On observability and optimal gain design for distributed linear filtering and prediction

Subhro Das

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

Abstract—**This paper presents a new approach to distributed linear filtering and prediction. The problem under consideration consists of a random dynamical system observed by a multi-agent network of sensors where the network is sparse. Inspired by the *consensus+innovations* type of distributed estimation approaches, this paper proposes a novel algorithm that fuses the concepts of consensus and innovations. The paper introduces a definition of distributed observability, required by the proposed algorithm, which is a weaker assumption than that of global observability and connected network assumptions combined together. Following first principles, the optimal gain matrices are designed such that the mean-squared error of estimation is minimized at each agent and the distributed version of the algebraic Riccati equation is derived for computing the gains.**

Index Terms—**Kalman filter, distributed estimation, multiagent networks, distributed algorithms, consensus**

I. INTRODUCTION

The Kalman filter [1], [2] is a celebrated solution for linear filtering and prediction of time-varying random fields observed with noisy measurements. The growing need for data privacy and robustness from centralized server failure, led to the development of distributed filtering and predictions algorithms. These new genre of distributed estimation algorithms [3], [4] further reduces the large communication and computation overload at the centralized processors. Thereby, they position themselves as a critical framework for several applications namely multi-agent control [5], indoor positioning & navigation, state estimation in power grid, spatio-temporal environment or field monitoring [6], connected vehicular network for traffic balancing & accident aversion, collaborative target tracking [7] etc.

In the literature, there are distributed estimators where the agents exchange information multiple number of times between each dynamics/observations iterates [8]–[14], so that average consensus occurs between observations. There are also single time-scale approaches [15]–[21] where the agents collaborate with their neighbors only once in between each dynamics/observation cycle. There are distributed Kalman filters [22], [23] where the agents communicate among themselves using the Gossip protocol [24], over a dynamic communication network [25] or over noisy/corrupted communication channels [26]–[28]. Some distributed estimators run a companion filter to estimate the global average of the pseudo-innovations [29], [30], a modified version of the innovations.

There is a gap in the literature in providing for a distributed state estimation algorithm that is as general and robust as its centralized counterpart, the Kalman filter. In some papers, there are stricter assumptions on the local observation model or the communication network among the agents, for example neighborhood observability [31], [32] or an undirected connected graph [15]. In practical distributed settings, these assumptions become a bottleneck. In other papers, there is an upper limit on the degree of instability on the system dynamics that a given observation-network model can handle with bounded MSE [33]. Such algorithms become inapplicable to distributed process control domains where the underlying systems are inherently unstable and needs to be stabilized by appropriate control input. If the estimation algorithm fails to track the unstable system, then the design of a stabilizable control input become infeasible. The biggest gap is in the optimality of the distributed state estimates which leverages the maximum information available from the local observations and the estimates obtained from neighbors. The literature lacks in providing the optimal gain matrix design such that the algorithm yields yield minimum MSE estimates.

This paper proposes a novel framework and an algorithm that addresses all these gaps in the distributed filtering and prediction literature. The algorithm is provided in Section IV, where the consensus on the state estimates are treated as innovations and the gain matrices are designed appropriately. Such optimal gains, provided in Section VI, yield field estimates with minimum mean-squared error (MMSE) at each agent under the assumption of distributed obervability at each agent. A new definition of distributed observability is introduced in Section III which is agnostic of the communication graph being directed or undirected and does not require the graph to be connected. First, we start with setting up the system-observation model framework in the following Section II.

II. SYSTEM-OBSERVATION-COMMUNICATION MODEL

The system under consideration follows a discrete-time, linear, and time-invariant state-space model

$$\boldsymbol{x}_k = F\boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1}, \tag{1}$$

where, $\boldsymbol{x}_k \in \mathbb{R}^n$ is the dynamic random state vector, $F \in \mathbb{R}^{n \times n}$ is the state transition matrix, $\boldsymbol{w}_k \in \mathbb{R}^n$ is the system noise at all time $t = kT$ where T is the discrete-time step

size and k is an integer time index. The system noise is white Gaussian noise with zero mean and covariance matrix Q_k, i.e., $\boldsymbol{w}_k \sim \mathcal{N}(0, Q_k)$. The initial condition of the system, $\boldsymbol{x}_0$ is also Gaussian, that follows $\boldsymbol{x}_0 \sim \mathcal{N}(\overline{\boldsymbol{x}}_0, P_0^+)$.

The dynamic random state (1) is observed by a multi-agent network of m agents (sensors). Each agent i observes only a few state variables and makes low dimensional measurements $\boldsymbol{z}_{i,k} \in \mathbb{R}^{p_i}$, such that $p_i << n, \forall i = 1, \ldots, m$. The observations of the agents in the cyber layer is represented by a linear and time-invariant model

$$\boldsymbol{z}_{i,k} = H_i \boldsymbol{x}_k + \boldsymbol{v}_{i,k}, \qquad i = 1, \ldots, m \qquad (2)$$

where, $H_i \in \mathbb{R}^{p_i \times n}$ is the measurement matrix and $\boldsymbol{v}_{i,k} \in \mathbb{R}^{p_i}$ is the measurement noise. The measurement noise, at each agent i, is also white Gaussian noise with zero mean and covariance matrix $R_{i,k}$, i.e., $\boldsymbol{v}_{i,k} \sim \mathcal{N}(0, R_{i,k})$. The system noise, the measurement noise, and the initial condition $\{\{\boldsymbol{w}_k\}, \{\boldsymbol{v}_{i,k}\}, \boldsymbol{x}_0\}_{\forall i,k \geq 0}$ are uncorrelated random sequences.

The agents in the network layer exchange their measurements and current estimates with their neighbors. Formally, the agent communication network is represented by a simple (no self-loops nor multiple edges) and *directed* graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{i : i = 1, \ldots, m\}$ is the set of agents and $\mathcal{E} = \{(i, j) : \exists \text{ an edge } j \rightarrow i\}$ is the set of local communication channels among the agents. We consider directed graph (one-way communications), that means our algorithm is easily extendable to undirected graphs (two-way communications) but the reverse is not always true. The adjacency matrix of $\mathcal{G}$ is denoted by $\mathcal{A} = [a_{ij}] \in \mathcal{R}^{m \times m}$, where,

$$a_{ij} = \begin{cases} 1, & \text{if } \exists \text{ an edge } j \rightarrow i \\ 0, & \text{otherwise.} \end{cases} \qquad (3)$$

For details on graphs refer to [34]. The communication network is sparse and time-invariant. For each agent i, let's define the open and closed neighborhoods as: $\Omega_i = \{j | (i, j) \in \mathcal{E}\}$, and $\overline{\Omega}_i = \{i\} \cup \{j | (i, j) \in \mathcal{E}\}$, respectively.

As in the case of a classical optimal Kalman filter, each agent i in the framework knows the system model, F and Q_k, the initial condition statistics, $\overline{\boldsymbol{x}}_0$ and P_0^+, the parameters of its and neighbors' measurement models, $\{H_j, R_{j,k} : j \in \overline{\Omega}_i\}$, and the communication network model, $\mathcal{G}$ along with the adjacency matrix $\mathcal{A}$. Note that the time-invariant state-space is chosen for notational simplicity. All the derivations, assumptions and the results in this paper also hold for a time-varying state-space model $(F_k, H_{i,k}, R_{i,k}, Q_k)$.

III. DISTRIBUTED OBSERVABILITY

Before we propose the distributed estimation algorithm and describe the design of the optimal gain, we introduce the notion of distributed observability, a measure of how well internal states of a system can be inferred from knowledge of its local measurements and interactions among agents in the network. Consider the physical system-observation-communication modeled by the ***state-space-network*** representaion (1)-(3). The local observability matrix $G_i \in \mathcal{R}^{np_i \times n}$ and

the global observability matrix $G \in \mathcal{R}^{(n\sum_{i=1}^{m} p_i) \times n}$ of the network are denoted by,

$$G_i = \begin{bmatrix} H_i \\ H_i F \\ H_i F^2 \\ \vdots \\ H_i F^{n-1} \end{bmatrix}, \forall i \in \mathcal{V}; \qquad G = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_m \end{bmatrix}. \qquad (4)$$

Let's define the ***connectivity matrix*** $\tilde{\mathcal{A}}$ of the network as,

$$\tilde{\mathcal{A}} = I_m + \mathcal{A} + \mathcal{A}^2 + \ldots + \mathcal{A}^{m-1}. \qquad (5)$$

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

Remark 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 (5), is a positive matrix, i.e., $\tilde{\mathcal{A}} > 0$. For a fully connected network, $I_m + \mathcal{A} > 0$.*

Remark 2. *Although most literature in distributed estimation requires a connected graph, but connected graph is not a necessary condition for the distributed observability (defined below) and is also not required for the distributed estimation algorithm proposed in this paper.*

Let the quantity $\tilde{\mathcal{A}}_i$ denote the ith row of the matrix $\tilde{\mathcal{A}}$ and the symbol $\bullet$ denote the face-splitting product of matrices (transposed Khatri–Rao product).

Definition 1 (Distributed Observability). *If the row rank of the distributed observability matrix $\mathcal{O}_i$, defined as,*

$$\mathcal{O}_i = \tilde{\mathcal{A}}_i \bullet G = \begin{bmatrix} \tilde{a}_{i,1} \\ \tilde{a}_{i,2} \\ \vdots \\ \tilde{a}_{i,m} \end{bmatrix} \bullet \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_m \end{bmatrix} = \begin{bmatrix} \tilde{a}_{i,1} \otimes G_1 \\ \tilde{a}_{i,2} \otimes G_2 \\ \vdots \\ \tilde{a}_{i,m} \otimes G_m \end{bmatrix} \qquad (6)$$

is equal to n, then the system is distributedly observable at agent i.

Note that, $\mathcal{O}_i \in \mathcal{R}^{(n\sum_{i=1}^{m} p_i) \times n}$. The requirement of invertibility (or full-rank) of the distributed observability Gramian, $\mathcal{O}_i^T \mathcal{O}_i$, is an equivalent alternative of the distributed observability definition, i.e., $\text{rank}(\mathcal{O}_i^T \mathcal{O}_i) = n$.

Assumption 1. *The state-space-network model (1)-(3) is distributedly observable at all agents in the network.*

This crucial assumption ensures that the proposed distributed estimator converges with bounded mean-squared error[1]. Note that there is no requirement for the system to be

---

[1] A slightly weaker criteria defined as distributed detectability suffices for the convergence of the proposed algorithm. It requires only the unstable states to be observable.

stable or the network to be connected. With the distributed observability definition & Assumption 1, we state the algorithm in the next section followed by the optimal gain design.

IV. DISTRIBUTED ESTIMATION ALGORITHM

At time k and agent i, let us denote the filter and prediction estimates of the system by $\widehat{\boldsymbol{x}}_{i,k}^+$ and $\widehat{\boldsymbol{x}}_{i,k}^-$, respectively, and the filter and prediction error covariance matrices by $\boldsymbol{P}_{i,k}^+$ and $\boldsymbol{P}_{i,k}^-$, respectively. The prediction and filtering updates of the distributed estimation algorithm at agent i are:

$$\widehat{\boldsymbol{x}}_{i,k}^- = F\widehat{\boldsymbol{x}}_{i,k-1}^+ \tag{7}$$

$$\boldsymbol{P}_{i,k}^- = F\boldsymbol{P}_{i,k-1}^+ F^T + Q_k \tag{8}$$

$$\mathcal{K}_{i,k} = \Sigma_{x,y_i}\Sigma_{y_i}^{-1} \tag{9}$$

$$\widehat{\boldsymbol{x}}_{i,k}^+ = \widehat{\boldsymbol{x}}_{i,k}^- + \sum_{j\in\Omega_i}B_{ij,k}(\widehat{\boldsymbol{x}}_{j,k}^- - \widehat{\boldsymbol{x}}_{i,k}^-) + \sum_{j\in\overline{\Omega}_i}M_{ij,k}(\boldsymbol{z}_{j,k} - H_j\widehat{\boldsymbol{x}}_{i,k}^-) \tag{10}$$

$$\boldsymbol{P}_{i,k}^+ = \boldsymbol{P}_{i,k}^- - \mathcal{K}_{i,k}\Sigma_{x,y_i}^T \tag{11}$$

where, $\mathcal{K}_{i,k} \in \mathbb{R}^{n\times\left(\sum_{j\in\overline{\Omega}_i}p_j+n|\Omega_i|\right)}$ is the distributed gain matrix. The covariance matrices Σ_{x,y_i} and Σ_{y_i} are derived in the optimal gain design Section VI. The local consensus weight matrices, $B_{ij,k} \in \mathbb{R}^{n\times n}$, and the local innovation weight matrices, $M_{ij,k} \in \mathbb{R}^{n\times p_j}$, are obtained from the distributed gain matrix $\mathcal{K}_{i,k}$ as shown in (15).

The equations (7)-(11) represents the proposed distributed estimation algorithm, where the minimized MSE prediction and filter estimates $\widehat{\boldsymbol{x}}_{i,k}^-$ and $\widehat{\boldsymbol{x}}_{i,k}^+$ are the conditional means,

$$\widehat{\boldsymbol{x}}_{i,k}^- = \mathbb{E}\left[\boldsymbol{x}_k \mid \{\boldsymbol{z}_{j,k-1}\}_{j\in\overline{\Omega}_i}, \{\widehat{\boldsymbol{x}}_{j,k-1}^+\}_{j\in\Omega_i}\right] \tag{12}$$

$$\widehat{\boldsymbol{x}}_{i,k}^+ = \mathbb{E}\left[\boldsymbol{x}_k \mid \{\boldsymbol{z}_{j,k}\}_{j\in\overline{\Omega}_i}, \{\widehat{\boldsymbol{x}}_{j,k}^-\}_{j\in\Omega_i}\right]. \tag{13}$$

The filter update (10) is visually similar to *consensus+innovations* type of algorithms[2]. But functionally this algorithm 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. To best represent this functionality, the filter update (10) is rewritten with the local innovation term, $\boldsymbol{y}_{i,k}$ at agent i, as:

$$\widehat{\boldsymbol{x}}_{i,k}^+ = \widehat{\boldsymbol{x}}_{i,k}^- + \mathcal{K}_{i,k}\underbrace{\begin{bmatrix} \boldsymbol{z}_{j_1,k} - H_{j_1}\widehat{\boldsymbol{x}}_{i,k}^- \\ \cdot\cdot \\ \boldsymbol{z}_{i,k} - H_i\widehat{\boldsymbol{x}}_{i,k}^- \\ \cdot\cdot \\ \boldsymbol{z}_{j_{|\overline{\Omega}_i|},k} - H_{j_{|\overline{\Omega}_i|}}\widehat{\boldsymbol{x}}_{i,k}^- \\ \widehat{\boldsymbol{x}}_{j_1,k}^- - \widehat{\boldsymbol{x}}_{i,k}^- \\ \vdots \\ \widehat{\boldsymbol{x}}_{j_{|\Omega_i|},k}^- - \widehat{\boldsymbol{x}}_{i,k}^- \end{bmatrix}}_{\boldsymbol{y}_{i,k}}, \quad \text{where,} \tag{14}$$

$$\mathcal{K}_{i,k} = \left[M_{ij_1,k}, \cdot\cdot, M_{ii,k}, \cdot\cdot, M_{ij_{|\overline{\Omega}_i|},k}, B_{ij_1,k}, \cdot\cdot\cdot, B_{ij_{|\Omega_i|},k}\right], \tag{15}$$

$\{j_1, \cdot\cdot, i, \cdot\cdot, j_{|\overline{\Omega}_i|}\} = \overline{\Omega}_i$ and $\{j_1, \cdot\cdot\cdot, j_{|\Omega_i|}\} = \Omega_i$.

The innovation sequences $\{\boldsymbol{y}_{i,k}\}_{\forall i,k\geq 0}$ are Gaussian random vectors, uncorrelated and are with zero mean, $\mathbb{E}[\boldsymbol{y}_{i,k}] = 0, \forall i,k \geq 0$. These innovation terms are key to the optimal design of the gain matrices.

[2]The distributed estimator is inspired by the pseudo-innovations, pseudo-observations and pseudo-state approaches that are summarized in [35], [36].

V. ERROR ANALYSIS

Now we derive the predictor and filter error terms, $\boldsymbol{\epsilon}_{i,k}^- \in \mathbb{R}^n$ and $\boldsymbol{\epsilon}_{i,k}^+ \in \mathbb{R}^n$, respectively, at each agent i,

$$\boldsymbol{\epsilon}_{i,k}^- = \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{i,k}^- \tag{16}$$

$$\boldsymbol{\epsilon}_{i,k}^+ = \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{i,k}^+. \tag{17}$$

The error processes $\boldsymbol{\epsilon}_{i,k}^-$ and $\boldsymbol{\epsilon}_{i,k}^+$ are unbiased, i.e, they are zero mean at all agents and for all time indices, $\mathbb{E}[\boldsymbol{\epsilon}_{i,k}^-] = 0$ and $\mathbb{E}[\boldsymbol{\epsilon}_{i,k}^+] = 0$, $\forall i,k \geq 0$. The error processes follows: $\boldsymbol{\epsilon}_{i,k}^- \sim \mathcal{N}(\boldsymbol{0}_n, \boldsymbol{P}_{i,k}^-)$ and $\boldsymbol{\epsilon}_{i,k}^+ \sim \mathcal{N}(\boldsymbol{0}_n, \boldsymbol{P}_{i,k}^+)$. This shows that the distributed prediction $\widehat{\boldsymbol{x}}_{i,k}^-$ and filtering $\widehat{\boldsymbol{x}}_{i,k}^+$ estimates provided by this algorithm are unbiased.

From equation (14) and using (2), the innovations are expanded as:

$$\boldsymbol{y}_{i,k} = \begin{bmatrix} H_{j_1}\boldsymbol{x}_k + \boldsymbol{v}_{j_1,k} - H_{j_1}\widehat{\boldsymbol{x}}_{i,k}^- \\ \cdot\cdot \\ H_i\boldsymbol{x}_k + \boldsymbol{v}_{i,k} - H_i\widehat{\boldsymbol{x}}_{i,k}^- \\ \cdot\cdot \\ H_{j_{|\overline{\Omega}_i|}}\boldsymbol{x}_k + \boldsymbol{v}_{j_{|\overline{\Omega}_i|},k} - H_{j_{|\overline{\Omega}_i|}}\widehat{\boldsymbol{x}}_{i,k}^- \\ \widehat{\boldsymbol{x}}_{j_1,k}^- - \boldsymbol{x}_k + \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{i,k}^- \\ \vdots \\ \widehat{\boldsymbol{x}}_{j_{|\Omega_i|},k}^- - \boldsymbol{x}_k + \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{i,k}^- \end{bmatrix} = \underbrace{\begin{bmatrix} H_{j_1} \\ \cdot\cdot \\ H_i \\ \cdot\cdot \\ H_{j_{|\overline{\Omega}_i|}} \\ \boldsymbol{0}_{n\times n} \\ \vdots \\ \boldsymbol{0}_{n\times n} \end{bmatrix}}_{\tilde{H}_i}\boldsymbol{\epsilon}_{i,k}^- + \underbrace{\begin{bmatrix} \boldsymbol{v}_{j_1,k} \\ \cdot\cdot \\ \boldsymbol{v}_{i,k} \\ \cdot\cdot \\ \boldsymbol{v}_{j_{|\overline{\Omega}_i|},k} \\ \boldsymbol{\epsilon}_{i,k}^- - \boldsymbol{\epsilon}_{j_1,k}^- \\ \vdots \\ \boldsymbol{\epsilon}_{i,k}^- - \boldsymbol{\epsilon}_{j_{|\Omega_i|},k}^- \end{bmatrix}}_{\delta_{i,k}}$$

where, $\tilde{H}_i \in \mathbb{R}^{\left(\sum_{j\in\overline{\Omega}_i}p_j+n|\Omega_i|\right)\times n}$ are the local innovation matrices and the $\delta_{i,k} \in \mathbb{R}^{\sum_{j\in\overline{\Omega}_i}p_j+n|\Omega_i|}$ are the local innovation noises at each agent i. In compact notation, the dynamics of the local innovations are represented by,

$$\boldsymbol{y}_{i,k} = \tilde{H}_i\boldsymbol{\epsilon}_{i,k}^- + \delta_{i,k}, \qquad \forall i,k \geq 0. \tag{18}$$

The local innovation noises $\delta_{i,k}$ are Gaussian random vectors with zero mean and let the variance be denoted by $\Delta_{i,k}$, i.e., $\delta_{i,k} \sim \mathcal{N}(\boldsymbol{0}, \Delta_{i,k})$. Using the equations (1), (7), (14) and (18) on the predictor and filter errors (16) and (17), their dynamics take the form:

$$\boldsymbol{\epsilon}_{i,k}^- = F\boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1} - F\widehat{\boldsymbol{x}}_{i,k-1}^+ = F\boldsymbol{\epsilon}_{i,k-1}^+ + \boldsymbol{w}_{k-1} \tag{19}$$

$$\boldsymbol{\epsilon}_{i,k}^+ = \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{i,k}^- - \mathcal{K}_{i,k}\boldsymbol{y}_{i,k} = (I_n - \mathcal{K}_{i,k}\tilde{H}_i)\boldsymbol{\epsilon}_{i,k}^- - \mathcal{K}_{i,k}\delta_{i,k}$$
$$= (F - \mathcal{K}_{i,k}\tilde{H}_i F)\boldsymbol{\epsilon}_{i,k-1}^+ + (I_n - \mathcal{K}_{i,k}\tilde{H}_i)\boldsymbol{w}_{k-1} - \mathcal{K}_{i,k}\delta_{i,k} \tag{20}$$

Given that the predictor and filter errors are zero-mean, the recursive updates of the evolution of their covariances, (8) and (11) are derived below.

$$\boldsymbol{P}_{i,k}^- = \mathbb{E}[\boldsymbol{\epsilon}_{i,k}^-\boldsymbol{\epsilon}_{i,k}^{-T}] = \mathbb{E}[(F\boldsymbol{\epsilon}_{i,k-1}^+ + \boldsymbol{w}_{k-1})(F\boldsymbol{\epsilon}_{i,k-1}^+ + \boldsymbol{w}_{k-1})^T]$$
$$= F\boldsymbol{P}_{i,k-1}^+ F^T + Q_k \tag{21}$$

$$\boldsymbol{P}_{i,k}^+ = \mathbb{E}[\boldsymbol{\epsilon}_{i,k}^+\boldsymbol{\epsilon}_{i,k}^{+T}] = \mathbb{E}[(\boldsymbol{\epsilon}_{i,k}^- - \mathcal{K}_{i,k}\boldsymbol{y}_{i,k})(\boldsymbol{\epsilon}_{i,k}^- - \mathcal{K}_{i,k}\boldsymbol{y}_{i,k})^T]$$
$$= \boldsymbol{P}_{i,k}^- - \mathcal{K}_{i,k}\Sigma_{x,y_i}^T - \Sigma_{x,y_i}\mathcal{K}_{i,k}^T + \mathcal{K}_{i,k}\Sigma_{y_i}\mathcal{K}_{i,k}^T$$
$$= \boldsymbol{P}_{i,k}^- - \mathcal{K}_{i,k}\Sigma_{x,y_i}^T \tag{22}$$

Note that the expectations of the cross-terms in (21) are zero, which can be shown by using techniques similar to the ones presented in [16]. The term $\mathbb{E}[\boldsymbol{\epsilon}_{i,k}^-\boldsymbol{y}_{i,k}^T] = \Sigma_{x,y_i}$ as shown later in (25). Finally, (22) was obtained by substituting $\mathcal{K}_{i,k}$ with $\Sigma_{x,y_i}\Sigma_{y_i}^{-1}$.

Similarly, the error cross covariances $\boldsymbol{P}^-_{ij,k}$ and $\boldsymbol{P}^+_{ij,k}$, needed for the optimal gain matrix design, are derived as,

$$\boldsymbol{P}^-_{ij,k} = \mathbb{E}[\boldsymbol{\epsilon}^-_{i,k}\boldsymbol{\epsilon}^{-T}_{j,k}] = \mathbb{E}\left[(F\boldsymbol{\epsilon}^+_{i,k-1}+\boldsymbol{w}_{k-1})(F\boldsymbol{\epsilon}^+_{j,k-1}+\boldsymbol{w}_{k-1})^T\right]$$
$$= F\boldsymbol{P}^+_{ij,k-1}F^T + Q_k \quad (23)$$
$$\boldsymbol{P}^+_{ij,k} = \mathbb{E}[\boldsymbol{\epsilon}^+_{i,k}\boldsymbol{\epsilon}^{+T}_{j,k}] = \mathbb{E}[(\boldsymbol{\epsilon}^-_{i,k} - \mathcal{K}_{i,k}\boldsymbol{y}_{i,k})(\boldsymbol{\epsilon}^-_{j,k} - \mathcal{K}_{j,k}\boldsymbol{y}_{j,k})^T]$$
$$= \boldsymbol{P}^-_{ij,k} - \mathcal{K}_{i,k}\Sigma^T_{x,y_i} - \Sigma_{x,y_j}\mathcal{K}^T_{j,k} + \mathcal{K}_{i,k}\Sigma_{y_{ij}}\mathcal{K}^T_{j,k}. \quad (24)$$

At iteration $k = 0$, initialize the estimate and the error covariances at each agent i with $\widehat{\boldsymbol{x}}^+_{i,0} = \overline{\boldsymbol{x}}_0$, $\boldsymbol{P}^+_{i,0} = P^+_0$, $\boldsymbol{P}^+_{ij,0} = P^+_0 \ \forall j \in \Omega_i$, and $\boldsymbol{P}^+_{ij,0} = \boldsymbol{0}_{n\times n} \ \forall j \notin \overline{\Omega}_i$. For ease of notation, we denote $\boldsymbol{P}^+_{ii,k}$ and $\boldsymbol{P}^-_{ii,k}$ by $\boldsymbol{P}^+_{i,k}$ and $\boldsymbol{P}^+_{i,k}$ respectively.

The convergence properties of the distributed estimator (7)-(11) is determined by the dynamics of the filter and prediction error processes, (20) and (19). If the error dynamics are asymptotically stable, then the error processes have asymptotically bounded error covariances that in turn guarantee the convergence of the distributed algorithm. Note that if the dynamics of the filter error processes, $\boldsymbol{\epsilon}^+_{i,k} \ \forall i$, are asymptotically stable, then the dynamics of the prediction error processes, $\boldsymbol{\epsilon}^+_{i,k} \ \forall i$ are also asymptotically stable. That is why the dynamics of only one of the error processes is typically studied and in this paper we consider the filter error processes.

For the distributed estimator to converge with bounded mean-squared error (MSE), the filter error (20) needs to be asymptotically stable, i.e., the spectral radius of the error's dynamics matrix has to be less than one, $\rho(F - \mathcal{K}_{i,k}\tilde{H}_iF) < 1$. Given that the *state-space-network* model (1)-(3) satisfies the Distributed Observability criteria (6), it guarantees that there exists gain matrices $\mathcal{K}_{i,k}$ at each agent i such that $\rho(F - \mathcal{K}_{i,k}\tilde{H}_iF) < 1$. This leads to the next section where a design of the optimal gain matrices are provided.

VI. OPTIMAL GAIN DESIGN

The asymptotic stability of the error dynamics guarantees convergence of distributed estimation algorithm (7)-(11) and bounded MSE, but we aim to to design the gain matrices $\mathcal{K}_{i,k}$ such that the MSE is not only bounded but also minimum.

Since the zero-mean innovation sequences $\{\boldsymbol{y}_{i,k}\}_{\forall i,k\geq 0}$ are Gaussian and uncorrelated, they are independent random vectors. By applying Gauss-Markov theorem on equation (15), the gain matrices that minimizes the MSE of the filter and prediction estimates are given by,

$$\mathcal{K}_{i,k} = \Sigma_{x,y_i}\Sigma^{-1}_{y_i}, \qquad \text{where,}$$
$$\Sigma_{x,y_i} = \mathbb{E}[(\boldsymbol{x}_k - \overline{\boldsymbol{x}}_k)\boldsymbol{y}^T_{i,k}] = \mathbb{E}[(\boldsymbol{x}_k - \widehat{\boldsymbol{x}}^-_{i,k} + \widehat{\boldsymbol{x}}^-_{i,k} - \overline{\boldsymbol{x}}_k)\boldsymbol{y}^T_{i,k}]$$
$$= \mathbb{E}[\boldsymbol{\epsilon}^-_{i,k}\boldsymbol{y}^T_{i,k}] = \mathbb{E}[\boldsymbol{\epsilon}^-_{i,k}(\tilde{H}_i\boldsymbol{\epsilon}^-_{i,k} + \delta_{i,k})^T] \quad (25)$$
$$= \boldsymbol{P}^-_{i,k}\tilde{H}^T_i + \Sigma_{\epsilon_i,\delta_i} \quad (26)$$
and,
$$\Sigma_{y_i} = \mathbb{E}[\boldsymbol{y}_{i,k}\boldsymbol{y}^T_{i,k}] = \mathbb{E}[(\tilde{H}_i\boldsymbol{\epsilon}^-_{i,k} + \delta_{i,k})(\tilde{H}_i\boldsymbol{\epsilon}^-_{i,k} + \delta_{i,k})^T]$$
$$= \tilde{H}_i\boldsymbol{P}^-_{i,k}\tilde{H}^T_i + \Delta_{i,k} + \tilde{H}_i\Sigma_{\epsilon_i,\delta_i} + \Sigma^T_{\epsilon_i,\delta_i}\tilde{H}^T_i. \quad (27)$$

The fact that $\mathbb{E}[(\widehat{\boldsymbol{x}}^-_{i,k} - \overline{\boldsymbol{x}}_k)\boldsymbol{y}^T_{i,k}] = 0$ was utilized to obtain (25). Now we employ the relations that $\mathbb{E}[\boldsymbol{\epsilon}^-_{i,k}\boldsymbol{v}^T_{j,k}] = $

$0 \ \forall \ j \in \overline{\Omega}_i$, to derive the two covariance quantities $\Sigma_{\epsilon_i,\delta_i}$ and $\Delta_{i,k}$.

$$\Sigma_{\epsilon_i,\delta_i} = \mathbb{E}[\boldsymbol{\epsilon}^-_{i,k}\delta^T_{i,k}]$$
$$= \left[\boldsymbol{0}_{n,p_{j_1}}\cdots\boldsymbol{0}_{n,p_i}\cdots\boldsymbol{0}_{n,p_{j_{|\overline{\Omega}_i|}}} \vdots (\boldsymbol{P}^-_{i,k}-\boldsymbol{P}^-_{ij_1,k})\cdots(\boldsymbol{P}^-_{i,k}-\boldsymbol{P}^-_{ij_{|\Omega_i|},k})\right] \quad (28)$$

$$\Delta_{i,k} = \mathbb{E}[\delta_{i,k}\delta^T_{i,k}]$$
$$= \text{blkdiag}\{\text{blkdiag}\{R_{j,k}\}_{j\in\overline{\Omega}_i}, \left[[\boldsymbol{P}^-_{i,k}-\boldsymbol{P}^-_{ji,k}-\boldsymbol{P}^-_{il,k}+\boldsymbol{P}^-_{jl,k}]_{j,l\in\Omega_i}\right]\} \quad (29)$$

where, blkdiag means a block-diagonal matrix. The only term left is the innovation cross-covariance, $\Sigma_{y_{ij}}$, which can be derived similarly using the steps described above as,

$$\Sigma_{y_{ij}} = \mathbb{E}[\boldsymbol{y}_{i,k}\boldsymbol{y}^T_{j,k}] = \mathbb{E}[(\tilde{H}_i\boldsymbol{\epsilon}^-_{i,k} + \delta_{i,k})(\tilde{H}_j\boldsymbol{\epsilon}^-_{j,k} + \delta_{j,k})^T]$$
$$= \tilde{H}_i\boldsymbol{P}^-_{ij,k}\tilde{H}^T_j + \Delta_{ij,k} + \tilde{H}_i\Sigma_{\epsilon_i,\delta_j} + \Sigma^T_{\epsilon_j,\delta_i}\tilde{H}^T_j, \quad (30)$$
$$\Sigma_{\epsilon_i,\delta_j} = \mathbb{E}[\boldsymbol{\epsilon}^-_{i,k}\delta^T_{j,k}]$$
$$= \left[\boldsymbol{0}_{n,p_{l_1}}\cdots\boldsymbol{0}_{n,p_j}\cdots\boldsymbol{0}_{n,p_{l_{|\overline{\Omega}_j|}}} \vdots (\boldsymbol{P}^-_{i,k}-\boldsymbol{P}^-_{il_1,k})\cdots(\boldsymbol{P}^-_{i,k}-\boldsymbol{P}^-_{il_{|\Omega_j|},k})\right] \quad (31)$$
$$\Delta_{ij,k} = \mathbb{E}[\delta_{i,k}\delta^T_{j,k}] \quad (32)$$
$$= \text{blkdiag}\{\text{blkdiag}\{R_{ql,k}\}_{q\in\overline{\Omega}_i,l\in\overline{\Omega}_j},$$
$$\left[[\boldsymbol{P}^-_{q,k} - \boldsymbol{P}^-_{ql,k} - \boldsymbol{P}^-_{lq,k} + \boldsymbol{P}^-_{l,k}]_{q\in\Omega_i,l\in\Omega_j}\right]\}. \quad (33)$$

Since the observation noises $\boldsymbol{v}_{i,k}$ are uncorrelated across different agents, $R_{ql,k} = R_{q,k}$ when $q = l$, otherwise $R_{ql,k} = \boldsymbol{0}, \ \forall q \neq l$. With $\Sigma_{\epsilon_i,\delta_i}$ and $\Delta_{i,k}$ from (28)-(29), the optimal gain matrices for the distributed estimator at each agent turns into:

$$\mathcal{K}_{i,k} = \left(\boldsymbol{P}^-_{i,k}\tilde{H}^T_i + \Sigma_{\epsilon_i,\delta_i}\right)\left(\tilde{H}_i\boldsymbol{P}^-_{i,k}\tilde{H}^T_i + \Delta_{i,k} + \tilde{H}_i\Sigma_{\epsilon_i,\delta_i}\right.$$
$$\left. + \Sigma^T_{\epsilon_i,\delta_i}\tilde{H}^T_i\right)^{-1} \quad (34)$$

Note that the gain matrices will be very sparse at each agent. To alleviate challenges in tracking of the complete network error covariances, [37] 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.

Combining equations (21) and (22), we get

$$\boldsymbol{P}^+_{i,k} = F\boldsymbol{P}^+_{i,k-1}F^T + Q_k - \mathcal{K}_{i,k}\Sigma^T_{x,y_i} \quad (35)$$

Equation (35), once substituted with (26)-(34) to express in terms of system parameters, will yield a recursive iteration of the filter error covariance matrix which is the distributed version of the discrete algebraic Riccati equation for the proposed distributed estimation algorithm.

Under the distributed observability Assumption 1, the Riccati equation has an asymptotic solution at each agent which is positive definite when started with a symmetric positive semi–definite matrix. This solution, which we will designate by $\boldsymbol{P}^+_{i,\infty}$ is the fixed point of equation (35). For the linear time invariant problems (assuming distributed observability), the steady state filter is asymptotically stable $\rho(F - \mathcal{K}_{i,\infty}\tilde{H}_iF) < 1$, i.e., the closed loop filter matrix $F - \mathcal{K}_{i,\infty}\tilde{H}_iF$ has all

poles inside the unit circle, regardless of F being or not asymptotically stable. To save on the storage burden, the steady state gain matrix $\mathcal{K}_{i,\infty}$ could be used at each agent i for all the iterations. This may not yield distributed estimates with minimum MSE, but will certainly provide estimates with bounded MSE.

VII. Conclusions

The primary contributions of this paper are: (i) providing a novel and meaningful definition for distributed observability; (ii) introducing a new class of distributed state estimation algorithm that treats consensus on neighbors' estimates as innovations; and (iii) designing the gain matrices for the distributed estimator such that the algorithm is optimal, i.e., it yields minimum MSE estimates at all agents. The algorithm, derivations and error analyses presented in this paper resolves most of the challenges related to convergence and optimality of distributed state estimation. The methodologies proposed in this paper have the potential to serve as the backbone for several downstream research challenges including sensor placement in a multi-agent network, adaptation to node or communication failures and other related research problems.

References

[1] Rudolph Emil Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[2] Rudolph E Kalman and Richard S Bucy, "New results in linear filtering and prediction theory," *Journal of basic Engineering*, vol. 83, no. 3, pp. 95–108, 1961.

[3] Chee-Yee Chong, "Forty years of distributed estimation: A review of noteworthy developments," in *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2017, pp. 1–10.

[4] Derui Ding, Qing-Long Han, Zidong Wang, and Xiaohua Ge, "A survey on model-based distributed control and filtering for industrial cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2483–2499, 2019.

[5] Udaya Ghai, Udari Madhushani, Naomi Leonard, and Elad Hazan, "A regret minimization approach to multi-agent contro," *arXiv preprint arXiv:2201.13288*, 2022.

[6] Zhuo Li, Keyou You, and Shiji Song, "Cooperative field prediction and smoothing via covariance intersection," *IEEE Transactions on Signal Processing*, vol. 69, pp. 797–808, 2021.

[7] Jing Yan, Ziqiang Xu, Xiaoyuan Luo, Cailian Chen, and Xinping Guan, "Feedback-based target localization in underwater sensor networks: A multisensor fusion approach," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 1, pp. 168–180, 2018.

[8] Reza Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *44th IEEE Conference on Decision and Control, and 8th European Control Conference*, 2005, pp. 8179–8184.

[9] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *46th IEEE Conference on Decision and Control*, 2007, pp. 5492–5498.

[10] U. A. Khan and José M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4919–4935, 2008.

[11] Ruggero Carli, Alessandro Chiuso, Luca Schenato, and Sandro Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622–633, 2008.

[12] Ioannis D Schizas, Georgios B Giannakis, Stergios I Roumeliotis, and Alejandro Ribeiro, "Consensus in ad hoc WSNs with noisy links - Part II: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, 2008.

[13] Alejandro Ribeiro, Ioannis D Schizas, S Roumeliotis, and Georgios B Giannakis, "Kalman filtering in wireless sensor networks," *IEEE Control Systems Magazine*, vol. 30, no. 2, pp. 66–86, 2010.

[14] David W Casbeer and Randy Beard, "Distributed information filtering using consensus filters," in *American Control Conference*. IEEE, 2009, pp. 1882–1887.

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

[16] Subhro Das and José M. F. Moura, "Consensus+innovations distributed Kalman filter with optimized gains," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 467–481, 2017.

[17] U. A. Khan and A. Jadbabaie, "Coordinated networked estimation strategies using structured systems theory," in *50th IEEE Conference on Decision and Control, and 11th European Control Conference*, 2011, pp. 2112–2117.

[18] M. Doostmohammadian and U. A. Khan, "On the genericity properties in distributed estimation: Topology design and sensor placement," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 195–204, 2013.

[19] Shinkyu Park and Nuno C Martins, "An augmented observer for the distributed estimation problem for LTI systems," in *American Control Conference*, 2012, pp. 6775–6780.

[20] Subhro Das and José M. F. Moura, "Distributed state estimation in multi-agent networks," in *38th IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 4246–4250.

[21] S. Das and José M. F. Moura, "Distributed Kalman filtering," in *21st European Signal Processing Conference*, 2013, pp. 1–5.

[22] Soummya Kar and José M. F. Moura, "Gossip and distributed Kalman filtering: Weak consensus under weak detectability," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1766–1784, 2011.

[23] Di Li, Soummya Kar, José M. F. Moura, H Vincent Poor, and Shuguang Cui, "Distributed Kalman filtering over massive data sets: analysis through large deviations of random Riccati equations," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1351–1372, 2015.

[24] A. G. Dimakis, S. Kar, José M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov 2010.

[25] Matthew D Howard and Zhihua Qu, "An optimal kalman-consensus filter for distributed implementation over a dynamic communication network," *IEEE Access*, vol. 9, pp. 66696–66706, 2021.

[26] Xingkang He, Karl Henrik Johansson, and Haitao Fang, "Distributed design of robust kalman filters over corrupted channels," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2422–2434, 2021.

[27] Anthony J Savas, Vaibhav Srivastava, and Naomi Ehrich Leonard, "On distributed linear filtering with noisy communication," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 2699–2704.

[28] Dan Zhang, Zhenhua Xu, Hamid Reza Karimi, and Qing-Guo Wang, "Distributed filtering for switched linear systems with sensor networks in presence of packet dropouts and quantization," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 10, pp. 2783–2796, 2017.

[29] Subhro Das and José M. F. Moura, "Distributed linear estimationof dynamic random fileds," in *51st Annual Allerton Conference on Communication, Control, and Computing*, 2013, pp. 1120–1125.

[30] S. Das and José M. F. Moura, "Distributed Kalman filtering and network tracking capacity," in *47th Asilomar Conference on Signals, Systems, and Computers*, 2013, pp. 629–633.

[31] R. Olfati-Saber, "Kalman-consensus filter: Optimality, stability, and performance," in *48th IEEE Conference on Decision and Control, and 28th Chinese Control Conference*, 2009, pp. 7036–7042.

[32] Federico S Cattivelli and Ali H Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069–2084, 2010.

[33] U. A. Khan and A. Jadbabaie, "Networked estimation under information constraints," *arXiv preprint arXiv:1111.4580*, 2011.

[34] Fan RK Chung, *Spectral Graph Teory*, vol. 92, American Mathematical Society, 1997.

[35] Subhro Das, "Distributed linear filtering and prediction of time-varying random fields," 2016.

[36] Subhro Das and José MF Moura, "Distributed estimation of dynamic fields over multi-agent networks," *arXiv preprint arXiv:1701.02710*, 2017.

[37] Eduardo Sebastián, Eduardo Montijano, and Carlos Sagüés, "All-in-one: Certifiable optimal distributed kalman filter under unknown correlations," *arXiv preprint arXiv:2105.15061*, 2021.