|_M := \sqrt{x^T M x}$ to denote its M-norm and $\lambda_{min}(M)$ to denote the smallest eigenvalue of M. For an identity matrix I and a scalar $\beta > 0$, we simply use $\|x\|_\beta$ to denote $\|x\|_{\beta I}$. For a given matrix A, its norm is

$$\|A\|_p := \sup_{x \neq 0} \left\{ \frac{\|Ax\|_p}{\|x\|_p} \right\}.$$

Specially, for an symmetric matrix A, $\|A\|_2$ denotes its spectral norm.

A function $f : \mathcal{R}^n \to \mathcal{R}$ is convex if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y), \;\; \forall x, y \in \mathcal{R}^n;$$

and it is strongly convex with modulus $\mu > 0$ if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\mu}{2} t(1-t)\|x - y\|^2, \;\; \forall x, y \in \mathcal{R}^n;$$

where $t \in [0, 1]$.

A multifunction $F : \mathcal{R}^n \rightrightarrows \mathcal{R}^n$ means that $F(x)$ is a set in $\mathcal{R}^n$ [31]. For a multifunction $F : \mathcal{R}^n \rightrightarrows \mathcal{R}^n$, we say that F is monotone if

$$(x - y)^T(\xi - \zeta) \geq 0, \qquad \forall \xi \in F(x), \; \forall \zeta \in F(y),$$

and strongly monotone with modulus $\mu > 0$ if

$$(x - y)^T(\xi - \zeta) \geq \mu\|x - y\|^2, \qquad \forall \xi \in F(x), \; \forall \zeta \in F(y).$$

It is well known, see, e.g., [31], that for a convex function f, ∂f, the subdifferential of f, is a monotone multifunction; and for a strongly convex function f with modulus $\mu > 0$, ∂f is a strongly monotone multifunction with modulus μ. That is, if f is convex, then

$$(\xi - \zeta)^T(x - y) \geq 0, \qquad \forall \xi \in \partial f(x), \; \forall \zeta \in \partial f(y);$$

and if f is strongly convex with modulus $\mu > 0$, then

$$(\xi - \zeta)^T(x - y) \geq \mu\|x - y\|^2, \qquad \forall \xi \in \partial f(x), \; \forall \zeta \in \partial f(y).$$

For a nonempty convex subset C in $\mathcal{R}^n$, its normal cone operator, denoted by N_C, is defined as

$$N_C(u) = \begin{cases} \{v : \langle u' - u, v\rangle \leq 0, \quad \forall u' \in C\}, & \text{if} \quad u \in C, \\ \varnothing, & \text{otherwise.} \end{cases} \tag{2.1}$$

For any two vectors a and b with the same dimensions, we have

$$2a^T b \leq \varsigma\|a\|^2 + \frac{1}{\varsigma}\|b\|^2, \tag{2.2}$$

for any positive scalar ς. Obviously, we also have

$$\|a+b\|^2 \leq (1+\varsigma)\|a\|^2 + \left(1+\frac{1}{\varsigma}\right)\|b\|^2, \ \forall \varsigma > 0.$$

These two inequalities will be frequently used in the proofs to be presented.

2.2 Optimality Characterization

In this subsection, we present some variational forms to characterize a solution point of the model (1.1) under discussion. These characterizations will be the basis of our convergence analysis to be conducted.

Recall that the Lagrangian function of (1.1) is

$$\mathcal{L}(x_1,x_2,x_3,\lambda) := \theta_1(x_1)+\theta_2(x_2)+\theta_3(x_3) - \lambda^T\left(A_1x_1+A_2x_2+A_3x_3-b\right),$$

where λ is the Lagrange multiplier. Solving (1.1) amounts to finding a saddle point of $\mathcal{L}(x_1,x_2,x_3,\lambda)$. That is, finding $(\hat{x}_1,\hat{x}_2,\hat{x}_3,\hat{\lambda}) \in \mathcal{X}_1\times\mathcal{X}_2\times\mathcal{X}_3\times\mathcal{R}^l$ such that the following inequalities hold:

$$\mathcal{L}(\hat{x}_1,\hat{x}_2,\hat{x}_3,\lambda) \leq \mathcal{L}(\hat{x}_1,\hat{x}_2,\hat{x}_3,\hat{\lambda}) \leq \mathcal{L}(x_1,x_2,x_3,\hat{\lambda}), \ \ \forall (x_1,x_2,x_3,\lambda)\in\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{X}_3\times\mathcal{R}^l. \quad (2.3)$$

Let the set of all saddle points of $\mathcal{L}(x_1,x_2,x_3,\lambda)$ be denoted by S. Throughout, S is assumed to be nonempty. Let $(\hat{x}_1,\hat{x}_2,\hat{x}_3,\hat{\lambda}) \in S$ be an arbitrary solution in S. Then, we have

$$\begin{cases} 0 \in \partial\theta_1(\hat{x}_1) - A_1^T\hat{\lambda} + N_{\mathcal{X}_1}(\hat{x}_1), & (2.4a)\\ 0 \in \partial\theta_2(\hat{x}_2) - A_2^T\hat{\lambda} + N_{\mathcal{X}_2}(\hat{x}_2), & (2.4b)\\ 0 \in \partial\theta_3(\hat{x}_3) - A_3^T\hat{\lambda} + N_{\mathcal{X}_3}(\hat{x}_3), & (2.4c)\\ 0 = \sum_{i=1}^{3} A_i\hat{x}_i - b, & (2.4d) \end{cases}$$

where the normal cone operator $N_{\mathcal{X}_i}$ is as that defined in (2.1).

Note that (2.4a)-(2.4d) can also be rewritten as

$$\begin{cases} (x_1'-\hat{x}_1)^T(\hat{\xi}_1 - A_1^T\hat{\lambda}) \geq 0, \ \forall x_1' \in \mathcal{X}_1, & (2.5a)\\ (x_2'-\hat{x}_2)^T(\hat{\xi}_2 - A_2^T\hat{\lambda}) \geq 0, \ \forall x_2' \in \mathcal{X}_2, & (2.5b)\\ (x_3'-\hat{x}_3)^T(\hat{\xi}_3 - A_3^T\hat{\lambda}) \geq 0, \ \forall x_3' \in \mathcal{X}_3, & (2.5c)\\ 0 = \sum_{i=1}^{3} A_i\hat{x}_i - b, & (2.5d) \end{cases}$$

where $\hat{\xi}_i \in \partial\theta_i(\hat{x}_i)$ for $i=1,2,3$. Thus, it follows from (2.5a)-(2.5c) that for $i=1,2,3$, we have

$$(x_i'-\hat{x}_i)^T\hat{\xi}_i \geq (A_ix_i' - A_i\hat{x}_1)^T\hat{\lambda}, \ \forall x_i' \in \mathcal{X}_i. \quad (2.6)$$

In our analysis, we also use the primal-dual gap function (see e.g., [3, 15, 21]) defined as

$$\begin{aligned} \mathcal{G}_{\mathcal{B}_{\mathbf{x}}\times\mathcal{B}_\lambda}(\mathbf{x},\lambda) \ := \ & \max_{\lambda'\in\mathcal{B}_\lambda}\{\theta_1(x_1)+\theta_2(x_2)+\theta_3(x_3) - (A_1x_1+A_2x_2+A_3x_3-b)^T\lambda'\} \\ & -\min_{\mathbf{x}'\in\mathcal{B}_{\mathbf{x}}}\{\theta_1(x_1')+\theta_2(x_2')+\theta_3(x_3') - (A_1x_1'+A_2x_2'+A_3x_3'-b)^T\lambda\}, \end{aligned}$$

where $\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}$ is a subset of $\mathcal{X} \times \mathcal{R}^l$. If $\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}$ contains a solution $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\lambda})$, then for any $(\mathbf{x}, \lambda) \in \mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}$, we have

$$
\begin{aligned}
&\mathcal{G}_{\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}}(\mathbf{x}, \lambda) \\
&\geq \left(\sum_{i=1}^{3} \theta_i(x_i) - \hat{\lambda}^T \left(\sum_{i=1}^{3} A_i x_i - b \right) \right) - \left(\sum_{i=1}^{3} \theta_i(\hat{x}_i) - \lambda^T \left(\sum_{i=1}^{3} A_i \hat{x}_i - b \right) \right) \\
&= \sum_{i=1}^{3} (\theta_i(x_i) - \theta_i(\hat{x}_i) - \hat{\lambda}^T (A_i x_i - A_i \hat{x}_i) \\
&\geq \sum_{i=1}^{3} (\theta_i(x_i) - \theta_i(\hat{x}_i) - \hat{\xi}_i^T (x_i - \hat{x}_i)) \\
&\geq 0,
\end{aligned} \tag{2.7}
$$

where the equality follows from (2.5d), the second inequality follows from (2.6), and the last inequality follows from the convexity of θ_i, $i = 1, 2, 3$.

The following lemma gives a sufficient condition for a point to be a saddle point. The assertion is an immediate conclusion of (2.7); we thus omit the proof.

Lemma 2.1. *Let $(\bar{\mathbf{x}}, \bar{\lambda})$ be an arbitrary point in $\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}$. If*

$$\mathcal{G}_{\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}}(\bar{\mathbf{x}}, \bar{\lambda}) \leq 0,$$

then $(\bar{\mathbf{x}}, \bar{\lambda})$ is a saddle point.

Based on this lemma, if $(\mathbf{x}, \lambda) \in \mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}$ and $\mathcal{G}_{\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\lambda}}(\bar{\mathbf{x}}, \bar{\lambda}) \leq \epsilon$ for some $\epsilon > 0$, then we call $(\mathbf{x}, \lambda)$ a saddle point of $\mathcal{L}(x_1, x_2, x_3, \lambda)$ with an accuracy of ϵ.

Finally, we use the notation

$$Q := \begin{pmatrix} \beta^2 A_2^T A_2 & 0 & 0 \\ 0 & \beta^2 A_3^T A_3 & 0 \\ 0 & 0 & I \end{pmatrix}, \tag{2.8}$$

where the matrices A_2 and A_3 are given in (1.1), and β is the penalty parameter in the direct extension of ADMM (1.2a)-(1.2d).

3 Convergence

In this section, we prove the convergence of the direct extension of ADMM (1.2a)-(1.2d) under the following assumption.

Assumption 3.1. *In (1.1), the functions θ_1 and θ_2 are convex; the function θ_3 is strongly convex with modulus $\mu_3 > 0$; A_2 and A_3 in (1.1) are full column rank matrices.*

Before the proof, let us present the variational characterization of the scheme (1.2a)-(1.2d). Indeed, similar as (2.4a)-(2.4d), the subproblems (1.2a)-(1.2c) can be characterized as

$$
\begin{cases}
0 \in \partial\theta_1(x_1^{k+1}) - A_1^T \lambda^k + \beta A_1^T (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b) + N_{\mathcal{X}_1}(x_1^{k+1}), \\
0 \in \partial\theta_2(x_2^{k+1}) - A_2^T \lambda^k + \beta A_2^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b) + N_{\mathcal{X}_2}(x_2^{k+1}), \\
0 \in \partial\theta_3(x_3^{k+1}) - A_3^T \lambda^k + \beta A_3^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) + N_{\mathcal{X}_3}(x_3^{k+1}).
\end{cases}
$$

Combining the step (1.2d) and (2.1), the $(k+1)$-th iterate generated by (1.2a)-(1.2d) satisfies $(\mathbf{x}^{k+1}, \lambda^{k+1}) \in \Omega$ and $\xi_i^{k+1} \in \partial\theta_i(x_i^{k+1})$ for $i = 1, 2, 3$, such that

$$\begin{cases} (x_1' - x_1^{k+1})^T \left(\xi_1^{k+1} - A_1^T\lambda^k + \beta A_1^T(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b)\right) \geq 0, \ \ \forall x_1' \in \mathcal{X}_1, & (3.2a) \\ (x_2' - x_2^{k+1})^T \left(\xi_2^{k+1} - A_2^T\lambda^k + \beta A_2^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b)\right) \geq 0, \ \ \forall x_2' \in \mathcal{X}_2, & (3.2b) \\ (x_3' - x_3^{k+1})^T \left(\xi_3^{k+1} - A_3^T\lambda^k + \beta A_3^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)\right) \geq 0, \ \ \forall x_3' \in \mathcal{X}_3, & (3.2c) \\ \lambda^{k+1} = \lambda^k - \beta\left(\sum_{i=1}^{3} A_ix_i^{k+1} - b\right). & (3.2d) \end{cases}$$

Now, we start proving the convergence of (1.2a)-(1.2d) under Assumption 3.1. First, we prove several lemmas.

Lemma 3.1. *Suppose Assumption 3.1 holds. For the iterative sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by the direct extension of ADMM (1.2a)-(1.2d), we have*

$$(A_3x_3^{k+1} - A_3x_3^k)^T(\lambda^{k+1} - \lambda^k) \geq \mu_3\|x_3^k - x_3^{k+1}\|^2, \tag{3.3}$$

and

$$\begin{aligned} &(A_2x_2^{k+1} - A_2x_2^k)^T(\lambda^{k+1} - \lambda^k) \\ &\geq -\left\{\frac{3}{8}\|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{4}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 + \frac{4}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2\right\}. \end{aligned} \tag{3.4}$$

Proof. It follows from (3.2d)

$$\sum_{j=1}^{3} A_ix_i^{k+1} - b = \frac{1}{\beta}(\lambda^k - \lambda^{k+1}). \tag{3.5}$$

Then (3.2c) can be rewritten into

$$(x_3' - x_3^{k+1})^T(\xi_3^{k+1} - A_3^T\lambda^{k+1}) \geq 0, \ \ \forall x_3' \in \mathcal{X}_3. \tag{3.6}$$

Setting $x_3' := x_3^k$ in (3.6) yields

$$(x_3^k - x_3^{k+1})^T(\xi_3^{k+1} - A_3^T\lambda^{k+1}) \geq 0,$$

and setting $x_3' := x_3^{k+1}$ and $x_2' := x_2^{k+1}$ for the kth iteration yields,

$$(x_3^{k+1} - x_3^k)^T(\xi_3^k - A_3^T\lambda^k) \geq 0.$$

Adding the above two inequalities, we have

$$(x_3^k - x_3^{k+1})^T\left((\xi_3^{k+1} - \xi_3^k) - A_3^T(\lambda^{k+1} - \lambda^k)\right) \geq 0.$$

By rearranging terms, (3.3) follows immediately from the above inequality and the convexity of θ_3. The inequality (3.3) is proved.

We now prove (3.4). Using (3.5) again, we can rewrite (3.2b) as

$$(x_2' - x_2^{k+1})^T\left(\xi_2^{k+1} - A_2^T\lambda^{k+1} + \beta A_2^T(A_3x_3^k - A_3x_3^{k+1})\right) \geq 0, \ \ \forall x_2' \in \mathcal{X}_2. \tag{3.7}$$

Then setting $x_2' := x_2^k$ in (3.7) and $x_2' := x_2^{k+1}$ for the kth iteration yields

$$(x_2^k - x_2^{k+1})^T \left(\xi_2^{k+1} - A_2^T \lambda^{k+1} + \beta A_2^T (A_3 x_3^k - A_3 x_3^{k+1})\right) \geq 0,$$

and

$$(x_2^{k+1} - x_2^k)^T \left(\xi_2^k - A_2^T \lambda^k + \beta A_2^T (A_3 x_3^{k-1} - A_3 x_3^k)\right) \geq 0,$$

respectively. Adding these two inequalities, we get

$$\begin{aligned}(x_2^k - x_2^{k+1})^T \Big((\xi_2^{k+1} - \xi_2^k) - A_2^T(\lambda^{k+1} - \lambda^k)\\ +\beta A_2^T (A_3 x_3^k - A_3 x_3^{k+1}) - \beta A_2^T (A_3 x_3^{k-1} - A_3 x_3^k)\Big) \geq 0.\end{aligned}$$

Rearranging terms, it follows that

$$\begin{aligned}&(A_2 x_2^{k+1} - A_2 x_2^k)^T (\lambda^{k+1} - \lambda^k)\\ &\geq \beta (A_2 x_2^{k+1} - A_2 x_2^k)^T \left\{(A_3 x_3^k - A_3 x_3^{k-1}) - (A_3 x_3^{k+1} - A_3 x_3^k)\right\}\\ &\geq -\left\{\frac{3}{16}\|A_2 x_2^{k+1} - A_2 x_2^k\|_\beta^2 + \frac{4}{3}\|A_3 x_3^k - A_3 x_3^{k-1}\|_\beta^2\right\}\\ &\quad -\left\{\frac{3}{16}\|A_2 x_2^{k+1} - A_2 x_2^k\|_\beta^2 + \frac{4}{3}\|A_3 x_3^{k+1} - A_3 x_3^k\|_\beta^2\right\}\\ &= -\left\{\frac{3}{8}\|A_2 x_2^{k+1} - A_2 x_2^k\|_\beta^2 + \frac{4}{3}\|A_3 x_3^k - A_3 x_3^{k-1}\|_\beta^2 + \frac{4}{3}\|A_3 x_3^{k+1} - A_3 x_3^k\|_\beta^2\right\},\end{aligned}$$

where the first inequality follows from the convexity of θ_2 and the second inequality from (2.2). The proof is complete. □

Lemma 3.2. *Suppose Assumption 3.1 holds. Let* $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ *be the sequence generated by the direct extension of ADMM (1.2a)-(1.2d). Then, for all* $(x_1', x_2', x_3', \lambda') \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{R}^l$*, we have*

$$\begin{aligned}&\left[\sum_{i=1}^{3} \theta_i(x_i^{k+1}) - \lambda'^T \left(\sum_{i=1}^{3} A_i x_i^{k+1} - b\right)\right] - \left[\sum_{i=1}^{3} \theta_i(x_i') - {\lambda^{k+1}}^T \left(\sum_{i=1}^{3} A_i x_i' - b\right)\right]\\ &\leq (\lambda^{k+1} - \lambda')^T \left(\sum_{i=1}^{3} A_i x_i^{k+1} - b\right)\\ &\quad +\beta (A_1 x_1' - A_1 x_1^{k+1} + A_2 x_2' - A_2 x_2^{k+1})^T (A_3 x_3^k - A_3 x_3^{k+1})\\ &\quad +\beta (A_1 x_1' - A_1 x_1^{k+1})^T (A_2 x_2^k - A_2 x_2^{k+1}) - \mu_3 \|x_3' - x_3^{k+1}\|^2.\end{aligned} \tag{3.8}$$

Proof. Using (3.2d) and (3.5), we know that (3.2a) is equivalent to

$$(x_1' - x_1^{k+1})^T \left(\xi_1^{k+1} - A_1^T \lambda^{k+1} + \beta A_1^T \left(\sum_{j=2}^{3} A_j (x_j^k - x_j^{k+1})\right)\right) \geq 0, \ \forall x_1' \in \mathcal{X}_1. \tag{3.9}$$

Rearranging terms yields

$$\begin{aligned}(x_1' - x_1^{k+1})^T \xi_1^{k+1} \quad &\geq \quad (A_1 x_1' - A_1 x_1^{k+1})^T \lambda^{k+1}\\ &\quad -\beta (A_1 x_1' - A_1 x_1^{k+1})^T \left(\sum_{j=2}^{3} A_j (x_j^k - x_j^{k+1})\right).\end{aligned}$$

It follows from the convexity of θ_1 that for all $x_1' \in \mathcal{R}^{n_1}$ and $\lambda' \in \mathcal{R}^l$, we have

$$\begin{aligned}
&\theta_1(x_1') - \theta_1(x_1^{k+1}) - (A_1x_1' - A_1x_1^{k+1})^T\lambda' \\
&\geq (x_1' - x_1^{k+1})^T\xi_1^{k+1} - (A_1x_1' - A_1x_1^{k+1})^T\lambda' \\
&\geq (A_1x_1' - A_1x_1^{k+1})^T(\lambda^{k+1} - \lambda') \\
&\quad -\beta(A_1x_1' - A_1x_1^{k+1})^T\left(\sum_{j=2}^{3} A_j(x_j^k - x_j^{k+1})\right).
\end{aligned} \tag{3.10}$$

Similarly, (3.2b) indicates that, for all $x_2' \in \mathcal{X}_2$, we have

$$(x_2' - x_2^{k+1})^T\xi_2^{k+1} \geq (A_2x_2' - A_2x_2^{k+1})^T\lambda^{k+1} - \beta(A_2x_2' - A_2x_2^{k+1})^T A_3(x_3^k - x_3^{k+1}).$$

In addition, it follows from the convexity of θ_2 that for all $x_2' \in \mathcal{R}^{n_2}$ and $\lambda' \in \mathcal{R}^l$, we have

$$\begin{aligned}
&\theta_2(x_2') - \theta_2(x_2^{k+1}) - (A_2x_2' - A_2x_2^{k+1})^T\lambda' \\
&\geq (x_2' - x_2^{k+1})^T\xi_2^{k+1} - (A_2x_2' - A_2x_2^{k+1})^T\lambda' \\
&\geq (A_2x_2' - A_2x_2^{k+1})^T(\lambda^{k+1} - \lambda') \\
&\quad -\beta(A_2x_2' - A_2x_2^{k+1})^T(A_3x_3^k - A_3x_3^{k+1}).
\end{aligned} \tag{3.11}$$

Then, (3.2c) is equivalent to,

$$(x_3' - x_3^{k+1})^T\xi_3^{k+1} \geq (A_3x_3' - A_3x_3^{k+1})^T\lambda^{k+1} \quad \forall x_3' \in \mathcal{X}_3.$$

Since θ_3 is strongly convex with modulus $\mu_3 > 0$, for all $x_3' \in \mathcal{R}^{n_3}$ and $\lambda' \in \mathcal{R}^l$, we have

$$\begin{aligned}
&\theta_3(x_3') - \theta_3(x_3^{k+1}) - (A_3x_3' - A_3x_3^{k+1})^T\lambda' \\
&\geq (x_3' - x_3^{k+1})^T\xi_3^{k+1} + \mu_3\|x_3' - x_3^{k+1}\|^2 - (A_3x_3' - A_3x_3^{k+1})^T\lambda' \\
&\geq (A_3x_3' - A_3x_3^{k+1})^T(\lambda^{k+1} - \lambda') + \mu_3\|x_3' - x_3^{k+1}\|^2.
\end{aligned} \tag{3.12}$$

Adding (3.10), (3.11), (3.12) and rearranging terms we obtain the assertion immediately. $\square$

Now, we need to further analyze the right-hand side of (3.8) and refine the assertion (3.8). This is done in the following lemma.

Lemma 3.3. *Suppose Assumption 3.1 holds. Let $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by the direct extension of ADMM (1.2a)-(1.2d). Then, for all $(x_1', x_2', x_3', \lambda') \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{R}^l$, we have*

$$\begin{aligned}
&2\beta\left[\sum_{i=1}^{3}\theta_i(x_i^{k+1}) - \lambda'^T\left(\sum_{i=1}^{3}A_ix_i^{k+1} - b\right)\right] - 2\beta\left[\sum_{i=1}^{3}\theta_i(x_i') - (\lambda^{k+1})^T\left(\sum_{i=1}^{3}A_ix_i' - b\right)\right] \\
&\leq \sum_{i=2}^{3}\left(\|A_ix_i^k - A_ix_i'\|_{\beta^2}^2 - \|A_ix_i^{k+1} - A_ix_i'\|_{\beta^2}^2\right) + (\|\lambda^k - \lambda'\|^2 - \|\lambda^{k+1} - \lambda'\|^2) \\
&\quad + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \\
&\quad - \frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 - \left(2\beta\mu_3\|x_3^{k+1} - x_3'\|^2 - \frac{13}{3}\|A_3x_3^{k+1} - A_3x_3'\|_{\beta^2}^2\right) \\
&\quad + \frac{13}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 - 2\beta\mu_3\|x_3^k - x_3^{k+1}\|^2 - \|\lambda^k - \lambda^{k+1}\|^2 \\
&\quad + 2\beta^2\left(\sum_{i=1}^{3}A_ix_i' - b\right)^T\left\{A_2x_2^k - A_2x_2^{k+1} + A_3x_3^k - A_3x_3^{k+1}\right\}.
\end{aligned} \tag{3.13}$$

Proof. Because of the identity

$$x^T y = \frac{1}{2}(\|x+y\|^2 - \|x\|^2 - \|y\|^2), \tag{3.14}$$

it follows that

$$\begin{aligned}
&\beta(A_2x_2' - A_2x_2^{k+1})^T(A_2x_2^k - A_2x_2^{k+1}) \\
&= \frac{1}{2}\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_\beta^2 + \|A_2x_2^{k+1} - A_2x_2'\|_\beta^2 - \|A_2x_2^k - A_2x_2'\|_\beta^2\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\beta(A_3x_3' - A_3x_3^{k+1})^T(A_3x_3^k - A_3x_3^{k+1}) \\
&= \frac{1}{2}\left\{\|A_3x_3^k - A_3x_3^{k+1}\|_\beta^2 + \|A_3x_3^{k+1} - A_3x_3'\|_\beta^2 - \|A_3x_3^k - A_3x_3'\|_\beta^2\right\}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\beta(A_1x_1' - A_1x_1^{k+1} + A_2x_2' - A_2x_2^{k+1})^T(A_3x_3^k - A_3x_3^{k+1}) \\
&= \beta\left[\left(\sum_{i=1}^3 A_ix_i' - b\right) - \left(\sum_{i=1}^3 A_ix_i^{k+1} - b\right)\right]^T(A_3x_3^k - A_3x_3^{k+1}) \\
&\quad -\beta(A_3x_3' - A_3x_3^{k+1})^T(A_3x_3^k - A_3x_3^{k+1}) \\
&= \beta\left[\left(\sum_{i=1}^3 A_ix_i' - b\right) - \left(\sum_{i=1}^3 A_ix_i^{k+1} - b\right)\right]^T(A_3x_3^k - A_3x_3^{k+1}) \\
&\quad -\frac{1}{2}\left\{\|A_3x_3^k - A_3x_3^{k+1}\|_\beta^2 + \|A_3x_3^{k+1} - A_3x_3'\|_\beta^2 - \|A_3x_3^k - A_3x_3'\|_\beta^2\right\} \\
&= \beta\left(\sum_{i=1}^3 A_ix_i' - b\right)^T(A_3x_3^k - A_3x_3^{k+1}) - (\lambda^k - \lambda^{k+1})^T(A_3x_3^k - A_3x_3^{k+1}) \\
&\quad -\frac{1}{2}\left\{\|A_3x_3^k - A_3x_3^{k+1}\|_\beta^2 + \|A_3x_3^{k+1} - A_3x_3'\|_\beta^2 - \|A_3x_3^k - A_3x_3'\|_\beta^2\right\} \\
&\le \beta\left(\sum_{i=1}^3 A_ix_i' - b\right)^T(A_3x_3^k - A_3x_3^{k+1}) - \frac{1}{2}\|A_3x_3^k - A_3x_3^{k+1}\|_\beta^2 - \mu_3\|x_3^k - x_3^{k+1}\|^2 \\
&\quad -\frac{1}{2}\left\{\|A_3x_3^{k+1} - A_3x_3'\|_\beta^2 - \|A_3x_3^k - A_3x_3'\|_\beta^2\right\},
\end{aligned} \tag{3.15}$$

where the third equality follows from (3.5) and the inequality follows from (3.3).

On the other hand, using (3.4)

$$\begin{aligned}
&-\beta\left[\left(\sum_{i=1}^3 A_ix_i^{k+1} - b\right) + (A_3x_3' - A_3x_3^{k+1})\right]^T(A_2x_2^k - A_2x_2^{k+1}) \\
&\le \frac{3}{8}\|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{4}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 + \frac{4}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2 \\
&\quad +\beta(A_3x_3^{k+1} - A_3x_3')^T(A_2x_2^k - A_2x_2^{k+1}) \\
&\le \frac{3}{8}\|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{4}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 + \frac{4}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2 \\
&\quad +\frac{13}{6}\|A_3x_3^{k+1} - A_3x_3'\|_\beta^2 + \frac{3}{26}\|A_2x_2^k - A_2x_2^{k+1}\|_\beta^2 \\
&= \frac{51}{104}\|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{4}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 + \frac{4}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2 \\
&\quad +\frac{13}{6}\|A_3x_3^{k+1} - A_3x_3'\|_\beta^2.
\end{aligned}$$

where the inequality follows from (2.2). Then

$$\begin{aligned}
&\beta(A_1x_1' - A_1x_1^{k+1})^T(A_2x_2^k - A_2x_2^{k+1})\\
&= \beta\left(\sum_{i=1}^3 A_ix_i' - b\right)^T (A_2x_2^k - A_2x_2^{k+1}) - \beta(A_2x_2' - A_2x_2^{k+1})^T(A_2x_2^k - A_2x_2^{k+1})\\
&\quad -\beta\left[\left(\sum_{i=1}^3 A_ix_i^{k+1} - b\right) + (A_3x_3' - A_3x_3^{k+1})\right]^T (A_2x_2^k - A_2x_2^{k+1})\\
&\le \beta\left(\sum_{i=1}^3 A_ix_i' - b\right)^T (A_2x_2^k - A_2x_2^{k+1}) - \frac{1}{2}\left\{\|A_2x_2^{k+1} - A_2x_2'\|_\beta^2 - \|A_2x_2^k - A_2x_2'\|_\beta^2\right\}\\
&\quad +\frac{4}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 - \frac{4}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2\\
&\quad -\frac{1}{104}\|A_2x_2^k - A_2x_2^{k+1}\|_\beta^2 + \frac{13}{6}\|A_3x_3^{k+1} - A_3x_3'\|_\beta^2.
\end{aligned}\tag{3.16}$$

Furthermore,

$$\begin{aligned}
(\lambda^{k+1} - \lambda')^T\left(\sum_{i=1}^3 A_ix_i^{k+1} - b\right) &= \frac{1}{\beta}(\lambda^{k+1} - \lambda')^T(\lambda^k - \lambda^{k+1})\\
&= \frac{1}{2\beta}\left\{\|\lambda^k - \lambda'\|^2 - \|\lambda^{k+1} - \lambda'\|^2\right\} - \frac{1}{2\beta}\|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned}\tag{3.17}$$

Thus, the assertion (3.13) follows directly from (3.15), (3.16), (3.17) and (3.8). □

Now, we are ready to prove the convergence of the direct extension of ADMM (3.2a)-(3.2d) under Assumption 3.1. The result is summarized in the following theorem.

Theorem 3.1. *Suppose Assumption 3.1 holds . Let $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by the direct extension of ADMM (1.2a)-(1.2d) with $\beta \in \left(0, \frac{6\mu_3}{13\|A_3^TA_3\|}\right)$. Then, the sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ converges to a saddle-point in S.*

Proof. Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\lambda})$ be a saddle point in S. Then, we have

$$\sum_{i=1}^3 A_i\hat{x}_i - b = 0. \tag{3.18}$$

Setting $(x_1', x_2', x_3', \lambda') := (\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\lambda})$ in (3.13) and using (3.18), we obtain

$$\begin{aligned}
&2\beta\left[\sum_{i=1}^3 \theta_i(x_i^{k+1}) - \hat{\lambda}^T\left(\sum_{i=1}^3 A_ix_i^{k+1} - b\right)\right] - 2\beta\sum_{i=1}^3 \theta_i(\hat{x}_i)\\
&\le \sum_{i=2}^3 \left(\|A_ix_i^k - A_i\hat{x}_i\|_{\beta^2}^2 - \|A_ix_i^{k+1} - A_i\hat{x}_i\|_{\beta^2}^2\right) + \left(\|\lambda^k - \hat{\lambda}\|^2 - \|\lambda^{k+1} - \hat{\lambda}\|^2\right)\\
&\quad +\frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2\\
&\quad -\frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 - \left(2\beta\mu_3 - \frac{13}{3}\beta^2\|A_3^TA_3\|\right)\|x_3^{k+1} - \hat{x}_3\|^2\\
&\quad -\left(2\beta\mu_3 - \frac{13}{3}\beta^2\|A_3^TA_3\|\right)\|x_3^{k+1} - x_3^k\|^2 - \|\lambda^k - \lambda^{k+1}\|^2,
\end{aligned}\tag{3.19}$$

where the second inequality follows from the Cauchy-Schwarz inequality. Since $\beta \in \left(0, \frac{6\mu_3}{13\|A_3^T A_3\|}\right)$, we have

$$\varrho_1 := 2\beta\mu_3 - \frac{13}{3}\beta^2\|A_3^T A_3\| > 0.$$

Then, (3.19) can be written as

$$\begin{aligned}
&2\beta\left[\sum_{i=1}^{3}\theta_i(x_i^{k+1}) - \hat{\lambda}^T\left(\sum_{i=1}^{3}A_i x_i^{k+1} - b\right)\right] - 2\beta\sum_{i=1}^{3}\theta_i(\hat{x}_i) \\
&\leq \sum_{i=2}^{3}\left(\|A_i x_i^k - A_i\hat{x}_i\|_{\beta^2}^2 - \|A_i x_i^{k+1} - A_i\hat{x}_i\|_{\beta}^2\right) + \left(\|\lambda^k - \hat{\lambda}\|^2 - \|\lambda^{k+1} - \hat{\lambda}\|^2\right) \\
&\quad + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \\
&\quad - \frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 - \varrho_1\|x_3^{k+1} - x_3^k\|^2 \\
&\quad - \varrho_1\|x_3^{k+1} - \hat{x}_3\|^2 - \|\lambda^k - \lambda^{k+1}\|^2. \qquad (3.20)
\end{aligned}$$

Since $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\lambda})$ is a saddle point, it follows from (2.7) that the left-hand side of (3.20) is nonegative. We thus have

$$\begin{aligned}
&\sum_{i=2}^{3}\|A_i x_i^{k+1} - A_i\hat{x}_i\|_{\beta^2}^2 + \|\lambda^{k+1} - \hat{\lambda}\|^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \\
&\leq \sum_{i=2}^{3}\|A_i x_i^k - A_i\hat{x}_i\|_{\beta^2}^2 + \|\lambda^k - \hat{\lambda}\|^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 \\
&\quad - \frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 - \varrho_1\|x_3^{k+1} - x_3^k\|^2 \\
&\quad - \varrho_1\|x_3^{k+1} - \hat{x}_3\|^2 - \|\lambda^k - \lambda^{k+1}\|^2. \qquad (3.21)
\end{aligned}$$

Recall the definition of Q in (2.8). Then, (3.21) can be rewritten as the compact form

$$\begin{aligned}
&\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \\
&\leq \|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 - \varrho_1\|x_3^{k+1} - x_3^k\|^2 \\
&\quad - \varrho_1\|x_3^{k+1} - \hat{x}_3\|^2 - \|\lambda^k - \lambda^{k+1}\|^2, \qquad (3.22)
\end{aligned}$$

where $\mathbf{w} := (x_2, x_3, \lambda)$. Thus we have

$$\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \leq \|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2, \qquad (3.23)$$

indicating that the sequence $\{\mathbf{w}^k\}$ is bounded. The relationship (3.2d) then further implies that $\{x_1^k\}$ is also bounded and hence the iterative sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by ADMM (1.2a)-(1.2d) is bounded. Rearranging terms for (3.22) yields

$$\begin{aligned}
&\frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \varrho_1\|x_3^{k+1} - x_3^k\|^2 + \varrho_1\|x_3^{k+1} - \hat{x}_3\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 \\
&\leq \left(\|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2\right) - \left(\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2\right).
\end{aligned}$$

Adding both sides of the above inequality for all k we get

$$\sum_{k=1}^{\infty}\left\{\frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \varrho_1\|x_3^{k+1} - x_3^k\|^2 + \varrho_1\|x_3^{k+1} - \hat{x}_3\|^2 + \|\lambda^k - \lambda^{k+1}\|^2\right\}$$
$$\leq \left(\|\mathbf{w}^1 - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^1 - A_3x_3^0\|_{\beta^2}^2\right),$$

and hence

$$\lim_{k\to\infty}\|\lambda^k - \lambda^{k+1}\| = \lim_{k\to\infty}\|A_2x_2^k - A_2x_2^{k+1}\| = \lim_{k\to\infty}\|x_3^k - x_3^{k+1}\| = \lim_{k\to\infty}\|x_3^{k+1} - \hat{x}_3\| = 0. \quad (3.24)$$

The boundedness indicates that there has at least one cluster point. Let $\tilde{\mathbf{w}} = (\tilde{x_1}, \tilde{x_2}, \tilde{x_3}, \tilde{\lambda})$ be an arbitrary cluster point of $\{\mathbf{w}^k\}$ and let $\{\mathbf{w}^{k_j}\}$ be the subsequence converging to $\tilde{\mathbf{w}}$. Taking limit in (3.9), (3.7) and (3.6), it follows from (3.24) that

$$\begin{cases} (x_1' - \tilde{x}_1)^T(\tilde{\xi}_1 - A_1^T\tilde{\lambda}) \geq 0, \ \forall x_1' \in \mathcal{X}_1, \\ (x_2' - \tilde{x}_2)^T(\tilde{\xi}_2 - A_2^T\tilde{\lambda}) \geq 0, \ \forall x_2' \in \mathcal{X}_2, \\ (x_3' - \tilde{x}_3)^T(\tilde{\xi}_3 - A_3^T\tilde{\lambda}) \geq 0, \ \forall x_3' \in \mathcal{X}_3. \end{cases}$$

For $\|\lambda^k - \lambda^{k+1}\| = \beta\|\sum_{j=1}^{3}(A_jx_j^{k+1} - b)\|$, by taking limit, we also have $\sum_{j=1}^{3}(A_j\tilde{x}_j - b) = 0$. Hence, it follows from (2.5a)-(2.5d) that $(\tilde{x_1}, \tilde{x_2}, \tilde{x_3}, \tilde{\lambda})$ is an saddle point in S and (3.23) means that the whole sequence has just one cluster point. Consequently, the iterative sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by the direct extension of ADMM (1.2a)-(1.2d) converges to a saddle point in S. The proof is complete. □

Remark 3.1. *For a specific application of the abstract model (1.1), it may not be difficult to determine the range for β. For example, for the stable principal component pursuit application in [34] (see also [33]), it is easy to see that its model corresponds to (1.1) with $A_3 = I$ and $\mu_3 = \frac{1}{\mu}$, where μ is a given constant in its model. Thus, to implement the scheme* (1.2a)-(1.2d) *for the model in [34], $\beta \in \left(0, \frac{6}{13\mu}\right)$ can sufficiently ensure the convergence.*

4 Ergodic Worst-case Convergence Rate

In this section, under Assumption 3.1, we establish the $O(1/t)$ worst-case convergence rate measured by the iteration complexity in the ergodic sense for the direct extension of ADMM (1.2a)-(1.2d). The main result is summarized in the following theorem.

Theorem 4.1. *Suppose Assumption 3.1 holds. After t iterations of the direct extension of ADMM* (1.2a)-(1.2d)*, we can find an approximate solution of* (1.1) *with an accuracy of $O(1/t)$.*

Proof. As proved in Theorem 3.1, the sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by the direct extension of ADMM (1.2a)-(1.2d) is convergent. Thus, there exists a constant $M > 0$ such that

$$\|A_ix_i^k\| \leq M, \ i = 1, 2, 3, \ \text{and} \ \|\lambda^k\| \leq M, \ \forall k \geq 0.$$

Then, it follows from Assumption 3.1 and (3.13) that for all $(x_1', x_2', x_3', \lambda') \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{R}^l$, we have

$$
\begin{aligned}
&2\beta \left[\sum_{i=1}^{3} \theta_i(x_i^{k+1}) - \lambda'^T \left(\sum_{i=1}^{3} A_i x_i^{k+1} - b \right) \right] - 2\beta \left[\sum_{i=1}^{3} \theta_i(x_i') - (\lambda^{k+1})^T \left(\sum_{i=1}^{3} A_i x_i' - b \right) \right] \\
&\le \sum_{i=2}^{3} \left(\|A_i x_i^k - A_i x_i'\|_{\beta^2}^2 - \|A_i x_i^{k+1} - A_i x_i'\|_{\beta}^2 \right) + (\|\lambda^k - \lambda'\|^2 - \|\lambda^{k+1} - \lambda'\|^2) \\
&\quad + \frac{8}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \frac{8}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\
&\quad + 2\beta^2 \left(\sum_{i=1}^{3} A_i x_i' - b \right)^T \left\{ A_2 x_2^k - A_2 x_2^{k+1} + A_3 x_3^k - A_3 x_3^{k+1} \right\}.
\end{aligned}
$$

Summing both sides of the above inequality from $k = 1$ to t yields

$$
\begin{aligned}
&2\beta \sum_{k=1}^{t} \left\{ \left[\sum_{i=1}^{3} \theta_i(x_i^{k+1}) - \lambda'^T \left(\sum_{i=1}^{3} A_i x_i^{k+1} - b \right) \right] - \left[\sum_{i=1}^{3} \theta_i(x_i') - (\lambda^{k+1})^T \left(\sum_{i=1}^{3} A_i x_i' - b \right) \right] \right\} \\
&\le \sum_{i=2}^{3} \|A_i x_i^1 - A_i x_i'\|_{\beta^2}^2 + \|\lambda^1 - \lambda'\|^2 + \frac{8}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 \\
&\quad + \beta^2 \left(\sum_{i=1}^{3} A_i x_i' - b \right)^T \left\{ A_2 x_2^0 - A_2 x_2^{t+1} + A_3 x_3^0 - A_3 x_3^{t+1} \right\} \\
&\le \sum_{i=2}^{3} \|A_i x_i^1 - A_i x_i'\|_{\beta^2}^2 + \|\lambda^1 - \lambda'\|^2 + \frac{8}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 \\
&\quad + 2\beta^2 \left\| \sum_{i=1}^{3} A_i x_i' - b \right\| \|A_2 x_2^0 - A_2 x_2^{t+1} + A_3 x_3^0 - A_3 x_3^{t+1}\| \\
&\le \sum_{i=2}^{3} \|A_i x_i^1 - A_i x_i'\|_{\beta^2}^2 + \|\lambda^1 - \lambda'\|^2 + \frac{8}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 + 8\beta^2 M \left\| \sum_{i=1}^{3} A_i x_i' - b \right\|, \qquad (4.1)
\end{aligned}
$$

where the second inequality follows from the Cauchy-Schwarz inequality and the last one is because of the triangle inequality and the boundedness of $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$. Define

$$
x_1^K := \frac{1}{t} \sum_{k=1}^{t} x_1^k, \; x_2^K := \frac{1}{t} \sum_{k=1}^{t} x_2^k, \; x_3^K := \frac{1}{t} \sum_{k=1}^{t} x_3^k, \text{ and } \lambda^K := \frac{1}{t} \sum_{k=1}^{t} \lambda^k.
$$

Then, it follows from the convexity of θ_1, θ_2 and θ_3 that

$$
\theta_1(x_1^K) \le \frac{1}{t} \sum_{k=1}^{t} \theta_1(x_1^k), \; \theta_2(x_2^K) \le \frac{1}{t} \sum_{k=1}^{t} \theta_2(x_2^k) \text{ and } \theta_3(x_3^K) \le \frac{1}{t} \sum_{k=1}^{t} \theta_2(x_3^k). \qquad (4.2)
$$

Combining (4.1) and (4.2) yields

$$
\begin{aligned}
&2\beta t \left\{ \left[\sum_{i=1}^{3} \theta_i(x_i^K) - \lambda'^T \left(\sum_{i=1}^{3} A_i x_i^K - b \right) \right] - \left[\sum_{i=1}^{3} \theta_i(x_i') - (\lambda^K)^T \left(\sum_{i=1}^{3} A_i x_i' - b \right) \right] \right\} \\
&\le \sum_{i=2}^{3} \|A_i x_i^1 - A_i x_i'\|_{\beta^2}^2 + \|\lambda^1 - \lambda'\|^2 + \frac{8}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 + 8\beta^2 M \left\| \sum_{i=1}^{3} A_i x_i' - b \right\|. \quad (4.3)
\end{aligned}
$$

Since $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ converges to a saddle point, $\{(x_1^t, x_2^t, x_3^t, \lambda^t)\}$ also converges to the same saddle point, and (4.3) implies that

$$\mathcal{G}_{\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_\lambda}(x_1^K, x_2^K, x_3^K, \lambda^K) \leq D,$$

where

$$D := \sup_{(\mathbf{x}', \lambda') \in \mathcal{B}_{\mathbf{x}} \times \mathcal{B}_\lambda} \left\{ \sum_{i=2}^{3} \|A_i x_i^1 - A_i x_i'\|_{\beta^2}^2 + \|\lambda^1 - \lambda'\|^2 \right. \\ \left. + \frac{8}{3}\|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 + 8\beta^2 M \left\| \sum_{i=1}^{3} A_i x_i' - b \right\| \right\}.$$

That is, $(\mathbf{x}^K, \lambda^K)$ is an approximate solution of (1.1) with an accuracy of $O(1/t)$. A worst-case $O(1/t)$ convergence rate measured by the iteration complexity in the ergodic sense is thus established for the direct extension of ADMM (1.2a)-(1.2d). This completes the proof. □

Remark 4.1. *The same convergence rate is claimed in [23] when both θ_2 and θ_3 are strongly convex; but it seems that a new proof is needed to obtain the result in [23].*

5 Non-ergodic Worst-case Convergence Rate

In this section, we establish the worst-case convergence rate measured by the iteration complexity in a non-ergodic sense for the direct extension of ADMM (1.2a)-(1.2d). Note that a non-ergodic worst convergence rate is generally stronger than its ergodic counterpart. Thus, we need to make the following additional assumption to derive a non-ergodic worst-case convergence rate.

Assumption 5.1. *In (1.1), the functions θ_1 and θ_2 are convex, and θ_3 is strongly convex with modulus $\mu_3 > 0$; $\partial\theta_3 + N_{\mathcal{X}_3}$ is Lipschitz continuous with constant L_3 and $\mu_3 > \frac{\sqrt{L_3}}{2\sqrt{\lambda_{min}(A_3 A_3^T)}}$; A_2 is full column rank matrices and A_3 is nonsingular.*

In Assumption 5.1, note that when $\mathcal{X}_3 = \mathcal{R}^{n_3}$, $\partial\theta_3 + N_{\mathcal{X}_3}$ reduces to $\partial\theta_3$ and "$\partial\theta_3 + N_{\mathcal{X}_3}$ is Lipschitz continuous" is equivalent to "$\partial\theta_3$ is Lipschitz continuous". For applications satisfying Assumption 5.1, see, e.g., [34].

First of all, we need to present the optimality measurement for an iterate generated by (1.2a)-(1.2d). Recall the optimality condition for the subproblems (1.2a)-(1.2c). We know that an iterate generated by (1.2a)-(1.2d) satisfies the following inequalities:

$$\begin{cases} (x_1' - x_1^{k+1})^T \left(\xi_1^{k+1} - A_1^T \lambda^{k+1}\right) \\ \quad + \beta A_1^T \left\{(A_2 x_2^k - A_2 x_2^{k+1}) + (A_3 x_3^k - A_3 x_3^{k+1})\right\} \geq 0, \ \ \forall x_1' \in \mathcal{X}_1, & (5.1a) \\ (x_2' - x_2^{k+1})^T \left(\xi_2^{k+1} - A_2^T \lambda^{k+1}\right) + \beta A_2^T (A_3 x_3^k - A_3 x_3^{k+1}) \geq 0, ; \ \forall x_2' \in \mathcal{X}_2, & (5.1b) \\ (x_3' - x_3^{k+1})^T \left(\xi_3^{k+1} - A_3^T \lambda^{k+1}\right) \geq 0, \ \ \forall x_3' \in \mathcal{X}_3, & (5.1c) \\ \lambda^{k+1} = \lambda^k - \beta \left(\sum_{j=1}^{3} A_i x_i^{k+1} - b \right). & (5.1d) \end{cases}$$

Therefore, if we have

$$\begin{cases} A_2x_2^k - A_2x_2^{k+1} = 0, & (5.2a)\\ A_3x_3^k - A_3x_3^{k+1} = 0, & (5.2b)\\ \lambda^k - \lambda^{k+1} = \sum_{j=1}^{3} A_i x_i^{k+1} - b = 0, & (5.2c) \end{cases}$$

then the iterate $(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ generated by (1.2a)-(1.2d) is an saddle point in S. Naturally, we can use the residual of (5.2a)-(5.2c) to measure the optimality of an iterate generated by (1.2a)-(1.2d).

Let us first prove one inequality in the following lemma, which will play an important role in our further analysis.

Lemma 5.1. *Suppose Assumption 5.1 holds. Let $\{(x_1^{k+1}, x_2^{k+2}, x_3^{k+3}, \lambda^{k+1})\}$ be the sequence generated by the direct extension of ADMM (1.2a)-(1.2d) with $\beta \in (0, 2\mu_3 - \frac{\sqrt{L_3}}{\sqrt{\lambda_{min}(A_3A_3^T)}})$. Then we have*

$$\begin{aligned} &\frac{1}{2\beta}\|\lambda^{k-1} - \lambda^k\|^2 + \frac{\beta}{2}\left\{\|A_3x_3^{k-1} - A_3x_3^k\|^2 + \|A_2x_2^{k-1} - A_2x_2^k\|^2\right\} \\ &\geq \frac{1}{2\beta}\|\lambda^k - \lambda^{k+1}\|^2 + \frac{\beta}{2}\left\{\|A_3x_3^k - A_3x_3^{k+1}\|^2 + \|A_2x_2^k - A_2x_2^{k+1}\|^2\right\}. \end{aligned} \tag{5.3}$$

Proof. Setting $x_1' := x_1^k$ in (5.1a) yields

$$(x_1^k - x_1^{k+1})^T \left(\xi_1^{k+1} - A_1^T\lambda^{k+1} + \beta A_1^T \left(\sum_{j=2}^{3} A_j(x_j^k - x_j^{k+1})\right)\right) \geq 0.$$

Rearranging terms yields

$$\begin{aligned} (x_1^k - x_1^{k+1})^T\xi_1^{k+1} \;\; \geq \;\; & (A_1x_1^k - A_1x_1^{k+1})^T\lambda^{k+1} \\ & -\beta(A_1x_1^k - A_1x_1^{k+1})^T \left(\sum_{j=2}^{3} A_j(x_j^k - x_j^{k+1})\right). \end{aligned}$$

Moreover, setting $x_1' := x_1^{k+1}$ for the kth iteration, we have

$$\begin{aligned} (x_1^{k+1} - x_1^k)^T\xi_1^k \;\; \geq \;\; & (A_1x_1^{k+1} - A_1x_1^k)^T\lambda^k \\ & -\beta(A_1x_1^{k+1} - A_1x_1^k)^T \left(\sum_{j=2}^{3} A_j(x_j^{k-1} - x_j^k)\right). \end{aligned}$$

Adding the above two inequalities

$$\begin{aligned} &(A_1x_1^{k+1} - A_1x_1^k)^T(\lambda^{k+1} - \lambda^k) \\ &\geq -\beta(A_1x_1^k - A_1x_1^{k+1})^T \left(\sum_{j=2}^{3} A_j(x_j^k - x_j^{k+1})\right) \\ &\quad -\beta(A_1x_1^{k+1} - A_1x_1^k)^T \left(\sum_{j=2}^{3} A_j(x_j^{k-1} - x_j^k)\right). \end{aligned} \tag{5.4}$$

Similarly, in (5.1b), setting $x_2' = x_2^k$ and setting $x_2' = x_2^{k+1}$ in the kth iteration, we have, after adding the resulting inequalities and rearranging terms,

$$
\begin{aligned}
&(A_2x_2^{k+1} - A_2x_2^k)^T(\lambda^{k+1} - \lambda^k) \\
\geq\ & -\beta(A_2x_2^k - A_2x_2^{k+1})^T(A_3(x_3^k - x_3^{k+1}) - \beta(A_2x_2^{k+1} - A_2x_2^k)^T(A_3(x_3^{k-1} - x_3^k)). \quad (5.5)
\end{aligned}
$$

Then, from (5.1c) we obtain

$$
(A_3x_3^{k+1} - A_3x_3^k)^T(\lambda^{k+1} - \lambda^k) \geq \mu_3\|x^{k+1} - x^k\|^2. \quad (5.6)
$$

Adding (5.4)-(5.6), we have

$$
\begin{aligned}
&\left(\sum_{j=1}^3 A_jx_j^{k+1} - \sum_{j=1}^3 A_jx_j^k\right)^T (\lambda^{k+1} - \lambda^k) \\
&\geq -\beta(A_1x_1^k - A_1x_1^{k+1})^T \left(\sum_{j=2}^3 A_j(x_j^k - x_j^{k+1})\right) - \beta(A_1x_1^{k+1} - A_1x_1^k)^T \left(\sum_{j=2}^3 A_j(x_j^{k-1} - x_j^k)\right) \\
&\quad -\beta(A_2x_2^k - A_2x_2^{k+1})^T(A_3(x_3^k - x_3^{k+1})) - \beta(A_2x_2^{k+1} - A_2x_2^k)^T(A_3(x_3^{k-1} - x_3^k)). \quad (5.7)
\end{aligned}
$$

Notice that (5.1d) is equivalent to

$$
\sum_{j=1}^3 A_jx_j^{k+1} - \sum_{j=1}^3 A_jx_j^k = \frac{1}{\beta}\Big((\lambda^k - \lambda^{k+1}) - (\lambda^{k-1} - \lambda^k)\Big).
$$

Then the left-hand term of (5.7) can be rewritten into

$$
\begin{aligned}
&\left(\sum_{j=1}^3 A_jx_j^{k+1} - \sum_{j=1}^3 A_jx_j^k\right)^T (\lambda^{k+1} - \lambda^k) \\
&= \frac{1}{\beta}\Big((\lambda^k - \lambda^{k+1}) - (\lambda^{k-1} - \lambda^k)\Big)^T (\lambda^{k+1} - \lambda^k) \\
&= \frac{1}{2\beta}\Big\{\|\lambda^{k-1} - \lambda^k\|^2 - \|\lambda^k - \lambda^{k+1}\|^2 - \|(\lambda^k - \lambda^{k+1}) - (\lambda^{k-1} - \lambda^k)\|^2\Big\}. \quad (5.8)
\end{aligned}
$$

We then treat the right-hand side of (5.7). These terms in the right-hand side of (5.7) can be written as

$$
\begin{aligned}
&-\beta(A_1x_1^k - A_1x_1^{k+1})^T \sum_{j=2}^3(A_jx_j^k - A_jx_j^{k+1}) - \beta(A_1x_1^{k+1} - A_1x_1^k)^T \sum_{j=2}^3(A_jx_j^{k-1} - A_jx_j^k) \\
&-\beta(A_2x_2^k - A_2x_2^{k+1})^T(A_3x_3^k - A_3x_3^{k+1}) - \beta(A_2x_2^{k+1} - A_2x_2^k)^T(A_3x_3^{k-1} - A_3x_3^k) \\
&= -\beta(A_3x_3^k - A_3x_3^{k+1})^T \sum_{j=1}^2(A_jx_j^k - A_jx_j^{k+1}) - \beta(A_1x_1^k - A_1x_1^{k+1})^T(A_2x_2^k - A_2x_2^{k+1}) \\
&-\beta(A_3x_3^{k-1} - A_3x_3^k)^T \sum_{j=1}^2(A_jx_j^{k+1} - A_jx_j^k) - \beta(A_1x_1^{k+1} - A_1x_1^k)^T(A_2x_2^{k-1} - A_2x_2^k). \quad (5.9)
\end{aligned}
$$

Using the identity (3.14), we have

$$-\beta(A_3x_3^k - A_3x_3^{k+1})^T\sum_{j=1}^{2}(A_jx_j^k - A_jx_j^{k+1})$$
$$= \frac{\beta}{2}\left\{\|A_3x_3^k - A_3x_3^{k+1}\|^2 + \left\|\sum_{j=1}^{2}A_j(x_j^k - x_j^{k+1})\right\|^2 - \left\|\sum_{j=1}^{3}A_j(x_j^k - x_j^{k+1})\right\|^2\right\} \quad (5.10)$$

and

$$-\beta(A_1x_1^k - A_1x_1^{k+1})^T(A_2x_2^k - A_2x_2^{k+1})$$
$$= \frac{\beta}{2}\left\{\|A_1x_1^k - A_1x_1^{k+1}\|^2 + \|A_2x_2^k - A_2x_2^{k+1}\|^2 - \left\|\sum_{j=1}^{2}A_j(x_j^k - x_j^{k+1})\right\|^2\right\}. \quad (5.11)$$

Then, using the Cauchy-Schwarz inequality, for all $\omega > 0$, we have

$$\begin{aligned}
&-\beta(A_3x_3^{k-1} - A_3x_3^k)^T\sum_{j=1}^{2}(A_jx_j^{k+1} - A_jx_j^k)\\
&= -\beta(A_3x_3^{k-1} - A_3x_3^k)^T\left\{\sum_{j=1}^{3}(A_jx_j^{k+1} - A_jx_j^k) - (A_3x_3^{k+1} - A_3x_3^k)\right\}\\
&= -(A_3x_3^{k-1} - A_3x_3^k)^T\left\{(\lambda^k - \lambda^{k+1}) - (\lambda^{k-1} - \lambda^k)\right\}\\
&\quad +\beta(A_3x_3^{k-1} - A_3x_3^k)^T(A_3x_3^{k+1} - A_3x_3^k)\\
&\ge -\frac{\omega}{2}\|A_3x_3^{k-1} - A_3x_3^k\|^2 - \frac{1}{2\omega}\|\lambda^k - \lambda^{k+1}\|^2 + \mu_3\|A_3x_3^{k-1} - A_3x_3^k\|\\
&\quad -\frac{\beta}{2}\|A_3x_3^{k-1} - A_3x_3^k\|^2 - \frac{\beta}{2}\|A_3x_3^k - A_3x_3^{k+1}\|^2, \qquad (5.12)
\end{aligned}$$

and

$$-\beta(A_1x_1^{k+1} - A_1x_1^k)^T(A_2x_2^{k-1} - A_2x_2^k) \ge -\frac{\beta}{2}\left\{\|A_1x_1^{k+1} - A_1x_1^k\|^2 + \|A_2x_2^{k-1} - A_2x_2^k\|^2\right\}. (5.13)$$

Substituting (5.10)-(5.13) into (5.9), we have

$$\begin{aligned}
&-\beta(A_1x_1^k - A_1x_1^{k+1})^T\sum_{j=2}^{3}(A_jx_j^k - A_jx_j^{k+1}) - \beta(A_1x_1^{k+1} - A_1x_1^k)^T\sum_{j=2}^{3}(A_jx_j^{k-1} - A_jx_j^k)\\
&-\beta(A_2x_2^k - A_2x_2^{k+1})^T(A_3x_3^k - A_3x_3^{k+1}) - \beta(A_2x_2^{k+1} - A_2x_2^k)^T(A_3x_3^{k-1} - A_3x_3^k)\\
&\ge \frac{\beta}{2}\left(\|A_2x_2^k - A_2x_2^{k+1}\|^2 - \|A_2x_2^{k-1} - A_2x_2^k\|^2\right) - \frac{1}{2\omega}\|\lambda^k - \lambda^{k+1}\|^2\\
&\quad + \left(\mu_3 - \frac{\beta}{2} - \frac{\omega}{2}\right)\|A_3x_3^{k-1} - A_3x_3^k\|^2 - \frac{\beta}{2}\left\|\sum_{j=1}^{3}A_j(x_j^k - x_j^{k+1})\right\|^2. \qquad (5.14)
\end{aligned}$$

It follows from the optimality conditions for x_3^{k+1}-subproblem and the optimality condition for the x_3^k-subproblem that

$$A_3^T\lambda^{k+1} \in \partial\theta_3(x_3^{k+1}) + N_{\mathcal{X}_3}(x_3^{k+1})$$

and

$$A_3^T\lambda^k \in \partial\theta_3(x_3^k) + N_{\mathcal{X}_3}(x_3^k),$$

and Assumption 5.1 means that

$$\|A_3^T\lambda^{k+1} - A_3^T\lambda^k\| \le L_3\|x_3^{k+1} - x_3^k\|.$$

Since A_3 is nonsingular, we further get

$$\|\lambda^{k+1} - \lambda^k\|^2 \le \frac{L_3}{\lambda_{\min}(A_3A_3^T)}\|x_3^{k+1} - x_3^k\|^2.$$

Combining (5.8) and (5.14), and setting $\omega = \frac{\sqrt{L_3}}{\sqrt{\lambda_{\min}(A_3A_3^T)}}$, it follows from (5.7)

$$\begin{aligned}
&\frac{1}{2\beta}\|\lambda^{k-1} - \lambda^k\|^2 + \frac{\beta}{2}\left\{\|A_3x_3^{k-1} - A_3x_3^k\|^2 + \|A_2x_2^{k-1} - A_2x_2^k\|^2\right\} \\
&\ge \frac{1}{2\beta}\|\lambda^k - \lambda^{k+1}\|^2 + \frac{\beta}{2}\left\{\|A_3x_3^k - A_3x_3^{k+1}\|^2 + \|A_2x_2^k - A_2x_2^{k+1}\|^2\right\} \\
&\quad + \left(\mu_3 - \frac{\sqrt{L_3}}{2\sqrt{\lambda_{\min}(A_3A_3^T)}}\right)\|A_3x_3^{k-1} - A_3x_3^k\|^2 \\
&\quad + \left(\mu_3 - \frac{\beta}{2} - \frac{\sqrt{L_3}}{2\sqrt{\lambda_{\min}(A_3A_3^T)}}\right)\|A_3x_3^k - A_3x_3^{k+1}\|^2 \\
&\ge \frac{1}{2\beta}\|\lambda^k - \lambda^{k+1}\|^2 + \frac{\beta}{2}\left\{\|A_3x_3^k - A_3x_3^{k+1}\|^2 + \|A_2x_2^k - A_2x_2^{k+1}\|^2\right\},
\end{aligned} \tag{5.15}$$

where the last inequality is because of $\beta \in (0, 2\mu_3 - \frac{\sqrt{L_3}}{\sqrt{\lambda_{\min}(A_3A_3^T)}})$. The proof is complete. □

Based on the above lemma, we are now ready to establish the worst-case convergence rate measured by the iteration complexity in a non-ergodic sense. We can actually measure this convergence rate by both $O(1/k)$ and $o(1/k)$ orders. Recall that we use the residual of (5.2a)-(5.2c) to measure the accuracy of an iterate of the scheme (3.2a)-(3.2d).

Theorem 5.1. *Let $\{\mathbf{w}^k := (x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by the direct extension of ADMM (3.2a)-(3.2d) with $\beta \in (0, 2\mu_3 - \frac{\sqrt{L_3}}{\sqrt{\lambda_{\min}(A_3A_3^T)}})$. Then, we have*

$$\frac{1}{2\beta}\|\lambda^t - \lambda^{t+1}\|^2 + \frac{\beta}{2}\{\|A_3x_3^t - A_3x_3^{t+1}\|^2 + \|A_2x_2^t - A_2x_2^{t+1}\|^2\} = O(1/t).$$

Proof. From (3.22), we have

$$\begin{aligned}
&\frac{1}{52}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \varrho_1\|x_3^{k+1} - x_3^k\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 \\
&\le \|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 - \|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2.
\end{aligned}$$

Setting

$$\mu = \min\left\{\frac{\beta}{26}, \frac{2\varrho_1}{\beta\|A_3^TA_3\|}, 2\beta\right\},$$

it follows that

$$\mu\left\{\frac{1}{2\beta}\|\lambda^k-\lambda^{k+1}\|^2+\frac{\beta}{2}\left(\|A_3x_3^k-A_3x_3^{k+1}\|^2+\|A_2x_2^k-A_2x_2^{k+1}\|^2\right)\right\}$$
$$\le\frac{1}{52}\|A_2x_2^k-A_2x_2^{k+1}\|_{\beta^2}^2+\varrho_1\|x_3^{k+1}-x_3^k\|^2+\|\lambda^k-\lambda^{k+1}\|^2$$
$$\le\|\mathbf{w}^k-\hat{\mathbf{w}}\|_Q^2-\|\mathbf{w}^{k+1}-\hat{\mathbf{w}}\|_Q^2+\frac{8}{3}\|A_3x_3^k-A_3x_3^{k-1}\|_{\beta^2}^2-\frac{8}{3}\|A_3x_3^{k+1}-A_3x_3^k\|_{\beta^2}^2. \tag{5.16}$$

Adding from $k=1$ to t, we have

$$\sum_{k=1}^{t}\frac{1}{2\beta}\|\lambda^k-\lambda^{k+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^k-A_3x_3^{k+1}\|^2+\|A_2x_2^k-A_2x_2^{k+1}\|^2\right\}$$
$$\le\frac{1}{\mu}\{\|\mathbf{w}^0-\hat{\mathbf{w}}\|_Q^2+\frac{8}{3}\|A_3x_3^1-A_3x_3^0\|_{\beta^2}^2\}.$$

Using Lemma 5.1 we have

$$\frac{1}{2\beta}\|\lambda^t-\lambda^{t+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^t-A_3x_3^{t+1}\|^2+\|A_2x_2^t-A_2x_2^{t+1}\|^2\right\}$$
$$\le\frac{1}{\mu t}\{\|\mathbf{w}^0-\hat{\mathbf{w}}\|_Q^2+\frac{8}{3}\|A_3x_3^1-A_3x_3^0\|_{\beta^2}^2\},$$

which means that

$$\frac{1}{2\beta}\|\lambda^t-\lambda^{t+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^t-A_3x_3^{t+1}\|^2+\|A_2x_2^t-A_2x_2^{t+1}\|^2\right\}=O(1/t).$$

The proof is complete. □

In Theorem 5.1, we show the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the direct extension of ADMM (3.2a)-(3.2d). Indeed, following the work [9], we can easily refine this $O(1/t)$ order to an $o(1/t)$ order, and thus establish a worst-case $o(1/t)$ convergence rate in a nonergodic sense for the scheme (3.2a)-(3.2d). We summarize it in the following theorem.

Theorem 5.2. *Let $\{\mathbf{w}^k:=(x_1^k,x_2^k,x_3^k,\lambda^k)\}$ be the sequence generated by the direct extension of ADMM (3.2a)-(3.2d) with $\beta\in(0,2\mu_3-\frac{\sqrt{L_3}}{\sqrt{\lambda_{\min}(A_3A_3^T)}})$. Then, we have*

$$\frac{1}{2\beta}\|\lambda^t-\lambda^{t+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^t-A_3x_3^{t+1}\|^2+\|A_2x_2^t-A_2x_2^{t+1}\|^2\right\}=o(1/t). \tag{5.17}$$

Proof. If adding (5.16) from $k=t$ to $2t$, we have

$$t\left\{\frac{1}{2\beta}\|\lambda^{2t}-\lambda^{2t+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^{2t}-A_3x_3^{2t+1}\|^2+\|A_2x_2^{2t}-A_2x_2^{2t+1}\|^2\right\}\right\}$$
$$\le\sum_{k=t}^{2t}\frac{1}{2\beta}\|\lambda^k-\lambda^{k+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^k-A_3x_3^{k+1}\|^2+\|A_2x_2^k-A_2x_2^{k+1}\|^2\right\}$$
$$=\sum_{k=1}^{2t}\frac{1}{2\beta}\|\lambda^k-\lambda^{k+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^k-A_3x_3^{k+1}\|^2+\|A_2x_2^k-A_2x_2^{k+1}\|^2\right\}$$
$$-\sum_{k=1}^{t}\frac{1}{2\beta}\|\lambda^k-\lambda^{k+1}\|^2+\frac{\beta}{2}\left\{\|A_3x_3^k-A_3x_3^{k+1}\|^2+\|A_2x_2^k-A_2x_2^{k+1}\|^2\right\}.$$

The limit of the last term in the right-hand side is 0. So, the assertion (5.17) is proved. □

6 Globally Linear Convergence

In this section, we show that the globally linear convergence of the direct extension of ADMM (1.2a)-(1.2d) can be ensured if adequate assumptions are made. Note that the same convergence rate was analyzed in [24] under stronger conditions. In addition to Assumption 3.1, we make the following assumptions in this section:

Assumption 6.1. *In (1.1), θ_1 is convex; θ_2 and θ_3 are strongly convex with modulus $\mu_2 > 0$ and $\mu_3 > 0$, respectively; $\partial\theta_1 + N_{\mathcal{X}_1}$ is Lipschitz continuous with constant L_1; A_2 and A_3 are full column rank matrices and A_1 is nonsingular.*

First of all, under Assumption 3.1, we can prove a result similar to (3.22). That is, there exists a constant $\zeta = \min\{\frac{1}{52}, \varrho_1, 1, \mu_2\} > 0$, such that

$$\begin{aligned} &\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \\ &\le \|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \zeta\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2\right. \\ &\quad \left. + \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\}. \end{aligned} \tag{6.1}$$

Then, let us explain the roadmap to prove the linear convergence rate of (1.2a)-(1.2d), from which the reason for making Assumption 6.1 is also clear. Indeed, for this purpose, according to (6.1), it is clear that we only need to bound $\|\lambda^{k+1} - \hat{\lambda}\|^2$ and $\|A_1x_1^{k+1} - A_1\hat{x}_1\|^2$, using the minus term in (6.1). We first consider $\|A_1x_1^{k+1} - A_1\hat{x}_1\|^2$. It follows from (3.18) and (3.5) that

$$A_1x_1^{k+1} - A_1\hat{x}_1 = \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) - (A_2x_2^{k+1} - A_2\hat{x}_2) - (A_3x_3^{k+1} - A_3\hat{x}_3).$$

Therefore,

$$\|A_1x_1^{k+1} - A_1\hat{x}_1\|^2 \le 3\left\{\frac{1}{\beta^2}\|\lambda^k - \lambda^{k+1}\|^2 + \|A_2x_2^{k+1} - A_2\hat{x}_2\|^2 + \|A_3x_3^{k+1} - A_3\hat{x}_3\|^2\right\}. \tag{6.2}$$

Then, we need to consider how to bound the term $\|\lambda^{k+1} - \hat{\lambda}\|^2$. Actually, Assumption 6.1 is for this purpose, as we will elucidate below.

Lemma 6.1. *Assume Assumption 6.1 holds. Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\lambda})$ be a saddle point in S; let $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by the direct extension of ADMM (1.2a)-(1.2d). Then, there exists a constant $\sigma > 0$ such that*

$$\|\lambda^{k+1} - \hat{\lambda}\|^2 \le \sigma\left(\|x_1^{k+1} - \hat{x}_1\|^2 + \|x_2^k - x_2^{k+1}\|^2 + \|x_3^k - x_3^{k+1}\|^2\right).$$

Proof. It follows from the optimality conditions for x_1^{k+1}-subproblem and the optimality condition for $\hat{x}_1$ that

$$A_1^T\lambda^{k+1} \in \partial\theta_1(x_1^{k+1}) + N_{\mathcal{X}_1}(x_1^{k+1}) + \beta A_1^T\sum_{i=2}^{3}(A_ix_i^k - A_ix_i^{k+1})$$

and

$$A_1^T\hat{\lambda} \in \partial\theta_1(\hat{x}_1) + N_{\mathcal{X}_1}(\hat{x}_1).$$

Using Assumption 6.1 and the Cauchy-Schwarz inequality, we get

$$\|A_1^T\lambda^{k+1} - A_1^T\hat{\lambda}\| \le L_1\|x_1^{k+1} - \hat{x}_1\| + \beta\|A_1\|\sum_{i=2}^{3}\left(\|A_i\|\|x_i^k - x_i^{k+1}\|\right),$$

and the assertion follows immediately from the non-singularity of A_1. □

Theorem 6.1. *Assume Assumption 6.1 holds. Then the sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by the direct extension of ADMM (1.2a)-(1.2d) converges globally linearly to a saddle point in S.*

Proof. Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{\lambda})$ be a saddle point in S. It follows from (6.2) that

$$\begin{aligned}
&\frac{\zeta}{4}\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\}\\
&\ge \frac{\zeta}{4}\left\{\|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\}\\
&\ge \frac{\zeta}{4}\left\{\beta^2\frac{1}{\beta^2}\|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\frac{1}{\|A_i^TA_i\|}\|A_ix_i^{k+1} - A_i\hat{x}_i\|^2\right\}\\
&\ge \frac{\zeta}{12}\min\left\{\beta^2, \frac{1}{\|A_2^TA_2\|}, \frac{1}{\|A_3^TA_3\|}\right\}\|A_1x_1^{k+1} - A_1\hat{x}_1\|^2\\
&\ge \frac{\zeta}{12}\min\left\{\beta^2, \frac{1}{\|A_2^TA_2\|}, \frac{1}{\|A_3^TA_3\|}\right\}\lambda_{\min}(A_1^TA_1)\|x_1^{k+1} - \hat{x}_1\|^2\\
&= \tau_1\|x_1^{k+1} - \hat{x}_1\|^2, \qquad (6.3)
\end{aligned}$$

where $\tau_1 = \frac{\zeta}{12}\min\left\{\beta^2, \frac{1}{\|A_2^TA_2\|}, \frac{1}{\|A_3^TA_3\|}\right\}\lambda_{\min}(A_1^TA_1)$.

On the other hand, we have

$$\begin{aligned}
&\frac{\zeta}{4}\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\}\\
&\ge \frac{\zeta}{4}\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2\right\}\\
&\ge \frac{\zeta}{4}\left\{\beta^2\lambda_{\min}(A_2^TA_2)\|x_2^k - x_2^{k+1}\|^2 + \|x_3^k - x_3^{k+1}\|^2\right\}\\
&\ge \frac{\zeta}{4}\min\left\{\beta^2\lambda_{\min}(A_2^TA_2), 1\right\}\left\{\|x_2^k - x_2^{k+1}\|^2 + \|x_3^k - x_3^{k+1}\|^2\right\}\\
&= \tau_2\left\{\|x_2^k - x_2^{k+1}\|^2 + \|x_3^k - x_3^{k+1}\|^2\right\}, \qquad (6.4)
\end{aligned}$$

where $\tau_2 = \frac{\zeta}{4}\min\left\{\beta^2\lambda_{\min}(A_2^TA_2), 1\right\}$.

Adding (6.3) and (6.4) and using Lemma 6.1, we have

$$\begin{aligned}
&\frac{\zeta}{2}\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\}\\
&\ge \tau_1\|x_1^{k+1} - \hat{x}_1\|^2 + \tau_2\left\{\|x_2^k - x_2^{k+1}\|^2 + \|x_3^k - x_3^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2\right\}\\
&\ge \min\{\tau_1, \tau_2\}\left\{\|x_1^{k+1} - \hat{x}_1\|^2 + \|x_2^k - x_2^{k+1}\|^2 + \|x_3^k - x_3^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2\right\}\\
&\ge \frac{\min\{\tau_1, \tau_2\}}{\sigma}\|\lambda^{k+1} - \hat{\lambda}\|^2\\
&= \tau\|\lambda^{k+1} - \hat{\lambda}\|^2, \qquad (6.5)
\end{aligned}$$

where $\tau = \frac{1}{\sigma}\min\{\tau_1, \tau_2\}$.

Moreover, we have

$$\begin{aligned}
&\frac{\zeta}{2}\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\} \\
&\geq \frac{\zeta}{2}\left\{\|x_3^k - x_3^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\} \\
&\geq \frac{\zeta}{2}\left\{\frac{3}{8\beta^2}\frac{8}{3}\|x_3^k - x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^{3}\frac{1}{\beta^2\|A_i^TA_i\|}\|A_ix_i^{k+1} - A_i\hat{x}_i\|_{\beta^2}^2\right\} \\
&\geq \frac{\zeta}{2}\min\left\{\frac{3}{8\beta^2\|A_3^TA_3\|}, \frac{1}{\beta^2\|A_2^TA_2\|}\right\}\left\{\frac{8}{3}\|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^{3}\|A_ix_i^{k+1} - A_i\hat{x}_i\|_{\beta^2}^2\right\} \\
&= \tau'\left\{\frac{8}{3}\|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^{3}\|A_ix_i^{k+1} - A_i\hat{x}_i\|_{\beta^2}^2\right\}, \qquad (6.6)
\end{aligned}$$

where $\tau' = \frac{\zeta}{2\beta^2}\min\left\{\frac{3}{8\|A_3^TA_3\|}, \frac{1}{\|A_2^TA_2\|}\right\}$.

It follows from (6.5) and (6.6) that

$$\begin{aligned}
&\zeta\left\{\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=2}^{3}\|x_i^{k+1} - \hat{x}_i\|^2\right\} \\
&\geq \tau\|\lambda^{k+1} - \hat{\lambda}\|^2 + \tau'\left\{\frac{8}{3}\|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^{3}\|A_ix_i^{k+1} - A_i\hat{x}_i\|_{\beta^2}^2\right\} \\
&\geq \min\{\tau, \tau'\}\left\{\|\lambda^{k+1} - \hat{\lambda}\|^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^{3}\|A_ix_i^{k+1} - A_i\hat{x}_i\|_{\beta^2}^2\right\} \\
&= \sigma'\left\{\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2\right\}, \qquad (6.7)
\end{aligned}$$

where Q is defined in (2.8) and its positive definiteness is ensured by Assumption 6.1 and $\sigma' = \min\{\tau, \tau'\}$.

Then, combining (6.1) and (6.7), we obtain

$$\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \leq \frac{1}{1+\sigma'}\left\{\|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{8}{3}\|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2\right\},$$

which implies the linear convergence rate of the sequence generated by the direct extension of ADMM (1.2a)-(1.2d) under the Q-norm. The proof is complete. □

Remark 6.1. *As just proved, Assumption 6.1 is sufficient to ensure the globally linear convergence of the direct extension of ADMM (1.2a)-(1.2d). Indeed, the same convergence rate could be ensured under other assumptions. For example, we can alter Assumption 6.1 as follows:*
"In (1.1), θ_1 is convex; θ_2 and θ_3 are strongly convex with modulus $\mu_2 > 0$ and $\mu_3 > 0$, respectively; and for $i = 2$ (Resp., $i = 3$), $\partial\theta_2 + N_{\mathcal{X}_2}$ (Resp. $\partial\theta_3 + N_{\mathcal{X}_3}$) is Lipschitz continuous with constant $L_2 > 0$ (Resp., $L_3 > 0$) and A_2 (Resp., A_3) is nonsingular, A_3 (Resp., A_2) is full column rank. "
Then, the globally linear convergence of the direct extension of ADMM (1.2a)-(1.2d) can be established analogously.

Remark 6.2. *Note that the direct extension of ADMM* (1.2a)-(1.2d) *is a splitting, thus inexact, version of the augmented Lagrangian method in [20, 30], which has been proved in [32] to be an application of the proximal point algorithm (PPA) in [25, 26]. Since the convergence rate of PPA is known to be only linear in [32], the linear convergence rate is the most we can expect for the direct extension of ADMM* (1.2a)-(1.2d)*. The linear convergence rate is indeed ideal for the generic convex model (1.1), and thus very strict conditions should be posed to gain this convergence rate. In fact, even for the original ADMM to tackle a two-block linearly constrained separable convex minimization model, its linear convergence rate can only be established for special cases (e.g., [1, 15]) or generic cases but with strong assumptions on the model (e.g., [7, 8]). Usually, the conditions that can ensure the linear convergence rate of* (1.2a)-(1.2d) *are too strict to be satisfied by the majority of the concrete applications of the abstract model (1.1). In this sense, discussing the global linear convergence rate for the direct extension of ADMM* (1.2a)-(1.2d) *only makes a theoretical sense.*

7 Conclusions

In this paper, we discuss the convergence of the direct extension of alternating direction method of multipliers (ADMM) for solving a three-block linearly constrained convex minimization model whose objective function is the sum of three functions without coupled variables. Our main result is that the direct extension of ADMM with a chosen penalty parameter is convergent when one function in the objective is strongly convex, together with some minor assumptions on the coefficient matrices in the constraints. Our condition differs from some existing conditions which require the strong convexity for two or three functions in its objective, and this condition can be easily satisfied by many concrete applications. Thus, we close the gap between the empirical efficiency of the direct extension of ADMM which has been long observed for some applications and the lack of reasonable theoretical justification. Our result also differs from some proved convergence results in the literature for various twisted schemes, because we work on the original scheme of the direct extension of ADMM. We also establish the worst-case convergence rate measured by the iteration complexity and the linear convergence rate in asymptotical sense under additional assumptions on the model under consideration. This deeper study on the convergence rate helps us better understand the theoretical aspects of the direct extension of ADMM, and thus sheds lights onto the convergence analysis study for its more-than-three-block counterparts whose objective function has at least four functions, which we believe is still open.

Last, we would like to mention that instead of the original scheme (1.2a)-(1.2d), we can also study the scheme

$$\begin{cases} x_1^{k+1} = \arg\min\left\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \lambda^k) \mid x_1 \in \mathcal{X}_1\right\}, & (7.1a)\\ x_2^{k+1} = \arg\min\left\{\mathcal{L}_\beta(x_1^{k+1}, x_2, x_3^k, \lambda^k) \mid x_2 \in \mathcal{X}_2\right\}, & (7.1b)\\ x_3^{k+1} = \arg\min\left\{\mathcal{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3, \lambda^k) \mid x_3 \in \mathcal{X}_3\right\}, & (7.1c)\\ \lambda^{k+1} = \lambda^k - \gamma\beta(\sum_{i=1}^{3} A_i x_i^{k+1} - b), & (7.1d) \end{cases}$$

where $\gamma \in (0, 1)$. Since the relaxation factor γ can be arbitrarily close to 1, asymptotically this scheme has no difference from the direct extension of ADMM (1.2a)-(1.2d). Similarly, we can discuss the convergence and estimate the convergence rate under different conditions for the scheme (7.1a)-(7.1d), and indeed, their proofs can be presented with easier notation than what we present in this paper.

References

[1] D. Boley, *Local linear convergence of ADMM on quadratic or linear programs*, SIAM J. Optim, 23(4) (2013), pp. 2183-2207.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Foundations and Trends in Machine Learning, Michael Jordan, Editor in Chief, 3 (2011), pp. 1–122.

[3] A. Chamboulle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120-145.

[4] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, *Latent variable graphical model selection via convex optimization*, Ann. Stat., 40 (2012), pp. 1935-1967.

[5] C. H. Chen, B. S. He, Y. Y. Ye, and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, Math. Program., Ser. A, DOI 10.1007/s10107-014-0826-5, 2014.

[6] C. H. Chen, Y. Shen, and Y. F. You, *On the convergence analysis of the alternating direction method of multipliers with three blocks, Abstract and Applied Analysis*, 2013(2013), Article ID 183961, 7 pages.

[7] E. Corman and X. M. Yuan, *A generalized proximal point algorithm and its convergence rate*, SIAM J. Optim., 24 (4) (2014), pp. 1614-1638.

[8] W. Deng and W. T. Yin, *On the global and linear convergence of the generalized alternating direction method of multipliers*, manuscript, 2012.

[9] W. Deng, M. J. Lai, Z. M. Peng, and W. T. Yin, *Parallel multi-block ADMM with $o(1/k)$ convergence*, manuscript, 2014.

[10] J. Eckstein and W. T. Yao, *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, RUTCOR Research Report, (2012).

[11] R. Glowinski, Numerical methods for nonlinear variational problems, Springer, 1984.

[12] R. Glowinski and A. Marrocco, *Sur l'approximation parélémENTS finis d'ordre un et larésolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, Revue Fr. Autom. Inform. Rech. Opér., Anal. Numér. 2 (1975), pp. 41–76.

[13] R. Glowinski, *On alternating directon methods of multipliers: a historical perspective*, Modeling, Simulation and Optimization for Science and Technology Computational Methods in Applied Sciences 34 (2014), pp. 59-82.

[14] D. R. Han and X. M. Yuan, *A note on the alternating direction method of multipliers*, J. Optim. Theory Appli., 155 (2013), pp. 227-238.

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

[16] D. R. Han, X. M. Yuan, and W. X. Zhang, *An augmented-Lagrangian-based parallel splitting method for separable convex programming with applications to image processing*, Math. Comput., 83 (2014), pp. 2263-2291.

[17] B. S. He, M. Tao, and X. M. Yuan, *Alternating direction method with Gaussian back substitution for separable convex programming*, SIAM J. Optim., 22 (2012), pp. 313-340.

[18] B. S. He, M. Tao, and X. M. Yuan, *Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming*, Math. Oper. Res., under revision.

[19] B. S. He, M. Tao, and X. M. Yuan, *A splitting method for separable convex programming*, IMA J. Numer. Anal., to appear.

[20] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4(1969), pp. 303-320.

[21] M. Hong and Z. Q. Luo, *On the linear convergence of alternating direction method of multipliers*, manuscript, 2012.

[22] M. Li, D. F. Sun, and K. C. Toh, *A convergent* 3*-block semi-proximal ADMM for convex minimization problems with one strongly convex block*, manuscript, 2014.

[23] T.Y. Lin, S.Q. Ma, and S.Z. Zhang, *On the convergence rate of multi-block ADMM*, manuscript, 2014.

[24] T.Y. Lin, S.Q. Ma, and S.Z. Zhang, *On the global linear convergence of the ADMM with multi-block variables*, manuscript, 2014.

[25] B. Martinet, *Regularization d'inequations variationelles par approximations successives*, Revue Francaise d'Informatique et de Recherche Opérationelle, 4 (1970), pp. 154-159.

[26] J. J. Moreau, *Proximité et dualit 'e dans un espace Hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273-299.

[27] G. J. McLachlan, Discriminant analysis and statistical pattern recognition, volume 544. Wiley-Interscience, 2004.

[28] K. Mohan, P. London, M. Fazel, D. Witten, and S. Lee, *Node-based learning of multiple gaussian graphical models*, avialable at: arXiv:/1303.5145, 2013.

[29] Y. G. Peng, A. Ganesh, J. Wright, W. L. Xu, and Y. Ma, *Robust alignment by sparse and low-rank decomposition for linearly correlated images*, IEEE Tran. Pattern Anal. Mach. Intel. 34(2012), pp. 2233-2246.

[30] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, 1969.

[31] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

[32] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Con. Optim., 14 (1976), pp. 877-898.

[33] M. Tao and X. M. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57-81.

[34] Z. Zhou, X. Li, J. Wright, E. J. Candes, and Y. Ma, *Stable principal component pursuit*, Proceedings of International Symposium on Information Theory, (2010).