$\mathcal{CIRFE}$: A Distributed Random Fields Estimator

Anit Kumar Sahu, *Student Member, IEEE*, Dusan Jakovetic, *Member, IEEE* and Soummya Kar, *Member, IEEE*

Abstract

This paper presents a communication efficient distributed algorithm, $\mathcal{CIRFE}$ of the *consensus+innovations* type, to estimate a high-dimensional parameter in a multi-agent network, in which each agent is interested in reconstructing only a few components of the parameter. This problem arises for example when monitoring the high-dimensional distributed state of a large-scale infrastructure with a network of limited capability sensors and where each sensor is only tasked with estimating some local components of the state. At each observation sampling epoch, each agent updates its local estimate of the parameter components in its interest set by simultaneously processing the latest locally sensed information (*innovations*) and the parameter estimates from other agents (*consensus*) in the local time-varying neighborhood of agents over a (possibly sparse) communication graph. Under minimal conditions on the inter-agent information exchange network and the sensing models, almost sure convergence of the estimate sequence at each agent to the components of the true parameter in its interest set is established. Furthermore, the paper characterizes the performance of $\mathcal{CIRFE}$ in terms of asymptotic covariance of the estimate sequences and specifically highlights the dependencies of the component wise asymptotic covariance in terms of the number of agents tasked with estimating it. Finally, simulation experiments demonstrate the efficacy of $\mathcal{CIRFE}$.

Index Terms

Distributed Estimation, Consensus Algorithms, Distributed Inference, Random Fields, Stochastic Approximation

1. INTRODUCTION

Large scale cyber-physical systems (CPS) have gained prominence recently. A large-scale CPS is constituted by a network of possibly heterogeneous entities, where each entity has sensing, computation and communication capabilities. A desirable trait of a CPS is to be energy efficient so as to have a longer operational life. Owing to the ad-hoc nature of the networked entities in a CPS and the deployment of such systems in a large-scale fashion spread over a large area, the sensing and the data collection are inherently distributed in such systems. Due to the large-scale deployment of such systems compounded with the inherent distributed nature of the data collection, a centralized coordinator based architecture is undesirable. A typical entity of a CPS with on board sensing and

A. K. Sahu and S. Kar are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA (email:anits@andrew.cmu.edu, soummyak@andrew.cmu.edu). D. Jakovetic is with the Department of Mathematics and Informatics, University of Novi Sad, Serbia (e-mail: djakovet@uns.ac.rs).

The work of A. K. Sahu and S. Kar was supported in part by NSF under grants CCF-1513936. The work of D. Jakovetic was supported by Ministry of Education, Science and Technological Development, Republic of Serbia, grant no. 174030.

computational capabilities is plagued by power hungry communication needs. Also, relevant to a large-scale CPS is the associated high-dimensional state of the CPS, which the network as a whole wants to reconstruct. Moreover, the high-dimensional state reconstruction needs to be done in a recursive fashion, which in essence points to the fact that the state estimate needs to be updated at each entity as and when a new datum is collected. The high-dimensional state of the CPS in conjunction with the power hungry communication bottleneck makes the problem at hand very challenging. In this paper, we are interested in distributed inference of the state of the aforementioned large-scale CPS like sensor networks monitoring a spatially distributed field tracking environmental modalities, or CPS where physical entities with sensing capabilities are deployed over large areas. An important example of the systems of interest is the smart grid–a large network of generators and loads instrumented with, for example, phasor measurement units (PMUs). Our goal is to reconstruct the physical field or state of the CPS that is represented by a parameter or a random field. The structure of the physical layer is reflected through the coupling among the observation sequences across different nodes. Suppose, for the purpose of illustration, corresponding to each field location, there is a low-power inexpensive sensor monitoring the location. The noisy sensor measurement at a location in the field is not only influenced by the parameter or field component at that location but possibly is a function of *neighboring* field components. As an example, in the smart grid context, a sensor at a node (location) may obtain a measurement of the power flowing into that node, which in turn is a function of the field components (e.g., voltages, angles) at that node and neighboring nodes. This coupling among parameter components in the measurements will be referred to as the *physical coupling* in the sequel. However, since the local measurement at a location is influenced by multiple field parameters, due to possible lack of observability or identifiability, in order to come up with a provably consistent estimates of the parameter components of interest, each agent exchanges information with its neighborhood which conforms to a pre-assigned inter-agent communication graph. The inter-agent communication graph forms the cyber layer of the system and is different from that of the physical layer, i.e., the coupling structure among the parameter components induced by the distributed measurement model. Due to the large-scale of the CPS and the high-dimensionality of the field, reconstructing the entire field at each agent may be too taxing and beyond the capability of the agents, and hence, agents may only be interested in estimating certain components of the parameter field locally; furthermore, the components of interest at a given agent, referred to as the *interest set* of the agent, varies from agent to agent. Broadly speaking, the observation model we adopt in this paper is of the form,

$$\mathbf{y}_n(t) = \mathbf{H}_n\boldsymbol{\theta}^* + \gamma_n(t),$$

where $\boldsymbol{\theta}^* \in \mathbb{R}^N$, and the dimension of $\boldsymbol{\theta}^*$ corresponds to the number of physical locations being monitored in the random field, $\mathbf{H}_n$ is a (fat) matrix that abstracts out the coupling in measurements for agent n with the state values at other nodes and $\gamma_n(t)$ represents the observation noise (to be specified later). Existing distributed estimation schemes, such as in [1]–[9] aim to reconstruct the entire parameter at each node of the networked setup, thus conforming to a homogeneous objective across all nodes. However, in this paper, we consider a distributed estimation scheme which allows agents to pursue a heterogeneous objective, in which agents' only estimate a few components of the parameter vector $\boldsymbol{\theta}^*$ corresponding to their *interest sets*. In other words, in our distributed estimation setup, agents may be interested in estimating different subsets of the parameter field, thus having heterogeneous estimation objectives. Distributed recursive inference schemes addressing estimation of a possibly high-dimensional parameter vector (see, for example [1]–[9]), tend to communicate at each (discrete or slotted) time step in its neighborhood and exchange estimates of the entire parameter vector at each time instant. Owing to the high-dimensionality of the state vector and limited storage and processing capabilities in the individual entities of a large-scale CPS,

exchanging high-dimensional estimates may be undesirable. Hence, in this paper, we consider a distributed estimator where each agent only infers a fraction of the field, but through cooperation and under the appropriate conditions generates provably consistent estimates of this fraction of the field. Instead of reconstructing the entire parameter at each agent, the distributed estimator we present is efficient from a computation and communication point of view, reconstructing a certain subset of the field based on the requirements of each agent, while adhering to a frugal communication protocol. In existing distributed estimation schemes such as in [1]–[9], the global model information in terms of the sensing models of all the agents are assumed to be inaccessible for an individual agent. However, the aforementioned setups subsume the knowledge of the dimension of the state vector to be estimated and hence adapt the storage requirements at each agent to cater to the exact dimension of the state vector. In contrast, in this paper, we present a distributed estimation algorithm of the *consensus* + *innovations* form ([1], [2]), namely $\mathcal{CIRFE}$, *consensus+innovations* Random Fields Estimator, where each agent reconstructs only a subset of the field by simultaneously processing information obtained from its neighbors (*consensus*) and the latest sensed information (*innovation*). It is of particular interest, that an individual agent in our setup does not have access to the global model information and any information about the dimension of the state vector corresponding to the field. The information about the state vector is constrained to neighborhoods where an individual agent has only information about the components of the state vector which potentially affect its own measurements, alleviating storage and model knowledge requirements.

Our main contributions are as follows:

Main Contribution 1: We propose a scheme, namely $\mathcal{CIRFE}$, where each entity reconstructs only a subset of the components of the state modeled by a vector parameter, and thereby also reducing the dimension of messages being communicated among the agents. Under mild conditions of the connectivity of the network, we establish consistency of the estimate sequence at each agent with respect to the components of the parameters in its interest set. The proposed scheme allows heterogeneity in terms of agents' objectives, while still allowing for inter-agent collaboration. Technically speaking, the heterogeneity in terms of agents' objectives seeks for consensus in agents' estimates only in terms of the components in their interest sets rather than the entire high-dimensional parameter vector as in the case of other distributed estimation algorithms proposed in [1]–[9].

Main Contribution 2: Technically, the consensus+innovations type approach that we employ for state reconstruction in the current setting constitutes a mixed time-scale stochastic approximation procedure [10]. We explicitly evaluate the asymptotic covariance of the component wise estimate sequences at each agent. The obtained asymptotic covariance is heterogeneous in terms of scaling of the variances of the components of the parameter based on the number of agents interested in reconstructing a particular component. Intuitively, the analysis in this paper explicitly establishes the fact that a component of the parameter vector which is in the interest set of a large number of nodes tends to generate an estimate sequence with a lower asymptotic variance as compared to a component of the state vector which is in the interest set of fewer number of agents. Finally, as a consequence of our estimator design, we show that the asymptotic covariance is independent of the inter-agent communication topology as long as the *time-averaged* network is connected. To the best of our knowledge, this is the first asymptotic covariance evaluation explicitly in terms of the number of agents interested in reconstructing entries of the state vector for distributed

estimation of high-dimensional fields, i.e., when each node is interested only in a subset of the vector parameter. As shown here, the all-nodes-interested-in-all parameters scenario, most commonly studied in the literature, is a special case of the result derived here.

Related Work: Relevant distributed estimation literature can be classified primarily into two types. The first type includes schemes which involve single snapshot data collection followed by inter-agent fusion through consensus type protocols (see, for example, [3], [4], [11]–[13]). The second type includes estimation schemes where the sensing and the processing of the information occur at the same time (see, for example [5]–[8], [14]–[19]). Representative approaches of this latter class are *consensus+innovations* type ([1], [9]) and the diffusion type ([20], [21]) algorithms. Distributed inference algorithms for random fields have been proposed in literature, see, for example [22], [23]. Reference [22] considers the estimation of a time-varying random field pertaining to a linear observation model, where each agent reconstructs only a few components of the field. However, in contrast with [22] where the incorporation of new sensed information is followed by multiple rounds of consensus, the proposed algorithm in this paper simultaneously fuses the neighborhood information and the current observation albeit for a static random field. In contrast with [23], where each agent tries to reconstruct the entire time-varying random field, the proposed algorithm reconstructs only a subset of the components of the entire field at each agent and the information exchange entails a low dimensional vector instead of the entire parameter. A random field estimation scheme in a fully distributed setup with arbitrary connected inter-agent communication topology where agents reconstruct only a subset of the physical field which, in turn, is coupled with the sensing field was also proposed in [24] (Chapter 3). The current work is inspired by [24] and generalizes the development in [24] in several fronts to achieve better estimate performance. In [24], a *single time-scale* consensus+innovations algorithm pertaining to a linear observation model was proposed and the consistency and asymptotic normality[1] of the estimator was established. By, a *single time-scale* consensus+innovations algorithm, we mean algorithms where the consensus and innovation potentials are controlled by the same time-decaying sequence, i.e., the information from the neighborhood and the newly sensed information at an agent are weighed equally across slotted time instants. The performance of the single time-scale version of the consensus+innovations distributed estimation algorithm in terms of asymptotic variance depends on the network topology and is thus affected when the connectivity of the network is relatively poor. In contrast with [24], we propose a *consensus + innovations* algorithm, where the *consensus* and *innovations* terms are weighed through different carefully crafted time-varying sequences, where the consensus among a pair of agents is limited to the common components of the state vector in their respective sets thus allowing heterogeneity in terms of agents' objectives. In this paper, we not only establish the consistency and asymptotic normality of the parameter estimate sequence but also, due to the employed mixed time stochastic approximation obtain the asymptotic covariance of the estimate sequences to be independent of the particular communication network instance. Recently, other works in literature have looked at similar setups, albeit in the distributed optimization framework [25], [26].

Paper Organization : The rest of the paper is organized as follows. Spectral graph theory and notation are discussed

[1]An estimate sequence is asymptotically normal if its $\sqrt{t}$ scaled error process, i.e., the difference between the sequence and the true parameter converges in distribution to a normal random variable.

next. The sensing model and the preliminaries are discussed in Section 2. Section 3 presents the proposed distributed estimation algorithm, while Section 4 and Section 6 concerns with the main results of the paper and the proof of the main results respectively. The simulation experiments for the proposed algorithm are presented in Section 5. Finally, Section 7 concludes the paper.

Notation. We denote by $\mathbb{R}$ the set of reals, and by $\mathbb{R}^k$ the k-dimensional Euclidean space. Vectors and matrices are in bold faces. We also denote by $\mathbf{A}_{ij}$ or $[\mathbf{A}]_{ij}$, the (i,j)-th entry of a matrix $\mathbf{A}$; $\mathbf{a}_i$ or $[\mathbf{a}]_i$ the i-th entry of a vector $\mathbf{a}$. The symbols $\mathbf{I}$ and $\mathbf{0}$ are the $k \times k$ identity matrix and the $k \times k$ zero matrix, respectively, the dimensions being clear from the context. The vector $\mathbf{e_i}$ is the i-th column of $\mathbf{I}$, also referred to as a canonical vector. The symbol $\top$ stands for matrix transpose. The operator $\otimes$ denotes the Kronecker product. The operator $||.||$ applied to a vector is the standard Euclidean $\mathcal{L}_2$ norm, while when applied to matrices stands for the induced $\mathcal{L}_2$ norm, which is equal to the spectral radius for symmetric matrices. The cardinality of a set $\mathcal{S}$ is $|\mathcal{S}|$. Finally, $diag(\mathbf{v})$ denotes the diagonal matrix with its diagonal elements as $\mathbf{v}$. All inequalities involving random variables are to be interpreted almost surely (a.s.).

Spectral Graph Theory. The inter-agent communication network is a simple[2] undirected graph $G = (V, E)$, where V denotes the set of agents or vertices with cardinality $|V| = N$, and E the set of edges with $|E| = M$. If there exists an edge between agents i and j, then $(i,j) \in E$. A path between agents i and j of length m is a sequence $(i = p_0, p_1, \cdots, p_m = j)$ of vertices, such that $(p_t, p_{t+1}) \in E$, $0 \leq t \leq m-1$. A graph is connected if there exists a path between all possible agent pairs. The neighborhood of an agent n is given by $\Omega_n = \{j \in V | (n,j) \in E\}$. The degree of agent n is given by $d_n = |\Omega_n|$. The structure of the graph is represented by the symmetric $N \times N$ adjacency matrix $\mathbf{A} = [A_{ij}]$, where $A_{ij} = 1$ if $(i,j) \in E$, and 0 otherwise. The degree matrix is given by the diagonal matrix $\mathbf{D} = diag(d_1 \cdots d_N)$. The graph Laplacian matrix is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$. The Laplacian is a positive semidefinite matrix, hence its eigenvalues can be ordered and represented as $0 = \lambda_1(\mathbf{L}) \leq \lambda_2(\mathbf{L}) \leq \cdots \lambda_N(\mathbf{L})$. Furthermore, a graph is connected if and only if $\lambda_2(\mathbf{L}) > 0$ (see [27] for instance).

2. SENSING MODEL AND PRELIMINARIES

Consider N physical agents monitoring a field over a large physical area. Each agent n is associated with a scalar state θ_n^*, which represents the field intensity parameter at its location. The agents are equipped with sensing capabilities. We assume each agent observes a time-series of measurements, given by noisy linear functions of its state and the states of *neighboring* agents. Due to this coupling in the observations, an agent should cooperate with neighbors to reconstruct its own state. For simplicity, we assume that the individual agent states are scalars. Our results can be generalized to vector valued states, though at the cost of extra notation. The observation at each agent is of the form:

$$\mathbf{y}_n(t) = \mathbf{H}_n \boldsymbol{\theta}^* + \gamma_n(t), \tag{1}$$

[2]A graph is said to be simple if it is devoid of self loops and multiple edges.

where $\mathbf{H}_n \in \mathbb{R}^{M_n \times N}$ is a sparsifying (to be clarified soon) sensing matrix, $\{\mathbf{y}_n(t)\}$ is a $\mathbb{R}^{M_n}$-valued observation sequence for the n-th agent and for each n where possibly $M_n \ll N$, $\{\gamma_n(t)\}$ is a zero-mean temporally independent and identically distributed (i.i.d.) noise sequence with nonsingular covariance matrix $\mathbf{R}_n$.

Assumption A1. *There exists $\epsilon_1 > 0$, such that, for all n, $\mathbb{E}_{\boldsymbol{\theta}}\left[\|\gamma_n(t)\|^{2+\epsilon_1}\right] < \infty$.*

The above assumption encompasses a broad class of noise distributions in the setup. The heterogeneity of the setup is exhibited in terms of the sensing matrix and the noise covariances at the agents. We now formalize an assumption on global model observability.

Assumption A2. *The matrix $\mathbf{G} = \sum_{n=1}^{N} \mathbf{H}_n^\top \mathbf{R}_n^{-1} \mathbf{H}_n$ is full rank.*

Assumption A2 is crucial for our distributed setup. It is to be noted that such an assumption is needed for even a setup with a centralized node which has access to all the data samples at each of the agent nodes at each time. Assumption A2 ensures that if a hypothetical fusion center could stack all the data samples together at any time t, it would have sufficient information so as to be able to unambiguously estimate the parameter of interest. Hence, the requirement for this assumption naturally extends to our distributed setup. As far as reconstructing the parameter $\boldsymbol{\theta}$ is concerned, there is an inherent scalability issue as the dimension of the parameter scales with the size of the network. Owing to the ad-hoc nature of setups as described above and observations being made at different agents in a sequential manner, one has to resort to recursive message-passing schemes while conforming to a communication protocol specified by a inter-agent communication graph. Given the possibly high-dimensional state of the field, it is not desirable and communication-wise feasible to exchange the high-dimensional data in the form of parameter estimates and for each agent to estimate the entire vector. Before, going over specifics of our algorithm, we next review recursive estimation both in the centralized and distributed setups.

A. Preliminaries

In this section, we go over the preliminaries of classical distributed estimation.

Distributed Estimation:

In the setup described above in (1), if a hypothetical fusion center having access to the data samples at all nodes at all times were to conduct the parameter estimation in a recursive manner, a (centralized) recursive least-squares type approach could be employed as follows:

$$\mathbf{x}_c(t+1) = \mathbf{x}_c(t) + \underbrace{\frac{a}{t+1} \sum_{n=1}^{N} \mathbf{H}_n^\top \boldsymbol{\Sigma}_n^{-1} \left(\mathbf{y}_n(t) - \mathbf{H}_n \mathbf{x}_c(t)\right)}_{\text{Global Innovation}}.$$

However, such a fusion center based scheme may not be implementable in our distributed multi-agent setting with time-varying sparse inter-agent interaction primarily due to the fact that the desired global innovation computation requires instantaneous access to the entire set of network sensed data at all times at the fusion center. Moreover, the fusion center intends to reconstruct the entire high-dimensional state and thus, maintains a N-dimensional estimate at all times. If in the case of a distributed setup, an agent n in the network were to replicate the centralized update

by replacing the global innovation in accordance with its local innovation, the update for the parameter estimate becomes

$$\widehat{\mathbf{x}}_n(t+1) = \widehat{\mathbf{x}}_n(t) + \underbrace{\frac{a}{t+1}\mathbf{H}_n^\top\left(\widehat{\mathbf{x}}_n(t)\right)\boldsymbol{\Sigma}_n^{-1}\left(\mathbf{y}_n(t) - \mathbf{H}_n\widehat{\mathbf{x}}_n(t)\right)}_{\text{Local Innovation}},$$

where $\{\widehat{\mathbf{x}}_n(t)\}$ represents the estimate sequence at agent n. The above update involves purely decentralized and independent local processing with no collaboration among the agents whatsoever. However, note that in the case when the data samples obtained at each agent lacks information about all the features, the parameter estimates would be erroneous and sub-optimal. As in the case of the fusion center based approach outlined above, each agent maintains a N-dimensional estimate at all times and hence the messages exchanged in the neighborhood are N-dimensional and could be very large depending on the size of the network. Hence, as a surrogate to the global innovation in the centralized recursions, the local estimators compute a local innovation based on the locally sensed data as an agent has access to information only in its neighborhood. The information loss at a node is compensated by incorporating an agreement or consensus potential into their updates which is then incorporated (see, for example [1], [28], [29]) as follows:

$$\mathbf{x}_n(t+1) = \mathbf{x}_n(t) - \underbrace{\frac{b}{(t+1)^{\delta_1}}\sum_{l\in\Omega_n(t)}\left(\mathbf{x}_n(t) - \mathbf{x}_l(t)\right)}_{\text{Neighborhood Consensus}} + \underbrace{\frac{a}{t+1}\boldsymbol{\Gamma}^{-1}\mathbf{H}_n^\top\boldsymbol{\Sigma}_n^{-1}\left(\mathbf{y}_n(t) - \mathbf{H}_n\mathbf{x}_n(t)\right)}_{\text{Local Innovation}}, \tag{2}$$

where $0 < \delta_1 < 1$, $\Omega_n(t)$ represents the neighborhood of agent n at time t and a, b are appropriately chosen positive constants. In the above scheme, the information exchange among agent nodes is limited to the parameter estimates. It has been shown in previous work that under appropriate conditions (see, for example [1]), the estimate sequence $\{\mathbf{x}_c(t)\}$ converges to $\boldsymbol{\theta}$ and is asymptotically normal, i.e.,

$$\sqrt{t+1}\left(\mathbf{x}_n(t) - \boldsymbol{\theta}\right) \xRightarrow{\mathcal{D}} \mathcal{N}\left(0, (N\boldsymbol{\Gamma})^{-1}\right),$$

where $\boldsymbol{\Gamma} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n$ and $\xRightarrow{\mathcal{D}}$ denotes convergence in distribution. The above established asymptotic normality also points to the conclusion that the MSE decays as $\Theta(1/t)$. For future reference, we will refer to the distributed estimation approach in (2) as the classical consensus+innovations approach. The aforementioned scheme, though optimal in terms of the asymptotic covariance entails the availability of global model information at each agent and exchange of the entire parameter estimate which in turn is N-dimensional among agents. Furthermore, due to the inherent spatial coupling in the observation sequence at each node with other nodes in its neighborhood, the availability of a particular entry of the state vector is localized to a small area. Hence, a large-scale deployment of such a system, would incorporate a significant delay for an agent to assimilate information about a particular entry of the state vector which is not local with respect to its neighborhood. Moreover, such a scheme requires the knowledge of the dimension of the state vector at each agent and storage of a high-dimensional local estimates, same as the size of the entire state vector. Such prior knowledge about attributes of the parameter such as dimension in conjunction with requirement for large memory at each agent might be practically infeasible owing to the ad-hoc nature and limited sensing, computation and storage capabilities of agents in a networked setup.

Thus, in both of the schemes above, specifically in the case which involves estimating a high-dimensional parameter,

it might not be practical to estimate the entire parameter at each agent. In such a high-dimensional parameter estimation scheme, it is highly favorable to estimate only a few entries of the parameter based on the requirements of each agent, which could potentially reduce the dimensions of messages being exchanged in the network thereby reducing the implementation complexity considerably.

3. $\mathcal{CIRFE}$: DISTRIBUTED RANDOM FIELDS ESTIMATION

In this section, we develop the algorithm $\mathcal{CIRFE}$. The parameter to be reconstructed which is the vector of states accumulated over the entire network is $\boldsymbol{\theta}^* \in \mathbb{R}^N$. The sparsifying nature of $\mathbf{H}_n$ in (1) is related to the coupling induced by the measurements in the field. To be specific, let us define $\widetilde{\mathcal{I}}_n$ as the set of agents whose states influence the measurement $\mathbf{y}_n(t)$ at agent n, i.e., $\widetilde{\mathcal{I}}_n$ collects the agents for which the corresponding columns of matrix $\mathbf{H}_n$ is non-zero. In what follows, we say an agent n is physically coupled to an agent l if the observation at agent n is influenced by the state component θ_l^*. Typically, $\widetilde{\mathcal{I}}_n$ is a small subset of the total number of agents N. Technically speaking, the above mentioned coupling induced by the measurements can be expressed in terms of an adjacency matrix, $\hat{\mathbf{A}}$, where $\hat{\mathbf{A}}_{nl} = 1$ if $l \in \widetilde{\mathcal{I}}_n$ and 0 otherwise. Now, that we have abstracted out the physical coupling (physical layer) in the networked system under consideration, we discuss about the communication layer (cyber layer), i.e., the inter-agent communication network and the associated communication protocol. Before getting into the communication protocol, we introduce *interest sets* of agents' around which the communication protocol is built. We intend to formulate a distributed estimation procedure, where every agent wants to reconstruct the states of a small subset of the agents, which we refer to as the *interest set* of the agent. In what follows, we point out that the n-th component of the field has a one-to-one correspondence with the n-th agent: this one-to-one correspondence is best illustrated by visualizing the agents to be (geographically) distributed in a field with θ_n^* representing the state of the field at the location of the n-th agent. Formally, the interest set of an agent is represented as $\mathcal{I}_n$. The interest set could vary from one agent to another. The interest sets can be arbitrary but need to satisfy the following assumption:

Assumption A3. *The set of agents physically coupled with agent n is a subset of the interest set of agent n, i.e., $\widetilde{\mathcal{I}}_n \subset \mathcal{I}_n$.*

We assume without loss of generality that $\widetilde{\mathcal{I}}_n$ and hence $\mathcal{I}_n$ is non-empty for all n. (For illustration, see below the example after Assumption A6). We number the nodes (equivalently, components of $\boldsymbol{\theta}$) in the interest sets of agents in increasing order. Thus, the interest set $\mathcal{I}_n$ at an agent n can be considered to be a vector with dimension $|\mathcal{I}_n|$. For example, $\mathcal{I}_n(r) = p$ indicates that agent p is the r-th agent in increasing order in the interest set $\mathcal{I}_n$. We also have that $\mathcal{I}_n^{-1}(p) = r$. Moreover, as each agent n is only interested in reconstructing the states of agents in its interest set, the estimate at agent n, $\mathbf{x}_n(t) \in \mathbb{R}^{|\mathcal{I}_n|}$, $\forall\ t$. At every time instant t, an agent n simultaneously fuses information received from the neighbors and the latest sensed information to update its parameter estimate. However, as the interest set of agents in the neighborhood might not be the same as that of the agent itself, the information received from the neighbors needs censoring. Let the message received from agent l at time t be denoted by $\mathbf{x}_l(t) \in \mathbb{R}^{|\mathcal{I}_l|}$,

where $l \in \Omega_n$. The censored message processed by agent n, $\mathbf{x}^r_{l,n}(t) \in \mathbb{R}^{|\mathcal{I}_n|}$ is generated as follows:

$$\mathbf{e}_j^\top \mathbf{x}^r_{l,n}(t) = \begin{cases} \mathbf{e}^\top_{\mathcal{I}_l^{-1}(\mathcal{I}_n(j))} \mathbf{x}_l(t) & \mathcal{I}_n(j) \in \mathcal{I}_l \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $\mathbf{e}_j$ and $\mathbf{e}_{\mathcal{I}_l^{-1}(\mathcal{I}_n(j))}$ are canonical vectors with $\mathbf{e}_j \in \mathbb{R}^{|\mathcal{I}_n|}$ and $\mathbf{e}_{\mathcal{I}_l^{-1}(\mathcal{I}_n(j))} \in \mathbb{R}^{|\mathcal{I}_l|}$. Agent n only wants to use estimates of those states from an agent in its neighborhood which are common to their interest sets. Formally, with agent l, agent n only wants to use estimates of the states in the set $\mathcal{I}_n \cap \mathcal{I}_l$. Similarly, while using the obtained estimate states from the neighbors, only those states in the set $\mathcal{I}_n \cap \mathcal{I}_l$ are updated. We also define the transformed estimate $\mathbf{x}^s_{l,n}(t) \in \mathbb{R}^{|\mathcal{I}_n|}$ at agent n, for each $l \in \Omega_n(t)$ as follows:

$$\mathbf{e}_j^\top \mathbf{x}^s_{l,n}(t) = \begin{cases} \mathbf{e}_j^\top \mathbf{x}_n(t) & \mathcal{I}_n(j) \in \mathcal{I}_l \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

where $j \in \{1, \cdots, |\mathcal{I}_n|\}$. The agent n also incorporates the latest sensed information $\mathbf{y}_n(t)$ while updating the parameter estimate at each sampling epoch and only retains the components of interest, i.e., those in $\mathcal{I}_n$. For a given vector $\mathbf{z} \in \mathbb{R}^{|\mathcal{I}_n|}$, let $\mathbf{z}^{\mathcal{P}_{\mathcal{I}_n}} \in \mathbb{R}^N$ be the vector whose j-th component is given by

$$\mathbf{e}_j^\top \mathbf{z}^{\mathcal{P}_{\mathcal{I}_n}} = \begin{cases} \mathbf{e}^\top_{\mathcal{I}_n^{-1}(j)} \mathbf{y} & \mathcal{I}_n(j) \in \mathcal{I}_l \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Finally, for a given vector $\mathbf{z} \in \mathbb{R}^N$, $\mathbf{z}_{\mathcal{I}_n}$ denotes the vector in $\mathbb{R}^{|\mathcal{I}_n|}$, where $\mathbf{e}_j^\top \mathbf{z}_{\mathcal{I}_n} = \mathbf{e}^\top_{\mathcal{I}_n(j)} \mathbf{z}$.
We now introduce the algorithm $\mathcal{CIRFE}$ for distributed parameter estimation:

$$\begin{aligned} \mathbf{x}_n(t+1) = \mathbf{x}_n(t) - \underbrace{\sum_{l \in \Omega_n(t)} \beta_t \left(\mathbf{x}^s_{l,n}(t) - \mathbf{x}^r_{l,n}(t) \right)}_{\text{Neighborhood Consensus}} \\ + \underbrace{\alpha_t \mathbf{H}_n^\top \mathbf{R}_n^{-1} \left(\mathbf{y}_n(t) - \mathbf{H}_n \mathbf{x}_n^{\mathcal{P}_{\mathcal{I}_n}}(t) \right)_{\mathcal{I}_n}}_{\text{Local Innovation}}, \end{aligned} \tag{6}$$

where $\Omega_n(t)$ represents the neighborhood of agent n at time t; and $\{\beta_t\}$ and $\{\alpha_t\}$ are the consensus and innovation weight sequences given by

$$\beta_t = \frac{\beta_0}{(t+1)^{\delta_1}}, \alpha_t = \frac{a}{t+1}, \tag{7}$$

where $a, b > 0$ and $0 < \delta_1 < 1/2 - 1/(2+\epsilon_1)$ and ϵ_1 was as defined in Assumption A1. It is to be noted that with the interest set of each agent being $\mathcal{I}_n = \{1, 2, \cdots, N\}$, we have that the update in (6) reduces to the classical consensus+innovations update for linear parameter estimation schemes (see, [28] for example). Thus, the classical consensus+innovations parameter estimation scheme, is strictly a special case of the update in (6). We formalize an assumption on the connectivity of the inter-agent communication graph before proceeding further.

Assumption A4. *The inter-agent communication graph is connected on average, i.e., $\lambda_2(\overline{\mathbf{L}}) > 0$, where $\overline{\mathbf{L}}$ denotes the mean of the sequence of identically and independently distributed (i.i.d) graph Laplacian sequence $\{\mathbf{L}(t)\}$.*

Remark 3.1. *In the parameter estimation scheme in* (6)*, an agent n uses only those components of its neighbor l's estimate $\mathbf{x}_l(t)$, which belong to its interest set $\mathcal{I}_n$. Thus, agents n and l combine components linearly which belong to $\mathcal{I}_n \cap \mathcal{I}_l$ and reject the rest of the components. From an implementation viewpoint, it is desirable for an agent l to only transmit those components to agent n which belong to $\mathcal{I}_n \cap \mathcal{I}_l$ instead of transmitting the entire*

$\mathbf{x}_l(t)$ *to agent* n *as the one which involves exchanging only those components which are common to the agents has lower communication overhead. In the former case, the receiving agent* n *will zero out the components it does not require, so both the transmission strategies would lead to the same update. Moreover, in the innovation term, where an agent* n *uses its own previous state to compute the innovation, an agent subsequently retains only the components of interest so as to keep the update economical in terms of size. We also emphasize here that the inter-agent communication graphs* $\{L(t)\}$ *and the physical adjacency matrix* $\hat{A}$ *induced by the measurement coupling may be structurally different.*

We now present a more compact representation of the $\mathcal{CIRFE}$ algorithm so as to be able to establish its asymptotic convergence properties. Let $\mathcal{I}$ denote a subset of $\{1, 2, \cdots, N\}$. Define the diagonal matrix $P_{\mathcal{I}}$ which selects the corresponding non-zero components of $\mathcal{I}$ from a $\mathbb{R}^{N^2}$ dimensional vector. For the estimate sequence $\{\mathbf{x}_n(t)\}$ at agent n, let $\{\widetilde{\mathbf{x}}_n(t)\} \in \mathbb{R}^N$ denote the auxiliary estimate sequence, where $\widetilde{\mathbf{x}}_n(t) = \mathbf{x}_n(t)^{\mathcal{P}_{\mathcal{I}_n}}$. With the above development in place, it is easy to see that, for $\mathbf{y} \in \mathbb{R}^N$, $\left(\mathbf{x}^r_{l,n}(t)\right)^{\mathcal{P}_{\mathcal{I}_n}} = P_{\mathcal{I}_n} P_{\mathcal{I}_l} \widetilde{\mathbf{x}}_l(t)$, $\left(\mathbf{x}^s_{l,n}(t)\right)^{\mathcal{P}_{\mathcal{I}_n}} = P_{\mathcal{I}_n} P_{\mathcal{I}_l} \widetilde{\mathbf{x}}_n(t)$ and $(\mathbf{x}_n(t))^{\mathcal{P}_{\mathcal{I}_n}} = P_{\mathcal{I}_n} \widetilde{\mathbf{x}}_n(t)$. The $\mathcal{CIRFE}$ update in (6) can then be written in terms of the auxiliary processes as follows:

$$\begin{aligned}\widetilde{\mathbf{x}}_n(t+1) = \widetilde{\mathbf{x}}_n(t) - \sum_{l \in \Omega_n(t)} \beta_t P_{\mathcal{I}_n} P_{\mathcal{I}_l} \left(\widetilde{\mathbf{x}}_n(t) - \widetilde{\mathbf{x}}_l(t)\right) \\ + \alpha_t P_{\mathcal{I}_n} \mathbf{H}_n^\top \mathbf{R}_n^{-1} \left(\mathbf{y}_n(t) - \mathbf{H}_n P_{\mathcal{I}_n} \widetilde{\mathbf{x}}_n(t)\right). \end{aligned} \tag{8}$$

We introduce the matrix $\mathbf{L}_{\mathcal{P}} \in \mathbb{R}^{N^2 \times N^2}$ so as to make the above representation more compact.

$$[\mathbf{L}_{\mathcal{P}}(t)]_{nl} = \begin{cases} -P_{\mathcal{I}_n} \sum_{r=1:r\neq n}^{N} \mathbf{L}_{nr}(t) P_{\mathcal{I}_r} & \text{if } n = l \\ \mathbf{L}_{nl}(t) P_{\mathcal{I}_l} P_{\mathcal{I}_n} & \text{otherwise,} \end{cases} \tag{9}$$

where $[\mathbf{L}_{\mathcal{P}}(t)]_{nl} \in \mathbb{R}^{N \times N}$ denotes the (n, l)-th sub-block of the block matrix $\mathbf{L}_{\mathcal{P}}$. It follows by elementary matrix multiplication properties that $\mathcal{P}\mathbf{L}_{\mathcal{P}}(t) = \mathbf{L}_{\mathcal{P}}(t)$. It is also to be noted that $\mathbf{L}_{\mathcal{P}}$ is a symmetric matrix. The matrix $\mathbf{L}_{\mathcal{P}}(t)$ at each time step t can be decomposed as follows:

$$\mathbf{L}_{\mathcal{P}}(t) = \overline{\mathbf{L}_{\mathcal{P}}} + \widetilde{\mathbf{L}_{\mathcal{P}}}(t), \tag{10}$$

where $\{\mathbf{L}_{\mathcal{P}}(t)\}$ is an i.i.d. sequence with mean $\overline{\mathbf{L}_{\mathcal{P}}}$ and $\widetilde{\mathbf{L}_{\mathcal{P}}}(t) = \mathbf{L}_{\mathcal{P}}(t) - \mathbb{E}\left[\mathbf{L}_{\mathcal{P}}(t)\right]$. Thus, we have that the residual sequence $\{\widetilde{\mathbf{L}_{\mathcal{P}}}(t)\}$ satisfies $\mathbb{E}\left[\widetilde{\mathbf{L}_{\mathcal{P}}}(t)\right] = \mathbf{0}$.
With the above development in place, the update in (8) can be written in a compact form as follows:

$$\widetilde{\mathbf{x}}(t+1) = \widetilde{\mathbf{x}}(t) - \beta_t \mathbf{L}_{\mathcal{P}}(t) \widetilde{\mathbf{x}}(t) + \alpha_t \mathcal{P} \mathbf{G}_H \mathbf{R}^{-1} \left(\mathbf{y}(t) - \mathbf{G}_H^\top \mathcal{P} \widetilde{\mathbf{x}}(t)\right), \tag{11}$$

where $\widetilde{\mathbf{x}}^\top(t) = \left[\widetilde{\mathbf{x}}_1^\top(t), \cdots, \widetilde{\mathbf{x}}_N^\top(t)\right]^\top$, $\mathbf{y}(t)^\top = [y_1(t)^\top \cdots y_N(t)^\top]^\top$, $\mathbf{R} = diag\left[\mathbf{R}_1, \cdots, \mathbf{R}_N\right]$, $\mathcal{P} = diag\left[\mathcal{P}_{\mathcal{I}_1}, \cdots, \mathcal{P}_{\mathcal{I}_N}\right]$, and $\mathbf{G}_H = diag[\mathbf{H}_1^\top, \mathbf{H}_2^\top, \cdots, \mathbf{H}_N^\top]$.

Remark 3.2. *The recursive update in* (11) *is of the stochastic approximation type. The stochastic approximation procedure, employed here is a mixed time-scale stochastic approximation as opposed to the classical single time-scale stochastic approximation ([10]). The above notion of mixed time-scale is very different from the more commonly studied two time-scale stochastic approximation (see, for instance [30]) in which a fast process is coupled with a slower dynamical system. The approach employed here is similar to the ones in [31] and [28] in*

which a single update procedure is influenced by multiple potentials with different time-decaying weights. Now, suppose that the interest set of each agent consists of all components of $\boldsymbol{\theta}^*$*, i.e., the update in* (11) *reduces to the classical consensus+innovations update in (2). A key technical step employed in the analysis of classical consensus+innovations procedures of the type in* (2) *(see, for example, [28]) consists of an approximation of the update in* (2) *to a single time-scale stochastic approximation procedure that is* asymptotically equivalent *to the former, in particular, that converges to the original iterate sequence at a rate faster than* $(t+1)^{0.5}$*. Typically, in the context of* (2) *the approximating single time-scale procedure is the network-averaged estimate sequence,* $\widetilde{\mathbf{x}}_{avg}(t) = \left(\frac{\mathbf{1}_N^\top}{N} \otimes \mathbf{I}_N\right) \widetilde{\mathbf{x}}(t)$*, and the analysis in [28] the fact that the Laplacian* $\mathbf{L}(t)$ *in* (2) *has a left eigen vector of* $\mathbf{1}_{N^2}$ *and that every agent is interested in estimating the entire parameter vector. However, in the context of the update in* (11)*, every agent is interested in only a few entries of the parameter which makes the characterization of asymptotic properties of the estimate sequences highly non-trivial and substantially different from prior work on consensus+innovations type estimation procedures [28] in which agents share the common objective of estimating all components of the parameter. However, in contrast to prior work on consensus+innovations type estimation procedures (see, for example [28]) in which agents share the common objective of estimating all components of the parameter, the analysis with heterogeneous agent objectives in* (11)*, in that each agent is interested in a different subset of components, requires new technical machinery. In particular, to obtain asymptotic properties of* (11)*, we develop a more generalized approximation of the mixed time-scale procedure to an appropriate single time-scale procedure that takes into account of the heterogeneity in agent objectives; this approximation and subsequent analysis require new technical tools that we develop in this paper.*

Define the subspace $\mathcal{S}_P \in \mathbb{R}^{N^2}$ by $\mathcal{S}_P = \left\{\mathbf{y} \in \mathbb{R}^{N^2} | \mathbf{y} = \mathcal{P}\mathbf{w}, \text{for some } \mathbf{w} \in \mathbb{R}^{N^2}\right\}$. We now formalize a key assumption relating the interest sets $\mathcal{I}_n$ to the network connectivity and global observability.

Assumption A5. *There exists a constant* $c_1 > 0$ *such that,*

$$\mathbf{y}^\top \left(\frac{\beta_0}{\alpha_0}\overline{\mathbf{L}_\mathcal{P}} + \mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\mathbf{G}_H^\top\mathcal{P}\right)\mathbf{y} \geq c_1 \left\|\mathbf{y}\right\|^2, \forall\ \mathbf{y} \in \mathcal{S}_P. \tag{12}$$

We formalize an assumption on the innovation gain sequence $\{\alpha_t\}$ before proceeding further.

Assumption A6. *Let* $\lambda_{min}(\cdot)$ *denote the smallest eigenvalue. We require that* a *satisfies,*

$$a \min\{\lambda_{min}\left(\sum_{n=1}^{N} \mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}\right), c_1, \beta_0^{-1}\} \geq 1,$$

where $\otimes$ *denotes the Kronecker product and* c_1 *is defined in* (12)*.*

It is to be noted that in Assumption A5, if $\mathcal{P} = \mathbf{I}_{N^2}$, then the subspace $\mathcal{S}_P$ reduces to $\mathbb{R}^{N^2}$ and the condition in (12) reduces to a commonly employed Lyapunov condition in classical consensus+innovations type inference procedures (see, for example, Lemma 6 in [1]) which, in turn, can be enforced by global observability and the mean connectivity of the network under consideration. However, in the case when $\mathcal{P} \neq \mathbf{I}_{N^2}$, the case considered

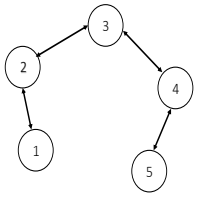

Fig. 1: A network example emphasizing the notion of structural observability.

in this paper, global observability and connectivity of the network is not sufficient to obtain the condition in (12). The insufficiency of global observability and connectivity of the network in order to enforce (12) can be attributed to heterogeneous objectives of the agents and censoring of messages at agents leading to an inherent information loss. Intuitively, such a condition calls for existence of information pathways between agents who share a particular component in their interest sets and the particular component in question to be observable at this set of agents collectively. As we show in the following (Lemma 3.3), a sufficient condition for Assumption A5 is that in addition to the global observability and the mean network connectedness, the induced subgraph for every entry of the vector $\boldsymbol{\theta}^*$ needs to be connected. The induced subgraph for the r-th entry is the set of agents and their associated links which have the r-th entry of $\boldsymbol{\theta}^*$ in their interest sets.

In the following, we will establish consistency of the $\mathcal{CIRFE}$ under Assumption A5. We now show by a simple example that, in general, Assumption A5 is stronger than mean connectivity and global observability. To this end, consider the simple network consisting of 5 nodes in Fig. 1.

Each node n corresponds to a physical component θ_n^*. Thus, $\boldsymbol{\theta}^* \in \mathbb{R}^5$. The solid lines connecting the nodes correspond to the inter-node communication pattern. Each node observes a noisy scalar functional. In particular, we assume

$$y_3(i) = \frac{1}{3}\left(\theta_2^* + \theta_3^* + \theta_4^*\right) + \gamma_3(i), \tag{13}$$

$$y_n(i) = \theta_n^* + \gamma_n(i), n = 1, 2, 4, 5. \tag{14}$$

Clearly, in this case, $G = \sum_{n=1}^{5} \mathbf{H}_n^\top \mathbf{H}_n$ is invertible and, as shown, the communication network is connected. Also, $\widetilde{\mathcal{I}_n} = \{n\}$ for $n = 1, 2, 4, 5$ and $\widetilde{\mathcal{I}}_3 = \{2, 3, 4\}$.

In case, every node wants to estimate the entire $\boldsymbol{\theta}^*$, then the above inference task reduces to the inference setup considered in [1], [29].

Consider the case, $\mathcal{I}_n = \widetilde{\mathcal{I}}_n$ for $n = 1, 2, 3, 4$, i.e., these nodes are interested in reconstructing only their own states and those who influence their observations. However, let $\mathcal{I}_5 = \{5, 1\}$, i.e., node 5 is interested in the state of node 1. This problem falls under the purview of $\mathcal{CIRFE}$. Clearly, Assumption A4 is satisfied. However, it can be shown by calculating the various terms, that assumption A5 is not satisfied and hence, convergence of $\mathcal{CIRFE}$ to desired values is not guaranteed. This shows that mean connectivity and global observability is not sufficient for assumption A5 in general. We provide an intuitive explanation, why the $\mathcal{CIRFE}$ is not expected to yield accurate estimates in this case and why the Lyapunov type requirement in assumption A5 is sufficient for $\mathcal{CIRFE}$'s desired

convergence.. Looking at Fig. 1, we note that the only node that observes (at least partially) the component θ_1^* is node 1, i.e., the influence of the state θ_1^* only affects the observations at node 1. Clearly, for node 5 to be able to reconstruct θ_1^*, it should be able to access information about θ_1^* from the allowed communication graph. Moreover, there is a path connecting node 1 to node 5. However, the other nodes in the path are not interested in reconstructing θ_1^*, so they do not participate in the exchange of information regarding θ_1^*. For example, node 2 ignores the estimate of θ_1^* at node 1 and similarly the others. As a result, the information about θ_1^* never reaches node 5, although the communication network is connected. Note that the induced subgraph of component 1 of $\boldsymbol{\theta}^*$ is disconnected, and it involves only nodes 1 and 5 and no links.

At the same time, it is easy to see that this problem is resolved if an extra communication link is added between nodes 1 and 5. Thus, we see that connectivity of the subgraph formed by those nodes interested in reconstructing θ_1 seems to facilitate proper information flow necessary for the desired convergence of $\mathcal{CIRFE}$. Based on this intuition, we formulate a general structural connectivity condition (see [24]) that guarantees the satisfaction of A5 which, in turn, will be used subsequently to derive the convergence of $\mathcal{CIRFE}$. We direct the reader to Lemma 3.4.1 in [24] for a proof.

Lemma 3.3 (Lemma 3.4.1 in [24])**.** *Let assumption A4 be satisfied and the global observability condition hold. For each component r of θ^*, define the subset $\mathcal{I}^r \subset [1, \cdots, N]$ by*

$$\mathcal{I}^r = \{n \in [1, \cdots, N] \mid r \in \mathcal{I}_n\} \tag{15}$$

Let $\overline{\mathcal{G}}$ denote the network graph corresponding to the mean Laplacian $\overline{\mathbf{L}}$, i.e., there is an edge between nodes n and l in $\overline{G}$ iff *the (n,l)-th entry in $\overline{\mathbf{L}}$ is non-zero. For each $1 \leq r \leq N$, denote the induced subgraph $\overline{\mathcal{G}}_r$ of $\overline{\mathcal{G}}$ with node set $\mathcal{I}^r$. Then, condition A2 is satisfied if $\overline{\mathcal{G}}_r$ is connected for all r.*

4. $\mathcal{CIRFE}$: MAIN RESULTS

In this section we formally state the main results concerning the distributed parameter estimation $\mathcal{CIRFE}$ algorithm. The proofs are relegated to Section 6The first result concerns with the consistency of the parameter estimate sequence at each agent n.

Theorem 4.1. *Consider the parameter estimate sequence $\{\widetilde{\mathbf{x}}(t)\}$ generated by the $\mathcal{CIRFE}$ algorithm according to* (11)*. Let Assumptions A1-A6 hold. Then, we have,*

$$\mathbb{P}_{\boldsymbol{\theta}^*}\left(\lim_{t\to\infty} \widetilde{\mathbf{x}}(t) = \mathcal{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right) = 1. \tag{16}$$

At this point, we note that the estimate sequence generated by $\mathcal{CIRFE}$ at any agent n is strongly consistent, i.e., $\mathbf{x}_n(t) \to \boldsymbol{\theta}^*_{\mathcal{I}_n}$ almost surely (a.s.) as $t \to \infty$. It is also to be noted that, owing to the heterogeneous objectives of the agents, the consensus in terms of the estimates sequences across any pair of agents is only limited to the common components of the parameter in their interest sets.

With the above development in place, we state a result which allows us to benchmark the asymptotic efficiency of the proposed algorithm.

Theorem 4.2. *Let the hypotheses of Theorem 4.1 hold. Then, the time-scaled sequence* $\sqrt{t+1}\left(\widetilde{\mathbf{x}}(t)-\mathcal{P}\left(\mathbf{1}_N\otimes\boldsymbol{\theta}^*\right)\right)$ *is asymptotically normal, i.e.,*

$$\sqrt{t+1}\left(\widetilde{\mathbf{x}}(t)-\mathcal{P}\left(\mathbf{1}_N\otimes\boldsymbol{\theta}^*\right)\right)\stackrel{\mathcal{D}}{\Longrightarrow}\mathcal{N}\left(\mathbf{0},\mathbf{S}_R\right), \tag{17}$$

where

$$\begin{aligned}\mathbf{S}_R&=\mathbf{P}\mathbf{M}\mathbf{P}^\top\\ [\mathbf{M}]_{ij}&=\left[\mathbf{P}\mathbf{Q}\left(\sum_{n=1}^N\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}\right)\mathbf{Q}\mathbf{P}\right]_{ij}\\ &\times\left([\boldsymbol{\Lambda}]_{ii}+[\boldsymbol{\Lambda}]_{jj}-1\right)^{-1},\end{aligned} \tag{18}$$

and $\mathbf{P}$ *and* $\boldsymbol{\Lambda}$ *are orthonormal and diagonal matrices such that* $\mathbf{P}^\top\mathbf{Q}\left(\sum_{n=1}^N\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}\right)\mathbf{P}=\boldsymbol{\Lambda}$*, in which,* $\mathbf{Q}=$ *diag* $\left[\frac{1}{Q_1},\frac{1}{Q_2},\cdots,\frac{1}{Q_N}\right]$*, with* Q_i *denoting the number of agents interested in the* i*-th entry of* $\boldsymbol{\theta}^*$*.*

Theorem 4.2 establishes the asymptotic normality of the time-scaled (auxilliary) estimate sequence. Noting that the estimate sequence $\{\mathbf{x}_n(t)\}$ is a linear transformation of the auxiliary estimate sequence, we conclude that $\sqrt{t+1}\left(\mathbf{x}_n(t)-\boldsymbol{\theta}^*_{\mathcal{I}_n}\right)$ is also asymptotically normal. It is also to be noted that, when the interest sets of each agent is the identity matrix, i.e., every agent is interested to reconstruct the entire parameter, the matrix $\mathbf{Q}$ reduces to $\frac{\mathbf{I}}{N}$ and the asymptotic covariance reduces to that of the classical consensus+innovations linear parameter estimation case (see [1] and the corresponding update in (2)). In this sense, the classical linear parameter estimation case is a special case of the problem being addressed here. It is to be noted that the case in which $\mathbf{Q}$ reduces to $\frac{1}{\widetilde{Q}}\mathbf{I}$ for some $\widetilde{Q}<N$ ($\widetilde{Q}<N$ agents interested in each entry of $\boldsymbol{\theta}^*$), the asymptotic variance reduces to,

$$\mathbf{S}_R=\frac{a\mathbf{I}}{2\widetilde{Q}}+\frac{\left(\frac{1}{N}\sum_{n=1}^N\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n+\frac{\mathbf{I}}{2a}\right)^{-1}}{\widetilde{Q}}.$$

The asymptotic covariance as derived in Theorem 4.2 explicitly showcases the heterogeneity in the scaling with respect to different components of the parameter through $\mathbf{Q}$, as different components have different cardinalities of interest sets.

5. Simulation Results

In this section, we demonstrate the efficiency of the proposed algorithm $\mathcal{CIRFE}$ through simulation experiments on a synthetic dataset. In particular, we construct a 10 node ring network, where every agent has exactly two nodes in its communication neighborhood. We number the nodes from 1 to 10. The neighbors for the i-th node in the communication graph are the nodes $(i-1)mod\ 10$ and $(i+1)mod\ 10$.

The physical coupling which affects each agent's observations is assumed to be an agent's 2-hop neighborhood. For instance, node 1's observations are affected by the value of the field at nodes 9, 10, 2 and 3. Thus, $\widetilde{\mathcal{I}}_1=\{9,10,2,3\}$. The interest set of each agent is taken to be all the field values which affects its observation. For instance, $\mathcal{I}_1=\{9,10,1,2,3\}$. We resort to a static Laplacian in the simulation setup here. We also note that in this case the inter-agent communication network is sparser than the physical network induced by measurement coupling. Each agent makes a scalar observation at each time. Hence, the observation matrix for each agent is given by a 5-sparse 10-dimensional row vector. To be specific, the observation matrices used in the simulation setup are given by $\mathbf{H}_1=[1.0,1.2,1.3,0,0,0,0,0,1.4,1.5]$, $\mathbf{H}_2=[1.5,1.0,1.2,1.3,0,0,0,0,0,1.4]$, $\mathbf{H}_3=$

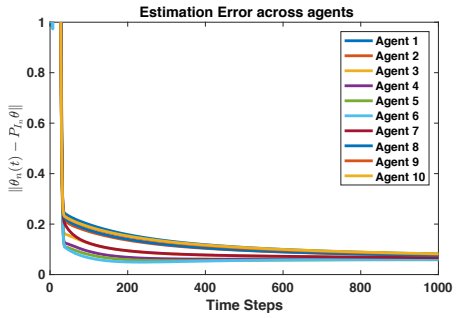

Fig. 2: Convergence of normalized estimation error at each agent

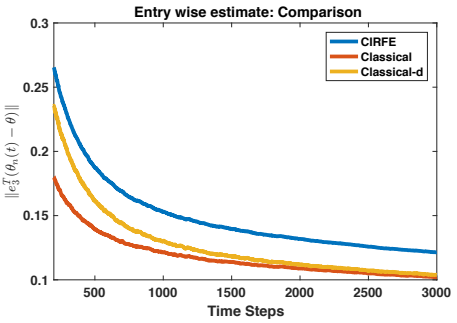

Fig. 3: Comparison of $e_3^\top \boldsymbol{\theta}$ estimation error

$[1.4, 1.5, 1.0, 1.2, 1.3, 0, 0, 0, 0, 0]$, $\mathbf{H}_4 = [0, 1.4, 1.5, 1.0, 1.2, 1.3, 0, 0, 0, 0]$, $\mathbf{H}_5 = [0, 0, 1.4, 1.5, 1.0, 1.2, 1.3, 0, 0, 0]$, $\mathbf{H}_6 = [0, 0, 0, 1.4, 1.5, 1.0, 1.2, 1.3, 0, 0]$, $\mathbf{H}_7 = [0, 0, 0, 0, 1.4, 1.5, 1.0, 1.2, 1.3, 0]$, $\mathbf{H}_8 = [0, 0, 0, 0, 0, 1.4, 1.5, 1.0, 1.2, 1.3]$, $\mathbf{H}_9 = [1.3, 0, 0, 0, 0, 0, 1.4, 1.5, 1.0, 1.2]$ and $\mathbf{H}_{10} = [1.2, 1.3, 0, 0, 0, 0, 0, 1.4, 1.5, 1.0]$. The noise covariance $\mathbf{R}$ is taken to be $\mathbf{I}_{10}$. The parameter capturing the field values is taken to be $\boldsymbol{\theta} = [1.2, 1.3, 1.4, 0.8, 0.7, 1.1, 0.9, 1.0, 1.8, 0.6]$. It can be seen that Assumption A5 is satisfied, by verifying Lemma 3.3 for the third parameter component θ_3^*.

We carry out 500 Monte-Carlo simulations for analyzing the convergence of the parameter estimates. The estimates are initialized as $\mathbf{x}_n(0) = \mathbf{0}$ for $n = 1, \cdots, 10$. The normalized error for the n-th agent at time t is given by the quantity $\|\mathbf{x}_n(t) - \mathcal{P}_{\mathcal{I}_n}\boldsymbol{\theta}\|/5$, as each agent's interest set has the cardinality of 5. Figure 2 shows the normalized error at every agent against the time index t. In Figures 3 and 4 we compare the performance of $\mathcal{CIRFE}$ to the classical distributed estimator in [1] (see (2) for the corresponding update), where each agent is interested in reconstructing the entire state or the parameter vector. We refer to the estimates of the distributed estimator in [1] as "classical" and "classical-d" (to be specified shortly) in the sequel. In Figures 3 and 4, "Classical-d" represents the case in the algorithm in [1], where an agent does not observe the entry to be estimated and entirely depends on the neighborhood communication to estimate the quantity of interest. We specifically study the estimation performance of the agents in the "Classical-d" case, as these are the agents that tend to increase the communication overhead considerably by being interested in estimates of components that they do not directly observe, relying on other agents possibly far off to obtain the desired information. Note that, in the current simulation setup, such class of agents do not exist for

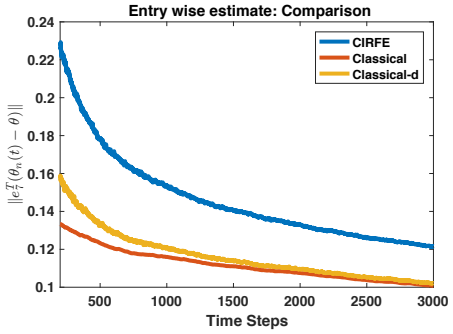

Fig. 4: Comparison of $e_7^\top \boldsymbol{\theta}$ estimation error

the proposed $\mathcal{CIRFE}$ algorithm. It can be observed from figures 3 and 4 that the estimation error in $\mathcal{CIRFE}$ is higher than that of the classical distributed estimator but at the same time exchanging 5-dimensional or even smaller dimensional messages as opposed to 10-dimensional messages in the case of the classical consensus+innovations estimator in [1]. This analysis brings about an inherent trade-off between estimation error and the dimension of the messages exchanged between agents. It is also to be noted that the agents in case of $\mathcal{CIRFE}$ store 5-dimensional vectors at each time step as opposed to 10-dimensional vectors in the case of the classical.

6. Proof of Main Results

Proof of Theorem 4.1. Define the sequence, $\{\widehat{\mathbf{x}}(t)\}$, as $\widehat{\mathbf{x}}(t) = \widetilde{\mathbf{x}}(t) - \mathcal{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)$. Then, we have,

$$\begin{aligned}\widehat{\mathbf{x}}(t+1) &= \widehat{\mathbf{x}}(t) - \left(\beta_t \overline{\mathbf{L}}_{\mathcal{P}} + \alpha_t \mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\mathbf{G}_H^\top\mathcal{P}\right)\widehat{\mathbf{x}}(t)\\ &- \beta_t \widetilde{\mathbf{L}}_{\mathcal{P}}(t)\widehat{\mathbf{x}}(t) + \alpha_t \mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\left(\mathbf{y}(t) - \mathbf{G}_H^\top\mathcal{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right). \end{aligned} \tag{19}$$

It is clear that $\{\widehat{\mathbf{x}}(t)\}$ is Markov with respect to its natural filtration $\{\mathcal{F}_t^{\widehat{\mathbf{X}}}\}$. Now, define the function $V : \mathbb{R}^{N^2} \longmapsto \mathbb{R}_+$ as, $V(\mathbf{y}) = \|\mathbf{y}\|^2$, for all $\mathbf{y}$. We note that

$$\mathbb{E}_{\theta^*}\left[V(\widehat{\mathbf{x}}(t+1)) \mid \mathcal{F}_t^{\widehat{\mathbf{x}}}\right] = \mathbb{E}_{\theta^*}\left[V(\widehat{\mathbf{x}}(t+1)) \mid \widehat{\mathbf{x}}(t)\right] \tag{20}$$

By basic algebraic manipulations, we have,

$$\begin{aligned}&\mathbb{E}_{\theta^*}\left[V(\widehat{\mathbf{x}}(t+1)) \mid \widehat{\mathbf{x}}(t)\right]\\ &\leq \widehat{\mathbf{x}}(t)^\top\left(\mathbf{I} - \beta_t\overline{\mathbf{L}}_{\mathcal{P}} - \alpha_t\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\mathbf{G}_H^\top\mathcal{P}\right)^2\widehat{\mathbf{x}}(t)\\ &+ \beta_t^2\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{L}}_{\mathcal{P}}(t)\widehat{\mathbf{x}}(t)\right\|^2\right]\\ &+ \alpha_t^2\mathbb{E}_{\theta^*}\left[\left\|\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\left(\mathbf{y}(t) - \mathbf{G}_H^\top\mathcal{P}\left(\mathbf{1}_N\otimes\boldsymbol{\theta}^*\right)\right)\right\|^2\right]. \end{aligned} \tag{21}$$

We note that $\beta_t\overline{\mathbf{L}}_{\mathcal{P}} + \alpha_t\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\mathbf{G}_H^\top\mathcal{P}$ is uniformly elliptic on the subspace $\mathcal{S}_{\mathcal{P}}$, and it is precisely the subspace where $\{\widehat{\mathbf{x}}(t)\}$ resides. We thus prove the result by showing convergence to zero of the sequence $\{\widehat{\mathbf{x}}(t)\}$ through the subspace $\mathcal{S}_{\mathcal{P}}$. To this end, using the fact, that, for $\mathbf{y} \in \mathcal{S}_{\mathcal{P}}$,

$$\mathbf{y}^\top\left(\frac{\beta_0}{\alpha_0}\mathbf{L}_{\mathcal{P}} + \mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\mathbf{G}_H^\top\mathcal{P}\right)\mathbf{y} \geq c_1\|\mathbf{y}\|^2, \text{ a.s.} \tag{22}$$

By choosing, t_1 sufficiently large, we have for $\widehat{\mathbf{x}}(t)^\top \in \mathcal{S}_{\mathcal{P}}$ for all $t \geq t_1$,

$$\begin{aligned}
&\widehat{\mathbf{x}}(t)^\top \left(\beta_t^2 \overline{\mathbf{L}}_{\mathcal{P}}^2 + \beta_t^2 \mathbb{E}_{\theta^*} \left\| \widetilde{\mathbf{L}}_{\mathcal{P}}(t) \right\|^2 - \beta_t \overline{\mathbf{L}}_{\mathcal{P}} \right) \widehat{\mathbf{x}}(t) \\
&\leq \left(c_1^{'} \beta_t^2 - c_3^{'} \beta_t \right) \|\widehat{\mathbf{x}}(t)\|^2 \leq 0,
\end{aligned} \tag{23}$$

where equality exists if $\widehat{\mathbf{x}}(t) = \mathcal{P}\left(\mathbf{1}_N \otimes \mathbf{a}\right)$, where $\mathbf{a} \in \mathbb{R}^N$. Thus, we obtain the following inequality:

$$\begin{aligned}
&\mathbb{E}_{\theta^*}\left[V(\widehat{\mathbf{x}}(t+1)) \mid \widehat{\mathbf{x}}(t) = \mathbf{y}\right] - V(\mathbf{y}) \leq c_{11}\alpha_t^2 \left(1 + \|\mathbf{y}\|^2\right) \\
&- \alpha_t c_{10} \|\mathbf{y}\|^2
\end{aligned} \tag{24}$$

for all $\mathbf{y} \in \mathcal{S}_{\mathcal{P}}$. Now, define the function $W : \mathbb{T}_+ \times \mathbb{R}^{N^2} \longmapsto \mathbb{R}_+$:

$$W(t, \mathbf{y}) = (1 + V(\mathbf{y})) \prod_{j=t}^{\infty} (1 + c_{11}\alpha_j^2). \tag{25}$$

From (24) it can be shown that, for $\mathbf{y} \in \mathcal{S}_{\mathcal{P}}$,

$$\begin{aligned}
&\mathbb{E}_{\theta^*}\left[W(t+1, \widehat{\mathbf{x}}(t+1)) \mid \widehat{\mathbf{x}}(t) = \mathbf{y}\right] - W(t, \mathbf{y}) \\
&\leq -\alpha_t c_{10} \|\mathbf{y}\|^2 \left(\prod_{j=t+1}^{\infty} (1 + c_{11}\alpha_j^2) \right) \\
&\leq -\alpha_t c_{10} \|\mathbf{y}\|^2
\end{aligned} \tag{26}$$

Now consider $\varepsilon > 0$, and let V_ε denote the set

$$V_\varepsilon = \{\mathbf{y} \in \mathbb{R}^{N^2} \mid \|\mathbf{y}\| \geq \varepsilon\} \cap \mathcal{S}_{\mathcal{P}} \tag{27}$$

Also, define τ_ε to be the exit time of the process $\{\widehat{\mathbf{x}}(t)\}$ from V_ε, i.e.,

$$\tau_\varepsilon = \inf\{i \in \mathbb{T}_+ \mid \widehat{\mathbf{x}}(t) \notin V_\varepsilon\} \tag{28}$$

We now show that $\tau_\varepsilon < \infty$ a.s. For mathematical simplicity, assume $\widehat{\mathbf{x}}(0) \in V_\varepsilon$. Consider the function

$$\widetilde{W}(t, \mathbf{y}) = W(t, \mathbf{y}) + c_{10}\varepsilon^2 \sum_{j=0}^{t-1} \alpha_j \tag{29}$$

By (26) it follows that, for $\mathbf{y} \in V_\varepsilon$,

$$\mathbb{E}_{\theta^*}\left[W(t+1, \widehat{\mathbf{x}}(t+1)) \mid \widehat{\mathbf{x}}(t) = \mathbf{y}\right] - W(t, \mathbf{y}) \leq -\alpha_t c_{10}\varepsilon^2 \tag{30}$$

and hence, it can be shown that, for $\mathbf{y} \in V_\varepsilon$,

$$\mathbb{E}_{\theta^*}\left[\widetilde{W}(t+1, \widehat{\mathbf{x}}(t+1)) \mid \widehat{\mathbf{x}}(t) = \mathbf{y}\right] - \widetilde{W}(t, \mathbf{y}) \leq 0 \tag{31}$$

Hence, we have that the stopped process $\{\widetilde{W}(\max\{t, \tau_\varepsilon\}, \widehat{\mathbf{x}}(\max\{t, \tau_\varepsilon\}))\}$ is a super martingale. Being nonnegative it converges a.s. as $t \to \infty$. By (30), we then conclude that the following term converges,

$$\lim_{t \to \infty} c_{10}\varepsilon^2 \sum_{j=0}^{(t \wedge \tau_\varepsilon)-1} \alpha_j \text{ converges a.s.} \tag{32}$$

Since, $\sum_{t \in \mathbb{T}_+} \alpha_t = \infty$, the above is possible, only if, $\tau_\varepsilon < \infty$ a.s.
We thus note, that the process $\{\widehat{\mathbf{x}}(t)\}$ leaves the set V_ε almost surely in finite time. Since, the process is constrained to lie in $\mathcal{S}_{\mathcal{P}}$ at all times, the finite time exit from V_ε suggests,

$$\mathbb{P}_{\theta^*}\left(\inf\{t \in \mathbb{T}_+ \mid \|\widehat{\mathbf{x}}(t)\| < \varepsilon\} < \infty\right) = 1 \tag{33}$$

Since $\varepsilon > 0$ is arbitrary, a subsequence almost surely converges to zero, and we have

$$\mathbb{P}_{\theta^*}\left(\liminf_{t\to\infty} \|\widehat{\mathbf{x}}(t)\| = 0\right) = 1 \tag{34}$$

Now going back to (24) and noting that $\{\widehat{\mathbf{x}}(t)\}$ takes values in $\mathcal{S}_{\mathcal{P}}$, we conclude that the process $\{V(\widehat{\mathbf{x}}(t))\}$ is a nonnegative supermartingale. Hence,

$$\mathbb{P}_{\theta^*}\left(\lim_{t\to\infty} V(\widehat{\mathbf{x}}(t)) \text{ exists}\right) = 1 \tag{35}$$

Also, by (34)

$$\mathbb{P}_{\theta^*}\left(\liminf_{t\to\infty} V(\widehat{\mathbf{x}}(t)) = 0\right) = 1 \tag{36}$$

and we conclude that

$$\mathbb{P}_{\theta^*}\left(\lim_{i\to\infty} \|\widehat{\mathbf{x}}(t)\| = 0\right) = 1 \tag{37}$$

□

Proof of Theorem 4.2. Let the number of agents interested in the i-th entry of $\boldsymbol{\theta}^*$ be Q_i. To get the vector of estimates of the i-th entry of $\boldsymbol{\theta}^*$, left multiply the selector matrix $\mathcal{S}_i \in \mathbb{R}^{Q_i \times N^2}$ and noting that $S_i \mathbf{L}_{\mathcal{P}}(t)\widetilde{\mathbf{x}}(t) = \mathbf{L}_{\mathcal{P},i}(t)\widetilde{\mathbf{x}}(i,t)$, where $\mathbf{L}_{\mathcal{P},i}(t) \in \mathbb{R}^{Q_i \times Q_i}$ is the subgraph induced by the interest sets for the i-th entry of $\boldsymbol{\theta}^*$, which is connected as a result of a sufficient condition which enforced Assumption A5 and $\widetilde{\mathbf{x}}(i,t) \in \mathbb{R}^{Q_i}$ is the vector of estimates for the i-th entry of $\boldsymbol{\theta}^*$.

A vector $\mathbf{z} \in \mathbb{R}^{N^2}$ may be decomposed as $\mathbf{z} = \mathbf{z}_{\mathcal{C}} + \mathbf{z}_{\mathcal{C}^\perp}$ with $\mathbf{z}_{\mathcal{C}}$ denoting its projection on the consensus or agreement subspace $\mathcal{C}$, $\mathcal{C} = \left\{\mathbf{z} \in \mathbb{R}^{N^2} | \mathbf{z} = \mathbf{1}_N \otimes \mathbf{a} \textit{ for some } \mathbf{a} \in \mathbb{R}^N\right\}$. We first prove the following Lemma regarding the mean connectedness of the subgraphs $\mathbf{L}_{\mathcal{P},i}(t)$.

Lemma 6.1. *Let $\{\mathbf{z}_t\}$ be an $\mathbb{R}^{N^2}$ valued $\mathcal{F}_t$-adapted process such that $\mathbf{z}_t \in \mathcal{C}^\perp$ for all t. Also, let $\{\mathbf{L}_t\}$ be an i.i.d. sequence of Laplacian matrices as in assumption A4 that satisfies*

$$\lambda_2\left(\overline{\mathbf{L}}\right) = \lambda_2\left(\mathbb{E}\left[\mathbf{L}_t\right]\right) > 0, \tag{38}$$

where $\mathbf{L}_t$ is $\mathcal{F}_{t+1}$-adapted and independent of $\mathcal{F}_t$ for all t.

$$\left\|\left(\mathbf{I}_{N^2} - \left(\mathbf{L}(t) \otimes \mathbf{I}_N\right)\right)\mathbf{z}_t\right\| \leq (1 - r_t)\left\|\mathbf{z}_t\right\|, \tag{39}$$

where $\{r_t\}$ is a $\mathbb{R}^+$ valued $\mathcal{F}_{t+1}$ process satisfying

$$\mathbb{E}\left[r_t | \mathcal{F}_t\right] \geq \underline{p}\beta_t \frac{\lambda_2\left(\overline{\mathbf{L}}\right)}{4|\mathcal{L}|}, \tag{40}$$

where $\mathcal{L}$ denotes the set of all possible Laplacians.

The following Lemmas will be used to quantify the rate of convergence of distributed vector or matrix valued recursions to their network-averaged behavior.

Lemma 6.2. *Let $\{z_t\}$ be an $\mathbb{R}^+$ valued $\mathcal{F}_t$-adapted process that satisfies*

$$z_{t+1} \leq \left(1 - r_1(t)\right) z_t + r_2(t) U_t (1 + J_t),$$

where $\{r_1(t)\}$ is an $\mathcal{F}_{t+1}$-adapted process, such that for all t, $r_1(t)$ satisfies $0 \leq r_1(t) \leq 1$ and

$$a_1 \leq \mathbb{E}\left[r_1(t) | \mathcal{F}_t\right] \leq \frac{1}{(t+1)^{\delta_1}}$$

with $a_1 > 0$ and $0 \leq \delta_1 < 1$. The sequence $\{r_2(t)\}$ is deterministic and $\mathbb{R}^+$ valued and satisfies $r_2(t) \leq \frac{a_2}{(t+1)^{\delta_2}}$ with $a_2 > 0$ and $\delta_2 > 0$. Further, let $\{U_t\}$ and $\{J_t\}$ be $\mathbb{R}^+$ valued $\mathcal{F}_t$ and $\mathcal{F}_{t+1}$ adapted processes, respectively, with $\sup_{t\geq 0} \|U_t\| < \infty$ a.s. The process $\{J_t\}$ is i.i.d. with J_t independent of $\mathcal{F}_t$ for each t and satisfies the moment condition $\mathbb{E}\left[\|J_t\|^{2+\epsilon_1}\right] < \kappa < \infty$ for some $\epsilon_1 > 0$ and a constant $\kappa > 0$. Then, for every δ_0 such that $0 \leq \delta_0 < \delta_2 - \delta_1 - \frac{1}{2+\epsilon_1}$, we have $(t+1)^{\delta_0} z_t \to 0$ a.s. as $t \to \infty$.

Lemma 6.3 (Lemma 4.1 in [28])**.** *Consider the scalar time-varying linear system*

$$u(t+1) \leq (1 - r_1(t))u(t) + r_2(t), \tag{41}$$

where $\{r_1(t)\}$ is a sequence, such that

$$\frac{a_1}{(t+1)^{\delta_1}} \leq r_1(t) \leq 1 \tag{42}$$

with $a_1 > 0, 0 \leq \delta_1 \leq 1$, whereas the sequence $\{r_2(t)\}$ is given by

$$r_2(t) \leq \frac{a_2}{(t+1)^{\delta_2}} \tag{43}$$

with $a_2 > 0, \delta_2 \geq 0$. Then, if $u(0) \geq 0$ and $\delta_1 < \delta_2$, we have

$$\lim_{t\to\infty} (t+1)^{\delta_0} u(t) = 0, \tag{44}$$

for all $0 \leq \delta_0 < \delta_2 - \delta_1$. Also, if $\delta_1 = \delta_2$, then the sequence $\{u(t)\}$ stays bounded, i.e. $\sup_{t\geq 0} \|u(t)\| < \infty$.

Proof of Lemma 6.1. Let $\mathcal{L}$ denote the set of possible Laplacian matrices which is necessarily finite. Since the set of Laplacians is finite, we have,

$$\underline{p} = \inf_{\mathbf{L}\in\mathcal{L}} p_{\mathbf{L}} > 0, \tag{45}$$

with $p_L = \mathbb{P}(\mathbf{L}(t) = \mathbf{L})$ for each $\mathbf{L} \in \mathcal{L}$ such that $\sum_{\mathbf{L}\in\mathcal{L}} p_{\mathbf{L}} = 1$. We also have that $\lambda_2\left(\overline{\mathbf{L}}\right) > 0$ implies that for every $\mathbf{z} \in \mathcal{C}^{\perp}$, where,

$$\mathcal{C} = \left\{\mathbf{x} | \mathbf{x} = \mathbf{1}_N \otimes \mathbf{a}, \mathbf{a} \in \mathbb{R}^N\right\}, \tag{46}$$

we have,

$$\sum_{\mathbf{L}\in\mathcal{L}} \mathbf{z}^\top \mathbf{L}\mathbf{z} \geq \sum_{\mathbf{L}\in\mathcal{L}} \mathbf{z}^\top p_{\mathbf{L}} \mathbf{L}\mathbf{z} = \mathbf{z}^\top \overline{\mathbf{L}}\mathbf{z} \geq \lambda_2\left(\overline{\mathbf{L}}\right) \|\mathbf{z}\|^2. \tag{47}$$

Owing to the finite cardinality of $\mathcal{L}$ and (47), we also have that for each $\mathbf{z} \in \mathcal{C}^{\perp}$,$\exists \mathbf{L}_{\mathbf{z}} \in \mathcal{L}$ such that,

$$\mathbf{z}^\top \mathbf{L}_{\mathbf{z}} \mathbf{z} \geq \frac{\lambda_2\left(\overline{\mathbf{L}}\right)}{|\mathcal{L}_t|} \|\mathbf{z}\|^2 \tag{48}$$

Moreover, since $\mathcal{L}$ is finite, the mapping $L_{\mathbf{z}} : \mathcal{C}^{\perp} \mapsto \mathcal{L}$ can be realized as a measurable function. For each, $\mathbf{L} \in \mathcal{L}$, the eigen values of $\mathbf{I}_{N^2} - \beta_t(\mathbf{L} \otimes \mathbf{I}_N)$ are given by N repetitions of 1 and $1 - \beta_t \lambda_n(\mathbf{L})$, where $2 \leq n \leq N$. Thus, for $t \geq t_0$, $\|\mathbf{I}_{N^2} - \beta_t(\mathbf{L} \otimes \mathbf{I}_N)\| \leq 1$ and $\|(\mathbf{I}_{N^2} - \beta_t(\mathbf{L} \otimes \mathbf{I}_N))\mathbf{z}\| \leq \|\mathbf{z}\|$. Hence, we can define a jointly measurable function $r_{\mathbf{L},\mathbf{z}}$ given by,

$$r_{\mathbf{L},\mathbf{z}} = \begin{cases} 1 & \text{if } t < t_0 \text{ or } \mathbf{z} = \mathbf{0} \\ 1 - \frac{\|(\mathbf{I}_{NM} - \beta_t(\mathbf{L}\otimes\mathbf{I}_M))\mathbf{z}\|}{\|\mathbf{z}\|} & \text{otherwise}, \end{cases} \tag{49}$$

which satisfies $0 \leq r_{\mathbf{L},\mathbf{z}} \leq 1$ for each $(\mathbf{L}, \mathbf{z})$. Define $\{r_t\}$ to be a $\mathcal{F}_{t+1}$ process given by, $r_t = r_{\mathbf{L},\mathbf{z}_t}$ for each t and $\|(\mathbf{I}_{N^2} - \beta_t (\mathbf{L} \otimes \mathbf{I}_N))\mathbf{z}_t\| = (1 - r_t)\|\mathbf{z}_t\|$ a.s. for each t. Then, we have,

$$\begin{aligned}
&\|(\mathbf{I}_{N^2} - \beta_t (\mathbf{L}_{\mathbf{z}_t} \otimes \mathbf{I}_N))\mathbf{z}_t\|^2 \\
&= \mathbf{z}_t^\top (\mathbf{I}_{N^2} - 2\beta_t (\mathbf{L}_{\mathbf{z}_t} \otimes \mathbf{I}_N))\mathbf{z}_t \\
&+ \mathbf{z}_t^\top \beta_t^2 (\mathbf{L}_{\mathbf{z}_t} \otimes \mathbf{I}_N)^2 \mathbf{z}_t \\
&\leq \left(1 - 2\beta_t \frac{\lambda_2(\overline{\mathbf{L}})}{|\mathcal{L}|}\right)\|\mathbf{z}_t\|^2 + c_1\beta_t^2\|\mathbf{z}_t\|^2 \\
&\leq \left(1 - \beta_t \frac{\lambda_2(\overline{\mathbf{L}})}{|\mathcal{L}|}\right)\|\mathbf{z}_t\|^2
\end{aligned} \tag{50}$$

where we have used the boundedness of the Laplacian matrix. With the above development in place, choosing an appropriate t_1 (making t_0 larger if necessary), for all $t \geq t_1$, we have,

$$\|(\mathbf{I}_{N^2} - \beta_t (\mathbf{L}_{\mathbf{z}_t} \otimes \mathbf{I}_N))\mathbf{z}_t\| \leq \left(1 - \beta_t \frac{\lambda_2(\overline{\mathbf{L}})}{4|\mathcal{L}|}\right)\|\mathbf{z}_t\|^2. \tag{51}$$

Then, from (51), we have,

$$\begin{aligned}
&\mathbb{E}\left[\|(\mathbf{I}_{N^2} - \beta_t (\mathbf{L}_{\mathbf{z}_t} \otimes \mathbf{I}_N))\mathbf{z}_t\| \,|\, \mathcal{F}_t\right] \\
&= \sum_{\mathbf{L} \in \mathcal{L}} p_{\mathbf{L}} (1 - r_{\mathbf{L},\mathbf{z}_t})\|\mathbf{z}_t\| \\
&\leq \left(1 - \left(\underline{p}\beta_t \frac{\lambda_2(\overline{\mathbf{L}})}{4|\mathcal{L}|} + \sum_{\mathbf{L} \neq \mathbf{L}_{\mathbf{z}_t}}\right)\right)\|\mathbf{z}_t\|.
\end{aligned} \tag{52}$$

Since, $\sum_{\mathbf{L} \neq \mathbf{L}_{\mathbf{z}_t}} p_{\mathbf{L}} r_{\mathbf{L},\mathbf{z}_t} \geq 0$, we have for all $t \geq t_1$,

$$\begin{aligned}
&(1 - \mathbb{E}[r_t|\mathcal{F}_t])\|\mathbf{z}_t\| \\
&= \mathbb{E}\left[\|(\mathbf{I}_{N^2} - \beta_t (\mathbf{L}_{\mathbf{z}_t} \otimes \mathbf{I}_N))\mathbf{z}_t\| \,|\, \mathcal{F}_t\right] \\
&\leq \left(1 - \underline{p}\beta_t \frac{\lambda_2(\overline{\mathbf{L}})}{4|\mathcal{L}|}\right)\|\mathbf{z}_t\|.
\end{aligned} \tag{53}$$

As $r_t = 1$ on the set $\{\mathbf{z}_t = 0\}$, we have that,

$$\mathbb{E}[r_t|\mathcal{F}_t] \geq \underline{p}\beta_t \frac{\lambda_2(\overline{\mathbf{L}})}{4|\mathcal{L}|}. \tag{54}$$

Thus, we have established that,

$$\|(\mathbf{I}_{N^2} - (\mathbf{L}(t) \otimes \mathbf{I}_N))\mathbf{z}_t\| \leq (1 - r_t)\|\mathbf{z}_t\|, \tag{55}$$

where $\{r_t\}$ is a $\mathbb{R}^+$ valued $\mathcal{F}_{t+1}$ process satisfying (54). □

With the above development in place, consider the residual process $\{\mathbf{x}^\dagger(t)\}$ given by $\mathbf{x}^\dagger(i,t) = \widetilde{\mathbf{x}}(i,t) - \mathbf{1}_{Q_i} \otimes \widetilde{\mathbf{x}}_{\text{avg,i}}(t)$, where i denotes the i-th entry of $\boldsymbol{\theta}^*$ and $\mathbf{x}^\dagger(t) = \left[\mathbf{x}^\dagger(1,t), \cdots, \mathbf{x}^\dagger(N,t)\right]^\top$. Thus, we have that the process $\{\mathbf{x}^\dagger(i,t)\}$ satisfies the recursion,

$$\mathbf{x}^\dagger(i, t+1) = (\mathbf{I}_{Q_i} - \mathbf{L}_{\mathcal{P},i}(t))\mathbf{x}^\dagger(i,t) + \alpha_t \widetilde{\mathbf{z}}(i,t), \tag{56}$$

where the process $\{\widetilde{\mathbf{z}}(i,t)\}$ is given by

$$\widetilde{\mathbf{z}}(i,t) = \left(\mathbf{I}_{Q_i} - \frac{1}{Q_i}\mathbf{1}_{Q_i}\mathbf{1}_{Q_i}^\top\right) \times \mathcal{S}_i \mathcal{P} \mathbf{G}_H \mathbf{R}^{-1}\left(\mathbf{y}(t) - \mathbf{G}_H^\top \mathcal{P} \widetilde{\mathbf{x}}(t)\right). \tag{57}$$

From (57), we also have,

$$\widetilde{\mathbf{z}}(i,t) = \overline{\mathbf{J}}_{i,t} + \overline{\mathbf{U}}_{i,t}, \tag{58}$$

where,

$$\begin{aligned}
\overline{\mathbf{J}}_{i,t} &= \left(\mathbf{I}_{Q_i} - \frac{1}{Q_i}\mathbf{1}_{Q_i}\mathbf{1}_{Q_i}^\top\right) \\
&\times \mathcal{S}_i\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\left(\mathbf{y}(t) - \mathbf{G}_H^\top\mathcal{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right) \\
\overline{\mathbf{U}}_t &= \left(\mathbf{I}_{Q_i} - \frac{1}{Q_i}\mathbf{1}_{Q_i}\mathbf{1}_{Q_i}^\top\right) \\
&\times \mathcal{S}_i\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\left(\mathbf{G}_H^\top\mathcal{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right) - \mathbf{G}_H^\top\mathcal{P}\widetilde{\mathbf{x}}(t)\right).
\end{aligned} \tag{59}$$

By Theorem 4.1, we also have that, the process $\{\widetilde{\mathbf{x}}(i,t)\}$ is bounded. Hence, there exists an $\mathcal{F}_t$-adapted process $\{\widetilde{U}_{i,t}\}$ such that $\left\|\overline{\mathbf{U}}_{i,t}\right\| \leq \widetilde{U}_{i,t}$ and $\sup_{t\geq 0}\widetilde{U}_{i,t} < \infty$ a.s.. Furthermore, denote the process $U_{i,t}$ as follows,

$$U_{i,t} = \max\left\{\widetilde{U}_{i,t}, \left\|\mathbf{I}_{Q_i} - \frac{1}{Q_i}\mathbf{1}_{Q_i}\mathbf{1}_{Q_i}^\top\right\|\right\}. \tag{60}$$

With the above development in place, we conclude,

$$\left\|\overline{\mathbf{U}}_{i,t}\right\| + \left\|\overline{\mathbf{J}}_{i,t}\right\| \leq U_{i,t}\left(1 + J_{i,t}\right), \tag{61}$$

where $J_{i,t} = \left\|\mathbf{y}(t) - \mathbf{G}_H^\top\mathcal{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right\|$ and $\mathbb{E}_{\boldsymbol{\theta}}\left[J_{i,t}^{2+\epsilon}\right] < \infty$. Then, from (39)-(56) we have,

$$\left\|\mathbf{x}^\dagger(i,t+1)\right\| \leq (1-r_t)\left\|\mathbf{x}^\dagger(i,t)\right\| + \alpha_t U_{i,t}(1+J_{i,t}), \tag{62}$$

which then falls under the purview of Lemma 6.2 and hence we have the assertion,

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\delta_0}\left(\widetilde{\mathbf{x}}(i,t) - \mathbf{1}_{Q_i} \otimes \widetilde{\mathbf{x}}_{\text{avg,i}}(t)\right) = 0\right) = 1, \tag{63}$$

where $0 < \delta_0 < 1 - \tau_1$ and hence δ_0 can be chosen to be $1/2 + \delta$, where $\delta > 0$ and we finally have,

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\frac{1}{2}+\delta}\left(\widetilde{\mathbf{x}}(t) - \mathbf{1}_N \otimes \widetilde{\mathbf{x}}_{\text{avg}}(t)\right) = 0\right) = 1, \tag{64}$$

as the above analysis can be repeated each entry i of the parameter of interest $\boldsymbol{\theta}^*$.

The proof of Theorem 4.2 needs the following Lemma from [32] concerning the asymptotic normality of the stochastic recursions.

Lemma 6.4 (Theorem 2.2 in [32])**.** *Let $\{\mathbf{z}_t\}$ be an $\mathbb{R}^k$-valued $\{\mathcal{F}_t\}$-adapted process that satisfies*

$$\begin{aligned}
\mathbf{z}_{t+1} &= \left(\mathbf{I}_k - \frac{1}{t+1}\Gamma_t\right)\mathbf{z}_t + (t+1)^{-1}\boldsymbol{\Phi}_t\mathbf{V}_t \\
&+ (t+1)^{-3/2}\mathbf{T}_t,
\end{aligned} \tag{65}$$

where the stochastic processes $\{\mathbf{V}_t\}, \{\mathbf{T}_t\} \in \mathbb{R}^k$ while $\{\boldsymbol{\Gamma}_t\}, \{\boldsymbol{\Phi}_t\} \in \mathbb{R}^{k\times k}$. Moreover, suppose for each t, $\mathbf{V}_{t-1}$ and $\mathbf{T}_t$ are $\mathcal{F}_t$-adapted, whereas the processes $\{\boldsymbol{\Gamma}_t\}$, $\{\boldsymbol{\Phi}_t\}$ are $\{\mathcal{F}_t\}$-adapted.
Also, assume that

$$\boldsymbol{\Gamma}_t \to \boldsymbol{\Gamma}, \boldsymbol{\Phi}_t \to \boldsymbol{\Phi}, \text{ and } \mathbf{T}_t \to 0 \text{ a.s. as } t \to \infty, \tag{66}$$

where $\boldsymbol{\Gamma}$ is a symmetric and positive definite matrix, and admits an eigen decomposition of the form $\mathbf{P}^\top\boldsymbol{\Gamma}\mathbf{P} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix and $\mathbf{P}$ is an orthogonal matrix. Furthermore, let the sequence $\{\mathbf{V}_t\}$ satisfy $\mathbb{E}\left[\mathbf{V}_t|\mathcal{F}_t\right] = 0$ for each t and suppose there exists a positive constant C and a matrix Σ such that $C > \left\|\mathbb{E}\left[\mathbf{V}_t\mathbf{V}_t^\top|\mathcal{F}_t\right] - \Sigma\right\| \to$

0 *a.s. as* $t \to \infty$ *and with* $\sigma_{t,r}^2 = \int_{\|\mathbf{V}_t\|^2 \geq r(t+1)} \|\mathbf{V}_t\|^2 \, d\mathbb{P}$*, let* $\lim_{t\to\infty} \frac{1}{t+1} \sum_{s=0}^{t} \sigma_{s,r}^2 = 0$ *for every* $r > 0$*. Then, we have,*

$$(t+1)^{1/2} \mathbf{z}_t \xRightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \mathbf{P}\mathbf{M}\mathbf{P}^\top\right), \tag{67}$$

where the (i,j)*-th entry of the matrix* $\mathbf{M}$ *is given by*

$$[\mathbf{M}]_{ij} = \left[\mathbf{P}^\top \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}^\top \mathbf{P}\right]_{ij} \left([\mathbf{\Lambda}]_{ii} + [\mathbf{\Lambda}]_{jj} - 1\right)^{-1}. \tag{68}$$

Multiplying the selection matrix, we have,

$$\begin{aligned}
&\widetilde{\mathbf{x}}(i,t+1) = \widetilde{\mathbf{x}}(i,t) - \mathbf{L}_{\mathcal{P},i}(t)\widetilde{\mathbf{x}}(i,t) + \alpha_t \mathcal{S}_i \mathcal{P}\mathbf{G}_H \mathbf{R}^{-1} \\
&\times \left(\mathbf{y}(t) - \mathbf{G}_H^\top \mathcal{P}\widetilde{\mathbf{x}}(t)\right) \\
&\Rightarrow \frac{\mathbf{1}_{Q_i}^\top}{Q_i}\widetilde{\mathbf{x}}(i,t+1) = \frac{\mathbf{1}_{Q_i}^\top}{Q_i}\widetilde{\mathbf{x}}(i,t) - \frac{\mathbf{1}_{Q_i}^\top}{Q_i}\mathbf{L}_{\mathcal{P},i}(t)\widetilde{\mathbf{x}}(i,t) \\
&+ \alpha_t \frac{\mathbf{1}_{Q_i}^\top}{Q_i} \mathcal{S}_i \mathcal{P}\mathbf{G}_H \mathbf{R}^{-1} \left(\mathbf{y}(t) - \mathbf{G}_H^\top \mathcal{P}\widetilde{\mathbf{x}}(t)\right) \\
&\Rightarrow \widetilde{\mathbf{x}}_{\text{avg},i}(t+1) = \widetilde{\mathbf{x}}_{\text{avg},i}(t) + \alpha_t \frac{\mathbf{1}_{Q_i}^\top}{Q_i} \mathcal{S}_i \mathcal{P}\mathbf{G}_H \mathbf{R}^{-1} \\
&\times \left(\mathbf{y}(t) - \mathbf{G}_H^\top \mathcal{P}\widetilde{\mathbf{x}}(t)\right),
\end{aligned} \tag{69}$$

where $\{\widetilde{\mathbf{x}}_{\text{avg},i}(t)\}$ is the averaged estimate sequence for the i-th entry of the parameter $\boldsymbol{\theta}^*$. Stacking, all such averages together we have,

$$\begin{aligned}
&\widetilde{\mathbf{x}}_{\text{avg}}(t+1) = \widetilde{\mathbf{x}}_{\text{avg}}(t) + \alpha_t \mathcal{S}_{\text{avg}} \mathcal{P}\mathbf{G}_H \mathbf{R}^{-1} \left(\mathbf{y}(t) - \mathbf{G}_H^\top \mathcal{P}\widetilde{\mathbf{x}}(t)\right) \\
&\Rightarrow \widetilde{\mathbf{x}}_{\text{avg}}(t+1) - \boldsymbol{\theta}^* = \left(\mathbf{I} - \alpha_t \mathbf{Q} \sum_{n=1}^{N} \mathcal{P}_{\mathcal{I}_n} \mathbf{H}_n^\top \mathbf{R}^{-1} \mathbf{H}_n \mathcal{P}_{\mathcal{I}_n}\right) \\
&\times \left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*\right) \\
&+ \alpha_t \mathcal{S}_{\text{avg}} \mathcal{P}\mathbf{G}_H \mathbf{R}^{-1} \gamma(t) \\
&+ \alpha_t \mathbf{Q} \sum_{n=1}^{N} \mathcal{P}_{\mathcal{I}_n} \mathbf{H}_n^\top \mathbf{R}_n^{-1} \mathbf{H}_n \left(\widetilde{\mathbf{x}}_n(t) - \mathcal{P}_{\mathcal{I}_n} \widetilde{\mathbf{x}}_{\text{avg}}(t)\right),
\end{aligned} \tag{70}$$

where $\mathcal{S}_{\text{avg}} = \left[\frac{\mathbf{1}_{Q_1}^\top}{Q_1}\mathcal{S}_1, \frac{\mathbf{1}_{Q_2}^\top}{Q_2}\mathcal{S}_2, \cdots, \frac{\mathbf{1}_{Q_N}^\top}{Q_N}\mathcal{S}_N\right]$ and $\mathbf{Q} = diag\left[\frac{1}{Q_1}, \frac{1}{Q_2}, \cdots, \frac{1}{Q_N}\right]$. In the above derivation, we make use of the fact that $\mathcal{S}_{\text{avg}} \mathcal{P}\mathbf{G}_H \mathbf{R}^{-1} \mathbf{G}_H^\top \mathcal{P}\mathbf{1}_N \otimes (\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*) = \mathbf{Q} \sum_{n=1}^{N} \mathcal{P}_n \mathbf{H}_n^\top \mathbf{R}^{-1} \mathbf{H}_n \left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*\right)$, which in turn follows from the fact that,

$$\begin{aligned}
&\mathcal{S}_{\text{avg}} = \mathbf{Q}\left[\mathcal{P}_{\mathcal{I}_1} \ \mathcal{P}_{\mathcal{I}_2} \cdots \mathcal{P}_{\mathcal{I}_N}\right] = \left[\mathbf{Q}\mathcal{P}_{\mathcal{I}_1} \ \mathbf{Q}\mathcal{P}_{\mathcal{I}_2} \cdots \mathbf{Q}\mathcal{P}_{\mathcal{I}_N}\right] \\
&\Rightarrow \mathcal{S}_{\text{avg}} \mathcal{P}\mathbf{G}_H \mathbf{R}^{-1} \mathbf{G}_H^\top \mathcal{P}\mathbf{1}_N \otimes (\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*) \\
&= \left[\mathbf{Q}\mathcal{P}_{\mathcal{I}_1} \ \mathbf{Q}\mathcal{P}_{\mathcal{I}_2} \cdots \mathbf{Q}\mathcal{P}_{\mathcal{I}_N}\right] \\
&\times \Big[\mathcal{P}_{\mathcal{I}_1} \mathbf{H}_1^\top \mathbf{R}_1^{-1} \mathbf{H}_1 \left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*\right) \cdots \\
&\mathcal{P}_{\mathcal{I}_N} \mathbf{H}_N^\top \mathbf{R}_N^{-1} \mathbf{H}_N \left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*\right)\Big]^\top \\
&= \mathbf{Q} \sum_{n=1}^{N} \mathcal{P}_{\mathcal{I}_n} \mathbf{H}_n^\top \mathbf{R}^{-1} \mathbf{H}_n \left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*\right).
\end{aligned} \tag{71}$$

Define, the residual sequence, $\{\mathbf{z}_t\}$, where $\mathbf{z}(t) = \widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}^*$, which can be then shown to satisfy the recursion

$$\mathbf{z}_{t+1} = \left(\mathbf{I}_N - \alpha_t \Gamma\right) \mathbf{z}_t + \alpha_t \mathbf{U}_t + \alpha_t \mathbf{J}_t, \tag{72}$$

where

$$\Gamma = \mathbf{Q}\sum_{n=1}^{N}\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}$$
$$\mathbf{U}_t = \mathbf{Q}\sum_{n=1}^{N}\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n\left(\widetilde{\mathbf{x}}_n(t) - \mathcal{P}_{\mathcal{I}_n}\widetilde{\mathbf{x}}_{\text{avg}}(t)\right)$$
$$\mathbf{J}_t = \mathcal{S}_{\text{avg}}\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\gamma(t). \tag{73}$$

We rewrite the recursion for $\{\mathbf{z}_t\}$ as follows:

$$\mathbf{z}_{t+1} = \left(\mathbf{I}_N - \alpha_t\Gamma_t\right)\mathbf{z}_t + (t+1)^{-3/2}\mathbf{T}_t + (t+1)^{-1}\mathbf{\Phi}_t\mathbf{V}_t, \tag{74}$$

where

$$\mathbf{\Gamma}_t = \mathbf{\Gamma} = \mathbf{Q}\sum_{n=1}^{N}\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}, \mathbf{\Phi}_t = a\mathbf{I}$$
$$\mathbf{T}_t = a(t+1)^{1/2}\mathbf{U}_t$$
$$= a\mathbf{Q}\sum_{n=1}^{N}\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n(t+1)^{0.5}\left(\widetilde{\mathbf{x}}_n(t) - \mathcal{P}_{\mathcal{I}_n}\widetilde{\mathbf{x}}_{\text{avg}}(t)\right) \xrightarrow{t\to\infty} 0$$
$$\mathbf{V}_t = \mathbf{J}_t = \mathcal{S}_{\text{avg}}\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\gamma(t),\ \mathbb{E}\left[\mathbf{V}_t|\mathcal{F}_t\right] = 0,$$
$$\mathbb{E}\left[\mathbf{V}_t\mathbf{V}_t^\top|\mathcal{F}_t\right] = \mathcal{S}_{\text{avg}}\mathcal{P}\mathbf{G}_H\mathbf{R}^{-1}\mathbf{G}_H^\top\mathcal{P}\mathcal{S}_{\text{avg}}$$
$$= \mathbf{Q}\left(\sum_{n=1}^{N}\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}\right)\mathbf{Q} \tag{75}$$

Due to the i.i.d nature of the noise process, we have the uniform integrability condition for the process $\{\mathbf{V}_t\}$. Hence, $\{\mathbf{x}_{\text{avg}}(t)\}$ falls under the purview of Lemma 6.4 and we thus conclude that

$$(t+1)^{1/2}\left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}\right) \xRightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{PMP}^\top), \tag{76}$$

in which,

$$[\mathbf{M}]_{ij} = \left[\mathbf{PQ}\left(\sum_{n=1}^{N}\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}\right)\mathbf{QP}\right]_{ij}$$
$$\times\left([\mathbf{\Lambda}]_{ii} + [\mathbf{\Lambda}]_{jj} - 1\right)^{-1}, \tag{77}$$

where $\mathbf{P}$ and $\mathbf{\Lambda}$ are orthonormal and diagonal matrices such that $\mathbf{P}^\top\mathbf{Q}\left(\sum_{n=1}^{N}\mathcal{P}_{\mathcal{I}_n}\mathbf{H}_n^\top\mathbf{R}_n^{-1}\mathbf{H}_n\mathcal{P}_{\mathcal{I}_n}\right)\mathbf{QP} = \mathbf{\Lambda}$. Now from (64), we have that the processes $\{\widetilde{\mathbf{x}}_n(t)\}$ and $\{\widetilde{\mathbf{x}}_{\text{avg}}(t)\}$ are indistinguishable in the $(t+1)^{1/2}$ time scale, which is formalized as follows:

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\lim_{t\to\infty}\left\|\sqrt{t+1}\left(\widetilde{\mathbf{x}}(t) - \boldsymbol{\theta}\right) - \sqrt{t+1}\left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}\right)\right\| = 0\right)$$
$$= \mathbb{P}_{\boldsymbol{\theta}}\left(\lim_{t\to\infty}\left\|\sqrt{t+1}\left(\widetilde{\mathbf{x}}(t) - \widetilde{\mathbf{x}}_{\text{avg}}(t)\right)\right\| = 0\right) = 1. \tag{78}$$

Thus, the difference of the sequences $\left\{\sqrt{t+1}\left(\widetilde{\mathbf{x}}_n(t) - \boldsymbol{\theta}\right)\right\}$ and $\left\{\sqrt{t+1}\left(\widetilde{\mathbf{x}}_{\text{avg}}(t) - \boldsymbol{\theta}\right)\right\}$ converges a.s. to zero as $t\to\infty$ and hence we have,

$$\sqrt{t+1}\left(\widetilde{\mathbf{x}}_n(t) - \boldsymbol{\theta}\right) \xRightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{PMP}^\top). \tag{79}$$

□

7. CONCLUSION

In this paper, we have proposed a *consensus+innovations* type algorithm, $\mathcal{CIRFE}$, for estimating a high-dimensional parameter or field that exhibits a cyber-physical flavor. In the proposed algorithm, every agent updates its estimate of a few components of the high-dimensional parameter vector by simultaneous processing of neighborhood information and local newly sensed information and in which the inter-agent collaboration is restricted to a possibly sparse communication graph. Under rather generic assumptions we establish the consistency of the parameter estimate sequence and characterize the asymptotic variance of the proposed estimator. A natural direction for future research consists of considering models with non-linear observation functions.

REFERENCES

[1] S. Kar and J. M. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, 2011.

[2] ——, "Consensus+ innovations distributed inference over networks: cooperation and sensing in networked systems," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 99–109, 2013.

[3] A. Das and M. Mesbahi, "Distributed linear parameter estimation in sensor networks based on Laplacian dynamics consensus algorithm," in *3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, vol. 2, Reston, VA, USA, 28-28 Sept. 2006, pp. 440–449.

[4] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in Ad Hoc WSNs with noisy links - part I: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, January 2008.

[5] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, July 2008.

[6] S. Stankovic, M. Stankovic, and D. Stipanovic, "Decentralized parameter estimation by consensus based stochastic approximation," in *46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, 12-14 Dec. 2007, pp. 1535–1540.

[7] I. Schizas, G. Mateos, and G. Giannakis, "Stability analysis of the consensus-based distributed LMS algorithm," in *Proceedings of the 33rd International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, Nevada, USA, April 1-4 2008, pp. 3289–3292.

[8] S. Ram, V. Veeravalli, and A. Nedic, "Distributed and recursive parameter estimation in parametrized linear state-space models," *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 488– 492, February 2010.

[9] A. K. Sahu and S. Kar, "Distributed sequential detection for Gaussian shift-in-mean hypothesis testing," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 89–103, 2016.

[10] M. B. Nevelson and R. Z. Khasminskiĭ, *Stochastic approximation and recursive estimation*. American Mathematical Society, 1973, vol. 47.

[11] D. Bertsekas, J. Tsitsiklis, and M. Athans, "Convergence theories of distributed iterative processes: A survey," *Technical Report for Information and Decision Systems, Massachusetts Inst. of Technology, Cambridge, MA*, 1984.

[12] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, January 2007.

[13] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, Jun. 2003.

[14] A. Nedić, A. Olshevsky, and C. A. Uribe, "Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs," *arXiv preprint arXiv:1410.1977*, 2014.

[15] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[16] S. Shahrampour and A. Jadbabaie, "Exponentially fast parameter estimation in networks using distributed dual averaging," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 6196–6201.

[17] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4583–4588, 2009.

[18] G. Mateos and G. B. Giannakis, "Distributed recursive least-squares: Stability and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3740–3754, 2012.

[19] Z. Weng and P. M. Djurić, "Efficient estimation of linear parameters from correlated node measurements over networks," *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1408–1412, 2014.

[20] F. S. Cattivelli and A. H. Sayed, "Diffusion lms strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.

[21] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, 2008.

[22] U. A. Khan and J. M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, p. 49194935, October 2008.

[23] S. Das and J. M. F. Moura, "Distributed Kalman filtering with dynamic observations consensus," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4458–4473, 2015.

[24] S. Kar, "Large scale networked dynamical systems: Distributed inference," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, 2010. [Online]. Available: http://gradworks.umi.com/34/21/3421734.html

[25] J. F. Mota, J. M. Xavier, P. M. Aguiar, and M. Püschel, "Distributed optimization with local domains: Applications in mpc and network flows," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 2004–2009, 2015.

[26] S. A. Alghunaim and A. H. Sayed, "Distributed coupled multi-agent stochastic optimization," *arXiv preprint arXiv:1712.08817*, 2017.

[27] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.

[28] S. Kar, J. M. F. Moura, and H. V. Poor, "Distributed linear parameter estimation: Asymptotically efficient adaptive strategies," *SIAM Journal on Control and Optimization*, vol. 51, no. 3, pp. 2200–2229, 2013.

[29] A. K. Sahu, S. Kar, J. M. Moura, and H. V. Poor, "Distributed constrained recursive nonlinear least-squares estimation: Algorithms and asymptotics," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 426–441, 2016.

[30] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge, UK: Cambridge University Press, 2008.

[31] S. B. Gelfand and S. K. Mitter, "Recursive stochastic algorithms for global optimization in $\mathbb{R}^d$," *SIAM J. Control Optim.*, vol. 29, no. 5, pp. 999–1018, September 1991.

[32] V. Fabian, "On asymptotic normality in stochastic approximation," *The Annals of Mathematical Statistics*, vol. 39, no. 4, pp. 1327–1332, August 1968.