

Distributed Scaled Proximal ADMM Algorithms for Cooperative Localization in WSNs

Mei Zhang, Zhiguo Wang, Feng Yin, *Senior Member, IEEE*, and Xiaojing Shen

August 8, 2023

Abstract

Distributed cooperative localization in wireless networks is a challenging problem since it typically requires solving a large-scale nonconvex and nonsmooth optimization problem. In this paper, we reformulate the classic cooperative localization problem as a smooth and constrained nonconvex minimization problem while its loss function is separable over nodes. By utilizing the structure of the reformulation, we propose two novel scaled proximal alternating direction method of multipliers (SP-ADMM) algorithms, which can be implemented in a distributed manner. Compared with the classic semi-definite programming relaxation techniques, the proposed algorithms can provide more accurate position estimates with significantly lower computation complexity. The associated theoretical analysis shows that our algorithms [globally converge to a KKT point](#) of the reformulated problem and a critical point of the original problem, with a favorable sublinear $\mathcal{O}(1/T)$ convergence rate, where T is the iteration counter. Numerical experiments have consistently shown that the proposed SP-ADMM algorithms are superior to state-of-the-art methods in terms of localization accuracy and computational complexity across all tested scenarios, varying network size, number of anchors, average number of neighbors, and noise variance levels.

Keywords— Proximal ADMM, distributed algorithm, global convergence rate, cooperative localization, wireless sensor network.

I. INTRODUCTION

Wireless sensor networks (WSNs) are widely used to deal with sensitive information in a variety of applications, including healthcare, military, Internet of Things, surveillance, and industrial [1]. In the aforementioned applications, all collected information is meaningful only when the locations of the sensor nodes are accurately known. Therefore, localization is an enabling technique for WSNs. The cooperative localization problem aims to determine multiple sensor locations with the aid of a relatively small portion

Mei Zhang, Zhiguo Wang (corresponding author), and Xiaojing Shen with Department of Mathematics, Sichuan University, Chengdu, Sichuan 610064, China. E-mail: zhang_mei@stu.scu.edu.cn, wangzhiguo@scu.edu.cn, shenxj@scu.edu.cn.

Feng Yin is with the School of Science and Engineering, Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China, and also with Shenzhen Research Institute of Big Data, Shenzhen 518172, China (e-mail: yinfeng@cuhk.edu.cn).

of anchors whose positions are precisely known and the relative noisy range measurements of adjacent nodes [2]. The range measurements can be measured in a variety of ways, such as time-of-arrival (TOA) [3], [4], time-difference-of-arrival (TDOA) [5], angle-of-arrival (AOA) [6] and received signal strength (RSS) [7], [8]. In recent years, there has been a growing interest in estimating target positions through optimization techniques due to the faster response speed of position estimation by these methods and convergence guarantee.

Nonconvex and nonsmooth is the major difficulty of the least squares objective function in the maximum likelihood framework [9]. It is hard to find an optimal solution with low computational complexity, and in consequence, most existing methods resort to developing approximate solutions through applying, for instance, the relaxation techniques, to the original nonconvex problem. A popular class of methods is based on semi-definite programming (SDP) relaxation [10]–[12], which can be solved by the interior-point algorithm. Although convex relaxation techniques guarantee convergence to a global minimum point, it is not necessarily a critical point of the original nonconvex formulation. In addition, the above convex relaxation methods are implemented in a centralized framework, since all measurements are collected and processed at a processing center. Centralized methods are vulnerable to the malfunction of any single node. There have been increasing efforts in developing distributed sensor network localization methods [13]–[15].

Distributed¹ methods are able to avoid the major drawbacks of the centralized counterparts for large-scale networks, thus making them well-suited for wireless sensor network localization. Distributed methods not just have lower computational and communication complexity due to scalability but are also more robust to node failures [16], [17]. This inspired the use of distributed optimization approaches to solve the target positioning problem in large-scale networks. Compared with the centralized SDP methods, the work [18] proposed a more practical alternative. By leveraging the concept of convex envelope, the authors developed a method that is both scalable and simple to implement. They also analyzed the total number of iterations required by the algorithm to approach the optimal function value with a high probability for the convexified problem. Despite these advances, it still can be difficult to infer where the solutions are the critical points of the original nonconvex problem (discussed in [19], [20]). In view of this, non-relaxed methods may be a workable viewpoint.

In this paper, we focus on the first-order method that solves the original nonconvex localization problem directly. In general, the design of this kind of algorithm first reformulates the problem by exploiting the

¹The term distributed in this paper refers to an implementation that sensor nodes can locate themselves and neighboring nodes based on the local network information containing neighbors, without whole network data [9], [13].

internal structure of the problem and then develops an efficient optimization algorithm to solve the reformulated problem. For example, a nonconvex sequential greedy (NCSG) optimization algorithm was proposed in [21]. It is also proved that this algorithm owns the convergence guarantee inherited from the non-linear Gauss-Seidel framework [22]. Notwithstanding, the limit point of the generated sequence converges to the KKT point of the reformulated problem, the connection with the original problem is yet not provided.

One efficient way is to reformulate the original localization problem as a two-block nonconvex optimization problem with linear equality constraints, and then a distributed ADMM method is proposed in [23], [24] to solve the resulting nonconvex reformulation. In [25], [26], they show that the ADMM algorithm may be unstable when the objective function is nonconvex and nonsmooth. In order to obtain a favorable initial point for the method developed in [23], the authors in [19] proposed a hybrid ADMM (ADMM-H) method that contains two stages. In the first stage, a reliable starting point of the target problem is given by exploiting the convex relaxation technique introduced in [18]. In the second stage, the solution will be constantly updated by adopting the method developed in [23]. While the simulations demonstrated faster convergence, there remain limitations both in the local minimum theoretical guarantee and practical implementation. This is discussed in detail by [20] and our Section V.

Recently, for the single source localization problem, the authors in [24] adopted a simple variational representation of the Euclidean norm, then derived an equivalent smooth reformulation with ball constraint. Later on, it was extended for the multi-source localization problem in [20], and they proposed an alternating minimization (AM) method to solve the reformulated constrained smooth and nonconvex optimization problem. There are two versions of the AM method in [20], the fully centralized (AM-FC) and the fully distributed (AM-FD). Due to the sequential² nature of the AM-FD method, it is impractical to apply it to large-scale networks. Therefore, the authors further propose a unifying AM (AM-U) algorithm to remedy this difficulty. The basic idea of AM-U is to divide the sensors in the network into several disjoint clusters and then apply the AM-FD method separately for each cluster. The convergence of the AM-FD can be guaranteed. Since the difficulty of nonconvex and nonsmooth, the convergence rate for the original problem is not provided.

In this work, we aim to exploit the advantages of [23] and [24] to develop an efficient distributed algorithm with lower computational complexity for the nonconvex and nonsmooth localization problem and establish a theoretical guarantee of its performance.

²In this paper, sequential means sensor nodes perform their update calculations in turn, that is, they have to wait for a part of neighbor nodes to complete the update before running their local update steps.

TABLE I
COMPARISONS OF DIFFERENT ALGORITHMS

Algorithm	Convex Relaxations	First order Method	Step Size	Convergence		Sensor i		
				Rate	Parallelized	Computational Complexity	Communication Cost	Storage Space
SDP [10]	✓	✗	-	✗	-	$O(n^3)$	nN_i	$\frac{1}{2} \sum_{i=1}^N N_i - \mathcal{E}_a ^* + n(N - m)$
SF [18]	✓	✓	fixed	✗ [*]	✓	$O(nN_i)$	nN_i	$2n + N_i + 1(\text{parallel method})$
AM-FD [20]	✗	✓	fixed	✗	✗	$O(nN_i)$	nN_i	$n + nN_i + N_i + 1$
ADMM-H [19]	hybrid [†]	✗	-	✗	✓	$O(n^3T_i + nN_i)^*$	$2nN_i + N_i$	$2N_i(N_i + 1)n^2 + 4nN_i + N_i + 9$
SP-ADMM (ours)	✗	✓	fixed	$\mathcal{O}(1/T)^{\ddagger}$	✓	$O(nN_i)$	$2nN_i$	$4nN_i + N_i + 3$ (Algorithm 2)

* is due to the convergence analysis of the SF for the convex relaxation problem rather than the original nonconvex problem.

† Here, “hybrid” means a two-stage algorithm, including a convex relaxation stage and a nonconvex stage.

* T_i refers to the number of iterations required by the nonconvex Newton algorithm to converge.

‡ T refers to the total number of iterations of the algorithm.

* $|\mathcal{E}_a|$ denotes the number of edges in which both nodes are anchors. The value of storage space refers to the minimum total storage space required by the SDP per step, as described in reference [27].

Our contributions are summarized as follows.

- By introducing an auxiliary variable for the nonsmooth Euclidean norm, we reformulate the classic cooperative localization problem as a smooth and constrained nonconvex minimization problem, whose loss function is separable over nodes and has two block variables that eventually leads to a nice optimization structure.
- To exploit the nice structure of the new reformulation, we proposed a scaled proximal ADMM (SP-ADMM) algorithm, which is suitable for distributed computing and parallel implementation. Moreover, to further reduce the storage space at each node, we proposed Algorithm 2 that is a simplified version of Algorithm 1. As shown in Table I, the proposed algorithms enjoy lower computation complexity and storage space when compared with the existing SDP relaxation method and ADMM-H method, respectively.
- By utilizing a novel potential function, [we demonstrates that the global convergence of the sequence generated by the proposed algorithm, that is, the whole sequence converges to a unique KKT point of the reformulated problem and a critical point of the original problem.](#) Remarkably, these algorithms also exhibit sublinear convergence, with a convergence rate of $\mathcal{O}(1/T)$, where T represents the iteration counter. To the best of our knowledge, this is the first result that shows the sublinear convergence rate of a distributed ADMM algorithm for a nonconvex and nonsmooth localization problem.
- Numerical experiments conducted on a variety of networks have consistently demonstrated that the proposed algorithms outperform existing methods in terms of both localization accuracy and computational efficiency. The experimental data encompasses scenarios with different network sizes,

number of anchors, average number of neighbor nodes, and variance levels of the measurement noise.

Notation: We use lowercase bold letters to denote vectors, $\mathbf{1}_{N_i}$ and $\mathbf{0}_{N_i}$ to represent N_i -dimensional column vector with all ones and all zeros, respectively. Capital bold letters represent matrices, specially, \mathbf{I}_{N_i} and \mathbf{O}_{N_i} denote $N_i \times N_i$ -dimensional identity matrix and zero matrix, respectively. $\|\mathbf{x}\|$ is the Euclidean norm of a real vector \mathbf{x} , \otimes denotes the Kronecker product. We use $\text{vec}(\mathbf{x}_i, i \in \mathcal{N})$ to denote the concatenated vector of \mathbf{x}_i for all $i \in \mathcal{N}$ and $\text{Diag}(\mathbf{z})$ to denote the diagonal matrix with the coefficients of \mathbf{z} along the diagonal. The projection operator of set \mathcal{B}^{N_i} is defined as:

$$\text{proj}_{\mathcal{B}^{N_i}}(\mathbf{u}_i^t) := \arg \min_{\mathbf{u}_i \in \mathcal{B}^{N_i}} \frac{1}{2} \|\mathbf{u}_i - \mathbf{u}_i^t\|^2.$$

Lastly, let $f : \mathcal{C} \rightarrow (-\infty, +\infty]$ be a proper closed and convex function and \mathbf{W} be a positive semi-definite matrix. Then the scaled proximal operator of f is given by

$$\text{prox}_f^{\mathbf{W}}(\mathbf{z}) := \arg \min_{\mathbf{v} \in \mathcal{C}} f(\mathbf{v}) + \frac{1}{2} \|\mathbf{v} - \mathbf{z}\|_{\mathbf{W}}^2,$$

where $\|\cdot\|_{\mathbf{W}}$ is the scaled norm induced by \mathbf{W} , i.e., $\|\mathbf{z}\|_{\mathbf{W}}^2 := \langle \mathbf{z}, \mathbf{W}\mathbf{z} \rangle$ for every \mathbf{z} in $\mathbb{R}^{(2N_i+1)n}$.

Synopsis: Section II introduces the reformulated smooth constrained nonconvex minimization problem. The proposed SP-ADMM algorithms are presented in Section III. Section IV presents the theoretical results of the convergence conditions and convergence rate of the SP-ADMM algorithm. The performance of the proposed SP-ADMM algorithm is illustrated in Section V, and the conclusion is given in Section VI.

II. PROBLEM FORMULATION

A. Problem Statement

The wireless sensor network that we consider throughout this paper is represented as an undirected and connected graph, $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, and the topology is assumed to be known. The node set $\mathcal{N} = \{1, 2, \dots, N\}$ consists of N nodes, some of which are anchors with known true positions collected in the set $\mathcal{A} = \{\mathbf{a}_{N-m+1}, \dots, \mathbf{a}_N\} \subset \mathcal{N}$. For each node $i \in \mathcal{N}$, we define $\mathcal{N}_i = \{j \mid (i, j) \in \mathcal{E}\}$ as the set of adjacent nodes to node i , and N_i as the cardinality of \mathcal{N}_i . The true position of node i is represented by $\mathbf{p}_i \in \mathbb{R}^n$ for $i \in \mathcal{N}$, and the collection of all node positions is denoted by $\mathbf{p} = \text{vec}(\mathbf{p}_i, i \in \mathcal{N}) \in \mathbb{R}^{nN}$. The available noisy range measurement between node i and its adjacent node $j \in \mathcal{N}_i$ is denoted as $d_{i,j}$, and we assume that $d_{i,j} = d_{j,i}$ following the convention in [3], [18]. Specifically, the noisy range measurement $d_{i,j}$ can be expressed as in [28], [29],

$$d_{i,j} = \|\mathbf{p}_i - \mathbf{p}_j\| + w_{i,j}, \quad i \in \mathcal{N}, j \in \mathcal{N}_i, \quad (1)$$

where $w_{i,j}$ are the zero-mean, independent, and identically-distributed Gaussian measurement noise terms.

Using these notations, the maximum likelihood estimator, as our baseline, can be obtained through solving the following nonconvex constrained optimization problem (following [23])

$$\arg \min_{\mathbf{p} \in \mathbb{R}^{nN}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{1}{2} (\|\mathbf{p}_i - \mathbf{p}_j\| - d_{i,j})^2 \quad (2a)$$

$$\text{subject to } \mathbf{p}_k = \mathbf{a}_k, \quad \forall k \in \mathcal{A}. \quad (2b)$$

B. Problem Reformulation

Our first step is to derive an equivalent smooth and constrained reformulation of problem (2), which provides the key insight toward algorithm design and its convergence of this paper. Notice that the objective function in problem (2) can be written explicitly as

$$\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left[\frac{1}{2} \|\mathbf{p}_i - \mathbf{p}_j\|^2 - d_{i,j} \underbrace{\|\mathbf{p}_i - \mathbf{p}_j\|}_{\text{nonsmooth}} + \frac{1}{2} d_{i,j}^2 \right]. \quad (3)$$

Obviously, there is a nonsmooth term in the objective function. Motivated by the recent works [20], [24], we apply the Cauchy-Schwartz inequality to obtain

$$\|\mathbf{p}_i - \mathbf{p}_j\| = \max_{\mathbf{u}_{i,j} \in \mathcal{B}} \mathbf{u}_{i,j}^T (\mathbf{p}_i - \mathbf{p}_j), \quad (4)$$

where $\mathcal{B} := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq 1\}$ is a unit ball in \mathbb{R}^n with the center at the origin and $\mathbf{u}_{i,j} \in \mathbb{R}^n$ is an auxiliary variable, then we have

$$-\|\mathbf{p}_i - \mathbf{p}_j\| = \min_{\mathbf{u}_{i,j} \in \mathcal{B}} -\mathbf{u}_{i,j}^T (\mathbf{p}_i - \mathbf{p}_j). \quad (5)$$

Substituting (5) into (3), problem (2) is rewritten as a minimization problem of a smooth function over a ball constraint set as follows

$$\arg \min_{\mathbf{p}, \mathbf{u}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left[\frac{1}{2} \|\mathbf{p}_i - \mathbf{p}_j\|^2 - d_{i,j} \mathbf{u}_{i,j}^T (\mathbf{p}_i - \mathbf{p}_j) \right] \quad (6a)$$

$$\text{subject to } \mathbf{u}_i \in \mathcal{B}^{N_i}, \quad \forall i \in \mathcal{N}, \quad (6b)$$

$$\mathbf{p}_k = \mathbf{a}_k, \quad \forall k \in \mathcal{A}. \quad (6c)$$

where $\mathbf{u} := \text{vec}(\mathbf{u}_i, i \in \mathcal{N})$, $\mathbf{u}_i := \text{vec}(\mathbf{u}_{i,j}, j \in \mathcal{N}_i)$, and $\mathcal{B}^{N_i} := \{(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_i}) \mid \mathbf{x}_j \in \mathcal{B}, j = 1, \dots, N_i\}$ is the Cartesian product of N_i balls \mathcal{B} . To decouple the loss function of problem (6) over nodes, we introduce auxiliary variables that duplicate the positions of neighboring nodes. Specifically, we define $\mathbf{z}_{i,j}^+ \in \mathbb{R}^n$ as a copy of the position \mathbf{p}_j for each neighbor $j \in \mathcal{N}_i$, i.e.,

$$\mathbf{z}_{i,j}^+ := \mathbf{p}_j, \quad j \in \mathcal{N}_i. \quad (7)$$

Similarly, we define $\mathbf{z}_{i,j}^- \in \mathbb{R}^n$ as a copy of the position \mathbf{p}_i assigned to neighbor $j \in \mathcal{N}_i$, i.e.,

$$\mathbf{z}_{i,j}^- := \mathbf{p}_i, \quad j \in \mathcal{N}_i. \quad (8)$$

By collecting all variables associated with node i together, we define a new variable \mathbf{z}_i as

$$\mathbf{z}_i := \begin{bmatrix} \mathbf{p}_i \\ \mathbf{z}_i^- \\ \mathbf{z}_i^+ \end{bmatrix} \in \mathbb{R}^{(2N_i+1)n}, \quad i \in \mathcal{N}. \quad (9)$$

Here, $\mathbf{z}_i^- := \text{vec}(\mathbf{z}_{i,j}^-, j \in \mathcal{N}_i) \in \mathbb{R}^{N_i n}$ is the collection of N_i copies of the position \mathbf{p}_i , and $\mathbf{z}_i^+ := \text{vec}(\mathbf{z}_{i,j}^+, j \in \mathcal{N}_i) \in \mathbb{R}^{N_i n}$ represents the collection of all replicas of \mathbf{p}_j from neighboring node $j \in \mathcal{N}_i$.

Using these notations, the objective function (6) can be written in a separable form as follows:

$$\sum_{i \in \mathcal{N}} \left[\underbrace{\frac{1}{2} \|\mathbf{Q}_i \mathbf{z}_i\|^2 - \mathbf{u}_i^T \mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i + \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i)}_{F_i(\mathbf{z}_i, \mathbf{u}_i)} \right], \quad (10)$$

where $\mathbf{D}_i := \mathbf{Diag}(\text{vec}(d_{i,j}, j \in \mathcal{N}_i)) \otimes \mathbf{I}_n \in \mathbb{R}^{N_i n \times N_i n}$ is the measurement matrix of node i and

$$\mathbf{Q}_i := [\mathbf{1}_{N_i}, \mathbf{O}_{N_i}, -\mathbf{I}_{N_i}] \otimes \mathbf{I}_n \in \mathbb{R}^{N_i n \times (1+2N_i)n}. \quad (11)$$

By using (9), we can rewrite (8) as a compact form

$$\mathbf{A}_i \mathbf{z}_i = \mathbf{0}, \quad (12)$$

where $\mathbf{A}_i := [\mathbf{1}_{N_i}, -\mathbf{I}_{N_i}, \mathbf{O}_{N_i}] \otimes \mathbf{I}_n \in \mathbb{R}^{N_i n \times (1+2N_i)n}$. Moreover, for a pair of connected sensors i and j , we can deduce from equation (8) that $\mathbf{z}_{j,i}^- = \mathbf{p}_j$. Combining this equation with (7), we obtain an additional constraint that applies to all connected sensors:

$$\mathbf{z} := \text{vec}(\mathbf{z}_i, i \in \mathcal{N}) \in \mathcal{Z}, \quad \mathcal{Z} := \{\mathbf{z} | \mathbf{z}_{i,j}^+ = \mathbf{p}_j = \mathbf{z}_{j,i}^-, \forall i \in \mathcal{N}, j \in \mathcal{N}_i\}. \quad (13)$$

Using (9), the set of anchors in (6c) can be expressed as

$$\mathbf{z} \in \mathcal{X} := \{\mathbf{z} | \mathbf{E}_i \mathbf{z}_i = \mathbf{a}_i, \forall i \in \mathcal{A}\}, \quad \mathbf{E}_i := [1, \mathbf{0}_{N_i}^T, \mathbf{0}_{N_i}^T] \otimes \mathbf{I}_n \in \mathbb{R}^{n \times (1+2N_i)n}. \quad (14)$$

Finally, using (10), (12), and (13), the nonconvex optimization problem (6) can be equivalently reformulated into the compact form

$$\arg \min_{\mathbf{z}, \mathbf{u}} \sum_{i \in \mathcal{N}} \left[\underbrace{\frac{1}{2} \|\mathbf{Q}_i \mathbf{z}_i\|^2 - \mathbf{u}_i^T \mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i + \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i)}_{F_i(\mathbf{z}_i, \mathbf{u}_i)} \right] \quad (15a)$$

$$\text{subject to } \mathbf{A}_i \mathbf{z}_i = \mathbf{0}, \quad i \in \mathcal{N}, \quad (15b)$$

$$\mathbf{z} \in \mathcal{X}, \quad \mathbf{z} \in \mathcal{Z}. \quad (15c)$$

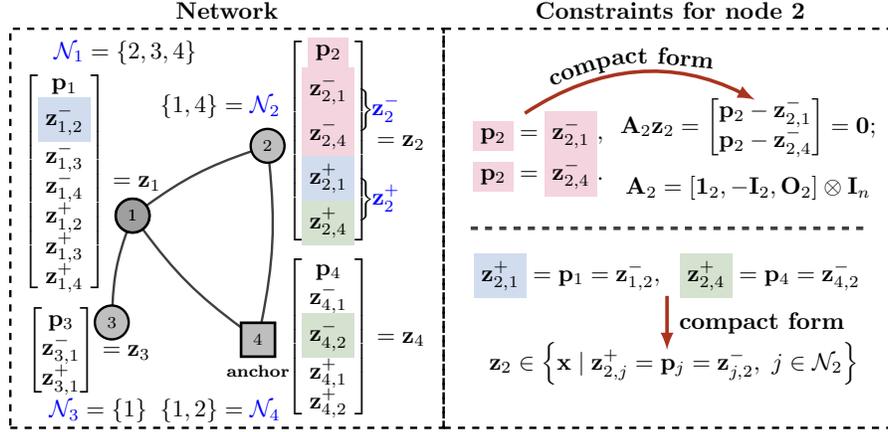


Fig. 1. An illustrating example of a connected wireless sensor network in 2-D space ($N = 4, m = 1$).

To show these notations clearly, we give an example of a connected wireless sensor network in 2-D space in Fig. 1.

Example 1 Fig.1 presents a wireless sensor network with 3 (inexact) positions of node $i = 1, 2, 3$ and one anchor of node 4. If we only focus on node 2, on the one hand, since $\mathcal{N}_2 = \{1, 4\}$ and $\mathbf{z}_{2,1}^-, \mathbf{z}_{2,4}^-$ are the copies of \mathbf{p}_2 , then we have

$$\mathbf{p}_2 = \mathbf{z}_{2,1}^-, \quad \mathbf{p}_2 = \mathbf{z}_{2,4}^-.$$

On the other hand, Fig. 1 presents that there exist the same partial elements in \mathbf{z}_2 , \mathbf{z}_1 , and \mathbf{z}_4 . Specifically, \mathbf{p}_1 is the element of both \mathbf{z}_1 and \mathbf{z}_2 ; \mathbf{z}_2 and \mathbf{z}_4 also contain \mathbf{p}_4 . To enforce this trivial observation, we obtain the following constraint

$$\mathbf{z}_{2,1}^+ = \mathbf{p}_1 = \mathbf{z}_{1,2}^-, \quad \mathbf{z}_{2,4}^+ = \mathbf{p}_4 = \mathbf{z}_{4,2}^-,$$

which is part of the constraint set \mathcal{Z} defined in (13).

Remark 1 The objective function $F_i(\mathbf{z}_i, \mathbf{u}_i)$ in problem (15) is convex with respect to \mathbf{z}_i when \mathbf{u}_i is fixed, however, it is NOT a jointly convex function for $(\mathbf{z}_i, \mathbf{u}_i)$. In addition, one can observe that the objective function is separable but the linear constraint \mathcal{Z} is NOT separable in \mathbf{z} . In Section III, we introduce a scaled proximal term to deal with this problem.

III. PROPOSED SP-ADMM ALGORITHMS

In this section, we present the proposed SP-ADMM algorithms for solving the nonconvex and nonsmooth optimization problem derived in (15). The proposed SP-ADMM algorithms are built upon the ADMM method [24] developed for large-scale nonconvex problems.

First, we introduce the augmented Lagrangian (AL) function of problem (15) as follows:

$$\mathcal{L}(\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda}) := \sum_{i \in \mathcal{N}} \mathcal{L}_i(\mathbf{z}_i, \mathbf{u}_i, \boldsymbol{\lambda}_i),$$

where

$$\mathcal{L}_i(\mathbf{z}_i, \mathbf{u}_i, \boldsymbol{\lambda}_i) := F_i(\mathbf{z}_i, \mathbf{u}_i) + \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i) + \langle \boldsymbol{\lambda}_i, \mathbf{A}_i \mathbf{z}_i \rangle + \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i\|^2, \quad i \in \mathcal{N}, \quad (16)$$

and $\boldsymbol{\lambda}_i := \text{vec}(\boldsymbol{\lambda}_{i,j}, j \in \mathcal{N}_i) \in \mathbb{R}^{N_i n}$ correspond to the Lagrangian multipliers, $c > 0$ is a penalty coefficient. We apply the following scaled proximal ADMM updates [30] to obtain

$$\mathbf{z}^{t+1} = \arg \min_{\substack{\mathbf{z} \in \mathcal{Z} \\ \mathbf{z} \in \mathcal{X}}} \sum_{i \in \mathcal{N}} \mathcal{L}_i(\mathbf{z}_i, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) + \frac{c}{2} \|\mathbf{z}_i - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2, \quad (17)$$

$$\mathbf{u}^{t+1} = \arg \min_{\mathbf{u}} \sum_{i \in \mathcal{N}} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i, \boldsymbol{\lambda}_i^t) + \frac{\rho}{2} \|\mathbf{u}_i - \mathbf{u}_i^t\|^2, \quad (18)$$

$$\boldsymbol{\lambda}_i^{t+1} = \boldsymbol{\lambda}_i^t + c \mathbf{A}_i \mathbf{z}_i^{t+1}, \quad \forall i \in \mathcal{N}, \quad (19)$$

where $\rho > 0$ is a penalty coefficient and t denotes the iteration step.

The AL function \mathcal{L} is separable in each of the variables $(\mathbf{z}_i, \mathbf{u}_i, \boldsymbol{\lambda}_i)$, $i \in \mathcal{N}$ and \mathcal{L}_i , $i \in \mathcal{N}$ is convex as a function separately of $(\mathbf{z}_i, \mathbf{u}_i)$ when the others are fixed. These features are key to distributed implementation. In addition, we remark that the (scaled) proximal term is critical in both the implementation efficiency and the theoretical analysis. It is used to ensure the following attractive properties:

- $\mathbf{z}_i \rightarrow \mathcal{L}_i(\mathbf{z}_i, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) + \frac{c}{2} \|\mathbf{z}_i - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2$ is strongly convex;
- $\mathbf{u}_i \rightarrow \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i, \boldsymbol{\lambda}_i^t) + \frac{\rho}{2} \|\mathbf{u}_i - \mathbf{u}_i^t\|^2$ is strongly convex;
- The sequence $\{(\mathbf{z}_i^t, \mathbf{u}_i^t)\}$ has lower computational complexity.

It should be noted that applying distinct penalty coefficients for \mathbf{z}_i and \mathbf{u}_i is necessary due to they control different aspects of the optimization problem. The proximal term controlled by ρ only affects the update for \mathbf{u} , whereas the AL term controlled by c affects the update for \mathbf{z} and requires the use of a dual multiplier $\boldsymbol{\lambda}$ to enforce the constraint. The choice of c can have a crucial impact on the convergence behavior of the algorithm. Moreover, using different penalty coefficients enables users to fine-tune the algorithm's parameters for optimal performance in practice presented in Section V.

To see the first point, we reformulate the objective function of problem (17). Combining (15a) and (16), and rearranging the above quadratic term, we have

$$\begin{aligned} \mathcal{L}_i(\mathbf{z}_i, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) + \frac{c}{2} \|\mathbf{z}_i - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 &= \frac{1}{2} \|\mathbf{z}_i\|_{\mathbf{W}_i}^2 - \langle \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t - \mathbf{A}_i^T \boldsymbol{\lambda}_i^t + c \mathbf{B}_i^T \mathbf{B}_i \mathbf{z}_i^t, \mathbf{z}_i \rangle \\ &\quad + \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i^t) + \frac{c}{2} \|\mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2, \end{aligned} \quad (20)$$

where

$$\mathbf{W}_i = \mathbf{Q}_i^T \mathbf{Q}_i + c \mathbf{A}_i^T \mathbf{A}_i + c \mathbf{B}_i^T \mathbf{B}_i. \quad (21)$$

Suppose $\mathbf{B}_i^T \mathbf{B}_i$ is chosen such that \mathbf{W}_i is a positive definite matrix, then the objective function of subproblem (17) is strongly convex, which ensures that the subproblem (17) has a unique solution. By completing the square of (20), we get

$$\mathcal{L}_i(\mathbf{z}_i, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) + \frac{c}{2} \|\mathbf{z}_i - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 = \frac{1}{2} \|\mathbf{z}_i - \tilde{\mathbf{z}}_i^{t+1}\|_{\mathbf{W}_i}^2 - \frac{1}{2} \|\tilde{\mathbf{z}}_i^{t+1}\|_{\mathbf{W}_i}^2 + \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i^t) + \frac{c}{2} \|\mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2, \quad (22)$$

where

$$\tilde{\mathbf{z}}_i^{t+1} = \mathbf{W}_i^{-1} (\mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t - \mathbf{A}_i^T \boldsymbol{\lambda}_i^t + c \mathbf{B}_i^T \mathbf{B}_i \mathbf{z}_i^t). \quad (23)$$

Substituting (22) into (17) and omitting the constant terms, the optimization problem (17) is equivalent to

$$\mathbf{z}^{t+1} = \arg \min_{\mathbf{z} \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}} \sum_{i \in \mathcal{N}} \frac{1}{2} \|\mathbf{z}_i - \tilde{\mathbf{z}}_i^{t+1}\|_{\mathbf{W}_i}^2. \quad (24)$$

Optimization problem (24) presents two challenges. Firstly, obtaining an analytic solution is difficult due to the linear constraints of both \mathcal{Z} and \mathcal{X} . Secondly, calculating the inverse matrix \mathbf{W}_i in (23) to obtain $\tilde{\mathbf{z}}_i^{t+1}$ can be computationally expensive. In the following, we demonstrate that an appropriate choice of $\mathbf{B}_i^T \mathbf{B}_i$ can overcome these challenges. To facilitate the calculation of the inverse of \mathbf{W}_i , we set

$$c \mathbf{B}_i^T \mathbf{B}_i = c |\mathbf{A}_i^T \mathbf{A}_i| + |\mathbf{Q}_i^T \mathbf{Q}_i|, \quad (25)$$

where $|\cdot|$ takes element-wise absolute value of a matrix. From the definition of \mathbf{A}_i and \mathbf{Q}_i in (12) and (11) respectively, we get

$$\mathbf{A}_i^T \mathbf{A}_i = \begin{bmatrix} N_i & -\mathbf{1}_{N_i}^T & \mathbf{0}_{N_i}^T \\ -\mathbf{1}_{N_i} & \mathbf{I}_{N_i} & \mathbf{0}_{N_i} \\ \mathbf{0}_{N_i} & \mathbf{0}_{N_i} & \mathbf{0}_{N_i} \end{bmatrix} \otimes \mathbf{I}_n, \quad (26)$$

$$\mathbf{Q}_i^T \mathbf{Q}_i = \begin{bmatrix} N_i & \mathbf{0}_{N_i}^T & -\mathbf{1}_{N_i}^T \\ \mathbf{0}_{N_i} & \mathbf{0}_{N_i} & \mathbf{0}_{N_i} \\ -\mathbf{1}_{N_i} & \mathbf{0}_{N_i} & \mathbf{I}_{N_i} \end{bmatrix} \otimes \mathbf{I}_n. \quad (27)$$

Substituting (26)-(27) into (25), we obtain

$$c\mathbf{B}_i^T\mathbf{B}_i = \begin{bmatrix} (c+1)N_i & c \cdot \mathbf{1}_{N_i}^T & \mathbf{1}_{N_i}^T \\ c \cdot \mathbf{1}_{N_i} & c \cdot \mathbf{I}_{N_i} & \mathbf{O}_{N_i} \\ \mathbf{1}_{N_i} & \mathbf{O}_{N_i} & \mathbf{I}_{N_i} \end{bmatrix} \otimes \mathbf{I}_n. \quad (28)$$

With the aid of (26)-(28), we can guarantee that \mathbf{W}_i defined in (21) is a positive definite diagonal matrix of the form

$$\mathbf{W}_i = 2 \cdot \mathbf{Diag} \left([(c+1)N_i, c \cdot \mathbf{1}_{N_i}^T, \mathbf{1}_{N_i}^T] \right) \otimes \mathbf{I}_n. \quad (29)$$

It makes $\tilde{\mathbf{z}}_i^{t+1}$ in (23) easy to compute. Moreover, by denoting $\tilde{\mathbf{z}}_i := [(\tilde{\mathbf{p}}_i)^T, (\tilde{\mathbf{z}}_i^-)^T, (\tilde{\mathbf{z}}_i^+)^T]^T$, where its last two parts are

$$\tilde{\mathbf{z}}_i^- := \text{vec} \left(\tilde{\mathbf{z}}_{i,j}^-, j \in \mathcal{N}_i \right), \quad \tilde{\mathbf{z}}_i^+ := \text{vec} \left(\tilde{\mathbf{z}}_{i,j}^+, j \in \mathcal{N}_i \right),$$

then \mathbf{z}_i^{t+1} can be deduced using the following remark.

Remark 2 Given \mathbf{B}_i , $i \in \mathcal{N}$ in (25), the analytic formula of the optimal solution \mathbf{z}_i^{t+1} in (24) is derived as follows

$$\mathbf{p}_i^{t+1} = \tilde{\mathbf{p}}_i^{t+1}, \quad (30)$$

$$(\mathbf{z}_i^-)^{t+1} = \text{vec} \left(\frac{c}{c+1} (\tilde{\mathbf{z}}_{i,j}^-)^{t+1} + \frac{1}{c+1} (\tilde{\mathbf{z}}_{j,i}^+)^{t+1}, j \in \mathcal{N}_i \right), \quad (31)$$

$$(\mathbf{z}_i^+)^{t+1} = \text{vec} \left(\frac{1}{c+1} (\tilde{\mathbf{z}}_{i,j}^+)^{t+1} + \frac{c}{c+1} (\tilde{\mathbf{z}}_{j,i}^-)^{t+1}, j \in \mathcal{N}_i \right). \quad (32)$$

The proof is shown in the Section 1 in the supplement material. From (30)-(32), one can see that \mathbf{z}_i^{t+1} can be obtained in a distributed fashion using only $\tilde{\mathbf{z}}_j^{t+1}$, $j \in \mathcal{N}_i$ from its neighbors.

To see the last point for \mathbf{u}_i , by omitting the terms that are irrelevant with \mathbf{u} , the optimization problem in (18) reduces to

$$\arg \min_{\mathbf{u}} \sum_{i \in \mathcal{N}} \left(-\langle \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i, \mathbf{z}_i^{t+1} \rangle + \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i) + \frac{\rho}{2} \|\mathbf{u}_i - \mathbf{u}_i^t\|^2 \right).$$

The above objective function is separable in \mathbf{u}_i and further performing projection yields

$$\mathbf{u}_i^{t+1} = \text{proj}_{\mathcal{B}^{N_i}} \left(\mathbf{u}_i^t + \frac{1}{\rho} \mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i^{t+1} \right), \quad (33)$$

which represents a projection onto the unit ball that can be reduced to

$$\mathbf{u}_{i,j}^{t+1} = \frac{\tilde{\mathbf{u}}_{i,j}^{t+1}}{\max\{1, \|\tilde{\mathbf{u}}_{i,j}^{t+1}\|\}}, \quad j \in \mathcal{N}_i, \quad (34)$$

where

$$\text{vec} \left(\tilde{\mathbf{u}}_{i,j}^{t+1}, j \in \mathcal{N}_i \right) = \mathbf{u}_i^t + \frac{1}{\rho} \mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i^{t+1}.$$

In summary, problem (17) and problem (18) can be equivalently rewritten as problem (24) and problem (33), respectively, and the corresponding distributed solutions are given in (30)-(32) and (34). An outline of the distributed procedure is described in Algorithm 1.

Algorithm 1: Distributed SP-ADMM algorithm

Input: parameter c, ρ , and initial values of $\mathbf{z}_i^0, \mathbf{u}_i^0, \boldsymbol{\lambda}_i^0 = \mathbf{0}$, $i = 1, \dots, N$.

Let $\mathbf{W}_i = 2 \cdot \text{Diag} \left([(c+1)N_i, c \cdot \mathbf{1}_{N_i}^T, \mathbf{1}_{N_i}^T] \right) \otimes \mathbf{I}_n$.

for $t \in \{0, \dots, T\}$ **do**

for client $i = 1, 2, \dots, N$ **in parallel do**

$$\tilde{\mathbf{z}}_i^{t+1} = \mathbf{W}_i^{-1} \left(\mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t - \mathbf{A}_i^T \boldsymbol{\lambda}_i^t + c \mathbf{B}_i^T \mathbf{B}_i \mathbf{z}_i^t \right) \quad (35)$$

$$\tilde{\mathbf{p}}_i^{t+1} = \mathbf{a}_i, \text{ if } i \in \mathcal{A} \quad (36)$$

end

communication:

\Rightarrow Broadcast $(\tilde{\mathbf{z}}_{i,j}^-)^{t+1}, (\tilde{\mathbf{z}}_{i,j}^+)^{t+1}$ to region $j \in \mathcal{N}_i$,

\Leftarrow Receive $(\tilde{\mathbf{z}}_{j,i}^-)^{t+1}, (\tilde{\mathbf{z}}_{j,i}^+)^{t+1}$ from region $j \in \mathcal{N}_i$.

for client $i = 1, 2, \dots, N$ **in parallel do**

$$\text{Update } \mathbf{z}_i^{t+1} \text{ via (30) – (32)} \quad (37)$$

$$\text{Update } \mathbf{u}_i^{t+1} \text{ via (34)}$$

$$\boldsymbol{\lambda}_i^{t+1} = \boldsymbol{\lambda}_i^t + c \mathbf{A}_i \mathbf{z}_i^{t+1} \quad (38)$$

end

end

Remark 3 We show that the SP-ADMM algorithm is closely related to the ADMM algorithm derived in [24] for single source localization. Specifically, let us set $\mathbf{p}^t = \mathbf{p}_i^t$, $\mathbf{a}_j = \mathbf{p}_j^t$ for all $t \geq 0$, $j \in \mathcal{N}_i = \mathcal{A}$ (single source localization). Then, by (30) and the definition of \mathbf{E}_i in (14), we obtain

$$\mathbf{p}^{t+1} = \tilde{\mathbf{p}}_i^{t+1} = \mathbf{E}_i \tilde{\mathbf{z}}_i^{t+1}, \quad i \notin \mathcal{A}. \quad (39)$$

Multiplying \mathbf{E}_i on both sides of (23) and rearranging terms yields

$$\tilde{\mathbf{p}}_i^{t+1} = \sum_{j \in \mathcal{N}_i} \frac{1}{2(c+1)N_i} \left[d_{ij} \mathbf{u}_{i,j}^t + (2c+1) (\mathbf{p}_i^t - \mathbf{p}_j^t) - \boldsymbol{\lambda}_{i,j}^t + (2c+1) \mathbf{p}_j^t + (\mathbf{z}_{i,j}^+)^t - c (\mathbf{p}_i^t - (\mathbf{z}_{i,j}^-)^t) \right]. \quad (40)$$

If $\mathbf{p}_i^t = (\mathbf{z}_{i,j}^-)^t$, which is $\mathbf{A}_i \mathbf{z}_i^t = \mathbf{0}$, $i \in \mathcal{N}$, then combining (39) with (40), replacing \mathbf{p}_j , $\mathbf{z}_{i,j}^+$ with \mathbf{a}_j , and rearranging terms again we have

$$\mathbf{p}^{t+1} = \frac{1}{m} \sum_{j=1}^m \left[\mathbf{a}_j + \frac{1}{2(c+1)m} (d_j \mathbf{u}_j^t + (2c+1) (\mathbf{p}^t - \mathbf{a}_j) - \boldsymbol{\lambda}_j^t) \right]. \quad (41)$$

Here we have replaced $d_{i,j}$, $\mathbf{u}_{i,j}^t$, $\boldsymbol{\lambda}_{i,j}^t$ by d_j , \mathbf{u}_j^t , $\boldsymbol{\lambda}_j^t$ since i is a fixed value in the single source localization problem. Hence, if \mathbf{z}_i^t satisfies the linear equality constraint (15b), then equation (41) is identical to the ADMM algorithm given in [[24], Equation (3.8)] except for the constant coefficients of the variables. Therefore, the proposed SP-ADMM algorithm can be regarded as an extended version of the ADMM algorithm in [[24] with extra capability to handle the multi-source localization problems.

However, in the case of multi-source localization, the ADMM algorithm can be significantly more complex than in single-source localization due to coupled variables and increased computational complexity. Coupled variables make subproblems in ADMM entangled and thus the convergence of the algorithm is hard to prove. Moreover, the increased number of variables in multi-source localization increases computational complexity, leading to slower convergence rates and reduced localization accuracy.

In Algorithm 1, we observed that in order to carry out the new round of iterations, node i has to store the following values: \mathbf{z}_i^t , \mathbf{u}_i^t , $\boldsymbol{\lambda}_i^t$, \mathbf{W}_i , $\mathbf{B}_i^T \mathbf{B}_i$, \mathbf{D}_i , \mathbf{Q}_i , \mathbf{A}_i , c , ρ , N_i . If we assume that the storage unit occupied by any real number is one, then node i requires $13n^2 N_i^2 + 10n^2 N_i + 2n^2 + 4nN_i + n + 3$ storage units in total. Note that although the matrices and parameters involved are fixed, the rest of the vectors are updated with each iteration. To reduce the storage space required by each sensor to run Algorithm 1, we transform the update step as follows.

Let us start by giving an explicit formula for the parts of the update variables in Algorithm 1. Using the form of \mathbf{A}_i , \mathbf{Q}_i , and \mathbf{D}_i , defined in (12), (11), and (10), respectively, we have

$$\begin{aligned} \mathbf{A}_i^T \boldsymbol{\lambda}_i^t &= \left[\sum_{j \in \mathcal{N}_i} (\boldsymbol{\lambda}_{i,j}^t)^T, \text{vec} (-\boldsymbol{\lambda}_{i,j}^t, j \in \mathcal{N}_i)^T, \mathbf{0}_{N_i}^T \right]^T, \\ \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t &= \left[\sum_{j \in \mathcal{N}_i} (d_{i,j} \mathbf{u}_{i,j}^t)^T, \mathbf{0}_{N_i}^T, \text{vec} (-d_{i,j} \mathbf{u}_{i,j}^t, j \in \mathcal{N}_i)^T \right]^T. \end{aligned} \quad (42)$$

Also, it follows from the definition (28) that

$$c\mathbf{B}_i^T \mathbf{B}_i \mathbf{z}_i^t = \left[\left((c+1)N_i \mathbf{p}_i^t + c \sum_{j \in \mathcal{N}_i} (\mathbf{z}_{i,j}^-)^t + \sum_{j \in \mathcal{N}_i} (\mathbf{z}_{i,j}^+)^t \right)^T, \text{vec} \left(c \left(\mathbf{p}_i^t + (\mathbf{z}_{i,j}^-)^t \right), j \in \mathcal{N}_i \right)^T, \right. \\ \left. \text{vec} \left(\mathbf{p}_i^t + (\mathbf{z}_{i,j}^+)^t, j \in \mathcal{N}_i \right)^T \right]^T. \quad (43)$$

By substituting (42)-(43) into (35) yields the following update

$$\tilde{\mathbf{p}}_i^{t+1} = \frac{1}{2(c+1)N_i} \sum_{j \in \mathcal{N}_i} \left[d_{i,j} \mathbf{u}_{i,j}^t - \boldsymbol{\lambda}_{i,j}^t + c \left(\mathbf{p}_i^t + (\mathbf{z}_{i,j}^-)^t \right) + \mathbf{p}_i^t + (\mathbf{z}_{i,j}^+)^t \right], \quad (44)$$

$$(\tilde{\mathbf{z}}_{i,j}^-)^{t+1} = \frac{1}{2c} \boldsymbol{\lambda}_{i,j}^t + \frac{1}{2} \left(\mathbf{p}_i^t + (\mathbf{z}_{i,j}^-)^t \right), \quad (45)$$

$$(\tilde{\mathbf{z}}_{i,j}^+)^{t+1} = -\frac{1}{2} d_{i,j} \mathbf{u}_{i,j}^t + \frac{1}{2} \left(\mathbf{p}_i^t + (\mathbf{z}_{i,j}^+)^t \right). \quad (46)$$

Using once again the form of \mathbf{Q}_i , \mathbf{D}_i and \mathbf{A}_i , it follows from the (34) and (38) that

$$\tilde{\mathbf{u}}_{i,j}^{t+1} = \mathbf{u}_{i,j}^t + \frac{d_{i,j}}{\rho} \left(\mathbf{p}_i^{t+1} - (\mathbf{z}_{i,j}^+)^{t+1} \right), \\ \mathbf{u}_{i,j}^{t+1} = \frac{1}{\max\{1, \|\tilde{\mathbf{u}}_{i,j}^{t+1}\|\}} \tilde{\mathbf{u}}_{i,j}^{t+1}, \quad (47)$$

$$\boldsymbol{\lambda}_{i,j}^{t+1} = \boldsymbol{\lambda}_{i,j}^t + c \left(\mathbf{p}_i^{t+1} - (\mathbf{z}_{i,j}^-)^{t+1} \right). \quad (48)$$

Next, we intend to remove $\tilde{\mathbf{z}}_i^{t+1}$ to reduce the storage space and computation required by Algorithm

1. Combining (30) with (36) and (44), we have

$$\mathbf{p}_i^{t+1} = \begin{cases} \frac{\sum_{j \in \mathcal{N}_i} (2d_{i,j} \mathbf{u}_{i,j}^t - 2\boldsymbol{\lambda}_{i,j}^t + \boldsymbol{\alpha}_{i,j}^t + \boldsymbol{\beta}_{i,j}^t)}{2(c+1)N_i}, & \text{if } i \notin \mathcal{A}, \\ \mathbf{a}_i, & \text{if } i \in \mathcal{A}. \end{cases} \quad (49)$$

where

$$\boldsymbol{\alpha}_{i,j}^t := \boldsymbol{\lambda}_{i,j}^t + c \left(\mathbf{p}_i^t + (\mathbf{z}_{i,j}^-)^t \right), \quad (50)$$

$$\boldsymbol{\beta}_{i,j}^t := -d_{i,j} \mathbf{u}_{i,j}^t + \mathbf{p}_i^t + (\mathbf{z}_{i,j}^+)^t. \quad (51)$$

By adding $\frac{1}{c}(\tilde{\mathbf{z}}_{j,i}^+)^{t+1}$ to the both side of (45) and using (31), we get

$$\frac{c+1}{c} (\mathbf{z}_{i,j}^-)^{t+1} = \frac{1}{2c} \boldsymbol{\lambda}_{i,j}^t + \frac{1}{2} \left(\mathbf{p}_i^t + (\mathbf{z}_{i,j}^-)^t \right) + \frac{1}{c} (\tilde{\mathbf{z}}_{j,i}^+)^{t+1}. \quad (52)$$

Then, substituting (46) into (52) and using (50)-(51), it yields

$$(\mathbf{z}_{i,j}^-)^{t+1} = \frac{1}{2(c+1)} \left(\boldsymbol{\alpha}_{i,j}^t + \boldsymbol{\beta}_{j,i}^t \right). \quad (53)$$

Similarly, we add $c(\tilde{\mathbf{z}}_{j,i}^-)^{t+1}$ to (46), then using (32) and (50)-(51) to obtain

$$(\mathbf{z}_{i,j}^+)^{t+1} = \frac{1}{2(c+1)} \left(\boldsymbol{\beta}_{i,j}^t + \boldsymbol{\alpha}_{j,i}^t \right). \quad (54)$$

Algorithm 2: Simplified SP-ADMM algorithm

Input: parameter c, ρ , and initial values of $\mathbf{z}^0 \in \mathcal{Z}$, $\mathbf{u}_{i,j}^0, \boldsymbol{\lambda}_{i,j}^0 = \mathbf{0}$, $i \in N, j \in \mathcal{N}_i$.

Compute $\{(\boldsymbol{\alpha}_{i,j}^0, \boldsymbol{\beta}_{i,j}^0)\}_{i \in N, j \in \mathcal{N}_i}$ defined in (50)-(51)

for $t \in \{0, \dots, T\}$ **do**

communications:

\Rightarrow Broadcast $\boldsymbol{\alpha}_{i,j}^t, \boldsymbol{\beta}_{i,j}^t$ to region $j \in \mathcal{N}_i$,

\Leftarrow Receive $\boldsymbol{\alpha}_{j,i}^t, \boldsymbol{\beta}_{j,i}^t$ from region $j \in \mathcal{N}_i$.

for client $i = 1, 2, \dots, N$ **in parallel do**

 Update \mathbf{p}_i^{t+1} via (49)

 Update $\mathbf{u}_{i,j}^{t+1}$ via (55)

 Update $\boldsymbol{\beta}_{i,j}^{t+1}$ via (56)

 Update $\boldsymbol{\alpha}_{i,j}^{t+1}$ via (58)

 Update $\boldsymbol{\lambda}_{i,j}^{t+1}$ via (57)

end

end

Applying (54) to (47) and (51), we get

$$\begin{aligned} \tilde{\mathbf{u}}_{i,j}^{t+1} &= \mathbf{u}_{i,j}^t + \frac{d_{i,j}}{\rho} \mathbf{p}_i^{t+1} - \frac{d_{i,j}}{2\rho(c+1)} (\boldsymbol{\beta}_{i,j}^t + \boldsymbol{\alpha}_{j,i}^t), \\ \mathbf{u}_{i,j}^{t+1} &= \frac{1}{\max\{1, \|\tilde{\mathbf{u}}_{i,j}^{t+1}\|\}} \tilde{\mathbf{u}}_{i,j}^{t+1}, \end{aligned} \quad (55)$$

$$\boldsymbol{\beta}_{i,j}^{t+1} = -d_{i,j} \mathbf{u}_{i,j}^{t+1} + \mathbf{p}_i^{t+1} + \frac{1}{2(c+1)} (\boldsymbol{\beta}_{i,j}^t + \boldsymbol{\alpha}_{j,i}^t). \quad (56)$$

Substituting (53) into (48) and (50), we have

$$\boldsymbol{\lambda}_{i,j}^{t+1} = \boldsymbol{\lambda}_{i,j}^t + c\mathbf{p}_i^{t+1} - \frac{c}{2(c+1)} (\boldsymbol{\alpha}_{i,j}^t + \boldsymbol{\beta}_{j,i}^t), \quad (57)$$

$$\boldsymbol{\alpha}_{i,j}^{t+1} = \boldsymbol{\lambda}_{i,j}^t + 2c\mathbf{p}_i^{t+1}. \quad (58)$$

The simplified version of the algorithm is summarized in Algorithm 2. From the updating steps of Algorithm 2, it can be seen that each node only needs to store the following values: $\boldsymbol{\alpha}_{i,j}^t, \boldsymbol{\beta}_{i,j}^t, \mathbf{u}_{i,j}^t, \boldsymbol{\lambda}_{i,j}^t, d_{i,j}, j \in \mathcal{N}_i$ and c, ρ, N_i . Compared with the storage space required by node i of Algorithm 1, Algorithm 2 demands only $4nN_i + N_i + 3$ storage units, which is significantly reduced compared to that required

by Algorithm 1. We remark that the update (49) and (55)-(58) of Algorithm 2 should be interpreted as an improved version that optimizes the storage requirement of Algorithm 1. Although Algorithm 2 offers superior storage efficiency, the presentation of Algorithm 1 still has value for several reasons. Firstly, Algorithm 1 serves as a foundation for understanding the principles and techniques behind Algorithm 2. By presenting Algorithm 1 first, we can demonstrate the motivation and rationale for the modifications made in Algorithm 2. Additionally, the convergence analysis for Algorithm 1 is more direct since Algorithm 2 streamlines the steps of Algorithm 1.

Remark 4 (Complexity Comparison) *For each iteration of Algorithm 1, the computation cost is derived by (35) that involves the inversion of diagonal matrix $\mathbf{W}_i \in \mathbb{R}^{n(2N_i+1) \times n(2N_i+1)}$, which is independent of the iteration index t , and thus needs to be computed only once. Besides, equation (34) requires calculating the Euclidean norm of $\mathbf{u}_i \in \mathbb{R}^{nN_i \times 1}$. Hence, the computational complexity of the proposed SP-ADMM algorithm is $\mathcal{O}(nN_i)$, which is lower than SDP relaxation method $\mathcal{O}(n^3)$ and ADMM-H, but equal to SF and AM-FD. The communication cost per iteration per node is directly proportional to the number of scalar variables that sensors need to transmit to their neighboring nodes [9]. Thus, the AM-FD, SDP, and SF algorithms have a communication cost of nN_i , while the proposed SP-ADMM requires sending two variables to its neighbor, resulting in a communication cost per iteration of $2nN_i$. It should be noted that the proposed SP-ADMM has lower communication cost than ADMM-H since the latter also needs to transmit a penalty parameter to its neighboring nodes at each iteration. The detailed comparisons between the proposed SP-ADMM algorithm with the other methods have been shown in Table I.*

IV. CONVERGENCE ANALYSIS

In this section, we present a proof for the convergence of the proposed Algorithm 1, which consists of three main steps. First, we provide an upper bound [31] for the difference $\mathcal{L}(\mathbf{z}^{t+1}, \mathbf{u}^{t+1}, \boldsymbol{\lambda}^{t+1}) - \mathcal{L}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$. Then, we construct a potential function ζ^t based on the AL function $\mathcal{L}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$, which monotonically decreases with each iteration and has a lower bound. [Employing the Kurdyka-Łojasiewicz property \[32\]–\[36\]](#), we establish the global convergence of the sequence generated by Algorithm 1. Next, we introduce the optimal gap function $\mathcal{F}(\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda})$ to show that the proposed Algorithm 1 converges to both a KKT point of problem (15) and a critical point of the original problem (2). Lastly, we define the ε -solution and prove that the algorithm sequence converges with a sublinear rate.

First, we bound the update step of the AL function $\mathcal{L}(\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda})$ at each iteration.

Lemma 1 Suppose $c\mathbf{B}_i^T\mathbf{B}_i$ takes the form of (25), and let $\{(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t)\}$ be a sequence generated by Algorithm 1. Then, for all $t \geq 1$, we have

$$\begin{aligned}
& \mathcal{L}(\mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}, \mathbf{u}^{t+1}) - \mathcal{L}(\mathbf{z}^t, \boldsymbol{\lambda}^t, \mathbf{u}^t) \\
& \leq \sum_{i \in \mathcal{N}} \left[-\frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 - \frac{\rho}{2} \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\|^2 \right. \\
& \quad + \frac{3(N_{\max} + 1)}{c} \|\mathbf{Q}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i(\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 \\
& \quad + 3c(N_{\max} + 1) \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1} - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \\
& \quad \left. + 3(1+c)(1+N_{\max}) \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \right], \tag{59}
\end{aligned}$$

where $N_{\max} := \max\{N_i, i \in \mathcal{N}\}$.

Proof: See the Section 2 in the supplement material. ■

Obviously, the right-hand-side (rhs) of (59) is a sum of three positive terms, no matter how large c and ρ are, there is no guarantee that $\mathcal{L}(\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda})$ decreases at each iteration step. Thus, it cannot be used as a potential function. In search for an appropriate potential function, we add the constraint violation and the proximal term to the AL function, then a novel potential function is designed as follows

$$\varsigma^t = \sum_{i \in \mathcal{N}} \frac{c}{2} \left[\kappa_1 \|\mathbf{A}_i \tilde{\mathbf{z}}_i^t\|^2 + \kappa_2 \|\mathbf{A}_i \mathbf{z}_i^t\|^2 + \frac{\rho}{2c} \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2 + (\kappa_1 + \kappa_2) \|\mathbf{z}_i^t - \mathbf{z}_i^{t-1}\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \right] + \mathcal{L}(\mathbf{z}^t, \boldsymbol{\lambda}^t, \mathbf{u}^t), \tag{60}$$

where $\kappa_1, \kappa_2, c, \rho > 0$ are some positive constants, which can be determined by the following Lemma.

Lemma 2 Suppose the sequence $\{(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t)\}$ is generated by Algorithm 1 and $c\mathbf{B}_i^T\mathbf{B}_i$ takes the form of (25). Then we have the following

$$\begin{aligned}
\varsigma^{t+1} - \varsigma^t & \leq \sum_{i \in \mathcal{N}} \left[-\frac{1}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{W}_i}^2 - \frac{\kappa_1 - 1}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{Q}_i^T \mathbf{Q}_i}^2 - \frac{c(\kappa_1 - 1)}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \right. \\
& \quad - \frac{\rho}{4} \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\|^2 - \frac{c(\kappa_1 - 6(N_{\max} + 1))}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1} - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \\
& \quad - \frac{c\kappa_1 - 6(N_{\max} + 1)}{2c} \|\mathbf{Q}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i(\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 \\
& \quad - \frac{c\kappa_1 - 6(1+c)(N_{\max} + 1)}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \\
& \quad - \left(\frac{c\kappa_2 \cdot \tilde{\tau}_{\min}}{2N_{\text{sum}}n(c+1)^2} - \frac{(N_{\max} + 1)c\kappa_1}{2} \right) \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|^2 \\
& \quad \left. - \left(\frac{\rho}{4} - d_{\max}^2(\kappa_1 + \kappa_2) \right) \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2 \right], \tag{61}
\end{aligned}$$

where $N_{\text{sum}} := \sum_{i \in \mathcal{N}} N_i$ is the total number of neighboring nodes, $d_{\max} := \max\{d_{i,j}, i \in \mathcal{N}, j \in \mathcal{N}_i\}$ is the maximum measurement distance and $\tilde{\tau}_{\min} := \min\{(c+1)^2 N_i^2 + c^2 N_i + N_i, i \in \mathcal{N}\}$.

Proof: See the Section 3 in the supplement material. ■

From the above analysis, it is clear that as long as κ_1, κ_2 and ρ are sufficiently large, the rhs of (61) is less than zero. As such, the potential function ς^t decreases at each iteration of SP-ADMM. Below, we derive the precise bounds for κ_1, κ_2 and ρ . First, following the first five rows on the rhs of (61), a sufficient condition for κ_1 is given below (for any given $c > 0$)

$$\kappa_1 \geq 6(N_{\max} + 1) \left(1 + \frac{1}{c}\right). \quad (62)$$

Second, for any given c and κ_1 , according to the sixth row on the rhs of (61), we need

$$\kappa_2 \geq \frac{N_{\text{sum}} n (c + 1)^2 (N_{\max} + 1) \kappa_1}{\tilde{\tau}_{\min}}. \quad (63)$$

Finally, given c, κ_1, κ_2 , and based on the last row on the rhs of (61), parameter ρ requires to satisfy

$$\rho \geq 4d_{\max}^2 (\kappa_1 + \kappa_2). \quad (64)$$

We conclude that if (62)-(64) are satisfied, then the potential function ς^t will decrease at every iteration.

Now, we are ready to establish the main result. To this end, let us define the function $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ as the optimal gap of problem (15) given by

$$\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) = \sum_{i \in \mathcal{N}} \left[\|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t - (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t))\|^2 + \|\mathbf{A}_i \mathbf{z}_i^t\|^2 + \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2 \right]. \quad (65)$$

Based on the function $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ defined above, Lemma 3 establishes its relationship with a KKT point of problem (15) and a critical point of problem (6).

Lemma 3 *When $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) = 0$, then $(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ is a KKT solution of problem (15) satisfying:*

$$\mathbf{z}^t \in \arg \min_{\substack{\mathbf{z} \in \mathcal{Z} \\ \mathbf{z} \in \mathcal{X}}} \sum_{i \in \mathcal{N}} F_i(\mathbf{z}_i, \mathbf{u}_i^t) + \langle \boldsymbol{\lambda}_i^t, \mathbf{A}_i \mathbf{z}_i \rangle, \quad (66a)$$

$$\mathbf{0} \in \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \partial \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i^t), \quad (66b)$$

$$\mathbf{0} = \mathbf{A}_i \mathbf{z}_i^t. \quad (66c)$$

Moreover, $(\mathbf{p}^t, \mathbf{u}^t)$ is a critical point of problem (6), and \mathbf{p}^t is a critical point of the original problem (2).

Proof: See the Section 4 in the supplement material. ■

Lemma 3 suggests that we can demonstrate the sublinear convergence of Algorithm 1 to a critical point of the original problem (2) by proving the sublinear convergence of the sequence $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}$ to a KKT

stationary point of problem (15). This convergence can be illustrated by invoking the following definition and lemma.

Definition 1 Given problem (15), we define $(\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda})$ as an ε -solution if $\|\mathbf{A}_i \mathbf{z}_i\| < \varepsilon$ and there exists a vector $\mathbf{v}_1 \in \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i, \mathbf{u}_i) + \mathbf{A}_i \boldsymbol{\lambda}_i + \partial \delta_{\mathcal{X} \times \mathcal{Z}}(\mathbf{z}_i)$ and a vector $\mathbf{v}_2 \in \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i, \mathbf{u}_i) + \partial \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i)$ such that $\|\mathbf{v}_1\| \leq \varepsilon, \|\mathbf{v}_2\| \leq \varepsilon$.

Lemma 4 Let parameters c and ρ satisfy (62)-(64), then for any $t > 1$, there exists a constant $M > 0$ such that we can find an $s \in \{1, 2, \dots, t-1\}$ satisfying $(\mathbf{z}_i^{s+1}, \mathbf{u}_i^{s+1}, \boldsymbol{\lambda}_i^{s+1})$ being a $M/\sqrt{t-1}$ -solution. This implies that an ε -solution can be obtained within M^2/ε^2 iterations.

Proof: See the Section 5 in the supplement material. ■

Theorem 1 For each nodes $i \in \mathcal{N}$, suppose the sequence $\{(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t)\}$ is generated by Algorithm 1 and the conditions (62)-(64) are satisfied by properly setting the parameters. Then we have

- **Lower Bound:**

$$\exists \underline{\zeta} > -\infty \text{ s.t. } \zeta^t \geq \underline{\zeta}, \forall t > 0.$$

- **Eventual Consensus:**

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{z}_i^{t+1} - \mathbf{z}_i^t &\rightarrow \mathbf{0}, & \lim_{t \rightarrow \infty} \boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t &\rightarrow \mathbf{0}, \\ \lim_{t \rightarrow \infty} \mathbf{u}_i^{t+1} - \mathbf{u}_i^t &\rightarrow \mathbf{0}, & \lim_{t \rightarrow \infty} \mathbf{A}_i \mathbf{z}_i^t &\rightarrow \mathbf{0}. \end{aligned}$$

- **Global Convergence:** the sequence $\{\mathbf{y}^t = (\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}$ has a finite length, i.e., $\sum_{t=1}^{\infty} \|\mathbf{y}^{t+1} - \mathbf{y}^t\| < \infty$, and it converges to a KKT point of problem (15).
- **Convergence to Stationary Points³:** $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) \rightarrow \mathbf{0}$ and the iteration sequence $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}$ converges to a KKT stationary point of problem (15).
- **Sublinear Convergence Rate:** For any given $\varepsilon_1 > 0$, suppose that $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ in (65) is less than ε_1 for the first time in the T -th iteration step, i.e.,

$$T := \arg \min_t \mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) \leq \varepsilon_1.$$

Then there exists a positive constant $\varepsilon_2 > 0$ (the specific form is defined in the (S.119) of the supplement material) such that $\varepsilon_1 \leq \frac{\varepsilon_2}{T-1}$, which means the convergence rate of function $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ is $\mathcal{O}(1/T)$.

³Given the linear constraint and nonsmooth term in the nonconvex optimization problem (15), we refer to the points that satisfy the KKT condition as the stationary points (or KKT points) of the problem (15), such as [26], [37]–[40].

Proof: See the Section 6 in the supplement material. ■

To our knowledge, Theorem 1 is the first result that shows the $\mathcal{O}(1/T)$ convergence rate of the distributed ADMM algorithm for the nonconvex and nonsmooth localization problem. Since Algorithm 2 is a simplified version of Algorithm 1 in Section III, the convergence result also holds for Algorithm 2. In the next section, numerical results will corroborate that the SP-ADMM algorithms can demonstrate more favorable convergence behavior than the benchmark methods.

V. NUMERICAL RESULTS

We consider the two-dimensional sensor network localization problem and evaluate the performance of the proposed Algorithm 1 against several other state-of-the-art methods, including SDP [10], SF [18], AM-FD [20] and ADMM-H [19]. We remark that the proposed Algorithm 2 only simplifies the storage space of Algorithm 1. Therefore, there is no difference in positioning accuracy between Algorithm 1 and Algorithm 2. Note also that the first two methods employ convex relaxation, the third one solves the original nonconvex problem directly, and the last one is a hybrid convex/nonconvex solver. Following [20], we will also examine the performance of SP-ADMM-NAG50 and AM-FD-NAG50, where SP-ADMM-NAG50 means that we run 50 steps of the NAG method [41] to obtain a good initial point, after that, the proposed SP-ADMM is used until convergence. For fairness, AM-FD-NAG50 uses the same initialization.

The criteria under which we compare the above algorithms are the practical running time counted as the maximal computation time among all parallel computing components and the averaged root mean squared error (RMSE) in a particular iteration, as described in [19], [23],

$$\text{RMSE}(t) = \sqrt{\sum_{i \in (\mathcal{N}/\mathcal{A})} \frac{\|\mathbf{p}_i^t - \mathbf{p}_i\|^2}{N - m}},$$

where \mathcal{N}/\mathcal{A} represents the set of agents with unknown positions, which has $N - m$ elements, \mathbf{p}_i is the true position of node i and \mathbf{p}_i^t is the estimated position of node i in the t -th iteration of the algorithm. In addition, to check the convergence of the proposed algorithm as stated in Theorem 1, let

$$\begin{aligned} S(t) &= \sum_{i \in \mathcal{N}} \|\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t\|^2, \\ U(t) &= \sum_{i \in \mathcal{N}} \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2, \\ P(t) &= \sum_{i \in \mathcal{N}} \|\mathbf{A}_i \mathbf{z}_i^t\|^2, \end{aligned} \tag{67}$$

TABLE II
BENCHMARK NETWORK AND ALGORITHM SETUP

Network Parameters					Method Parameters						
					ADMM-H						AM-FD
N	m	C_{range}	σ_{add}	D_{avg}	ϵ_c	ζ_c	τ_c	θ_c	δ_c	λ_{max}	\mathbf{u}^0
500	10	0.3	0.02	14.15	0.002	0.25	0.008	0.98	1.01	10^3	$\mathbf{0}$
1000	20	0.1	0.007	11.09	0.003	0.07	0.002	0.98	1.01	10^3	$\mathbf{0}$

where $S(t)$ and $U(t)$ are defined as the stationarity gap while $P(t)$ is defined as the feasibility gap for problem (15). The MATLAB code that implements the proposed SP-ADMM is available at: <https://github.com/zm-stu/SP-ADMM>.

A. Benchmark Network

In the experiment, various methods are applied to the benchmark network data collected by the Stanford Computational Imaging Lab [42]. Table II lists the details of the two networks and the specific parameters used for both the ADMM-H method and AM-FD method, which refer to the settings in [19] and [20]. C_{range} and $D_{\text{avg}} := \frac{1}{N} \sum_{i \in \mathcal{N}} N_i$ in Table II represent the communication range and the average number of neighbors, respectively. For the proposed algorithm, parameters are set to $\rho = c = 0.11$, $\mathbf{u}^0 = \mathbf{0}$ for $N = 500$ networks and $\rho = c = 0.0197$, $\mathbf{u}^0 = \mathbf{0}$ for $N = 1000$ networks. The initial point \mathbf{z}^0 for both networks is selected from a uniform distribution, $\mathbf{Unif}(-1, 1)^{n(4|\mathcal{E}|+N)}$. Following [15], we consider two different kinds of measurement noise: one is an additive white Gaussian noise (AWGN) with standard deviation $\sigma_{i,j} = \sigma_{\text{add}}$, while the other is the range dependent Gaussian noise, namely a zero mean Gaussian distribution with range dependent variance $\sigma_{i,j}^2 = \sigma_{\text{add}} \|\mathbf{p}_i - \mathbf{p}_j\|^2$. The purpose of including the range-dependent Gaussian noise experiment is to evaluate the robustness of our proposed algorithm.

Convergence performance: Let us first examine the convergence behaviors of the proposed algorithm, SDP [10], SF [18], AM-FD [20], and ADMM-H [19]. Fig. 2(a) and Fig. 2(c) display the RMSE versus the iteration number when we use AWGN and range-dependent Gaussian noise, respectively. As shown in Fig. 2(a) that ADMM-H based on a two-stage approach performs the best and its RMSE approaches the Cramer–Rao Lower Bound (CRLB) [43], but it is less advantageous in terms of the running time (see Table III). We note that the AM-FD-NAG50 and SP-ADMM-NAG50 show lower RMSE than both SF

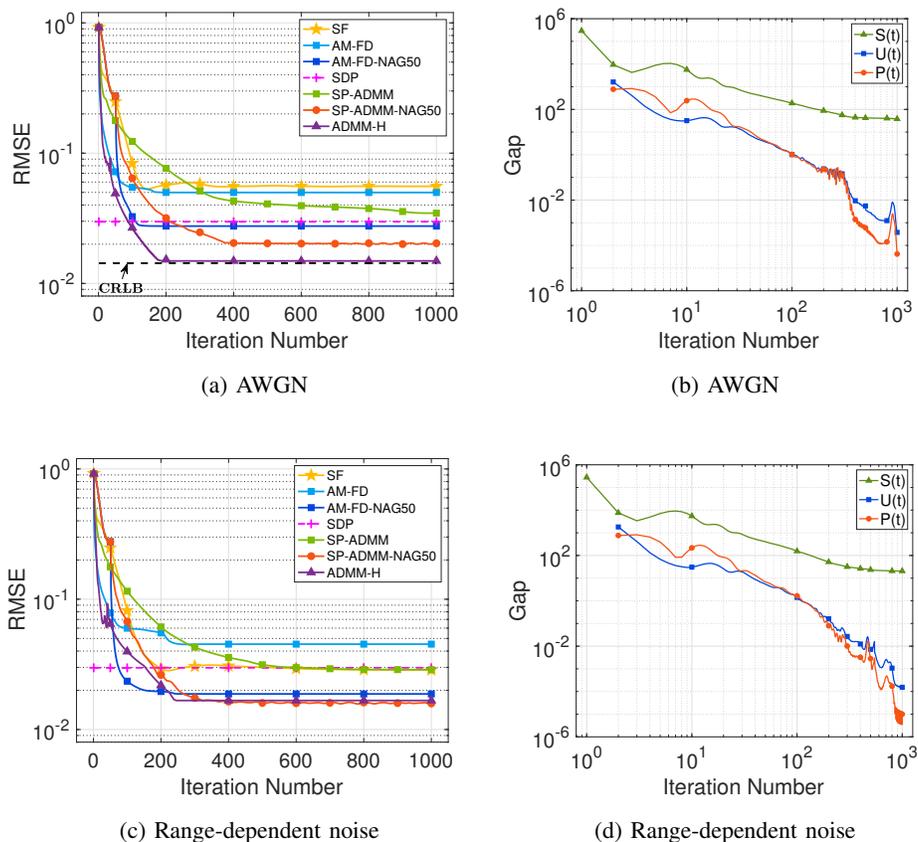
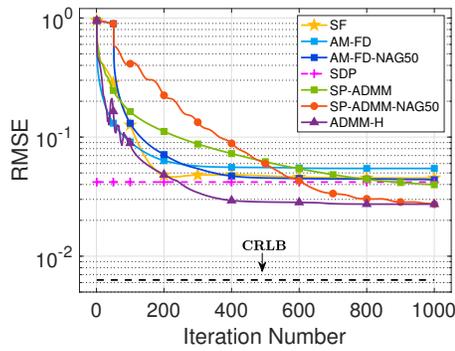


Fig. 2. Performance with benchmark network ($N = 500, m = 10$). Measurement noise: AWGN (first row) and Range-dependent Gaussian noise $\sigma_{i,j}^2 = \sigma_{\text{add}} \|\mathbf{p}_i - \mathbf{p}_j\|^2$ (second row). RMSE value (left column); feasibility gap and stationarity gap (right column).

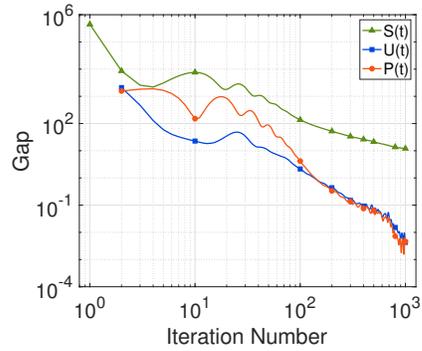
and SDP, illustrating that non-relaxed problems generally result in better location estimation than that of the relaxed problems. When we focus on the first-order methods, Fig. 2(a) and Fig. 2(c) demonstrate that the SP-ADMM-NAG50 algorithm achieves higher accuracy with fewer iterations than other methods for both AWGN and range-dependent Gaussian noise scenarios. Furthermore, we see that AM-FD-NAG50 outperforms AM-FD, which reflects the advantage of using the NAG method to provide an initial value that is beneficial for the nonconvex localization problem (2).

In Fig. 3, we plot the RMSE versus the iteration number for a large network with ($N = 1000, m = 20$). Further comparing Fig. 3(a) (Fig. 3(c)) with Fig. 2(a) (Fig. 2(c)), we observe that the performance of all the methods is degraded. Intriguingly, Fig. 3(c) shows that the proposed method achieves the best performance under range-dependent Gaussian noise. It implies that the proposed method may be more suitable for a large network in reality, possibly due to the noise and error elasticity of ADMM [9].

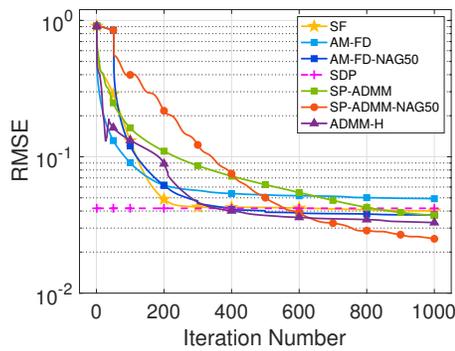
Gap: To further examine the convergence of the SP-ADMM algorithm, we plot in Fig. 2(b) (Fig.



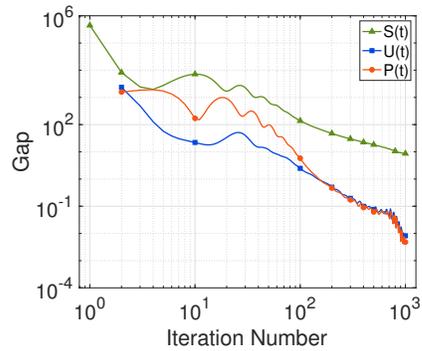
(a) AWGN



(b) AWGN



(c) Range-dependent noise



(d) Range-dependent noise

Fig. 3. Performance with benchmark network ($N=1000$, $m=20$). Measurement noise: AWGN (first row) and Range-dependent Gaussian noise $\sigma_{i,j}^2 = \sigma_{\text{add}} \|\mathbf{p}_i - \mathbf{p}_j\|^2$ (second row). RMSE value (left column); feasibility gap and stationarity gap (right column).

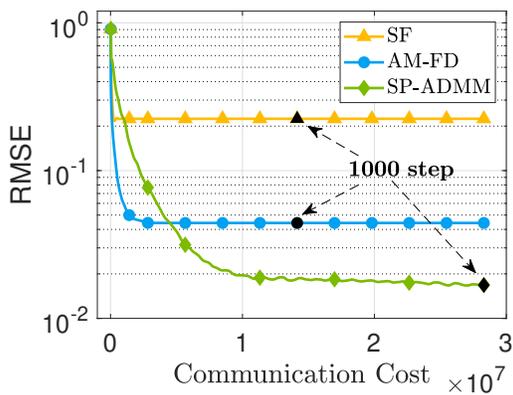
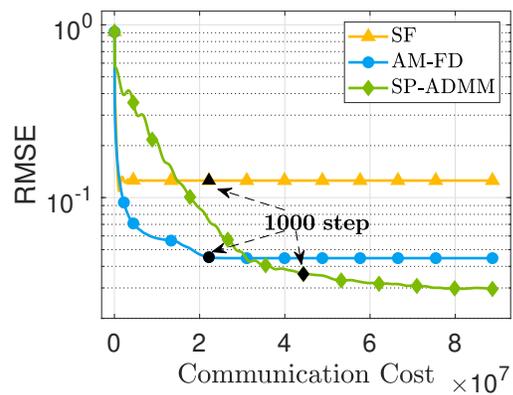
(a) $N = 500, m = 10$ (b) $N = 1000, m = 20$

Fig. 4. Comparison of RMSE and communication cost for different methods in the benchmark network with AWGN noise.

TABLE III
COMPARISONS OF RUNNING TIME

Algorithm	RMSE (AWGN)	Run time (seconds)	
		Parallelized	Sequential(Per Step)
Benchmark Network ($N = 500$, step = 1000, CRLB ≈ 0.014)			
SF	9.85e02	0.05401	1.24e03 (1.2428)
AM-FD	4.99e-02	- ^a	6.19e02 (0.6193)
AM-FD-NAG50	2.75e-02	-	6.75e02 (0.6753)
SDP	2.97e-02	-	1.98e02 (6.1815)
ADMM-H	1.49e-02	0.1518	3.50e03 (3.4957)
SP-ADMM-NAG50	2.03e-02	0.01091	6.64e02 (0.6642)
SP-ADMM	2.98e-02	0.01784	6.30e02 (0.6298)
Benchmark Network ($N = 1000$, step = 1000, CRLB ≈ 0.005)			
SF	7.41e-02	0.045	6.09e03 (6.085)
AM-FD	5.07e-02	-	1.43e02 (0.1430)
AM-FD-NAG50	4.17e-02	-	1.45e02 (0.1455)
SDP	4.19e-02	-	1.56e03 (23.9601)
ADMM-H	2.73e-02	0.152	1.06e04 (10.5978)
SP-ADMM-NAG50	2.94e-02	0.034	9.63e01 (0.0965)
SP-ADMM	3.73e-02	0.033	9.04e01 (0.0905)

^a These symbols “-” in the table mean that the methods cannot be implemented in parallel.

2(d)) and Fig. 3(b) (Fig. 3(d)) the curves of the optimal gap of problem (15) versus the iteration number. Observe explicitly that the performance gap of our proposed SP-ADMM algorithm reduces as the iteration number increases. This result verifies the efficacy of the Theorem 1, which states that the iterative sequence $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}$ generated by Algorithm 1 converges to a KKT stationary point of problem (15) and the convergence rate is $\mathcal{O}(1/T)$.

Running time: Table III shows the running time of the methods in Fig. 2 and Fig. 3. It can be seen that the proposed SP-ADMM algorithm is most efficient in computation time than other methods. Note that while SF has a competitive running time compared to ours, its RMSE is bigger than ours. In addition, although the SDP solver here uses SeDuMi [44], it still consumed a lot of computational time, especially when the network size N is large. SDP terminates when the norm of the constraint gap reaches the order of 10^{-7} , while the other methods perform 1000 iterations. In contrast, AM-FD and AM-FD-NAG50 require modestly increased computation time as the network size grows. Nevertheless, compared

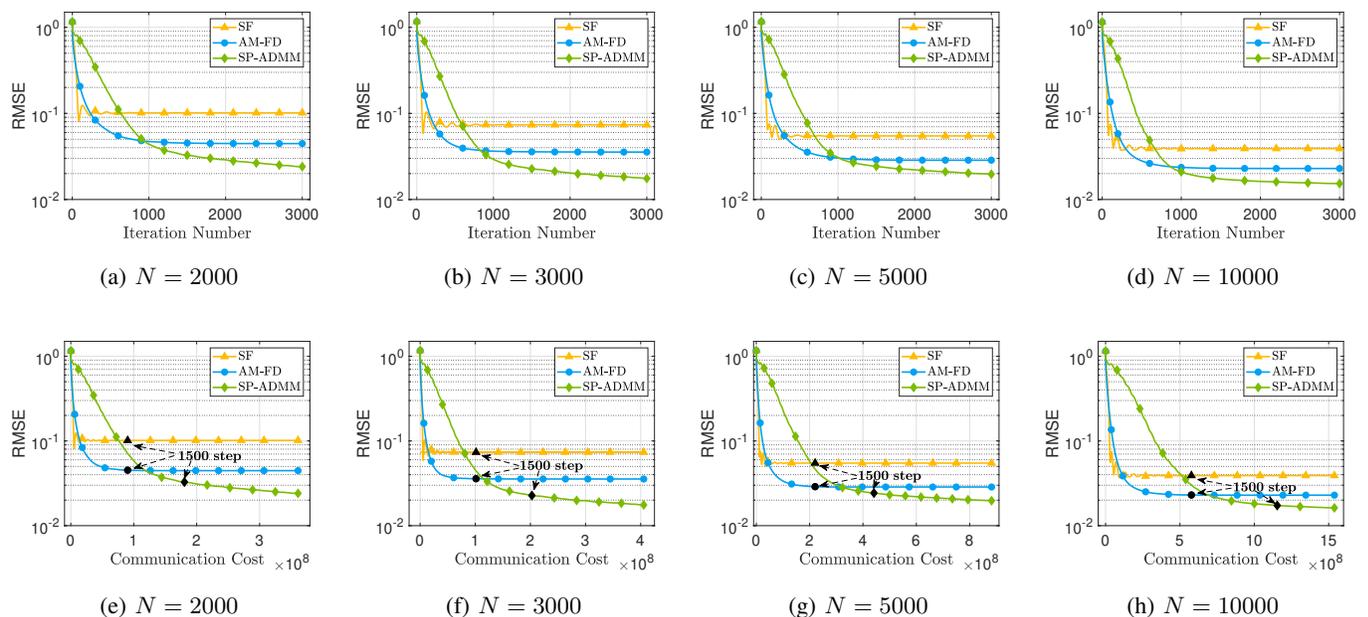


Fig. 5. Convergence performance of different methods with various sizes. (a)-(d) RMSE value vs. iteration number; (e)-(h) RMSE value vs. communication cost.

to other parallel methods, AM-FD and AM-FD-NAG50 are not the most sensible choices for large-scale networks and networks with a large average number of neighboring nodes, causing extra delays by the sequential structure under distributed networks. As a result, the structure of parallel implementation would be more efficient. Lastly, we note that the application of the ADMM-H method to large-scale networks is impractical due to the unaffordable sequential running time. Moreover, it can be seen from Table II that the ADMM-H has more parameters to adjust for different networks.

Communication cost: To evaluate the efficiency of various communication methods, we performed a comparison of the RMSE of different methods under benchmark networks as communication costs increased, as depicted in Fig. 4. Our results reveal that although the proposed SP-ADMM incurs a higher communication cost than the SF and AM-FD algorithms during the same iteration (i.e., 1000 steps), it achieves higher accuracy at the same communication cost once convergence has been reached. Additionally, Fig. 4 also indicates that the AM-FD and SF methods were incapable of meeting higher accuracy requirements, even with sufficient communication resources.

B. Synthesized Network

In this subsection, we evaluate the performance of the proposed SP-ADMM algorithm in comparison to ADMM-H [19], SDP [10], SF [18], and AM-FD [20] under varying factors that could potentially

TABLE IV
SYNTHESIZED NETWORK AND ALGORITHM SETUP

Network Parameters					Method Parameters						
					ADMM-H						AM-FD
N	m	C_{range}	σ_{add}	D_{avg}	ϵ_c	ζ_c	θ_c	δ_c	λ_{max}	τ_c	\mathbf{u}^0
Random (Changing m)											
495	5	0.3	0.02	12.88							
515	20	0.3	0.02	16.48	0.002	0.25	0.008	0.98	1.01	10^3	$\mathbf{0}$
525	30	0.3	0.02	18.50							

impact localization accuracy.

1) **Influence of the Number of Sensors (N):** We initiated our investigation by evaluating the performance of random networks with sizes $N=2000, 3000, 5000,$ and 10000 . Similar to the benchmark networks, the anchor number was set to 2% of the network size, and the value of σ_{add} of AWGN was chosen to be 7% of the communication range. The results of our experiments are presented in Fig. 5, where Fig. 5(a)-(d) depicts RMSE versus iteration steps and Fig. 5(e)-(h) illustrates RMSE versus communication loss. Similar to the results obtained for the benchmark networks shown in Fig. 4, our findings demonstrate that the proposed SP-ADMM algorithm incurs a higher communication cost compared to the SF and AM-FD algorithms during the same number of iterations (i.e., 1500 steps). However, once convergence has been achieved, the SP-ADMM algorithm provides higher accuracy at the same communication cost. Moreover, Fig. 5(e)-(h) show that even with sufficient communication resources, the AM-FD and SF methods are incapable of meeting higher accuracy requirements.

2) **Influence of the Number of Anchors (m):** The parameters for the random networks and ADMM-H method are summarized in Table IV. The three networks of the experiment are formed by randomly removing 5 anchors, randomly generating 10 anchors, and randomly generating 20 anchors on the benchmark network ($N = 500, m = 10$). Fig. 6 explores the influence of the number of anchor nodes on positioning accuracy. In all three networks, the parameters of our algorithm were set to $\rho = c = 0.11, \mathbf{u}^0 = \mathbf{0}$, and \mathbf{z}^0 from a distribution $\mathbf{Unif}(-1, 1)^{n(4|\mathcal{E}|+N)}$. As expected, the localization accuracy of most methods improves as the number of anchors increases. This is due to the anchors know their true locations, thus can provide more accurate position estimates of the neighboring nodes. In addition, we can see that except for the ADMM-H method, the SP-ADMM-NAG works better than other

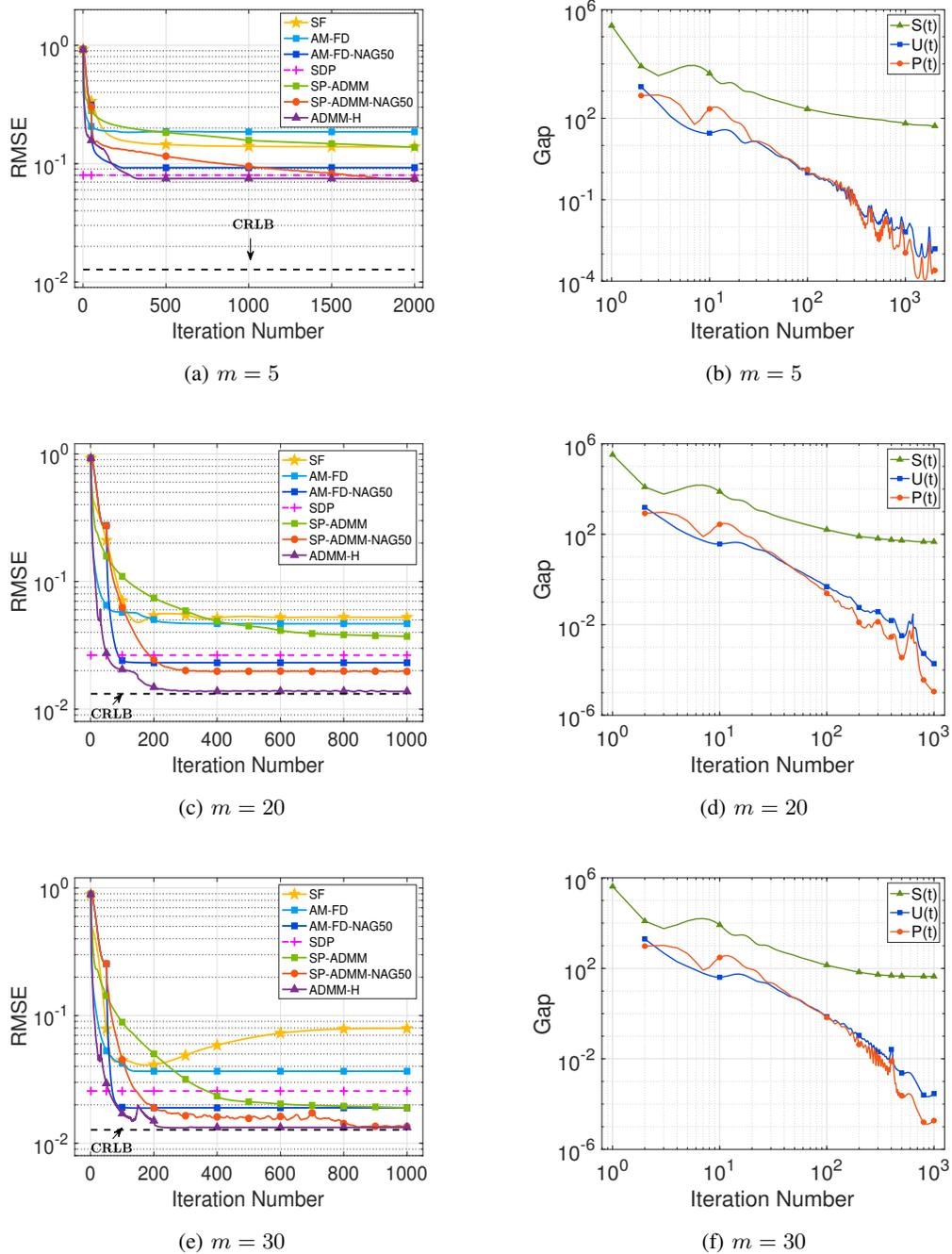


Fig. 6. Convergence performance under a different number of anchors with AWGN for the $N - m = 490$ nodes network. RMSE value (left column); primal feasibility and stationarity gap (right column).

methods and gets closer to CRLB when increasing the number of anchors. A closer inspection shows that for the AM-FD (light blue line), which is also a nonconvex relaxation method, the proposed SP-ADMM (green line) performs better even though it converges not so fast at the beginning. AM-FD-NAG50 (dark

TABLE V
COMPARISONS OF RUNNING TIME

Algorithm	RMSE	Run time (seconds)	
		Parallelized	Sequential(Per Step)
Random ($N = 495, m = 5, \text{step} = 2000, \text{CRLB} \approx 0.016$)			
SF	4.64e-01	0.0402	1.97e02(0.0984)
AM-FD	1.86e-01	- ^b	1.23e02 (0.0613)
AM-FD-NAG50	9.25e-02	-	1.24e02 (0.0623)
SDP	7.99e-02	-	1.59e02 (4.9609)
ADMM-H	7.47e-02	0.1412	6.27e03 (3.1338)
SP-ADMM-NAG50	7.45e-02	0.0256	1.16e02 (0.0581)
SP-ADMM	1.38e-01	0.0251	1.11e02 (0.0554)
Random ($N = 510, m = 20, \text{step} = 1000, \text{CRLB} \approx 0.013$)			
SF	9.40e-02	0.04971	1.39e03 (1.3411)
AM-FD	4.67e-02	-	7.20e01 (0.07202)
AM-FD-NAG50	2.31e-02	-	7.66e01 (0.0768)
SDP	2.65e-02	-	3.24e02 (10.4481)
ADMM-H	1.38e-02	1.31	3.56e03 (3.5575)
SP-ADMM-NAG50	1.97e-02	0.03232	8.10e01 (0.0881)
SP-ADMM	3.72e-02	0.0388	8.29e01 (0.0811)
Random ($N = 520, m = 30, \text{step} = 1000, \text{CRLB} \approx 0.013$)			
SF	7.94e-02	0.0411	1.81e03 (1.8142)
AM-FD	3.67e-02	-	7.51e01(0.0752)
AM-FD-NAG50	1.90e-02	-	8.20e01 (0.0821)
SDP	2.56e-02	-	5.57e02 (18.5801)
ADMM-H	1.33e-02	1.41	3.51e03 (3.5150)
SP-ADMM-NAG50	1.35e-02	0.0343	8.54e01 (0.0854)
SP-ADMM	1.90e-02	0.0351	8.51e01 (0.0811)

^b These symbols “-” in the table mean that the methods cannot be implemented in parallel.

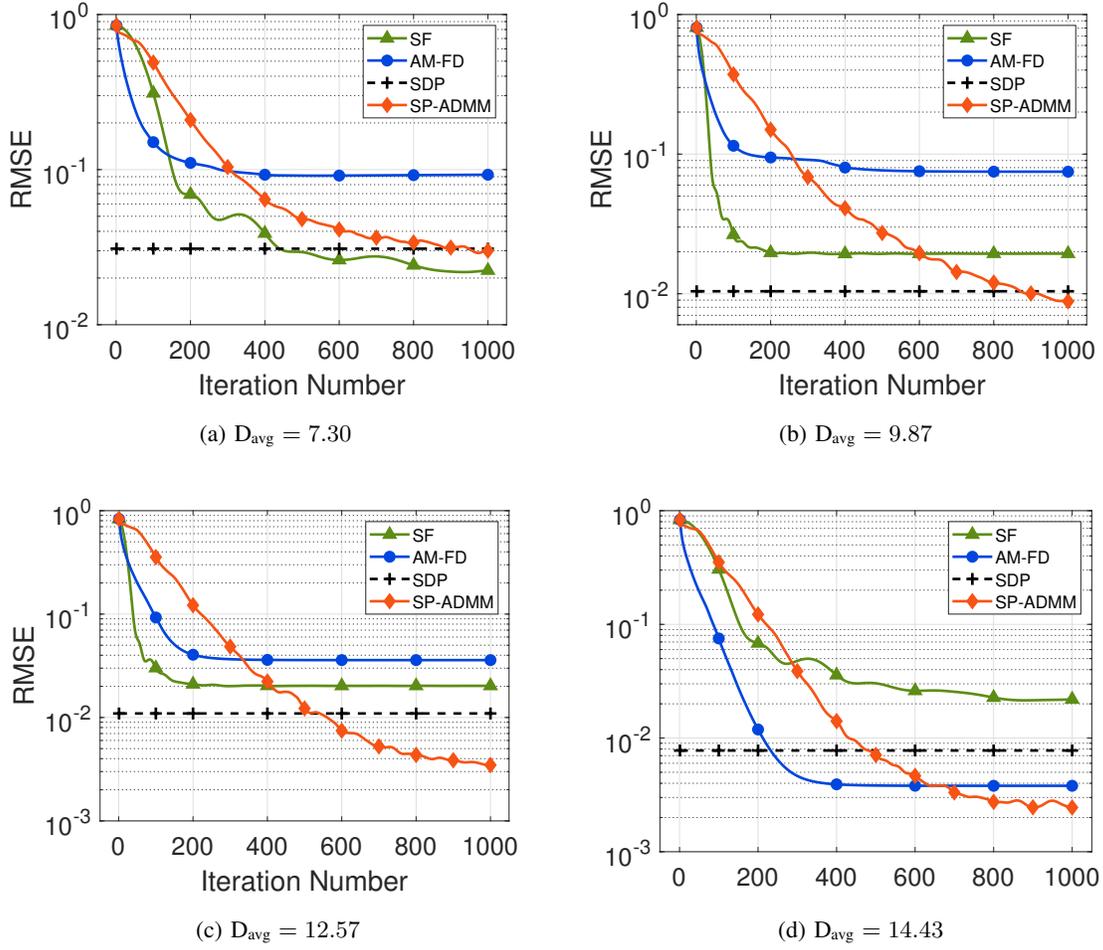


Fig. 7. Convergence performance under different average number of neighboring nodes (D_{avg}). Here, we adopt $\sigma_{\text{add}} = 0.02$; $N = 108, m = 8$.

blue line) and SP-ADMM-NAG50 (red line) also have such performance. The running time is shown in Table V, and we find that the SP-ADMM is most computationally time efficient due to the parallel implementation. The results here are similar to those in Table III and consistent with Table I.

3) **Influence of the Average Number of Neighbors (D_{avg}):** Fig. 7 investigates the influence of the average number of neighbors on the performance of four algorithms: the proposed SP-ADMM algorithm, SF, SDP, and AM-FD. The network ($N = 108, m = 8$) randomly placed in the 2D $[0, 1] \times [0, 1]$ area, and we set the parameters $\rho = c = 0.0265$, $\mathbf{z}^0 = \mathbf{0}$, and $\mathbf{u}^0 = 0.5 \cdot \mathbf{1}$. As shown in Fig. 7, the RMSE values of most algorithms decrease as the average number of neighbors increases. The reason may be that each node communicates with more neighbors at each iteration, which leads to the position estimated by different nodes for the same node reaching consistency faster. Notably, the SP-ADMM algorithm

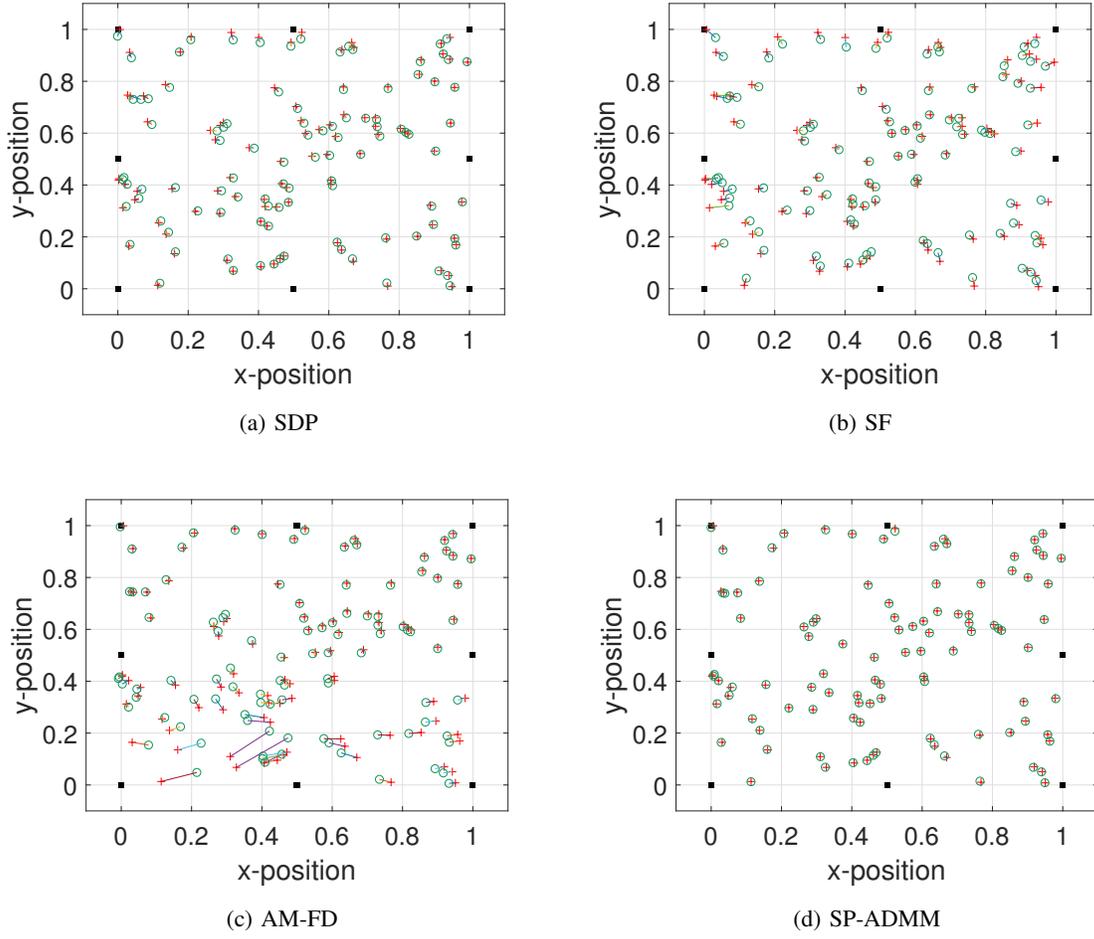


Fig. 8. Position estimates obtained by the SDP, SF, AM-FD, and the proposed SP-ADMM algorithm in 2-D plane. Here, we adopt $\sigma_{\text{add}} = 0.02$, $N = 108$, $m = 8$, $D_{\text{avg}} = 12.57$ ($C_{\text{range}} = 0.23$). Here, the anchors are marked by \blacksquare , the true agent positions by $+$ and the estimated ones by \circ .

consistently achieves the lowest RMSE in all scenarios. Fig. 8 visualizes the position estimates obtained in Fig. 7(c). It clearly shows that the positions estimated by our approach are very close to the true agents' locations.

4) **Influence of the Measurement Noise Variance** ($\sigma_{i,j}$): To test the robustness of the proposed algorithm against different measurement noise profiles, the localization scenarios are further extended to sensor networks with different measurement noise variances $\sigma_{i,j}^2 = \sigma_{\text{add}} \|\mathbf{p}_i - \mathbf{p}_j\|^2$. The corresponding convergence results we obtained are depicted in Fig. 9. Both the network and our proposed algorithm used the same parameters as in Fig. 7(c). Note that Fig. 7(c) shows the test results for $\sigma_{\text{add}} = 0.02$, while we show some other configurations in Fig. 9. Here, we have similar observations as in Fig. 7. Although the performance of these methods demonstrates certain loss with the increase of the noise variance, the

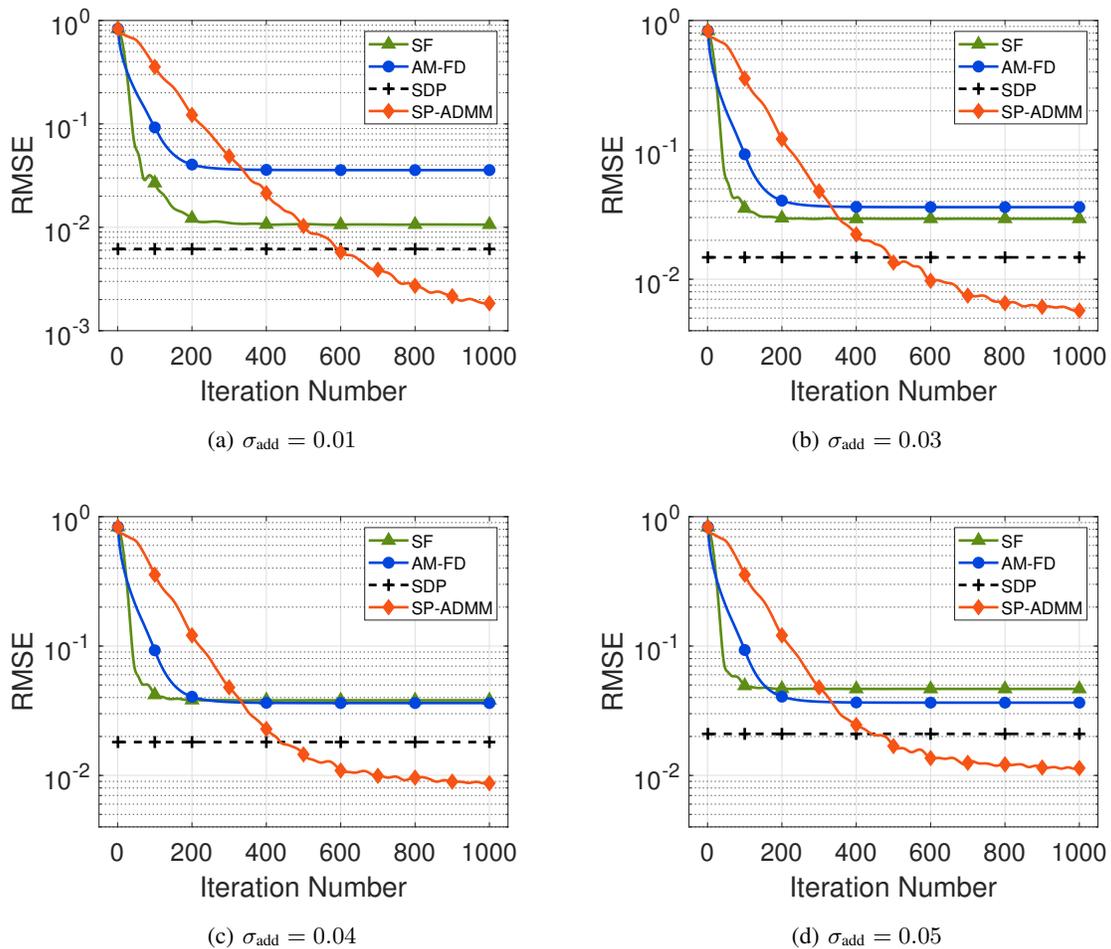


Fig. 9. Convergence performance under different measurement noise variances. Here, we let $N = 108$, $m = 8$; communication range is set to 0.23.

localization error of our proposed SP-ADMM algorithm is consistently lower in all scenarios. Moreover, one can see that the non-relaxed methods are relatively insensitive to measurement noise compared with the other convex relaxation methods. Perhaps the reason that the solution of the convex relaxation method is an approximation of the original problem.

We also conducted an experiment to investigate the influence of range-dependent noise on the positioning effectiveness of various methods in extreme cases. To this end, we randomly generated a set of 100 nodes and 8 anchors within a $[0, 1] \times [0, 1]$ square. The communication distance is set to 0.23, and the network topology is depicted in Fig. 10(a). The distance measurements between sensors are subject to Gaussian-distributed noise with zero mean and variance $\sigma_{i,j}^2 = \sigma_{\text{add}} \|\mathbf{p}_i - \mathbf{p}_j\|^2$. The initial positions are drawn from a uniform distribution $\text{Unif}(-1, 1)^{200}$, and we calculated the corresponding

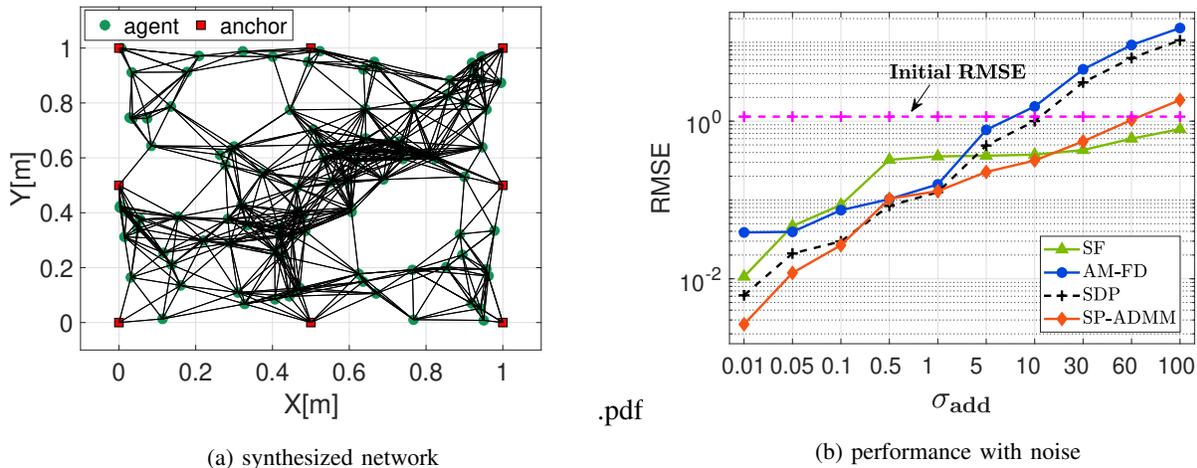


Fig. 10. Influence of range-dependent measurement noise $w_{i,j}$ on positioning accuracy across various methods. $w_{i,j} \sim \text{Gauss}(0, \sigma_{\text{add}} \|\mathbf{p}_i - \mathbf{p}_j\|^2)$

Root-Mean-Squared-Error (RMSE) value as follows:

$$\text{RMSE}(0) = \sqrt{\frac{1}{N_{MC}} \frac{1}{100} \sum_{k=1}^{N_{MC}} \sum_{i \in (\mathcal{N}/\mathcal{A})}^{100} (\mathbf{p}_i^0 - \mathbf{p}_i)^T (\mathbf{p}_i^0 - \mathbf{p}_i)},$$

where N_{MC} denotes the count of independent Monte Carlo trials, while $\mathbf{p}_i^{0,k}$ symbolizes the initial position vector of the i -th node in the k -th Monte Carlo trial. For Fig. 10, we set $N_{MC} = 50$. When the RMSE of the algorithm is larger than the initial RMSE, it means that the algorithm has failed to improve the accuracy of the sensor positions and may have even made it worse. From Fig. 10, one can observe that the proposed SP-ADMM failed to provide appropriate position estimates when the variance of the range-noise is bigger than 60. Additionally, Fig. 10 shows that the RMSEs of all methods increase as the noise variance becomes larger. Nevertheless, the proposed SP-ADMM shows the slowest performance degradation, indicating that the proposed method is more robust than AM-FD method and SDP method.

C. Evaluation of the Proposed SP-ADMM Algorithm Versus Penalty Parameters

We simulated the proposed SP-ADMM algorithm under different values of the penalty parameters c, ρ , as well as the initial value \mathbf{u}^0 . The different network parameters ($C_{\text{scale}}, C_{\text{range}}, N_{\text{max}} = \{N_i, i \in \mathcal{N}\}$) are also considered. Fig. 12 summarizes the RMSE after 1000 iterations. Fig. 11 and Fig. 12 not only show the sensitivity of the proposed SP-ADMM algorithm to parameters ρ, c , and \mathbf{u}^0 , but also provide practical advice that can assist users, using the proposed SP-ADMM algorithm, to reduce the time in adjusting parameters ρ, c, \mathbf{u}^0 on different networks.

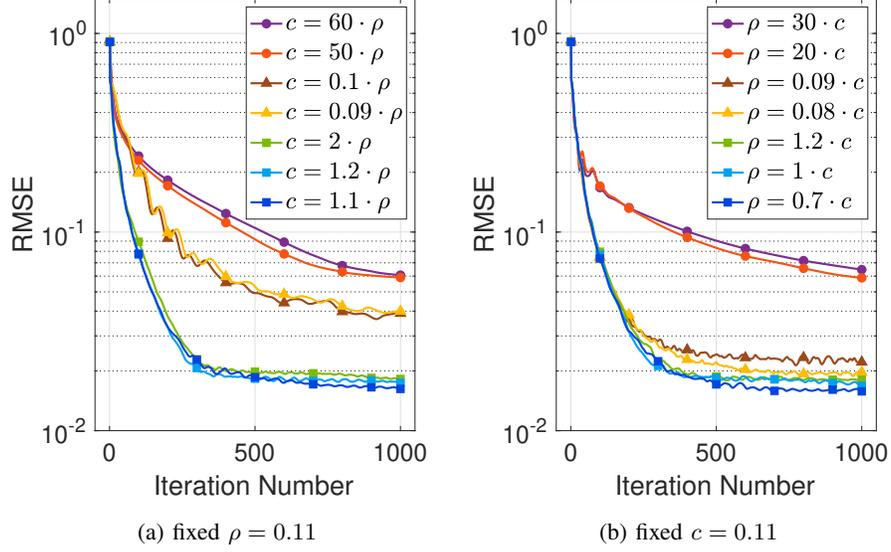


Fig. 11. RMSE curves achieved by our proposed methods with different penalty coefficients under the benchmark network ($N = 500, m = 10$) with AWGN noise.

1) **The Selection Strategy of ρ and c :** Theorem 1 guarantees that the proposed Algorithms will converge if ρ is selected in accordance with (62)-(64) for a given c . To corroborate this claim, we present in Fig. 11 the RMSE with iterations for different values of ρ , while maintaining $c = 0.1$ under the benchmark network ($N=500, m=10$). Our experimental results demonstrate that, in practical implementation, the selection of proximal penalty coefficients impacts the convergence rate and final accuracy of the proposed algorithm. Specifically, for a benchmark network with $N = 500$ and $m = 10$, we observed that choosing a smaller value of c or ρ within an appropriate range resulted in the fastest convergence and highest accuracy. For instance, as depicted in Fig. 11(a) (Fig. 11(b)), when c (ρ) is in the range of $[1.1\rho, 2\rho]$ ($[0.7c, 1.2c]$), the algorithm's performance improves with decreasing values of the parameter c . However, selecting a value of c or ρ that is too small or too large may significantly slow down the convergence rate and compromise the final accuracy, as evidenced by the yellow, brown, red, and purple lines in the two subfigures of Fig. 11. The appropriate range of the parameter values is typically selected empirically based on the specific data used.

2) **Localization Performance with u and c Selection:** Fig. 12(a) and Fig. 12(b) validate this statement. C_{scale} denotes the side length of the square area where the network is deployed, for example, $C_{\text{scale}} = 10$ indicates that the network deployed over $[0, 10] \times [0, 10]$ area. Fig. 12(a)-(b) employ the same sensor network as Figure , except for the variance of the measured noise $\sigma_{i,j}^2$, which is set to $\sigma_{i,j}^2 = 0.02\|\mathbf{p}_i -$

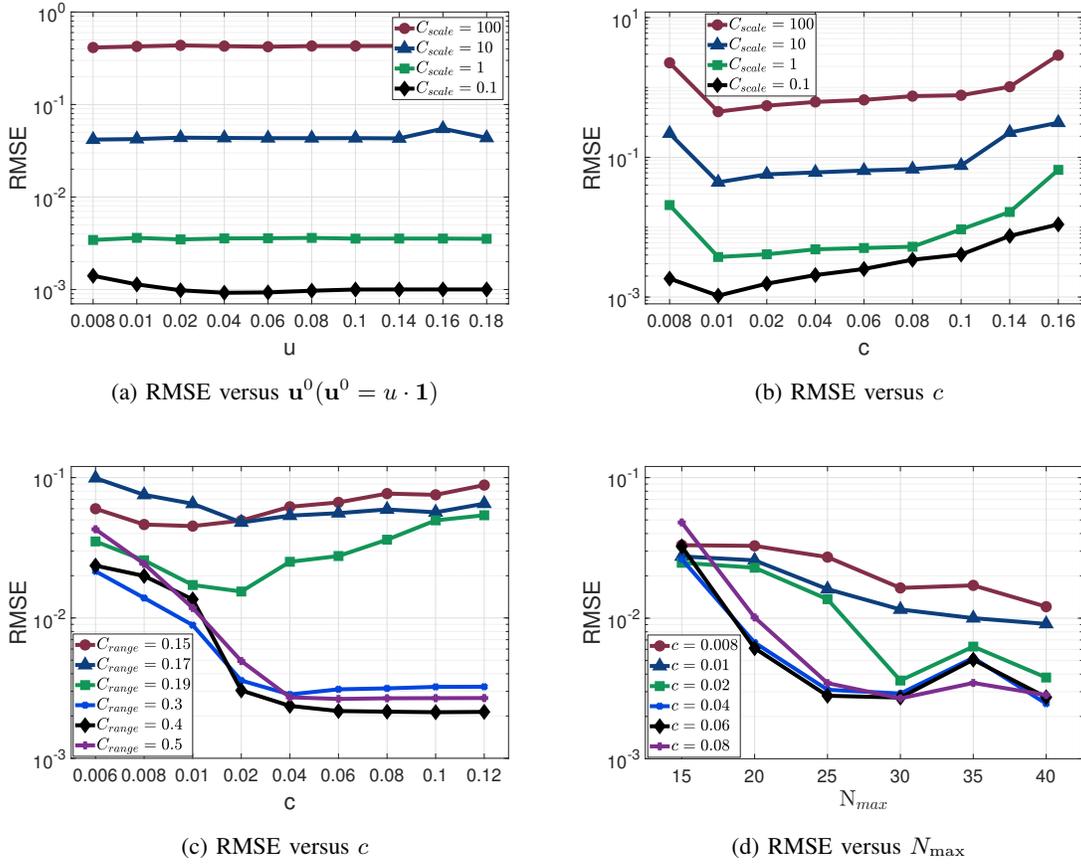


Fig. 12. Localization accuracy of the SP-ADMM algorithm at iteration step $T = 1000$. Measurement noise variance $\sigma_{i,j}^2 = 0.02\|\mathbf{p}_i - \mathbf{p}_j\|^2$; C_{scale} denotes the side length of the square area where the network is deployed; \mathbf{u}_i initialized as $\mathbf{u}_i^0 = u \cdot \mathbf{1}_{N_i}, \forall i \in \mathcal{N}$; C_{range} refers to communication range; $N_{\max} = \max\{N_i, i \in \mathcal{N}\}$. (a) penalty parameter c is fixed at 0.0265; (b)(c)(d) initial parameter u is fixed at 0.5.

$\mathbf{p}_j\|^2$. We can see that the curves at different C_{scale} show tiny fluctuations for varying u than in changing c . This is because the convergence rate is strongly dependent on the penalty parameter c : please see the supplementary material for the specific form of the convergence rate. These results indicate that the performance of the proposed SP-ADMM is not sensitive to initial parameter \mathbf{u} but sensitive to penalty parameter c . Besides, Fig. 12(b) shows that the appropriate interval value of c for different C_{scale} is consistent, and cannot be too large or too small. Therefore, parameter c should be adjusted more carefully when the proposed SP-ADMM algorithm is applied to different networks.

3) *The Selection Strategy of c based on the Communication Range*: In this part, we focus on how the parameter c should be chosen for networks with different communication ranges. The network composed of $N = 108, m = 8$ nodes located over $[0, 1] \times [0, 1]$ square area and the measurement noise

with variance $\sigma_{i,j}^2 = 0.02\|\mathbf{p}_i - \mathbf{p}_j\|^2$. Fig. 12(c) depicts the evolution of RMSE under networks at different communication ranges. Under the same number of iterative steps, it can be seen from Fig. 12(c) that networks with large C_{range} can obtain a lower RMSE by choosing a large c , networks with smaller C_{range} can reach lower RMSE by selecting smaller c .

These results are consistent with (62)-(64) derived from Lemma 2. In detail, the maximum measurement distance d_{max} , the maximum number of adjacent nodes N_{max} and the total number of adjacent nodes N_{sum} all increase as C_{range} increases. Therefore, according to (62)-(64), ρ needs to be increased to ensure the convergence of the proposed algorithm. Since we set $\rho = c$ in all experiments, it is reasonable for c to larger. Moreover, we also obtain similar results for the benchmark network, such as $\rho = c = 0.11$ ($C_{\text{range}} = 0.3, N = 500$) is greater than $\rho = c = 0.0197$ ($C_{\text{range}} = 0.1, N = 1000$). These results serve as the rule-of-thumb for selecting the parameter c of the proposed SP-ADMM algorithm when it is applied to different networks.

4) **Convergence rate with N_{max}** : Fig. 12(d) displays the RMSE versus the maximum number of adjacent nodes N_{max} for different values of c . The sensor network graph is the same as Fig. 12(c) at $C_{\text{range}} = 0.25$. As seen in Fig. 12(d), a lower RMSE is obtained at a larger N_{max} , which means that for the same penalty c , the proposed algorithm converges faster on a network with a larger N_{max} . This is indeed consistent with our convergence rate of Theorem 1. It can be seen from the supplement that the coefficient ϵ_2 of the sublinear convergence rate is inversely proportional to N_{max} when c, ρ, κ_1 and κ_2 are fixed.

D. Summary

The proposed algorithm has been evaluated against SDP, SF, AM-FD, and ADMM-H methods in terms of RMSE, running time, and communication cost across a range of network scenarios. These scenarios include different number of anchors and sensor nodes, average number of neighboring nodes, and different variance levels of the measurement noise under both benchmark and synthesized network configurations. Furthermore, we have included an empirical investigation on the impact of the penalty parameters ρ and c for our proposed SP-ADMM algorithm.

The results indicated that the proposed algorithm surpasses the SDP, SF, and AM-FD methods in terms of localization accuracy, while exhibiting lower computational complexity and communication cost than the ADMM-H method. Although the communication cost of the proposed SP-ADMM is higher than that of the SF and AM-FD algorithms per iteration, it provides higher accuracy at the same communication cost once convergence has been reached. Overall, the proposed SP-ADMM is a promising solution for cooperative localization problem in wireless sensor networks.

One possible future direction is to explore selective communication strategy among nodes to reduce the communication burden. This involves nodes randomly or sequentially choosing a subset of neighboring nodes to communicate with during each iteration. Additionally, addressing outliers in realistic measurement data and fusing other types of measurements are also important areas for future research.

VI. CONCLUSION

In this paper, we have proposed novel distributed parallel SP-ADMM algorithms for wireless sensor network localization in Gaussian measurement noise. Our proposed algorithms directly tackle the challenging nonconvex and nonsmooth problem without resorting to convex relaxation. We have shown that Algorithm 1 has a low computational complexity (in Table I) and proved that the algorithm converges to a critical point of the original problem (in Lemma 3, 4, and Theorem 1) at rate $\mathcal{O}(1/T)$. Moreover, we have proposed Algorithm 2, which is an improved version of Algorithm 1 in terms of storage space. Simulation and experimental results not only have shown that the proposed SP-ADMM algorithm is robust and performs well in the different average number of neighboring nodes, measurement noise variance, coverage area, and the number of anchors and non-anchors, but also provided suggestions of the parameters for migrating the proposed algorithm to different networks.

REFERENCES

- [1] P. K. Singh, B. K. Bhargava, M. Paprzycki, N. C. Kaushal, and W.-C. Hong, *Handbook of wireless sensor networks: issues and challenges in current Scenario's*. Springer, 2020, vol. 1132.
- [2] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–450, 2009.
- [3] F. Yin, C. Fritsche, D. Jin, F. Gustafsson, and A. M. Zoubir, "Cooperative localization in WSNs using Gaussian mixture modeling: Distributed ECM algorithms," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1448–1463, 2015.
- [4] Y.-M. Pun and A. M.-C. So, "Local strong convexity of source localization and error bound for target tracking under time-of-arrival measurements," *IEEE Transactions on Signal Processing*, vol. 70, pp. 190–201, 2021.
- [5] N. Okello, F. Fletcher, D. Musicki, and B. Ristic, "Comparison of recursive algorithms for emitter localisation using TDOA measurements from a pair of UAVs," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 3, pp. 1723–1732, 2011.
- [6] S. Xu and K. Doğançay, "Optimal sensor placement for 3-D angle-of-arrival target localization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1196–1211, 2017.
- [7] F. Yin, Y. Zhao, F. Gunnarsson, and F. Gustafsson, "Received-signal-strength threshold optimization using Gaussian processes," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2164–2177, 2017.
- [8] D. Jin, F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Bayesian cooperative localization using received signal strength with unknown path loss exponent: Message passing approaches," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1120–1135, 2020.

- [9] A. Simonetto and G. Leus, "Distributed maximum likelihood sensor network localization," *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1424–1437, 2014.
- [10] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *ACM Transactions on Sensor Networks (TOSN)*, vol. 2, no. 2, pp. 188–220, 2006.
- [11] P. Biswas, T.-C. Liang, K.-C. Toh, Y. Ye, and T.-C. Wang, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," *IEEE Transactions on Automation Science and Engineering*, vol. 3, no. 4, pp. 360–371, 2006.
- [12] Z. Wang, S. Zheng, Y. Ye, and S. Boyd, "Further relaxations of the semidefinite programming approach to sensor network localization," *SIAM Journal on Optimization*, vol. 19, no. 2, pp. 655–673, 2008.
- [13] S. Srirangarajan, A. H. Tewfik, and Z.-Q. Luo, "Distributed sensor network localization using SOCP relaxation," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4886–4895, 2008.
- [14] P. Di Lorenzo and G. Scutari, "NEXT: In-network nonconvex optimization," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 120–136, 2016.
- [15] D. Jin, F. Yin, A. M. Zoubir, and H. C. So, "Exploiting sparsity of ranging biases for NLOS mitigation," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3782–3795, 2021.
- [16] R. M. Buehrer, H. Wymeersch, and R. M. Vaghefi, "Collaborative sensor network localization: Algorithms and practical issues," *Proceedings of the IEEE*, vol. 106, no. 6, pp. 1089–1114, 2018.
- [17] M. Hong, D. Hajinezhad, and M.-M. Zhao, "Prox-PDA: The proximal primal-dual algorithm for fast distributed nonconvex optimization and learning over networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1529–1538.
- [18] C. Soares, J. Xavier, and J. Gomes, "Simple and fast convex relaxation method for cooperative localization in sensor networks using range measurements," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4532–4543, 2015.
- [19] N. Piovesan and T. Erseghe, "Cooperative localization in WSNs: A hybrid convex/nonconvex solution," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 162–172, 2016.
- [20] E. Gur, S. Sabach, and S. Shtern, "Alternating minimization based first-order method for the wireless sensor network localization problem," *IEEE Transactions on Signal Processing*, vol. 68, pp. 6418–6431, 2020.
- [21] Q. Shi, C. He, H. Chen, and L. Jiang, "Distributed wireless sensor network localization via sequential greedy optimization algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3328–3340, 2010.
- [22] D. Bertsekas, P. Tsitsiklis, and N. John, "Parallel and distributed computation: Numeral methods," 1989.
- [23] T. Erseghe, "A distributed and maximum-likelihood sensor network localization algorithm based upon a nonconvex problem formulation," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 4, pp. 247–258, 2015.
- [24] D. R. Luke, S. Sabach, M. Teboulle, and K. Zatlaway, "A simple globally convergent algorithm for the nonsmooth nonconvex single source localization problem," *Journal of Global Optimization*, vol. 69, no. 4, pp. 889–909, 2017.
- [25] Z. Wang, J. Zhang, T.-H. Chang, J. Li, and Z.-Q. Luo, "Distributed stochastic consensus optimization with momentum for nonconvex nonsmooth problems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4486–4501, 2021.
- [26] J. Zhang and Z.-Q. Luo, "A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization," *SIAM Journal on Optimization*, vol. 30, no. 3, pp. 2272–2302, 2020.
- [27] L. Ding, A. Yurtsever, V. Cevher, J. A. Tropp, and M. Udell, "An optimal-storage approach to semidefinite programming using approximate complementarity," *SIAM Journal on Optimization*, vol. 31, no. 4, pp. 2695–2725, 2021.
- [28] S. Yousefi, X.-W. Chang, and B. Champagne, "Cooperative localization of mobile nodes in NLOS," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication*. IEEE, 2014, pp. 275–279.

- [29] R. M. Vaghefi and R. M. Buehrer, "Cooperative localization in NLOS environments using semidefinite programming," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1382–1385, 2015.
- [30] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [31] W. Fan, J. Liang, G. Lu, X. Fan, and H. C. So, "Spectrally-agile waveform design for wideband mimo radar transmit beam pattern synthesis via majorization-admm," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1563–1578, 2021.
- [32] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.
- [33] K. Guo, D. Han, and T.-T. Wu, "Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints," *International Journal of Computer Mathematics*, vol. 94, no. 8, pp. 1653–1669, 2017.
- [34] J. Bolte, S. Sabach, M. Teboulle, and Y. Vaisbourd, "First order methods beyond convexity and lipschitz gradient continuity with applications to quadratic inverse problems," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2131–2151, 2018.
- [35] R. I. Boğ and D.-K. Nguyen, "The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates," *Mathematics of Operations Research*, vol. 45, no. 2, pp. 682–712, 2020.
- [36] M. Yashtini, "Convergence and rate analysis of a proximal linearized admm for nonconvex nonsmooth optimization," *Journal of Global Optimization*, vol. 84, no. 4, pp. 913–939, 2022.
- [37] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.
- [38] S. Kumar, R. Jain, and K. Rajawat, "Asynchronous optimization over heterogeneous networks via consensus admm," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 1, pp. 114–129, 2016.
- [39] K. Sun and X. A. Sun, "A two-level distributed algorithm for general constrained non-convex optimization with global convergence," *arXiv preprint arXiv:1902.07654*, 2019.
- [40] X. Zhang, J. Ma, Z. Cheng, S. Huang, C. W. de Silva, and T. H. Lee, "Improved hierarchical admm for nonconvex cooperative distributed model predictive control," *arXiv preprint arXiv:2011.00463*, 2020.
- [41] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate $O(1/k^2)$," in *Dokl. akad. nauk Sssr*, vol. 269, 1983, pp. 543–547.
- [42] Y. Ye, "Computational optimization laboratory.stanford university." [Online]. Available: <https://web.stanford.edu/~yyye/Col>
- [43] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O’dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on signal processing*, vol. 51, no. 8, pp. 2137–2148, 2003.
- [44] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.
- [45] A. Beck, *First-order methods in optimization*. Society for Industrial and Applied Mathematics, 2017.
- [46] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Mathematical Programming*, vol. 162, no. 1, pp. 165–199, 2017.

Supplementary Materials of Manuscript "Distributed Scaled Proximal ADMM Algorithms for Cooperative Localization in WSNs"

Mei Zhang, Zhiguo Wang, Feng Yin, and Xiaojing Shen

1. Proof of the Remark 2

Recall that

$$\mathcal{Z} = \left\{ \mathbf{z} \mid \mathbf{z}_{i,j}^+ = \mathbf{p}_j = \mathbf{z}_{j,i}^-, \forall i \in \mathcal{N}, j \in \mathcal{N}_i \right\},$$

and \mathbf{W}_i is a diagonal matrix defined in (29). Substituting them into (24), then the optimization (24) is rewritten as

$$\min_{\mathbf{z} \in \mathcal{X}} \sum_{i \in \mathcal{N}} \left[(c+1) N_i \|\mathbf{p}_i - \tilde{\mathbf{p}}_i^{t+1}\|^2 + \sum_{j \in \mathcal{N}_i} c \|\mathbf{z}_{i,j}^- - (\tilde{\mathbf{z}}_{i,j}^-)^{t+1}\|^2 + \sum_{j \in \mathcal{N}_i} \|\mathbf{z}_{i,j}^+ - (\tilde{\mathbf{z}}_{i,j}^+)^{t+1}\|^2 \right]. \quad (\text{S.1})$$

From the definition of \mathcal{X} in (14) and (S.1), we immediately obtain the optimal solution for \mathbf{p}_i^{t+1} as follows

$$\mathbf{p}_i^{t+1} = \begin{cases} \tilde{\mathbf{p}}_{i,i}^{t+1}, & \text{for } i \notin \mathcal{A}, \\ \mathbf{a}_i, & \text{for } i \in \mathcal{A}. \end{cases} \quad (\text{S.2})$$

Combining (S.2) with (36) yields

$$\mathbf{p}_i^{t+1} = \tilde{\mathbf{p}}_i^{t+1}, \forall i \in \mathcal{N}. \quad (\text{S.3})$$

The terms in problem (S.1) related to $\mathbf{z}_{i,j}^-$ are

$$c \left\| \mathbf{z}_{i,j}^- - (\tilde{\mathbf{z}}_{i,j}^-)^{t+1} \right\|^2 + \left\| \mathbf{z}_{i,j}^- - (\tilde{\mathbf{z}}_{j,i}^+)^{t+1} \right\|^2, \forall i \in \mathcal{N}, j \in \mathcal{N}_i.$$

By the optimality condition, we have

$$c \left((\mathbf{z}_{i,j}^-)^{t+1} - (\tilde{\mathbf{z}}_{i,j}^-)^{t+1} \right) + \left((\mathbf{z}_{i,j}^-)^{t+1} - (\tilde{\mathbf{z}}_{j,i}^+)^{t+1} \right) = \mathbf{0}.$$

After rearranging the terms, we obtain

$$(\mathbf{z}_{i,j}^-)^{t+1} = \frac{1}{c+1} \left(c \cdot (\tilde{\mathbf{z}}_{i,j}^-)^{t+1} + (\tilde{\mathbf{z}}_{j,i}^+)^{t+1} \right). \quad (\text{S.4})$$

Similarly, for $i \in \mathcal{N}, j \in \mathcal{N}_i$, we can obtain the closed form solution for $\mathbf{z}_{i,j}^+$ as follows

$$(\mathbf{z}_{i,j}^+)^{t+1} = \frac{1}{c+1} \left((\tilde{\mathbf{z}}_{i,j}^+)^{t+1} + c \cdot (\tilde{\mathbf{z}}_{j,i}^-)^{t+1} \right), \forall i \in \mathcal{N}, j \in \mathcal{N}_i. \quad (\text{S.5})$$

This completes the proof of Remark 2 by combining (9) and (S.3)-(S.5).

2. Proof of the Lemma 1

In (22), we take $\mathbf{z}_i = \mathbf{z}_i^{t+1}$ and $\mathbf{z}_i = \mathbf{z}_i^t$, it yields that

$$\mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) + \frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 - \mathcal{L}_i(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) = \frac{1}{2} \|\mathbf{z}_i^{t+1} - \tilde{\mathbf{z}}_i^{t+1}\|_{\mathbf{W}_i}^2 - \frac{1}{2} \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|_{\mathbf{W}_i}^2. \quad (\text{S.6})$$

Since \mathbf{z}_i^{t+1} is the optimal solution from Remark 2 and $\mathbf{z}_i^t \in \mathcal{Z} \cap \mathcal{X}$, $\mathbf{z}_i^{t+1} \in \mathcal{Z} \cap \mathcal{X}$, then we have

$$\frac{1}{2} \|\mathbf{z}_i^{t+1} - \tilde{\mathbf{z}}_i^{t+1}\|_{\mathbf{W}_i}^2 \leq \frac{1}{2} \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|_{\mathbf{W}_i}^2. \quad (\text{S.7})$$

Substituting (S.7) into (S.6), we obtain

$$\mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) - \mathcal{L}_i(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) \leq -\frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2. \quad (\text{S.8})$$

Using the \mathbf{u} update in (18) with the same technique, we can obtain the following inequality

$$\mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^t) - \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t) \leq -\frac{\rho}{2} \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\|^2. \quad (\text{S.9})$$

Recall that $\boldsymbol{\lambda}_i^{t+1} = \boldsymbol{\lambda}_i^t + c\mathbf{A}_i \mathbf{z}_i^{t+1}$ in (38), we have the trivial equality

$$\mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}) - \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^t) = \frac{1}{c} \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2. \quad (\text{S.10})$$

Applying (23), we have

$$\mathbf{A}_i^T \boldsymbol{\lambda}_i^t = \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t - \mathbf{W}_i \tilde{\mathbf{z}}_i^{t+1} + c\mathbf{B}_i^T \mathbf{B}_i \mathbf{z}_i^t. \quad (\text{S.11})$$

Both sides plus the term $c\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^{t+1}$, by (38), it yields

$$\begin{aligned} \mathbf{A}_i^T \boldsymbol{\lambda}_i^{t+1} &= \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t - \mathbf{W}_i \tilde{\mathbf{z}}_i^{t+1} + c\mathbf{B}_i^T \mathbf{B}_i \mathbf{z}_i^t + c\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^{t+1} \\ &= \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t - \mathbf{Q}_i^T \mathbf{Q}_i \tilde{\mathbf{z}}_i^{t+1} - c\mathbf{A}_i^T \mathbf{A}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1}) - c\mathbf{B}_i^T \mathbf{B}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t), \end{aligned} \quad (\text{S.12})$$

where the last equality dues to the definition of \mathbf{W}_i in (21). Let $\tilde{\sigma}_{min}$ denote the smallest non-zero eigenvalue of $\mathbf{A}_i^T \mathbf{A}_i$, we have

$$\tilde{\sigma}_{min} \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2 \leq \|\mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t)\|^2.$$

From the definition of $\mathbf{A}_i^T \mathbf{A}_i$ in (26), it derives $\tilde{\sigma}_{min} = 1$. The above inequality combined with (S.12) implies that

$$\begin{aligned} &\|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2 \\ &\leq \|\mathbf{Q}_i^T \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1}) - \mathbf{Q}_i^T \mathbf{Q}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - c\mathbf{A}_i^T \mathbf{A}_i [\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1} - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^t)] \\ &\quad - c\mathbf{B}_i^T \mathbf{B}_i [\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})]\|^2 \\ &\leq 3(N_i + 1) \|\mathbf{Q}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 + 3c^2(N_i + 1) \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1} - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \\ &\quad + 3c^2 \|\mathbf{B}_i^T \mathbf{B}_i\| \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2, \end{aligned} \quad (\text{S.13})$$

where the last inequality holds because of the triangle inequality and $\|\mathbf{Q}_i^T \mathbf{Q}_i\| = \|\mathbf{A}_i^T \mathbf{A}_i\| = N_i + 1$ from (26)-(27), $\|\mathbf{A}_i^T \mathbf{A}_i\|$ denotes the spectral norm of a matrix $\mathbf{A}_i^T \mathbf{A}_i$. Along with the form of the matrix $c\mathbf{B}_i^T \mathbf{B}_i$ in (28), we have

$$\|\mathbf{B}_i^T \mathbf{B}_i\| \leq \frac{(1+c)(1+N_{\max})}{c}, \quad (\text{S.14})$$

where $N_{\max} := \max\{N_i, i \in \mathcal{N}\}$. Substituting (S.14) into (S.13) and the resulting applied to (S.10), we have

$$\begin{aligned} & \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}^{t+1}, \boldsymbol{\lambda}_i^{t+1}) - \mathcal{L}(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^t) \\ & \leq \frac{3(N_{\max} + 1)}{c} \|\mathbf{Q}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i(\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 + 3c(N_{\max} + 1) \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1} - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \\ & \quad + 3(1+c)(1+N_{\max}) \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2. \end{aligned} \quad (\text{S.15})$$

Combing (S.8), (S.9), (S.15), moreover, taking summation over $i \in \mathcal{N}$, we obtain the final result (59).

3. Proof of the Lemma 2

The following Lemma shows the descent of the desired term.

Lemma 5 *Suppose $c\mathbf{B}_i^T \mathbf{B}_i$ takes the form of (25), then the following is true for Algorithm 1*

$$\begin{aligned} & \sum_{i \in \mathcal{N}} \frac{c}{2} \|\mathbf{A}_i \tilde{\mathbf{z}}_i^{t+1}\|^2 + \frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \\ & \leq \sum_{i \in \mathcal{N}} \left[\frac{c}{2} \|\mathbf{A}_i \tilde{\mathbf{z}}_i^t\|^2 + \frac{c}{2} \|\mathbf{z}_i^t - \mathbf{z}_i^{t-1}\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 - \frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 - \frac{1}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{Q}_i^T \mathbf{Q}_i}^2 \right. \\ & \quad + \frac{d_{\max}^2}{2} \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2 - \frac{1}{2} \|\mathbf{Q}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i(\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 \\ & \quad - \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1} - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 + \frac{c(N_i + 1)}{2} \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|^2 \\ & \quad \left. - \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \right], \end{aligned}$$

where $d_{\max} := \max\{d_{ij}, i \in \mathcal{N}, j \in \mathcal{N}_i\}$.

Proof: According to (S.12), we get

$$\begin{aligned} \langle \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle & = \langle -\mathbf{W}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) + \mathbf{Q}_i^T \mathbf{D}_i(\mathbf{u}_i^t - \mathbf{u}_i^{t-1}), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle \\ & \quad + \langle c\mathbf{A}_i^T \mathbf{A}_i(\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) + c\mathbf{B}_i^T \mathbf{B}_i(\mathbf{z}_i^t - \mathbf{z}_i^{t-1}), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle. \end{aligned} \quad (\text{S.16})$$

Let us bound the left-hand side (lhs) and the rhs of (S.16) separately. First, the lhs of (S.16) can be expressed as

$$\langle \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle = \langle c\mathbf{A}_i^T \mathbf{A}_i \tilde{\mathbf{z}}_i^{t+1}, \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle,$$

where the equality dues to dual update in (38). Furthermore,

$$\begin{aligned}
& \langle c\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^{t+1}, \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle \\
&= \langle c\mathbf{A}_i^T \mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle - \langle c\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t, \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle \\
&= -\frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 + \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 + \frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \\
&\quad + \frac{c}{2} \|\mathbf{A}_i (\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^t)\|^2 - \frac{c}{2} \|\mathbf{A}_i (\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1})\|^2 + \frac{c}{2} \|\mathbf{A}_i \tilde{\mathbf{z}}_i^{t+1}\|^2 - \frac{c}{2} \|\mathbf{A}_i \tilde{\mathbf{z}}_i^t\|^2, \tag{S.17}
\end{aligned}$$

where the final equality dues to $\langle \mathbf{a}, \mathbf{b} \rangle = -\frac{1}{2} \|\mathbf{a} - \mathbf{b}\|^2 + \frac{1}{2} \|\mathbf{a}\|^2 + \frac{1}{2} \|\mathbf{b}\|^2$ for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{(2N_i+1)n}$. And from the compatibility of the norm that

$$\frac{c}{2} \|\mathbf{A}_i (\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1})\|^2 \leq \frac{c(N_i+1)}{2} \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|^2, \tag{S.18}$$

where $\|\mathbf{A}_i^T \mathbf{A}_i\| = N_i + 1$ due to (26).

Second, we have rewritten the first row on the rhs of the (S.16):

$$\begin{aligned}
& \langle -\mathbf{W}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) + \mathbf{Q}_i^T \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1}), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle \\
&= -\|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t\|_{\mathbf{W}_i}^2 - \frac{1}{2} \|\mathbf{Q}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 \\
&\quad + \frac{1}{2} \|\mathbf{Q}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t)\|^2 + \frac{1}{2} \|\mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2. \tag{S.19}
\end{aligned}$$

For the second row on the rhs of the (S.16), we have

$$\begin{aligned}
& \langle c\mathbf{A}_i^T \mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) + c\mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^t - \mathbf{z}_i^{t-1}), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle \\
&= \frac{c}{2} \|\mathbf{A}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t)\|^2 + \frac{c}{2} \|\mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2 - \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t - (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \\
&\quad + \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 + \frac{c}{2} \|\mathbf{z}_i^t - \mathbf{z}_i^{t-1}\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 - \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2. \tag{S.20}
\end{aligned}$$

Recall $\mathbf{W}_i = \mathbf{Q}_i^T \mathbf{Q}_i + c\mathbf{A}_i^T \mathbf{A}_i + c\mathbf{B}_i^T \mathbf{B}_i$. Combing (S.19) and (S.20), then the rhs of (S.16) can be expressed as

$$\begin{aligned}
& \langle -\mathbf{W}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) + \mathbf{Q}_i^T \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1}), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle \\
&\quad + \langle c\mathbf{A}_i^T \mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) + c\mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^t - \mathbf{z}_i^{t-1}), \tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t \rangle \\
&= -\frac{1}{2} \|\mathbf{Q}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 - \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t - (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 \\
&\quad - \frac{1}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t\|_{\mathbf{W}_i}^2 - \frac{c}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 + \frac{c}{2} \|\mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2 \\
&\quad + \frac{c}{2} \|\mathbf{z}_i^t - \mathbf{z}_i^{t-1}\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 + \frac{1}{2} \|\mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2. \tag{S.21}
\end{aligned}$$

From the optimality conditions for the strongly convex optimization problem (24), for any $\mathbf{z}_i \in \mathcal{Z} \cap \mathcal{X}$ we have

$$\begin{aligned} \sum_{i \in \mathcal{N}} \langle \mathbf{W}_i(\mathbf{z}_i^{t+1} - \tilde{\mathbf{z}}_i^{t+1}), \mathbf{z}_i^{t+1} - \mathbf{z}_i \rangle &\leq 0, \\ \sum_{i \in \mathcal{N}} \langle \mathbf{W}_i(\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^t), \mathbf{z}_i^t - \mathbf{z}_i \rangle &\leq 0. \end{aligned}$$

Plugging $\mathbf{z}_i = \mathbf{z}_i^t$ into the first inequality and $\mathbf{z}_i = \mathbf{z}_i^{t+1}$ into the second, adding the resulting inequalities, we have

$$\sum_{i \in \mathcal{N}} \langle \mathbf{W}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t + \mathbf{z}_i^t - \mathbf{z}_i^{t+1}), \mathbf{z}_i^t - \mathbf{z}_i^{t+1} \rangle \leq 0.$$

Rearranging the above inequality, we have

$$\sum_{i \in \mathcal{N}} \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|_{\mathbf{W}_i}^2 \leq \sum_{i \in \mathcal{N}} \langle \mathbf{W}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t), \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rangle. \quad (\text{S.22})$$

Using $\langle \mathbf{a}, \mathbf{b} \rangle = -\frac{1}{2}\|\mathbf{a} - \mathbf{b}\|^2 + \frac{1}{2}\|\mathbf{a}\|^2 + \frac{1}{2}\|\mathbf{b}\|^2$, then (S.22) becomes

$$\sum_{i \in \mathcal{N}} \frac{1}{2} \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|_{\mathbf{W}_i}^2 \leq \sum_{i \in \mathcal{N}} \frac{1}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t\|_{\mathbf{W}_i}^2 - \frac{1}{2} \|\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t - (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2. \quad (\text{S.23})$$

Substituting (S.18) (S.23) into (S.17) (S.21), respectively, again using $\mathbf{W}_i = \mathbf{Q}_i^T \mathbf{Q}_i + c\mathbf{A}_i^T \mathbf{A}_i + c\mathbf{B}_i^T \mathbf{B}_i$ and combining (S.16), we obtain the result. \blacksquare

Note that it remains to bound the term $\|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|^2$ in Lemma 5. Next, we establish a simple Lemma to ensure that the term decreases.

Lemma 6 Let $\{(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t)\}$ be the sequence generated by Algorithm 1, $c\mathbf{B}_i^T \mathbf{B}_i$ takes the form of (25).

Then we have

$$\begin{aligned} &\sum_{i \in \mathcal{N}} \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 + \frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \\ &\leq \sum_{i \in \mathcal{N}} \left[\frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^t\|^2 + \frac{c}{2} \|\mathbf{z}_i^t - \mathbf{z}_i^{t-1}\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 + \frac{d_{\max}^2}{2} \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2 - \frac{c \cdot \tilde{\tau}_{\min}}{2N_{\text{sum}}n(c+1)^2} \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|^2 \right], \end{aligned}$$

where $\tilde{\tau}_{\min} := \min\{(c+1)^2 N_i^2 + c^2 N_i + N_i, i \in \mathcal{N}\}$ and $N_{\text{sum}} := \sum_{i \in \mathcal{N}} N_i$.

Proof: Using (S.12) again, we can get

$$\begin{aligned} \langle \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t), \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rangle &= \langle -\mathbf{W}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t), \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rangle + c \|\mathbf{A}_i(\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2 \\ &\quad + \langle \mathbf{Q}_i^T \mathbf{D}_i(\mathbf{u}_i^t - \mathbf{u}_i^{t-1}) + c\mathbf{B}_i^T \mathbf{B}_i(\mathbf{z}_i^t - \mathbf{z}_i^{t-1}), \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rangle. \quad (\text{S.24}) \end{aligned}$$

First, the lhs of (S.24) can be expressed as

$$\begin{aligned} \langle \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t), \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rangle &= \langle c\mathbf{A}_i \mathbf{z}_i^{t+1}, \mathbf{A}_i \mathbf{z}_i^{t+1} - \mathbf{A}_i \mathbf{z}_i^t \rangle \\ &= \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 - \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^t\|^2 + \frac{c}{2} \|\mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2. \end{aligned} \quad (\text{S.25})$$

where the first equality dues to (38), and the second equality dues to $\langle \mathbf{a}, \mathbf{a} - \mathbf{b} \rangle = \frac{1}{2} \|\mathbf{a} - \mathbf{b}\|^2 + \frac{1}{2} \|\mathbf{a}\|^2 - \frac{1}{2} \|\mathbf{b}\|^2$ for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{(2N_i+1)n}$.

Next, we deal with the rhs of the (S.24). By (S.22) yields

$$\sum_{i \in \mathcal{N}} \langle -\mathbf{W}_i (\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t), \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rangle \leq - \sum_{i \in \mathcal{N}} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{W}_i}^2. \quad (\text{S.26})$$

And using the Cauchy–Schwarz inequality, we have

$$\begin{aligned} &\langle \mathbf{Q}_i^T \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1}) + c\mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^t - \mathbf{z}_i^{t-1}), \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rangle \\ &\leq \frac{1}{2} \|\mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 + \frac{1}{2} \|\mathbf{Q}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2 + \frac{c}{2} \|\mathbf{z}_i^t - \mathbf{z}_i^{t-1}\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 + \frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2. \end{aligned} \quad (\text{S.27})$$

Combing (S.24)-(S.27) and using $\mathbf{W}_i = c\mathbf{B}_i^T \mathbf{B}_i + c\mathbf{A}_i^T \mathbf{A}_i + \mathbf{Q}_i^T \mathbf{Q}_i$, we have

$$\begin{aligned} &\sum_{i \in \mathcal{N}} \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 + \frac{c}{2} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \\ &\leq \sum_{i \in \mathcal{N}} \left[\frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^t\|^2 + \frac{c}{2} \|\mathbf{z}_i^t - \mathbf{z}_i^{t-1}\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 + \frac{1}{2} \|\mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 - \frac{c}{2} \|\mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2 \right]. \end{aligned} \quad (\text{S.28})$$

Hence, we will use the dual residual $\sum_{i \in \mathcal{N}} \|\mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|^2$ to bound $\sum_{i \in \mathcal{N}} \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|$. Recall the definition of \mathbf{z}_i in (9) and \mathbf{A}_i in (12), we have

$$\mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) = \left[\mathbf{p}_i^{t+1} - \mathbf{p}_i^t - \left((\mathbf{z}_{i,j}^-)^{t+1} - (\mathbf{z}_{i,j}^-)^t \right) \right]_{j \in \mathcal{N}_i},$$

then it yields

$$\mathbf{1}_{N_i n}^T \cdot \mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) = \sum_{j \in \mathcal{N}_i} \left[\mathbf{p}_i^{t+1} - \mathbf{p}_i^t - \left((\mathbf{z}_{i,j}^-)^{t+1} - (\mathbf{z}_{i,j}^-)^t \right) \right]. \quad (\text{S.29})$$

Substituting (30)-(31) into (S.29), rearranging the terms, and summing the above relation over $i \in \mathcal{N}$, we obtain

$$\begin{aligned} \sum_{i \in \mathcal{N}} \mathbf{1}_{N_i n}^T \cdot \mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) &= \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left[\frac{c}{c+1} \left(\tilde{\mathbf{p}}_i^{t+1} - \mathbf{p}_i^t - \left((\tilde{\mathbf{z}}_{i,j}^-)^{t+1} - (\mathbf{z}_{i,j}^-)^t \right) \right) \right. \\ &\quad \left. + \frac{1}{c+1} \left(\tilde{\mathbf{p}}_i^{t+1} - \mathbf{p}_i^t - \left((\tilde{\mathbf{z}}_{j,i}^+)^{t+1} - (\mathbf{z}_{i,j}^-)^t \right) \right) \right]. \end{aligned} \quad (\text{S.30})$$

For the first row on the rhs of the (S.30), using (12) again, we have

$$\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{c}{c+1} \left(\tilde{\mathbf{p}}_i^{t+1} - \mathbf{p}_i^t - \left((\tilde{\mathbf{z}}_{i,j}^-)^{t+1} - (\mathbf{z}_{i,j}^-)^t \right) \right) = \sum_{i \in \mathcal{N}} \mathbf{1}_{N_i n}^T \cdot \frac{c}{c+1} \mathbf{A}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t). \quad (\text{S.31})$$

For the second row on the rhs of the (S.30), since \mathbf{z}_i^{t+1} is an optimal solution of (24) for all $t \geq 1$, we have $\mathbf{z}_i^t \in \mathcal{Z}$, then we can deduce that

$$(\mathbf{z}_{i,j}^-)^t = (\mathbf{z}_{j,i}^+)^t, \quad \forall i \in \mathcal{N}, j \in \mathcal{N}_i. \quad (\text{S.32})$$

Hence, together with (S.32), the second row on the rhs of the (S.30) can be expressed as

$$\begin{aligned} & \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{1}{c+1} \left(\tilde{\mathbf{p}}_i^{t+1} - \mathbf{p}_i^t - \left((\tilde{\mathbf{z}}_{j,i}^+)^{t+1} - (\mathbf{z}_{i,j}^-)^t \right) \right) \\ &= \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{1}{c+1} \left(\tilde{\mathbf{p}}_i^{t+1} - \mathbf{p}_i^t - \left((\tilde{\mathbf{z}}_{j,i}^+)^{t+1} - (\mathbf{z}_{j,i}^+)^t \right) \right) \\ &= \sum_{i \in \mathcal{N}} \mathbf{1}_{N_i n}^T \cdot \frac{1}{c+1} \mathbf{Q}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t), \end{aligned} \quad (\text{S.33})$$

where the last step is due to (11) and rearranges the terms. Substituting (S.31) and (S.33) into (S.30), we can get

$$\left\| \sum_{i \in \mathcal{N}} \mathbf{1}_{N_i n}^T \cdot \mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) \right\|^2 = \frac{1}{(c+1)^2} \left\| \sum_{i \in \mathcal{N}} \mathbf{1}_{N_i n}^T \cdot [c\mathbf{A}_i + \mathbf{Q}_i] (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t) \right\|^2. \quad (\text{S.34})$$

Let $\mathbf{A} := \text{diag}(\mathbf{A}_i, i \in \mathcal{N})$ denote the block diagonal matrix whose diagonal coefficients correspond to the $\mathbf{A}_i, i \in \mathcal{N}$ and $\mathbf{Q} := \text{diag}(\mathbf{Q}_i, i \in \mathcal{N})$, $\tilde{\mathbf{z}}^{t+1} := \text{vec}(\tilde{\mathbf{z}}_i^{t+1}, i \in \mathcal{N})$, then (S.34) can be rewritten as a compact form:

$$\left\| \mathbf{1}_{N_{\text{sum}} n}^T \cdot \mathbf{A} (\mathbf{z}^{t+1} - \mathbf{z}^t) \right\|^2 = \frac{1}{(c+1)^2} \left\| \mathbf{1}_{N_{\text{sum}} n}^T \cdot [c\mathbf{A} + \mathbf{Q}] (\tilde{\mathbf{z}}^{t+1} - \mathbf{z}^t) \right\|^2, \quad (\text{S.35})$$

where $N_{\text{sum}} := \sum_{i \in \mathcal{N}} N_i$. Due to the Cauchy-Schwarz inequality, we can upper bound the lhs of (S.35)

$$N_{\text{sum}} n \cdot \left\| \mathbf{A} (\mathbf{z}^{t+1} - \mathbf{z}^t) \right\|^2 \geq \left\| \mathbf{1}_{N_{\text{sum}} n}^T \cdot \mathbf{A} (\mathbf{z}^{t+1} - \mathbf{z}^t) \right\|^2. \quad (\text{S.36})$$

For the rhs of (S.35), we have

$$\left\| \mathbf{1}_{N_{\text{sum}} n}^T \cdot [c\mathbf{A} + \mathbf{Q}] (\tilde{\mathbf{z}}^{t+1} - \mathbf{z}^t) \right\|^2 \geq \tilde{\tau}_{\min} \left\| \tilde{\mathbf{z}}^{t+1} - \mathbf{z}^t \right\|^2, \quad (\text{S.37})$$

where $\tilde{\tau}_{\min}$ is defined as the smallest non-zero eigenvalue of $[\mathbf{1}_{N_{\text{sum}} n}^T \cdot (c\mathbf{A} + \mathbf{Q})]^T [\mathbf{1}_{N_{\text{sum}} n}^T \cdot (c\mathbf{A} + \mathbf{Q})]$.

From the definition of \mathbf{A}_i and \mathbf{Q}_i in (12) and (11), it derives $\tilde{\tau}_{\min} = \min\{(c+1)^2 N_i^2 + c^2 N_i + N_i, i \in \mathcal{N}\}$.

Substituting (S.36) and (S.37) into (S.35), we obtain

$$\frac{c}{2} \left\| \mathbf{A} (\mathbf{z}^{t+1} - \mathbf{z}^t) \right\|^2 \geq \frac{c \cdot \tilde{\tau}_{\min}}{2N_{\text{sum}} n (c+1)^2} \left\| \tilde{\mathbf{z}}^{t+1} - \mathbf{z}^t \right\|^2, \quad (\text{S.38})$$

which can be rewritten as a summation over network nodes:

$$\sum_{i \in \mathcal{N}} \frac{c}{2} \left\| \mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t) \right\|^2 \geq \sum_{i \in \mathcal{N}} \frac{c \cdot \tilde{\tau}_{\min}}{2N_{\text{sum}} n (c+1)^2} \left\| \tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t \right\|^2. \quad (\text{S.39})$$

Finally, we substitute (S.39) to (S.28), and the proof is complete. \blacksquare

Using Lemma 1, Lemma 5, Lemma 6 and the definition of the potential function ζ^t in (60), we can get the desired result of (61).

4. Proof of the Lemma 3

If $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) = 0$, it implies that, for any node $i \in \mathcal{N}$,

$$\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t - (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t)) = 0, \quad (\text{S.40})$$

$$\mathbf{A}_i \mathbf{z}_i^t = 0, \quad (\text{S.41})$$

$$\mathbf{u}_i^t - \mathbf{u}_i^{t-1} = 0. \quad (\text{S.42})$$

According to the definition of the projection operator, we have

$$\text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t - (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t)) = \arg \min_{\substack{\mathbf{z} \in \mathcal{Z} \\ \mathbf{z} \in \mathcal{X}}} \|\mathbf{z}_i - \mathbf{z}_i^t + (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t)\|^2, \quad (\text{S.43})$$

Using (S.40), we deduce that \mathbf{z}_i^t is an optimal solution for problem (S.43). Therefore, based on the first order necessary condition of (S.43), we obtain

$$\langle \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t, \mathbf{x} - \mathbf{z}_i^t \rangle \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}, \mathbf{x} \in \mathcal{Z}. \quad (\text{S.44})$$

Since $F_i(\mathbf{z}_i, \mathbf{u}_i) + \langle \boldsymbol{\lambda}_i, \mathbf{A}_i \mathbf{z}_i \rangle$ is convex over \mathcal{X} and \mathcal{Z} for any fixed \mathbf{u}_i , then the above expression (S.44) implies that \mathbf{z}_i^t is also sufficient for \mathbf{z}_i^t to minimize $F_i(\mathbf{z}_i, \mathbf{u}_i) + \langle \boldsymbol{\lambda}_i, \mathbf{A}_i \mathbf{z}_i \rangle$ over \mathcal{X} and \mathcal{Z} , i.e.,

$$\mathbf{z}_i^t \in \arg \min_{\substack{\mathbf{z} \in \mathcal{Z} \\ \mathbf{z} \in \mathcal{X}}} F_i(\mathbf{z}_i, \mathbf{u}_i^t) + \langle \boldsymbol{\lambda}_i^t, \mathbf{A}_i \mathbf{z}_i \rangle.$$

Thanks to the separable property of the objective function about node i , we know \mathbf{z}^t shall satisfy the KKT condition

$$\mathbf{z}^t \in \arg \min_{\substack{\mathbf{z} \in \mathcal{Z} \\ \mathbf{z} \in \mathcal{X}}} \sum_{i \in \mathcal{N}} F_i(\mathbf{z}_i, \mathbf{u}_i^t) + \langle \boldsymbol{\lambda}_i^t, \mathbf{A}_i \mathbf{z}_i \rangle. \quad (\text{S.45})$$

Similarly, the primal variable \mathbf{u}_i is updated as follows

$$\mathbf{u}_i^t = \text{proj}_{\mathcal{B}^{N_i}}(\mathbf{u}_i^{t-1} - \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t)).$$

Combing it with (S.42), we have

$$0 = \|\mathbf{u}_i^{t-1} - \mathbf{u}_i^t\|^2 = \|\mathbf{u}_i^{t-1} - \text{proj}_{\mathcal{B}^{N_i}}(\mathbf{u}_i^{t-1} - \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t))\|^2,$$

Therefore \mathbf{u}_i^{t-1} is the optimal solution of the following problem

$$\min_{\mathbf{u}_i \in \mathcal{B}^{N_i}} \|\mathbf{u}_i - \mathbf{u}_i^{t-1} + \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t)\|^2. \quad (\text{S.46})$$

Based on the optimality condition of problem (S.46) and equation (S.42), we have

$$0 \leq \langle \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t), \mathbf{u}_i - \mathbf{u}_i^t + \mathbf{u}_i^t - \mathbf{u}_i^{t-1} \rangle = \langle \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t), \mathbf{u}_i - \mathbf{u}_i^t \rangle, \quad \forall \mathbf{u}_i \in \mathcal{B}^{N_i}. \quad (\text{S.47})$$

Since $F_i(\mathbf{z}_i, \mathbf{u}_i)$ is convex for any fixed \mathbf{z}_i , it implies that

$$\mathbf{u}_i^t \in \arg \min_{\mathbf{u}_i \in \mathcal{B}^{N_i}} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t), \quad (\text{S.48})$$

which is also written as

$$\mathbf{0} \in \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \partial \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i^t). \quad (\text{S.49})$$

Combing (S.45), (S.49) and (S.41), one can observe that $(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ satisfy the KKT condition (66).

Next, we show that if $(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ is a KKT solution of Problem (15), which satisfies (66), then $(\mathbf{p}^t, \mathbf{u}^t)$ is a critical point of the nonconvex problem (6) satisfying:

$$\mathbf{0} = \sum_{j \in \mathcal{N}_i} (\mathbf{p}_i^t - \mathbf{p}_j^t - d_{i,j} \mathbf{u}_{i,j}^t + \mathbf{p}_i^t - \mathbf{p}_j^t + d_{j,i} \mathbf{u}_{j,i}^t), \quad \forall i \in (\mathcal{N}/\mathcal{A}), \quad (\text{S.50a})$$

$$\mathbf{0} \in -d_{i,j} (\mathbf{p}_i^t - \mathbf{p}_j^t) + \partial \delta_{\mathcal{B}}(\mathbf{u}_{i,j}^t), \quad \forall i \in \mathcal{N}, j \in \mathcal{N}_i, \quad (\text{S.50b})$$

$$\mathbf{p}_j^t = \mathbf{a}_j, \quad \forall j \in \mathcal{A}. \quad (\text{S.50c})$$

Firstly, since \mathbf{z}^t is an optimal solution of (66a), we have $\mathbf{z}^t \in \mathcal{X}$, thus (S.50c) is proved.

Secondly, we prove $(\mathbf{p}^t, \mathbf{u}^t)$ satisfies (S.50b). Since \mathbf{z}^t is an optimal solution of (66a), we have $\mathbf{z}^t \in \mathcal{Z}$. By combining this with the KKT condition (66c), we can deduce that

$$(\mathbf{z}_{i,j}^+)^t = \mathbf{p}_j^t = (\mathbf{z}_{j,i}^-)^t, \quad \forall i \in \mathcal{N}, \quad \forall j \in \mathcal{N}_i. \quad (\text{S.51})$$

By using the definition of $F_i(\mathbf{z}_i^t, \mathbf{u}_i^t)$ and (S.51), we obtain

$$\nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) = -\mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i^t = \text{vec} \left(-d_{i,j} (\mathbf{p}_i^t - (\mathbf{z}_{i,j}^+)^t), j \in \mathcal{N}_i \right) = \text{vec}(-d_{i,j} (\mathbf{p}_i^t - \mathbf{p}_j^t), j \in \mathcal{N}_i), \quad (\text{S.52})$$

where

$$\begin{aligned} \mathbf{D}_i &= \text{Diag}(\text{vec}(d_{i,j}, j \in \mathcal{N}_i)) \otimes \mathbf{I}_n, \quad \mathbf{Q}_i = [\mathbf{1}_{N_i}, \mathbf{0}_{N_i}, -\mathbf{I}_{N_i}] \otimes \mathbf{I}_n, \\ \mathbf{z}_i^t &= \left[(\mathbf{p}_i^t)^T, \text{vec} \left((\mathbf{z}_{i,j}^-)^t, j \in \mathcal{N}_i \right)^T, \text{vec} \left((\mathbf{z}_{i,j}^+)^t, j \in \mathcal{N}_i \right)^T \right]^T. \end{aligned} \quad (\text{S.53})$$

By combining (S.52) with KKT condition (66b), we have that $(\mathbf{z}_i^t, \mathbf{u}_{i,j}^t)$ satisfies the critical condition (S.50b).

Finally, let us prove (S.50a). Recall that

$$F_i(\mathbf{z}_i, \mathbf{u}_i) = \frac{1}{2} \|\mathbf{Q}_i \mathbf{z}_i\|^2 - \mathbf{u}_i^T \mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i, \quad (\text{S.54})$$

$$\mathcal{X} = \{\mathbf{z} \mid \mathbf{E}_i \mathbf{z}_i = \mathbf{a}_i, \forall i \in \mathcal{A}\}, \quad \mathbf{E}_i = [1, \mathbf{0}_{N_i}^T, \mathbf{0}_{N_i}^T] \otimes \mathbf{I}_n, \quad \mathcal{Z} = \{\mathbf{z}_{i,j}^+ = \mathbf{p}_j = \mathbf{z}_{j,i}^-, \forall i \in \mathcal{N}, j \in \mathcal{N}_i\}.$$

Substituting (S.53) and (S.54) into (66a), then the optimization problem (66a) is rewritten as

$$\begin{aligned}
& \arg \min_{\substack{\mathbf{z} \in \mathcal{Z} \\ \mathbf{z} \in \mathcal{X}}} \sum_{i \in \mathcal{N}} F_i(\mathbf{z}_i, \mathbf{u}_i) + \langle \boldsymbol{\lambda}_i, \mathbf{A}_i \mathbf{z}_i \rangle \\
&= \arg \min_{\substack{\mathbf{z}_{i,j}^+ = \mathbf{z}_{j,i}^- \\ \mathbf{p}_i = \mathbf{a}_i, i \in \mathcal{A}}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left(\frac{1}{2} \|\mathbf{p}_i - \mathbf{z}_{i,j}^+\|^2 - d_{i,j} \mathbf{u}_{i,j}^T (\mathbf{p}_i - \mathbf{z}_{i,j}^+) + \boldsymbol{\lambda}_{i,j}^T (\mathbf{p}_i - \mathbf{z}_{i,j}^-) \right) \\
&= \arg \min_{\mathbf{p}_i, \mathbf{z}_{i,j}^-} \sum_{i \in (\mathcal{N}/\mathcal{A})} \sum_{j \in \mathcal{N}_i} \left(\frac{1}{2} \|\mathbf{p}_i - \mathbf{z}_{j,i}^-\|^2 - d_{i,j} \mathbf{u}_{i,j}^T (\mathbf{p}_i - \mathbf{z}_{j,i}^-) + \boldsymbol{\lambda}_{i,j}^T (\mathbf{p}_i - \mathbf{z}_{i,j}^-) \right) \\
&\quad + \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{N}_i} \left(\frac{1}{2} \|\mathbf{a}_i - \mathbf{z}_{j,i}^-\|^2 - d_{i,j} \mathbf{u}_{i,j}^T (\mathbf{a}_i - \mathbf{z}_{j,i}^-) + \boldsymbol{\lambda}_{i,j}^T (\mathbf{a}_i - \mathbf{z}_{i,j}^-) \right). \tag{S.55}
\end{aligned}$$

where the last equality holds by substituting the constraints $\mathbf{z}_{i,j}^+ = \mathbf{z}_{j,i}^-$, $\forall i \in \mathcal{N}, j \in \mathcal{N}_i$ and $\mathbf{p}_i = \mathbf{a}_i, i \in \mathcal{A}$ into the objective function of the first equality. Using the KKT condition (66a), we know that \mathbf{z}^t is an optimal solution of problem (S.55). By using the optimality condition of (S.55) about variables \mathbf{p}_i and $\mathbf{z}_{i,j}^-$, we have that

$$\begin{aligned}
\mathbf{0} &= \sum_{j \in \mathcal{N}_i} \left(\mathbf{p}_i^t - (\mathbf{z}_{j,i}^-)^t - d_{i,j} \mathbf{u}_{i,j}^t + \boldsymbol{\lambda}_{i,j}^t \right), \quad \forall i \in (\mathcal{N}/\mathcal{A}), \\
\mathbf{0} &= \sum_{j \in \mathcal{N}_i} \left((\mathbf{z}_{i,j}^-)^t - \mathbf{p}_j^t + d_{j,i} \mathbf{u}_{j,i}^t - \boldsymbol{\lambda}_{i,j}^t \right), \quad \forall i \in (\mathcal{N}/\mathcal{A}),
\end{aligned}$$

Adding the above equalities together, we get

$$\mathbf{0} = \sum_{j \in \mathcal{N}_i} \left(\mathbf{p}_i^t - (\mathbf{z}_{j,i}^-)^t - d_{i,j} \mathbf{u}_{i,j}^t + (\mathbf{z}_{i,j}^-)^t - \mathbf{p}_j^t + d_{j,i} \mathbf{u}_{j,i}^t \right), \quad \forall i \in (\mathcal{N}/\mathcal{A}), \tag{S.56}$$

By using the KKT condition (66c), we have $(\mathbf{z}_{i,j}^-)^t = \mathbf{p}_i^t$, $(\mathbf{z}_{j,i}^-)^t = \mathbf{p}_j^t$. Then (S.56) can be rewritten as

$$\mathbf{0} = \sum_{j \in \mathcal{N}_i} \left(\mathbf{p}_i^t - \mathbf{p}_j^t - d_{i,j} \mathbf{u}_{i,j}^t + \mathbf{p}_i^t - \mathbf{p}_j^t + d_{j,i} \mathbf{u}_{j,i}^t \right), \quad \forall i \in (\mathcal{N}/\mathcal{A}). \tag{S.57}$$

which shows that \mathbf{z}^t satisfy (S.50a).

We have thus established that if $(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ is a KKT solution of problem (15), which satisfies (66), then $(\mathbf{p}^t, \mathbf{u}^t)$ is a critical point of problem (6). Furthermore, we can demonstrate below that \mathbf{p}^t also qualifies as a critical point of the original problem (2).

Recall that the function $F_i(\mathbf{z}_i, \mathbf{u}_i)$ is defined as

$$F_i(\mathbf{z}_i, \mathbf{u}_i) = \frac{1}{2} \|\mathbf{Q}_i \mathbf{z}_i\|^2 - \mathbf{u}_i^T \mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i = \sum_{j \in \mathcal{N}_i} \frac{1}{2} \left[\|\mathbf{p}_i - \mathbf{p}_j\|^2 - d_{i,j} \mathbf{u}_{i,j}^T (\mathbf{p}_i - \mathbf{p}_j) \right], \quad \forall i \in \mathcal{N}. \tag{S.58}$$

Therefore, combining (S.48) with (S.58) yields that $\mathbf{u}_{i,j}^t$ is the optimal solution of the following problem

$$\min_{\mathbf{u}_{i,j} \in \mathcal{B}} -d_{i,j} \mathbf{u}_{i,j}^T (\mathbf{p}_i^t - \mathbf{p}_j^t).$$

From this, we can derive that

$$\mathbf{u}_{j,i}^t = -\mathbf{u}_{i,j}^t, \forall i \in \mathcal{N}, j \in \mathcal{N}_i. \quad (\text{S.59})$$

Using (S.59) and (S.50), we obtain that

$$\mathbf{p}_i^t - \mathbf{p}_j^t \in \partial \delta_{\mathcal{B}}(\mathbf{u}_{i,j}^t) \text{ for all } i \in \mathcal{N}, j \in \mathcal{N}_i.$$

Since $\delta_{\mathcal{B}}(\cdot)$ is a proper, lower semicontinuous, and convex function, it follows from [[45] Theorem 4.20, p. 104] that

$$\mathbf{u}_{i,j}^t \in \partial \delta_{\mathcal{B}}^*(\mathbf{p}_i^t - \mathbf{p}_j^t), \text{ for all } i \in \mathcal{N}, j \in \mathcal{N}_i.$$

Here, $\delta_{\mathcal{B}}^*$ represents the Fenchel conjugate of $\delta_{\mathcal{B}}$. From [[45], Example 2.31, p. 28], we also find that $\delta_{\mathcal{B}}^*(\cdot) = \|\cdot\|$. Hence $\mathbf{u}_{i,j}^t \in \partial \|\cdot\|(\mathbf{p}_i^t - \mathbf{p}_j^t)$ for all $i \in \mathcal{N}$ and $j \in \mathcal{N}_i$. Substituting this fact, along with (S.59) and (S.50c), into (S.50a), we can get

$$\mathbf{0} = \sum_{j \in \mathcal{N}_i} (\mathbf{p}_i^t - \mathbf{p}_j^t) - d_{i,j} \mathbf{u}_{i,j}^t, \quad \forall i \in (\mathcal{N}/\mathcal{A}), \quad (\text{S.60a})$$

$$\mathbf{u}_{i,j}^t \in \partial \|\cdot\|(\mathbf{p}_i^t - \mathbf{p}_j^t), \quad \forall i \in \mathcal{N}, j \in \mathcal{N}_i, \quad (\text{S.60b})$$

$$\mathbf{p}_j^t = \mathbf{a}_j, \quad \forall j \in \mathcal{A}. \quad (\text{S.60c})$$

Therefore, from (S.60) we know that \mathbf{p}^t is a critical point of the original nonconvex problem (2).

5. Proof of the Lemma 4

According to equations (62)-(64), it can be observed that the value of ζ^t decreases after each iteration. Combining this observation with the first part of Theorem 1, it can be concluded that for any $t > 1$, there exists an $s \in 1, 2, \dots, t-1$ such that $\zeta^s - \zeta^{s+1} \leq \frac{\zeta^0 - \underline{\zeta}}{t-1}$. By utilizing equation (61), we can let

$$K = (\zeta^0 - \underline{\zeta}) \cdot \max \left\{ \frac{1}{2}, \frac{\kappa_1 - 1}{2}, \frac{c(\kappa_1 - 1)}{2}, \frac{\rho}{4}, \frac{c(\kappa_1 - 6(N_{\max} + 1))}{2}, \frac{c\kappa_1 - 6(N_{\max} + 1)}{2c}, \frac{c\kappa_2 \cdot \tilde{\tau}_{\min}}{2N_{\text{sum}}n(c+1)^2} - \frac{(N_{\max} + 1)c\kappa_1}{2}, \frac{\rho}{4} - d_{\max}^2(\kappa_1 + \kappa_2) \right\}.$$

and we obtain

$$\begin{aligned} \|\mathbf{z}^{s+1} - \mathbf{z}_i^s\|^2 &< \frac{K}{t-1}, \quad \|\mathbf{u}_i^{s+1} - \mathbf{u}_i^s\|^2 < \frac{K}{t-1}, \quad \|\tilde{\mathbf{z}}_i^{s+1} - \mathbf{z}_i^{s+1} - (\tilde{\mathbf{z}}_i^s - \mathbf{z}_i^s)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 < \frac{K}{t-1}, \\ \|\mathbf{Q}_i(\tilde{\mathbf{z}}_i^{s+1} - \tilde{\mathbf{z}}_i^s) - \mathbf{D}_i(\mathbf{u}_i^s - \mathbf{u}_i^{s-1})\|^2 &< \frac{K}{t-1}, \quad \|\tilde{\mathbf{z}}_i^{s+1} - \mathbf{z}_i^s - (\tilde{\mathbf{z}}_i^s - \mathbf{z}_i^{s-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 < \frac{K}{t-1}. \end{aligned} \quad (\text{S.61})$$

From problem (17), we have

$$\mathbf{z}_i^{s+1} = \arg \min_{\mathbf{z}} F_i(\mathbf{z}_i, \mathbf{u}_i^s) + \langle \boldsymbol{\lambda}_i^s, \mathbf{A}_i \mathbf{z}_i \rangle + \frac{c}{2} \|\mathbf{z}_i - \mathbf{z}_i^s\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 + \delta_{\mathcal{X} \times \mathcal{Z}}(\mathbf{z}_i).$$

Then by applying the optimality condition, we can get

$$\mathbf{0} \in \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i^s) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^s + c \mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^{s+1} - \mathbf{z}_i^s) + \partial \delta_{\mathcal{X} \times \mathcal{Z}}(\mathbf{z}_i^{s+1}).$$

Letting

$$\mathbf{v}_1 = \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i^{s+1}) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^{s+1} - \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i^s) - \mathbf{A}_i^T \boldsymbol{\lambda}_i^s - c \mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^{s+1} - \mathbf{z}_i^s),$$

then we have

$$\begin{aligned} \|\mathbf{v}_1\| &= \left\| -\mathbf{Q}_i^T \mathbf{D}_i (\mathbf{u}_i^{s+1} - \mathbf{u}_i^s) + \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{s+1} - \boldsymbol{\lambda}_i^s) + c \mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^{s+1} - \mathbf{z}_i^s) \right\| \\ &\leq d_{\max} \sqrt{N_i + 1} \|\mathbf{u}_i^{s+1} - \mathbf{u}_i^s\| + \sqrt{N_i + 1} \|\boldsymbol{\lambda}_i^{s+1} - \boldsymbol{\lambda}_i^s\| + (1+c)(1+N_{\max}) \|\mathbf{z}_i^{s+1} - \mathbf{z}_i^s\|, \end{aligned} \quad (\text{S.62})$$

where the last inequality holds due to $\|\mathbf{Q}_i^T \mathbf{Q}_i\| = \|\mathbf{A}_i^T \mathbf{A}_i\| = N_i + 1$ and (S.14). Substituting (S.61) and (S.13) into (S.62), we have

$$\begin{aligned} \|\mathbf{v}_1\| &\leq \frac{\sqrt{K}}{\sqrt{t-1}} \left(d_{\max} \sqrt{N_i + 1} + \sqrt{N_i + 1} (\sqrt{N_i + 1} + c \sqrt{N_i + 1} + (1+c)(1+N_{\max})) \right. \\ &\quad \left. + (1+c)(1+N_{\max}) \right) \\ &\leq \frac{\sqrt{M_1}}{\sqrt{t-1}}, \end{aligned}$$

where $M_1 := K(d_{\max} \sqrt{N_i + 1} + \sqrt{N_i + 1} (\sqrt{N_i + 1} + c \sqrt{N_i + 1} + (1+c)(1+N_{\max})) + (1+c)(1+N_{\max}))^2 > 0$. Recall the \mathbf{u} update step, we have

$$\mathbf{u}_i^{s+1} = \arg \min_{\mathbf{u}} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i) + \frac{\rho}{2} \|\mathbf{u}_i - \mathbf{u}_i^s\|^2 + \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i),$$

By the optimality condition,

$$\mathbf{0} \in \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i^{s+1}) + \rho(\mathbf{u}_i^{s+1} - \mathbf{u}_i^s) + \partial \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i^{s+1}).$$

Letting

$$\mathbf{v}_2 = \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i^s) - \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i^{s+1}) - \rho(\mathbf{u}_i^{s+1} - \mathbf{u}_i^s) = -\rho(\mathbf{u}_i^{s+1} - \mathbf{u}_i^s), \quad (\text{S.63})$$

then we have

$$\mathbf{v}_2 \in \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^{s+1}, \mathbf{u}_i^{s+1}) + \partial \delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i^{s+1}). \quad (\text{S.64})$$

By the inequality (S.61), we get

$$\|\mathbf{v}_2\| \leq \frac{\sqrt{K}}{\sqrt{t-1}} \cdot \rho = \frac{M_2}{\sqrt{t}}, \quad (\text{S.65})$$

where $M_2 := K\rho^2 > 0$. Then the Lemma 1 holds for $\mathbf{v}_1, \mathbf{v}_2$ and $M := \max\{M_1, M_2\}$ and $(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ is a $M/\sqrt{t-1}$ -solution. It indicates that the proposed algorithm converges to a KKT stationary point of problem (15) at a sublinear rate.

6. Proof of the Theorem 1

Lemma 7 Suppose $c, \rho, \kappa_1, \kappa_2$ are chosen according to (62)-(64). Then the following statement holds true

$$\exists \underline{\varsigma} > -\infty \text{ s.t. } \varsigma^t \geq \underline{\varsigma}, \forall t > 0.$$

Proof: By using the update step (38) of the dual variable $\boldsymbol{\lambda}_i^{t+1}$, we have

$$\begin{aligned} & \mathcal{L}(\mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}, \mathbf{u}^{t+1}) \\ &= \sum_{i \in \mathcal{N}} F_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}) + \langle \boldsymbol{\lambda}_i^{t+1}, \mathbf{A}_i \mathbf{z}_i^{t+1} \rangle + \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 \\ &= \sum_{i \in \mathcal{N}} F_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}) + \frac{1}{c} \langle \boldsymbol{\lambda}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t \rangle + \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 \\ &= \sum_{i \in \mathcal{N}} \left[F_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}) + \frac{1}{2c} (\|\boldsymbol{\lambda}_i^{t+1}\|^2 - \|\boldsymbol{\lambda}_i^t\|^2 + \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2) + \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 \right]. \end{aligned} \quad (\text{S.66})$$

Adding (S.66) from $t = 1$ to T , we obtain

$$\begin{aligned} & \sum_{t=1}^T \mathcal{L}(\mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}, \mathbf{u}^{t+1}) \\ &= \sum_{t=1}^T \sum_{i \in \mathcal{N}} \left[F_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}) + \frac{c}{2} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 + \frac{1}{2c} (\|\boldsymbol{\lambda}_i^{t+1}\|^2 - \|\boldsymbol{\lambda}_i^t\|^2 + \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2) \right]. \end{aligned} \quad (\text{S.67})$$

From the definition of F_i it follows that

$$\begin{aligned} \sum_{i \in \mathcal{N}} F_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}) &= \sum_{i \in \mathcal{N}} \frac{1}{2} \|\mathbf{Q}_i \mathbf{z}_i^{t+1}\|^2 - (\mathbf{u}_i^{t+1})^t \mathbf{D}_i \mathbf{Q}_i \mathbf{z}_i^{t+1} \\ &= \sum_{i \in \mathcal{N}} \frac{1}{2} \|\mathbf{Q}_i \mathbf{z}_i^{t+1} - \mathbf{D}_i \mathbf{u}_i^{t+1}\|^2 - \frac{1}{2} \|\mathbf{D}_i \mathbf{u}_i^{t+1}\|^2 \\ &\geq \sum_{i \in \mathcal{N}} -\frac{d_{\max}^2}{2} \|\mathbf{u}_i^{t+1}\|^2 > -\infty, \forall t > 0, \end{aligned} \quad (\text{S.68})$$

where the last inequality is due to $\mathbf{u}_i^t \in \mathcal{B}^{N_i}, \forall i \in \mathcal{N}, t > 0$. Therefore, according to the definition (60) of potential function, (S.67) and (S.68), it shows that

$$\sum_{t=1}^T \varsigma^t > -\infty, \forall T > 0. \quad (\text{S.69})$$

When $c, \rho, \kappa_1, \kappa_2$ satisfy (62)-(64), Lemma 2 show that the potential function ς^t decreases at each iteration of Algorithm 1. Hence, by (S.69) we can infer that $\varsigma^t \geq \underline{\varsigma} > -\infty, \forall t > 0$. ■

Lemma 8 *Suppose the parameters c, ρ, κ_1 and κ_2 satisfy (62)-(64), then the iterative sequence $\{(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t)\}$ of Algorithm 1 satisfies*

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{z}_i^{t+1} - \mathbf{z}_i^t &\rightarrow \mathbf{0}, \quad \lim_{t \rightarrow \infty} \mathbf{A}_i \mathbf{z}_i^t \rightarrow \mathbf{0}, \\ \lim_{t \rightarrow \infty} \boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t &\rightarrow \mathbf{0}, \quad \lim_{t \rightarrow \infty} \mathbf{u}_i^{t+1} - \mathbf{u}_i^t \rightarrow \mathbf{0}, \forall i \in \mathcal{N}. \end{aligned}$$

Proof: Since \mathbf{W}_i is a positive definite diagonal matrix, combining (61) and Lemma 7, we have

$$\lim_{t \rightarrow \infty} \mathbf{z}_i^{t+1} - \mathbf{z}_i^t \rightarrow \mathbf{0}. \quad (\text{S.70})$$

Using the update step of $\boldsymbol{\lambda}_i^{t+1}$ (cf.(38)), we have from (S.10), (S.13) and Lemma 7 that

$$\lim_{t \rightarrow \infty} \mathbf{A}_i \mathbf{z}_i^{t+1} \rightarrow \mathbf{0}, \quad \lim_{t \rightarrow \infty} \boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t \rightarrow \mathbf{0}. \quad (\text{S.71})$$

By the inequality (61) and Lemma 7, we further obtain

$$\lim_{t \rightarrow \infty} \mathbf{u}_i^{t+1} - \mathbf{u}_i^t \rightarrow \mathbf{0}. \quad (\text{S.72})$$

■

Lemma 9 (Sufficient decrease condition) *Suppose the sequence $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}_{t \geq 1}$ is generated by Algorithm 1, $c\mathbf{B}_i^T \mathbf{B}_i$ takes the form of (25), and the conditions (62)-(64) are satisfied. Then we have*

$$\zeta^{t+1} - \zeta^t \leq \sum_{i \in \mathcal{N}} \left[-\min\{c, 1\} \cdot \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|^2 - \frac{\rho}{4} \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\|^2 - C_0 \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2 - C_1 \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|^2 \right],$$

where $C_0 := \frac{\min\{c\kappa_1 - 6(N_{\max} + 1), c(c\kappa_1 - 6(1+c)(N_{\max} + 1))\}}{\max\{3(N_i + 1), 3c(1+c)(1+N_{\max})\}}$ and $C_1 := \frac{c\kappa_2 \tilde{\tau}_{\min}}{2N_{\text{sum}}n(c+1)^2} - \frac{(N_{\max} + 1)c\kappa_1}{2}$. $d_{\max} := \max\{d_{i,j}, i \in \mathcal{N}, j \in \mathcal{N}_i\}$, $N_{\max} := \max\{N_i, i \in \mathcal{N}\}$, $N_{\text{sum}} := \sum_{i \in \mathcal{N}} N_i$ is the total number of neighboring nodes and $\tilde{\tau}_{\min} := \min\{(c+1)^2 N_i^2 + c^2 N_i + N_i, i \in \mathcal{N}\}$.

Proof: Substituting (S.14) into (S.13), we have

$$\begin{aligned} \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2 &\leq \max\{3(N_i + 1), 3c(1+c)(1+N_{\max})\} \cdot \left(\|\mathbf{Q}_i(\tilde{\mathbf{z}}_i^{t+1} - \tilde{\mathbf{z}}_i^t) - \mathbf{D}_i(\mathbf{u}_i^t - \mathbf{u}_i^{t-1})\|^2 \right. \\ &\quad \left. + \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1} - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^t)\|_{\mathbf{A}_i^T \mathbf{A}_i}^2 + \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t - (\tilde{\mathbf{z}}_i^t - \mathbf{z}_i^{t-1})\|_{\mathbf{B}_i^T \mathbf{B}_i}^2 \right), \forall i \in \mathcal{N}. \end{aligned} \quad (\text{S.73})$$

Combining (S.73), (29) with (61), we can derive the result. ■

Lemma 10 (Bounded sequence) *Suppose that conditions (62)-(64) are satisfied, the graph obtained by the network is connected, and there is at least one anchor sensor. Then the sequence $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t)\}_{t \geq 1}$ generated by Algorithm 1 is bounded.*

Proof: We establish the boundedness of $\{\mathbf{u}^t\}_{t \geq 1}$, $\{\mathbf{z}^t\}_{t \geq 1}$, $\{\boldsymbol{\lambda}^t\}_{t \geq 1}$, and $\{\tilde{\mathbf{z}}^t\}_{t \geq 1}$ sequentially:

1) It is evident from Algorithm 1 that $\{\mathbf{u}^t\}_{t \geq 1} \subset \mathcal{B}^{N_{N_i}}$, where \mathcal{B} represents the unit ball constraint. Hence, the sequence $\{\mathbf{u}^t\}_{t \geq 1}$ is bounded.

2) Using the update rule of λ_i^{t+1} , we obtain

$$c \sum_{t=1}^{\infty} \sum_{i \in \mathcal{N}} \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 = \sum_{t=1}^{\infty} \sum_{i \in \mathcal{N}} \|\lambda_i^{t+1} - \lambda_i^t\|^2 \leq \frac{1}{C_0} \sum_{t=1}^{\infty} (\varsigma^t - \varsigma^{t+1}) \leq \frac{1}{C_0} (\varsigma^1 - \underline{\varsigma}) < \infty. \quad (\text{S.74})$$

where the first inequality is due to Lemma 9 and the second inequality comes from Lemma 7. Consequently, $\{\|\mathbf{A}_i \mathbf{z}_i^t\|\}_{t \geq 1}$ is bounded for all $i \in \mathcal{N}$. Similarly, by using Lemma 9 and Lemma 7 again, we can show that $\{\|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|\}_{t \geq 1}$ is also bounded for all $i \in \mathcal{N}$. Recall that

$$\mathbf{Q}_i = [\mathbf{1}_{N_i}, \mathbf{0}_{N_i}, -\mathbf{I}_{N_i}] \otimes \mathbf{I}_n, \quad \mathbf{A}_i = [\mathbf{1}_{N_i}, -\mathbf{I}_{N_i}, \mathbf{0}_{N_i}] \otimes \mathbf{I}_n, \quad \forall i \in \mathcal{N}, \quad (\text{S.75})$$

then with simple algebraic manipulation, we can derive

$$[\mathbf{0}_{N_i}, \mathbf{0}_{N_i}, -\mathbf{I}_{N_i}] \otimes \mathbf{I}_n \cdot \mathbf{Q}_i^T = \mathbf{I}_{N_i n}, \quad [\mathbf{0}_{N_i}, \mathbf{0}_{N_i}, -\mathbf{I}_{N_i}] \otimes \mathbf{I}_n \cdot \mathbf{A}_i^T = \mathbf{0}_{N_i n}, \quad \forall i \in \mathcal{N}. \quad (\text{S.76})$$

In addition, substituting (29) into the update step of $\tilde{\mathbf{z}}_i^{t+1}$, we have

$$\mathbf{Q}_i^T \mathbf{Q}_i \mathbf{z}_i^t = -\mathbf{W}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t) + \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t - c \mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t - c \mathbf{A}_i^T \lambda_i^t, \quad \forall i \in \mathcal{N}. \quad (\text{S.77})$$

Multiplying both sides of (S.77) by $[\mathbf{0}_{N_i}, \mathbf{0}_{N_i}, -\mathbf{I}_{N_i}] \otimes \mathbf{I}_n$ and taking the norm, we obtain

$$\|\mathbf{Q}_i \mathbf{z}_i^t\| = \|[\mathbf{0}_{N_i}, \mathbf{0}_{N_i}, -\mathbf{I}_{N_i}] \otimes \mathbf{I}_n \cdot \mathbf{W}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t) + \mathbf{D}_i \mathbf{u}_i^t\| \leq 2\|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\| + d_{\max} \|\mathbf{u}_i^t\|,$$

where $d_{\max} = \max\{d_{i,j}, i \in \mathcal{N}, j \in \mathcal{N}_i\}$. The first equality is due to (S.76), while the second inequality follows the triangle inequality and (29). Since the right-hand side (rhs) of the above inequality is bounded, it follows that $\{\|\mathbf{Q}_i \mathbf{z}_i^t\|\}_{t \geq 1}$ is bounded for all $i \in \mathcal{N}$. Recall the definition of $\mathbf{z}_i = [\mathbf{p}_i^T, (\mathbf{z}_i^-)^T, (\mathbf{z}_i^+)^T]^T$ and $\mathbf{Q}_i, \mathbf{A}_j$ in (S.75), we have

$$\mathbf{Q}_i \mathbf{z}_i^t = \text{vec} \left(\mathbf{p}_i^t - (\mathbf{z}_{i,j}^+)^t, j \in \mathcal{N}_i \right), \quad \mathbf{A}_j \mathbf{z}_j^t = \text{vec} \left(\mathbf{p}_j^t - (\mathbf{z}_{j,i}^-)^t, i \in \mathcal{N}_j \right). \quad (\text{S.78})$$

By using the triangle inequality, for all $t \geq 0$, we have

$$\|(\mathbf{z}_{i,j}^+)^t\| \leq \|\mathbf{p}_i^t\| + \|\mathbf{p}_i^t - (\mathbf{z}_{i,j}^+)^t\|, \quad \|\mathbf{p}_j^t\| \leq \|(\mathbf{z}_{j,i}^-)^t\| + \|\mathbf{p}_j^t - (\mathbf{z}_{j,i}^-)^t\|. \quad (\text{S.79})$$

Recall the update step of \mathbf{z}_i^t , we have

$$(\mathbf{z}_{i,j}^+)^t = (\mathbf{z}_{j,i}^-)^t, \quad \forall i \in \mathcal{N}, j \in \mathcal{N}_i, \quad \text{and} \quad \mathbf{p}_i^t = \mathbf{a}_i, \quad \forall i \in \mathcal{A} \subseteq \mathcal{N}. \quad (\text{S.80})$$

Since we have at least one anchor, combining (S.79) with (S.80), we can get

$$\|\mathbf{p}_j^t\| \leq \|\mathbf{a}_i\| + \|\mathbf{p}_i^t - (\mathbf{z}_{i,j}^+)^t\| + \|\mathbf{p}_j^t - (\mathbf{z}_{j,i}^-)^t\|, \quad \forall j \in \mathcal{N}_i, i \in \mathcal{A}. \quad (\text{S.81})$$

Therefore, the sequence $\{\|\mathbf{p}_j^t\|\}_{t \geq 1}$ is bounded for all $j \in \mathcal{N}_i, i \in \mathcal{A}$ due to the rhs of (S.81) being bounded, which is a consequence of the boundedness of $\{\|\mathbf{Q}_i \mathbf{z}_i^t\|\}_{t \geq 1}$ and $\{\|\mathbf{A}_i \mathbf{z}_i^t\|\}_{t \geq 1}$ for all $i \in \mathcal{N}$ (see (S.78)). Using the definition of \mathbf{z}_j^t and the triangle inequality, we have

$$\|\mathbf{z}_j^t\| \leq \sum_{i \in \mathcal{N}_j} 3\|\mathbf{p}_j^t\| + \|\mathbf{p}_j^t - (\mathbf{z}_{j,i}^-)^t\| + \|\mathbf{p}_j^t - (\mathbf{z}_{j,i}^+)^t\|, \quad \forall j \in \mathcal{N}. \quad (\text{S.82})$$

Since the rhs of (S.82) is bounded, then the sequence $\{\|\mathbf{z}_j^t\|\}_{t \geq 1}$ is bounded for all $j \in \mathcal{N}_i, i \in \mathcal{A}$. By applying a similar argument as in the derivation of (S.79)-(S.82), we can show that \mathbf{z}_i^t is bounded in a neighborhood of the nodes $j \in \mathcal{N}_i, i \in \mathcal{A}$. Given that the graph is connected, we can thus conclude that the sequence $\{\mathbf{z}_i^t\}_{t \geq 1}$ is bounded for all $i \in \mathcal{N}$.

3) Rearranging the terms of (S.77) and taking the norm yields

$$\|\boldsymbol{\lambda}_i^t\| \leq \|\mathbf{A}_i^T \boldsymbol{\lambda}_i^t\| \leq \|\mathbf{Q}_i^T \mathbf{Q}_i \mathbf{z}_i^t\| + \|\mathbf{W}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t)\| + \|\mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t\| + \|c \mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t\|, \quad (\text{S.83})$$

where the first inequality holds due to the smallest eigenvalue of $\mathbf{A}_i \mathbf{A}_i^T$ being 1 from (S.75), and the second inequality follows from the triangle inequality. Since the rhs of (S.83) is bounded, it follows that $\{\boldsymbol{\lambda}_i^t\}_{t \geq 1}$ is bounded for all $i \in \mathcal{N}$.

4) Since $\|\tilde{\mathbf{z}}_i^{t+1}\| \leq \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\| + \|\mathbf{z}_i^t\|$, and the rhs is bounded, it follows that the sequence $\{\tilde{\mathbf{z}}_i^t\}_{t \geq 1}$ is bounded for all $i \in \mathcal{N}$. ■

Lemma 11 (Subgradient bound) *Let $\{\mathbf{y}^t = (\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}_{t \geq 1}$ be a sequence generated by Algorithm 1, κ_1, κ_2, c , and ρ are parameters used in the potential function ς^t . Then we have*

$$\mathbf{v}^{t+1} := (\mathbf{v}_{\mathbf{z}^{t+1}}, \mathbf{v}_{\mathbf{u}^{t+1}}, \mathbf{v}_{\boldsymbol{\lambda}^{t+1}}, \mathbf{v}_{\tilde{\mathbf{z}}^{t+1}}, \mathbf{v}_{\mathbf{u}^t}, \mathbf{v}_{\mathbf{z}^t}) \in \partial \varsigma^{t+1}, \quad \forall t \geq 0,$$

where

$$\begin{aligned} \mathbf{v}_{\mathbf{z}_i^{t+1}} &:= \mathbf{W}_i (\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}) + \mathbf{Q}_i^T \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t+1}) + (\kappa_2 + 1) \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t) \\ &\quad + (c \mathbf{A}_i^T \mathbf{A}_i + \mathbf{Q}_i^T \mathbf{Q}_i + c(\kappa_1 + \kappa_2) \mathbf{B}_i^T \mathbf{B}_i) (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t), \\ \mathbf{v}_{\mathbf{u}_i^{t+1}} &:= \frac{\rho}{2} (\mathbf{u}_i^t - \mathbf{u}_i^{t+1}), \quad \mathbf{v}_{\boldsymbol{\lambda}_i^{t+1}} := \frac{1}{c} (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t), \quad \mathbf{v}_{\tilde{\mathbf{z}}_i^{t+1}} := \kappa_1 \mathbf{A}_i^T (c \mathbf{A}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1}) + \boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t), \\ \mathbf{v}_{\mathbf{u}_i^t} &:= \frac{\rho}{2} (\mathbf{u}_i^t - \mathbf{u}_i^{t+1}), \quad \mathbf{v}_{\mathbf{z}_i^t} := c(\kappa_1 + \kappa_2) \mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^t - \mathbf{z}_i^{t+1}), \quad i \in \mathcal{N}. \end{aligned} \quad (\text{S.84})$$

Moreover, for every $t \geq 0$, it holds that

$$\begin{aligned} \|\mathbf{v}^{t+1}\| &\leq \sum_{i \in \mathcal{N}} \alpha_i \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\| + \sum_{i \in \mathcal{N}} \left(\rho + d_{\max} \sqrt{N_i + 1} \right) \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\| + \sum_{i \in \mathcal{N}} \beta_i \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\| \\ &\leq C_2 \|\mathbf{y}^{k+1} - \mathbf{y}^k\|, \end{aligned} \quad (\text{S.85})$$

where

$$\begin{aligned}\alpha_i &:= (c+1)(N_i+1) + 2(\kappa_1 + \kappa_2)(1+c)(1+N_{\max}) + \frac{2N_i(c+1)^2 \sqrt{nNN_{\text{sum}}(N_i+1)}}{\sqrt{\tilde{\tau}_{\min}}}, \\ \beta_i &:= \sqrt{N_i+1}(\kappa_1+1+\kappa_2) + \frac{1}{c}, \quad C_2 := \sqrt{3N} \cdot \max\{\alpha_i, \rho + d_{\max}\sqrt{N_i+1}, \beta_i, i \in \mathcal{N}\}.\end{aligned}\quad (\text{S.86})$$

Proof: By taking partial derivatives of ς^{t+1} with respect to $\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}, \tilde{\mathbf{z}}_i^{t+1}, \mathbf{u}_i^t, \mathbf{z}_i^t$ and using the update step of $\boldsymbol{\lambda}_i^{t+1}$, we obtain

$$\begin{aligned}\nabla_{\mathbf{z}_i^{t+1}} \varsigma^{t+1} &= \nabla_{\mathbf{z}_i} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}) + \kappa_2 \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t) + c(\kappa_1 + \kappa_2) \mathbf{B}_i^T \mathbf{B}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t), \\ \partial_{\mathbf{u}_i^{t+1}} \varsigma^{t+1} &= \partial_{\mathbf{u}_i} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}) + \frac{\rho}{2} (\mathbf{u}_i^{t+1} - \mathbf{u}_i^t), \\ \nabla_{\boldsymbol{\lambda}_i^{t+1}} \varsigma^{t+1} &= \nabla_{\boldsymbol{\lambda}_i} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}), \\ \nabla_{\tilde{\mathbf{z}}_i^{t+1}} \varsigma^{t+1} &= \mathbf{v}_{\tilde{\mathbf{z}}_i^{t+1}}, \quad \nabla_{\mathbf{u}_i^t} \varsigma^{t+1} = \mathbf{v}_{\mathbf{u}_i^t}, \quad \nabla_{\mathbf{z}_i^t} \varsigma^{t+1} = \mathbf{v}_{\mathbf{z}_i^t},\end{aligned}\quad (\text{S.87})$$

where $\mathbf{v}_{\tilde{\mathbf{z}}_i^{t+1}}, \mathbf{v}_{\mathbf{u}_i^t}$, and $\mathbf{v}_{\mathbf{z}_i^t}$ are defined in (S.84). Using the update step of $\tilde{\mathbf{z}}_i^{t+1}$, we have

$$\begin{aligned}\nabla_{\mathbf{z}_i} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}) &= \mathbf{W}_i (\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}) + \mathbf{Q}_i^T \mathbf{D}_i (\mathbf{u}_i^t - \mathbf{u}_i^{t+1}) + \mathbf{A}_i^T (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t) \\ &\quad + (c\mathbf{A}_i^T \mathbf{A}_i + \mathbf{Q}_i^T \mathbf{Q}_i) (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t).\end{aligned}\quad (\text{S.88})$$

Combining (S.88) with (S.87) yields $\mathbf{v}_{\mathbf{z}_i^{t+1}} = \nabla_{\mathbf{z}_i^{t+1}} \varsigma^{t+1}$. By the optimality condition of \mathbf{u} subproblem (18), we have

$$\mathbf{0} \in \partial_{\mathbf{u}_i} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^t) + \rho (\mathbf{u}_i^{t+1} - \mathbf{u}_i^t) = \partial_{\mathbf{u}_i} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}) + \rho (\mathbf{u}_i^{t+1} - \mathbf{u}_i^t).\quad (\text{S.89})$$

Combining (S.89) with (S.87) yields $\mathbf{v}_{\mathbf{u}_i^{t+1}} \in \partial_{\mathbf{u}_i^{t+1}} \varsigma^{t+1}$. Applying the update formula of $\boldsymbol{\lambda}_i^{t+1}$, yields

$$\mathbf{v}_{\boldsymbol{\lambda}_i^{t+1}} = \frac{1}{c} (\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t) = \mathbf{A}_i \mathbf{z}_i^{t+1} = \nabla_{\boldsymbol{\lambda}_i} \mathcal{L}_i(\mathbf{z}_i^{t+1}, \mathbf{u}_i^{t+1}, \boldsymbol{\lambda}_i^{t+1}) = \nabla_{\boldsymbol{\lambda}_i^{t+1}} \varsigma^{t+1}.$$

By using the expressions for \mathbf{v}^{t+1} and the triangle inequality, we obtain the following bound

$$\begin{aligned}\|\mathbf{v}^{t+1}\| &\leq \|\mathbf{W}_i\| \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\| + (2c(\kappa_1 + \kappa_2) \|\mathbf{B}_i^T \mathbf{B}_i\| + c\|\mathbf{A}_i^T \mathbf{A}_i\| + \|\mathbf{Q}_i^T \mathbf{Q}_i\|) \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\| \\ &\quad + \left(\sigma_{\max}^{1/2}(\mathbf{Q}_i \mathbf{Q}_i^T) \|\mathbf{D}_i\| + \rho \right) \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\| + \left(\sigma_{\max}^{1/2}(\mathbf{A}_i \mathbf{A}_i^T) (\kappa_2 + 1 + \kappa_1) + \frac{1}{c} \right) \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|,\end{aligned}\quad (\text{S.90})$$

where $\sigma_{\max}(\mathbf{Q}_i \mathbf{Q}_i^T)$ and $\sigma_{\max}(\mathbf{A}_i \mathbf{A}_i^T)$ denote the largest eigenvalue of $\mathbf{Q}_i \mathbf{Q}_i^T$ and $\mathbf{A}_i \mathbf{A}_i^T$, respectively.

By taking the square root of both sides of the inequality (S.38), and combining the result using the fact that $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{N} \|\mathbf{x}\|_2, \forall \mathbf{x} \in \mathbb{R}^N$, we obtain

$$\sum_{i \in \mathcal{N}} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\| \leq \frac{(c+1) \sqrt{nNN_{\text{sum}}}}{\sqrt{\tilde{\tau}_{\min}}} \sum_{i \in \mathcal{N}} \|\mathbf{A}_i (\mathbf{z}_i^{t+1} - \mathbf{z}_i^t)\|.\quad (\text{S.91})$$

Furthermore, using (29) and (S.75), we have

$$\|\mathbf{W}_i\| = 2(c+1)N_i, \sigma_{\max}(\mathbf{A}_i \mathbf{A}_i^T) = \sigma_{\max}(\mathbf{Q}_i \mathbf{Q}_i^T) = N_i + 1.\quad (\text{S.92})$$

Substituting (S.14), (S.91), and (S.92) into (S.90), we get

$$\|\mathbf{v}^{t+1}\| \leq \sum_{i \in \mathcal{N}} \alpha_i \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\| + \sum_{i \in \mathcal{N}} \left(\rho + d_{\max} \sqrt{N_i + 1} \right) \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\| + \sum_{i \in \mathcal{N}} \beta_i \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|, \quad (\text{S.93})$$

where α_i and β_i are defined in (S.86). By the fact that $\|\mathbf{x}\|_1 \leq \sqrt{3N} \|\mathbf{x}\|_2$, $\forall \mathbf{x} \in \mathbb{R}^{3N}$, we obtain

$$\|\mathbf{v}^{t+1}\| \leq C_2 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|,$$

where constant C_2 is defined in (S.86). This concludes the proof. \blacksquare

In the following, we denote by $\omega(\{\mathbf{x}^t\}_{t \geq 1})$ the set of limit points of the sequence $\{\mathbf{x}^t\}_{t \geq 1}$, and we define $\text{crit } \mathcal{L} := \{\mathbf{y} : \mathbf{0} \in \partial \mathcal{L}(\mathbf{y})\}$ as the set of all critical points of \mathcal{L} .

Lemma 12 (Properties of limit point set) *Let $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}_{t \geq 1}$ be a sequence generated by Algorithm*

1. Suppose that conditions (62)-(64) are satisfied, and let $\Omega := \omega(\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})\}_{t \geq 1})$, we have the following results:

- 1) *the set Ω is nonempty and compact;*
- 2) $\lim_{t \rightarrow \infty} \text{dist} \left[(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1}), \Omega \right] = 0$;
- 3) $\Omega \subseteq \{(\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda}, \mathbf{z}, \mathbf{u}, \mathbf{z}) : (\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda}) \in \text{crit } \mathcal{L}\}$;
- 4) *any critical point of \mathcal{L} is a KKT point of problem (15);*
- 5) *the potential function ζ^t is finite and constant on Ω .*

Proof: We prove the results item by item below.

- 1) Since $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t)\}_{t \geq 1}$ is bounded by Lemma 10, and thus Ω is nonempty and bounded. By the definition of Ω , it is closed and therefore compact.
- 2) As a consequence of the limit point definition.
- 3) Let $(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*)$ be a limit point of the sequence $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}_{t \geq 1}$, which exists since the sequence $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}_{t \geq 1}$ is bounded. Consequently, there exists a subsequence $\{(\mathbf{z}^{t_k}, \mathbf{u}^{t_k}, \boldsymbol{\lambda}^{t_k})\}_{k \geq 1}$ of $\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}_{t \geq 1}$ that converges to $(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*)$. Note that Lemma 9 with Lemma 7 implies

$$\sum_{i \in \mathcal{N}} \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|^2 \rightarrow 0, \quad \sum_{i \in \mathcal{N}} \|\mathbf{u}_i^{t+1} - \mathbf{u}_i^t\|^2 \rightarrow 0, \quad \sum_{i \in \mathcal{N}} \|\mathbf{z}_i^{t+1} - \mathbf{z}_i^t\|^2 \rightarrow 0, \quad \text{as } t \rightarrow \infty,$$

which means $\{(\tilde{\mathbf{z}}^{t_k}, \mathbf{u}^{t_k-1}, \mathbf{z}^{t_k-1})\}_{k \geq 1}$ converges to $(\mathbf{z}^*, \mathbf{u}^*, \mathbf{z}^*)$. Hence, $(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{z}^*, \mathbf{u}^*, \mathbf{z}^*) \in \Omega$, and from the continuity of ζ^t , it follows that

$$\lim_{k \rightarrow \infty} \zeta^{t_k} = \zeta^*.$$

On the other hand, from Lemma 9, Lemma 11, and Lemma 7, we know that $\mathbf{v}^{t_k} \in \partial \zeta^{t_k}$ and $\mathbf{v}^{t_k} \rightarrow \mathbf{0}$ as $k \rightarrow \infty$. The closeness property of $\partial \zeta$ (see [[32], Remark 1 (ii)]) implies that

$$\mathbf{0} \in \partial \zeta^*. \quad (\text{S.94})$$

In addition, according to (S.87) we also have

$$\partial\zeta^* = \partial\mathcal{L}(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*). \quad (\text{S.95})$$

Combining (S.94) with (S.95), we conclude that $(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*)$ is a critical point of \mathcal{L} .

4) Let $(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*) \in \text{crit } \mathcal{L}$. Thanks to the separable property of \mathcal{L} about node i , we have

$$\mathbf{0} = \nabla_{\mathbf{z}_i} \mathcal{L}_i(\mathbf{z}_i^*, \mathbf{u}_i^*, \boldsymbol{\lambda}_i^*) = \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^*, \mathbf{u}_i^*) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^* + c \mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^*, \quad (\text{S.96})$$

$$\mathbf{0} \in \partial_{\mathbf{u}_i} \mathcal{L}_i(\mathbf{z}_i^*, \mathbf{u}_i^*, \boldsymbol{\lambda}_i^*) = \nabla_{\mathbf{u}_i} F_i(\mathbf{z}_i^*, \mathbf{u}_i^*) + \partial\delta_{\mathcal{B}^{N_i}}(\mathbf{u}_i^*), \quad (\text{S.97})$$

$$\mathbf{0} = \nabla_{\boldsymbol{\lambda}_i} \mathcal{L}_i(\mathbf{z}_i^*, \mathbf{u}_i^*, \boldsymbol{\lambda}_i^*) = \mathbf{A}_i \mathbf{z}_i^*. \quad (\text{S.98})$$

Thus, it follows from (S.96) and (S.98) that

$$\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^*, \mathbf{u}_i^*) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^* = \mathbf{0}. \quad (\text{S.99})$$

Since $F_i(\mathbf{z}_i, \mathbf{u}_i) + \langle \boldsymbol{\lambda}_i, \mathbf{A}_i \mathbf{z}_i \rangle$ is convex over \mathcal{X} and \mathcal{Z} for any fixed \mathbf{u}_i , then the above expression (S.99) implies that \mathbf{z}_i^* also satisfies

$$\mathbf{z}_i^* \in \arg \min_{\substack{\mathbf{z}_i \in \mathcal{Z} \\ \mathbf{x} \in \mathcal{X}}} \lim F_i(\mathbf{z}_i, \mathbf{u}_i^*) + \langle \boldsymbol{\lambda}_i^*, \mathbf{A}_i \mathbf{z}_i \rangle. \quad (\text{S.100})$$

Combining (S.97), (S.98), and (S.100), we obtain $(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*)$ as a KKT point of problem (15).

5) The sequence $\{\zeta^t\}_{t \geq 1}$ decreases by Lemma 9 and is bounded from below by Lemma 7, thereby implying its convergence to some finite limit ζ^* . It follows that ζ^t is constant on Ω . ■

To achieve our main goal to establish the global convergence of the whole sequence, we recall the following key result obtained in the existing literature [32]–[36].

Lemma 13 (Uniform KL property) *Let Ω be a compact set and let $\sigma : \mathbb{R}^d \rightarrow (-\infty, \infty]$ a proper and lower semicontinuous function. Assume that σ is constant on Ω , and satisfies the KL property at each point of Ω . Then there exist $\varepsilon > 0, \eta > 0$ and $\varphi \in \Phi_\eta$ such that for all \bar{u} in Ω , and all u in the following intersection*

$$\left\{ u \in \mathbb{R}^d : \text{dist}(u, \Omega) < \varepsilon \right\} \cap [\sigma(\bar{u}) < \sigma(u) < \sigma(\bar{u}) + \eta],$$

one has,

$$\varphi'(\sigma(u) - \sigma(\bar{u})) \text{dist}(0, \partial\sigma(u)) \geq 1. \quad (\text{S.101})$$

We can now conveniently summarize our convergence results.

Lemma 14 (Global convergence) *Let $\{\mathbf{y}^t = (\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)\}_{t \geq 1}$ be a sequence generated by Algorithm 1. Suppose that conditions (62)-(64) are satisfied, then the following statements are true:*

1) *the sequence $\{\mathbf{y}^t\}_{t \geq 1}$ is bounded and has finite length, namely,*

$$\sum_{t=1}^{\infty} \|\mathbf{y}^{t+1} - \mathbf{y}^t\| < +\infty;$$

2) *the sequence $\{\mathbf{y}^t\}_{t \geq 0}$ converges to a KKT point of the problem (15).*

Proof: As in Lemma 12, we denote by $\Omega := \omega(\{(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})\}_{t \geq 1})$, which is a nonempty and compact set. From the proof of Lemma 12 and the continuity of ζ^t , it follows that

$$\lim_{t \rightarrow \infty} \zeta^t = \zeta^*, \text{ for all } (\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*, \mathbf{z}^*, \mathbf{u}^*, \mathbf{z}^*) \in \Omega. \quad (\text{S.102})$$

Moreover, from Lemma 9, we have

$$C_3 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2 \leq \zeta^t - \zeta^{t+1}, \quad \forall t \geq 1. \quad (\text{S.103})$$

where $C_3 := \min\{c, 1, \frac{\rho}{4}, C_0\}$. We consider two cases.

1) If there exists an integer $\bar{t} \geq 0$ such that $\zeta^{\bar{t}} = \zeta^*$, then using the decreasing property obtained by Lemma 9 and (S.103), we have

$$C_3 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2 \leq \zeta^t - \zeta^{t+1} \leq \zeta^{\bar{t}} - \zeta^* = 0, \quad t \geq \bar{t}.$$

Thus, $\mathbf{y}^{t+1} = \mathbf{y}^t$ for any $t \geq \bar{t}$, and it is clear that $\sum_{t=1}^{\infty} \|\mathbf{y}^{t+1} - \mathbf{y}^t\| < +\infty$ holds.

2) Since ζ^t is non-increasing from Lemma 9, we have $\zeta^t \geq \zeta^*$ for all t . Therefore, let us assume that $\zeta^t > \zeta^*$ for all t . Again from (S.102), we know that for any $\eta > 0$, there exists $t_0 > 0$ such that

$$\zeta^t < \zeta^* + \eta, \quad t \geq t_0.$$

According to Lemma 12, we know that $\lim_{t \rightarrow \infty} \text{dist}[(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1}), \Omega] = 0$. This implies that for any $\varepsilon > 0$, there exists $t_1 \geq 1$ such that $\text{dist}[(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1}), \Omega] < \varepsilon$ for all $t \geq t_1$. Summing up all these facts, we obtain the following inequalities for any $\eta, \varepsilon > 0$

$$\text{dist}[(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t, \tilde{\mathbf{z}}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1}), \Omega] < \varepsilon, \quad \text{and} \quad \zeta^* < \zeta^t < \zeta^* + \eta, \quad \text{for all } t \geq t_2 := \max\{t_0, t_1\}.$$

According to Lemma 12, it has been established that ζ is a constant on Ω . Moreover, we note that ζ^t is semi-algebraic (a polynomial function) and satisfies the KL property (see [[34], Theorem 6.1]). Therefore, we can apply Lemma 13 with Ω . Consequently, for any $t \geq t_2$, we have

$$\varphi'(\zeta^t - \zeta^*) \cdot \text{dist}(\mathbf{0}, \partial \zeta^t) \geq 1. \quad (\text{S.104})$$

Due to the concavity of φ , we get

$$\varphi(\zeta^t - \zeta^*) - \varphi(\zeta^{t+1} - \zeta^*) \geq \varphi'(\zeta^t - \zeta^*)(\zeta^t - \zeta^{t+1}). \quad (\text{S.105})$$

From (S.85) of Lemma 11, we also have

$$\text{dist}(\mathbf{0}, \partial\zeta^t) \leq C_2 \|\mathbf{y}^t - \mathbf{y}^{t-1}\|, \quad C_2 > 0.$$

By combining this with (S.105), (S.104), and $\varphi'(\zeta^t - \zeta^*) > 0$, we obtain

$$\begin{aligned} \zeta^t - \zeta^{t+1} &\leq \frac{\varphi(\zeta^t - \zeta^*) - \varphi(\zeta^{t+1} - \zeta^*)}{\varphi'(\zeta^t - \zeta^*)} \\ &\leq C_2 \|\mathbf{y}^t - \mathbf{y}^{t-1}\| \cdot [\varphi(\zeta^t - \zeta^*) - \varphi(\zeta^{t+1} - \zeta^*)]. \end{aligned} \quad (\text{S.106})$$

For convenience, we define the following for two arbitrary nonnegative integers p and q

$$\Delta_{p,q} := \varphi(\zeta^p - \zeta^*) - \varphi(\zeta^q - \zeta^*).$$

Combining (S.106) with (S.103), we conclude that for any $t \geq t_2$, the following inequality holds

$$C_3 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2 \leq \zeta^t - \zeta^{t+1} \leq C_2 \|\mathbf{y}^t - \mathbf{y}^{t-1}\| \Delta_{t,t+1},$$

Using the fact that $2\sqrt{ab} \leq a + b$ for all $a, b \geq 0$, we can infer

$$2\|\mathbf{y}^{t+1} - \mathbf{y}^t\| \leq \|\mathbf{y}^t - \mathbf{y}^{t-1}\| + \frac{C_2}{C_3} \Delta_{t,t+1}. \quad (\text{S.107})$$

Let us now prove that for any $t > t_2$, the following inequality holds

$$\sum_{k=t_2+1}^t \|\mathbf{y}^{k+1} - \mathbf{y}^k\| \leq \|\mathbf{y}^{t_2+1} - \mathbf{y}^{t_2}\| + \frac{C_2}{C_3} \Delta_{t_2,t+1}.$$

Summing up (S.107) for $k = t_2 + 1, t_2 + 2, \dots, t$ yields

$$\begin{aligned} 2 \sum_{k=t_2+1}^t \|\mathbf{y}^{k+1} - \mathbf{y}^k\| &\leq \sum_{k=t_2+1}^t \|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \frac{C_2}{C_3} \sum_{k=t_2+1}^t \Delta_{k,k+1} \\ &\leq \sum_{k=t_2+1}^{t+1} \|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \frac{C_2}{C_3} \sum_{k=t_2+1}^t \Delta_{k,k+1} \\ &\leq \sum_{k=t_2+1}^t \|\mathbf{y}^{k+1} - \mathbf{y}^k\| + \|\mathbf{y}^{t_2+1} - \mathbf{y}^{t_2}\| + \frac{C_2}{C_3} \Delta_{t_2+1,t+1}, \end{aligned}$$

where the last inequality follows from the fact that $\Delta_{p,q} + \Delta_{q,r} = \Delta_{p,r}$ for arbitrary nonnegative integers p, q , and r . Since $\varphi \geq 0$, we have for any $t > t_2$ that

$$\sum_{k=t_2+1}^t \|\mathbf{y}^{k+1} - \mathbf{y}^k\| \leq \|\mathbf{y}^{t_2+1} - \mathbf{y}^{t_2}\| + \frac{C_2}{C_3} \cdot \varphi(\zeta^{t_2+1} - \zeta^*).$$

As the right hand-side of the inequality above does not depend on t at all, it immediately follows that the sequence $\{\mathbf{y}^t\}_{t \in \mathbb{N}}$ has finite length, that is

$$\sum_{t=1}^{\infty} \|\mathbf{y}^{t+1} - \mathbf{y}^t\| < \infty.$$

This means that the sequence $\{\mathbf{y}^t\}_{t \geq 1}$ is a Cauchy sequence and hence a convergent sequence.

Thanks to Lemma 12, there exists $\mathbf{y}^* \in \text{crit}\mathcal{L}$ such that $\lim_{t \rightarrow \infty} \mathbf{y}^t = \mathbf{y}^*$ and \mathbf{y}^* is also a KKT point of problem (15). We complete the proof. \blacksquare

Lemma 15 *Suppose $c\mathbf{B}_i^T\mathbf{B}_i$ takes the form of (25), and the sequence $\{(\mathbf{z}_i^t, \mathbf{u}_i^t, \boldsymbol{\lambda}_i^t)\}$ is generated by the Algorithm 1. Then we have*

$$\begin{aligned} & \sum_{i \in \mathcal{N}} \|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t - (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t))\|^2 \\ & \leq \sum_{i \in \mathcal{N}} \left[\sigma_1 \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|^2 + \sigma_2 \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|^2 + 3c^2 (N_{\max} + 1)^2 \|\mathbf{A}_i \mathbf{z}_i^t\|^2 \right], \end{aligned}$$

where

$$\begin{aligned} \sigma_1 &= 3(2(c+1)N_{\max} - 1)^2 + 6 \max \left\{ (1+c)^2, \left(1 + \frac{1}{c}\right)^2 \right\}, \\ \sigma_2 &= \frac{3}{2} \max \left\{ (1+c)^2, \left(1 + \frac{1}{c}\right)^2 \right\}. \end{aligned} \tag{S.108}$$

Proof: Using (35) and the definition of the F_i given in problem (15), we have

$$\begin{aligned} \nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t &= \mathbf{Q}_i^T \mathbf{Q}_i \mathbf{z}_i^t - \mathbf{Q}_i^T \mathbf{D}_i \mathbf{u}_i^t + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t \\ &= -\mathbf{W}_i(\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t) + c\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t. \end{aligned}$$

Hence, with the above equality, we have

$$\begin{aligned} & \|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t - (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t))\| \\ &= \|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1}) + \text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1}) - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t + (\mathbf{W}_i(\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t) + c\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t))\| \\ &\leq \|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1})\| + \|(\mathbf{I}_i - \mathbf{W}_i)(\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t) - c\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t\|, \end{aligned} \tag{S.109}$$

where the inequality dues to the triangle inequality and the nonexpansive property of the projection operator [46]. By completing the square and using the Cauchy-Schwarz inequality, we further obtain

$$\begin{aligned} & \|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t - (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t))\|^2 \\ & \leq 3\|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1})\|^2 + 3c^2 \|\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t\|^2 + 3\|(\mathbf{I}_i - \mathbf{W}_i)(\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t)\|^2. \end{aligned} \tag{S.110}$$

Now, let us consider solving $\text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1})$. Since

$$\begin{aligned} \text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1}) &= \arg \min_{\mathbf{z}} \frac{1}{2} \sum_{i \in \mathcal{N}} \|\mathbf{z}_i - \tilde{\mathbf{z}}_i^{t+1}\|^2 \\ &\text{subject to } \mathbf{z} \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}. \end{aligned}$$

By a similar argument as the proof of Remark 2, we can derive

$$\text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1}) = \widetilde{\mathbf{W}}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^{t+1}) + \tilde{\mathbf{z}}_i^{t+1}, \quad (\text{S.112})$$

where $\tilde{\mathbf{x}}_{i,i}^{t+1} = \mathbf{a}_i$ if $i \in \mathcal{A}$ and

$$\widetilde{\mathbf{W}}_i = \frac{1}{2} \cdot \mathbf{Diag} \left(\left[0, (c+1) \cdot \mathbf{1}_{N_i}^T, \frac{c+1}{c} \cdot \mathbf{1}_{N_i}^T \right] \right) \otimes \mathbf{I}_n. \quad (\text{S.113})$$

Then, by (S.112) and Cauchy–Schwartz inequality, we have

$$\begin{aligned} \|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\tilde{\mathbf{z}}_i^{t+1})\|^2 &= \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1} - \widetilde{\mathbf{W}}_i (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t + \mathbf{z}_i^t - \mathbf{z}_i^{t+1})\|^2 \\ &\leq 2 \left\| \left(\mathbf{I}_i + \widetilde{\mathbf{W}}_i \right) (\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}) \right\|^2 + 2 \|\widetilde{\mathbf{W}}_i (\mathbf{z}_i^t - \mathbf{z}_i^{t+1})\|^2 \\ &\leq 2 \|\mathbf{I}_i + \widetilde{\mathbf{W}}_i\|^2 \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|^2 + 2 \|\widetilde{\mathbf{W}}_i\|^2 \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|^2. \end{aligned} \quad (\text{S.114})$$

For the third part of (S.110), by using the Cauchy–Schwartz inequality again

$$\|(\mathbf{I}_i - \mathbf{W}_i) (\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t)\|^2 \leq \|\mathbf{I}_i - \mathbf{W}_i\|^2 \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|^2. \quad (\text{S.115})$$

Substituting (S.114)-(S.115) into (S.110), we have

$$\begin{aligned} &\|\mathbf{z}_i^t - \text{proj}_{\mathcal{X}, \mathcal{Z}}(\mathbf{z}_i^t - (\nabla_{\mathbf{z}_i} F_i(\mathbf{z}_i^t, \mathbf{u}_i^t) + \mathbf{A}_i^T \boldsymbol{\lambda}_i^t))\|^2 \\ &\leq \left(6 \|\mathbf{I}_i + \widetilde{\mathbf{W}}_i\|^2 + 3 \|\mathbf{I}_i - \mathbf{W}_i\|^2 \right) \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|^2 + 6 \|\widetilde{\mathbf{W}}_i\| \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|^2 + 3c^2 \|\mathbf{A}_i^T \mathbf{A}_i \mathbf{z}_i^t\|^2 \\ &= \sigma_1 \|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^{t+1}\|^2 + \sigma_2 \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|^2 + 3c^2 (N_{\max} + 1)^2 \|\mathbf{A}_i \mathbf{z}_i^t\|^2, \end{aligned}$$

where the last equality follows from (26), (29) and (S.113), and σ_1, σ_2 are defined by (S.108). Hence we finish the proof of Lemma 15. \blacksquare

Note that it follows from (65), Lemma 8, and Lemma 15 that the sequence $\{(\mathbf{z}_i, \mathbf{u}_i, \boldsymbol{\lambda}_i)\}$ generated by SP-ADMM algorithm converges to critical point of the original problem (15) when the parameters satisfy (62)-(64).

Finally, we prove the third part of the Theorem 1. Let us scale the upper bound of $\mathcal{F}(\mathbf{z}, \mathbf{u}, \boldsymbol{\lambda})$ even further. Using the Cauchy-Schwarz inequality, we have

$$\|\mathbf{A}_i \mathbf{z}_i^t\|^2 = \|\mathbf{A}_i (\mathbf{z}_i^t - \mathbf{z}_i^{t+1} + \mathbf{z}_i^{t+1})\|^2 \leq 2 \|\mathbf{A}_i (\mathbf{z}_i^t - \mathbf{z}_i^{t+1})\|^2 + 2 \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2. \quad (\text{S.116})$$

Together (65) with Lemma 15 and (S.116), we have

$$\begin{aligned}
& \mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) \tag{S.117} \\
& \leq \sum_{i \in \mathcal{N}} \sigma_1 \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|^2 + \sigma_2 \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|^2 + \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2 + 2\sigma_3 \|\mathbf{A}_i(\mathbf{z}_i^t - \mathbf{z}_i^{t+1})\|^2 + 2\sigma_3 \|\mathbf{A}_i \mathbf{z}_i^{t+1}\|^2 \\
& \leq \sum_{i \in \mathcal{N}} \sigma_1 \|\tilde{\mathbf{z}}_i^{t+1} - \mathbf{z}_i^t\|^2 + (\sigma_2 + 2\sigma_3(N_{\max} + 1)) \|\mathbf{z}_i^t - \mathbf{z}_i^{t+1}\|^2 + \|\mathbf{u}_i^t - \mathbf{u}_i^{t-1}\|^2 + \frac{2\sigma_3}{c^2} \|\boldsymbol{\lambda}_i^{t+1} - \boldsymbol{\lambda}_i^t\|^2,
\end{aligned}$$

where σ_1 and σ_2 are defined in (S.108), $\sigma_3 = 3c^2(N_{\max} + 1)^2 + 1$. The last inequality holds due to the update step of $\boldsymbol{\lambda}_i^{t+1}$ and $\|\mathbf{A}_i^T \mathbf{A}_i\| = N_i + 1$. Matching the bounds of Lemma 9 with those of (S.117), we obtain

$$\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) \leq \epsilon(\varsigma^t - \varsigma^{t+1}), \tag{S.118}$$

where $\epsilon = \frac{\min\{c, 1, \frac{\rho}{4}, C_0, C_1, \frac{\rho}{4} - d_{\max}^2(\kappa_1 + \kappa_2)\}}{\max\{\sigma_1, \sigma_2 + 2\sigma_3(1 + N_{\max}), 1, \frac{2\sigma_3}{c^2}\}}$. Suppose that $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ reaches the lower bound ϵ_1 for the first time in step T , with Lemma 7 and add up the inequality (S.118) of the previous T step, we can get

$$\begin{aligned}
\epsilon_1 & \leq \frac{1}{T-1} \sum_{t=1}^T \mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t) \\
& \leq \frac{1}{T-1} \epsilon(\varsigma^1 - \varsigma^{T+1}) \\
& \leq \frac{1}{T-1} \epsilon(\varsigma^1 - \underline{\varsigma}) = \frac{\epsilon_2}{T-1}. \tag{S.119}
\end{aligned}$$

because of $\epsilon_2 = \epsilon(\varsigma^1 - \underline{\varsigma}) > 0$ is a constant, we can deduce that $\mathcal{F}(\mathbf{z}^t, \mathbf{u}^t, \boldsymbol{\lambda}^t)$ converges at $\mathcal{O}(1/T)$ rate.