Control Theory and Technology (2025) 23:161–175
https://doi.org/10.1007/s11768-025-00253-x

RESEARCH ARTICLE

Stability analysis of distributed Kalman filtering algorithm for stochastic regression model

Siyu Xie[1] · Die Gan[2] · Zhixin Liu[3,4]

Received: 18 June 2024 / Revised: 25 November 2024 / Accepted: 20 December 2024 / Published online: 22 April 2025

Abstract

The work proposes a distributed Kalman filtering (KF) algorithm to track a time-varying unknown signal process for a stochastic regression model over network systems in a cooperative way. We provide the stability analysis of the proposed distributed KF algorithm without independent and stationary signal assumptions, which implies that the theoretical results are able to be applied to stochastic feedback systems. Note that the main difficulty of stability analysis lies in analyzing the properties of the product of non-independent and non-stationary random matrices involved in the error equation. We employ analysis techniques such as stochastic Lyapunov function, stability theory of stochastic systems, and algebraic graph theory to deal with the above issue. The stochastic spatio-temporal cooperative information condition shows the cooperative property of multiple sensors that even though any local sensor cannot track the time-varying unknown signal, the distributed KF algorithm can be utilized to finish the filtering task in a cooperative way. At last, we illustrate the property of the proposed distributed KF algorithm by a simulation example.

Keywords Distributed Kalman filtering algorithm · Stochastic cooperative information condition · Sensor networks · L_p-exponential stability · Stochastic regression model

1 Introduction

Nowadays, more and more data can be collected through sensor networks, and how to use the collected data to estimate or track an unknown parameter process has obtained plenty of research attention [1–8]. Basically, there exist two different ways to process the data, i.e., the centralized and distributed methods. For the centralized processing method, measurements or estimates from all sensors over the network need to be transferred to a fusion center, which may have no feasibility due to energy consumption, limited communication capabilities, privacy considerations or packet losses. Moreover, this method has poor robustness because when the fusion center is damaged the whole network will collapse. Because of these limitations, the distributed processing method arises, where each sensor will estimate the unknown parameter vectors by the local noisy observations and the data from the neighboring sensors. Compared with the centralized method, the distributed method is more robust and scalable [4].

Note that different types of distributed estimation algorithms can be obtained by combining different cooperative

This work was supported in part by Sichuan Science and Technology Program under Grant No. 2025ZNSFSC151, in part by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDA27030201, and the Natural Science Foundation of China under Grant No. U21B6001, in part by the Natural Science Foundation of Tianjin under Grant No. 24JCQNJC01930.

✉ Die Gan
gandie@amss.ac.cn

Siyu Xie
syxie@uestc.edu.cn

Zhixin Liu
lzx@amss.ac.cn

1 School of Aeronautics and Astronautic, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China

2 College of Artificial Intelligence, Nankai University, Tianjin 300350, China

3 State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Beijing 100190, China

4 School of Mathematical Sciences, University of Academy of Sciences, Beijing 100049, China

strategies with different algorithms, such as incremental least mean squares (LMS) [4], consensus LMS [9, 10], diffusion LMS [5, 6, 11], consensus and diffusion stochastic gradient (SG) [12], consensus least squares (LS) [7, 8], incremental LS [13], diffusion LS [14–18], distributed KF [19–32], and so on. References [9–11, 18, 33, 34] established the theoretical results on stability and performance for the distributed LMS and LS algorithms without requiring the stationarity and independency conditions for the regression vectors. Since the KF algorithm would be optimal when the noise and the parameter variation are white and Gaussian noises, and the linearized observation matrices in the extended KF, which are widely used to estimate the state of a nonlinear engineering system, are often stochastic, here we focus on the KF algorithm for a stochastic regression model with stochastic observation vector in this work. Another reason why we consider this question is that the stability results in the existing work are far from satisfactory, as it is difficult to apply to non-stationary and non-independent signals generated from practical stochastic feedback systems.

In fact, plenty of research was focused on distributed KF algorithms where observation matrices of the system are deterministic. For example, [21] studied a distributed KF based on consensus strategies, and [22] presented a hybrid distributed information fusion algorithm and established convergence results under very mild topology conditions. Moreover, [23] introduced a distributed KF based on the covariance intersection method, and analyzed stability properties. In addition, [24] developed a consensus and innovations type Kalman filter, and [25] proposed a gossip-based distributed KF for deterministic time-varying observation matrices, and provided the error reduction rate. Furthermore, [26] and [27] considered consensus Kalman filters where the communication channels have random failures. A distributed KF algorithm with data packet losses for linear time-invariant discrete-time systems was studied in [28], and a distributed KF for deterministic time-varying observation matrices with mild assumption on local observability and network topology was studied in [29]. Moreover, [30] provided a boundedness analysis of the error covariance matrix for deterministic time-varying observation matrices, [31] developed a distributed filtering algorithm to estimate a sparse signal sequence for the dynamic model with deterministic observation matrix, and [32] presented distributed filtering algorithms based on tunable weights under attack for the deterministic time-invariant observation matrix.

To our knowledge, a first attempt to consider distributed KF algorithms with general random coefficients for the dynamic system was made in [35], where local innovation pairs are diffused to collectively track the unknown parameters. However, the proposed distributed KF algorithm is required to exchange a lot of information since it needs to diffuse L times for each time iteration, where L is larger than or equal to the diameter of the sensor network topology which increases as the network grows. Several communication cycles are required for information fusion at each time step. Notice that in making several communication cycles in one time instant may achieve the optimal (in the MMSE sense), or nearly optimal, state estimates. However, this greatly increases the communication complexity of the algorithm.

We will investigate a well-known distributed time-varying stochastic linear regression model in this work, and provide a theoretical analysis for the stability of our proposed distributed KF algorithm which is developed by the covariance intersection fusion rule (cf., [23, 26, 30]) and based on diffusion strategy. Note also that it only needs to diffuse one time for each time iteration, which greatly reduces the communication complexity compared with [35]. The main contributions of this work are summarized as the following three-fold:

- The stability analysis of the distributed KF algorithm can be provided without imposing the commonly used assumptions such as stationarity and independence on the stochastic regression vectors (cf., [14–17]), which shows that the theoretical results are expected to be applied to distributed adaptive control problems.
- Under a stochastic cooperative information assumption, we present the stability analysis of the distributed KF algorithm, which is a temporal and spatial union information condition on the random regression vectors, and implies that the networked system can finish the tracking task collaboratively even though no local node can due to a lack of necessary excitations.
- The proposed distributed KF algorithm in this paper focuses on tracking unknown time-varying signals where the observations is taken from the widely used time-varying linear regression model, while most of the existing literature focuses on the study of distributed KF algorithms where the observation matrices are deterministic, see e.g., [21–32].

For the remainder of the paper, the problem formulation is given in Sect. 2. The error equations, mathematical definitions, and assumptions are provided in Sect. 3. Sections 4 and 5 provide the main theoretical results and proofs, respectively. A case study is given in Sect. 6 followed by the conclusions in Sect. 7.

2 Problem formulation

2.1 Graph theory

Here we assume that the sensor network contains n vertexes and we model the network topology as a directed graph $\mathcal{G}$. We denote $\mathcal{V} = \{1, \ldots, n\}$ as the set of vertexes, and denote $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ as the set of directed arrows. Also, an arrow (i, j) is directed from the tail i to the head j. For a vertex in the network, the number of head (tail) ends adjacent to a vertex is its indegree (outdegree). A path from j_1 to j_t is a sequence of sensors $j_1, j_2, \ldots, j_t$ $(t \geq 2)$, such that $(j_i, j_{i+1}) \in \mathcal{E}$ for $i = 1, \ldots, t-1$. If there exists a path between any two vertexes in the digraph, we know that the digraph $\mathcal{G}$ is strongly connected. The diameter $D_{\mathcal{G}}$ of the graph $\mathcal{G}$ is defined as the maximum shortest path length between any two vertexes.

Here we use $\mathcal{A} = \{a_{ij}\}_{n\times n}$ to describe the structure of $\mathcal{G}$, which is called the weighted adjacency matrix, and we know that $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise. If $\sum_{j=1}^{n} a_{ji} = \sum_{j=1}^{n} a_{ij} = 1, \forall i = 1, \ldots, n$, the graph $\mathcal{G}$ is balanced. Note also that the weighted adjacency matrix $\mathcal{A}$ of a directed graph may be asymmetric. The vertex i is used to denote the ith sensor, and the arrow (i, j) is used to denote the communication from the vertex i to the vertex j. We use $\mathcal{N}_i = \{\ell \in V|(\ell, i) \in \mathcal{E}\}$, to denote the neighboring set of the vertex i, and any neighboring nodes have the ability to send data over a directed arrow between them.

2.2 Observation model

The main task of the paper is to develop a distributed method to estimate or track an unknown time-varying parameter vector $\boldsymbol{\zeta}_k$ by cooperating with each other in networked systems. Assume that each sensor i at time instant k collects the measurement $\{o_{k,i}, \boldsymbol{\varphi}_{k,i}\}$ that follows the widely used time-varying and stochastic regression model:

$$o_{k,i} = \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \boldsymbol{\zeta}_k + n_{k,i}, \tag{1}$$

where $k \geq 0$ is the time instant, $i \in \{1, \ldots, n\}$ is the ith sensor, $(\cdot)^{\mathrm{T}}$ is the transpose operator, $o_{k,i} \in \mathbb{R}$ is the scalar measurement, $\boldsymbol{\varphi}_{k,i} \in \mathbb{R}^m$ is the stochastic regressor signal, $n_{k,i} \in \mathbb{R}$ is the measurement noise, and $\boldsymbol{\zeta}_k \in \mathbb{R}^m$ is the time-varying parameter to be estimated by all sensors in the network. The variation of $\boldsymbol{\zeta}_k$ is denoted by $\boldsymbol{\delta}_k$, i.e.,

$$\boldsymbol{\delta}_k = \boldsymbol{\zeta}_k - \boldsymbol{\zeta}_{k-1}, \quad k \geq 1, \tag{2}$$

where $\boldsymbol{\delta}_k \in \mathbb{R}^m$ is an undefined vector. To simplify the notations, here we focus on the case where $o_{k,i}$ is a scalar and $\boldsymbol{\varphi}_{k,i}$ is a column vector. While if $o_{k,i}$ is a vector and $\boldsymbol{\varphi}_{k,i}$ is a matrix, we can do a similar analysis and get the same results. By choosing $\boldsymbol{\varphi}_{k,i}$ and $o_{k,i}$ appropriately, we can see that many technical problem formulations fit the structure (1), for examples, collaborative spectral sensing, signal processing, target localization, and so on, [4].

Note that while most literature (cf., [21–32]) focused on deterministic observation vectors or matrices, here we consider the case where the observation vector $\boldsymbol{\varphi}_{k,i}$ in (1) is stochastic. Although the Eq. (2) is a simplified system model compared to the general linear time-varying system models considered in e.g., [21–32], this is the first step for us to consider distributed KF algorithms for the dynamical system with random coefficients, i.e., $\boldsymbol{\varphi}_{k,i}$, which has practical importance for stochastic feedback systems. For instance, if $\boldsymbol{\varphi}_{k,i} = [o_{k,i}, \cdots, o_{k-p,i}, x_{k,i}, \cdots, x_{k-q,i}]^{\mathrm{T}}$ which consists of current and past input–output data of the systems, and $x_{k,i}$ is the input signal at time k, then the model (1) can be reduced to the well-known autoregressive model with exogenous inputs (ARX) with time-varying coefficients. Also, it is easy to see that $\boldsymbol{\varphi}_{k,i}$ is stochastic and cannot satisfy the independent and identically distributed condition. Also, the general linear time-varying system model will be considered in future work.

Tracking or estimating a time-varying signal is a critical problem in control engineering, system identification, signal processing, and so on. Different recursive algorithms were derived in existing work (see e.g., [36, 37]), which usually have the following form:

$$\widehat{\boldsymbol{\zeta}}_{k+1,i} = \widehat{\boldsymbol{\zeta}}_{k,i} + \boldsymbol{L}_{k,i}(o_{k,i} - \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \widehat{\boldsymbol{\zeta}}_{k,i}),$$

where $\widehat{\boldsymbol{\zeta}}_{k,i} \in \mathbb{R}^m$ is the estimate of node i at time instant k, $\boldsymbol{L}_{k,i} \in \mathbb{R}^m$ is called the adaptation gain that requires to be designed. Note that $\boldsymbol{L}_{k,i}$ cannot tend to zero as k tends to infinity since the unknown parameter to be estimated is time-varying.

The most common ways of selecting $\boldsymbol{L}_{k,i}$ can obtain the LMS algorithm, i.e., $\boldsymbol{L}_{k,i} = \mu \boldsymbol{\varphi}_{k,i}$, where $\mu > 0$ is the step size; and the recursive LS algorithm with forgetting factor, i.e.,

$$\boldsymbol{L}_{k,i} = \frac{\Omega_{k,i} \boldsymbol{\varphi}_{k,i}}{\alpha + \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \Omega_{k,i} \boldsymbol{\varphi}_{k,i}},$$

$$\Omega_{k+1,i} = \frac{1}{\alpha}\left[\Omega_{k,i} - \frac{\Omega_{k,i} \boldsymbol{\varphi}_{k,i} \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \Omega_{k,i}}{\alpha + \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \Omega_{k,i} \boldsymbol{\varphi}_{k,i}}\right],$$

where an initial matrix $\Omega_{0,i} \in \mathbb{R}^{m\times m}$ is positive definite and $\alpha \in (0, 1)$ is a forgetting factor; and the KF algorithm, i.e.,

$$\begin{aligned} \boldsymbol{L}_{k,i} &= \frac{\Omega_{k,i} \boldsymbol{\varphi}_{k,i}}{r_i + \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \Omega_{k,i} \boldsymbol{\varphi}_{k,i}}, \\ \Omega_{k+1,i} &= \Omega_{k,i} - \frac{\Omega_{k,i} \boldsymbol{\varphi}_{k,i} \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \Omega_{k,i}}{r_i + \boldsymbol{\varphi}_{k,i}^{\mathrm{T}} \Omega_{k,i} \boldsymbol{\varphi}_{k,i}} + Q, \end{aligned} \tag{3}$$

where $r_i \in \mathbb{R}$ and $Q \in \mathbb{R}^{m\times m}$ may be viewed as *a priori* estimates for the variances of $n_{k,i}$ and δ_k, and $r_i > 0$, $Q > 0$ hold. For simplicity of discussion, here we take r_i and Q as constants. The KF algorithm would be optimal when the parameter variation and the noise are white Gaussian noises. Therefore, we focus on the KF algorithm in this work.

As far as we know, the best result that guarantees the stability of the KF algorithm for each sensor i in the network and allows $\{\varphi_{k,i}\}$ to be a large class of stochastic signals was given in [36], which assumed that $\{\varphi_{k,i}, \mathcal{F}_{k,i}\}$ is an adapted process[1] and satisfies the following individual excitation condition, namely, for each i, there exists an integer $\kappa > 0$ such that $\{\alpha_{k,i}, k \geq 0\} \in \mathcal{S}^0(\alpha)$ for some $\alpha \in (0, 1)$, and

$$\alpha_{k,i} \triangleq \lambda_{\min}\left\{\mathbb{E}\left[\frac{1}{\kappa+1}\sum_{j=k\kappa+1}^{(k+1)\kappa}\frac{\varphi_{j,i}\varphi_{j,i}^{\mathrm{T}}}{1+\|\varphi_{j,i}\|^2}\Big|\mathcal{F}_{k\kappa,i}\right]\right\}, \tag{4}$$

and $\mathcal{S}^0(\alpha)$ is defined in Definition 3. Note that $\lambda_{\max}\{\cdot\}$ and $\lambda_{\min}\{\cdot\}$ are defined as the largest and the smallest eigenvalues of the matrix, and $\mathbb{E}[\cdot|\mathcal{F}_{k\kappa,i}]$ denotes the conditional mathematical expectation operator. For the case that stochastic regression vectors $\{\varphi_{k,i}\}$ are high-dimensional or sparse, it is difficult or even impossible to make the condition (4) satisfied. We can improve this situation by designing a distributed KF algorithm in which the measurements of nodes are exchanged in a sensor network.

2.3 Distributed KF algorithm

In the following part, we present the distributed KF algorithm (see Algorithm 1).

Algorithm 1 can be derived from some existing literature for distributed Kalman filters (cf., [23, 26, 30]) by assuming that the observation and states obey (1) and (2), respectively. In fact, Algorithm 1 is designed by using the structure of the standard KF algorithm (3) and the covariance intersection fusion rule in [38]. In Step 1, each sensor i first uses the Kalman filter update equations (3) to obtain the prediction estimate $\bar{\zeta}_{k+1,i}$ and prediction covariance matrix $\bar{\Omega}_{k+1,i}$. Since the estimates and covariance matrices from different sensors may contain complementary information, combining these two kinds of information together may help to achieve a more accurate estimation of the unknown parameter. We denote $x_{k+1,i} \triangleq \Omega_{k+1,i}^{-1}\widehat{\zeta}_{k+1,i}$ and $\bar{x}_{k+1,i} \triangleq \bar{\Omega}_{k+1,i}^{-1}\bar{\zeta}_{k+1,i}$. Thus in Step 2, each sensor i combines the inverse covariance matrix $\bar{\Omega}_{k+1,\ell}^{-1}$ and information vector $\bar{x}_{k+1,\ell}$ from its neighboring sensors in a convex manner to obtain the matrix $\Omega_{k+1,i}^{-1}$ and vector $x_{k+1,i}$, that is $\Omega_{k+1,i}^{-1} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{\Omega}_{k+1,\ell}^{-1}$ and $x_{k+1,i} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{x}_{k+1,i}$, which correspond to (7) and (8).

[1] $\mathcal{F}_{k,i}$ is any family of non-decreasing σ-algebras.

Algorithm 1 Distributed KF algorithm

Initialization: For each sensor $i \in \{1, \cdots, n\}$, we select an arbitrary initial column vector $\widehat{\zeta}_{0,i} \in \mathbb{R}^m$ and an arbitrary initial positive definite matrix $\Omega_{0,i} \in \mathbb{R}^{m\times m}$.

Output: The estimates $\{\widehat{\zeta}_{k+1,i}\}_{i=1}^n$, for each time instant $k = 0, 1, 2, \cdots$ and for each sensor $i = 1, \cdots, n$

Step 1 Adapt:

$$\bar{\zeta}_{k+1,i} = \widehat{\zeta}_{k,i} + \frac{\Omega_{k,i}\varphi_{k,i}}{r_i + \varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}\varphi_{k,i}}(o_{k,i} - \varphi_{k,i}^{\mathrm{T}}\widehat{\zeta}_{k,i}), \tag{5}$$

$$\bar{\Omega}_{k+1,i} = \Omega_{k,i} - \frac{\Omega_{k,i}\varphi_{k,i}\varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}}{r_i + \varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}\varphi_{k,i}} + Q, \tag{6}$$

Step 2 Combine:

$$\Omega_{k+1,i}^{-1} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{\Omega}_{k+1,\ell}^{-1}, \tag{7}$$

$$\widehat{\zeta}_{k+1,i} = \Omega_{k+1,i}\sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{\Omega}_{k+1,\ell}^{-1}\bar{\zeta}_{k+1,\ell}, \tag{8}$$

where $r_i \in \mathbb{R}$, $Q \in \mathbb{R}^{m\times m}$ and $r_i > 0$, $Q > 0$.

Note that the main contribution of this paper is to provide a theoretical analysis of Algorithm 1 without independent and stationary signal assumptions on regression signals $\varphi_{k,i}$. Note that if $\mathcal{A} = I_n$, the proposed distributed KF algorithm will degenerate to the non-cooperative KF algorithm (3).

3 Some preliminaries

3.1 Error equation

Before analyzing the distributed KF algorithm, we first need to derive the estimation error equation. For the sensor i, define the following two estimation errors:

$$\widetilde{\zeta}_{k,i} = \zeta_k - \widehat{\zeta}_{k,i}, \quad \widetilde{\bar{\zeta}}_{k,i} = \zeta_k - \bar{\zeta}_{k,i}.$$

Then from (7) and (8), we have

$$\begin{aligned}\widetilde{\zeta}_{k+1,i} =& \zeta_{k+1} - \Omega_{k+1,i}\sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{\Omega}_{k+1,\ell}^{-1}\bar{\zeta}_{k+1,\ell}\\ =& \Omega_{k+1,i}\sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{\Omega}_{k+1,\ell}^{-1}\zeta_{k+1}\\ &- \Omega_{k+1,i}\sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{\Omega}_{k+1,\ell}^{-1}\bar{\zeta}_{k+1,\ell}\\ =& \Omega_{k+1,i}\sum_{\ell\in\mathcal{N}_i} a_{\ell i}\bar{\Omega}_{k+1,\ell}^{-1}\widetilde{\bar{\zeta}}_{k+1,\ell}.\end{aligned} \tag{9}$$

From (1), (2) and (5), we can also obtain that

$$\widetilde{\bar{\zeta}}_{k+1,i} = \zeta_{k+1} - \bar{\zeta}_{k+1,i}$$

$$
\begin{aligned}
&=\zeta_k + \delta_{k+1} - \widehat{\zeta}_{k,i} \\
&\quad - \frac{\Omega_{k,i}\varphi_{k,i}}{r_i + \varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}\varphi_{k,i}}(o_{k,i} - \varphi_{k,i}^{\mathrm{T}}\widehat{\zeta}_{k,i}) \\
&=\widetilde{\zeta}_{k,i} + \delta_{k+1} - \frac{\Omega_{k,i}\varphi_{k,i}}{r_i + \varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}\varphi_{k,i}}(\varphi_{k,i}^{\mathrm{T}}\zeta_k \\
&\quad - \varphi_{k,i}^{\mathrm{T}}\widehat{\zeta}_{k,i} + n_{k,i}) \\
&=\Big(I_m - \frac{\Omega_{k,i}\varphi_{k,i}\varphi_{k,i}^{\mathrm{T}}}{r_i + \varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}\varphi_{k,i}\varphi_{k,i}^{\mathrm{T}}}\Big)\widetilde{\zeta}_{k,i} \\
&\quad - \frac{\Omega_{k,i}\varphi_{k,i}}{r_i + \varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}\varphi_{k,i}} n_{k,i} + \delta_{k+1}. \qquad (10)
\end{aligned}
$$

Denote

$$L_{k,i} = \frac{\Omega_{k,i}\varphi_{k,i}}{r_i + \varphi_{k,i}^{\mathrm{T}}\Omega_{k,i}\varphi_{k,i}},$$

we have

$$\widetilde{\widetilde{\zeta}}_{k+1,i} = (I_m - L_{k,i}\varphi_{k,i}^{\mathrm{T}})\widetilde{\zeta}_{k,i} - L_{k,i}n_{k,i} + \delta_{k+1}. \qquad (11)$$

To proceed with our analysis, we present the following notations to write the above error equation into a vector form:

where the notation col$\{\cdots\}$ represents a column vector of the corresponding vectors, the notation diag$\{\cdots\}$ represents a block matrix formed in a diagonal manner of the corresponding matrices or vectors, $\mathcal{A}$ is the adjacency matrix, and $\otimes$ represents the Kronecker product.

By (1) and (2), we have

$$\boldsymbol{O}_k = \boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{Z}_k + \boldsymbol{N}_k, \qquad (12)$$

and

$$\boldsymbol{\Delta}_k = \boldsymbol{Z}_k - \boldsymbol{Z}_{k-1}, \quad k \geq 1. \qquad (13)$$

For Algorithm 1, we have

$$
\begin{cases}
\bar{\boldsymbol{Z}}_{k+1} = \widehat{\boldsymbol{Z}}_k + \boldsymbol{L}_k(\boldsymbol{O}_k - \boldsymbol{\Phi}_k^{\mathrm{T}}\widehat{\boldsymbol{Z}}_k), \\
\bar{\boldsymbol{\Omega}}_{k+1} = \boldsymbol{\Omega}_k - \boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{\Omega}_k + \boldsymbol{Q}_{\mathrm{diag}}, \\
\mathrm{vec}\{\boldsymbol{\Omega}_{k+1}^{-1}\} = \mathscr{A}^{\mathrm{T}}\mathrm{vec}\{\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\}, \\
\widehat{\boldsymbol{Z}}_{k+1} = \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\bar{\boldsymbol{Z}}_{k+1},
\end{cases} \qquad (14)
$$

where the notation vec$\{\cdot\}$ represents an operator that stacks the blocks of a block diagonal matrix on top of each other.

By $\widetilde{\boldsymbol{Z}}_k = \boldsymbol{Z}_k - \widehat{\boldsymbol{Z}}_k$ and $\widetilde{\widetilde{\boldsymbol{Z}}}_k = \boldsymbol{Z}_k - \bar{\boldsymbol{Z}}_k$, we know from (11) that

$$\widetilde{\widetilde{\boldsymbol{Z}}}_{k+1} = (I_{mn} - \boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}})\widetilde{\boldsymbol{Z}}_k - \boldsymbol{L}_k\boldsymbol{N}_k + \boldsymbol{\Delta}_{k+1},$$

and by (9), we have the following error equation of Algorithm 1,

$$
\begin{aligned}
\widetilde{\boldsymbol{Z}}_{k+1} &= \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\widetilde{\widetilde{\boldsymbol{Z}}}_{k+1} \\
&= \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn} - \boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}})\widetilde{\boldsymbol{Z}}_k \\
&\quad - \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\boldsymbol{L}_k\boldsymbol{N}_k \\
&\quad + \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\boldsymbol{\Delta}_{k+1}. \qquad (15)
\end{aligned}
$$

We will analyze the stability of the above distributed KF algorithm under non-independent and correlated signal assumptions in Sect. 4. Note that the tracking error from the distributed filtering error equation (15) hinges on the exponential stability of its homogeneous equation, i.e.,

$$\widetilde{\boldsymbol{Z}}_{k+1} = \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn} - \boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}})\widetilde{\boldsymbol{Z}}_k,$$

by Propositions 2.1 and 2.2 in [36], i.e., the stochastic internal-external stability results. Also, the exponential stability of the homogeneous equation depends essentially on the properties of product of random matrices $\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn} - \boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}})$.

3.2 Some definitions

Here we need some definitions on the stability of random matrices from [36]. The Euclidean norm of a matrix $Y \in \mathbb{R}^{m\times n}$ is denoted as $\|Y\| = (\lambda_{\max}\{YY^{\mathrm{T}}\})^{\frac{1}{2}}$, and the L_p-norm of a random matrix X is defined as $\|X\|_{L_p} = \{\mathbb{E}[\|X\|^p]\}^{\frac{1}{p}}$, where $\mathbb{E}[\cdot]$ denotes the mathematical expectation operator.

Definition 1 Suppose that $\{X_k, k \geq 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$ is a random matrix sequence. The sequence $\{X_k\}$ is called L_p-stable if $\sup_{k\geq 0}\mathbb{E}[\|X_k\|^p] < \infty$, holds for some $p > 0$.

Definition 2 For a square random matrix sequence $X = \{X_k, k \geq 0\}$, if $X \in S_p(\alpha)$ for $p \geq 0$, where

$$S_p(\alpha) = \Big\{X : \Big\|\prod_{t=s+1}^{k}(I - X_t)\Big\|_{L_p} \leq C\alpha^{k-s}, \quad \forall k \geq s+1, \forall s \geq 0, \quad \text{for constant } C > 0\Big\}, \qquad (16)$$

then $\{I - X_k, k \geq 0\}$ is called L_p-exponentially stable with the parameter $\alpha \in [0, 1)$.

Remark 1 For the sequence $\{\eta_k\}$ generated by $\eta_{k+1} = (I - X_k)\eta_k + \varepsilon_{k+1}$, $k \geq 0$, we know from [36] that $\{X_k, k \geq 0\} \in S_p(\alpha)$ is in some sense a necessary and sufficient condition for the stability of η_k. Note that the analysis of the product

Table 1 Vector form and Dimension

Vector form	Dimension	Vector form	Dimension
$\boldsymbol{O}_k \triangleq \mathrm{col}\{o_{k,1},\ldots,o_{k,n}\}$	$\mathbb{R}^n$	$\widetilde{\boldsymbol{Z}}_k \triangleq \mathrm{col}\{\widetilde{\zeta}_{k,1},\ldots,\widetilde{\zeta}_{k,n}\}$ where $\widetilde{\zeta}_{k,i}=\zeta_k-\widehat{\zeta}_{k,i}$	$\mathbb{R}^{mn}$
$\boldsymbol{Z}_k \triangleq \mathrm{col}\{\underbrace{\zeta_k,\ldots,\zeta_k}_{n}\}$	$\mathbb{R}^{mn}$	$\widetilde{\bar{\boldsymbol{Z}}}_k \triangleq \mathrm{col}\{\widetilde{\bar{\zeta}}_{k,1},\ldots,\widetilde{\bar{\zeta}}_{k,n}\}$ where $\widetilde{\bar{\zeta}}_{k,i}=\zeta_k-\bar{\zeta}_{k,i}$	$\mathbb{R}^{mn}$
$\boldsymbol{\Phi}_k \triangleq \mathrm{diag}\{\boldsymbol{\varphi}_{k,1},\ldots,\boldsymbol{\varphi}_{k,n}\}$	$\mathbb{R}^{mn\times n}$	$\boldsymbol{L}_k \triangleq \mathrm{diag}\{L_{k,1},\ldots,L_{k,n}\}$	$\mathbb{R}^{mn\times n}$
$\boldsymbol{N}_k \triangleq \mathrm{col}\{n_{k,1},\ldots,n_{k,n}\}$	$\mathbb{R}^n$	$\boldsymbol{\Omega}_k \triangleq \mathrm{diag}\{\Omega_{k,1},\ldots,\Omega_{k,n}\}$	$\mathbb{R}^{mn\times mn}$
$\boldsymbol{\Delta}_k \triangleq \mathrm{col}\{\underbrace{\boldsymbol{\delta}_k,\ldots,\boldsymbol{\delta}_k}_{n}\}$	$\mathbb{R}^{mn}$	$\boldsymbol{Q}_{\mathrm{diag}} \triangleq \mathrm{diag}\{\underbrace{Q,\ldots,Q}_{n}\}$	$\mathbb{R}^{mn\times mn}$
$\widehat{\boldsymbol{Z}}_k \triangleq \mathrm{col}\{\widehat{\zeta}_{k,1},\ldots,\widehat{\zeta}_{k,n}\}$	$\mathbb{R}^{mn}$	$\bar{\boldsymbol{\Omega}}_k \triangleq \mathrm{diag}\{\bar{\Omega}_{k,1},\ldots,\bar{\Omega}_{k,n}\}$	$\mathbb{R}^{mn\times mn}$
$\bar{\boldsymbol{Z}}_k \triangleq \mathrm{col}\{\bar{\zeta}_{k,1},\ldots,\bar{\zeta}_{k,n}\}$	$\mathbb{R}^{mn}$	$\mathscr{A} \triangleq \mathcal{A}\otimes I_m$	$\mathbb{R}^{mn\times mn}$

of random matrices is a mathematically difficult task. However, we may transfer the analysis of the product of random matrices to the analysis of a certain class of scalar sequences for linear random equations generated from adaptive filtering algorithms. Also, the corresponding scalar sequence can be studied based on some information assumption on the regressor signals.

Thus, we introduce a subclass of $S_1(\alpha)$, which will be used to introduce the stochastic cooperative information condition in the following part.

Definition 3 For a scalar random sequence $x=\{x_k, k\geq 0\}$ and $\alpha\in(0,1)$, we define

$$S^0(\alpha)=\Big\{x: x_k\in[0,1], \mathbb{E}\big[\prod_{t=s+1}^{k}(1-x_t)\big]\leq C\alpha^{k-s},$$
$$\forall k\geq s+1, \forall s\geq 0, \quad \text{for constant } C>0\Big\}.$$

3.3 Assumptions

We need the following widely used network topology assumption for the stability analysis.

Assumption 1 (Network topology condition) The digraph $\mathcal{G}$ is balanced and strongly connected.

Remark 2 By Assumption 1, we know that when s is larger than or equal to the graph's diameter, i.e., $s\geq D_{\mathcal{G}}$, then each element of the matrix $\mathcal{A}^s=\underbrace{\mathcal{A}\cdots\mathcal{A}}_{s}$ will be positive.

In the following part, we will denote that $\mathcal{F}_k=\sigma\{\boldsymbol{\varphi}_{j,i},\boldsymbol{\delta}_j, n_{j-1,i}, j\leq k, i=1,\ldots,n\}$.

Assumption 2 (Stochastic cooperative information condition) There exists an integer $\kappa>0$ such that $\{\alpha_k, k\geq 0\}\in S^0(\alpha)$ for some $\alpha\in(0,1)$, where

$$\alpha_k\triangleq\lambda_{\min}\Bigg\{\mathbb{E}\Bigg[\frac{1}{n(\kappa+1)}\sum_{i=1}^{n}\sum_{j=k\kappa+1}^{(k+1)\kappa}\frac{\boldsymbol{\varphi}_{j,i}\boldsymbol{\varphi}_{j,i}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{j,i}\|^2}\Big|\mathcal{F}_{k\kappa}\Bigg]\Bigg\}. \tag{17}$$

Remark 3 Almost all the existing literature on the stability and performance analyses on distributed adaptive filters requires some stringent assumptions on the regressors, such as independence and stationarity (see e.g., [14–17]), which cannot be satisfied for signals generated by stochastic systems with feedback loops. In fact, Assumption 2 contains not only temporal union information but also spatial union information of all the sensors, which is more general than the independent or stationary signal conditions. It can also be considered as an extension of the excitation condition (4) from a single sensor to the network. This conditional mathematical expectation-based information condition for an individual sensor was first given in [37] and then refined in [36], which is often used for exponential stability (see [36]) of the adaptive filtering algorithms.

Note that Assumption 2 shows that to track an unknown and time-varying parameters, the regressor vectors $\{\boldsymbol{\varphi}_{k,i}\}$ may have some sort of "persistent excitations" in the sense that the prediction of the "future" is non-degenerate given the "past". Moreover, under Assumption 2, the distributed KF algorithm can be shown to have the capability to accomplish the tracking task cooperatively even if any sensor cannot track the unknown signal individually.

4 Main results

4.1 One basic result

We first present a basic result on how to convert the investigation of $S_p(\cdot)$ to that of a sequence in $S^0(\cdot)$ before giving the stability analysis of the proposed distributed KF algorithm. Here we introduce the Lyapunov equation as follows,

$$\bar{\boldsymbol{\Omega}}_{k+1}=(I_{mn}-\boldsymbol{A}_k)\boldsymbol{\Omega}_k(I_{mn}-\boldsymbol{A}_k)^{\mathrm{T}}+\boldsymbol{Q}_k, \tag{18}$$

and

$$\mathrm{vec}\{\boldsymbol{\Omega}_k^{-1}\}=\mathscr{A}^{\mathrm{T}}\mathrm{vec}\{\bar{\boldsymbol{\Omega}}_k^{-1}\}, \quad k\geq 0, \tag{19}$$

where $\boldsymbol{A}_k, \boldsymbol{Q}_k, \boldsymbol{\Omega}_k, \bar{\boldsymbol{\Omega}}_k \in \mathbb{R}^{mn\times mn}$, and $\boldsymbol{\Omega}_0 > 0$. The following theorem transfers the research of $\{\boldsymbol{A}_k\}$ to the research of a scalar random sequence in $S^0(\alpha)$, and also gives conditions for the L_p-exponential stability of (15).

Theorem 1 *Assume that $\{\boldsymbol{A}_k\}$ is a sequence of random matrices, and $\{\boldsymbol{Q}_k\}$ is a sequence of positive definite random matrices. Let $\{\boldsymbol{\Omega}_k\}$ and $\{\bar{\boldsymbol{\Omega}}_k\}$ recursively be generated by (18) and (19), we have for all $t > s$,*

$$\begin{aligned}&\left\|\prod_{k=s}^{t-1}\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn}-\boldsymbol{A}_k)\right\|^2\\&\leq\left\{\prod_{k=s}^{t-1}\left(1-\frac{1}{1+\|\boldsymbol{Q}_k^{-1}\bar{\boldsymbol{\Omega}}_{k+1}\|}\right)\right\}\left\{\|\boldsymbol{\Omega}_t\|\cdot\|\boldsymbol{\Omega}_s^{-1}\|\right\}.\end{aligned}\qquad(20)$$

Thus, if $\{\boldsymbol{\Omega}_k\}$ satisfies the following two assumptions:

1. $\left\{\frac{1}{1+\|\boldsymbol{Q}_k^{-1}\bar{\boldsymbol{\Omega}}_{k+1}\|}\right\}\in S^0(\alpha)$, *for some* $\alpha\in[0,1)$;
2. $\sup_{t\geq s\geq 0}\|(\|\boldsymbol{\Omega}_t\|\cdot\|\boldsymbol{\Omega}_s^{-1}\|)\|_{L_p}<\infty$, *for some* $p\geq 1$,

then

$$\{I_{mn}-\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn}-\boldsymbol{A}_k)\}\in S_p(\alpha^{1/2p}).\qquad(21)$$

Remark 4 Note that the proof of Theorem 1 is given in Sect. 5. This result does not require that $\boldsymbol{A}_k$ is non-negative and definite, and we may convert the verification of (21) to the following two tasks: 1) to prove that a given random scalar sequence belongs to $S^0(\alpha)$; and 2) to verify the L_p stability of a certain process. By Remark 1, we know that $\{I_{mn}-\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn}-\boldsymbol{A}_k)\}\in S_p(\alpha^{1/2p})$ is in some sense the necessary and sufficient condition for the stability of $\widetilde{\boldsymbol{Z}}_k$. Therefore, by Theorem 1, we now proceed to provide the stability result of the proposed distributed KF algorithm in Sect. 4.2.

4.2 Stability of the distributed KF algorithm

For convenience, we denote

$$\begin{aligned}&\boldsymbol{A}_k\triangleq\boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}},\quad\boldsymbol{R}\triangleq\mathrm{diag}\{r_1,\ldots,r_n\}\otimes I_m,\\&\boldsymbol{Q}_k\triangleq\boldsymbol{R}\boldsymbol{L}_k\boldsymbol{L}_k^{\mathrm{T}}+\boldsymbol{Q}_{\mathrm{diag}}.\end{aligned}\qquad(22)$$

By (14), we have

$$\begin{aligned}\bar{\boldsymbol{\Omega}}_{k+1}=&(I_{mn}-\boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}})\boldsymbol{\Omega}_k(I_{mn}-\boldsymbol{L}_k\boldsymbol{\Phi}_k^{\mathrm{T}})^{\mathrm{T}}\\&+\boldsymbol{R}\boldsymbol{L}_k\boldsymbol{L}_k^{\mathrm{T}}+\boldsymbol{Q}_{\mathrm{diag}}\\=&(I_{mn}-\boldsymbol{A}_k)\boldsymbol{\Omega}_k(I_{mn}-\boldsymbol{A}_k)^{\mathrm{T}}+\boldsymbol{Q}_k,\end{aligned}\qquad(23)$$

which has the same form of (18) and (19). Therefore, before applying Theorem 1, we first establish the boundedness result of $\{\boldsymbol{\Omega}_k\}$.

Lemma 2 *For $\{\boldsymbol{\Omega}_k\}$ generated by (14), if Assumptions 1 and 2 are satisfied, there exists a positive constant ε^* such that for any $\varepsilon\in[0,\varepsilon^*)$,*

$$\sup_{k\geq 0}\mathbb{E}[\exp(\varepsilon\|\boldsymbol{\Omega}_k\|)]<\infty.\qquad(24)$$

We put the proof of Lemma 2 in Sect. 5. The following result can be obtained by Lemma 2 directly, and we omit the proof here.

Corollary 1 *For $\{\boldsymbol{\Omega}_k\}$ generated by (14), if Assumptions 1 and 2 are satisfied, for any $p > 0$, then we have*

$$\sup_{k\geq 0}\mathbb{E}[\|\boldsymbol{\Omega}_k\|^p]<\infty.\qquad(25)$$

Before applying Theorem 1, we also need to verify that the given random scalar sequence is in $S^0(\alpha)$.

Lemma 3 *For $\{\boldsymbol{\Omega}_k\}$ generated by (14), if Assumptions 1 and 2 are satisfied, for any $\mu\in(0,1]$, then there exists a constant $\alpha\in(0,1)$ such that*

$$\left\{\frac{\mu}{1+\|\boldsymbol{Q}_{\mathrm{diag}}^{-1}\|\cdot\|\bar{\boldsymbol{\Omega}}_{k+1}\|}\right\}\in S^0(\alpha).\qquad(26)$$

The proof of Lemma 3 is presented in Sect. 5. By Lemmas 2 and 3, the assumptions 1) and 2) in Theorem 1 can be easily verified. Thus, we can establish the L_p-exponential stability of (15). Furthermore, we can get an upper bound of the tracking error for Algorithm 1. Here the following notation $\log(\cdot)$ is denoted as the logarithmic operator based on the natural number e.

Theorem 4 *Consider the stochastic observation model (1) and the distributed KF algorithm (14). Denote $\pi_k=\|\boldsymbol{N}_k\|+\|\boldsymbol{\Delta}_{k+1}\|$. Under Assumptions 1 and 2, if for some $\beta>2$ and $r\geq 1$,*

$$\tau_r\triangleq\sup_k\|\pi_k\log^{\beta}(e+\pi_k)\|_{L_r}<\infty,\qquad(27)$$

then for any $p<r$, the tracking error $\{\widetilde{\boldsymbol{Z}}_k, k\geq 0\}$ is L_p-stable, and

$$\limsup_{k\to\infty}\|\widetilde{\boldsymbol{Z}}_k\|_{L_p}\leq c[\tau_r\log^{1+\beta/2}(e+\tau_r^{-1})],$$

where $c>0$ is a finite constant depending on $\{\boldsymbol{\Phi}_k\}$, $\boldsymbol{R}$, $\boldsymbol{Q}_{\mathrm{diag}}$ and p.

Remark 5 The proof of Theorem 4 is given in Sect. 5. By Theorem 4, we can see that the tracking error $\widetilde{\boldsymbol{Z}}_k$ is positively related to the parameter variation $\boldsymbol{\Delta}_{k+1}$ and the noise $\boldsymbol{N}_k$. Here we just require the moment assumptions on the parameter variation and the measurement noise, and do not require independency, stationarity or Gaussian property. Note that if $\boldsymbol{\Phi}_k$ and $\{\boldsymbol{N}_k, \boldsymbol{\Delta}_k\}$ are assumed to be independent, then condition (27) can be replaced by $\sup_k \|\pi_k\|_{L_r} < \infty$ which is a natural condition for the L_p-stability.

Remark 6 For the case that the communication topology has a certain uncertainty, to be specific, the switching communication graphs are governed by a homogeneous irreducible and aperiodic Markov chain whose states belong to a finite set. If all possible digraphs are balanced and the union of those digraphs is strongly connected, then we can obtain a similar result as that of Theorem 4 since the key inequality (34) may still hold, which will be investigated in our future work.

5 Proofs of main results

5.1 Proof of Theorem 1

We need the following lemmas for the proof.

Lemma 5 *[39] For any matrices X, Y, Z and D with proper dimensions, and the relevant matrices are all invertible, then $(X+YDZ)^{-1} = X^{-1} - X^{-1}Y(D^{-1}+ZX^{-1}Y)^{-1}ZX^{-1}$.*

Lemma 6 *[18] Assume that $\mathcal{A} = \{a_{ij}\} \in \mathbb{R}^{n\times n}$ is the adjacency matrix, and denote $\mathscr{A} = \mathcal{A}\otimes I_m$. Let $\bar{\boldsymbol{\Omega}}_{k+1}$ and $\boldsymbol{\Omega}_{k+1}$ be defined in (14), then for any $k \geq 1$,*

$$\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\mathscr{A} \leq \boldsymbol{\Omega}_{k+1}^{-1}, \quad \mathscr{A}\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}} \leq \bar{\boldsymbol{\Omega}}_{k+1}. \tag{28}$$

The proof of Theorem 1 is provided here:

Proof Consider the following equation for $t > s$,

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn} - \boldsymbol{A}_k)\boldsymbol{x}_k, \quad k \in [s, t-1], \tag{29}$$

where $\boldsymbol{x}_s \in \mathbb{R}^{mn}$ is taken to be deterministic and $\|\boldsymbol{x}_s\| = 1$. Then

$$\boldsymbol{x}_t = \prod_{k=s}^{t-1}\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn} - \boldsymbol{A}_k)\boldsymbol{x}_s. \tag{30}$$

Next, we consider the following Lyapunov function: $V_k = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\Omega}_k^{-1}\boldsymbol{x}_k$. Denote $\boldsymbol{B}_k = I - \boldsymbol{A}_k$, then by (29), Lemmas 5 and 6, we have

$$\begin{aligned} V_{k+1} &= \boldsymbol{x}_{k+1}^{\mathrm{T}}\boldsymbol{\Omega}_{k+1}^{-1}\boldsymbol{x}_{k+1} \\ &= \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{B}_k^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\mathscr{A}\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\boldsymbol{B}_k\boldsymbol{x}_k, \end{aligned} \tag{31}$$

and by (23) and (28), we have

$$\begin{aligned} &\boldsymbol{B}_k^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\mathscr{A}\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\boldsymbol{B}_k \\ &\leq \boldsymbol{B}_k^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\boldsymbol{B}_k = \boldsymbol{B}_k^{\mathrm{T}}(\boldsymbol{B}_k\boldsymbol{\Omega}_k\boldsymbol{B}_k^{\mathrm{T}} + \boldsymbol{Q}_k)^{-1}\boldsymbol{B}_k \\ &= \boldsymbol{\Omega}_k^{-1} - (\boldsymbol{\Omega}_k + \boldsymbol{\Omega}_k\boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{Q}_k^{-1}\boldsymbol{B}_k\boldsymbol{\Omega}_k)^{-1} \\ &= \boldsymbol{\Omega}_k^{-\frac{1}{2}}(I_{mn} - [I_{mn} + \boldsymbol{\Omega}_k^{\frac{1}{2}}\boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{Q}_k^{-1}\boldsymbol{B}_k\boldsymbol{\Omega}_k^{\frac{1}{2}}]^{-1})\boldsymbol{\Omega}_k^{-\frac{1}{2}} \\ &\leq (1 - [1 + \|\boldsymbol{Q}_k^{-1}\boldsymbol{B}_k\boldsymbol{\Omega}_k\boldsymbol{B}_k^{\mathrm{T}}\|]^{-1})\boldsymbol{\Omega}_k^{-1}, \end{aligned} \tag{32}$$

which yields

$$V_{k+1} \leq \left(1 - \frac{1}{1 + \|\boldsymbol{Q}_k^{-1}\bar{\boldsymbol{\Omega}}_{k+1}\|}\right)V_k.$$

Thus,

$$V_t \leq \prod_{k=s}^{t-1}\left(1 - \frac{1}{1 + \|\boldsymbol{Q}_k^{-1}\bar{\boldsymbol{\Omega}}_{k+1}\|}\right)V_s.$$

Hence we have

$$\begin{aligned} &\left\|\prod_{k=s}^{t-1}\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\mathrm{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn} - \boldsymbol{A}_k)\right\|^2 \\ &= \max_{\|\boldsymbol{x}_s\|=1}\|\boldsymbol{x}_t\|^2 = \max_{\|\boldsymbol{x}_s\|=1}\|\boldsymbol{x}_t\boldsymbol{\Omega}_t^{-\frac{1}{2}}\boldsymbol{\Omega}_t^{\frac{1}{2}}\|^2 \\ &\leq \max_{\|\boldsymbol{x}_s\|=1}\|\boldsymbol{x}_t\boldsymbol{\Omega}_t^{-\frac{1}{2}}\|^2\|\boldsymbol{\Omega}_t^{\frac{1}{2}}\|^2 = \max_{\|\boldsymbol{x}_s\|=1}V_t\|\boldsymbol{\Omega}_t\| \\ &\leq \left\{\prod_{k=s}^{t-1}\left[1 - \frac{1}{1 + \|\boldsymbol{Q}_k^{-1}\bar{\boldsymbol{\Omega}}_{k+1}\|}\right]\right\}\left\{\|\boldsymbol{\Omega}_t\| \max_{\|\boldsymbol{x}_s\|=1}V_s\right\} \\ &\leq \left\{\prod_{k=s}^{t-1}\left[1 - \frac{1}{1 + \|\boldsymbol{Q}_k^{-1}\bar{\boldsymbol{\Omega}}_{k+1}\|}\right]\right\}\left\{\|\boldsymbol{\Omega}_t\| \cdot \|\boldsymbol{\Omega}_s^{-1}\|\right\}, \end{aligned} \tag{33}$$

which completes the proof. □

5.2 Proof of Lemma 2

The following lemma investigates the property of $\{\boldsymbol{\Omega}_k\}$ which will be used for the proof of Lemma 2. Note that the notation $\mathrm{Tr}(\cdot)$ is denoted as the trace of the corresponding matrix.

Lemma 7 *Let Assumption 1 be satisfied and $\{\boldsymbol{\Omega}_k\}$ be generated by (14). Then*

$$T_{s+1} \leq (1 - b_{s+1})T_s + d, \tag{34}$$

where

$$T_s = \sum_{k=(s-1)\kappa'+D_{\mathcal{G}}}^{s\kappa'-1}\mathrm{Tr}(\boldsymbol{\Omega}_{k+1}), \; T_0 = 0, \; b_{s+1} = \frac{a_{\min}^2 c_{s+1}^1}{n\kappa c_{s+1}^2},$$

$$c_{s+1}^{1} = \mathrm{Tr}\Big(\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa' Q\Big)^{2}\sum_{j=1}^{n}\sum_{k=s\kappa'+D_{\mathcal{G}}}^{(s+1)\kappa'-1}\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^{2}}\Big),$$

$$c_{s+1}^{2} = \sum_{j=1}^{n}(r_j+1)\cdot\Big(1+\lambda_{\max}\Big\{\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa' Q\Big\}\Big) \cdot \mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa' Q\Big),$$

$$d = \frac{3}{2}nh(\kappa'+1)\mathrm{Tr}(Q),$$

and $a_{\min} = \min_{i,j\in\mathcal{V}} a_{ij}^{(D_{\mathcal{G}})} > 0$, $\kappa' = \kappa + D_{\mathcal{G}}$, *and* κ *is the constant appearing in Assumption 2.*

Proof Let $a_{ij}^{(k)}$ denote the i-th row j-th column element of the matrix $\mathcal{A}^k, k \geq 1$, where $a_{ij}^{(1)} = a_{ij}$. By (14) and the inequality $\big(\sum_{j=1}^{n} a_{ij}\boldsymbol{A}_j\big)^{-1} \leq \sum_{j=1}^{n} a_{ij}\boldsymbol{A}_j^{-1}$ with $\boldsymbol{A}_j > 0$ [40], we know that for any $k \in [s\kappa'+D_{\mathcal{G}}, (s+1)\kappa']$

$$\begin{aligned}
\Omega_{k,i} &= \Big\{\sum_{j=1}^{n} a_{ji}\bar{\Omega}_{k,j}^{-1}\Big\}^{-1} \leq \sum_{j=1}^{n} a_{ji}\bar{\Omega}_{k,j}\\
&= \sum_{j=1}^{n} a_{ji}(\bar{\Omega}_{k,j}-Q)+Q\\
&= \sum_{j=1}^{n} a_{ji}(\Omega_{k-1,j}^{-1}+r_j^{-1}\boldsymbol{\varphi}_{k-1,j}\boldsymbol{\varphi}_{k-1,j}^{\mathrm{T}})^{-1}+Q\\
&\leq \sum_{j=1}^{n} a_{ji}\Omega_{k-1,j}+Q\\
&\leq \sum_{j=1}^{n} a_{ji}\Big(\sum_{t=1}^{n} a_{tj}\Omega_{k-2,t}\Big)+2Q\\
&= \sum_{j=1}^{n} a_{ji}^{(2)}\Omega_{k-2,j}+2Q\\
&\leq \cdots \leq \sum_{j=1}^{n} a_{ji}^{(k-s\kappa')}\Omega_{s\kappa',j}+(k-s\kappa')Q\\
&\leq \sum_{j=1}^{n} a_{ji}^{(k-s\kappa')}\Omega_{s\kappa',j}+\kappa' Q.
\end{aligned} \tag{35}$$

Hence by the matrix inversion formula, i.e., Lemma 5, it follows that for any $k \in [s\kappa'+D_{\mathcal{G}}, (s+1)\kappa']$,

$$\begin{aligned}
&\Omega_{k+1,i}\\
&= \Big\{\sum_{j=1}^{n} a_{ji}\bar{\Omega}_{k+1,j}^{-1}\Big\}^{-1}\\
&= \Big\{\sum_{j=1}^{n} a_{ji}[(\Omega_{k,j}^{-1}+r_j^{-1}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}})^{-1}+Q]^{-1}\Big\}^{-1}\\
&\leq \sum_{j=1}^{n} a_{ji}(\Omega_{k,j}^{-1}+r_j^{-1}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}})^{-1}+Q\\
&\leq \sum_{j=1}^{n} a_{ji}\Big[\Big(\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big)^{-1} + r_j^{-1}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}\Big]^{-1}+Q\\
&= Q+\sum_{j=1}^{n} a_{ji}\Big[\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\\
&\quad -\Big(\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big)\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}\\
&\quad \cdot\frac{\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q}{r_j+\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}\Big(\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big)\boldsymbol{\varphi}_{k,j}}\Big]\\
&= \sum_{j=1}^{n} a_{ji}^{(k-s\kappa'+1)}\Omega_{s\kappa',j}+(\kappa'+1)Q\\
&\quad -\sum_{j=1}^{n} a_{ji}\Big(\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big)\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}\\
&\quad \cdot\frac{\Big(\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big)}{r_j+\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}\Big(\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big)\boldsymbol{\varphi}_{k,j}}\\
&\leq \sum_{j=1}^{n} a_{ji}^{(k-s\kappa'+1)}\Omega_{s\kappa',j}+(\kappa'+1)Q\\
&\quad -\sum_{j=1}^{n} a_{ji}\Big(\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big)\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^{2}}\\
&\quad \cdot\frac{\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q}{(r_j+1)\Big(1+\lambda_{\max}\Big\{\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big\}\Big)}.
\end{aligned} \tag{36}$$

From Assumption 1 and Remark 2, it is clear that $a_{ji}^{(D_{\mathcal{G}})} \geq a_{\min} > 0$, where $a_{\min} = \min_{i,j\in\mathcal{V}} a_{ij}^{(D_{\mathcal{G}})} > 0$, and $D_{\mathcal{G}}$ is denoted by the diameter of the graph $\mathcal{G}$. Hence, we have $a_{ji}^{(k)} \geq a_{\min}$ for any $k > D_{\mathcal{G}}$.

By C_r- and Schwarz inequalities, it is easy to obtain that $\sum_{j=1}^{n} a_j b_j \leq \sum_{j=1}^{n} a_j \sum_{j=1}^{n} b_j$, where the constants $a_j \geq 0, b_j \geq 0$. Furthermore, by choosing $a_j = \frac{c_j}{d_j}, b_j = d_j$ with $c_j \geq 0, d_j > 0$, then we can conclude that $\sum_{j=1}^{n}\frac{c_j}{d_j} \geq \frac{\sum_{j=1}^{n} c_j}{\sum_{j=1}^{n} d_j}$. For $k \in [s\kappa'+D_{\mathcal{G}}, (s+1)\kappa']$, we have by using the above inequalities that

$$\begin{aligned}
&\mathrm{Tr}(\boldsymbol{\Omega}_{k+1})\\
&= \mathrm{Tr}\Big(\sum_{i=1}^{n}\Omega_{k+1,i}\Big)\\
&\leq \mathrm{Tr}\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}^{(k-s\kappa'+1)}\Omega_{s\kappa',j}\Big)+n(\kappa'+1)\mathrm{Tr}(Q)\\
&\quad -\mathrm{Tr}\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}\Big[\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big]\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^{2}}\\
&\quad \cdot\frac{\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q}{(r_j+1)\Big(1+\lambda_{\max}\Big\{\sum_{t=1}^{n} a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa' Q\Big\}\Big)}\Big)
\end{aligned}$$

$$
\begin{aligned}
=&\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}\Big)+n(\kappa'+1)\mathrm{Tr}(Q)\\
&-\sum_{j=1}^{n}\frac{1}{(r_j+1)\Big(1+\lambda_{\max}\Big\{\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big\}\Big)}\\
&\cdot\mathrm{Tr}\Big(\Big[\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big]\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\\
&\cdot\Big[\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big]\Big)\\
\le&\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})+n(\kappa'+1)\mathrm{Tr}(Q)\\
&-\frac{1}{\sum_{j=1}^{n}(r_j+1)\Big(1+\lambda_{\max}\Big\{\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big\}\Big)}\\
&\cdot\mathrm{Tr}\Big(\sum_{j=1}^{n}\Big[\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big]\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\\
&\cdot\Big[\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big]\Big)\\
\le&\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})+n(\kappa'+1)\mathrm{Tr}(Q)\\
&-\frac{\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}\Big)}{\sum_{j=1}^{n}(r_j+1)\sum_{j=1}^{n}\Big(1+\lambda_{\max}\Big\{\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big\}\Big)}\\
&\cdot\frac{1}{\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)}\\
&\cdot\mathrm{Tr}\Big(\sum_{j=1}^{n}\Big[\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big]\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\\
&\cdot\Big[\sum_{t=1}^{n}a_{tj}^{(k-s\kappa')}\Omega_{s\kappa',t}+\kappa'Q\Big]\Big)\\
\le&\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})+n(\kappa'+1)\mathrm{Tr}(Q)\\
&-\frac{\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})}{n\sum_{j=1}^{n}(r_j+1)\cdot\Big(1+\lambda_{\max}\Big\{\sum_{t=1}^{n}\Omega_{s\kappa',t}+\kappa'Q\Big\}\Big)}\\
&\cdot\frac{a_{\min}^2\mathrm{Tr}\Big(\Big(\sum_{t=1}^{n}\Omega_{s\kappa',t}+\kappa'Q\Big)^2\sum_{j=1}^{n}\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\Big)}{\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)}.
\end{aligned}\tag{37}
$$

Summing both sides of (37), we obtain that

$$
\begin{aligned}
&T_{s+1}\\
&=\sum_{k=s\kappa'+D_{\mathcal{G}}}^{(s+1)\kappa'-1}\mathrm{Tr}(\boldsymbol{\Omega}_{k+1})\le\kappa\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})+n\kappa(\kappa'+1)\mathrm{Tr}(Q)\\
&\quad-\frac{a_{\min}^2\kappa\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})}{n\kappa\sum_{j=1}^{n}(r_j+1)\cdot\Big(1+\lambda_{\max}\Big\{\sum_{t=1}^{n}\Omega_{s\kappa',t}+\kappa'Q\Big\}\Big)}\\
&\quad\cdot\frac{\mathrm{Tr}\Big(\Big(\sum_{t=1}^{n}\Omega_{s\kappa',t}+\kappa'Q\Big)^2\sum_{j=1}^{n}\sum_{k=s\kappa'+D_{\mathcal{G}}}^{(s+1)\kappa'-1}\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\Big)}{\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)}\\
&\quad\le\kappa\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})-b_{s+1}\kappa\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})+n\kappa(\kappa'+1)\mathrm{Tr}(Q).
\end{aligned}
$$

Again we have

$$
\begin{aligned}
&h\mathrm{Tr}(\boldsymbol{\Omega}_{s\kappa'})\\
&=\sum_{k=(s-1)\kappa'+D_{\mathcal{G}}}^{s\kappa'-1}\sum_{j=1}^{n}\mathrm{Tr}(\Omega_{s\kappa',j})\\
&\le\sum_{k=(s-1)\kappa'+D_{\mathcal{G}}}^{s\kappa'-1}\sum_{j=1}^{n}\mathrm{Tr}\Big(\sum_{t=1}^{n}a_{t,j}^{(s\kappa'-k)}\Omega_{k+1,t}+(s\kappa'-k)Q\Big)\\
&=T_s+\frac{1}{2}n\kappa(\kappa'+1)\mathrm{Tr}(Q),
\end{aligned}\tag{38}
$$

and

$$
\begin{aligned}
T_{s+1}&\le(1-b_{s+1})T_s+\frac{3}{2}n\kappa(\kappa'+1)\mathrm{Tr}(Q)\\
&=(1-b_{s+1})T_s+d,\quad s\ge0,
\end{aligned}\tag{39}
$$

which completes the proof. □

Proof of Lemma 2 Denote $\mathcal{H}_s=\mathcal{F}_{s\kappa'-1}$. Then it is clear that T_s and b_s are $\mathcal{H}_s$-measurable, and

$$
b_{s+1}\in\Big[0,\frac{a_{\min}^2}{\sum_{i=1}^{n}(r_i+1)}\Big].\tag{40}
$$

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

$$
\begin{aligned}
&\mathbb{E}[b_{s+1}|\mathcal{H}_s]\\
&\ge\frac{a_{\min}^2n(1+\kappa)\alpha_s'\mathrm{Tr}\Big(\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)^2\Big)}{n\kappa\Big(\sum_{i=1}^{n}r_i+n\Big)(1+\kappa'\|Q\|)\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)}\\
&\ge\frac{a_{\min}^2\alpha_s'\Big[\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)\Big]^2}{m\Big(\sum_{i=1}^{n}r_i+n\Big)(1+\kappa'\|Q\|)\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)}\\
&=\frac{a_{\min}^2\alpha_s'\mathrm{Tr}\Big(\sum_{j=1}^{n}\Omega_{s\kappa',j}+\kappa'Q\Big)}{m\Big(\sum_{i=1}^{n}r_i+n\Big)(1+\kappa'\|Q\|)}\\
&\ge\frac{a_{\min}^2\kappa'\|Q\|\alpha_s'}{m\Big(\sum_{i=1}^{n}r_i+n\Big)(1+\kappa'\|Q\|)},
\end{aligned}\tag{41}
$$

where

$$
\alpha_s'=p\lambda_{\min}\Big\{\mathbb{E}\Big[\sum_{j=1}^{n}\sum_{k=s\kappa'+D_{\mathcal{G}}}^{(s+1)\kappa'-1}\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^{\mathrm{T}}}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\Big|\mathcal{H}_s\Big]\Big\}
$$

with $p=\frac{1}{n(1+\kappa)}$. By this condition and applying Lemmas 2.1–2.3 in [36], it is easy to see that $\{b_{k+1}\}\in S^0(\gamma)$ where $\gamma\in[0,1)$. Consequently, using the definition of $S^0(\cdot)$, we know that

$$
\mathbb{E}\Big[\sum_{k=s}^{t}(1-b_{k+1})\Big]\le C\gamma^{t-s+1},\forall t\ge s\ge0,\tag{42}
$$

where $C > 0$ and $\gamma \in [0, 1)$ are two constants.

From Lemma 7, we know that for any $\varepsilon > 0$, $\exp(\varepsilon T_{s+1}) \leq \exp((1 - b_{s+1})\varepsilon T_s) \cdot \exp(d\varepsilon)$. By the inequality $\exp(ax) \leq a\exp(x) + 1$, $0 < a < 1$, $x > 0$, we get

$$\exp(\varepsilon T_{s+1}) \leq \exp(d\varepsilon) \cdot [(1 - b_{s+1})\exp(\varepsilon T_s) + 1]. \tag{43}$$

From this it is clear that if ε^* is chosen to be small enough such that $\exp(d\varepsilon)\gamma < 1$, then we have by (42) and (43) that

$$\sup_{s \geq 0} \mathbb{E}[\exp(\varepsilon T_s)] < \infty, \forall \varepsilon \in (0, \varepsilon^*).$$

This completes the proof. □

5.3 Proof of Lemma 3

Denote

$$x_s = \frac{\kappa(1 + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\boldsymbol{Q}_{\text{diag}}\|) + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| T_s}{\mu},$$

where T_s is defined in Lemma 7. Then we have

$$x_{s+1} \leq (1 - b_{s+1})x_s + \frac{\kappa(1 + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\boldsymbol{Q}_{\text{diag}}\|) + d\|\boldsymbol{Q}_{\text{diag}}^{-1}\|}{\mu}.$$

It is easy to see from (41), Assumption 2 and Lemma 2.3 in [36] that Lemma 3.1 in [36] is applicable to the above equation. Hence we know that

$$\left\{\frac{1}{x_s}\right\} \in S^0(\gamma),$$

for some $\gamma \in (0, 1)$. Note that

$$x_s = \sum_{k=(s-1)\kappa' + D_{\mathcal{G}}}^{s\kappa' - 1} \frac{1 + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \|\boldsymbol{Q}_{\text{diag}}\| + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \text{Tr}(\boldsymbol{\Omega}_{k+1})}{\mu}.$$

Similar to the proof of Lemma 5 in [37], it follows that

$$\left\{\frac{\mu}{1 + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\boldsymbol{Q}_{\text{diag}}\| + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \text{Tr}(\boldsymbol{\Omega}_k)}\right\} \in S^0(\alpha)$$

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

$$\left\{\frac{\mu}{1 + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\boldsymbol{Q}_{\text{diag}}\| + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\boldsymbol{\Omega}_k\|}\right\} \in S^0(\alpha).$$

Since $(\bar{\boldsymbol{\Omega}}_{k+1} - \boldsymbol{Q}_{\text{diag}})^{-1} = \boldsymbol{\Omega}_k^{-1} + \boldsymbol{R}^{-1}\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\text{T}}$, we have

$$\bar{\boldsymbol{\Omega}}_{k+1} \leq \boldsymbol{\Omega}_k + \boldsymbol{Q}_{\text{diag}},$$

$$\|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\bar{\boldsymbol{\Omega}}_{k+1}\| \leq \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\boldsymbol{\Omega}_k\| + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\boldsymbol{Q}_{\text{diag}}\|.$$

By this and the property of $S^0(\alpha)$, we can obtain that

$$\left\{\frac{\mu}{1 + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\bar{\boldsymbol{\Omega}}_{k+1}\|}\right\} \in S^0(\alpha)$$

holds for some $\alpha \in (0, 1)$.

5.4 Proof of Theorem 4

By (23) and the definition of $\boldsymbol{Q}_k$ in (22),

we obtain that $\boldsymbol{Q}_k \geq \boldsymbol{Q}_{\text{diag}}$ and $\bar{\boldsymbol{\Omega}}_k \geq \boldsymbol{Q}_{\text{diag}}$. Hence, by the definition of $\bar{\boldsymbol{\Omega}}_k^{-1}$, we know that

$$\|\boldsymbol{\Omega}_k^{-1}\| \leq \|\bar{\boldsymbol{\Omega}}_k^{-1}\| \leq \|\boldsymbol{Q}_{\text{diag}}^{-1}\|.$$

By Theorem 1, we have for $t > s$,

$$\begin{aligned}&\left\|\prod_{k=s}^{t-1} \boldsymbol{\Omega}_{k+1}\mathscr{A}^{\text{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}(I_{mn} - \boldsymbol{A}_k)\right\| \\ &\leq \left\{\prod_{k=s}^{t-1}\left(1 - \frac{1}{1 + \|\boldsymbol{Q}_{\text{diag}}^{-1}\| \cdot \|\bar{\boldsymbol{\Omega}}_{k+1}\|}\right)^{\frac{1}{2}}\right\} \\ &\quad \cdot \left\{\|\boldsymbol{\Omega}_t\|^{\frac{1}{2}} \cdot \|\boldsymbol{Q}_{\text{diag}}^{-1}\|^{\frac{1}{2}}\right\}.\end{aligned} \tag{44}$$

Note that

$$\|\boldsymbol{L}_k\| \leq \frac{\|\boldsymbol{\Omega}_k\|^{\frac{1}{2}}}{2\sqrt{r_{\min}}}, \tag{45}$$

and

$$\begin{aligned}&\|\boldsymbol{\Omega}_{k+1}\mathscr{A}^{\text{T}}\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\| \\ &\leq \|\boldsymbol{\Omega}_{k+1}\| \cdot \|\bar{\boldsymbol{\Omega}}_{k+1}^{-1}\| \leq \|\boldsymbol{\Omega}_{k+1}\| \cdot \|\boldsymbol{Q}_{\text{diag}}^{-1}\|\end{aligned} \tag{46}$$

hold, where $r_{\min} = \min_{i=1,\ldots,n}\{r_1, \ldots, r_n\}$. By the error equation (15), we have

$$\begin{aligned}\widetilde{\boldsymbol{Z}}_{k+1} &= \prod_{i=0}^{k} \boldsymbol{\Omega}_{i+1}\mathscr{A}^{\text{T}}\bar{\boldsymbol{\Omega}}_{i+1}^{-1}(I_{mn} - \boldsymbol{A}_i)\widetilde{\boldsymbol{Z}}_0 \\ &\quad + \sum_{i=0}^{k}\Bigg[\prod_{j=i+1}^{k} \boldsymbol{\Omega}_{j+1}\mathscr{A}^{\text{T}}\bar{\boldsymbol{\Omega}}_{j+1}^{-1}(I_{mn} - \boldsymbol{A}_j) \\ &\quad \cdot \boldsymbol{\Omega}_{i+1}\mathscr{A}^{\text{T}}\bar{\boldsymbol{\Omega}}_{i+1}^{-1} \cdot (-\boldsymbol{L}_i\boldsymbol{N}_i + \boldsymbol{\Delta}_{i+1})\Bigg].\end{aligned}$$

By (44)–(46), we obtain the following inequality for $p < r$

$$\|\widetilde{\boldsymbol{Z}}_{k+1}\|_{L_p}$$

$$
\begin{aligned}
&\leq \Big\| \prod_{i=0}^{k} \boldsymbol{\Omega}_{i+1} \mathscr{A}^{\mathrm{T}} \bar{\boldsymbol{\Omega}}_{i+1}^{-1} (I_{mn} - \boldsymbol{A}_i) \widetilde{\boldsymbol{Z}}_0 \Big\|_{L_p} \\
&\quad + \| \boldsymbol{Q}_{\mathrm{diag}}^{-1} \|^{\frac{3}{2}} \sum_{i=0}^{k} \Big\| \prod_{j=i+1}^{k} \Big(1 - \frac{1}{2(1 + \| \boldsymbol{Q}_{\mathrm{diag}}^{-1} \| \cdot \| \bar{\boldsymbol{\Omega}}_{j+1} \|)} \Big) \\
&\quad \cdot \| \boldsymbol{\Omega}_{k+1} \|^{\frac{1}{2}} \| \boldsymbol{\Omega}_{i+1} \| \Big(1 + \frac{\| \boldsymbol{\Omega}_i \|^{\frac{1}{2}}}{2\sqrt{r_{\min}}} \Big) \pi_i \Big\|_{L_p} \\
&\leq \Big\| \prod_{i=0}^{k} \boldsymbol{\Omega}_{i+1} \mathscr{A}^{\mathrm{T}} \bar{\boldsymbol{\Omega}}_{i+1}^{-1} (I_{mn} - \boldsymbol{A}_i) \widetilde{\boldsymbol{Z}}_0 \Big\|_{L_p} \\
&\quad + \| \boldsymbol{Q}_{\mathrm{diag}}^{-1} \|^{\frac{3}{2}} \sup_{i \geq 0} \| \boldsymbol{\Omega}_i \|_{L_q} \\
&\quad \cdot \sum_{i=0}^{k} \Big\| \prod_{j=i+1}^{k} \Big(1 - \frac{1}{2(1 + \| \boldsymbol{Q}_{\mathrm{diag}}^{-1} \| \cdot \| \bar{\boldsymbol{\Omega}}_{j+1} \|)} \Big) \\
&\quad \cdot \| \boldsymbol{\Omega}_{k+1} \|^{\frac{1}{2}} \Big(1 + \frac{\| \boldsymbol{\Omega}_i \|^{\frac{1}{2}}}{2\sqrt{r_{\min}}} \Big) \pi_i \Big\|_{L_r}.
\end{aligned} \tag{47}
$$

By Lemma 2 and the Schwarz inequality, we have

$$
\begin{aligned}
&\sup_{k \geq i} \mathbb{E}[\exp(\varepsilon \| \boldsymbol{\Omega}_{k+1} \|^{\frac{1}{2}} \cdot \| \boldsymbol{\Omega}_i \|^{\frac{1}{2}})] \\
&\leq \sup_{k \geq i} \{ \mathbb{E}[\exp(\varepsilon \| \boldsymbol{\Omega}_{k+1} \|)] \}^{\frac{1}{2}} \cdot \{ \mathbb{E}[\exp(\varepsilon \| \boldsymbol{\Omega}_i \|)] \}^{\frac{1}{2}} \\
&< \infty.
\end{aligned}
$$

By Lemmas 2, 3 and the condition (27), we can see that the rest part of the proof can be obtained by following the proof of Theorem 4.1 in [36], we omit the details of the proof here.

6 Simulation results

A simulation example is given here to show the cooperative property of networked systems: even though no individual nodes can track the unknown time-varying parameter vectors, the networked system can still accomplish the tracking task collaboratively. Besides, the performance properties of the distributed KF algorithm are given in contrast to other distributed adaptive filtering algorithms.

A networked system composed of $n = 6$ nodes is considered here, and $\mathcal{A}$ is chosen as follows,

$$
\mathcal{A} = \begin{pmatrix}
4/5 & 1/5 & 0 & 0 & 0 & 0 \\
0 & 4/5 & 1/5 & 0 & 0 & 0 \\
0 & 0 & 4/5 & 1/5 & 0 & 0 \\
0 & 0 & 0 & 4/5 & 1/5 & 0 \\
0 & 0 & 0 & 0 & 4/5 & 1/5 \\
1/5 & 0 & 0 & 0 & 0 & 4/5
\end{pmatrix}.
$$

Hence the corresponding graph is directed, balanced, and strongly connected.

An unknown 4-dimensional time-varying signal $\boldsymbol{\zeta}_k$ will be tracked, and we assume that the parameter variation in (2) obeys a Gaussian distribution, i.e., $\boldsymbol{\delta}_k \sim N(0, 0.04, 4, 1)$. The stochastic measurement noises $\{n_{k,i}, k \geq 1, i = 1, 2, 3, 4, 5, 6\}$ in (1) are chosen to be independent identically distributed (i.i.d.) with $n_{k,i} \sim N(0, 0.01, 1, 1)$. The regression vectors $\boldsymbol{\varphi}_{k,i} (i = 1, 2, 3, 4, 5, 6)$ are given by the a state space model as follows:

$$
\begin{cases}
\boldsymbol{x}_{k,i} = A_i \boldsymbol{x}_{k-1,i} + B_i \epsilon_{k,i}, \\
\boldsymbol{\varphi}_{k,i} = C_i \boldsymbol{x}_{k,i},
\end{cases} \tag{48}
$$

where $\{\epsilon_{k,i}\}$ $(i = 1, 2, 3, 4, 5, 6)$ are i.i.d. with $\epsilon_{k,i} \sim N(0, 0.4, 1, 1)$, and

$$
\begin{aligned}
&A_1 = A_3 = A_5 = \mathrm{diag}\{1/2, 2/3, 3/4, 2/5\}, \\
&A_2 = A_4 = \mathrm{diag}\{1/3, 1/2, 5/6, 3/5\}, \\
&A_6 = [\mathrm{col}\{3/4, 3/4, 3/4, 3/4\}, \boldsymbol{0}_{4 \times 3}], \\
&B_1 = B_5 = (1, 0, 0, 0)^{\mathrm{T}}, \ B_2 = B_6 = (0, 1, 0, 0)^{\mathrm{T}}, \\
&B_3 = (0, 0, 1, 0)^{\mathrm{T}}, \ B_4 = (0, 0, 0, 1)^{\mathrm{T}}, \\
&C_1 = C_5 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ C_2 = C_6 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
&C_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ C_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned}
$$

It is easy to check that Assumption 2 is satisfied.

For numerical simulations, let $\boldsymbol{x}_{0,i} = (1, 1, 1, 1)^{\mathrm{T}}$, $\boldsymbol{\zeta}_0 = (1, 1, 1, 1)^{\mathrm{T}}$, $\hat{\boldsymbol{\zeta}}_{0,i} = (0, 0, 0, 0)^{\mathrm{T}}$, $\Omega_{0,i} = I_4$, $r_i = 0.1$ $(i = 1, 2, 3, 4, 5, 6)$ and $Q = 0.1 \times I_4$. Here we will run the simulation process 200 times with the same initial states. Thus, for each node i, we can obtain the following 200 sequences:

$$
\begin{aligned}
&\{ \| \hat{\boldsymbol{\zeta}}_{k,i}^{j} - \boldsymbol{\zeta}_k^{j} \|^2, k = 1, \ldots, 1000 \}, \\
&\quad i = 1, \ldots, 6, \ j = 1, \ldots, 200,
\end{aligned}
$$

where j means the j-th simulation result. Then

$$
\frac{1}{200} \sum_{j=1}^{200} \| \hat{\boldsymbol{\zeta}}_{k,i}^{j} - \boldsymbol{\zeta}_k^{j} \|^2, \ \ i = 1, \ldots, 6, \ k = 1, \ldots, 1000
$$

is used to approximate the tracking error of sensor i in Fig. 1, and

$$
\frac{1}{200n} \sum_{j=1}^{200} \sum_{i=1}^{n} \| \hat{\boldsymbol{\zeta}}_{k,i}^{j} - \boldsymbol{\zeta}_k^{j} \|^2, \ \ k = 1, \ldots, 1000
$$

is used to approximate the average tracking errors over the whole network in Figs. 2 and 3.

Figure 1 shows the tracking errors of the six sensors using the non-cooperative traditional KF algorithm and the proposed distributed KF algorithm. From Fig. 1, we can see that if we use the non-cooperative one to track $\boldsymbol{\zeta}_k$, the errors of all six sensors are large since all sensors do not satisfy the information assumption (4), while the tracking errors of all six sensors in the distributed KF algorithm fall into the small neighborhood of 0 because all sensors cooperatively satisfy Assumption 2.

For the same regression vectors and initial settings as above, Fig. 2 compares our distributed KF algorithm ($r_i = 0.1$ $(i = 1, 2, 3, 4, 5, 6)$ and $Q = 0.1 \times I_4$) with three types of distributed LMS algorithms (i.e., distributed normalized LMS, combination then adaption (CTA) type distributed LMS and adaption then combination (ATC) type distributed LMS, see e.,g, [10, 11]. Both the step sizes of CTA type and ATC type distributed LMS algorithms are chosen to be $\mu = 0.8$. Also, the two step sizes of distributed normalized LMS algorithm are chosen to be $\mu = 0.5$ and $\nu = 0.4$. As depicted in Fig. 2, at time instant $k = 400$, the distributed KF algorithm exhibits a tracking error of approximately 0.011, significantly outperforming the ATC type, CTA type, and distributed normalized LMS algorithms, which respectively show tracking errors of approximately 2.106, 2.141, and 2.191. Furthermore, at time $k = 600$, the distributed KF algorithm maintains its precision with a tracking error of approximately 0.016, while the ATC type, CTA type, and distributed normalized LMS algorithm exhibit tracking errors of approximately 2.029, 2.064, and 2.143, respectively, underscoring the superiority of the distributed KF algorithm in terms of tracking error.

Moreover, for the same regression vectors and initial settings as above, Figs. 3 and 4 show the impact of measurement noise $n_{k,i}$ and parameter variation $\boldsymbol{\delta}_k$ on the performance of the distributed KF algorithm with $r_i = 0.1$ $(i = 1, 2, 3, 4, 5, 6)$ and $Q = 0.1 \times I_4$, respectively. From Fig. 3, we see that a direct correlation between the magnitude of noise variance and the size of the tracking error. That is, the tracking error of the distributed KF algorithm is positively related to the noise variance. From Fig. 4, it is evident that our algorithm can achieve effective tracking on different parameter variations, thereby demonstrating its robustness.

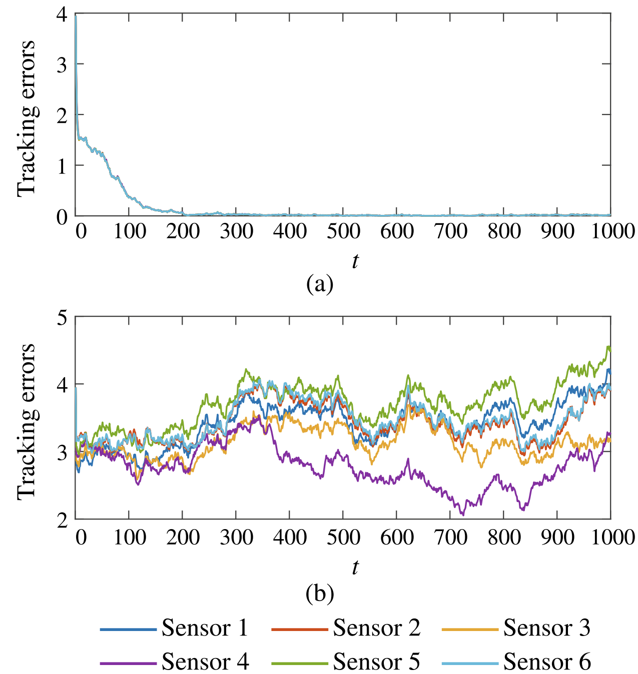


Fig. 1 Tracking errors of the six sensors by using **a** non-cooperative KF algorithm and **b** distributed KF algorithm

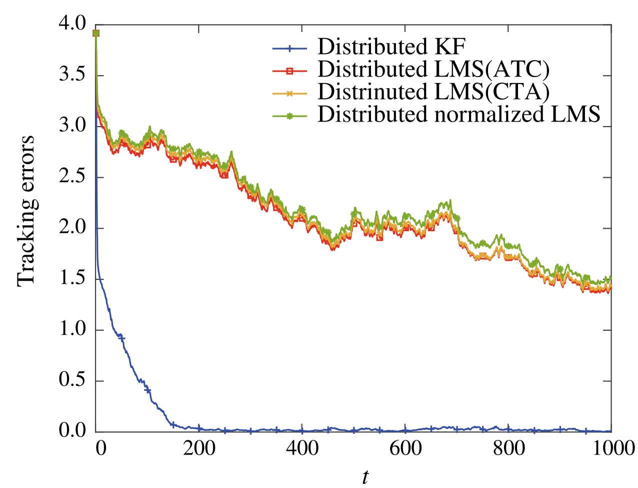


Fig. 2 Tracking errors of several distributed algorithms

7 Concluding remarks

We studied the stability of a distributed KF algorithm in this paper, which can be used to track a time-varying parameter vector cooperatively in sensor networks for stochastic regression models. Here we need no independency, no stationarity and no Gaussian property for the stability analysis. Thus it allows the given theory to be applied to stochastic feedback systems, which lays a theoretical foundation for future research on some related issues that combine learning, communication and control. In addition, the stochastic cooperative information condition shows the cooperative property in the sense that even though any sensor cannot track the parameters due to a lack of necessary excitations, the distributed KF algorithm can collaboratively accomplish the tracking task. Some relevant research topics should be considered, for example, to consider more general system models

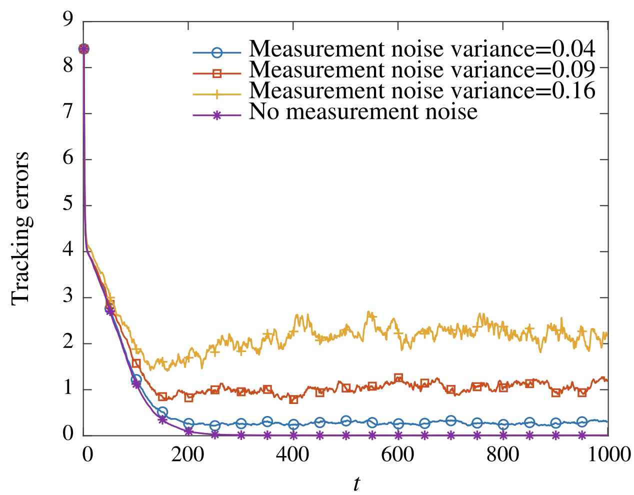


Fig. 3 Tracking errors of distributed KF algorithms on different measurement noises

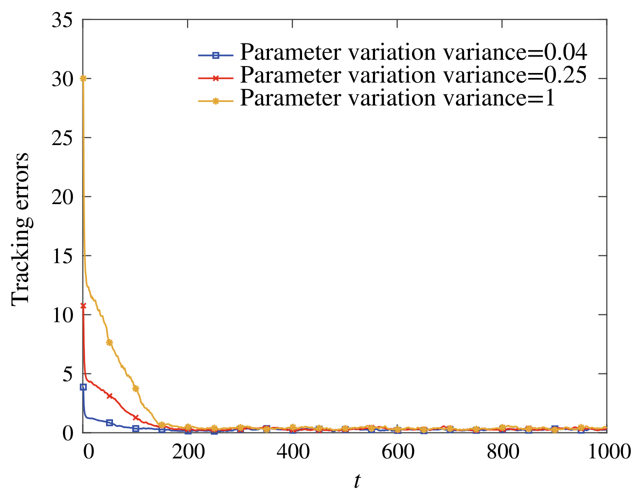


Fig. 4 Tracking errors of distributed KF algorithms on different parameter variations

with random coefficients, to combine distributed estimation or filtering methods with control problems, and so on.

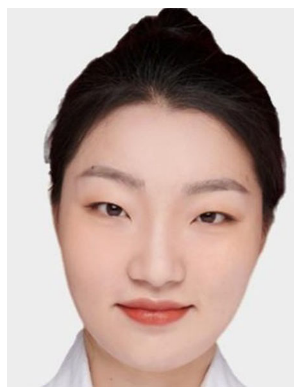
Data Availability Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. Hu, C., Hu, X., & Hong, Y. (2019). Distributed adaptive Kalman filter based on variational Bayesian technique. *Control Theory and Technology, 17*(1), 37–47.
2. Mahmoud, M. S., & Khalid, H. M. (2013). Distributed Kalman filtering: a bibliographic review. *IET Control Theory & Applications, 7*(4), 483–501.
3. Wang, J., & Li, T. (2021). Distributed multi-area state estimation for power systems with switching communication graphs. *IEEE Transactions on Smart Grid, 12*(1), 787–797.
4. Sayed, A. H. (2014). Adaptive networks. *Proceedings of the IEEE, 102*(4), 460–497.
5. Piggott, M. J., & Solo, V. (2016). Diffusion LMS with correlated regressors I: Realization-wise stability. *IEEE Transactions on Signal Processing, 64*(21), 5473–5484.
6. Nosrati, H., Shamsi, M., Taheri, S. M., & Sedaaghi, M. H. (2015). Adaptive networks under non-stationary conditions: Formulation, performance analysis, and application. *IEEE Transactions on Signal Processing, 63*(16), 4300–4314.
7. Gratton, C., Venkategowda, N. K., Arablouei, R., & Werner, S. (2019). Consensus-based distributed total least-squares estimation using parametric semidefinite programming. *In: IEEE International Conference on Acoustics, Speech and Signal Processing,* (ICASSP), pp. 5227–5231. Brighton, United Kingdom.
8. Mateos, G., Schizas, I. D., & Giannakis, G. B. (2009). Distributed recursive least-squares for consensus-based in-network adaptive estimation. *IEEE Transactions on Signal Processing, 57*(11), 4583–4588.
9. Xie, S., & Guo, L. (2018). A necessary and sufficient condition for stability of LMS-based consensus adaptive filters. *Automatica, 93*, 12–19.
10. Xie, S., & Guo, L. (2018). Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition. *SIAM Journal on Control and Optimization, 56*(5), 3404–3431.
11. Xie, S., & Guo, L. (2018). Analysis of distributed adaptive filters based on diffusion strategies over sensor networks. *IEEE Transactions on Automatic Control, 63*(11), 3643–3658.
12. Gan, D., & Liu, Z. (2024). Convergence of the distributed SG algorithm under cooperative excitation condition. *IEEE Transactions on Neural Networks and Learning Systems, 35*(5), 7087–7101.
13. Plata-Chaves, J., Bogdanovic, N., & Berberidis, K. (2013). Distributed incremental-based RLS for node-specific parameter estimation over adaptive networks. *IEEE Transactions on Signal Processing, 62*(20), 5382–5397.
14. Cattivelli, F. S., Lopes, C. G., & Sayed, A. H. (2008). Diffusion recursive least-squares for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing, 56*(5), 1865–1877.
15. Bertrand, A., Moonen, M., & Sayed, A. H. (2011). Diffusion bias-compensated RLS estimation over adaptive networks. *IEEE Transactions on Signal Processing, 59*(11), 5212–5224.
16. Arablouei, R., Doğançay, K., Werner, S., & Huang, Y.-F. (2014). Adaptive distributed estimation based on recursive least-squares and partial diffusion. *IEEE Transactions on Signal Processing, 62*(14), 3510–3522.
17. Vahidpour, V., Rastegarnia, A., Khalili, A., & Sanei, S. (2017). Analysis of partial diffusion recursive least squares adaptation over noisy links. *IET Signal Processing, 11*(6), 749–757.
18. Xie, S., Zhang, Y., & Guo, L. (2021). Convergence of a distributed least squares. *IEEE Transactions on Automatic Control, 66*(10), 4952–4959.
19. He, X., Xue, W., Fang, H., & Hu, X. (2020). Consistent Kalman filters for nonlinear uncertain systems over sensor networks. *Control Theory and Technology, 18*(4), 399–408.
20. Li, W., Jia, Y., & Du, J. (2016). Event-triggered Kalman consensus filter over sensor networks. *IET Control Theory & Applications, 10*(1), 103–110.
21. Carli, R., Chiuso, A., Schenato, L., & Zampieri, S. (2008). Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications, 26*(4), 622–633.
22. Wang, S., & Ren, W. (2018). On the convergence conditions of distributed dynamic state estimation using sensor networks: A unified framework. *IEEE Transactions on Control Systems Technology, 26*(4), 1300–1316.

23. Battistelli, G., & Chisci, L. (2014). Kullback-leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica, 50*(3), 707–718.
24. Das, S., & Moura, J. M. (2017). Consensus + innovations distributed Kalman filter with optimized gains. *IEEE Transactions on Signal Processing, 65*(2), 467–481.
25. Ma, K., Wu, S., Wei, Y., & Zhang, W. (2017). Gossip-based distributed tracking in networks of heterogeneous agents. *IEEE Communications Letters, 21*(4), 801–804.
26. Liu, Q., Wang, Z., He, X., & Zhou, D. (2018). On Kalman-consensus filtering with random link failures over sensor networks. *IEEE Transactions on Automatic Control, 63*(8), 2701–2708.
27. Battilotti, S., Cacace, F., d'Angelo, M., & Germani, A. (2018). Distributed Kalman filtering over sensor networks with unknown random link failures. *IEEE Control Systems Letters, 2*(4), 587–592.
28. Zhou, J., Gu, G., & Chen, X. (2019). Distributed Kalman filtering over wireless sensor networks in the presence of data packet drops. *IEEE Transactions on Automatic Control, 64*(4), 1603–1610.
29. Li, C., Dong, H., Li, J., & Wang, F. (2019). Distributed Kalman filtering for sensor network with balanced topology. *Systems & Control Letters, 131*, 104500.
30. He, X., Hu, C., Hong, Y., Shi, L., & Fang, H.-T. (2020). Distributed Kalman filters with state equality constraints: Time-based and event-triggered communications. *IEEE Transactions on Automatic Control, 65*(1), 28–43.
31. Vahidpour, V., Rastegarnia, A., Latifi, M., Khalili, A., & Sanei, S. (2020). Performance analysis of distributed Kalman filtering with partial diffusion over noisy network. *IEEE Transactions on Aerospace and Electronic Systems, 56*(3), 1767–1782.
32. Chen, S., Chen, X., Pei, Z., Zhang, X., & Fang, H. (2016). Distributed filtering algorithm based on tunable weights under untrustworthy dynamics. *IEEE/CAA Journal of Automatica Sinica, 3*(2), 225–232.
33. Gan, D., & Liu, Z. (2022). Distributed order estimation of ARX model under cooperative excitation condition. *SIAM Journal on Control and Optimization, 60*(3), 1519–1545.
34. Gan, D., & Liu, Z. (2023). Distributed sparse identification for stochastic dynamic systems under cooperative non-persistent excitation condition. *Automatica, 151*, 110958.
35. Gan, D., Xie, S., & Liu, Z. (2020). Stability of the distributed Kalman filter with general random coefficients. *Science China Information Sciences, 64*, 172204.
36. Guo, L. (1994). Stability of recursive stochastic tracking algorithms. *SIAM Journal on Control and Optimization, 32*(5), 1195–1225.
37. Guo, L. (1990). Estimating time-varying parameters by the Kalman filter based algorithm: stability and convergence. *IEEE Transactions on Automatic Control, 35*(2), 141–147.
38. Julier, S. J., & Uhlmann, J. K. (2001). *General Decentralized Data Fusion with Covariance Intersection*. Orlando, FL: CRC Press.
39. Guo, L. (2020). *Time-varying Stochastic Systems, Stability and Adaptive Theory* (2nd ed.). Beijing: Science Press.
40. Horn, R. A., & Johnson, C. R. (2013). *Matrix Analysis*. New York: Cambridge University Press.

Siyu Xie received the B.Sc. degree in information and computing science (systems control) from Beihang University in 2013, and Ph.D. degree in control theory from Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2018. She was a postdoctoral fellow from 2019 to 2022 at the Department of Electrical and Computer Engineering, Wayne State University, USA. She is currently a Professor in the School of Aeronautics and Astronautics with University of Electronic Science and Technology of China. She received the IEEE CSS Beijing Chapter Young Author Prize in 2015, China National Scholarship for graduate students in 2017, the President Scholarship of Chinese Academy of Sciences in 2018, and Excellent Doctoral dissertation of Chinese Academy of Sciences in 2019. Her research interests include networked systems, distributed optimization problems for power systems, distributed adaptive filters, identification, and distributed control.



Die Gan received the B.Sc. degree in mathematics from Shandong University, Jinan, China, in 2017, and the Ph.D. degree in systems the ory 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. She 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. Her research interests include system identification, machine learning, and distributed adaptive filters.

Zhixin Liu received the B.Sc. degree in Mathematics from Shandong University, Jinan, China, in 2002, and the Ph.D. degree in control theory from Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2007. She is currently a professor of AMSS, CAS. She had visiting positions at KTH Royal Institute of Technology, University of New South Wales, Canberra and University of Maryland, College Park. Her current research interests are complex systems and multi-agent systems. She is a co-author of the SIGEST paper in SIREV in 2014 and a co-recipient of the Best Theory Paper Award at The 13rd World Congress on Intelligent Control and Automation (WCICA).