Iteration Complexity Analysis of Multi-Block ADMM for a Family of Convex Minimization without Strong Convexity

Tianyi Lin[*] Shiqian Ma[*] Shuzhong Zhang[†]

May 6, 2015

Abstract

The alternating direction method of multipliers (ADMM) is widely used in solving structured convex optimization problems due to its superior practical performance. On the theoretical side however, a counterexample was shown in [7] indicating that the multi-block ADMM for minimizing the sum of N ($N \geq 3$) convex functions with N block variables linked by linear constraints may diverge. It is therefore of great interest to investigate further sufficient conditions on the input side which can guarantee convergence for the multi-block ADMM. The existing results typically require the strong convexity on parts of the objective. In this paper, we present convergence and convergence rate results for the multi-block ADMM applied to solve certain N-block ($N \geq 3$) convex minimization problems *without requiring strong convexity*. Specifically, we prove the following two results: (1) the multi-block ADMM returns an ϵ-optimal solution within $O(1/\epsilon^2)$ iterations by solving an associated perturbation to the original problem; (2) the multi-block ADMM returns an ϵ-optimal solution within $O(1/\epsilon)$ iterations when it is applied to solve a certain *sharing problem*, under the condition that the augmented Lagrangian function satisfies the Kurdyka-Łojasiewicz property, which essentially covers most convex optimization models except for some pathological cases.

Keywords: Alternating Direction Method of Multipliers (ADMM), Convergence Rate, Regularization, Kurdyka-Łojasiewicz property, Convex Optimization

1 Introduction

We consider 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 $f_i : \mathbf{R}^{n_i} \to \mathbf{R}$ are closed convex functions. One effective way to solve (1.1), whenever applicable, is the so-called Alternating Direction Method of Multipliers (ADMM). The ADMM is closely related to the Douglas-Rachford [11] and

[*]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.

Peaceman-Rachford [32] operator splitting methods that date back to 1950s. These operator splitting methods were further studied later in [30, 15, 17, 12]. 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 [5, 13] for more information.

The ADMM is constructed under an augmented Lagrangian framework, where 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 ADMM for solving (1.1), the following updating procedure is implemented:

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

Note that the 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 [16, 14]. There are also some recent works that study the convergence rate properties of ADMM when $N = 2$ (see, e.g., [23, 31, 10, 2, 22]).

However, the convergence of multi-block ADMM (1.2) (we call (1.2) *multi-block ADMM* when $N \geq 3$) has remained unclear for a long time. Recently, Chen et al. [7] constructed a counterexample to show the failure of ADMM (1.2) when $N \geq 3$. Notwithstanding its theoretical convergence assurance, the multi-block ADMM (1.2) has been applied very successfully to solve problems with $N \, (N \geq 3)$ block variables; for example, see [35, 33]. It is thus of great interest to further study sufficient conditions that can guarantee the convergence of multi-block ADMM. Some recent works on studying the sufficient conditions guaranteeing the convergence of multi-block ADMM are described briefly as follows. Han and Yuan [18] showed that the multi-ADMM (1.2) converges if all the functions $f_1, \ldots, f_N$ are strongly convex and γ is restricted to certain region. This condition is relaxed in [8, 28] to allow only $N-1$ functions to be strongly convex and γ is restricted to certain region. Especially, Lin, Ma and Zhang [28] proved the sublinear convergence rate under such conditions. Closely related to [8, 28], Cai, Han and Yuan [6] and Li, Sun and Toh [27] proved that for $N = 3$, convergence of multi-block ADMM can be guaranteed under the assumption that only one function among f_1, f_2 and f_3 is required to be strongly convex, and γ is restricted in certain region. In addition to strong convexity of $f_2, \ldots, f_N$, by assuming further conditions on the smoothness of the functions and some rank conditions on the matrices in the linear constraints, Lin, Ma and Zhang [29] proved the globally linear convergence of multi-block ADMM. Note that the above mentioned works all require that (parts of) the objective function is strongly convex. Without assuming strong convexity, Hong and Luo [25] studied a variant of ADMM (1.2) with small stepsize in updating the Lagrangian multiplier. Specifically, [25] 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 proven 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 and convergence rate of some other variants of ADMM (1.2), and we refer the interested readers to [20, 21, 19, 9, 34, 24, 36] for the details of these variants. However, it is observed by many researchers that modified versions of ADMM though with convergence guarantee, often perform slower than the multi-block ADMM with no convergent guarantee (see [34]). Therefore, in this paper, we focus on studying the sufficient conditions that guarantee the convergence of the direct extension of ADMM, i.e., the multi-block ADMM (1.2) and studying its convergence rate.

Our contribution. The main contribution in this paper lies in the following. First, we show that the ADMM (1.2) when $N \geq 3$ returns an ϵ-optimal solution within $O(1/\epsilon^2)$ iterations, with the condition that γ depends on ϵ. Here we do not assume strong convexity of any objective function f_i. It should be pointed out that our result does not contradict the counterexample proposed in [7] since we apply the ADMM (1.2) to an associated perturbed problem of (1.1) rather than (1.1) itself. Secondly, we show that the ADMM (1.2) when $N \geq 3$ returns an ϵ-optimal solution within $O(1/\epsilon)$ iterations under the condition that the augmented Lagrangian $\mathcal{L}_\gamma$ is a Kurdyka-Łojasiewicz (KL) function [3, 4], ∇f_N is Lipschitz continuous, $A_N = I$, and γ is sufficiently large. To the best of our knowledge, the convergence rate results given in this paper are the first sublinear convergence rate results for the unmodified multi-block ADMM without assuming any strong convexity of the objective function (note that although without assuming strong convexity, [25] studies a variant of the multi-block ADMM). In this sense, the results presented in this paper complement with the existing results in the literature.

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 $O(1/\epsilon^2)$ iteration complexity of ADMM (1.2) by introducing an associated problem of (1.1). In Section 4, we prove the $O(1/\epsilon)$ iteration complexity of ADMM (1.2) with Kurdyka-Łojasiewicz (KL) property.

2 Preliminaries

We denote $\Omega = \mathcal{X}_1 \times \ldots \times \mathcal{X}_N \times \mathbf{R}^p$ and the optimal set of (1.1) as Ω^*, and the following assumption is made throughout this paper.

Assumption 2.1 *The optimal set* Ω^* *for problem* (1.1) *is non-empty.*

According to the first-order optimality conditions for (1.1), solving (1.1) is equivalent to finding

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

such that the following holds:

$$\begin{cases} (x_i - x_i^*)^\top (g_i(x_i^*) - A_i^\top \lambda^*) \geq 0, & \forall x_i \in \mathcal{X}_i, \\ A_1 x_1^* + \cdots + A_N x_N^* - b = 0, \end{cases} \tag{2.1}$$

for $i = 1, 2, \ldots, N$.

In this paper, we analyze the iteration complexity of ADMM (1.2) under two scenarios. The conditions of the two scenarios are listed in Table 1. The following assumption is only used in Scenario 2.

Assumption 2.2 *We assume that $\mathcal{X}_N = \mathbf{R}^{n_N}$. We also assume that f_i has a finite lower bound, i.e., $\inf_{x_i \in \mathcal{X}_i} f_i(x_i) \geq f_i^* > -\infty$ for $i = 1, 2, \ldots, N$. Moreover, it is assumed that $f_i + \mathbf{1}_{\mathcal{X}_i}$ is a coercive function for $i = 1, 2, \ldots, N-1$, where $\mathbf{1}_{\mathcal{X}_i}$ denotes the indicator function of $\mathcal{X}_i$, i.e.,*

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

Furthermore, we assume that $\mathcal{L}_\gamma$ is a KL function (will be defined later).

Scenario	Lipschitz Continuous	Matrices	Additional Assumption	Iteration Complexity
1	—	—	$\frac{\epsilon}{2} \leq \gamma \leq \epsilon$	$O(1/\epsilon^2)$
2	∇f_N	$A_N = I$	$\gamma > \sqrt{2}L$ and Assumption 2.2	$O(1/\epsilon)$

Table 1: Two Scenarios Leading to Sublinear Convergence

Remark 2.3 *Some remarks are in order here regarding the conditions in Scenario 2. Note that it is not very restrictive to require $f_i + \mathbf{1}_{\mathcal{X}_i}$ to be a coercive function. In fact, many functions used as regularization terms including ℓ_1-norm, ℓ_2-norm, ℓ_∞-norm for vectors and nuclear norm for matrices are all coercive functions; assuming the compactness of $\mathcal{X}_i$ also leads to the coerciveness of $f_i + \mathbf{1}_{\mathcal{X}_i}$. Moreover, the assumptions $A_N = I$ and ∇f_N is Lipschitz continuous actually cover many interesting applications in practice. For example, many problems arising from machine learning, statistics, image processing and so on always have the following structure:*

$$\min \ f_1(x_1) + \cdots + f_{N-1}(x_{N-1}) + f_N(b - A_1 x_1 - \cdots - A_{N-1} x_{N-1}), \tag{2.2}$$

where f_N denotes a loss function on data fitting, which is usually a smooth function, and $f_1, \ldots, f_{N-1}$ are regularization terms to promote certain structures of the solution. This problem is usually referred as sharing problem (see, e.g., [5, 26]). (2.2) *can be reformulated as*

$$\begin{array}{ll} \min & f_1(x_1) + \cdots + f_{N-1}(x_{N-1}) + f_N(x_N) \\ \text{s.t.} & A_1 x_1 + \cdots + A_{N-1} x_{N-1} + x_N = b, \end{array} \tag{2.3}$$

which is in the form of (1.1) *and can be solved by ADMM (see [5, 26]). Note that $A_N = I$ in* (2.3) *and it is very natural to assume that ∇f_N is Lipschitz continuous. Thus the conditions in Scenario 2 are satisfied.*

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

$$u = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}, u^k = \begin{pmatrix} x_1^k \\ \vdots \\ x_N^k \end{pmatrix}, u^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_N^* \end{pmatrix}, w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, w^k = \begin{pmatrix} u^k \\ \lambda^k \end{pmatrix}, w^* = \begin{pmatrix} u^* \\ \lambda^* \end{pmatrix}.$$

We denote by $f(u) \equiv f_1(x_1) + \cdots + f_N(x_N)$ the objective function of problem (1.1); $\mathbf{1}_{\mathcal{X}}$ is the indicator function of $\mathcal{X}$; ∇f is the gradient of f; $\|x\|$ denotes the Euclidean norm of x.

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

$$(w_1 - w_2)^\top (w_3 - w_4) = \frac{1}{2}\left(\|w_1 - w_4\|^2 - \|w_1 - w_3\|^2\right) + \frac{1}{2}\left(\|w_3 - w_2\|^2 - \|w_4 - w_2\|^2\right), \quad (2.4)$$

$$(w_1 - w_2)^\top (w_3 - w_1) = \frac{1}{2}\left(\|w_2 - w_3\|^2 - \|w_1 - w_2\|^2 - \|w_1 - w_3\|^2\right). \quad (2.5)$$

3 Iteration Complexity of ADMM: Associated Perturbation

In this section, we prove the $O(1/\epsilon^2)$ iteration complexity of ADMM (1.2) under the conditions in Scenario 1 of Table 1. Indeed, given $\epsilon > 0$ sufficiently small and initial point u^0, we introduce an associated perturbed problem of (1.1), i.e.,

$$\begin{array}{ll} \min & f_1(x_1) + \tilde{f}_2(x_2) + \cdots + \tilde{f}_N(x_N) \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + \cdots + A_N x_N = b \\ & x_i \in \mathcal{X}_i, i = 1, \ldots, N, \end{array} \quad (3.1)$$

where $\tilde{f}_i(x_i) = f_i(x_i) + \frac{\mu}{2}\left\|A_i x_i - A_i x_i^0\right\|^2$ for $i = 2, \ldots, N$, and $\mu = \epsilon(N-2)(N+1)$. Note $\tilde{f}_i$ are not necessarily strongly convex. We prove that the ADMM (1.2) for associated perturbed problem (3.1) returns an ϵ-optimal solution of the original problem (1.1), in terms of both objective value and constraint violation, within $O(1/\epsilon^2)$ iterations.

The ADMM for solving (3.1) can be summarized as (note that some constant terms in the subproblems are discarded):

$$x_1^{k+1} := \operatorname*{argmin}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \frac{\gamma}{2}\left\|A_1 x_1 + \sum_{j=2}^{N} A_j x_j^k - b - \frac{1}{\gamma}\lambda^k\right\|^2, \quad (3.2)$$

$$x_i^{k+1} := \operatorname*{argmin}_{x_i \in \mathcal{X}_i} \tilde{f}_i(x_i) + \frac{\gamma}{2}\left\|\sum_{j=1}^{i-1} A_j x_j^{k+1} + A_i x_i + \sum_{j=i+1}^{N} A_j x_j^k - b - \frac{1}{\gamma}\lambda^k\right\|^2, \ i = 2, \ldots, N, \quad (3.3)$$

$$\lambda^{k+1} := \lambda^k - \gamma\left(A_1 x_1^{k+1} + A_2 x_2^{k+1} + \cdots + A_N x_N^{k+1} - b\right). \quad (3.4)$$

The first-order optimality conditions for (3.2)-(3.3) are given respectively by $x_i^{k+1} \in \mathcal{X}_i$ and

$$(x_1 - x_1^{k+1})^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^k + \gamma A_1^\top \left(A_1 x_1^{k+1} + \sum_{j=2}^N A_j x_j^k - b \right) \right] \geq 0, \tag{3.5}$$

$$(x_i - x_i^{k+1})^\top \left[g_i(x_i^{k+1}) + \mu A_i^\top A_i \left(x_i^{k+1} - x_i^0 \right) - A_i^\top \lambda^k + \gamma A_i^\top \left(\sum_{j=1}^i A_j x_j^{k+1} + \sum_{j=i+1}^N A_j x_j^k - b \right) \right] \geq 0, \tag{3.6}$$

hold for any $x_i \in \mathcal{X}_i$ and $g_i \in \partial f_i$, a subgradient of f_i, for $i = 1, 2, \ldots, N$. Moreover, by combining with (3.4), (3.5)-(3.6) can be rewritten as

$$(x_1 - x_1^{k+1})^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} + \gamma A_1^\top \left(\sum_{j=2}^N A_j (x_j^k - x_j^{k+1}) \right) \right] \geq 0, \tag{3.7}$$

$$(x_i - x_i^{k+1})^\top \left[g_i(x_i^{k+1}) + \mu A_i^\top A_i \left(x_i^{k+1} - x_i^0 \right) - A_i^\top \lambda^{k+1} + \gamma A_i^\top \left(\sum_{j=i+1}^N A_j (x_j^k - x_j^{k+1}) \right) \right] \geq 0. \tag{3.8}$$

Lemma 3.1 *Let $(x_1^{k+1}, x_2^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1}) \in \Omega$ be generated by the ADMM (1.2) from given $(x_2^k, \ldots, x_N^k, \lambda^k)$. For any $u^* = (x_1^*, x_2^*, \ldots, x_N^*) \in \Omega^*$ and $\lambda \in \mathbf{R}^p$, it holds true under conditions in Scenario 1 that*

$$\begin{aligned}
& f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ \vdots \\ x_N^* - x_N^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \lambda^{k+1} \\ -A_2^\top \lambda^{k+1} \\ \vdots \\ -A_N^\top \lambda^{k+1} \\ \sum_{i=1}^N A_i x_i^{k+1} - b \end{pmatrix} \\
& + \frac{1}{2\gamma} \left(\left\| \lambda - \lambda^k \right\|^2 - \left\| \lambda - \lambda^{k+1} \right\|^2 \right) + \frac{\epsilon (N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2 \\
& + \frac{\gamma}{2} \sum_{i=1}^{N-1} \left(\left\| \sum_{j=1}^i A_j x_j^* + \sum_{j=i+1}^N A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^i A_j x_j^* + \sum_{j=i+1}^N A_j x_j^{k+1} - b \right\|^2 \right) \\
\geq \; & 0.
\end{aligned} \tag{3.9}$$

Proof. Note that combining (3.7)-(3.8) yields

$$\begin{aligned}
& \begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ \vdots \\ x_N - x_N^{k+1} \end{pmatrix}^\top \left[\begin{pmatrix} g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} \\ g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} \\ \vdots \\ g_N(x_N^{k+1}) - A_N^\top \lambda^{k+1} \end{pmatrix} + \begin{pmatrix} 0 \\ \mu A_2^\top (A_2 x_2^{k+1} - A_2 x_2^0) \\ \vdots \\ \mu A_N^\top (A_N x_N^{k+1} - A_N x_N^0) \end{pmatrix} + H \begin{pmatrix} x_2^k - x_2^{k+1} \\ \vdots \\ x_N^k - x_N^{k+1} \end{pmatrix} \right] \\
\geq \; & 0,
\end{aligned} \tag{3.10}$$

where $H \in \mathbf{R}^{(\sum_{i=1}^N n_i)\times(\sum_{i=2}^N n_i)}$ is defined as follow:

$$H := \begin{pmatrix} \gamma A_1^\top A_2 & \gamma A_1^\top A_3 & \cdots & \gamma A_1^\top A_N \\ 0 & \gamma A_2^\top A_3 & \cdots & \gamma A_2^\top A_N \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma A_{N-1}^\top A_N \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

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

$$(x_i - x_i^{k+1})^\top A_i^\top \left(\sum_{j=i+1}^N A_j(x_j^k - x_j^{k+1})\right), \ i = 1, 2, \dots, N-1.$$

For $i = 1, 2, \dots, N-1$, we have,

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

where in the second equality we applied the identity (2.4).

Therefore, we have

$$\begin{aligned}
&\begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ \vdots \\ x_N - x_N^{k+1} \end{pmatrix}^\top \begin{pmatrix} \gamma A_1^\top A_2 & \gamma A_1^\top A_3 & \cdots & \gamma A_1^\top A_N \\ 0 & \gamma A_2^\top A_3 & \cdots & \gamma A_2^\top A_N \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma A_{N-1}^\top A_N \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_2^k - x_2^{k+1} \\ \vdots \\ x_N^k - x_N^{k+1} \end{pmatrix} \\
\leq\ & \frac{\gamma}{2} \sum_{i=1}^{N-1} \left(\left\| \sum_{j=1}^{i} A_j x_j + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 \right) \\
& + \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2 + \frac{\gamma}{2} \sum_{i=2}^{N-1} \left\| \sum_{j=1}^{i-1} A_j x_j + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2 .
\end{aligned} \tag{3.11}$$

Combining (3.4), (3.10) and (3.11), it holds for any $\lambda \in \mathbf{R}^p$ that

$$\begin{aligned}
&\begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ \vdots \\ x_N - x_N^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} \\ g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} \\ \vdots \\ g_N(x_N^{k+1}) - A_N^\top \lambda^{k+1} \\ \sum_{i=1}^N A_i x_i^{k+1} - b \end{pmatrix} + \frac{1}{\gamma} \left(\lambda - \lambda^{k+1} \right)^\top \left(\lambda^{k+1} - \lambda^k \right) \\
& + \mu \sum_{i=2}^{N} \left(x_i - x_i^{k+1} \right)^\top A_i^\top A_i \left(x_i^{k+1} - x_i^0 \right) + \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2 + \frac{\gamma}{2} \sum_{i=2}^{N-1} \left\| \sum_{j=1}^{i-1} A_j x_j + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2 \\
& + \frac{\gamma}{2} \sum_{i=1}^{N-1} \left(\left\| \sum_{j=1}^{i} A_j x_j + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 \right) \\
\geq\ & 0.
\end{aligned} \tag{3.12}$$

Using (2.5), we have

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

and

$$\begin{aligned}
& \mu \left(x_i - x_i^{k+1} \right)^\top A_j^\top A_j \left(x_i^{k+1} - x_i^0 \right) \\
=\ & \frac{\mu}{2} \left(\left\| A_i x_i - A_i x_i^0 \right\|^2 - \left\| A_i x_i^{k+1} - A_i x_i^0 \right\|^2 - \left\| A_i x_i - A_i x_i^{k+1} \right\|^2 \right) \\
\leq\ & \frac{\mu}{2} \left\| A_i x_i - A_i x_i^0 \right\|^2 - \frac{\mu}{2} \left\| A_i x_i - A_i x_i^{k+1} \right\|^2 .
\end{aligned}$$

Letting $u = u^*$ in (3.12), and invoking the convexity of f_i that

$$f_i(x_i^*) - f_i(x_i^{k+1}) \geq (x_i^* - x_i^{k+1})^\top g_i(x_i^{k+1}), \quad i = 1, 2, \ldots, N$$

and

$$\begin{aligned}\frac{\gamma}{2}\sum_{i=2}^{N-1}\left\|\sum_{j=1}^{i-1}A_jx_j^*+\sum_{j=i}^{N}A_jx_j^{k+1}-b\right\|^2 &= \frac{\gamma}{2}\sum_{i=2}^{N-1}\left\|\sum_{j=i}^{N}A_j(x_j^{k+1}-x_j^*)\right\|^2\\ &\le \frac{\gamma(N+1)(N-2)}{2}\sum_{i=2}^{N}\left\|A_ix_i^{k+1}-A_ix_i^*\right\|^2,\end{aligned}$$

we obtain,

$$\begin{aligned}& f(u^*)-f(u^{k+1})+\begin{pmatrix} x_1^*-x_1^{k+1}\\ x_2^*-x_2^{k+1}\\ \vdots\\ x_N^*-x_N^{k+1}\\ \lambda-\lambda^{k+1}\end{pmatrix}^\top\begin{pmatrix} -A_1^\top\lambda^{k+1}\\ -A_2^\top\lambda^{k+1}\\ \vdots\\ -A_N^\top\lambda^{k+1}\\ \sum_{i=1}^N A_ix_i^{k+1}-b\end{pmatrix}\\ & +\frac{1}{2\gamma}\left(\left\|\lambda-\lambda^k\right\|^2-\left\|\lambda-\lambda^{k+1}\right\|^2\right)+\frac{\mu}{2}\sum_{i=2}^{N}\left(\|A_ix_i^*-A_ix_i^0\|^2-\left\|A_ix_i^*-A_ix_i^{k+1}\right\|^2\right)\\ & +\frac{\gamma}{2}\sum_{i=1}^{N-1}\left(\left\|\sum_{j=1}^{i}A_jx_j^*+\sum_{j=i+1}^{N}A_jx_j^k-b\right\|^2-\left\|\sum_{j=1}^{i}A_jx_j^*+\sum_{j=i+1}^{N}A_jx_j^{k+1}-b\right\|^2\right)\\ & +\frac{\gamma(N+1)(N-2)}{2}\sum_{i=2}^{N}\left\|A_ix_i^*-A_ix_i^{k+1}\right\|^2\\ \ge\ & 0.\end{aligned}$$

This together with the facts that $\mu=\epsilon(N-2)(N+1)$ and $\gamma\le\epsilon$ implies that

$$\frac{\gamma(N+1)(N-2)}{2}\sum_{j=2}^{N}\left\|A_jx_j^*-A_jx_j^{k+1}\right\|^2-\frac{\mu}{2}\sum_{j=2}^{N}\left\|A_jx_j^*-A_jx_j^{k+1}\right\|^2\le 0,$$

which further implies the desired inequality (3.9). □

Now we are ready to prove the $O(1/\epsilon^2)$ iteration complexity of the ADMM for (1.1) in an ergodic case.

Theorem 3.2 *Let $(x_1^{k+1},x_2^{k+1},\ldots,x_N^{k+1},\lambda^{k+1})\in\Omega$ be generated by ADMM (3.2)-(3.4) from given $(x_2^k,\ldots,x_N^k,\lambda^k)$. For any integer $t>0$, let $\bar{u}^t=(\bar{x}_1^t,\bar{x}_2^t,\ldots,\bar{x}_N^t)$ and $\bar{\lambda}^t$ be defined as*

$$\bar{x}_i^t=\frac{1}{t+1}\sum_{k=0}^{t}x_i^{k+1},\ i=1,2,\ldots,N,\quad \bar{\lambda}^t=\frac{1}{t+1}\sum_{k=0}^{t}\lambda^{k+1}.$$

For any $(u^*, \lambda^*) \in \Omega^*$, by defining $\rho := \|\lambda^*\| + 1$, it holds in Scenario 1 that,

$$
\begin{aligned}
0 &\leq f(\bar{u}^t) - f(u^*) + \rho \left\| \sum_{i=1}^N A_i \bar{x}_i^t - b \right\| \\
&\leq \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^N A_j (x_j^0 - x_j^*) \right\|^2 + \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2 .
\end{aligned}
$$

This also implies that when $t = O(1/\epsilon^2)$, $\bar{u}^t = (\bar{x}_1^t, \bar{x}_2^t, \ldots, \bar{x}_N^t)$ is an ϵ-optimal solution to the original problem (1.1), i.e., both the error of the objective function value and the residual of the equality constraint satisfy that

$$
|f(\bar{u}^t) - f(u^*)| = O(\epsilon), \qquad \textit{and} \qquad \left\| \sum_{i=1}^N A_i \bar{x}_i^t - b \right\| = O(\epsilon). \tag{3.13}
$$

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

$$
\begin{aligned}
& f(u^*) - f(\bar{u}^t) + \lambda^\top \left(\sum_{i=1}^N A_i \bar{x}_i^t - b \right) \\
= \; & f(u^*) - f(\bar{u}^t) + \begin{pmatrix} x_1^* - \bar{x}_1^t \\ x_2^* - \bar{x}_2^t \\ \vdots \\ x_N^* - \bar{x}_N^t \\ \lambda - \bar{\lambda}^t \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \bar{\lambda}^t \\ -A_2^\top \bar{\lambda}^t \\ \vdots \\ -A_N^\top \bar{\lambda}^t \\ \sum_{i=1}^N A_i \bar{x}_i^t - b \end{pmatrix} \\
\geq \; & \frac{1}{t+1} \sum_{k=0}^t \left[f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ \vdots \\ x_N^* - x_N^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \lambda^{k+1} \\ -A_2^\top \lambda^{k+1} \\ \vdots \\ -A_N^\top \lambda^{k+1} \\ \sum_{i=1}^N A_i x_i^{k+1} - b \end{pmatrix} \right] \\
\geq \; & \frac{1}{t+1} \sum_{k=0}^t \left[\frac{1}{2\gamma} \left(\left\| \lambda - \lambda^{k+1} \right\|^2 - \left\| \lambda - \lambda^k \right\|^2 \right) - \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2 \right. \\
& \left. + \frac{\gamma}{2} \sum_{i=1}^{N-1} \left(\left\| \sum_{j=1}^i A_j x_j^* + \sum_{j=i+1}^N A_j x_j^{k+1} - b \right\|^2 - \left\| \sum_{j=1}^i A_j x_j^* + \sum_{j=i+1}^N A_j x_j^k - b \right\|^2 \right) \right] \\
\geq \; & -\frac{1}{2\gamma(t+1)} \left\| \lambda - \lambda^0 \right\|^2 - \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=1}^i A_j x_j^* + \sum_{j=i+1}^N A_j x_j^0 - b \right\|^2 \\
& - \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2 .
\end{aligned} \tag{3.14}
$$

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

$$0 \geq f(u^*) - f(\bar{u}^t) + (\lambda^*)^\top \left(\sum_{i=1}^N A_i \bar{x}_i^t - b \right).$$

Moreover, since $\rho := \|\lambda^*\| + 1$, by applying Cauchy-Schwarz inequality, we obtain

$$0 \leq f(\bar{u}^t) - f(u^*) + \rho \left\| \sum_{i=1}^N A_i \bar{x}_i^t - b \right\|. \tag{3.15}$$

By setting $\lambda = -\rho \left(\sum_{i=1}^N A_i \bar{x}_i^t - b \right) / \left\| \sum_{i=1}^N A_i \bar{x}_i^t - b \right\|$ in (3.14), and noting that $\|\lambda\| = \rho$, we obtain

$$\begin{aligned} & f(\bar{u}^t) - f(u^*) + \rho \left\| \sum_{i=1}^N A_i \bar{x}_i^t - b \right\| \\ \leq \quad & \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^N A_j (x_j^0 - x_j^*) \right\|^2 + \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2. \end{aligned} \tag{3.16}$$

When $t = O(1/\epsilon^2)$, and together with the condition that $\frac{\epsilon}{2} \leq \gamma \leq \epsilon$, we have

$$\frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^N A_j (x_j^0 - x_j^*) \right\|^2 + \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2 = O(\epsilon). \tag{3.17}$$

We now define the function

$$v(\xi) = \min\{ f(u) | \sum_{i=1}^N A_i x_i - b = \xi, x_i \in \mathcal{X}_i, i = 1, 2, \ldots, N \}.$$

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

$$v(\xi) \geq v(0) + \langle \lambda^*, \xi \rangle \geq f(u^*) - \|\lambda^*\| \|\xi\|. \tag{3.18}$$

Let $\bar{\xi} = \sum\limits_{i=1}^N A_i \bar{x}_i^t - b$, we have $f(\bar{u}^t) \geq v(\bar{\xi})$. Therefore, combining (3.15), (3.17) and (3.18), we get

$$\begin{aligned} & -\|\lambda^*\| \|\bar{\xi}\| \leq f(\bar{u}^t) - f(u^*) \\ \leq \quad & \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^N A_j (x_j^0 - x_j^*) \right\|^2 + \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2 - \rho \|\bar{\xi}\| \\ \leq \quad & C\epsilon - \rho \|\bar{\xi}\|, \end{aligned}$$

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

$$\| \sum_{i=1}^N A_i \bar{x}_i^t - b \| = \|\bar{\xi}\| \leq C\epsilon. \tag{3.19}$$

Moreover, by combining (3.15) and (3.19), one obtains that

$$-\rho C\epsilon \leq -\rho\|\bar{\xi}\| \leq f(\bar{u}^t) - f(u^*) \leq (1-\rho)C\epsilon. \tag{3.20}$$

Finally, we note that (3.19), (3.20) imply (3.13). $\square$

4 Iteration Complexity of ADMM: Kurdyka-Łojasiewicz Property

In this section, we prove an $O(1/\epsilon)$ iteration complexity of ADMM (1.2) under the conditions in Scenario 2 of Table 1. Indeed, we prove that the ADMM for the original problem (1.1) returns an ϵ-optimal solution within $O(1/\epsilon)$ iterations in Scenario 2.

Under the conditions in Scenario 2, the multi-block ADMM (1.2) for solving (1.1) can be rewritten as:

$$x_1^{k+1} := \operatorname*{argmin}_{x_1\in\mathcal{X}_1} f_1(x_1) + \frac{\gamma}{2}\left\|A_1x_1 + \sum_{j=2}^{N-1} A_jx_j^k + x_N^k - b - \frac{1}{\gamma}\lambda^k\right\|^2, \tag{4.1}$$

$$x_i^{k+1} := \operatorname*{argmin}_{x_i\in\mathcal{X}_i} f_i(x_i) + \frac{\gamma}{2}\left\|\sum_{j=1}^{i-1} A_jx_j^{k+1} + A_ix_i + \sum_{j=i+1}^{N-1} A_jx_j^k + x_N^k - b - \frac{1}{\gamma}\lambda^k\right\|^2, \quad i = 2,\ldots,N-1, \tag{4.2}$$

$$x_N^{k+1} := \operatorname{argmin} f_N(x_N) + \frac{\gamma}{2}\left\|\sum_{j=1}^{N-1} A_jx_j^{k+1} + x_N - b - \frac{1}{\gamma}\lambda^k\right\|^2, \tag{4.3}$$

$$\lambda^{k+1} := \lambda^k - \gamma\left(A_1x_1^{k+1} + A_2x_2^{k+1} + \cdots + A_{N-1}x_{N-1}^{k+1} + x_N^{k+1} - b\right). \tag{4.4}$$

The first-order optimality conditions for (4.1)-(4.3) are given respectively by $x_i^{k+1} \in \mathcal{X}_i, i = 1,\ldots,N-1$, and

$$g_1(x_1^{k+1}) - A_1^\top\lambda^k + \gamma A_1^\top\left(A_1x_1^{k+1} + \sum_{j=2}^{N-1} A_jx_j^k + x_N^k - b\right) = 0, \tag{4.5}$$

$$g_i(x_i^{k+1}) - A_i^\top\lambda^k + \gamma A_i^\top\left(\sum_{j=1}^{i} A_jx_j^{k+1} + \sum_{j=i+1}^{N-1} A_jx_j^k + x_N^k - b\right) = 0, \tag{4.6}$$

$$\nabla f_N(x_N^{k+1}) - \lambda^k + \gamma\left(\sum_{j=1}^{N-1} A_jx_j^{k+1} + x_N^{k+1} - b\right) = 0, \tag{4.7}$$

where $g_i \in \partial\left(f_i + \mathbf{1}_{\mathcal{X}_i}\right)$ is a subgradient of $f_i + \mathbf{1}_{\mathcal{X}_i}$ for $i = 1,2,\ldots,N-1$. Moreover, by combining with

(4.4), (4.5)-(4.7) can be rewritten as

$$g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} + \gamma A_1^\top \left(\sum_{j=2}^{N-1} A_j(x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right) = 0, \tag{4.8}$$

$$g_i(x_i^{k+1}) - A_i^\top \lambda^{k+1} + \gamma A_i^\top \left(\sum_{j=i+1}^{N-1} A_j(x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right) = 0, \tag{4.9}$$

$$\nabla f_N(x_N^{k+1}) - \lambda^{k+1} = 0. \tag{4.10}$$

Note that in Scenario 2 we require that $\mathcal{L}_\gamma$ is a Kurdyka-Łojasiewicz (KL) function. Let us first introduce the notion of the KL function and the KL property, which can be found, e.g., in [3, 4]. We denote $\operatorname{dist}(x, S) := \inf\{\|y - x\| : y \in S\}$ as the distance from x to S. Let $\eta \in (0, +\infty]$. We further denote Φ_η to be the class of all concave and continuous functions $\varphi : [0, \eta) \to \mathbf{R}_+$ satisfying the following conditions:

1. $\varphi(0) = 0$;
2. φ is C^1 on $(0, \eta)$ and continuous at 0;
3. for all $s \in (0, \eta) : \varphi'(s) > 0$.

Definition 4.1 *Let $f : \Omega \to (-\infty, +\infty]$ be proper and lower semicontinuous.*

1. *The function f has Kurdyka-Łojasiewicz (KL) property at $w_0 \in \{w \in \Omega : \partial f(w) \neq \emptyset\}$ if there exists $\eta \in (0, +\infty]$, a neighbourhood W_0 of w_0 and a function $\varphi \in \Phi_\eta$ such that for all*

$$\bar{w}_0 \in W \cap \{w \in \Omega : f(w) < f(w_0) < f(w) + \eta\},$$

   the following inequality holds,

$$\varphi'(f(\bar{w}_0) - f(w_0)) \operatorname{dist}(0, \partial f(\bar{w}_0)) \geq 1. \tag{4.11}$$

2. *The function f is a KL function if f satisfies the KL property at each point of $\Omega \cap \{\partial f(w) \neq \emptyset\}$.*

Remark 4.1 *It is important to remark that most convex functions from practical applications satisfy the KL property; see Section 5.1 of [4]. In fact, convex functions that do not satisfy the KL property exist (see [3] for a counterexample) but they are rare and difficult to construct. Indeed, $\mathcal{L}_\gamma$ will be a KL function if each f_i satisfies growth condition, or uniform convexity, or they are general convex semialgebraic or real analytic functions. We refer the interested readers to [1] and [4] for more information.*

The following result, which is called *uniformized KL property*, is from Lemma 6 of [4].

Lemma 4.2 *[Lemma 6 [4]] Let Ω be a compact set and $f : \mathbf{R}^n \to (-\infty, \infty]$ be a proper and lower semi-continuous function. Assume that f is constant on Ω and satisfies the KL property at each point of Ω. Then, there exists $\epsilon > 0$, $\eta > 0$ and $\varphi \in \Phi_\eta$ such that for all $\bar{u}$ in Ω and all u in the intersection:*

$$\{u \in \mathbf{R}^n : \operatorname{dist}(u, \Omega) < \epsilon\} \cap \{u \in \mathbf{R}^n : f(\bar{u}) < f(u) < f(\bar{u}) + \eta\},$$

the following inequality holds,

$$\varphi'\left(f(u)-f(\bar{u})\right)\operatorname{dist}\left(0,\partial f(u)\right)\geq 1.$$

We now give a formal definition of the limit point set. Let the sequence $w^k=\left(x_1^k,\ldots,x_N^k,\lambda^k\right)$ be a sequence generated by the multi-ADMM (1.2) from a starting point $w^0=\left(x_1^0,\ldots,x_N^0,\lambda^0\right)$. The set of all limit points is denoted by $\Omega(w^0)$, i.e.,

$$\Omega(w^0)=\left\{\bar{w}\in\mathbf{R}^{n_1}\times\cdots\times\mathbf{R}^{n_N}\times\mathbf{R}^p:\exists\text{ an infinite sequence }\{k_l\}_{l=1,\ldots}\text{ such that }w^{k_l}\to\bar{w}\text{ as }l\to\infty\right\}.$$

In the following we present the main results in this section. Specifically, Theorem 4.3 gives the convergence of the multi-ADMM (1.2), and we include its proof in the Appendix. Theorem 4.5 shows that the whole sequence generated by the multi-ADMM (1.2) converges.

Theorem 4.3 *Under the conditions in Scenario 2 of Table 1, then:*

1. *$\Omega(w^0)$ is a non-empty set, and any point in $\Omega(w^0)$ is a stationary point of $\mathcal{L}_\gamma(x_1,\ldots,x_N,\lambda)$;*
2. *$\Omega(w^0)$ is a compact and connected set;*
3. *The function $\mathcal{L}_\gamma(x_1,\ldots,x_N,\lambda)$ is finite and constant on $\Omega(w^0)$.*

Remark 4.4 *In Theorem 4.3, we do not require $\mathcal{L}_\gamma$ to be a KL function, which is only required in Theorem 4.5 (see next).*

Theorem 4.5 *Suppose that $\mathcal{L}_\gamma(x_1,\ldots,x_N,\lambda)$ is a KL function. Let the sequence $w^k=\left(x_1^k,\ldots,x_N^k,\lambda^k\right)$ be generated by the multi-block ADMM (1.2). Let $w^*=(x_1^*,\ldots,x_N^*,\lambda^*)\in\Omega(w^0)$, the sequence $w^k=\left(x_1^k,\ldots,x_N^k,\lambda^k\right)$ has a finite length, i.e.,*

$$\sum_{k=0}^{\infty}\left(\sum_{i=1}^{N-1}\|A_ix_i^k-A_ix_i^{k+1}\|+\|x_N^k-x_N^{k+1}\|+\|\lambda^k-\lambda^{k+1}\|\right)\leq G,\tag{4.12}$$

where the constant G is given by

$$G:=2\left(\sum_{i=1}^{N-1}\|A_ix_i^0-A_ix_i^1\|+\|x_N^0-x_N^1\|+\|\lambda^0-\lambda^1\|\right)+\frac{2M\gamma(1+L^2)}{\gamma^2-2L^2}\varphi\left(\mathcal{L}_\gamma(w^1)-\mathcal{L}_\gamma(w^*)\right),$$

and

$$M=\max\left(\gamma\sum_{i=1}^{N-1}\left\|A_i^\top\right\|,\frac{1}{\gamma}+1+\sum_{i=1}^{N-1}\left\|A_i^\top\right\|\right)>0,$$

and the whole sequence $\left(A_1x_1^k,A_2x_2^k,\ldots,A_{N-1}x_{N-1}^k,x_N^k,\lambda^k\right)$ converges to $\left(A_1x_1^*,\ldots,A_{N-1}x_{N-1}^*,x_N^*,\lambda^*\right)$.

Proof. The proof of this theorem is almost identical to the proof of Theorem 1 in [4], by utilizing the uniformized KL property (Lemma 4.2), and the facts that $\Omega(w^0)$ is compact, $\mathcal{L}_\gamma(w)$ is constant (proved in Theorem 4.3), with function Ψ replaced by $\mathcal{L}_\gamma$ and some other minor changes. We thus omit the proof for succinctness. $\square$

Based on Theorem 4.5, we prove a key lemma for analyzing the iteration complexity for the ADMM.

Lemma 4.6 *Let* $(x_1^{k+1}, x_2^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by the multi-ADMM* (4.1)-(4.4) *(or equivalently,* (1.2)*) from given* $(x_2^k, \ldots, x_N^k, \lambda^k)$*. For any* $u^* = (x_1^*, x_2^*, \ldots, x_N^*) \in \Omega^*$ *and* $\lambda \in \mathbf{R}^p$*, it holds in Scenario 2 that*

$$
\begin{aligned}
& f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ \vdots \\ x_N^* - x_N^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \lambda^{k+1} \\ -A_2^\top \lambda^k + 1 \\ \vdots \\ -\lambda^{k+1} \\ \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \end{pmatrix} \\
& + \frac{\gamma}{2}\left(\left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 - \left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 \right) \\
& + \frac{1}{2\gamma}\left(\left\| \lambda - \lambda^k \right\|^2 - \left\| \lambda - \lambda^{k+1} \right\|^2 \right) + \gamma D(N-2)\left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right) \\
\geq \ & 0,
\end{aligned} \tag{4.13}
$$

where D *is a constant.*

Proof. Note that combining (4.9)-(4.10) yields

$$
\begin{aligned}
& \begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ \vdots \\ x_N - x_N^{k+1} \end{pmatrix}^\top \left[\begin{pmatrix} g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} \\ g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} \\ \vdots \\ \nabla f_N(x_N^{k+1}) - \lambda^{k+1} \end{pmatrix} + \begin{pmatrix} \gamma A_1^\top A_2 & \gamma A_1^\top A_3 & \cdots & \gamma A_1^\top \\ 0 & \gamma A_2^\top A_3 & \cdots & \gamma A_2^\top \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma A_{N-1}^\top \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_2^k - x_2^{k+1} \\ \vdots \\ x_N^k - x_N^{k+1} \end{pmatrix} \right] \\
\geq \ & 0,
\end{aligned} \tag{4.14}
$$

where $x_i \in \mathcal{X}_i$ and $g_i \in \partial(f_i + \mathbf{1}_{\mathcal{X}_i})$ is a subgradient of $f_i + \mathbf{1}_{\mathcal{X}_i}$ for $i = 1, 2, \ldots, N-1$.

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

$$
\left(x_i - x_i^{k+1}\right)^\top A_i^\top \left(\sum_{j=i+1}^{N-1} A_j (x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right), \ i = 1, 2, \ldots, N-1.
$$

For the first term, we have (similar to Lemma 3.1)

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

For $i = 2, 3, \ldots, N-1$, we have,

$$
\begin{aligned}
& (x_i - x_i^{k+1})^\top A_i^\top \left[\sum_{j=i+1}^{N-1} A_j (x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right] \\
\leq \; & \left\| A_i x_i - A_i x_i^{k+1} \right\| \left[\sum_{j=i+1}^{N-1} \left\| A_j x_j^k - A_j x_j^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right] \\
\leq \; & \left\| A_i x_i - A_i x_i^{k+1} \right\| \left[\sum_{j=1}^{N-1} \left\| A_j x_j^k - A_j x_j^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ \vdots \\ x_N - x_N^{k+1} \end{pmatrix}^\top \begin{pmatrix} \gamma A_1^\top A_2 & \gamma A_1^\top A_3 & \cdots & \gamma A_1^\top \\ 0 & \gamma A_2^\top A_3 & \cdots & \gamma A_2^\top \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma A_{N-1}^\top \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_2^k - x_2^{k+1} \\ \vdots \\ x_N^k - x_N^{k+1} \end{pmatrix} \\
\leq \; & \frac{\gamma}{2} \left(\left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 - \left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 \right) + \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2 \\
& + \gamma \left(\sum_{i=2}^{N-1} \left\| A_i x_i - A_i x_i^{k+1} \right\| \right) \left[\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right]. \qquad (4.15)
\end{aligned}
$$

Combining (4.4), (4.14) and (4.15), it holds for any $\lambda \in \mathbf{R}^p$ that

$$\begin{aligned}
& \begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ \vdots \\ x_N - x_N^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} \\ g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} \\ \vdots \\ \nabla f_N(x_N^{k+1}) - \lambda^{k+1} \\ \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \end{pmatrix} + \frac{1}{\gamma} \left(\lambda - \lambda^{k+1}\right)^\top \left(\lambda^{k+1} - \lambda^k\right) \\
& + \frac{\gamma}{2} \left(\left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 - \left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 \right) + \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2 \\
& + \gamma \left(\sum_{i=2}^{N-1} \left\| A_i x_i - A_i x_i^{k+1} \right\| \right) \left[\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right] \\
\geq \; & 0. \qquad (4.16)
\end{aligned}$$

Using (2.5), we have

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

Letting $u = u^*$ in (4.16), and invoking the convexity of f_i, we obtain

$$\begin{aligned}
& f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ \vdots \\ x_N^* - x_N^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \lambda^{k+1} \\ -A_2^\top \lambda^{k+1} \\ \vdots \\ -\lambda^{k+1} \\ \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \end{pmatrix} + \frac{1}{2\gamma} \left(\left\| \lambda - \lambda^k \right\|^2 - \left\| \lambda - \lambda^{k+1} \right\|^2 \right) \\
& + \frac{\gamma}{2} \left(\left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 - \left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 \right) \\
& + \gamma \left(\sum_{i=2}^{N-1} \left\| A_i x_i^* - A_i x_i^{k+1} \right\| \right) \left[\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right] \\
\geq \; & 0.
\end{aligned}$$

From Theorem 4.5 we know that the whole sequence $\left(A_1 x_1^k, A_2 x_2^k, \ldots, A_{N-1} x_{N-1}^k, x_N^k, \lambda^k\right)$ converges to $\left(A_1 x_1^*, \ldots, A_{N-1} x_{N-1}^*, x_N^*, \lambda^*\right)$. Therefore, there exists a constant $D > 0$ such that

$$\left\| A_i x_i^k - A_i x_i^{k+1} \right\| \leq D, \qquad (4.17)$$

for any $k \geq 0$ and any $i = 2, 3, \ldots, N-1$. This implies (4.13). □

Now, we are ready to prove the $O(1/\epsilon)$ iteration complexity of the multi-block ADMM for (1.1).

Theorem 4.7 *Let $(x_1^{k+1}, x_2^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1}) \in \Omega$ be generated by ADMM (4.1)-(4.4) from given $(x_2^k, \ldots, x_N^k, \lambda^k)$. For any integer $t > 0$, let $\bar{u}^t = (\bar{x}_1^t, \bar{x}_2^t, \ldots, \bar{x}_N^t)$ and $\bar{\lambda}^t$ be defined as*

$$\bar{x}_i^t = \frac{1}{t+1} \sum_{k=0}^{t} x_i^{k+1}, \; i = 1, 2, \ldots, N, \quad \bar{\lambda}^t = \frac{1}{t+1} \sum_{k=0}^{t} \lambda^{k+1}.$$

For any $(u^*, \lambda^*) \in \Omega^*$, by defining $\rho := \|\lambda^*\| + 1$, it holds in Scenario 2 that,

$$\begin{aligned} 0 &\leq f(\bar{u}^t) - f(u^*) + \rho \left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| \\ &\leq \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \left\| \sum_{i=2}^{N-1} A_i (x_i^0 - x_j^*) + (x_N^0 - x_N^*) \right\|^2 + \frac{\gamma D G}{t+1}. \end{aligned}$$

Note this also implies that when $t = O(1/\epsilon)$, $\bar{u}^t = (\bar{x}_1^t, \bar{x}_2^t, \ldots, \bar{x}_N^t)$ is an ϵ-optimal solution to the original problem (1.1), i.e., both the error of the objective function value and the residual of the equality constraint satisfy that

$$|f(\bar{u}^t) - f(u^*)| = O(\epsilon), \quad \text{and} \quad \left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| = O(\epsilon). \tag{4.18}$$

Proof. Because $(u^k, \lambda^k) \in \Omega$, it holds that $(\bar{u}^t, \bar{\lambda}^t) \in \Omega$ for all $t \geq 0$. By Lemma 4.6 and invoking the convexity of function $f(\cdot)$, we have

$$\begin{aligned} & f(u^*) - f(\bar{u}^t) + \lambda^\top \left(\sum_{i=1}^{N-1} A_i \bar{x}_i^t + \bar{x}_N^t - b \right) \\ =& f(u^*) - f(\bar{u}^t) + \begin{pmatrix} x_1^* - \bar{x}_1^t \\ x_2^* - \bar{x}_2^t \\ \vdots \\ x_N^* - \bar{x}_N^t \\ \lambda - \bar{\lambda}^t \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \bar{\lambda}^t \\ -A_2^\top \bar{\lambda}^t \\ \vdots \\ -\bar{\lambda}^t \\ \sum_{i=1}^{N-1} A_i \bar{x}_i^t + \bar{x}_N^t - b \end{pmatrix} \\ \geq& \frac{1}{t+1} \sum_{k=0}^{t} \left[f(u^*) - f(u^{k+1}) + \begin{pmatrix} x_1^* - x_1^{k+1} \\ x_2^* - x_2^{k+1} \\ \vdots \\ x_N^* - x_N^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \lambda^{k+1} \\ -A_2^\top \lambda^{k+1} \\ \vdots \\ -\lambda^{k+1} \\ \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \end{pmatrix} \right] \\ \geq& \frac{1}{t+1} \sum_{k=0}^{t} \left[\frac{1}{2\gamma} \left(\left\| \lambda - \lambda^{k+1} \right\|^2 - \left\| \lambda - \lambda^k \right\|^2 \right) - \gamma D (N-2) \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right) \right. \\ & \left. + \frac{\gamma}{2} \left(\left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 - \left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 \right) \right] \end{aligned}$$

$$
\begin{aligned}
&\geq -\frac{1}{2\gamma(t+1)}\left\|\lambda-\lambda^0\right\|^2 - \frac{\gamma}{2(t+1)}\left\|A_1x_1^* + \sum_{i=2}^{N-1}A_ix_i^0 + x_N^0 - b\right\|^2 \\
&\quad -\frac{\gamma D(N-2)}{t+1}\sum_{k=0}^{t}\left(\sum_{i=1}^{N-1}\left\|A_ix_i^k - A_ix_i^{k+1}\right\| + \left\|x_N^k - x_N^{k+1}\right\|\right) \\
&\geq -\frac{1}{2\gamma(t+1)}\left\|\lambda-\lambda^0\right\|^2 - \frac{\gamma}{2(t+1)}\left\|A_1x_1^* + \sum_{i=2}^{N-1}A_ix_i^0 + x_N^0 - b\right\|^2 - \frac{\gamma DG(N-2)}{t+1},
\end{aligned}
$$

where the last inequality holds due to Theorem 4.5. Note that this inequality holds for all $\lambda \in \mathbf{R}^p$. From the optimal condition (2.1) we obtain

$$
0 \geq f(u^*) - f(\bar{u}^t) + (\lambda^*)^\top\left(\sum_{i=1}^{N-1}A_i\bar{x}_i^t + \bar{x}_N^t - b\right).
$$

Moreover, since $\rho := \|\lambda^*\| + 1$, $\|\lambda - \lambda_0\|^2 \leq 2(\rho^2 + \|\lambda^0\|^2)$ for all $\|\lambda\| \leq \rho$, and $\sum_{i=1}^{N-1}A_ix_i^* + x_N^* = b$, we obtain

$$
\begin{aligned}
0 &\leq f(\bar{u}^t) - f(u^*) + \rho\left\|\sum_{i=1}^{N-1}A_i\bar{x}_i^t + \bar{x}_N^t - b\right\| \\
&\leq \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)}\left\|\sum_{i=2}^{N-1}A_i(x_i^0 - x_i^*) + (x_N^0 - x_N^*)\right\|^2 + \frac{\gamma DG(N-2)}{t+1}. \qquad (4.19)
\end{aligned}
$$

When $t = O(1/\epsilon)$, we have

$$
\frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)}\left\|\sum_{i=2}^{N-1}A_i(x_i^0 - x_i^*) + (x_N^0 - x_N^*)\right\|^2 + \frac{\gamma DG(N-2)}{t+1} = O(\epsilon). \qquad (4.20)
$$

By the same argument as in the proof for Theorem 3.2, (4.18) follows from (4.20).

□

Acknowledgements

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

References

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the kurdyka-lojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[2] 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.

[3] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010.

[4] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearization minimization for nonconvex and nonsmooth problems. *Math. Program. Ser. A*, 146:459–494, 2014.

[5] 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.

[6] X. Cai, D. Han, and X. Yuan. The direct extension of admm for three-block separable convex minimization models is convergent when one function is strongly convex. *Optimization Online*, 2014.

[7] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming, DOI 10.1007/s10107-014-0826-5*, 2014.

[8] C. Chen, Y. Shen, and Y. You. On the convergence analysis of the alternating direction method of multipliers with three blocks. In *Abstract and Applied Analysis*, volume 2013. Hindawi Publishing Corporation, 2013.

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

[10] 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.

[11] 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.

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

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

[14] 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.

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

[16] 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.

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

[18] 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.

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

[20] 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.

[21] 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.

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

[23] 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.

[24] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo. A block successive upper bound minimization method of multipliers for linearly constrained convex optimization. 2014.

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

[26] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *preprint available at http://arxiv.org/pdf/1410.1390v1.pdf*, 2014.

[27] M. Li, D. Sun, and K-C. Toh. A convergent 3-block semi-proximal admm for convex minimization problems with one strongly convex block. *Asia-Pacific Journal of Operational Research*, 2015.

[28] T. Lin, S. Ma, and S. Zhang. On the convergence rate of multi-block admm. *preprint http://arxiv.org/abs/1408.4265*, 2014.

[29] T. Lin, S. Ma, and S. Zhang. On the global linear convergence of the ADMM with multi-block variables. *submitted*, May 2014.

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

[31] 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.

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

[33] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[34] D. Sun, K.-C. Toh, and L. Yang. A convergent proximal alternating direction method of multipliers for conic programming with 4-block constraints. *SIAM J. Optimization*, 2014.

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

[36] X. Wang, M. Hong, S. Ma, and Z.-Q. Luo. Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. *preprint*, 2013.

A Proof of Theorem 4.3

We first prove a key lemma in the proof of Theorem 4.3.

Lemma A.1 *The following holds in Scenario 2,*

1. *The iterative gap of dual variable can be bounded by that of primal variable, i.e.,*

$$\nabla f_N(x_N^{k+1}) = \lambda^{k+1}, \tag{A.1}$$

and

$$\left\|\lambda^{k+1} - \lambda^k\right\|^2 \le L^2 \left\|x_N^{k+1} - x_N^k\right\|, \tag{A.2}$$

where L satisfies that

$$\|\nabla f_N(x) - \nabla f_N(y)\| \le L \|x - y\|.$$

2. *The augmented Lagrangian L_γ has a sufficient decrease in each iteration, i.e.,*

$$\begin{aligned} & \mathcal{L}_\gamma(x_1^k, \ldots, x_{N+1}^k, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, \ldots, x_{N+1}^{k+1}, \lambda^{k+1}) \\ \ge\ & \frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \left(\sum_{i=1}^{N-1} \left\|A_i x_i^k - A_i x_i^{k+1}\right\|^2 + \left\|x_N^k - x_N^{k+1}\right\|^2 + \left\|\lambda^k - \lambda^{k+1}\right\|^2 \right). \end{aligned} \tag{A.3}$$

3. *The augmented Lagrangian $\mathcal{L}_\gamma(w^k)$ is uniformly lower bounded, and it holds true that*

$$\sum_{k=0}^{\infty} \left(\sum_{i=1}^{N-1} \left\|A_i x_i^{k+1} - A_i x_i^k\right\|^2 + \left\|x_N^{k+1} - x_N^k\right\|^2 + \left\|\lambda^{k+1} - \lambda^k\right\|^2 \right) \le \frac{2\gamma(1+L^2)}{\gamma^2 - 2L^2} \left(\mathcal{L}_\gamma(w^0) - L^*\right) \tag{A.4}$$

where L^* is the uniformly lower bound of $\mathcal{L}_\gamma(w^k)$, and hence

$$\lim_{k\to\infty} \left(\sum_{i=1}^{N-1} \left\|A_i x_i^k - A_i x_i^{k+1}\right\|^2 + \left\|x_N^k - x_N^{k+1}\right\|^2 + \left\|\lambda^k - \lambda^{k+1}\right\|^2 \right) = 0. \tag{A.5}$$

Moreover, $\left\{\left(x_1^k, x_2^k, \ldots, x_N^k, \lambda^k\right) : k = 0, 1, \ldots\right\}$ is a bounded sequence.

4. *There exists a upper bound for a subgradient of augmented Lagrangian $\mathcal{L}_\gamma$ in each iteration. Indeed, we define*

$$R_i^{k+1} = \gamma A_i^\top \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) - \gamma A_i^\top \left(\sum_{j=i+1}^{N-1} A_j (x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right)$$

and

$$R_N^{k+1} = \gamma \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right), \quad R_\lambda^{k+1} = b - \sum_{i=1}^{N-1} A_i x_i^{k+1} - x_N^{k+1}$$

for each positive integer k, and $i = 1, 2, \ldots, N$. Then $\left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1}\right) \in \partial \mathcal{L}_\gamma(w^{k+1})$. Moreover, it holds that

$$\begin{aligned} & \left\| \left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1}\right) \right\| \\ \leq & \sum_{i=1}^{N} \left\| R_i^{k+1} \right\| + \left\| R_\lambda^{k+1} \right\| \\ \leq & M \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_i^k - x_i^{k+1} \right\| + \left\| \lambda^k - \lambda^{k+1} \right\| \right), \quad \forall k \geq 0, \end{aligned} \tag{A.6}$$

where M is a constant defined as

$$M = \max \left(\gamma \sum_{i=1}^{N-1} \left\| A_i^\top \right\|, \frac{1}{\gamma} + 1 + \sum_{i=1}^{N-1} \left\| A_i^\top \right\| \right) > 0. \tag{A.7}$$

Proof of Lemma A.1.

1. (A.1) follows from (4.10) directly. Then we consider the inequality (A.2). It follows from (A.1) and the fact that ∇f_N is Lipschitz continuous with L that

$$\left\| \lambda^{k+1} - \lambda^k \right\|^2 = \left\| \nabla f(x_N^{k+1}) - \nabla f(x_N^k) \right\|^2 \leq L^2 \left\| x_N^{k+1} - x_N^k \right\|^2.$$

2. Multiply both sides of (4.5) by $x_1^k - x_1^{k+1}$, and invoking the convexity of f_1, we have

$$\begin{aligned} 0 = & \left(x_1^k - x_1^{k+1}\right)^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^k + \gamma A_1^\top \left(A_1 x_1^{k+1} + \sum_{j=2}^{N-1} A_j x_j^k + x_N^k - b \right) \right] \\ \leq & f(x_1^k) - f(x_1^{k+1}) - \left(A_1 x_1^k - A_1 x_1^{k+1} \right)^\top \lambda^k \\ & + \gamma \left(A_1 x_1^k - A_1 x_1^{k+1} \right)^\top \left(A_1 x_1^{k+1} + \sum_{j=2}^{N-1} A_j x_j^k + x_N^k - b \right) \end{aligned}$$

$$
\begin{aligned}
&= \left(f(x_1^k) - A_1 x_1^k + \frac{\gamma}{2} \left\| \sum_{j=1}^{N-1} A_j x_j^k + x_N^k - b \right\|^2 \right) - \frac{\gamma}{2} \left\| A_1 x_1^k - A_1 x_1^{k+1} \right\|^2 \\
&\quad - \left(f(x_1^{k+1}) - A_1 x_1^{k+1} + \frac{\gamma}{2} \left\| A_1 x_1^{k+1} + \sum_{j=2}^{N-1} A_j x_j^k + x_N^k - b \right\| \right) \\
&= \mathcal{L}_\gamma(x_1^k, \ldots, x_N^k, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^k, \ldots, x_N^k, \lambda^k) - \frac{\gamma}{2} \left\| A_1 x_1^k - A_1 x_1^{k+1} \right\|^2
\end{aligned} \tag{A.8}
$$

where the second equality holds due to (2.5).

For $i = 2, 3, \ldots, N$, we can derive from (4.6) and (4.7) that

$$
\begin{aligned}
& \mathcal{L}_\gamma \left(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i^k, \ldots, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \ldots, x_i^{k+1}, x_{i+1}^k, \ldots, \lambda^k \right) \\
\geq \;& \frac{\gamma}{2} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 .
\end{aligned} \tag{A.9}
$$

Summing (A.8) and (A.9) over $i = 2, \ldots, N$, we have

$$
\begin{aligned}
& \mathcal{L}_\gamma \left(x_1^k, \ldots, x_N^k, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^k \right) \\
\geq \;& \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \frac{\gamma}{2} \left\| x_N^k - x_N^{k+1} \right\|^2 .
\end{aligned} \tag{A.10}
$$

On the other hand, it follows from (A.1) that

$$
\begin{aligned}
& \mathcal{L}_\gamma \left(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1} \right) \\
= \;& \frac{1}{\gamma} \left\| \lambda^k - \lambda^{k+1} \right\|^2 \geq -\frac{L^2}{\gamma} \left\| x_N^k - x_N^{k+1} \right\|^2 .
\end{aligned} \tag{A.11}
$$

Combining (A.10) and (A.11) yields

$$
\begin{aligned}
& \mathcal{L}_\gamma \left(x_1^k, \ldots, x_N^k, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1} \right) \\
\geq \;& \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \frac{\gamma^2 - 2L^2}{2\gamma} \left\| x_N^k - x_N^{k+1} \right\|^2 \\
= \;& \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \left\| x_N^k - x_N^{k+1} \right\|^2 + \frac{L^2(\gamma^2 - 2L^2)}{2\gamma(1+L^2)} \left\| x_N^k - x_N^{k+1} \right\|^2 \\
\geq \;& \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \left(\left\| x_N^k - x_N^{k+1} \right\|^2 + \left\| \lambda^k - \lambda^{k+1} \right\|^2 \right) \\
\geq \;& \frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \left\| x_N^k - x_N^{k+1} \right\|^2 + \left\| \lambda^k - \lambda^{k+1} \right\|^2 \right)
\end{aligned} \tag{A.12}
$$

where the last inequality holds due to the fact that

$$
\frac{\gamma}{2} \geq \frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)}.
$$

3. Note that

$$\begin{aligned}
&\mathcal{L}_\gamma\left(x_1^{k+1},\dots,x_N^{k+1},\lambda^{k+1}\right)\\
=\;& \sum_{i=1}^{N-1} f_i(x_i^{k+1}) + f_N(x_N^{k+1}) - \left\langle \lambda^{k+1}, \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\rangle + \frac{\gamma}{2}\left\| \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2.
\end{aligned}$$

It follows from (A.1) and the fact that ∇f_N is Lipschitz continuous with constant L that,

$$\begin{aligned}
& f_N\left(b - \sum_{i=1}^{N-1} A_i x_i^{k+1}\right)\\
\leq\;& f_N(x_N^{k+1}) + \left\langle \nabla f_N(x_N^{k+1}), \left(b - \sum_{i=1}^{N-1} A_i x_i^{k+1} - x_N^{k+1}\right)\right\rangle + \frac{L}{2}\left\| b - \sum_{i=1}^{N-1} A_i x_i^{k+1} - x_N^{k+1}\right\|^2\\
=\;& f_N(x_N^{k+1}) - \left\langle \nabla f_N(x_N^{k+1}), \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right\rangle + \frac{L}{2}\left\| \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right\|^2\\
=\;& f_N(x_N^{k+1}) - \left\langle \lambda^{k+1}, \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right\rangle + \frac{L}{2}\left\| \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right\|^2.
\end{aligned}$$

This implies that there exists $L^* > -\infty$, such that

$$\begin{aligned}
& \mathcal{L}_\gamma(x_1^{k+1},\dots,x_N^{k+1},\lambda^{k+1})\\
\geq\;& \sum_{i=1}^{N-1} f_i(x_i^{k+1}) + f_N\left(b - \sum_{i=1}^{N-1} A_i x_i^{k+1}\right) + \frac{\gamma - L}{2}\left\| \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right\|^2\\
>\;& L^*, \qquad\qquad (\text{A.13})
\end{aligned}$$

where the last inequality holds since $\gamma > L$ and $\inf_{\mathcal{X}_i} f_i > f_i^*$ for $i = 1, 2, \dots, N$.
Therefore, it directly follows from (A.3) and $\gamma > \sqrt{2}L$ that,

$$\frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \sum_{k=0}^{K} \left(\sum_{i=1}^{N-1} \|A_i x_i^{k+1} - A_i x_i^k\|^2 + \|x_N^{k+1} - x_N^k\|^2 + \|\lambda^{k+1} - \lambda^k\|^2 \right) \leq \mathcal{L}_\gamma(w^0) - L^*.$$

Letting $K \to \infty$, we have

$$\frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \sum_{k=0}^{\infty} \left(\sum_{i=1}^{N-1} \|A_i x_i^{k+1} - A_i x_i^k\|^2 + \|x_N^{k+1} - x_N^k\|^2 + \|\lambda^{k+1} - \lambda^k\|^2 \right) \leq \mathcal{L}_\gamma(w^0) - L^*,$$

which implies (A.4) and (A.5).

It also follows from (A.13), (A.3) and $\gamma > \sqrt{2}L$ that $\mathcal{L}_\gamma(w^0) - f_N^* \geq \sum_{i=1}^{N-1} f_i(x_i^{k+1})$. This implies that $\left\{\left(x_1^k, x_2^k, \dots, x_{N-1}^k\right) : k = 0, 1, \dots\right\}$ is a bounded sequence by using the coerciveness of $f_i + \mathbf{1}_{\mathcal{X}_i}, i = 1, 2, \dots, N-1$. The boundedness of $\left(x_N^k, \lambda^k\right)$ can be obtained by using (4.4), (A.2) and (A.5).

4. From the definition of $\mathcal{L}_\gamma$, it is clear that for $i = 1, \ldots, N-1$,

$$g_i(x_i^{k+1}) - A_i^\top \lambda^{k+1} + \gamma A_i^\top \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) \in \partial_{x_i} \mathcal{L}_\gamma(w^{k+1}),$$

and

$$\nabla f(x_N^{k+1}) - \lambda^{k+1} + \gamma \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) = \nabla_{x_N} \mathcal{L}_\gamma(w^{k+1}),$$

and

$$b - \sum_{i=1}^{N-1} A_i x_i^{k+1} - x_N^{k+1} = \nabla_\lambda \mathcal{L}_\gamma(w^{k+1}),$$

where $g_i \in \partial \left(f_i + \mathbf{1}_{\mathcal{X}_i} \right)$ for $i = 1, 2, \ldots, N-1$. Since (4.8), (4.9), and (4.10) imply that

$$g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} = -\gamma A_1^\top \left(\sum_{j=2}^{N-1} A_j (x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right), \tag{A.14}$$

$$g_i(x_i^{k+1}) - A_i^\top \lambda^{k+1} = -\gamma A_i^\top \left(\sum_{j=i+1}^{N-1} A_j (x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right), \tag{A.15}$$

$$\nabla f_N(x_N^{k+1}) - \lambda^{k+1} = 0, \tag{A.16}$$

we have

$$R_i^{k+1} = \gamma A_i^\top \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) - \gamma A_i^\top \left(\sum_{j=i+1}^{N-1} A_j (x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right) \in \partial_{x_i} \mathcal{L}_\gamma(w^{k+1}),$$

$$R_N^{k+1} = \gamma \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) = \nabla_{x_N} \mathcal{L}_\gamma(w^{k+1}),$$

$$R_\lambda^{k+1} = b - \sum_{i=1}^{N-1} A_i x_i^{k+1} - x_N^{k+1} = \nabla_\lambda \mathcal{L}_\gamma(w^{k+1}),$$

for $i = 1, 2, \ldots, N-1$. This implies that $\left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1} \right) \in \partial \mathcal{L}_\gamma(w^{k+1})$.

We now need to estimate the norms of $R_i^{k+1}, 1 \le i \le N-1$ and R_N^k and R_λ^k. It holds true that,

$$\begin{aligned} \left\| R_i^{k+1} \right\| &\le \gamma \left\| A_i^\top \right\| \left(\sum_{j=i+1}^{N-1} \left\| A_j x_j^k - A_j x_j^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right) + \gamma \left\| A_i^\top \right\| \left\| \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\| \\ &\le \gamma \left\| A_i^\top \right\| \left(\sum_{j=1}^{N-1} \left\| A_j x_j^k - A_j x_j^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right) + \left\| A_i^\top \right\| \left\| \lambda^k - \lambda^{k+1} \right\| \end{aligned}$$

and

$$\left\| R_N^{k+1} \right\| \le \gamma \left\| \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\| = \left\| \lambda^k - \lambda^{k+1} \right\|,$$

and

$$\left\|R_\lambda^{k+1}\right\| \quad = \quad \left\|\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right\| = \frac{1}{\gamma}\left\|\lambda^k - \lambda^{k+1}\right\|.$$

Therefore, we arrive at (A.6) where M is defined in (A.7).

□

Proof of Theorem 4.3.

1. It has been proven in Lemma A.1 that $\left\{\left(x_1^k, x_2^k, \ldots, x_N^k, \lambda^k\right) : k = 0, 1, \ldots\right\}$ is a bounded sequence. Therefore, we conclude that $\Omega(w^0)$ is non-empty by the Bolzano-Weierstrass Theorem. Let $w^* = (x_1^*, \ldots, x_N^*, \lambda^*) \in \Omega(w^0)$ be a limit point of $\{w^k = \left(x_1^k, \ldots, x_N^k, \lambda^k\right) : k = 0, 1, \ldots\}$. Then there exists a subsequence $\left\{w^{k_q} = \left(x_1^{k_q}, \ldots, x_N^{k_q}, \lambda^{k_q}\right) : q = 0, 1, \ldots\right\}$ such that $w^{k_q} \to w^*$ as $q \to \infty$. Since $f_i, i = 1, \ldots, N-1$, are lower semi-continuous, we obtain that

$$\liminf_{q\to\infty} f_i(x_i^{k_q}) \geq f_i(x_i^*), \quad i = 1, 2, \ldots, N. \tag{A.17}$$

From the iterative step (4.1)-(4.4), we have for any integer k and any $i = 1, \ldots, N-1$,

$$x_i^{k+1} := \underset{x_i \in \mathcal{X}_i}{\operatorname{argmin}} \ \mathcal{L}_\gamma(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \ldots, x_N^k; \lambda^k).$$

Letting $x_i = x_i^*$ in the above, we get

$$\mathcal{L}_\gamma(x_1^{k+1}, \ldots, x_i^{k+1}, x_{i+1}^k, \ldots, x_N^k; \lambda^k) \leq \mathcal{L}_\gamma(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i^*, x_{i+1}^k, \ldots, x_N^k; \lambda^k),$$

i.e.,

$$\begin{aligned} & f_i(x_i^{k+1}) - \left\langle \lambda^k, A_i x_i^{k+1} \right\rangle + \frac{\gamma}{2}\left\|\sum_{j=1}^{i} A_j x_j^{k+1} + \sum_{j=i+1}^{N-1} A_j x_j^k + x_N^k - b\right\|^2 \\ \leq \quad & f_i(x_i^*) - \left\langle \lambda^k, A_i x_i^* \right\rangle + \frac{\gamma}{2}\left\|\sum_{j=1}^{i-1} A_j x_j^{k+1} + A_i x_i^* + \sum_{j=i+1}^{N-1} A_j x_j^k + x_N^k - b\right\|^2. \end{aligned}$$

Choosing $k = k_q - 1$ in the above inequality and letting q go to $+\infty$, we obtain

$$\limsup_{q\to+\infty} f_i(x_i^{k_q}) \leq \limsup_{q\to+\infty}\left(\frac{\gamma}{2}\left\|A_i x_i^{k_q} - A_i x_i^*\right\|^2 - \left\langle \lambda^k, A_i x_i^{k_q} - A_i x_i^* \right\rangle\right) + f_i(x_i^*), \tag{A.18}$$

for $i = 1, 2, \ldots, N-1$. Here we have used the facts that both the sequence $\{w^k : k = 0, 1, \ldots\}$ is bounded, and γ is finite, and that the distance between two successive iterates tends to zero (A.5), and the fact that

$$\sum_{j=1}^{i} A_j x_j^{k+1} + \sum_{j=i+1}^{N-1} A_j x_j^k + x_N^k - b = \sum_{j=i+1}^{N-1}\left(A_j x_j^k - A_j x_j^{k+1}\right) + \left(x_N^k - x_N^{k+1}\right) + \frac{1}{\gamma}(\lambda^k - \lambda^{k+1}).$$

From (A.5) we also have $x_i^{k_q-1} \to x_i^*$ as $q \to \infty$, hence (A.18) reduces to

$$\limsup_{q\to\infty} f_i(x_i^{k_q}) \le f_i(x_i^*).$$

Therefore, combining with (A.17), $f_i(x_i^{k_q})$ tends to $f_i(x_i^*)$ as $q \to \infty$. Therefore, we can conclude that

$$\begin{aligned}\lim_{q\to\infty} \mathcal{L}_\gamma(w^{k_q}) &= \lim_{q\to\infty} \left(\sum_{i=1}^{N} f_i(x_i^{k_q}) - \left\langle \lambda^{k_q}, \sum_{i=1}^{N-1} A_i x_i^{k_q} + x_N^{k_q} - b \right\rangle + \frac{\gamma}{2} \left\| \sum_{i=1}^{N-1} A_i x_i^{k_q} + x_N^{k_q} - b \right\|^2 \right) \\ &= \sum_{i=1}^{N} f_i(x_i^*) - \left\langle \lambda^*, \sum_{i=1}^{N-1} A_i x_i^* + x_N^* - b \right\rangle + \frac{\gamma}{2} \left\| \sum_{i=1}^{N-1} A_i x_i^* + x_N^* - b \right\|^2 \\ &= \mathcal{L}_\gamma(w^*).\end{aligned}$$

On the other hand, it follows from (A.5) and (A.6) that

$$\left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1}\right) \in \partial \mathcal{L}_\gamma(w^{k+1}) \tag{A.19}$$

$$\left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1}\right) \to (0, \ldots, 0), \quad k \to \infty. \tag{A.20}$$

It implies that $(0, \ldots, 0) \in \partial \mathcal{L}_\gamma(x_1^*, \ldots, x_N^*, \lambda^*)$ due to the closeness of $\partial \mathcal{L}_\gamma$. Therefore, $w^* = (x_1^*, \ldots, x_N^*, \lambda^*)$ is a critical point of $\mathcal{L}_\gamma(x_1, \ldots, x_N, \lambda)$.

2. The proof for this assertion directly follows from Lemma 5 and Remark 5 of [4]. We omit the proof here for succinctness.

3. We define that L^* is the finite limit of $\mathcal{L}_\gamma(x_1^k, \ldots, x_N^k, \lambda^k)$ as k goes to infinity, i.e.,

$$L^* = \lim_{k\to\infty} \mathcal{L}_\gamma(x_1^k, \ldots, x_N^k, \lambda^k).$$

Take $w^* \in \Omega(w^0)$. There exists a subsequence w^{k_q} converging to w^* as q goes to infinity. Since we have proven that

$$\lim_{q\to\infty} \mathcal{L}_\gamma(w^{k_q}) = \mathcal{L}_\gamma(w^*),$$

and $\mathcal{L}_\gamma(w^k)$ is a non-increasing sequence, we conclude that $\mathcal{L}_\gamma(w^*) = L^*$, hence the restriction of $\mathcal{L}_\gamma(x_1, \ldots, x_N, \lambda)$ to $\Omega(w^0)$ equals L^*.