CrossMark

Iteration Complexity Analysis of Multi-block ADMM for a Family of Convex Minimization Without Strong Convexity

Tianyi Lin[1] · Shiqian Ma[1] · Shuzhong Zhang[2]

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 Chen et al. (Math Program 155(1):57–79, 2016.) 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 provide two different ways related to multi-block ADMM that can find an ϵ-optimal solution and do not require strong convexity of the objective function. 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; this case can be seen as using multi-block ADMM to solve a modified 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; this case can be seen as applying multi-block ADMM to solving a special class of problems.

✉ Shiqian Ma
sqma@se.cuhk.edu.hk

[1] Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China

[2] Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, USA

1 Introduction

We consider the following multi-block convex minimization problem:

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

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. Problem (1.1) is a general convex minimization problem with linear equality constraints, and there exist many algorithms for solving this problem. For example, the augmented Lagrangian method can be used to solve it. In this paper, we consider the case that the functions $f_i, i = 1, \ldots, N$ have certain structures so that it is beneficial to handle them separately, rather than treat them as a whole. One effective way to solve (1.1) that can take advantage of its separable structure is the so-called Alternating Direction Method of Multipliers (ADMM). The ADMM is closely related to the Douglas–Rachford [12] and Peaceman–Rachford [34] operator splitting methods that date back to 1950s. These operator splitting methods were further studied later in [13,16,18,32]. 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 [5,15] 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:

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

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 [14,17]. There are also some recent works that study the convergence rate properties of ADMM when $N = 2$ (see, e.g., [2,11,23,24,33]).

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 lack of 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 [35,37]. 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 [19] showed that the multi-block 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 [8,31] to allow only $N - 1$ functions to be strongly convex and γ is restricted to a certain region. Especially, Lin et al. [31] proved the sublinear convergence rate under such conditions. Cai et al. [6] and Li et al. [28] 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 to be in a 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 et al. [30] proved the globally linear convergence of multi-block ADMM. Davis and Yin [9] studied a variant of the 3-block ADMM (see Algorithm 8 in [9]) which requires that f_1 is strongly convex and γ is smaller than a certain bound to guarantee the convergence. More recently, Lin, Ma and Zhang [29] proved that the 3-block ADMM is globally convergent for any $\gamma > 0$ when it is applied to solve a class of regularized least squares decomposition problems. Note that the above mentioned works all require that (parts of) the objective function is strongly convex. Without assuming strong convexity, Hong and Luo [26] studied a variant of ADMM (1.2) with small stepsize in updating the Lagrange multiplier. Specifically, [26] proposes to replace the last equation in (1.2) to $\lambda^{k+1} := \lambda^k - \alpha\gamma(\sum_{j=1}^{N} A_j x_j^{k+1} - b)$, 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 [10,20–22,25,36,38] 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. The authors of [36] made the following statement (see page 885 in [36]): "However, to the best of our knowledge, up to now the dilemma is that at least for convex conic programming, the modified versions, though with a convergence guarantee, often perform two to three times slower than the multiblock ADMM with no convergent guarantee." Wang et al. [38] compared the multi-block ADMM with several modifications of it, and the numerical experiments on basis pursuit and latent variable Gaussian graphical model selection problems show that the multi-block ADMM always outperforms its modifications tested (see Section 3 of [38]). These results motivated us to further study whether ADMM can be guaranteed to converge when it is applied to solve certain class of problems, although it is not necessarily convergent in general.

Our contributions. The main contributions of this paper are as follows. First, we show that if ADMM (1.2) is applied to solve an associated perturbed problem of (1.1) rather than (1.1) itself, then it returns an ϵ-optimal solution to (1.1) 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. Note that our result in this part can be seen as a perturbation approach. Instead of modifying the ADMM procedure itself as most papers in the literature propose, we propose

to modify the problem. Such approach is not uncommon. The perturbation technique applied to resolve the cycling problem in the simplex method for linear programming is one famous example of this type. Modifying the algorithm sometimes requires additional computational efforts. Moreover, as shown in [36] and [38], modified versions of ADMM perform worse than the original ADMM for some problems. Our approach maintains the effectiveness of ADMM, at least for certain problems discussed in [36] and [38], because we do not modify the algorithm itself. Secondly, we show that the ADMM (1.2) returns an ϵ-optimal solution within $O(1/\epsilon)$ iterations when it is applied to solve the so-called *sharing* problems, under the condition that the augmented Lagrangian $\mathcal{L}_\gamma$ is a Kurdyka–Łojasiewicz (KL) function [3,4], and γ is sufficiently large. Interestingly, it turns out that ADMM possesses nice convergence properties when it is applied to solve sharing problems. In a recent work by Hong et al. [27], global convergence of ADMM was established even when it is applied to solve a nonconvex sharing problem. In this paper, we show that by combining the KL property and the convexity of the problem, we can establish a sublinear convergence rate of ADMM for solving sharing problem.

Organization. The rest of this paper is organized as follows. In Sect. 2 we provide some preliminaries for our convergence rate analysis. In Sect. 3, we prove the $O(1/\epsilon^2)$ iteration complexity of ADMM (1.2) by introducing an associated problem of (1.1). In Sect. 4, we prove the $O(1/\epsilon)$ iteration complexity of ADMM (1.2) with Kurdyka–Łojasiewicz (KL) property.

2 Preliminaries

We denote $\Omega = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \mathbf{R}^p$ 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:

$$\begin{cases} (x_i - x_i^*)^\top (g_i(x_i^*) - A_i^\top \lambda^*) \geq 0, & \forall x_i \in \mathcal{X}_i, \quad i = 1, 2, \ldots, N \\ A_1 x_1^* + \cdots + A_N x_N^* - b = 0. \end{cases} \tag{2.1}$$

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, we assume 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).

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

Table 1 Two scenarios leading to sublinear convergence

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	$O(1/\epsilon)$

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:

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

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., [5,27]). (2.2) can be reformulated as

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

which is in the form of (1.1) and can be solved by ADMM (see [5,27]). 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.

Notation For simplicity, we use u to denote the stacked vectors $x_1, \ldots, x_N$, i.e., $u = [x_1; x_2; \ldots; x_N]$, $u^k = [x_1^k; x_2^k; \ldots; x_N^k]$, $u^* = [x_1^*, x_2^*; \ldots; x_N^*]$ and denote $w = [u; \lambda]$. 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,

$$\begin{aligned} (w_1 - w_2)^\top (w_3 - w_4) &= \frac{1}{2} \left(\|w_1 - w_4\|^2 - \|w_1 - w_3\|^2 \right) \\ &\quad + \frac{1}{2} \left(\|w_3 - w_2\|^2 - \|w_4 - w_2\|^2 \right) \end{aligned} \qquad (2.4)$$

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

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 $u^0 := (x_1^0; \ldots; x_N^0)$, we introduce an associated perturbed problem of (1.1), i.e.,

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

where $\tilde{f}_i(x_i) = f_i(x_i) + \frac{\mu}{2} \|A_i x_i - A_i x_i^0\|^2$ 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):

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

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

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

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

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

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

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

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

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

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* (3.2)–(3.4) *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 the conditions in Scenario 1 that*

$$f(u^*) - f\left(u^{k+1} \right) + \sum_{i=1}^N \left(x_i^* - x_i^{k+1} \right)^\top \left(-A_i^\top \lambda^{k+1} \right) + (\lambda - \lambda^{k+1})^\top \left(\sum_{i=1}^N A_i x_i^{k+1} - b \right)$$
$$+ \frac{1}{2\gamma} \left(\left\| \lambda - \lambda^k \right\|^2 - \left\| \lambda - \lambda^{k+1} \right\|^2 \right) + \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^N \left\| A_i x_i^* - A_i x_i^0 \right\|^2$$

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

$$\geq 0. \tag{3.9}$$

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

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

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

$$\left. - \left(- \sum_{j=i+1}^{N} A_j x_j^k \right) \right]$$

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

$$+ \frac{1}{2} \left(\left\| \sum_{j=1}^{i-1} A_j x_j + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2 - \left\| \sum_{j=1}^{i-1} A_j x_j + A_i x_i^{k+1} + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 \right)$$

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

$$+ \frac{1}{2} \left\| \sum_{j=1}^{i-1} A_j x_j + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2,$$

where in the second equality we applied the identity (2.4). Summing the above inequality for $i = 1, \ldots, N - 1$ and using (3.4) yield,

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

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

$$+ \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2 + \frac{\gamma}{2} \sum_{i=2}^{N-1} \left\| \sum_{j=1}^{i-1} A_j x_j + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2. \tag{3.10}$$

Adding (3.7) and (3.8), and combining the resulting inequality with (3.4) and (3.10) yields that the following inequality holds for any $\lambda \in \mathbf{R}^p$:

$$
\sum_{i=1}^{N} \left(x_i - x_i^{k+1} \right)^\top \left(g_i(x_i^{k+1}) - A_i^\top \lambda^{k+1} \right) + (\lambda - \lambda^{k+1})^\top \left(\sum_{i=1}^{N} A_i x_i^{k+1} - b \right)
$$
$$
+ \frac{1}{\gamma} (\lambda - \lambda^{k+1})^\top (\lambda^{k+1} - \lambda^k)
$$
$$
+ \mu \sum_{i=2}^{N} \left(x_i - x_i^{k+1} \right)^\top A_i^\top A_i \left(x_i^{k+1} - x_i^0 \right) + \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2
$$
$$
+ \frac{\gamma}{2} \sum_{i=2}^{N-1} \left\| \sum_{j=1}^{i-1} A_j x_j + \sum_{j=i}^{N} A_j x_j^{k+1} - b \right\|^2
$$
$$
+ \frac{\gamma}{2} \sum_{i=1}^{N-1} \left(\left\| \sum_{j=1}^{i} A_j x_j + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 \right)
$$
$$
\geq 0. \tag{3.11}
$$

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

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

Letting $u = u^*$ in (3.11), and invoking the convexity of f_i that

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

and

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

we obtain,

$$
f(u^*) - f(u^{k+1}) + \sum_{i=1}^{N} \left(x_i^* - x_i^{k+1} \right)^\top \left(-A_i^\top \lambda^{k+1} \right) + (\lambda - \lambda^{k+1})^\top \left(\sum_{i=1}^{N} A_i x_i^{k+1} - b \right)
$$

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

$$
+ \frac{\mu}{2} \sum_{i=2}^{N} \left(\left\| A_i x_i^* - A_i x_i^0 \right\|^2 - \left\| A_i x_i^* - A_i x_i^{k+1} \right\|^2 \right)
$$

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

$$
+ \frac{\gamma(N+1)(N-2)}{2} \sum_{i=2}^{N} \left\| A_i x_i^* - A_i x_i^{k+1} \right\|^2
$$

$$
\geq 0.
$$

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

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

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*

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

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

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

$$
\leq \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^0 - x_j^*) \right\|^2
$$

$$
+ \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^{N} \left\| A_i x_i^* - A_i x_i^0 \right\|^2.
$$

This also implies that when $t = O(1/\epsilon^2)$, $\bar{u}^t$ 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

$$|f(\bar{u}^t) - f(u^*)| = O(\epsilon), \quad and \quad \left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| = O(\epsilon). \tag{3.12}$$

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

$$
\begin{aligned}
& f(u^*) - f(\bar{u}^t) + \lambda^\top \left(\sum_{i=1}^{N} A_i \bar{x}_i^t - b \right) \\
&= f(u^*) - f(\bar{u}^t) + \sum_{i=1}^{N} \left(x_i^* - \bar{x}_i^t \right)^\top \left(-A_i^\top \bar{\lambda}^t \right) + (\lambda - \bar{\lambda}^t)^\top \left(\sum_{i=1}^{N} A_i \bar{x}_i^t - b \right) \\
&\geq \frac{1}{t+1} \sum_{k=0}^{t} \left[f(u^*) - f(u^{k+1}) \right. \\
&\quad \left. + \sum_{i=1}^{N} \left(x_i^* - x_i^{k+1} \right)^\top \left(-A_i^\top \lambda^{k+1} \right) + (\lambda - \lambda^{k+1})^\top \left(\sum_{i=1}^{N} A_i x_i^{k+1} - b \right) \right] \\
&\geq \frac{1}{t+1} \sum_{k=0}^{t} \left[\frac{1}{2\gamma} \left(\left\| \lambda - \lambda^{k+1} \right\|^2 - \left\| \lambda - \lambda^k \right\|^2 \right) \right. \\
&\quad - \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^{N} \left\| A_i x_i^* - A_i x_i^0 \right\|^2 \\
&\quad \left. + \frac{\gamma}{2} \sum_{i=1}^{N-1} \left(\left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^{k+1} - b \right\|^2 - \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^k - b \right\|^2 \right) \right] \\
&\geq -\frac{1}{2\gamma(t+1)} \left\| \lambda - \lambda^0 \right\|^2 - \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=1}^{i} A_j x_j^* + \sum_{j=i+1}^{N} A_j x_j^0 - b \right\|^2 \\
&\quad - \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^{N} \left\| A_i x_i^* - A_i x_i^0 \right\|^2. \tag{3.13}
\end{aligned}
$$

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

$$0 \leq f(\bar{u}^t) - f(u^*) + \rho \left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\|. \tag{3.14}$$

By setting $\lambda = -\rho(\sum_{i=1}^{N} A_i \bar{x}_i^t - b)/\|\sum_{i=1}^{N} A_i \bar{x}_i^t - b\|$ in (3.13), and noting that $\|\lambda\| = \rho$, we obtain

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

$$\leq \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^0 - x_j^*) \right\|^2$$

$$+ \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^{N} \left\| A_i x_i^* - A_i x_i^0 \right\|^2. \tag{3.15}$$

When $t = O(1/\epsilon^2)$, together with the condition that $\frac{\epsilon}{2} \leq \gamma \leq \epsilon$, it is implied that the right hand side of (3.15) is of order $O(\epsilon)$.

We now define the function

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

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 and the Cauchy–Schwarz inequality, it holds that

$$v(\xi) \geq v(0) + \langle \lambda^*, \xi \rangle \geq f(u^*) - \|\lambda^*\|\|\xi\|. \tag{3.16}$$

Let $\bar{\xi} = \sum_{i=1}^{N} A_i \bar{x}_i^t - b$, we have $f(\bar{u}^t) \geq v(\bar{\xi})$. Therefore, by combining (3.14), (3.16), and the fact that the right hand side (thus the left hand side) of (3.15) is of order $O(\epsilon)$, we get

$$-\|\lambda^*\|\|\bar{\xi}\| \leq f(\bar{u}^t) - f(u^*)$$

$$\leq \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \sum_{i=1}^{N-1} \left\| \sum_{j=i+1}^{N} A_j(x_j^0 - x_j^*) \right\|^2$$

$$+ \frac{\epsilon(N-2)(N+1)}{2} \sum_{i=2}^{N} \left\| A_i x_i^* - A_i x_i^0 \right\|^2 - \rho\|\bar{\xi}\|$$

$$\leq C\epsilon - \rho\|\bar{\xi}\|,$$

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

$$\left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| = \|\bar{\xi}\| \leq C\epsilon, \tag{3.17}$$

where C is a constant depending on $\rho, \lambda^0, x^0, x^*$ and N. Moreover, by combining (3.14) and (3.17), one obtains that

$$-\rho C\epsilon \leq -\rho\|\bar{\xi}\| \leq f(\bar{u}^t) - f(u^*) \leq (1-\rho)C\epsilon. \tag{3.18}$$

Finally, we note that (3.17), (3.18) imply (3.12). $\qquad\square$

Remark 3.3 We remark that by applying ADMM to solve the associated perturbed problem (3.1), we can find an ϵ-optimal solution to the original problem (1.1), i.e., an approximate solution to (1.1). However, the assumptions required in this section do not ensure that the

exact optimal solution to (1.1) can be found, because that would require $\gamma \to 0$ and thus $\epsilon \to 0$.

3.1 Solving the Counterexample in [7]

We now implement our algorithm in this section to solve a perturbed version of the counterexample constructed by Chen et al. [7]. As we will see, after modifying the problem by adding the perturbation terms, the multi-block ADMM is always able to find an ϵ-optimal solution. The counterexample constructed in [7] is to find a solution to the linear system $A_1 x_1 + A_2 x_2 + A_3 x_3 = 0$, where $A_1 = (1, 1, 1)^\top$, $A_2 = (1, 1, 2)^\top$ and $A_3 = (1, 2, 2)^\top$. It is easy to see that the unique solution of this linear system is $(x_1, x_2, x_3) = (0, 0, 0)$. The associated perturbed problem to this example is given by

$$\begin{aligned} \min \quad & \frac{\mu}{2}\|A_2 x_2 - A_2 x_2^0\|^2 + \frac{\mu}{2}\|A_3 x_3 - A_3 x_3^0\|^2 \\ \text{s.t.} \quad & A_1 x_1 + A_2 x_2 + A_3 x_3 = 0, \end{aligned} \tag{3.19}$$

where $\mu = 4\epsilon$. To satisfy the conditions in Scenario 1, we chose $\gamma = \epsilon$. We randomly generated $x^0 = (x_1^0, x_2^0, x_3^0)$ using the *randn* function in Matlab with random seed set to 0. We plot the residual of the equality constraint and the three components of x^t and the average of x^k for $k = 0, \ldots, t$ in Figs. 1 and 2 for $\epsilon = 10^{-2}$ and $\epsilon = 10^{-3}$ respectively, where t denotes the iteration number. From Figs. 1 and 2 we can see that ADMM for solving the perturbed problem is able to decrease the residual of the equality constraint to the order of $O(\epsilon)$ in $O(1/\epsilon^2)$ iterations, which matches our theoretical results in Theorem 3.2. We also see that both x^t and the average of x^k for $k = 0, \ldots, t$ converge to zero, which is the unique solution to the original problem. As a result, our approach successfully eliminates the risk of divergence of the multi-block ADMM.

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. Under the conditions in Scenario 2, since $\mathcal{X}_N = \mathbf{R}^{n_N}$, f_N is differentiable and $A_N = I$, (1.1) becomes the so-called sharing problem. It turns out that ADMM has nice convergence properties when it is applied to solve sharing problems. In a recent work by Hong, Luo and Razaviyayn [27], it was shown that ADMM is globally convergent under certain conditions for solving a nonconvex sharing problem. For convex sharing problem, Corollary 3.1 in [27] gives conditions that guarantee the global convergence of ADMM. These conditions are the same as our conditions required in Scenario 2. For example, both papers require $\gamma > \sqrt{2}L$. The difference is that global convergence of ADMM for solving (both nonconvex and convex) sharing problem is shown in [27], and we show that by using the KL property, we are able to prove the sublinear convergence rate of ADMM for convex sharing problem.

The first-order optimality conditions for the subproblems in (1.2) are given by $x_i^{k+1} \in \mathcal{X}_i$, $i = 1, \ldots, N-1$, and for any $x_i \in \mathcal{X}_i$,

$$g_i\left(x_i^{k+1}\right) - A_i^\top \lambda^k + \gamma A_i^\top \left(\sum_{j=1}^{i} A_j x_j^{k+1} + \sum_{j=i+1}^{N-1} A_j x_j^k + x_N^k - b\right) = 0, \quad i = 1, \ldots, N-1, \tag{4.1}$$
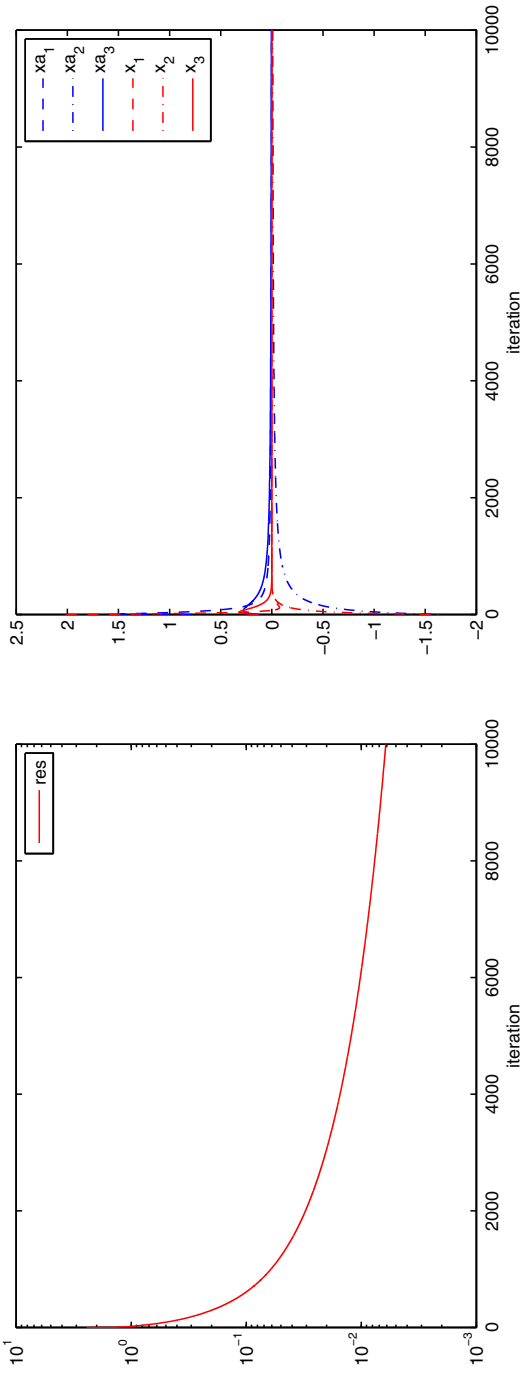
Fig. 1 ADMM for solving the associated perturbed problem (3.19) with $\epsilon = 10^{-2}$. Left: the residual of the equality constraint; Right: x^t and the average of x^k for $k = 0, \ldots, t$ (denoted as xa) as defined in Theorem 3.2
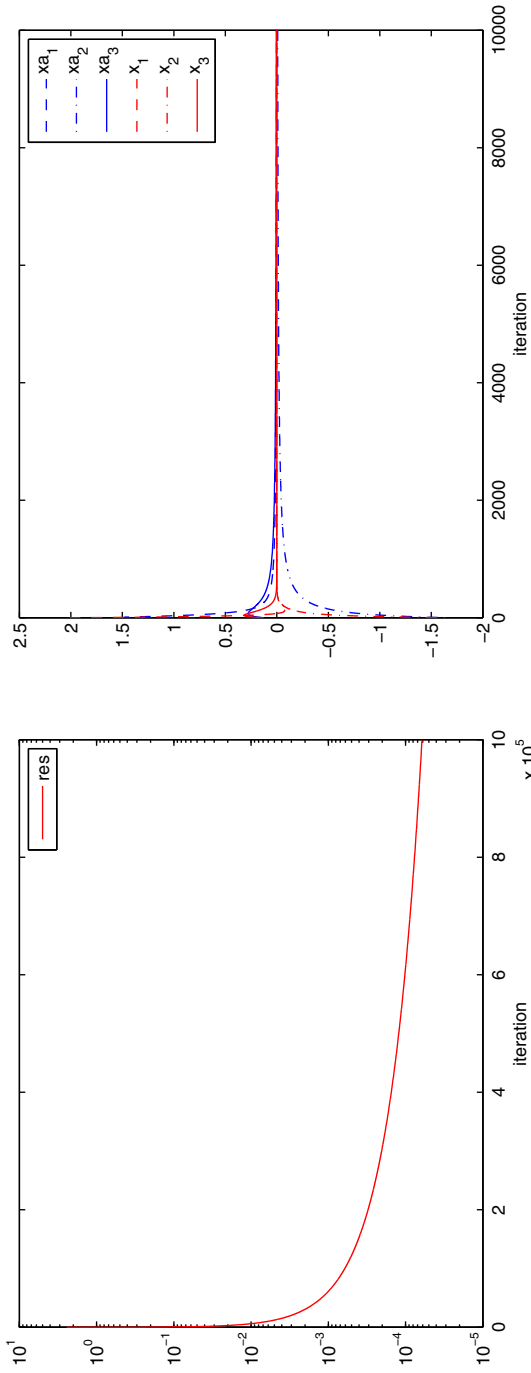
Fig. 2 ADMM for solving the associated perturbed problem (3.19) with $\epsilon = 10^{-3}$. *Left*: the residual of the equality constraint; *Right*: x^t and the average of x^k for $k = 0, \ldots, t$ (denoted as xa) as defined in Theorem 3.2

$$\nabla f_N \left(x_N^{k+1} \right) - \lambda^k + \gamma \left(\sum_{j=1}^{N-1} A_j x_j^{k+1} + x_N^{k+1} - b \right) = 0, \tag{4.2}$$

where $g_i \in \partial (f_i + \mathbf{1}_{\mathcal{X}_i})$ is a subgradient of $f_i + \mathbf{1}_{\mathcal{X}_i}$, $i = 1, 2, \ldots, N - 1$. The last equation in (1.2) that updates λ^k now becomes

$$\lambda^{k+1} := \lambda^k - \gamma \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} + \cdots + A_{N-1} x_{N-1}^{k+1} + x_N^{k+1} - b \right). \tag{4.3}$$

Moreover, by combining with (4.3), (4.1) and (4.2) can be rewritten as

$$g_i(x_i^{k+1}) - A_i^\top \lambda^{k+1} + \gamma A_i^\top \left(\sum_{j=i+1}^{N-1} A_j (x_j^k - x_j^{k+1}) + (x_N^k - x_N^{k+1}) \right) = 0, \ i = 1, \ldots, N-1, \tag{4.4}$$

$$\nabla f_N(x_N^{k+1}) - \lambda^{k+1} = 0. \tag{4.5}$$

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 $\mathrm{dist}(x, S) := \inf\{\|y - x\| : y \in S\}$ 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 KL property at $w_0 \in \{w \in \Omega : \partial f(w) \neq \emptyset\}$ if there exists $\eta \in (0, +\infty]$, a neighborhood 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,

$$\varphi'(f(\bar{w}_0) - f(w_0))\mathrm{dist}(0, \partial f(\bar{w}_0)) \geq 1. \tag{4.6}$$

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 [1,4] 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:*

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

the following inequality holds,

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

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-block 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.,

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

In the following we present the main results in this section. Specifically, Theorem 4.3 gives the convergence of the multi-block ADMM (1.2), and we include its proof in the "Appendix". Theorem 4.5 shows that the whole sequence generated by the multi-block 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 $w^* = \left(x_1^*, \ldots, x_N^*, \lambda^*\right) \in \Omega(w^0)$, then the sequence $w^k = \left(x_1^k, \ldots, x_N^k, \lambda^k\right)$ has a finite length, i.e.,*

$$\sum_{k=0}^{\infty}\left(\sum_{i=1}^{N-1}\left\|A_i x_i^k - A_i x_i^{k+1}\right\| + \left\|x_N^k - x_N^{k+1}\right\| + \left\|\lambda^k - \lambda^{k+1}\right\|\right) \leq G, \qquad (4.7)$$

where the constant G is given by

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

and

$$M = \max\left(\gamma \sum_{i=1}^{N-1}\left\|A_i^\top\right\|, \frac{1}{\gamma} + 1 + \sum_{i=1}^{N-1}\left\|A_i^\top\right\|\right) > 0, \qquad (4.8)$$

and 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 $\left(A_1 x_1^*, \ldots, A_{N-1} x_{N-1}^*, x_N^*, \lambda^*\right)$.

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. □

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

Lemma 4.6 *Suppose that* $\mathcal{L}_\gamma(x_1, \ldots, x_N, \lambda)$ *is a KL function. Let* $(x_1^{k+1}, x_2^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by the multi-block 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 under conditions in Scenario 2 that*

$$
\begin{aligned}
&f(u^*) - f(u^{k+1}) + \sum_{i=1}^{N-1} \left(x_i^* - x_i^{k+1}\right)^\top \left(-A_i^\top \lambda^{k+1}\right) - \left(x_N^* - x_N^{k+1}\right)^\top \lambda^{k+1} \\
&+ (\lambda - \lambda^{k+1})^\top \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right) \\
&+ \frac{\gamma}{2} \left(\left\|A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b\right\|^2 - \left\|A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b\right\|^2\right) \\
&+ \frac{1}{2\gamma} \left(\left\|\lambda - \lambda^k\right\|^2 - \left\|\lambda - \lambda^{k+1}\right\|^2\right) \\
&+ \gamma D(N-2) \left(\sum_{i=1}^{N-1} \left\|A_i x_i^k - A_i x_i^{k+1}\right\| + \left\|x_N^k - x_N^{k+1}\right\|\right) \\
&\geq 0,
\end{aligned}
\tag{4.9}
$$

where D is a constant that will be specified later.

Proof Note that similar to the proof of Lemma 3.1, we can prove that, for any $x_1 \in \mathcal{X}_1$,

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

For any $x_i \in \mathcal{X}_i, i = 2, 3, \ldots, N-1$, we have,

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

Therefore, for any $x_i \in \mathcal{X}_i, i = 1, \ldots, N - 1$, it holds that

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

$$
\leq \frac{\gamma}{2} \left(\left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 - \left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 \right)
$$

$$
+ \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2
$$

$$
+ \gamma \left(\sum_{i=2}^{N-1} \left\| A_i x_i - A_i x_i^{k+1} \right\| \right) \left[\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right]. \quad (4.10)
$$

Combining (4.3), (4.4), (4.5) and (4.10), it holds for any $\lambda \in \mathbf{R}^p$ that

$$
\sum_{i=1}^{N-1} \left(x_i - x_i^{k+1} \right)^\top \left(g_i(x_i^{k+1}) - A_i^\top \lambda^{k+1} \right) + \left(x_N - x_N^{k+1} \right)^\top \left(\nabla f_N(x_N^{k+1}) - \lambda^{k+1} \right)
$$

$$
+ (\lambda - \lambda^{k+1})^\top \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) + \frac{1}{\gamma} (\lambda - \lambda^{k+1})^\top (\lambda^{k+1} - \lambda^k)
$$

$$
+ \frac{\gamma}{2} \left(\left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 - \left\| A_1 x_1 + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 \right)
$$

$$
+ \frac{1}{2\gamma} \left\| \lambda^{k+1} - \lambda^k \right\|^2
$$

$$
+ \gamma \left(\sum_{i=2}^{N-1} \left\| A_i x_i - A_i x_i^{k+1} \right\| \right) \left[\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right]
$$

$$
\geq 0. \quad (4.11)
$$

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 (4.11), and invoking the convexity of f_i, we obtain

$$
f(u^*) - f(u^{k+1}) + \sum_{i=1}^{N-1} \left(x_i^* - x_i^{k+1} \right)^\top \left(-A_i^\top \lambda^{k+1} \right) + \left(x_N^* - x_N^{k+1} \right)^\top (-\lambda^{k+1})
$$

$$
+ \left(\lambda - \lambda^{k+1} \right)^\top \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) + \frac{1}{2\gamma} \left(\left\| \lambda - \lambda^k \right\|^2 - \left\| \lambda - \lambda^{k+1} \right\|^2 \right)
$$

$$
+ \frac{\gamma}{2} \left(\left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 - \left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 \right)
$$

$$
+ \gamma \left(\sum_{i=2}^{N-1} \left\| A_i x_i^* - A_i x_i^{k+1} \right\| \right) \left[\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right] \geq 0.
$$

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

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

for any $k \geq 0$ and any $i = 2, 3, \ldots, N - 1$. This implies (4.9). $\square$

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

Theorem 4.7 *Suppose that* $\mathcal{L}_\gamma(x_1, \ldots, x_N, \lambda)$ *is a KL function. Let* $(x_1^{k+1}, x_2^{k+1}, \ldots, x_N^{k+1},$ $\lambda^{k+1}) \in \Omega$ *be generated by the multi-block ADMM* (1.2) *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*

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

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

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

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

Note this also implies that when $t = O(1/\epsilon)$, $\bar{u}^t$ *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*

$$|f(\bar{u}^t) - f(u^*)| = O(\epsilon), \quad and \quad \left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| = O(\epsilon). \tag{4.12}$$

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 for any $\lambda \in \mathbf{R}^p$,

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

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

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

$$\geq \frac{1}{t+1} \sum_{k=0}^{t} \left(f(u^*) - f(u^{k+1}) + \sum_{i=1}^{N-1} \left(x_i^* - x_i^{k+1} \right)^\top \left(-A_i^\top \lambda^{k+1} \right) \right.$$

$$
+ \left(x_N^* - x_N^{k+1} \right)^\top \left(-\lambda^{k+1} \right)
$$

$$
+ (\lambda - \lambda^{k+1})^\top \left(\sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right) \Bigg)
$$

$$
\geq \frac{1}{t+1} \sum_{k=0}^{t} \Bigg[\frac{1}{2\gamma} \left(\left\| \lambda - \lambda^{k+1} \right\|^2 - \left\| \lambda - \lambda^k \right\|^2 \right)
$$

$$
- \gamma D(N-2) \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right)
$$

$$
+ \frac{\gamma}{2} \left(\left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2 - \left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^k + x_N^k - b \right\|^2 \right) \Bigg]
$$

$$
\geq -\frac{1}{2\gamma(t+1)} \left\| \lambda - \lambda^0 \right\|^2 - \frac{\gamma}{2(t+1)} \left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^0 + x_N^0 - b \right\|^2
$$

$$
- \frac{\gamma D(N-2)}{t+1} \sum_{k=0}^{t} \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_N^k - x_N^{k+1} \right\| \right)
$$

$$
\geq -\frac{1}{2\gamma(t+1)} \left\| \lambda - \lambda^0 \right\|^2 - \frac{\gamma}{2(t+1)} \left\| A_1 x_1^* + \sum_{i=2}^{N-1} A_i x_i^0 + x_N^0 - b \right\|^2 - \frac{\gamma D G(N-2)}{t+1},
$$

$$
\tag{4.13}
$$

where the last inequality holds due to Theorem 4.5. From the optimality conditions (2.1) we obtain

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

Since $\rho := \|\lambda^*\| + 1$, by the Cauchy–Schwarz inequality, we obtain

$$
0 \leq f(\bar{u}^t) - f(u^*) + \rho \left\| \sum_{i=1}^{N-1} A_i \bar{x}_i^t + \bar{x}_N^t - b \right\|.
\tag{4.14}
$$

By setting $\lambda = -\rho(\sum_{i=1}^{N-1} A_i \bar{x}_i^t + \bar{x}_N^t - b)/\| \sum_{i=1}^{N-1} A_i \bar{x}_i^t + \bar{x}_N^t - b\|$ in (4.13), and noting that $\|\lambda\| = \rho$, we obtain

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

$$
\leq \frac{\rho^2 + \|\lambda^0\|^2}{\gamma(t+1)} + \frac{\gamma}{2(t+1)} \left\| \sum_{i=2}^{N-1} A_i(x_i^0 - x_i^*) + (x_N^0 - x_N^*) \right\|^2 + \frac{\gamma D G(N-2)}{t+1}.
$$

$$
\tag{4.15}
$$

When $t = O(1/\epsilon)$, we have that the right hand side of (4.15) is of order $O(\epsilon)$. By the same arguments as in the proof for Theorem 3.2, (4.12) follows from (4.14) and (4.15). □

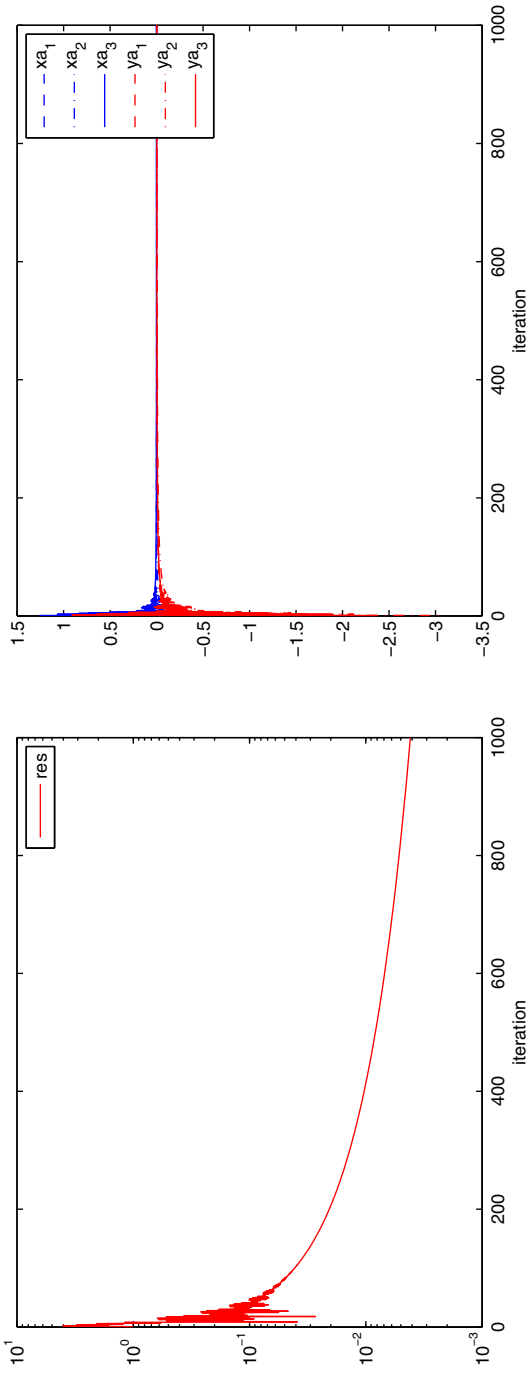**Fig. 3** ADMM for solving the sharing problem (4.16) with $\gamma = 1.01 * \sqrt{2} * L$. Left: the residual of the equality constraint; Right: the average of x^k and y^k for $k = 0, \ldots, t$ (denoted as xa and ya respectively)
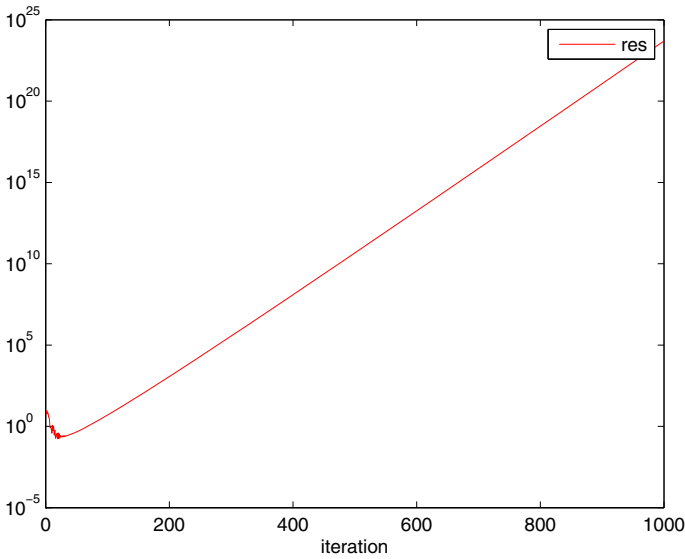
Fig. 4 The residual of the equality constraint of ADMM for solving the sharing problem (4.16) with $\gamma = 0.01 * \sqrt{2} * L$

4.1 Divergence Counterexample Revisited

In this subsection we shall test the same divergence counterexample as considered in Sect. 3.1, but now equivalently reformulated in a way to satisfy the conditions required in Scenario 2. The example in [7] can obviously be formulated equivalently as:

$$\begin{aligned} \min \ & \tfrac{1}{2}\|y\|^2 \\ \text{s.t.} \ & A_1 x_1 + A_2 x_2 + A_3 x_3 + y = 0, \end{aligned} \qquad (4.16)$$

where A_1, A_2, A_3 are the same as in Sect. 3.1. Note that (4.16) is in the form of the sharing problem with four block variables x_1, x_2, x_3 and y, and it satisfies assumptions in Scenario 2. Note that (4.16) corresponds to $f_4(\cdot) = \tfrac{1}{2}\|\cdot\|^2$ and the Lipschitz constant $L = 1$ for ∇f_4. It is easy to see that the unique solution to (4.16) is $x_1 = x_2 = x_3 = 0$ and $y = (0, 0, 0)^\top$. We then apply ADMM (1.2) to solve (4.16) with $\gamma = 1.01 * \sqrt{2} * L$ so that the assumption on γ in Scenario 2 is satisfied. Similarly as the tests in Sect. 3.1, we plot the residual of the equality constraint and the components of the average of x^k and y^k for $k = 0, \ldots, t$ in Fig. 3, where t denotes the iteration number. From Fig. 3 we again observed the convergence of ADMM. We then tested the same problem using a smaller γ, which was set to $0.01 * \sqrt{2} * L$. The residual of the equality constraint is plotted in Fig. 4. Figure 4 shows that after a few number of iterations, ADMM diverges. This indicates that if the assumption $\gamma > \sqrt{2} * L$ is violated, then it is possible that ADMM diverges for solving the sharing problem (4.16).

Appendix: Proof of Theorem 4.3

We first prove the following lemma.

Lemma 4.8 *The following results hold under the conditions in Scenario 2.*

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

$$\|\lambda^{k+1} - \lambda^k\|^2 \leq L^2 \|x_N^{k+1} - x_N^k\|, \tag{4.17}$$

   where L is the Lipschitz constant for ∇f_N.
2. *The augmented Lagrangian $\mathcal{L}_\gamma$ has a sufficient decrease in each iteration, i.e.,*

$$\mathcal{L}_\gamma \left(x_1^k, \ldots, x_{N+1}^k; \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \ldots, x_{N+1}^{k+1}; \lambda^{k+1} \right)$$
$$\geq \frac{\gamma^2 - 2L^2}{2\gamma(1 + L^2)} \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \left\| x_N^k - x_N^{k+1} \right\|^2 + \left\| \lambda^k - \lambda^{k+1} \right\|^2 \right). \tag{4.18}$$

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

$$\sum_{k=0}^{\infty} \left(\sum_{i=1}^{N-1} \left\| A_i x_i^{k+1} - A_i x_i^k \right\|^2 + \left\| x_N^{k+1} - x_N^k \right\|^2 + \left\| \lambda^{k+1} - \lambda^k \right\|^2 \right)$$
$$\leq \frac{2\gamma(1 + L^2)}{\gamma^2 - 2L^2} \left(\mathcal{L}_\gamma(w^0) - L^* \right) \tag{4.19}$$

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

$$\lim_{k \to \infty} \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \left\| x_N^k - x_N^{k+1} \right\|^2 + \left\| \lambda^k - \lambda^{k+1} \right\|^2 \right) = 0. \tag{4.20}$$

   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 an upper bound for a subgradient of augmented Lagrangian $\mathcal{L}_\gamma$ in each iteration. In fact, we define*

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

   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

$$\left\| \left(R_1^{k+1}, \dots, R_N^{k+1}, R_\lambda^{k+1} \right) \right\| \leq \sum_{i=1}^{N} \left\| R_i^{k+1} \right\| + \left\| R_\lambda^{k+1} \right\|$$

$$\leq M \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\| + \left\| x_i^k - x_i^{k+1} \right\| + \left\| \lambda^k - \lambda^{k+1} \right\| \right), \quad \forall k \geq 0, \tag{4.21}$$

where M is a constant defined in (4.8).

Proof of Lemma 4.8. 1. (4.17) follows directly from (4.5).
2. From (4.1), by invoking the convexity of f_i, we have for $i = 1, \dots, N-1$:

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

$$\leq f_i \left(x_i^k \right) - f_i \left(x_i^{k+1} \right) - \left(A_i x_i^k - A_i x_i^{k+1} \right)^\top \lambda^k$$

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

$$= \mathcal{L}_\gamma \left(x_1^{k+1}, \dots, x_{i-1}^k, x_i^k, \dots, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \dots, x_i^{k+1}, x_{i+1}^k, \dots, \lambda^k \right)$$

$$- \frac{\gamma}{2} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2. \tag{4.22}$$

Similarly, from (4.2) we can prove that

$$\mathcal{L}_\gamma \left(x_1^{k+1}, \dots, x_{N-1}^{k+1}, x_N^k, \dots; \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \dots, x_N^{k+1}; \lambda^k \right) \geq \frac{\gamma}{2} \left\| x_N^k - x_N^{k+1} \right\|^2. \tag{4.23}$$

Summing (4.22) over $i = 1, \dots, N-1$ and (4.23), we have

$$\mathcal{L}_\gamma \left(x_1^k, \dots, x_N^k, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \dots, x_N^{k+1}, \lambda^k \right)$$

$$\geq \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \frac{\gamma}{2} \left\| x_N^k - x_N^{k+1} \right\|^2. \tag{4.24}$$

On the other hand, it follows from (4.2) and (4.17) that

$$\mathcal{L}_\gamma \left(x_1^{k+1}, \dots, x_N^{k+1}, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^{k+1}, \dots, x_N^{k+1}, \lambda^{k+1} \right)$$

$$= \frac{1}{\gamma} \left\| \lambda^k - \lambda^{k+1} \right\|^2 \geq -\frac{L^2}{\gamma} \left\| x_N^k - x_N^{k+1} \right\|^2. \tag{4.25}$$

Combining (4.24) and (4.25) yields

$$
\mathcal{L}_\gamma\left(x_1^k, \ldots, x_N^k, \lambda^k\right) - \mathcal{L}_\gamma\left(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1}\right)
$$

$$
\geq \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \frac{\gamma^2 - 2L^2}{2\gamma} \left\| x_N^k - x_N^{k+1} \right\|^2
$$

$$
\geq \frac{\gamma}{2} \sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \left(\left\| x_N^k - x_N^{k+1} \right\|^2 + \left\| \lambda^k - \lambda^{k+1} \right\|^2 \right)
$$

$$
\geq \frac{\gamma^2 - 2L^2}{2\gamma(1+L^2)} \left(\sum_{i=1}^{N-1} \left\| A_i x_i^k - A_i x_i^{k+1} \right\|^2 + \left\| x_N^k - x_N^{k+1} \right\|^2 + \left\| \lambda^k - \lambda^{k+1} \right\|^2 \right).
$$

$$(4.26)$$

3. It follows from (4.2) and the fact that ∇f_N is Lipschitz continuous with constant L that,

$$
f_N\left(b - \sum_{i=1}^{N-1} A_i x_i^{k+1}\right)
$$

$$
\leq f_N\left(x_N^{k+1}\right) + \left\langle \nabla f_N\left(x_N^{k+1}\right), \left(b - \sum_{i=1}^{N-1} A_i x_i^{k+1} - x_N^{k+1}\right)\right\rangle
$$

$$
+ \frac{L}{2} \left\| b - \sum_{i=1}^{N-1} A_i x_i^{k+1} - x_N^{k+1} \right\|^2
$$

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

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

$$
\mathcal{L}_\gamma\left(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1}\right)
$$

$$
\geq \sum_{i=1}^{N-1} f_i(x_i^{k+1}) + f_N\left(b - \sum_{i=1}^{N-1} A_i x_i^{k+1}\right) + \frac{\gamma - L}{2} \left\| \sum_{i=1}^{N-1} A_i x_i^{k+1} + x_N^{k+1} - b \right\|^2
$$

$$
> L^*,
$$

$$(4.27)$$

where the last inequality holds since $\gamma > L$ and $\inf_{\mathcal{X}_i} f_i > f_i^*$ for $i = 1, 2, \ldots, N$. Therefore, it directly follows from (4.18) 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 $K \to \infty$ gives (4.19) and (4.20).
It also follows from (4.27), (4.18) 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 $\{(x_1^k, x_2^k, \ldots, x_{N-1}^k) : k = 0, 1, \ldots\}$ is a bounded sequence by using the coerciveness of $f_i + \mathbf{1}_{\mathcal{X}_i}$, $i = 1, 2, \ldots, N-1$. The boundedness of (x_N^k, λ^k) can be obtained by using (4.3), (4.17) and (4.20).

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

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

and

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

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 $i = 1, 2, \ldots, N-1$.
Combining these relations with (4.4) and (4.5) yields that

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

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

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

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

for $i = 1, 2, \ldots, N-1$. Therefore, $\left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1}\right) \in \partial \mathcal{L}_\gamma(w^{k+1})$.
We now need to bound the norms of R_i^{k+1}, $i = 1, \ldots, N-1$, R_N^k and R_λ^k. It holds that

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

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

$$\leq \gamma \left\|A_i^\top\right\| \left(\sum_{j=1}^{N-1} \left\|A_j x_j^k - A_j x_j^{k+1}\right\| + \left\|x_N^k - x_N^{k+1}\right\|\right) + \left\|A_i^\top\right\| \left\|\lambda^k - \lambda^{k+1}\right\|$$

and

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

These relations immediately imply (4.21). $\square$

Proof of Theorem 4.3. 1. It has been proven in Lemma 4.8 that $\{(x_1^k, x_2^k, \ldots, x_N^k, \lambda^k) : k = 0, 1, \ldots\}$ 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 $\{w^k = (x_1^k, \ldots, x_N^k, \lambda^k) : k = 0, 1, \ldots\}$. 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

$$\liminf_{q \to \infty} f_i(x_i^{k_q}) \geq f_i(x_i^*), \quad i = 1, 2, \ldots, N. \tag{4.28}$$

From (1.2), we have for any integer k and any $i = 1, \ldots, N-1$,

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

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

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

i.e.,

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

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

$$\limsup_{q \to +\infty} f_i \left(x_i^{k_q}\right) \leq \limsup_{q \to +\infty} \left(\frac{\gamma}{2} \left\| A_i x_i^{k_q} - A_i x_i^* \right\|^2 + \left\langle \lambda^k, A_i x_i^{k_q} - A_i x_i^* \right\rangle\right) + f_i \left(x_i^*\right), \tag{4.29}$$

for $i = 1, 2, \ldots, N-1$. Here we have used the facts that 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 (4.20), and the fact that

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

From (4.20) we also have $x_i^{k_q-1} \to x_i^*$ as $q \to \infty$, hence (4.29) reduces to $\limsup_{q \to \infty} f_i(x_i^{k_q}) \leq f_i(x_i^*)$. Therefore, combining with (4.28), $f_i(x_i^{k_q})$ tends to $f_i(x_i^*)$ as $q \to \infty$. Hence, we can conclude that

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

$$\left. + \frac{\gamma}{2} \left\| \sum_{i=1}^{N-1} A_i x_i^{k_q} + x_N^{k_q} - b \right\|^2 \right)$$

$$= \sum_{i=1}^N f_i\left(x_i^*\right) - \left\langle \lambda^*, \sum_{i=1}^{N-1} A_i x_i^* + x_N^* - b \right\rangle + \frac{\gamma}{2} \left\| \sum_{i=1}^{N-1} A_i x_i^* + x_N^* - b \right\|^2$$

$$= \mathcal{L}_\gamma(w^*).$$

On the other hand, it follows from (4.20) and (4.21) that

$$\left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1} \right) \in \partial \mathcal{L}_\gamma(w^{k+1}) \qquad (4.30)$$

$$\left(R_1^{k+1}, \ldots, R_N^{k+1}, R_\lambda^{k+1} \right) \to (0, \ldots, 0), \quad k \to \infty. \qquad (4.31)$$

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, $w^* = \left(x_1^*, \ldots, x_N^*, \lambda^* \right)$ 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 $\tilde{L}$ is the finite limit of $\mathcal{L}_\gamma(x_1^k, \ldots, x_N^k, \lambda^k)$ as k goes to infinity, i.e.,

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

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 $\mathcal{L}_\gamma(w^*) = \tilde{L}$, hence the restriction of $\mathcal{L}_\gamma(x_1, \ldots, x_N, \lambda)$ to $\Omega(w^0)$ equals $\tilde{L}$.

$\square$

References

1. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Lojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
2. Boley, D.: Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. SIAM J. Optim. **23**(4), 2183–2207 (2013)
3. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Am. Math. Soc. **362**(6), 3319–3363 (2010)
4. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearization minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1–2), 459–494 (2014)
5. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
6. Cai, X., Han, D., Yuan, X.: The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex. Preprint http://www.optimization-online.org/DB_FILE/2014/11/4644.pdf (2014)
7. Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. 155(1), 57–79 (2016)

8. Chen, C., Shen, Y., You, Y.: On the convergence analysis of the alternating direction method of multipliers with three blocks. In: Abstract and Applied Analysis, Article ID 183961 (2013)
9. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. Technical report, UCLA CAM Report 15-13 (2015)
10. Deng, W., Lai, M., Peng, Z., Yin, W.: Parallel multi-block ADMM with $o(1/k)$ convergence. Technical report, UCLA CAM 13-64 (2013)
11. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. J. Sci. Comput. **66**(3), 889–916 (2016)
12. Douglas, J., Rachford, H.H.: On the numerical solution of the heat conduction problem in 2 and 3 space variables. Trans. Am. Math. Soc. **82**, 421–439 (1956)
13. Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization. Ph.D. thesis, Massachusetts Institute of Technology (1989)
14. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
15. Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives. Pac. J. Optim. **11**(4), 619–644 (2015)
16. Fortin, M., Glowinski, R.: Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. North-Holland Pub. Co., Amsterdam (1983)
17. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems. North-Holland, Amsterdam (1983)
18. Glowinski, R., Le Tallec, P.: Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics. SIAM, Philadelphia (1989)
19. Han, D., Yuan, X.: A note on the alternating direction method of multipliers. J. Optim. Theory Appl. **155**(1), 227–238 (2012)
20. He, B., Hou, L., Yuan, X.: On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. SIAM J. Optim. **25**(4), 2274–2312 (2013)
21. He, B., Tao, M., Yuan, X.: Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J. Optim. **22**, 313–340 (2012)
22. He, B., Tao, M., Yuan, X.: Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming. Preprint http://www.optimization-online.org/DB_FILE/2012/09/3611.pdf (2012)
23. He, B., Yuan, X.: On the $O(1/n)$ convergence rate of Douglas–Rachford alternating direction method. SIAM J. Numer. Anal. **50**, 700–709 (2012)
24. He, B., Yuan, X.: On nonergodic convergence rate of Douglas–Rachford alternating direction method of multipliers. Numer. Math. **130**(3), 567–577 (2015)
25. Hong, M., Chang, T.-H., Wang, X., Razaviyayn, M., Ma, S., Luo, Z.-Q.: A block successive upper bound minimization method of multipliers for linearly constrained convex optimization (2014). Preprint ArXiv:1401.7079
26. Hong, M., Luo, Z.: On the linear convergence of the alternating direction method of multipliers (2012). Preprint ArXiv:1208.3922
27. Hong, M., Luo, Z.-Q., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM J. Optim. **26**(1), 337–364 (2016)
28. Li, M., Sun, D., Toh, K.-C.: A convergent 3-block semi-proximal ADMM for convex minimization with one strongly convex block. Asia Pac. J. Oper. Res. **32**, 1550024 (2015)
29. Lin, T., Ma, S., Zhang, S.: Global convergence of unmodified 3-block ADMM for a class of convex minimization problems (2015). Preprint ArXiv:1505.04252
30. Lin, T., Ma, S., Zhang, S.: On the global linear convergence of the ADMM with multi-block variables. SIAM J. Optim. **25**(3), 1478–1497 (2015)
31. Lin, T., Ma, S., Zhang, S.: On the sublinear convergence rate of multi-block ADMM. J. Oper. Res. Soc. China **3**(3), 251–274 (2015)
32. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
33. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM J. Optim. **23**, 475–507 (2013)
34. Peaceman, D.H., Rachford, H.H.: The numerical solution of parabolic elliptic differential equations. SIAM J. Appl. Math. **3**, 28–41 (1955)
35. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2233–2246 (2012)

36. Sun, D., Toh, K.-C., Yang, L.: A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type of constraints. SIAM J. Optim. **25**(2), 882–915 (2015)
37. Tao, M., Yuan, X.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**, 57–81 (2011)
38. Wang, X., Hong, M., Ma, S., Luo, Z.-Q.: Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. Pac. J. Optim. **11**(4), 645–667 (2015)