Distributed Design of Robust Kalman Filters Over Corrupted Channels

Xingkang He, Karl Henrik Johansson, *Fellow, IEEE*, and Haitao Fang

***Abstract*—We study distributed filtering for a class of uncertain systems over corrupted communication channels. We propose a distributed robust Kalman filter with stochastic gains, through which upper bounds of the conditional mean square estimation errors are calculated online. We present a robust collective observability condition, under which the mean square error of the distributed filter is proved to be uniformly upper bounded if the network is strongly connected. For better performance, we modify the filer by introducing a switching fusion scheme based on a sliding window. It provides a smaller upper bound of the conditional mean square error. Numerical simulations are provided to validate the theoretical results and show that the filter scales to large networks.**

***Index Terms*—Sensor network, distributed filtering, robust Kalman filter, corrupted channel.**

I. Introduction

In recent years, networked state estimation problems for sensor networks are drawing more and more attention due to their many applications [1]–[3]. Compared to the centralized methods, distributed algorithms, implemented at each sensor, are more resilient to network vulnerabilities, require less energy-consuming communication, and are able to perform parallel processing. Thus, a growing number of researchers are focusing on the study of distributed state estimation problems [4]–[8]. System uncertainties and communication imperfections pose, however, great challenges to the implementation and use of existing distributed filters. Thus, it is important to study distributed robust filters for real-time state estimation of uncertain systems.

System uncertainties exist in most applications in both the dynamics and measurements. Multiplicative noise arises in many situations [9]. When system dynamics suffer multiplicative noise, it is challenging to design effective filters due to the state-dependent uncertainty. The authors in [10] studied centralized estimation problems for systems with multiplicative noise and parameter uncertainties. In [11], distributed fusion estimation for systems with multiplicative and correlated noise was studied. In [12], the authors studied distributed filtering for systems with multiplicative noise in the dynamics when the network is given by a complete graph. Measurement degradation usually comes from sensor or communication limitations [12]–[14]. A detailed study on Kalman filters with measurement degradations was given in [15]. In [13], a distributed filter was proposed for a state-saturated system with degraded measurements and quantization effects. A robust estimation problem based on randomly dropped measurements was studied in [16]. A distributed robust filter was provided in [17] for a class of linear systems with uncertain measurements. Moreover, to deal with random changes in model structures and parameters in the real systems, some robust filtering approaches were proposed for systems with unknown parameters under non-Gaussian measurement noise [18]–[20] and for nonlinear uncertain Markov jump systems [21], [22]. Most of the above results were studied in a centralized framework, and for the distributed algorithms, few connections between filter performance and system uncertainties were provided.

In the literature of distributed estimation over sensor networks [12]–[14], [23]–[28], a common assumption is that the communications between sensors are noise-free. This is, however, difficult to fulfill in practice [29]. Uncertainty induced by channel noise makes it more challenging to design and analyze distributed filters. The authors in [30] investigated the design of distributed filters with constant filtering gains and fusion weights, and gave conditions to ensure the boundedness of the mean square error (MSE). In [23], a distributed filter was proposed by combining a diffusion step with the Kalman filter. The filter performance was analyzed under the assumption that each sub-system is observable, which is a restrictive condition for high-dimensional systems. Time-varying distributed filters can achieve better performance than static [31]–[33]. However, authors of [31]–[33] all assumed perfect communication. Although [34] studied the case that the state estimates suffer channel noise, the parameter matrices were required to be perfectly transmitted. The design of distributed robust filters exposed to corrupted communication channels needs further investigation.

The main contributions of this article are summarized in the following.

- For systems suffering multiplicative stable noise and measurements exposed to fading and additive noise, we design a robust distributed Kalman filter able to handle corrupted communication channels (Algorithm 1). The filter is shown

Manuscript received June 4, 2020; revised September 25, 2020 and January 10, 2021; accepted March 31, 2021. Date of publication April 5, 2021; date of current version April 28, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gwo Giun Lee. This work was supported in part by the Knut and Alice Wallenberg Foundation, in part by the Swedish Research Council, and in part by the National Natural Science Foundation of China under Grant 11931018. *(Corresponding author: Xingkang He.)*

Xingkang He and Karl Henrik Johansson are with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden, and also with Digital Futures (e-mail: xingkang@kth.se; kallej@kth.se).

Haitao Fang is with the LSC, NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China and also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: htfang@iss.ac.cn).

Digital Object Identifier 10.1109/TSP.2021.3070779

TABLE I
MAIN SYMBOLS IN THIS ARTICLE: k AND m STAND FOR TIME INSTANTS, i AND j STAND FOR SENSOR LABELS

symbol	meaning	symbol	meaning	symbol	meaning	symbol	meaning
x_k	state	w_k	process noise	$y_{k,i}$	measurement	$v_{k,i}$	measurement noise
ϵ_k	multip. noise	$\gamma_{k,i}$	fading factor	N	sensor number	A_k	system matrix
$C_{k,i}$	measurement matrix	F_k	multip. noise matrix	$\mathcal{V}$	node set	$\mathcal{E}$	link set
$\mathcal{V}=[a_{i,j}]$	adjacency matrix	$\mathcal{N}_i$	neighbor set	$\varepsilon_{k,i,j}$	channel noise	$\mathcal{D}_{k,i,j}$	channel noise
Q_k	cov. bound of w_k	μ_k	cov. bound of ϵ_k	$\varphi_{k,i}$	cov. bound of $\gamma_{k,i}$	$R_{k,i}$	cov. bound of $v_{k,i}$
$\Upsilon_{i,j}$	bound of $\varepsilon_{k,i,j}$	$\mathcal{D}_{i,j}$	bound of $\mathcal{D}_{k,i,j}$	P_0	bound of $E\{x_0x_0^T\}$	$\tau_{k,i}$	mean of $\gamma_{k,i}$
$\hat{x}_{k,i}$	fused estimate	$\bar{x}_{k,i}$	predicted estimate	$\tilde{x}_{k,i}$	updated estimate	$\hat{\tilde{x}}_{k,i,j}$	corrupted $\tilde{x}_{k,i}$
$\mathcal{W}_k$	channel noise σ-algebra	$K_{k,i}$	filter gain	$W_{k,i,j}$	fusion weight	Π_k	bound of $E\{x_kx_k^T\}$
$\bar{P}_{k,i}$	parameter in prediction	$\tilde{P}_{k,i}$	parameter in update	$P_{k,i}$	parameter in fusion	$\bar{\tilde{P}}_{k,i,j}$	corrupted $\tilde{P}_{k,j}$
Δ_i	optimization interval	L	window length	$\Phi_{k,m}$	transition matrix	$\bar{N}$	observability parameter

to be conditionally consistent in the sense that the MSE is conditionally bounded.

- We extend traditional collective observability to robust collective observability, under which the MSE of the distributed robust Kalman filter is proved to be uniformly upper bounded for any strongly connected network (Theorem III.1).
- We modify the proposed distributed robust Kalman filer by introducing a switching fusion scheme based on a sliding window and past state estimates (Algorithm 2). Adaptive covariance intersection (CI) weights are obtained by solving semi-definite programming (SDP) problems at the preset intervals. It is proved that the modified filter inherits the main properties of the distributed robust Kalman filter (Theorem IV.1), but in addition provides a smaller upper bound of the conditional MSE.

This article presents significant contributions compared to the existing literature. In particular, first, compared to [12]–[14], [23]–[27] where the communications are required to be noise-free, or [34] where the transmitted state estimates suffer channel noise, this article studies a more general case of channel corruption. We allow that both the transmitted estimates and parameter matrices can be polluted by channel noise. Second, thisarticle does not make the assumption that the nominal systems have to be stable [12], [13], [25] or that each sub-system is observable [14], [23], [24]. Moreover, different from [12], [13], [25], the design of the filters in this article is based on the information from the local sensor and the neighbor communications. Third, compared with the existing results [12], [13], [25]–[27], [34], using the neighbor estimates in a sliding window, the switching fusion scheme of this article can utilize the state estimates more efficiently.

The remainder of this article is organized as follows: Section 2 presents the problem formulation. The filter design and performance analysis are given in Section 3. Section 4 provides the modified filter based on a sliding-window method. After Section 5 gives numerical simulations, Section 6 concludes this article.

Notations: Superscript T represents transpose. The notation $A \geq B$ ($A > B$), where A and B are real symmetric matrices, means that $A - B$ is a positive semidefinite (positive definite) matrix. We denote $\mathbf{1}_n$ an n-dimensional vector with all elements one, I_n the identity matrix with n rows and columns, $\mathbb{R}^n$ the set of n-dimensional real vectors, and $\mathbb{N}$ the set of natural numbers. The operator $E\{x\}$ denotes the mathematical expectation of the stochastic vector x, and $\mathrm{Cov}\{x\} = E\{(x - E\{x\})(x - E\{x\})^T\}$. We use $\mathrm{blockdiag}\{\cdot\}$ and $\mathrm{diag}\{\cdot\}$ to represent the diagonalizations of square matrix elements and scalar elements, respectively. The trace of matrix P is denoted by $\mathrm{Tr}(P)$. For a real-valued matrix A, $\rho(A)$ denotes the spectral radius and $\|A\|_2 = \sqrt{\rho(A^TA)}$. The scalar $\lambda_{\max}(B)$ is the maximal eigenvalue of the real-valued symmetric matrix B, and $\sigma(\cdot)$ is the minimal σ-algebra operator generated by a collection of subsets. For reading convenience, main symbols of this article are provided in Table I.

II. PROBLEM FORMULATION

This section presents a motivating example followed by some preliminaries together with the problem formulation.

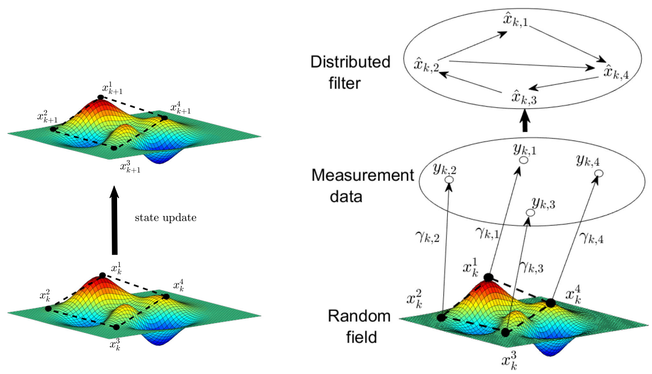
A. Motivating Example

In a spatially distributed physical system, let the state vector consist of elements over a large geographical area. The evolution of the state is related to spatial and temporal dynamics. Sensors located at different positions can collaborate based on their intermittent measurements of partial elements of the state. The state and the measurements are polluted by noise. A random dynamic field driven by noise w_k and monitored by a sensor network is shown in Fig. 1, cf. [35]. The variable x_k^i stands for the temperature in station i at time k. Colors represent values of x_k^i. The problem considered in this article is how to design a distributed robust filter based on the corrupted measurements $y_{k,i}, k \in \mathbb{N}, i = 1, \ldots, 4$, and the collaboration of the sensors, such that the overall temperature field x_k can be effectively estimated by each sensor.

B. Preliminaries

Consider the system dynamics

$$x_{k+1} = (A_k + F_k\epsilon_k)x_k + w_k, \tag{1}$$



(a) Evolution of the field (b) Distributed sensing and estimation

Fig. 1. A random temperature field over a geographical area. The evolution of the field is driven by some stochastic process w_k. The right figure illustrates that sensors obtain corrupted measurements of the temperature state, and communicate with other sensors over a network to achieve an estimate of the overall state.

where $x_k \in \mathbb{R}^n$ denotes the system state vector, $w_k \in \mathbb{R}^n$ the independent process noise with zero mean, $\epsilon_k \in \mathbb{R}$ the independent multiplicative noise also with zero mean. The matrices F_k, $k \in \mathbb{N}$, are non-singular matrices.

The system state is monitored by a sensor network with N sensors

$$y_{k,i} = \gamma_{k,i} C_{k,i} x_k + v_{k,i}, i = 1, \ldots, N, \tag{2}$$

where $y_{k,i} \in \mathbb{R}^{m_i}$ stands for the measurement vector of sensor i, $v_{k,i} \in \mathbb{R}^{m_i}$ the independent measurement noise with zero mean and $\gamma_{k,i} \in \mathbb{R}$ the independent random fading factor in the interval [0,1] with $E\{\gamma_{k,i}\} = \tau_{k,i}$, where $0 < \tau_{k,i} \leq 1$ is a known scalar, all at time $k = 1, 2, \ldots$. The matrices A_k, F_k, and $C_{k,i}$ have appropriate dimensions and are known to sensor i.

We model the sensor communications as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, which consists of nodes $\mathcal{V} = \{1, 2, \ldots, N\}$, links $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and the weighted adjacency matrix $\mathcal{A} = [a_{i,j}]$, where $a_{i,i} > 0, a_{i,j} \geq 0, \sum_{j \in \mathcal{V}} a_{i,j} = 1$. If $a_{i,j} > 0, j \neq i$, there is a link $(j, i) \in \mathcal{E}$, through which node i can directly receive messages from node j. In this case, node j is called a (in-)neighbor of node i and node i is called a out-neighbor of node j. The (in-)neighbor set of node i, including itself, is denoted by $\mathcal{N}_i$. The graph $\mathcal{G}$ is called strongly connected if for any two nodes i_1, i_l, there exists a directed path from i_l to $i_1 : (i_l, i_{l-1}), \ldots, (i_3, i_2), (i_2, i_1)$. Let $\{\tilde{x}_{k,j}, \tilde{P}_{k,j}\}$ be the pair that node j communicates to its out-neighbor nodes at time k, where $\tilde{x}_{k,j} \in \mathbb{R}^n$ and $\tilde{P}_{k,j} \in \mathbb{R}^{n \times n}$. Due to channel noise, the pair $\{\tilde{\tilde{x}}_{k,i,j}, \tilde{\tilde{P}}_{k,i,j}\}$ received by node i from node j is

$$\begin{aligned} \tilde{\tilde{x}}_{k,i,j} &= \tilde{x}_{k,j} + \varepsilon_{k,i,j}, j \in \mathcal{N}_i \\ \tilde{\tilde{P}}_{k,i,j} &= \tilde{P}_{k,j} + \mathcal{D}_{k,i,j}, j \in \mathcal{N}_i, \end{aligned} \tag{3}$$

where $\varepsilon_{k,i,j} \in \mathbb{R}^n$ and $\mathcal{D}_{k,i,j} \in \mathbb{R}^{n \times n}$ are the channel noise processes. If $\tilde{P}_{k,j}$ is symmetric, $\mathcal{D}_{k,i,j}$ is reasonably assumed to be symmetric. Because it is sufficient to transmit the upper triangular part of the symmetric matrix $\tilde{P}_{k,j}$. In Lemma III.2, we will show that the transmitted matrix $\tilde{P}_{k,j}$ is indeed symmetric.

Let $(\Omega, \mathcal{F}, P)$ be the basic probability space, and $\mathcal{F}_k$ be a filtration of the σ-algebra $\mathcal{F}$. A discrete-time sequence $\{\xi_k\}$ is said to be adapted to the family of σ-algebras $\{\mathcal{F}_k\}$ if ξ_k is measurable to $\mathcal{F}_k$. We refer the reader [36] for details. We require the following assumption.

Assumption II.1: The following conditions on noise and initial estimates hold.

1) The initial state x_0, its estimates $\hat{x}_{0,i}$, and the noise ϵ_k, w_k, $\gamma_{k+1,i}, v_{k+1,i}, \varepsilon_{k+1,i,j}, \mathcal{D}_{k+1,i,j}$ are independent both in time and space, for all $i, j \in \mathcal{V}, k = 0, 1, \ldots$.
2) There exist known matrices $Q_k, R_{k+1,i}, P_0$ and scalars $\mu_k, \varphi_{k+1,i}$, such that for all $i \in \mathcal{V}$, and $k = 0, 1, \ldots$,

$$E\{w_k w_k^T\} \leq Q_k, \quad \inf_{k \in \mathbb{N}} Q_k > 0, \quad E\{x_0 x_0^T\} \leq P_0$$

$$E\{\epsilon_k^2\} \leq \mu_k, \qquad \mathrm{Cov}\{\gamma_{k+1,i}\} \leq \varphi_{k+1,i}$$

$$E\{v_{k+1,i} v_{k+1,i}^T\} \leq R_{k+1,i}$$

$$\sup_{k \in \mathbb{N}} \left[\tau_{k+1,i}^2 C_{k+1,i}^T R_{k+1,i}^{-1} C_{k+1,i}\right] < \infty$$

$$E\{(\hat{x}_{0,i} - x_0)(\hat{x}_{0,i} - x_0)^T\} \leq P_{0,i}.$$

3) There exist positive semi-definite matrices $\Upsilon_{i,j}$ and $\mathcal{D}_{i,j}$ such that for all $i \in \mathcal{V}$, $j \in \mathcal{N}_i$, and $k = 1, 2, \ldots$,

$$\sup\{\varepsilon_{k,i,j} \varepsilon_{k,i,j}^T\} \leq \Upsilon_{i,j}, -\mathcal{D}_{i,j} \leq \mathcal{D}_{k,i,j} \leq \mathcal{D}_{i,j},$$

where the channel noise $\varepsilon_{k,i,j}$ and $\mathcal{D}_{k,i,j}$ are in (3).

Note that the exact covariance information of the stochastic uncertainties is not required. Bounds and statistics are known only to individual sensors. Thus, the conditions in 2) of Assumption II.1 are milder than [12], [13], [25], where each sensor was assumed to have full knowledge on the statistics of the system.

Let $\hat{x}_k$ be the estimate of the system state x_k. Due to unknown correlation between sensor estimates, the MSE of each sensor can not be obtained in a distributed manner [14], [32], [33], [37]. We introduce the following definitions to consider the bounds of MSE.

Definition II.1: [38] (Consistency) The pair $\{\hat{x}_k, P_k\}$ is consistent if there is a deterministic sequence $\{P_k\}$ such that $E\{(\hat{x}_k - x_k)(\hat{x}_k - x_k)^T\} \leq P_k$.

Definition II.2: (Conditional consistency) The pair $\{\hat{x}_k, P_k\}$ is conditionally consistent if there is a sequence $\{P_k\}$, such that $E\{(\hat{x}_k - x_k)(\hat{x}_k - x_k)^T | \mathcal{K}_k\} \leq P_k$, where $\mathcal{K}_k$ is a σ-algebra and P_k is measurable to $\mathcal{K}_k$.

Note that the consistency defined above is different from the one in parameter identification, which instead is on asymptotic convergence to the true parameters. The consistency definition we use in this article [26], [27], [32], [33] provides two benefits. First, the estimation error of each sensor can be evaluated online by utilizing some probability inequalities [39]. Second, a CI-based fusion method can be utilized in the filter design. We introduce conditional consistency in Definition II.2 to cope with channel noise. The idea is to use that the pair $\{\hat{x}_k, E\{P_k\}\}$ is consistent, if $\{\hat{x}_k, P_k\}$ is conditionally consistent.

C. Problem

In this article, we consider a three-step distributed filtering structure. Each sensor $i \in \mathcal{V}$, executes a state prediction, measurement update and local fusion at each time:

$$\begin{aligned} \bar{x}_{k,i} &= A_{k-1}\hat{x}_{k-1,i} \\ \tilde{x}_{k,i} &= \bar{x}_{k,i} + K_{k,i}(y_{k,i} - \tau_{k,i}C_{k,i}\bar{x}_{k,i}) \\ \hat{x}_{k,i} &= \sum_{j\in\mathcal{N}_i} W_{k,i,j}\hat{\tilde{x}}_{k,i,j}, \end{aligned} \tag{4}$$

where $\bar{x}_{k,i}$, $\tilde{x}_{k,i}$, and $\hat{x}_{k,i}$ are the state estimates in prediction, update, and fusion of sensor i at time k, respectively. Moreover, $\hat{\tilde{x}}_{k,i,j}$ given in (3) is the noisy estimate received by sensor i from sensor j. Besides, $K_{k,i}$ is the filtering gain parameter matrix, $W_{k,i,j}$ is the local fusion parameter matrix. Both $K_{k,i}$ and $W_{k,i,j}$ remain to be designed.

Different from the existing results [12], [30], [33], [40], measurements and measurement matrices are not transmitted in our setting. The advantages of this protocol lie in several aspects including privacy, security and energy saving.

In this article, we consider three essential subproblems:

(**a**) How to design the parameters $K_{k,i}$ and $W_{k,i,j}$ in the distributed filter (4), such that the filter is conditionally consistent? (Lemmas III.2 and III.3)

(**b**) Which conditions on system structure and noise statistics enable the mean square estimation error to be bounded? (Theorem III.1)

(**c**) How to improve the performance of the filter (4) when past estimates are available? (Algorithm 2, Proposition IV.1, and Theorem IV.1)

III. Distributed Robust Kalman Filter Design

In this section, we first provide a distributed design of the filter gain $K_{k,i}$ and the fusion weight $W_{k,i,j}$ of the filter (4). Then we present our proposed distributed robust Kalman filter (DRKF) algorithm. Finally, it is shown that the algorithm gives bounded MSE.

Lemma III.1: Under Assumption II.1, it holds that $E\{x_k x_k^T\} \leq \Pi_k, \forall k \in \mathbb{N}$, where Π_k is recursively calculated through $\Pi_{k+1} = A_k\Pi_k A_k^T + \mu_k F_k \Pi_k F_k^T + Q_k$, with $\Pi_0 = P_0$, in which P_0, μ_k, and Q_k are in Assumption II.1.

Proof: See Appendix A. ■

Lemma III.1 provides an upper bound of the mean square of the system state $x(t)$, which is accessible to each sensor based on its system knowledge and useful in the algorithm design and analysis as follows. Similar approaches are found in [10], [41]. By employing the CI-method [38], the following lemma provides a choice for the fusion weight $W_{k,i,j}$ giving conditional consistency.

Lemma III.2: Consider system (1)–(2) satisfying Assumption II.1. For the filter (4) with $k \geq 1$ and $i \in \mathcal{V}$, if $K_{k,i}$ is adapted to the channel noise σ-algebra $\mathcal{W}_k = \sigma(\mathcal{D}_{t,i,j}, 1 \leq t \leq k, i, j, \in \mathcal{V})$, and

$$W_{k,i,j} = a_{i,j}P_{k,i}(\bar{\tilde{P}}_{k,i,j} + \mathcal{D}_{i,j} + \Upsilon_{i,j})^{-1}, \tag{5}$$

then the pairs $\{\bar{x}_{k,i}, \bar{P}_{k,i}\}$, $\{\tilde{x}_{k,i}, \tilde{P}_{k,i}\}$, $\{\hat{x}_{k,i}, P_{k,i}\}$ are all conditionally consistent given $\mathcal{W}_k$, where

$$\begin{aligned} \bar{P}_{k,i} &= A_{k-1}P_{k-1,i}A_{k-1}^T + \mu_{k-1}F_{k-1}\Pi_{k-1}F_{k-1}^T + Q_{k-1} \\ \tilde{P}_{k,i} &= (I - \tau_{k,i}K_{k,i}C_{k,i})\bar{P}_{k,i}(I - \tau_{k,i}K_{k,i}C_{k,i})^T \\ &\quad + K_{k,i}\left(R_{k,i} + \varphi_{k,i}C_{k,i}\Pi_k C_{k,i}^T\right)K_{k,i}^T \\ \bar{\tilde{P}}_{k,i,j} &= \tilde{P}_{k,j} + \mathcal{D}_{k,i,j}, j \in \mathcal{N}_i \\ P_{k,i} &= \left(\sum_{j\in\mathcal{N}_i} a_{i,j}(\bar{\tilde{P}}_{k,i,j} + \mathcal{D}_{i,j} + \Upsilon_{i,j})^{-1}\right)^{-1}. \end{aligned}$$

Proof: See Appendix B. ■

Note that the design of the fusion weight $W_{k,i,j}$ in Lemma III.2 is fully distributed, and it depends on the communication noise bounds, i.e., $\Upsilon_{i,j}$ and $\mathcal{D}_{k,i,j}$, which is an extension to [12], [13], [25], [26], [34]. In the following lemma, we design the filter gain $K_{k,i}$ of filter (4) such that the bound of the conditional MSE, i.e., $\tilde{P}_{k,i}$, is minimized at each measurement update.

Lemma III.3: The optimal solution $K_{k,i}^* := \arg\min_{K_{k,i}} \mathrm{Tr}\{\tilde{P}_{k,i}\}$ is given by

$$K_{k,i}^* = \tau_{k,i}\bar{P}_{k,i}C_{k,i}^T\Xi_{k,i}^{-1},$$

where $\Xi_{k,i} = \tau_{k,i}^2 C_{k,i}\bar{P}_{k,i}C_{k,i}^T + R_{k,i} + \varphi_{k,i}C_{k,i}\Pi_k C_{k,i}^T$. Furthermore, $K_{k,i}^*$ is adapted to the channel noise σ-algebra $\mathcal{W}_k$ in Lemma III.2.

Proof: See Appendix C. ■

The designed filter gain in Lemma III.3 inherits the gain of the optimal centralized robust filters in [10], [41], but here it is stochastic and adapted to the channel noise σ-algebra $\mathcal{W}_k = \sigma(\mathcal{D}_{t,i,j}, 1 \leq t \leq k, i, j, \in \mathcal{V})$. With the filter parameters $K_{k,i}$ and $W_{k,i,j}$ given in Lemmas III.2 and III.3, respectively, we obtain the DRKF given in Algorithm 1. Different from [14], [37], the implementation of this algorithm only depends on the local measurement information $\{y_{k,i}, C_{k,i}, R_{k,i}, \varphi_{k,i}, \tau_{k,i}\}$ and the estimate pairs $\{\hat{\tilde{x}}_{k,i,j}, \bar{\tilde{P}}_{k,i,j}, j \in \mathcal{N}_i\}$ from neighbors. Thus, it obeys a fully distributed design and implementation. For sensor i, the computational complexity of Algorithm 1 at each time is $O(\max\{n^3 d_i, m_i^3\})$, where d_i is the cardinality of the set $\mathcal{N}_i$, and n and m_i are the dimensions of the system state and sensor measurement, respectively. The overall computational complexity for all sensors is consequently $O(\max\{Nn^3 d_i, Nm_i^3\})$. Thus, the algorithm is scalable to large networks. The performance of the algorithm is degraded if the upper bounds in Assumption II.1 are not tight. In systems with measurement outliers [19], Algorithm 1 can be adapted to estimate the state by developing appropriate scheme for discarding the measurement outliers.

Next we find mild conditions to guarantee boundedness of the MSE for Algorithm 1. For $j > k$, we denote the transition matrix by $\Phi_{j,k} = A_{j-1}\Phi_{j-1,k}$, where $\Phi_{k,k} = I_n$. We assume robust collective observability in the following.

Algorithm 1: Distributed Robust Kalman Filter (DRKF).

Initial setting:
$\{\hat{x}_{0,i}, P_{0,i}, \Pi_0, \mathcal{D}_{i,j}, \Upsilon_{i,j}, j \in \mathcal{N}_i, i \in \mathcal{V}\}$.
Prediction: For each sensor i:
$\bar{x}_{k,i} = A_{k-1}\hat{x}_{k-1,i}$,
$\bar{P}_{k,i} = A_{k-1}P_{k-1,i}A_{k-1}^T + \mu_{k-1}F_{k-1}\Pi_{k-1}F_{k-1}^T + Q_{k-1}$,
$\Pi_k = A_{k-1}\Pi_{k-1}A_{k-1}^T + \mu_{k-1}F_{k-1}\Pi_{k-1}F_{k-1}^T + Q_{k-1}$.
Update: For each sensor i:
$\tilde{x}_{k,i} = \bar{x}_{k,i} + K_{k,i}(y_{k,i} - \tau_{k,i}C_{k,i}\bar{x}_{k,i})$,
$K_k = \tau_{k,i}\bar{P}_{k,i}C_{k,i}^T(\tau_{k,i}^2 C_{k,i}\bar{P}_{k,i}C_{k,i}^T + R_{k,i} + \varphi_{k,i}C_{k,i}\Pi_k C_{k,i}^T)^{-1}$
$\tilde{P}_{k,i} = (I - \tau_{k,i}K_{k,i}C_{k,i})\bar{P}_{k,i}$.
Fusion: For each sensor i:
$\hat{x}_{k,i} = P_{k,i}\sum_{j\in\mathcal{N}_i} a_{i,j}(\bar{\tilde{P}}_{k,i,j} + \mathcal{D}_{i,j} + \Upsilon_{i,j})^{-1}\hat{\tilde{x}}_{k,i,j}$,
$P_{k,i} = (\sum_{j\in\mathcal{N}_i} a_{i,j}(\bar{\tilde{P}}_{k,i,j} + \mathcal{D}_{i,j} + \Upsilon_{i,j})^{-1})^{-1}$,
where $\hat{\tilde{x}}_{k,i,j}$ and $\bar{\tilde{P}}_{k,i,j}$ are given in (3).

Assumption III.1: (Robust collective observability) There exists an integer $\bar{N} > 0$ and a constant $\alpha > 0$ such that for $k \in \mathbb{N}$,

$$\sum_{i=1}^{N}\sum_{j=k}^{k+\bar{N}} \Phi_{j,k}^T \bar{C}_{j,i}^T \tilde{R}_{j,i}^{-1}\bar{C}_{j,i}\Phi_{j,k} \geq \alpha I_n, \tag{6}$$

where

$$\bar{C}_{j,i} = \tau_{j,i}C_{j,i}, \quad j \in \mathbb{N}, \quad i \in \mathcal{V}$$
$$\tilde{R}_{j,i} = R_{j,i} + \varpi_j\varphi_{j,i}C_{j,i}C_{j,i}^T$$
$$\varpi_j = \|P_0\|_2\prod_{i=0}^{j-1}\bar{\alpha}_i + \sum_{s=1}^{j}\left(\bar{q}_{s-1}\prod_{l=s}^{j}\bar{\alpha}_l\right) + \bar{q}_j$$
$$\bar{\alpha}_j = \|A_j\|_2^2 + \mu_j\|F_j\|_2^2$$
$$\bar{q}_j = \|Q_j\|_2.$$

Assumption III.1 is based on the system structure and noise statistics. It can be regarded as a distributed version of the observability condition with multiplicative noise in [41]. The condition does not require that each sub-system is observable [14], [23], [24]. Moreover, if $\varphi_{k,i} \equiv 0$, $\forall k \in \mathbb{N}, i \in \mathcal{V}$, Assumption III.1 corresponds to the collective observability condition for time-varying stochastic systems in [26].

A requirement on the multiplicative noise ϵ_k is needed. Recall that μ_k is the bound of the variance of ϵ_k. Denote the time sequence of non-zero multiplicative noise by

$$\mathbb{K}_T = \left\{k_t = \min_{\mu_k>0} k | k \geq k_{t-1}, k, t \in \mathbb{N}\right\}. \tag{7}$$

Assumption III.2: There exist positive scalars λ_1, λ_2, M and $\varrho \in (0,1)$, such that

$$\lambda_1 I_n \leq A_k A_k^T \leq \lambda_2 I_n, k \in \mathbb{N} \tag{8}$$

$$\prod_{t=s}^{l}\rho_{k_t} \leq M\varrho^{l-s}, 0 \leq s \leq l < \infty \tag{9}$$

$$\sup_{t\in\mathbb{N}}\|\mu_{k_{t+1}}F_{k_{t+1}}\mathcal{Q}_{k_{t+1},k_t}F_{k_{t+1}}^T\|_2 < \infty, \tag{10}$$

where $k_t \in \mathbb{K}_T$ in (7) and

$$\rho_{k_t} = \frac{\mu_{k_{t+1}}}{\mu_{k_t}}\|F_{k_{t+1}}\Phi_{k_{t+1},k_t}F_{k_t}^{-1}\|_2^2 + \mu_{k_{t+1}}\|F_{k_{t+1}}\Phi_{k_{t+1},k_t}\|_2^2$$

$$\mathcal{Q}_{k_{t+1},k_t} = \sum_{k=k_t}^{k_{t+1}}\Phi_{k_{t+1},k}Q_k\Phi_{k_{t+1},k}^T.$$

Compared to [12], [13], [25], (8) is a milder condition as it permits the nominal system to be unstable. If $\{k|\mu_k > 0, k \in \mathbb{N}\}$ is finite or even empty, (9) and (10) can still be made satisfied by replacing the points $\mu_k = 0$ with sufficiently small positive $\bar{\mu}_k$. For further analysis, we need Lemmas III.4–III.5.

Lemma III.4: If Assumption III.1 holds, then

$$\sum_{i=1}^{N}\sum_{j=k}^{k+\bar{N}}\Phi_{j,k}^T\bar{C}_{j,i}^T\bar{R}_{j,i}^{-1}\bar{C}_{j,i}\Phi_{j,k} \geq \alpha I_n, \tag{11}$$

where $\bar{R}_{k,i} := R_{k,i} + \varphi_{k,i}C_{k,i}\Pi_k C_{k,i}^T$.

Proof: See Appendix D. ■

Different from (6) in Assumption III.1, (11) in Lemma III.4 utilizes Π_k given by Lemma III.1. We note that Lemma III.4 is provided for the proof of Theorem III.1.

Lemma III.5: If Assumption III.2 holds, then

$$\sup_{k\in\mathbb{N}}\{\mu_k F_k \Pi_k F_k^T\} < \infty.$$

Proof: See Appendix E. ■

Lemma III.5 is useful in the proof of the following theorem. Next we state our main result on Algorithm 1: the estimation MSE of $e_{k,i} := \hat{x}_{k,i} - x_k$ is bounded.

Theorem III.1: Suppose system (1)–(2) satisfies Assumptions II.1, III.1–III.2 and that $\mathcal{G}$ is strongly connected. Then, the estimation MSE for Algorithm 1 is uniformly bounded for all sensors, i.e., there exists a positive scalar η such that

$$\sup_{k\geq N+\bar{N}}\lambda_{max}\left(E\{e_{k,i}e_{k,i}^T\}\right) \leq \frac{\eta}{\alpha}, \forall i \in \mathcal{V},$$

where α is given in Assumption III.1.

Proof: See Appendix F. ■

Theorem III.1 states that a larger α can lead to a smaller upper bound of the MSE. Thus, increasing observability ($\bar{C}_{k,i}$) and reducing noise interference ($\bar{R}_{k,i}$) can both contribute to improving estimation performance of the DRKF in Algorithm 1.

IV. DRKF With a Sliding Window

In this section, we modify the DRKF algorithm to include also past estimates received from neighbors. The presented DRKF with sliding-window fusion (DRKF-SWF) algorithm is shown to give bounded MSE. In the numerical simulation in next section, it is shown to sometimes outperform the DRKF algorithm.

Since the estimates $\{\hat{\tilde{x}}_{k,i,j}, \bar{\tilde{P}}_{k,i,j}, j \in \mathcal{N}_i\}$ have been corrupted by the channel noise through (3), designing a distributed filter simply based on the latest estimates may lead to performance degradation if these estimates have been seriously deteriorated. In this case, we fuse the past estimates received from neighbors. This leads to a better estimate than that of simply fusing current estimates. To decide which past estimates to use, a sliding window with length $L \geq 1$ is introduced. For $l = 0, \ldots, L$, we denote

$$\begin{aligned} \check{x}_{k-l,j} &:= \hat{\bar{x}}_{k-l,i,j} \\ \check{P}_{k-l,j} &:= \bar{\bar{P}}_{k-l,i,j} + \mathcal{D}_{i,j} + \Upsilon_{i,j}. \end{aligned} \tag{12}$$

By Lemma III.2, $\{\check{x}_{k,j}, \check{P}_{k-l,j}\}$ is conditionally consistent given the channel noise σ-algebra $\mathcal{W}_k = \sigma(\mathcal{D}_{t,i,j}, 1 \leq t \leq k, i, j, \in \mathcal{V})$. Sensor i has the available messages $\{\check{x}_{l,j}, \check{P}_{l,j}\}_{l=k-L+1}^{k}$ from sensor j. We denote

$$\begin{aligned} (\check{x}_{k,j}^1, \check{P}_{k,j}^1) &:= (f_0(\check{x}_{k,j}), g_0(\check{P}_{k,j})) := (\check{x}_{k,j}, \check{P}_{k,j}) \\ (\check{x}_{k,j}^2, \check{P}_{k,j}^2) &:= (f_1(\check{x}_{k-1,j}), g_1(\check{P}_{k-1,j})) \\ &\vdots \\ (\check{x}_{k,j}^L, \check{P}_{k,j}^L) &:= (f_{L-1}(\check{x}_{k-L+1,j}), g_{L-1}(\check{P}_{k-L+1,j})), \end{aligned} \tag{13}$$

where for $l = 1, \ldots, L-1$,

$$\begin{aligned} f_l(\check{x}_{k-l,j}) &= f_1(f_{l-1}(\check{x}_{k-l,j})) \\ g_l(\check{P}_{k-l,j}) &= g_1(g_{l-1}(\check{P}_{k-l,j})) \\ f_1(\check{x}_{k-l,j}) &= A_{k-l}\check{x}_{k-l,j} \\ g_1(\check{P}_{k-l,j}) &= A_{k-l}\check{P}_{k-l,j}A_{k-l}^T + Q_{k-l} \\ &\quad + \mu_{k-l}F_{k-l}\Pi_{k-l}F_{k-l}^T. \end{aligned} \tag{14}$$

At time k, based on the local knowledge and the information received from neighbors, sensor i can fuse the messages $\{\check{x}_{l,j}, \check{P}_{l,j}, j \in \mathcal{V}_i\}_{l=k-L+1}^{k}$ to obtain a better estimate of x_k. By (21), $\{\check{x}_{l,j}, \check{P}_{l,j}, j \in \mathcal{V}_i\}_{l=k-L+1}^{k}$ are all conditionally consistent given $\mathcal{W}_k = \sigma(\mathcal{D}_{t,i,j}, 1 \leq t \leq k, i, j, \in \mathcal{V})$.

Let

$$\hat{x}_{k,i} = P_{k,i} \sum_{s=1}^{L} \sum_{j \in \mathcal{N}_i} a_{i,j,k}^s (\check{P}_{k,j}^s)^{-1} \check{x}_{k,j}^s \tag{15}$$

$$P_{k,i} = \left(\sum_{s=1}^{L} \sum_{j \in \mathcal{N}_i} a_{i,j,k}^s (\check{P}_{k,j}^s)^{-1} \right)^{-1}, \tag{16}$$

where $a_{i,j,k}^s$ is element (i,j) of $\bar{\mathcal{A}}_k \in \mathbb{R}^{N \times NL}$ which is the CI fusion weight matrix for $\{\check{x}_{k,j}^s, \check{P}_{k,j}^s, j \in \mathcal{V}_i\}_{l=k-L+1}^{k}$. In the following, the design of $\bar{\mathcal{A}}_k$ is studied. By the proof of Lemma III.2 and (13), $\{\check{x}_{k,j}^s, \check{P}_{k,j}^s, j \in \mathcal{V}_i\}_{l=k-L+1}^{k}$ are conditionally consistent given $\mathcal{W}_k = \sigma(\mathcal{D}_{t,i,j}, 1 \leq t \leq k, i, j, \in \mathcal{V})$. The design of $\bar{\mathcal{A}}_k$ is given by solving the following optimization problem.

$$\begin{aligned} &\underset{a_{i,j,k}^s, j \in \mathcal{N}_i}{\text{minimize}} \quad \mathrm{Tr}(\mathcal{J}_{k,i}^{-1}) \\ &\text{subject to} \\ &\qquad \mathcal{J}_{k,i} > 0 \\ &\qquad 0 \leq a_{i,j,k}^s \leq 1, \\ &\qquad \sum_{s=1}^{L} \sum_{j \in \mathcal{N}_i} a_{i,j,k}^s = 1 \end{aligned} \tag{17}$$

where $\mathcal{J}_{k,i} = \sum_{s=1}^{L} \sum_{j \in \mathcal{N}_i} a_{i,j,k}^s (\check{P}_{k,j}^s)^{-1} - \sum_{j \in \mathcal{N}_i} a_{i,j} \check{P}_{k,j}^{-1}$. The optimal solution to (17) is denoted by $\bar{a}_{i,j,k}^s, j \in \mathcal{N}_i, s = 1, \ldots, L$. According to [26], the problem in (17) is convex and equivalent to an SDP problem, which can be effectively solved by many existing algorithms if the problem is feasible. If the problem is infeasible, we use the same fusion approach as Algorithm 1, i.e., $\bar{\mathcal{A}}_k = (\mathcal{A}\ 0^{N \times (N-1)L})$. The feasibility of the SDP is equivalent to the feasibility test problem of linear matrix inequality [42]. Due to resource constraints, it may be undesirable to solve the online optimization problem (17) at each time. Suppose sensor i has the ability to solve (17) at time instants $\{k_s\}_{s=1}^{\infty}$, subject to

$$\mod(k_s, \Delta_i) = 0,$$

where $\mod(a,b)$ is the remainder operator of a/b and $\Delta_i \in \mathbb{Z}^+$ is the time interval length within which sensor i can not solve the optimization problem. In other words, at time instants $\{k_s\}_{s=1}^{\infty}$, each sensor employs (15) to obtain a fused estimate, and for other instants, it utilizes the fusion methods in Algorithm 1 based on the latest estimates from its neighbors. We provide the DRKF-SWF in Algorithm 2. Compared with [12], [13], [25]–[27], [34], Algorithm 2 utilizes the past information more efficiently and considers the limitation of step-wise optimization. The computational burden of Algorithm 2, in addition to that of Algorithm 1, is that it solves the SDP convex optimization problem (17) for every Δ_i. Thus, also Algorithm 2 scales to large networks, as such optimization problems are easy to solve. The difficulty in the implementation of Algorithm 2 is that solving the optimization problem (17) needs more computational resources if the dimension of the system state increases.

The following lemma shows that Algorithm 2 is conditionally consistent given the channel noise σ-algebra $\mathcal{W}_k$.

Lemma IV.1: Consider system (1)–(2) satisfying Assumption II.1. Then for Algorithm 2, the pairs $\{\bar{x}_{k,i}, \bar{P}_{k,i}\}$, $\{\tilde{x}_{k,i}, \tilde{P}_{k,i}\}$, and $\{\hat{x}_{k,i}, P_{k,i}\}$ are conditionally consistent given $\mathcal{W}_k$.

Proof: Similar to the proof of Lemma III.2 but considering the CI fusion in (15) and the fact that $K_{k,i}$ is adapted to $\mathcal{W}_k = \sigma(\mathcal{D}_{t,i,j}, 1 \leq t \leq k, i, j, \in \mathcal{V})$. ■

Lemma IV.1, corresponding to Lemma III.2, shows that Algorithm 2 shares the same conditional consistency as Algorithm 1. Algorithm 2 is better than Algorithm 1 in the following sense.

Algorithm 2: Distributed Robust Kalman Filter With Sliding-Window Fusion (DRKF-SWF).

Initial setting:
$\{L, \Delta_i, \hat{x}_{0,i}, P_{0,i}, \Pi_0, \mathcal{D}_{i,j}, \Upsilon_{i,j}, j \in \mathcal{N}_i, i \in \mathcal{V}\}$.
Prediction: Same as Algorithm 1.
Update: Same as Algorithm 1.
Local Fusion: For each sensor i:
if $\mod(k, \Delta_i) = 0$ and (17) has a feasible solution:
$\hat{x}_{k,i} = P_{k,i} \sum_{s=1}^{L} \sum_{j\in\mathcal{N}_i} \bar{a}^s_{i,j,k} (\check{P}^s_{k,j})^{-1} \check{x}^s_{k,j}$
$P_{k,i} = \left(\sum_{s=1}^{L} \sum_{j\in\mathcal{N}_i} \bar{a}^s_{i,j,k} (\check{P}^s_{k,j})^{-1} \right)^{-1}$,
where $\check{P}^s_{k,j}$ and $\check{x}^s_{k,j}$ are given in (13), and $\{\bar{a}^s_{i,j,k}\}_{s=1}^{L}$ are given by solving (17);
else
$\hat{x}_{k,i} = P_{k,i} \sum_{j\in\mathcal{N}_i} a_{i,j} (\bar{\tilde{P}}_{k,i,j} + \mathcal{D}_{i,j} + \Upsilon_{i,j})^{-1} \hat{\tilde{x}}_{k,i,j}$
$P_{k,i} = \left(\sum_{j\in\mathcal{N}_i} a_{i,j} (\bar{\tilde{P}}_{k,i,j} + \mathcal{D}_{i,j} + \Upsilon_{i,j})^{-1} \right)^{-1}$,
where $\hat{\tilde{x}}_{k,i,j}$ and $\bar{\tilde{P}}_{k,i,j}$ are given in (3).

Proposition IV.1: Consider system (1)–(2) satisfying Assumption II.1. Under the same initial setting and the channel noise σ-algebra $\mathcal{W}_k = \sigma(\mathcal{D}_{t,i,j}, 1 \le t \le k, i, j, \in \mathcal{V})$, for Algorithms 1–2, it holds that

$$P^B_{k,i} \le P^A_{k,i}, \tag{18}$$

where $P^A_{k,i}$ and $P^B_{k,i}$ are the $P_{k,i}$ matrix of Algorithm 1 and Algorithm 2, respectively.

Proof: If $\mod(k, \Delta_i) = 0$ and (17) is feasible, the constraint of (17) $\mathcal{J}_{k,i} > 0$ ensures that Algorithm 2 has a smaller $P_{k,i}$. Otherwise, the fusion scheme of Algorithm 2 is the same as Algorithm 1, which also ensures (18). ■

Proposition IV.1 shows that compared to Algorithm 1, Algorithm 2 has a smaller upper bound of the MSE. A larger window parameter L can lead to a smaller objective function of (17), but the computation will increase as well. Also, the time length Δ_i influences the estimation performance, since a larger Δ_i means that sensor i does not solve the optimization problem (17) for a longer time interval. The parameters L and Δ_i can be chosen based on the computational and communication ability of the sensor network. Furthermore, let T be the time length of interest, then Algorithm 2 degenerates to Algorithm 1 if $\Delta_i > T$. The boundedness of the MSE for Algorithm 2 is presented in the following.

Theorem IV.1: Suppose system (1)–(2) satisfies Assumptions II.1, III.1–III.2 and that $\mathcal{G}$ is strongly connected. Then, the estimation MSE for Algorithm 2 is uniformly bounded for all sensors,, i.e., there exists a positive scalar $\tilde{\eta}$ such that

$$\sup_{k\ge N+\bar{N}} \lambda_{max}\left(E\{e_{k,i}e^T_{k,i}\}\right) \le \frac{\tilde{\eta}}{\alpha}, \forall i \in \mathcal{V},$$

where α is given in Assumption III.1.

Proof: It follows from Lemma IV.1 and the proof of Theorem III.1. ■

Theorem IV.1, corresponding to Theorem III.1, shows that Algorithm 2 shares the same MSE boundedness as Algorithm 1 under mild conditions.

V. Numerical Simulations

In this section, we study two examples to validate the effectiveness of the proposed algorithms and the theoretical results developed in the paper.

A. Example 1

For the temperature field in Fig. 1, we suppose that the initial state x_0 and sensor measurement noise are generated by independent standard normal distributions. The fading factors $\gamma_{k,i}$ follow independent uniform distributions, $i = 1, 2, 3, 4$. The time sequence $\{t_k\}$ lies in the interval [0,10] with uniform sampling step 0.1, thus $k = 0, 1, \ldots, 100$. The matrices and scalars in (1) are assumed to be

$$\begin{aligned}
A_k &= \begin{pmatrix} 0.8 \times (1 + 0.01t_k) & 0.01 \\ 0.1 & 0.98 \end{pmatrix} \\
F_k &= I_4, Q_k = 0.1 \times I_2, P_0 = I_2, \mu_k = 0.1 \times (t_k + 2)^{-1} \\
R_{k,1} &= 0.07, R_{k,2} = 0.08, R_{k,3} = R_{k,4} = 0.09 \\
\tau_{k,1} &= 0.85, \varphi_{k,1} = 0.8 \times 10^{-3}, C_{k,1} = \begin{pmatrix} 0 & 1 \end{pmatrix} \\
\tau_{k,2} &= 0.15, \varphi_{k,2} = 0.8 \times 10^{-3}, C_{k,2} = \begin{pmatrix} 0 & 1 \end{pmatrix} \\
\tau_{k,3} &= 0.20, \varphi_{k,3} = 0.8 \times 10^{-3}, C_{k,3} = \begin{pmatrix} 0 & 1 \end{pmatrix} \\
\tau_{k,4} &= 0.85, \varphi_{k,4} = 0.8 \times 10^{-3}, C_{k,4} = \begin{pmatrix} 1 & 0 \end{pmatrix}.
\end{aligned} \tag{19}$$

The initial setting of the filters is $\hat{x}_{i,0} = \mathbf{1}_2$ and $P_{i,0} = 100 \times I_2, \forall i \in \mathcal{V}$. The weighted adjacency matrix is

$$\mathcal{A} = [a_{i,j}] = \begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0.3 & 0.4 & 0 & 0.3 \end{pmatrix}.$$

The channel noise is assumed to be mutually independent and uniformly distributed over $[-1, 1]$. We choose $\Upsilon_{i,j} = \mathcal{D}_{i,j} = I_2, i, j \in \mathcal{V}$.

We conduct Monte Carlo experiments, in which $N_t = 100$ runs are performed. We denote

$$\begin{aligned}
\mathrm{MSE_{k,i}} &= \frac{1}{N_t} \sum_{j=1}^{N_t} (\hat{x}^j_{k,i} - x^j_k)^T (\hat{x}^j_{k,i} - x^j_k) \\
\mathrm{Tr}(P_{k,i}) &= \frac{1}{N_t} \sum_{j=1}^{N_t} \mathrm{Tr}(P^j_{k,i}),
\end{aligned} \tag{20}$$

where $\hat{x}^j_{k,i}$ and $P^j_{k,i}$ are the state estimate and parameter matrix of the jth run of sensor i.

We show how $\mathrm{Tr}(P_{k,i})$ is an upper bound of $\mathrm{MSE_{k,i}}$. Fig. 2 shows that this holds for each sensor. Let $\mathrm{MSE_k} = \frac{1}{|\mathcal{V}|} \sum_{i\in\mathcal{V}} \mathrm{MSE_{k,i}}$, $\mathrm{Tr}(P_k) = \frac{1}{|\mathcal{V}|} \sum_{i\in\mathcal{V}} \mathrm{Tr}(P_{k,i})$. To illustrate the relationship between the initial conditions and the output of the DRKF, we provide Table II, where $\mathrm{MSE_{max}} =$

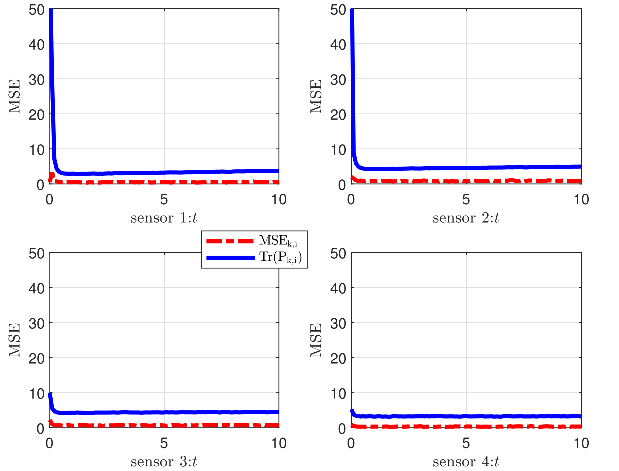


Fig. 2. Consistent estimates of DRKF.

TABLE II
$\mathrm{MSE}_{\max}$ AND $P_{\max}$ OF THE DRKF WITH DIFFERENT INITIAL QUANTITIES

Case number	$P_{0,i}$	Π_0	$\mathcal{D}_{i,j}$	$\Upsilon_{i,j}$	$\mathrm{MSE}_{\max}$	$P_{\max}$
1	$100I_2$	I_2	I_2	I_2	0.74	4.15
2	$500I_2$	I_2	I_2	I_2	0.75	4.15
3	$100I_2$	$5I_2$	I_2	I_2	0.73	4.16
4	$100I_2$	I_2	$5I_2$	I_2	0.89	9.38
5	$100I_2$	I_2	I_2	$5I_2$	0.90	9.38

$\max_{k=51,\ldots,100} \mathrm{MSE}_k$, and $P_{\max} = \max_{k=51,\ldots,100} P_k$. Here we just consider $k \in \{51, \ldots, 100\}$, since the estimation error after $k = 51$ is relatively steady. Table II shows that $P_{0,i}$ and Π_0 have little influence on the output of the DRKF, but $\mathcal{D}_{i,j}$ and $\mp_{i,j}$ affect $\mathrm{MSE}_{\max}$ and $P_{\max}$, as expected.

We compare the proposed DRKF algorithm with centralized Kalman filter (CKF), centralized robust Kalman filter (CRKF) [10], [41], collaborative scalar-gain estimator (CSGF) [30], and distributed state estimation with consensus on the posteriors (DSEA-CP) [32]. The centralized filters CKF and CRKF utilize the observations of all sensors without suffering communication noise. Moreover, for the considered scenario, CRKF is the optimal robust filter in the sense that its filter gain ensures the minimum bound of MSE [10], [41]. The MSE of these algorithms are shown in Fig. 3, which indicates that the DRKF achieves better estimation accuracy than CSGF, DSEA-CP, and DRKF. Fig. 4 shows that DRKF-SWF provides bounded mean square estimation errors and consistent estimates. By setting $\Delta_i = \Delta$, $i \in \mathcal{V}$, Fig. 5 shows that DRKF-SWF with sliding-window length $L = 2$ provides smaller upper bounds than the DRKF by decreasing the interval length Δ_i.

B. Example 2

Consider the undirected network with 50 sensors in Fig. 6. The weights of the adjacency matrix are given by

$$a_{i,j} = \frac{1}{\max\{d_i, d_j\}}, \quad i \in \mathcal{V}, j \in \mathcal{N}_i, j \neq i$$

$$a_{i,i} = 1 - \sum_{j \in \mathcal{N}_i, j \neq i} a_{i,j}.$$

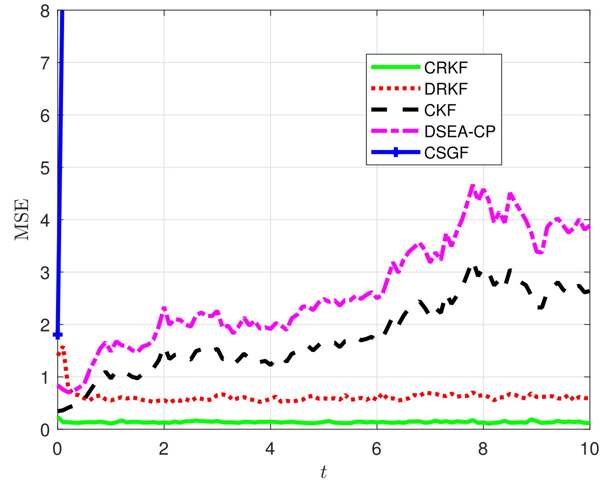


Fig. 3. Comparison of tracking performance for the proposed filter DRKF together with filters from the literature.

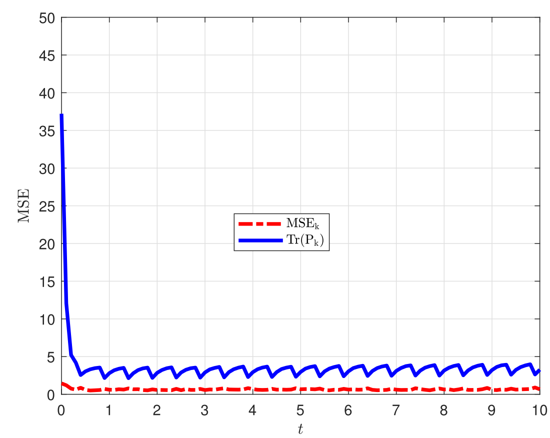


Fig. 4. Consistent estimates of DRKF-SWF with $L = 2$ and $\Delta_i = 5$.

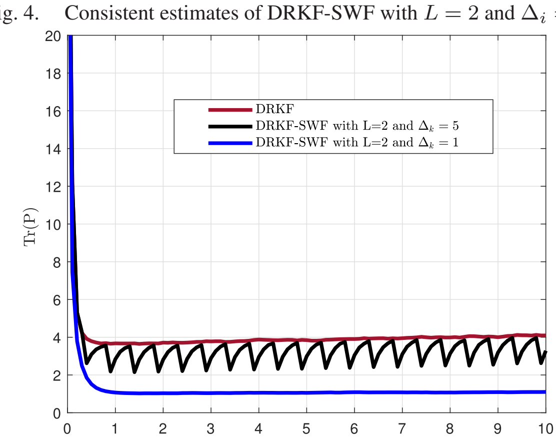


Fig. 5. Comparison between DRKF and DRKF-SWF.

where d_i is the cardinality of the set $\mathcal{N}_i$. We assume $A_k = \begin{pmatrix} 1.05 & 0.1 \\ -0.1 & 0.98 \end{pmatrix}$, $\mu_k = 0$, $R_{k,i} = 1, i \in \{1, \ldots, 50\}$. For each sensor, the pair of measurement vector and fading statistics are randomly chosen out of the four combinations in (19). The rest of the simulation settings are the same as in Example 1. Fig. 7 shows the bounded MSE and its upper bound, which verifies the estimation consistency of Algorithm 1. In Fig. 8, we compare the estimation performance of the DRKF with the four algorithms

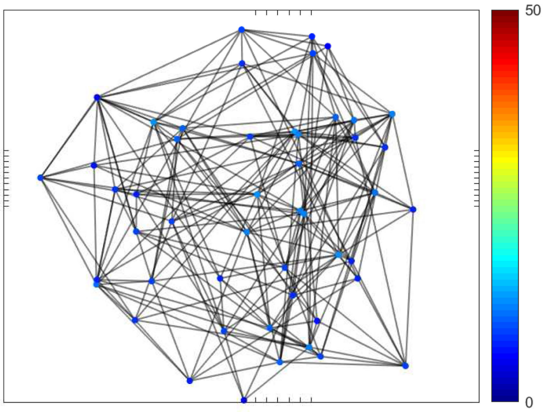


Fig. 6. A sensor network with 50 nodes.

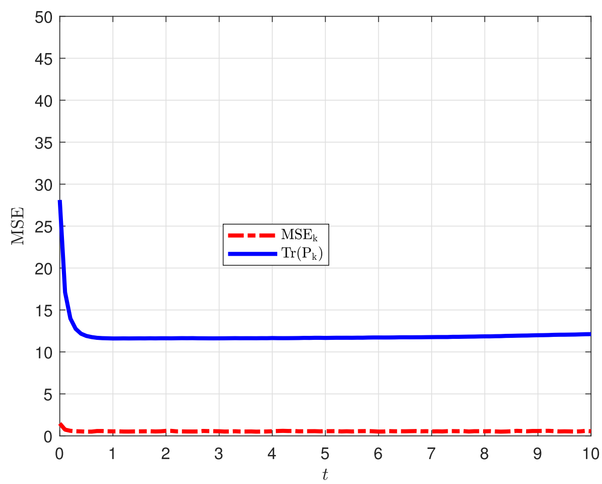


Fig. 7. The consistency of DRKF.

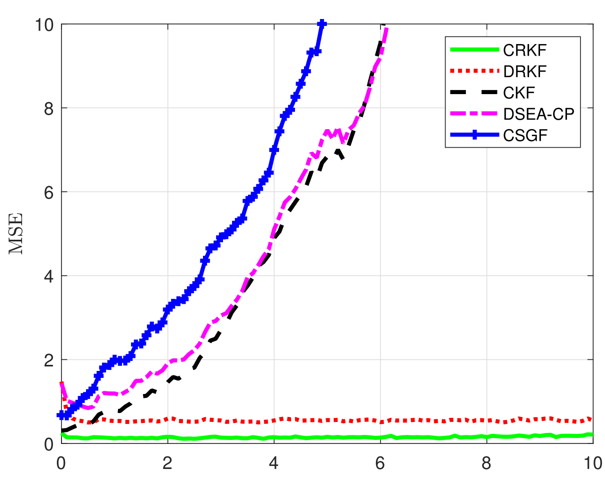


Fig. 8. Comparison between DRKF with filters from the literature.

mentioned in Example 1. The result shows that the proposed DRKF achieves better performance than the CSGF, DSEA-CP, and CKF, whose estimation errors are diverging fast due to the instability of the system dynamics (i.e., $\rho(A_k) = 1.02 > 1$). The performance of Algorithm 1, i.e., DRKF, is close to CRKF.

VI. Conclusion

This article studied a distributed robust state estimation problem for a class of discrete-time stochastic systems with multiplicative noise and degraded measurements over corrupted communication channels. Employing local imprecise statistics, we first proposed a three-step DRKF. Then, under some mild conditions, we proved that its MSE is uniformly upper bounded by a constant matrix after a finite transient time. The finite time interval is related to the collective observability and the network size. A switching fusion scheme based on a sliding-window fusion method was proposed as a DRKF-SWF algorithm to obtain a smaller upper bound of the MSE. By considering extended computational ability of the sensors, the DRKF-SWF shows that better performance can be achieved.

APPENDIX

A. Proof of Lemma 3.1

We use an inductive method to prove this lemma. At the initial time, it follows from Assumption II.1 that $E\{x_0x_0^T\} \leq P_0 = \Pi_0$. Suppose at time k that $E\{x_kx_k^T\} \leq \Pi_k, \forall k \geq 0$. According to (1), x_k is adapted to $\mathcal{F}_{k-1}$. By Assumption 2.1, we have $E\{\epsilon_k x_k\} = 0$, and $E\{w_k\epsilon_k\} = 0$. For $E\{\epsilon_k^2 x_k x_k^T\}$, it holds that $E\{\epsilon_k^2 x_k x_k^T\} = E\{\epsilon_k^2\}E\{x_kx_k^T\}$, then

$$
\begin{aligned}
&E\{x_{k+1}x_{k+1}^T\}\\
&= E\{(A_k + F_k\epsilon_k)x_kx_k^T(A_k + F_k\epsilon_k)^T\} + E\{w_kw_k^T\}\\
&\quad + E\{(A_k + F_k\epsilon_k)x_kw_k^T\} + E\{w_kx_k^T(A_k + F_k\epsilon_k)^T\}\\
&\leq A_kE\{x_kx_k^T\}A_k^T + E\{\epsilon_k^2\}F_kE\{x_kx_k^T\}F_k^T + E\{w_kw_k^T\}\\
&\leq A_k\Pi_kA_k^T + \mu_kF_k\Pi_kF_k^T + Q_k = \Pi_{k+1}.
\end{aligned}
$$

Hence, we obtain $E\{x_{k+1}x_{k+1}^T\} \leq \Pi_{k+1}$.

B. Proof of Lemma 3.2

Regarding the filtering structure in (4), for proof convenience, we denote the state estimation errors by $\bar{e}_{k,i} = \bar{x}_{k,i} - x_k, \tilde{e}_{k,i} = \tilde{x}_{k,i} - x_k, \bar{\tilde{e}}_{k,i,j} = \hat{\tilde{x}}_{k,i,j} - x_k$, and $e_{k,i} = \hat{x}_{k,i} - x_k$, respectively. Then it is straightforward to obtain the dynamics of these estimation errors as follows

$$
\begin{aligned}
\bar{e}_{k,i} &= A_{k-1}e_{k-1,i} - w_{k-1} - \epsilon_{k-1}F_{k-1}x_{k-1}\\
\tilde{e}_{k,i} &= (I_n - \tau_{k,i}K_{k,i}C_{k,i})\bar{e}_{k,i}\\
&\quad + K_{k,i}(v_{k,i} + (\gamma_{k,i} - \tau_{k,i})C_{k,i}x_k)\\
\bar{\tilde{e}}_{k,i,j} &= \tilde{e}_{k,j} + \varepsilon_{k,i,j}, j \in \mathcal{N}_i\\
e_{k,i} &= \sum_{j\in\mathcal{N}_i} W_{k,i,j}\bar{\tilde{e}}_{k,i,j}.
\end{aligned}
\tag{21}
$$

First, we make a conjecture that if $\{\hat{x}_{t-1,i}, P_{t-1,i}\}, t \geq 1$ is conditionally consistent given the channel noise σ-algebra $\mathcal{W}_{t-1}$, i.e., $E\{e_{t-1,i}e_{t-1,i}^T|\mathcal{W}_{t-1}\} \leq P_{t-1,i}$, then the pairs $\{\bar{x}_{t,i}, \bar{P}_{t,i}\}, \{\tilde{x}_{t,i}, \tilde{P}_{t,i}\}, \{\hat{x}_{t,i}, P_{t,i}\}$ are all conditionally consistent given $\mathcal{W}_k$. In the following, we prove the conjecture.

Suppose at time $t=k$, the pair $\{\hat{x}_{k-1,i}, P_{k-1,i}\}$, $k\geq 1$, is conditionally consistent given $\mathcal{W}_{k-1}$. According to Assumption II.1 and (21), we have $E\{\epsilon_{k-1}e_{k-1,i}x_{k-1}|\mathcal{W}_k\}=0$ and $E\{w_{k-1}e_{k-1,i}|\mathcal{W}_k\}=0$. It follows that

$$\begin{aligned}
&E\{\bar{e}_{k,i}\bar{e}_{k,i}^T|\mathcal{W}_k\}\\
&\leq A_{k-1}E\{e_{k-1,i}e_{k-1,i}^T|\mathcal{W}_{k-1}\}A_{k-1}^T+Q_{k-1}\\
&\quad+\mu_{k-1}F_{k-1}E\{x_{k-1}x_{k-1}^T\}F_{k-1}^T\\
&\leq A_{k-1}P_{k-1,i}A_{k-1}^T+\mu_{k-1}F_{k-1}\Pi_{k-1}F_{k-1}^T+Q_{k-1}=\bar{P}_{k,i}.
\end{aligned} \tag{22}$$

In the measurement update, according to (21), we have $\tilde{e}_{k,i}=(I_n-\tau_{k,i}K_{k,i}C_{k,i})\bar{e}_{k,i}+K_{k,i}v_{k,i}+(\gamma_{k,i}-\tau_{k,i})K_{k,i}C_{k,i}x_k$. By Assumption II.1, $E\{\bar{e}_{k,i}\gamma_{k,i}|\mathcal{W}_k\}=0$ and $E\{\bar{e}_{k,i}v_{k,i}^T|\mathcal{W}_k\}=0$. Since $v_{k,i}$ and $\gamma_{k,i}$ are mutually independent and $K_{k,i}$ is adapted to $\mathcal{W}_k$, we have

$$\begin{aligned}
&E\{\tilde{e}_{k,i}\tilde{e}_{k,i}^T|\mathcal{W}_k\}\\
&\leq (I_n-\tau_{k,i}K_{k,i}C_{k,i})E\{\bar{e}_{k,i}\bar{e}_{k,i}^T|\mathcal{W}_k\}(I_n-\tau_{k,i}K_{k,i}C_{k,i})^T\\
&\quad+\varphi_{k,i}K_{k,i}C_{k,i}\Pi_kC_{k,i}^TK_{k,i}^T+K_{k,i}R_{k,i}K_{k,i}^T\\
&\leq (I_n-\tau_{k,i}K_{k,i}C_{k,i})\bar{P}_{k,i}(I_n-\tau_{k,i}K_{k,i}C_{k,i})^T\\
&\quad+K_{k,i}\left(\varphi_{k,i}C_{k,i}\Pi_kC_{k,i}^T+R_{k,i}\right)K_{k,i}^T=\tilde{P}_{k,i}.
\end{aligned} \tag{23}$$

Note that the communication channels are imperfect and the messages received by each sensor are polluted by the channel noise through (3). According to Assumption II.1 and (21), we have

$$\begin{aligned}
&E\{\bar{\bar{e}}_{k,i,j}\bar{\bar{e}}_{k,i,j}^T|\mathcal{W}_k\}\\
&=E\{(\tilde{x}_{k,j}+\varepsilon_{k,i,j}-x_k)(\tilde{x}_{k,j}+\varepsilon_{k,i,j}-x_k)^T|\mathcal{W}_k\}\\
&\leq E\{(\tilde{x}_{k,j}-x_k)(\tilde{x}_{k,j}-x_k)^T|\mathcal{W}_k\}+E\{\varepsilon_{k,i,j}\varepsilon_{k,i,j}^T|\mathcal{W}_k\}\\
&\leq \tilde{P}_{k,j}+\sup\{\varepsilon_{k,i,j}\varepsilon_{k,i,j}^T\}\\
&\leq \tilde{P}_{k,j}+\Upsilon_{i,j}\\
&\leq \tilde{P}_{k,j}+\mathcal{D}_{k,i,j}+\mathcal{D}_{i,j}+\Upsilon_{i,j}=\bar{\tilde{P}}_{k,i,j}+\mathcal{D}_{i,j}+\Upsilon_{i,j},
\end{aligned}$$

where $\bar{\tilde{P}}_{k,i,j}$ is the received matrix by sensor i from sensor j. In the local fusion step, $e_{k,i}=\sum_{j\in\mathcal{N}_i}W_{k,i,j}\bar{\bar{e}}_{k,i,j}$. Given $W_{k,i,j}$ in (5), according to (23) and the consistent estimation of the CI method [38], we have $E\{e_{k,i}e_{k,i}^T|\mathcal{W}_k\}\leq P_{k,i}$.

Thus, the conjecture holds. Then the conclusion is obtained based on the conjecture and the initial estimation condition in Assumption II.1.

C. Proof of Lemma III.3

According to Lemma III.2, we have

$$\begin{aligned}
\tilde{P}_{k,i}&=(I_n-\tau_{k,i}K_{k,i}C_{k,i})\bar{P}_{k,i}(I_n-\tau_{k,i}K_{k,i}C_{k,i})^T\\
&\quad+K_{k,i}\left(\varphi_{k,i}C_{k,i}\Pi_kC_{k,i}^T+R_{k,i}\right)K_{k,i}^T\\
&=\bar{P}_{k,i}-\tau_{k,i}K_{k,i}C_{k,i}\bar{P}_{k,i}-\tau_{k,i}\bar{P}_{k,i}C_{k,i}^TK_{k,i}^T\\
&\quad+\tau_{k,i}^2K_{k,i}C_{k,i}\bar{P}_{k,i}C_{k,i}^TK_{k,i}^T\\
&\quad+K_{k,i}\left(\varphi_{k,i}C_{k,i}\Pi_kC_{k,i}^T+R_{k,i}\right)K_{k,i}^T\\
&=\bar{P}_{k,i}-\tau_{k,i}K_{k,i}C_{k,i}\bar{P}_{k,i}-\tau_{k,i}\bar{P}_{k,i}C_{k,i}^TK_{k,i}^T\\
&\quad+K_{k,i}\Xi_{k,i}K_{k,i}^T\\
&=(K_{k,i}-K_{k,i}^*)\Xi_{k,i}(K_{k,i}-K_{k,i}^*)^T\\
&\quad+(I-\tau_{k,i}K_{k,i}^*C_{k,i})\bar{P}_{k,i},
\end{aligned} \tag{24}$$

where $K_{k,i}^*=\tau_{k,i}\bar{P}_{k,i}C_{k,i}^T\Xi_{k,i}^{-1}$ and $\Xi_{k,i}=\tau_{k,i}^2C_{k,i}\bar{P}_{k,i}C_{k,i}^T+R_{k,i}+\varphi_{k,i}C_{k,i}\Pi_kC_{k,i}^T$. Thus, (24) shows that $\mathrm{Tr}(\tilde{P}_{k,i})$ is minimized when $K_{k,i}=K_{k,i}^*=\tau_{k,i}\bar{P}_{k,i}C_{k,i}^T\Xi_{k,i}^{-1}$. As a result, $\tilde{P}_{k,i}=(I-\tau_{k,i}K_{k,i}C_{k,i})\bar{P}_{k,i}$. Since $K_{k,i}^*$ is a measurable function of $\bar{P}_{k,i}$, which is adapted to $\mathcal{W}_k$, also, $K_{k,i}^*$ is adapted to $\mathcal{W}_k$.

D. Proof of Lemma III.4

According to Lemma III.1, we have $\Pi_{k+1}=A_k\Pi_kA_k^T+\mu_kF_k\Pi_kF_k^T+Q_k$. Taking the 2-norm of both sides yields $\|\Pi_{k+1}\|_2\leq\|\Pi_k\|_2(\|A_k\|_2^2+\mu_k\|F_k\|_2^2)+\|Q_k\|_2$. Denote $\|A_k\|_2^2+\mu_k\|F_k\|_2^2=:\bar{\alpha}_k$ and $\|Q_k\|_2=:\bar{q}_k$. Then, $\|\Pi_{k+1}\|_2\leq\varpi_{k+1}$, where $\varpi_{k+1}=\|P_0\|_2\prod_{i=0}^k\bar{\alpha}_i+\sum_{s=1}^k(\bar{q}_{s-1}\prod_{j=s}^k\bar{\alpha}_j)+\bar{q}_k$. It follows that $\bar{R}_{k,i}=:R_{k,i}+\varphi_{k,i}C_{k,i}\Pi_kC_{k,i}^T\leq R_{k,i}+\varpi_k\varphi_{k,i}C_{k,i}C_{k,i}^T=\tilde{R}_{k,i}$. If (6) is satisfied, (11) holds.

E. Proof of Lemma III.5

According to Lemma III.1 and Assumption III.2, we have $\Pi_{k_{t+1}}=\Phi_{k_{t+1},k_t}\Pi_{k_t}\Phi_{k_{t+1},k_t}^T+\mathcal{Q}_{k_{t+1},k_t}+\mu_{k_t}\Phi_{k_{t+1},k_t}F_{k_t}\Pi_{k_t}F_{k_t}^T\Phi_{k_{t+1},k_t}^T$. Multiplying from left by $\mu_{k_{t+1}}F_{k_{t+1}}$ and from right by $F_{k_{t+1}}^T$ yields

$$\begin{aligned}
&\mu_{k_{t+1}}F_{k_{t+1}}\Pi_{k_{t+1}}F_{k_{t+1}}^T\\
&=\mu_{k_{t+1}}F_{k_{t+1}}\Phi_{k_{t+1},k_t}\Pi_{k_t}\Phi_{k_{t+1},k_t}^TF_{k_{t+1}}^T\\
&\quad+\mu_{k_{t+1}}F_{k_{t+1}}\mu_{k_t}\Phi_{k_{t+1},k_t}F_{k_t}\Pi_{k_t}F_{k_t}^T\Phi_{k_{t+1},k_t}^TF_{k_{t+1}}^T\\
&\quad+\mu_{k_{t+1}}F_{k_{t+1}}\mathcal{Q}_{k_{t+1},k_t}F_{k_{t+1}}^T,
\end{aligned}$$

where $\mathcal{Q}_{k_{t+1},k_t}=\sum_{k=k_t}^{k_{t+1}}\Phi_{k_{t+1},k}Q_k\Phi_{k_{t+1},k}^T$. Denote $\mu_{k_t}F_{k_t}\Pi_{k_t}F_{k_t}^T=:\Theta_{k_t}$, then we have

$$\begin{aligned}&\Theta_{k_{t+1}}\\&=\frac{\mu_{k_{t+1}}}{\mu_{k_t}}F_{k_{t+1}}\Phi_{k_{t+1},k_t}F_{k_t}^{-1}\Theta_{k_t}F_{k_t}^{-T}\Phi_{k_{t+1},k_t}^{T}F_{k_{t+1}}^{T}\\&\quad+\mu_{k_{t+1}}F_{k_{t+1}}\Phi_{k_{t+1},k_t}\Theta_{k_t}\Phi_{k_{t+1},k_t}^{T}F_{k_{t+1}}^{T}\\&\quad+\mu_{k_{t+1}}F_{k_{t+1}}\mathcal{Q}_{k_{t+1},k_t}F_{k_{t+1}}^{T}. \end{aligned}\tag{25}$$

Taking 2-norm of both sides of (25) yields

$$\begin{aligned}&\|\Theta_{k_{t+1}}\|_2\\&\leq\|\frac{\mu_{k_{t+1}}}{\mu_{k_t}}F_{k_{t+1}}\Phi_{k_{t+1},k_t}F_{k_t}^{-1}\Theta_{k_t}F_{k_t}^{-T}\Phi_{k_{t+1},k_t}^{T}F_{k_{t+1}}^{T}\|_2\\&\quad+\|\mu_{k_{t+1}}F_{k_{t+1}}\Phi_{k_{t+1},k_t}\Theta_{k_t}\Phi_{k_{t+1},k_t}^{T}F_{k_{t+1}}^{T}\|_2\\&\quad+\mu_{k_{t+1}}\|F_{k_{t+1}}\mathcal{Q}_{k_{t+1},k_t}F_{k_{t+1}}^{T}\|_2\\&\leq\rho_{k_t}\|\Theta_{k_t}\|_2+\mu_{k_{t+1}}\|F_{k_{t+1}}\mathcal{Q}_{k_{t+1},k_t}F_{k_{t+1}}^{T}\|_2. \end{aligned}\tag{26}$$

According to [43], conditions (9) and (10) now give $\sup_{k_t\in\mathbb{N}}\|\Theta_{k_t}\|_2<\infty$, i.e., Θ_k is uniformly upper bounded.

F. *Proof of Theorem III.1*

Introduce

$$\begin{aligned}S_{k,i}&:=P_{k,i}^{-1}\\\tilde{Q}_k&:=\mu_kF_k\Pi_kF_k^T+Q_k\\G_{k,i}&:=\sum_{j\in\mathcal{N}_i}a_{i,j}\bar{C}_{k,j}^T\bar{R}_{k,j}^{-1}\bar{C}_{k,j}\\\bar{R}_{k,i}&:=R_{k,i}+\varphi_{k,i}C_{k,i}\Pi_kC_{k,i}^T.\end{aligned}$$

By Assumption II.1,

$$\begin{aligned}&\bar{\tilde{P}}_{k,i,j}+\mathcal{D}_{i,j}+\Upsilon_{i,j}\\&=\tilde{P}_{k,j}+\mathcal{D}_{k,i,j}+\mathcal{D}_{i,j}+\Upsilon_{i,j}\\&\geq\tilde{P}_{k,j}+\Upsilon_{i,j}\geq\tilde{P}_{k,j}.\end{aligned}$$

As $\inf_{k\in\mathbb{N}}Q_k>0$, and $\sup_{k\in\mathbb{N}}[\tau_{k,i}^2C_{k,i}^TR_{k,i}^{-1}C_{k,i}]<\infty$ in Assumption II.1, there exists a scalar $\vartheta_0>0$ such that $\bar{\tilde{P}}_{k,i,j}+\mathcal{D}_{i,j}+\Upsilon_{i,j}\leq(1+\vartheta_0)\tilde{P}_{k,j}$.

According to Algorithm 1 and Lemma III.5,

$$\begin{aligned}S_{k,i}&=\sum_{j\in\mathcal{N}_i}a_{i,j}(\bar{\tilde{P}}_{k,i,j}+\mathcal{D}_{i,j}+\Upsilon_{i,j})^{-1}\\&\geq\sum_{j\in\mathcal{N}_i}\frac{a_{i,j}}{1+\vartheta_0}\left(A_{k-1}S_{k-1,j}^{-1}A_{k-1}^T+\tilde{Q}_{k-1}\right)^{-1}+\frac{G_{k,i}}{1+\vartheta_0}\\&\geq\bar{\eta}A_{k-1}^{-T}\left(\sum_{j\in\mathcal{N}_i}a_{i,j}S_{k-1,j}\right)A_{k-1}^{-1}+\frac{G_{k,i}}{1+\vartheta_0},\end{aligned}\tag{27}$$

where $0<\bar{\eta}<1$. This inequality is obtained by Lemma 1 in [32] using Assumption III.2 and $\frac{1}{1+\vartheta_0}<1$. Let $a_{ij,k}$ be the (i,j)th element of $\mathcal{A}^k$. By recursively applying (27) $k\geq N+\bar{N}$ times, we have

$$S_{k,i}\geq\bar{\eta}^k\Phi_{k,0}^{-T}\left[\sum_{j\in\mathcal{V}}a_{ij,k}S_{0,j}\right]\Phi_{k,0}^{-1}+\frac{\bar{S}_{k,i}}{1+\vartheta_0},\tag{28}$$

where

$$\bar{S}_{k,i}=\sum_{s=1}^{k}\bar{\eta}^{s-1}\Phi_{k,k-s+1}^{-T}\left[\sum_{j\in\mathcal{V}}a_{ij,s}\tilde{S}_{k-s+1,j}\right]\Phi_{k,k-s+1}^{-1},$$

with $\tilde{S}_{k,j}=\bar{C}_{k,j}^T\bar{R}_{k,j}^{-1}\bar{C}_{k,j}$. Since the first term of the right-hand side of (28) is positive definite, it follows that

$$S_{k,i}\geq\frac{\bar{S}_{k,i}}{1+\vartheta_0},\forall k\geq N+\bar{N}.\tag{29}$$

Since $\mathcal{G}$ is strongly connected, $a_{ij,s}>0$ for $s\geq N-1$ [26]. Supposing $\bar{L}=N+\bar{N}$, we obtain

$$\begin{aligned}\bar{S}_{k,i}&\geq\sum_{s=1}^{\bar{L}}\bar{\eta}^{s-1}\Phi_{k,k-s+1}^{-T}\left[\sum_{j\in\mathcal{V}}a_{ij,s}\tilde{S}_{k-s+1,j}\right]\Phi_{k,k-s+1}^{-1}\\&\geq a_{\min}\bar{\eta}^{\bar{L}-1}\sum_{s=N}^{\bar{L}}\Phi_{k,k-s+1}^{-T}\left[\sum_{j\in\mathcal{V}}\tilde{S}_{k-s+1,j}\right]\Phi_{k,k-s+1}^{-1}\\&=a_{\min}\bar{\eta}^{\bar{L}-1}\sum_{j=1}^{N}\sum_{s=N}^{\bar{L}}\Phi_{k,k-s+1}^{-T}\tilde{S}_{k-s+1,j}\Phi_{k,k-s+1}^{-1},\end{aligned}\tag{30}$$

where $a_{\min}=\min_{i,j\in\mathcal{V},s\in\{N,\ldots,\bar{L}\}}a_{ij,s}>0$.

According to Assumption III.2, there exists a scalar $\beta>0$, such that $\Phi_{k,k-\bar{L}+1}^{-T}\Phi_{k,k-\bar{L}+1}^{-1}\geq\beta I_n,\forall k\geq 0$. From Lemma III.4 and $\bar{L}=N+\bar{N}$, it holds that

$$\begin{aligned}&\sum_{j=1}^{N}\sum_{s=N}^{\bar{L}}\Phi_{k,k-s+1}^{-T}\tilde{S}_{k-s+1,j}\Phi_{k,k-s+1}^{-1}\\&=\Phi_{k,k-\bar{L}+1}^{-T}\\&\quad\times\sum_{j=1}^{N}\left[\sum_{s=k-\bar{L}+1}^{k-N+1}\Phi_{s,k-\bar{L}+1}^{T}\tilde{S}_{k-\bar{L}+1,j}\Phi_{s,k-\bar{L}+1}\right]\Phi_{k,k-\bar{L}+1}^{-1}\\&\geq\alpha\Phi_{k,k-\bar{L}+1}^{-T}\Phi_{k,k-\bar{L}+1}^{-1}\geq\alpha\beta I_n,\forall k\geq N+\bar{N}.\end{aligned}\tag{31}$$

Summing up (30) and (31) yields

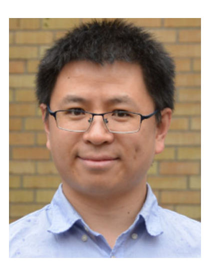
$$\bar{S}_{k,i}\geq a_{\min}\bar{\eta}^{\bar{L}-1}\alpha\beta I_n,\forall k\geq N+\bar{N}.\tag{32}$$

Let $S_*(\alpha)=a_{\min}\bar{\eta}^{\bar{L}-1}\alpha\beta I_n$. In light of (29), it holds that $P_{k,i}^{-1}=S_{k,i}\geq S_*(\alpha),\ \forall k\geq N+\bar{N}$. Hence, $\sup_{k\geq\bar{L}}P_{k,i}\leq S_*^{-1}(\alpha)$. Since the filter is conditionally consistent, $\sup_{k\geq\bar{L}}E\{(\hat{x}_{k,i}-x_k)(\hat{x}_{k,i}-x_k)^T|\mathcal{W}_k\}\leq S_*^{-1}(\alpha)$. Taking mathematical expectation of both sides and denoting $\eta=\frac{\bar{\eta}^{1-\bar{L}}}{a_{\min}}>0$, the conclusion of the theorem holds.

References

[1] B. S. Rao and H. F. Durrant-Whyte, “Fully decentralised algorithm for multisensor Kalman filtering,” in *IEE Proc. D. (Control Theory Appl.)*, vol. 138, 1991, pp. 413–420.

[2] J. Y. Yu, M. J. Coates, M. G. Rabbat, and S. Blouin, "A distributed particle filter for bearings-only tracking on spherical surfaces," *IEEE Signal Process. Lett.*, vol. 23, no. 3, pp. 326–330, Mar. 2016.
[3] A. A. Saucan, M. J. Coates, and M. Rabbat, "A multisensor multi-Bernoulli filter," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5495–5509, Oct. 2017.
[4] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. IEEE Conf. Decis. Control*, 2007, pp. 5492–5498.
[5] R. Olfati-Saber, "Kalman-consensus filter: Optimality, stability, and performance," in *Proc. IEEE Conf. Decis. Control Chin. Control Conf.*, 2009, pp. 7036–7042.
[6] U. A. Khan and J. M. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4919–4935, Oct. 2008.
[7] S. Kar and J. M. Moura, "Gossip and distributed Kalman filtering: Weak consensus under weak detectability," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1766–1784, Apr. 2011.
[8] S. D. Gupta, M. Coates, and M. Rabbat, "Error propagation in gossip-based distributed particle filters," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 3, pp. 148–163, Sep. 2015.
[9] V. Tuzlukov, *Signal Processing Noise*. Boca Raton, FL, USA: CRC Press, 2002.
[10] F. Yang, Z. Wang, and Y. Hung, "Robust Kalman filtering for discrete time-varying uncertain systems with multiplicative noises," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1179–1183, Jul. 2002.
[11] J. Feng, Z. Wang, and M. Zeng, "Distributed weighted robust Kalman filter fusion for uncertain systems with autocorrelated and cross-correlated noises," *Inf. Fusion*, vol. 14, no. 1, pp. 78–86, 2013.
[12] Y. Liu, Z. Wang, X. He, and D. Zhou, "Minimum-variance recursive filtering over sensor networks with stochastic sensor gain degradation: Algorithms and performance analysis," *IEEE Control Netw. Syst.*, vol. 3, no. 3, pp. 265–274, Sep. 2016.
[13] C. Wen, Z. Wang, Q. Liu, and F. E. Alsaadi, "Recursive distributed filtering for a class of state-saturated systems with fading measurements and quantization effects," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 6, pp. 930–941, Jun. 2018.
[14] W. Yang, G. Chen, X. Wang, and L. Shi, "Stochastic sensor activation for distributed state estimation over a sensor network," *Automatica*, vol. 50, no. 8, pp. 2070–2076, 2014.
[15] S. Dey, A. S. Leong, and J. S. Evans, "Kalman filtering with faded measurements," *Automatica*, vol. 45, no. 10, pp. 2223–2233, 2009.
[16] T. Zhou, "Robust recursive state estimation with random measurement droppings," *IEEE Trans. Autom. Control*, vol. 61, no. 1, pp. 156–171, Jan. 2016.
[17] V. Ugrinovskii, "Distributed robust filtering with H_∞ consensus of estimates," *Automatica*, vol. 47, no. 1, pp. 1–13, 2011.
[18] V. Stojanovic and D. Prsic, "Robust identification for fault detection in the presence of noises: Application to hydraulic servo drives," *Nonlinear Dyn.*, vol. 100, no. 3, pp. 2299–2313.
[19] V. Stojanovic and N. Nedic, "Joint state and parameter robust estimation of stochastic nonlinear systems," *Int. J. Robust Nonlinear Control*, vol. 26, no. 14, pp. 3058–3074, 2016.
[20] V. Stojanovic, S. He, and B. Zhang, "State and parameter joint estimation of linear stochastic systems in presence of faults and non-Gaussian noises," *Int. J. Robust Nonlinear Control*, vol. 30, no. 16, pp. 6683–6700, 2020.
[21] Y. Yin, P. Shi, F. Liu, K. L. Teo, and C.-C. Lim, "Robust filtering for nonlinear nonhomogeneous Markov jump systems by fuzzy approximation approach," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1706–1716, Sep. 2015.
[22] X. Dong, S. He, and V. Stojanovic, "Robust fault detection filter design for a class of discrete-time conic-type non-linear Markov jump systems with jump fault signals," *IET Control Theory Appl.*, vol. 14, no. 14, pp. 1912–1919, 2020.
[23] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, Sep. 2010.
[24] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Consensus based overlapping decentralized estimation with missing observations and communication faults," *Automatica*, vol. 45, no. 6, pp. 1397–1406, 2009.
[25] W. Li, Y. Jia, and J. Du, "Distributed filtering for discrete-time linear systems with fading measurements and time-correlated noise," *Digit. Signal Process.*, vol. 60, pp. 211–219, 2017.
[26] X. He, W. Xue, and H. Fang, "Consistent distributed state estimation with global observability over sensor network," *Automatica*, vol. 92, pp. 162–172, 2018.
[27] X. He, C. Hu, Y. Hong, L. Shi, and H. Fang, "Distributed Kalman filters with state equality constraints: Time-based and event-triggered communications," *IEEE Trans. Autom. Control*, vol. 65, no. 1, pp. 28–43, Jan. 2020.
[28] X. He, W. Xue, X. Zhang, and H. Fang, "Distributed filtering for uncertain systems under switching sensor networks and quantized communications," *Automatica*, vol. 114, 2020, Art no. 108842.
[29] S. Kar and J. M. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
[30] U. A. Khan and A. Jadbabaie, "Collaborative scalar-gain estimators for potentially unstable social dynamics with limited communication," *Automatica*, vol. 50, no. 7, pp. 1909–1914, 2014.
[31] A. Speranzon, C. Fischione, K. H. Johansson, and A. Sangiovanni-Vincentelli, "A distributed minimum variance estimator for sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 609–621, May 2008.
[32] G. Battistelli and L. Chisci, "Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.
[33] S. Wang and W. Ren, "On the convergence conditions of distributed dynamic state estimation using sensor networks: A unified framework," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 4, pp. 1300–1316, Jul. 2018.
[34] H. Ji, F. L. Lewis, Z. Hou, and D. Mikulski, "Distributed information-weighted Kalman consensus filter for sensor networks," *Automatica*, vol. 77, pp. 18–30, 2017.
[35] S. Das and J. M. Moura, "Consensus innovations distributed Kalman filter with optimized gains," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 467–481, Jan. 2017.
[36] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*. New York, NY, USA: Springer Science & Business, 2012.
[37] W. Yang, C. Yang, H. Shi, L. Shi, and G. Chen, "Stochastic link activation for distributed filtering under sensor power constraint," *Automatica*, vol. 75, pp. 109–118, 2017.
[38] W. Niehsen, "Information fusion based on fast covariance intersection filtering," in *Proc. Int. Conf. Inf. Fusion*, 2002, pp. 901–904.
[39] A. Poznyak, *Advanced Mathematical Tools for Automatic Control Engineers*: Volume 2: *Stochastic Systems*, vol. 2. Elsevier, 2009.
[40] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.
[41] J. Tugnait, "Stability of optimum linear estimators of stochastic signals in white multiplicative noise," *IEEE Trans. Autom. Control*, vol. 26, no. 3, pp. 757–761, Jun. 1981.
[42] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, vol. 15. Philadelphia, PA, USA: SIAM, 1994.
[43] S. N. Elaydi, *An Introduction to Difference Equations*. New York, NY, USA: Springer, 2005.



Xingkang He received the B.S. degree from the School of Mathematics, Hefei University of Technology, Hefei, China, in 2013 and the Ph.D. degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2018. Since October 2018, he has been a Postdoctoral Researcher with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include networked estimation theories, security of cyber-physical systems, event-based estimation and control, and autonomous vehicles. He was the recipient of Best Paper Award in 2018 IEEE Data Driven Control and Learning Systems Conference.

Karl Henrik Johansson (Fellow, IEEE) received the M.Sc. and Ph.D. degrees from Lund University, Lund, Sweden. He is currently a Professor with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden and the Director of digital futures. He has held visiting positions with UC Berkeley, Berkeley, CA, USA, Caltech, Pasadena, CA, USA, Nanyang Technological University, Singapore, HKUST Institute of Advanced Studies, Hong Kong, and NTNU, Trondheim, Norway. His research interests include networked control systems and cyber-physical systems with applications in transportation, energy, and automation networks. He is a Member of the Swedish Research Council's Scientific Council for Natural Sciences and Engineering Sciences. He was on the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and is currently the Vice-President of the European Control Association. He was the recipient of several best paper awards and other distinctions from IEEE, IFAC, and ACM, the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He was awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar with the Knut and Alice Wallenberg Foundation. He is a Fellow of the Royal Swedish Academy of Engineering Sciences, and he is also a IEEE Control Systems Society Distinguished Lecturer.

Haitao Fang received the B.S. degree in probability and statistics from Peking University, Beijing, China, in 1990, the M.S. degree in applied mathematics from Tsinghua University, Beijing, China, in 1993, and the Ph.D. degree in applied mathematics from the Peking University, in 1996 . From 1996 to 1998, he was a Postdoc with the Institute of Systems Science, Chinese Academy of Sciences (CAS), Beijing, China, and in 1998, joined the Institute as an Assistant Professor. He is currently with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, CAS, as a Full Professor. His research interests include dynamical behavior analysis of networked systems and distributed system estimation, optimization, and control. He is currently the Deputy Editor-in-Chief of the *Journal of Control Theory and Applications* and an Associate Editor for the *Acta Mathematicae Applicatae Sinica*.