Convergence analysis of the direct extension of ADMM for multiple-block separable convex minimization

Min Tao¹ · Xiaoming Yuan²

Abstract Recently, the alternating direction method of multipliers (ADMM) has found many efficient applications in various areas; and it has been shown that the convergence is not guaranteed when it is directly extended to the multiple-block case of separable convex minimization problems where there are $m \geq 3$ functions without coupled variables in the objective. This fact has given great impetus to investigate various conditions on both the model and the algorithm’s parameter that can ensure the convergence of the direct extension of ADMM (abbreviated as “e-ADMM”). Despite some results under very strong conditions (e.g., at least $(m−1)$ functions should be strongly convex) that are applicable to the generic case with a general $m$, some others concentrate on the special case of $m = 3$ under the relatively milder condition that only one function is assumed to be strongly convex. We focus on extending the convergence analysis from the case of $m = 3$ to the more general case of $m \geq 3$. That is, we show the convergence of e-ADMM for the case of $m \geq 3$ with the assumption

Communicated by: Karsten Urban

M. Tao: This author was supported by the Natural Science Foundation of China: NSFC-11301280 and the sponsorship of Jiangsu overseas research and training program for university prominent young and middle-aged teachers and presidents.

X. Yuan: This author was supported by the General Research Fund from Hong Kong Research Grants Council: 12313516.

Min Tao
taom@nju.edu.cn

Xiaoming Yuan
xmyuan@hkbu.edu.hk

¹ Department of Mathematics, National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China

² Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
of only \((m - 2)\) functions being strongly convex; and establish its convergence rates in different scenarios such as the worst-case convergence rates measured by iteration complexity and the globally linear convergence rate under stronger assumptions. Thus the convergence of e-ADMM for the general case of \(m \geq 4\) is proved; this result seems to be still unknown even though it is intuitive given the known result of the case of \(m = 3\). Even for the special case of \(m = 3\), our convergence results turn out to be more general than the existing results that are derived specifically for the case of \(m = 3\).

**Keywords**  
Convex programming · Alternating direction method of multipliers · Multiple-block separable models · Convergence analysis

**Mathematics Subject Classification (2010)**  
90C25 · 90C33 · 65K05

### 1 Introduction

We consider a canonical convex minimization model with separable structure and linear constraints, whose objective function is the sum of \(m\) functions without coupled variables:

\[
\min \left\{ \sum_{i=1}^{m} \theta_i(x_i) \mid \sum_{i=1}^{m} A_i x_i = b, \ x_i \in \mathcal{X}_i, \ i = 1, 2, \ldots, m \right\}, \tag{1.1}
\]

where \(\theta_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R} (i = 1, 2, \ldots, m)\) are closed proper convex functions (not necessarily smooth); \(A_i \in \mathbb{R}^{l \times n_i} (i = 1, 2, \ldots, m)\); \(\mathcal{X}_i \subseteq \mathbb{R}^{n_i} (i = 1, 2, \ldots, m)\) are nonempty closed convex sets; \(b \in \mathbb{R}^{l}\) and \(\sum_{i=1}^{m} n_i = n\). The solution set of (1.1) is assumed to be nonempty throughout our discussion. To mention a few applications of the model (1.1) with \(m \geq 3\), we refer to, e.g., the robust principal component analysis model with noisy and missing data in [25], the moving objective detection problem in [27], the doubly nonnegative semidefinite programming in [24] and the joint day-ahead power procurement and load scheduling problem in [26].

Let the augmented Lagrangian function of (1.1) be

\[
L_{\beta}(x_1, x_2, \ldots, x_m, z) := \sum_{i=1}^{m} \theta_i(x_i) - z^\top \left( \sum_{i=1}^{m} A_i x_i - b \right) + \frac{\beta}{2} \left\| \sum_{i=1}^{m} A_i x_i - b \right\|_2^2
\]

with \(z \in \mathbb{R}^{l}\) the Lagrange multiplier and \(\beta > 0\) the penalty parameter. We focus on the following iterative scheme with \(m \geq 3\):

\[
\begin{align*}
&x_1^{k+1} = \arg \min \{ L_{\beta}(x_1, x_2^k, \ldots, x_m^k, z) \mid x_1 \in \mathcal{X}_1 \}, \quad \quad \tag{1.2a} \\
x_2^{k+1} = \arg \min \{ L_{\beta}(x_1^{k+1}, x_2, \ldots, x_m^k, z) \mid x_2 \in \mathcal{X}_2 \}, \quad \quad \tag{1.2b} \\
&\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \_quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad
which starts from a given iterate \((x_2^k, \ldots, x_m^k, z^k) \in X_2 \times \ldots \times X_m \times \mathbb{R}^l\). When \(m = 2\), the scheme (1.2) reduces to the alternating direction method of multipliers (ADMM) originally proposed in [13]. The convergence of ADMM has been well studied in the literature, see [10, 11, 16, 17]. Recently, ADMM has found many applications in a wide range of areas; we refer to, e.g., [3, 8, 14] for its review. The scheme (1.2) can be regarded as a direct extension of the ADMM (abbreviated as “e-ADMM”) for solving (1.1). Despite the inertia in algorithmic design and the numerical efficiency in empirical implementation (e.g., in [22, 25, 27]), it was shown in [6] that the e-ADMM (1.2) is not necessarily convergent when \(m = 3\). By mathematical induction, it is easy to prove that the same conclusion for the general case of \(m \geq 3\). This rather surprising fact has immediately given impetus to investigate various conditions to ensure the convergence of the e-ADMM (1.2) with \(m \geq 3\).

In the literature, there are some results for the generic case with a general \(m \geq 3\), it was shown in [15] that the e-ADMM (1.2) is convergent if all the functions \(\theta_i\) are strongly convex. In [20], the global convergence of (1.2) was shown under the conditions that \((m - 1)\) of the functions \(\theta_i\) are strongly convex. In [21], the linear convergence of (1.2) was shown under the condition that at least \((m - 1)\) of the functions \(\theta_i\) are strongly convex together with other assumptions such as \(\nabla \theta_i\) are Lipschitz continuous and \(A_i\) are full row/column rank. In addition, the authors of [18] showed that the linear convergence can be guaranteed if the step size of the last step (1.2e) for updating the multiplier \(z^{k+1}\) is shrunk by a sufficiently small factor and a certain error bound condition is satisfied.

For the special case of \(m = 3\), there is a richer set of literature. The first one is [5], which shows the convergence of (1.2) under the assumption that two functions of \(\theta_i\) are strongly convex. Still requiring the strong convexity of two functions, the work [20] proves some refined convergence results such as the \(O(1/t)\) ergodic convergence rate and \(o(1/t)\) non-ergodic convergence rate measured by the iteration complexity, where \(t\) is the iteration number. Later, the results in [5, 20] were improved in [4, 19], in which the convergence of (1.2) was obtained with only one strongly convex function for the case \(m = 3\). According to the results in [6], the strong convexity of at least one function seems minimal for the special case of \(m = 3\) of (1.1) to ensure the convergence of the e-ADMM (1.2); and the work in [4, 19] verifies this conclusion positively.

Given the mentioned results for the case of \(m = 3\), by analogy, can we claim that we need \((m - 2)\) strongly convex functions amid \(\theta_i\)’s to ensure the convergence of the e-ADMM (1.2) for the generic case with a general \(m\) that can be larger than 3? Our main goal in this paper is to answer this question affirmatively. As we shall show, though the answer seems to be intuitive because of the known result for the special case of \(m = 3\), technically the extension from \(m = 3\) to \(m \geq 3\) is highly nontrivial. One may ask if we can only require \((m - 3)\) of the functions \(\theta_i\) to be strongly convex to ensure the convergence of (1.2) when \(m \geq 4\). In Section 6.2, we give an example to show that in general it is not guaranteed and thus verify the rationale of considering the convergence of (1.2) with \(m \geq 3\) with the assumption of \((m - 2)\) functions being strongly convex. We also refer to the recent work [7] for a more general study in the operator context and it includes the case of (1.1) with \(m - 2\) strongly convex.
functions as a special case. But the resulting algorithm (see Algorithm 9 in [7]) for this special case is not the same as the e-ADMM (1.2) under our consideration, e.g., for the subproblems accompanying the strongly convex functions, their objectives do not involve augmented Lagrangian terms and they are solved in parallel.

In addition to the strong convexity of some or all the functions in the objective of (1.1), it is worthwhile to mention that the penalty parameter $\beta$ in (1.2) should be appropriately restricted to theoretically ensure the convergence, see, e.g., all the work [4, 5, 15, 19–21]. Note that such a requirement of the penalty parameter is usually conservative because it is used to sufficiently ensure the convergence. In numerical implementation, it can be appropriately relaxed to result in faster convergence. According to Theorem 4.1 in [6], even all the functions $\theta_i$’s in the objective of (1.1) are strongly convex, the scheme (1.2) with $m = 3$ may be divergent if the penalty parameter $\beta$ is not well restricted. Similarly, in Section 6.3, we show that the scheme (1.2) could be divergent even when $(m - 2)$ functions are strongly convex while the $\beta$ is not restricted appropriately. Therefore, to discuss the convergence of e-ADMM (1.2) with $m \geq 3$ for the more difficult case where only some of the functions $\theta_i$’s are assumed to be strongly convex, we shall also restrict the penalty parameter into certain intervals. Indeed, as we shall elucidate, the range of $\beta$, which is eligible to the case with a generic $m$, is even larger than those in [4, 19] when it reduces to the special case of $m = 3$. That is, we shall prove the convergence for the e-ADMM (1.2) by requiring only $(m - 2)$ strongly convex functions and a larger range of $\beta$ for $m \geq 3$. Moreover, we shall establish the worst-case $O(1/t)$ convergence rate in the ergodic sense for $m \geq 3$; and explore some stronger conditions that can ensure the globally linear convergence for $m \geq 3$. Thus, compared with existing work in the same category such as [4, 5, 15, 19–21], the convergence results in this paper are more general and they are proved under weaker conditions.

The rest of this paper is organized as follows. We summarize some notations, present the assumptions for future discussion and recall some known results in Section 2. In Section 3, we prove the convergence of e-ADMM (1.2) under certain assumptions; this is the main result of this paper. Then, we establish the worst-case convergence rate measured by the iteration complexity in Section 4. In Section 5, we show that the e-ADMM (1.2) can be guaranteed to be globally linear convergent if further conditions are posed. In Section 6, we show that the convergence of e-ADMM (1.2) may not be guaranteed if there are no appropriate assumptions on the model (1.1) or the penalty parameter $\beta$ in (1.2). Some examples are also constructed. Finally, we draw some conclusions in Section 7.

## 2 Preliminaries

In this section, we define some notations to be used; present some assumptions on the model (1.1) under which our convergence analysis will be conducted; and show the optimality condition of the model (1.1) in the variational inequality context.
2.1 Notation

The domain of a function $f$ is denoted by $\text{dom}(f)$ and the set of all relative interior points of a given nonempty convex set $\Omega$ by $\text{ri}(\Omega)$. Given a vector $x \in \mathbb{R}^n$, the notation $x_{[i:j]}$ ($1 \leq i \leq j \leq n$) denotes the subvector of $x$ consisting of the $i$-th up to the $j$-th entries of $x$. If $i = j$, $x_{[i]}$ just denotes the $i$-th entry of $x$. For any given vector $x$ and a symmetric positive semi-definite matrix $M$ with appropriate dimensionality, we use $\|x\|^2_M$ to denote $x^\top M x$. For a symmetric matrix $M$, let $\|M\| := \sqrt{\|M^\top M\|}$ and $\rho(M)$ to denote its spectral radius, i.e., the maximal absolute value of its eigenvalues. We use the notation $\text{diag}(\cdot)$ to denote a diagonal matrix. A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is strongly convex with modulus $\mu > 0$ if it satisfies

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\mu}{2} t(1-t)\|x-y\|^2, \quad \forall x, y \in \mathbb{R}^n, \quad (2.1)$$

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

Then, based on the coefficient matrices $A_i$ in (1.1) and the penalty parameter $\beta$ in (1.2), we define some matrices to simplify our notations in later analysis. More specifically, for $m \geq 3$, let the block triangular matrices $M$, $N$ and block diagonal matrix $Q$ be respectively defined as:

$$M = \begin{pmatrix}
0 & A_1 \top & A_2 & \cdots & A_m \top & A_m & 0 \\
A_1 & A_2 & A_3 & \cdots & A_m & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & A_{m-1} \top & A_m & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 
\end{pmatrix}, \quad N = \begin{pmatrix}
0 & 0 & 0 & \cdots & 0 \\
0 & A_2 \top & A_3 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & A_{m} \top & A_m & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 
\end{pmatrix},$$

and

$$Q := \begin{pmatrix}
\beta A_2 \top & A_2 & \cdots & 0 & 0 \\
0 & \beta A_3 \top & A_3 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \cdots & \beta A_m \top & A_m \\
0 & \cdots & \cdots & 0 & \frac{1}{\beta} I 
\end{pmatrix}, \quad (2.2)$$

Note that both $M$ and $N$ are in the space $(n + l) \times (n + l)$; and $Q$ in $(\sum_{i=2}^{m} n_i + l) \times (\sum_{i=2}^{m} n_i + l)$. Also, the matrix $Q$ defined in (2.3) is positive definite if $A_i$ ($i = 2, \ldots, m$) are all assumed to be full column rank and $\beta > 0$.

2.2 Assumptions

Then, we present the assumptions on the model (1.1) to conduct the convergence analysis for the e-ADMM (1.2) with a general $m \geq 3$. 

 Springer
**Assumption 2.1** In (1.1) with $m \geq 3$, the functions $\theta_1$ and $\theta_2$ are convex; the functions $\theta_3, \ldots, \theta_m$ are strongly convex with the modulus $\mu_i > 0$ ($i = 3, \ldots, m$); $A_i$ ($i = 1, \ldots, m$) are full column rank matrices.

**Assumption 2.2** There exists $u' = (x'_1, \ldots, x'_m) \in ri(dom(\theta_1) \times dom(\theta_2) \times \cdots \times dom(\theta_m)) \cap F$, where

$$F := \left\{ u = (x_1, \ldots, x_m) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \mid \sum_{i=1}^{m} A_i x_i = b \right\}.$$ 

Note that both $\theta_1$ and $\theta_2$ are assumed to be convex; but we also say that they both satisfy the inequality (2.1) with $\mu = 0$ as long as there is no confusion. This helps us present the analysis in a unified notation. We first present an elementary lemma and its proof is omitted.

**Lemma 2.1** Let $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be strongly convex with the modulus $\mu$ and $X$ be a closed convex set in $D$. If $x^*$ is a minimizer of $f$ on $X$, then we have $f(x^*) \leq f(x) - \frac{\mu}{2} \|x^* - x\|^2$ for all $x \in X$.

### 2.3 Optimality condition of (1.1) as a variational inequality

In the following, we characterize the optimality condition of the model (1.1) as a variational inequality. The variational inequality representation plays a crucial role in our convergence analysis to be conducted.

First, let $W := \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m \times \mathcal{Z}$ and the Lagrangian function of (1.1) be

$$L(x_1, x_2, \ldots, x_m, z) := \sum_{i=1}^{m} \theta_i(x_i) - z^\top \left( \sum_{i=1}^{m} A_i x_i - b \right), \quad (2.4)$$

with $z$ the Lagrange multiplier. Under Assumption 2.2, it follows from [23, Corollary 28.2.2] and [23, Corollary 28.3.1] that the set of saddle points of $L(x_1, x_2, \ldots, x_m, z)$, denoted by $W^*$, is nonempty due to the nonempty assumption on the solution set of (1.1). Then, solving (1.1) amounts to finding a saddle point of $L(x_1, x_2, \ldots, x_m, z)$. Hence, according to its definition, finding a saddle point of $L(x_1, x_2, \ldots, x_m, z)$ can be further explained as solving the following system:

\[
\begin{align*}
  x_1^* &= \arg\min_{x_1 \in \mathcal{X}_1} L(x_1, x_2^*, x_3^*, \ldots, x_m^*, z^*), \\
  x_2^* &= \arg\min_{x_2 \in \mathcal{X}_2} L(x_1^*, x_2, x_3^*, \ldots, x_m^*, z^*), \\
  x_3^* &= \arg\min_{x_3 \in \mathcal{X}_3} L(x_1^*, x_2^*, x_3, \ldots, x_m^*, z^*), \\
  & \quad \vdots \\
  x_m^* &= \arg\min_{x_m \in \mathcal{X}_m} L(x_1^*, x_2^*, x_3^*, \ldots, x_m, z^*), \\
  z^* &= \arg\max_{z} L(x_1^*, x_2^*, x_3^*, \ldots, x_m^*, z).
\end{align*}
\]  

(2.5a) (2.5b) (2.5c) (2.5d) (2.5e)
Invoking Lemma 2.1 and Assumption 2.1, for \( i = 3, \ldots, m \), the \( x_i \)-subproblem in (2.5) is

\[
\theta_i(x_i) - (z^*)^\top A_i x_i \geq \theta_i(x_i^*) - (z^*)^\top (A_i x_i^*) + \frac{\mu_i}{2} \| x_i^* - x_i \|^2, \quad \forall x_i \in X_i.
\]

Obviously, the above inequality is also true for \( i = 1, 2 \) if we set \( \mu_i = 0 \) for \( i = 1, 2 \). Therefore, the optimality condition of the model (1.1) can be characterized by finding \( w^* \in \mathcal{W}^* \) such that:

\[
\begin{align*}
\theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^\top (-A_1^\top z^*) &\geq 0, \\
\theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^\top (-A_2^\top z^*) &\geq 0, \\
& \quad \quad \cdots \\
\theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^\top (-A_i^\top z^*) &\geq \frac{\mu_i}{2} \| x_i - x_i^* \|^2, \\
& \quad \quad \cdots \\
\theta_m(x_m) - \theta_m(x_m^*) + (x_m - x_m^*)^\top (-A_m^\top z^*) &\geq \frac{\mu_m}{2} \| x_m - x_m^* \|^2, \\
\sum_{i=1}^m A_i x_i^* - b &= 0,
\end{align*}
\]

(2.6)

where \( \mu_i \) is the strongly convex modulus of \( \theta_i \) for \( i = 3, \ldots, m \). More compactly, the system (2.6) can be written as the variational inequality:

\[
\text{VI} (\mathcal{W}, F, \theta) \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \geq \sum_{i=3}^m \frac{\mu_i}{2} \| x_i - x_i^* \|^2, \quad \forall w \in \mathcal{W},
\]

(2.7a)

where

\[
u = \left( \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array} \right), \quad w = \left( \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \\ z \end{array} \right) \quad \text{and} \quad \theta(u) = \sum_{i=1}^m \theta_i(x_i), \quad F(w) = \left( \begin{array}{c} -A_1^\top z \\ -A_2^\top z \\ \vdots \\ -A_m^\top z \\ \sum_{i=1}^m A_i x_i - b \end{array} \right).
\]

(2.7b)

Note that \( u \) collects all the primal variables in (1.1) and it is a sub-vector of \( w \). Since the variable \( x_1 \) is not involved in the iteration of the e-ADMM (1.2), we denote by

\[
v = (x_2, x_3, \ldots, x_m, z)
\]

all the primal and dual variables that are essentially involved in the iteration (1.2). Moreover, the solution set of \( \text{VI} (\mathcal{W}, F, \theta) \), i.e., \( \mathcal{W}^* \), is also convex due to Theorem 2.3.5 in [9]. Accordingly, we also use the notation

\[
\mathcal{V}^* = \{ (x_2^*, \ldots, x_m^*, z^*) \mid (x_1^*, x_2^*, \ldots, x_m^*, z^*) \in \mathcal{W}^* \}.
\]

2.4 Several elementary lemmas

In the following, we present several elementary lemmas that will be frequently used in the analysis to come.
Lemma 2.2 [12] Let \( a \) and \( b \) be two vectors with the same dimension. Then, we have
\[
\|a + b\|^2 \leq 2(\|a\|^2 + \|b\|^2).
\]

Lemma 2.3 [12] Let \( a \) and \( b \) be two vectors with the same dimension. Then, for any positive scalar \( \kappa \), it holds that
\[
a^\top b \geq -\frac{1}{2} \left( \frac{1}{\kappa} \|a\|^2 + \kappa \|b\|^2 \right).
\]

Lemma 2.4 [12] For given vectors \( u^i \in \mathbb{R}^n \), \( i = 1, \ldots, m \), we have
\[
\| \sum_{i=1}^{m} u^i \|^2 \leq m \sum_{j=1}^{m} \| u^j \|^2.
\]

Proof Invoking Cauchy-Schwarz inequality, for real numbers \( a_i \) and \( b_i \) (\( i = 1, \ldots, m \)), we have
\[
\left( \sum_{i=1}^{m} a_i b_i \right)^2 \leq \left( \sum_{i=1}^{m} a_i^2 \right) \left( \sum_{i=1}^{m} b_i^2 \right).
\]
The assertion follows from the above inequality directly.

Lemma 2.5 The mapping \( F(w) \) defined in (2.7b) satisfies
\[
(w' - w)^\top (F(w') - F(w)) = 0, \quad \forall w', w \in \mathbb{R}^{n+l}.
\]

Proof It is trivial and thus omitted.

3 Convergence

In this section, we prove the convergence of the e-ADMM (1.2) for \( m \geq 3 \) under the mentioned assumptions on the model (1.1) with a certain restriction on the penalty parameter \( \beta \). This is the main result of this paper. As mentioned, the proof is highly nontrivial. So we organize the discussion into several subsections. The roadmap of the proof is also reflected by the titles of these subsections.

3.1 Discerning the difference of an iterate from a solution point

We intend to observe the iterate \( w^{k+1} \) generated by the e-ADMM (1.2) and quantify its difference from a solution point in \( \mathcal{W}^* \) in terms of the variational inequality representation (2.7) of the optimality condition. Since the iterate \( w^{k+1} \) generated by the scheme (1.2) can be expressed in the variational inequality form, it is possible to compare it with the variational inequality representation (2.7) and so discern the difference of the iterate \( w^{k+1} \) from a solution point in \( \mathcal{W}^* \). More precisely, we can
show that this difference can be measured by some crossing terms and hence we need to carefully analyze these crossing terms. The following lemma follows from the first-order optimality conditions of the subproblems in the e-ADMM (1.2).

**Lemma 3.1** Suppose Assumptions 2.1 and 2.2 hold. Let \( \{w^k\} \) be the sequence generated by the e-ADMM (1.2). Then, we have \( x_i^{k+1} \in X_i \) \((i = 1, \ldots, m)\) and

\[
\theta_i(x_i) - \theta_i(x_i^{k+1}) - (x_i - x_i^{k+1})^\top A_i^\top z^{k+1} - \beta \sum_{j=i+1}^{m} A_j(x_j^k - x_j^{k+1}) \geq \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2, \quad (3.1)
\]

with \( \mu_1 = 0, \mu_2 = 0 \) and \( \mu_i > 0 \) for \( i = 3, \ldots, m \).

**Proof** According to the optimality condition of the \( x_i \)-subproblem (1.2c), we have \( x_i^{k+1} \in X_i \) such that

\[
\theta_i(x_i) - \theta_i(x_i^{k+1}) - (x_i - x_i^{k+1})^\top A_i^\top z^{k+1} - \beta \left( \sum_{j=1}^{i} A_j x_j^{k+1} + \sum_{j=i+1}^{m} A_j x_j^k - b \right) \geq \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2, \forall x_i \in X_i.
\]

Substituting the Eq. (1.2e) into the last inequality, we obtain the assertion (3.1).

Recall the characterization of \( W^* \) in (2.7). The following lemma reflects the discrepancy of \( w^{k+1} \) from a solution point in \( W^* \).

**Lemma 3.2** Suppose Assumptions 2.1 and 2.2 hold. Let \( \{w^k\} \) be the sequence generated by the e-ADMM (1.2a)–(1.2e). Then, we have

\[
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^\top F(w^{k+1}) + \frac{1}{\beta} (z - z^{k+1})^\top (z^{k+1} - z^k) + \beta (w - w^{k+1})^\top M(w - w^{k+1}) \geq \sum_{i=3}^{m} \frac{\mu_i}{2} \|x_i^{k+1} - x_i\|^2, \forall w \in W. \quad (3.2)
\]

**Proof** First, it follows from Eq. (1.2e) that

\[
\sum_{i=1}^{m} A_i x_i^{k+1} - b - \frac{1}{\beta} (z^k - z^{k+1}) = 0.
\]
Thus we have
\[
(z - z^{k+1})^\top \left[ \sum_{i=1}^{m} A_i x_i^{k+1} - b - \frac{1}{\beta} (z^k - z^{k+1}) \right] \geq 0, \quad \forall z \in \mathbb{R}^l. \quad (3.3)
\]

Combining the inequalities (3.1) for \( i = 1, \ldots, m \), with the above inequality, we have
\[
\begin{cases}
\theta_1(x_1) - \theta_1(x_1^{k+1}) - (x_1 - x_1^{k+1})^\top A_1^\top [z^{k+1} - \beta \sum_{j=2}^{m} A_j (x_j^k - x_j^{k+1})] \geq 0, \\
\theta_2(x_2) - \theta_2(x_2^{k+1}) - (x_2 - x_2^{k+1})^\top A_2^\top [z^{k+1} - \beta \sum_{j=3}^{m} A_j (x_j^k - x_j^{k+1})] \geq 0, \\
\vdots \\
\theta_i(x_i) - \theta_i(x_i^{k+1}) - (x_i - x_i^{k+1})^\top A_i^\top [z^{k+1} - \beta \sum_{j=i+1}^{m} A_j (x_j^k - x_j^{k+1})] \geq 0, \\
(z - z^{k+1})^\top \left[ \sum_{i=1}^{m} A_i x_i^{k+1} - b - \frac{1}{\beta} (z^k - z^{k+1}) \right] \geq 0, \quad \forall w \in \mathcal{W}. 
\end{cases} \quad (3.4)
\]

Adding all these inequalities together and using the definitions of \( F \) in (2.7b) and \( M \) in (2.2), we immediately obtain the assertion (3.2).

3.2 Replacing the crossing terms by summable quadratic terms

According to Lemma 3.2 and the optimality condition (2.7), it is clear that our emphasis should be analyzing the crossing term
\[
\frac{1}{\beta} (z - z^{k+1})^\top (z^{k+1} - z^k) + \beta (w - w^{k+1})^\top M (w^k - w^{k+1}) \quad (3.5)
\]

which gives the difference of the iterate \( w^{k+1} \) from a solution point in \( \mathcal{W}^* \). As we shall show later, the first term in (3.5) can be handled easily, whereas the second one should be sophisticatedly treated. This is indeed the most technical part in the paper.

We start from the following lemma.

Lemma 3.3 Suppose Assumptions 2.1 and 2.2 hold. For the iterative sequence \( \{w^k\} \) generated by the e-ADMM (1.2), we have
\[
\begin{align*}
(z^{k+1} - z^k, A_i x_i^{k+1} - A_i x_i^k) & \geq -\beta (A_i x_i^{k+1} - A_i x_i^k)^\top \left[ \sum_{j=i+1}^{m} A_j (x_j^{k+1} - x_j^k) - \sum_{j=i+1}^{m} A_j (x_j^k - x_j^{k-1}) \right] \\
& \quad + \mu_i \|x_i^k - x_i^{k+1}\|^2, \quad i = 1, \ldots, m, 
\end{align*}
\]

where \( \mu_1 = 0, \mu_2 = 0 \) and \( \mu_i > 0 \) \( (i = 3, \ldots, m) \).
Proof Setting $x_i := x_i^k$ in (3.1), we get

$$
\theta_i(x_i^k) - \theta_i(x_i^{k+1}) - (x_i^k - x_i^{k+1})^\top A_i^\top \left[ z_i^{k+1} + \beta \sum_{j=i+1}^m A_j (x_j^{k+1} - x_j^k) \right] \geq \frac{\mu_i}{2} \| x_i^{k+1} - x_i^k \|^2.
$$

Setting $x_i := x_i^{k+1}$ in (3.1) with the index $k$ replaced with $k - 1$, we have

$$
\theta_i(x_i^{k+1}) - \theta_i(x_i^k) - (x_i^{k+1} - x_i^k)^\top A_i^\top \left[ z_i^k + \beta \sum_{j=i+1}^m A_j (x_j^k - x_j^{k-1}) \right] \geq \frac{\mu_i}{2} \| x_i^{k+1} - x_i^k \|^2.
$$

Adding the above two inequalities, we obtain that for $i = 1, \ldots, m$, it holds

$$
\langle A_i x_i^{k+1} - A_i x_i^k, z_i^{k+1} - z_i^k \rangle \geq -\beta (A_i x_i^{k+1} - A_i x_i^k)^\top \left[ \sum_{j=i+1}^m A_j (x_j^{k+1} - x_j^k) - \sum_{j=i+1}^m A_j (x_j^k - x_j^{k-1}) \right].
$$

Note we use the convention $\sum_{i=m+1}^m a_i = 0$. The assertion (3.6) is proved.

Lemma 3.4 Suppose Assumptions 2.1 and 2.2 hold. For the iterative sequence $\{w^k\}$ generated by the e-ADMM (1.2), we have

$$
\langle z^k - z^{k+1}, \sum_{i=2}^m A_i (x_i^k - x_i^{k+1}) \rangle \geq -\beta \sum_{i=2}^{m-1} (A_i x_i^{k+1} - A_i x_i^k)^\top \left[ \sum_{j=i+1}^m A_j (x_j^{k+1} - x_j^k) \right.
$$

$$
\left. - \sum_{j=i+1}^m A_j (x_j^k - x_j^{k-1}) \right] + \sum_{i=3}^m \mu_i \| x_i^{k+1} - x_i^k \|^2.
$$

(3.7)

Proof Adding inequalities (3.6) from $i = 2$ to $m$ together, the assertion (3.7) follows immediately.

In the following lemma, we use the results in Lemmas 3.2 and 3.4; and represent the difference between the iterate $w_i^{k+1}$ from a solution point in $W^*$ by some quadratic terms (see (3.9) and (3.10)) and crossing terms in terms of only $A_i x_i^{k+1}$ (see (3.11) and (3.12)). This refined treatment turns out to be more convenient for successive operations over different subproblems; and it is the key to the proof of the main convergence results to be conducted.
Lemma 3.5 Suppose Assumptions 2.1 and 2.2 hold. Let \( \{w^k\} \) be the sequence generated by the e-ADMM (1.2). Then, for any \( w \in \mathcal{W} \), we have

\[
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^\top F(w) + \beta \left( \sum_{i=1}^{m} A_i x_i - b \right)^\top \sum_{i=1}^{m} A_i (x_i^k - x_i^{k+1}) \\
\geq \frac{\beta}{2} \sum_{i=2}^{m} \Delta(A_i x_i^{k+1}, A_i x_i^k, A_i x_i) + \frac{1}{2\beta} \Delta(z^{k+1}, z^k, z) \\
+ \sum_{i=3}^{m} \left( \mu_i \| x_i^{k+1} - x_i^k \|^2 + \frac{\mu_i}{2} \| x_i^{k+1} - x_i^k \|^2 \right) \\
+ \Upsilon_1^{k+1}(x_i^{k+1}, x_i^k, x_i^{k-1}) + \Upsilon_2^{k+1}(x_i^{k+1}, x_i^k, x_i),
\]

where

\[
\Delta(A_i x_i^{k+1}, A_i x_i^k, A_i x_i) := \| A_i x_i^{k+1} - A_i x_i \|^2 - \| A_i x_i^k - A_i x_i \|^2, \\
\Delta(z^{k+1}, z^k, z) := \| z^{k+1} - z \|^2 - \| z^k - z \|^2 + \| z^k - z^{k+1} \|^2,
\]

\[
\Upsilon_1^{k+1}(x_i^{k+1}, x_i^k, x_i^{k-1}) \\
:= -\beta \sum_{i=2}^{m-1} (A_i x_i^{k+1} - A_i x_i^k)^\top \left( \sum_{j=i+1}^{m} A_j (x_j^{k+1} - x_j^k) - \sum_{j=i+1}^{m} A_j (x_j^k - x_j^{k-1}) \right),
\]

\[
\Upsilon_2^{k+1}(x_i^{k+1}, x_i^k, x_i) := \beta \sum_{i=3}^{m} \sum_{j=2}^{i-1} (A_i x_i^{k+1} - A_i x_i)^\top (A_j x_j^{k+1} - A_j x_j^k).
\]

Proof First, using the definitions of \( M \) and \( N \) in (2.2), it holds that

\[
\beta (w - w^{k+1})^\top M (w^k - w^{k+1}) + \beta (w - w^{k+1})^\top N (w^k - w^{k+1}) \\
= \beta \left( \sum_{i=1}^{m} [(A_i x_i - b) - (A_i x_i^{k+1} - b)] \right)^\top \sum_{i=2}^{m} (A_i x_i^k - A_i x_i^{k+1}).
\]
Substituting (3.13) into the left-hand side of (3.2), we obtain

$$
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + \frac{1}{\beta}(z - z^{k+1})^T (z^{k+1} - z^k) 
+ \beta \left( \sum_{i=1}^m A_i x_i - b \right)^T \sum_{i=2}^m (A_i x_i^k - A_i x_i^{k+1}) 
- \beta \left( \sum_{i=1}^m A_i x_i^{k+1} - b \right)^T \sum_{i=2}^m (A_i x_i^k - A_i x_i^{k+1}) 
- \beta (w - w^{k+1})^T N(w^k - w^{k+1}) \geq \sum_{i=3}^m \frac{\mu_i}{2} \|x_i^k - x_i^{k+1}\|^2, \forall w \in \mathcal{W}. \quad (3.14)
$$

On the other hand, using the definition of $\triangle(z^{k+1}, z^k, z)$ in (3.10), we have

$$
\frac{1}{\beta}(z - z^{k+1})^T (z^{k+1} - z^k) = -\frac{1}{2\beta} \triangle(z^{k+1}, z^k, z).
$$

Substituting the above identity into (3.14) and using (1.2e), we obtain

$$
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + \beta \left( \sum_{i=1}^m A_i x_i - b \right)^T \sum_{i=2}^m (A_i x_i^k - A_i x_i^{k+1}) 
\geq \sum_{i=3}^m \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 + \beta (w - w^{k+1})^T N(w^k - w^{k+1}) + \frac{1}{2\beta} \triangle(z^{k+1}, z^k, z) 
+(z^k - z^{k+1})^T \sum_{i=2}^m (A_i x_i^k - A_i x_i^{k+1}), \forall w \in \mathcal{W}. \quad (3.15)
$$

Next, substituting (3.7) into the last term of the right-hand of (3.15), it yields

$$
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + \beta \left( \sum_{i=1}^m A_i x_i - b \right)^T \sum_{i=2}^m (A_i x_i^k - A_i x_i^{k+1}) 
\geq \sum_{i=3}^m \left[ \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 + \mu_i \|x_i^{k+1} - x_i^k\|^2 \right] + \beta (w - w^{k+1})^T N(w^k - w^{k+1}) 
+ \frac{1}{2\beta} \triangle(z^{k+1}, z^k, z) - \beta \sum_{i=2}^{m-1} (A_i x_i^{k+1} - A_i x_i^k)^T \left[ \sum_{j=i+1}^m A_j (x_j^{k+1} - x_j^k) - \sum_{j=i+1}^m A_j (x_j^i - x_j^{i-1}) \right], \forall w \in \mathcal{W}. \quad (3.16)
$$
On the other hand, it follows from (3.9) and the definition of the matrix $N$ in (2.2) that

$$
\beta(w - w^{k+1})^\top N(w^k - w^{k+1}) = \frac{\beta}{2} \sum_{i=2}^{m} \Delta(A_i x_i^{k+1}, A_i x_i^k, A_i x_i) + \beta \sum_{i=3}^{m} \sum_{j=2}^{i-1} (A_i x_i^{k+1} - A_i x_i)^\top (A_j x_j^{k+1} - A_j x_j^k).
$$

(3.17)

Substituting (3.17) into (3.16), we get

$$
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^\top F(w^{k+1}) + \beta \left( \sum_{i=1}^{m} A_i x_i - b \right)^\top \sum_{i=2}^{m} (A_i x_i^k - A_i x_i^{k+1}) \geq \sum_{i=3}^{m} \left[ \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 + \mu_i \|x_i^{k+1} - x_i^k\|^2 \right] + \frac{\beta}{2} \sum_{i=2}^{m} \Delta(A_i x_i^{k+1}, A_i x_i^k, A_i x_i) + \frac{1}{2\beta} \Delta(z^{k+1}, z, z) - \beta \sum_{i=2}^{m-1} (A_i x_i^{k+1} - A_i x_i^k)^\top \left( \sum_{j=i+1}^{m} A_j (x_j^{k+1} - x_j^k) - \sum_{j=i+1}^{m} A_j (x_j^k - x_j^{k-1}) \right) + \beta \sum_{i=3}^{m} \sum_{j=2}^{i-1} (A_i x_i^{k+1} - A_i x_i)^\top (A_j x_j^{k+1} - A_j x_j^k), \forall w \in \mathcal{W}.
$$

(3.18)

Finally, the assertion (3.8) follows from Lemma 2.5 and inequality (3.18) immediately.

For succinctness, we temporarily skip the superscripts and the variables for $\Upsilon_i$ ($i = 1, 2$). The next lemma focuses on analyzing the crossing terms $\Upsilon_1$ and $\Upsilon_2$ in the right-hand side of (3.8); and finding their lower bounds representable by negative quadratic terms. The purpose of doing so is that the difference between the iterate $w^{k+1}$ and a solution point in $\mathcal{W}^*$ can be completely represented by quadratic terms in a unified way. More specially, we decompose $\Upsilon_1$ and $\Upsilon_2$ into following several terms:

$$
\Upsilon_1^{(1)} := -\beta (A_2 x_2^{k+1} - A_2 x_2^k)^\top \left( \sum_{j=3}^{m} A_j (x_j^{k+1} - x_j^k) - \sum_{j=3}^{m} A_j (x_j^k - x_j^{k-1}) \right)
$$

(3.19)

$$
\Upsilon_1^{(2)} := -\beta \sum_{i=3}^{m-1} (A_i x_i^{k+1} - A_i x_i^k)^\top \sum_{j=i+1}^{m} A_j (x_j^{k+1} - x_j^k)
$$

(3.20)
\[ \Upsilon_{1}^{(22)} := \beta \sum_{i=3}^{m-1} (A_{i}x_{i}^{k+1} - A_{i}x_{i}^{k})^\top \sum_{j=i+1}^{m} A_{j}(x_{j}^{k} - x_{j}^{k-1}) \] (3.21)

\[ \Upsilon_{2}^{(1)} := \beta (A_{2}x_{2}^{k+1} - A_{2}x_{2}^{k})^\top \sum_{i=3}^{m} (A_{i}x_{i}^{k+1} - A_{i}x_{i}) \] (3.22)

\[ \Upsilon_{2}^{(2)} := \beta \sum_{j=3}^{m} \sum_{i=j+1}^{m} (A_{i}x_{i}^{k+1} - A_{i}x_{i})^\top (A_{j}x_{j}^{k+1} - A_{j}x_{j}) \] (3.23)

Then, we take a further analysis for each smaller term to get their lower bounds.

**Lemma 3.6** Suppose Assumptions 2.1 and 2.2 hold. Let \( \{w^{k}\} \) be the sequence generated by the e-ADMM (1.2). Then, for any \( w \in \mathcal{W} \), we have the following assertions:

1) For any scalars \( a, b > 0 \), it holds that

\[ \Upsilon_{1}^{(1)} \geq \beta \left( -(a + b) \| A_{2}x_{2}^{k} - A_{2}x_{2}^{k+1} \|^2 - \frac{1}{4a} \sum_{i=3}^{m} (A_{i}x_{i}^{k+1} - A_{i}x_{i}^{k})^2 \right. \]

\[ \left. - \frac{m - 2}{4b} \sum_{i=3}^{m} \| A_{i}x_{i}^{k} - A_{i}x_{i}^{k-1} \|^2 \right). \] (3.24)

2) The following identity holds:

\[ \Upsilon_{1}^{(21)} = -\frac{\beta}{2} \left( \| \sum_{i=3}^{m} (A_{i}x_{i}^{k+1} - A_{i}x_{i}^{k}) \|^2 - \sum_{i=3}^{m} \| A_{i}x_{i}^{k+1} - A_{i}x_{i}^{k} \|^2 \right). \] (3.25)

3) It holds that

\[ \Upsilon_{1}^{(22)} \geq -\frac{\beta}{2} \sum_{i=3}^{m-1} (m - i) \| A_{i}x_{i}^{k+1} - A_{i}x_{i}^{k} \|^2 - \frac{\beta}{2} \sum_{i=4}^{m} (i - 3) \| A_{i}x_{i}^{k} - A_{i}x_{i}^{k-1} \|^2. \] (3.26)

4) For any scalar \( \delta > 0 \), it holds that

\[ \Upsilon_{2}^{(1)} \geq -\beta \left( \frac{m - 2}{2\delta} \sum_{i=3}^{m} \| A_{i}x_{i}^{k+1} - A_{i}x_{i} \|^2 + \frac{\delta}{2} \| A_{2}x_{2}^{k+1} - A_{2}x_{2}^{k} \|^2 \right). \] (3.27)
5) It holds that

$$\Upsilon^{(2)}_2 \geq - \frac{\beta}{2} \left( \sum_{i=3}^{m} (i - 3) \| A_i x_i^{k+1} - A_i x_i \|^2 + \sum_{i=3}^{m-1} (m - i) \| A_i x_i^{k+1} - A_i x_i \|^2 \right). \quad (3.28)$$

6) $\Upsilon_1$ and $\Upsilon_2$ defined respectively in (3.11) and (3.12) satisfy the following equations:

$$\Upsilon_1 = \Upsilon_1^{(1)} + \Upsilon_1^{(21)} + \Upsilon_1^{(22)} \quad \text{and} \quad \Upsilon_2 = \Upsilon_2^{(1)} + \Upsilon_2^{(2)}.$$

(3.29)

Proof

1) Using Cauchy-Schwarz inequality, for any positive scalars $a$ and $b$, we have

$$\Upsilon_1^{(1)} = - \beta (A_2 x_2^{k+1} - A_2 x_2^k)^\top \sum_{j=3}^{m} A_j (x_j^{k+1} - x_j^k) + \beta (A_2 x_2^{k+1} - A_2 x_2^k)^\top \sum_{j=3}^{m} A_j (x_j^k - x_j^{k-1})$$

$$\geq \beta \left( -a \| A_2 x_2^k - A_2 x_2^{k+1} \|^2 - \frac{1}{4a} \sum_{i=3}^{m} (A_i x_i^{k+1} - A_i x_i^k) \|^2 \right.$$

$$- b \| A_2 x_2^k - A_2 x_2^{k+1} \|^2 - \frac{1}{4b} \sum_{i=3}^{m} (A_i x_i^k - A_i x_i^{k-1}) \|^2 \bigg)$$

$$\geq \beta \left( -a \| A_2 x_2^k - A_2 x_2^{k+1} \|^2 - \frac{1}{4a} \sum_{i=3}^{m} (A_i x_i^{k+1} - A_i x_i^k) \|^2 \right.$$

$$+ \beta \left( -b \| A_2 x_2^k - A_2 x_2^{k+1} \|^2 - \frac{m-2}{4b} \sum_{i=3}^{m} \| A_i x_i^k - A_i x_i^{k-1} \|^2 \bigg),$$

where the last inequality follows from Lemma 2.4. Then, the inequality (3.24) follows directly.

2) Invoking the identity $x^\top y = \frac{1}{2} (\| x + y \|^2 - \| x \|^2 - \| y \|^2)$, we know

$$\Upsilon_1^{(21)} = - \beta \frac{1}{2} \sum_{i=3}^{m-1} \left( \left\| \sum_{j=i}^{m} (A_j x_j^{k+1} - A_j x_j^k) \right\|^2 - \left\| A_i x_i^{k+1} - A_i x_i^k \right\|^2 - \left\| \sum_{j=i+1}^{m} (A_j x_j^{k+1} - A_j x_j^k) \right\|^2 \right)$$

$$= - \beta \left( \left\| \sum_{i=3}^{m} (A_i x_i^{k+1} - A_i x_i^k) \right\|^2 - \sum_{i=3}^{m} \| A_i x_i^{k+1} - A_i x_i^k \|^2 \right).$$

Then, the inequality (3.25) is proved.
3) Using Cauchy-Schwarz inequality, we have

\[
\gamma_1^{(22)} \geq -\frac{\beta}{2} \sum_{i=3}^{m-1} \sum_{j=i+1}^{m} \left( \| A_i x_i^{k+1} - A_i x_i^k \|^2 + \| A_j x_j^k - A_j x_j^{k-1} \|^2 \right)
\]

\[
= -\frac{\beta}{2} \sum_{i=3}^{m-1} \sum_{j=i+1}^{m} \| A_i x_i^{k+1} - A_i x_i^k \|^2 - \frac{\beta}{2} \sum_{i=3}^{m-1} \sum_{j=i+1}^{m} \| A_j x_j^k - A_j x_j^{k-1} \|^2
\]

\[
= -\frac{\beta}{2} \sum_{i=3}^{m-1} (m-i) \| A_i x_i^{k+1} - A_i x_i^k \|^2 - \frac{\beta}{2} \sum_{j=4}^{m} \sum_{i=3}^{j-1} \| A_j x_j^k - A_j x_j^{k-1} \|^2
\]

\[
= -\frac{\beta}{2} \sum_{i=3}^{m-1} (m-i) \| A_i x_i^{k+1} - A_i x_i^k \|^2 - \frac{\beta}{2} \sum_{i=4}^{m} \sum_{i=3}^{i-3} \| A_i x_i^k - A_i x_i^{k-1} \|^2
\]

Thus, we obtain the inequality (3.26).

4) Using Cauchy-Schwarz inequality, for any positive scalar \( \delta \), we have

\[
\gamma_2^{(1)} \geq -\beta \left( \frac{1}{2\delta} \sum_{i=3}^{m} (A_i x_i^{k+1} - A_i x_i) \|^2 + \frac{\delta}{2} \| A_2 x_2^{k+1} - A_2 x_2^k \|^2 \right)
\]

\[
\geq -\beta \left( \frac{m-2}{2\delta} \sum_{i=3}^{m} \| A_i x_i^{k+1} - A_i x_i \|^2 + \frac{\delta}{2} \| A_2 x_2^{k+1} - A_2 x_2^k \|^2 \right),
\]

where the second inequality follows directly from Lemma 2.4. Then, the inequality (3.27) follows directly.

5) Again, using Cauchy-Schwarz inequality, it yields

\[
\gamma_2^{(2)} \geq -\frac{\beta}{2} \sum_{j=3}^{m-1} \sum_{i=j+1}^{m} \left( \| A_j x_j^{k+1} - A_j x_j^k \|^2 + \| A_i x_i^{k+1} - A_i x_i \|^2 \right)
\]

\[
= -\frac{\beta}{2} \sum_{j=3}^{m-1} \sum_{i=j+1}^{m} \| A_j x_j^{k+1} - A_j x_j^k \|^2 - \frac{\beta}{2} \sum_{i=4}^{m} \sum_{j=3}^{i-1} \| A_i x_i^{k+1} - A_i x_i \|^2
\]

\[
= -\frac{\beta}{2} \left( \sum_{i=3}^{m-1} (m-i) \| A_i x_i^{k+1} - A_i x_i^k \|^2 + \sum_{i=3}^{m} (i-3) \| A_i x_i^{k+1} - A_i x_i \|^2 \right),
\]

where the second equality follows from changing the sum order for the second term (the same technique as the proof for (3.26)). Thus, the inequality (3.28) is proved.
6) The assertion (3.29) follows from the definitions of $\gamma_1$, $\gamma_2$, $\gamma_1^{(1)}$, $\gamma_2^{(2)}$, $\gamma_1^{(2)}$, and $\gamma_2^{(1)}$ (see (3.11), (3.12) and (3.19)–(3.23)), and some elementary calculations.

With the previously proved lemmas, we can derive a favorable relationship for two consecutive iterates about their respective differences from a solution point in $\mathcal{W}^*$. This relationship is reflected by an inequality that is completely representable by quadratic terms without any crossing terms. It is thus easy to show that the sequence generated by the e-ADMM (1.2) is Fejér monotone with respect to $\mathcal{W}^*$.

**Lemma 3.7** Suppose Assumptions 2.1 and 2.2 hold. Let $\{w^k\}$ be the sequence generated by the e-ADMM (1.2). For arbitrary positive scalars $a$, $b$, $\delta$, and any $w^* \in \mathcal{W}^*$, we have

$$
\frac{\beta}{2} \sum_{i=2}^{m} \|A_i x_i^{k+1} - A_i x_i^*\|^2 + \frac{1}{2\beta} \|z^{k+1} - z^*\|^2 + \beta \sum_{i=3}^{m} \left[ \frac{(i-3)}{2} + \frac{m-2}{4b} \right] \|A_i x_i^{k+1} - A_i x_i^*\|^2
$$

$$
\leq \frac{\beta}{2} \sum_{i=2}^{m} \|A_i x_i^k - A_i x_i^*\|^2 + \frac{1}{2\beta} \|z^k - z^*\|^2 + \beta \sum_{i=3}^{m} \left[ \frac{(i-3)}{2} + \frac{m-2}{4b} \right] \|A_i x_i^k - A_i x_i^{k-1}\|^2
$$

$$
- \sum_{i=2}^{m} C_i \|A_i x_i^{k+1} - A_i x_i^*\|^2 - \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 - \sum_{i=3}^{m} \zeta_i \|A_i x_i^{k+1} - A_i x_i^*\|^2,
$$

(3.30)

where

$$
C_2 := (\frac{1}{2} - (a + b) - \frac{\delta}{2})\beta,
$$

(3.31)

$$
C_i := \frac{\mu_i}{\|A_i^j A_i\|} - \beta \left( \frac{1}{4a} + \frac{1}{4b} \right) (m - 2) + \frac{3m-i-7}{2}, \quad i = 3, \ldots, m,
$$

(3.32)

and

$$
\zeta_i := \frac{\mu_i}{\|A_i^j A_i\|} - \beta \left( \frac{m-2}{28} + \frac{(i-3)}{2} \right), \quad i = 3, \ldots, m.
$$

(3.33)

**Proof** First, substituting (3.24)–(3.28) into (3.8) and invoking (3.29), we derive that

$$
\frac{\beta}{2} \sum_{i=2}^{m} \|A_i x_i^{k+1} - A_i x_i^*\|^2 + \frac{1}{2\beta} \|z^{k+1} - z\|^2 + \beta \sum_{i=3}^{m} \left[ \frac{(i-3)}{2} + \frac{m-2}{4b} \right] \|A_i x_i^{k+1} - A_i x_i^*\|^2
$$

$$
\leq \frac{\beta}{2} \sum_{i=2}^{m} \|A_i x_i^k - A_i x_i^*\|^2 + \frac{1}{2\beta} \|z^k - z\|^2 + \beta \sum_{i=3}^{m} \left[ \frac{(i-3)}{2} + \frac{m-2}{4b} \right] \|A_i x_i^k - A_i x_i^{k-1}\|^2
$$

$$
- C_2 \|A_2 x_2^{k+1} - A_2 x_2^*\|^2 - \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 + \frac{\beta}{2} (1 + \frac{1}{2a}) \sum_{i=3}^{m} (A_i x_i^{k+1} - A_i x_i^*)^2
$$
\[
- \sum_{i=3}^{m} \left\{ \mu_i \|x_i - x_{i+1}\|^2 - \frac{\beta}{2} \left[ \frac{m-2}{2b} + 2m - i - 5 \right] \|A_i x_i - A_i x_{i+1}\|^2 \right\} \\
- \sum_{i=3}^{m} \left\{ \frac{\mu_i}{2} \|x_i - x_{i+1}\|^2 - \beta \left[ \frac{m-2}{2\delta} + \frac{(i-3)}{2} \right] \|A_i x_{i+1} - A_i x_i\|^2 \right\} \\
+ \left\{ \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^\top F(w) + \beta \left( \sum_{i=1}^{m} A_i x_i - b \right)^\top \sum_{i=1}^{m} A_i (x_i - x_{i+1}) \right\}.
\]

(3.34)

Invoking the Cauchy-Schwarz inequality and \( a > 0 \), we have

\[
\frac{\beta}{2} (1 + \frac{1}{2a}) \| \sum_{i=3}^{m} (A_i x_{i+1} - A_i x_i) \|^2 \\
\leq \frac{\beta}{2} (1 + \frac{1}{2a}) (m-2) \sum_{i=3}^{m} \|A_i x_{i+1} - A_i x_i\|^2.
\]

(3.35)

Then, using (2.7a), we get

\[
\theta(u^*) - \theta(u^{k+1}) + (w^* - w^{k+1})^\top F(w^*) + \beta \left( \sum_{i=1}^{m} A_i x_i^* - b \right)^\top \sum_{i=1}^{m} A_i (x_i^* - x_{i+1}) \\
\leq -\sum_{i=3}^{m} \frac{\mu_i}{2} \|x_i^* - x_{i+1}\|^2.
\]

(3.36)

Setting \( w := w^* \in \mathcal{W}^* \) in (3.34) and combining (3.35) and (3.36), we obtain the assertion (3.30)--(3.33) directly. \( \square \)

3.3 Main result

In this subsection, we prove the convergence of the e-ADMM (1.2) with \( m \geq 3 \) under Assumptions 2.1 and 2.2. This is the main result of this paper. As indicated by Theorem 4.1 in [6], the penalty parameter \( \beta \) must be appropriately restricted to guarantee the convergence of the e-ADMM (1.2) even all functions are assumed to be strongly convex. Therefore, in the following theorem we first present a range for \( \beta \) to ensure the convergence of the e-ADMM (1.2) with \( m \geq 3 \) under our assumptions.

We target a larger range for \( \beta \) by choosing a specific choice of \( a, b \) and \( \delta \) which are involved in the definitions of \( C_i \) (\( i = 2, \ldots, m \)) and \( \zeta_i \) (\( i = 3, \ldots, m \)) in (3.31)--(3.33) to ensure that all these coefficients are positive. With the positiveness of these coefficients, as we shall show in the proof, we can establish the contractive property of the sequence generated by the e-ADMM (1.2). It is noticed that the range of \( \beta \) relies on \( m \), thus it is impossible to get an optimal range of \( \beta \) for all \( m \geq 3 \). However, we provide a specific choice for the parameters \( a, b \) and \( \delta \) to ensure the upper bound of the range for \( \beta \) is linear with \( m \). Moreover, the coefficient in the linear term for \( m \) is as small as possible. It turns out to be better than the existing literatures.
Lemma 3.8 Suppose Assumptions 2.1 and 2.2 hold. When $\beta$ is restricted by

$$\beta \in \left(0, \min_{3 \leq i \leq m} \frac{\mu_i}{\max\{4m - 10, 3m - 6.5\} \left\|A_i^\top A_i\right\|}\right),$$

(3.37)

there always exists a sufficient small positive scalar $\epsilon'$ such that

$$\hat{C}_2 := \frac{\epsilon'}{10(m - 2 + \epsilon')} \beta > 0,$$

(3.38)

$$\hat{C}_i := \frac{\mu_i}{\left\|A_i^\top A_i\right\|} - \beta (4m - 10) > 0, \ i = 3, \ldots, m,$$

(3.39)

$$\hat{\zeta}_i := \frac{\mu_i}{\left\|A_i^\top A_i\right\|} - \beta (3m - 6.5 + 2.5\epsilon') > 0, \ i = 3, \ldots, m.$$ 

(3.40)

Proof When $\beta$ is restricted to the range (3.37), there always exists a sufficient small positive scalar $\epsilon'$ such that $\hat{\zeta}_i$ defined in (3.40) with $i = 3, \ldots, m$ are all positive. With the purpose of finding a larger range of $\beta$ while ensuring the positiveness of all the coefficients in (3.31)–(3.33), we choose $a = b$ and then it results in the following equations:

$$C_2 = \left(\frac{1}{2} - 2a - \frac{\delta}{2}\right) \beta,$$

(3.41)

$$C_i = \frac{\mu_i}{\left\|A_i^\top A_i\right\|} - \beta \left(\frac{m - 2}{2a} + \frac{3m - i - 7}{2}\right), \ i = 3, \ldots, m,$$

(3.42)

$$\zeta_i = \frac{\mu_i}{\left\|A_i^\top A_i\right\|} - \beta \left[\frac{m - 2 + (i - 3)}{2\delta} + \frac{i - 3}{2}\right], \ i = 3, \ldots, m.$$ 

(3.43)

To ensure $C_2$ in (3.41) is positive, we probe the choice of $\delta$ as

$$\delta = (1 - 4a) \frac{m - 2}{m - (2 - \epsilon')}$$

so that the numerator $m - 2$ in (3.42) can be canceled. This particular choice also allows us to derive a range of $\beta$ whose upper bound can be represented by some linear terms of $m$. Indeed, with the mentioned probe, we have

$$C_2 := \frac{1 - 4a}{2} \frac{\epsilon'}{m - 2 + \epsilon'} \beta,$$

(3.44)

$$C_i := \frac{\mu_i}{\left\|A_i^\top A_i\right\|} - \beta \left(\frac{m - 2}{2a} + \frac{3m - i - 7}{2}\right), \ i = 3, \ldots, m,$$

(3.45)

$$\zeta_i := \frac{\mu_i}{\left\|A_i^\top A_i\right\|} - \beta \left[\frac{m - 2 + \epsilon'}{2(1 - 4a)} + \frac{i - 3}{2}\right], \ i = 3, \ldots, m.$$ 

Further probing different values of $a$, we choose $a = \frac{1}{3}$ in (3.43)–(3.45). At the same time, we choose $i = 3$ in (3.44) and $i = m$ in (3.45) to get a uniform lower-bound
for $C_i$ ($i = 3, \ldots, m$) and $\zeta_i$ ($i = 3, \ldots, m$), respectively. Then, the definitions in (3.43)–(3.45) can be specified accordingly as

$$
\hat{C}_2 = \frac{\epsilon'}{10(m - 2 + \epsilon')} \beta,
$$

$$
C_i \geq \hat{C}_i = \frac{\mu_i}{\|A_i^\top A_i\|} - \beta(4m - 10), \quad i = 3, \ldots, m,
$$

$$
\zeta_i \geq \hat{\zeta}_i = \frac{\mu_i}{\|A_i^\top A_i\|} - \beta(3m - 6.5 + 2.5\epsilon'), \quad i = 3, \ldots, m.
$$

We note that $\hat{C}_i$ ($i = 2, \ldots, m$) and $\hat{\zeta}_i$ ($i = 3, \ldots, m$) are all positive when $\beta$ is restricted to the range (3.37).

Note that the coefficients $\hat{C}_i$ ($i = 2, \ldots, m$) and $\hat{\zeta}_i$ ($i = 3, \ldots, m$) are unrelated to $a$, $b$ and $\delta$. Now, we are in the stage to prove the convergence of the e-ADMM (1.2) with the restriction (3.37) on $\beta$. Let us define a potential function $\Phi(v^{k+1}, v^k, v)$ as

$$
\Phi(v^{k+1}, v^k, v) := \frac{1}{2} \|v^{k+1} - v\|^2_Q + \beta \sum_{i=3}^m \left[ \frac{(i - 3)}{2} + \frac{5(m - 2)}{4} \right] \|A_i x_i^{k+1} - A_i x_i^k\|^2,
$$

(3.46)

where the matrix $Q$ is defined in (2.3). Define a block diagonal matrix as follows:

$$
\tilde{Q} = \text{diag} \left( 2\hat{C}_2 A_2^\top A_2, \cdots, 2\hat{C}_m A_m^\top A_m, \frac{1}{\beta} I \right).
$$

Then, the matrix $\tilde{Q}$ is positive definite when $\beta$ satisfying (3.37) according to Lemma 3.8. With the specific choices of $a$, $b$ and $\delta$ (i.e., $a = b = 1/5$ and $\delta := \frac{m-2}{5(m-2)}$), we have shown that the coefficients $C_i$ ($i = 2, \ldots, m$) and $\zeta_i$ ($i = 3, \ldots, m$) defined in (3.31)–(3.33) have positive lower bounds: $\hat{C}_i$ ($i = 2, \ldots, m$) and $\hat{\zeta}_i$ ($i = 3, \ldots, m$) defined respectively in (3.38)–(3.40). Then, substituting the mentioned values of $a$, $b$ and $\delta$ into (3.30)–(3.33), we get

$$
\Phi(v^{k+1}, v^k, v^*) \leq \Phi(v^k, v^{k-1}, v^*) - \frac{1}{2} \|v^k - v^{k+1}\|^2_{\tilde{Q}}.
$$

(3.47)

**Theorem 3.1** Suppose Assumptions 2.1 and 2.2 hold. Let $\{w^k\}$ be the sequence generated by the e-ADMM (1.2) with $\beta$ restricted in (3.37). Then, the sequence $\{w^k\}$ converges to a solution point in $W^*$.

**Proof** It follows from (3.47) and the definition of $\Phi$ in (3.46) that the sequence defined by

$$
\left\{ \frac{1}{2} \|v^{k+1} - v^*\|^2_Q + \beta \sum_{i=3}^m \left[ \frac{(i - 3)}{2} + \frac{5(m - 2)}{4} \right] \|A_i x_i^{k+1} - A_i x_i^k\|^2 \right\}
$$
is non-increasing. It implies that the sequence \( \{v^k\} \) is bounded under the assumption that \( A_i \ (i = 2, \ldots, m) \) are full column rank. The relationship (1.2e) and the fact that \( A_1 \) is full column rank further imply that the sequence \( \{x^k_i\} \) is also bounded. Hence, the iterative sequence \( \{w^k\} \) generated by the scheme (1.2a)–(1.2e) is bounded.

Summarizing (3.47) for all \( k \) and rearranging the terms, we get

\[
\frac{1}{2} \sum_{k=1}^{\infty} \|v^k - v^{k+1}\|_Q^2 \leq \left( \frac{1}{2}\|v^1 - v^*\|_Q^2 + \beta \sum_{i=3}^{m} \left( \frac{(i-3)}{2} + \frac{5(m - 2)}{4} \right) \|A_i x^0_i - A_i x^1_i\|_Q^2 \right),
\]

which implies

\[
\lim_{k \to \infty} \|z^k - z^{k+1}\| = 0 \quad \text{and} \quad \lim_{k \to \infty} \|A_i x^k_i - A_i x^{k+1}_i\| = 0, \ i = 2, \ldots, m. \quad (3.48)
\]

Moreover, the boundedness of the sequence indicates that the sequence \( \{w^k\} \) has at least one cluster point. Let \( w^\infty \) be an arbitrary cluster point of \( \{w^k\} \) and \( \{w^kj\} \) be the subsequence converging to \( w^\infty \). Then, the sequence \( \{v^kj\} \) converges to \( v^\infty \); and the whole sequence \( \{v^k\} \) has only one cluster point \( v^\infty \) because of (3.47). On the other hand, it follows from (1.2e) and the fact that \( A_1 \) is full column rank that

\[
A_1^{k+1} = (A_1^\top A_1)^{-1} A_1^\top \left( b - \sum_{i=2}^{m} A_i x^k_i + \frac{z^k - z^{k+1}}{\beta} \right) .
\]

Then, the sequence \( \{x^k_i\} \) has only one cluster point, say \( x^\infty_i := (A_1^\top A_1)^{-1} A_1^\top \left( b - \sum_{i=2}^{m} A_i x^\infty_i \right) \), by combining the above equation with \( v^k \to v^\infty \). Thus, the sequence \( \{w^k\} \) converges to \( w^\infty \). Let \( j \to +\infty \) in (3.4) by replacing \( w^{k+1} \) with \( w^{kj} \) and using (3.48), we have

\[
\begin{align*}
\theta_1(x_1) - \theta_1(x^\infty_1) + (x_1 - x^\infty_1)^\top (A_1^\top z^\infty) & \geq 0, \\
\theta_2(x_2) - \theta_2(x^\infty_2) + (x_2 - x^\infty_2)^\top (A_2^\top z^\infty) & \geq 0, \\
\vdots & \vdots \\
\theta_i(x_i) - \theta_i(x^\infty_i) + (x_i - x^\infty_i)^\top (A_i^\top z^\infty) & \geq \frac{\mu_i}{2} \|x_i - x^\infty_i\|_2^2, \ i = 3, \ldots, m, \quad \forall \ w \in \mathcal{W}.
\end{align*}
\]

According to the optimality condition (2.7), we know \( w^\infty \in \mathcal{W}^* \). Consequently, the sequence \( \{w^k\} \) generated by the e-ADMM (1.2) with \( \beta \) restricted in (3.37) converges to a solution point in \( \mathcal{W}^* \).

Remark 3.1 It can be seen from (3.47) that the sequence \( \{A_i x^k_i\} \ (i = 1, \ldots, m) \) converges to \( \{A_i x^\infty_i\} \ (i = 1, \ldots, m) \) even without the full column rank assumptions on \( A_i \)'s \( (i = 1, \ldots, m) \).
Remark 3.2 We have shown the convergence of the e-ADMM (1.2) when $\beta$ is restricted in the range (3.37), for the generic case with a general $m \geq 3$. Indeed, when the special case of $m = 3$ is considered, the range (3.37) reduces to

$$\beta \in \left( 0, \frac{2}{5 \| A_3^\top A_3 \|} \right),$$

which is still larger than some ones in the literature that are only eligible to the special case of $m = 3$, e.g., the range $\left( 0, \frac{6}{17 \| A_3^\top A_3 \|} \right)$ proposed in [4].

4 Ergodic convergence rate

In [16, 17], some worst-case $O(1/t)$ convergence rates measured by the iteration complexity were established for the original ADMM scheme which corresponds to the scheme (1.2) with $m = 2$. Since then, there are some works focusing on investigating the convergence rates in the same nature for various splitting methods in the literature. This kind of convergence rate means a global estimate on the convergence speed for the algorithm under discussion. In this section, we establish a worst-case $O(1/t)$ convergence rate measured by the iteration complexity for the e-ADMM (1.2) with $m \geq 3$ under Assumptions 2.1 and 2.2. Compared with (3.37), the restriction on $\beta$ to ensure the $O(1/t)$ convergence rate is slightly more restrictive. In order to establish the ergodic convergence rate, we still require the positiveness of $C_2$ defined in (3.31), $C_i (i = 3, \ldots, m)$ in (3.32) and $\tilde{\zeta}_i (i = 3, \ldots, m)$ in (4.6). Note that $\tilde{\zeta}_i$ in (4.6) is deferent from $\zeta_i$ is in (3.33); and the difference results in a more restrictive range of $\beta$ as to be shown later.

The following lemma will be used to prove a worst-case $O(1/t)$ convergence rate for the e-ADMM (1.2) with $m \geq 3$.

Lemma 4.1 Suppose Assumptions 2.1 and 2.2 hold. Let $\{w^k\}$ be the sequence generated by the e-ADMM (1.2) with $m \geq 3$. If $\beta$ is restricted by

$$\beta \in \left( 0, \min_{3 \leq i \leq m} \frac{\mu_i}{\left( \frac{13 + \sqrt{33}}{4} m - \frac{17 + \sqrt{33}}{2} \right) \| A_i^\top A_i \|} \right),$$

then we have

$$\Theta(v^{k+1}, v^k, w) \leq \Theta(v^k, v^{k-1}, w) + \Xi(w^{k+1}, w^k, w),$$

where

$$\Theta(v^{k+1}, v^k, w) := \frac{1}{2} \| v^{k+1} - v \|_Q^2 + \beta \sum_{i=3}^m \tau_i \| A_i x_i^{k+1} - A_i x_i^k \|^2,$$

$$\tau_i := \frac{(i-3)}{2} + \frac{(7 + \sqrt{33}) (m-2)}{8},$$

$$\Xi(w^{k+1}, w^k, w) := \frac{1}{2} \| v^{k+1} - v \|_Q^2 - \beta \sum_{i=3}^m \tau_i \| A_i x_i^{k+1} - A_i x_i^k \|^2.$$
and
\[
\Xi(w^{k+1}, w^k, w) := \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^\top F(w) \\
+ \beta \left( \sum_{i=1}^m A_i x_i - b \right) \left( \sum_{i=2}^m A_i (x_i^k - x_i^{k+1}) \right). \tag{4.4}
\]

**Proof** First, substituting (3.35) into (3.34) and then merging the coefficients of similar items, we get
\[
\begin{align*}
\frac{\beta}{2} \sum_{i=2}^m \|A_i x_i^{k+1} - A_i x_i^k\|^2 + \frac{1}{2\beta} \|z^{k+1} - z\|^2 \\
+ \beta \sum_{i=3}^m \left\{ \frac{(i-3)}{2} + \frac{(m-2)}{4b} \right\} \|A_i x_i^{k+1} - A_i x_i^k\|^2 \\
\leq \frac{\beta}{2} \sum_{i=2}^m \|A_i x_i^k - A_i x_i\|^2 + \frac{1}{2\beta} \|z^k - z\|^2 \\
+ \beta \sum_{i=3}^m \left\{ \frac{(i-3)}{2} + \frac{(m-2)}{4b} \right\} \|A_i x_i^k - A_i x_i^{k-1}\|^2 \\
- \sum_{i=2}^m C_i \|A_i x_i^{k+1} - A_i x_i^k\|^2 - \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 - \sum_{i=3}^m \tilde{\zeta}_i \|A_i x_i^{k+1} - A_i x_i\|^2 \\
+ \Xi(w^{k+1}, w^k, w), \tag{4.5}
\end{align*}
\]

where \(C_2, C_i \ (i = 3, \ldots, m), \Xi(w^{k+1}, w^k, w)\) are defined respectively in (3.31), (3.32), (4.4), and denote
\[
\tilde{\zeta}_i := \frac{\mu_i}{2\|A_i\|} - \beta \left[ \frac{m-2}{2} + \frac{(i-3)}{2} \right], \ i = 3, \ldots, m. \tag{4.6}
\]

The heuristics of the following part is similar as that of Lemma 3.8. We skip the details for succinctness. Setting \(a = b = \frac{7 - \sqrt{33}}{8} \) and \(\delta = \frac{\sqrt{33} - 5}{2} \left( \frac{m-2}{m-2+\epsilon} \right) (\epsilon > 0)\), we get
\[
\begin{align*}
C_2 &= \frac{\sqrt{33} - 5}{4} \frac{\epsilon'}{(m-2+\epsilon')\beta}, \\
C_i &= \frac{\mu_i}{\|A_i\|} - \left( \frac{13 + \sqrt{33}}{4} m - \frac{14 + i + \sqrt{33}}{2} \right) \beta, \ i = 3, \ldots, m, \\
\text{and} \\
\tilde{\zeta}_i &= \frac{\mu_i}{2\|A_i\|} - \left( \frac{5 + \sqrt{33}}{8} m - \frac{5 + \sqrt{33}}{4} + \frac{i-3}{2} + \frac{5 + \sqrt{33}}{8} \epsilon' \right) \beta, \ i = 3, \ldots, m. \tag{4.7}
\end{align*}
\]
Let $\epsilon' \to 0$. Then, we derive that $C_i > 0$, $(i = 2, \ldots, m)$ and $\tilde{\zeta}_i > 0$ $(i = 3, \ldots, m)$ when $\beta$ satisfies (4.1). Thus, the assertion (4.2) follows from (4.5) immediately.

Based on Lemma 4.1, we now establish a worst-case $O(1/t)$ convergence rate in the ergodic sense for the e-ADMM (1.2). For this analysis, the quality of an iterate is measured by the feasibility violation and the decrease of the objective function. Let us define

$$x_{i,t}^{k+1} := \frac{1}{t} \sum_{k=1}^{t} x_{i}^{k+1}, \quad i = 1, \ldots, m; \quad u_{i,t}^{k+1} := \frac{1}{t} \sum_{k=1}^{t} u_{i}^{k+1} \quad \text{and} \quad w_{i,t}^{k+1} := \frac{1}{t} \sum_{k=1}^{t} w_{i}^{k+1}. \quad (4.8)$$

Obviously, $w_{i,t}^{k+1} \in \mathcal{W}$ because of the convexity of $X_i$ $(i = 1, \ldots, m)$. Note that we are considering the case of $m \geq 3$. Hence, the interval (4.1) is included in the restriction of $\beta$ (3.37). Then, invoking Theorem 3.1, the sequence $\{\frac{1}{2} \|v^k - v^*\|_Q^2\}$ is bounded and thus there exists a constant $\kappa$ such that

$$\|A_ix_i^k\| \leq \kappa, \quad \forall \ i = 1, \ldots, m, \quad \text{and} \quad \|z^k\| \leq \kappa, \quad \forall \ k \geq 0. \quad (4.9)$$

**Theorem 4.1** Suppose Assumptions 2.1 and 2.2 hold. For $t$ iterations generated by the e-ADMM (1.2) with $\beta$ restricted in (4.1), the following assertions hold.

1) For $\bar{C} := \beta \sum_{i=3}^{m} \tau_i \|A_ix_i^1 - A_ix_i^0\|^2$ and $\tau_i$ is defined in (4.3), we have

$$\theta(u_{i,t}^{k+1}) - \theta(u) + (u_{i,t}^{k+1} - w)^{\top} F(w)$$

$$\leq \frac{1}{t} \left[ 2\beta \kappa (m - 1) \sum_{i=1}^{m} A_ix_i - b \right] + \frac{1}{2} \|v^1 - v\|_Q^2 + \bar{C}. \quad (4.10)$$

2) There exists a constant $\bar{c}_1 > 0$ such that

$$\| \sum_{i=1}^{m} A_ix_{i,t}^{k+1} - b \| \leq \frac{\bar{c}_1}{t^2}. \quad (4.11)$$

3) There exists a constant $\bar{c}_2 > 0$ such that

$$|\theta(u_{i,t}^{k+1}) - \theta(u^*)| \leq \frac{\bar{c}_2}{t}. \quad (4.12)$$

**Proof** 1) First, it follows from the assertion (4.2) that for all $w \in \mathcal{W}$, we have

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^{\top} F(w) + \beta \left( \sum_{i=1}^{m} A_ix_i - b \right)^{\top} \sum_{i=2}^{m} A_i(x_i^k - x_i^{k+1})$$

$$\geq \Theta(v^{k+1}, v^k, v) - \Theta(v^k, v^{k-1}, v).$$

 Springer
Summarizing both sides of the above inequalities from \( k = 1, 2, \ldots, t \), we have

\[
t\theta(u) - \sum_{k=1}^{t} \theta(u^{k+1}) + \left( tw - \sum_{k=1}^{t} w^{k+1} \right)^{T} F(w) + \beta \left( \sum_{i=1}^{m} A_{i}x_{i} - b \right)^{T} \sum_{i=2}^{m} A_{i} (x_{i} - x_{i}^{t+1}) \geq \Theta(v^{t+1}, v', v) - \Theta(v^{1}, v^{0}, v).
\] (4.13)

Then, it follows from the convexity of \( \theta_{i} \) \((i = 1, \ldots, m)\) that

\[
\theta(uk^{+1}) \leq \frac{1}{t} \sum_{k=1}^{t} \theta(u^{k+1}).
\] (4.14)

Combining (4.9), (4.13) and (4.14), we have

\[
\theta(u^{k+1}) - \theta(u) + (w^{k+1} - w)^{T} F(w) \leq \frac{1}{t} \left( \Theta(v^{1}, v^{0}, v) + 2\beta \kappa (m-1) \parallel \sum_{i=1}^{m} A_{i}x_{i} - b \parallel \right).
\] (4.15)

Thus, the assertion (4.10) follows from the above inequality and the definition of \( \Theta(v^{1}, v^{0}, v) \) directly.

2) Let us define \( \bar{c}_{1} = \frac{2}{\beta_{2}} (\parallel z^{1} - z^{*} \parallel^{2} + \parallel z^{k+1} - z^{*} \parallel^{2}) \). Then, we have

\[
\left\| \sum_{i=1}^{m} A_{i}x_{i}^{k+1} - b \right\|^{2} = \left\| \frac{1}{t} \sum_{k=1}^{t} \left( \sum_{i=1}^{m} A_{i}x_{i}^{k+1} - b \right) \right\|^{2} = \frac{1}{t} \sum_{k=1}^{t} \left[ \frac{1}{\beta} (z^{k} - z^{k+1}) \right]^{2} = \frac{1}{t} \parallel z^{1} - z^{*+1} \parallel^{2} \leq \frac{2}{t^{2} \beta^{2}} \left( \parallel z^{1} - z^{*} \parallel^{2} + \parallel z^{k+1} - z^{*} \parallel^{2} \right) = \bar{c}_{1} / t^{2},
\]

where the equalities follow from (4.8), (1.2e) and Lemma 2.2, respectively. The assertion (4.11) is proved immediately.

3) It follows from \( L(u^{k+1}_{t}, z^{*}) \geq L(u^{*}, z^{*}) \) with \( L \) defined in (2.4) that

\[
\theta(u^{k+1}_{t}) - \theta(u^{*}) \geq (z^{*}, \sum_{i=1}^{m} A_{i}x_{i}^{k+1} - b) \geq -\frac{1}{2} \left( \frac{1}{t} \parallel z^{*} \parallel^{2} + t \parallel \sum_{i=1}^{m} A_{i}x_{i}^{k+1} - b \parallel^{2} \right) \geq -\frac{1}{2t} (\parallel z^{*} \parallel^{2} + \bar{c}_{1}),
\] (4.16)

where the inequalities follow from the fact \( L(u^{k+1}_{t}, z^{*}) \geq L(u^{*}, z^{*}) \), Lemma 2.3 and (4.11), respectively. On the other hand, setting \( w := w^{*} \) in (4.15), we obtain

\[
\theta(u^{k+1}_{t}) - \theta(u^{*}) + (w^{k+1}_{t} - w^{*})^{T} F(w^{*}) \leq \frac{1}{t} \Theta(v^{1}, v^{0}, v^{*}).
\]
Invoking the definition of $F$ in (2.7b), we have

$$(w_{k+1} - w^*)^T F(w^*) = -\langle z^*, \sum_{i=1}^{m} A_i x_{i,t} - b \rangle \geq -\frac{1}{2t} (\|z^*\|^2 + \bar{c}_1),$$

where the proof of the last inequality is similar to (4.16). Combining these two inequalities above, we get

$$\theta(u_{k+1}^t) - \theta(u^*) \leq \frac{1}{t} (\Theta(v^1, v^0, v^*) + \frac{1}{2} (\|z^*\|^2 + \bar{c}_1)).$$

(4.17)

The inequalities (4.16) and (4.17) indicate that the assertion (4.12) holds by setting $\bar{c}_2 := \Theta(v^1, v^0, v^*) + \frac{1}{2} (\|z^*\|^2 + \bar{c}_1)$. □

For a compact set $D \subset \mathcal{W}$ containing a solution point of the variational inequality (2.7), let us define

$$\tilde{d} := \sup \left[ 2\beta \kappa (m - 1) \| \sum_{i=1}^{m} A_i x_i - b \| + \frac{1}{2} \| v^1 - v \|^2 \right] w \in D.$$

Then, for the first $t$ iterations of the e-ADMM (1.2), the point $w_{k+1}^t$ defined in (4.8) satisfies

$$\sup_{w \in D} \left[ \theta(u_{k+1}^t) - \theta(u) + (u_{k+1}^t - w)^T F(w) \right] \leq \frac{\tilde{d} + \bar{c}}{t}. $$

(4.18)

On the other hand, invoking Theorem 3.1 and Stolz-Cesàro Theorem (see, e.g. [1]), the sequence $\{w_{k+1}^t\}$ converges to the same saddle point $w^\infty$ as the sequence $\{w_k\}$ does. Therefore, it implies that $w_{k+1}^t$ is an approximated solution of (1.1) with an accuracy of $O(1/t)$ in the sense of (4.18). Theorem 4.1 also indicates a worst-case $O(1/t)$ convergence rate of the e-ADMM (1.2) in the sense of the feasibility violation (4.11) and the decrease of the objective function (4.12).

Remark 4.1 For the special case of $m = 3$, the restriction of $\beta$ (4.1) reduces to

$$\beta \in \left( 0, \frac{\sqrt{33} - 5}{2\|A_3^\top A_3\|} \right),$$

which is larger than that in [4] for the special case of $m = 3$:

$$\beta \in \left( 0, \frac{6}{17\|A_3^\top A_3\|} \right).$$

For $m \geq 4$, we can derive a less restrictive range for $\beta$:

$$\beta \in \left( 0, \min_{3 \leq i \leq m} \frac{\mu_i}{\max\{4.5m - 11\|A_i^\top A_i\|\}} \right).$$

(4.19)
Indeed, setting \(a = b = \frac{1}{6}\) and \(\delta = \frac{1}{3} (m-2)/\epsilon' > 0\). Let \(\epsilon' \to 0^+\), we know that \(C_i > 0\) \((i = 2, \ldots, m)\) and \(\tilde{\zeta}_i > 0\) \((i = 3, \ldots, m)\) defined in (3.31)–(3.32), (4.6) when \(\beta\) satisfies (4.19). Thus, the inequality (4.2) holds with \(\tau_i = \frac{(i-3)}{2} + \frac{3(m-2)}{2}\) when \(\beta\) is restricted to (4.19). Thus, Theorem 4.1 also holds and it ensures a worst-case \(O(1/t)\) ergodic convergence rate in senses of both the variational inequality characterization (4.10) and the combination of the feasibility violation (4.11) and the decrease of the objective function (4.12).

5 Globally linear convergence under stronger conditions

In this section, we show that it is possible to theoretically derive the globally linear convergence for the e-ADMM (1.2) with \(m \geq 3\). The results in Section 3 are useful for this purpose. Note that the globally linear convergence is a strong result and thus more restrictive assumptions are needed to ensure this result. We refer to [21] for some existing results about the linear convergence of the e-ADMM (1.2) under some conditions stronger than what we shall present now. Our assumptions to ensure the globally linear convergence of the e-ADMM (1.2) are listed below.

**Assumption 5.1** In (1.1), \(\theta_1\) is convex and \(\theta_i\) \((i = 2, \ldots, m)\) are strongly convex with modulus \(\mu_i\). Moreover, one of the following conditions hold:

1) One of \(\nabla\theta_i\) \((i = 2, \ldots, m)\) is Lipschitz continuous with constant \(L_i\), the corresponding \(A_i\) is full row rank and the corresponding \(X_i = \Re^{n_i}\);

2) \(\nabla\theta_1\) is Lipschitz continuous with constant \(L_1\), \(A_1\) is nonsingular and \(X_1 = \Re^{n_1}\).

First of all, with Assumption 5.1, we can prove a result similar to (3.30), i.e.,

\[
\Phi(v(k+1), v^k, v^*) \leq \Phi(v^k, v^{k-1}, v^*) - \sum_{i=2}^{m} \hat{C}_i \|A_i x_i^{k+1} - A_i x_i^k\|^2 - \frac{1}{2\beta} \|z^k - z^{k+1}\|^2 \\
- \sum_{i=3}^{m} \hat{\zeta}_i \|A_i x_i^{k+1} - A_i x_i^*\|^2 - \mu_2 \|x_2^{k+1} - x_2^*\|^2 - \mu_2 \|x_2^{k+1} - x_2^*\|^2,
\]

where \(\Phi(v(k+1), v^k, v)\), and \(\hat{C}_i\), \(\hat{\zeta}_i\) are defined in (3.46), (3.38)–(3.40), respectively. Recall that when the penalty parameter \(\beta\) is restricted into (3.37), we know that \(\hat{C}_i\), \(\hat{\zeta}_i\) in (3.38)–(3.40) are positive. Thus, there exists a constant

\[
\varsigma := \min \left\{ \min_{3 \leq i \leq m} \{\hat{\zeta}_i\}, \min_{2 \leq i \leq m} \{\hat{C}_i\}, \frac{1}{2} \max_{1 \leq i \leq m} \frac{\max\{4m - 10, 3m - 6.5\}}{\mu_i} \|A_i^\top A_i\|, \mu_2 \right\} > 0
\]
such that
\[
\Phi(v^{k+1}, v^k, v^*) \leq \Phi(v^k, v^{k-1}, v^*)
\]
\[
-\zeta \left( \sum_{i=2}^{m} \| A_i x_i^{k+1} - A_i x_i^k \|^2 + \| z^k - z^{k+1} \|^2 + \sum_{i=3}^{m} \| A_i x_i^{k+1} - A_i x_i^* \|^2 + \| x_2^{k+1} - x_2^k \|^2 + \| x_2^k - x_2^{k+1} \|^2 \right).
\]
\(5.1\)

Indeed, according to (5.1) and the definition of \(\Phi(v^{k+1}, v^k, v)\) in (3.46), it is clear that we only need to bound the terms \(\| z^{k+1} - z^* \|^2\) in terms of the minus term in (5.1). As we show below, this is exactly why we need to assume Assumption 5.1.

**Lemma 5.1** Suppose Assumption 5.1 holds. Let \(w^*\) be a saddle point in \(W^*\) and \(\{w^k\}\) be the sequence generated by the e-ADMM (1.2) with \(m \geq 3\). Then, there exists a constant \(\sigma_1 > 0\) such that
\[
\| z^{k+1} - z^* \|^2 \leq \sigma_1 \left( \sum_{i=2}^{m} \| A_i x_i^{k+1} - A_i x_i^k \|^2 + \| x_2^{k+1} - x_2^k \|^2 + \| x_2^k - x_2^{k+1} \|^2 \right).
\]
\(5.2\)

**Proof** We consider two cases.

**Case I:** “One of \(\nabla \theta_i (i = 2, \ldots, m)\) is Lipschitz continuous with constant \(L_i\), the corresponding \(A_i\) is full row rank and the corresponding \(X_i = \mathbb{R}^{n_i}\).” The optimality conditions for the \(x_i\)- and \(x_i^*\)-subproblems are respectively:
\[
A_i^T z_i^{k+1} = \nabla \theta_i (x_i^{k+1}) + \beta A_i^T \sum_{j=i+1}^{m} (A_j x_j^k - A_j x_j^{k+1}),
\]
and
\[
- A_i^T z_i^* = -\nabla \theta_i (x_i^*).
\]
Adding these two equalities, we get
\[
\sqrt{\lambda_{\min}(A_i A_i^T)} \| z_i^{k+1} - z^* \| \leq \| A_i^T z_i^{k+1} - A_i^T z_i^* \| \leq \| \nabla \theta_i (x_i^{k+1}) - \nabla \theta_i (x_i^*) \|
\]
\[
+ \beta \| A_i \| \sum_{j=i+1}^{m} \| A_j x_j^k - A_j x_j^{k+1} \|
\]
\[
\leq L_i \| x_i^{k+1} - x^* \| + \beta \| A_i \| \sum_{j=i+1}^{m} \| A_j x_j^k - A_j x_j^{k+1} \|, \quad i = 2, \ldots, m.
\]
\(5.3\)

Then, there exists a constant \(\sigma_1 > 0\) such that conclusion (5.2) follows.
Case II: “$\nabla \theta_1$ is Lipschitz continuous with constant $L_1$, $A_1$ is nonsingular and $X_1 = \mathbb{R}^{n_1}$.” It follows from (1.2e) and the last equation in (2.6) that

$$A_1x_{k+1} - A_1x^* = \frac{1}{\beta}(z^k - z^{k+1}) - \sum_{i=2}^m (A_ix_{i+1} - A_ix^*).$$

Then, because $A_1$ is nonsingular, we have

$$\sqrt{\lambda_{\min}(A_1^T A_1)} \|x_{1+1}^k - x^*\| \leq \frac{1}{\beta} \|z^k - z^{k+1}\| + \sum_{i=3}^m \|A_ix_{i+1}^{k+1} - A_ix^*\|$$

On the other hand, similarly as (5.3), we get

$$\sqrt{\lambda_{\min}(A_1^T A_1)} \|z^{k+1} - z^*\| \leq L_1 \|x_{1+1}^k - x^*\| + \beta \|A_1\| \sum_{j=2}^m \|A_jx_j^k - A_jx_j^{k+1}\|.$$  

Combining these two inequalities, the conclusion (5.2) follows immediately. \[\square\]

Now, we can derive the globally linear convergence of the e-ADMM (1.2) under Assumptions 5.1. To compare with Theorems 3.1-3.3 in [21], we just show the linear convergence of the sequence

$$\{(A_1x_1^{k+1}, A_2x_2^{k+1}, \ldots, A_mx_m^{k+1}, z^{k+1})\}. \quad (5.4)$$

If further assumptions are assumed such as that all $A_i$ ($i = 1, \ldots, m$) are assumed to be full column rank, it is trivial to derive the linear convergence of the sequence $\{(x_1^{k+1}, x_2^{k+1}, \ldots, x_m^{k+1}, z^{k+1})\}$. We skip the details for succinctness.

**Theorem 5.1** Suppose Assumption 5.1 holds. Let $\{(x_1^{k+1}, x_2^{k+1}, \ldots, x_m^{k+1}, z^{k+1})\}$ be the sequence generated by the e-ADMM (1.2) with $m \geq 3$ and the restriction of $\beta$ (3.37). Then the sequence $\{(A_1x_1^{k+1}, A_2x_2^{k+1}, \ldots, A_mx_m^{k+1}, z^{k+1})\}$ converges linearly to a point in $\{(A_1x_1^*, A_2x_2^*, \ldots, A_mx_m^*, z^*)\}$ if $v^* \in V^*$, and $A_1x_1^* = b - \sum_{i=2}^m A_iy_i^*$.  

**Proof** First, the sequence (5.4) is bounded due to (5.1). Let $(A_1x_1^*, A_2x_2^*, \ldots, A_mx_m^*, z^*)$ be a cluster point of the sequence (5.4). Then, it can be shown that the sequence (5.4) has only one limit point by combining (5.1) with Lemma 5.1. Moreover, $v^* := (x_2^*, \ldots, x_m^*, z^*) \in V^*$ and $A_1x_1^* = b - \sum_{i=2}^m A_iy_i^*$. It follows from (5.1) that there exists a positive scalar $\sigma'$ such that

$$\Phi(v^{k+1}, v^k, v^*) \leq \sigma' \left( \sum_{i=2}^m \|A_ix_{i+1}^{k+1} - A_ix_i^k\|^2 + \|z^k - z^{k+1}\|^2 + \sum_{i=3}^m \|A_ix_{i+1}^{k+1} - A_ix_i^*\|^2 \right)$$

$$+ \|x_2^{k+1} - x_2^*\|^2 + \|x_m^{k+1} - x_m^*\|^2) \right).$$
Then, combining (5.1) with the above inequality, we obtain

$$\Phi(v^{k+1}, v^k, v^*) \leq \frac{\sigma'}{\zeta + \sigma'} \Phi(v^k, v^{k-1}, v^*).$$

It implies the Q-linearly convergence rate of the sequence \(\{\Phi(v^{k+1}, v^k, v^*)\}\). Thus, we know that the sequences \(\{\|A_i x_i^{k+1} - A_i x_i^*\|^2\} (i = 2, \ldots, m), \{\|z^{k+1} - z^*\|^2\}\) and \(\{\|A_i x_i^{k+1} - A_i x_i^k\|^2\} (i = 3, \ldots, m)\) all converges R-linearly. Recall that

$$\|z^k - z^{k+1}\|^2 \leq 2(\|z^k - z^*\|^2 + \|z^{k+1} - z^*\|^2).$$

The sequence \(\|z^k - z^{k+1}\|^2\) also converges R-linearly. Finally, it follows from (1.2e) that

$$A_1 x_1^{k+1} - A_1 x_1^* = \frac{1}{\beta}(z^k - z^{k+1}) - m \sum_{i=2}^m (A_i x_i^{k+1} - A_i x_i^*).$$

Then, combining with the above equality, and using Lemmas 2.2 and 2.4, we have

$$\|A_1 x_1^{k+1} - A_1 x_1^*\|^2 \leq 2\left(\frac{1}{\beta} \|z^k - z^{k+1}\|^2 + \| \sum_{i=2}^m (A_i x_i^{k+1} - A_i x_i^*) \|^2 \right)$$

$$\leq 2\left(\frac{1}{\beta} \|z^k - z^{k+1}\|^2 + (m - 1) \sum_{i=2}^m \|A_i x_i^{k+1} - A_i x_i^*\|^2 \right).$$

(5.5)

Therefore, the sequence \(\{\|A_1 x_1^{k+1} - A_1 x_1^*\|^2\}\) also converges R-linearly because of the R-linear convergence of the sequences \(\{\|z^k - z^{k+1}\|^2\}\) and \(\{\|A_i x_i^{k+1} - A_i x_i^*\|^2\}\) \((i = 2, \ldots, m)\). The proof is complete.

**Remark 5.1** In [21], three scenarios are considered to ensure the linear convergence of the e-ADMM (1.2). We list them in Table 1. Note that all the cases in [21] additionally require \(\lambda_i = \eta_i^{n_i} (i = 1, \ldots, m)\). Scenario 1 in Table 2 of [21] is included in our Assumption 5.1; while we can easily establish the linear convergence of the e-ADMM with (3.37) for Scenarios 2 and 3 in Table 2 of [21] by following the roadmap of the proof of Theorem 5.1. For succinctness, we omit the proof details for Scenarios 2 and 3 in Table 2 of [21]. Now we elaborate on the difference of the restrictions on \(\beta\) in Table 1. Note that the denominator of the upper bound for \(\beta\) in (3.37) is a linear function of \(m\) while that in [21] is quadratic. So it is not hard to see that our restriction on \(\beta\) (3.37) is less restrictive than those in [21]. More specifically, for example, if we consider the case of \(m = 15\) and \(\mu_i \equiv \mu (i = 2, \ldots, m)\), then the ranges of \(\beta\) in [21] for Scenarios 1 and 2 are \((0, \frac{4}{3})\) and \((0, \frac{4}{15})\), respectively, while those in (3.37) for both cases are \((0, \frac{4}{5})\). The difference of \(\beta\)'s range becomes more apparent for larger values of \(m\).
Table 1 Comparison of assumptions and restrictions on $\beta$ with [21]

| Scenario | Assumptions in [21] | Assumption 5.1 |
|----------|----------------------|----------------|
| 1.       | $X_i = \mathbb{R}^{n_i}$ ($i = 1, \ldots, n$) | General nonempty closed convex sets |
|          | $\theta_2, \ldots, \theta_m$ are strongly convex, \n|          | $\nabla \theta_m$ is Lipschitz continuous, \n|          | $A_m$ is full row rank, \n|          | $\left(0, \min \left\{ \min_{2 \leq i \leq m-1} \frac{4 \mu_i}{(2m-i)(i-1)|A_i A_i^\top|}, \frac{4 \mu_m}{(m+1)(m-2)|A_m A_m^\top|} \right\} \right)$; \n|          | $\left(0, \min_{3 \leq i \leq m} \left\{ \frac{\mu_i}{\max(4m-10,3m-6.5)|A_i A_i^\top|} \right\} \right)$; \n| 2.       | $\theta_1, \ldots, \theta_m$ are strongly convex, \n|          | $\nabla \theta_1, \ldots, \nabla \theta_m$ are Lipschitz continuous, \n|          | $\left(0, \min \left\{ \min_{2 \leq i \leq m-1} \frac{4 \mu_i}{3(2m-i)(i-1)|A_i A_i^\top|}, \frac{4 \mu_m}{(3m^2-3m-2)|A_m A_m^\top|} \right\} \right)$; \n|          | $\left(0, \min_{3 \leq i \leq m} \left\{ \frac{\mu_i}{\max(4m-10,3m-6.5)|A_i A_i^\top|} \right\} \right)$; \n| 3.       | $\theta_2, \ldots, \theta_m$ are strongly convex, \n|          | $\nabla \theta_1, \ldots, \nabla \theta_m$ are Lipschitz continuous, \n|          | $A_1$ is full column rank, \n|          | $\left(0, \min \left\{ \min_{2 \leq i \leq m-1} \frac{4 \mu_i}{3(2m-i)(i-1)|A_i A_i^\top|}, \frac{4 \mu_m}{(3m^2-3m-2)|A_m A_m^\top|} \right\} \right)$.
6 Two assertions

In this section, we construct some examples to show that the e-ADMM (1.2) with \( m \geq 3 \) are divergent if the model (1.1) or the penalty parameter \( \beta \) in (1.2) is not appropriately assumed. These examples exclude the hope of ensuring the convergence of (1.2) under too mild assumptions and to some extent justify the rationale of our assumptions for discussing the convergence of the e-ADMM (1.2) with \( m \geq 3 \). In particular, we verify the following assertions.

- The e-ADMM (1.2) may be divergent for solving (1.1) if \((m - 3)\) functions are strongly convex for any penalty parameter \( \beta > 0 \) with \( m \geq 4 \);
- The e-ADMM (1.2) may be divergent for solving (1.1) if \((m - 2)\) functions are strongly convex without any restriction on the penalty parameter \( \beta \) with \( m \geq 3 \).

6.1 Application of e-ADMM to a linear homogeneous equation

Inspired by [6], we consider a linear homogeneous equation with \( m \) variables

\[
\sum_{i=1}^{m} A_i x_i = 0, \quad \text{(6.1)}
\]

where \( A_i \in \mathbb{R}^m \) \((i = 1, \ldots, m)\) are all column vectors and the matrix \( A := [A_1, \ldots, A_m] \) is assumed to be nonsingular. Obviously, the equation (6.1) has the unique solution \( x_i = 0 \) \((i = 1, \ldots, m)\). The equation (6.1) is a special case of the model (1.1) where the objective function is null, \( b \) is the all-zero vector in \( \mathbb{R}^m \) and \( X_i = \mathbb{R} \) for \( i = 1, \ldots, m \).

Applying the e-ADMM (1.2) with \( \beta > 0 \) to (6.1), we obtain

\[
\begin{align*}
-A_1^T z^k + \beta A_1^T (A_1 x_1^{k+1} + A_2 x_2^k + \cdots + A_m x_m^k) &= 0, \\
-A_2^T z^k + \beta A_2^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + \cdots + A_m x_m^k) &= 0, \\
& \quad \cdots \quad \cdots \quad \cdots \\
-A_m^T z^k + \beta A_m^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + \cdots + A_m x_m^{k+1}) &= 0, \\
z^{k+1} &= z^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + \cdots + A_m x_m^{k+1}).
\end{align*}
\]

Introducing the new variable \( \mu^k := z^k / \beta \), the scheme (6.2) can be rewritten as

\[
\begin{pmatrix}
x_2^{k+1} \\
\vdots \\
x_m^{k+1} \\
\mu^{k+1}
\end{pmatrix}
= S
\begin{pmatrix}
x_2^k \\
\vdots \\
x_m^k \\
\mu^k
\end{pmatrix}, \quad \text{where } S = L^{-1} R,
\]

(6.3)
with
\[
L = \begin{pmatrix}
A^T A_2 & 0 & \cdots & 0 & 0 & 0_{1 \times m} \\
A_3^T A_2 & A_3^T A_3 & \cdots & 0 & 0 & 0_{1 \times m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
A_{m-1}^T A_2 & \cdots & \cdots & A_{m-1}^T A_m & 0 & 0_{1 \times m} \\
A_m^T A_2 & A_m^T A_3 & \cdots & A_m^T A_{m-1} & A_m^T A_m & 0_{1 \times m} \\
A_2 & A_3 & \cdots & A_{m-1} & A_m & I_{m \times m}
\end{pmatrix}
\] (6.4)

and
\[
R = \begin{pmatrix}
0 & -A_2^T A_3 & \cdots & \cdots & -A_2^T A_m & A_2^T \\
0 & 0 & \cdots & \cdots & -A_3^T A_m & A_3^T \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \cdots & -A_{m-1}^T A_m & A_{m-1}^T \\
0 & 0 & \cdots & \cdots & 0 & A_m^T \\
0_{m \times 1} & 0_{m \times 1} & 0_{m \times 1} & \cdots & 0_{m \times 1} & I_{m \times m}
\end{pmatrix}
\]

\[
-\frac{1}{A_1^T A_1} \begin{pmatrix}
A_2^T A_1 \\
A_3^T A_1 \\
\vdots \\
A_{m-1}^T A_1 \\
A_m^T A_1 \\
A_1^T
\end{pmatrix}
\begin{pmatrix}
-A_1^T A_2, -A_1^T A_3, \cdots, -A_1^T A_{m-1}, -A_1^T A_m, A_1^T
\end{pmatrix}.
\] (6.5)

Therefore, the e-ADMM (1.2) for solving (6.1) is divergent if the spectral radius of \(S\), denoted by \(\rho(S)\), is strictly larger than 1. Note that \(\rho(S)\) is independent of \(\beta\). That is, when the e-ADMM (1.2) is applied to the special problem (6.1), the convergence is independent of the value of \(\beta\).

Based on the analysis above, the divergence of (1.2) with \(m = 3\) for any \(\beta > 0\) has been illustrated in [6] by the example with \(A\) defined as
\[
A = (A_1, A_2, A_3) = \begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 2 \\
1 & 2 & 2
\end{pmatrix}.
\] (6.6)

For this case, we have \(\rho(S) = 1.0278 > 1\) where \(S\) is the corresponding matrix given in (6.3). We can extend the assertion to the more general case of \(m \geq 3\). Indeed, the following theorem can be easily proved by mathematical induction; thus we omit the proof.

**Theorem 6.1** For model (1.1) with \(m \geq 3\), the e-ADMM (1.2) is not necessarily convergent for any \(\beta > 0\).

**Proof** Indeed, we can use mathematical induction on \(m\) to show that for any \(m \geq 3\), there exists one specific matrix \(A^{(m)} \in \mathbb{R}^{m \times m}\) such that the corresponding iterative
matrix, i.e., $S$ in (6.3), satisfies $\rho(S) > 1$ when e-ADMM is applied to (6.1) with $A := A^{(m)}$.

6.2 The divergence of (1.2) with $(m - 3)$ strongly convex functions for any $\beta > 0$ with $m \geq 4$

Let us consider the problem

$$\min_x 0.05x_4^2$$

$$s.t. \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = 0$$

which corresponds to the model (1.1) with one strongly convex function in its objective. Let us denote the model matrix of the linear constraint in (6.7) by $\hat{A} := [\hat{A}_1, \hat{A}_2, \hat{A}_3, \hat{A}_4]$. Applying the e-ADMM (1.2) with $\beta > 0$ to (6.7), we obtain

$$\begin{cases}
-\hat{A}_1^T \mu^k + \hat{A}_1^T (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^k + \hat{A}_3 x_3^k + \hat{A}_4 x_4^k) = 0, \\
-\hat{A}_2^T \mu^k + \hat{A}_2^T (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^{k+1} + \hat{A}_3 x_3^k + \hat{A}_4 x_4^k) = 0, \\
-\hat{A}_3^T \mu^k + \hat{A}_3^T (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^{k+1} + \hat{A}_3 x_3^{k+1} + \hat{A}_4 x_4^k) = 0, \\
-\hat{A}_4^T \mu^k + \hat{A}_4^T (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^{k+1} + \hat{A}_3 x_3^{k+1} + \hat{A}_4 x_4^{k+1}) + \frac{0.1}{\beta} x_4^{k+1} = 0,
\end{cases}$$

Consequently, we have

$$\begin{cases}
\hat{A}_1^T \mu^k / (1 + \frac{0.1}{\beta}), \\
\mu^{k+1} = \mu^k - (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^{k+1} + \hat{A}_3 x_3^{k+1} + \hat{A}_4 x_4^{k+1}),
\end{cases}$$

Moreover, if we set $\hat{\mu} := \mu_{[1;3]}$, and recall the definition of $A$ in (6.6), we get

$$\begin{cases}
-\hat{A}_3^T \hat{\mu}^k + \hat{A}_3^T (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^k + \hat{A}_3 x_3^k) = 0, \\
-\hat{A}_3^T \hat{\mu}^k + \hat{A}_3^T (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^{k+1} + \hat{A}_3 x_3^k) = 0, \\
-\hat{A}_3^T \hat{\mu}^k + \hat{A}_3^T (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^{k+1} + \hat{A}_3 x_3^{k+1}) = 0, \\
\hat{\mu}^{k+1} = \hat{\mu}^k - (\hat{A}_1 x_1^{k+1} + \hat{A}_2 x_2^{k+1} + \hat{A}_3 x_3^{k+1}), \\
\hat{\mu}^{k+1} = \hat{\mu}^k - x_4^{k+1}.
\end{cases}$$

Also, we denote $y^T := (x_2^T, x_3^T, \hat{\mu}^T)$. Then, the iterative scheme (6.8) is

$$\begin{pmatrix} y_4^{k+1} \\ \mu_{[4]}^{k+1} \\ x_4^{k+1} \end{pmatrix} = \begin{pmatrix} S & 0_{5 \times 1} & 0_{5 \times 1} \\ 0_{1 \times 5} & 1 & -1 \\ 0_{1 \times 5} & 1/(1 + \frac{0.1}{\beta}) & 0 \end{pmatrix} \begin{pmatrix} y_4^k \\ \mu_{[4]}^k \\ x_4^k \end{pmatrix},$$

(6.9)
where $S$ is the matrix given in (6.3) when the e-ADMM is applied to (6.1) with $A$ defined in (6.6). Let $\hat{S}(\beta)$ be the coefficient matrix given in (6.9), which is clearly dependent on $\beta$. Then, we have $\rho(\hat{S}(\beta)) = \rho(S) = 1.0278 > 1$ for any $\beta > 0$. In fact, we have $\rho(\hat{S}(\beta)) = \rho(S)$ because the absolute value or the modulus of the maximum eigenvalue of the $2 \times 2$ submatrix in the lower right corner of the coefficient matrix in (6.9) is less than 1. Hence, the e-ADMM (1.2) may be divergent for any penalty parameter $\beta > 0$ if $m = 4$ and there is one strongly convex function. We extend the conclusion to the general case of $m \geq 4$ in the following theorem.

**Theorem 6.2** For model (1.1) with $m \geq 4$, the e-ADMM (1.2) is not necessarily convergent for an arbitrarily fixed $\beta > 0$ if there are $(m - 3)$ strongly convex functions in the objective of (1.1).

**Proof** For any $m \geq 4$, we consider the following convex programming:

$$
\min_{\tilde{A}_1, \ldots, \tilde{A}_m} \frac{\alpha_i}{2} \sum_{i=4}^{m} x_i^2 \quad \text{s.t.} \quad \sum_{i=1}^{m} \tilde{A}_i x_i = 0, \quad (6.10)
$$

where $\tilde{A} := [\tilde{A}_1, \ldots, \tilde{A}_m] \in \mathbb{R}^{m \times m}$ or rewritten as

$$
\tilde{A} = \begin{pmatrix} A & 0_{3 \times (m-3)} \\ 0_{(m-3) \times 3} & I_{(m-3) \times (m-3)} \end{pmatrix},
$$

where the $3 \times 3$ matrix $A$ is defined in (6.6). Applying the e-ADMM (1.2) with $\beta > 0$ to (6.10), we obtain

$$
\begin{bmatrix}
-\tilde{A}_1^T \mu^k + \tilde{A}_1^T (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^k + \cdots + \tilde{A}_m x_m^k) = 0, \\
-\tilde{A}_2^T \mu^k + \tilde{A}_2^T (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^{k+1} + \cdots + \tilde{A}_m x_m^k) = 0, \\
-\tilde{A}_3^T \mu^k + \tilde{A}_3^T (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^{k+1} + \cdots + \tilde{A}_m x_m^k) = 0, \\
\vdots \\
-\tilde{A}_m^T \mu^k + \tilde{A}_m^T (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^{k+1} + \cdots + \tilde{A}_m x_m^{k+1}) + \sigma_i x_i^{k+1} = 0,
\end{bmatrix}
\mu^{k+1} = \mu^k - (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^{k+1} + \tilde{A}_3 x_3^{k+1} + \cdots + \tilde{A}_m x_m^{k+1}).
$$

Note that we have

$$
\tilde{A}_j^T \tilde{A}_i = 0, \quad \text{when} \quad \begin{cases} i = 1, 2, 3 \text{ and } j > 3, \\
\text{i, j > 3 and } i \neq j, \\
\text{j = 1, 2, 3 and } i > 3.
\end{cases}
$$

Consequently, it implies that

$$
\begin{bmatrix}
-\tilde{A}_1^T \mu^k + \tilde{A}_1^T (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^k + \tilde{A}_3 x_3^k) = 0, \\
-\tilde{A}_2^T \mu^k + \tilde{A}_2^T (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^{k+1} + \tilde{A}_3 x_3^k) = 0, \\
-\tilde{A}_3^T \mu^k + \tilde{A}_3^T (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^{k+1} + \tilde{A}_3 x_3^{k+1}) = 0, \\
\end{bmatrix}
\mu^{k+1} = \mu^k - (\tilde{A}_1 x_1^{k+1} + \tilde{A}_2 x_2^{k+1} + \tilde{A}_3 x_3^{k+1} + \cdots + \tilde{A}_m x_m^{k+1}).
$$
Moreover, setting $\hat{\mu} := \mu_{[1:3]}$ and recall the definition of $A$ in (6.6), we get

$$
\begin{align*}
- A_1^T \hat{\mu}^k + A_1^T (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k) &= 0, \\
- A_2^T \hat{\mu}^k + A_2^T (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k) &= 0, \\
- A_3^T \hat{\mu}^k + A_3^T (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k) &= 0,
\end{align*}
$$

(6.11)

Let us denote $y^T := (x_2^T, x_3^T, \hat{\mu}^T)$. Then, the iterative scheme (6.11) can be written as

$$
\begin{pmatrix}
y^{k+1} \\
\mu_{[4]}^{k+1} \\
\vdots \\
\mu_{[m]}^{k+1} \\
x_4^{k+1} \\
\vdots \\
x_m^{k+1}
\end{pmatrix} =
\begin{pmatrix}
S & 0_{5 \times (m-3)} & 0_{5 \times (m-3)} \\
0_{(m-3) \times 5} & I_{(m-3) \times (m-3)} & -I_{(m-3) \times (m-3)} \\
0_{(m-3) \times 5} & D_{(m-3) \times (m-3)} & 0_{(m-3) \times (m-3)}
\end{pmatrix}
\begin{pmatrix}
y^k \\
\mu_{[4]}^k \\
\vdots \\
\mu_{[m]}^k \\
x_4^k \\
\vdots \\
x_m^k
\end{pmatrix},
$$

(6.12)

where

$$
D_{(m-3) \times (m-3)} = \text{diag}(1/(1 + \sigma_4/\beta), \ldots, 1/(1 + \sigma_m/\beta)).
$$

It can be easily shown that the absolute value of the maximal eigenvalue of

$$
\begin{pmatrix}
I_{(m-3) \times (m-3)} & -I_{(m-3) \times (m-3)} \\
D_{(m-3) \times (m-3)} & 0_{(m-3) \times (m-3)}
\end{pmatrix}
$$

is less than 1. Thus, for the coefficient matrix in (6.12), denoted by $\tilde{S}(\beta)$, we have

$$
\rho(\tilde{S}(\beta)) = \rho(S) = 1.0278 > 1, \quad \forall \beta > 0.
$$

The proof is complete.

\[\square\]

### 6.3 The divergence of (1.2) with $(m - 2)$ strongly convex functions and without restriction on $\beta$ for $m \geq 3$

Then, we show that the e-ADMM (1.2) may be divergent even if there are $(m - 2)$ strongly convex functions in the objective of (1.1) while there is no restriction on $\beta$ for the case of $m \geq 3$. We consider the model

$$
\min \frac{1}{2} a_3 x_3^2 \\
s.t. \ A_1 x_1 + A_2 x_2 + A_3 x_3 = 0,
$$

(6.13)
where $A_i \in \mathbb{R}^3$ ($i = 1, 2, 3$) are all column vectors, and the matrix $A := [A_1, A_2, A_3]$ is assumed to be nonsingular, and the scalar $a_3$ is positive. It is easy to see that each iteration of (1.2) applied to (6.13) can be characterized by a matrix iteration (6.3) with the iterative matrix $S$ defined as:

$$
\hat{S} = \hat{L}^{-1}R,
$$

where the matrix $R$ is defined in (6.5) with $m = 3$, and $\hat{L}$ is defined below:

$$
\hat{L} = \begin{pmatrix}
A_2^T A_2 & 0 & 0 \\
A_3^T A_2 & A_3^T A_3 + a_3/\beta & 0 \\
A_2 & A_3 & I_{3\times 3}
\end{pmatrix}.
$$

Specifically, let us take

$$
A = [A_1, A_2, A_3] = \begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 2 \\
1 & 2 & 2
\end{pmatrix}
$$

and $a_3 = 0.05$.

With simple calculations, if we take $\beta = 1$, then we have $\rho(\hat{S}) = 1.0259$ for the matrix $\hat{S}$ given in (6.14). On the other hand, for this example, we have $m = 3$, $f_3(x_3) = 0.025x_3^2$ and $\mu_3 = 0.05$. According to Theorem 3.1, the e-ADMM (1.2) is guaranteed to be convergent when $\beta \in (0, 0.05/22.5)$. Indeed, we calculate that $\rho(\hat{S}) = 0.9781$ when $\beta = 0.0022 \in (0, 0.05/22.5)$. Now, we extend the conclusion to the general case of $m \geq 3$. This theorem also shows that it is necessary to appropriately restrict the penalty parameter $\beta$ when discussing the convergence of the e-ADMM (1.2)

**Theorem 6.3** For model (1.1) with $m \geq 3$, the e-ADMM (1.2) is not necessarily convergent for all $\beta > 0$ when there are $(m - 2)$ strongly convex functions in its objective.

**Proof** In the following, we show that for any $\beta > 0$, there exist examples such that the e-ADMM (1.2) is divergent. First, it follows from Theorem 6.1 that there exists a specific matrix $A^{(m)} \in \mathbb{R}^{m \times m}$ with $m \geq 3$ such that the e-ADMM (1.2) is divergent when it is applied to the equation (6.1) with $A := A^{(m)}$. It implies that the corresponding matrix $S^{(m)}$ given in (6.3) has $\rho(S^{(m)}) > 1$. Recall the matrices $L$ and $R$ composing $S^{(m)}$ by $S^{(m)} = L^{-1}R$ (see (6.4) and (6.5)). Then, with this specific choice of $A^{(m)}$, we consider the following problem:

$$
\min_{\sigma} \frac{\sigma}{2} \sum_{i=3}^{m} x_i^2 \\
\text{s.t. } \sum_{i=1}^{m} A_i^{(m)} x_i = 0,
$$

where $A^{(m)} = [A_1^{(m)}, \ldots, A_m^{(m)}]$, $\sigma > 0$ and there are $(m - 2)$ strongly convex functions in its objective. One can show that each iteration of (1.2) applied to (6.15)
can be characterized by a matrix iteration (6.3) with the iterative matrix \( \tilde{S}^{(m)} \) defined as \( \tilde{S}^{(m)} = \tilde{L}^{-1} R \) with \( R \) define in (6.5) and \( \tilde{L} \) defined by

\[
\tilde{L} = \begin{pmatrix}
A_2^T A_2 & 0 & \cdots & 0 & 0 & 0 \\
A_3^T A_2 & A_3^T A_3 + \sigma/\beta & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
A_{m-1}^T A_2 & \cdots & \cdots & A_{m-1}^T A_{m-1} + \sigma/\beta & 0 & 0 \\
A_m^T A_2 & A_m^T A_3 & \cdots & A_m^T A_{m-1} & A_m^T A_m + \sigma/\beta & 0 \\
A_2 & A_3 & \cdots & A_{m-1} & A_m & I_{m \times m}
\end{pmatrix}
\]

Set \( \beta = 1 \) in the e-ADMM (1.2). Thus, there exists a positive scalar \( \tilde{\sigma}_1 \) such that \( \sigma \| L^{-1} \| < 1 \) when \( \sigma \in (0, \tilde{\sigma}_1) \). Then, let \( E := \tilde{L} - L = \text{diag}(0, \sigma, \cdots, \sigma, 0_{1 \times m}) \), we have

\[
\| I - (I + L^{-1} E)^{-1} \| = \| \sum_{k=1}^{\infty} (-1)^{k+1} (L^{-1} E)^k \| \leq \sum_{k=1}^{\infty} \| L^{-1} E \|^k
\]

\[
= \frac{\| L^{-1} E \|}{1 - \| L^{-1} E \|} \leq \| E \| \| L^{-1} \| = \sigma \| L^{-1} \|.
\]

Next, define \( \Delta_S := \tilde{S}^{(m)} - S^{(m)} \). Then, we have

\[
\| \Delta_S \| = \| - (I - (I + L^{-1} E)^{-1}) L^{-1} R \| \leq \| I - (I + L^{-1} E)^{-1} \| \| L^{-1} R \|
\]

\[
\leq \sigma \| L^{-1} \| \| L^{-1} R \|. \quad (6.16)
\]

Moreover, there exist a sufficient small scalar \( 0 < \varepsilon < 1 \) and a positive constant \( \tilde{\sigma}_2 \) such that \( \| \Delta_S \| < \varepsilon \) when \( \sigma \in (0, \tilde{\sigma}_2) \). Then, invoking Theorem 2 in [2], there exist a positive scalar \( \kappa \) dependent on \( S^{(m)} = L^{-1} R \) and an eigenvalue of \( \tilde{S}^{(m)} \) such that

\[
| \lambda_{\text{max}}(S^{(m)}) - \lambda(\tilde{S}^{(m)}) | \leq \kappa \| \Delta_S \|^{\frac{1}{2m-1}} \leq \sigma^{\frac{1}{2m-1}} \kappa (\| L^{-1} \| \| L^{-1} R \|)^{\frac{1}{2m-1}}, \quad (6.17)
\]

where \( \lambda_{\text{max}}(\cdot) \) represents the eigenvalue with the maximum absolute value, \( \lambda(\tilde{S}^{(m)}) \) denotes one of the eigenvalues of \( \tilde{S}^{(m)} \) and the last inequality is due to (6.16). Then, the right-hand side of the above inequality only depends on \( \sigma \) since \( \kappa (\| L^{-1} \| \| L^{-1} R \|)^{\frac{1}{2m-1}} \) is a constant. Therefore, there exists a sufficient small \( \hat{\sigma} \) (e.g., \( 0 < \hat{\sigma} < \min(\tilde{\sigma}_1, \tilde{\sigma}_2) \)) such that \( \rho(S^{(m)}) - (\kappa \hat{\sigma} \| L^{-1} \| \| L^{-1} R \|)^{\frac{1}{2m-1}} > 1 \) whenever \( \rho(S^{(m)}) > 1 \). As a consequence, we have \( \rho(\tilde{S}^{(m)}) > 1 \) due to (6.17). This implies that the e-ADMM (1.2) with \( \beta = 1 \) is divergent when solving (6.15) with setting \( \sigma := \hat{\sigma} \). Indeed, for any \( \beta > 0 \), we can construct one specific problem defined in (6.15), i.e., finding a appropriate \( \sigma \), such that the e-ADMM (1.2) with this \( \beta \) is divergent. Note (6.15) is a special case of (1.1) with \( (m - 2) \) strongly convex functions in its objective.

\[ \square \]

7 Conclusions

In this paper, we conduct the convergence analysis for the direct extension of ADMM (“e-ADMM”) for solving a separable convex minimization model whose objective
function is the sum of \( m \) functions without coupled variables. We extend some existing results for the special case of \( m = 3 \) to the general case of \( m \geq 3 \), and prove the convergence of the e-ADMM when \((m - 2)\) functions are assumed to be strongly convex and the penalty parameter is appropriately restricted. Even for the special case of \( m = 3 \), it turns out that the penalty parameter in our analysis is less restricted than some existing results that are analyzed specifically for this special case. The worst-case convergence rate measured by iteration complexity and globally linear convergence are also derived under some additional assumptions. Finally, we show that the conditions, i.e., \( m - 2 \) strongly convex functions and \( \beta \) is appropriately restricted, are necessary to ensure the convergence of the e-ADMM for the general problem (1.1).

References

1. Ash, J.M., Berele, A., Catoiu, S.: Plausible and Genuine Extensions of L’Hôpital’s Rule. Math. Mag. 85, 52–60 (2012)
2. Bhatia, R., Elsner, L., Krause, G.: Bounds for the Variation of the Roots of a Polynomial and the Eigenvalues of a Matrix. Linear Algebra Appl. 142, 195–209 (1990)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach Learn. 3, 1–122 (2010)
4. Cai, X.J., Han, D.R., Yuan, X.M.: The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex. Comput. Optim. Appl. 66, 39–73 (2017)
5. 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, Article ID 183961, 7 pages
6. 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., Ser. A 155, 57–79 (2016)
7. Davis, D., Yin, W.T.: A three-operator splitting scheme and its optimization applications, available at https://arxiv.org/pdf/1504.01032v1.pdf
8. Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. Pacific. J. Optim. 11, 619–644 (2015)
9. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I, Springer Series in Operations Research, Springer-Verlag (2003)
10. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems, pp. 299–331. North-Holland Publishing Co., Amsterdam (1983)
11. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. 2, 17–40 (1976)
12. Garling, D.J.H.: Inequalities: A Journey into Linear Analysis. Cambridge University, United Kindom (2007)
13. Glowinski, R., Marrocco, A.: Sur l’approximation par éléments finis d’ordre un et l’approximation par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. Revue Fr. Autom. Inform. Rech. Opér., Anal Numér 2, 41–76 (1975)
14. Glowinski, R.: On alternating direction methods of multipliers: A historical perspective. In: Fitzgibbon, W., Kuznetsov, Y.A., Neittaanmaki, P., Pironneau, O. (eds.) Modeling, Simulation and Optimization for Science and Technology, pp. 59–82. Springer, Netherlands (2014)
15. Han, D.R., Yuan, X.M.: A note on the alternating direction method of multipliers. J. Optim. Theory Appl. 155, 227–238 (2013)
16. He, B.S., Yuan, X.M.: On the \( O(1/n) \) convergence rate of Douglas-Rachford alternating direction method. SIAM J. Numer. Anal. 50, 700–709 (2012)
17. He, B.S., Yuan, X.M.: On nonergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. Numer. Math. 130, 567–577 (2015)
18. Hong, M., Luo, Z.Q.: On the linear convergence of alternating direction method of multipliers. Math. Program. 162, 165–199 (2017)

19. 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), 19 (2015)

20. 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)

21. 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)

22. 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 Tran. Pattern Anal. Mach. Intel. 34, 2233–2246 (2012)

23. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)

24. Sun, D.F., Toh, K.C., Yang, L.Q.: A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type constraints. SIAM J. Optim. 25, 882–915 (2015)

25. 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)

26. Wang, X.F., Hong, M.Y., Chang, T.H., Razaviyayn, M., Luo, Z.Q.: Joint day-ahead power procurement and load scheduling using stochastic alternating direction method of multipliers. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 7754–7758 (2014)

27. 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)