

On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function

Xingju Cai¹ · Deren Han¹ · Xiaoming Yuan²

Received: 14 February 2016 / Published online: 27 July 2016
© Springer Science+Business Media New York 2016

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. We focus on the three-block case of such a model whose objective function is the sum of three functions, and discuss the convergence of the direct extension of ADMM. We show that when one function in the objective is strongly convex, the penalty parameter and the operators in the linear equality constraint are appropriately restricted, it is sufficient to guarantee the convergence of the direct extension of ADMM. We further estimate the worst-case convergence rate measured by the iteration complexity in both the ergodic and nonergodic senses, and derive the globally linear convergence in asymptotical sense under some additional conditions.

✉ Xiaoming Yuan
xmyuan@hkbu.edu.hk

Xingju Cai
caixingju@nynu.edu.cn

Deren Han
handeren@nynu.edu.cn

¹ School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, People's Republic of China

² Department of Mathematics, Hong Kong Baptist University, Hong Kong, People's Republic of China

Keywords 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 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 [34], the latent variable Gaussian graphical model selection in [4, 28], the stable principal component pursuit with nonnegative constraint in [38], the quadratic discriminant analysis model in [27] and the quadratic conic programming in [22].

Inspired by the mentioned applications, we focus on a three-block linearly constrained separable convex minimization model as follows. Let \mathcal{X}_i with $i = 1, 2, 3$ and \mathcal{Z}^* be real finite-dimensional Euclidean spaces, each equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$; $\theta_i : \mathcal{X}_i \rightarrow (-\infty, +\infty]$ with $i = 1, 2, 3$ be closed proper convex functions; $A_i : \mathcal{X}_i \rightarrow \mathcal{Z}^*$ with $i = 1, 2, 3$ be linear operators; and $b \in \mathcal{Z}^*$ be a given vector. We aim at finding a solution point of the model

$$\min \left\{ \sum_{i=1}^3 \theta_i(x_i) \mid \sum_{i=1}^3 A_i x_i = b, x_i \in \mathcal{X}_i, i = 1, 2, 3 \right\}. \quad (1.1)$$

The solution set of (1.1) is assumed to be nonempty throughout our discussion. In our analysis, we also use the notations $\mathbf{x} := (x_1, x_2, x_3)$ and $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$. Let the augmented Lagrangian function of (1.1) be

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

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

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

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, 34, 38]), 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. Numerically the original scheme (1.2a)–(1.2d) usually performs better than all the twisted variants with provable convergence (see, e.g., [16, 34]); and it is the most convenient scheme to be implemented compared with its variants. In [21], the authors suggested attaching a shrinkage factor to the Lagrange-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 function in the objective of (1.1) is strongly convex. Indeed, a relaxation factor in $(0, \frac{1+\sqrt{5}}{2})$ which was initiated in [11] can be attached to the Lagrange-multiplier updating step (1.2d) provided that the proximal terms and the penalty parameter β are chosen judiciously.

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 only of theoretical interests and they are usually too strict to be satisfied by the mentioned applications.

We are thus interested in understanding the gap between the empirical efficiency of (1.2a)–(1.2d) and the lack of theoretical conditions that can both ensure the convergence of (1.2a)–(1.2d) and be satisfied by some applications of the abstract model (1.1). As [22], we independently show that the convergence of (1.2a)–(1.2d) can be

ensured when one function in the objective of (1.1) is strongly convex, the penalty parameter β is appropriately restricted and some assumptions on the linear operators A_j 's hold—some conditions that hold for some concrete applications of (1.1). Note that the restriction onto β in [22] is determined through checking the positive definiteness of some operators, and it is generally more restrictive than ours because it targets a setting more general than (1.2a)–(1.2d) with larger step sizes for updating z and semi-proximal terms for regularizing the x_j -subproblems. Moreover, it worths to mention that such a condition on β is sufficient to theoretically ensure the convergence of (1.2a)–(1.2d). It is thus usually conservative and can be enlarged in practice to result in better numerical performance. Another purpose of this paper 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 globally linear convergence rate in asymptotical sense under some additional assumptions. We refer to [6, 21, 23, 24] for some convergence rate analysis for (1.2a)–(1.2d).

The rest of this paper is organized as follows. Some necessary preliminaries for further analysis are provided in Sect. 2. In Sect. 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 Sects. 4 and 5, respectively. In Sect. 6, under some additional conditions, we show that the scheme (1.2a)–(1.2d) is globally linear convergent. Finally, some concluding remarks are made in Sect. 7.

2 Preliminaries

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

2.1 Notation

We use calligraphic letters (e.g., \mathcal{H} , \mathcal{X}) to denote spaces, each equipped with an inner product $\langle \cdot, \cdot \rangle$ and an associated norm $\| \cdot \|$. All vectors are column vectors. For two arbitrary vectors $x \in \mathcal{X}_1$ and $y \in \mathcal{X}_2$, we simply use $u = (x, y)$ to denote their adjunction. That is, (x, y) denotes the vector $(x^*, y^*)^*$. For a linear operator A_i , we denote by A_i^* its adjoint operator.

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

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

Specially, for a symmetric linear operator A , $\|A\|$ denotes its spectral norm.

We say a function $f : \mathcal{H} \rightarrow (-\infty, \infty]$ is convex if

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

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

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

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

A multifunction $F : \mathcal{H} \rightrightarrows \mathcal{H}$ means that $F(x)$ is a set in \mathcal{H} , see, e.g., [32]. For a multifunction $F : \mathcal{H} \rightrightarrows \mathcal{H}$, we say that F is monotone if

$$\langle x - y, \xi - \zeta \rangle \geq 0, \quad \forall \xi \in F(x), \quad \forall \zeta \in F(y);$$

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

$$\langle x - y, \xi - \zeta \rangle \geq \mu\|x - y\|^2, \quad \forall \xi \in F(x), \quad \forall \zeta \in F(y).$$

It is well known, see, e.g., [32], 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 for any $x, y \in \text{dom } f$

$$\langle \xi - \zeta, x - y \rangle \geq 0, \quad \forall \xi \in \partial f(x), \quad \forall \zeta \in \partial f(y);$$

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

$$\langle \xi - \zeta, x - y \rangle \geq \mu\|x - y\|^2, \quad \forall \xi \in \partial f(x), \quad \forall \zeta \in \partial f(y),$$

where

$$\text{dom } f := \{x \in \mathcal{H} \mid f(x) < \infty\}$$

is the domain of $f : \mathcal{H} \rightarrow (-\infty, \infty]$. An equivalent inequality for the strong convexity of f is that

$$f(x) \geq f(y) + \langle \xi, x - y \rangle + \frac{\mu}{2}\|x - y\|^2, \quad \forall \xi \in \partial f(y), \quad \forall x, y \in \text{dom } f.$$

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

$$2|\langle a, b \rangle| \leq \varsigma\|a\|^2 + \frac{1}{\varsigma}\|b\|^2, \tag{2.1}$$

for any positive scalar ς . Obviously, we also have

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

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

$$\mathcal{L}(x_1, x_2, x_3, z) := \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3) - \langle z, A_1x_1 + A_2x_2 + A_3x_3 - b \rangle,$$

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

$$\begin{aligned} \mathcal{L}(\hat{x}_1, \hat{x}_2, \hat{x}_3, z) &\leq \mathcal{L}(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{z}) \leq \mathcal{L}(x_1, x_2, x_3, \hat{z}), \\ \forall (x_1, x_2, x_3, z) &\in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{Z}. \end{aligned}$$

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

$$\begin{cases} 0 \in \partial\theta_1(\hat{x}_1) - A_1^*\hat{z}, & (2.2a) \\ 0 \in \partial\theta_2(\hat{x}_2) - A_2^*\hat{z}, & (2.2b) \\ 0 \in \partial\theta_3(\hat{x}_3) - A_3^*\hat{z}, & (2.2c) \\ 0 = \sum_{i=1}^3 A_i\hat{x}_i - b. & (2.2d) \end{cases}$$

It follows from (2.2a)–(2.2c) that, for $i = 1, 2, 3$, we have

$$\theta_i(x'_i) \geq \theta_i(\hat{x}_i) + \langle \hat{z}, A_ix'_i - A_i\hat{x}_i \rangle, \quad \forall x'_i \in \mathcal{X}_i$$

if θ_i is convex; and

$$\theta_i(x'_i) \geq \theta_i(\hat{x}_i) + \langle \hat{z}, A_ix'_i - A_i\hat{x}_i \rangle + \frac{\mu_i}{2}\|x'_i - \hat{x}_i\|^2, \quad \forall x'_i \in \mathcal{X}_i \quad (2.3)$$

if θ_i is strongly convex with the modulus $\mu_i > 0$.

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

$$\mathcal{G}_{\mathcal{B}_x \times \mathcal{B}_z}(\mathbf{x}, z) := \max_{z' \in \mathcal{B}_z} \{ \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3) - \langle A_1x_1 + A_2x_2 + A_3x_3 - b, z' \rangle \} \\ - \min_{x' \in \mathcal{B}_x} \{ \theta_1(x'_1) + \theta_2(x'_2) + \theta_3(x'_3) - \langle A_1x'_1 + A_2x'_2 + A_3x'_3 - b, z \rangle \},$$

where $\mathcal{B}_x \times \mathcal{B}_z$ is a subset of $\mathcal{X} \times \mathcal{Z}$. If $\mathcal{B}_x \times \mathcal{B}_z$ contains a solution $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{z})$, then for any $(\mathbf{x}, z) \in \mathcal{B}_x \times \mathcal{B}_z$, we have

$$\mathcal{G}_{\mathcal{B}_x \times \mathcal{B}_z}(\mathbf{x}, z) \geq \left(\sum_{i=1}^3 \theta_i(x_i) - \left\langle \hat{z}, \sum_{i=1}^3 A_i x_i - b \right\rangle \right) \\ - \left(\sum_{i=1}^3 \theta_i(\hat{x}_i) - \left\langle z, \sum_{i=1}^3 A_i \hat{x}_i - b \right\rangle \right) \\ = \sum_{i=1}^3 (\theta_i(x_i) - \theta_i(\hat{x}_i) - \langle \hat{z}, A_i x_i - A_i \hat{x}_i \rangle) \\ \geq \sum_{i=1}^3 \frac{\mu_i}{2} \|x_i - \hat{x}_i\|^2 \\ \geq 0, \tag{2.4}$$

where the equality follows from (2.2d), the second inequality follows from (2.3), and the last inequality follows from the convexity of $\theta_i, i = 1, 2, 3$.

The following lemma gives a sufficient condition for a point to be a saddle point of $\mathcal{L}(x_1, x_2, x_3, z)$. The assertion is an immediate conclusion of (2.4); we thus omit the proof.

Lemma 2.1 *Let $(\bar{\mathbf{x}}, \bar{z})$ be an arbitrary point in interior of $\mathcal{B}_x \times \mathcal{B}_z$. If*

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

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

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

Finally, we use the notation

$$Q := \begin{pmatrix} \beta^2 A_2^* A_2 & 0 & 0 \\ 0 & \beta^2 A_3^* A_3 & 0 \\ 0 & 0 & I \end{pmatrix}, \tag{2.5}$$

where the operators 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_1 and A_2 in (1.1) are of full column rank.

Let us first present the variational characterization of the scheme (1.2a)–(1.2d). Indeed, similar as (2.2a)–(2.2d), the subproblems (1.2a)–(1.2c) can be characterized as

$$\begin{cases} 0 \in \partial\theta_1(x_1^{k+1}) - A_1^*z^k + \beta A_1^*(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b), & (3.1a) \\ 0 \in \partial\theta_2(x_2^{k+1}) - A_2^*z^k + \beta A_2^*(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b), & (3.1b) \\ 0 \in \partial\theta_3(x_3^{k+1}) - A_3^*z^k + \beta A_3^*(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b). & (3.1c) \end{cases}$$

Combining it with the step (1.2d), the $(k + 1)$ th iterate generated by (1.2a)–(1.2d) satisfies that for any $x'_1 \in \mathcal{X}_1$, $x'_2 \in \mathcal{X}_2$, and $x'_3 \in \mathcal{X}_3$,

$$\begin{cases} \theta_1(x'_1) - \theta_1(x_1^{k+1}) - \langle A_1x'_1 - A_1x_1^{k+1}, z^k - \beta(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b) \rangle \geq 0, & (3.2a) \\ \theta_2(x'_2) - \theta_2(x_2^{k+1}) - \langle A_2x'_2 - A_2x_2^{k+1}, z^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b) \rangle \geq 0, & (3.2b) \\ \theta_3(x'_3) - \theta_3(x_3^{k+1}) - \langle A_3x'_3 - A_3x_3^{k+1}, z^{k+1} \rangle \geq \frac{\mu_3}{2} \|x'_3 - x_3^{k+1}\|^2, & (3.2c) \\ z^{k+1} = z^k - \beta \left(\sum_{i=1}^3 A_i x_i^{k+1} - b \right). & (3.2d) \end{cases}$$

Now, we start to prove 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 $\{x_1^k, x_2^k, x_3^k, z^k\}$ generated by the direct extension of ADMM (1.2a)–(1.2d), we have*

$$\langle A_3x_3^{k+1} - A_3x_3^k, z^{k+1} - z^k \rangle \geq \mu_3 \|x_3^k - x_3^{k+1}\|^2, \tag{3.3}$$

and

$$\begin{aligned} & \langle A_2x_2^{k+1} - A_2x_2^k, z^{k+1} - z^k \rangle \\ & \geq - \left\{ \frac{3}{10} \|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{5}{3} \|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 + \frac{5}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2 \right\}. \end{aligned} \tag{3.4}$$

Proof Setting $x'_3 := x_3^k$ in (3.2c) yields

$$\theta_3(x_3^k) - \theta_3(x_3^{k+1}) - \langle A_3x_3^k - A_3x_3^{k+1}, z^{k+1} \rangle \geq \frac{\mu_3}{2} \|x_3^k - x_3^{k+1}\|^2;$$

and setting $x'_3 := x_3^{k+1}$ for the k th iteration yields

$$\theta_3(x_3^{k+1}) - \theta_3(x_3^k) - \langle A_3x_3^{k+1} - A_3x_3^k, z^k \rangle \geq \frac{\mu_3}{2} \|x_3^k - x_3^{k+1}\|^2.$$

Adding the above two inequalities, we have

$$\langle A_3(x_3^{k+1} - x_3^k), z^{k+1} - z^k \rangle \geq \mu_3 \|x_3^k - x_3^{k+1}\|^2,$$

and the inequality (3.3) is proved.

We now prove (3.4). It follows from (3.2d) that

$$\sum_{i=1}^3 A_i x_i^{k+1} - b = \frac{1}{\beta} (z^k - z^{k+1}). \tag{3.5}$$

Using (3.5), we can rewrite (3.2b) as

$$\theta_2(x'_2) - \theta_2(x_2^{k+1}) - \langle A_2x'_2 - A_2x_2^{k+1}, z^{k+1} - \beta(A_3x_3^k - A_3x_3^{k+1}) \rangle \geq 0, \quad \forall x'_2 \in \mathcal{X}_2. \tag{3.6}$$

Then, setting $x'_2 := x_2^k$ and $x'_2 := x_2^{k+1}$ in (3.6) for the k th iteration yields

$$\theta_2(x_2^k) - \theta_2(x_2^{k+1}) - \langle A_2x_2^k - A_2x_2^{k+1}, z^{k+1} - \beta(A_3x_3^k - A_3x_3^{k+1}) \rangle \geq 0,$$

and

$$\theta_2(x_2^{k+1}) - \theta_2(x_2^k) - \langle A_2x_2^{k+1} - A_2x_2^k, z^k - \beta(A_3x_3^{k-1} - A_3x_3^k) \rangle \geq 0,$$

respectively. Adding these two inequalities, we get

$$\langle A_2x_2^{k+1} - A_2x_2^k, z^{k+1} - z^k - \beta(A_3x_3^k - A_3x_3^{k+1}) \rangle + \beta \langle A_3x_3^{k-1} - A_3x_3^k \rangle \geq 0.$$

Rearranging terms, it follows that

$$\begin{aligned} & \langle A_2x_2^{k+1} - A_2x_2^k, z^{k+1} - z^k \rangle \\ & \geq \beta \langle A_2x_2^{k+1} - A_2x_2^k, (A_3x_3^k - A_3x_3^{k+1}) - (A_3x_3^{k-1} - A_3x_3^k) \rangle \\ & \geq - \left\{ \frac{3}{20} \|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{5}{3} \|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 \right\} \\ & \quad - \left\{ \frac{3}{20} \|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{5}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2 \right\} \\ & = - \left\{ \frac{3}{10} \|A_2x_2^{k+1} - A_2x_2^k\|_\beta^2 + \frac{5}{3} \|A_3x_3^k - A_3x_3^{k-1}\|_\beta^2 + \frac{5}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_\beta^2 \right\}, \end{aligned}$$

where the second inequality follows from (2.1). The proof is complete. □

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

$$\begin{aligned} & \left[\sum_{i=1}^3 \theta_i(x_i^{k+1}) - \left\langle z', \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle \right] - \left[\sum_{i=1}^3 \theta_i(x'_i) - \left\langle z^{k+1}, \sum_{i=1}^3 A_i x'_i - b \right\rangle \right] \\ & \leq \left\langle z^{k+1} - z', \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle \\ & \quad + \beta \left\langle A_1 x'_1 - A_1 x_1^{k+1} + A_2 x'_2 - A_2 x_2^{k+1}, A_3 x_3^k - A_3 x_3^{k+1} \right\rangle \\ & \quad + \beta \left\langle A_1 x'_1 - A_1 x_1^{k+1}, A_2 x_2^k - A_2 x_2^{k+1} \right\rangle - \frac{\mu_3}{2} \|x'_3 - x_3^{k+1}\|^2. \end{aligned} \tag{3.7}$$

Proof Using (3.2d) and rearranging terms, it follows from (3.2a)–(3.2c) that

$$\left\{ \begin{aligned} & \theta_1(x'_1) - \theta_1(x_1^{k+1}) - \left\langle A_1 x'_1 - A_1 x_1^{k+1}, z^{k+1} + \beta \left(A_2 (x_2^{k+1} - x_2^k) + A_3 (x_3^{k+1} - x_3^k) \right) \right\rangle \geq 0, \end{aligned} \right. \tag{3.8a}$$

$$\left\{ \begin{aligned} & \theta_2(x'_2) - \theta_2(x_2^{k+1}) - \left\langle A_2 x'_2 - A_2 x_2^{k+1}, z^{k+1} + \beta A_3 (x_3^{k+1} - x_3^k) \right\rangle \geq 0, \end{aligned} \right. \tag{3.8b}$$

$$\left\{ \begin{aligned} & \theta_3(x'_3) - \theta_3(x_3^{k+1}) - \left\langle A_3 x'_3 - A_3 x_3^{k+1}, z^{k+1} \right\rangle \geq \frac{\mu_3}{2} \|x'_3 - x_3^{k+1}\|^2. \end{aligned} \right. \tag{3.8c}$$

Adding (3.8a)–(3.8c) and rearranging terms, we obtain the assertion (3.7) immediately. □

Now, we need to further analyze the right-hand side of (3.7) and refine the assertion (3.7). This is done in the following lemma.

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

$$\begin{aligned} & 2\beta \left[\sum_{i=1}^3 \theta_i(x_i^{k+1}) - \left\langle z', \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle \right] - 2\beta \left[\sum_{i=1}^3 \theta_i(x'_i) - \left\langle z^{k+1}, \sum_{i=1}^3 A_i x'_i - b \right\rangle \right] \\ & \leq \sum_{i=2}^3 \left(\|A_i x_i^k - A_i x'_i\|_{\beta^2}^2 - \|A_i x_i^{k+1} - A_i x'_i\|_{\beta^2}^2 \right) + \left(\|z^k - z'\|^2 - \|z^{k+1} - z'\|^2 \right) \\ & \quad + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\ & \quad - \frac{1}{40} \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 - \left(\beta \mu_3 \|x_3^{k+1} - x'_3\|^2 - \frac{8}{3} \|A_3 x_3^{k+1} - A_3 x'_3\|_{\beta^2}^2 \right) \end{aligned}$$

$$\begin{aligned}
 & + \frac{17}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 - 2\beta\mu_3 \|x_3^k - x_3^{k+1}\|^2 - \|z^k - z^{k+1}\|^2 \\
 & + 2\beta^2 \left\langle \sum_{i=1}^3 A_i x'_i - b, A_2x_2^k - A_2x_2^{k+1} + A_3x_3^k - A_3x_3^{k+1} \right\rangle.
 \end{aligned} \tag{3.9}$$

Proof Because of the identity

$$\langle x, y \rangle = \frac{1}{2} \left(\|x\|^2 + \|y\|^2 - \|x - y\|^2 \right), \tag{3.10}$$

it follows that

$$\begin{aligned}
 & \beta \left\langle A_2x'_2 - A_2x_2^{k+1}, A_2x_2^k - A_2x_2^{k+1} \right\rangle \\
 & = \frac{1}{2} \left\{ \|A_2x_2^k - A_2x_2^{k+1}\|_{\beta}^2 + \|A_2x_2^{k+1} - A_2x'_2\|_{\beta}^2 - \|A_2x_2^k - A_2x'_2\|_{\beta}^2 \right\}
 \end{aligned}$$

and

$$\begin{aligned}
 & \beta \left\langle A_3x'_3 - A_3x_3^{k+1}, A_3x_3^k - A_3x_3^{k+1} \right\rangle \\
 & = \frac{1}{2} \left\{ \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta}^2 + \|A_3x_3^{k+1} - A_3x'_3\|_{\beta}^2 - \|A_3x_3^k - A_3x'_3\|_{\beta}^2 \right\}.
 \end{aligned}$$

Consequently, we have

$$\begin{aligned}
 & \beta \langle A_1x'_1 - A_1x_1^{k+1} + A_2x'_2 - A_2x_2^{k+1}, A_3x_3^k - A_3x_3^{k+1} \rangle \\
 & = \beta \left\langle \left(\sum_{i=1}^3 A_i x'_i - b \right) - \left(\sum_{i=1}^3 A_i x_i^{k+1} - b \right), A_3x_3^k - A_3x_3^{k+1} \right\rangle \\
 & \quad - \beta \left\langle A_3x'_3 - A_3x_3^{k+1}, A_3x_3^k - A_3x_3^{k+1} \right\rangle \\
 & = \beta \left\langle \left(\sum_{i=1}^3 A_i x'_i - b \right) - \left(\sum_{i=1}^3 A_i x_i^{k+1} - b \right), A_3x_3^k - A_3x_3^{k+1} \right\rangle \\
 & \quad - \frac{1}{2} \left\{ \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta}^2 + \|A_3x_3^{k+1} - A_3x'_3\|_{\beta}^2 - \|A_3x_3^k - A_3x'_3\|_{\beta}^2 \right\} \\
 & = \beta \left\langle \sum_{i=1}^3 A_i x'_i - b, A_3x_3^k - A_3x_3^{k+1} \right\rangle - \langle z^k - z^{k+1}, A_3x_3^k - A_3x_3^{k+1} \rangle \\
 & \quad - \frac{1}{2} \left\{ \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta}^2 + \|A_3x_3^{k+1} - A_3x'_3\|_{\beta}^2 - \|A_3x_3^k - A_3x'_3\|_{\beta}^2 \right\} \\
 & \leq \beta \left\langle \sum_{i=1}^3 A_i x'_i - b, A_3x_3^k - A_3x_3^{k+1} \right\rangle - \frac{1}{2} \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta}^2 - \mu_3 \|x_3^k - x_3^{k+1}\|^2 \\
 & \quad - \frac{1}{2} \left\{ \|A_3x_3^{k+1} - A_3x'_3\|_{\beta}^2 - \|A_3x_3^k - A_3x'_3\|_{\beta}^2 \right\},
 \end{aligned} \tag{3.11}$$

where the third equality follows from (3.5) and the inequality follows from (3.3). On the other hand, using (3.4), we have

$$\begin{aligned}
 & -\beta \left\langle \left(\sum_{i=1}^3 A_i x_i^{k+1} - b \right) + (A_3 x'_3 - A_3 x_3^{k+1}), A_2 x_2^k - A_2 x_2^{k+1} \right\rangle \\
 & \leq \frac{3}{10} \|A_2 x_2^{k+1} - A_2 x_2^k\|_\beta^2 + \frac{5}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_\beta^2 + \frac{5}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_\beta^2 \\
 & \quad + \beta \left\langle A_3 x_3^{k+1} - A_3 x'_3, A_2 x_2^k - A_2 x_2^{k+1} \right\rangle \\
 & \leq \frac{3}{10} \|A_2 x_2^{k+1} - A_2 x_2^k\|_\beta^2 + \frac{5}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_\beta^2 + \frac{5}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_\beta^2 \\
 & \quad + \frac{4}{3} \|A_3 x_3^{k+1} - A_3 x'_3\|_\beta^2 + \frac{3}{16} \|A_2 x_2^k - A_2 x_2^{k+1}\|_\beta^2 \\
 & = \frac{39}{80} \|A_2 x_2^{k+1} - A_2 x_2^k\|_\beta^2 + \frac{5}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_\beta^2 + \frac{5}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_\beta^2 \\
 & \quad + \frac{4}{3} \|A_3 x_3^{k+1} - A_3 x'_3\|_\beta^2,
 \end{aligned}$$

where the inequality follows from (2.1). Then, it holds that

$$\begin{aligned}
 & \beta \left\langle A_1 x'_1 - A_1 x_1^{k+1}, A_2 x_2^k - A_2 x_2^{k+1} \right\rangle \\
 & = \beta \left\langle \sum_{i=1}^3 A_i x'_i - b, A_2 x_2^k - A_2 x_2^{k+1} \right\rangle - \beta \left\langle A_2 x'_2 - A_2 x_2^{k+1}, A_2 x_2^k - A_2 x_2^{k+1} \right\rangle \\
 & \quad - \beta \left\langle \left(\sum_{i=1}^3 A_i x_i^{k+1} - b \right) + (A_3 x'_3 - A_3 x_3^{k+1}), A_2 x_2^k - A_2 x_2^{k+1} \right\rangle \\
 & \leq \beta \left\langle \sum_{i=1}^3 A_i x'_i - b, A_2 x_2^k - A_2 x_2^{k+1} \right\rangle \\
 & \quad - \frac{1}{2} \left\{ \|A_2 x_2^{k+1} - A_2 x'_2\|_\beta^2 - \|A_2 x_2^k - A_2 x'_2\|_\beta^2 \right\} \\
 & \quad + \frac{5}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_\beta^2 - \frac{5}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_\beta^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_\beta^2 \\
 & \quad - \frac{1}{80} \|A_2 x_2^k - A_2 x_2^{k+1}\|_\beta^2 + \frac{4}{3} \|A_3 x_3^{k+1} - A_3 x'_3\|_\beta^2. \tag{3.12}
 \end{aligned}$$

Furthermore, since

$$\begin{aligned}
 & \left\langle z^{k+1} - z', \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle = \frac{1}{\beta} \left\langle z^{k+1} - z', z^k - z^{k+1} \right\rangle \\
 & = \frac{1}{2\beta} \left\{ \|z^k - z'\|^2 - \|z^{k+1} - z'\|^2 - \|z^k - z^{k+1}\|^2 \right\}, \tag{3.13}
 \end{aligned}$$

we obtain the assertion (3.9) by combining (3.11), (3.12), (3.13), and (3.7). □

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 $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ be the sequence generated by the direct extension of ADMM (1.2a)–(1.2d) with $\beta \in \left(0, \frac{6\mu_3}{17\|A_3^*A_3\|}\right)$. Then, the sequence $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ converges to a saddle-point in S .*

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

$$\sum_{i=1}^3 A_i \hat{x}_i - b = 0. \tag{3.14}$$

Setting $(x'_1, x'_2, x'_3, z') := (\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{z})$ in (3.9) and using (3.14), we obtain

$$\begin{aligned} & 2\beta \left[\sum_{i=1}^3 \theta_i(x_i^{k+1}) - \left\langle \hat{z}, \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle \right] - 2\beta \sum_{i=1}^3 \theta_i(\hat{x}_i) \\ & \leq \sum_{i=2}^3 \left(\|A_i x_i^k - A_i \hat{x}_i\|_{\beta^2}^2 - \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 \right) + \left(\|z^k - \hat{z}\|^2 - \|z^{k+1} - \hat{z}\|^2 \right) \\ & \quad + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\ & \quad - \frac{1}{40} \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 - \left(\beta\mu_3 - \frac{8}{3}\beta^2 \|A_3^*A_3\| \right) \|x_3^{k+1} - \hat{x}_3\|^2 \\ & \quad - \left(2\beta\mu_3 - \frac{17}{3}\beta^2 \|A_3^*A_3\| \right) \|x_3^{k+1} - x_3^k\|^2 - \|z^k - z^{k+1}\|^2. \end{aligned} \tag{3.15}$$

Since $\beta \in \left(0, \frac{6\mu_3}{17\|A_3^*A_3\|}\right)$, we have

$$\varrho_1 := \beta\mu_3 - \frac{17}{6}\beta^2 \|A_3^*A_3\| > 0.$$

Then, (3.15) can be written as

$$\begin{aligned} & 2\beta \left[\sum_{i=1}^3 \theta_i(x_i^{k+1}) - \left\langle \hat{z}, \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle \right] - 2\beta \sum_{i=1}^3 \theta_i(\hat{x}_i) \\ & \leq \sum_{i=2}^3 \left(\|A_i x_i^k - A_i \hat{x}_i\|_{\beta^2}^2 - \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 \right) + \left(\|z^k - \hat{z}\|^2 - \|z^{k+1} - \hat{z}\|^2 \right) \\ & \quad + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{40}\|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 - \varrho_1\|x_3^{k+1} - x_3^k\|^2 \\
& - \varrho_1\|x_3^{k+1} - \hat{x}_3\|^2 - \|z^k - z^{k+1}\|^2.
\end{aligned} \tag{3.16}$$

Since $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{z})$ is a saddle point, it follows from (2.4) that the left-hand side of (3.16) is nonnegative. We thus have

$$\begin{aligned}
& \sum_{i=2}^3 \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 + \|z^{k+1} - \hat{z}\|^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\
& \leq \sum_{i=2}^3 \|A_i x_i^k - A_i \hat{x}_i\|_{\beta^2}^2 + \|z^k - \hat{z}\|^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 \\
& \quad - \frac{1}{40} \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 - \varrho_1 \|x_3^{k+1} - x_3^k\|^2 \\
& \quad - \varrho_1 \|x_3^{k+1} - \hat{x}_3\|^2 - \|z^k - z^{k+1}\|^2.
\end{aligned} \tag{3.17}$$

Recall the definition of Q in (2.5). Then, (3.17) can be rewritten as the compact form

$$\begin{aligned}
& \|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\
& \leq \|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 \\
& \quad - \frac{1}{40} \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 - \varrho_1 \|x_3^{k+1} - x_3^k\|^2 \\
& \quad - \varrho_1 \|x_3^{k+1} - \hat{x}_3\|^2 - \|z^k - z^{k+1}\|^2,
\end{aligned} \tag{3.18}$$

where $\mathbf{w} := (x_2, x_3, z)$.

Thus we have

$$\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \leq \|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2, \tag{3.19}$$

indicating that the sequence $\{\mathbf{w}^k\}$ is bounded. Rearranging terms for (3.18) yields

$$\begin{aligned}
& \frac{1}{40} \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 + \varrho_1 \|x_3^{k+1} - x_3^k\|^2 + \varrho_1 \|x_3^{k+1} - \hat{x}_3\|^2 + \|z^k - z^{k+1}\|^2 \\
& \leq \left(\|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 \right) \\
& \quad - \left(\|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \right).
\end{aligned}$$

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

$$\sum_{k=1}^{\infty} \left\{ \frac{1}{40} \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 + \varrho_1 \|x_3^{k+1} - x_3^k\|^2 + \varrho_1 \|x_3^{k+1} - \hat{x}_3\|^2 + \|z^k - z^{k+1}\|^2 \right\} \leq \left(\|\mathbf{w}^1 - \hat{\mathbf{w}}\|_{\mathcal{Q}}^2 + \frac{10}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 \right),$$

and hence

$$\begin{aligned} \lim_{k \rightarrow \infty} \|z^k - z^{k+1}\| &= \lim_{k \rightarrow \infty} \|A_2 x_2^k - A_2 x_2^{k+1}\| = \lim_{k \rightarrow \infty} \|x_3^k - x_3^{k+1}\| \\ &= \lim_{k \rightarrow \infty} \|x_3^{k+1} - \hat{x}_3\| = 0. \end{aligned} \tag{3.20}$$

The full column rank of A_1 assumed in Assumption 3.1, (3.20), and (1.2d) then imply that $\{x_1^k\}$ is also bounded. Hence the iterative sequence $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ generated by ADMM (1.2a)–(1.2d) is bounded.

Then, the boundedness of the sequence indicates that it has at least one cluster point. Let $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{z})$ be an arbitrary cluster point of $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ and $\{(x_1^{k_j}, x_2^{k_j}, x_3^{k_j}, z^{k_j})\}$ be the subsequence converging to $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{z})$. Taking the limit along the subsequence in (3.1a)–(3.1c) and using (3.20), we have

$$\begin{cases} 0 \in \partial\theta_1(\tilde{x}_1) - A_1^* \tilde{z}, \\ 0 \in \partial\theta_2(\tilde{x}_2) - A_2^* \tilde{z}, \\ 0 \in \partial\theta_3(\tilde{x}_3) - A_3^* \tilde{z}. \end{cases}$$

Moreover, since $\|z^k - z^{k+1}\| = \beta \left\| \sum_{j=1}^3 A_j x_j^{k+1} - b \right\|$, by taking the limit, we have

$\sum_{j=1}^3 A_j \tilde{x}_j - b = 0$. Hence, it follows from (2.2a)–(2.2d) that $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{z})$ is a saddle point in S and (3.19) means that the whole sequence has only one cluster point. Consequently, the iterative sequence $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ 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 some specific applications of the abstract model (1.1), it may not be difficult to determine the range for β . For example, for the application of moving object detection by detecting contiguous outliers studied in [37], it is easy to see that its model corresponds to (1.1) with $A_3 = I$ and $\mu_3 = 1$. Thus, choosing $\beta \in (0, \frac{6}{17})$ can sufficiently ensure the convergence if the scheme (1.2a)–(1.2d) is implemented to solve the model in [37].

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 $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ generated by the direct extension of ADMM (1.2a)–(1.2d) is convergent. Thus, there exists a constant $M > 0$ such that

$$\|A_i x_i^k\| \leq M, \quad i = 1, 2, 3, \quad \text{and} \quad \|z^k\| \leq M, \quad \forall k \geq 0.$$

Then, it follows from Assumption 3.1 and (3.9) that for all $(x'_1, x'_2, x'_3, z') \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{Z}$, we have

$$\begin{aligned} & 2\beta \left[\sum_{i=1}^3 \theta_i(x_i^{k+1}) - \left\langle z', \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle \right] \\ & - 2\beta \left[\sum_{i=1}^3 \theta_i(x'_i) - \left\langle z^{k+1}, \sum_{i=1}^3 A_i x'_i - b \right\rangle \right] \\ & \leq \sum_{i=2}^3 \left(\|A_i x_i^k - A_i x'_i\|_{\beta^2}^2 - \|A_i x_i^{k+1} - A_i x'_i\|_{\beta^2}^2 \right) + \left(\|z^k - z'\|^2 - \|z^{k+1} - z'\|^2 \right) \\ & \quad + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\ & \quad + 2\beta^2 \left\langle \sum_{i=1}^3 A_i x'_i - b, A_2 x_2^k - A_2 x_2^{k+1} + A_3 x_3^k - A_3 x_3^{k+1} \right\rangle. \end{aligned}$$

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

$$\begin{aligned} & 2\beta \sum_{k=1}^t \left\{ \left[\sum_{i=1}^3 \theta_i(x_i^{k+1}) - \left\langle z', \sum_{i=1}^3 A_i x_i^{k+1} - b \right\rangle \right] \right. \\ & \quad \left. - \left[\sum_{i=1}^3 \theta_i(x'_i) - \left\langle z^{k+1}, \sum_{i=1}^3 A_i x'_i - b \right\rangle \right] \right\} \\ & \leq \sum_{i=2}^3 \|A_i x_i^1 - A_i x'_i\|_{\beta^2}^2 + \|z^1 - z'\|^2 + \frac{10}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 \\ & \quad + 2\beta^2 \left\langle \sum_{i=1}^3 A_i x'_i - b, A_2 x_2^0 - A_2 x_2^{t+1} + A_3 x_3^0 - A_3 x_3^{t+1} \right\rangle \end{aligned}$$

$$\begin{aligned}
 &\leq \sum_{i=2}^3 \|A_i x_i^1 - A_i x'_i\|_{\beta^2}^2 + \|z^1 - z'\|^2 + \frac{10}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 \\
 &\quad + 2\beta^2 \left\| \sum_{i=1}^3 A_i x'_i - b \right\| \left\| A_2 x_2^0 - A_2 x_2^{t+1} + A_3 x_3^0 - A_3 x_3^{t+1} \right\| \\
 &\leq \sum_{i=2}^3 \|A_i x_i^1 - A_i x'_i\|_{\beta^2}^2 + \|z^1 - z'\|^2 \\
 &\quad + \frac{10}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 + 8\beta^2 M \left\| \sum_{i=1}^3 A_i x'_i - b \right\|, \tag{4.1}
 \end{aligned}$$

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

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

Then, it follows from the convexity of $\theta_1, \theta_2,$ and θ_3 that

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

Combining (4.1) and (4.2) yields

$$\begin{aligned}
 &2\beta t \left\{ \left[\sum_{i=1}^3 \theta_i(x_i^K) - \left\langle z', \sum_{i=1}^3 A_i x_i^K - b \right\rangle \right] - \left[\sum_{i=1}^3 \theta_i(x'_i) - \left\langle z^1, \sum_{i=1}^3 A_i x'_i - b \right\rangle \right] \right\} \\
 &\leq \sum_{i=2}^3 \|A_i x_i^1 - A_i x'_i\|_{\beta^2}^2 + \|z^1 - z'\|^2 + \frac{10}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 + 8\beta^2 M \left\| \sum_{i=1}^3 A_i x'_i - b \right\|. \tag{4.3}
 \end{aligned}$$

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

$$2\beta t \mathcal{G}_{\mathcal{B}_x \times \mathcal{B}_z} (x_1^K, x_2^K, x_3^K, z^K) \leq D,$$

where

$$\begin{aligned}
 D := \sup_{(x', z') \in \mathcal{B}_x \times \mathcal{B}_z} &\left\{ \sum_{i=2}^3 \|A_i x_i^1 - A_i x'_i\|_{\beta^2}^2 + \|z^1 - z'\|^2 \right. \\
 &\left. + \frac{10}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 + 8\beta^2 M \left\| \sum_{i=1}^3 A_i x'_i - b \right\| \right\}.
 \end{aligned}$$

That is, (\mathbf{x}^K, z^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. \square

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 more assumptions to derive a non-ergodic worst-case convergence rate. We first make the following additional assumption, and refer to, e.g. [38], for some applications that this assumption holds.

Assumption 5.1 In (1.1), θ_1 and θ_2 are convex, θ_3 is smooth and $\nabla\theta_3$ is Lipschitz continuous with constant $L_3 > 0$, θ_3 is strongly convex with modulus $\mu_3 > L_3/2$; A_1 and A_2 are of full column rank.

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:

$$\begin{cases} \theta_1(x'_1) - \theta_1(x_1^{k+1}) - \left(A_1x'_1 - A_1x_1^{k+1}, z^{k+1} + \beta \left(A_2(x_2^{k+1} - x_2^k) + A_3(x_3^{k+1} - x_3^k) \right) \right) \geq 0, & (5.1a) \\ \theta_2(x'_2) - \theta_2(x_2^{k+1}) - \left(A_2x'_2 - A_2x_2^{k+1}, z^{k+1} + \beta A_3(x_3^{k+1} - x_3^k) \right) \geq 0, & (5.1b) \\ \theta_3(x'_3) - \theta_3(x_3^{k+1}) - \left(A_3x'_3 - A_3x_3^{k+1}, z^{k+1} \right) \geq \frac{\mu_3}{2} \|x'_3 - x_3^{k+1}\|^2, & (5.1c) \\ z^{k+1} = z^k - \beta \left(\sum_{i=1}^3 A_i x_i^{k+1} - b \right). & (5.1d) \end{cases}$$

Therefore, if we have

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

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

Define

$$\beta_{\max} := \min \left\{ \frac{2\mu_3 - L_3}{\|A_3^* A_3\|}, \frac{6\mu_3}{17 \|A_3^* A_3\|} \right\}. \tag{5.3}$$

It follows from Assumption 5.1 that $\beta_{\max} > 0$. To derive the non-ergodic worst-case convergence rate, let us first prove a key inequality that will play an important role in our further analysis.

Lemma 5.1 *Suppose Assumption 5.1 holds. Let $\{(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, z^{k+1})\}$ be the sequence generated by the direct extension of ADMM (1.2a)–(1.2d) with the restriction $\beta \in (0, \beta_{\max})$. Then we have*

$$\begin{aligned} & \frac{1}{2\beta} \|z^{k-1} - z^k\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^{k-1} - A_3 x_3^k\|^2 + \|A_2 x_2^{k-1} - A_2 x_2^k\|^2 \right\} \\ & \geq \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^k - A_3 x_3^{k+1}\|^2 + \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 \right\}. \end{aligned}$$

Proof Setting $x'_1 := x_1^k$ in (5.1a) yields

$$\begin{aligned} & \theta_1(x_1^k) - \theta_1(x_1^{k+1}) - \left\langle A_1 x_1^k - A_1 x_1^{k+1}, z^{k+1} \right. \\ & \quad \left. + \beta \left(A_2(x_2^{k+1} - x_2^k) + A_3(x_3^{k+1} - x_3^k) \right) \right\rangle \geq 0. \end{aligned}$$

Rearranging terms, we have

$$\begin{aligned} & \theta_1(x_1^k) - \theta_1(x_1^{k+1}) \geq \left\langle A_1 x_1^k - A_1 x_1^{k+1}, z^{k+1} \right\rangle \\ & \quad - \beta \left\langle A_1 x_1^k - A_1 x_1^{k+1}, \sum_{j=2}^3 A_j(x_j^k - x_j^{k+1}) \right\rangle. \end{aligned}$$

Moreover, setting $x'_1 := x_1^{k+1}$ for the k th iteration, we obtain

$$\begin{aligned} & \theta_1(x_1^{k+1}) - \theta_1(x_1^k) \geq \left\langle A_1 x_1^{k+1} - A_1 x_1^k, z^k \right\rangle \\ & \quad - \beta \left\langle A_1 x_1^{k+1} - A_1 x_1^k, \sum_{j=2}^3 A_j(x_j^{k-1} - x_j^k) \right\rangle. \end{aligned}$$

Adding the above two inequalities and rearranging terms, we have

$$\begin{aligned} & \left\langle A_1 x_1^{k+1} - A_1 x_1^k, z^{k+1} - z^k \right\rangle \geq -\beta \left\langle A_1 x_1^k - A_1 x_1^{k+1}, \sum_{j=2}^3 A_j(x_j^k - x_j^{k+1}) \right\rangle \\ & \quad - \beta \left\langle A_1 x_1^{k+1} - A_1 x_1^k, \sum_{j=2}^3 A_j(x_j^{k-1} - x_j^k) \right\rangle. \end{aligned} \tag{5.4}$$

Similarly, setting $x'_2 = x_2^k$ and $x'_2 = x_2^{k+1}$ in the k th iteration in (5.1b), and adding the resulting inequalities, we get the following inequality:

$$\begin{aligned} \left\langle A_2 x_2^{k+1} - A_2 x_2^k, z^{k+1} - z^k \right\rangle &\geq -\beta \left\langle A_2 x_2^k - A_2 x_2^{k+1}, A_3 (x_3^k - x_3^{k+1}) \right\rangle \\ &\quad - \beta \left\langle A_2 x_2^{k+1} - A_2 x_2^k, A_3 (x_3^{k-1} - x_3^k) \right\rangle. \end{aligned} \quad (5.5)$$

Then, it follows from (5.1c) that

$$\left\langle A_3 x_3^{k+1} - A_3 x_3^k, z^{k+1} - z^k \right\rangle \geq \mu_3 \|x_3^{k+1} - x_3^k\|^2. \quad (5.6)$$

Adding (5.4)–(5.6), we have

$$\begin{aligned} &\left\langle \sum_{j=1}^3 A_j x_j^{k+1} - \sum_{j=1}^3 A_j x_j^k, z^{k+1} - z^k \right\rangle \\ &\geq -\beta \left\langle A_1 x_1^k - A_1 x_1^{k+1}, \sum_{j=2}^3 A_j (x_j^k - x_j^{k+1}) \right\rangle \\ &\quad - \beta \left\langle A_1 x_1^{k+1} - A_1 x_1^k, \sum_{j=2}^3 A_j (x_j^{k-1} - x_j^k) \right\rangle \\ &\quad - \beta \left\langle A_2 x_2^k - A_2 x_2^{k+1}, A_3 (x_3^k - x_3^{k+1}) \right\rangle - \beta \left\langle A_2 x_2^{k+1} - A_2 x_2^k, A_3 (x_3^{k-1} - x_3^k) \right\rangle \\ &\quad + \mu_3 \|x_3^{k+1} - x_3^k\|^2. \end{aligned} \quad (5.7)$$

Further, because of the identity (3.5), we know that

$$\sum_{j=1}^3 A_j x_j^{k+1} - \sum_{j=1}^3 A_j x_j^k = \frac{1}{\beta} \left((z^k - z^{k+1}) - (z^{k-1} - z^k) \right).$$

We thus can rewrite the left-hand term of (5.7) as

$$\begin{aligned} &\left\langle \sum_{j=1}^3 A_j x_j^{k+1} - \sum_{j=1}^3 A_j x_j^k, z^{k+1} - z^k \right\rangle \\ &= \frac{1}{\beta} \left\langle (z^k - z^{k+1}) - (z^{k-1} - z^k), z^{k+1} - z^k \right\rangle \\ &= \frac{1}{2\beta} \left\{ \|z^{k-1} - z^k\|^2 - \|z^k - z^{k+1}\|^2 - \|(z^k - z^{k+1}) - (z^{k-1} - z^k)\|^2 \right\}. \end{aligned} \quad (5.8)$$

Now, we treat the right-hand side of (5.7). Indeed, these terms can be written as

$$\begin{aligned}
 & -\beta \left\langle A_1x_1^k - A_1x_1^{k+1}, \sum_{j=2}^3 (A_jx_j^k - A_jx_j^{k+1}) \right\rangle - \beta \left\langle A_1x_1^{k+1} - A_1x_1^k, \sum_{j=2}^3 (A_jx_j^{k-1} - A_jx_j^k) \right\rangle \\
 & -\beta \left\langle A_2x_2^k - A_2x_2^{k+1}, A_3x_3^k - A_3x_3^{k+1} \right\rangle - \beta \left\langle A_2x_2^{k+1} - A_2x_2^k, A_3x_3^{k-1} - A_3x_3^k \right\rangle \\
 & = -\beta \left\langle A_3x_3^k - A_3x_3^{k+1}, \sum_{j=1}^2 (A_jx_j^k - A_jx_j^{k+1}) \right\rangle - \beta \left\langle A_1x_1^k - A_1x_1^{k+1}, A_2x_2^k - A_2x_2^{k+1} \right\rangle \\
 & -\beta \left\langle A_3x_3^{k-1} - A_3x_3^k, \sum_{j=1}^2 (A_jx_j^{k+1} - A_jx_j^k) \right\rangle - \beta \left\langle A_1x_1^{k+1} - A_1x_1^k, A_2x_2^{k-1} - A_2x_2^k \right\rangle. \tag{5.9}
 \end{aligned}$$

Using the identity (3.10), we have

$$\begin{aligned}
 & -\beta \left\langle A_3x_3^k - A_3x_3^{k+1}, \sum_{j=1}^2 (A_jx_j^k - A_jx_j^{k+1}) \right\rangle \\
 & = \frac{\beta}{2} \left\{ \|A_3x_3^k - A_3x_3^{k+1}\|^2 + \left\| \sum_{j=1}^2 A_j (x_j^k - x_j^{k+1}) \right\|^2 - \left\| \sum_{j=1}^3 A_j (x_j^k - x_j^{k+1}) \right\|^2 \right\} \tag{5.10}
 \end{aligned}$$

and

$$\begin{aligned}
 & -\beta \left\langle A_1x_1^k - A_1x_1^{k+1}, A_2x_2^k - A_2x_2^{k+1} \right\rangle \\
 & = \frac{\beta}{2} \left\{ \|A_1x_1^k - A_1x_1^{k+1}\|^2 + \|A_2x_2^k - A_2x_2^{k+1}\|^2 - \left\| \sum_{j=1}^2 A_j (x_j^k - x_j^{k+1}) \right\|^2 \right\}. \tag{5.11}
 \end{aligned}$$

Then, using (2.1) and (3.3), we have that for all $\omega > 0$

$$\begin{aligned}
 & -\beta \left\langle A_3x_3^{k-1} - A_3x_3^k, \sum_{j=1}^2 (A_jx_j^{k+1} - A_jx_j^k) \right\rangle \\
 & = -\beta \left\langle A_3x_3^{k-1} - A_3x_3^k, \sum_{j=1}^3 (A_jx_j^{k+1} - A_jx_j^k) - (A_3x_3^{k+1} - A_3x_3^k) \right\rangle \\
 & = -\left\langle A_3x_3^{k-1} - A_3x_3^k, (z^k - z^{k+1}) - (z^{k-1} - z^k) \right\rangle \\
 & \quad + \beta \left\langle A_3x_3^{k-1} - A_3x_3^k, A_3x_3^{k+1} - A_3x_3^k \right\rangle \\
 & \geq -\frac{\omega}{2} \|x_3^{k-1} - x_3^k\|^2 - \frac{1}{2\omega} \|A_3^*z^k - A_3^*z^{k+1}\|^2 + \mu_3 \|x_3^{k-1} - x_3^k\|^2 \\
 & \quad - \frac{\beta}{2} \|A_3x_3^{k-1} - A_3x_3^k\|^2 - \frac{\beta}{2} \|A_3x_3^k - A_3x_3^{k+1}\|^2, \tag{5.12}
 \end{aligned}$$

and

$$\begin{aligned}
 & -\beta \left\langle A_1 x_1^{k+1} - A_1 x_1^k, A_2 x_2^{k-1} - A_2 x_2^k \right\rangle \geq \\
 & -\frac{\beta}{2} \left\{ \|A_1 x_1^{k+1} - A_1 x_1^k\|^2 + \|A_2 x_2^{k-1} - A_2 x_2^k\|^2 \right\}. \quad (5.13)
 \end{aligned}$$

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

$$\begin{aligned}
 & -\beta \left\langle A_1 x_1^k - A_1 x_1^{k+1}, \sum_{j=2}^3 (A_j x_j^k - A_j x_j^{k+1}) \right\rangle \\
 & -\beta \left\langle A_1 x_1^{k+1} - A_1 x_1^k, \sum_{j=2}^3 (A_j x_j^{k-1} - A_j x_j^k) \right\rangle \\
 & -\beta \left\langle A_2 x_2^k - A_2 x_2^{k+1}, A_3 x_3^k - A_3 x_3^{k+1} \right\rangle - \beta \left\langle A_2 x_2^{k+1} - A_2 x_2^k, A_3 x_3^{k-1} - A_3 x_3^k \right\rangle \\
 & \geq \frac{\beta}{2} \left(\|A_2 x_2^k - A_2 x_2^{k+1}\|^2 - \|A_2 x_2^{k-1} - A_2 x_2^k\|^2 \right) \\
 & -\frac{1}{2\omega} \|A_3^* z^k - A_3^* z^{k+1}\|^2 - \frac{\beta}{2} \|A_3 x_3^{k-1} - A_3 x_3^k\|^2 \\
 & + \left(\mu_3 - \frac{\omega}{2} \right) \|x_3^{k-1} - x_3^k\|^2 - \frac{\beta}{2} \left\| \sum_{j=1}^3 A_j (x_j^k - x_j^{k+1}) \right\|^2. \quad (5.14)
 \end{aligned}$$

Using the identities $A_3^* z^{k+1} = \nabla \theta_3(x_3^{k+1})$ and $A_3^* z^k = \nabla \theta_3(x_3^k)$, and the Lipschitz continuity of $\nabla \theta_3$, we obtain

$$\|A_3^* z^{k+1} - A_3^* z^k\| \leq L_3 \|x_3^{k+1} - x_3^k\|.$$

Combining (5.8) and (5.14), and setting $\omega = L_3$, it follows from (5.7) that

$$\begin{aligned}
 & \frac{1}{2\beta} \|z^{k-1} - z^k\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^{k-1} - A_3 x_3^k\|^2 + \|A_2 x_2^{k-1} - A_2 x_2^k\|^2 \right\} \\
 & \geq \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^k - A_3 x_3^{k+1}\|^2 + \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 \right\} \\
 & + \left(\mu_3 - \frac{L_3}{2} \right) \|x_3^k - x_3^{k+1}\|^2 \\
 & + \left(\mu_3 - \frac{\beta \|A_3^* A_3\|}{2} - \frac{L_3}{2} \right) \|x_3^{k-1} - x_3^k\|^2 \\
 & \geq \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^k - A_3 x_3^{k+1}\|^2 + \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 \right\},
 \end{aligned}$$

where the last inequality is due to the fact that $\beta \leq \frac{2\mu_3 - L_3}{\|A_3^* A_3\|}$. The proof is complete. \square

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 $O(1/t)$ and $o(1/t)$ orders. Recall that we use the residual of (5.2)–(5.2c) to measure the accuracy of an iterate of the scheme (3.2a)–(3.2d).

Theorem 5.1 *Suppose Assumption 5.1 holds. Let $\{(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, z^{k+1})\}$ be the sequence generated by the direct extension of ADMM (3.2a)–(3.2d) with the restriction $\beta \in (0, \beta_{\max})$. Then, we have*

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

Proof First, it follows from (3.18) that

$$\begin{aligned} & \frac{1}{40} \|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \varrho_1 \|x_3^{k+1} - x_3^k\|^2 + \|z^k - z^{k+1}\|^2 \\ & \leq \|w^k - \hat{w}\|_Q^2 - \|w^{k+1} - \hat{w}\|_Q^2 + \frac{10}{3} \|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \frac{10}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2. \end{aligned} \tag{5.15}$$

Setting

$$\mu = \min \left\{ \frac{\beta}{20}, \frac{2\varrho_1}{\beta \|A_3^* A_3\|} \right\},$$

it follows that

$$\begin{aligned} & \mu \left\{ \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left(\|A_3x_3^k - A_3x_3^{k+1}\|^2 + \|A_2x_2^k - A_2x_2^{k+1}\|^2 \right) \right\} \\ & \leq \frac{1}{40} \|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \varrho_1 \|x_3^{k+1} - x_3^k\|^2 + \|z^k - z^{k+1}\|^2 \\ & \leq \|w^k - \hat{w}\|_Q^2 - \|w^{k+1} - \hat{w}\|_Q^2 + \frac{10}{3} \|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \frac{10}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2. \end{aligned} \tag{5.16}$$

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

$$\begin{aligned} & \sum_{k=1}^t \left\{ \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left(\|A_3x_3^k - A_3x_3^{k+1}\|^2 + \|A_2x_2^k - A_2x_2^{k+1}\|^2 \right) \right\} \\ & \leq \frac{1}{\mu} \left\{ \|w^0 - \hat{w}\|_Q^2 + \frac{10}{3} \|A_3x_3^1 - A_3x_3^0\|_{\beta^2}^2 \right\}. \end{aligned}$$

Then, using Lemma 5.1, we obtain

$$\frac{1}{2\beta} \|z^t - z^{t+1}\|^2 + \frac{\beta}{2} \left(\|A_3x_3^t - A_3x_3^{t+1}\|^2 + \|A_2x_2^t - A_2x_2^{t+1}\|^2 \right)$$

$$\leq \frac{1}{\mu t} \left\{ \|\mathbf{w}^0 - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3 x_3^1 - A_3 x_3^0\|_{\beta^2}^2 \right\},$$

which means that

$$\frac{1}{2\beta} \|z^t - z^{t+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^t - A_3 x_3^{t+1}\|^2 + \|A_2 x_2^t - A_2 x_2^{t+1}\|^2 \right\} = O(1/t).$$

The proof is complete. \square

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 *Suppose Assumption 5.1 holds. Let $\{(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, z^{k+1})\}$ be the sequence generated by the direct extension of ADMM (3.2a)–(3.2d) with the restriction $\beta \in (0, \beta_{\max})$. Then, we have*

$$\frac{1}{2\beta} \|z^t - z^{t+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^t - A_3 x_3^{t+1}\|^2 + \|A_2 x_2^t - A_2 x_2^{t+1}\|^2 \right\} = o(1/t). \quad (5.17)$$

Proof Taking $k = t$ to $2t$ in (5.16) and adding the resulting inequalities, we obtain

$$\begin{aligned} & t \left\{ \frac{1}{2\beta} \|z^{2t} - z^{2t+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^{2t} - A_3 x_3^{2t+1}\|^2 + \|A_2 x_2^{2t} - A_2 x_2^{2t+1}\|^2 \right\} \right\} \\ & \leq \sum_{k=t}^{2t} \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^k - A_3 x_3^{k+1}\|^2 + \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 \right\} \\ & = \sum_{k=1}^{2t} \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^k - A_3 x_3^{k+1}\|^2 + \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 \right\} \\ & \quad - \sum_{k=1}^t \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} \left\{ \|A_3 x_3^k - A_3 x_3^{k+1}\|^2 + \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 \right\}. \end{aligned}$$

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

6 Globally linear convergence

In this section, we show that it is possible to theoretically derive the globally linear convergence of the direct extension of ADMM (1.2a)–(1.2d) provided that some strengthen enough conditions are assumed.

6.1 Linear convergence under stronger conditions

In this subsection, we prove the linear convergence of the direct extension of ADMM under the following conditions. We refer to [24] for some existing analysis under stronger conditions.

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; $\nabla\theta_1$ is Lipschitz continuous with constant L_1 ; A_2 and A_3 are of full column rank; and A_1 is nonsingular.

First of all, under Assumption 3.1, we can prove a result similar to (3.18). That is, there exists a constant $\zeta = \min\{\frac{1}{40}, \varrho_1, \mu_2\} > 0$, such that

$$\begin{aligned} & \|w^{k+1} - \hat{w}\|_Q^2 + \frac{10}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \\ & \leq \|w^k - \hat{w}\|_Q^2 + \frac{10}{3} \|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 - \zeta \left\{ \|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 \right. \\ & \quad \left. + \|z^k - z^{k+1}\|^2 + \sum_{i=2}^3 \|x_i^{k+1} - \hat{x}_i\|^2 \right\}. \end{aligned} \tag{6.1}$$

Then, let us explain the roadmap to prove the linear convergence rate of (1.2a)–(1.2d). Indeed, according to (6.1), it is clear that we only need to find bounds for the terms $\|z^{k+1} - \hat{z}\|^2$ and $\|A_1x_1^{k+1} - A_1\hat{x}_1\|^2$ because of the minus term in (6.1). We first consider $\|A_1x_1^{k+1} - A_1\hat{x}_1\|^2$. It follows from (3.14) and (3.5) that

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

Therefore, we have

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

Then, we need to consider how to bound the term $\|z^{k+1} - \hat{z}\|^2$ to derive the linear convergence of the sequence generated by the scheme (1.2a)–(1.2d). As we show below, this is exactly why we need to assume Assumption 6.1.

Lemma 6.1 *Suppose Assumption 6.1 holds. Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{z})$ be a saddle point in S ; let $\{(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, z^{k+1})\}$ be the sequence generated by the direct extension of ADMM (1.2a)–(1.2d). Then, there exists a constant $\sigma > 0$ such that*

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

Proof It follows from the optimality conditions for the x_1 - and \hat{x}_1 -subproblems that

$$A_1^* z^{k+1} = \nabla \theta_1(x_1^{k+1}) + \beta A_1^* \sum_{i=2}^3 (A_i x_i^k - A_i x_i^{k+1})$$

and

$$A_1^* \hat{z} = \nabla \theta_1(\hat{x}_1).$$

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

$$\|A_1^* z^{k+1} - A_1^* \hat{z}\| \leq L_1 \|x_1^{k+1} - \hat{x}_1\| + \beta \|A_1\| \sum_{i=2}^3 (\|A_i\| \|x_i^k - x_i^{k+1}\|),$$

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

Now, we can derive the linear convergence of the sequence generated by the scheme (1.2a)-(1.2d) under Assumptions 3.1 and 6.1.

Theorem 6.1 *Suppose Assumptions 3.1 and 6.1 hold. Then the sequence $\{(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, z^{k+1})\}$ generated by the direct extension of ADMM (1.2a)–(1.2d) with $\beta \in (0, \frac{6\mu_3}{17\|A_3^* A_3\|})$ converges globally linearly to a saddle point in S .*

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

$$\begin{aligned} & \frac{\zeta}{4} \left\{ \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 + \sum_{i=2}^3 \|x_i^{k+1} - \hat{x}_i\|^2 \right\} \\ & \geq \frac{\zeta}{4} \left\{ \|z^k - z^{k+1}\|^2 + \sum_{i=2}^3 \|x_i^{k+1} - \hat{x}_i\|^2 \right\} \\ & \geq \frac{\zeta}{4} \left\{ \beta^2 \frac{1}{\beta^2} \|z^k - z^{k+1}\|^2 + \sum_{i=2}^3 \frac{1}{\|A_i^* A_i\|} \|A_i x_i^{k+1} - A_i \hat{x}_i\|^2 \right\} \\ & \geq \frac{\zeta}{12} \min \left\{ \beta^2, \frac{1}{\|A_2^* A_2\|}, \frac{1}{\|A_3^* A_3\|} \right\} \|A_1 x_1^{k+1} - A_1 \hat{x}_1\|^2 \\ & \geq \frac{\zeta}{12} \min \left\{ \beta^2, \frac{1}{\|A_2^* A_2\|}, \frac{1}{\|A_3^* A_3\|} \right\} \lambda_{\min}(A_1^* A_1) \|x_1^{k+1} - \hat{x}_1\|^2 \\ & = \tau_1 \|x_1^{k+1} - \hat{x}_1\|^2, \end{aligned} \tag{6.3}$$

where $\tau_1 = \frac{\zeta}{12} \min \left\{ \beta^2, \frac{1}{\|A_2^* A_2\|}, \frac{1}{\|A_3^* A_3\|} \right\} \lambda_{\min}(A_1^* A_1)$. In addition, we have

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

where $\tau_2 = \frac{\zeta}{4} \min \{ \beta^2 \lambda_{\min}(A_2^*A_2), 1 \}$. Adding (6.3) and (6.4), and using Lemma 6.1, we have

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

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

$$\begin{aligned}
 & \frac{\zeta}{2} \left\{ \|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 + \sum_{i=2}^3 \|x_i^{k+1} - \hat{x}_i\|^2 \right\} \\
 & \geq \frac{\zeta}{2} \left\{ \|x_3^k - x_3^{k+1}\|^2 + \sum_{i=2}^3 \|x_i^{k+1} - \hat{x}_i\|^2 \right\} \\
 & \geq \frac{\zeta}{2} \left\{ \frac{3}{10\beta^2} \frac{10}{3} \|x_3^k - x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^3 \frac{1}{\beta^2 \|A_i^*A_i\|} \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 \right\} \\
 & \geq \frac{\zeta}{2\beta^2} \min \left\{ \frac{1}{\|A_2^*A_2\|}, \frac{3}{10\|A_3^*A_3\|} \right\} \left\{ \frac{10}{3} \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^3 \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 \right\} \\
 & = \tau' \left\{ \frac{10}{3} \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^3 \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 \right\}, \tag{6.6}
 \end{aligned}$$

where $\tau' = \frac{\zeta}{2\beta^2} \min \left\{ \frac{1}{\|A_2^*A_2\|}, \frac{3}{10\|A_3^*A_3\|} \right\}$.

Then, it follows from (6.5) and (6.6) that

$$\begin{aligned}
 & \zeta \left\{ \|A_2x_2^k - A_2x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 + \sum_{i=2}^3 \|x_i^{k+1} - \hat{x}_i\|^2 \right\} \\
 & \geq \tau \|z^{k+1} - \hat{z}\|^2 + \tau' \left\{ \frac{10}{3} \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^3 \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 \right\} \\
 & \geq \min \{ \tau, \tau' \} \left\{ \|z^{k+1} - \hat{z}\|^2 + \frac{10}{3} \|A_3x_3^k - A_3x_3^{k+1}\|_{\beta^2}^2 + \sum_{i=2}^3 \|A_i x_i^{k+1} - A_i \hat{x}_i\|_{\beta^2}^2 \right\} \\
 & = \sigma' \left\{ \|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \right\}, \tag{6.7}
 \end{aligned}$$

where Q is defined in (2.5) and its positive definiteness is ensured by Assumption 6.1 and $\sigma' = \min \{ \tau, \tau' \}$. Finally, combining (6.1) and (6.7), we obtain

$$\begin{aligned}
 & \|\mathbf{w}^{k+1} - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3x_3^{k+1} - A_3x_3^k\|_{\beta^2}^2 \\
 & \leq \frac{1}{1 + \sigma'} \left\{ \|\mathbf{w}^k - \hat{\mathbf{w}}\|_Q^2 + \frac{10}{3} \|A_3x_3^k - A_3x_3^{k-1}\|_{\beta^2}^2 \right\},
 \end{aligned}$$

which implies the linear convergence rate of the generated sequence under the Q -norm. The proof is complete. □

6.2 Linear convergence under error bound conditions

As mentioned, the linear convergence of the scheme (1.2a)–(1.2d) has been well investigated under some error bound conditions in [21], which has also inspired the work [15] for deriving the linear convergence of the original two-block ADMM scheme in the quadratic programming context, which was further extended to the general case that the objective functions possess polyhedra subdifferentials [35]. In this subsection, we provide an alternative proof for the globally linear convergence of (1.2a)–(1.2d) under some error bound conditions. The new conditions differ from Assumption 6.1 in that no smoothness or strong convexity assumption on the objective function is assumed.

Assumption 6.2 Let $\mathcal{U} := \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{Z}$ and let \tilde{U} be a bounded set in \mathcal{U} . Then there exists $\eta > 0$ such that

$$\text{dist}(u, S) \leq \eta \text{dist}(0, F(u)), \quad \forall u \in \tilde{U}, \tag{6.8}$$

where $u := (x_1, x_2, x_3, z)$, and

$$F(u) := \begin{pmatrix} \partial\theta_1(x_1) - A_1^*z \\ \partial\theta_2(x_2) - A_2^*z \\ \partial\theta_3(x_3) - A_3^*z \\ \sum_{i=1}^3 A_i x_i - b \end{pmatrix}. \tag{6.9}$$

Obviously, using the definition in (6.9) and the optimality condition (2.2a)–(2.2d), we see that a point $u \in S$ if and only if $0 \in F(u)$. Using the result [36, Theorem 3.3], it is clear that the error bound condition (6.8) holds if F is a piecewise linear multifunction, which is true for the case where θ_i 's are, e.g., the l_1 - or l_2 -terms. For the definition of piecewise linear multifunction or polyhedral multifunction, see, e.g., [31, 36].

Recall that under Assumption 3.1, we show that there exists a constant $\zeta = \min\{\frac{1}{40}, \varrho_1\} > 0$ such that

$$\begin{aligned} & \|w^{k+1} - \hat{w}\|_Q^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\ & \leq \|w^k - \hat{w}\|_Q^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \zeta \left\{ \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 \right. \\ & \quad \left. + \|z^k - z^{k+1}\|^2 + \|x_3^{k+1} - \hat{x}_3\|^2 \right\}, \end{aligned} \tag{6.10}$$

where $w := (x_2, x_3, z)$ and Q is defined by (2.5), see (3.18). To prove the linear convergence rate of the scheme (1.2a)–(1.2d) under Assumption 6.2, it follows from (6.10) that we only need to bound the term $\|w^{k+1} - \hat{w}\|^2$ because of the minus term in (6.10). Then, under Assumption 6.2, the key is to find a bound of the term $\text{dist}(0, F(u^{k+1}))$. Therefore, we first prove the following lemma.

Lemma 6.2 *Let $F(u)$ be defined in (6.9) and $\{u^k := (x_1^k, x_2^k, x_3^k, z^k)\}$ be the sequence generated by the direct extension of ADMM (1.2a)–(1.2d). Then, there exists a constant $\sigma > 0$ such that*

$$\text{dist}^2(0, F(u^{k+1})) \leq \sigma \left\{ \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 \right\}, \tag{6.11}$$

where

$$\sigma = \max \left\{ \frac{1}{\beta^2}, \beta^2(2\|A_1\|^2 + \|A_2\|^2)\|A_3\|^2, 2\|A_1\|^2 \right\}. \tag{6.12}$$

Proof It follows from the optimality condition of the x_1 -subproblem that

$$0 \in \partial\theta_1(x_1^{k+1}) - A_1^*z^k + \beta A_1^* (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b),$$

which can be alternatively written as (using (1.2d))

$$\beta A_1^* \left[A_2 \left(x_2^{k+1} - x_2^k \right) + A_3 \left(x_3^{k+1} - x_3^k \right) \right] \in \partial \theta_1 \left(x_1^{k+1} \right) - A_1^* z^{k+1}.$$

Consequently, we have

$$\begin{aligned} \text{dist}^2(0, \partial \theta_1(x_1^{k+1}) - A_1^* z^{k+1}) &\leq \|\beta A_1^* [A_2(x_2^{k+1} - x_2^k) + A_3(x_3^{k+1} - x_3^k)]\|^2 \\ &\leq 2\beta^2 \|A_1\|^2 \{ \|A_2(x_2^{k+1} - x_2^k)\|^2 + \|A_3(x_3^{k+1} - x_3^k)\|^2 \}. \end{aligned} \quad (6.13)$$

Moreover, the optimality condition of the x_2 -subproblem and (1.2d) indicate that

$$\beta A_2^* A_3(x_3^{k+1} - x_3^k) \in \partial \theta_2(x_2^{k+1}) - A_2^* z^{k+1}.$$

Hence, we obtain

$$\text{dist}^2(0, \partial \theta_2(x_2^{k+1}) - A_2^* z^{k+1}) \leq \|\beta A_2^* A_3(x_3^{k+1} - x_3^k)\|^2 \leq \beta^2 \|A_2\|^2 \|A_3(x_3^{k+1} - x_3^k)\|^2. \quad (6.14)$$

Further, the optimality condition of the x_3 -subproblem and (1.2d) indicate that

$$0 \in \partial \theta_3(x_3^{k+1}) - A_3^* z^{k+1},$$

which means

$$\text{dist}^2(0, \partial \theta_3(x_3^{k+1}) - A_3^* z^{k+1}) = 0. \quad (6.15)$$

In addition, it follows from (1.2d) that

$$\frac{1}{\beta} (z^k - z^{k+1}) = A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b. \quad (6.16)$$

Finally, combining (6.13)–(6.16), we obtain

$$\begin{aligned} \text{dist}^2(0, F(u^{k+1})) &= \text{dist}^2(0, \partial \theta_1(x_1^{k+1}) - A_1^* z^{k+1}) + \text{dist}^2(0, \partial \theta_2(x_2^{k+1}) - A_2^* z^{k+1}) \\ &\quad + \text{dist}^2(0, \partial \theta_3(x_3^{k+1}) - A_3^* z^{k+1}) + \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2 \\ &\leq 2\beta^2 \|A_1\|^2 \|A_2(x_2^{k+1} - x_2^k)\|^2 + \frac{1}{\beta^2} \|z^{k+1} - z^k\|^2 \\ &\quad \beta^2 (2\|A_1\|^2 + \|A_2\|^2) \|A_3(x_3^{k+1} - x_3^k)\|^2. \end{aligned} \quad (6.17)$$

Therefore, the assertion (6.11) follows immediately from (6.17) and the definition of σ in (6.12). \square

Now, we show the globally linear convergence of the direct extension of ADMM (1.2a)–(1.2d) under Assumptions 3.1 and 6.2.

Theorem 6.2 *Suppose Assumptions 3.1 and 6.2 hold. Then the sequence $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ generated by the direct extension of ADMM (1.2a)–(1.2d) with $\beta \in \left(0, \frac{6\mu_3}{17\|A_3^*A_3\|}\right)$ converges globally linearly to a saddle point in S .*

Proof Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{z})$ be a saddle point in S . According to Theorem 3.1, the whole sequence $\{(x_1^k, x_2^k, x_3^k, z^k)\}$ converges to a saddle point in S . Thus it is bounded. Because of Assumption 6.2, we know that

$$\text{dist}(u^{k+1}, S) \leq \eta \text{dist}(0, F(u^{k+1})), \quad \forall k \geq 0, \tag{6.18}$$

which, together with (6.11), implies that

$$\text{dist}^2(u^{k+1}, S) \leq \eta^2 \sigma \left\{ \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 + \|z^k - z^{k+1}\|^2 \right\}. \tag{6.19}$$

Let us define

$$M = \begin{pmatrix} 0 & 0 \\ 0 & Q \end{pmatrix},$$

with Q given in (2.5). Then, combining (6.19) with (6.10), we have

$$\begin{aligned} & \|u^{k+1} - \hat{u}\|_M^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\ & \leq \|u^k - \hat{u}\|_M^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \zeta \left\{ \|A_2 x_2^k - A_2 x_2^{k+1}\|_{\beta^2}^2 + \|x_3^k - x_3^{k+1}\|^2 \right. \\ & \quad \left. + \|z^k - z^{k+1}\|^2 \right\} \\ & \leq \|u^k - \hat{u}\|_M^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \frac{\zeta}{3} \|x_3^k - x_3^{k+1}\|^2 - \frac{2\zeta}{3\sigma\eta^2} \text{dist}^2(u^{k+1}, S). \end{aligned} \tag{6.20}$$

It follows from (6.18) that

$$\text{dist}_M^2(u^{k+1}, S) \leq \eta^2 \|M\| \text{dist}^2(0, F(u^{k+1})), \quad \forall k \geq 0.$$

Let $u^* \in S$ such that $\text{dist}_M(u^k, S) = \|u^k - u^*\|_M$. Since (6.20) holds for any saddle point in S , we have

$$\begin{aligned} & \|u^{k+1} - u^*\|_M^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\ & \leq \text{dist}_M(u^k, S) + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \frac{\zeta}{3} \|x_3^k - x_3^{k+1}\|^2 - \frac{2\zeta}{3\sigma\eta^2} \text{dist}^2(u^{k+1}, S). \end{aligned} \tag{6.21}$$

Setting

$$\bar{\mu} := \min \left\{ \frac{\zeta}{10\beta^2 \|A_3\|^2}, \frac{\zeta}{3\sigma\eta^2 \|M\|} \right\} \quad \text{and} \quad \gamma := \frac{\zeta}{3(1 + \bar{\mu})\sigma\eta^2},$$

we obtain the following inequality:

$$\begin{aligned} & (1 + \bar{\mu}) \left\{ \text{dist}_M^2(u^{k+1}, S) + \text{dist}_{\gamma I}^2(u^{k+1}, S) + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \right\} \\ & \leq \|u^{k+1} - u^*\|_M^2 + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \\ & \quad + \bar{\mu} \left\{ \text{dist}_M^2(u^{k+1}, S) + \frac{10}{3} \|A_3 x_3^{k+1} - A_3 x_3^k\|_{\beta^2}^2 \right\} \\ & \quad + (1 + \bar{\mu})\gamma \text{dist}^2(u^{k+1}, S) \\ & \leq \|u^k - u^*\|_M^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 - \left(\frac{\zeta}{3} - \frac{10\beta^2 \|A_3\|^2 \bar{\mu}}{3} \right) \|x_3^k - x_3^{k+1}\|^2 \\ & \quad - \left(\frac{\zeta}{3\sigma\eta^2} - \bar{\mu} \|M\| \right) \text{dist}^2(u^{k+1}, S) \\ & \leq \|u^k - u^*\|_M^2 + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 \\ & = \text{dist}_M^2(u^k, S) + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2 \\ & \leq \text{dist}_{\gamma I + M}^2(u^k, S) + \frac{10}{3} \|A_3 x_3^k - A_3 x_3^{k-1}\|_{\beta^2}^2, \end{aligned} \tag{6.22}$$

where the second inequality follows from (6.21) and the third one from the definition of $\bar{\mu}$. Thus, the inequality (6.22) implies the linear convergence rate of the sequence generated by the direct extension of ADMM measured by the sum of the distance under the $(\gamma I + M)$ -norm. The proof is complete. \square

Remark 6.1 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]; and it has been proved in [33] that the augmented Lagrangian method is an application of the proximal point algorithm (PPA) in [25, 26]. Since the convergence rate of PPA is known to be only linear in [33], 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. We show that the convergence of the direct extension of ADMM can be ensured if one function in the objective is strongly convex, the penalty parameter is appropriately restricted and some assumptions on the operators in the constraints hold. This convergence result can well explain why the direct extension of ADMM is convergent for some applications. We also establish the worst-case convergence rate measured by the iteration complexity and the globally linear convergence rate in asymptotical sense under additional assumptions. This is a deeper study on the convergence of the direct extension of ADMM that may help us better understand its theoretical aspects. We would like to mention that instead of the original scheme (1.2a)–(1.2d), we can study the scheme

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

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.

Acknowledgements Xingju Cai is supported by the NSFC Grant 11401315 and the NSF from Jiangsu province BK20140914. Deren Han is supported by a project funded by PAPD of Jiangsu Higher Education Institutions and the NSFC Grants 11371197 and 11431002. Xiaoming Yuan was supported by the NSFC/RGC Joint Research Scheme: N_PolyU504/14.

References

1. Boley, D.: Local linear convergence of ADMM on quadratic or linear programs. *SIAM J. Optim.* **23**(4), 2183–2207 (2013)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. In: Jordan, M. (ed.) *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122 (2011)
3. Chamboulle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
4. Chandrasekaran, V., Parrilo, P.A., Willsky, A.S.: Latent variable graphical model selection via convex optimization. *Ann. Stat.* **40**, 1935–1967 (2012)
5. 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.* **155**, 57–79 (2016)

6. Chen, C.H., Shen, Y., You, Y.F.: On the convergence analysis of the alternating direction method of multipliers with three blocks. *Abstr. Appl. Anal.* **2013** (2013), Article ID 183961
7. Corman, E., Yuan, X.M.: A generalized proximal point algorithm and its convergence rate. *SIAM J. Optim.* **24**(4), 1614–1638 (2014)
8. Deng, W., Yin, W.T.: On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.* **66**(3), 889–916 (2016)
9. Deng, W., Lai, M.J., Peng, Z.M., Yin, W.T.: Parallel multi-block ADMM with $o(1/k)$ convergence. *Manuscript* (2014)
10. Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives. *Pac. J. Optim.* **11**, 619–644 (2015)
11. Glowinski, R.: *Numerical Methods for Nonlinear Variational Problems*. Springer, New York (1984)
12. Glowinski, R., Marrocco, A.: Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue Fr. Autom. Inf. Rech. Opér., Anal. Numér.* **2**, 41–76 (1975)
13. Glowinski, R.: On alternating direction methods of multipliers: a historical perspective. *Model. Simul. Optim. Sci. Technol. Comput. Methods Appl. Sci.* **34**, 59–82 (2014)
14. Han, D.R., Yuan, X.M.: A note on the alternating direction method of multipliers. *J. Optim. Theory Appl.* **155**, 227–238 (2013)
15. Han, D.R., Yuan, X.M.: Local linear convergence of the alternating direction method of multipliers for quadratic programs. *SIAM J. Numer. Anal.* **51**, 3446–3457 (2013)
16. 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. Comput.* **83**, 2263–2291 (2014)
17. He, B.S., Tao, M., Yuan, X.M.: Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM J. Optim.* **22**, 313–340 (2012)
18. 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. *Math. Oper. Res.* (under revision)
19. He, B.S., Tao, M., Yuan, X.M.: A splitting method for separable convex programming. *IMA J. Numer. Anal.* **35**, 394–426 (2015)
20. Hestenes, M.R.: Multiplier and gradient methods. *J. Optim. Theory Appl.* **4**, 303–320 (1969)
21. Hong, M.Y., Luo, Z.Q.: On the linear convergence of alternating direction method of multipliers. *Math. Program.* (to appear)
22. 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.* **32**, 1550024 (2015). 19 pages
23. Lin, T.Y., Ma, S.Q., Zhang, S.Z.: On the sublinear convergence rate of multi-block ADMM. *J. Oper. Res. Soc. China* **3**, 251–274 (2015)
24. Lin, T.Y., Ma, S.Q., Zhang, S.Z.: On the global linear convergence of the ADMM with multi-block variables. *SIAM J. Optim.* **25**, 1478–1497 (2015)
25. Martinet, B.: Regularization d'inequations variationnelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationnelle* **4**, 154–159 (1970)
26. Moreau, J.J.: Proximité et dualité dans un espace Hilbertien. *Bull. Soc. Math. France* **93**, 273–299 (1965)
27. McLachlan, G.J.: *Discriminant Analysis and Statistical Pattern Recognition*, vol. 544. Wiley, New York (2004)
28. Mohan, K., London, P., Fazel, M., Witten, D., Lee, S.: Node-based learning of multiple gaussian graphical models. *J. Mach. Learn. Res.* **15**, 445–488 (2014)
29. Peng, Y.G., Ganesh, A., Wright, J., Xu, W.L., Ma, Y.: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2233–2246 (2012)
30. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) *Optimization*, pp. 283–298. Academic Press, New York (1969)
31. Robinson, S.M.: Some continuity properties of polyhedral multifunctions. *Math. Program. Stud.* **14**, 206–214 (1981)
32. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
33. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898 (1976)

34. Tao, M., Yuan, X.M.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.* **21**, 57–81 (2011)
35. 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.* **54**(2), 625–640 (2016)
36. Zheng, X.Y., Ng, K.F.: Metric subregularity of piecewise linear multifunctions and applications to piecewise linear multiobjective optimization. *SIAM J. Optim.* **24**, 154–174 (2014)
37. Zhou, X., Yang, C., Yu, W.: Moving object detection by detecting contiguous outliers in the Low-Rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 597–610 (2013)
38. Zhou, Z., Li, X., Wright, J., Candes, E.J., Ma, Y.: Stable principal component pursuit. In: *Proceedings of International Symposium on Information Theory*. Austin (2010)