Analysis of the Compressed Distributed Kalman Filter Over Markovian Switching Topology

Rongjiang Li, Die Gan, *Member, IEEE*, Siyu Xie, Haibo Gu, *Member, IEEE*, and Jinhu Lü, *Fellow, IEEE*

***Abstract*—This article investigates the distributed estimation problem of an unknown high-dimensional sparse state vector for a stochastic dynamic system. The communication topology randomly switches, and the switching law is governed by a time-homogeneous Markovian chain. By means of the compressed sensing (CS) theory and a diffusion strategy, we propose a compressed distributed Kalman filter (CDKF). That is, each sensor first compresses the original high-dimensional regression data. Then, the covariance intersection fusion rule is utilized to obtain a distributed Kalman filter (DKF) estimate in the compressed low-dimensional space. Afterward, the original high-dimensional sparse state vector can be well recovered by a reconstruction technique. In terms of stability analysis, one of the main difficulties lies in analyzing the product of nonindependent and nonstationary random matrices in the context of time-varying communication topologies. Relying on the stochastic stability theory, the Markov chain theory, and the CS theory, we establish the upper bound for the estimation error under the compressed cooperative excitation condition, which is much weaker than the traditional uncompressed collective observability conditions used in the existing literature. Finally, we provide a simulation example to illustrate the performance of the proposed algorithm.**

***Index Terms*—Compressed sensing (CS), distributed Kalman filter (DKF), Markovian switching topology, sparse state estimation, stochastic dynamic system.**

I. Introduction

Sensor networks are comprised of numerous spatially dispersed sensor nodes, which can share local information with neighbors to collaborate on complicated tasks. As one of these essential tasks, distributed state estimation over sensor networks has aroused extensive research interest in many areas, including spacecraft navigation, environmental monitoring, and so on. In the absence of a fusion center, distributed estimation algorithms have advantages over centralized ones in terms of robustness and scalability [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. Among them, most results are obtained over static and undirected networks. In reality, however, communication topologies may exhibit randomly switching or asymmetric properties due to link failures or reconstruction and sensor heterogeneity [16]. In [17] and [18], the randomly switching process is modeled as an independent and identically distributed (i.i.d.) process. However, it is common for randomly switching topologies to exhibit temporal correlation [19]. Consequently, some studies on distributed algorithms over sensor networks have modeled communication topologies as Markovian switching topologies to capture this correlation [19], [20], [21], [22]. For example, [20], [21], [22] investigated distributed estimation algorithms for deterministic or temporally independent observation matrices over sensor networks subject to Markovian switching topologies.

In many cases, the unknown state vectors to be estimated can be sparse, such as speech signals, image signals, solar waves, and so on [23], [24], [25], [26]. Given the prevalence of sparsity, efforts have been made to utilize sparsity as a prior to improve the estimation performance. One approach for sparse state estimation involves incorporating a regularization term into the cost function. For instance, Liu et al. [26] utilized ℓ_0- or ℓ_1-norm as the sparsity penalties and presented a recursive least squares-based distributed adaptive filter by assuming that observation matrices are i.i.d.. Similarly, Gan and Liu [27] proposed a distributed sparse identification algorithm by incorporating a ℓ_1-regularization term into the ℓ_2-estimation error. We remark that the literature mentioned above is concerned with the sparsity of unknown state vectors. Nevertheless, the sparsity of regression vectors also deserves attention.

Another approach for estimating sparse signals is the compressed sensing (CS) theory [28]. It facilitates the insufficient excitation (i.e., the degeneration of covariance matrices of regression vectors) triggered by the sparsity of the signal. For example, Xu et al. [29] designed a distributed compressed estimation algorithm on the basis of the CS technique and provided a simulation example to illustrate the advantages

Received 25 September 2024; revised 11 November 2024; accepted 13 November 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3305600, and in part by the National Natural Science Foundation of China under Grant 62141604 and Grant 62103015. This article was recommended by Associate Editor W. Xia. *(Corresponding author: Jinhu Lü.)*

Rongjiang Li is with the Key Laboratory of Systems and Control and the 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: lirongjiang@amss.ac.cn).

Die Gan is with the College of Artificial Intelligence, Nankai University, Tianjin 300350, China (e-mail: gandie@amss.ac.cn).

Siyu Xie is with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: syxie@uestc.edu.cn).

Haibo Gu and Jinhu Lü are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China, and also with Zhongguancun Laboratory, Beijing 100094, China (e-mail: guhaibo@amss.ac.cn; jhlu@iss.ac.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2024.3507275.

Digital Object Identifier 10.1109/TCYB.2024.3507275

of the proposed algorithm with regard to convergence rate and mean square error performance. In [30], they presented a compressed-combine-reconstruct-adaptive algorithm (CCRA) wherein the CS technique was applied to the diffusion stage. Then, the stability analysis was conducted under the independent assumptions of the regression vectors. We remark that most of the existing stability analysis for stochastic dynamic systems critically depends on such stringent assumptions as independence and stationarity [26], [30], [31], with only a few exceptions. For instance, Gan and Liu [32] and Xie and Guo [33] integrated the CS technique into the distributed least squares algorithm and the distributed normalized least mean squares algorithm, respectively. After that, the relatively elegant stability analysis was conducted.

Regarding the Kalman filter, its distributed forms are widely studied over sensor networks because of its optimality in the minimum mean square error sense when noise processes obey a Gaussian distribution. Most distributed Kalman filters (DKFs) were proposed for deterministic observation matrices or regression vectors (see, e.g., [3], [4], [5], and [6]). For example, Sebastián et al. [3] introduced the first event-triggered and certifiably optimal DKF for deterministic fixed observation matrices and proved the global asymptotic stability of the estimator and optimality under positive certification. In [4], a distributed Kalman-consensus filter was proposed for the deterministic time-invariant observation matrix, and its robustness margins were investigated. As for the deterministic time-varying observation matrices, Ma et al. [5] proposed a gossip-based DKF and provided a theoretical analysis of Lyapunov stability and convergence speed. Yang et al. [6] designed a DKF over delaying sensor networks and derived sufficient conditions for convergence on estimation errors and the boundedness of error covariances. It is worth mentioning that [34] designed a compressed Kalman filter for general sparse dynamic systems wherein regression vectors were stochastic. Since it was designed for the single sensor case, it cannot be directly applied to sensor networks. Besides, the asymmetric and switching characteristics of communication topologies make this state estimation problem more challenging.

In light of the above discussions, this article proposes a compressed DKF (CDKF) to cooperatively estimate the sparse state vector. In order to better characterize possible link failures or reconstructions in actual sensor networks, we model the communication topologies among sensors as Markovian switching topologies. As for the theoretical analysis, we establish the stability of CDKF under the compressed cooperative excitation condition. The main contributions are summarized in the following four aspects.

1) The distributed state estimation problem over Markovian switching topologies is investigated for stochastic dynamic systems, while the majority of existing literature focuses on the case where regression vectors are deterministic [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14].
2) A CDKF is proposed to cooperatively estimate high-dimensional sparse state vectors by taking advantage of the CS theory and the diffusion strategy.
3) The upper bound for the estimation error is established without independent and stationary signal assumptions, as commonly used in [26], [30], and [31]. Thus, the stability results of the CDKF are anticipated to be applicable to stochastic feedback systems.
4) The stability analysis is established under the compressed cooperative excitation condition, which is much weaker than the uncompressed ones in [3], [9], [10], and [35]. This suggests that the proposed algorithm may succeed in the sparse estimation task even if traditional uncompressed DKFs would fail due to insufficient excitation.

The remainder of this article is arranged as follows. The problem formulation is presented in Section II, and the main results are stated in Section III. In Section IV, an illustrative example is provided to demonstrate the performance of the algorithm. Finally, concluding remarks are made in Section V.

Notations: For a vector $a \in \mathbb{R}^n$, its ith element is denoted as $a_{(i)}$. The notation $\|a\|_{\ell_0}$ represents the number of nonzero elements in a. A vector a is called s-sparse if it has at most s nonzero elements (i.e., $\|a\|_{\ell_0} \leq s$). Additionally, the ℓ_1-norm $\|a\|_{\ell_1}$ is defined by $\|a\|_{\ell_1} = \sum_{i=1}^{n} |a_{(i)}|$. For two real symmetric matrices $A, B \in \mathbb{R}^{n\times n}$, $A > B$ ($A \geq B$) means that $A - B$ is a positive definite (semi-definite) matrix. The spectral norm of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as $\|A\| = \{\lambda_{\max}(AA^{\mathrm{T}})\}^{(1/2)}$ with $\lambda_{\max}(\cdot)$ being the maximum eigenvalue of A. The notation $(\cdot)^{\mathrm{T}}$ stands for the transpose operator. Also, $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of the matrix, and the n-dimensional square identity matrix is denoted by I_n. We use $\mathrm{tr}(A)$ to represent the trace of the corresponding matrix A. For a random matrix B, its L_p-norm is defined as $\|B\|_p = \{\mathbb{E}[\|B\|^p]\}^{(1/p)}$, $p \geq 1$ with $\mathbb{E}[\,\cdot\,]$ being the expectation operator. Notations $\mathbb{E}[\cdot|\cdot]$, $\mathbb{P}\{\cdot\}$, and $\mathbb{P}\{\cdot|\cdot\}$ describe the conditional expectation operator, probability measure, and conditional probability measure, respectively.

II. Problem Formulation

A. System Model

In a sensor network of n sensors, the observation model of each sensor i at time k is described by the following stochastic regression model:

$$z_{k,i} = h_{k,i}^{\mathrm{T}} x_k + v_{k,i}, \ \ k \geq 0, \ \ i = 1, \ldots, n \tag{1}$$

where $z_{k,i} \in \mathbb{R}$ is the observation, $h_{k,i} \in \mathbb{R}^m$ is the stochastic regression vector, $v_{k,i} \in \mathbb{R}$ is the local observation noise, and $x_k \in \mathbb{R}^m$ is the time-varying state of interest. The variation of the unknown state vector x_k is described as follows:

$$x_{k+1} = x_k + \omega_{k+1}, \ \ k \geq 0 \tag{2}$$

where $\omega_{k+1} \in \mathbb{R}^m$ represents the process noise.

In practical scenarios, such as high-dimensional data classification and channel estimation [23], [24], [25], [26], sparsity appears not only in the state x_k of interest but also in the regression vector $h_{k,i}$. Here, we concentrate on the case where $h_{k,i}$ and x_k are $3s$-sparse and s-sparse, respectively (i.e., they contain at most $3s$ and s nonzero elements, respectively).

Remark 1: Compared with the vast majority of research [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14] that focuses on deterministic regression vectors or observation matrices, here, the regression vector $h_{k,i}$ is assumed to be stochastic, and thus the regression model (1) admits feedback control. For instance, if $h_{k,i} = [z_{k-1,i}, \dots, z_{k-p,i}, u_{k,i}, \dots, u_{k-q,i}]$ with the input signal $u_{k,i}$ being the control law $u_{k,i} = f(z_{j,i}, j \le k-1)$, then the regression model (1) can be reduced to the autoregressive model with exogenous inputs. Note that $h_{k,i}$ composed of current and past input-output data is stochastic and fails to meet strict assumptions commonly found in the study of stochastic dynamic systems, such as i.i.d. condition. This fact calls for further investigation of assumptions imposed on regression vectors of stochastic dynamic systems.

Given that the regression vector $h_{k,i}$ is stochastic, we specify a simpler model (1) and (2) compared to the linear system model in [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], and [36] regarding the state evolution matrix. In further work, more general linear time-varying system models will be considered for broader range of applications.

B. Markovian Switching Topologies

The communication links between sensors are modeled by a weighted digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \dots, n\}$ is the set of nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of directed edges, and $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{n\times n}$ is the weighted adjacency matrix. The ordered pair $(i, j) \in \mathcal{E}$ if and only if there is a communication link from node i to node j. The notation $\mathcal{N}_i = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}\}$ represents the set of neighbors of the node i and the node i is also included. All elements in $\mathcal{A}$ are non-negative, and $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise. Here, we assume that $\mathcal{A}$ is doubly stochastic (i.e., each row and column sum to 1). Note that $\mathcal{A}$ may be asymmetric. A directed path from node i_1 to i_ℓ is an ordered sequence of nodes $i_1, i_2, \dots, i_\ell$ such that $(i_t, i_{t+1}) \in \mathcal{E}$ for $t = 1, \dots, \ell - 1$. A digraph is strongly connected if there is a directed path between any pair of distinct nodes.

Affected by uncertainties, such as link failures and packet losses, the communication topology is no longer fixed. Thus, the communication topology is considered to be stochastically time-varying, and it can be described by a stochastically switching digraph, which stochastically switches among $\bar{s}$ different digraphs $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{\bar{s}}$. At time k, the communication topology is denoted by $\mathcal{G}_{m(k)} = (\mathcal{V}, \mathcal{E}_{m(k)}, \mathcal{A}_{m(k)})$, where the switching process $m(k)$ is driven by a time-homogeneous Markov chain taking values on a finite set $\mathbb{S} = \{1, 2, \dots, \bar{s}\}$.

This article aims to provide a distributed algorithm to cooperatively estimate the time-varying sparse state vector over the Markovian switching topology. Moreover, this article attempts to find relatively mild assumptions to establish upper bounds on the estimation error.

C. Compressed Distributed Kalman Filter

According to the covariance intersection fusion rule, several DKFs are proposed to cooperatively estimate the unknown state vector via local communications and interactions among sensors [9], [10], [11]. In detail, each sensor generates the local estimates based on its measurements and then fuses the estimate among neighbors. Note that our setting is based on the sparsity of the state vector and the switching characteristics of communication topologies. Thus, the analysis methods in the above literature cannot be directly applied. Now, we propose the CDKF over the Markovian switching directed topologies, which relies on the CS theory to reduce the computational cost and improve the estimation performance in sparse and high-dimensional scenarios (see Algorithm 1).

Specifically, at every time instant k, the compressed regression vector $\varphi_{k,i} = Dh_{k,i} \in \mathbb{R}^d$ is obtained with the help of a prespecified sensing matrix $D \in \mathbb{R}^{d\times m}$ (with $d \ll m$.[1]) For the generation of the sensing matrix D, please refer to Appendix A for more details. Then in step 2, we adopt the adapt-then-combine strategy to update a low-dimensional estimate $\hat{\zeta}_{k,i}$ for the compressed state vector $\zeta_k = Dx_k \in \mathbb{R}^d$. Note that $\zeta_{k+1} = \zeta_k + \bar{\omega}_{k+1}$, where $\bar{\omega}_k = D\omega_k$. Also, the original model (1) can be rewritten as follows:

$$\begin{aligned} z_{k,i} &= h_{k,i}^{\mathrm{T}} x_k + v_{k,i} \\ &= \varphi_{k,i}^{\mathrm{T}} \zeta_k + h_{k,i}^{\mathrm{T}} x_k - \varphi_{k,i}^{\mathrm{T}} \zeta_k + v_{k,i} \\ &= \varphi_{k,i}^{\mathrm{T}} \zeta_k + h_{k,i}^{\mathrm{T}} \big[I_m - D^{\mathrm{T}} D\big] x_k + v_{k,i} \\ &\triangleq \varphi_{k,i}^{\mathrm{T}} \zeta_k + \bar{v}_{k,i}. \end{aligned} \tag{3}$$

In this case, we regard $\bar{v}_{k,i} = h_{k,i}^{\mathrm{T}}[I_m - D^{\mathrm{T}}D]x_k + v_{k,i}$ as the new "noise" term. Hence, for the adaptation step, every sensor performs the Kalman iteration based on the new "measurements" $\{z_{k,i}, \varphi_{k,i}\}_{i\in\mathcal{V}}$ with $r_i > 0$ and $\Sigma > 0$ being arbitrarily chosen. Then, for the combination step, the diffusion strategy is utilized by exchanging estimates with neighboring sensors and fusing the collected estimates through a convex combination. Finally, in step 3, we tackle convex optimization problem (8) to recover a high-dimensional estimate $\{\hat{x}_{k+1,i}\}_{i\in\mathcal{V}}$ for the original state vector x_k. Here, $\bar{C}$ represents the bound of the estimation error $\|\tilde{\zeta}_{k,i}\|$.

III. Main Results

In this section, we will establish the exponential stability and the estimation error bound for the proposed CDKF (i.e., Algorithm 1) without requiring independent and stationary assumptions on the system signals. For this purpose, we first derive the compressed error equation for Algorithm 1.

A. Compressed Error Equation

For the sensor i, introduce the following two compressed estimation errors: $\tilde{\zeta}_{k,i} = \zeta_k - \hat{\zeta}_{k,i}$ and $\check{\zeta}_{k,i} = \zeta_k - \bar{\zeta}_{k,i}$. Then from (6) and (7), we have

$$\begin{aligned} \tilde{\zeta}_{k+1,i} &= \zeta_{k+1} - P_{k+1,i} \sum_{l\in\mathcal{N}_{i,m(k)}} a_{li,m(k)} \bar{P}_{k+1,l}^{-1} \bar{\zeta}_{k+1,l} \\ &= P_{k+1,i} \sum_{l\in\mathcal{N}_{i,m(k)}} a_{li,m(k)} \bar{P}_{k+1,l}^{-1} \zeta_{k+1} \\ &\quad - P_{k+1,i} \sum_{l\in\mathcal{N}_{i,m(k)}} a_{li,m(k)} \bar{P}_{k+1,l}^{-1} \bar{\zeta}_{k+1,l} \\ &= P_{k+1,i} \sum_{l\in\mathcal{N}_{i,m(k)}} a_{li,m(k)} \bar{P}_{k+1,l}^{-1} \check{\zeta}_{k+1,l}. \end{aligned} \tag{9}$$

[1]The relation $d \ll m$ means that d is much smaller than m.

Algorithm 1 CDKF

Input: $\{h_{k,i}, z_{k,i}\}_{i\in\mathcal{V}}$, $k = 0, 1, 2, \ldots$; sensing matrix D
Output: $\{\hat{x}_{k+1,i}\}_{i\in\mathcal{V}}$, $k = 0, 1, 2, \ldots$
for every sensor $i \in \mathcal{V}$ **do**
Initialization: Begin with an initial value $\hat{\zeta}_{0,i}$ and an initial positive definite matrix $P_{0,i} > 0$.
for each time $k = 0, 1, 2, \ldots$ **do**
Step 1. Compression: $\varphi_{k,i} = Dh_{k,i}$.
Step 2. Estimation in a low-dimensional dimension.
i) Adaptation process

$$L_{k,i} = P_{k,i}\varphi_{k,i}\left(r_i + \varphi_{k,i}^{\mathrm{T}} P_{k,i}\varphi_{k,i}\right)^{-1}$$

$$\bar{\zeta}_{k+1,i} = \hat{\zeta}_{k,i} + L_{k,i}\left(z_{k,i} - \varphi_{k,i}^{\mathrm{T}}\hat{\zeta}_{k,i}\right) \tag{4}$$

$$\bar{P}_{k+1,i} = P_{k,i} - L_{k,i}\varphi_{k,i}^{\mathrm{T}} P_{k,i} + \Sigma. \tag{5}$$

ii) Combination process

$$P_{k+1,i}^{-1} = \sum_{l\in\mathcal{N}_{i,m(k)}} a_{li,m(k)}\bar{P}_{k+1,l}^{-1} \tag{6}$$

$$\hat{\zeta}_{k+1,i} = P_{k+1,i}\sum_{l\in\mathcal{N}_{i,m(k)}} a_{li,m(k)}\bar{P}_{k+1,l}^{-1}\bar{\zeta}_{k+1,l}. \tag{7}$$

Step 3. Reconstruction:

$$\hat{x}_{k+1,i} = \arg\min_{x\in\mathcal{X}}\|x\|_{\ell_1} \tag{8}$$

where $\mathcal{X} = \{x \in \mathbb{R}^m | \|Dx - \hat{\zeta}_{k+1,i}\| \le \bar{C}\}$.

From (3) and (4), we can obtain that

$$\begin{aligned}\check{\zeta}_{k+1,i} &= \zeta_{k+1} - \bar{\zeta}_{k+1,i}\\ &= \zeta_k + \bar{\omega}_{k+1} - \hat{\zeta}_{k,i} - L_{k,i}\left(z_{k,i} - \varphi_{k,i}^{\mathrm{T}}\hat{\zeta}_{k,i}\right)\\ &= \tilde{\zeta}_{k,i} + \bar{\omega}_{k+1} - L_{k,i}\left(\varphi_{k,i}^{\mathrm{T}}\zeta_k - \varphi_{k,i}^{\mathrm{T}}\hat{\zeta}_{k,i} + \bar{v}_{k,i}\right)\\ &= \left(I_d - L_{k,i}\varphi_{k,i}^{\mathrm{T}}\right)\tilde{\zeta}_{k,i} - L_{k,i}\bar{v}_{k,i} + \bar{\omega}_{k+1}.\end{aligned} \tag{10}$$

For convenience, the following notations are introduced to write the above compressed error equation in a compact form: With notations in Table I, we rewrite (4)–(7) in Algorithm 1 as follows:

$$\begin{cases}\bar{\Lambda}_{k+1} = \hat{\Lambda}_k + L_k\left(Z_k - \Xi_k^{\mathrm{T}}\hat{\Lambda}_k\right)\\ \bar{P}_{k+1} = P_k - L_k\Lambda_k^{\mathrm{T}}P_k + \bar{\Sigma}\\ \mathrm{vec}\{P_{k+1}^{-1}\} = \mathscr{A}_{m(k)}^{\mathrm{T}}\mathrm{vec}\{\bar{P}_{k+1}^{-1}\}\\ \hat{\Lambda}_{k+1} = P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}\bar{\Lambda}_{k+1}\end{cases} \tag{11}$$

where $\mathrm{vec}\{\cdot\}$ stands for the operator which stacks the blocks of the block diagonal matrix on top of each other. Also, on account of $\tilde{\Lambda}_k = \Lambda_k - \hat{\Lambda}_k$ and $\check{\Lambda}_k = \Lambda_k - \bar{\Lambda}_k$, applying (10) yields that

$$\check{\Lambda}_{k+1} = \left(I_{dn} - L_k\Xi_k^{\mathrm{T}}\right)\tilde{\Lambda}_k - L_k\bar{V}_k + \bar{\Omega}_{k+1}.$$

Then according to (9), we obtain the compressed error equation as follows:

$$\begin{aligned}\tilde{\Lambda}_{k+1} &= P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}\check{\Lambda}_{k+1}\\ &= P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}\left(I_{dn} - L_k\Xi_k^{\mathrm{T}}\right)\tilde{\Lambda}_k\\ &\quad - P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}L_k\bar{V}_k\\ &\quad + P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}\bar{\Omega}_{k+1}.\end{aligned} \tag{12}$$

TABLE I
SOME NOTATIONS

Notation	Definition	Dimension
Z_k	$\mathrm{col}\{z_{k,1},\ldots,z_{k,n}\}$	$n\times 1$
X_k	$\mathrm{col}\{x_k,\ldots,x_k\}$	$mn\times 1$
V_k	$\mathrm{col}\{v_{k,1},\ldots,v_{k,n}\}$	$n\times 1$
$\bar{V}_k$	$\mathrm{col}\{\bar{v}_{k,1},\ldots,\bar{v}_{k,n}\}$	$n\times 1$
Ξ_k	$\mathrm{diag}\{\varphi_{k,1},\ldots,\varphi_{k,n}\}$	$dn\times n$
Ω_k	$\mathrm{col}\{\omega_k,\ldots,\omega_k\}$	$mn\times 1$
$\bar{\Omega}_k$	$\mathrm{col}\{\bar{\omega}_k,\ldots,\bar{\omega}_k\}$	$dn\times 1$
L_k	$\mathrm{diag}\{L_{k,1},\ldots,L_{k,n}\}$	$dn\times n$
P_k	$\mathrm{diag}\{P_{k,1},\ldots,P_{k,n}\}$	$dn\times dn$
$\bar{P}_k$	$\mathrm{diag}\{\bar{P}_{k,1},\ldots,\bar{P}_{k,n}\}$	$dn\times dn$
$\bar{\Sigma}$	$\mathrm{diag}\{\Sigma,\ldots,\Sigma\}$	$dn\times dn$
Λ_k	$\mathrm{col}\{\zeta_k,\ldots,\zeta_k\}$	$dn\times 1$
$\hat{\Lambda}_k$	$\mathrm{col}\{\hat{\zeta}_{k,1},\ldots,\hat{\zeta}_{k,n}\}$	$dn\times 1$
$\bar{\Lambda}_k$	$\mathrm{col}\{\bar{\zeta}_{k,1},\ldots,\bar{\zeta}_{k,n}\}$	$dn\times 1$
$\tilde{\Lambda}_k$	$\mathrm{col}\{\tilde{\zeta}_{k,1},\ldots,\tilde{\zeta}_{k,n}\}$	$dn\times 1$
$\check{\Lambda}_k$	$\mathrm{col}\{\check{\zeta}_{k,1},\ldots,\check{\zeta}_{k,n}\}$	$dn\times 1$
$\mathscr{A}_k$	$\mathcal{A}_k\otimes I_m$	$mn\times mn$

B. Some Definitions

The compressed error equation (12) can be regarded as the vector random linear equation $\Theta_{k+1} = A_k\Theta_k + \rho_{k+1}$, $k \ge 0$ where $\{A_k, \ k \ge 0\}$ is a sequence of $dn \times dn$ random matrices. Roughly speaking, the stability of Θ_k is tied to the stability of its homogeneous part, that is, $\Theta_{k+1} = A_k\Theta_k$, which relies on the product of random matrices. To this end, we introduce some definitions on the stability of random matrices.

Definition 1 [37]: For a sequence of $d \times d$ random matrices $B = \{B_k, \ k \ge 0\}$, we say that $\{I - B_k, \ k \ge 0\}$ is L_p-exponentially stable ($p \ge 1$) with parameter $\tau \in [0, 1)$ if B belongs to the set

$$\mathcal{S}_p(\tau) = \left\{B : \left\|\prod_{t=s+1}^{k}(I - B_t)\right\|_p \le M\tau^{k-s}\right.$$
$$\left.\forall k \ge s+1 \quad \forall s \ge 0, \text{ for some } M > 0\right\}.$$

Definition 2 [37]: For a scalar sequence $\beta = \{\beta_k, k \ge 0\}$, we define

$$\mathcal{S}^0(\tau) = \left\{\beta : \beta_k \in [0,1], \mathbb{E}\left[\prod_{t=s+1}^{k}(1 - \beta_t)\right] \le M\tau^{k-s}\right.$$
$$\left.\forall k \ge s+1 \quad \forall s \ge 0, \ \text{for some } M > 0\right\}$$

where $\tau \in [0, 1)$.

As demonstrated in [37] and [38], $\{I - A_k, k \ge 0\} \in \mathcal{S}_p(\cdot)$ is the sufficient and necessary condition for the stability of Θ_k in a sense. Verifying whether the random matrix sequence $\{I - A_k, k \ge 0\}$ belongs to the set $\mathcal{S}_p(\cdot)$ poses a significant mathematical challenge, particularly without the assumptions of independence and stationarity for the signals. It is worth noting that for random linear equations arising from adaptive filtering algorithms, including the Kalman filter algorithm, the investigation of $\mathcal{S}_p(\cdot)$ can be simplified to that of a scalar sequence in $\mathcal{S}^0(\cdot)$. This corresponding scalar sequence can then be studied based on excitation or observability conditions imposed on the regression vectors [37], [38]. For further analysis, the lemma on the set $\mathcal{S}^0(\cdot)$ is introduced as follows.

Lemma 1 [37]: For two scalar sequences $\iota = \{\iota_k,\ k \geq 0\}$ and $\nu = \{\nu_k,\ k \geq 0\}$:

1) If $0 \leq \iota_k \leq \nu_k \leq 1$ holds for any $k \geq 0$, then $\{\iota_k\} \in \mathcal{S}^0(\tau)$ implies $\{\nu_k\} \in \mathcal{S}^0(\tau)$.
2) Let $\{\iota_k\} \in \mathcal{S}^0(\tau)$ and $\iota_k \leq \iota^* < 1\ \forall k \geq 0$ where ι^* is a constant. Then for any $\varepsilon \in (0,1)$, $\{\varepsilon\iota_k\} \in \mathcal{S}^0(\tau^{(1-\iota^*)\varepsilon})$.
3) Let $\iota = \{\iota_k, \mathcal{H}_k\}$ and $\nu = \{\nu_k, \mathcal{H}_k\}$ be adapted processes, such that $\iota_k \in [0,1]$, $\mathbb{E}[\iota_{k+1}|\mathcal{H}_k] \geq \nu_k$, $k \geq 0$. Then $\{\nu_k\} \in \mathcal{S}^0(\tau)$ implies that $\{\iota_k\} \in \mathcal{S}^0(\sqrt{\tau})$.

C. Assumptions

For stability analysis, we need the following assumptions:

Assumption 1 [Restricted Isometry Property (RIP)]: The sensing matrix $D \in \mathbb{R}^{d\times m}$ satisfies the RIP with order $4s$ where the $4s$-restricted isometry constant is denoted as δ_{4s} (see Definition 3).

Assumption 2 (Compressed Cooperative Excitation Condition): For the adapted sequences $\{\varphi_{k,i}, \mathcal{F}_k, k \geq 0\}_{i\in\mathcal{V}}$, there exists an integer $h > 0$ such that $\{\tau_k, k \geq 0\} \in \mathcal{S}^0(\tau)$ for some $\tau \in (0,1)$, where τ_k is defined by

$$\tau_k \triangleq \lambda_{\min}\left(\mathbb{E}\left[\frac{1}{n(1+h)}\sum_{i=1}^{n}\sum_{j=kh+1}^{(k+1)h}\frac{\varphi_{j,i}\varphi_{j,i}^{\mathrm{T}}}{1+\|\varphi_{j,i}\|^2}\middle|\mathcal{F}_{kh}\right]\right)$$

with $\mathcal{F}_k = \sigma\{h_{j,i}, \omega_j, v_{j-1,i}, j \leq k, i \in \mathcal{V}\}$.

Remark 2: Essentially, Assumption 2 is a stochastic collective observability condition for the compressed regression vectors. To clarify, we give some intuitive explanations in terms of "excitation," "collective," and "compressed," with the latter two characteristics further illustrated in the simulation section.

1) *Excitation:* Assumption 2 says that the smallest eigenvalue τ_k is not "too small."
2) *Collective:* Compared to the excitation condition for the single case in [34] and [37], Assumption 2 contains not only temporal union information but also spatial union information of all the sensors, which implies that multiple sensors can cooperate to accomplish the estimation task even if any individual sensor cannot. Similar collective conditions are commonly exerted on the deterministic coefficients when studying the stability of the uncompressed DKF (cf. [3], [9], [10], [35]).
3) *Compressed:* Compared to collective observability conditions in [3], [9], [10], and [35], Assumption 2 is assumed for the compressed regression vectors $\{\varphi_{k,i}\}$ rather than the original ones $\{h_{k,i}\}$. Therefore, Assumption 2 is much weaker than that in [3], [9], [10], and [35], which implies that the CDKF may still get the compressed estimation results stably even if the uncompressed DKFs cannot fulfill the estimation tasks.

Assumption 3 (Markovian Switching Network Topology Assumption): The following assumptions are imposed on the network topology:

1) The union of all possible digraphs $\{\mathcal{G}_1,\ldots,\mathcal{G}_{\bar{s}}\}$ is strongly connected.
2) The Markov chain $\{m(k),\ k \geq 0\}$ is irrespective to $\mathcal{F}_k$. It is ergodic with the transition probability matrix $P = [p_{ab}] \in \mathbb{R}^{\bar{s}\times\bar{s}}$ where $p_{ab} = \mathbb{P}\{m(k+1) = b|m(k) = a\}$.

Remark 3: Some remarks on the Markovian graph model are given below.

1) Roughly speaking, Assumption 3 1) is quite weak in the sense that it is permissible for the sensor network to disconnect at any time k.
2) Assuming ergodicity, each state of the Markov chain can be reached from any other state in the state space in a positive probability. In other words, there exists an integer $q_0 \geq 0$ such that

$$\mathbb{P}\{m(k+q_0) = b|m(k) = a\} > 0 \tag{13}$$

holds for all k and all states $a, b \in \mathbb{S}$.

D. Stability Results

Note that the theoretical analysis is based on the sparsity in the system model and the switching characteristics of communication topologies. Therefore, the analysis methods in [9], [10], [11], [33], and [34] cannot be directly applied. Now, we list the sketch of stability analysis for Algorithm 1 as follows.

1) In Lemma 2 we derive conditions for the exponential stability of the homogeneous part of the compressed error equation (12).
2) In Lemma 4 we investigate the properties of $\{P_k\}$ based on Lemma 3, thus verifying the conditions in Lemma 2.
3) According to Lemmas 2 and 4, we obtain the exponential stability of the homogeneous part of the compressed error equation (12) in Theorem 1. Also, we establish an upper bound for the estimation error of the compressed state vector ζ_k.
4) Based on Theorem 1 and Lemma 6, we establish the upper bound for the estimation error of the original state vector x_k in Theorem 2.
5) According to Theorem 2, we derive the estimation error bound for the original state vector in probability in Corollary 1.

For Algorithm 1, it is clear that by (11), we have

$$\begin{cases}\bar{P}_{k+1} = \left(I_{dn} - L_k\Xi_k^{\mathrm{T}}\right)P_k\left(I_{dn} - L_k\Xi_k^{\mathrm{T}}\right)^{\mathrm{T}} + RL_kL_k^{\mathrm{T}} + \bar{\Sigma} \\ \mathrm{vec}\left\{P_k^{-1}\right\} = \mathscr{A}_{m(k-1)}^{\mathrm{T}}\mathrm{vec}\left\{\bar{P}_k^{-1}\right\}\end{cases} \tag{14}$$

where $P_0 > 0$. Based on this equation, the following lemma from [38] outlines conditions for the L_p-exponential stability of the homogeneous part of the compressed error equation (12). For simplicity, denote $\Sigma_k \triangleq RL_kL_k^{\mathrm{T}} + \bar{\Sigma}$ and $R \triangleq \mathrm{diag}\{r_1,\ldots,r_n\}\otimes I_m$. The lemma is then stated as follows.

Lemma 2: For the random recursive equation (14), we have for all $t > s$

$$\left\|\prod_{k=s}^{t-1} P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}\left(I_{dn} - L_k\Xi_k^{\mathrm{T}}\right)\right\|^2 \leq \left\{\prod_{k=s}^{t-1}\left(1 - \frac{1}{1 + \left\|\Sigma_k^{-1}\bar{P}_{k+1}\right\|}\right)\right\}\cdot\left\{\|P_t\|\cdot\|\bar{P}_s^{-1}\|\right\}.$$

Moreover, if $\{P_k\}$ satisfies the following two assumptions: 1) $\{1/(1 + \|\Sigma_k^{-1}\bar{P}_{k+1}\|)\} \in \mathcal{S}^0(\tau)$, for some $\tau \in [0, 1)$ and 2) $\sup_{t\geq s\geq 0}\|(\|P_t\|\cdot\|\bar{P}_s^{-1}\|)\|_p < \infty$, for some $p \geq 1$, then we have

$$\left\{I_{dn} - P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}\left(I_{dn} - L_k\Xi_k^{\mathrm{T}}\right)\right\} \in \mathcal{S}_p\left(\tau^{1/2p}\right). \quad (15)$$

Remark 4: Intuitively, this lemma implies that the investigation of (15) can be simplified to verify whether a certain scalar sequence belongs to the set $\mathcal{S}^0(\cdot)$ and whether a certain process is bounded in the sense of L_p-norm.

Before verifying the two assumptions on $\{P_k\}$ in Lemma 2, we first analyze the following properties of $\{P_k\}$. For simplicity of notation, we denote A_k^t as the matrix representing the product of weighted adjacency matrices during the time interval $[k, t]$, expressed as $\mathscr{A}_{m(k)}\mathscr{A}_{m(k+1)}\ldots\mathscr{A}_{m(t)}$. The element in row i, column j of A_k^t is denoted as $A_k^t(i, j)$. Additionally, we denote $\Psi_i(\rho h', k-1) = \sum_{j=1}^n A_{\rho h'}^{k-1}(j, i)\ P_{\rho h',j} + h'\Sigma$, $\Psi'(\rho h') = \sum_{j=1}^n P_{\rho h',j} + h'\Sigma$, and $z_\rho = \lfloor([\rho h' + n\bar{s}q_0]/h)\rfloor + 1$. The operator $\lfloor a\rfloor$ rounds a to the nearest integer less than or equal to a.

Lemma 3: Let $\{P_k\}$ be generated by (11). Then we have

$$T_{\rho+1} \leq \left(1 - b_{\rho+1}\right)T_\rho + d'$$

where

$$T_\rho = \sum_{k=z_{\rho-1}h+1}^{(z_{\rho-1}+1)h}\mathrm{tr}(P_{k+1}),\ \ T_0 = 0$$

$$c_{\rho+1}^1 = \mathrm{tr}\left(\sum_{k=z_\rho h+1}^{(z_\rho+1)h}\sum_{j=1}^n\Psi_j^2\left(\rho h', k-1\right)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1 + \left\|\varphi_{k,j}\right\|^2}\right)$$

$$c_{\rho+1}^2 = \sum_{j=1}^n\left(r_j + 1\right)\cdot\left(1 + \lambda_{\max}\left(\Psi'\left(\rho h'\right)\right)\right)\cdot\mathrm{tr}\left(\Psi'\left(\rho h'\right)\right)$$

$$b_{\rho+1} \triangleq \frac{c_{\rho+1}^1}{nhc_{\rho+1}^2}$$

$$d' = \frac{3}{2}nh\left(h' + 1\right)\mathrm{tr}(\Sigma)$$

and $h' = h + n\bar{s}q_0$ with h, q_0 being constants defined in Assumption 2 and Remark 3, respectively.

The proof of Lemma 3 is listed in Appendix B. Using this lemma, we verify the two assumptions on $\{P_k\}$ in Lemma 2 as follows.

Lemma 4: For $\{P_k\}$ and $\{\bar{P}_k\}$ generated by (11), provided that Assumptions 2 and 3 are satisfied, then we obtain the following.

1) There exists a positive constant ε^* such that for any $\varepsilon \in [0, \varepsilon^*)$, $\sup_{k\geq 0}\mathbb{E}[\exp(\varepsilon\|P_k\|)] < \infty$.
2) For any $\mu \in (0, 1]$, there exists a constant $\tau \in (0, 1)$ such that $\{\mu/(1 + \|\bar{\Sigma}^{-1}\|\cdot\|\bar{P}_{k+1}\|)\} \in \mathcal{S}^0(\tau)$.

Proof: We first prove Lemma 4 1): by (13) in Remark 3, we can see that there exists a positive constant p_0 such that for all ρ

$$\begin{aligned}
&\mathbb{P}\big\{m(\rho + n\bar{s}q_0) = \bar{s}, m(\rho + (n\bar{s} - 1)q_0) = \bar{s} - 1, \ldots\\
&\quad m(\rho + ((n-1)\bar{s} + 1)q_0) = 1\\
&\quad m(\rho + 2\bar{s}q_0) = \bar{s}, m(\rho + (2\bar{s} - 1)q_0) = \bar{s} - 1, \ldots\\
&\quad m(\rho + (\bar{s} + 1)q_0) = 1\\
&\quad m(\rho + \bar{s}q_0) = \bar{s}, m(\rho + (\bar{s} - 1)q_0) = \bar{s} - 1, \ldots\\
&\quad m(\rho + q_0) = 1\big|m(\rho)\big\}\\
&= \mathbb{P}\big\{m(\rho + n\bar{s}q_0) = \bar{s}\big|m(\rho + (n\bar{s} - 1)q_0) = \bar{s} - 1\big\}\ldots\\
&\quad\mathbb{P}\big\{m(\rho + ((n-1)\bar{s} + 1)q_0) = 1\big|m(\rho + ((n-1)\bar{s})q_0) = \bar{s}\big\}\\
&\quad\ldots\mathbb{P}\big\{m(\rho + q_0) = 1\big|m(\rho)\big\}\\
&\geq p_0 > 0.
\end{aligned} \quad (16)$$

By (16), we know that the Markov chain $\{m(k), k \geq 0\}$ can visit all states in $\mathbb{S}$ with n times in a positive probability during the time interval $[\rho + q_0, \rho + n\bar{s}q_0]$.

We are now in a position to analyze the term $\mathbb{E}[b_{\rho+1}|\mathcal{F}_{z_\rho h}]$. By Assumption 3 and [39, Lemma 5.1], we know that for $k \in [z_\rho h + 1, (z_\rho + 1)h]$, there exists positive constants σ_0 and σ such that the following inequality holds:

$$\begin{aligned}
\mathbb{E}\left[A_{\rho h'+1}^{k-1}(u, j)|\mathcal{F}_k'\right] &= \mathbb{E}\left[A_{\rho h'+1}^{k-1}(u, j)\middle|m\left(\rho h'\right)\right]\\
&\geq \sigma_0\mathbb{E}\left[A_{\rho h'+q_0}^{k-1}(u, j)\middle|m\left(\rho h'\right)\right] \geq \sigma, a.s
\end{aligned}$$

where $\mathcal{F}_k'$ is a σ-algebra generated by $\{m(1), \ldots, m(\rho h')\}$ and $\mathcal{F}_k$. Clearly, $\sigma \in (0, 1]$. Then, by [40], it yields that

$$\begin{aligned}
&\mathbb{E}\left[\Psi_j^2(k-1)\middle|\mathcal{F}_k'\right]\\
&= \mathbb{E}\left[\left(\sum_{l=1}^n A_{\rho h'}^{k-1}(l, j)\ P_{\rho h',l} + h'\Sigma\right)^2\middle|\mathcal{F}_k'\right]\\
&\geq \left\{\mathbb{E}\left[\sum_{l=1}^n A_{\rho h'}^{k-1}(l, j)\ P_{\rho h',l} + h'\Sigma\middle|\mathcal{F}_k'\right]\right\}^2\\
&= \left\{\sum_{l=1}^n\mathbb{E}\left[A_{\rho h'}^{k-1}(l, j)\ \middle|\mathcal{F}_k'\right]P_{\rho h',l} + h'\Sigma\right\}^2\\
&\geq \left(\sigma\sum_{l=1}^n P_{\rho h',l} + h'\Sigma\right)^2\\
&\geq \sigma^2\left(\sum_{l=1}^n P_{\rho h',l} + h'\Sigma\right)^2 \triangleq \sigma^2\left(\Psi'(\rho h')\right)^2.
\end{aligned}$$

By $\mathcal{F}_{z_\rho h} \subset \mathcal{F}_k$ and $\varphi_{k,j} \in \mathcal{F}_k$, we conclude that

$$\begin{aligned}
&\mathbb{E}\left[\Psi_j^2(k-1)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1 + \left\|\varphi_{k,j}\right\|^2}\middle|\mathcal{F}_{z_\rho h}'\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\Psi_j^2(k-1)\middle|\mathcal{F}_k'\right]\cdot\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1 + \left\|\varphi_{k,j}\right\|^2}\middle|\mathcal{F}_{z_\rho h}'\right]\\
&\geq \sigma^2\mathbb{E}\left[\left(\Psi'\left(\rho h'\right)\right)^2\cdot\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1 + \left\|\varphi_{k,j}\right\|^2}\middle|\mathcal{F}_{z_\rho h}'\right].
\end{aligned}$$

From the above analysis, we can obtain that

$$\begin{aligned}&\mathbb{E}\Big[b_{\rho+1}\Big|\mathcal{F}'_{z_\rho h}\Big]\\&\geq\frac{\sigma^2}{nhc_{\rho+1}^2}\\&\quad\cdot\operatorname{tr}\left(\sum_{k=z_\rho h+1}^{(z_\rho+1)h}\sum_{j=1}^{n}\mathbb{E}\left[\left(\Psi'(\rho h')\right)^2\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\left\|\varphi_{k,j}\right\|^2}\middle|\mathcal{F}'_{z_\rho h}\right]\right)\\&\geq\frac{1}{nhc_{\rho+1}^2}\sigma^2 n(1+h)\tau'_\rho\operatorname{tr}\Big(\big(\Psi'(\rho h')\big)^2\Big)\end{aligned}$$

where

$$\tau'_\rho=\lambda_{\min}\left(\mathbb{E}\left[\frac{1}{n(1+h)}\sum_{j=1}^{n}\sum_{k=z_\rho h+1}^{(z_\rho+1)h}\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\middle|\mathcal{F}_{z_\rho h}\right]\right).$$

It is clear that T_ρ and b_ρ are $\bar{\mathcal{F}}_{z_\rho h}$-measurable, and

$$b_{\rho+1}\in\left[0,\left(\sum_{i=1}^{n}(r_i+1)\right)^{-1}\right].$$

By the inequality $\operatorname{tr}(B^2)\geq m^{-1}(\operatorname{tr}(B))^2$ with $B\in\mathbb{R}^{m\times m}$ being the positive definite matrix, we have

$$\mathbb{E}\big[b_{\rho+1}\big|\bar{\mathcal{F}}_{z_\rho h}\big]\geq\frac{\sigma^2 h'\|\Sigma\|\tau'_\rho}{d\big(\sum_{i=1}^{n}(r_i+1)\big)(1+h'\|\Sigma\|)}.\tag{17}$$

Then by Assumption 2 and applying Lemma 1, an trivial verification yields that $\{b_{\rho+1}\}\in\mathcal{S}^0(\chi)$ for some $\chi\in[0,1)$. Consequently, by the definition of $\mathcal{S}^0(\cdot)$, we can obtain that

$$\mathbb{E}\left[\sum_{k=s}^{t}(1-b_{k+1})\right]\leq C\chi^{t-s+1}\quad\forall t\geq s\geq 0$$

for some constants $C>0$ and $\chi\in[0,1)$. According to Lemma 3, it is derived that for $\forall\varepsilon>0$

$$\exp\big(\varepsilon T_{\rho+1}\big)\leq\exp\big((1-b_{\rho+1})\varepsilon T_\rho\big)\cdot\exp\big(d'\varepsilon\big).$$

By the following inequality:

$$\exp(\beta\Upsilon)-1\leq\beta\exp(\Upsilon),\ \ 0<\beta<1,\ \ \Upsilon>0$$

we get

$$\exp\big(\varepsilon T_{\rho+1}\big)\leq\big[\big(1-b_{\rho+1}\big)\exp\big(\varepsilon T_\rho\big)+1\big]\cdot\exp\big(d'\varepsilon\big).$$

From this, let ε be sufficiently small to ensure that $\exp(d'\varepsilon)\chi<1$ holds. It yields that

$$\sup_{\rho\geq 0}\mathbb{E}\big[\exp\big(\varepsilon T_\rho\big)\big]<\infty$$

which implies that Lemma 4 1) holds.

Now, we prove Lemma 4 2): denote $y_\rho=\mu^{-1}[h(1+\|\bar{\Sigma}^{-1}\|\cdot\|\bar{\Sigma}\|)+\|\bar{\Sigma}^{-1}\|T_\rho]$ with T_ρ being defined in Lemma 3. It can be seen that

$$\begin{aligned}y_{\rho+1}\leq&\ \big(1-b_{\rho+1}\big)y_\rho\\&+\mu^{-1}\Big[h\Big(1+\|\bar{\Sigma}^{-1}\|\cdot\|\bar{\Sigma}\|\Big)+d'\|\bar{\Sigma}^{-1}\|\Big].\end{aligned}$$

It is easy to see from (17), Assumption 2 and Lemma 1 2) that [37, Lemma 3.1] is applicable to the above equation. Hence, we know that $\{1/y_\rho\}\in\mathcal{S}^0(\gamma)$, for some $\gamma\in(0,1)$. Note that

$$y_\rho=\sum_{k=z_{\rho-1}h+1}^{(z_{\rho-1}+1)h}\mu^{-1}\Big[1+\Big\|\bar{\Sigma}^{-1}\Big\|\cdot\|\bar{\Sigma}\|+\Big\|\bar{\Sigma}^{-1}\Big\|\operatorname{tr}(P_{k+1})\Big].$$

Similar to the proof in Lemma 5 of [41], it follows that:

$$\left\{\mu\Big[1+\Big\|\bar{\Sigma}^{-1}\Big\|\cdot\|\bar{\Sigma}\|+\Big\|\bar{\Sigma}^{-1}\Big\|\cdot\operatorname{tr}(P_k)\Big]^{-1}\right\}\in\mathcal{S}^0(\tau)$$

holds for some $\tau\in(0,1)$. Then we know that

$$\left\{\mu\Big[1+\Big\|\bar{\Sigma}^{-1}\Big\|\cdot\|\bar{\Sigma}\|+\Big\|\bar{\Sigma}^{-1}\Big\|\cdot\|P_k\|\Big]^{-1}\right\}\in\mathcal{S}^0(\tau).$$

Since $(\bar{P}_{k+1}-\bar{\Sigma})^{-1}=P_k^{-1}+R^{-1}\Xi_k\Xi_k^{\mathrm{T}}$, we have $\bar{P}_{k+1}\leq P_k+\bar{\Sigma}$, and $\|\bar{\Sigma}^{-1}\|\cdot\|\bar{P}_{k+1}\|\leq\|\bar{\Sigma}^{-1}\|\cdot\|P_k\|+\|\bar{\Sigma}^{-1}\|\cdot\|\bar{\Sigma}\|$. By this and Lemma 1 1), we can obtain that

$$\Big\{\mu/\Big(1+\|\bar{\Sigma}^{-1}\|\cdot\|\bar{P}_{k+1}\|\Big)\Big\}\in\mathcal{S}^0(\tau)$$

holds for some $\tau\in(0,1)$, which is the desired conclusion. ∎

Following Lemma 4, the assumptions on $\{P_k\}$ in Lemma 2 are satisfied. Furthermore, it is worth mentioning that under the Markovian switching topologies, the adjacency matrix is asymmetric and random, which gives rise to new challenges for the stability analysis. By virtue of stochastic stability theory, Markov chain theory, and CS theory, we establish the following stability results over Markovian switching topologies.

Theorem 1: Consider the system model (1) and (2) and the compressed error equation (12). Suppose that for some $q\geq(1/2)$ and for any $i\in\mathcal{V}$, $\sup_{k\geq 0}\|\varphi_{k,i}\|_{2q}<\infty$. Also suppose that Assumptions 1–3 hold.

1) $\{I_{dn}-P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}(I_{dn}-L_k\Xi_k^{\mathrm{T}})\}$ is L_p exponentially stable ($p\geq 1$).
2) Given that

$$\sigma_r\triangleq\sup_k\Big\|\Delta_k\log^\beta(e+\Delta_k)\Big\|_r<\infty$$

hold for some $r\geq 1$, $\beta>2$, where $\Delta_k=([3\delta_{4s}]/[\sqrt{1-\delta_{4s}}])\|X_k\|+\|V_k\|+\sqrt{1+\delta_{4s}}\|\Omega_{k+1}\|$, then for $p=(q^{-1}+r^{-1})^{-1}$, the compressed estimation error $\{\tilde{\Lambda}_k,k\geq 0\}$ is L_p-stable and

$$\limsup_{k\to\infty}\|\tilde{\Lambda}_k\|_p\leq C\tag{18}$$

where $C\triangleq c[\sigma_r\log^\beta(e+\sigma_r^{-1})]$ with c being a positive constant.

Proof: According to Lemmas 2 and 4, it follows immediately that for some $p\geq 1$ and $\tau\in[0,1)$, we have:

$$\Big\{I_{dn}-P_{k+1}\mathscr{A}_{m(k)}^{\mathrm{T}}\bar{P}_{k+1}^{-1}\big(I_{dn}-L_k\Xi_k^{\mathrm{T}}\big)\Big\}\in\mathcal{S}_p\Big(\tau^{1/2p}\Big).$$

Additionally, by the compressed error equation (12), it is yielded that

$$\tilde{\Lambda}_{k+1} = \prod_{i=0}^{k} P_{i+1}\mathscr{A}_{m(i)}^{\mathrm{T}}\bar{P}_{i+1}^{-1}\left(I_{dn} - L_i\Xi_i^{\mathrm{T}}\right)\tilde{\Lambda}_0 + \sum_{i=0}^{k}\left[\prod_{j=i+1}^{k} P_{j+1}\mathscr{A}_{m(j)}^{\mathrm{T}}\bar{P}_{j+1}^{-1}\left(I_{dn} - L_j\Xi_j^{\mathrm{T}}\right) \cdot P_{i+1}\mathscr{A}_{m(i)}^{\mathrm{T}}\bar{P}_{i+1}^{-1}\cdot\left(-L_i\bar{V}_i + \bar{\Omega}_{i+1}\right)\right].$$

Next, by Minkowski inequality, we obtain that

$$\left\|\tilde{\Lambda}_{k+1}\right\|_p \le \left\|\prod_{i=0}^{k} P_{i+1}\mathscr{A}_{m(i)}^{\mathrm{T}}\bar{P}_{i+1}^{-1}\left(I_{dn} - L_i\Xi_i^{\mathrm{T}}\right)\tilde{\Lambda}_0\right\|_p + \sum_{i=0}^{k}\left\|\prod_{j=i+1}^{k} P_{j+1}\mathscr{A}_{m(j)}^{\mathrm{T}}\bar{P}_{j+1}^{-1}\left(I_{dn} - L_j\Xi_j^{\mathrm{T}}\right) \cdot P_{i+1}\mathscr{A}_{m(i)}^{\mathrm{T}}\bar{P}_{i+1}^{-1}\cdot\left(-L_i\bar{V}_i + \bar{\Omega}_{i+1}\right)\right\|_p \triangleq T_1 + \sum_{i=0}^{k} T_{2,i}.$$

In the following, we divide our analysis for $T_{2,i}$ into three steps:

$$T_{2,i} \le \left\|\left(\left\|\prod_{j=i+1}^{k} P_{j+1}\mathscr{A}_{m(j)}^{\mathrm{T}}\bar{P}_{j+1}^{-1}\left(I_{dn} - L_j\Xi_j^{\mathrm{T}}\right)\right\| \cdot \left\|P_{i+1}\mathscr{A}_{m(i)}^{\mathrm{T}}\bar{P}_{i+1}^{-1}\right\| \cdot \left\|-L_i\bar{V}_i + \bar{\Omega}_{i+1}\right\|\right)\right\|_p \triangleq \|(I_1 \cdot I_2 \cdot I_3)\|_p. \tag{19}$$

By (14), we obtain that $\Sigma_k \ge \bar{\Sigma}$ and $\bar{P}_k \ge \bar{\Sigma}$. Then, it follows that $\|P_k^{-1}\| \le \|\bar{P}_k^{-1}\| \le \|\bar{\Sigma}^{-1}\|$. According to Lemma 2, we obtain that for $k > i$

$$I_1 \le \left\{\prod_{j=i+1}^{k}\left(1 - \frac{1}{1 + \left\|\bar{\Sigma}^{-1}\right\| \cdot \|\bar{P}_{j+1}\|}\right)^{\frac{1}{2}}\right\} \cdot \left\{\|P_{k+1}\|^{\frac{1}{2}} \cdot \left\|\bar{P}_{i+1}^{-1}\right\|^{\frac{1}{2}}\right\} \le \left\{\prod_{j=i+1}^{k}\left(1 - \frac{1}{2\left(1 + \left\|\bar{\Sigma}^{-1}\right\| \cdot \|\bar{P}_{j+1}\|\right)}\right)\right\} \cdot \left\{\|P_{k+1}\|^{\frac{1}{2}} \cdot \left\|\bar{\Sigma}^{-1}\right\|^{\frac{1}{2}}\right\}. \tag{20}$$

The next thing to do is estimate I_2. Clearly, $\mathscr{A}_{m(i)}$ is a doubly stochastic matrix and $\|\mathscr{A}_{m(i)}\| = 1$, yielding that

$$I_2 \le \|P_{i+1}\| \cdot \|\mathscr{A}_{m(i)}\| \cdot \left\|\bar{P}_{i+1}^{-1}\right\| \le \|P_{i+1}\| \cdot \left\|\bar{\Sigma}^{-1}\right\|. \tag{21}$$

It remains to estimate I_3 in (19). For simplicity, denote $U_k \triangleq \mathrm{col}\{u_{k,1}, \ldots, u_{k,n}\} \in \mathbb{R}^n$ where $u_{k,i} = h_{k,i}^{\mathrm{T}}[I_m - D^{\mathrm{T}}D]x_k$. Since $\|h_{k,i}\|_{\ell_0} \le 3s$ and $\|x_k\|_{\ell_0} \le s$, we define index sets of nonzero elements as L_{3s} and L_s, respectively. Without loss of generality, we use $h_{k,i,4s}$ and $x_{k,4s}$ to denote vectors which are formed by elements of $h_{k,i}$ and x_k indexed by $L_{4s} = L_{3s} \cup L_s$. Also, the notation D_{4s} represents the submatrix which is constructed by columns of D indexed by L_{4s}. By Assumption 1, any eigenvalue of $D_{4s}^{\mathrm{T}}D_{4s}$ is in $[1 - \delta_{4s}, 1 + \delta_{4s}]$. Hence, it can be derived that

$$\begin{aligned}
\|u_{k,i}\| &= \left\|h_{k,i}^{\mathrm{T}}\left[I_m - D^{\mathrm{T}}D\right]x_k\right\| \\
&= \left\|h_{k,i,4s}^{\mathrm{T}}\left[I_{4s} - D_{4s}^{\mathrm{T}}D_{4s}\right]x_{k,4s}\right\| \\
&\le \left\|h_{k,i,4s}^{\mathrm{T}}\left[(1+\delta_{4s})I_{4s} - D_{4s}^{\mathrm{T}}D_{4s}\right]x_{k,4s}\right\| + \delta_{4s}\|h_{k,i,4s}x_{k,4s}\| \\
&\le \|h_{k,i,4s}\| \cdot \left\|\left[(1+\delta_{4s})I_{4s} - D_{4s}^{\mathrm{T}}D_{4s}\right]\right\| \cdot \|x_{k,4s}\| \\
&\quad + \delta_{4s}\|h_{k,i,4s}\| \cdot \|x_{k,4s}\| \\
&\le 2\delta_{4s}\|h_{k,i,4s}\| \cdot \|x_{k,4s}\| + \delta_{4s}\|h_{k,i,4s}\| \cdot \|x_{k,4s}\| \\
&= 3\delta_{4s}\|h_{k,i}\| \cdot \|x_k\| \\
&\le \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\varphi_{k,i}\| \cdot \|x_k\|
\end{aligned}$$

and $\|U_k\| \le ([3\delta_{4s}]/[\sqrt{1-\delta_{4s}}])\|\Xi_k\| \cdot \|X_k\|$. Notice that $\|L_i\| \le ([\|P_i\|^{(1/2)}]/[2\sqrt{r_{\min}}])$, where $r_{\min} = \min_{i\in\mathcal{V}}\{r_1, \ldots, r_n\}$ and thus, we have

$$\begin{aligned}
I_3 &\le \|L_i\| \cdot (\|U_i\| + \|V_i\|) + \|\bar{\Omega}_{i+1}\| \\
&\le \frac{\|P_i\|^{\frac{1}{2}}}{2\sqrt{r_{\min}}}\left(\frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\Xi_i\| \cdot \|X_i\| + \|V_i\|\right) \\
&\quad + \sqrt{1+\delta_{4s}}\|\Omega_{i+1}\| \\
&\le \left(1 + \frac{\|P_i\|^{\frac{1}{2}}\max\{1, \|\Xi_i\|\}}{2\sqrt{r_{\min}}}\right)\xi_i.
\end{aligned} \tag{22}$$

Denote $p = (q^{-1} + r^{-1})^{-1}$. Then combining (20)–(22) and applying Hölder inequality, we get that

$$\begin{aligned}
T_{2,i} &\le \left\|\bar{\Sigma}^{-1}\right\|^{\frac{3}{2}}\sum_{i=0}^{k}\left\|\prod_{j=i+1}^{k}\left(1 - \frac{1}{2\left(1 + \|\bar{\Sigma}^{-1}\| \cdot \|\bar{P}_{j+1}\|\right)}\right) \cdot \|P_{k+1}\|^{\frac{1}{2}} \cdot \|P_{i+1}\|\left(1 + \frac{\|P_i\|^{\frac{1}{2}}\max\{1, \|\Xi_i\|\}}{2\sqrt{r_{\min}}}\right)\xi_i\right\|_p \\
&\le \left\|\bar{\Sigma}^{-1}\right\|^{\frac{3}{2}}\sup_{i\ge 0}\|P_i\|_{2q} \cdot \max\left\{1, \sup_{i\ge 0}\|\Xi_i\|_{2q}\right\} \\
&\quad \cdot \sum_{i=0}^{k}\left\|\prod_{j=i+1}^{k}\left(1 - \frac{1}{2\left(1 + \|\bar{\Sigma}^{-1}\| \cdot \|\bar{P}_{j+1}\|\right)}\right) \cdot \|P_{k+1}\|^{\frac{1}{2}}\left(1 + \frac{\|P_i\|^{\frac{1}{2}}}{2\sqrt{r_{\min}}}\right)\xi_i\right\|_r.
\end{aligned}$$

By Schwarz inequality and Lemma 4, we have

$$\begin{aligned}
&\sup_{k\ge i}\mathbb{E}\left[\exp\left(\varepsilon\|P_{k+1}\|^{\frac{1}{2}} \cdot \|P_i\|^{\frac{1}{2}}\right)\right] \\
&\le \sup_{k\ge i}\left\{\mathbb{E}\left[\exp(\varepsilon\|P_{k+1}\|)\right]\right\}^{\frac{1}{2}} \cdot \left\{\mathbb{E}\left[\exp(\varepsilon\|P_i\|)\right]\right\}^{\frac{1}{2}} < \infty.
\end{aligned}$$

The rest of the proof is analogous to [37, Th. 4.1] and therefore, further details are not provided here. By Lemma 4, we draw the desired conclusion. ■

Remark 5: From (18) and the definition of Δ_k, it is seen that the upper bound of compressed estimation error is composed of three components: the first component is related to the compression error, which has a positive correlation with δ_{4s}; the second component is associated with the magnitude of system noise v_k; and the last one depends on the magnitude of the state variation ω_k. Regarding the first part, it can be small since the RIP constant can be small. For instance, let the sensing matrix D be a Gaussian or Bernoulli random matrix. Then the RIP constant δ_{4s} for D can be arbitrarily small when the inequality $d \geq 480s\log(m/4s)/\delta_{4s}^3$ holds with d and m being dimensions of D. As a result, assuming that the system noise v_k and the parameter variation ω_k are also minimal, the upper bound of the estimation error C will be small.

Apart from the compressed estimation error, we derive the estimation error bound for the state of interest based on the CS theory.

Theorem 2: Suppose that $D \in \mathbb{R}^{d\times m}$ satisfies 4sth RIP with s satisfying $\delta_{3s} + 3\delta_{4s} < 2$. Under the same conditions as used in Theorem 1, the upper bound for the estimation error of the original state vector satisfies

$$\lim_{k\to\infty}\sup \left\|x_k - \hat{x}_{k,i}\right\|_p \leq C_s C$$

with C_s being defined in Lemma 6 in Appendix A.

Proof: By Theorem 1, the upper bound of compressed estimation error is given

$$\lim_{k\to\infty}\sup \|\tilde{\zeta}_{k+1,i}\|_p = \lim_{k\to\infty}\sup \|\zeta_{k+1} - \hat{\zeta}_{k+1,i}\|_p = \lim_{k\to\infty}\sup \|Dx_{k+1} - \hat{\zeta}_{k+1,i}\|_p \leq C.$$

Let $\bar{C} = C$ in (8) of Algorithm 1, we can obtain that the recovered state vector $\hat{x}_{k+1,i}$ obeys

$$\lim_{k\to\infty}\sup \|x_{k+1} - \hat{x}_{k+1,i}\|_p \leq C_s C.$$

This proves the theorem. ■

Ultimately, we present the probability that the upper bound for the estimation error of the original high-dimensional state vector stays in a certain range.

Corollary 1: Under the same conditions as used in Theorem 2, for any given constant $\varepsilon > 0$ and $\gamma \in (0, 1)$, there exists time instant T_ε

$$\mathbb{P}\left\{\left\|x_k - \hat{x}_{k,i}\right\| \leq \eta(C_s C + \varepsilon)^{1-\gamma}\right\} \geq 1 - \frac{(C_s C + \varepsilon)^\gamma}{\eta}$$

holds with $\eta = \max\{1, 2(C_s C + \varepsilon)^\gamma\}$.

Proof: Consider the special case of $p = 1$ in Theorem 2, and we have $\lim\sup_{k\to\infty} \mathbb{E}[\|x_k - \hat{x}_{k,i}\|] \leq C_s C$. In other words, $\forall\varepsilon > 0$, $\exists\, T_\varepsilon \in \mathbb{R}$, s.t.$\forall t \geq T_\varepsilon$, $\mathbb{E}[\|x_k - \hat{x}_{k,i}\|] \leq C_s C + \varepsilon$. Then by Markov inequality, we have for any $t \geq T_\varepsilon$

$$\mathbb{P}\Big(\|x_k - \hat{x}_{k,i}\| \geq \eta(C_s C + \varepsilon)^{1-\gamma}\Big) \leq \frac{\mathbb{E}\big[\|x_k - \hat{x}_{k,i}\|\big]}{\eta(C_s C + \varepsilon)^{1-\gamma}} \leq \frac{(C_s C + \varepsilon)^\gamma}{\eta} \leq \frac{1}{2}.$$

This completes the proof. ■

Remark 6: Note that C_s may depend only on the RIP constant δ_{4s}. It immediately follows that CC_s can be sufficiently small for a sufficiently small δ_{4s} and in consequence, η becomes close to 1. That is, the estimation error will likely remain in the small range of zero with a high probability.

According to Theorems 1 and 2, it is clear that for the s-sparse unknown state vectors x_k, the upper bound of the estimation error is positively related to the s-restricted isometry constant δ_{4s} that increases with s. Furthermore, we can see that the stability results of the CDKF over the Markovian switching topologies are established without requiring the independence or stationarity conditions of the signals. Consequently, our results are likely to be applicable to the feedback systems.

IV. Simulation

To demonstrate the efficacy of the CDKF for estimating the sparse state vector, a simulation example is given as follows: the sensor network composed of four sensors is represented by three switched digraphs $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ in Fig. 1. The corresponding weighted adjacency matrices are given by

$$\mathcal{A}_1 = \begin{bmatrix} 0.8 & 0.1 & 0.1 & 0 \\ 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.2 & 0 & 0.6 & 0.2 \end{bmatrix},\ \mathcal{A}_2 = \begin{bmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.3 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0.5 \end{bmatrix}$$

$$\mathcal{A}_3 = \begin{bmatrix} 0.8 & 0 & 0 & 0.2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.2 & 0 & 0.6 & 0.2 \end{bmatrix}.$$

In addition, the switching signal $m(k)$ is governed by a time-homogeneous Markov chain with probability transmission matrix being

$$P = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.5 & 0.1 & 0.4 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}.$$

Obviously, Assumption 3 on communication topologies is satisfied. Four sensors cooperate to estimate a 2-sparse 80-D state vector x_k and the positions of the two nonzero elements are given randomly. Here, we focus on the case where two elements of ω_k in (2) follow $1/k^2 \cdot \mathcal{N}(0, 0.1^2)$.

It is assumed that the noise sequence $\{v_{k,i} \in \mathbb{R}\}$ is i.i.d. with Gaussian distribution $\mathcal{N}(0, 0.2^2)$. The stochastic regression vectors $\{h_{k,i} \in \mathbb{R}^{80}\}$ are generated according to the following expression:

$$h_{k,i} = \Bigg[0, \ldots, 0, \underbrace{1.1^k + \sum_{p=0}^{k-1} 1.1^p o_{k-p,i}}_{i^{th}}, 0, \ldots, 0\Bigg]^{\mathrm{T}}$$

where $o_{p,i} \sim \mathcal{N}(0, 0.1^2)$. The sensing matrix D is given by the Gaussian matrix $D \sim \mathcal{N}(0, 1/3, 3, 80)$. Obviously, compressed regression vectors $\varphi_{k,i} = Dh_{k,i} \in \mathbb{R}^3$ satisfy Assumption 2, while uncompressed high-dimensional ones $h_{k,i}$ fail to meet the traditional excitation condition in [10].

To demonstrate the estimation performance of Algorithm 1, we repeat CDKF, DKF in [10], compressed consensus normalized least mean squares algorithm (CC-NLMSs) in [33], and CCRA in [30] for 200 times with the same initial values. As for the compressed algorithms (i.e., CDKF, CC-NLMS, and CCRA), we use the Gaussian matrix $D \sim \mathcal{N}(0, 1/3, 3, 80)$

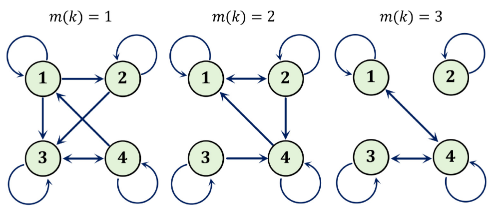


Fig. 1. Topological structure of the sensor network.

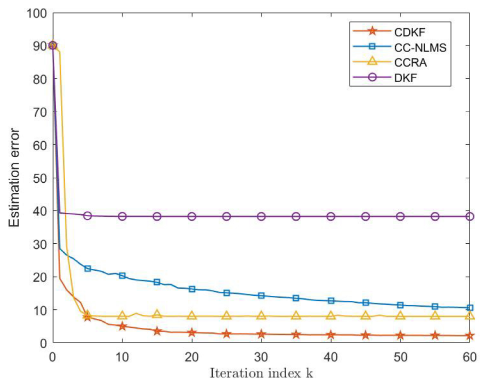


Fig. 2. Estimation errors for different algorithms.

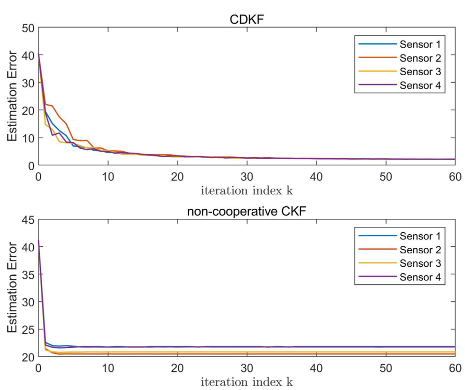


Fig. 3. Estimation errors of sensors for cooperative and noncooperative algorithms.

as the sensing matrix and resort to the orthogonal matching pursuit algorithm [42] to perform the reconstruction step in Algorithm 1. Fig. 2 demonstrates that the estimation error of CDKF is apparently smaller than that of CC-NLMS while the estimation error of the DKF stays large. Also, we compare the CDKF with the noncooperative compressed Kalman filter (noncooperative CKF) (i.e., $\mathcal{A} = I_4$) in Fig. 3 to demonstrate the cooperative effect of sensors. We can see that the estimation error for the CDKF falls into a small neighborhood of zero, while for the noncooperative case, the estimation error of every sensor is large.

V. Concluding Remarks

This article investigates the distributed state estimation problems of the sparse state vector over Markovian switching topologies. According to the compression-estimation-reconstruction scheme, we present the CDKF to estimate the sparse state of interest cooperatively. It is shown that the proposed compressed algorithm may effectively accomplish the estimation task, even when the uncompressed one falls short in accurately estimating unknown sparse state vectors because of inadequate excitation. As for stability analysis, the estimation error bound is established under a compressed cooperative excitation condition. Here, we require no independent or stationary assumptions. Thus, our stability analysis is applicable to feedback systems, which facilitates further research on problems pertaining to the combination of distributed state estimation and control. In addition, further research is necessary to explore how to introduce an error feedback scheme to lower the compression error.

Appendix A
Compressed Sensing Theory

CS is a signal processing technique that allows efficient sensing and reconstruction of a approximately sparse signal. It considers the recovery of a signal $x \in \mathbb{R}^m$ from a noisy observation

$$z = Dx + \varepsilon \tag{23}$$

where $D \in \mathbb{R}^{d\times m}(d \ll m)$ is the sensing matrix and $\varepsilon \in \mathbb{R}^d$ is the noise. It is undoubtedly challenging to deal with the underdetermined linear system (23). However, the sparsity of a signal can be beneficial to the recovery problem. Accordingly, the true signal x is supposed to be s-sparse, that is, $\|x\|_{\ell_0} \le s$ for some $s \le d \ll m$.

The reconstruction performance will greatly depend on 1) the generation of the sensing matrix D and 2) the design of the signal reconstruction algorithm.

1) *Generation of the Sensing Matrix:* To ensure the accurate recovery of the sparse signal x, Candès et al. [43] introduced the following concept as a requirement for the sensing matrix D.

Definition 3 (RIP): Let $D \in \mathbb{R}^{d\times m}$ be the sensing matrix and s ($1 \le s \le m$) be an integer. We say that the matrix D satisfies the RIP of order s if there exists a constant $\delta_s \in [0, 1)$, which is the smallest quantity such that

$$(1-\delta_s)\|a\|^2 \le \|D_L a\|^2 \le (1+\delta_s)\|a\|^2 \quad \forall a \in \mathbb{R}^{\#L} \tag{24}$$

holds for every submatrix D_L which is formed by columns of D corresponding to the index set L with $\#L \le s$. The notation $\#L$ represents the cardinality of the set L.

Remark 7: Loosely speaking, the concept of RIP is a characterization of the near orthogonality of the matrix, at least for sparse vectors. An equivalent formulation of (24) is $1-\delta_s \le \lambda_{\min}(D_L^{\mathrm{T}} D_L) \le \lambda_{\max}(D_L^{\mathrm{T}} D_L) \le 1+\delta_s$. Considerable progress has been made in generating the sensing matrix satisfying the RIP [44]. For example, for the Gaussian or Bernoulli random matrix satisfying the RIP, the following theoretical result was established in [45]:

Lemma 5 [45]: For given d, m, and $0 < \delta < 1$, suppose that the sensing matrix $D \in \mathbb{R}^{d\times m}$ is a Gaussian or Bernoulli random matrix, then there exist positive constants c_1, c_2 which only relates to δ such that for a prespecified δ and any $s \leq c_1 d/\log(m/s)$, the probability that RIP holds is no less than $1 - 2\exp(-c_2 d)$.

Remark 8: As mentioned in [45], when c_1 is sufficiently small, c_2 enables to be larger than 0. Let $c_1 = \delta^3/120$ and $d \geq 120 s \log(m/s)/\delta^3$. Then, the probability that the sensing matrix D satisfies RIP is not less than $1 - 2\exp(-c_2 d)$.

2) *Design of the Signal Reconstruction Algorithm:* In view of the sparsity of the signal, the recovery problem of the observation model (23) can be recast as follows:

$$\min_{x\in\mathbb{R}^m} \|x\|_{\ell_1}, \ \text{s.t.}\ \|z - Dx\| \leq \bar{C} \tag{25}$$

where $\bar{C}$ measures the magnitude of noise with $\|\varepsilon\| \leq \bar{C}$. The following lemma from [43] measures the deviation between the recovered and true signals.

Lemma 6: Let s satisfy $\delta_{3s} + 3\delta_{4s} < 2$ where δ_{3s} and δ_{4s} are defined in Definition 3. Then the recovered signal x^* derived from solving the convex optimization problem (25) obeys

$$\left\|x - x^*\right\| \leq C_s \bar{C}$$

where the constant C_s may only depend on δ_{4s}.

Remark 9: As illustrated in the proof procedure in [43], the constant C_s can be taken as $4/(\sqrt{3(1-\delta_{4s})} - \sqrt{1+\delta_{3s}})$, which is positively related to δ_{4s}. In Lemma 6, it is demonstrated that the recovery of the signal is stable. That is, small noise ε can only result in small derivations between the recovered and true signals. Furthermore, the signal can be exactly recovered when the noise is zero.

Appendix B
Proof of Lemma 3

Combining (5) and (6) and the well-known matrix inversion formula (see, e.g., [46, Theorem 1.1.17]), we have for any $k \in [z_\rho h + 1, (z_\rho + 1)h]$

$$\begin{aligned} P_{k,i} &= \left\{\sum_{j=1}^{n} a_{ji,m(k-1)} \bar{P}_{k,j}^{-1}\right\}^{-1} \leq \sum_{j=1}^{n} a_{ji,m(k-1)} \bar{P}_{k,j} \\ &= \sum_{j=1}^{n} a_{ji,m(k-1)} \big(\bar{P}_{k,j} - \Sigma\big) + \Sigma \\ &= \sum_{j=1}^{n} a_{ji,m(k-1)} \left(P_{k-1,j}^{-1} + r_j^{-1}\varphi_{k-1,j}\varphi_{k-1,j}^{\mathrm{T}}\right)^{-1} + \Sigma \\ &\leq \sum_{j=1}^{n} a_{ji,m(k-1)} P_{k-1,j} + \Sigma. \end{aligned} \tag{26}$$

Then, recursively, we have

$$\begin{aligned} P_{k,i} &\leq \sum_{j=1}^{n} a_{ji,m(k-1)} \left(\sum_{t=1}^{n} a_{tj,m(k-2)} P_{k-2,t}\right) + 2\Sigma \\ &= \sum_{j=1}^{n} A_{k-2}^{k-1}(j,i)\ P_{k-2,j} + 2\Sigma \\ &\leq \cdots \leq \sum_{j=1}^{n} A_{\rho h'}^{k-1}(j,i)\ P_{\rho h',j} + \big(k - \rho h'\big)\Sigma \\ &\leq \sum_{j=1}^{n} A_{\rho h'}^{k-1}(j,i)\ P_{\rho h',j} + h'\Sigma \triangleq \Psi_i\big(\rho h', k-1\big). \end{aligned}$$

Hence, by the above inequality and the matrix inversion formula, we obtain

$$\begin{aligned} &\sum_{j=1}^{n} a_{ji,m(k)} \left(P_{k,j}^{-1} + r_j^{-1}\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}\right)^{-1} \\ &\leq \sum_{j=1}^{n} a_{ji,m(k)} \left(\Psi_j^{-1}\big(\rho h', k-1\big) + r_j^{-1}\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}\right)^{-1} \\ &= \sum_{j=1}^{n} a_{ji,m(k)} \Psi_j\big(\rho h', k-1\big) \\ &\quad - \sum_{j=1}^{n} a_{ji,m(k)} \frac{\Psi_j\big(\rho h', k-1\big)\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}\Psi_j\big(\rho h', k-1\big)}{r_j + \varphi_{k,j}^{\mathrm{T}}\Psi_j(\rho h', k-1)\varphi_{k,j}} \\ &= \Psi_i\big(\rho h', k\big) \\ &\quad - \sum_{j=1}^{n} a_{ji,m(k)} \frac{\Psi_j\big(\rho h', k-1\big)\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}\Psi_j\big(\rho h', k-1\big)}{r_j + \varphi_{k,j}^{\mathrm{T}}\Psi_j(\rho h', k-1)\varphi_{k,j}} \\ &\leq \Psi_i\big(\rho h', k\big) \\ &\quad - \sum_{j=1}^{n} a_{ji,m(k)} \frac{\Psi_j\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\Psi_j\big(\rho h', k-1\big)}{\big(r_j+1\big)\big(1+\lambda_{\max}\big(\Psi_j(\rho h', k-1)\big)\big)}. \end{aligned}$$

Hence, by the above inequality and (26), we have

$$\begin{aligned} P_{k+1,i} &\leq \Sigma + \Psi_i\big(\rho h', k\big) \\ &\quad - \sum_{j=1}^{n} a_{ji,m(k)} \frac{\Psi_j\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\Psi_j\big(\rho h', k-1\big)}{\big(r_j+1\big)\big(1+\lambda_{\max}\big(\Psi_j(\rho h', k-1)\big)\big)}. \end{aligned} \tag{27}$$

For constants $a_j, b_j \geq 0$, by C_r- and Schwartz inequalities, it is obvious that $\sum_{j=1}^{m} a_j b_j \leq \sum_{j=1}^{m} a_j \sum_{j=1}^{m} b_j$. Moreover, when setting $a_j = (c_j/d_j)$, $b_j = d_j$ with c_j and d_j being positive constants, we have $\sum_{j=1}^{m}(c_j/d_j) \geq ([\sum_{j=1}^{m} c_j]/[\sum_{j=1}^{m} d_j])$. Recall the definitions of $\Psi_i(\rho h', k)$ and $\Psi'(\rho h')$. Hence, $\Psi_i(\rho h', k) \leq \Psi'(\rho h')$ holds for any $k \geq 0$ and $i \in \mathcal{V}$. Also we have

$$\begin{aligned} &\operatorname{tr}\left(\sum_{i=1}^{n} \Psi_i\big(\rho h', k\big)\right) \\ &= \operatorname{tr}\left(\sum_{i=1}^{n}\left(\sum_{j=1}^{n} A_{\rho h'}^{k}(j,i) P_{\rho h',j} + h'\Sigma\right)\right) \\ &= \operatorname{tr}\left(nh'\Sigma + \sum_{j=1}^{n} P_{\rho h',j}\right) \\ &= nh'\operatorname{tr}(\Sigma) + \operatorname{tr}\big(P_{\rho h'}\big). \end{aligned} \tag{28}$$

For $k \in [z_\rho h + 1, (z_\rho + 1)h]$, we have by (27) and (28)

$$
\begin{aligned}
\operatorname{tr}(P_{k+1}) &= \operatorname{tr}\left(\sum_{i=1}^{n} P_{k+1,i}\right)\\
&\leq n\operatorname{tr}(\Sigma) + \operatorname{tr}\left(\sum_{i=1}^{n} \Psi_i\big(\rho h', k\big)\right)\\
&\quad - \operatorname{tr}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji,m(k)} \frac{\Psi_j^2\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}}{\big(r_j+1\big)\big(1+\lambda_{\max}\big(\Psi_j(\rho h', k-1)\big)\big)}\right)\\
&= n\big(h'+1\big)\operatorname{tr}(\Sigma) + \operatorname{tr}\big(P_{\rho h'}\big)\\
&\quad - \sum_{j=1}^{n} \frac{\operatorname{tr}\left(\Psi_j^2\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\right)}{\big(r_j+1\big)\big(1+\lambda_{\max}\big(\Psi_j(\rho h', k-1)\big)\big)}\\
&\leq n\big(h'+1\big)\operatorname{tr}(\Sigma) + \operatorname{tr}\big(P_{\rho h'}\big)\\
&\quad - \frac{\sum_{j=1}^{n}\operatorname{tr}\left(\Psi_j^2\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\right)}{\sum_{j=1}^{n}\big(r_j+1\big)\big(1+\lambda_{\max}\big(\Psi_j(\rho h', k-1)\big)\big)}\\
&\leq n\big(h'+1\big)\operatorname{tr}(\Sigma) + \operatorname{tr}\big(P_{\rho h'}\big) - \frac{\operatorname{tr}\big(P_{\rho h'}\big)}{\operatorname{tr}(\Psi'(\rho h'))}\\
&\quad \cdot \frac{\sum_{j=1}^{n}\operatorname{tr}\left(\Psi_j^2\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\right)}{\sum_{j=1}^{n}\big(r_j+1\big)\cdot\sum_{j=1}^{n}\big(1+\lambda_{\max}\big(\Psi_j(\rho h', k-1)\big)\big)}\\
&\leq n\big(h'+1\big)\operatorname{tr}(\Sigma) + \operatorname{tr}\big(P_{\rho h'}\big) - \frac{\operatorname{tr}\big(P_{\rho h'}\big)}{\operatorname{tr}(\Psi'(\rho h'))}\\
&\quad \cdot \frac{\sum_{j=1}^{n}\operatorname{tr}\left(\Psi_j^2\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\right)}{n\sum_{j=1}^{n}\big(r_j+1\big)(1+\lambda_{\max}(\Psi'(\rho h')))}.
\end{aligned}
$$

Adding both sides of the above inequality yields

$$
\begin{aligned}
T_{\rho+1} &= \sum_{k=z_\rho h+1}^{(z_\rho+1)h} \operatorname{tr}(P_{k+1})\\
&\leq nh\big(h'+1\big)\operatorname{tr}(\Sigma) + h\operatorname{tr}\big(P_{\rho h'}\big) - h\operatorname{tr}\big(P_{\rho h'}\big)\\
&\quad \cdot \frac{\operatorname{tr}\left(\sum_{k=z_\rho h+1}^{(z_\rho+1)h}\sum_{j=1}^{n}\Psi_j^2\big(\rho h', k-1\big)\frac{\varphi_{k,j}\varphi_{k,j}^{\mathrm{T}}}{1+\|\varphi_{k,j}\|^2}\right)}{nh\sum_{j=1}^{n}\big(r_j+1\big)(1+\lambda_{\max}(\Psi'(\rho h')))\cdot\operatorname{tr}(\Psi'(\rho h'))}\\
&= \big(1-b_{\rho+1}\big)h\operatorname{tr}\big(P_{\rho h'}\big) + nh\big(h'+1\big)\operatorname{tr}(\Sigma).
\end{aligned}
$$

Again, for the first term of the above equation, since $P_{\rho h',j} \leq \sum_{t=1}^{n} A_{k+1}^{\rho h'-1}(t,j)P_{k+1,t} + (\rho h' - k - 1)\Sigma$, we obtain

$$
\begin{aligned}
h\operatorname{tr}\big(P_{\rho h'}\big) &= \sum_{k=z_{\rho-1}h+1}^{(z_{\rho-1}+1)h}\sum_{j=1}^{n}\operatorname{tr}\big(P_{\rho h',j}\big)\\
&\leq T_\rho + \frac{1}{2}nh\big(h'+1\big)\operatorname{tr}(\Sigma)
\end{aligned}
$$

and

$$
\begin{aligned}
T_{\rho+1} &\leq \big(1-b_{\rho+1}\big)T_\rho + \frac{3}{2}nh\big(h'+1\big)\operatorname{tr}(\Sigma)\\
&= \big(1-b_{\rho+1}\big)T_\rho + d', \quad \rho \geq 0.
\end{aligned}
$$

References

[1] B. Lian, Y. Wan, Y. Zhang, M. Liu, F. L. Lewis, and T. Chai, "Distributed Kalman consensus filter for estimation with moving targets," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 5242–5254, Jun. 2022.

[2] Y. Xu, Y. Deng, Z. Huang, M. Lin, and P. Shi, "Distributed state estimation over sensor networks with substate decomposition approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 1, pp. 527–537, Jan./Feb. 2023.

[3] E. Sebastián, E. Montijano, and C. Sagüés, "ECO-DKF: Event-triggered and certifiable optimal distributed Kalman filter under unknown correlations," *IEEE Trans. Autom. Control*, vol. 69, no. 4, pp. 2613–2620, Apr. 2024.

[4] B. Lian, F. L. Lewis, G. A. Hewer, K. Estabridis, and T. Chai, "Robustness analysis of distributed Kalman filter for estimation in sensor networks," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 12479–12490, Nov. 2022.

[5] K. Ma, S. Wu, Y. Wei, and W. Zhang, "Gossip-based distributed tracking in networks of heterogeneous agents," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 801–804, Apr. 2017.

[6] H. Yang, H. Li, Y. Xia, and L. Li, "Distributed Kalman filtering over sensor networks with transmission delays," *IEEE Trans. Cybern.*, vol. 51, no. 11, pp. 5511–5521, Nov. 2021.

[7] J. Zhou, G. Gu, and X. Chen, "Distributed Kalman filtering over wireless sensor networks in the presence of data packet drops," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1603–1610, Apr. 2019.

[8] D. Li, S. Kar, F. E. Alsaadi, A. M. Dobaie, and S. Cui, "Distributed Kalman filtering with quantized sensing state," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5180–5193, Oct. 2015.

[9] 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.

[10] Q. Liu, Z. Wang, X. He, and D. H. Zhou, "On Kalman-consensus filtering with random link failures over sensor networks," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2701–2708, Aug. 2018.

[11] X. He, C. Hu, Y. Hong, L. Shi, and H.-T. 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.

[12] W. Chen, Z. Wang, D. Ding, X. Yi, and Q.-L. Han, "Distributed state estimation over wireless sensor networks with energy harvesting sensors," *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 3311–3324, May 2023.

[13] R. Olfati-Saber, "Kalman-consensus filter: Optimality, stability, and performance," in *Proc. 48th IEEE Conf. Decision Control (CDC), 28th Chin. Control Conf. (CCC)*, 2009, pp. 7036–7042.

[14] C. Li, H. Dong, J. Li, and F. Wang, "Distributed Kalman filtering for sensor network with balanced topology," *Syst. Control Lett.*, vol. 131, Sep. 2019, Art. no. 104500.

[15] D. Gan, S. Xie, and Z. Liu, "Stability of the distributed Kalman filter using general random coefficients," *Sci. China Inf. Sci.*, vol. 64, no. 7, 2021, Art. no. 172204.

[16] D. Ding, Z. Wang, and Q.-L. Han, "A scalable algorithm for event-triggered state estimation with unknown parameters and switching topologies over sensor networks," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 4087–4097, Sep. 2020.

[17] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Trans. Autom. Control*, vol. 50, no. 11, pp. 1867–1872, Nov. 2005.

[18] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[19] I. Matei, N. Martins, and J. S. Baras, "Almost sure convergence to consensus in Markovian random graphs," in *Proc. 47th IEEE Conf. Decision Control(CDC)*, 2008, pp. 3535–3540.

[20] F. Zhu and J. H. Park, "An adaptive smooth event-triggered approach to recursive estimation for nonlinear time-varying systems with uncertain communication topologies," *IEEE Trans. Control Netw. Syst.*, vol. 11, no. 2, pp. 902–914, Jun. 2024.

[21] Q. Zhang and J.-F. Zhang, "Distributed parameter estimation over unreliable networks with Markovian switching topologies," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2545–2560, Oct. 2012.

[22] Q. Liu, Z. Wang, X. He, and D. Zhou, "Event-based distributed filtering over Markovian switching topologies," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1595–1602, Apr. 2019.

[23] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[24] S. Vinga, "Structured sparsity regularization for analyzing high-dimensional omics data," *Brief. Bioinf.*, vol. 22, no. 1, pp. 77–87, 2021.

[25] M. O. Khalifa, A. H. Abdelhafiz, and A. Zerguine, "Sparse channel estimation using adaptive filtering and compressed sampling," in *Proc. Int. Conf. Comput., Elect. Electron. Eng. (ICCEEE)*, 2013, pp. 144–147.

[26] Z. Liu, Y. Liu, and C. Li, "Distributed sparse recursive least-squares over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1386–1395, Mar. 2014.

[27] D. Gan and Z. Liu, "Distributed sparse identification for stochastic dynamic systems under cooperative nonpersistent excitation condition," *Automatica*, vol. 151, May 2023, Art. no. 110958.

[28] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[29] S. Xu, R. C. de Lamare, and H. V. Poor, "Distributed compressed estimation based on compressive sensing," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1311–1315, Sep. 2015.

[30] L. Li and D. Li, "A distributed estimation method over network based on compressed sensing," *Int. J. Distrib. Sens. Netw.*, vol. 15, no. 4, pp. 1–10, 2019.

[31] P. Di Lorenzo and A. H. Sayed, "Sparse distributed learning based on diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1419–1433, Mar. 2013.

[32] D. Gan and Z. Liu, "Performance analysis of the compressed distributed least squares algorithm," *Syst. Control Lett.*, vol. 164, Jun. 2022, Art. no. 105228.

[33] S. Xie and L. Guo, "Analysis of compressed distributed adaptive filters," *Automatica*, vol. 112, Feb. 2020, Art. no. 108707.

[34] R. Li, D. Gan, S. Xie, and J. Lü, "Stability and performance analysis of the compressed Kalman filter algorithm for sparse stochastic systems," *Sci. China Technol. Sci.*, vol. 67, no. 2, pp. 380–394, 2024.

[35] 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.

[36] S. Zhang, L. Peng, and X. Chang, "Optimal energy allocation based on SINR under DoS attack," *Neurocomputing*, vol. 570, Feb. 2024, Art. no. 127116.

[37] L. Guo, "Stability of recursive stochastic tracking algorithms," *SIAM J. Control Optim.*, vol. 32, no. 5, pp. 1195–1225, 1994.

[38] S. Xie, D. Gan, and Z. Liu, "Stability analysis of distributed Kalman filtering algorithm for stochastic regression model," 2024, *arXiv:2411.01198*.

[39] D. Gan, S. Xie, Z. Liu, and J. Lü, "Stability of FFLS-based diffusion adaptive filter under cooperative excitation condition," *IEEE Trans. Autom. Control*, vol. 69, no. 11, pp. 7479–7492, Nov. 2024.

[40] M. Rahman and M. Ahsanullah, "A note on the expected values of powers of a matrix," *Can. J. Statist., Revue Canadienne de Statistique*, vol. 1, no. 1, pp. 123–125, 1973.

[41] L. Guo, "Estimating time-varying parameters by the Kalman filter-based algorithm: Stability and convergence," *IEEE Trans. Autom. Control*, vol. 35, no. 2, pp. 141–147, Feb. 1990.

[42] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[43] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

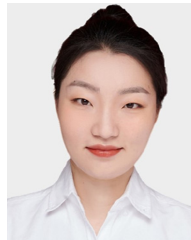
[44] Y. V. Parkale and S. L. Nalbalwar, "Sensing matrices in compressed sensing," in *Proc. Comput. Eng. Technol. (ICCET)*, 2020, pp. 113–123.

[45] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, pp. 253–263, Jan. 2008.

[46] L. Guo, *Time-Varying Stochastic Systems, Stability and Adaptive Theory*, 2nd ed. Beijing, China: Science Press, 2020.

Rongjiang Li received the B.S. degree in statistics from Wuhan University, Wuhan, China, in 2019, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2024.

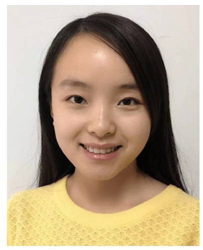
She is currently a Postdoctoral Research Associate with AMSS, CAS. Her current research focuses on system identification, distributed adaptive filter, and game theory.



Die Gan (Member, IEEE) received the B.S. degree in mathematics from Shandong University, Jinan, China, in 2017, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2022.

She was a Postdoctoral Fellow from 2022 to 2024 with Zhongguancun Laboratory, Beijing. She is currently an Associate Professor with the College of Artificial Intelligence, Nankai University, Tianjin, China. Her research interests include system identification, machine learning, and distributed adaptive filters.

Dr. Gan received the China National Scholarship for graduate students in 2020, the Outstanding Graduate Title of Beijing in 2022, and the First Youth Excellent Paper Award of the Beijing Mathematical Society in 2023.



Siyu Xie received the B.S. degree in information and computing science (systems control) from Beihang University, Beijing, China, in 2013, and the Ph.D. degree in control theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, in 2018.

She was a Postdoctoral Fellow from 2019 to 2022 with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA. She is currently a Professor with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, China. Her research interests include networked systems, distributed optimization problems for power systems, distributed adaptive filters, identification, and distributed control.

Prof. Xie received the IEEE CSS Beijing Chapter Young Author Prize in 2015, the China National Scholarship for graduate students in 2017, the President Scholarship of Chinese Academy of Sciences in 2018, and the Excellent Doctoral dissertation of Chinese Academy of Sciences in 2019.

Haibo Gu (Member, IEEE) received the M.Sc. degree in mathematics from Inner Mongolia University, Hohhot, China, in 2016, and the Ph.D. degree in system theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2019.

From 2019 to 2021, he was a Postdoctoral Fellow of Zhuoyue Program with Beihang University, Beijing, where he is currently an Associate Professor with the School of Automation Science and Electrical Engineering. His current research interests include complex networks, multiagent systems, network dynamics, and cooperative control.

Jinhu Lü (Fellow, IEEE) received the Ph.D. degree in applied mathematics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2002.

He was a Professor with RMIT University, Melbourne, VIC, Australia, and a Visiting Fellow with Princeton University, Princeton, NJ, USA. He is currently a Vice President of Scientific Research with Beihang University, Beijing. He is/was the Chief Scientist of the National Key Research and Development Program of China and the Leading Scientist of the Innovative Research Groups, National Natural Science Foundation of China, Beijing. His current research interests include cooperation1 control, industrial Internet, complex networks, and big data.

Prof. Lü was a recipient of the prestigious Ho Leung Ho Lee Foundation Award in 2015, the National Innovation Competition Award in 2020, the State Natural Science Award three times from the Chinese Government in 2008, 2012, and 2016, respectively, the Australian Research Council Future Fellowships Award in 2009, the National Natural Science Fund for Distinguished Young Scholars Award, and the Highly Cited Researcher Award in engineering from 2014 to 2020, and from 2022 to 2023. He was the General Co-Chair of the 43rd Annual Conference of the IEEE Industrial Electronics Society in 2017. He is/was an Editor in various ranks for 16 SCI journals, including the Co-Editor-in-Chief of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. He is the Fellow of IEEE/CAA/ORSC/CICC/CSIAM.