AN ALTERNATIVE APPROACH FOR DISTRIBUTED PARAMETER ESTIMATION UNDER GAUSSIAN SETTINGS

Subhro Das
MIT-IBM Watson AI Lab, IBM Research
Cambridge, MA, USA
subhro.das@ibm.com

ABSTRACT
This paper takes a different approach for the distributed linear parameter estimation over a multi-agent network. The parameter vector is considered to be stochastic with a Gaussian distribution. The sensor measurements at each agent are linear and corrupted with additive white Gaussian noise. Under such settings, this paper presents a novel distributed estimation algorithm that fuses the concepts of consensus and innovations by incorporating the consensus terms of neighboring estimates into the innovation terms. Under the assumption of distributed parameter observability, introduced in this paper, we design the optimal gain matrices such that the distributed estimates are consistent and achieve fast convergence.

Index Terms—parameter estimation, distributed algorithms, multiagent system, consensus, consistency

1. INTRODUCTION
Distributed parameter estimation has been studied significantly over the last two decades [1–3] for application in cyber-physical systems and wireless sensor networks. In distributed inference, there exists a class of consensus+innovations distributed estimation algorithms that generalize classical distributed consensus by combining cooperation among agents (consensus) with assimilation of their observations (innovations) [4, 5]. The problem is to estimate an underlying parameter vector by a large-scale network of inexpensive and low-power sensors, such that the measurements are sparse and corrupted by noise.

Here we point out that distributed estimation framework is fundamentally different from federated learning [6], where a federated (fusion) server exists that performs some sort of aggregation (averaging in most cases) of the local parameters. However, in the distributed setting the data is distributed across several agents and there is no fusion center. Each agent computes their own local estimates and shares those estimates with their neighboring agents in the network.

Although the primary focus has been on deterministic parameter [7, 8], in this paper we consider a stochastic parameter that follows a Gaussian distribution. Here, we introduce a new definition of distributed parameter observability [9] which is a weaker criteria than the distributed observability conditions considered in the literature. The key contribution of this work is the novel distributed algorithm that leverages the estimates from its neighbors and performs consensus within the innovation process. Such fusion is different from the consensus on pseudo-innovations in [10, 11]. This technique facilitates the design of the optimal gain matrices, in contrast to scalar weights considered in literature, that ensures fast convergence and yields consistent estimates.

Some of the important questions [12, 13] in distributed algorithms include the algorithms’ convergence pattern, consensus among the agents on their parameter estimates, distributed versus centralized performance comparison, and the rate of convergence. We discuss all these properties of the distributed estimator in the following sections.

2. GAUSSIAN PARAMETER ESTIMATION
2.1. Problem Statement
Consider the problem of estimating a random vector parameter \( \theta \in \mathbb{R}^n \) by a multi-agent network of \( m \) agents. The prior probability distribution of the parameter is Gaussian, \( \theta \sim \mathcal{N}(\bar{\theta}, \Sigma_\theta) \) where \( \bar{\theta} \in \mathbb{R}^n \) and \( \Sigma_\theta \in \mathbb{R}^{n \times n} \).

Each agent \( i \) in the network observes only a few variables and makes low dimensional measurements \( z_{i,k} \in \mathbb{R}^p \), such that \( p_i \ll n, \forall i = 1, \ldots, m \) at iteration \( k \). The measurements are i.i.d and corrupted with noise, which we assume to be Gaussian. The observations of the agents in the cyber layer is represented by a linear and time-invariant model

\[
z_{i,k} = H_i \theta + v_{i,k}, \quad i = 1, \ldots, m \tag{1}
\]

where, \( H_i \in \mathbb{R}^{p_i \times n} \) is the measurement matrix and \( v_{i,k} \in \mathbb{R}^{p_i} \) is the measurement noise. The Gaussian measurement noise, at each agent \( i \), has zero mean and covariance matrix \( R_{i,k} \), i.e., \( v_{i,k} \sim \mathcal{N}(0_{p_i}, R_{i,k}) \). The measurement noises \( \{v_{i,k}\}_{i,k \geq 0} \) are uncorrelated random sequences.

The communication network helps the agents to share their measurements and current estimates with their neigh-
Readers are referred to [14] for detailed information about the distributed estimation algorithm to benchmark and compare the results on the distributed estimates. In problem scenarios where local communication channels among the agents, the communication network can be defined by a simple (no self-loops nor multiple edges) and directed graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{i : i = 1, \ldots, m\}$ is the set of agents and $\mathcal{E} = \{(i, j) : \exists \text{ an edge } j \to i\}$ is the set of local communication channels among the agents. The communication graph is sparse and time-invariant. Note that the framework is based on a directed graph (one-way communications), that means the proposed approach is easily extendable to undirected graphs (two-way communications). The reverse extension is not always true. Let’s denote the adjacency matrix of $G$ by $A = [a_{ij}] \in \mathbb{R}^{m \times m}$, where

\[ a_{ij} = \begin{cases} 
1, & \text{if } \exists \text{ an edge } j \to i \\
0, & \text{otherwise}. 
\end{cases} \]

(2)

The equations (4)-(6) represent the centralized filter. The estimates $x_k^*$ are essentially conditional means given the observations $z_k$, i.e.,

\[ x_k^* = \mathbb{E}\{\theta \mid \{z_k\}_{k=0}^\infty\}. \]

(7)

To analyze the convergence properties of the centralized estimator, we derive the expressions for the centralized estimator error $e_k^* \in \mathbb{R}^n$ and innovation $y_k^* \in \mathbb{R}^p$, sequences.

\[ e_k^* = \theta - x_k^* \]  
\[ y_k^* = z_k - H x_k^* = H \theta + v_k \]

(8)  
(9)

The aggregated measurements for a centralized estimation then the centralized estimator can be initialized with $x_0 = \mathbb{E}\theta$ and $P_0 = I_n$. The centralized estimation updates for all $k \geq 0$ are:

\[ K_k^* = P_k^* H^T \left( H P_k^* H^T + R_k \right)^{-1} \]  
\[ x_{k+1}^* = x_k^* + K_k^* (z_k - H x_k^*) \]  
\[ P_{k+1}^* = (I_n - K_k^* H) P_k^* \]

(4)  
(5)  
(6)

The theorems (4)-(6) represent the centralized filter. The centralized observability matrix $\bar{K}_k^*$ is the set of $\mathcal{V}$.

The centralized process follows the dynamics, utilizing (5) and (10),

\[ e_{k+1}^* = \theta - x_{k+1}^* - \theta^* = K_k^* (z_k - H x_k^*) = e_k^* - K_k^* y_k^* \]  
\[ = e_k^* - K_k^* (H e_k^* + v_k) = (I_n - K_k^* H) e_k^* - K_k^* v_k \]

(11)  
(12)

The convergence properties of the centralized estimator (4)-(6) is determined by the dynamics of the error process, (12). If the error dynamics is asymptotically stable, then the error processes achieves asymptotic convergence that in turn the convergence of the estimation algorithm. The error dynamics is asymptotically stable if the spectral radius of the error’s dynamics matrix is less than one, i.e., $\rho(I_n - K_k^* H) < 1$.

**Definition 1** (Centralized Observability). If the row rank of the centralized observability matrix $H$ is equal to $n$, then the parameter is observable under the measurement model (3). This criteria is akin to the condition,

\[ \text{rank} \left( H^T H \right) = \text{rank} \left( \sum_{i=1}^m H_i^T H_i \right) = n. \]

(13)

Given that the observation model (3) satisfies the centralized observability criteria (13), it guarantees that there exists gain matrices $K_k^*$ such that $\rho(I_n - K_k^* H) < 1, \forall k \geq 0$.

The design of the optimal centralized gain matrices that ensures unbiasedness and consistency of the centralized estimates while ensuring fastest convergence, is obtained by applying Gauss-Markov theorem on

\[ x_{k+1}^* = x_k^* + K_k^* y_k^*. \]

(14)

Note that both the centralized estimator error $e_k^*$ and innovation $y_k^*$ sequences are zero-mean, i.e., $\mathbb{E}\{e_k^*\} = 0$ and $\mathbb{E}\{y_k^*\} = 0$. Then the error covariance is $P_k^* = \mathbb{E}\{e_k^* e_k^{T}\}$. The centralized gain is constructed as:

\[ K_k^* = \Sigma_{\theta y_k^*} \Sigma_{y_k^*}^{-1}, \]  
\[ \Sigma_{\theta y_k^*} = \mathbb{E}\{(\theta - \bar{\theta})(y_k^* - \mathbb{E}[y_k^*])^T\} = \mathbb{E}[(\theta - x_k^* + x_k^* - \bar{\theta})(y_k^* - \mathbb{E}[y_k^*])^T] = \mathbb{E}[e_k^* (H e_k^* + v_k)] = P_k^* H^T \]  
\[ \Sigma_{y_k^*} = \mathbb{E}\{(y_k^* - \mathbb{E}[y_k^*])(y_k^* - \mathbb{E}[y_k^*])^T\} = \mathbb{E}[(H e_k^* + v_k)(H e_k^* + v_k)^T] = H P_k^* H^T + R_k. \]

(15)  
(16)  
(17)
Replacing the covariance expressions from (16)-(17) into (15) yield optimal centralized estimator gains (4). Leveraging the uncorrelatedness of the noise, innovation and error sequences it can be shown that \( \mathbb{E}[\{x_k^q - \tilde{\theta} y_k^q\}^T]\) and \( \mathbb{E}[\epsilon_k^q v_k^q]^T = 0 \) which are used in the derivation of the gain.

Now we derive the recursive update of the centralized error covariance \( P_{k}^* \), starting with (12).

\[
P_{k+1}^* = \mathbb{E}[\epsilon_{k+1}^q \epsilon_{k+1}^q]^T
= \mathbb{E}[(I_n - K_k^* H) \epsilon_k^q (I_n - K_k^* H) \epsilon_k^q]^T
= P_k^* - K_k^* H P_k^* H^T K_k^* + K_k^*(H P_k^* H^T + R_k) K_k^*^T
= (I_n - K_k^* H) P_k^*,
\]

which is the update equation (6). This completes the derivation of all the processes of the centralized estimator.

Since the centralized framework satisfies the centralized observability criteria (13), the gain matrices \( K_k^* \) in (4) are such that \( \rho(I_n - K_k^* H) < 1 \) \( \forall k \geq 0 \). In the limit, the centralized error covariance then decays down to a zero matrix,

\[
\lim_{k \to \infty} P_k^* = \lim_{k \to \infty} (I_n - K_k^* H) P_{k-1}^* = \left( \lim_{k \to \infty} \Pi_{k=0}^{k-1} (I_n - K_k^* H) \right) P_0^* = 0_{n \times n}.
\]

Thus, as \( k \to \infty \) the centralized estimate \( x_k^q \) converges to the true parameter vector \( \theta \), i.e., \( \mathbb{P} \left[ \lim_{k \to \infty} x_k^q = \theta \right] = 1 \). Hence the centralized estimator is a consistent estimator.

3. DISTRIBUTED PARAMETER ESTIMATION

In this paper, a distributed approach is proposed that tracks the centralized approach at each agent. For each agent to be able to successfully track the centralized estimator, a key condition is the distributed observability that needs to be satisfied.

3.1. Distributed observability for parameter estimation

This paper introduces a new notion of distributed observability for parameter estimation by extending the distributed state observability definition provided in [9]. The distributed observability at each agent is a measure of how well the parameter can be inferred from knowledge of its local measurements and interactions among agents in the network.

Consider the observation-communication model represented by (1) and (2). Connectivity matrix \( \tilde{A} \) of the network is defined in [9], in terms of adjacency matrix \( A \), as \( \tilde{A} = I_m + A + A^2 + \ldots + A^{n-1} \).

As described in [9], the element \( [A^q]_{i,j} \) of the matrix, \( A^q \) \( \forall q \in \mathbb{Z}^+ \), gives the number of directed walks of length \( q \) from agent \( j \) to agent \( i \). Then, the connectivity matrix is a non-negative matrix\(^1\), \( \tilde{A} \geq 0 \), and its elements \( \tilde{a}_{i,j} = \tilde{a}_{i,j} \) denote the number of walks (of length \( < m \)) from \( j \) to \( i \).

\(^1\)If there exists \( i,j \) such that \( \tilde{a}_{i,j} = 0 \), then there doesn’t exist any path from \( j \) to \( i \) and would imply that the graph is not connected. The agent communication network, i.e., the directed graph, is connected if the connectivity matrix, defined in (20), is a positive matrix, i.e., \( \tilde{A} > 0 \). For a fully connected network, \( I_m + A > 0 \).

3.2. Distributed Parameter Estimation Algorithm

At time \( k \) and agent \( i \), let \( x_{i,k} \in \mathbb{R}^n \) denote the distributed estimate of \( \theta \), \( P_{i,k} \in \mathbb{R}^{n \times n} \) denote the error covariance, \( \kappa_{i,k} \in \mathbb{R}^{n \times n} \) denote the cross-covariance between estimation errors of agent \( i \) & agent \( j \), and \( K_{i,k} \in \mathbb{R}^{m \times n \times n} \) denote the gain matrix. At iteration \( k = 0 \), initialize the estimate and the error covariance\(^2\) at each agent \( i \) with \( x_{i,0} = \tilde{\theta}, P_{i,0} = \Sigma_0, P_{i,j,0} = \Sigma_{i,j} \forall j \in \Omega_i, \) and \( P_{j,i,0} = P_{0,n} \forall k < j \notin \Omega_i \). The update equations of the distributed estimation algorithm at each agent \( i \) for all \( k \geq 0 \) are:

\[
\kappa_{i,k} = \Sigma_{\theta,\theta_i}^{-1},
\]

\[
x_{i,k+1} = x_{i,k} + \sum_{j \in \Omega_i} B_{i,j,k}(x_{j,k} - x_{i,k}) - \sum_{j \in \Omega_i} M_{i,j,k}(x_{j,k} - H_j x_{i,k}),
\]

\[
P_{i,k+1} = P_{i,k} - K_{i,k} \Sigma_{\theta,\theta_i}^{-1},
\]

\[
P_{i,j,k} = P_{i,k} - K_{i,k} \Sigma_{\theta,\theta_i} - \Sigma_{\theta,\theta_i} K_{i,j,k}^T + K_{i,k} \Sigma_{\theta,\theta_i} K_{j,k}^T,
\]

where, \( j \in \Omega_i \), the covariance matrices \( \Sigma_{\theta,\theta_i}, \Sigma_{\theta,\theta_j}, \) and \( \Sigma_{\theta,\theta_j} \) are derived as part of the optimal gain design in subsection 4.2. The parameter estimate update step (24) is of canonical form, defined in (20), is a positive matrix, i.e., \( \tilde{A} > 0 \). For a fully connected network, \( I_m + A > 0 \).

\(^2\)If the prior statistics of the parameter, \( \tilde{\theta} \) and \( \Sigma_0 \), are unknown, then the centralized estimator can be initialized with \( x_{i,0} = \theta_0, P_{i,0} = I_n, \), and \( P_{i,j,0} = I_n \forall i = 1, \ldots, m \) and \( j \in \Omega_i \). Note that we denote \( P_{i,k} \) by \( P_t \) for ease of notation.

Let the \( \tilde{A} \), denote the \( \theta \)-th row of the matrix \( \tilde{A} \) and the symbol \( \bullet \) denote the face-splitting product of matrices (transposed Khatri–Rao product).

**Definition 2 (Distributed Observability - Parameter estimation).** If the row rank of the distributed observability matrix \( \Omega_i \), defined as,

\[
\Omega_i = \tilde{A}_i \bullet H = \begin{bmatrix} \tilde{a}_{i,1} & H_1 \\ \vdots & \vdots \\ \tilde{a}_{i,m} & H_m \end{bmatrix} = \begin{bmatrix} \tilde{a}_{i,1} H_1 \\ \vdots \\ \tilde{a}_{i,m} H_m \end{bmatrix}
\]

is equal to \( n \), then the parameter is distributedly observable at agent \( i \). This condition is equivalent to,

\[
\text{rank}(\Omega_i^T \Omega_i) = \text{rank} \left( \sum_{j=1}^{m} \tilde{a}_{i,j}^2 H_j^T H_j \right) = n
\]

Most papers in distributed parameter estimation assumes that both the centralized observability (13) and the connected graph conditions hold true. In contrast, this distributed observability for parameter estimation is a weaker assumption and does not require the graph to be connected. The distributed parameter estimator proposed in this paper needs the underlying framework to only satisfy the distributed observability criteria and is thus a more broadly applicable approach.
consensus+innovations type, where $B_{ij,k} \in \mathbb{R}^{n \times n}$ are the local consensus weight matrices and $M_{ij,k} \in \mathbb{R}^{n \times p_i}$ are the local innovation weight matrices.

It is to be noted that not only the observations $\{z_{j,k}\}_{j \in \Omega_i}$ but also the parameter estimates $\{x_{j,k}\}_{j \in \Omega_i}$, from the neighbors can be considered as additional information at each agent at each iteration. Based on this characterization, the notions of consensus and innovations are fused together into innovations $y_{i,k} \in \mathbb{R}^{(\sum_{j \in \Omega_i} p_i + n|\Omega_i|)}$ at each agent $i$ as:

$$x_{i,k+1} = x_{i,k} + K_{i,k} y_{i,k},$$

where

$$K_{i,k} = \left[ M_{ij,k}, \ldots, M_{ik,k}, \ldots, M_{ij,\Omega_i,k}, B_{ij,k}, \ldots, B_{ij,\Omega_i,k} \right],$$

(28)

$\{j_1, \ldots, j_{|\Omega_i|}\} = \Omega_i$ and $\{j_1, \ldots, j_{|\Omega_i|}\} = \Omega_i$. The local consensus $B_{ij,k}$ and local innovation $M_{ij,k}$ weight matrices in the parameter estimate update step (24) are obtained from the distributed gain matrix $K_{i,k}$ by utilizing the expression in (28). The equations (23)-(25) along with (28) represents the proposed distributed parameter estimation algorithm that fuses the concepts of consensus and innovations by treating the consensus on the state estimates as innovations along with the local innovations of the agent and its neighbors. In the next section, we investigate the convergence properties of this algorithm and provide a design of the optimal gain matrix that ensures fast convergence.

4. CONVERGENCE ANALYSIS

Similar to the centralized estimator, the distributed estimates $x_{i,k+1}$ are conditional means of the parameter $\theta$ given the observations $\{z_{j,k}\}_{j \in \Omega_i}$ and the neighbors’ parameter estimates $\{x_{j,k}\}_{j \in \Omega_i}$, i.e.,

$$x_{i,k+1} = \mathbb{E} \left[ \theta \mid \{z_{j,k}\}_{j \in \Omega_i}, \{x_{j,k}\}_{j \in \Omega_i}, k = 0, \ldots, K \right].$$

(29)

This relation implies that the innovation sequences $\{y_{i,k}\}_{k \geq 0}$ are Gaussian random vectors, uncorrelated and are with zero mean, $\mathbb{E}[y_{i,k}] = 0, \forall k \geq 0$. Building upon these, we analyze the error processes of the distributed estimator.

4.1. Error Analysis

The distributed estimation error terms, $\epsilon_{i,k} \in \mathbb{R}^n$ at each agent $i$, are defined as,

$$\epsilon_{i,k} = \theta - x_{i,k}.$$  

(30)

The error processes $\epsilon_{i,k}$ are unbiased, i.e., zero mean at all agents and for all time indices, $\mathbb{E}[\epsilon_{i,k}] = 0, \forall i, k \geq 0$. The error processes follow: $\epsilon_{i,k} \sim \mathcal{N}(0, P_{i,k})$. We now derive the innovations in terms of the error process and vice versa. First, using (27) and (1), the innovation terms expand as:

$$y_{i,k} = \begin{bmatrix} H_{ij_1} \theta + \nu_{ij_1,k} - H_{ij_1} x_{i,k} \\ \vdots \\ H_{ij_{|\Omega_i|}} \theta + \nu_{ij_{|\Omega_i|},k} - H_{ij_{|\Omega_i|}} x_{i,k} \end{bmatrix} = \begin{bmatrix} H_{ij_1} \theta \\ \vdots \\ H_{ij_{|\Omega_i|}} \theta \end{bmatrix} + \begin{bmatrix} \nu_{ij_1,k} \\ \vdots \\ \nu_{ij_{|\Omega_i|},k} \end{bmatrix},$$

$$y_{i,k} = \begin{bmatrix} \tilde{H}_{ij_1} \epsilon_{i,k} + \delta_{ij_1,k} \\ \vdots \\ \tilde{H}_{ij_{|\Omega_i|}} \epsilon_{i,k} + \delta_{ij_{|\Omega_i|},k} \end{bmatrix},$$

(31)

where

$$\tilde{H}_i \in \mathbb{R}^{(\sum_{j \in \Omega_i} p_i + n|\Omega_i|) \times n}$$

are the local innovation matrices and the $\delta_{ij,k} \in \mathbb{R}^{(\sum_{j \in \Omega_i} p_i + n|\Omega_i|)}$ are the local innovation noises at each agent $i$. The dynamics of the local innovation processes are represented in compact notation by (31). The local innovation noises $\delta_{ij,k}$ are Gaussian random vectors with zero mean and let the variance be denoted by $\Delta_{ij,k}$, i.e., $\delta_{ij,k} \sim \mathcal{N}(0, \Delta_{ij,k})$.

The distributed error process $\epsilon_{i,k}$ follows the dynamics, utilizing (27) and (31),

$$\epsilon_{i,k+1} = \theta - x_{i,k+1} = \theta - (x_{i,k} + K_{i,k} y_{i,k}) = \epsilon_{i,k} - K_{i,k} y_{i,k},$$

(32)

$$= \left( I_n - K_{i,k} \tilde{H}_i \right) \epsilon_{i,k} - K_{i,k} \delta_{i,k},$$

(33)

and, the evolution of the distributed error covariance $P_{i,k}$ takes the form, using (32),

$$P_{i,k+1} = \mathbb{E}[\epsilon_{i,k+1} \epsilon_{i,k+1}^T] = \mathbb{E} \left[ (\epsilon_{i,k} - K_{i,k} y_{i,k}) (\epsilon_{i,k} - K_{i,k} y_{i,k})^T \right],$$

$$= P_{i,k} + K_{i,k} \Sigma_{\theta y_i} K_{i,k}^T - K_{i,k} \Sigma_{\theta y_i} K_{i,k}^T,$$

(34)

The term $\mathbb{E}[\epsilon_{i,k} y_{i,k}^T]$ is shown later in (36) and by substituting $K_{i,k}$ with $\Sigma_{\theta y_i} \Sigma_{y_i}^{-1}$ we obtain equation (34). Similarly, the error cross covariance $P_{ij,k}$ is derived as,

$$P_{ij,k+1} = \mathbb{E}[\epsilon_{i,k+1} \epsilon_{j,k+1}^T] = \mathbb{E} \left[ (\epsilon_{i,k} - K_{i,k} y_{i,k}) (\epsilon_{j,k} - K_{j,k} y_{j,k})^T \right],$$

$$= P_{ij,k} + K_{ij,k} \Sigma_{\theta y_i} K_{ij,k}^T + K_{ij,k} \Sigma_{\theta y_j} K_{ij,k}^T,$$

(35)

In the above derivation, $\mathbb{E}[\epsilon_{i,k} y_{i,k}^T]$ also reduces to $\Sigma_{\theta y_i}$ and the relation holds true when $i$ and $j$ are interchanged. The designs of the gain matrices and the covariances are provided in the next subsection (36)-(43). Equations (34) and (35) are the update steps for the distributed error covariances of the proposed distributed parameter estimation algorithm.

4.2. Optimal Gain Design

Following along the same lines of centralized optimal gain matrix design, we state that the optimal distributed gain ma-
trices that ensure unbiasedness and consistency of the distributed estimates while ensuring fast convergence, are obtained by following the first principles of Gauss-Markov theorem. Since the zero-mean innovation sequences \( \{ y_{i,k} \}_{i,k \geq 0} \) are Gaussian and uncorrelated, they are independent random vectors. Applying Gauss-Markov theorem on (27), the optimal gain matrices for the distributed parameter estimates are:

\[
K_{i,k} = \sum \theta_{y_i} \sum_{y_i}^{-1}, \quad \text{where,}
\]

\[
\sum \theta_{y_i} = E[\theta - \bar{\theta}] y_i^T, \quad \sum_{y_i}^{-1} = \sum [\bar{y}_i, \bar{y}_i] = E[y_i, y_i^T], \quad \text{for } i = 1, 2, \ldots
\]

\[
= E[y_i, (H_0 y_i + \delta_i)^T] = K_{i,k} y_i + \sum_{i,j} \epsilon_{ij}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}^T
\]

The innovation cross-covariance, \( \epsilon_{ij} \), is derived similarly using the steps described above:

\[
\epsilon_{ij} = E[y_i, \epsilon_j] = E[y_i, (H_0 y_i + \delta_i)y_j^T] = \sum_{i,j} y_{i,j} + \sum_{i,j} \epsilon_{ij}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}^T
\]

\[
\epsilon_{ij} = E[y_i, \epsilon_j] = E[y_i, (H_0 y_i + \delta_i)y_j^T] = \sum_{i,j} y_{i,j} + \sum_{i,j} \epsilon_{ij}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}^T
\]

\[
\epsilon_{ij} = E[y_i, \epsilon_j] = E[y_i, (H_0 y_i + \delta_i)y_j^T] = \sum_{i,j} y_{i,j} + \sum_{i,j} \epsilon_{ij}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}^T
\]

\[
\epsilon_{ij} = E[y_i, \epsilon_j] = E[y_i, (H_0 y_i + \delta_i)y_j^T] = \sum_{i,j} y_{i,j} + \sum_{i,j} \epsilon_{ij}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}^T
\]

Note that the gain matricesare very sparse at each agent. To alleviate challenges in tracking of the complete network error covariances, [19] presents a certifiable optimal distributed filter that performs optimal fusion of estimates under unknown correlations by a particular tight Semidefinite Programming (SDP) relaxation. Further, given that the matrices does not depend on the measurements, they all can be pre-computed and stored at each agent.

4.3. Asymptotic properties

The convergence properties of distributed parameter estimation algorithm (23)-(25) can be investigated through the lens of the behavior of the error dynamics (33). If the error dynamics are asymptotically stable, i.e., the spectral radius of the error’s dynamics matrix is less than one, \( \rho \left( I_n - K_{i,k} H_i \right) < 1, \forall i \), then the error processes have asymptotically decaying error covariances that in turn guarantee the convergence of the distributed algorithm.

Given that the observation-network model (1)-(2) satisfies the Distributed Observability criteria (21) for parameter estimation, it guarantees that there exists gain matrices \( K_{i,k} \) at each agent \( i \) such that

\[
\rho \left( I_n - K_{i,k} H_i \right) < 1, \forall i. \tag{45}
\]

Note that there may exist multiple realizations of the distributed gain matrices \( K_{i,k} \in \mathbb{R}^{n \times (\sum_{i \in n} |\pi_i|)} \) satisfying (45) and for all such realizations the convergence of the distributed algorithm is guaranteed. The optimal design of the distributed gain matrices, provided in this paper (23) (28) and (44), not only guarantees convergence by satisfying (45) but also ensures fastest convergence.

From equations (34) and (37), we study the asymptotic behavior of the distributed estimation error covariance,

\[
\lim_{k \to \infty} P_{i,k+1} = \lim_{k \to \infty} \left( P_{i,k} - K_{i,k} \left( P_{i,k} H_i^T + \sum_{i,j} \epsilon_{ij} \right) \right). \tag{46}
\]

We can prove (by contradiction) that the design of the gain matrices \( K_{i,k} \) are such that \( \rho \left( I_n - K_{i,k} H_i + \sum_{i,j} \epsilon_{ij} \right) < 1, \forall i \) at all time indices \( k \geq 0 \). Let’s assume that the spectral radius is not less than unity for some \( i = \ell \) and \( k = \kappa \), then we replace \( K_{i,k} \) with

\[
K_{i,k} = \tilde{K}_{i,k} = \tilde{M}_{i,j,k}, \quad \forall j \in \Omega, \tag{47}
\]

where, \( \tilde{M}_{i,j,k} \) is the optimal design of the gain matrices \( \tilde{K}_{i,k} \), which can be derived similarly using the steps described above:

\[
\tilde{M}_{i,j,k} = \sum \theta_{y_i} \sum_{y_i}^{-1}, \quad \text{where,}
\]

\[
\sum \theta_{y_i} = E[\theta - \bar{\theta}] y_i^T, \quad \sum_{y_i}^{-1} = \sum [\bar{y}_i, \bar{y}_i] = E[y_i, y_i^T], \quad \text{for } i = 1, 2, \ldots
\]

\[
= E[y_i, (H_0 y_i + \delta_i)^T] = K_{i,k} y_i + \sum_{i,j} \epsilon_{ij}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}, \quad \epsilon_{ij} = \sum_{i,j} y_{i,j}^T
\]

\[
= \tilde{M}_{i,j,k} = \sum \theta_{y_i} \sum_{y_i}^{-1}, \quad \text{where,}
\]

\[
\sum \theta_{y_i} = E[\theta - \bar{\theta}] y_i^T, \quad \sum_{y_i}^{-1} = \sum [\bar{y}_i, \bar{y}_i] = E[y_i, y_i^T], \quad \text{for } i = 1, 2, \ldots
\]
becomes less than unity, given that \(\rho \left( I_n - K_i, k \tilde{H}_i \right) < 1\) is true because of the distributed observability assumption.

Since the spectral radius of the product terms in the right hand side of (46) are all less than unity, then in the limit, as \(k \to \infty\), the distributed estimation error covariance \(P_i, k\) reduces to zero, i.e., \(\lim_{k \to \infty} P_i, k+1 = 0, \forall i\). Hence, similar to its centralized counterpart, the proposed distributed estimator is also a consistent estimator where the distributed estimates \(x_{i, k}\) converge to the true parameter vector \(\theta\), i.e.,
\[
\mathbb{P} \left( \lim_{k \to \infty} x_{i, k} = \theta \right) = 1 \text{ at all the agents } i \text{ in the network.}
\]

5. CONCLUSIONS

This paper considers the distributed parameter estimation problem where the parameter is stochastic and Gaussian distributed. The primary contributions are: (a) introducing a new distributed parameter estimation algorithm that incorporates consensus on neighbors’ estimates as innovations; (b) defining a new distributed parameter observability criteria that is a weaker assumption compared to the assumptions in literature; and, (c) designing the gain matrices, instead of scalar weights, for the distributed estimator such that the algorithm is optimal, i.e., it yields consistent estimates at all agents and achieves fast convergence. The framework and the methodology proposed in this paper can serve as a backbone for several downstream research problems in distributed (stochastic) parameter estimation, especially for optimal sensor placement, adaptation to node or communication failures.

6. REFERENCES

[1] C.-Y. Chong, “Forty years of distributed estimation: A review of noteworthy developments,” in Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2017.

[2] D. Ding, Q.-L. Han, Z. Wang, and X. Ge, “A survey on model-based distributed control and filtering for industrial cyber-physical systems,” IEEE Transactions on Industrial Informatics, 2019.

[3] X. Ge, Q.-L. Han, X.-M. Zhang, L. Ding, and F. Yang, “Distributed event-triggered estimation over sensor networks: A survey,” IEEE Trans. on Cybernetics, 2019.

[4] S. Kar, J.M.F. Moura, and K. Ramanan, “Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication,” IEEE Transactions on Information Theory, 2012.

[5] S. Kar and J.M.F. Moura, “Consensus+innovations distributed inference over networks: cooperation and sensing in networked systems,” IEEE Signal Processing Magazine, 2013.

[6] J. Konečný, B. McMahan, F. Yu, P. Richtárik, A.T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” arXiv preprint arXiv:1610.05492, 2016.

[7] A. Ribeiro and G.B. Giannakis, “Distributed estimation in gaussian noise for bandwidth-constrained wireless sensor networks,” in Asilomar Conference on Signals, Systems and Computers, 2004.

[8] S. Stanković, M. Stankovic, and journal=IEEE Transactions on Automatic Control year=2010 title=Decentralized parameter estimation by consensus based stochastic approximation .

[9] S. Das, “On observability and optimal gain design for distributed linear filtering and prediction,” arXiv preprint arXiv:2203.03521, 2022.

[10] S. Das and J.M.F. Moura, “Distributed linear estimation of dynamic random fields,” in Allerton Conference on Communication, Control, and Computing, 2013.

[11] S. Das and J.M.F. Moura, “Distributed Kalman filtering and network tracking capacity,” in Asilomar Conference on Signals, Systems, and Computers, 2013.

[12] K.R. Rad and A. Tahbaz-Salehi, “Distributed parameter estimation in networks,” in IEEE Conference on Decision and Control (CDC), 2010.

[13] P. Bianchi, G. Fort, W. Hachem, and J. Jakubowicz, “Convergence of a distributed parameter estimator for sensor networks with local averaging of the estimates,” in IEEE ICASSP, 2011.

[14] F.R.K. Chung, Spectral Graph Theory, vol. 92, American Mathematical Society, 1997.

[15] S. Das and J.M.F. Moura, “Distributed state estimation in multi-agent networks,” in IEEE ICASSP, 2013.

[16] S. Das and J.M.F. Moura, “Distributed Kalman filtering,” in European Signal Processing Conference, 2013.

[17] S. Das and J.M.F. Moura, “Distributed Kalman filtering with dynamic observations consensus,” IEEE Transactions on Signal Processing, 2015.

[18] S. Das and J.M.F. Moura, “Consensus+innovations distributed Kalman filter with optimized gains,” IEEE Transactions on Signal Processing, 2017.

[19] E. Sebastián, E. Montijano, and C. Sagüés, “All-in-one: Certifiable optimal distributed kalman filter under unknown correlations,” arXiv, arXiv:2105.15061, 2021.