DISTRIBUTED RANDOM FIELDS ESTIMATION: A CONSENSUS+INNOVATIONS APPROACH

Anit Kumar Sahu, Soummya Kar and José M. F. Moura

Department of Electrical and Computer Engineering,
Carnegie Mellon University, Pittsburgh PA 15213
{anits,soummyak}@andrew.cmu.edu, moura@ece.cmu.edu

ABSTRACT

This paper presents a distributed algorithm to estimate a high dimensional parameter in a multi-agent network, but when each agent is only interested in reconstructing a few components of the parameter. The distributed recursive estimator is of the *consensus+innovations* type. 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 estimate by simultaneously processing the latest locally sensed information (*innovations*) and the parameter estimates from other agents (*consensus*) in the local neighborhood conforming to a pre-specified inter-agent communication topology. This paper establishes the consistency and the order-optimal convergence rate at each agent of the estimates of the components of the state an agent is interested to reconstruct and also characterizes the asymptotic variance of the proposed distributed estimator.

***Index Terms*—** Distributed Estimation, Consensus Algorithms, Distributed Inference, Random Fields.

1. INTRODUCTION

Large scale cyber-physical systems (CPS) have gained prominence recently. We are interested in distributed inference of the state of these 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, phase 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 given by the spatial correlation between the components of the parameter. Moreover, corresponding to each field location, there is a low-power inexpensive sensor monitoring the location. However, in order to come up with a provably consistent estimate sequence, each agent exchanges information with its neighborhood which conforms to an 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. 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. Instead, in this paper, we consider an 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. Hence, instead of reconstructing the entire parameter at each agent, the distributed estimator we present is efficient from a computation and communication point of view favorable to only reconstruct a certain subset of the field based on the requirements of each agent. In particular, we present a distributed estimation algorithm of the *consensus + innovations* form ([1, 2]), namely $\mathcal{CIRFE}$, 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*).

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]). 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]). Representative approaches of this latter class are *consensus + innovations* type ([1, 9]) and the diffusion type ([10, 11]) algorithms. Distributed inference algorithms for random fields have been proposed in literature. (see, for example [12, 13]) Reference [12] 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 [12] where the incorporation of new sensed information is followed by multiple rounds of consensus, the proposed algorithm in this paper pertaining to a non-linear observation model simultaneously fuses the neighborhood information and the current observation albeit for a static random field. In contrast with [13], 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. 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 [14] (Chapter 3). In [14], a single time-scale consen-

This work was supported in part by NSF under grant CCF-1513936.

sus+innovations algorithm pertaining to a linear observation model was proposed and the consistency and asymptotic normality[1] of the estimator was established. Recently, other works in literature have looked at similar setups, albeit in the distributed optimization framework [15]. In contrast with [14], we propose a *consensus + innovations* algorithm pertaining to a non-linear observation model, where the *consensus* and *innovations* terms are weighed through different time-varying sequences. In this paper, we not only establish the consistency and asymptotic normality of the parameter estimate sequence but also the order-optimality of the convergence of the parameter estimates to the true underlying physical field.

The rest of the paper is organized as follows. Spectral graph theory and notation are discussed next. The sensing model and the preliminaries are discussed in Section 2. Section 3 presents the proposed distributed estimation algorithm, while Section 4 concerns with the main results of the paper. Finally, Section 5 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; $\mathbf{A}_{ij}$ 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 $\mathbf{I}$. 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 equivalent to the spectral radius for symmetric matrices. The cardinality of a set $\mathcal{S}$ is $|\mathcal{S}|$. 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 [16] for instance).

[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.

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

2. SENSING MODEL AND PRELIMINARIES

Consider N physical agents monitoring a random field over a large physical area. Each agent n is associated with a scalar state $\boldsymbol{\theta}_n^*$, which represents the field intensity parameter at its location. The agents are equipped with sensing capabilities. The underlying field is spatially correlated so that the observation of agent n is correlated to the observations of its neighbors. 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. It can be easily generalized to vector valued states, though at the cost of extra notation. The observation at each agent is of the form:

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

where $\mathbf{f}_n(\cdot) : \mathbb{R}^N \to \mathbb{R}^{M_n}$ is a sparsifying nonlinear function, $\{\mathbf{y}_n(t)\}$ is a $\mathbb{R}^{M_n}$-valued observation sequence for the n-th agent and for each n where $M_n < N$, $\{\gamma_n(t)\}$ is a zero-mean temporally independent and identically distributed (i.i.d.) noise sequence with nonsingular covariance matrix $\mathbf{R}_n$. 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 $\mathbf{f}_n(\cdot)$ is related to the physical coupling in the random field. To be specific, let us define $\widetilde{\mathcal{I}}_n$ as the set of agents whose field locations are spatially correlated with the field location at agent n. Typically, $\widetilde{\mathcal{I}}_n$ is a small subset of the total number of agents N. Technically speaking, the above mentioned physical coupling 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. In addition to the physical coupling, the agents exchange information among themselves which conforms to a pre-specified possibly sparse inter-agent communication graph. The physical coupling represented as a graph in addition to the inter-agent communication graph makes the network under consideration a cyber-physical system. We formalize assumptions on the inter-agent communication graph and smoothness of the sensing functions before proceeding further.

Assumption A1. *The inter-agent communication graph, modeling the information exchange among the agents, is connected, i.e.* $\lambda_2(\mathbf{L}) > 0$*, where* $\mathbf{L}$ *denotes the associated graph Laplacian matrix.*

Also, as the communication graph is constructed separately from the physical coupling, the inter-agent communication structure is possibly different from and not compatible with the physical coupling.

Assumption A2. *The sensing function* $\mathbf{f}_n(.)$ *for each* n *is continuously differentiable and Lipschitz continuous, i.e., for each agent* n*, there exists a constant* $k_n > 0$ *such that* $\|\mathbf{f}_n(\boldsymbol{\theta}) - \mathbf{f}_n(\boldsymbol{\theta}^*)\| \leq k_n \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|$*, for all* $\boldsymbol{\theta}, \boldsymbol{\theta}^* \in \mathbb{R}^N$.

Smoothness conditions on the sensing functions, such as the one imposed by assumption A2 is common in the literature addressing statistical inference algorithms in non-linear settings.

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. 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., $\widetilde{\mathcal{I}}_n \subset \mathcal{I}_n$.*

Furthermore, we assume that the interest set of every agent n is non-empty. The vector of states $\boldsymbol{\theta}^*$ is rendered locally unobservable at all the agents. Hence, the following global observability condition is needed so as to enable each agent to obtain a consistent estimate of its interest set. We formalize the global observability assumption as follows:

Assumption A4. *The sensing model is globally observable, i.e., any pair $\boldsymbol{\theta}, \acute{\boldsymbol{\theta}}$ of possible parameter instances in $\mathbb{R}^N$ satisfies $\sum_{n=1}^{N} \left\| \mathbf{f}_n(\boldsymbol{\theta}) - \mathbf{f}_n(\acute{\boldsymbol{\theta}}) \right\|^2 = 0$ if and only if $\boldsymbol{\theta} = \acute{\boldsymbol{\theta}}$.*

Global observability corresponds to the centralized setting, where an estimator has access to the agents' observations at all times. The global observability assumption corresponds to the fact that if there was a centralized estimator with simultaneous access to all the agents' measurements, this centralized estimator would be able to reasonably estimate the underlying parameter.

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

In this section, we develop the algorithm $\mathcal{CIRFE}$. Before introducing the algorithm we introduce some more notation. We number the nodes in the interest sets of agents in increasing order. Thus, the interest set $\mathcal{I}_n$ at each 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 to reconstruct the states of agents in its interest set, the estimate sequence 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, $\mathbf{x}_{l,n}^r(t) \in \mathbb{R}^{|\mathcal{I}_n|}$ is generated as follows:

$$\mathbf{e}_j^\top \mathbf{x}_{l,n}^r(t) = \begin{cases} \mathbf{e}_{\mathcal{I}_l^{-1}(\mathcal{I}_n(j))}^\top \mathbf{x}_l(t) & \mathcal{I}_n(j) \in \mathcal{I}_l \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Agent n only wants to use estimates of the states from the agents in its neighborhood which is 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. The transformed estimate $\mathbf{x}_{l,n}^s(t) \in \mathbb{R}^{|\mathcal{I}_n|}$ is generated as follows:

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

The agent n also incorporates the latest sensed information while updating the parameter estimate at each sampling epoch. However, the latest sensed information might include other components of $\boldsymbol{\theta}^*$. Hence, an agent only retains the components of interest, i.e., those in $\mathcal{I}_n$. For a given vector $\mathbf{y} \in \mathbb{R}^{|\mathcal{I}_n|}$, let $\mathbf{y}^{\mathcal{P}_n} \in \mathbb{R}^N$ be the vector whose j-th component is given by

$$\mathbf{e}_j^\top \mathbf{y}^{\mathcal{P}_n} = \begin{cases} \mathbf{e}_{\mathcal{I}_n^{-1}(j)}^\top \mathbf{y} & \mathcal{I}_n(j) \in \mathcal{I}_l \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Finally, for a given vector $\mathbf{y} \in \mathbb{R}^N$, $\mathbf{y}_{\mathcal{I}_n}$ denotes the vector in $\mathbb{R}^{|\mathcal{I}_n|}$, where $\mathbf{e}_j^\top \mathbf{y}_{\mathcal{I}_n} = \mathbf{e}_{\mathcal{I}_n(j)}^\top \mathbf{y}$.

We now introduce the algorithm $\mathcal{CIRFE}$ for distributed parameter estimation:

$$\mathbf{x}_n(t+1) = \mathbf{x}_n(t) \underbrace{- b\alpha_t \sum_{l \in \Omega_n} \left(\mathbf{x}_{l,n}^s(t) - \mathbf{x}_{l,n}^r(t) \right)}_{\text{Neighborhood Consensus}} + \underbrace{\alpha_t \left(\nabla \mathbf{f}_n \left(\mathbf{x}_n^{\mathcal{P}_n}(t) \right) \mathbf{R}_n^{-1} \left(\mathbf{y}_n(t) - \mathbf{f}_n \left(\mathbf{x}_n^{\mathcal{P}_n}(t) \right) \right) \right)_{\mathcal{I}_n}}_{\text{Local Innovation}}, \tag{5}$$

where $\nabla \mathbf{f}(\cdot)$ denotes the gradient of $\mathbf{f}$ and $\alpha_t = a/(t+1)$. We now present a more compact representation of the $\mathcal{CIRFE}$ algorithm so as to be able to establish its consistency and characterize its asymptotic variance. Let $\mathcal{I}$ denote a subset of $\{1, 2, \cdots, N\}$. Define the diagonal matrix $P_{\mathcal{I}}$ which selects the corresponding components of $\mathcal{I}$ from a 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 $\widetilde{\mathbf{x}}_n(t) = \mathbf{x}_n(t)^{\mathcal{P}_n}$. With the above development in place, it is easy to see that, for $\mathbf{y} \in \mathbb{R}^N$, $\left(\mathbf{x}_{l,n}^r(t)\right)^{\mathcal{P}_n} = P_{\mathcal{I}_n} P_{\mathcal{I}_l} \widetilde{\mathbf{x}}_l(t)$, $\left(\mathbf{x}_{l,n}^s(t)\right)^{\mathcal{P}_n} = P_{\mathcal{I}_n} P_{\mathcal{I}_l} \widetilde{\mathbf{x}}_n(t)$ and $\left(\mathbf{x}_n(t)\right)^{\mathcal{P}_n} = P_{\mathcal{I}_n} \widetilde{\mathbf{x}}_n(t)$. The $\mathcal{CIRFE}$ update in (5) can then be written in terms of auxiliary states as follows:

$$\begin{aligned} \widetilde{\mathbf{x}}_n(t+1) &= \widetilde{\mathbf{x}}_n(t) - b\alpha_t \sum_{l \in \Omega_n} P_{\mathcal{I}_n} P_{\mathcal{I}_l} \left(\widetilde{\mathbf{x}}_n(t) - \widetilde{\mathbf{x}}_l(t) \right) \\ &+ \alpha_t P_{\mathcal{I}_n} \left(\nabla \mathbf{f}_n \left(P_{\mathcal{I}_n} \widetilde{\mathbf{x}}_n(t) \right) \mathbf{R}_n^{-1} \left(\mathbf{y}_n(t) - \mathbf{f}_n \left(P_{\mathcal{I}_n} \widetilde{\mathbf{x}}_n(t) \right) \right) \right). \end{aligned} \tag{6}$$

We introduce the matrix $\mathbf{L}_P \in \mathbb{R}^{N^2 \times N^2}$ so as to compactify the above representation further.

$$[\mathbf{L}_P]_{nl} = \begin{cases} -P_{\mathcal{I}_n} \sum_{i=1:i \neq n}^{N} L_{ni} P_{\mathcal{I}_r} & \text{if } n = l \\ L_{nl} P_{\mathcal{I}_l} & \text{otherwise,} \end{cases} \tag{7}$$

where $[\mathbf{L}_P]_{nl} \in \mathbb{R}^{N \times N}$ denotes the (n,l)-th sub-block of the block matrix $\mathbf{L}_P$. With the above development in place, the update in (6) can be written in a compact form as follows:

$$\begin{aligned} \widetilde{\mathbf{x}}(t+1) &= \widetilde{\mathbf{x}}(t) - b\alpha_t \mathbf{L}_P \widetilde{\mathbf{x}}(t) \\ &+ \alpha_t \mathbf{P} \mathbf{G}(\mathbf{P}\widetilde{\mathbf{x}}(t)) \mathbf{R}^{-1} \left(\mathbf{y}(t) - \mathbf{f}\left(\mathbf{P}\widetilde{\mathbf{x}}(t) \right) \right), \end{aligned} \tag{8}$$

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

Define the subspace $\mathcal{S}_P \in \mathbb{R}^{N^2}$ by $\mathcal{S}_P = \left\{\mathbf{y} \in \mathbb{R}^{N^2} | \mathbf{y} = \mathbf{P}\mathbf{w}, \text{for some } \mathbf{w} \in \mathbb{R}^{N^2}\right\}$. 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,*

$$\mathbf{y}^\top \left(\mathbf{L}_P + \mathbf{G}\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)\mathbf{R}^{-1}\mathbf{G}\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)^\top\right)\mathbf{y} \geq c_1 \left\|\mathbf{y}\right\|^2, \forall\, \mathbf{y} \in \mathcal{S}_P. \quad (9)$$

If the interest set $\mathcal{I}_n = \mathbf{I}_N$ for all n in case of a linear observation model $\mathbf{f}(\mathbf{1}_N \otimes \boldsymbol{\theta}) = \mathbf{G}_H^\top \mathbf{1}_N \otimes \boldsymbol{\theta}$, then the matrix $\mathbf{L}_P + \mathbf{G}\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)\mathbf{R}^{-1}\mathbf{G}\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)^\top$ and the subspace $\mathcal{S}_P$ reduce to $\mathbf{L} \otimes \mathbf{I}_N + \mathbf{P}\mathbf{G}_H\mathbf{R}^{-1}\mathbf{G}_H^\top\mathbf{P}$ and $\mathbb{R}^{N^2}$ respectively, for which c_1 can be evaluated in a straightforward manner. In general, global observability and network connectivity is not enough to enforce Assumption A4. A sufficient condition to enforce Assumption A4 is that the induced subgraph for every entry of the vector $\boldsymbol{\theta}^*$ needs to be connected. To be specific, an induced subgraph for the r-th entry is the set of agents and their associated links which have r-th entry of $\boldsymbol{\theta}^*$ in their interest sets. We formalize an assumption which would be instrumental in establishing the consistency of the estimate sequence.

Assumption A6. *The following aggregate strict monotonicity condition holds: there exists a constant $c_1 > 0$ such that for each pair $\boldsymbol{\theta}, \acute{\boldsymbol{\theta}}$ we have that*

$$\sum_{n=1}^{N}\left(\boldsymbol{\theta} - \acute{\boldsymbol{\theta}}\right)^\top P_{\mathcal{I}_n}\left(\nabla f_n(\boldsymbol{\theta})\right)\mathbf{R}_n^{-1}\left(f_n(\boldsymbol{\theta}) - f_n(\acute{\boldsymbol{\theta}})\right) \geq c_1\left\|\boldsymbol{\theta} - \acute{\boldsymbol{\theta}}\right\|^2. \quad (10)$$

The parameter update proposed in this paper is recursive in nature. The above assumption puts a condition on the sensing functions so as to ensure the existence of stochastic Lyapunov function, thus, aiding establishing the convergence of the $\mathcal{CIRFE}$ algorithm.

Finally, let $\{\lambda_1, \lambda_2, \cdots, \lambda_{N_0}\}$ denote the set of non-zero eigenvalues of $\mathbf{L}_P + \mathbf{G}\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)\mathbf{R}^{-1}\mathbf{G}^\top\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)$ arranged in a non-decreasing order. Then, the corresponding set of orthonormal eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{N_0}\}$ span the subspace $\mathcal{S}_P$.

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

In this section we formally state the main results concerning the distributed parameter estimation $\mathcal{CIRFE}$ algorithm. The 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* (8)*. Let Assumptions A1-A4 hold. Then, we have,*

$$\mathbb{P}_{\boldsymbol{\theta}^*}\left(\lim_{t\to\infty}(t+1)^\tau\left\|\widetilde{\mathbf{x}}(t) - \mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right\| = 0\right) = 1, \quad (11)$$

for all $\tau \in [0, 1/2)$. In other words, the parameter estimate sequence $\{\mathbf{x}_n(t)\}$ are consistent for each n, and hence we have,

$$\mathbb{P}_{\boldsymbol{\theta}^*}\left(\lim_{t\to\infty}(t+1)^\tau\left\|\mathbf{x}_n(t) - \boldsymbol{\theta}^*_{\mathcal{I}_n}\right\| = 0\right) = 1, \quad (12)$$

for all $\tau \in [0, 1/2)$.

The convergence in Theorem 4.1 is order-optimal. Standard arguments in estimation theory show that in general there exists no $\tau \geq 1/2$ such that any centralized estimator $\{\widetilde{x}^c(t)\}$ satisfies $(t+1)^\tau\|\widetilde{x}^c(t) - \mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\| \to 0$ a.s. as $t \to \infty$. The next result concerns with the characterization of the asymptotic variance of the proposed estimator.

Theorem 4.2. *Let the hypotheses of Theorem 4.1 hold. In addition, assume that $(t+1)\alpha_t > \frac{1}{2\lambda_1}$. Then, the time-scaled sequence $\sqrt{t+1}\left(\widetilde{\mathbf{x}}(t) - \mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)$ is asymptotically normal, i.e.,*

$$\sqrt{t+1}\left(\widetilde{\mathbf{x}}(t) - \mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}\left(\mathbf{0}, \mathbf{S}_R\right), \quad (13)$$

where

$$\mathbf{S}_R = \sum_{k=1}^{N_0}\frac{a^2 s_R(k)}{2a\lambda_k - 1}\mathbf{u}_k\mathbf{u}_k^\top$$
$$s_R(k) = \mathbf{u}_k^\top\mathbf{G}\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)\mathbf{R}^{-1}\mathbf{G}^\top\left(\mathbf{P}\left(\mathbf{1}_N \otimes \boldsymbol{\theta}^*\right)\right)\mathbf{u}_k. \quad (14)$$

Theorem 4.2 establishes the asymptotic normality of the time-scaled estimate sequence. Having established the auxiliary estimate sequence $\{\widetilde{\mathbf{x}}(t)\}$ and noting that the estimate sequence $\{\mathbf{x}_n(t)\}$ is a linear transformation of the auxiliary estimate sequence, we conclude that $\sqrt{t+1}\left(\mathbf{x}(t) - \boldsymbol{\theta}^*_{\mathcal{I}_n}\right)$ is also asymptotically normal.

5. CONCLUSION

In this paper, we have proposed a *consensus + innovations* type algorithm, $\mathcal{CIRFE}$, for estimating a random field which exhibits a cyber-physical flavor. In the proposed algorithm, every agent updates its estimate of a few components of a high-dimensional field parameter vector by simultaneous processing of neighborhood information and local newly sensed information and where 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 and non-Gaussian noise.

6. 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] S. Kar and J. M. F. Moura, "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, S. Kar, J. M. F. Moura, and H. V. Poor, "Distributed constrained recursive nonlinear least-squares estimation: Algorithms and asymptotics," *arXiv preprint arXiv:1602.00382*, 2016.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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

[15] 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.

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