On the Convergence Rate of Multi-Block ADMM

Tianyi Lin* Shiqian Ma* Shuzhong Zhang†

March 30, 2014

Abstract

The alternating direction method of multipliers (ADMM) is widely used in solving structured convex optimization problems. Despite of its success in practice, the convergence properties of the standard ADMM for minimizing the sum of N ($N \geq 3$) convex functions with N block variables linked by linear constraints, have remained unclear for a very long time. In this paper, we present convergence and convergence rate results for the standard ADMM applied to solve N-block ($N \geq 3$) convex minimization problem, under the condition that one of these functions is convex (not necessarily strongly convex) and the other $N-1$ functions are strongly convex. Specifically, in that case the ADMM is proven to converge with rate $O(1/t)$ in a certain ergodic sense, and $o(1/t)$ in non-ergodic sense, where t denotes the number of iterations.

Keywords: Alternating Direction Method of Multipliers, Convergence Rate, Convex Optimization

1 Introduction

We consider solving the following multi-block convex minimization problem:

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + \cdots + f_N(x_N) \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + \cdots + A_N x_N = b \\ & x_i \in \mathcal{X}_i,\ i = 1, \ldots, N, \end{array} \tag{1.1}$$

where $A_i \in \mathbf{R}^{p \times n_i}$, $b \in \mathbf{R}^p$, $\mathcal{X}_i \subset \mathbf{R}^{n_i}$ are closed convex sets, and $f_i : \mathbf{R}^{n_i} \to \mathbf{R}^p$ are closed convex functions. One recently popular way to solve (1.1), when the functions f_i's are of special structures, is to apply the ADMM. The ADMM is closely related to the Douglas-Rachford [6] and Peaceman-Rachford [22] operator splitting methods that date back to 1950s. These operator splitting methods were further studied later in [20, 10, 12, 7]. The ADMM has been revisited recently due to its success in solving problems with special structures arising from compressed sensing, machine learning, image processing, and so on; see the recent survey papers [2, 8] for more information.

*Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China.

†Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, USA.

ADMM for solving (1.1) is based on an augmented Lagrangian method framework. The augmented Lagrangian function for (1.1) is defined as

$$\mathcal{L}_\gamma(x_1,\ldots,x_N;\lambda) := \sum_{j=1}^N f_j(x_j) - \left\langle \lambda, \sum_{j=1}^N A_j x_j - b \right\rangle + \frac{\gamma}{2}\left\| \sum_{j=1}^N A_j x_j - b \right\|^2,$$

where λ is the Lagrange multiplier and $\gamma > 0$ is a penalty parameter. In a typical iteration of the standard ADMM for solving (1.1), the following updating procedure is implemented:

$$\left\{\begin{array}{lcl} x_1^{k+1} & := & \mathrm{argmin}_{x_1\in\mathcal{X}_1} \ \mathcal{L}_\gamma(x_1, x_2^k, \ldots, x_N^k; \lambda^k) \\ x_2^{k+1} & := & \mathrm{argmin}_{x_2\in\mathcal{X}_2} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2, x_3^k, \ldots, x_N^k; \lambda^k) \\ & \vdots & \\ x_N^{k+1} & := & \mathrm{argmin}_{x_N\in\mathcal{X}_N} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, \ldots, x_{N-1}^{k+1}, x_N; \lambda^k) \\ \lambda^{k+1} & := & \lambda^k - \gamma\left(\sum_{j=1}^N A_j x_j^{k+1} - b\right). \end{array}\right. \tag{1.2}$$

Note that the standard ADMM (1.2) minimizes in each iteration the augmented Lagrangian function with respect to $x_1,\ldots,x_N$ alternatingly in a Gauss-Seidel manner. The ADMM (1.2) for solving two-block convex minimization problems (i.e., $N=2$) has been studied extensively in the literature. The global convergence of ADMM (1.2) when $N=2$ has been shown in [11, 9]. There are also some very recent works that study the convergence rate properties of ADMM when $N=2$ (see, e.g., [18, 21, 5, 1, 17]).

However, the convergence of ADMM (1.2) when $N\geq 3$ has remained unclear for a long time. Some recent progresses on solving this problem are the followings. In [13], it is shown that ADMM (1.2) converges if all the functions $f_1,\ldots,f_N$ are further assumed to be strongly convex. In a very recent work by Chen et al. [3], a counter-example is constructed that shows the failure of ADMM (1.2) when $N\geq 3$. In fact, it is shown in [3] that for a given γ, it is possible to construct a counter-example such that ADMM fails. The authors in [3] also give some sufficient condition to guarantee the convergence of ADMM (1.2). The sufficient condition given in [3] for $N=3$, is that there are two matrices that are orthogonal, i.e., $A_1^\top A_2 = 0$, or $A_2^\top A_3 = 0$, or $A_1^\top A_3 = 0$. This condition is essentially very restrictive, and rules out many interesting applications. In a recent work by Hong and Luo [19], a variant of ADMM (1.2) with small step size in updating the Lagrange multiplier is studied. Specifically, [19] proposes to replace the last equation in (1.2) to

$$\lambda^{k+1} := \lambda^k - \alpha\gamma\left(\sum_{j=1}^N A_j x_j^{k+1} - b\right),$$

where $\alpha > 0$ is a small step size. Linear convergence of this variant is proved under the assumption that the objective function satisfies certain error bound conditions. However, it is noted that the selection of α is in fact bounded by some parameters associated with the error bound conditions to guarantee the convergence. Therefore, it might be difficult to choose α in practice. There are also studies on the convergence rate of some other variants of ADMM (1.2), and we refer the interested readers to [15, 16, 14, 4] for details of these variants. In this paper, we focus on the *original standard ADMM* (1.2).

Our contribution. The main contribution in this paper lies in the followings. We show that the standard ADMM (1.2) when $N \geq 3$ converges with rate $O(1/t)$ in ergodic sense and $o(1/t)$ in non-ergodic sense, under the assumption that $f_2, \ldots, f_N$ are strongly convex and f_1 is convex but not necessarily strongly convex. It should be pointed out that our assumption is weaker than the one used in [13], in which all the functions are required to be strongly convex. Moreover, unlike the sufficient condition suggested in [3], we do not make any assumption on the matrices $A_1, \ldots, A_N$. To the best of our knowledge, the convergence rate results given in this paper are the first sublinear convergence rate results for standard ADMM when $N \geq 3$.

Organization. The rest of this paper is organized as follows. In Section 2 we provide some preliminaries for our convergence rate analysis. In Section 3, we prove the convergence rate of ADMM (1.2) in the ergodic sense. In Section 4, we prove the convergence rate of ADMM (1.2) in the non-ergodic sense. Section 5 draws some conclusions and points out some future directions.

2 Preliminaries

We will only prove the convergence results of ADMM for $N = 3$, because all the analysis can be extended to arbitrary N easily. As a result, for the ease of presentation and succinctness, we assume $N = 3$ in the rest of this paper. We will present the results for general N but omit the proofs.

We restate the problem (1.1) for $N = 3$ as

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + f_3(x_3) \\ \text{s.t.} & A_1x_1 + A_2x_2 + A_3x_3 = b \\ & x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, x_3 \in \mathcal{X}_3. \end{array} \tag{2.1}$$

The ADMM for solving (2.1) can be summarized as (note that some constant terms in the three subproblems are discarded):

$$x_1^{k+1} := \operatorname*{argmin}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \frac{\gamma}{2}\|A_1x_1 + A_2x_2^k + A_3x_3^k - b - \frac{1}{\gamma}\lambda^k\|^2 \tag{2.2}$$

$$x_2^{k+1} := \operatorname*{argmin}_{x_2 \in \mathcal{X}_2} f_2(x_2) + \frac{\gamma}{2}\|A_1x_1^{k+1} + A_2x_2 + A_3x_3^k - b - \frac{1}{\gamma}\lambda^k\|^2 \tag{2.3}$$

$$x_3^{k+1} := \operatorname*{argmin}_{x_3 \in \mathcal{X}_3} f_3(x_3) + \frac{\gamma}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3 - b - \frac{1}{\gamma}\lambda^k\|^2 \tag{2.4}$$

$$\lambda^{k+1} := \lambda^k - \gamma\left(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b\right). \tag{2.5}$$

The first-order optimality conditions for (2.2)-(2.4) are given respectively by $x_i^{k+1} \in \mathcal{X}_i, i = 1, 2, 3$, and

$$(x_1 - x_1^{k+1})^\top \left[g_1(x_1^{k+1}) - A_1^\top\lambda^k + \gamma A_1^\top(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b)\right] \geq 0, \quad \forall x_1 \in \mathcal{X}_1, \tag{2.6}$$

$$(x_2 - x_2^{k+1})^\top \left[g_2(x_2^{k+1}) - A_2^\top\lambda^k + \gamma A_2^\top(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b)\right] \geq 0, \quad \forall x_2 \in \mathcal{X}_2, \tag{2.7}$$

$$(x_3 - x_3^{k+1})^\top \left[g_3(x_3^{k+1}) - A_3^\top\lambda^k + \gamma A_3^\top(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)\right] \geq 0, \quad \forall x_3 \in \mathcal{X}_3, \tag{2.8}$$

where $g_i \in \partial f_i$ is the subdifferential of f_i for $i = 1, 2, 3$. Moreover, by combining with (2.5), (2.6)-(2.8) can be rewritten as

$$(x_1 - x_1^{k+1})^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} + \gamma A_1^\top A_2(x_2^k - x_2^{k+1}) + \gamma A_1^\top A_3(x_3^k - x_3^{k+1}) \right] \geq 0, \quad \forall x_1 \in \mathcal{X}_1, \quad (2.9)$$

$$(x_2 - x_2^{k+1})^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} + \gamma A_2^\top A_3(x_3^k - x_3^{k+1}) \right] \geq 0, \quad \forall x_2 \in \mathcal{X}_2, \quad (2.10)$$

$$(x_3 - x_3^{k+1})^\top \left[g_3(x_3^{k+1}) - A_3^\top \lambda^{k+1} \right] \geq 0, \quad \forall x_3 \in \mathcal{X}_3. \quad (2.11)$$

We denote $\Omega = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathbf{R}^p$ and the optimal set of (2.1) as Ω^*, and the following assumption is made throughout this paper.

Assumption 2.1 *The optimal set Ω^* for problem (2.1) is non-empty.*

According to the first-order optimality conditions for (2.1), solving (2.1) is equivalent to finding

$$(x_1^*, x_2^*, x_3^*, \lambda^*) \in \Omega^*$$

such that the following holds:

$$\begin{cases} (x_1 - x_1^*)^\top (g_1(x_1^*) - A_1^\top \lambda^*) \geq 0, \forall x_1 \in \mathcal{X}_1, \\ (x_2 - x_2^*)^\top (g_2(x_2^*) - A_2^\top \lambda^*) \geq 0, \forall x_2 \in \mathcal{X}_2, \\ (x_3 - x_3^*)^\top (g_3(x_3^*) - A_3^\top \lambda^*) \geq 0, \forall x_3 \in \mathcal{X}_3, \\ A_1 x_1^* + A_2 x_2^* + A_3 x_3^* - b = 0, \end{cases} \quad (2.12)$$

where $g_i(x_i^*) \in \partial f_i(x_i^*)$, $i = 1, 2, 3$.

Furthermore, the following condition is assumed in our subsequent analysis.

Assumption 2.2 *The functions f_2 and f_3 are strongly convex with parameters $\sigma_2 > 0$ and $\sigma_3 > 0$, respectively; i.e., the following two inequalities hold:*

$$f_2(y) \geq f_2(x) + (y - x)^\top g_2(x) + \frac{\sigma_2}{2} \|y - x\|^2, \quad \forall x, y \in \mathcal{X}_2, \quad (2.13)$$

$$f_3(y) \geq f_3(x) + (y - x)^\top g_3(x) + \frac{\sigma_3}{2} \|y - x\|^2, \quad \forall x, y \in \mathcal{X}_3, \quad (2.14)$$

or equivalently,

$$(y - x)^\top (g_2(y) - g_2(x)) \geq \sigma_2 \|y - x\|^2, \quad \forall x, y \in \mathcal{X}_2, \quad (2.15)$$

$$(y - x)^\top (g_3(y) - g_3(x)) \geq \sigma_3 \|y - x\|^2, \quad \forall x, y \in \mathcal{X}_3, \quad (2.16)$$

where $g_2(x) \in \partial f_2(x)$ and $g_3(x) \in \partial f_3(x)$ are the subdifferentials of f_2 and f_3 respectively.

In our analysis, the following two well-known identities are used frequently,

$$(w_1 - w_2)^\top (w_3 - w_4) = \frac{1}{2} \left(\|w_1 - w_4\|^2 - \|w_1 - w_3\|^2 \right) + \frac{1}{2} \left(\|w_3 - w_2\|^2 - \|w_4 - w_2\|^2 \right), \quad (2.17)$$

$$(w_1 - w_2)^\top (w_3 - w_1) = \frac{1}{2} \left(\|w_2 - w_3\|^2 - \|w_1 - w_2\|^2 - \|w_1 - w_3\|^2 \right). \quad (2.18)$$

Notations. For simplicity, we use the following notation to denote the stacked vectors or tuples:

$$u = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, u^k = \begin{pmatrix} x_1^k \\ x_2^k \\ x_3^k \end{pmatrix}, u^* = \begin{pmatrix} x_1^* \\ x_2^* \\ x_3^* \end{pmatrix}.$$

We denote by $f(u) \equiv f_1(x_1)+f_2(x_2)+f_3(x_3)$ the objective function of problem (2.1); g_i is a subgradiatent of f_i; $\lambda_{\max}(B)$ denotes the largest eigenvalue of a real symmetric matrix B; $\|x\|$ denotes the Euclidean norm of x.

3 Ergodic Convergence Rate of ADMM

In this section, we prove the $O(1/t)$ convergence rate of ADMM (2.2)-(2.5) in the ergodic sense.

Lemma 3.1 *Assume that* $\gamma \le \min\left\{\frac{\sigma_2}{2\lambda_{\max}(A_2^\top A_2)}, \frac{\sigma_3}{2\lambda_{\max}(A_3^\top A_3)}\right\}$, *where* σ_2 *and* σ_3 *are defined in Assumption 2.2. Let* $(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by ADMM from given* $(x_2^k, x_3^k, \lambda^k)$. *Then, for any* $u^* = (x_1^*, x_2^*, x_3^*) \in \Omega^*$ *and* $\lambda \in \mathbf{R}^p$, *it holds that*

$$\begin{aligned} & f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ x_3^* - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \lambda^{k+1} \\ -A_2^\top \lambda^{k+1} \\ -A_3^\top \lambda^{k+1} \\ A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \end{pmatrix} \\ & + \frac{1}{2\gamma}\left(\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2\right) + \frac{\gamma}{2}\left(\|A_1 x_1^* + A_2 x_2^* + A_3 x_3^k - b\|^2 - \|A_1 x_1^* + A_2 x_2^* + A_3 x_3^{k+1} - b\|^2\right) \\ & + \frac{\gamma}{2}\left(\|A_1 x_1^* + A_2 x_2^k + A_3 x_3^k - b\|^2 - \|A_1 x_1^* + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2\right) \\ \ge\ & \frac{\gamma}{2}\|A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b\|^2. \end{aligned} \tag{3.1}$$

Proof. Note that combining (2.9)-(2.11) yields

$$\begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ x_3 - x_3^{k+1} \end{pmatrix}^\top \left[\begin{pmatrix} g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} \\ g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} \\ g_3(x_3^{k+1}) - A_3^\top \lambda^{k+1} \end{pmatrix} + \begin{pmatrix} \gamma A_1^\top A_2 & \gamma A_1^\top A_3 \\ 0 & \gamma A_2^\top A_3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_2^k - x_2^{k+1} \\ x_3^k - x_3^{k+1} \end{pmatrix} \right] \ge 0. \tag{3.2}$$

The key step in our proof is to bound the following two terms

$$(x_1 - x_1^{k+1})^\top A_1^\top (A_2(x_2^k - x_2^{k+1}) + A_3(x_3^k - x_3^{k+1})) \quad \text{and} \quad (x_2 - x_2^{k+1})^\top A_2^\top A_3 (x_3^k - x_3^{k+1}).$$

For the first term, we have

$$
\begin{aligned}
& (x_1 - x_1^{k+1})^\top A_1^\top \left[A_2(x_2^k - x_2^{k+1}) + A_3(x_3^k - x_3^{k+1})\right] \\
= & \left[(A_1x_1 - b) - (A_1x_1^{k+1} - b)\right]^\top \left[(-A_2x_2^{k+1} - A_3x_3^{k+1}) - (-A_2x_2^k - A_3x_3^k)\right] \\
= & \frac{1}{2}\left(\|A_1x_1 + A_2x_2^k + A_3x_3^k - b\|^2 - \|A_1x_1 + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2\right) \\
& + \frac{1}{2}\left(\|A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2 - \|A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b\|^2\right) \\
= & \frac{1}{2}\left(\|A_1x_1 + A_2x_2^k + A_3x_3^k - b\|^2 - \|A_1x_1 + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2\right) + \frac{1}{2\gamma^2}\|\lambda^{k+1} - \lambda^k\|^2 \\
& - \frac{1}{2}\|A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b\|^2,
\end{aligned}
$$

where in the second equality we used the identity (2.17), and the last equality follows from the updating formula for λ^{k+1} in (2.5).

For the second term, we have

$$
\begin{aligned}
& (x_2 - x_2^{k+1})^\top A_2^\top A_3(x_3^k - x_3^{k+1}) \\
= & ((A_1x_1 + A_2x_2 - b) - (A_1x_1 + A_2x_2^{k+1} - b))^\top((-A_3x_3^{k+1}) - (-A_3x_3^k)) \\
= & \frac{1}{2}\left(\|A_1x_1 + A_2x_2 + A_3x_3^k - b\|^2 - \|A_1x_1 + A_2x_2 + A_3x_3^{k+1} - b\|^2\right) \\
& + \frac{1}{2}\left(\|A_1x_1 + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2 - \|A_1x_1 + A_2x_2^{k+1} + A_3x_3^k - b\|^2\right) \\
\leq & \frac{1}{2}\left(\|A_1x_1 + A_2x_2 + A_3x_3^k - b\|^2 - \|A_1x_1 + A_2x_2 + A_3x_3^{k+1} - b\|^2\right) \\
& + \frac{1}{2}\|A_1x_1 + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2,
\end{aligned}
$$

where in the second equality we applied the identity (2.17).

Therefore, we have

$$
\begin{aligned}
& (x_1 - x_1^{k+1})^\top \gamma A_1^\top (A_2(x_2^k - x_2^{k+1}) + A_3(x_3^k - x_3^{k+1})) + (x_2 - x_2^{k+1})^\top \gamma A_2^\top A_3(x_3^k - x_3^{k+1}) \\
\leq & \frac{\gamma}{2}\left(\|A_1x_1 + A_2x_2^k + A_3x_3^k - b\|^2 - \|A_1x_1 + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2\right) \\
& + \frac{\gamma}{2}\left(\|A_1x_1 + A_2x_2 + A_3x_3^k - b\|^2 - \|A_1x_1 + A_2x_2 + A_3x_3^{k+1} - b\|^2\right) \\
& + \frac{1}{2\gamma}\|\lambda^{k+1} - \lambda^k\|^2 + \frac{\gamma}{2}\|A_1x_1 + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2 \\
& - \frac{\gamma}{2}\|A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b\|^2. \qquad (3.3)
\end{aligned}
$$

Combining (3.3), (3.2) and (2.5), it holds for any $\lambda \in \mathbf{R}^p$ that

$$
\begin{aligned}
&\begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ x_3 - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} \\ g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} \\ g_3(x_3^{k+1}) - A_3^\top \lambda^{k+1} \\ A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \end{pmatrix} + \frac{1}{\gamma}(\lambda - \lambda^{k+1})^\top (\lambda^{k+1} - \lambda^k) \\
&+ \frac{1}{2\gamma}\|\lambda^{k+1} - \lambda^k\|^2 + \frac{\gamma}{2}\|A_1 x_1 + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2 \\
&+ \frac{\gamma}{2}\left(\|A_1 x_1 + A_2 x_2^k + A_3 x_3^k - b\|^2 - \|A_1 x_1 + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2\right) \\
&+ \frac{\gamma}{2}\left(\|A_1 x_1 + A_2 x_2 + A_3 x_3^k - b\|^2 - \|A_1 x_1 + A_2 x_2 + A_3 x_3^{k+1} - b\|^2\right) \\
\geq\ & \frac{\gamma}{2}\|A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b\|^2. \qquad (3.4)
\end{aligned}
$$

Using the identity

$$
\frac{1}{\gamma}\left(\lambda - \lambda^{k+1}\right)^\top (\lambda^{k+1} - \lambda^k) + \frac{1}{2\gamma}\|\lambda^{k+1} - \lambda^k\|^2 = \frac{1}{2\gamma}\left(\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2\right),
$$

letting $u = u^*$ in (3.4), and applying the facts that (invoking (2.13) and (2.14))

$$
\begin{aligned}
f_2(x_2^*) - f_2(x_2^{k+1}) - \frac{\sigma_2}{2}\|x_2^* - x_2^{k+1}\|^2 &\geq (x_2^* - x_2^{k+1})^\top g_2(x_2^{k+1}), \\
f_3(x_3^*) - f_3(x_3^{k+1}) - \frac{\sigma_3}{2}\|x_3^* - x_3^{k+1}\|^2 &\geq (x_3^* - x_3^{k+1})^\top g_3(x_3^{k+1}),
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{\gamma}{2}\|A_1 x_1^* + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2 \\
=\ & \frac{\gamma}{2}\|A_2(x_2^{k+1} - x_2^*) + A_3(x_3^{k+1} - x_3^*)\|^2 \\
\leq\ & \gamma(\lambda_{\max}(A_2^\top A_2)\|x_2^{k+1} - x_2^*\|^2 + \lambda_{\max}(A_3^\top A_3)\|x_3^{k+1} - x_3^*\|^2),
\end{aligned}
$$

we obtain,

$$
\begin{aligned}
& f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ x_3^* - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \lambda^{k+1} \\ -A_2^\top \lambda^{k+1} \\ -A_3^\top \lambda^{k+1} \\ A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \end{pmatrix} \\
&+ \frac{1}{2\gamma}\left(\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2\right) + \frac{\gamma}{2}\left(\|A_1 x_1^* + A_2 x_2^* + A_3 x_3^k - b\|^2 - \|A_1 x_1^* + A_2 x_2^* + A_3 x_3^{k+1} - b\|^2\right) \\
&+ \left(\gamma\lambda_{\max}(A_2^\top A_2) - \frac{\sigma_2}{2}\right)\|x_2^{k+1} - x_2^*\|^2 + \left(\gamma\lambda_{\max}(A_3^\top A_3) - \frac{\sigma_3}{2}\right)\|x_3^{k+1} - x_3^*\|^2 \\
&+ \frac{\gamma}{2}\left(\|A_1 x_1^* + A_2 x_2^k + A_3 x_3^k - b\|^2 - \|A_1 x_1^* + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2\right) \\
\geq\ & \frac{\gamma}{2}\|A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b\|^2.
\end{aligned}
$$

This together with the facts that $\gamma\lambda_{\max}(A_2^\top A_2) - \frac{\sigma_2}{2} \le 0$ and $\gamma\lambda_{\max}(A_3^\top A_3) - \frac{\sigma_3}{2} \le 0$ implies the desired inequality (3.1). $\square$

Now, we are ready to present the $O(1/t)$ ergodic convergence rate of the ADMM.

Theorem 3.2 *Assume that* $\gamma \le \min\left\{\frac{\sigma_2}{2\lambda_{\max}(A_2^\top A_2)}, \frac{\sigma_3}{2\lambda_{\max}(A_3^\top A_3)}\right\}$. *Let* $(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by ADMM* (2.2)-(2.5) *from given* $(x_2^k, x_3^k, \lambda^k)$. *For any integer* $t > 0$, *let* $\bar{u}^t = (\bar{x}_1^t, \bar{x}_2^t, \bar{x}_3^t)$ *and* $\bar{\lambda}^t$ *be defined as*

$$\bar{x}_1^t = \frac{1}{t+1}\sum_{k=0}^t x_1^{k+1}, \quad \bar{x}_2^t = \frac{1}{t+1}\sum_{k=0}^t x_2^{k+1}, \quad \bar{x}_3^t = \frac{1}{t+1}\sum_{k=0}^t x_3^{k+1}, \quad \bar{\lambda}^t = \frac{1}{t+1}\sum_{k=0}^t \lambda^{k+1}.$$

Then, for any $(u^*, \lambda^*) \in \Omega^*$, *by defining* $\rho := \|\lambda^*\| + 1$, *we have*

$$\begin{aligned} 0 &\le f(\bar{u}^t) - f(u^*) + \rho\|A_1\bar{x}_1^t + A_2\bar{x}_1^t + A_3\bar{x}_3^t - b\| \\ &\le \frac{\gamma}{2(t+1)}\|A_3x_3^* - A_3x_3^0\|^2 + \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)}\|A_1x_1^* + A_2x_2^0 + A_3x_3^0 - b\|^2. \end{aligned}$$

Note that this also implies that both the error of the objective function value and the residual of the equality constraint converge to 0 *with convergence rate* $O(1/t)$, *i.e.,*

$$|f(\bar{u}^t) - f(u^*)| = O(1/t), \quad \text{and} \quad \|A_1\bar{x}_1^t + A_2\bar{x}_1^t + A_3\bar{x}_3^t - b\| = O(1/t). \tag{3.5}$$

Proof. Because $(u^k, \lambda^k) \in \Omega$, it holds that $(\bar{u}^t, \bar{\lambda}^t) \in \Omega$ for all $t \ge 0$. By Lemma 3.1 and invoking the convexity of function $f(\cdot)$, we have

$$\begin{aligned} & f(u^*) - f(\bar{u}^t) + \lambda^\top(A_1\bar{x}_1^t + A_2\bar{x}_2^t + A_3\bar{x}_3^t - b) \\ = \;& f(u^*) - f(\bar{u}^t) + \begin{pmatrix} x_1^* - \bar{x}_1^t \\ x_2^* - \bar{x}_2^t \\ x_3^* - \bar{x}_3^t \\ \lambda - \bar{\lambda}^t \end{pmatrix}^\top \begin{pmatrix} -A_1^\top\bar{\lambda}^t \\ -A_2^\top\bar{\lambda}^t \\ -A_3^\top\bar{\lambda}^t \\ A_1\bar{x}_1^t + A_2\bar{x}_2^t + A_3\bar{x}_3^t - b \end{pmatrix} \\ \ge \;& \frac{1}{t+1}\sum_{k=0}^t \left[f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ x_3^* - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top\lambda^{k+1} \\ -A_2^\top\lambda^{k+1} \\ -A_3^\top\lambda^{k+1} \\ A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b \end{pmatrix} \right] \\ \ge \;& \frac{1}{t+1}\sum_{k=0}^t \Bigg[\frac{1}{2\gamma}(\|\lambda - \lambda^{k+1}\|^2 - \|\lambda - \lambda^k\|^2) \\ & + \frac{\gamma}{2}\left(\|A_1x_1^* + A_2x_2^* + A_3x_3^{k+1} - b\|^2 - \|A_1x_1^* + A_2x_2^* + A_3x_3^k - b\|^2\right) \\ & + \frac{\gamma}{2}\left(\|A_1x_1^* + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2 - \|A_1x_1^* + A_2x_2^k + A_3x_3^k - b\|^2\right) \Bigg] \\ \ge \;& -\frac{1}{2\gamma(t+1)}\|\lambda - \lambda^0\|^2 - \frac{\gamma}{2(t+1)}\|A_1x_1^* + A_2x_2^* + A_3x_3^0 - b\|^2 - \frac{\gamma}{2(t+1)}\|A_1x_1^* + A_2x_2^0 + A_3x_3^0 - b\|^2. \end{aligned}$$

Note that this inequality holds for all $\lambda \in \mathbf{R}^p$. From weak duality of (2.1) we obtain

$$0 \geq f(u^*) - f(\bar{u}^t) + (\lambda^*)^\top (A_1\bar{x}_1^t + A_2\bar{x}_1^t + A_3\bar{x}_3^t - b).$$

Moreover, since $\rho := \|\lambda^*\| + 1$, $\|\lambda - \lambda_0\|^2 \leq 2(\rho^2 + \|\lambda^0\|^2)$ for all $\|\lambda\| \leq \rho$, and $A_1x_1^* + A_2x_2^* + A_3x_3^* = b$, we obtain

$$\begin{aligned} 0 &\leq f(\bar{u}^t) - f(u^*) + \rho\|A_1\bar{x}_1^t + A_2\bar{x}_1^t + A_3\bar{x}_3^t - b\| \\ &\leq \frac{\gamma}{2(t+1)}\|A_3(x_3^* - x_3^0)\|^2 + \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)}\|A_2(x_2^* - x_2^0) + A_3(x_3^* - x_3^0)\|^2. \end{aligned} \tag{3.6}$$

We now define the function

$$v(\xi) = \min\{f(u) | A_1x_1 + A_2x_2 + A_3x_3 - b = \xi, x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, x_3 \in \mathcal{X}_3\}.$$

It is easy to verify that v is convex, $v(0) = f(u^*)$, and $\lambda^* \in \partial v(0)$. Therefore, from the convexity of v, it holds that

$$v(\xi) \geq v(0) + \langle \lambda^*, \xi \rangle \geq f(u^*) - \|\lambda^*\|\|\xi\|. \tag{3.7}$$

Let $\bar{\xi} = A_1\bar{x}_1 + A_2\bar{x}_2 + A_3\bar{x}_3 - b$, we have $f(\bar{u}^t) \geq v(\bar{\xi})$. Therefore, by denoting the constant

$$C := \frac{\gamma}{2}\|A_3x_3^* - A_3x_3^0\|^2 + \frac{\|\lambda^0\|^2}{\gamma} + \frac{\gamma}{2}\|A_1x_1^* + A_2x_2^0 + A_3x_3^0 - b\|^2,$$

and combining (3.6) and (3.7), we can get

$$\frac{C + \rho^2/\gamma}{t+1} - \rho\|\bar{\xi}\| \geq f(\bar{u}^t) - f(u^*) \geq -\|\lambda^*\|\|\bar{\xi}\|,$$

which, by using $\rho = \|\lambda^*\| + 1$, yields,

$$\|A_1\bar{x}_1 + A_2\bar{x}_2 + A_3\bar{x}_3 - b\| = \|\bar{\xi}\| \leq \frac{C + \rho^2/\gamma}{t+1}. \tag{3.8}$$

Moreover, by combining (3.6) and (3.8), one obtains that

$$-\frac{\rho C + \rho^3/\gamma}{t+1} \leq f(\bar{u}^t) - f(u^*) \leq \frac{C + \rho^2/\gamma}{t+1}. \tag{3.9}$$

(3.8) and (3.9) imply (3.5) immediately. □

Therefore, we have established the $O(1/t)$ convergence rate of the ADMM (2.2)-(2.5) in an ergodic sense. Our proof is readily extended to the case of N-block ADMM (1.2). The following theorem shows the $O(1/t)$ convergence rate of N-block ADMM (1.2). We omit the proof here for the sake of succinctness.

Theorem 3.3 *Assume that*

$$\gamma \leq \min_{i=2,\cdots,N-1} \left\{ \frac{2\sigma_i}{(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)}, \frac{2\sigma_N}{(N-2)(N+1)\lambda_{\max}(A_N^\top A_N)} \right\},$$

where σ_i is the strong convexity parameter of f_i, $i = 2, \ldots, N$. Let $(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \cdots, x_N^{k+1}, \lambda^{k+1}) \in \Omega$ be generated by the N-block ADMM (1.2). For any integer $t > 0$, we define

$$\bar{x}_i^t = \frac{1}{t+1} \sum_{k=0}^{t} x_i^{k+1}, 1 \le i \le N, \qquad \bar{\lambda}^t = \frac{1}{t+1} \sum_{k=0}^{t} \lambda^{k+1}.$$

Then, for $\rho := \|\lambda^*\| + 1$, it holds that

$$\sum_{i=1}^{N} (f_i(\bar{x}_i^t) - f_i(x_i^*)) + \rho \left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| \le \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{m=i+1}^{N} A_m (x_m^0 - x_m^*) \right\|^2 + \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)}.$$

Similarly as Theorem 3.2, this also implies that N-block ADMM (1.2) converges with rate $O(1/t)$ in terms both error of objective function value and the residual of the equality constraints, i.e., it holds that

$$|f(\bar{u}^t) - f(u^*)| = O(1/t), \quad \text{and} \quad \left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| = O(1/t).$$

4 Non-Ergodic Convergence Rate of ADMM

In this section, we prove an $o(1/k)$ non-ergodic convergence rate for ADMM (2.2)-(2.5).

Let us first observe the following (see also Lemma 4.1 in [13]). Suppose at the $(k+1)$-th iteration of ADMM (2.2)-(2.5), we have

$$\begin{cases} A_2 x_2^{k+1} - A_2 x_2^k = 0, \\ A_3 x_3^{k+1} - A_3 x_3^k = 0, \\ A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b = 0. \end{cases} \tag{4.1}$$

Then, (2.9)-(2.11) would immediately lead to

$$\begin{cases} (x_1 - x_1^{k+1})^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} \right] \ge 0, & \forall x_1 \in \mathcal{X}_1, \\ (x_2 - x_2^{k+1})^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} \right] \ge 0, & \forall x_2 \in \mathcal{X}_2, \\ (x_3 - x_3^{k+1})^\top \left[g_3(x_3^{k+1}) - A_3^\top \lambda^{k+1} \right] \ge 0, & \forall x_3 \in \mathcal{X}_3. \end{cases}$$

In other words, if (4.1) is satisfied, then $(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ would have been already an optimal solution for (2.1). It is therefore natural to introduce a residual for the linear system (4.1) as an optimality measure. Below is such a measure, to be denoted by R_{k+1}:

$$R_{k+1} := \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2 + 2\|A_2 x_2^{k+1} - A_2 x_2^k\|^2 + 3\|A_3 x_3^{k+1} - A_3 x_3^k\|^2. \tag{4.2}$$

In the sequel, we will show that R_k converges to 0 at the rate $o(1/k)$. Note that this gives the convergence rate of ADMM (2.2)-(2.5) in non-ergodic sense.

We first show that R_k is non-increasing.

Lemma 4.1 *Assume* $\gamma \le \min\{\frac{\sigma_2}{\lambda_{\max}(A_2^\top A_2)}, \frac{\sigma_3}{\lambda_{\max}(A_3^\top A_3)}\}$. *Let the sequence* $\{x_1^k, x_2^k, x_3^k, \lambda^k\}$ *be generated by ADMM* (2.2)-(2.5). *It holds that* R_k *defined in* (4.2) *is non-increasing, i.e.,*

$$R_{k+1} \le R_k, \;\; k = 0, 1, 2, ... \tag{4.3}$$

Proof. Letting $x_1 = x_1^k$ in (2.6) yields,

$$(x_1^k - x_1^{k+1})^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^k + \gamma A_1^\top (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b)\right] \ge 0,$$

with $g_1 \in \partial f_1$, which further implies that

$$\begin{aligned}
& (x_1^{k+1} - x_1^k)^\top g_1(x_1^{k+1}) \\
\le\; & (x_1^k - x_1^{k+1})^\top (-A_1^\top \lambda^k) + (x_1^k - x_1^{k+1})^\top \left[\gamma A_1^\top (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b)\right] \\
=\; & (A_1 x_1^k - A_1 x_1^{k+1})^\top (-\lambda^k) + \gamma (A_1 x_1^k - A_1 x_1^{k+1})^\top (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b) \\
=\; & (A_1 x_1^k - A_1 x_1^{k+1})^\top (-\lambda^k) + \frac{\gamma}{2} \Big(\|A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b\|^2 \\
& - \|A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b\|^2 - \|A_1 x_1^k - A_1 x_1^{k+1}\|^2 \Big),
\end{aligned} \tag{4.4}$$

where the last equality is due to the identity (2.18). Letting $x_1 = x_1^{k+1}$ in (2.9) with $k+1$ changed to k yields,

$$(x_1^{k+1} - x_1^k)^\top \left[g_1(x_1^k) - A_1^\top \lambda^k + \gamma A_1^\top A_2 (x_2^{k-1} - x_2^k) + \gamma A_1^\top A_3 (x_3^{k-1} - x_3^k)\right] \ge 0,$$

which further implies that

$$\begin{aligned}
& (x_1^k - x_1^{k+1})^\top g_1(x_1^k) \\
\le\; & (x_1^{k+1} - x_1^k)^\top (-A_1^\top \lambda^k) + \gamma (x_1^{k+1} - x_1^k)^\top \left[A_1^\top A_2 (x_2^{k-1} - x_2^k) + A_1^\top A_3 (x_3^{k-1} - x_3^k)\right] \\
=\; & (A_1 x_1^{k+1} - A_1 x_1^k)^\top (-\lambda^k) + \gamma (A_1 x_1^{k+1} - A_1 x_1^k)^\top \left[A_2 (x_2^{k-1} - x_2^k) + A_3 (x_3^{k-1} - x_3^k)\right] \\
\le\; & (A_1 x_1^{k+1} - A_1 x_1^k)^\top (-\lambda^k) + \frac{\gamma}{2} \left(\|A_1 x_1^{k+1} - A_1 x_1^k\|^2 + \|A_2 (x_2^{k-1} - x_2^k) + A_3 (x_3^{k-1} - x_3^k)\|^2 \right) \\
\le\; & (A_1 x_1^{k+1} - A_1 x_1^k)^\top (-\lambda^k) \\
& + \frac{\gamma}{2} \left(\|A_1 x_1^{k+1} - A_1 x_1^k\|^2 + 2\|A_2 (x_2^{k-1} - x_2^k)\|^2 + 2\|A_3 (x_3^{k-1} - x_3^k)\|^2 \right).
\end{aligned} \tag{4.5}$$

Combining (4.4) and (4.5) gives

$$(x_1^{k+1} - x_1^k)^\top \left[g_1(x_1^{k+1}) - g_1(x_1^k)\right] \tag{4.6}$$

$$\le \frac{\gamma}{2} \Big(\|A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b\|^2 - \|A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b\|^2 \tag{4.7}$$

$$+ 2\|A_2 (x_2^{k-1} - x_2^k)\|^2 + 2\|A_3 (x_3^{k-1} - x_3^k)\|^2 \Big). \tag{4.8}$$

Letting $x_2 = x_2^k$ in (2.7) yields,

$$(x_2^k - x_2^{k+1})^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^k + \gamma A_2^\top (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b)\right] \ge 0,$$

which further implies that

$$
\begin{aligned}
& (x_2^{k+1} - x_2^k)^\top g_2(x_2^{k+1}) \\
\leq\ & (x_2^k - x_2^{k+1})^\top (-A_2^\top \lambda^k) + (x_2^k - x_2^{k+1})^\top \left[\gamma A_2^\top (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b)\right] \\
=\ & (A_2 x_2^k - A_2 x_2^{k+1})^\top (-\lambda^k) + \gamma (A_2 x_2^k - A_2 x_2^{k+1})^\top (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b) \\
=\ & (A_2 x_2^k - A_2 x_2^{k+1})^\top (-\lambda^k) + \frac{\gamma}{2} \Big(\|A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b\|^2 \\
& - \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b\|^2 - \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 \Big), \qquad (4.9)
\end{aligned}
$$

where the last equality is due to the identity (2.18). Letting $x_2 = x_2^{k+1}$ in (2.10) with $k+1$ changed to k yields,

$$
(x_2^{k+1} - x_2^k)^\top \left[g_2(x_2^k) - A_2^\top \lambda^k + \gamma A_2^\top A_3 (x_3^{k-1} - x_3^k) \right] \geq 0,
$$

which further implies that

$$
\begin{aligned}
& (x_2^k - x_2^{k+1})^\top g_2(x_2^k) \\
\leq\ & (x_2^{k+1} - x_2^k)^\top (-A_2^\top \lambda^k) + \gamma (x_2^{k+1} - x_2^k)^\top \left[A_2^\top A_3 (x_3^{k-1} - x_3^k) \right] \\
=\ & (A_2 x_2^{k+1} - A_2 x_2^k)^\top (-\lambda^k) + \gamma (A_2 x_2^{k+1} - A_2 x_2^k)^\top (A_3 x_3^{k-1} - A_3 x_3^k) \\
\leq\ & (A_2 x_2^{k+1} - A_2 x_2^k)^\top (-\lambda^k) + \frac{\gamma}{2} \Big(\|A_2 x_2^{k+1} - A_2 x_2^k\|^2 + \|A_3 x_3^{k-1} - A_3 x_3^k\|^2 \Big). \qquad (4.10)
\end{aligned}
$$

Combining (4.9) and (4.10) gives

$$
\begin{aligned}
& (x_2^{k+1} - x_2^k)^\top \left[g_2(x_2^{k+1}) - g_2(x_2^k) \right] \\
\leq\ & \frac{\gamma}{2} \Big(\|A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b\|^2 - \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b\|^2 \\
& + \|A_3 x_3^{k-1} - A_3 x_3^k\|^2 \Big). \qquad (4.11)
\end{aligned}
$$

Letting $x_3 = x_3^k$ in (2.11) and $x_3 = x_3^{k+1}$ in (2.11) with $k+1$ changed to k, and adding the two resulting inequalities, yields,

$$
\begin{aligned}
& (x_3^{k+1} - x_3^k)^\top \left[g_3(x_3^{k+1}) - g_3(x_3^k) \right] \\
\leq\ & (x_3^{k+1} - x_3^k)^\top (A_3^\top \lambda^{k+1} - A_3^\top \lambda^k) \\
=\ & (A_3 x_3^{k+1} - A_3 x_3^k)^\top (\lambda^{k+1} - \lambda^k) \\
=\ & \gamma (A_3 x_3^k - A_3 x_3^{k+1})^\top \Big(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \Big) \\
=\ & \frac{\gamma}{2} \Big(\|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b\|^2 - \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2 \\
& - \|A_3 x_3^k - A_3 x_3^{k+1}\|^2 \Big), \qquad (4.12)
\end{aligned}
$$

where the last equality is due to the identity (2.18).

Combining (4.7), (4.11) and (4.12) yields,

$$
\begin{aligned}
& (x_1^{k+1}-x_1^k)^\top \left[g_1(x_1^{k+1})-g_1(x_1^k)\right] + (x_2^{k+1}-x_2^k)^\top \left[g_2(x_2^{k+1})-g_2(x_2^k)\right] \\
& +(x_3^{k+1}-x_3^k)^\top \left[g_3(x_3^{k+1})-g_3(x_3^k)\right] \\
\leq\ & \frac{\gamma}{2}\left[\|A_1x_1^k+A_2x_2^k+A_3x_3^k-b\|^2 - \|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2 + 2\|A_2(x_2^{k-1}-x_2^k)\|^2\right. \\
& +2\|A_3(x_3^{k-1}-x_3^k)\|^2 + \|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2 - \|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^k-b\|^2 \\
& +\|A_3x_3^{k-1}-A_3x_3^k\|^2 + \|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^k-b\|^2 \\
& \left.-\|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b\|^2 - \|A_3x_3^k-A_3x_3^{k+1}\|^2\right] \\
=\ & \frac{\gamma}{2}\left[\|A_1x_1^k+A_2x_2^k+A_3x_3^k-b\|^2 + 2\|A_2(x_2^{k-1}-x_2^k)\|^2 + 3\|A_3(x_3^{k-1}-x_3^k)\|^2\right. \\
& \left.-\|A_3x_3^k-A_3x_3^{k+1}\|^2 - \|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b\|^2\right] \\
=\ & \frac{\gamma}{2}\left[R_k - R_{k+1} + 2\|A_2x_2^{k+1}-A_2x_2^k\|^2 + 2\|A_3x_3^k-A_3x_3^{k+1}\|^2\right]. \qquad (4.13)
\end{aligned}
$$

Note that (2.15) and (2.16) imply that

$$
\begin{aligned}
(x_2^{k+1}-x_2^k)^\top \left[g_2(x_2^{k+1})-g_2(x_2^k)\right] &\geq \sigma_2\|x_2^{k+1}-x_2^k\|^2, \\
(x_3^{k+1}-x_3^k)^\top \left[g_3(x_3^{k+1})-g_3(x_3^k)\right] &\geq \sigma_3\|x_3^{k+1}-x_3^k\|^2.
\end{aligned} \qquad (4.14)
$$

Combining (4.13) and (4.14), and the fact that $\gamma \leq \min\left\{\frac{\sigma_2}{\lambda_{\max}(A_2^\top A_2)}, \frac{\sigma_3}{\lambda_{\max}(A_3^\top A_3)}\right\}$, it is easy to see that $R_{k+1} \leq R_k$ for $k = 0, 1, 2, ...$ □

We are now ready to present the $o(1/k)$ non-ergodic convergence rate of the ADMM (2.2)-(2.5).

Theorem 4.2 *Assume* $\gamma \leq \min\left\{\frac{\sigma_2}{2\lambda_{\max}(A_2^\top A_2)}, \frac{\sigma_3}{2\lambda_{\max}(A_3^\top A_3)}\right\}$. *Let the sequence* $\{x_1^k, x_2^k, x_3^k, \lambda^k\}$ *be generated by ADMM* (2.2)-(2.5). *Then* $\sum_{k=1}^\infty R_k < +\infty$ *and* $R_k = o(1/k)$.

Proof. Combining (4.11) and (4.12) yields

$$\begin{aligned}
& (x_2^{k+1}-x_2^k)^\top \left[g_2(x_2^{k+1})-g_2(x_2^k)\right] + (x_3^{k+1}-x_3^k)^\top \left[g_3(x_3^{k+1})-g_3(x_3^k)\right] \\
\leq\ & \frac{\gamma}{2}\Big[\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2 - \|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^k-b\|^2 \\
& +\|A_3x_3^{k-1}-A_3x_3^k\|^2 + \|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^k-b\|^2 \\
& -\|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b\|^2 - \|A_3x_3^k-A_3x_3^{k+1}\|^2\Big] \\
=\ & \frac{\gamma}{2}\Big[\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2 + \|A_3x_3^{k-1}-A_3x_3^k\|^2 - \|A_3x_3^k-A_3x_3^{k+1}\|^2 \\
& -\|A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b\|^2\Big] \\
=\ & \frac{\gamma}{2}\Big[\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2 + \|A_3x_3^{k-1}-A_3x_3^k\|^2 - \|A_3x_3^k-A_3x_3^{k+1}\|^2 - R_{k+1} \\
& +2\|A_2x_2^{k+1}-A_2x_2^k\|^2 + 3\|A_3x_3^k-A_3x_3^{k+1}\|^2\Big] \\
\leq\ & \frac{\gamma}{2}\Big[\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2 + \|A_3x_3^{k-1}-A_3x_3^k\|^2 - \|A_3x_3^k-A_3x_3^{k+1}\|^2 - R_{k+1}\Big] \\
& +\gamma\lambda_{\max}(A_2^\top A_2)\|x_2^{k+1}-x_2^k\|^2 + \frac{3\gamma}{2}\lambda_{\max}(A_3^\top A_3)\|x_3^{k+1}-x_3^k\|^2.
\end{aligned}$$

Using (4.14) and the assumption that $\gamma \leq \min\left\{\frac{\sigma_2}{2\lambda_{\max}(A_2^\top A_2)}, \frac{\sigma_3}{2\lambda_{\max}(A_3^\top A_3)}\right\}$, we obtain

$$R_{k+1} \leq \|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2 + \|A_3x_3^{k-1}-A_3x_3^k\|^2 - \|A_3x_3^k-A_3x_3^{k+1}\|^2. \tag{4.15}$$

From the optimality conditions (2.12) and the convexity of f, it follows that

$$f(u^*)-f(u^{k+1}) \leq (x_1^*-x_1^{k+1})^\top(A_1^\top\lambda^*) + (x_2^*-x_2^{k+1})^\top(A_2^\top\lambda^*) + (x_3^*-x_3^{k+1})^\top(A_3^\top\lambda^*). \tag{4.16}$$

By combining (3.1) and (4.16), we have

$$\begin{aligned}
& \begin{pmatrix} x_1^*-x_1^{k+1} \\ x_2^*-x_2^{k+1} \\ x_3^*-x_3^{k+1} \end{pmatrix}^\top \begin{pmatrix} A_1^\top(\lambda^*-\lambda^{k+1}) \\ A_2^\top(\lambda^*-\lambda^{k+1}) \\ A_3^\top(\lambda^*-\lambda^{k+1}) \end{pmatrix} \\
& +\frac{1}{2\gamma}\|\lambda^k-\lambda^{k+1}\|^2 + \frac{\gamma}{2}\left(\|A_1x_1^*+A_2x_2^*+A_3x_3^k-b\|^2 - \|A_1x_1^*+A_2x_2^*+A_3x_3^{k+1}-b\|^2\right) \\
& +\frac{\gamma}{2}\left(\|A_1x_1^*+A_2x_2^k+A_3x_3^k-b\|^2 - \|A_1x_1^*+A_2x_2^{k+1}+A_3x_3^{k+1}-b\|^2\right) \\
\geq\ & \frac{\gamma}{2}\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2.
\end{aligned} \tag{4.17}$$

Note that the first term in (4.17) is equal to

$$\begin{aligned}
& -(A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b)^\top(\lambda^*-\lambda^{k+1}) \\
=\ & \frac{1}{\gamma}\left(\lambda^{k+1}-\lambda^k\right)^\top\left(\lambda^*-\lambda^{k+1}\right) \\
=\ & \frac{1}{2\gamma}\left(\|\lambda^*-\lambda^k\|^2 - \|\lambda^{k+1}-\lambda^k\|^2 - \|\lambda^*-\lambda^{k+1}\|^2\right).
\end{aligned}$$

Therefore, (4.17) can be rearranged as

$$\begin{aligned}
&\frac{1}{\gamma^2}\left(\|\lambda^*-\lambda^k\|^2-\|\lambda^*-\lambda^{k+1}\|^2\right)+(\|A_1x_1^*+A_2x_2^*+A_3x_3^k-b\|^2-\|A_1x_1^*+A_2x_2^*+A_3x_3^{k+1}-b\|^2)\\
&+\left(\|A_1x_1^*+A_2x_2^k+A_3x_3^k-b\|^2-\|A_1x_1^*+A_2x_2^{k+1}+A_3x_3^{k+1}-b\|^2\right)\\
\geq\ &\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2. \qquad (4.18)
\end{aligned}$$

By (4.15) and (4.18) we get that

$$\begin{aligned}
&\sum_{k=1}^{\infty}R_{k+1}\\
\leq\ &\sum_{k=1}^{\infty}\left[\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2+\|A_3x_3^{k-1}-A_3x_3^k\|^2-\|A_3x_3^k-A_3x_3^{k+1}\|^2\right]\\
\leq\ &\sum_{k=1}^{\infty}\|A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b\|^2+\|A_3x_3^0-A_3x_3^1\|^2\\
\leq\ &\|A_3x_3^0-A_3x_3^1\|^2+\sum_{k=1}^{\infty}\left[\left(\|A_1x_1^*+A_2x_2^*+A_3x_3^k-b\|^2-\|A_1x_1^*+A_2x_2^*+A_3x_3^{k+1}-b\|^2\right)\right.\\
&+\left(\|A_1x_1^*+A_2x_2^k+A_3x_3^k-b\|^2-\|A_1x_1^*+A_2x_2^{k+1}+A_3x_3^{k+1}-b\|^2\right)\\
&\left.+\frac{1}{\gamma^2}\left(\|\lambda^*-\lambda^k\|^2-\|\lambda^*-\lambda^{k+1}\|^2\right)\right]\\
\leq\ &\|A_3x_3^0-A_3x_3^1\|^2+\|A_1x_1^*+A_2x_2^*+A_3x_3^1-b\|^2+\|A_1x_1^*+A_2x_2^1+A_3x_3^1-b\|^2+\frac{1}{\gamma^2}\|\lambda^*-\lambda^1\|^2.
\end{aligned}$$

Note that we have proved that R_k is monotonically non-increasing, and $\sum_{k=1}^{\infty}R_k<+\infty$. As observed in Lemma 1.2 of [4], one has

$$kR_{2k}\leq R_k+R_{k+1}+\cdots+R_{2k}\to 0,\ \text{as}\ k\to\infty,$$

and therefore $R_k=o(1/k)$. □

Note that our analysis can be extended to N-block ADMM (1.2) easily. The results are summarized in the following theorem and the proof is omitted for the sake of succinctness.

Theorem 4.3 *Assume that*

$$\gamma\leq\min_{i=2,\cdots,N-1}\left\{\frac{2\sigma_i}{(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)},\frac{2\sigma_N}{(N-2)(N+1)\lambda_{\max}(A_N^\top A_N)}\right\}.$$

Let $(x_1^{k+1},x_2^{k+1},x_3^{k+1},\cdots,x_N^{k+1},\lambda^{k+1})\in\Omega$ be generated by ADMM (1.2). Then $\sum_{k=1}^{\infty}R_k<+\infty$ and $R_k=o(1/k)$, where R_k is defined as

$$R_{k+1}:=\left\|\sum_{i=1}^{N}A_ix_i^{k+1}-b\right\|^2+\sum_{i=2}^{N}\frac{(2N-i)(i-1)}{2}\|A_ix_i^k-A_ix_i^{k+1}\|^2.$$

5 Conclusions

In this paper, we analyzed the sublinear convergence rate of the standard multi-block ADMM in both ergodic and non-ergodic sense. These are the first sublinear convergence rate results for standard multi-block ADMM. Using the techniques developed in this paper, we can also analyze the convergence rate of some variants of the standard multi-block ADMM such as the ones studied in [14] and [4], where the primal variables are updated in a Jacobian manner; we plan to pursue this direction of research in the future.

Acknowledgements

Research of Shiqian Ma was supported in part by the Hong Kong Research Grants Council (RGC) Early Career Scheme (ECS) (Project ID: CUHK 439513). Research of Shuzhong Zhang was supported in part by the National Science Foundation under Grant Number CMMI-1161242.

References

[1] D. Boley. Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. *SIAM Journal on Optimization*, 23(4):2183–2207, 2013.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[3] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Preprint*, 2013.

[4] W. Deng, M. Lai, Z. Peng, and W. Yin. Parallel multi-block admm with $o(1/k)$ convergence. *Preprint*, 2014.

[5] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. Technical report, Rice University CAAM, 2012.

[6] J. Douglas and H. H. Rachford. On the numerical solution of the heat conduction problem in 2 and 3 space variables. *Transactions of the American Mathematical Society*, 82:421–439, 1956.

[7] J. Eckstein. *Splitting methods for monotone operators with applications to parallel optimization*. PhD thesis, Massachusetts Institute of Technology, 1989.

[8] J. Eckstein. Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results. *Preprint*, 2012.

[9] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[10] M. Fortin and R. Glowinski. *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*. North-Holland Pub. Co., 1983.

[11] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*. North-Hollan, Amsterdam, 1983.

[12] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM, Philadelphia, Pennsylvania, 1989.

[13] D. Han and X. Yuan. A note on the alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 155(1):227–238, 2012.

[14] B. He, L. Hou, and X. Yuan. On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. *Preprint*, 2013.

[15] B. He, M. Tao, and X. Yuan. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM Journal on Optimization*, 22:313–340, 2012.

[16] B. He, M. Tao, and X. Yuan. Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming. *Preprint*, 2013.

[17] B. He and X. Yuan. On nonergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Preprint*, 2012.

[18] B. He and X. Yuan. On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50:700–709, 2012.

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

[20] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16:964–979, 1979.

[21] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM Journal on Optimization*, 23:475–507, 2013.

[22] D. H. Peaceman and H. H. Rachford. The numerical solution of parabolic elliptic differential equations. *SIAM Journal on Applied Mathematics*, 3:28–41, 1955.