Optimal discrete-time distributed Kalman filter with reduced communication

Stefano Battilotti, *Senior Member, IEEE*, Alessandro Borri, *Senior Member, IEEE*, Filippo Cacace, *Senior Member, IEEE*, and Massimiliano d'Angelo, *Member, IEEE*

***Abstract*— This paper proposes and analyzes a distributed filter where the consensus term is a virtual output rather than the local state estimate. This feature allows for reducing the data transmitted among nodes at each intermediate step, namely instead of exchanging a vector of the dimension of the state, nodes exchange a vector of the dimension of the rank of the total output matrix. The main finding is that the convergence to the performance of the centralized Kalman filter and mean square boundedness of the estimation error are not lost despite an increase in the number of consensus steps. Simulations show that the total communication overhead is reduced without performance degradation with respect to the original distributed filter, where nodes exchange local state estimates.**

***Index Terms*— Distributed filtering; Network analysis; Stochastic systems.**

I. Introduction

DISTRIBUTED filtering algorithms are an active area of research and one of the main applications of wireless sensor networks (WSNs), consisting of numerous nodes distributed across different geographical locations, each operating under low power constraints and having limited computational capabilities [1]. In the context of distributed estimation for WSNs, each sensor possesses a subset of observations, which can either be transmitted to a central node (the so called fusion center) or shared among nodes to cooperatively perform large-scale sensing tasks that cannot be accomplished by individual devices [2]. The availability of low-cost sensors and the diffusion of wireless networks has contributed in recent years to the development of many applications, such as environmental monitoring [3], airborne target tracking [4], space situation awareness [5], spacecraft navigation [6], among many others. The large majority of these distributed filtering algorithms trace their origins back to the seminal work [7] on the consensus-based paradigm, and they can be broadly classified into three main categories [2]: state estimate fusion [8], [9], measurement vector fusion [10]–[13], and information vector fusion [14]–[17]. Together with the main problem, many research endeavors have tackled the distributed estimation problem under several assumptions and constraints or exploiting spatial correlation to reduce transmission requirements, such as bandwidth and energy constraints [18]–[20], [20], rate constraints [21], quantization [22], [23], reduced-order sensor observations [24].

In this paper, we are concerned with stochastic linear time-invariant discrete-time systems on bi-directional links, and we use a novel approach that mixes state and measurement vector fusion. Algorithms in discrete-time can be further categorized into two classes based on the absence or presence of an inner time scale between two instants of time. This inner cycle, often referred to as "consensus step iterations", is able to recover some desired performance but at the cost of computational complexity and communication time [14], [15], [25]. We investigate the latter case (presence of the inner time scale), wherein the filter structure at a node consists of a prediction with a local correction term (derived from the node's own sensor measurements) and a subsequent filter equation that incorporates the consensus step iterations computed using the (previous) estimates from neighboring nodes. In the recent work [26], it has been theoretically proved that, with a large number of consensus step iterations, the local filters at each node can recover the performance of the centralized optimal filter, namely the Kalman Filter which ideally collects the measurements from all the nodes. In the spirit of [27] (which refers however to complete graphs), an intriguing question arises: can we achieve optimal performance by exchanging "less information" among neighboring nodes during each iteration of the consensus step? This paper focuses on determining whether it is sufficient to exchange a lower-dimensional vector than the entire system state prediction to attain the optimal performance of the (ideal) centralized Kalman filter. The paper answers positively to this question, and addresses the following three contributions:

- We establish that achieving optimal performance only necessitates the exchange of a vector with dimension equal to the rank of the global output matrix (namely the matrix that collects all the output maps of the sensing nodes), which is typically less than the dimension of the entire state.
- We present the design of the output map required to execute the consensus step iterations at a node.
- We exhibit a bound on the number of consensus step

Submitted on xx-xx-2023.

S. Battilotti is with Dipartimento Ingegneria Informatica, Automatica e Gestionale (DIAG), Sapienza Università di Roma, Via Ariosto 25, 00185, Rome, Italy (e-mail: battilotti@diag.uniroma1.it).

A. Borri and M. d'Angelo are with the Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", National Research Council of Italy (CNR-IASI), Via dei Taurini 19, 00185, Rome, Italy, and Via Giosuè Carducci 32, 67100, L'Aquila, Italy (e-mail: {alessandro.borri, massimiliano.dangelo}@iasi.cnr.it).

F. Cacace is with Università Campus Bio-Medico di Roma, Via Álvaro del Portillo, 21, 00128, Rome, Italy (e-mail: f.cacace@ieee.org).

iterations that ensures stability of the estimation error.

The proposed filter not only serves as an alternative to the filter [26], which exchanges state estimates among nodes, to reduce the communication, but it also stands as the sole algorithm feasible in scenarios where exchanging the full state estimates is not possible.

The structure of the paper is outlined as follows. In Section II, we introduce the problem along with some preliminary concepts and assumptions. Section III presents the novel distributed filter with reduced communication. Section IV provides proofs of its capability to attain optimal performance of the centralized Kalman filter arbitrarily close when the consensus step iterations are sufficiently large. Section V presents a numerical example of the proposed method to demonstrate its performance with some comparisons. The concluding remarks are provided in Section VI.

Notation and preliminaries. A Gaussian random variable with mean m and covariance Ψ is indicated as $\text{Gauss}(m, \Psi)$. $\mathbf{1}_N$ denotes a vector with N entries 1. $U_N = \mathbf{1}_N \mathbf{1}_N^\top$ is the square matrix of size N with entries 1. For a square matrix M, $\sigma(M)$ denotes the set of its eigenvalues and $\rho(M)$ its spectral radius, $\rho(M) = \max_{\lambda_i \in \sigma(M)} |\lambda_i|$. When $\rho(M) < 1$ we say that M is Schur. $M_1 > M_2$ (resp. $M_1 \geqslant M_2$) denotes that $M_1 - M_2$ is positive definite (resp. positive semi-definite). Given any matrix A, then $\|A\|$ denotes the matrix operator norm, $\otimes$ is the Kronecker product and $A^{[i]}$ is the i-th Kronecker power. The vectorization or stack operation of an $m \times n$ matrix A is the mn column vector $\text{st}(A) = (a_{1,1}, \ldots, a_{m,1}, a_{1,2}, \ldots, a_{m,n})^\top$. Given $v = (v_1^\top, \ldots, v_n^\top)^\top \in \mathbb{R}^{n^2}$, with $v_i \in \mathbb{R}^n$, $A = \text{st}^{-1}(v) \in \mathbb{R}^{n \times n}$ is the inverse operation. Given matrices M_i, $i = 1, \ldots, n$, of suitable size we define $\text{row}_i(M_i) = [M_1 \ \ldots \ M_n]$, $\text{col}_i(M_i) = (\text{row}_i(M_i^\top))^\top$. Finally, $\text{diag}_i(M_i)$ denotes block diagonal composition.

II. Problem Statement and Assumptions

A. Network preliminaries

We consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the vertices $\mathcal{V} = \{1, 2 \ldots, N\}$ represent the N nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges of the graph. The presence of an edge (i, j) in $\mathcal{G}$ implies that nodes i and j can exchange information between them. The graph is undirected, that is, $(j, i) \in \mathcal{E} \iff (i, j) \in \mathcal{E}$. Two nodes i and j, with $i \neq j$, are neighbors to each other if $(i, j) \in \mathcal{E}$. The set of neighbors of node i is $\mathcal{N}_i := \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$, $\overline{\mathcal{N}}_i = \{i\} \cup \mathcal{N}_i$ is the set of neighbors including i itself. A path is a sequence of connected edges in a graph. A graph is connected if there is a path between every pair of vertices. The symmetric adjacency matrix $\mathcal{A} \in \{0, 1\}^{N \times N}$ of $\mathcal{G}$ has the (i, j)-th entry 1 if $(i, j) \in \mathcal{E}$ and 0 otherwise. The degree matrix $\mathcal{D}$ of $\mathcal{G}$ is a diagonal matrix whose i-th entry is the number of edges of the i-th node. The Laplacian of an undirected $\mathcal{G}$ is the symmetric matrix $\mathcal{L} \in \mathbb{R}^{N \times N}$ defined by $\mathcal{L} = -\mathcal{A} + \mathcal{D}$. When the graph is connected, $0 = \lambda_1(\mathcal{L}) < \lambda_2(\mathcal{L}) \leqslant \cdots \leqslant \lambda_N(\mathcal{L})$, where $\lambda_i(\mathcal{L}) \in \sigma(\mathcal{L})$. An eigenvector associated to $\lambda_1(\mathcal{L})$ is $\mathbf{1}_N$.

B. Dynamical system

We consider a system described by the discrete-time linear time-invariant model

$$x(k+1) = Ax(k) + Fn^s(k), \tag{1}$$

where as usual $x(k) \in \mathbb{R}^n$ is the state and $n^s(k) \in \mathbb{R}^p$ is the state noise such that $n^s(k) \sim \text{Gauss}(0, I_p)$, thus $Q = FF^\top > 0$ is the covariance matrix of the process noise. The distributed estimate of $x(k)$ is obtained through a network modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which the nodes in $\mathcal{S} \subseteq \mathcal{V}$ are sensors with sensing capabilities modeled by the measurement equations

$$y_i(k) = C_i x(k) + G_i n_i(k), \quad i \in \mathcal{S} \subseteq \mathcal{V} \tag{2}$$

where $y_i(k) \in \mathbb{R}^{q_i}$, $n_i(k) \in \mathbb{R}^{n_{q_i}}$, $n_i(k) \sim \text{Gauss}(0, I_{n_{q_i}})$, thus $R_i = G_i G_i^\top \in \mathbb{R}^{q_i \times q_i}$, $R_i > 0$ is the covariance matrix of the measurement noise of node i. As usual, the measurement noises $n_i(k)$ at each time $k \geqslant 0$ of the sensing nodes are mutually independent and independent from $n_s(k)$ and $x(0)$. We denote $C = \text{col}_i(C_i) \in \mathbb{R}^{q \times n}$, where $q = \sum_i q_i$ and $r = \text{rank}(C)$. The matrix $R = \text{diag}_i(R_i) \in \mathbb{R}^{q \times q}$, $R > 0$, is the overall covariance of the measurement errors across the nodes.

Remark 1: We emphasize that, although C is in $\mathbb{R}^{q \times n}$ where q is typically much larger than n, namely $q > n$, still the rank of C is commonly less than n, namely $r < n$.

When R is non singular, the asymptotic estimation error covariance P of the centralized Kalman filter satisfies [26]

$$P = (I_n - PS)\left(APA^\top + Q\right)(I_n - PS)^\top + PSP \tag{3}$$

$$S = C^\top R^{-1} C \tag{4}$$

$$R = \text{diag}_i(R_i), \tag{5}$$

with the gain of the centralized Kalman filter given by

$$K = PC^\top R^{-1}. \tag{6}$$

C. Goal and assumptions

Our aim is to design a distributed algorithm where at each time unit a node can interact with its neighbors an arbitrary but bounded number of times such that the estimation error at each node approximates arbitrarily well the asymptotic variance of the estimation error of the centralized Kalman filter that collects all the measurements $y_i(k)$ of the network. In particular, we present a distributed filter that retains the features of [26] while reducing communication requirements. We have the following assumptions.

Assumption 1: The graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is connected.

Assumption 2: (A, Q) is controllable and (A, C) is observable.

Assumption 2 is the standard global observability and controllability assumption, and we stress that nothing is required on (A, C_i).

Furthermore, we assume that the nodes only know their local information matrices and no global knowledge is required. In particular, we have the following assumption.

Assumption 3: The i-th node knows A, Q, C_i and R_i.

Remark 2: We note that, since $S = \sum_{i \in \mathcal{S}} C_i^\top R_i^{-1} C_i$ (and N) is the sum of the terms $C_i^\top R_i^{-1} C_i$ (respectively 1) across

the graph, S and N can be computed offline at each node by means of distributed algorithms, see [26], [28], [29].
The final assumption is necessary for determining the choice of an output consensus matrix, which plays the pivotal role in reducing the communication workload.

Assumption 4: Each node knows the rank r of the global output matrix C.
It is clear that the latter hypothesis is weaker than assuming the knowledge at each node of the entire C (which is still a plausible assumption in the context of *global information* of the network). Moreover, this assumption is typically applicable in scenarios where the nature of the measurements is already known (*e.g.* in a 2D target tracking problem where only positions in both dimensions are measured we have $r = 2$).

III. Distributed Kalman Filter with Reduced Communication

In this section, we propose a new filter design aimed at reducing communication among nodes. We introduce the filter in Section III-A, and we discuss the design of the critical output consensus matrices $\bar{C}_i$ in Section III-B. As detailed in Remark 3, the key distinction from the filter proposed in [26] lies in the fact that the exchange between neighbors involves the estimates of some *predicted output*, which has the same dimensionality as the output, rather than the estimates of the state, which has the dimension of the system's state.

A. Filter

We denote by $\hat{\mathbf{x}}_i(k)$ the estimate of $x(k)$ and by $\hat{\mathbf{x}}_i(k|k-1)$ the prediction of $x(k)$ at node $i = 1, \ldots, N$ and time k. The reduced-order distributed Kalman filter with $\gamma \in \mathbb{N}$ consensus steps iterates over time at each node $i \in \mathcal{V}$ is the following operations.

1) Local prediction:

$$\hat{\mathbf{x}}_i(k+1|k) = A\hat{\mathbf{x}}_i(k) \tag{7}$$

2) Local correction:

$$\begin{aligned} \xi_i(k+1,0) &= \hat{\mathbf{x}}_i(k+1|k) \\ &\quad + K_i(y_i(k+1) - C_i\hat{\mathbf{x}}_i(k+1|k)) \end{aligned} \tag{8}$$

$$K_i = NPC_i^\top R_i^{-1}, \tag{9}$$

where P solves (3).

3) Consensus iterations. Set $h = 0$ and iterate while $h \leqslant \gamma$:

3.1 Compute

$$\bar{y}_i^j(k+1,h) = \bar{C}_j\xi_i(k+1,h), \tag{10}$$

for $j \in \mathcal{N}_i \cup \{i\}$, where $\bar{C}_j$ is a matrix of rank r that we call *output consensus matrix of node j*, discussed in the next Section III-B, and send it to the neighbor $j \in \mathcal{N}_i$.

3.2 Receive $\bar{y}_j^i(k+1,h)$, with $j \in \mathcal{N}_i$, and compute

$$\begin{aligned} \xi_i(k+1,h+1) &= \xi_i(k+1,h) \\ &\quad + \alpha\bar{K}_i \sum_{j \in \mathcal{N}_i} (\bar{y}_j^i(k+1,h) - \bar{y}_i^i(k+1,h)) \end{aligned} \tag{11}$$

where $\bar{K}_i = PS\bar{C}_i^\top(\bar{C}_i\bar{C}_i^\top)^{-1}$, and α is a design parameter.

3.3 Set $h \leftarrow h+1$.

4) Set $\hat{\mathbf{x}}_i(k+1) = \xi_i(k+1,\gamma)$.

At each node the estimate is initialized as

$$\hat{\mathbf{x}}_i(0) = \mathbb{E}[x(0)] + K_i(y_i(0) - C_i\mathbb{E}[x(0)]), \tag{12}$$

where K_i is defined in (9).

We note that, in order to implement the filter described above, Assumption 3 is required, whereas we additionally assume that Assumption 4 holds if a specific protocol for the selection of the output consensus matrix is implemented (see next Subsection III-B). Also, we remark again that because of Remark 2, the terms N and S (and thus P) are known locally at the nodes.

Remark 3: The proposed filter differs from the distributed Kalman filter of [26] at step 3), eqs. (10) and (11). In the distributed Kalman filter of [26], the exchange among neighbors involves the estimates $\xi_i(h)$ (dimension n), in (10)-(11) $\bar{y}_i^j(k+1,h)$ (dimension r). This results in reduced communication requirements at each consensus iteration with respect to filters that exchange the whole state estimate like [26]. The reduced order filter is still able to recover the optimal performance of the centralized Kalman filter as proved in Section IV.

We clarify the trade-off between the amount of information exchanged and estimation accuracy in Proposition 2 of Section IV-A. The critical choice of $\bar{C}_i$ in (10) and the corresponding gain $\bar{K}_i$ in (11) is discussed in the next section.

B. Choice of the output consensus matrix

In this section we explain how to choose the crucial output consensus matrix $\bar{C}_i$ of Eq. (10) at the i-th node. We present two protocols for the selection of $\bar{C}_i$ that yield the *same* filter.

1) Homogeneous selection: Since $C \in \mathbb{R}^{q\times n}$ and recalling that $\text{rank}(C) = r$, we can represent the matrix C without loss of generality as

$$C = \begin{bmatrix} \bar{C} \\ M_{\bar{C}}\bar{C} \end{bmatrix}$$

where $\text{rank}(\bar{C}) = r$, $\bar{C} \in \mathbb{R}^{r\times n}$ and some $M_{\bar{C}} \in \mathbb{R}^{(q-r)\times r}$. For each $i \in \mathcal{V}$, the choice of the output consensus matrix is

$$\bar{C}_i := \bar{C}, \tag{13}$$

where $\bar{C}$ is obtained through the following algorithm.

Algorithm for homogeneous selection

Each node $i \in \mathcal{V}$ sets $\mathcal{I}_i(0) = C_i$ and

1. at time $t = 0$, sends $\mathcal{I}_i(0)$ to the neighbors $j \in \mathcal{N}_i$;
2. at time $0 < t \leqslant N-1$, receives $\mathcal{I}_j(t-1)$ from the neighbors $j \in \mathcal{N}_i$ and sends $\mathcal{I}_i(t) = \text{st}_{j\in\bar{\mathcal{N}}_i}(\mathcal{I}_j(t-1))$ to the neighbors $j \in \mathcal{N}_i$;
3. at time $t = N$, removes redundant rows from $\mathcal{I}_i(t-1)$, sorts the rows and defines $\bar{C}$ as the first r independent rows extracted from the resulting matrix.

With this choice, $\bar{K}_i = \bar{K} = PS\bar{C}^\top(\bar{C}\bar{C}^\top)^{-1}$. Moreover,

$$\begin{aligned}\bar{K}\bar{C} &= PS\bar{C}^\top(\bar{C}\bar{C}^\top)^{-1}\bar{C} = KC\bar{C}^\top(\bar{C}\bar{C}^\top)^{-1}\bar{C} \\ &= K\begin{bmatrix} \bar{C} \\ M_{\bar{C}}\bar{C}\end{bmatrix}\bar{C}^\top(\bar{C}\bar{C}^\top)^{-1}\bar{C} = KC. \end{aligned} \tag{14}$$

In essence, the protocol involves exchanging $\mathcal{I}_i(0)$ throughout the network to calculate a common $\bar{C}$ (a form of row-wise ordering is necessary). This protocol requires N pre-processing steps and a computational storage capacity to retain the values of $\mathcal{I}_i(N-1)$. This protocol does not need Assumption 4.

2) Heterogeneous selection: For each $i \in \mathcal{V}$, the output consensus matrix is $\bar{C}_i$ and it is computed through the algorithm for heterogeneous selection below.

Algorithm for heterogeneous selection

Each node $i \in \mathcal{V}$ sets $\mathcal{I}_i(0) = C_i$ and

1. at time $t = 0$, sends $\mathcal{I}_i(0)$ to the neighbors $j \in \mathcal{N}_i$;
2. at time $0 < t \leqslant N-1$, while $\text{rank}(\mathcal{I}_i(t-1)) < r$, sends $\mathcal{I}_i(t) = \text{st}_{j\in\bar{\mathcal{N}}_i}(\mathcal{I}_j(t-1))$ to the neighbors $j \in \mathcal{N}_i$;
3. defines $\bar{C}_i$ as the first r independent rows extracted from $\mathcal{I}_i(t)$;
4. sends $\bar{C}_i$ to the neighbors $j \in \mathcal{N}_i$.

The protocol involves exchanging the matrices C_i throughout the network until a matrix $\bar{C}_i \in \mathbb{R}^{r\times n}$ of rank r is obtained. With this protocol, each node has a distinct $\bar{C}_i$, which is communicated to the neighbors. This protocol does not rely on any row-wise ordering and the computational storage to retain the values of $\mathcal{I}_i(t-1)$ is less than the previous protocol, but it needs Assumption 4 for the stopping condition.

It is worth emphasizing again that the choice between the two protocols for selecting the output consensus matrix has a direct impact on the real-world implementation. This decision involves a trade-off between computational pre-processing power and the flexibility of transmissions, as discussed above. However, from a theoretical standpoint, both protocols yield the same algorithm as shown in the next proposition.

Proposition 1: The estimate provided by the distributed Kalman filter with reduced communication of Section III-A is identical, regardless of whether the output consensus matrix $\bar{C}_i$ for each node $i \in \mathcal{V}$ is determined with homogeneous or heterogeneous selection.

Proof: The difference between the two possible selections of $\bar{C}_i$ is in the consensus term of equation (11), namely

$$\Delta = \bar{K}_i\sum_{j\in\mathcal{N}_i}(\bar{y}^i_j(k+1,h) - \bar{y}^i_i(k+1,h)). \tag{15}$$

It is important to note that, regardless of the chosen protocol, Δ is the same. In fact, with homogeneous selection we have

$$\begin{aligned}\Delta_{\text{ho}} &= \bar{K}_i\sum_{j\in\mathcal{N}_i}(\bar{y}^i_j(k+1,h) - \bar{y}^i_i(k+1,h)) \\ &= \bar{K}\bar{C}\sum_{j\in\mathcal{N}_i}(\xi_j(k+1,h) - \xi_i(k+1,h)), \end{aligned} \tag{16}$$

where $\bar{K} = PS\bar{C}^\top(\bar{C}\bar{C}^\top)^{-1}$. With heterogenous selection we have

$$\Delta_{\text{he}} = \bar{K}_i\bar{C}_i\sum_{j\in\mathcal{N}_i}(\xi_j(k+1,h) - \xi_i(k+1,h)), \tag{17}$$

where $\bar{K}_i = PS\bar{C}_i^\top(\bar{C}_i\bar{C}_i^\top)^{-1}$. By the same argument as in Section III-B.1, $\bar{K}\bar{C} = \bar{K}_i\bar{C}_i = KC$ and thus $\Delta_{\text{ho}} = \Delta_{\text{he}}$. ■

IV. Stability and Optimality Analysis

In this section we prove that the distributed Kalman filter (7)–(12) approximates the centralized Kalman filter when increasing the number γ of consensus iterations and we provide a lower bound on γ for the stability of the estimation error.

Lemma 1: Let $\varepsilon(k) = \text{col}_i(x(k) - \hat{\mathbf{x}}_i(k))$ be the overall estimation error of (7)–(12). The overall error covariance matrix $P_\gamma(k) = \mathbb{E}[\varepsilon(k)\varepsilon^\top(k)]$ evolves with

$$P_\gamma(k+1) = \Theta^\gamma\Big(\text{diag}_i(A_i)P_\gamma(k)\text{diag}_i(A_i^\top) + \Psi_M\Big)\Theta^{\gamma\top}, \tag{18}$$

$$\Theta = I_{nN} - \alpha\mathcal{L}\otimes KC \tag{19}$$

$$A_i = (I_n - K_iC_i)A \tag{20}$$

$$\tilde{D} = \text{diag}_i(I_n - K_iC_i) \tag{21}$$

$$\Psi_M = \tilde{D}(U_N\otimes Q)\tilde{D}^\top + N^2\text{diag}_i(PS_iP) \tag{22}$$

Proof. With Proposition 1 in mind, by defining $\varepsilon_i(k) = x(k) - \hat{\mathbf{x}}_i(k)$ and the intermediate estimation error $\varepsilon_i(k,h) = x(k) - \xi_i(k,h)$, we have

$$\begin{aligned}&\varepsilon_i(k+1,0) = x(k+1) - \xi_i(k+1,0) = A\varepsilon_i(k) \\ &+ Fn^s(k) - K_i(C_i(x(k+1) - A\hat{\mathbf{x}}_i(k)) + G_in_i(k)) \\ &= (I_n - K_iC_i)(A\varepsilon_i(k) + Fn^s(k)) - K_iG_in_i(k+1), \end{aligned} \tag{23}$$

$$\bar{y}_i(k+1,h) = \bar{C}_i(x(k+1) - \varepsilon_i(k+1,h)). \tag{24}$$

By using (24) and letting $d_i = |\mathcal{N}_i|$, (11) becomes

$$\begin{aligned}\xi_i(k+1,h+1) &= (I_n - \alpha d_iKC)\,\xi_i(k+1,h) \\ &+ \alpha d_iKCx(k+1) - \alpha KC\sum_{j\in\mathcal{N}_i}\varepsilon_j(k+1,h), \end{aligned} \tag{25}$$

$$\begin{aligned}\varepsilon_i(k+1,h+1) &= (I_n - \alpha d_iKC)\,\varepsilon_i(k+1,h) \\ &+ \alpha KC\sum_{j\in\mathcal{N}_i}\varepsilon_j(k+1,h). \end{aligned} \tag{26}$$

The overall intermediate estimation error $\varepsilon(k,h) = \text{col}_i(\varepsilon_i(k,h))$ is transformed at each consensus step as

$$\begin{aligned}\varepsilon(k+1,h+1) &= \text{diag}_i\,(I_n - \alpha d_iKC)\,\varepsilon(k+1,h) \\ &\quad + (\alpha\mathcal{A}\otimes KC)\varepsilon(k+1,h) \\ &= (I_{nN} - \alpha D\otimes KC + \alpha\mathcal{A}\otimes KC)\,\varepsilon(k+1,h) \\ &= \Theta\,\varepsilon(k+1,h). \end{aligned} \tag{27}$$

From (23) and (27) we obtain

$$\begin{aligned}\varepsilon(k+1) = \Theta^\gamma\Big(&\text{diag}_i(A_i)\varepsilon(k) - \text{col}_i((I_n - K_iC_i)Fn^s(k)) \\ &- \text{col}_i(K_iG_in_i(k+1))\Big). \end{aligned} \tag{28}$$

Consequently, $P_\gamma(k) = \mathbb{E}[\varepsilon(k)\varepsilon^\top(k)]$ evolves with

$$\begin{aligned}P_\gamma(k+1) &= \Theta^\gamma \mathrm{diag}_i(A_i) P_\gamma(k) \mathrm{diag}_i(A_i^\top)\Theta^{\gamma\top} \\ &+ \Theta^\gamma \mathrm{col}_i(I_n - K_iC_i) Q \mathrm{row}_i(I_n - K_iC_i)^\top \Theta^{\gamma\top} \\ &+ \Theta^\gamma \mathrm{diag}_i(K_iR_iK_i^\top)\Theta^{\gamma\top} \\ &= \Theta^\gamma\Big(\mathrm{diag}_i(A_i)P_\gamma(k)\mathrm{diag}_i(A_i^\top) + \Psi_M\Big)\Theta^{\gamma\top}. \end{aligned} \tag{29}$$

We need some preliminary results, proved in Appendix I-A and Appendix I-B.

Lemma 2: If Assumption 2 holds, then

$$\rho(I_n - PS) = 1. \tag{30}$$

Lemma 3: Let Assumption 1 and 2 hold and $\alpha\lambda_k < 1$. Let r be the rank of C and w_j, $v_j^\top$ respectively the right and left eigenvectors of $I_n - PS$ associated to the eigenvalue $\mu_1 = 1$. Then, the following limit holds

$$\lim_{\gamma\to\infty} \Theta^\gamma = \Pi_\lambda + \Pi_\mu, \tag{31}$$

where

$$\Pi_\lambda := \frac{1}{N}(U_N \otimes I_n)$$

$$\Pi_\mu := \left(I_N - \frac{1}{N}U_N\right)\otimes M, \quad M = \sum_{j=1}^{n-r} w_j v_j^\top \in \mathbb{R}^{n\times n}.$$

We note that Π_λ and Π_μ are orthogonal projectors, and in particular, when the rank of C is n then $M = 0$.

A. Optimality result

In this section we establish the fact that the distributed filter III-A recovers the performance of the centralized Kalman filter, namely the (ideal) filter that aggregates all measurements, given that the number γ of the consensus step iterations is sufficiently large.

Theorem 1: If Assumptions 1 and 2 hold and the gain α in (11) is such that $\alpha\lambda_k < 1$ $\forall \lambda_k \in \sigma(\mathcal{L})$, then the estimation error covariance at each node of the filter (7)–(12) tends to the covariance of the centralized Kalman filter as $\gamma \to \infty$:

Proof. Let $P_\infty(k)$ be the covariance of the estimation error for $\gamma \to \infty$. Equation (18) becomes

$$\begin{aligned}P_\infty(k+1) &= (\Pi_\lambda + \Pi_\mu)\,\mathrm{diag}_i(A_i)P_\infty(k)\mathrm{diag}_i(A_i^\top)\,(\Pi_\lambda+\Pi_\mu)^\top \\ &+ \tilde{Q} \end{aligned}\tag{32}$$

$$\begin{aligned}\tilde{Q} &= (\Pi_\lambda + \Pi_\mu)\left(\tilde{D}(U_N\otimes Q)\tilde{D}^\top \right.\\ &\left. + \mathrm{diag}_i(N^2 PS_iP)\right)(\Pi_\lambda+\Pi_\mu)^\top. \end{aligned}\tag{33}$$

The proof develops along the following steps.

Step 1. Based on the properties $MPS_i = 0$, and $S_iM = 0$, we show that

$$\tilde{Q} = U_N \otimes ((I_n - PS)Q(I_n - PS)^\top + PSP), \tag{34}$$

i.e., all the blocks of $\tilde{Q}$ are identical. This is easily proved by showing that the rows and columns of Π_μ sum up to 0 and the following properties hold:

$$\Pi_\mu \mathrm{diag}_i(A_i) = \Pi_\mu(I_N \otimes A) \tag{35}$$

$$\Pi_\mu \tilde{D} = \Pi_\mu \tag{36}$$

$$\Pi_\mu(U_N\otimes Q) = 0 \tag{37}$$

$$\Pi_\mu \mathrm{diag}_i(N^2PS_iP) = 0. \tag{38}$$

Step 2. Let $\tilde{A} = (\Pi_\lambda + \Pi_\mu)\mathrm{diag}_i(A_i)$. Eq. (32) is rewritten

$$P_\infty(k+1) = \tilde{A}P_\infty(k)\tilde{A}^\top + \tilde{Q}. \tag{39}$$

The next step is to prove that $\rho(\tilde{A}) < 1$. This guarantees that $\lim_{k\to\infty} P_\infty(k)$ exists. In fact, with $X(k) = \mathrm{st}(P_\infty(k))$ one obtains

$$X(k+1) = \tilde{A}^{[2]}X(k) + \mathrm{st}\,\tilde{Q}, \tag{40}$$

where $\rho(\tilde{A}^{[2]}) < 1$ whenever $\rho(\tilde{A}) < 1$, and the asymptotic value is

$$\lim_{k\to\infty} P_\infty(k) = \mathrm{st}^{-1}\left(\left(I_{(nN)^2} - \tilde{A}^{[2]}\right)^{-1}\mathrm{st}\,\tilde{Q}\right). \tag{41}$$

In order to prove $\rho(\tilde{A}^{[2]}) < 1$ we show that $\sigma(\tilde{A}) = \sigma(\bar{A}) \cup \sigma(M\bar{A})$, where

$$\bar{A} = (I_n - KC)A = (I_n - PS)A = \frac{1}{N}\sum_i A_i. \tag{42}$$

is Schur. Since U_N is symmetric there is an orthonormal basis $\{u_k\}$ of $\mathbb{R}^N$ of right eigenvectors of U_N, $k = 1,\dots,N$. Let $u_1 = \frac{1}{\sqrt{N}}\mathbf{1}_N$. Direct computation yields

$$\tilde{A} = \frac{1}{\sqrt{N}}u_1\otimes \mathrm{row}_i(A_i) + \left(I_N - \frac{1}{N}U_N\right)\otimes(M\bar{A}).$$

Let $\bar{\lambda} \in \sigma(\bar{A})$, $\bar{A}\zeta_j = \bar{\lambda}_j\zeta_j$. Since $(I_N - \frac{1}{N}U_M)u_1 = (I_N - u_1u_1^\top)u_1 = 0$,

$$\begin{aligned}\tilde{A}(u_1\otimes\zeta_j) &= \frac{1}{\sqrt{N}}(u_1\otimes \mathrm{row}_i(A_i))(u_1\otimes\zeta_j) \\ &= u_1 \otimes \left(\frac{1}{N}\sum_{i=1}^N A_i\zeta_j\right) \\ &= u_1\otimes\bar{A}\zeta_j = \bar{\lambda}_j(u_1\otimes\zeta_j), \end{aligned}\tag{43}$$

that proves $\sigma(\bar{A}) \subset \sigma(\tilde{A})$. The remaining $(N-1)n$ eigenvalues of $\tilde{A}$ are exactly the eigenvalues of $M\bar{A}$, each with algebraic and geometric multiplicity $N-1$. To see this, let $\tilde{\lambda}_j \in \sigma(M\bar{A})$, $\xi_j^\top M\bar{A} = \tilde{\lambda}_j\xi_j^\top$. Since $u_k^\top u_1 = 0$ and $u_k^\top(I_N - \frac{1}{N}U_N) = u_k^\top$, for $k = 2,\dots,N$,

$$(u_k^\top\otimes\xi_j^\top)\tilde{A} = u_k^\top\otimes\xi_j^\top M\bar{A} = \tilde{\lambda}_j(u_k^\top\otimes\xi_k^\top). \tag{44}$$

Since $\rho(\bar{A}) < 1$ (Lemma 2) we are left to prove $\rho(M\bar{A}) < 1$. This can be proved by a Lyapunov argument with the functional $V(k) = x^\top(k)Px(k)$ on the auxiliary dynamic system $x(k+1) = (M\bar{A})^\top x(k)$. By re-writing the Riccati equation (3) as

$$\bar{A}P\bar{A}^\top - P = -(I_n - PS)Q(I_n - PS)^\top - PSP, \tag{45}$$

and the property $M(I_n - PS) = M$, we obtain

$$\begin{aligned}V(k+1) - V(k) &= x^\top(k)(M\bar{A}P\bar{A}^\top M^\top - P)x(k) \\ &= x^\top(k)(-P + MPM^\top - MQM^\top)x(k). \end{aligned}\tag{46}$$

We now prove that the matrix on the right-hand side is negative definite. Define

$$Z = I_n - PS - M = \sum_{j=n-r+1}^{n}\lambda_j w_j v_j^\top. \tag{47}$$

Clearly, $MZ = 0$ and since $\rho(I_n - PS) = 1$ (Lemma 2) and the eigenvalue 1 is in M, $\rho(Z) < 1$. Recalling that $SPM^\top = 0$ we have

$$\begin{aligned} ZPM^\top &= P(I_n - SP)M^\top - MPM^\top \\ &= PM^\top - MPM^\top \qquad (48) \\ Z^2PM^\top &= ZPM^\top - ZMPM^\top = ZPM^\top \qquad (49) \\ ZPM^\top - Z^2PM^\top &= (I_n - Z)ZPM^\top = 0. \qquad (50) \end{aligned}$$

Since $\rho(Z) < 1$, $(I_n - Z)$ is non-singular and we conclude that $ZPM^\top = 0$ and $PM^\top = MPM^\top = MP$. Consequently, if $\Delta_k V = V(k+1) - V(k)$,

$$\begin{aligned} P - MPM^\top &= P - MP - PM^\top + MPM^\top \\ &= (I_n - M)P(I_n - M)^\top \qquad (51) \\ \Delta_k V &= -x^\top(k)((I_n - M)P(I_n - M)^\top \\ &\quad + MQM^\top)x(k), \qquad (52) \end{aligned}$$

where it is easy to check that the sum of the two semi-definite positive matrices on the right-hand side yields a positive definite matrix. This proves $\rho(M\tilde{A}) < 1$ and consequently $\rho(\tilde{A}) < 1$ and the existence of $\lim_{k\to\infty} P_\infty(k)$.
Step 3. The final step is to prove that $U_N \otimes P$ is the asymptotic value of (39) and it satisfies (41). This is easily verified, since

$$\begin{aligned} &\tilde{A}(U_N \otimes P)\tilde{A}^\top + \tilde{Q} = \Pi_\lambda \text{diag}_i(A_i)(U_N \otimes P)\text{diag}_i(A_i^\top)\Pi_\lambda \\ &\quad + \Pi_\lambda \text{diag}_i(A_i)(U_N \otimes P)(I_N \otimes A^\top)\Pi_\mu^\top \\ &\quad + \Pi_\mu(I_N \otimes A)(U_N \otimes P)\text{diag}_i(A_i)\Pi_\lambda \\ &\quad + \Pi_\mu(I_N \otimes A)(U_N \otimes P)(I_N \otimes A^\top)\Pi_\mu^\top + \tilde{Q} \\ &= \Pi_\lambda \text{diag}_i(A_i)(U_N \otimes P)\text{diag}_i(A_i^\top)\Pi_\lambda + \tilde{Q} \\ &= U_N \otimes \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_i P A_j^\top\right) + \tilde{Q} \\ &= U_N \otimes \big((I_n - PS)(APA^\top + Q)(I_n - PS)^\top \\ &\quad + Q + PSP\big) = U_N \otimes P, \qquad (53) \end{aligned}$$

where we have used the property

$$\Pi_\mu(I_N \otimes A)(U_N \otimes P) = \Pi_\mu(U_N \otimes AP) = 0. \qquad (54)$$

□

Remark 4: Due to Assumption 1, $\lambda_k \geqslant 0$, thus $\alpha\lambda_k \geqslant 0$ for all positive α.

B. Bound on the number γ of consensus steps

We provide a lower bound for γ that ensures mean square boundedness of the estimation error.

Theorem 2: If Assumptions 1 and 2 hold and α is such that $\alpha\lambda_k < 1$ for all $\lambda_k \in \sigma(\mathcal{L})$, then the estimation error of the filter (7)–(12) is asymptotically unbiased in time and mean square bounded if and only if γ is such that $\rho(\Theta^\gamma \text{diag}_i(A_i)) < 1$, where Θ, A_i are defined in (19)–(20). A sufficient lower bound is any γ such that

$$\theta^\gamma \leqslant \frac{(1-\sqrt{\mu})^2}{(1+N\|P\|\|P^{\frac{1}{2}}SP^{-\frac{1}{2}}\|)^2(\|P^{\frac{1}{2}}AP^{-\frac{1}{2}}\|)^2}, \qquad (55)$$

with P solution of (3), any $\mu \in (0,1)$ such that $\bar{A}P\bar{A}^\top \leqslant \mu P$, $\bar{A}$ as in (42) and $\theta := 1 - 2\alpha(1 - \cos(\pi/N))$. Moreover in these cases,

$$\begin{aligned} &\lim_{k\to\infty} P_\gamma(k) = P_\gamma^\infty = \\ &\text{st}^{-1}\left(\left(I_{n^2N^2} - (\Theta^\gamma \text{diag}_i(A_i))^{[2]}\right)^{-1} \Theta^{\gamma[2]}\,\text{st}(\Psi_M)\right) \qquad (56) \end{aligned}$$

with Ψ_M defined in (22).
The proof of (56) is easily obtained from (18) by using the same transformation as in (41). The proof of (55) is similar to the proof of Theorem 5 of [30] (see Section IV-C) by taking into account also Lemma 2 and it exploits the bound $\lambda_2(\mathcal{L}) \geqslant 2(1 - \cos(\pi/N))$ [31] that ensures $|\theta| < 1$ whenever $\alpha\lambda_k < 1$. P, S, N and A are locally known, then (55) can be computed at each node. We emphasize that the hypothesis $\forall k: \ \alpha\lambda_k < 1$ is crucial for the mean square boundedness of the estimation error. Finally, the next proposition provides conditions under which the algorithm proposed in this paper reduces the communication burden with respect to [26]. The proof is immediate.

Proposition 2: Let $P_{\gamma_f}^\infty$ be the covariance of the estimation error of the filter in [26] with γ_f consensus steps, that can be obtained from (56) by replacing Θ with $\Theta_f = (I_N - \alpha\mathcal{L}) \otimes I_n$ and $\gamma = \gamma_f$. Let γ_r be the smallest integer such that $\text{tr}(P_{\gamma_r}^\infty) \leqslant \text{tr}(P_{\gamma_f}^\infty)$. If $\gamma_r \leqslant \gamma_f \frac{n}{r}$ then the filter (7)–(12) has mean square error not larger than the one in [26] with reduced communication.

Remark 5: Eq. (56) is also useful to assess the robustness of the filter to permanent link failures (changes in the underlying graph) that leave the graph connected. In this case the eigenvalues of $\mathcal{L}$ may change, and in general $\text{tr}(P_{\gamma_r}^\infty)$ increases when $\lambda_2(\mathcal{L})$ decreases.

V. NUMERICAL EXAMPLE

Consider the network of Fig. 1 consisting of a connected graph with $N = 5$ nodes and 5 edges, where nodes 1 and 5 have measurement matrices C_1 and C_5 respectively, while the remaining nodes have communication capabilities only ($q_2 = q_3 = q_4 = 0$). The system in the form of (1)–(2) has

$$A = \begin{pmatrix} -1.15 & 0.65 & -0.1 & -0.75 & 1.35 \\ -0.65 & 0.85 & -0.7 & -0.65 & 0.65 \\ -0.85 & 0.15 & 0.2 & -0.85 & 0.85 \\ 1.45 & -0.75 & -1.8 & -0.15 & -0.25 \\ 0.45 & -0.45 & -1.2 & -0.75 & 0.95 \end{pmatrix}$$

$$\begin{aligned} C_1 &= \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad C_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\ Q &= FF^\top = \text{diag}(1, 0.5, 0.7, 0.3, 0.4) \\ R &= \text{diag}(C_1C_1^\top, C_5C_5^\top) = \text{diag}(0.32, 0.94). \end{aligned}$$

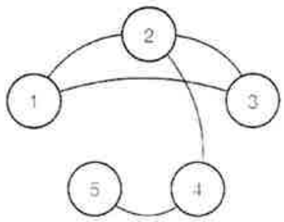

Fig. 1. Network topology of the example.

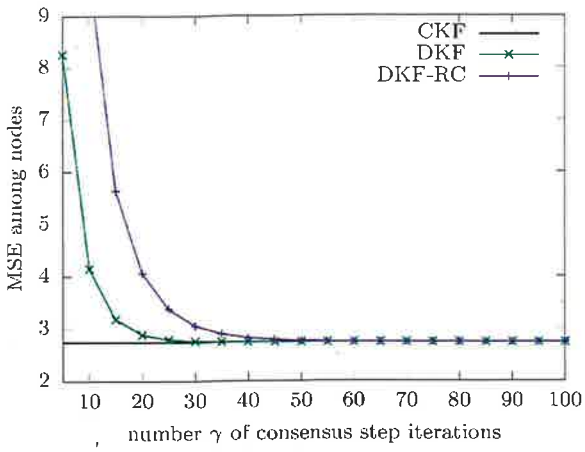

Fig. 2. Performance in terms of MSE for the distributed Kalman filter with reduced communication (DKF-RC) alongside the previous distributed Kalman filter of [26] (DKF), together with the ideal optimal performance of the centralized Kalman filter (CKF).

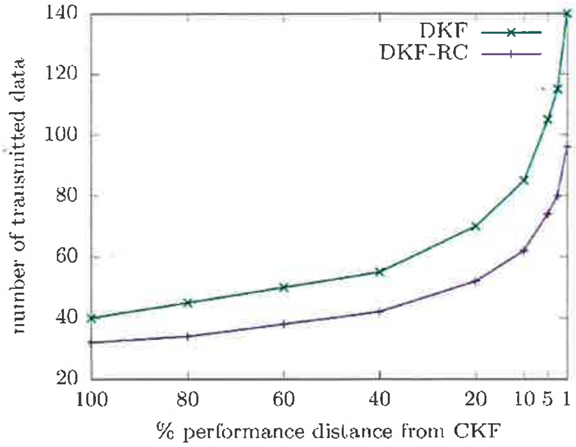

Fig. 3. Amount of communication data transmitted by a node at each time $k \geqslant 0$ necessary to achieve a specific percentage distance from the performance of the centralized Kalman filter (CKF).

Note that the system is globally observable, that is, the couple (A, C) with $C = \text{col}(C_1, C_5)$ is observable, but the pairs (A, C_1) and (A, C_5) are not individually observable. We compute the mean square error (MSE) of (7)–(12) as the norm of the estimation error averaged over all the nodes and times (avoiding transient effects), over a time horizon $T = 200$ points and for several values of the consensus step iterations γ. We note that in this example the *output consensus matrix* (described in Section III-B) is $C \in \mathbb{R}^{2\times 5}$, thus the exchanged information among the nodes $\bar{y}_i^j$ is in $\mathbb{R}^2$ while the information exchange by the filter in [26] is the estimate $\hat{\mathbf{x}}_i \in \mathbb{R}^5$. Figure 2 shows the results of Theorem 2 highlighting the mean square boundedness with finite γ and comparing the MSE of (7)–(12) with the filter in [26] when varying γ. We note that in this example the necessary and sufficient condition in order to obtain mean square boundedness of the proposed filter is $\gamma = 3$, while for the filter in [26] is $\gamma = 2$. Also, the sufficient condition (55) for both filters is $\gamma = 310$. It is evident that, albeit at a slower rate, the distributed filter with reduced communication recovers the performance of the centralized optimal filter as gamma increases. This is reasonable since the distributed filter with reduced communication exchanges "less" information at each instant of the consensus step iterations, hence necessitating a higher number γ of consensus step iterations to attain the desired level of performance. Figure 3 illustrates the results of Proposition 2. In particular, it shows the amount of data transmitted[1] by a node at each time $k \geqslant 0$ necessary to achieve a specific percentage distance from the performance of the centralized Kalman filter (CKF). This distance is calculated as $(\text{MSE}_{\text{DKF}} - \text{MSE}_{\text{CKF}})/\text{MSE}_{\text{CKF}}$. For example, when targeting a 5% offset from the performance of the centralized Kalman filter, the distributed Kalman filter of [26] transmits 105 data, while the distributed Kalman filter with reduced communication needs 74 transmitted data.

[1]the amount of communication data is computed as $\gamma \cdot \text{card}(v)$, where $\text{card}(v)$ is the cardinality of the vector v transmitted at each consensus step ($v = \hat{\mathbf{x}}_i$ in the case of [26], $v = \bar{y}_i^j$ in the proposed filter).

VI. Conclusions

The results presented in this paper are useful to achieve mean square stability of the estimation error for a distributed filter with bandwidth limitations for the nodes. Another interesting application is to cooperative filtering of systems with infinite-dimensional state, since in that case the estimates of the state are hard to exchange. It is also of interest to extend these results to time-varying networks. Another topic that deserves further investigation is the optimal choice of the output consensus matrix with respect to the trade-off between exchanged information and performance of the filter.

Appendix I
Complementary Proofs

A. Proof of Lemma 2

Since Assumption 2 holds, there exists $P = P^\top > 0$ that satisfies the Riccati equation (3). From the definition $S = C^\top R^{-1} C$ in (4) it is immediate to see that $I_n - KC = I_n - PS$. In order to prove $\rho(I_n - PS) = 1$ we prove that $z(k+1) = (I_n - PS)z(k)$ is simply stable. With the Lyapunov function $V(k) = z^\top(k)P^{-1}z(k)$ we obtain $V(k+1) - V(k) = -z^\top(2S - SPS)P)z(k)$, hence the second point of the Lemma is proved by showing that $S \geqslant SPS$. From the Riccati equation (3) it follows that $P - PSP \geqslant 0$ that implies $I_n - P^{\frac{1}{2}}SP^{\frac{1}{2}} \geqslant 0$. Let $S_1 = C^\top R^{-\frac{1}{2}}$, $S = S_1 S_1^\top$. Since

$$\sigma(I_n - P^{\frac{1}{2}}S_1 S_1^\top P^{\frac{1}{2}}) \supseteq \sigma(I_q - S_1^\top P S_1) \tag{57}$$

it follows that $I_q - S_1^\top P S_1 \geqslant 0$. Pre- and post-multiplying by S_1 and $S_1^\top$ respectively, we obtain $S - SPS \geqslant 0$. □

B. Proof of Lemma 3

Let w_j and $v_j^\top$ be the right and left eigenvectors of $I_n - PS$ with corresponding eigenvalue μ_j where S is defined in (4). It is always possible to normalize these eigenvectors so that $\|w_j\| = 1$, $v_j^\top w_j = 1$ and $v_j^\top w_i = 0$ when $i \neq j$. From Lemma 2 it follows $|\mu_j| \leqslant 1$. When C is not full rank, $r < n$, without loss of generality let us number $w_1, \ldots, w_{n-r}$ the eigenvectors associated to $\mu_1 = 1$, and similarly for $v_j^\top$. Let λ_k, $k = 1, \dots, N$ be the eigenvalues of $\mathcal{L}$. $\mathcal{L}$ is symmetric, and we denote u_k the corresponding left and right normalized eigenvectors, with $\lambda_1 = 1$, $u_1 = \frac{1}{\sqrt{N}}\mathbf{1}_N$. w_j, $v_j^\top$ are the right and left eigenvectors of both $I_n - PS$ and $I_n - \alpha\lambda_k PS$,

$$(I_n - \alpha\lambda_k PS)w_j = (1 + (\mu_j - 1)\alpha\lambda_k)w_j = \mu_{k,j}w_j, \quad (58)$$

where, again $|\mu_{k,j}| \leqslant 1$ because $\alpha\lambda_k < 1$. When $\mu_1 = 1$, $\mu_{k_j} = 1$, that is, $\rho(I_n - \alpha\lambda_k PS) = 1$. Recalling that $KC = PS$ we obtain

$$\begin{aligned}\Theta(u_k \otimes w_j) =& (I_{nN} - \alpha\mathcal{L} \otimes PS)(u_k \otimes w_j)\\ =& u_k \otimes w_j - \alpha\lambda_k u_k \otimes PSw_j\\ =& u_k \otimes (I_n - \alpha\lambda_k PS)w_j = \mu_{k,j}(u_k \otimes w_{k,j}).\end{aligned}$$

This proves that $\sigma(\Theta) = \cup_{k=1}^N \sigma(I_n - \alpha\lambda_k PS)$ and $\rho(\Theta) = 1$. Since all the remaining eigenvalues of Θ are inside the unit circle, $\lim_{\gamma\to\infty}\Theta^\gamma$ exists and can be represented as the projection of the subspace associated to the eigenvalue 1. This subspace is generated by both $\lambda_1 = 0$ and the unitary eigenvalues of $I_n - \alpha\lambda_k PS$ for $\lambda_k \neq 0$, that correspond to $\mu_1 = 1$. With $\lambda_1 = 0$ we obtain that $1 \in \sigma(\Theta)$ with multiplicity n and the corresponding eigenvectors can be represented as $u_1 \otimes e_j$, where e_j are the versors of the Euclidean basis of $\mathbb{R}^n$. Thus, the projector on the autospace corresponding to $\lambda_1 = 0$ is

$$\Pi_\lambda := (u_1 \otimes I_n)(u_1^\top \otimes I_n) = \frac{1}{N}(U_N \otimes I_n). \quad (59)$$

The projector on the autospace of the unitary eigenvalue of Θ associated to $\mu_1 = 1$ is

$$\begin{aligned}\Pi_\mu :=& \sum_{k=2}^{N}\sum_{j=1}^{n-r}(u_k \otimes w_j)(u_k^\top \otimes v_j^\top)\\ =& \left(\sum_{k=2}^{N} u_k u_k^\top\right) \otimes \left(\sum_{j=1}^{n-r} w_j v_j^\top\right) = \left(I_N - \frac{1}{N}U_N\right) \otimes M,\end{aligned} \quad (60)$$

where the last passage follows from $\sum_{k=1}^N u_k u_k^\top = I_N$. From the property $\frac{1}{N}U_N = \frac{1}{N^2}U_N^2$ it is easy to verify that these projectors are orthogonal, $\Pi_\lambda^2 = \Pi_\lambda$, $\Pi_\mu^2 = \Pi_\mu$, and $\Pi_\mu\Pi_\lambda = \Pi_\lambda\Pi_\mu = 0$, thus $\lim_{\gamma\to\infty}\Theta^\gamma = \Pi_\lambda + \Pi_\mu$, that is (31). □

REFERENCES

[1] I.F. Akyildiz, W.n Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.

[2] S. He, H.-S. Shin, S. Xu, and A. Tsourdos. Distributed estimation over a low-cost sensor network: A review of state-of-the-art. *Information Fusion*, 54:21–43, 2020.

[3] B. Silva, R.M. Fisher, A. Kumar, and G.P. Hancke. Experimental link quality characterization of wireless sensor networks for underground monitoring. *IEEE Trans. on Industrial Informatics*, 11(5):1099–1110, 2015.

[4] T. Sun and M. Xin. Multiple UAV target tracking using consensus-based distributed high degree cubature information filter. In *AIAA guidance, navigation, and control conference*, page 0347, 2015.

[5] B. Jia, K.D. Pham, E. Blasch, D. Shen, Z. Wang, and G. Chen. Cooperative space object tracking using space-based optical sensors via consensus-based filters. *IEEE Trans. on Aerospace and Electronic Systems*, 52(4):1908–1936, 2016.

[6] T.T. Vu and A.R. Rahmani. Distributed consensus-based Kalman filter estimation and control of formation flying spacecraft: Simulation and validation. In *AIAA Guidance, Navigation, and Control Conference*, page 1553, 2015.

[7] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. In *Proc. of the 44th IEEE Conf. on Decision and Control*, pages 8179–8184. IEEE, 2005.

[8] S.P. Talebi, S. Werner, V. Gupta, and Y.-F. Huang. On stability and convergence of distributed filters. *IEEE Signal Processing Letters*, 28:494–498, 2021.

[9] W. Yang, Y. Zhang, G. Chen, C. Yang, and L. Shi. Distributed filtering under false data injection attacks. *Automatica*, 102:34–44, 2019.

[10] X. Ge, Q.-L. Han, and Z. Wang. A dynamic event-triggered transmission scheme for distributed set-membership estimation over wireless sensor networks. *IEEE Trans. on Cybernetics*, 49(1):171–183, 2017.

[11] M. Kamgarpour and C. Tomlin. Convergence properties of a decentralized Kalman filter. In *Proc. of the 47th IEEE Conf. on Decision and Control*, pages 3205–3210. IEEE, 2008.

[12] C. Wan, Y. Gao, X.R. Li, and E. Song. Distributed filtering over networks using greedy gossip. In *2018 21st International Conference on Information Fusion*, pages 1968–1975. IEEE, 2018.

[13] Z. Wu, M. Fu, Y. Xu, and R. Lu. A distributed Kalman filtering algorithm with fast finite-time convergence for sensor networks. *Automatica*, 95:63–72, 2018.

[14] G. Battistelli and L. Chisci. Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica*, 50(3):707–718, 2014.

[15] A.T. Kamal, J.A. Farrell, and A.K. Roy-Chowdhury. Information weighted consensus filters and their application in distributed camera networks. *IEEE Trans. on Automatic Control*, 58(12):3112–3125, 2013.

[16] S.P. Talebi and S. Werner. Distributed Kalman filtering and control through embedded average consensus information fusion. *IEEE Trans. on Automatic Control*, 64(10):4396–4403, 2019.

[17] S. Wang and W. Ren. On the convergence conditions of distributed dynamic state estimation using sensor networks: A unified framework. *IEEE Trans. on Control Systems Technology*, 26(4):1300–1316, 2017.

[18] A. Ribeiro and G.B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-part i: Gaussian case. *IEEE Trans. on signal processing*, 54(3):1131–1143, 2006.

[19] A. Ribeiro and G.B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-part ii: Unknown probability density function. *IEEE Trans. on Signal Processing*, 54(7):2784–2796, 2006.

[20] Bo Chen, Wen-An Zhang, Li Yu, Guoqiang Hu, and Haiyu Song. Distributed fusion estimation with communication bandwidth constraints. *IEEE Trans. on Automatic Control*, 60(5):1398–1403, 2014.

[21] J. Li and G. AlRegib. Rate-constrained distributed estimation in wireless sensor networks. *IEEE Trans. on Signal Processing*, 55(5):1634–1643, 2007.

[22] J.A. Gubner. Distributed estimation and quantization. *IEEE Trans. on Information Theory*, 39(4):1456–1459, 1993.

[23] J. Fang and H. Li. Hyperplane-based vector quantization for distributed estimation in wireless sensor networks. *IEEE Trans. on Information Theory*, 55(12):5682–5699, 2009.

[24] I.D. Schizas, G.B. Giannakis, and Z.-Q. Luo. Distributed estimation using reduced-dimensionality sensor observations. *IEEE Trans. on Signal Processing*, 55(8):4284–4299, 2007.

[25] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, and A. Graziano. Consensus-based linear and nonlinear filtering. *IEEE Trans. on Automatic Control*, 60(5):1410–1415, 2015.

[26] S. Battilotti, F. Cacace, and M. d'Angelo. A stability with optimality analysis of consensus-based distributed filters for discrete-time linear systems. *Automatica*, 129:109589, 2021.

[27] J. Speyer. Computation and transmission requirements for a decentralized linear-quadratic-gaussian control problem. *IEEE Trans. on Automatic Control*, 24(2):266–269, 1979.

[28] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proc. of the 44th IEEE Symposium on Foundations of Computer Science, 2003.*, pages 482–491. IEEE, 2003.

[29] S. Battilotti, F. Cacace, M. d'Angelo, and A. Germani. Asymptotically optimal consensus-based distributed filtering of continuous-time linear systems. *Automatica*, 122:109189, 2020.

[30] S. Battilotti, F. Cacace, and M. d'Angelo. Distributed optimal control of discrete-time linear systems over networks. *IEEE Trans. on Control of Network Systems*, 2023.

[31] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge university press, 2012.