# On the Global Linear Convergence of the ADMM with Multi-Block Variables

Tianyi Lin* Shiqian Ma* Shuzhong Zhang†

May 31, 2014

**Abstract**

The alternating direction method of multipliers (ADMM) has been widely used for solving structured convex optimization problems. In particular, the ADMM can solve convex programs that minimize the sum of $N$ convex functions with $N$-block variables linked by some linear constraints. While the convergence of the ADMM for $N = 2$ was well established in the literature, it remained an open problem for a long time whether or not the ADMM for $N \geq 3$ is still convergent. Recently, it was shown in [3] that without further conditions the ADMM for $N \geq 3$ may actually fail to converge. In this paper, we show that under some easily verifiable and reasonable conditions the global linear convergence of the ADMM when $N \geq 3$ can still be assured, which is important since the ADMM is a popular method for solving large scale multi-block optimization models and is known to perform very well in practice even when $N \geq 3$. Our study aims to offer an explanation for this phenomenon.

Keywords: Alternating Direction Method of Multipliers, Global Linear Convergence, Convex Optimization

## 1 Introduction

In this paper, we consider the global linear convergence of the standard alternating direction method of multipliers (ADMM) for solving convex minimization problems with $N$-block variables when $N \geq 3$. The problem under consideration can be formulated as

$$\begin{array}{ll} \min & \tilde{f}_1(x_1) + \tilde{f}_2(x_2) + \cdots + \tilde{f}_N(x_N) \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + \cdots + A_N x_N = b, \\ & x_i \in \mathcal{X}_i, i = 1, \ldots, N, \end{array} \tag{1.1}$$

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 $\tilde{f}_i : \mathbf{R}^{n_i} \to \mathbf{R}^p$ are closed convex functions. Note that the convex constraint $x_i \in \mathcal{X}_i$ can be incorporated into the objective using an

*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.

indicator function, i.e., (1.1) can be rewritten as

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + \cdots + f_N(x_N) \\ \text{s.t.} & A_1x_1 + A_2x_2 + \cdots + A_Nx_N = b, \end{array} \tag{1.2}$$

where $f_i(x_i) := \tilde{f}_i(x_i) + \mathbf{1}_i(x_i)$ and

$$\mathbf{1}_i(x_i) := \begin{cases} 0 & \text{if } x_i \in \mathcal{X}_i \\ +\infty & \text{otherwise.} \end{cases}$$

We thus consider the equivalent reformulation (1.2) throughout this paper for the ease of presentation. For given $(x_2^k, \ldots, x_N^k; \lambda^k)$, a typical iteration of the ADMM for solving (1.2) can be summarized as:

$$\begin{cases} x_1^{k+1} & := & \operatorname{argmin}_{x_1} \ \mathcal{L}_\gamma(x_1, x_2^k, \ldots, x_N^k; \lambda^k) \\ x_2^{k+1} & := & \operatorname{argmin}_{x_2} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2, x_3^k, \ldots, x_N^k; \lambda^k) \\ & \vdots & \\ x_N^{k+1} & := & \operatorname{argmin}_{x_N} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, \ldots, x_{N-1}^{k+1}, x_N; \lambda^k) \\ \lambda^{k+1} & := & \lambda^k - \gamma \left( \sum_{i=1}^N A_i x_i^{k+1} - b \right), \end{cases} \tag{1.3}$$

where

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

denotes the augmented Lagrangian function of (1.2) with $\lambda$ being the Lagrange multiplier and $\gamma > 0$ being a penalty parameter. It is noted that in each iteration, the ADMM updates the primal variables $x_1, \ldots, x_N$ in a Gauss-Seidel manner.

When $N = 2$, the ADMM (1.3) was shown to be equivalent to the Douglas-Rachford operator splitting method that dated back to 1950s for solving variational problems arising from PDEs [5, 9]. The convergence of the ADMM (1.3) when $N = 2$ was thus established in the context of operator splitting methods [16, 8]. Recently, ADMM has been revisited due to its success in solving structured convex optimization problems arising from sparse and low-rank optimization and related problems (we refer the readers to some recent survey papers for more details, see, e.g., [2, 6]). In [16], Lions and Mercier showed that the Douglas-Rachford operator splitting method converges linearly under the assumption that some involved monotone operator is both coercive and Lipschitz. Eckstein and Bertsekas [7] showed the linear convergence of the ADMM (1.3) with $N = 2$ for solving linear programs, which depends on a bound on the largest iterate in the course of the algorithm. In a recent work by Deng and Yin [4], a generalized ADMM was proposed in which some proximal terms were added to the two subproblems in (1.3), and it was shown that this generalized ADMM converges linearly under certain assumptions on the strong convexity of functions $f_1$ and $f_2$, and the rank of $A_1$ and $A_2$. For instance, one sufficient condition suggested in [4] that guarantees the linear convergence of the generalized ADMM is that $f_1$ and $f_2$ are both strongly convex, $\nabla f_2$ is Lipschitz continuous and $A_2$ is of full row rank. Han and Yuan [11] and Boley [1] both studied the local linear convergence of ADMM (1.3) when $N = 2$ for solving quadratic programs. The result in [11] was based on some error bound condition [17], and the one given in [1] was obtained by first writing the ADMM as a matrix recurrence and then performing a spectral analysis on the recurrence. Moreover, it was shown that the ADMM (1.3) when $N = 2$ converges sublinearly under the simple convexity assumption both in ergodic and non-ergodic sense [13, 18, 12]. It should be noted that all the convergence results on the ADMM (1.3) discussed above are for the case $N = 2$.

While the convergence properties of the ADMM when $N = 2$ have been well studied, its convergence when $N \geq 3$ has remained unclear for a very long time. The following includes some recent progresses on this direction. In a recent work by Chen et al. [3], a counter-example was given which shows that without further conditions the ADMM for $N \geq 3$ may actually fail to converge. Existing works that study sufficient conditions ensuring the convergence of ADMM when $N \geq 3$ are briefly summarized as follows. Han and Yuan [10] proved the global convergence of ADMM (1.3) under the condition that $f_1, \ldots, f_N$ are all strongly convex and $\gamma$ is restricted to certain region. Hong and Luo [14] proposed to adopt a small step size when updating the Lagrange multiplier $\lambda^k$ in (1.3), i.e., they suggested that the update for $\lambda^k$, i.e.,

$$\lambda^{k+1} := \lambda^k - \gamma \left( \sum_{i=1}^{N} A_i x_i^{k+1} - b \right), \tag{1.4}$$

be changed to

$$\lambda^{k+1} := \lambda^k - \alpha\gamma \left( \sum_{i=1}^{N} A_i x_i^{k+1} - b \right), \tag{1.5}$$

where $\alpha > 0$ is a small step size. It was shown in [14] that this variant of ADMM converges linearly under the assumption that certain error bound condition holds and $\alpha$ is bounded by some constant that is related to the error bound condition. In a very recent work by Lin, Ma and Zhang [15], it was shown that the ADMM (1.3) possesses sublinear convergence rate in both ergodic and non-ergodic sense under the conditions that $f_2, \ldots, f_N$ are strongly convex and $\gamma$ is restricted to certain region.

**Our contribution.** In this paper, we show the global linear convergence of ADMM (1.3) when $N \geq 3$. It should be noted that the linear convergence results in [16, 4, 11, 1] are for the case $N = 2$, while ours consider the case when $N \geq 3$. Moreover, compared with the *local* linear convergence results in [11] and [1] for $N = 2$, we prove the *global* linear convergence for $N \geq 3$. Furthermore, our result is for the *original standard multi-block ADMM* (1.3), while the one presented in [14] is a variant of (1.3) which replaces (1.4) with (1.5). To the best of our knowledge, our results in this paper are the first global linear convergence results for the *original standard multi-block ADMM* (1.3) when $N \geq 3$.

The rest of this paper is organized as follows. In Section 2, we provide some preliminaries and prove three technical lemmas for the subsequent analysis. In Section 3, we prove the global linear convergence of ADMM (1.3) under three different scenarios. Finally, we conclude the paper in Section 4.

## 2 Preliminaries and Technical Lemmas

We use $\Omega^* \subset \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_N \times \mathbf{R}^p$ to denote the set of primal-dual optimal solutions of (1.2). Note that according to the first-order optimality conditions for (1.2), solving (1.2) is equivalent to finding

$$(x_1^*, \cdots, x_N^*, \lambda^*) \in \Omega^*$$

such that the followings hold:

$$A_i^\top \lambda^* \in \partial f_i(x_i^*), i = 1, 2, \cdots, N, \tag{2.1}$$

$$\sum_{i=1}^N A_i x_i^* - b = 0. \tag{2.2}$$

We thus make the following assumption throughout this paper.

**Assumption 2.1** *The optimal set* $\Omega^*$ *for problem* (1.2) *is non-empty.*

In our analysis, the following well-known identity is used frequently:

$$(w_1 - w_2)^\top (w_3 - w_4) = \frac{1}{2}\left(\|w_1 - w_4\|^2 - \|w_1 - w_3\|^2\right) + \frac{1}{2}\left(\|w_2 - w_3\|^2 - \|w_2 - w_4\|^2\right). \tag{2.3}$$

**Notations.** We use $g_i$ to denote a subgradient of $f_i$; $\lambda_{\max}(B)$ and $\lambda_{\min}(B)$ denote respectively the largest and smallest eigenvalues of a real symmetric matrix $B$; $\|x\|$ denotes the Euclidean norm of $x$. We use $\sigma_i > 0$ to denote the convexity parameter of $f_i$, i.e., the following inequalities hold for $i = 1, \ldots, N$:

$$(x - y)^\top (g_i(x) - g_i(y)) \geq \sigma_i \|x - y\|^2, \quad \forall x, y \in \mathcal{X}_i, \tag{2.4}$$

where $g_i(x) \in \partial f_i(x)$ is the subdifferential of $f_i$. Note that $f_i$ is strongly convex if and only if $\sigma_i > 0$, and if $f_i$ is convex but not strongly convex, then $\sigma_i = 0$.

In this paper, we consider three scenarios that lead to global linear convergence of ADMM (1.3). The conditions of the three scenarios are listed in Table 1.

| scenario | strongly convex | Lipschitz continuous | full row rank | full column rank |
|---|---|---|---|---|
| 1 | $f_2, \cdots, f_N$ | $\nabla f_N$ | $A_N$ | — |
| 2 | $f_1, \cdots, f_N$ | $\nabla f_1, \cdots, \nabla f_N$ | — | — |
| 3 | $f_2, \cdots, f_N$ | $\nabla f_1, \cdots, \nabla f_N$ | — | $A_1$ |

Table 1: Three scenarios leading to global linear convergence

We remark here that when $N = 2$, the three scenarios listed in Table 1 actually reduce to the same conditions considered by Deng and Yin as scenarios 1, 4 and 3, respectively in [4]. We also remark here that since we incorporated the indicator functions into the objective function in (1.2), scenario 1 actually requires that there is no constraint $x_N \in \mathcal{X}_N$; scenarios 2 and 3 require that there is no constraint $x_i \in \mathcal{X}_i$, $i = 1, \ldots, N$.

The first-order optimality conditions for the $N$ subproblems in (1.3) are given by

$$A_i^\top \lambda^k - \gamma A_i^\top \left(\sum_{j=1}^i A_j x_j^{k+1} + \sum_{j=i+1}^N A_j x_j^k - b\right) \in \partial f_i(x_i^{k+1}), \quad i = 1, 2, \cdots, N, \tag{2.5}$$

where we have adopted the convention $\sum_{j=N+1}^{N} a_j = 0$. By combining with the updating formula for $\lambda^k$ (1.4), (2.5) can be rewritten as

$$A_i^\top \lambda^{k+1} - \gamma A_i^\top \left[ \sum_{j=i+1}^{N} A_j(x_j^k - x_j^{k+1}) \right] \in \partial f_i(x_i^{k+1}), \quad i = 1, 2, \cdots, N. \tag{2.6}$$

Before we present the linear convergence of ADMM (1.3), we prove the following three technical lemmas that will be used in subsequent analysis.

**Lemma 2.2** *Let $(x_1^*, \ldots, x_N^*, \lambda^*) \in \Omega^*$. The sequence $\{x_1^k, x_2^k, \ldots, x_N^k, \lambda^k\}$ generated via ADMM (1.3) satisfies,*

$$\begin{aligned} &\left( \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^k) \right\|^2 + \frac{1}{2\gamma} \|\lambda^* - \lambda^k\|^2 \right) - \left( \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1}) \right\|^2 + \frac{1}{2\gamma} \|\lambda^* - \lambda^{k+1}\|^2 \right) \\ &\geq \sum_{i=1}^{N-1} \left[ \left( \sigma_i - \frac{\gamma(2N-i)(i-1)}{4} \lambda_{\max}(A_i^\top A_i) \right) \|x_i^{k+1} - x_i^*\|^2 \right] \\ &\quad + \left( \sigma_N - \frac{\gamma(N+1)(N-2)}{4} \lambda_{\max}(A_N^\top A_N) \right) \|x_N^{k+1} - x_N^*\|^2 + \frac{\gamma}{2} \left\| A_1 x_1^{k+1} + \sum_{j=2}^{N} A_j x_j^k - b \right\|^2. \end{aligned} \tag{2.7}$$

*Proof.* Combining (2.6), (2.1) and (2.4) yields,

$$(x_i^{k+1} - x_i^*)^\top A_i^\top \left( \lambda^{k+1} - \lambda^* - \gamma \sum_{j=i+1}^{N} A_j(x_j^k - x_j^{k+1}) \right) \geq \sigma_i \|x_i^{k+1} - x_i^*\|^2, \quad i = 1, \ldots, N. \tag{2.8}$$

From (1.4) and (2.2), it is easy to obtain

$$\sum_{i=1}^{N} A_i(x_i^{k+1} - x_i^*) = \frac{1}{\gamma}(\lambda^k - \lambda^{k+1}). \tag{2.9}$$

Summing (2.8) over $i = 1, \cdots, N$ and using (2.9), we can get

$$\frac{1}{\gamma}(\lambda^k - \lambda^{k+1})^\top(\lambda^{k+1} - \lambda^*) + \gamma \sum_{i=1}^{N-1} (x_i^* - x_i^{k+1})^\top A_i^\top \left[ \sum_{j=i+1}^{N} A_j(x_j^k - x_j^{k+1}) \right] \geq \sum_{i=1}^{N} \sigma_i \|x_i^{k+1} - x_i^*\|^2. \tag{2.10}$$

By adopting the convention $\sum_{i=1}^{0} a_i = 0$, we have that

$$
\begin{aligned}
& \sum_{i=1}^{N-1} (x_i^* - x_i^{k+1})^\top A_i^\top \left[ \sum_{j=i+1}^{N} A_j (x_j^k - x_j^{k+1}) \right] \\
= & \sum_{i=1}^{N-1} \left[ \left( \sum_{j=1}^{i} A_j x_j^* - b \right) - \left( \sum_{j=1}^{i-1} A_j x_j^* + A_i x_i^{k+1} - b \right) \right]^\top \left[ \left( - \sum_{j=i+1}^{N} A_j x_j^{k+1} \right) - \left( - \sum_{j=i+1}^{N} A_j x_j^k \right) \right] \\
= & \sum_{i=1}^{N-1} \left[ \frac{1}{2} \left( \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 \right) \right. \\
& \left. + \frac{1}{2} \left( \left\| \sum_{j=1}^{i-1} A_j x_j^* + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2 - \left\| \sum_{j=1}^{i-1} A_j x_j^* + A_i x_i^{k+1} + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 \right) \right] \\
\leq & \frac{1}{2} \sum_{i=1}^{N-1} \left( \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 \right) \\
& + \frac{1}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=1}^{i-1} A_j x_j^* + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2 - \frac{1}{2} \left\| A_1 x_1^{k+1} + \sum_{j=2}^{N} A_j x_j^k - b \right\|^2 \\
= & \frac{1}{2} \sum_{i=1}^{N-1} \left( \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 \right) \\
& + \frac{1}{2\gamma^2} \|\lambda^{k+1} - \lambda^k\|^2 + \frac{1}{2} \sum_{i=2}^{N-1} \left\| \sum_{j=1}^{i-1} A_j x_j^* + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2 - \frac{1}{2} \left\| A_1 x_1^{k+1} + \sum_{j=2}^{N} A_j x_j^k - b \right\|^2, \quad (2.11)
\end{aligned}
$$

where in the second equality we have used the identity (2.3), and the last equality follows from (1.4). By combining (2.10) and (2.11), we have

$$
\begin{aligned}
& \frac{\gamma}{2} \sum_{i=1}^{N-1} \left( \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 \right) \\
& + \frac{1}{\gamma} (\lambda^k - \lambda^{k+1})^\top (\lambda^{k+1} - \lambda^*) + \frac{1}{2\gamma} \|\lambda^{k+1} - \lambda^k\|^2 + \frac{\gamma}{2} \sum_{i=2}^{N-1} \left\| \sum_{j=1}^{i-1} A_j x_j^* + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2 \quad (2.12) \\
\geq & \sum_{i=1}^{N} \sigma_i \|x_i^{k+1} - x_i^*\|^2 + \frac{\gamma}{2} \left\| A_1 x_1^{k+1} + \sum_{j=2}^{N} A_j x_j^k - b \right\|^2 .
\end{aligned}
$$

Using again (2.2), we obtain

$$\left\|\sum_{j=1}^{i-1} A_j x_j^* + \sum_{j=i}^{N} A_j x_j^{k+1} - b\right\|^2 = \left\|\sum_{j=i}^{N} A_j (x_j^{k+1} - x_j^*)\right\|^2 \leq (N-i+1)\sum_{j=i}^{N} \lambda_{\max}(A_j^\top A_j)\|x_j^{k+1} - x_j^*\|^2,$$

where the inequality follows from the convexity of $\|\cdot\|^2$. Therefore, we have

$$\begin{aligned}
&\sum_{i=2}^{N-1}\left\|\sum_{j=1}^{i-1} A_j x_j^* + \sum_{j=i}^{N} A_j x_j^{k+1} - b\right\|^2 \\
&\leq \sum_{i=2}^{N-1}\left((N-i+1)\sum_{j=i}^{N}\lambda_{\max}(A_j^\top A_j)\|x_j^{k+1} - x_j^*\|^2\right) \qquad (2.13)\\
&= \sum_{i=2}^{N-1}\frac{(2N-i)(i-1)}{2}\lambda_{\max}(A_i^\top A_i)\|x_i^{k+1} - x_i^*\|^2 + \frac{(N+1)(N-2)}{2}\lambda_{\max}(A_N^\top A_N)\|x_N^{k+1} - x_N^*\|^2.
\end{aligned}$$

By combining (2.12) and (2.13) and using the identity

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

we have

$$\begin{aligned}
&\frac{\gamma}{2}\sum_{i=1}^{N-1}\left(\left\|\sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^k - b\right\|^2 - \left\|\sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b\right\|^2\right) \\
&+ \frac{1}{2\gamma}\left(\|\lambda^* - \lambda^k\|^2 - \|\lambda^* - \lambda^{k+1}\|^2\right) \\
&\geq \sum_{i=1}^{N-1}\left[\left(\sigma_i - \frac{\gamma(2N-i)(i-1)}{4}\lambda_{\max}(A_i^\top A_i)\right)\|x_i^{k+1} - x_i^*\|^2\right] \\
&+ \left(\sigma_N - \frac{\gamma(N+1)(N-2)}{4}\lambda_{\max}(A_N^\top A_N)\right)\|x_N^{k+1} - x_N^*\|^2 + \frac{\gamma}{2}\left\|A_1 x_1^{k+1} + \sum_{j=2}^{N} A_j x_j^k - b\right\|^2,
\end{aligned}$$

which further implies (2.7) by using (2.2). □

**Remark 2.3** *We note here that* (2.7) *can be equivalently rearranged as*

$$\begin{aligned}
&\left(\gamma\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N}A_j(x_j^*-x_j^k)\right\|^2+\frac{1}{2\gamma}\|\lambda^*-\lambda^k\|^2\right)-\left(\gamma\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N}A_j(x_j^*-x_j^{k+1})\right\|^2+\frac{1}{2\gamma}\|\lambda^*-\lambda^{k+1}\|^2\right)\\
\geq&\sum_{i=1}^{N-1}\left[\left(\sigma_i-\frac{\gamma(2N-i)(i-1)}{4}\lambda_{\max}(A_i^\top A_i)\right)\|x_i^{k+1}-x_i^*\|^2\right]\\
&+\left(\sigma_N-\frac{\gamma(N+1)(N-2)}{4}\lambda_{\max}(A_N^\top A_N)\right)\|x_N^{k+1}-x_N^*\|^2+\frac{\gamma}{2}\left\|A_1x_1^{k+1}+\sum_{j=2}^{N}A_jx_j^k-b\right\|^2\\
&+\frac{\gamma}{2}\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N}A_j(x_j^*-x_j^k)\right\|^2-\frac{\gamma}{2}\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N}A_j(x_j^*-x_j^{k+1})\right\|^2.
\end{aligned}\tag{2.14}$$

*Both* (2.7) *and* (2.14) *will be used in subsequent analysis. In scenario 1, we will use* (2.7) *to show that*

$$\frac{\gamma}{2}\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N}A_j(x_j^*-x_j^k)\right\|^2+\frac{1}{2\gamma}\|\lambda^*-\lambda^k\|^2$$

*converges to zero linearly; in scenarios 2 and 3, we will use* (2.14) *to show that*

$$\gamma\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N}A_j(x_j^*-x_j^k)\right\|^2+\frac{1}{2\gamma}\|\lambda^*-\lambda^k\|^2$$

*converges to zero linearly.*

The next lemma considers the convergence of $\{x_1^k,\ldots,x_N^k,\lambda^k\}$ under conditions listed in scenarios 2 and 3 in Table 1.

**Lemma 2.4** *Assume that the conditions listed in scenario 2 or scenario 3 in Table 1 hold. Moreover, we assume that $\gamma$ satisfies the following conditions:*

$$\gamma<\min_{i=2,\ldots,N-1}\left\{\frac{4\sigma_i}{(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)},\frac{4\sigma_N}{(N+1)(N-2)\lambda_{\max}(A_N^\top A_N)}\right\}.\tag{2.15}$$

*Then $(x_1^k,\ldots,x_N^k,\lambda^k)$ generated by ADMM* (1.3) *converges to some $(x_1^*,\ldots,x_N^*,\lambda^*)\in\Omega^*$.*

*Proof.* Note that the conditions listed in scenarios 2 and 3 in Table 1 both require that $f_2,\ldots,f_N$ are strongly convex. Denote the right hand side of inequality (2.7) by $\xi^k$. It follows from (2.15) and (2.7) that $\xi^k\geq 0$ and $\sum_{k=0}^{\infty}\xi^k<+\infty$, which further implies that $\xi^k\to 0$. Hence, for any $(x_1^*,\ldots,x_N^*,\lambda^*)\in\Omega^*$, we have $x_i^k-x_i^*\to 0$ for $i=2,\ldots,N$, and $A_1x_1^{k+1}+\sum_{j=2}^{N}A_jx_j^k-b\to 0$, which also implies that $A_1x_1^k-A_1x_1^*\to 0$. In scenario 2, it is assumed that $f_1$ is strongly convex. Thus $\sigma_1>0$ and (2.7) implies that $x_1^k - x_1^* \to 0$. In scenario 3, it is assumed that $A_1$ is of full column rank. It thus follows from $A_1 x_1^k - A_1 x_1^* \to 0$ that $x_1^k - x_1^* \to 0$.

Moreover, when (2.15) holds, it follows from (2.7) that $\frac{\gamma}{2}\sum_{i=1}^{N-1}\|\sum_{j=i+1}^{N} A_j(x_j^* - x_j^k)\|^2 + \frac{1}{2\gamma}\|\lambda^* - \lambda^k\|^2$ is non-increasing and upper bounded. It thus follows that $\|\lambda^* - \lambda^k\|^2$ converges and $\{\lambda^k\}$ is bounded. Therefore, $\{\lambda^k\}$ has a converging subsequence $\{\lambda^{k_j}\}$. Let $\bar{\lambda} = \lim_{j\to\infty}\{\lambda^{k_j}\}$. By passing the limit in (2.6), it holds that $A_i^\top \bar{\lambda} = \nabla f_i(x_i^*)$ for $i = 1, 2, \ldots, N$. Thus, $(x_1^*, \ldots, x_N^*, \bar{\lambda}) \in \Omega^*$ and we can just let $\lambda^* = \bar{\lambda}$. Since $\|\lambda^* - \lambda^k\|^2$ converges and $\lambda^{k_j} \to \lambda^*$, we conclude that $\lambda^k \to \lambda^*$. $\square$

Before proceeding to the next lemma, we define a constant $\kappa$ that will be used subsequently.

**Definition 2.1** *We define a constant $\kappa$ as follows.*

- *(i). If the matrix $[A_1, \cdots, A_N]$ is of full row rank, then $\kappa := \lambda_{\min}^{-1}([A_1, \cdots, A_N][A_1, \cdots, A_N]^\top) > 0$.*
- *(ii). Otherwise, assume $\mathrm{rank}([A_1, \cdots, A_N]) = r < p$. Without loss of generality, assuming that the first $r$ rows of $[A_1, \cdots, A_N]$ (denoted by $[A_1^r, \cdots, A_N^r]$) are linearly independent, we have*

$$[A_1, \cdots, A_N] = \begin{bmatrix} I \\ B \end{bmatrix} [A_1^r, \cdots, A_N^r], \tag{2.16}$$

*where $I \in \mathbf{R}^{r\times r}$ is the identity matrix and $B \in \mathbf{R}^{(p-r)\times r}$. Let $E := (I + B^\top B)[A_1^r, \cdots, A_N^r]$. It is easy to see that $E$ has full row rank. Then $\kappa$ is defined as $\kappa := \lambda_{\min}^{-1}(EE^\top)\lambda_{\max}(I + B^\top B) > 0$.*

The next lemma concerns bounding $\|\lambda^{k+1} - \lambda^*\|^2$ using terms related to $x_i^k - x_i^*$, $i = 1, \ldots, N$.

**Lemma 2.5** *Let $(x_1^*, \ldots, x_N^*, \lambda^*) \in \Omega^*$. Assume that the conditions listed in scenario 2 or scenario 3 in Table 1 hold, and $\gamma$ satisfies (2.15). Suppose $\nabla f_i$ is Lipschitz continuous with constant $L_i$ for $i = 1, \ldots, N$, and the initial Lagrange multiplier $\lambda^0$ is in the range space of $[A_1, \cdots, A_N]$ (note that letting $\lambda^0 = 0$ suffices). It holds that*

$$\begin{aligned}\|\lambda^{k+1} - \lambda^*\|^2 \le & \sum_{i=1}^{N} \left(2\kappa L_i^2\right)\|x_i^{k+1} - x_i^*\|^2 \\ & + \sum_{i=1}^{N-1}\left(4\kappa\gamma^2\lambda_{\max}(A_i^\top A_i)\right)\left(\left\|\sum_{j=i+1}^{N} A_j(x_j^k - x_j^*)\right\|^2 + \left\|\sum_{j=i+1}^{N} A_j(x_j^{k+1} - x_j^*)\right\|^2\right),\end{aligned} \tag{2.17}$$

*where $\kappa > 0$ is defined in Definition 2.1.*

*Proof.* We first show the following inequality

$$\|\lambda^{k+1} - \lambda^*\|^2 \le \kappa \cdot \left\| \begin{bmatrix} A_1^\top \\ \vdots \\ A_N^\top \end{bmatrix} (\lambda^{k+1} - \lambda^*) \right\|^2. \tag{2.18}$$

In case (i), $[A_1, \cdots, A_N]$ has full row rank, so (2.18) holds trivially. Now we consider case (ii). By the updating formula of $\lambda^{k+1}$ (1.4) and (2.2), we know that if the initial Lagrange multiplier $\lambda^0$ is in the range space of $[A_1, \cdots, A_N]$, then $\lambda^k, k = 1, 2, \cdots$, always stay in the range space of $[A_1, \cdots, A_N]$, so does $\lambda^*$. Therefore, from (2.16), we can get

$$\lambda^{k+1} = \begin{bmatrix} I \\ B \end{bmatrix} \lambda_r^{k+1}, \quad \lambda^* = \begin{bmatrix} I \\ B \end{bmatrix} \lambda_r^*, \quad \begin{bmatrix} A_1^\top \\ \vdots \\ A_N^\top \end{bmatrix} (\lambda^{k+1} - \lambda^*) = \begin{bmatrix} (A_1^r)^\top \\ \vdots \\ (A_N^r)^\top \end{bmatrix} (I + B^\top B)(\lambda_r^{k+1} - \lambda_r^*),$$

where $\lambda_r^{k+1}$ and $\lambda_r^*$ denote the first $r$ rows of $\lambda^{k+1}$ and $\lambda^*$, respectively. Since $E := (I+B^\top B)[A_1^r, \cdots, A_N^r]$ has full row rank, it now follows that

$$\left\| \begin{bmatrix} A_1^\top \\ \vdots \\ A_N^\top \end{bmatrix} (\lambda^{k+1} - \lambda^*) \right\|^2 = \|E^\top(\lambda_r^{k+1} - \lambda_r^*)\|^2 \geq \lambda_{\min}(EE^\top)\|\lambda_r^{k+1} - \lambda_r^*\|^2 \geq \frac{\lambda_{\min}(EE^\top)}{\lambda_{\max}(I + B^\top B)}\|\lambda^{k+1} - \lambda^*\|^2,$$

which implies (2.18).

Using the optimality conditions (2.6), and the Lipschitz continuity of $\nabla f_i, i = 1, \cdots, N$, we have

$$\left\| \begin{bmatrix} A_1^\top \\ A_2^\top \\ \vdots \\ A_N^\top \end{bmatrix} (\lambda^{k+1} - \lambda^*) + \begin{bmatrix} -\gamma A_1^\top \\ 0 \\ \vdots \\ 0 \end{bmatrix} \left( \sum_{j=2}^N A_j(x_j^k - x_j^{k+1}) \right) + \cdots + \begin{bmatrix} 0 \\ \vdots \\ -\gamma A_{N-1}^\top \\ 0 \end{bmatrix} (A_N(x_N^k - x_N^{k+1})) \right\|^2$$
$$= \sum_{i=1}^N \|\nabla f_i(x_i^{k+1}) - \nabla f_i(x_i^*)\|^2 \leq \sum_{i=1}^N L_i^2 \|x_i^{k+1} - x_i^*\|^2,$$

which together with (2.18) implies that

$$\begin{aligned}
& \|\lambda^{k+1} - \lambda^*\|^2 \\
\leq & \kappa \cdot \left\| \begin{bmatrix} A_1^\top \\ A_2^\top \\ \vdots \\ A_N^\top \end{bmatrix} (\lambda^{k+1} - \lambda^*) \right\|^2 \\
\leq & 2\kappa \left( \left\| \begin{bmatrix} -\gamma A_1^\top \\ 0 \\ \vdots \\ 0 \end{bmatrix} \left( \sum_{j=2}^N A_j(x_j^k - x_j^{k+1}) \right) + \cdots + \begin{bmatrix} 0 \\ \vdots \\ -\gamma A_{N-1}^\top \\ 0 \end{bmatrix} (A_N(x_N^k - x_N^{k+1})) \right\|^2 + \sum_{i=1}^N L_i^2 \|x_i^{k+1} - x_i^*\|^2 \right) \\
\leq & 2\kappa\gamma^2 \sum_{i=1}^{N-1} \lambda_{\max}(A_i^\top A_i) \left\| \sum_{j=i+1}^N A_j(x_j^k - x_j^{k+1}) \right\|^2 + 2\kappa \sum_{i=1}^N L_i^2 \|x_i^{k+1} - x_i^*\|^2 \\
\leq & 4\kappa\gamma^2 \sum_{i=1}^{N-1} \lambda_{\max}(A_i^\top A_i) \left( \left\| \sum_{j=i+1}^N A_j(x_j^k - x_j^*) \right\|^2 + \left\| \sum_{j=i+1}^N A_j(x_j^{k+1} - x_j^*) \right\|^2 \right) + 2\kappa \sum_{i=1}^N L_i^2 \|x_i^{k+1} - x_i^*\|^2.
\end{aligned}$$

□

## 3 Global Linear Convergence of the ADMM

In this section, we prove the global linear convergence of the ADMM (1.3) under the three scenarios listed in Table 1. We note the following inequality,

$$\sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1}) \right\|^2 \leq \sum_{i=2}^{N} \left[ \frac{(2N-i)(i-1)}{2} \lambda_{\max}(A_i^\top A_i) \|x_i^* - x_i^{k+1}\|^2 \right], \tag{3.1}$$

which follows from the convexity of $\|\cdot\|^2$. We shall use this inequality in our subsequent analysis.

### 3.1 $Q$-linear convergence under scenario 1

**Theorem 3.1** *Suppose that the conditions listed in scenario 1 in Table 1 hold. If $\gamma$ satisfies* (2.15)*, then it holds that*

$$\begin{aligned} &\left( \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^k) \right\|^2 + \frac{1}{2\gamma} \|\lambda^* - \lambda^k\|^2 \right) \\ \geq &(1+\delta_1) \left( \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1}) \right\|^2 + \frac{1}{2\gamma} \|\lambda^* - \lambda^{k+1}\|^2 \right), \end{aligned} \tag{3.2}$$

*where*

$$\delta_1 := \min_{i=2,\ldots,N-1} \left\{ \frac{4\sigma_i - \gamma(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)}{\gamma(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)}, \frac{4\gamma\sigma_N - \gamma^2(N+1)(N-2)\lambda_{\max}(A_N^\top A_N)}{2\lambda_{\min}^{-1}(A_N A_N^\top) L_N^2 + \gamma^2 N(N-1)\lambda_{\max}(A_N^\top A_N)} \right\}. \tag{3.3}$$

*Note that it follows from* (2.15) *that $\delta_1 > 0$. As a result of* (3.2)*, we conclude that*

$$\left( \sum_{j=2}^{N} A_j x_j^k, \sum_{j=3}^{N} A_j x_j^k, \ldots, \sum_{j=N}^{N} A_j x_j^k, \lambda^k \right)$$

*converges $Q$-linearly.*

*Proof.* Because $\nabla f_N$ is Lipschitz continuous with constant $L_N$, by setting $i = N$ in (2.6) and (2.1), we get

$$\|A_N^\top(\lambda^{k+1} - \lambda^*)\|^2 = \|\nabla f_N(x_N^{k+1}) - \nabla f_N(x_N^*)\|^2 \leq L_N^2 \|x_N^{k+1} - x_N^*\|^2,$$

which implies

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \lambda_{\min}^{-1}(A_N A_N^\top) L_N^2 \|x_N^{k+1} - x_N^*\|^2, \tag{3.4}$$

due to the fact that $A_N$ is of full row rank.

By combining (2.7), (3.3), (3.1) and (3.4), it follows that (note that we do not assume that $f_1$ is strongly convex, and thus $\sigma_1 = 0$),

$$
\begin{aligned}
&\left(\frac{\gamma}{2}\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N} A_j(x_j^* - x_j^k)\right\|^2 + \frac{1}{2\gamma}\|\lambda^* - \lambda^k\|^2\right) - \left(\frac{\gamma}{2}\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1})\right\|^2 + \frac{1}{2\gamma}\|\lambda^* - \lambda^{k+1}\|^2\right) \\
\geq& \sum_{i=2}^{N-1}\left[\left(\sigma_i - \frac{\gamma(2N-i)(i-1)}{4}\lambda_{\max}(A_i^\top A_i)\right)\|x_i^{k+1} - x_i^*\|^2\right] \\
&+ \left(\sigma_N - \frac{\gamma(N+1)(N-2)}{4}\lambda_{\max}(A_N^\top A_N)\right)\|x_N^{k+1} - x_N^*\|^2 \\
\geq& \delta_1\left[\sum_{i=2}^{N}\left[\frac{\gamma(2N-i)(i-1)}{4}\lambda_{\max}(A_i^\top A_i)\|x_i^* - x_i^{k+1}\|^2\right] + \frac{\lambda_{\min}^{-1}(A_N A_N^\top)L_N^2}{2\gamma}\|x_N^* - x_N^{k+1}\|^2\right] \\
\geq& \delta_1\left[\frac{\gamma}{2}\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1})\right\|^2 + \frac{1}{2\gamma}\|\lambda^* - \lambda^{k+1}\|^2\right],
\end{aligned}
$$

which further implies (3.2). □

## 3.2 $Q$-linear convergence under scenario 2

**Theorem 3.2** *Suppose that the conditions listed in scenario 2 in Table 1 hold. If $\gamma$ satisfies*

$$\gamma < \min_{i=2,\cdots,N-1}\left\{\frac{4\sigma_i}{3(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)}, \frac{4\sigma_N}{(3N^2-3N-2)\lambda_{\max}(A_N^\top A_N)}\right\}, \tag{3.5}$$

*then it holds that*

$$
\begin{aligned}
&\left(\gamma\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N} A_j(x_j^* - x_j^k)\right\|^2 + \frac{1}{2\gamma}\|\lambda^* - \lambda^k\|^2\right) \\
\geq&(1+\delta_2)\left(\gamma\sum_{i=1}^{N-1}\left\|\sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1})\right\|^2 + \frac{1}{2\gamma}\|\lambda^* - \lambda^{k+1}\|^2\right),
\end{aligned} \tag{3.6}
$$

*where*

$$\delta_2 := \min\left\{\frac{\sigma_1\gamma}{\kappa L_1^2}, \delta_3, \delta_4, \delta_5\right\}, \tag{3.7}$$

*and*

$$
\begin{cases}
\delta_3 := \min\limits_{i=2,\dots,N-1}\left\{\dfrac{4\sigma_i\gamma - 3\gamma^2(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)}{2\gamma^2(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i) + 4\kappa L_i^2}\right\}, \\
\delta_4 := \min\limits_{i=1,\dots,N-1}\left\{\dfrac{1}{4\kappa\lambda_{\max}(A_i^\top A_i)}\right\}, \\
\delta_5 := \dfrac{4\sigma_N\gamma - (3N^2-3N-2)\gamma^2\lambda_{\max}(A_N^\top A_N)}{2\gamma^2 N(N-1)\lambda_{\max}(A_N^\top A_N) + 4\kappa L_N^2},
\end{cases} \tag{3.8}
$$

*where $\kappa$ is defined in Definition 2.1. Note that it follows from* (3.5) *that $\delta_2 > 0$. As a result of* (3.6)*, we conclude that*

$$\left(\sum_{j=2}^{N} A_j x_j^k, \sum_{j=3}^{N} A_j x_j^k, \ldots, \sum_{j=N}^{N} A_j x_j^k, \lambda^k\right)$$

*converges Q-linearly.*

*Proof.* By combining (2.17) and (3.1), we have

$$\begin{aligned}
&(1+\delta_2)\left(\gamma \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j (x_j^* - x_j^{k+1}) \right\|^2 \right) + \delta_2 \left( \frac{1}{2\gamma} \|\lambda^* - \lambda^{k+1}\|^2 \right) \\
\leq &(1+\delta_2) \sum_{i=2}^{N} \left[ \frac{\gamma(2N-i)(i-1)}{2} \lambda_{\max}(A_i^\top A_i) \|x_i^* - x_i^{k+1}\|^2 \right] \\
&+ \frac{\delta_2}{2\gamma} \left( \sum_{i=1}^{N-1} \left(4\kappa\gamma^2 \lambda_{\max}(A_i^\top A_i)\right) \left( \left\| \sum_{j=i+1}^{N} A_j (x_j^k - x_j^*) \right\|^2 + \left\| \sum_{j=i+1}^{N} A_j (x_j^{k+1} - x_j^*) \right\|^2 \right) \right. \\
&\left. + \sum_{i=1}^{N} \left(2\kappa L_i^2\right) \|x_i^{k+1} - x_i^*\|^2 \right) \\
\leq &\sum_{i=1}^{N-1} \left[ \left( \sigma_i - \frac{\gamma(2N-i)(i-1)}{4} \lambda_{\max}(A_i^\top A_i) \right) \|x_i^{k+1} - x_i^*\|^2 \right] \\
&+ \left( \sigma_N - \frac{\gamma(N+1)(N-2)}{4} \lambda_{\max}(A_N^\top A_N) \right) \|x_N^{k+1} - x_N^*\|^2 \\
&+ \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j (x_j^* - x_j^k) \right\|^2 + \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j (x_j^* - x_j^{k+1}) \right\|^2,
\end{aligned} \tag{3.9}$$

where the last inequality follows from the definition of $\delta_2$ in (3.7). Finally we note that combining (3.9) with (2.14) yields (3.6). ☐

### 3.3 $Q$-linear convergence under scenario 3

**Theorem 3.3** *Suppose that the conditions listed in scenario 3 in Table 1 hold. If $\gamma$ satisfies* (3.5)*, then it holds that*

$$\begin{aligned}
&\left(\gamma \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j (x_j^* - x_j^k) \right\|^2 + \frac{1}{2\gamma} \|\lambda^* - \lambda^k\|^2 \right) \\
\geq &(1+\delta_6) \left( \gamma \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j (x_j^* - x_j^{k+1}) \right\|^2 + \frac{1}{2\gamma} \|\lambda^* - \lambda^{k+1}\|^2 \right),
\end{aligned} \tag{3.10}$$

*where*

$$\delta_6 := \min\left\{\frac{\gamma^2}{4\kappa\gamma^2(N-1)\lambda_{\max}(A_1^\top A_1)+4\kappa L_1^2\lambda_{\min}^{-1}(A_1^\top A_1)}, \delta_3, \delta_4, \delta_5\right\}, \tag{3.11}$$

*with $\delta_3$, $\delta_4$ and $\delta_5$ defined in* (3.8)*. Note that it follows from* (3.5) *that $\delta_6 > 0$. As a result of* (3.10)*, we conclude that*

$$\left(\sum_{j=2}^N A_j x_j^k, \sum_{j=3}^N A_j x_j^k, \ldots, \sum_{j=N}^N A_j x_j^k, \lambda^k\right)$$

*converges Q-linearly.*

*Proof.* Since $A_1$ is of full column rank, it is easy to verify that

$$\begin{aligned}
\lambda_{\min}(A_1^\top A_1)\|x_1^{k+1}-x_1^*\|^2 \leq& \|A_1(x_1^{k+1}-x_1^*)\|^2 \\
=& \left\|\left(A_1x_1^{k+1}+\sum_{j=2}^N A_j x_j^k - b\right) - \left(\sum_{j=2}^N A_j(x_j^k - x_j^*)\right)\right\|^2 \\
\leq& 2\left\|A_1x_1^{k+1}+\sum_{j=2}^N A_j x_j^k - b\right\|^2 + 2\left\|\sum_{j=2}^N A_j(x_j^k - x_j^*)\right\|^2.
\end{aligned} \tag{3.12}$$

Combining (3.12) and (2.17) yields,

$$\begin{aligned}
&\frac{1}{2\gamma}\|\lambda^*-\lambda^{k+1}\|^2 \\
\leq& \sum_{i=2}^N \left(\frac{\kappa L_i^2}{\gamma}\right)\|x_i^*-x_i^{k+1}\|^2 \\
&+ \frac{2}{\gamma}(\kappa L_1^2\lambda_{\min}^{-1}(A_1^\top A_1))\left(\left\|A_1x_1^{k+1}+\sum_{j=2}^N A_j x_j^k - b\right\|^2 + \left\|\sum_{j=2}^N A_j(x_j^k - x_j^*)\right\|^2\right) \\
&+ \sum_{i=1}^{N-1}\left(2\kappa\gamma(N-1)\lambda_{\max}(A_i^\top A_i)\right)\left(\left\|\sum_{j=i+1}^N A_j(x_j^k-x_j^*)\right\|^2 + \left\|\sum_{j=i+1}^N A_j(x_j^{k+1}-x_j^*)\right\|^2\right).
\end{aligned} \tag{3.13}$$

Combining (3.13), (3.1) and (3.11) yields,

$$
\begin{aligned}
&(1+\delta_6)\gamma \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1}) \right\|^2 + \delta_6 \frac{1}{2\gamma} \|\lambda^* - \lambda^{k+1}\|^2 \\
\leq & \sum_{i=1}^{N-1} \left[ \left( \sigma_i - \frac{\gamma(2N-i)(i-1)}{4} \lambda_{\max}(A_i^\top A_i) \right) \|x_i^{k+1} - x_i^*\|^2 \right] \\
&+ \left( \sigma_N - \frac{\gamma(N+1)(N-2)}{4} \lambda_{\max}(A_N^\top A_N) \right) \|x_N^{k+1} - x_N^*\|^2 \\
&+ \frac{\gamma}{2} \left\| A_1 x_1^{k+1} + \sum_{j=2}^{N} A_j x_j^k - b \right\|^2 \\
&+ \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^k) \right\|^2 + \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^* - x_j^{k+1}) \right\|^2,
\end{aligned}
$$

which together with (2.14) implies (3.10). □

### 3.4 $R$-linear Convergence

From the results in Theorems 3.1, 3.2 and 3.3, we have the following immediate corollary on the $R$-linear convergence of ADMM (1.3).

**Corollary 3.4** *Under the same conditions in Theorem 3.1, or Theorem 3.2, or Theorem 3.3, $x_N^k$, $\lambda^k$ and $A_i x_i^k, i = 1, \ldots, N-1$ converge $R$-linearly. Moreover, if $A_i, i = 1, 2, \cdots, N-1$ are further assumed to be of full column rank, then $x_i^k, i = 1, 2, \cdots, N-1$ converge $R$-linearly.*

*Proof.* Note that under all the three scenarios, we have shown that the sequence

$$
\left( \sum_{j=2}^{N} A_j x_j^k, \sum_{j=3}^{N} A_j x_j^k, \ldots, \sum_{j=N}^{N} A_j x_j^k, \lambda^k \right)
$$

converges $Q$-linearly. It follows that $\lambda^k$ and $\sum_{j=i+1}^{N} A_j x_j^k, i = 1, \cdots, N-1$ converge $R$-linearly, since any part of a $Q$-linear convergent quantity converges $R$-linearly. It also implies that $A_2 x_2^k, \ldots, A_N x_N^k$ converge $R$-linearly. It now follows from (2.9) that $A_1 x_1^k$ converges $R$-linearly. By setting $i = N$ in (2.8), one obtains,

$$
(x_N^{k+1} - x_N^*)^\top A_N^\top (\lambda^{k+1} - \lambda^*) \geq \sigma_N \|x_N^{k+1} - x_N^*\|^2,
$$

which implies that

$$
\|x_N^{k+1} - x_N^*\| \|A_N\| \|\lambda^{k+1} - \lambda^*\| \geq \sigma_N \|x_N^{k+1} - x_N^*\|^2,
$$

i.e.,

$$\|x_N^{k+1} - x_N^*\| \le \frac{\|A_N\|}{\sigma_N} \|\lambda^{k+1} - \lambda^*\|.$$

The $R$-linear convergence of $x_N^k$ then follows from the fact that $\lambda^k$ converges $R$-linearly. □

Now we make some remarks on the convergence results presented in this section.

**Remark 3.5** *If we incorporate the indicator function into the objective function in* (1.2)*, then its subgradient cannot be Lipschitz continuous on the boundary of the constraint set. Therefore, scenarios 2 and 3 can only occur if the constraint sets $\mathcal{X}_i$'s are actually the whole space. However, scenario 1 does allow most of the constraint sets to exist; essentially, it only requires that $x_N$ is unconstrained, and all other blocks of variables can be constrained. It remains an interesting question to figure out if the linear convergence rate still holds if all blocks of variables are constrained.*

**Remark 3.6** *Finally, we remark that the scenario 1 in Table 1 also gives rise to a linear convergence rate of the ADMM for convex optimization with* inequality *constraints:*

$$\begin{array}{ll} \min & \tilde{f}_1(x_1) + \tilde{f}_2(x_2) + \cdots + \tilde{f}_N(x_N) \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + \cdots + A_N x_N \le b \\ & x_i \in \mathcal{X}_i,\ i = 1, 2, \ldots, N. \end{array}$$

*In that case, by introducing a slack variable $x_0$ with the constraint $x_0 \in \mathbf{R}^p_+$, the corresponding ADMM becomes*

$$\left\{ \begin{array}{lcl} x_0^{k+1} & := & \operatorname{argmin}_{x_0 \in \mathbf{R}^p_+} \mathcal{L}_\gamma(x_0, x_1^k, \ldots, x_N^k; \lambda^k) = \left( -\sum_{i=1}^N A_i x_i^k + b + \frac{1}{\gamma}\lambda^k \right)_+, \\ x_i^{k+1} & := & \operatorname{argmin}_{x_i \in \mathcal{X}_i} \mathcal{L}_\gamma(x_0^{k+1}, x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \cdots, x_N^k; \lambda^k),\ i = 1, 2, ..., N, \\ \lambda^{k+1} & := & \lambda^k - \gamma \left( x_0^{k+1} + \sum_{i=1}^N A_i x_i^{k+1} - b \right), \end{array} \right.$$

*where*

$$\mathcal{L}_\gamma(x_0, x_1, \ldots, x_N; \lambda) := \sum_{i=1}^N \tilde{f}_i(x_i) - \left\langle \lambda, x_0 + \sum_{i=1}^N A_i x_i - b \right\rangle + \frac{\gamma}{2} \left\| x_0 + \sum_{i=1}^N A_i x_i - b \right\|^2.$$

*Suppose that the functions $\tilde{f}_i$, $i = 2, \ldots, N$ are all strongly convex, and $\nabla \tilde{f}_N$ is Lipschitz continuous, $x_N \in \mathcal{X}_N$ does not present and $A_N$ has full row rank, Theorem 3.1 assures that the above ADMM algorithm converges globally linearly.*

## 4 Conclusions

In this paper we proved that the original ADMM for convex optimization with multi-block variables is linearly convergent under some conditions. In particular, we presented three scenarios under which a linear convergence rate holds for the ADMM; these conditions can be considered as extensions of the ones discussed in [4] for the 2-block ADMM. Convergence and complexity analysis for multi-block ADMM are important because the ADMM is widely used and acknowledged to be an efficient and effective practical solution method for large scale convex optimization models arising from image processing, statistics, machine learning, and so on.

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

[5] 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.

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

[7] J. Eckstein and D. P. Bertsekas. An alternating direction method for linear programming. Technical report, MIT Laboratory for Information and Decision Systems, 1990.

[8] 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.

[9] 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.

[10] 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.

[11] D. Han and X. Yuan. Local linear convergence of the alternating direction method of multipliers for quadratic programs. *SIAM J. Numer. Anal.,*, 51(6):3446–3457, 2013.

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

[13] 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.

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

[15] T. Lin, S. Ma, and S. Zhang. On the convergence rate of multi-block ADMM. *submitted*, March 2014.

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

[17] Z.-Q. Luo and P. Tseng. Error bounds and the convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM J. Optim.*, 2:43–54, 1992.

[18] 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.