

## Distributed Augmented Lagrangian Method for Link-Based Resource Sharing Problems of Multi-Agent Systems

Ananduta, Wicak; Nedic, Angelia; Ocampo-Martinez, Carlos

**DOI**

[10.1109/TAC.2021.3092561](https://doi.org/10.1109/TAC.2021.3092561)

**Publication date**

2021

**Document Version**

Accepted author manuscript

**Published in**

IEEE Transactions on Automatic Control

**Citation (APA)**

Ananduta, W., Nedic, A., & Ocampo-Martinez, C. (2021). Distributed Augmented Lagrangian Method for Link-Based Resource Sharing Problems of Multi-Agent Systems. *IEEE Transactions on Automatic Control*, 67(6), 3067-3074. <https://doi.org/10.1109/TAC.2021.3092561>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Distributed Augmented Lagrangian Method for Link-Based Resource Sharing Problems of Multi-Agent Systems

Wicak Ananduta, Angelia Nedić, and Carlos Ocampo-Martinez

**Abstract**—A multi-agent optimization problem motivated by the management of energy systems is discussed. The associated cost function is separable and convex although not necessarily strongly convex and there exist edge-based coupling equality constraints. In this regard, we propose a distributed algorithm based on solving the dual of the augmented problem. Furthermore, we consider that the communication network might be time-varying and the algorithm might be carried out asynchronously. The time-varying nature and the asynchronicity are modeled as random processes. Then, we show the convergence and the convergence rate of the proposed algorithm under the aforementioned conditions.

**Index Terms**—multi-agent optimization, stochastic time-varying network, asynchronous method

## I. INTRODUCTION

We consider an optimization problem of multi-agent systems. Specifically, the agents in the network cooperatively optimize a separable convex cost function subject to convex local constraints and equality coupling constraints. Furthermore, the set of decision variables of each agent is partitioned into shared decisions, i.e., variables that are involved in the coupling constraints and can be shared with other agents, and private decisions, i.e., variables that must be kept private and cannot be shared with other agents. The objective of the paper is to develop a distributed method for this problem. In addition, we also consider imperfect operation where the communication network might be time-varying and the algorithm might not be carried out synchronously.

The optimization problem considered is mainly motivated by the economic dispatch of large-scale energy systems [1] and belongs to a subclass of network flow problems [2], where we seek an optimal flow of certain goods from some sources to some sinks. Specifically, it is a convex network flow problem with a particular control structure, where each node has a computational unit and these units cooperatively solve the problem of the network. As a convex network flow problem, it represents an optimization problem of flow-based networks, such as electrical [1], thermal energy [3] and water networks [4]. Moreover, the problem also represents a convex relaxation of network flow problems of indivisible goods.

W. Ananduta is with the Delft Center for Systems and Control, TU Delft, Netherlands. C. Ocampo-Martinez is with Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Barcelona, Spain. A. Nedić is with School of Electrical, Computer and Energy Engineering, Arizona State University, Arizona, USA (emails: w.ananduta@tudelft.nl, carlos.ocampo@upc.edu, angelia.nedich@asu.edu)

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675318 (INCITE).

Solutions to a convex relaxation can be used as a lower bound of the optimal solution to the original problem.

The first challenge of developing a distributed method to solve the problem considered is how to deal with the coupling constraints. To address this issue, we employ the Lagrangian relaxation [5]–[7]. The main idea of this concept is to relax the coupling constraints such that the relaxed problem is decomposable. In this regard, Lagrange multipliers associated with the coupling constraints are introduced. In the dual problem, we aim to maximize such multipliers. Many distributed optimization methods, including those for energy management problems, e.g., [8]–[11], are developed based on solving the dual problem. Such distributed algorithms are iterative and require the exchanging of information. In particular, the agents that are coupled through link-based constraints must communicate certain information at each iteration. In this article, we consider the case when the information exchange process might be imperfect. In particular, we study the possibility of having a randomly time-varying communication network and asynchronous updates, which are relevant to the applications that we consider [12].

Therefore, in this paper, we develop a distributed optimization algorithm suitable for the previously explained problem. The algorithm is based on the Lagrange dual approach. Furthermore, we also consider that the communication network might be stochastically time-varying and the algorithm can be implemented asynchronously. Then, we show analytically that the sequence generated by the proposed algorithm converges to an optimal solution almost surely with the rate of  $\mathcal{O}(1/k)$ .

Now, we position the contributions of this work with respect to the existing literature. As previously mentioned, the problem that we consider is suitable to be decomposed using the Lagrangian dual approach. Although it is possible to reformulate the problem into a consensus-based problem [13], the latter approach can become impractical when the number of agents is large because the information that must be exchanged is unnecessarily large. In order to deal with a larger class of cost functions, particularly those that are not necessarily strongly convex, we consider augmenting the problem. In this regard, the proposed algorithm is more closely related to the accelerated distributed augmented Lagrangian (ADAL) method, discussed in [14], [15], than to the alternating direction method of multipliers (ADMM) [5], [7]. Similar to the ADAL method, we use some information from the neighbors in the local optimization step and require a convex combination step to update the primal variable. Differently, in the proposed method, each agent only performs

a convex combination step to update the shared variables instead of all the decisions. Moreover, since we consider a different augmented Lagrangian function, the condition of the step size, which guarantees convergence, is also different. Finally, we also note that the ADAL method in [14], [15] considers perfect communication, i.e., with a fixed neighbor-to-neighbor communication graph.

In this work, we are interested in developing a distributed method that works over a time-varying communication graph and asynchronous updates, implying imperfect information exchanges. To that end, we consider the communication network and asynchronous updates as random processes, similarly to the work in [16]–[19]. It is important to note that the distributed algorithm developed in this manuscript is different from those in [16]–[19] as they consider the ADMM approach and, to the best of our knowledge, the ADAL approach that we consider has not been employed on stochastically time-varying networks. We note that, unlike the ADAL approach, distributed ADMM methods require the introduction of auxiliary variables associated with the coupled variables, as shown in [17], [18]. In addition, for energy management problems, distributed methods that have been proposed, e.g., [9]–[11], [21]–[26] typically assume a perfect communication process, i.e., the necessary information required to execute the updates is available at each iteration. Therefore, for the considered applications, this technical note provides a more resilient distributed method than those in the aforementioned papers in dealing with potential communication problems.

## II. PROBLEM FORMULATION

### A. Preliminaries

*Notations:* We consider all vectors as column vectors. A stack of column vectors  $x_i$ , for all  $i \in \mathcal{N} = \{1, 2, \dots, n\}$ , denoted by  $[x_i]_{i \in \mathcal{N}}$ , is also a column vector. The inner product of vectors  $x$  and  $y$  is denoted by  $\langle x, y \rangle$ . The Euclidean norm of vector  $x$  is denoted by  $\|x\|_2$ . Moreover, for a diagonal matrix  $D \in \mathbb{R}^{d \times d}$ , the square of a weighted vector norm induced by  $D$  is denoted by  $\|\cdot\|_D^2$ , i.e.,  $\|x\|_D^2 = \langle x, Dx \rangle$ , for a vector  $x$ . The all-ones vector with the size of  $n$  is denoted by  $\mathbf{1}_n$  whereas the identity matrix with the size  $n \times n$  is denoted by  $I_n$ . Furthermore, the block-diagonal operator, which construct a block diagonal matrix of the arguments, is denoted by  $\text{blkdiag}(\cdot)$ .

*Definition 1: (Convexity)* A differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex, if, for any  $x, y \in \mathbb{R}^n$ , it holds that

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle.$$

*Definition 2: (Strong convexity [27, Theorem 5.24.iii])* A differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is strongly convex with strong convexity constant  $m > 0$ , if, for any  $x, y \in \mathbb{R}^n$ , it holds that

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq m\|y - x\|^2.$$

### B. Multi-agent optimization

We consider a group of interconnected agents that is represented by an undirected graph  $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ , where  $\mathcal{N} =$

$\{1, \dots, n\}$  denotes the set of agents and  $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$  denotes the set of links that connect the agents, i.e.,  $\{i, j\} \in \mathcal{E}$  means that agent  $i$  is coupled with agent  $j$  in a constraint. Furthermore, denote the set of neighbors of agent  $i$  by  $\mathcal{N}_i = \{j : \{i, j\} \in \mathcal{E}\}$ . The optimization problem that all agents consider to solve cooperatively is

$$\begin{aligned} & \underset{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}}{\text{minimize}} && \sum_{i=1}^n (f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i)) \end{aligned} \quad (1a)$$

$$\text{subject to } \mathbf{v}_i^j + \mathbf{v}_j^i = 0, \quad \forall j \in \mathcal{N}_i, i \in \mathcal{N}, \quad (1b)$$

where each agent  $i$  has private/local decisions denoted by  $\mathbf{u}_i \in \mathbb{R}^{n_i^p}$  and a shared decision denoted by  $\mathbf{v}_i^j \in \mathbb{R}^{n_s}$ , for every neighbor  $j \in \mathcal{N}_i$ . The vector  $\mathbf{v}_i$  collects all the shared decisions of agent  $i$ , i.e.,  $\mathbf{v}_i = [\mathbf{v}_i^j]_{j \in \mathcal{N}_i}$ . For each agent  $i$ , the cost function in (1a) is divided into two parts:  $f_i^p$  and  $f_i^s$ , which depend on  $\mathbf{u}_i$  and  $\mathbf{v}_i$ , respectively. Furthermore, the decisions of agent  $i$  ( $\mathbf{u}_i, \mathbf{v}_i$ ) are constrained by the local set  $\mathcal{C}_i$ . Moreover, the shared decisions of agent  $i$  are also coupled with the shared decisions of its neighbors through the equality constraints (1b). Additionally, we suppose that the following assumptions hold.

*Assumption 1:* The functions  $f_i^p : \mathbb{R}^{n_i^p} \rightarrow \mathbb{R}$  and  $f_i^s : \mathbb{R}^{n_s|\mathcal{N}_i|} \rightarrow \mathbb{R}$ , for each  $i \in \mathcal{N}$ , are differentiable and convex. Moreover,  $f_i^p(\mathbf{u}_i)$ , for each  $i \in \mathcal{N}$ , is strongly convex with strong convexity constant, denoted by  $m_i$ .  $\square$

*Assumption 2:* The set  $\mathcal{C}_i$ , for each  $i \in \mathcal{N}$ , is polyhedral and compact.  $\square$

*Assumption 3:* The feasible set of Problem (1) is non-empty.  $\square$

*Remark 1:* By Assumption 1, the cost function is continuous. Based on the Weierstrass theorem, since the problem is feasible (Assumption 3) and  $\mathcal{C}_i$ , for each  $i \in \mathcal{N}$ , is compact, the optimal value is finite and the problem has a solution.  $\square$

*Remark 2:* As practical examples, we refer to [21]–[26] for energy management problems that consider the same problem structure, i.e., polyhedral and compact local set constraints and edge-based coupling constraints.  $\square$

### C. Time-varying communication and asynchronicity

We model the communication network as a random graph [17]. To that end, let the communication network be described as an undirected graph  $\mathcal{G}^c(k) = (\mathcal{N}, \mathcal{E}^c(k))$ , where  $\mathcal{E}^c(k) \subseteq \mathcal{E}$  denotes the set of communication links that are active at iteration  $k-1$ , i.e.,  $\{i, j\} \in \mathcal{E}^c(k)$  means that agents  $i$  and  $j$  can exchange information between each other. Thus, the random model of the communication network is defined in Assumption 4.

*Assumption 4:* The set  $\mathcal{E}^c(k)$  is an independent and identically distributed random variable. Furthermore, any communication link between two coupled agents  $i$  and  $j$ , where  $\{i, j\} \in \mathcal{E}$ , is active with a positive probability denoted by  $\beta_{ij}$ , i.e.,  $\mathbb{P}(\{i, j\} \in \mathcal{E}^c(k)) = \beta_{ij} > 0$ .  $\square$

Moreover, we also allow asynchronous updates, i.e., not all agents might update their decisions at each iteration. The asynchronous updates are also modeled as a random process, as follows. Denote the set of agents that are active and update

their primal and dual variables at iteration  $k-1$  by  $\mathcal{A}(k)$ . Then, we consider the following assumption.

**Assumption 5:** The set  $\mathcal{A}(k) \subseteq \mathcal{N}$  is an independent and identically distributed random variable. Moreover, an agent  $i \in \mathcal{N}$  is active and updates its primal and dual variables at iteration  $k$  with a positive probability denoted by  $\gamma_i$ , i.e.,  $\mathbb{P}(i \in \mathcal{A}(k)) = \gamma_i > 0$ .  $\square$

### III. PROPOSED METHOD

#### A. Algorithm design

We consider the augmented problem of (1) in the following form:

$$\begin{aligned} & \underset{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}}{\text{minimize}} \quad \sum_{i=1}^n \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) + \sum_{j \in \mathcal{N}_i} \|\mathbf{v}_i^j + \mathbf{v}_j^i\|_2^2 \right) \\ & \text{subject to} \quad \mathbf{v}_i^j + \mathbf{v}_j^i = 0, \quad \forall j \in \mathcal{N}_i, i \in \mathcal{N}. \end{aligned} \quad (2)$$

We use the dual approach to decompose Problem (2). To this end, we denote the decisions of all agents by  $\mathbf{u} = [\mathbf{u}_i]_{i \in \mathcal{N}}$  and  $\mathbf{v} = [\mathbf{v}_i]_{i \in \mathcal{N}}$  and we introduce the Lagrangian of the augmented problem (2), denoted by  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$ , as follows:

$$\begin{aligned} L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}) = & \sum_{i \in \mathcal{N}} \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) + \right. \\ & \left. \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^j, \mathbf{v}_i^j + \mathbf{v}_j^i \rangle + \|\mathbf{v}_i^j + \mathbf{v}_j^i\|_2^2 \right) \right), \end{aligned} \quad (3)$$

where the coupled constraints in (1b) are relaxed and  $\boldsymbol{\lambda}_i^j \in \mathbb{R}^{n_s}$ , for all  $j \in \mathcal{N}_i$ , are the Lagrange multipliers associated to them. Note that, for convenience, the Lagrange multipliers are compactly written as  $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_i]_{i \in \mathcal{N}}$ , where  $\boldsymbol{\lambda}_i = [\boldsymbol{\lambda}_i^j]_{j \in \mathcal{N}_i}$ . Now, we introduce the dual function, denoted by  $q(\boldsymbol{\lambda})$ , as follows:

$$q(\boldsymbol{\lambda}) = \underset{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}}{\text{minimize}} \quad L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}). \quad (4)$$

Since each  $\mathcal{C}_i$  is assumed to be compact (Assumption 2) and the Lagrangian function is continuous (Assumption 1), by the Weierstrass theorem it follows that a minimizer in (4) exists and the value  $q(\boldsymbol{\lambda})$  is finite for every  $\boldsymbol{\lambda}$ . Hence, the domain of  $q(\boldsymbol{\lambda})$  is the entire space of  $\boldsymbol{\lambda}$ . We also know from the duality theory that  $q(\boldsymbol{\lambda})$  is concave and continuous.

The dual problem associated with (2) is stated as follows:

$$\underset{\boldsymbol{\lambda}}{\text{maximize}} \quad q(\boldsymbol{\lambda}). \quad (5)$$

Note that the dual optimal value is finite. Furthermore, the strong duality holds and the set of dual optimal points is non-empty since, in the primal problem (1), the cost function is convex and the constraints are linear [28, Proposition 5.2.1]. In other word, there exists a saddle point of the Lagrangian function  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$ , i.e., a point  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*) \in \prod_{i \in \mathcal{N}} \mathcal{C}_i \times \mathbb{R}^{\sum_{i \in \mathcal{N}} n_s |\mathcal{N}_i|}$  such that, for any  $(\mathbf{u}, \mathbf{v}) \in \prod_{i \in \mathcal{N}} \mathcal{C}_i$  and  $\boldsymbol{\lambda} \in \mathbb{R}^{\sum_{i \in \mathcal{N}} n_s |\mathcal{N}_i|}$ , it holds that

$$L(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}) \leq L(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}^*). \quad (6)$$

The dual function  $q(\boldsymbol{\lambda})$  has separable constraints and all the terms in the Lagrangian function are also separable, except for the quadratic term  $\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \|\mathbf{v}_i^j + \mathbf{v}_j^i\|_2^2$ . In

---

#### Algorithm 1: Distributed augmented Lagrangian

---

**Initialization:** For each agent  $i \in \mathcal{N}$ ,  $\mathbf{v}_i(0) = \mathbf{v}_{i0} \in \mathbb{R}^{|\mathcal{N}_i| n_s}$  and  $\boldsymbol{\lambda}_i(0) = \boldsymbol{\lambda}_{i0} \in \mathbb{R}^{|\mathcal{N}_i| n_s}$ .

**Iteration:** For each agent  $i \in \mathcal{N}$ ,

- 1) Update  $\mathbf{u}_i(k+1)$  and  $\hat{\mathbf{v}}_i(k)$  according to

$$\begin{aligned} & (\mathbf{u}_i(k+1), \hat{\mathbf{v}}_i(k)) \\ & = \arg \min_{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i} f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) + \\ & \quad \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^j(k) + \boldsymbol{\lambda}_j^i(k), \mathbf{v}_i^j \rangle + \|\mathbf{v}_i^j + \mathbf{v}_j^i(k)\|_2^2 \right). \end{aligned} \quad (8)$$

- 2) Update  $\mathbf{v}_i^j(k+1)$ , for all  $j \in \mathcal{N}_i$ , as follows:

$$\mathbf{v}_i^j(k+1) = \eta_i^j \hat{\mathbf{v}}_i^j(k) + (1 - \eta_i^j) \mathbf{v}_i^j(k). \quad (9)$$

- 3) Send  $\mathbf{v}_i^j(k+1)$  to and receive  $\mathbf{v}_j^i(k+1)$  from neighbors  $j \in \mathcal{N}_i$ .

- 4) Update the dual variables  $\boldsymbol{\lambda}_i^j(k+1)$ , for all  $j \in \mathcal{N}_i$ , according to

$$\boldsymbol{\lambda}_i^j(k+1) = \boldsymbol{\lambda}_i^j(k) + \eta_i^j \left( \mathbf{v}_i^j(k+1) + \mathbf{v}_j^i(k+1) \right). \quad (10)$$

- 5) Send  $\boldsymbol{\lambda}_i^j(k+1)$  to and receive  $\boldsymbol{\lambda}_j^i(k+1)$  from neighbors  $j \in \mathcal{N}_i$ .
- 

this regard, each agent will use the information from its neighbors as a way to approximate that quadratic term and decompose  $q(\boldsymbol{\lambda})$ . For each agent  $i \in \mathcal{N}$ , denote by  $\tilde{\mathbf{v}}_i^j$  the information associated to  $\mathbf{v}_j^i$  from neighbor  $j \in \mathcal{N}_i$ . Thus, the minimization on the right hand side of (4) is approximated by

$$\begin{aligned} & \underset{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}}{\text{minimize}} \quad \sum_{i \in \mathcal{N}} \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) + \right. \\ & \quad \left. \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^j + \boldsymbol{\lambda}_j^i, \mathbf{v}_i^j \rangle + \|\mathbf{v}_i^j + \tilde{\mathbf{v}}_i^j\|_2^2 \right) \right). \end{aligned} \quad (7)$$

We are in the position to state the proposed distributed approach, which is shown in Algorithm 1. In step 1 of Algorithm 1, each agent updates the local decisions  $\mathbf{u}_i(k+1)$  and an auxiliary variable, which is denoted by  $\hat{\mathbf{v}}_i(k)$  and used to update the shared decisions, by solving the decomposed problem (7), where  $\tilde{\mathbf{v}}_i^j = \mathbf{v}_j^i(k)$ . Then, the update of  $\mathbf{v}_i(k+1)$  by (9), where  $\eta_i^j \in (0, 1)$ , uses a convex combination of  $\hat{\mathbf{v}}_i(k)$  and the value at the previous iteration  $\mathbf{v}_i(k)$ . Meanwhile, the dual variables are updated by (10), using the step size  $\eta_i^j$ , for all  $j \in \mathcal{N}_i$ . We will discuss the choice of  $\eta_i^j$  later in the convergence analysis.

Now, we consider the time-varying nature of the communication network. Based on Assumptions 4 and 5, agent  $i$  can only exchange information to its neighbor  $j \in \mathcal{N}_i$  if both agents are active and the link  $\{i, j\}$  is also active. In this regard, for each agent  $i$ , we denote the set of coupled neighbors with which agent  $i$  can exchange information by  $\mathcal{A}_i(k) = \{j \in \mathcal{N}_i \cap \mathcal{A}(k) : \{i, j\} \in \mathcal{E}^c(k)\}$ . In this situation, an active agent  $i \in \mathcal{A}(k+1)$  might not have  $\mathbf{v}_j^i(k)$  and  $\boldsymbol{\lambda}_j^i(k)$

**Algorithm 2:** Distributed augmented Lagrangian with imperfect communication

**Initialization:** For each agent  $i \in \mathcal{N}$ ,  $\mathbf{v}_i(0) = \mathbf{v}_{i0} \in \mathbb{R}^{|\mathcal{N}_i|n_s}$  and  $\boldsymbol{\lambda}_i(0) = \boldsymbol{\lambda}_{i0} \in \mathbb{R}^{|\mathcal{N}_i|n_s}$ . Moreover,  $\mathbf{z}_i^j(0) = \mathbf{v}_j^i(0)$  and  $\boldsymbol{\xi}_i^j(0) = \boldsymbol{\lambda}_j^i(0)$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ .

**Iteration:** For each agent  $i \in \mathcal{A}(k+1)$ ,

- 1) Update  $(\mathbf{u}_i(k+1), \hat{\mathbf{v}}_i(k))$  according to

$$\begin{aligned} & (\mathbf{u}_i(k+1), \hat{\mathbf{v}}_i(k)) \\ &= \arg \min_{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i} f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) + \\ & \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^j(k) + \boldsymbol{\xi}_i^j(k), \mathbf{v}_i^j(k) \rangle + \|\mathbf{v}_i^j(k) + \mathbf{z}_i^j(k)\|_2^2 \right). \end{aligned} \quad (11)$$

- 2) Update  $\mathbf{v}_i^j(k+1)$ , for all  $j \in \mathcal{N}_i$ , as follows:

$$\mathbf{v}_i^j(k+1) = \begin{cases} \eta_i^j \hat{\mathbf{v}}_i^j(k) + (1 - \eta_i^j) \mathbf{v}_i^j(k), \\ \mathbf{v}_i^j(k), & \text{otherwise.} \end{cases} \quad \forall j \in \mathcal{A}_i(k+1), \quad (12)$$

- 3) Send  $\mathbf{v}_i^j(k+1)$  to and receive  $\mathbf{v}_j^i(k+1)$  from neighbor  $j \in \mathcal{A}_i(k+1)$ .
- 4) Update the auxiliary and dual variables  $\mathbf{z}_i(k+1)$  and  $\boldsymbol{\lambda}_i(k+1)$  according to:

$$\begin{aligned} \mathbf{z}_i^j(k+1) &= \begin{cases} \mathbf{v}_j^i(k+1), & \forall j \in \mathcal{A}_i(k+1), \\ \mathbf{z}_i^j(k), & \text{otherwise,} \end{cases} \quad (13) \\ \boldsymbol{\lambda}_i^j(k+1) &= \begin{cases} \boldsymbol{\lambda}_i^j(k) + \eta_i^j (\mathbf{v}_i^j(k+1) + \mathbf{z}_i^j(k+1)), \\ \boldsymbol{\lambda}_i^j(k), & \text{otherwise.} \end{cases} \quad \forall j \in \mathcal{A}_i(k+1), \end{aligned} \quad (14)$$

- 5) Send  $\boldsymbol{\lambda}_i^j(k+1)$  to and receive  $\boldsymbol{\lambda}_j^i(k+1)$  from neighbors  $j \in \mathcal{A}_i(k+1)$ .
- 6) Update the auxiliary variable  $\boldsymbol{\xi}_i(k+1)$  according to

$$\boldsymbol{\xi}_i^j(k+1) = \begin{cases} \boldsymbol{\lambda}_j^i(k+1), & \forall j \in \mathcal{A}_i(k+1), \\ \boldsymbol{\xi}_i^j(k), & \text{otherwise.} \end{cases} \quad (15)$$

For agent  $i \notin \mathcal{A}(k+1)$ ,  $\mathbf{u}_i(k+1) = \mathbf{u}_i(k)$ ,  $\mathbf{v}_i(k+1) = \mathbf{v}_i(k)$ ,  $\mathbf{z}_i(k+1) = \mathbf{z}_i(k)$ ,  $\boldsymbol{\lambda}_i(k+1) = \boldsymbol{\lambda}_i(k)$ , and  $\boldsymbol{\xi}_i(k+1) = \boldsymbol{\xi}_i(k)$ .

at iteration  $k$ . Therefore, it needs to track  $\mathbf{v}_i^j(k)$  and  $\boldsymbol{\lambda}_i^j(k)$ . In this regard, this information is captured by the auxiliary variables  $\mathbf{z}_i^j(k)$  and  $\boldsymbol{\xi}_i^j(k)$ , for all  $j \in \mathcal{N}_i$ , respectively. The proposed distributed method follows Algorithm 2.

**Remark 3:** In order to initialize the auxiliary variables  $\mathbf{z}_i(0)$  and  $\boldsymbol{\xi}_i(0)$ , either agent  $i \in \mathcal{N}$  receives  $\mathbf{v}_j^i(0)$  and  $\boldsymbol{\lambda}_j^i(0)$  from all neighbors  $j \in \mathcal{N}_i$  or it is set such that, for each  $i \in \mathcal{N}$ ,  $\mathbf{v}_i(0) = \mathbf{z}_i(0) = v_0 \mathbf{1}_{|\mathcal{N}_i|n_s}$  and  $\boldsymbol{\lambda}_i(0) = \boldsymbol{\xi}_i(0) = \lambda_0 \mathbf{1}_{|\mathcal{N}_i|n_s}$ , for any  $v_0, \lambda_0 \in \mathbb{R}$ .  $\square$

**Remark 4:** The case where the algorithm is performed under perfect communication, as stated in Algorithm 1, can be considered as a special case of Algorithm 2 where  $\beta_{ij} = 1$ , for all  $\{i, j\} \in \mathcal{E}$  and  $\gamma_i = 1$ , for all  $i \in \mathcal{N}$ .  $\square$

## B. Convergence statement

The convergence of the sequence produced by Algorithm 2 is stated in Theorem 1, as follows.

**Theorem 1:** Let Assumptions 1-5 hold. Furthermore, let the sequence  $\{(\mathbf{u}(k), \mathbf{v}(k), \hat{\mathbf{v}}(k), \boldsymbol{\lambda}(k))\}$  be generated by Algorithm 2. If  $\eta_i^j = \eta_j^i = \eta_{ij} \in (0, \frac{1}{4})$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ , then, with probability 1,

- a. (Feasibility)  $\lim_{k \rightarrow \infty} \|\mathbf{v}_i^j(k) + \mathbf{v}_j^i(k)\|_2^2 = 0$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ ,
- b. (Primal and dual variable convergence) There exists a saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$  (see (3)), denoted by  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*)$ , such that  $\lim_{k \rightarrow \infty} \mathbf{u}(k) = \mathbf{u}^*$ ,  $\lim_{k \rightarrow \infty} \mathbf{v}(k) = \mathbf{v}^*$ , and  $\lim_{k \rightarrow \infty} \boldsymbol{\lambda}(k) = \boldsymbol{\lambda}^*$ .  $\square$

*Proof:* See Section IV-B.  $\blacksquare$

Notice that, if the dual variables  $\boldsymbol{\lambda}_i(0)$ , for all  $i \in \mathcal{N}$ , are initialized such that  $\boldsymbol{\lambda}_{i0} = \boldsymbol{\lambda}_{j0}^i$ , then, we have that  $\boldsymbol{\lambda}_i^j(k) = \boldsymbol{\lambda}_j^i(k)$ , for all  $k \geq 0$ , since  $\eta_i^j = \eta_j^i = \eta_{ij}$ . In this setup, the second round of communication (Step 5) in Algorithm 2 is not necessary and each agent  $i \in \mathcal{A}(k+1)$  can update  $\boldsymbol{\xi}_i^j(k+1) = \boldsymbol{\lambda}_j^i(k+1)$ , for all  $j \in \mathcal{A}_i(k+1)$ , and  $\boldsymbol{\xi}_i^j(k+1) = \boldsymbol{\xi}_i^j(k)$ , otherwise.

We also state the convergence rate of Algorithm 2 in terms of the ergodic average of the primal and auxiliary variables, which are defined, for all  $i \in \mathcal{N}$  and  $k \geq 1$ , as follows:

$$\bar{\mathbf{u}}_i(k) = \sum_{\ell=0}^{k-1} \frac{\mathbf{u}_i(\ell)}{k}, \quad \bar{\mathbf{v}}_i(k) = \sum_{\ell=0}^{k-1} \frac{\mathbf{v}_i(\ell)}{k}, \quad \bar{\hat{\mathbf{v}}}_i(k) = \sum_{\ell=0}^{k-1} \frac{\hat{\mathbf{v}}_i(\ell)}{k}, \quad (16)$$

**Theorem 2:** Let Assumptions 1-5 hold. Furthermore, let the sequence  $\{(\mathbf{u}(k), \mathbf{v}(k), \hat{\mathbf{v}}(k), \boldsymbol{\lambda}(k))\}$  be generated by Algorithm 2 with  $\eta_i^j = \eta_j^i = \eta_{ij} \in (0, \frac{1}{4})$ , for all  $j \in \mathcal{N}_i$ . Then, the ergodic average of the primal variables (16) converge to a solution to Problem (1) with the convergence rate  $\mathcal{O}(\frac{1}{k})$ .

*Proof:* See Section IV-C.  $\blacksquare$

## IV. CONVERGENCE AND RATE ANALYSIS

This section is devoted to proving Theorems 1 and 2. Prior to that, we establish some intermediate results that are useful for proving these theorems.

### A. Intermediate Results

**Lemma 1:** Let Assumptions 1-3 hold. Furthermore, let  $(\mathbf{u}_i(k+1), \hat{\mathbf{v}}_i(k))$  be the attainer of the local optimization in (8) and  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*)$  be a saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$  as defined in (3). Then, it holds that

$$\begin{aligned} 0 &\leq \sum_{i \in \mathcal{N}} \left( -m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2 \right. \\ &+ \sum_{j \in \mathcal{N}_i} \langle \boldsymbol{\lambda}_i^{j*} - \boldsymbol{\lambda}_i^j(k), \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k) \rangle \\ &- \sum_{j \in \mathcal{N}_i} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k)\|_2^2 \\ &\left. - 2 \sum_{j \in \mathcal{N}_i} \langle \mathbf{v}_j^i(k) - \hat{\mathbf{v}}_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle \right). \end{aligned} \quad (17)$$

*Proof:* Since  $(\mathbf{u}_i^*, \mathbf{v}_i^*) \in \mathcal{C}_i$ , the optimality condition [29, Theorem 20] of the local optimization in (8) yields the following relation:

$$\begin{aligned} 0 \leq & \langle \nabla f_i^P(\mathbf{u}_i(k+1)), \mathbf{u}_i^* - \mathbf{u}_i(k+1) \rangle \\ & + \langle \nabla f_i^S(\hat{\mathbf{v}}_i(k)), \mathbf{v}_i^* - \hat{\mathbf{v}}_i(k) \rangle \\ & + \sum_{j \in \mathcal{N}_i} \langle \lambda_j^j(k) + \lambda_j^i(k), \mathbf{v}_i^{j*} - \hat{\mathbf{v}}_i^j(k) \rangle \\ & + \sum_{j \in \mathcal{N}_i} 2 \langle \hat{\mathbf{v}}_i^j(k) + \mathbf{v}_j^i(k), \mathbf{v}_i^{j*} - \hat{\mathbf{v}}_i^j(k) \rangle. \end{aligned} \quad (18)$$

Now, we consider the second inequality in (6), which implies that  $(\mathbf{u}^*, \mathbf{v}^*) = \arg \min_{(\mathbf{u}, \mathbf{v}) \in \mathcal{C}_i, i \in \mathcal{N}} L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}^*)$ . Based on the optimality condition of this minimization and the fact that  $(\mathbf{u}_i(k+1), \hat{\mathbf{v}}_i(k)) \in \mathcal{C}_i$ , we obtain that

$$\begin{aligned} 0 \leq & \sum_{i \in \mathcal{N}} \left( \langle \nabla f_i^P(\mathbf{u}_i^*), \mathbf{u}_i(k+1) - \mathbf{u}_i^* \rangle \right. \\ & + \langle \nabla f_i^S(\mathbf{v}_i^*), \hat{\mathbf{v}}_i(k) - \mathbf{v}_i^* \rangle \\ & + \sum_{j \in \mathcal{N}_i} \langle \lambda_i^{j*} + \lambda_j^{i*}, \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle. \end{aligned} \quad (19)$$

By summing up (18) over all agents  $i \in \mathcal{N}$  and combining with (19), we obtain that

$$\begin{aligned} 0 \leq & \sum_{i \in \mathcal{N}} \left( \langle \nabla f_i^P(\mathbf{u}_i^*) - \nabla f_i^P(\mathbf{u}_i(k+1)), \mathbf{u}_i(k+1) - \mathbf{u}_i^* \rangle \right. \\ & + \langle \nabla f_i^S(\mathbf{v}_i^*) - \nabla f_i^S(\hat{\mathbf{v}}_i(k)), \hat{\mathbf{v}}_i(k) - \mathbf{v}_i^* \rangle \\ & + \sum_{j \in \mathcal{N}_i} \langle \lambda_i^{j*} + \lambda_j^{i*} - \lambda_i^j(k) - \lambda_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle \\ & \left. - 2 \sum_{j \in \mathcal{N}_i} \langle \hat{\mathbf{v}}_i^j(k) + \mathbf{v}_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle \right). \end{aligned} \quad (20)$$

Applying the convexity and strong convexity relations (cf. Definitions 1-2 for  $f_i^S(\mathbf{v}_i)$  and  $f_i^P(\mathbf{u}_i)$ , for all  $i \in \mathcal{N}$ , and adding the term  $\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} 2 \langle \hat{\mathbf{v}}_i^j(k) - \hat{\mathbf{v}}_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle = 0$  to (20), it follows that

$$\begin{aligned} 0 \leq & \sum_{i \in \mathcal{N}} \left( -m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2 \right. \\ & + \sum_{j \in \mathcal{N}_i} \langle \lambda_i^{j*} + \lambda_j^{i*} - \lambda_i^j(k) - \lambda_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle \\ & - 2 \sum_{j \in \mathcal{N}_i} \langle \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle \\ & \left. - 2 \sum_{j \in \mathcal{N}_i} \langle \mathbf{v}_j^i(k) - \hat{\mathbf{v}}_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle \right). \end{aligned} \quad (21)$$

Now, we consider the second term on the right-hand side of the inequality, i.e.,  $\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \langle \lambda_i^{j*} + \lambda_j^{i*} - \lambda_i^j(k) - \lambda_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle$ . By considering the summation over all links and since at each link there exist two inner products associated to both agents coupled by that link, that term is equivalent to  $\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \langle \lambda_i^{j*} - \lambda_i^j(k), \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k) \rangle$ . Similarly, the third term on the right-hand side of the inequality (21), i.e.,  $2 \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \langle \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k), \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^{j*} \rangle$  is equivalent to  $\sum_{\{i,j\} \in \mathcal{E}} 2 \langle \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k), \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k) \rangle = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k)\|_2^2$ . Thus, we obtain the desired inequality (17). ■

Now, we define the auxiliary variables,  $\tilde{\lambda}_i^j(k)$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ , as follows:

$$\tilde{\lambda}_i^j(k) = \lambda_i^j(k) + (1 - \eta_i^j)(\mathbf{v}_i^j(k) + \mathbf{v}_j^i(k)), \quad (22)$$

and obtain a useful estimate in Lemma 2.

*Lemma 2:* Let Assumptions 1-3 hold. Furthermore, let  $(\mathbf{u}_i(k+1), \hat{\mathbf{v}}_i(k))$  be the attainer of the local optimization in (8),  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*)$  be a saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$  as defined in (3), and  $\tilde{\lambda}_i^j(k)$  be defined as in (22). Then, it holds that

$$\begin{aligned} & \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \langle \tilde{\lambda}_i^j(k) - \lambda_i^{j*}, \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k) \rangle \\ & + 2 \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \langle \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^j(k), \mathbf{v}_i^j(k) - \mathbf{v}_i^{j*} \rangle \\ & \leq \sum_{i \in \mathcal{N}} \left( -m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2 \right. \\ & \quad - \sum_{j \in \mathcal{N}_i} \frac{3}{2} \|\hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^j(k)\|_2^2 \\ & \quad \left. - \sum_{j \in \mathcal{N}_i} \frac{\eta_i^j + \eta_j^i - (\eta_i^j + \eta_j^i)^2}{2} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k)\|_2^2 \right). \end{aligned} \quad (23)$$

*Proof:* We omit the proof due to the space constraint. It follows a similar line of analysis as that of the proof of Lemma 3 in [14]. ■

As the next building block to show the convergence result, we define a Lyapunov function, denoted by  $V(k)$ . For any given saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$  (see (3)), denoted by  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*)$ , we can construct  $V(k)$  as follows:

$$V(k) = \|\mathbf{v}(k) - \mathbf{v}^*\|_H^2 + \frac{1}{2} \|\tilde{\boldsymbol{\lambda}}(k) - \boldsymbol{\lambda}^*\|_H^2, \quad (24)$$

where  $\tilde{\boldsymbol{\lambda}} = [\tilde{\lambda}_i(k)]_{i \in \mathcal{N}}$ ,  $\tilde{\lambda}_i(k) = [\tilde{\lambda}_i^j(k)]_{j \in \mathcal{N}_i}$ ,  $\tilde{\lambda}_i^j(k)$  is defined in (22),  $H = \text{blkdiag}(\{H_i\}_{i \in \mathcal{N}})$  and  $H_i = \text{blkdiag}(\{(\eta_i^j)^{-1} I_{n_s}\}_{j \in \mathcal{N}_i})$ , for all  $i \in \mathcal{N}$ . Now, we show that  $\{V(k)\}$  is non-increasing under Algorithm 1 and obtain an estimate that will be used in the main theorems.

*Lemma 3:* Let Assumptions 1-3 hold. Furthermore, let the sequence  $\{\mathbf{u}(k), \mathbf{v}(k), \hat{\mathbf{v}}(k), \boldsymbol{\lambda}(k)\}$  be generated by Algorithm 1,  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*)$  be a saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$  as defined in (3), and  $V(k)$  be defined in (24). If  $\eta_i^j = \eta_{ij} \in (0, \frac{1}{4})$ , then  $\{V(k)\}$  is a monotonically non-increasing sequence and the following inequality holds:

$$\begin{aligned} & V(k+1) - V(k) \\ & \leq - \sum_{i \in \mathcal{N}} m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2 \\ & \quad - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left( \frac{3}{2} - \eta_{ij} \right) \|\hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^j(k)\|_2^2 \\ & \quad - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{\eta_{ij} - (2\eta_{ij})^2}{2} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k)\|_2^2. \end{aligned} \quad (25)$$

*Proof:* Firstly, notice that  $\tilde{\lambda}_i^j(k+1)$  can be expressed as  $\tilde{\lambda}_i^j(k+1) = \tilde{\lambda}_i^j(k) + \eta_i^j(\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k))$ . Thus, we have that

$$\begin{aligned} & \|\tilde{\lambda}_i^j(k+1) - \lambda_i^{j*}\|_2^2 \\ & = \|\tilde{\lambda}_i^j(k) - \lambda_i^{j*}\|_2^2 + 2\eta_i^j \langle \tilde{\lambda}_i^j(k) - \lambda_i^{j*}, \hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k) \rangle \\ & \quad + \|\eta_i^j(\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k))\|_2^2. \end{aligned} \quad (26)$$

Moreover, we also have that

$$\begin{aligned} \|\mathbf{v}_i^j(k+1) - \mathbf{v}_i^{j*}\|_2^2 &= \|\mathbf{v}_i^j(k) - \mathbf{v}_i^{j*}\|_2^2 + \|\eta_i^j(\hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^j(k))\|_2^2 \\ &\quad + 2\eta_i^j \langle \hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^j(k), \mathbf{v}_i^j(k) - \mathbf{v}_i^{j*} \rangle. \end{aligned} \quad (27)$$

By using the expression of  $V(k+1)$  from (26)-(27) and using the inequality in (23), we obtain that:

$$\begin{aligned} V(k+1) - V(k) &\leq \|\hat{\mathbf{v}}(k) - \mathbf{v}(k)\|_{H^{-1}}^2 + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{\eta_i^j}{2} \|(\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k))\|_2^2 \\ &\quad - \frac{3}{2} \|\hat{\mathbf{v}}(k) - \mathbf{v}(k)\|_2^2 - \sum_{i \in \mathcal{N}} m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2 \\ &\quad - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{2\eta_i^j - (2\eta_{ij}^j)^2}{2} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k)\|_2^2. \end{aligned}$$

Thus, the inequality (25) follows and  $V(k)$  is monotonically non-increasing if  $\eta_i^j = \eta_j^i = \eta_{ij} \in (0, \frac{1}{4})$ . ■

The function  $V(k)$  is used to construct a Lyapunov function for Algorithm 2. Moreover, the estimate obtained in Lemma 3 will also be used to obtain the result in Lemma 4. Therefore, now consider the function  $\tilde{V}(k)$ , defined as follows. For any saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$  (see (3)), denoted by  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*)$ , we have that

$$\tilde{V}(k) = \|\mathbf{v}(k) - \mathbf{v}^*\|_H^2 + \frac{1}{2} \|\boldsymbol{\nu}(k) - \boldsymbol{\lambda}^*\|_H^2, \quad (28)$$

where  $\boldsymbol{\nu}(k) = [\boldsymbol{\nu}_i(k)]_{i \in \mathcal{N}}$ ,  $\boldsymbol{\nu}_i(k) = [\boldsymbol{\nu}_i^j(k)]_{j \in \mathcal{N}_i}$ ,

$$\boldsymbol{\nu}_i^j(k) = \boldsymbol{\lambda}_i^j(k) + (1 - \eta_i^j)(\mathbf{v}_i^j(k) + \mathbf{z}_i^j(k)),$$

for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ ,  $\tilde{H} = \text{blkdiag}(\{\tilde{H}_i\}_{i \in \mathcal{N}})$ , and  $\tilde{H}_i = \text{blkdiag}(\{(\alpha_{ij}\eta_i^j)^{-1}I_{n_s}\}_{j \in \mathcal{N}_i})$ , for all  $i \in \mathcal{N}$ , where  $\alpha_{ij} = \beta_{ij}\gamma_i\gamma_j \in (0, 1]$ . Recall that  $\beta_{ij}$  and  $\gamma_i$  are the probability of link  $\{i, j\}$  being active and agent  $i$  being active, respectively.

**Lemma 4:** Let Assumptions 1-5 hold. Furthermore, let the sequence  $\{\mathbf{u}(k), \mathbf{v}(k), \hat{\mathbf{v}}(k), \boldsymbol{\lambda}(k)\}$  be generated by Algorithm 2,  $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*)$  be a saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$  as defined in (3), and  $\tilde{V}(k)$  be defined as in (28). If  $\eta_i^j = \eta_j^i = \eta_{ij} \in (0, \frac{1}{4})$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ , then the sequence  $\{\tilde{V}(k)\}$  is a non-negative supermartingale and it holds with probability 1 that

$$\begin{aligned} \mathbb{E}(\tilde{V}(k+1)|\mathcal{F}(k)) - \tilde{V}(k) &\leq - \sum_{i \in \mathcal{N}} m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2 \\ &\quad - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left(\frac{3}{2} - \eta_{ij}\right) \|\hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^j(k)\|_2^2 \\ &\quad - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{\eta_{ij} - (2\eta_{ij})^2}{2} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_j^i(k)\|_2^2 \leq 0. \end{aligned} \quad (29)$$

*Proof:* Since  $\tilde{V}(k)$  is a sum of norms and  $\eta_i^j$  and  $\alpha_{ij}$  are positive, the sequence  $\{\tilde{V}(k)\}$  is clearly non-negative. Denote by  $\mathcal{F}(k)$  the filtration up to and including the iteration  $k$ , i.e.,  $\mathcal{F}(k) = \{\mathcal{A}(\ell), \mathcal{E}^c(\ell), \mathbf{u}(\ell), \mathbf{v}(\ell), \boldsymbol{\lambda}(\ell), \mathbf{z}(\ell), \boldsymbol{\xi}(\ell), \ell =$

$0, 1, \dots, k\}$ . Now, we show that the conditional expectation of the sequence with respect to  $\mathcal{F}(k)$  is always non-increasing. Based on Assumptions 4 and 5, a proper initialization in Algorithm 2, and the update rules (12), (13), and (14), the variables  $\mathbf{v}_i^j(k+1)$ ,  $\mathbf{z}_i^j(k+1) = \mathbf{v}_j^i(k+1)$ , and  $\boldsymbol{\lambda}_i^j(k+1)$ , for each  $j \in \mathcal{N}_i$ , are only updated when agents  $i$  and  $j$  are active and link  $\{i, j\}$  is active. Therefore, we can denote the probability of  $\mathbf{v}_i^j(k+1)$ ,  $\mathbf{z}_i^j(k+1) = \mathbf{v}_j^i(k+1)$ , and  $\boldsymbol{\lambda}_i^j(k+1)$  being updated by  $\alpha_{ij} = \beta_{ij}\gamma_i\gamma_j \in (0, 1]$ , whereas, with probability  $1 - \alpha_{ij}$ , they are not updated and the values remain the same as  $\mathbf{v}_i^j(k)$ ,  $\mathbf{z}_i^j(k) = \mathbf{v}_j^i(k)$ , and  $\boldsymbol{\lambda}_i^j(k)$ . Thus, we also observe that  $\boldsymbol{\nu}_i^j(k+1) = \tilde{\boldsymbol{\lambda}}_i^j(k+1) = \boldsymbol{\lambda}_i^j(k+1) + (1 - \eta_i^j)(\mathbf{v}_i^j(k+1) + \mathbf{v}_j^i(k+1))$  with probability  $\alpha_{ij}$  or the value  $\boldsymbol{\nu}_i^j(k) = \tilde{\boldsymbol{\lambda}}_i^j(k)$  is kept with probability  $1 - \alpha_{ij}$ . Hence, we obtain, with probability 1, that

$$\begin{aligned} \mathbb{E}(\tilde{V}(k+1)|\mathcal{F}(k)) - \tilde{V}(k) &= -\tilde{V}(k) \\ &\quad + \mathbb{E}\left(\|\mathbf{v}(k+1) - \mathbf{v}^*\|_H^2 + \frac{1}{2}\|\boldsymbol{\nu}(k+1) - \boldsymbol{\lambda}^*\|_H^2 \middle| \mathcal{F}(k)\right) \\ &= \|\mathbf{v}(k+1) - \mathbf{v}^*\|_H^2 - \|\mathbf{v}(k) - \mathbf{v}^*\|_H^2 \\ &\quad + \frac{1}{2}\|\tilde{\boldsymbol{\lambda}}(k+1) - \boldsymbol{\lambda}^*\|_H^2 - \frac{1}{2}\|\tilde{\boldsymbol{\lambda}}(k) - \boldsymbol{\lambda}^*\|_H^2. \end{aligned}$$

Notice that since the scalings of the remaining quadratic terms do not involve  $\alpha_{ij}$ , we can use the weighted vector norm induced by  $H$ . Based on the definition of  $V(k)$  given in (24), we obtain with probability 1 that

$$\mathbb{E}(\tilde{V}(k+1)|\mathcal{F}(k)) - \tilde{V}(k) = V(k+1) - V(k).$$

Therefore, by applying (25) to this relation, the desired relations in (29) follow, with probability 1, when  $\eta_i^j = \eta_j^i = \eta_{ij} \in (0, \frac{1}{4})$ . Thus, (29) also shows that the sequence  $\{\tilde{V}(k)\}$  is non-negative supermartingale. ■

## B. Proof of Theorem 1

Now, we are ready to prove Theorem 1. Recall the function  $\tilde{V}(k)$  defined in (28) and the inequality (29) in Lemma 4. Rearranging and iterating (29), for  $\ell = 0, \dots, k$ , and taking the total expectation, we have that

$$\begin{aligned} &\sum_{\ell=0}^k \sum_{i \in \mathcal{N}} \mathbb{E}(m_i \|\mathbf{u}_i(\ell+1) - \mathbf{u}_i^*\|_2^2) \\ &\quad + \sum_{\ell=0}^k \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left(\frac{3}{2} - \eta_{ij}\right) \mathbb{E}(\|\hat{\mathbf{v}}_i^j(\ell) - \mathbf{v}_i^j(\ell)\|_2^2) \\ &\quad + \sum_{\ell=0}^k \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{\eta_{ij} - (2\eta_{ij})^2}{2} \mathbb{E}(\|\hat{\mathbf{v}}_i^j(\ell) + \hat{\mathbf{v}}_j^i(\ell)\|_2^2) \\ &\leq \sum_{\ell=0}^k \mathbb{E}(\tilde{V}(\ell) - \tilde{V}(\ell+1)) \\ &= \tilde{V}(0) - \mathbb{E}(\tilde{V}(k+1)) \leq \tilde{V}(0), \end{aligned}$$

where the last inequality is obtained by dropping the non-positive term  $-\mathbb{E}(\tilde{V}(k+1))$ . The above inequalities imply that  $\{\mathbb{E}(m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2)\}$ ,

for all  $i \in \mathcal{N}$ , is summable and converges to 0. Similarly,  $\{\mathbb{E}(\|\hat{\mathbf{v}}_i^j(k) - \mathbf{v}_i^j(k)\|_2^2)\}$ , and  $\{\mathbb{E}(\|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_i^j(k)\|_2^2)\}$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ , are also summable and converge to 0. Using the Markov inequality, for any  $\varepsilon > 0$ , we have that  $\limsup_{k \rightarrow \infty} \mathbb{P}(\Psi(\mathbf{u}, \mathbf{v}, \hat{\mathbf{v}}) \geq \varepsilon) \leq \limsup_{k \rightarrow \infty} \frac{1}{\varepsilon} \mathbb{E}(\Psi(\mathbf{u}, \mathbf{v}, \hat{\mathbf{v}})) = 0$ , where  $\Psi(\mathbf{u}, \mathbf{v}, \hat{\mathbf{v}}) = \sum_{i \in \mathcal{N}} m_i \|\mathbf{u}_i(k+1) - \mathbf{u}_i^*\|_2^2 + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left(\frac{3}{2} - \eta_{ij}\right) \|\mathbf{v}_i^j(k) - \hat{\mathbf{v}}_i^j(k)\|_2^2 + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{\eta_{ij} - (2\eta_{ij})^2}{2} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_i^j(k)\|_2^2$ . Thus, it holds with probability 1 that

$$\lim_{k \rightarrow \infty} \|\mathbf{u}_i(k) - \mathbf{u}_i^*\|_2^2 = 0, \quad \forall i \in \mathcal{N}, \quad (30)$$

$$\lim_{k \rightarrow \infty} \|\mathbf{v}_i^j(k) - \hat{\mathbf{v}}_i^j(k)\|_2^2 = 0, \quad \forall j \in \mathcal{N}_i, \quad \forall i \in \mathcal{N}, \quad (31)$$

$$\lim_{k \rightarrow \infty} \|\hat{\mathbf{v}}_i^j(k) + \hat{\mathbf{v}}_i^j(k)\|_2^2 = 0, \quad \forall j \in \mathcal{N}_i, \quad \forall i \in \mathcal{N}, \quad (32)$$

Moreover, based on (31) and (32), it follows with probability 1 that

$$\lim_{k \rightarrow \infty} \|\mathbf{v}_i^j(k) + \mathbf{v}_i^j(k)\|_2^2 = 0, \quad \forall j \in \mathcal{N}_i, \quad \forall i \in \mathcal{N}. \quad (33)$$

Based on (29) and the martingale convergence theorem, the sequences  $\{\|\mathbf{v}(k) - \mathbf{v}^*\|_H^2\}$  and  $\{\|\boldsymbol{\nu}(k) - \boldsymbol{\lambda}^*\|_H^2\}$  are bounded with probability 1, i.e., there exist accumulation points of the sequences  $\{\mathbf{v}(k)\}$  and  $\{\boldsymbol{\nu}(k)\}$ . Furthermore,  $\{\boldsymbol{\lambda}(k)\}$  is also bounded with probability 1 and has accumulation points due to the boundedness of  $\{\boldsymbol{\nu}(k)\}$ , the relation in (33), and the fact that  $\mathbf{z}_i^j(k) = \mathbf{v}_i^j(k)$ , for each  $k \in \mathbb{Z}_{\geq 0}$ , which follows from the initialization of  $\mathbf{z}_i^j(k)$  in Algorithm 2 and the update rule (13).

Let  $\{(\mathbf{v}(k_\ell), \boldsymbol{\lambda}(k_\ell))\}$  be a convergent subsequence and assume that  $(\mathbf{v}^a, \boldsymbol{\lambda}^a)$  is its limit point. Therefore, due to the initialization of the variables in Algorithm 2 and the update rules (13) and (15), it follows that  $\lim_{\ell \rightarrow \infty} \mathbf{z}_i^j(k_\ell) = \lim_{\ell \rightarrow \infty} \mathbf{v}_i^j(k_\ell) = \mathbf{v}_i^{ja}$  and  $\lim_{\ell \rightarrow \infty} \boldsymbol{\xi}_i^j(k_\ell) = \lim_{\ell \rightarrow \infty} \boldsymbol{\lambda}_i^j(k_\ell) = \boldsymbol{\lambda}_i^{ja}$  with probability 1, for each  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ .

Now, we need to show that  $(\mathbf{u}^*, \mathbf{v}^a, \boldsymbol{\lambda}^a)$  is a saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$ , i.e.,  $(\mathbf{u}^*, \mathbf{v}^a, \boldsymbol{\lambda}^a)$  satisfies the inequalities in (6). Based on (33),  $\mathbf{v}_i^{ja} + \mathbf{v}_i^{ja} = \lim_{\ell \rightarrow \infty} (\mathbf{v}_i^j(k_\ell) + \mathbf{v}_i^j(k_\ell)) = 0$ , with probability 1, for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ . Thus, we have that, for any  $\boldsymbol{\lambda} \in \mathbb{R}^{\sum_{i \in \mathcal{N}} n_i |\mathcal{N}_i|}$ ,  $L(\mathbf{u}^*, \mathbf{v}^a, \boldsymbol{\lambda}) = L(\mathbf{u}^*, \mathbf{v}^a, \boldsymbol{\lambda}^a)$ , satisfying the first inequality in (6). Now, we show the second inequality in (6). Consider the update step (11), for all  $i \in \mathcal{N}$ , i.e.,

$$\begin{aligned} (\mathbf{u}(k+1), \hat{\mathbf{v}}(k)) = & \arg \min_{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}} \sum_{i \in \mathcal{N}} \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) \right. \\ & \left. + \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^j(k) + \boldsymbol{\xi}_i^j(k), \mathbf{v}_i^j \rangle + \|\mathbf{v}_i^j + \mathbf{z}_i^j(k)\|_2^2 \right) \right). \end{aligned}$$

By substituting  $k$  with  $k_\ell$  and taking the limit as  $\ell$  goes to infinity on both sides of the equality, it holds with probability

1 that

$$\begin{aligned} (\mathbf{u}^*, \mathbf{v}^a) = & \lim_{\ell \rightarrow \infty} \arg \min_{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}} \sum_{i \in \mathcal{N}} \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) \right. \\ & \left. + \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^j(k_\ell) + \boldsymbol{\xi}_i^j(k_\ell), \mathbf{v}_i^j \rangle + \|\mathbf{v}_i^j + \mathbf{z}_i^j(k_\ell)\|_2^2 \right) \right) \\ = & \arg \min_{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}} \sum_{i \in \mathcal{N}} \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) \right. \\ & \left. + \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^{ja} + \boldsymbol{\lambda}_j^{ia}, \mathbf{v}_i^j \rangle + \|\mathbf{v}_i^j + \mathbf{v}_j^{ia}\|_2^2 \right) \right) \\ = & \arg \min_{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}} \sum_{i \in \mathcal{N}} \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) \right. \\ & \left. + \sum_{j \in \mathcal{N}_i} \langle \boldsymbol{\lambda}_i^{ja}, \mathbf{v}_i^j + \mathbf{v}_j^{ia} \rangle \right). \quad (34) \end{aligned}$$

The left-hand side of the first equality is obtained by using  $\lim_{\ell \rightarrow \infty} (\mathbf{u}(k_\ell + 1), \hat{\mathbf{v}}(k_\ell)) = (\mathbf{u}^*, \mathbf{v}^a)$ , with probability 1, due to (30) and (31), which implies that  $\lim_{\ell \rightarrow \infty} \hat{\mathbf{v}}(k_\ell) = \mathbf{v}^a$ , with probability 1. The second equality is obtained since  $\lim_{\ell \rightarrow \infty} \mathbf{z}_i^j(k_\ell) = \mathbf{v}_i^{ja}$  and  $\lim_{\ell \rightarrow \infty} \boldsymbol{\xi}_i^j(k_\ell) = \boldsymbol{\lambda}_i^{ja}$ , with probability 1, for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ . Then, the last equality holds since the term  $\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \|\mathbf{v}_i^j + \mathbf{v}_j^{ia}\|_2^2$  is zero at  $(\mathbf{u}^*, \mathbf{v}^a)$  due to the fact that  $\mathbf{v}_i^{ja} + \mathbf{v}_j^{ia} = 0$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ . Additionally,  $\mathbf{v}^a$  is also an attainer of  $\min_{\mathbf{v}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \|\mathbf{v}_i^j + \mathbf{v}_j^{ia}\|_2^2$  since  $\mathbf{v}_i^{ja} + \mathbf{v}_j^{ia} = 0$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ . Therefore, the pair  $(\mathbf{u}^*, \mathbf{v}^a)$  also minimizes  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}^a)$ , i.e.,  $(\mathbf{u}^*, \mathbf{v}^a) \in \arg \min_{(\mathbf{u}_i, \mathbf{v}_i) \in \mathcal{C}_i, i \in \mathcal{N}} \sum_{i \in \mathcal{N}} \left( f_i^p(\mathbf{u}_i) + f_i^s(\mathbf{v}_i) + \sum_{j \in \mathcal{N}_i} \left( \langle \boldsymbol{\lambda}_i^{ja}, \mathbf{v}_i^j + \mathbf{v}_j^{ia} \rangle + \|\mathbf{v}_i^j + \mathbf{v}_j^{ia}\|_2^2 \right) \right)$ , where the cost function in the minimization is obtained by adding the quadratic term  $\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \|\mathbf{v}_i^j + \mathbf{v}_j^{ia}\|_2^2$  to the cost function on the right-hand side of the last equality in (34). Hence, the preceding relation implies the second inequality in (6). Thus,  $(\mathbf{u}^*, \mathbf{v}^a, \boldsymbol{\lambda}^a)$  is a saddle point of  $L(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$ . Finally, we can set  $\mathbf{v}^* = \mathbf{v}^a$  and  $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^a$  in  $\tilde{V}(k)$  (see (28)). Since the subsequence of  $\tilde{V}(k_\ell)$  converges to 0 with probability 1 and  $\tilde{V}(k)$  is non-negative supermartingale, the entire sequence  $\{(\mathbf{v}(k), \boldsymbol{\lambda}(k))\}$  converges to  $(\mathbf{v}^a, \boldsymbol{\lambda}^a)$  with probability 1.

### C. Proof of Theorem 2

By rearranging the summation of (29) over  $\ell = 0, \dots, k-1$  and taking the total expectation, we have that

$$\begin{aligned} & \sum_{\ell=0}^{k-1} \sum_{i \in \mathcal{N}} \mathbb{E} (m_i \|\mathbf{u}_i(\ell+1) - \mathbf{u}_i^*\|_2^2) \\ & + \sum_{\ell=0}^{k-1} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left( \frac{3}{2} - \eta_{ij} \right) \mathbb{E} \left( \|\hat{\mathbf{v}}_i^j(\ell) - \mathbf{v}_i^j(\ell)\|_2^2 \right) \\ & + \sum_{\ell=0}^{k-1} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{\eta_{ij} - (2\eta_{ij})^2}{2} \mathbb{E} \left( \|\hat{\mathbf{v}}_i^j(\ell) + \hat{\mathbf{v}}_i^j(\ell)\|_2^2 \right) \quad (35) \\ & \leq \sum_{\ell=0}^{k-1} \mathbb{E} \left( \tilde{V}(\ell) - \tilde{V}(\ell+1) \right) = \tilde{V}(0) - \mathbb{E} \left( \tilde{V}(k) \right) \\ & \leq \frac{1}{\alpha} \|\mathbf{v}(0) - \mathbf{v}^*\|_H^2 + \frac{1}{2\alpha} \|\tilde{\boldsymbol{\lambda}}(0) - \boldsymbol{\lambda}^*\|_H^2, \end{aligned}$$

where the last inequality is obtained by dropping the non-positive term  $-\mathbb{E}(\tilde{V}(k))$  and by defining  $\underline{\alpha} = \min_{\{i,j\} \in \mathcal{E}} \gamma_i \gamma_j \beta_{ij}$ . Furthermore, due to the convexity of the squared of the Euclidean norm, it follows that, for  $k \geq 1$ ,  $k\mathbb{E}(\|\bar{\mathbf{u}}_i(k) - \mathbf{u}_i^*\|_2^2) \leq \sum_{\ell=0}^{k-1} \mathbb{E}(\|\mathbf{u}_i(\ell+1) - \mathbf{u}_i^*\|_2^2)$ ,  $k\mathbb{E}(\|\hat{\mathbf{v}}_i^j(k-1) - \bar{\mathbf{v}}_i^j(k-1)\|_2^2) \leq \sum_{\ell=0}^{k-1} \mathbb{E}(\|\hat{\mathbf{v}}_i^j(\ell) - \mathbf{v}_i^j(\ell)\|_2^2)$ ,  $k\mathbb{E}(\|\hat{\mathbf{v}}_i^j(k-1) + \bar{\mathbf{v}}_i^j(k-1)\|_2^2) \leq \sum_{\ell=0}^{k-1} \mathbb{E}(\|\hat{\mathbf{v}}_i^j(\ell) + \bar{\mathbf{v}}_i^j(\ell)\|_2^2)$ . By applying the above relations to (35) and using the fact that  $m_i > 0$ , for all  $i \in \mathcal{N}$ , and  $\frac{3}{2} - \eta_{ij} > 0$ ,  $\frac{\eta_{ij} - (2\eta_{ij})^2}{2} > 0$ , for all  $j \in \mathcal{N}_i$  and  $i \in \mathcal{N}$ , we have the desired convergence rate, i.e., for  $k \geq 1$ ,

$$\begin{aligned} & \sum_{i \in \mathcal{N}} \mathbb{E}(m_i \|\bar{\mathbf{u}}_i(k) - \mathbf{u}_i^*\|_2^2) \\ & + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \left( \frac{3}{2} - \eta_{ij} \right) \mathbb{E}(\|\hat{\mathbf{v}}_i^j(k-1) - \bar{\mathbf{v}}_i^j(k-1)\|_2^2) \\ & + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_i} \frac{\eta_{ij} - (2\eta_{ij})^2}{2} \mathbb{E}(\|\hat{\mathbf{v}}_i^j(k-1) + \bar{\mathbf{v}}_i^j(k-1)\|_2^2) \\ & \leq \frac{1}{\underline{\alpha}k} (\|\mathbf{v}(0) - \mathbf{v}^*\|_H^2 + \frac{1}{2} \|\tilde{\boldsymbol{\lambda}}(0) - \boldsymbol{\lambda}^*\|_H^2). \end{aligned} \quad (36)$$

*Remark 5:* The inequality (36) implies that if the activation probabilities of agents and links are larger, then the convergence is achieved faster.  $\square$

## V. CONCLUSION AND FUTURE WORK

This technical note discusses a distributed algorithm for a multi-agent convex optimization problem with edge-based coupling constraints, which is related to energy management problems. The proposed method works asynchronously over time-varying communication networks. We model the asynchronicity and the time-varying nature of the communication network as random processes and show the convergence and the rate of the proposed algorithm. As future work, we consider generalizing the problem that can be dealt with, such as by introducing global objectives of control, coupling inequality constraints, or non-convex coupling constraints, which is relevant to the optimal power flow problems. Moreover, we also consider the implementation of inexact minimization to the proposed algorithm to reduce computational burden.

## REFERENCES

- [1] D. K. Molzahn, F. Dörfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, "A survey of distributed optimization and control algorithms for electric power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2941–2962, 2017.
- [2] D. P. Bertsekas, *Network optimization: continuous and discrete models*. Athena Scientific Belmont, MA, 1998.
- [3] V. Rostampour and T. Keviczky, "Probabilistic energy management for building climate comfort in smart thermal grids with seasonal storage systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3687–3697, 2019.
- [4] J. Grosso, C. Ocampo-Martínez, V. Puig, and B. Joseph, "Chance-constrained model predictive control for drinking water networks," *Journal of Process Control*, vol. 24, no. 5, pp. 504–516, 2014.
- [5] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2010, vol. 25, no. 3.
- [7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [8] A. G. Bakirtzis and P. N. Biskas, "A decentralized solution to the dc-opf of interconnected power systems," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1007–1013, 2003.
- [9] M. Kraning, E. Chu, J. Lavaei, and S. Boyd, "Dynamic network energy management via proximal message passing," *Foundations and Trends in Optimization*, vol. 1, no. 2, pp. 73–126, 2014.
- [10] T. Wang, D. O'Neill, and H. Kamath, "Dynamic control and optimization of distributed energy resources in a microgrid," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2884–2894, 2015.
- [11] C. A. Hans, P. Braun, J. Raisch, L. Grune, and C. Reincke-Collon, "Hierarchical distributed model predictive control of interconnected microgrids," *IEEE Trans. Sustainable Energy*, vol. 10, no. 1, pp. 407–416, 2019.
- [12] X. Ge, F. Yang, and Q.-L. Han, "Distributed networked control systems: A brief overview," *Information Sciences*, vol. 380, pp. 117–131, 2017.
- [13] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.
- [14] N. Chatzipanagiotis, D. Dentecheva, and M. M. Zavlanos, "An augmented Lagrangian method for distributed optimization," *Mathematical Programming*, vol. 152, no. 1, pp. 405–434, 2015.
- [15] S. Lee, N. Chatzipanagiotis, and M. M. Zavlanos, "Complexity certification of a distributed augmented lagrangian method," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 827–834, 2018.
- [16] S. Tsai, Y. Tseng, and T. Chang, "Communication-efficient distributed demand response: A randomized ADMM approach," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1085–1095, 2017.
- [17] E. Wei and A. Ozdaglar, "On the  $O(1/k)$  convergence of asynchronous distributed alternating direction method of multipliers," pp. 1–30, 2013, arXiv:1307.8254.
- [18] T. Chang, "A proximal dual consensus ADMM method for multi-agent constrained optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3719–3734, 2016.
- [19] M. Hong and T. Chang, "Stochastic proximal gradient consensus over random networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2933–2948, 2017.
- [20] T. Chang, M. Hong, W. Liao, and X. Wang, "Asynchronous distributed admm for large-scale optimization—part i: Algorithm and convergence analysis," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3118–3130, 2016.
- [21] S. Kar, G. Hug, J. Mohammadi, and J. M. F. Moura, "Distributed state estimation and energy management in smart grids: A consensus+innovations approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1022–1038, 2014.
- [22] G. Hug, S. Kar, and C. Wu, "Consensus + innovations approach for distributed multiagent coordination in a microgrid," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1893–1903, 2015.
- [23] K. Baker, J. Guo, G. Hug, and X. Li, "Distributed MPC for efficient coordination of storage and renewable energy sources across control areas," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 992–1001, 2016.
- [24] A. Kargarian, J. Mohammadi, J. Guo, S. Chakrabarti, M. Barati, G. Hug, S. Kar, and R. Baldick, "Toward distributed/decentralized dc optimal power flow implementation in future electric power systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2574–2594, 2018.
- [25] T. Sousa, T. Soares, P. Pinson, F. Moret, T. Baroche, and E. Sorin, "Peer-to-peer and community-based markets: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 104, pp. 367–378, 2019.
- [26] W. Ananduta, J. M. Maestre, C. Ocampo-Martínez, and H. Ishii, "Resilient distributed model predictive control for energy management of interconnected microgrids," *Optimal Control Applications and Methods*, vol. 41, no. 1, pp. 146–169, 2020.
- [27] A. Beck, *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, 2017.
- [28] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995.
- [29] A. Nedić, *Lecture Notes Optimization I*. Hamilton Institute, 2008.