SIAM J. CONTROL OPTIM.
Vol. 63, No. 1, pp. 650–675

DISTRIBUTED EXTENDED STOCHASTIC GRADIENT ALGORITHM FOR JOINT IDENTIFICATION OF SYSTEM PARAMETERS AND NOISE MODEL PARAMETERS*

DIE GAN†, REN YAN‡, SHUNING CHEN‡, AND ZHIXIN LIU§

Abstract. In this paper, we study the distributed identification problem of dynamic systems described by stochastic regression models with colored noise. To deal with challenges brought by the colored noise, we provide estimates for the unknown noise over a previous time period, and incorporate them with observed signals to construct extended regression vectors. Based on this, we develop a distributed extended stochastic gradient algorithm to estimate unknown parameter matrices by integrating the diffusion strategy of extended regression vectors with the consensus strategy of neighbors' estimates. We introduce cooperative nonpersistent excitation conditions to reflect the temporal and spatial union information condition, under which the almost sure convergence of the proposed distributed algorithm is established. Our results are obtained without assuming the independence, stationarity or persistent excitation conditions for the regressors and the Gaussian property for the system noises. Finally, numerical results are provided to verify the effectiveness of our proposed algorithm.

Key words. distributed identification, almost sure convergence, stochastic gradient algorithm, colored noise, ARMAX model, sensor network

MSC codes. 93B30, 68W15, 93A16

DOI. 10.1137/24M1643621

1. Introduction.

Over the last decade, distributed identification algorithms over sensor networks have attracted considerable interests. Compared with centralized algorithms with a fusion center, all the sensors in distributed ones cooperatively accomplish the identification task by communicating with theirs local neighbors. Distributed algorithms have advantages of robustness, saving communication and computational costs, and protecting privacy and, therefore, are applied in many practical scenarios including resource allocation, consensus seeking, and target tracking [28, 22, 16].

In the current literature, there are several local information fusion strategies to design the distributed adaptive estimation and filtering algorithms, where consensus [4, 23, 5, 25] and diffusion [24, 4, 9, 1] strategies are commonly used. Based on these strategies, different distributed estimation and filtering algorithms are designed for both time-invariant and time-varying systems, and corresponding analysis for the

*Received by the editors March 4, 2024; accepted for publication (in revised form) October 7, 2024; published electronically February 18, 2025.
https://doi.org/10.1137/24M1643621
Funding: This work was supported by the Natural Science Foundation of China under grants T2293772 and U21B6001, and by the Strategic Priority Research Program of the Chinese Academy of Sciences under grant XDA27030201.
†College of Artificial Intelligence, Nankai University, Tianjin, People's Republic of China (gandie@amss.ac.cn).
‡Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, 100190, People's Republic of China (yanren2021@amss.ac.cn, chenshuning@amss.ac.cn).
§Corresponding author. Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100190, People's Republic of China (lzx@amss.ac.cn).

performance of distributed algorithms is also investigated under some signal conditions. For the deterministic signals, Javed, Poveda, and Chen in [17] studied the stability of the cooperative gradient algorithm by assuming regression vectors to satisfy a cooperative persistent excitation (PE) condition. Chen et al. in [7] investigated the convergence of a distributed adaptive identification algorithm under a cooperative PE condition. Vahidpour et al. in [24] analyzed the mean-square steady-state performance of a partial diffusion Kalman filtering algorithm under the ergodicity condition of deterministic observation matrices.

The external interface is prevalent in practical situations, and some research has been conducted for the distributed identification problem of dynamic systems subject to noises. For example, Bertrand, Moonen, and Sayed in [4] studied the mean-square performance of a diffusion least squares (LS) algorithm with the independent and stationary regression vectors. Schizas, Mateos, and Giannakis in [23] established the stability of the consensus-based least mean squares (LMS) for the strictly stationary and ergodic regression vectors. Barani, Savadi, and Yazdi in [2] presented the convergence analysis of the diffusion stochastic gradient descent algorithm for independent and identically distributed (i.i.d.) signals. We see that most existing results on theoretical analysis of distributed estimation algorithms of dynamic stochastic systems are obtained by relying on independence or stationarity assumptions of regression vectors, due to mathematical challenges involved in analyzing product of random matrices. Such statistical assumptions are too stringent to be satisfied since the practical regression signals are often correlated caused by multipath effect or feedback. In order to remove such assumptions on regression signals, some efforts have been made, e.g., [11] for distributed stochastic gradient (SG) algorithms, [26] for LS-based distributed identification algorithms, [9, 10] for distributed order estimation and distributed sparse identification problems. We have added a specific expression of noise through the backward shift operator in line 162 In these studies, the system noises are often assumed to be a martingale difference sequence. However, in many practical applications such as target position tracking and speech signal extracting, it is not appropriate to model them as white noises because of the colored property of measurement noises [30]. Moreover, when collecting observed signals from spatially distributed sensors, the noise model is more complicated due to the sensors' environment [29]. The colored noise is often modeled as an autoregressive process in signal processing and statistical modeling, including the well-known autoregressive moving average system with exogenous inputs (ARMAX) model. So far very few investigations focus on the distributed estimation problems of dynamic systems with colored measurement noises, e.g., [20, 27, 29], whereas the conditions are imposed on regression signals such as requiring the coefficient matrices to be deterministic or assuming the covariance matrix of regressors to be time invariant. This motivates us to study the distributed identification problem for stochastic models with colored noise under general signal conditions.

We note that the SG algorithm is widely used to estimate time-invariant parameters in the fields of system identification and adaptive control due to its simplicity and low computational complexity, and the SG-based distributed algorithms attract considerable attention (cf. [5, 13]). In this paper, we develop a distributed identification algorithm based on the SG algorithm with colored noise to estimate unknown time-invariant parameters, and analyze performance of the proposed distributed algorithm. Compared with the case where the noise is a martingale difference sequence, in the case of colored noise, we need to identify unknown parameters in both the system and noise model simultaneously. To deal with this issue, we provide estimates of the noise over a past period of time, and incorporate them with the observed signals to build extended regression vectors. Based on this, we propose a novel distributed extended SG algorithm where the extended regression vectors are diffused over sensor networks; the consensus strategy of neighbors' estimates is also employed. We introduce a cooperative non-PE condition, a temporal and spatial union information condition on the stochastic regression vectors, to reflect the cooperative effect of multiple sensors. The challenges in theoretical analysis lie in analyzing properties of the product of nonindependent and nonstationary random matrices. By virtue of the powerful mathematical tools such as the martingale theory, strictly positive-real theory, and algebraic graph theory, we establish almost sure convergence of the proposed algorithm under the cooperative excitation condition. Different from existing results in the literature, e.g., [4, 23, 2], our theoretical results are obtained without relying on the independence or stationarity assumptions of the regression vector, which makes our theory applicable to stochastic feedback systems.

The rest of the paper is organized as follows. In section 2, we give some preliminaries on matrices, graphs, and stochastic processes. In section 3, we introduce the system model and the distributed extended SG algorithm. In section 4, we present the convergence results of the distributed algorithm proposed in this paper, with their proofs given in section 5. In section 6, we provide a numerical example to verify the effectiveness of the proposed algorithm. The concluding remarks are made in section 7.

2. Preliminaries.

In this section, we introduce some notations and preliminary results on matrices, graphs, and stochastic processes.

2.1. Matrix theory.

Let $\mathbb{R}^p$ and $\mathbb{R}^{p\times q}$ represent the set of p-dimensional real column vectors and $p\times q$-dimensional real matrices, respectively. For a matrix $A\in\mathbb{R}^{p\times q}$, $\|A\|$ refers to its Euclidean norm, i.e., $\|A\|=\sqrt{\lambda_{\max}(AA^T)}$, where $(\cdot)^T$ represents the transpose operator and $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of the matrix. Similarly, the minimum eigenvalue of a matrix is denoted as $\lambda_{\min}(\cdot)$. The condition number of an invertible matrix A is defined as $\|A\|\cdot\|A^{-1}\|$. For two symmetric matrices A and B, $A\geq B$ ($A>B, A\leq B, A<B$) means that $A-B$ is a positive semidefinite (positive definite, negative semidefinite, negative definite) matrix. The p-dimensional square identity matrix is denoted by $\boldsymbol{I}_p$. $\mathrm{Tr}(\cdot)$ denotes the trace operator of a square matrix. For l matrices $A_1,\cdots,A_l$ with the same dimension $p\times q$, the notation $\mathrm{col}\{A_1,\cdots,A_l\}\in\mathbb{R}^{pl\times q}$ means stacking l matrices into a column form. The notation $\mathrm{diag}\{A_1,\cdots,A_l\}\in\mathbb{R}^{pl\times ql}$ means constructing a block diagonal matrix with $A_1,\cdots,A_l$ as diagonal elements. The Kronecker product of matrix A and B is denoted by $A\otimes B$. For two positive scalar sequences $\{a_k\}$ and $\{b_k\}$, $a_k=O(b_k)$ means there exists a positive constant C independent of k such that $a_k\leq Cb_k$ holds for any $k\in\mathbb{N}$ with $\mathbb{N}$ being the set of natural numbers. A matrix is said to be nonnegative if all of its elements are nonnegative.

2.2. Graph theory.

Consider a graph $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{A})$ over n sensors, where $\mathcal{V}=\{1,2,\cdots,n\}$ is the sensor set, the edge set $\mathcal{E}$ refers to the communication between sensors, and $\mathcal{A}$ is the adjacency matrix. Denote the element in the ith row and jth column of $\mathcal{A}$ as a_{ij}. If sensor j can receive information from sensor i, then $(i,j)\in\mathcal{E}$ and $a_{ij}>0$, otherwise $(i,j)\notin\mathcal{E}$ and $a_{ij}=0$. In this paper, we assume that the topology of the sensor network is undirected, that is, $a_{ij}>0$ if and only if $a_{ji}>0$. The neighbor set of sensor i is defined as $\mathcal{N}_i=\{j|(j,i)\in\mathcal{E}\}$. A path of graph $\mathcal{G}$ with length h is a sequence of edges $(n_0,n_1),(n_1,n_2),\cdots,(n_{h-1},n_h)$ belonging to $\mathcal{E}$ with $\{n_0,n_1,n_2,\dots,n_h\}\subset\mathcal{V}$. A graph is said to be connected if there exists at least one path for any two nodes. The diameter of a graph is defined as the maximum value of the shortest path length between any two nodes, which is denoted by $D(\mathcal{G})$. The matrix $\mathcal{A}$ is nonnegative, and it is called row stochastic if $\sum_{j=1}^{n} a_{ij} = 1$ holds for all $i \in \{1, 2, \ldots, n\}$. Also, if $\sum_{i=1}^{n} a_{ij} = 1$ holds for all $j \in \{1, 2, \ldots, n\}$, then $\mathcal{A}$ is column stochastic. A matrix is called doubly stochastic if it is both row and column stochastic. For convenience of analysis, the adjacency matrix considered in this paper is symmetric and doubly stochastic. Then the Laplacian matrix of the graph can be written as $\mathcal{L} = \boldsymbol{I}_n - \mathcal{A}$. We first give a lemma about the properties of a Laplacian matrix.

LEMMA 2.1 (see [18]). *The Laplacian matrix $\mathcal{L}$ has at least one zero eigenvalue, with other eigenvalues positive and less than or equal to* 2. *Moreover, if the graph $\mathcal{G}$ is connected, then $\mathcal{L}$ has only one zero eigenvalue.*

2.3. Stochstic process. Let $(\Omega, \mathscr{F}, P)$ be a probability space, and $\{\mathscr{F}_k, k \geq 0\}$ be a nondecreasing family of sub-σ-algebras of $\mathscr{F}$. For the adapted sequence $\{X_k, \mathscr{F}_k\}_{k=0}^{\infty}$ on the probability space $(\Omega, \mathscr{F}, P)$, $\{X_k\}$ is a martingale difference sequence if it satisfies $\mathbb{E}|X_k| < \infty$ and $\mathbb{E}(X_k|\mathscr{F}_{k-1}) = 0$ almost surely (a.s.) $\forall k \geq 0$, where $\mathbb{E}(\cdot)$ and $\mathbb{E}(\cdot|\cdot)$ denote the expectation operator and conditional mathematical expectation, respectively. For a martingale difference sequence, we have the following martingale estimation theorem and martingale convergence theorem.

LEMMA 2.2 (martingale estimation theorem [6]). *Let $\{X_k, \mathscr{F}_k\}$ be a matrix martingale difference sequence and $\{M_k, \mathscr{F}_k\}$ be an adapted sequence of random matrices. If $\sup_k \mathbb{E}(\|X_{k+1}\|^{\alpha}|\mathscr{F}_k) < \infty$ a.s. for some $\alpha \in (0, 2]$, then as $t \to \infty$*

$$\sum_{k=0}^{t} M_k X_{k+1} = O\left(s_t(\alpha) \log^{\frac{1}{\alpha}+\eta}\left(s_t^{\alpha}(\alpha) + e\right)\right) \text{ a.s.} \quad \forall \eta > 0,$$

where $s_t(\alpha) = (\sum_{k=0}^{t} \|M_k\|^{\alpha})^{\frac{1}{\alpha}}$.

LEMMA 2.3 (martingale convergence theorem [6]). *Let $\{X_k, \mathscr{F}_k\}$ be a martingale difference sequence. Then as $t \to \infty$, $\xi_t = \sum_{k=1}^{t} X_k$ converges on the set S, where*

$$S = \left\{\sum_{k=1}^{\infty} \mathbb{E}\left(|X_k|\middle|\mathscr{F}_{k-1}\right) < \infty\right\}.$$

3. Problem formulation.

3.1. System description. This paper considers a sensor network consisting of n sensors. For each sensor i, we assume that its observation model is described by the following stochastic regression model:

$$y_{k+1}^{i} = \theta_0^T \phi_k^{0,i} + \epsilon_{k+1}^{i}, \quad k \geq 0, \tag{3.1}$$

where y_k^i denotes the m_1-dimensional measurement signal, $\phi_k^{0,i}$ is the m_2-dimensional stochastic regression vector of sensor i at the time instant k, $\theta_0 \in \mathbb{R}^{m_2 \times m_1}$ is the unknown system parameter matrix, $\{\epsilon_k^i\}$ is the m_1-dimensional colored noise which is driven by a martingale difference sequence $\{\omega_k^i, \mathscr{F}_k\}$, i.e., for any $k \geq 1$, $\mathbb{E}(\omega_k^i \mid \mathscr{F}_{k-1}) = 0$ holds, and

$$\epsilon_{k+1}^{i} = \omega_{k+1}^{i} + C_1 \omega_k^i + \cdots + C_r \omega_{k-r+1}^{i}, \tag{3.2}$$

where $\{\mathscr{F}_k\}$ is a family of nondecreasing σ-algebras, $\omega_k^i \equiv 0$ if $k \leq 0$, and $C_j \in \mathbb{R}^{m_1 \times m_1} (j = 1, \ldots, r)$ are unknown matrices.

For convenience of expression, we denote

$$C(z)=\boldsymbol{I}_{m_1}+C_1z+C_2z^2+\cdots+C_rz^r \tag{3.3}$$

with z being the shift-back operator, i.e., $zx_t=x_{t-1}$; then we have $\epsilon^i_{k+1}=C(z)\omega^i_{k+1}$.

Denote

$$\begin{aligned}\theta^T&=[\theta_0^T,C_1,\dots,C_r]\in\mathbb{R}^{m_1\times m},\\ \varphi_k^{0,i}&=[(\phi_k^{0,i})^T,(\omega_k^i)^T,\dots,(\omega_{k-r+1}^i)^T]^T\in\mathbb{R}^m\end{aligned} \tag{3.4}$$

with $m=m_2+m_1r$. Then, system (3.1) can be equivalently transformed into the following expression:

$$y^i_{k+1}=\theta^T\varphi_k^{0,i}+\omega^i_{k+1},\quad k\geq 0. \tag{3.5}$$

In this paper, our purpose is to design distributed identification algorithms for each sensor to estimate the unknown parameter matrix θ by using the information of itself and neighboring sensors.

The stochastic regression model (3.1) or (3.5), which includes the well-known ARMAX model, is widely used in the field of adaptive control and statistical learning. For distributed identification of such models, current research mainly focuses on the case of white noise [11, 26], and the algorithms and theoretical analysis for systems with colored noise are very few. However, in practical scenarios such as target position tracking and speech signal extraction, it may be more appropriate to characterize the external interference as colored noises (cf. [30]). This motivates us to study the distributed identification problem of stochastic regression models with colored noise.

3.2. Distributed extended SG algorithm. Differently from the model in [11] where all elements in $\varphi_k^{0,i}$ are known, the stochastic regression vector $\varphi_k^{0,i}$ in (3.4) is not completely available since it contains unknown noises. Thus, the algorithm in [11] cannot be used directly. To solve this problem, we give the following estimates for unknown noises $\omega^i_k,\dots,\omega^i_{k-r+1}$ in $\varphi_k^{0,i}$:

$$\hat{\omega}^i_t=y^i_t-(\hat{\theta}^i_{t-1})^T\varphi^i_{t-1},\quad t=k-r+2,\dots,k+1, \tag{3.6}$$

where $\hat{\theta}^i_{t-1}$ and φ^i_{t-1} are the estimates of unknown coefficients and the extended regression vector of sensor i at the time instant $t-1$ $(k-r+2\leq t\leq k+1)$ defined as follows:

$$\varphi^i_{k+1}=[(\phi^{0,i}_{k+1})^T,(\hat{\omega}^i_{k+1})^T,\dots,(\hat{\omega}^i_{k+2-r})^T]^T. \tag{3.7}$$

Based on this, we replace unknown noises $\omega^i_k,\dots,\omega^i_{k-r+1}$ in the regression vector $\varphi_k^{0,i}$ with the corresponding estimates. Thus, we propose the following distributed extended SG algorithm for the stochastic regression model (3.5) with colored noise to estimate the unknown parameter matrix θ. The details can be found in Algorithm 3.1.

In Algorithm 3.1, the term x^i_k can be regarded as the adaptive weighting coefficient between the estimates of the sensor i and its neighbors, and the term $1/r^i_k$ can be regarded as an adaptive gain of the innovation. Here, we consider a two time-scale scheme with communication rate $Q\geq 1$ in Step 1 of Algorithm 3.1, i.e., each sensor diffuses the norm of the extended regression vector with its neighbors for $Q\geq 1$ times to form the term x^i_k. The communication rate Q is important to guarantee convergence of the product of random matrices (see Theorem 17 in [11] for details of the proof), and this diffusion strategy is widely used in the design of the distributed

Algorithm 3.1. Distributed extended SG algorithm.

Input: the measurement $\{y_k^i\}_{k\geq 1}$, stochastic regression vectors $\{\phi_k^{0,i}\}_{k\geq 0}$, $i \in \{1,2,\ldots,n\}$, the step sizes μ,ν.

Output: estimates $\{\hat{\theta}_k^i\}_{k\geq 1}$, $i \in \{1,2,\ldots,n\}$.

Intialization: For each sensor $i \in \{1,\ldots,n\}$, begin with any initial estimates $\hat{\theta}_0^i \in \mathbb{R}^{m\times m_1}$, $r_0^i = 1$, $\hat{\omega}_j^i = 0$, $j \leq 0$, $\varphi_0^i = [(\phi_0^{0,i})^T, \underbrace{0,\ldots,0}_{r}]^T$.

for each time instant $k = 0,1,2,\ldots$ **do**

  for each sensor $i = 1,2,\ldots,n$ **do**

   Step 1. Diffuse the norm of extended regression vector:

   Let $x_k^i(0) = \frac{\|\varphi_k^i\|^2}{r_k^i}$.

   for $q = 1,2,\ldots,Q$ with $Q \geq 1$ **do**

    $x_k^i(q) = \sum_{j\in\mathcal{N}_i} a_{ij} x_k^j(q-1)$.

   end for

   Let $x_k^i = x_k^i(Q)$.

   Step 2. Update the estimate $\hat{\theta}_{k+1}^i$:

$$z_k^i = x_k^i \sum_{l\in\mathcal{N}_i} a_{li}(\hat{\theta}_k^i - \hat{\theta}_k^l),$$

$$\hat{\theta}_{k+1}^i = \underbrace{\hat{\theta}_k^i + \mu\frac{\varphi_k^i}{r_k^i}\left((y_{k+1}^i)^T - (\varphi_k^i)^T\hat{\theta}_k^i\right)}_{\text{Extended SG algorithm}} \underbrace{-\mu\nu\sum_{j\in\mathcal{N}_i} a_{ij}(z_k^i - z_k^j)}_{\text{Consensus-based item}},$$

   where μ and ν are two step sizes lying in $(0,1)$.

   Step 3. Update the extended regression vector φ_{k+1}^i and r_{k+1}^i:

$$\hat{\omega}_t^i = y_t^i - (\hat{\theta}_{t-1}^i)^T\varphi_{t-1}^i,\ \ t = k+2-r,\ldots,k+1,$$

$$\varphi_{k+1}^i = [(\phi_{k+1}^{0,i})^T, (\hat{\omega}_{k+1}^i)^T, \ldots, (\hat{\omega}_{k+2-r}^i)^T]^T,$$

$$r_{k+1}^i = r_k^i + \|\varphi_{k+1}^i\|^2.$$

  end for

end for

algorithms; see, e.g., [3, 12, 15, 21, 19]. In Step 2, the update of estimates consists of two parts: the first part is the extended SG algorithm which tries to minimize the prediction error, while the second part $\mu\nu\sum_{j\in\mathcal{N}_i} a_{ij}(z_k^i - z_k^j)$ can be regarded as the result of minimizing the weighted distance between the estimates of the sensor i and its neighbors. In Step 3, $r_k^i = 1 + \sum_{t=1}^k \|\varphi_t^i\|^2$, which increases with k, can eliminate the influence of noise on the estimation error.

For convenience of analysis, we introduce the following notations:

$$\begin{aligned}
&\boldsymbol{Y}_k = [y_k^1,\ldots,y_k^n] && (m_1\times n),\\
&\boldsymbol{\Phi}_k^0 = \mathrm{diag}\{\varphi_k^{0,1},\ldots,\varphi_k^{0,n}\} && (mn\times n),\\
&\boldsymbol{\Phi}_k = \mathrm{diag}\{\varphi_k^1,\ldots,\varphi_k^n\} && (mn\times n),\\
&\boldsymbol{W}_k = [\omega_k^1,\ldots,\omega_k^n] && (m_1\times n),\\
&\boldsymbol{\Theta} = \mathrm{col}\{\underbrace{\theta,\ldots,\theta}_{n}\} && (mn\times m_1),
\end{aligned}$$

$$\begin{aligned}
&\hat{\boldsymbol{\Theta}}_k = \mathrm{col}\{\hat{\theta}_k^1, \ldots, \hat{\theta}_k^n\} && (mn \times m_1),\\
&\widetilde{\boldsymbol{\Theta}}_k = \mathrm{col}\{\widetilde{\theta}_k^1, \ldots, \widetilde{\theta}_k^n\},\ \widetilde{\theta}_k^i = \theta - \hat{\theta}_k^i && (mn \times m_1),\\
&\boldsymbol{R}_k = \mathrm{diag}\{r_k^1, \ldots, r_k^n\} && (n \times n),\\
&\boldsymbol{A}_k = \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{\Phi}_k^T && (mn \times mn),\\
&\mathscr{L} = \mathcal{L} \otimes \boldsymbol{I}_m, \mathcal{L} \text{ is the Laplacian matrix} && (mn \times mn),\\
&\boldsymbol{X}_k(Q) = \mathrm{diag}\{x_k^1(Q), \ldots, x_k^n(Q)\} && (n \times n),\\
&\boldsymbol{G}_k = \boldsymbol{A}_k + \nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L} && (mn \times mn).
\end{aligned}$$

By the notations above, we can rewrite the system model (3.5) and the distributed extended SG algorithm (i.e., Algorithm 3.1) into the following matrix form:

$$\boldsymbol{Y}_{k+1} = \boldsymbol{\Theta}^T \boldsymbol{\Phi}_k^0 + \boldsymbol{W}_{k+1}, \tag{3.8}$$

$$\hat{\boldsymbol{\Theta}}_{k+1} = \hat{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1}(\boldsymbol{Y}_{k+1}^T - \boldsymbol{\Phi}_k^T \hat{\boldsymbol{\Theta}}_k) - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\hat{\boldsymbol{\Theta}}_k. \tag{3.9}$$

Denote the difference between the regressors $\varphi_k^{0,i}$ and φ_k^i as

$$\psi_k^{\zeta,i} := \mathrm{col}\{\underbrace{0, \ldots, 0}_{m_2}, \zeta_k^i, \ldots, \zeta_{k+1-r}^i\} = \varphi_k^i - \varphi_k^{0,i} \in \mathbb{R}^m, \tag{3.10}$$

where ζ_k^i represents the difference between noise ω_k^i and its estimate $\hat{\omega}_k^i$, i.e.,

$$\zeta_k^i = \hat{\omega}_k^i - \omega_k^i = y_k^i - (\hat{\theta}_{k-1}^i)^T \varphi_{k-1}^i - \omega_k^i \in \mathbb{R}^{m_1}. \tag{3.11}$$

Let $\boldsymbol{\zeta}_k = [\zeta_k^1, \ldots, \zeta_k^n] \in \mathbb{R}^{m_1 \times n}$ and $\boldsymbol{\Psi}_k^\zeta = \mathrm{diag}\{\psi_k^{\zeta,1}, \ldots, \psi_k^{\zeta,n}\} \in \mathbb{R}^{mn \times n}$. Then, we can straightly derive that

$$\boldsymbol{\Psi}_k^\zeta = \boldsymbol{\Phi}_k - \boldsymbol{\Phi}_k^0, \tag{3.12}$$

$$\boldsymbol{\zeta}_k = \boldsymbol{Y}_k - \hat{\boldsymbol{\Theta}}_{k-1}^T \boldsymbol{\Phi}_{k-1} - \boldsymbol{W}_k. \tag{3.13}$$

By (3.8), (3.9), and (3.12), we have

$$\begin{aligned}
\hat{\boldsymbol{\Theta}}_{k+1} &= \hat{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1}({\boldsymbol{\Phi}_k^0}^T \boldsymbol{\Theta} + \boldsymbol{W}_{k+1}^T - \boldsymbol{\Phi}_k^T \hat{\boldsymbol{\Theta}}_k) - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\hat{\boldsymbol{\Theta}}_k\\
&= \hat{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1}(\boldsymbol{\Phi}_k^T \boldsymbol{\Theta} - {\boldsymbol{\Psi}_k^\zeta}^T \boldsymbol{\Theta} + \boldsymbol{W}_{k+1}^T - \boldsymbol{\Phi}_k^T \hat{\boldsymbol{\Theta}}_k) - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\hat{\boldsymbol{\Theta}}_k\\
&= \hat{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1}(\boldsymbol{\Phi}_k^T \widetilde{\boldsymbol{\Theta}}_k - {\boldsymbol{\Psi}_k^\zeta}^T \boldsymbol{\Theta} + \boldsymbol{W}_{k+1}^T) - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\hat{\boldsymbol{\Theta}}_k.
\end{aligned}$$

Hence, by the fact $\mathscr{L}\boldsymbol{\Theta} = 0$, we can obtain the following estimation error equation:

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = (I - \mu \boldsymbol{G}_k)\widetilde{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} {\boldsymbol{\Psi}_k^\zeta}^T \boldsymbol{\Theta} - \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{W}_{k+1}^T. \tag{3.14}$$

Recursively define the state transition matrix $\boldsymbol{\Pi}(k,j)$ as follows:

$$\boldsymbol{\Pi}(k+1,j) = (\boldsymbol{I}_{mn} - \mu \boldsymbol{G}_k)\boldsymbol{\Pi}(k,j),\ \boldsymbol{\Pi}(j,j) = \boldsymbol{I}_{mn}. \tag{3.15}$$

4. The main results.

Before establishing the convergence results of Algorithm 3.1, we need to introduce some assumptions.

4.1. Assumptions.

Assumption 4.1. The graph $\mathcal{G}$ of the sensor network is connected.

Remark 4.1. Denote $\mathcal{A}^Q = \underbrace{\mathcal{A}\cdots\mathcal{A}}_{Q} = [a_{ij}^{(Q)}]$, i.e., $a_{ij}^{(Q)}$ is the ith row, jth column entry of the matrix $\mathcal{A}^Q$. By the theory of the product of stochastic matrices, if Assumption 4.1 is satisfied and Q is no less than the diameter of the graph, then we have $a_{ij}^{(Q)} > 0$ for all i and j.

To deal with the effect of the colored noise, we first introduce the concept of strictly positive-real (SPR).

Definition 4.1. *A matrix $H(z)$ of rational functions with real coefficients is called SPR if $H(z)$ has no poles in $|z| < 1$ and*

$$H(e^{j\lambda}) + H^T(-e^{j\lambda}) > 0 \ \ \forall \lambda \in [0, 2\pi].$$

A common property of SPR is listed in the lemma below.

Lemma 4.2 (see [6]). *Let $H(z)$ be a matrix of rational functions with real coefficients. The sequences $\{u_k\}$ and $\{v_k\}$ are connected by the transfer matrix $H(z)$, i.e., $v_k = H(z)u_k, k \geq 0$. If $H(z)$ is SPR, then there exists a constant $c > 0$ such that*

$$\sum_{k=0}^{t} u_k^T v_k \geq c \sum_{k=0}^{t} (\|u_k\|^2 + \|v_k\|^2) \ \forall\, t \geq 0. \tag{4.1}$$

Assumption 4.2. The transfer matrix $C(z) - \frac{\mu(1+4\nu)}{2} I_{m_1}$ is SPR, where $C(z)$ is defined in (3.3).

Remark 4.2. The SPR property guarantees stability of the noise part in a certain sense [6, 8], which is widely used in studying identification and adaptive control problems and plays a key role in dealing with colored noise, as shown by Lemma 4.6 in the next section.

We also need the following assumption on the noise for subsequent theoretical analysis.

Assumption 4.3. Assume that $\{\boldsymbol{W}_k, \mathscr{F}_k\}$ is a martingale difference sequence, where $\mathscr{F}_k = \sigma\{\varphi_j^i, \omega_j^i,\ j \leq k,\ i = 1, \ldots, n\}$, and there exists a positive constant c_0 (which may depend on sample ω) such that $\sup_k \mathbb{E}(\|\boldsymbol{W}_{k+1}\|^2 | \mathscr{F}_k) \leq c_0$ holds.

In addition, we need to introduce an excitation condition on the regression vector φ_k^i for convergence analysis.

Assumption 4.4 (cooperative nonpersistent excitation condition I). There exist two positive constants N and K_0 such that for $k \geq K_0$, the inequality

$$\frac{\lambda_{\max}^{(k)}}{\lambda_{\min}^{(k)}} \leq N \left(\log(\|\boldsymbol{R}_k\|)\right)^{\frac{1}{3}} \quad \text{a.s.} \tag{4.2}$$

holds, where $\lambda_{\max}^{(k)}$ and $\lambda_{\min}^{(k)}$ represent the maximum and minimum eigenvalues of the matrix $\frac{n}{m}\boldsymbol{I}_m + \sum_{i=1}^{n}\sum_{j=1}^{k} \varphi_j^i (\varphi_j^i)^T$, respectively, and $\|\boldsymbol{R}_k\| \to \infty$ as $k \to \infty$.

Remark 4.3. In fact, Assumption 4.4 is a temporal and spatial union information condition on the regression vectors. It is clear that Assumption 4.4 is much weaker than the classical PE condition in the multiple sensor case [7] and includes many typical cases such as the i.i.d. signals and the stationary ergodic signals. Moreover, Assumption 4.4 illustrates the cooperative effect of multiple sensors in the sense that the whole sensor network can cooperatively finish the estimation task, even if any individual sensor cannot due to lack of necessary information; see the simulation results in section 6.

A natural problem is to ask whether Assumption 4.4 imposed on φ_k^i can be transformed into conditions on $\varphi_k^{0,i}$. Denote

$$
\begin{aligned}
r_k^{0,i} &= 1+\sum_{j=1}^{k}\|\boldsymbol{\varphi}_j^{0,i}\|^2,\\
\boldsymbol{R}_k^0 &= \operatorname{diag}\{r_k^{0,1},\dots,r_k^{0,n}\},\\
\lambda_{\max}^{(0,k)} &= \lambda_{\max}\left(\frac{n}{m}\boldsymbol{I}_m+\sum_{i=1}^{n}\sum_{j=1}^{k}\varphi_j^{0,i}(\varphi_j^{0,i})^T\right),\\
\lambda_{\min}^{(0,k)} &= \lambda_{\min}\left(\frac{n}{m}\boldsymbol{I}_m+\sum_{i=1}^{n}\sum_{j=1}^{k}\varphi_j^{0,i}(\varphi_j^{0,i})^T\right).
\end{aligned}
$$

Then, we introduce the following excitation condition on $\{\varphi_k^{0,i}\}$ by using $\boldsymbol{R}_k^0$, $\lambda_{\max}^{(0,k)}$, and $\lambda_{\min}^{(0,k)}$.

Assumption 4.5 (cooperative nonpersistent excitation condition II). There exist two positive constants N and K_0 such that for $k\geq K_0$, the inequality

$$
\frac{\lambda_{\max}^{(0,k)}}{\lambda_{\min}^{(0,k)}}\leq N\left(\log(\|\boldsymbol{R}_k^0\|)\right)^{\frac{1}{3}}\quad \text{a.s.} \tag{4.3}
$$

holds, and $\|\boldsymbol{R}_k^0\|\to\infty$ as $k\to\infty$.

For the regression vectors $\varphi_k^{0,i}$, we define some notations similar to the extended regressors φ_k^i:

$$
\begin{aligned}
x_k^{0,i}(Q) &= \sum_{j=1}^{n}a_{ij}^{(Q)}\frac{\|\boldsymbol{\varphi}_k^{0,j}\|^2}{r_k^{0,j}}\in\mathbb{R},\\
\boldsymbol{A}_k^0 &= \boldsymbol{\Phi}_k^0(\boldsymbol{R}_{\boldsymbol{k}}^{\boldsymbol{0}})^{-1}{\boldsymbol{\Phi}_k^0}^T\in\mathbb{R}^{mn\times mn},\\
\boldsymbol{X}_k^0(Q) &= \operatorname{diag}\{x_k^{0,1}(Q),\dots,x_k^{0,n}(Q)\}\in\mathbb{R}^{n\times n},\\
\boldsymbol{G}_k^0 &= \boldsymbol{A}_k^0+\nu\mathscr{L}(\boldsymbol{X}_k^0(Q)\otimes\boldsymbol{I}_m)\mathscr{L}\in\mathbb{R}^{mn\times mn}.
\end{aligned}
$$

Also, we define the matrix $\boldsymbol{\Pi}^0(k,j)$ in a similar way as $\boldsymbol{\Pi}(k,j)$, i.e.,

$$
\boldsymbol{\Pi}^0(k+1,j)=(\boldsymbol{I}_{mn}-\mu\boldsymbol{G}_k^0)\boldsymbol{\Pi}^0(k,j),\ \ \boldsymbol{\Pi}^0(j,j)=\boldsymbol{I}_{mn}. \tag{4.4}
$$

LEMMA 4.3. [11]. *Suppose that Assumption* 4.1 *is satisfied. If* $\mu>0,\nu>0$*, and* $\mu(1+4\nu)\leq 1$*, then we have*

$$
0\leq\mu\boldsymbol{G}_k\leq\boldsymbol{I}_{mn}.
$$

From Lemma 4.3, we have $\|\boldsymbol{\Pi}(k,j)\|\leq 1$. Similarly, we can prove that $\|\boldsymbol{\Pi}^0(k,j)\|\leq 1$.

4.2. Convergence results. To analyze the convergence of Algorithm 3.1, we first introduce the following basic lemmas.

LEMMA 4.4 (see [14]). *Suppose that* $\{d_k\}_{k=1}^{\infty}$ *is a sequence of nonnegative real numbers. Then, we have the following results:*

$$\text{(i)} \sum_{k=1}^{\infty} \frac{d_k}{D_k^{\alpha}} < \infty \quad \forall\, \alpha > 1,$$
$$\text{(ii)} \sum_{k=1}^{\infty} \frac{d_k}{D_k} = \infty \textit{ iff } \lim_{k\to\infty} D_k = \infty,$$

where $D_t = 1 + \sum_{k=1}^{t} d_k$.

LEMMA 4.5 (Kronecker lemma [6]). *Let* $a_1, a_2, \ldots$ *and* $0 < b_1 < b_2 < \cdots$ *be real sequences such that* $\{b_k\}$ *increases to infinity as* $k \to \infty$. *If* $\sum_{k=1}^{\infty} \frac{a_k}{b_k}$ *converges, then* $\lim_{t\to\infty} \frac{1}{b_t} \sum_{k=1}^{t} a_k = 0$.

Compared with the convergence analysis in [11], we see that the main challenge lies in analyzing the property of $\mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{\Psi}_k^{\zeta^T} \boldsymbol{\Theta}$ in the error equation (3.14). To deal with this term, by the definition of $\psi_k^{\zeta,i}$ in (3.10), we first establish the boundedness of $\{\mathrm{Tr}(\sum_{j=0}^{k} \boldsymbol{\zeta}_{j+1} \boldsymbol{R}_j^{-1} \boldsymbol{\zeta}_{j+1}^T)\}$ using the SPR property and martingale theory in the following lemma.

LEMMA 4.6. *If step sizes* μ, ν *in Algorithm* 3.1 *satisfy* $\mu(1 + 4\nu) \le 1$, *then under Assumptions* 4.1, 4.2, *and* 4.3, $\mathrm{Tr}(\widetilde{\boldsymbol{\Theta}}_k^T \widetilde{\boldsymbol{\Theta}}_k)$ *and* $\mathrm{Tr}\left(\sum_{j=0}^{k} \boldsymbol{\zeta}_{j+1} \boldsymbol{R}_j^{-1} \boldsymbol{\zeta}_{j+1}^T\right)$ *are* a.s. bounded.

The proof of Lemma 4.6 is very complicated, and we present it in subsection 5.1.

By the definition of $\boldsymbol{\Psi}_k^{\zeta}, \boldsymbol{\zeta}_k$, and $\boldsymbol{R}_k^{-1}$ and Lemma 4.6, we have

$$\begin{aligned} \mathrm{Tr}\left(\sum_{k=0}^{\infty} \boldsymbol{\Psi}_k^{\zeta} \boldsymbol{R}_k^{-1} \boldsymbol{\Psi}_k^{\zeta^T}\right) &= \sum_{i=1}^{n} \sum_{k=0}^{\infty} \sum_{j=0}^{r-1} \frac{\mathrm{Tr}\left(\zeta_{k-j}^i (\zeta_{k-j}^i)^T\right)}{r_k^i} \\ &= \sum_{i=1}^{n} \sum_{k=0}^{\infty} \sum_{j=0}^{r-1} \frac{(\zeta_{k-j}^i)^T \zeta_{k-j}^i}{r_{k-j-1}^i} < \infty, \end{aligned} \tag{4.5}$$

which gives the boundedness property of $\mathrm{Tr}\left(\sum_{k=0}^{\infty} \boldsymbol{\Psi}_k^{\zeta} \boldsymbol{R}_k^{-1} \boldsymbol{\Psi}_k^{\zeta^T}\right)$.

Remark 4.4. We note that if Assumption 4.3 is relaxed to the weaker noise condition

$$\sup_k \mathbb{E}\left(\|\boldsymbol{R}_k^{-\frac{\epsilon}{2}} \boldsymbol{W}_{k+1}^T\|^2 | \mathscr{F}_k\right) < \infty$$

with $\epsilon \in (0, 1)$, then under Assumptions 4.1 and 4.2, we see that Lemma 4.6 still holds by Lemma 4.4 and following the proof line of Lemma 4.6.

In the following, we will provide a sufficient condition for the convergence of the distributed extended SG algorithm.

THEOREM 4.7. *Suppose that the condition number of* $\boldsymbol{R}_k$ *is bounded* (*i.e., there exists a positive constant* γ *which may depend on the sample* ω *such that* $\sup_k \max_{1\le i\le n}$

$r_k^i/\min_{1\le i\le n} r_k^i \le \gamma\big)$. *Under the conditions of Lemma* 4.6, *if* $\boldsymbol{\Pi}(k,0) \xrightarrow[k\to\infty]{} 0$ *a.s., then* $\hat{\boldsymbol{\Theta}}_k \xrightarrow[k\to\infty]{} \boldsymbol{\Theta}$ *a.s. for any initial value* $\hat{\boldsymbol{\Theta}}_0$.

Proof. By (3.14), it can be recursively derived that

$$\begin{aligned}\widetilde{\boldsymbol{\Theta}}_{k+1} = {}& \boldsymbol{\Pi}(k+1,0)\widetilde{\boldsymbol{\Theta}}_0 + \mu\sum_{j=0}^{k}\boldsymbol{\Pi}(k+1,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{\Psi}_j^{\zeta^T}\boldsymbol{\Theta}\\ &-\mu\sum_{j=0}^{k}\boldsymbol{\Pi}(k+1,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T.\end{aligned}\tag{4.6}$$

Since $\boldsymbol{\Pi}(k,0) \xrightarrow[k\to\infty]{} 0$, we only need to prove that the last two terms on the right side of the above equation will converge to zero. Denote $(\boldsymbol{\Pi}(\cdot,\cdot))^T$ as $\boldsymbol{\Pi}^T(\cdot,\cdot)$; then for any k,

$$\begin{aligned} mn &= \mathrm{Tr}\Big(\boldsymbol{\Pi}(k,k)\boldsymbol{\Pi}^T(k,k)\Big)\\ &\ge \mathrm{Tr}\left(\sum_{j=0}^{k-1}\Big[\boldsymbol{\Pi}(k,j+1)\boldsymbol{\Pi}^T(k,j+1)-\boldsymbol{\Pi}(k,j)\boldsymbol{\Pi}^T(k,j)\Big]\right)\\ &= \mathrm{Tr}\left(\sum_{j=0}^{k-1}\boldsymbol{\Pi}(k,j+1)\Big[\boldsymbol{I}_{mn}-\boldsymbol{\Pi}(j+1,j)\cdot\boldsymbol{\Pi}^T(j+1,j)\Big]\boldsymbol{\Pi}^T(k,j+1)\right)\\ &\ge \mathrm{Tr}\left(\sum_{j=0}^{k-1}\boldsymbol{\Pi}(k,j+1)\Big[\mu\boldsymbol{G}_j+\mu\boldsymbol{G}_j(\boldsymbol{I}_{mn}-\mu\boldsymbol{G}_j)\Big]\boldsymbol{\Pi}^T(k,j+1)\right)\\ &\ge \mathrm{Tr}\left(\sum_{j=0}^{k-1}\boldsymbol{\Pi}(k,j+1)\mu\boldsymbol{G}_j\boldsymbol{\Pi}^T(k,j+1)\right)\\ &\ge \mathrm{Tr}\left(\sum_{j=0}^{k-1}\boldsymbol{\Pi}(k,j+1)\mu\boldsymbol{A}_j\boldsymbol{\Pi}^T(k,j+1)\right)\\ &\ge \mu\sum_{j=0}^{k-1}\|\boldsymbol{\Pi}(k,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-\frac{1}{2}}\|^2.\end{aligned}$$

Then we have

$$\sum_{j=0}^{k-1}\|\boldsymbol{\Pi}(k,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-\frac{1}{2}}\|^2 \le \frac{mn}{\mu}.\tag{4.7}$$

From (4.5) and (4.7), it can be obtained by Hölder's inequality that

$$\begin{aligned}&\left\|\sum_{j=0}^{k}\boldsymbol{\Pi}(k+1,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{\Psi}_j^{\zeta^T}\right\|\\ &\quad\le\left(\sum_{j=M+1}^{k}\|\boldsymbol{\Pi}(k,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-\frac{1}{2}}\|^2\right)^{\frac{1}{2}}\left(\sum_{j=M+1}^{k}\|\boldsymbol{R}_j^{-\frac{1}{2}}\boldsymbol{\Psi}_j^{\zeta^T}\|^2\right)^{\frac{1}{2}}\end{aligned}$$

$$
\begin{aligned}
&+\left\|\sum_{j=0}^{M}\mathbf{\Pi}(k+1,j+1)\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\mathbf{\Psi}_j^{\zeta T}\right\| \\
&\leq\sqrt{\frac{mn}{\mu}}\left(\operatorname{Tr}\sum_{j=M+1}^{k}\mathbf{\Psi}_k^{\zeta}\boldsymbol{R}_k^{-1}\mathbf{\Psi}_k^{\zeta T}\right)^{\frac{1}{2}}+\left\|\sum_{j=0}^{M}\mathbf{\Pi}(k+1,j+1)\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\mathbf{\Psi}_j^{\zeta T}\right\| \\
&\longrightarrow 0,\ \text{ as } k\to\infty \text{ and } \ \text{then } M\to\infty.
\end{aligned}
$$

Then according to the sufficiency of Theorem 14 in [11], we have

$$
\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^{T}\longrightarrow 0,\ \text{ as } k\to\infty.
$$

This completes the proof of this theorem. ∎

According to [11], we know that if the cooperative non-PE condition I (i.e., Assumption 4.4) on the estimated regression vector $\{\varphi_k^i\}$ is satisfied, then we have $\mathbf{\Pi}(k,0)\xrightarrow[k\to\infty]{}0$ under some mild conditions. Furthermore, by Theorem 4.7, convergence of the distributed extended SG algorithm for the model (3.5) with colored noise can be obtained. It is natural to ask whether convergence of the algorithm can be obtained under the cooperative excitation condition on $\varphi_k^{0,i}$, i.e., Assumption 4.5. To answer this question, we first provide a lemma to guarantee boundedness of the condition number of $\boldsymbol{R}_k$.

LEMMA 4.8. *Under conditions of Lemma* 4.6*, if* $\|\boldsymbol{R}_k^0\|\xrightarrow[k\to\infty]{}\infty$*, and the condition number of* $\boldsymbol{R}_k^0$ *is bounded, then the condition number of* $\boldsymbol{R}_k$ *is bounded.*

Proof. By (4.5) we have for any $i\in\{1,\dots,n\}$

$$
\sum_{j=1}^{\infty}\frac{\|\psi_j^{\zeta,i}\|^2}{r_j^i}=\sum_{j=1}^{\infty}\|\psi_j^{\zeta,i}(r_j^i)^{-\frac{1}{2}}\|^2\leq\sum_{j=1}^{\infty}\|\mathbf{\Psi}_j^{\zeta}\boldsymbol{R}_j^{-\frac{1}{2}}\|^2<\infty. \tag{4.8}
$$

By the condition that the condition number of $\boldsymbol{R}_k^0$ is bounded for all $k\geq 0$, we see that there exists a positive constant γ_0 such that $\|\boldsymbol{R}_k^0\|\|(\boldsymbol{R}_k^0)^{-1}\|\leq\gamma_0$. Combining this with the condition $\|\boldsymbol{R}_k^0\|\to\infty$, we have $\|(\boldsymbol{R}_k^0)^{-1}\|\xrightarrow[k\to\infty]{}0$, which implies that for any i, $r_k^{0,i}\xrightarrow[k\to\infty]{}\infty$. We now show that $r_k^i\xrightarrow[k\to\infty]{}\infty$ by reduction to absurdity. According to the definition of r_k^i, we see that r_k^i is nondecreasing with k. Thus, the sequence r_k^i has a limit. If $r_k^i\to c<\infty$, then by (4.8), we have $\sum_{j=1}^{\infty}\|\psi_j^{\zeta,i}\|^2<\infty$. Hence from (3.10) we can get

$$
\begin{aligned}
r_k^{0,i}&=r_k^i-2\sum_{j=1}^{k}(\varphi_j^i)^T\psi_j^{\zeta,i}+\sum_{j=1}^{k}\|\psi_j^{\zeta,i}\|^2 \\
&\leq r_k^i+2\sqrt{r_k^i}\left(\sum_{j=1}^{k}\|\psi_j^{\zeta,i}\|^2\right)^{\frac{1}{2}}+\sum_{j=1}^{k}\|\psi_j^{\zeta,i}\|^2<\infty,
\end{aligned} \tag{4.9}
$$

which contradicts $r_k^{0,i}\to\infty$. Hence for any i, we have $r_k^i\to\infty$. Then by (4.8) and Lemma 4.5, we have $\frac{\sum_{j=1}^{k}\|\psi_j^{\zeta,i}\|^2}{r_k^i}\xrightarrow[k\to\infty]{}0$. For any $i\in\{1,\dots,n\}$, by the definition of r_k^i, we can derive that

$$\left|\frac{\sum_{j=1}^{k}(\varphi_j^i)^T\psi_j^{\zeta,i}}{r_k^i}\right| \le \frac{(\sum_{j=1}^{k}\|\varphi_j^i\|^2)^{\frac{1}{2}}(\sum_{j=1}^{k}\|\psi_j^{\zeta,i}\|^2)^{\frac{1}{2}}}{r_k^i} \le \left(\frac{\sum_{j=1}^{k}\|\psi_j^{\zeta,i}\|^2}{r_k^i}\right)^{\frac{1}{2}}. \tag{4.10}$$

Hence by (4.10), for any i, we have

$$\lim_{k\to\infty}\frac{r_k^{0,i}}{r_k^i} = \lim_{k\to\infty}\frac{r_k^i - 2\sum_{j=1}^{k}(\varphi_j^i)^T\psi_j^{\zeta,i} + \sum_{j=1}^{k}\|\psi_j^{\zeta,i}\|^2}{r_k^i} = 1. \tag{4.11}$$

Then we can obtain the following inequalities:

$$\limsup_{k\to\infty}\frac{\|\boldsymbol{R}_k\|}{\|\boldsymbol{R}_k^0\|} = \limsup_{k\to\infty}\frac{\max_i r_k^i}{\max_i r_k^{0,i}} \le \limsup_{k\to\infty}\frac{r_k^{i'}}{r_k^{0,i'}} \le \limsup_{k\to\infty}\sum_{i=1}^{n}\frac{r_k^i}{r_k^{0,i}} = n, \tag{4.12}$$

$$\limsup_{k\to\infty}\frac{\|\boldsymbol{R}_k^{-1}\|}{\|(\boldsymbol{R}_k^0)^{-1}\|} = \limsup_{k\to\infty}\frac{\min_i r_k^{0,i}}{\min_i r_k^i} \le \limsup_{k\to\infty}\frac{r_k^{0,i''}}{r_k^{i''}} \le \limsup_{k\to\infty}\sum_{i=1}^{n}\frac{r_k^{0,i}}{r_k^i} = n, \tag{4.13}$$

where $i' = i'(k) = \arg\max_i r_k^i \in \{1,2,\dots,n\}$, $i'' = i''(k) = \arg\min_i r_k^i \in \{1,2,\dots,n\}$.

Hence we have

$$\sup_k \|\boldsymbol{R}_k\|\|(\boldsymbol{R}_k)^{-1}\| = \sup_k \|\boldsymbol{R}_k^0\|\|(\boldsymbol{R}_k^0)^{-1}\| \cdot \frac{\|\boldsymbol{R}_k\|}{\|\boldsymbol{R}_k^0\|}\cdot\frac{\|\boldsymbol{R}_k^{-1}\|}{\|(\boldsymbol{R}_k^0)^{-1}\|} < \infty, \tag{4.14}$$

which completes the proof of this lemma. ∎

THEOREM 4.9. *Under the conditions of Lemma* 4.6*, if the condition number of* $\boldsymbol{R}_k^0$ *is bounded, then as* $k\to\infty$*,* $\boldsymbol{\Pi}(k,0)\to 0$ *if and only if* $\boldsymbol{\Pi}^0(k,0)\to 0$.

The detailed proof of Theorem 4.9 is given in subsection 5.2. It provides a necessary and sufficient condition for $\boldsymbol{\Pi}(k,0)\to 0$, and contributes to the following theorem for convergence of the algorithm under cooperative non-PE condition II (i.e., Assumption 4.5).

THEOREM 4.10. *Suppose that there exist* $i_1, i_2 \in \{1,\dots,n\}$ *such that* $r_k^{0,i_1} = O(r_{k-1}^{0,i_1})$*,* $r_k^{0,i_2} \xrightarrow[k\to\infty]{} \infty$*. Under the conditions of Theorem* 4.9*, if Assumption* 4.5 *is further satisfied and the communication rate* $Q \ge D(\mathcal{G})$*, then we have* $\widetilde{\boldsymbol{\Theta}}_k \xrightarrow[k\to\infty]{} 0$ *a.s..*

Proof. Using Theorem 17 in [11], we have $\boldsymbol{\Pi}^0(k,0) \xrightarrow[k\to\infty]{} 0$, which implies that $\boldsymbol{\Pi}(k,0) \xrightarrow[k\to\infty]{} 0$ by Theorem 4.9. Hence by Theorem 4.7, we can obtain $\hat{\boldsymbol{\Theta}}_k \to \boldsymbol{\Theta}$, $k\to\infty$ a.s., which completes the proof. ∎

5. Proofs of Lemma 4.6 and Theorem 4.9.

5.1. Proof of Lemma 4.6.

Proof. By (3.5)–(3.7) and the definition ζ_k^i in (3.11), we have

$$\begin{aligned} C(z)\zeta_k^i &= C(z)(y_k^i - (\hat{\theta}_{k-1}^i)^T\varphi_{k-1}^i - \omega_k^i) \\ &= y_k^i - C(z)\omega_k^i - (\hat{\theta}_{k-1}^i)^T\varphi_{k-1}^i + (C(z)-\boldsymbol{I}_{m_1})(y_k^i - (\hat{\theta}_{k-1}^i)^T\varphi_{k-1}^i) \\ &= \theta_0^T\phi_{k-1}^{0,i} - (\hat{\theta}_{k-1}^i)^T\varphi_{k-1}^i + C_1\hat{\omega}_{k-1}^i + \cdots + C_r\hat{\omega}_{k-r}^i \\ &= \theta^T\varphi_{k-1}^i - (\hat{\theta}_{k-1}^i)^T\varphi_{k-1}^i = (\widetilde{\theta}_{k-1}^i)^T\varphi_{k-1}^i. \end{aligned} \tag{5.1}$$

By the assumption that the transfer matrix $C(z) - \frac{\mu(1+4\nu)}{2}\boldsymbol{I}_{m_1}$ is SPR and Lemma 4.2, we see that there exists a positive constant $\bar{c} > 0$ such that

$$a_k^i \triangleq 2\mu \sum_{j=1}^{k} (\zeta_j^i)^T \left[(\widetilde{\theta}_{j-1}^i)^T \varphi_{j-1}^i - \frac{\mu(1+4\nu)(1+\bar{c})}{2} \zeta_j^i \right] \geq 0,\ i = 1, \ldots, n. \tag{5.2}$$

Let $b_k = \sum_{i=1}^{n} \frac{a_k^i}{r_{k-1}^i} \geq 0$. Since r_k^i is nondecreasing with k, we can obtain that

$$\begin{aligned} b_{k+1} - b_k &= \sum_{i=1}^{n} \left(\frac{a_{k+1}^i}{r_k^i} - \frac{a_k^i}{r_{k-1}^i} \right) \leq \sum_{i=1}^{n} \frac{a_{k+1}^i - a_k^i}{r_k^i} \\ &= 2\mu \sum_{i=1}^{n} \frac{1}{r_k^i} \left((\zeta_{k+1}^i)^T \left[(\widetilde{\theta}_k^i)^T \varphi_k^i - \frac{\mu(1+4\nu)(1+\bar{c})}{2} \zeta_{k+1}^i \right] \right) \\ &= 2\mu \mathrm{Tr} \left(\widetilde{\boldsymbol{\Theta}}_k^T \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{\zeta}_{k+1}^T - \frac{\mu(1+4\nu)(1+\bar{c})}{2} \boldsymbol{\zeta}_{k+1} \boldsymbol{R}_k^{-1} \boldsymbol{\zeta}_{k+1}^T \right). \end{aligned} \tag{5.3}$$

By (3.9) and (3.13), we have

$$\hat{\boldsymbol{\Theta}}_{k+1} = \hat{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} (\boldsymbol{\zeta}_{k+1}^T + \boldsymbol{W}_{k+1}^T) - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m) \mathscr{L} \hat{\boldsymbol{\Theta}}_k.$$

By $\widetilde{\boldsymbol{\Theta}}_{k+1} = \boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}_{k+1}$ and $\mathscr{L}\boldsymbol{\Theta} = 0$, we can derive the following recursive estimation error equation:

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = (\boldsymbol{I}_{mn} - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}) \widetilde{\boldsymbol{\Theta}}_k - \mu \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} (\boldsymbol{\zeta}_{k+1}^T + \boldsymbol{W}_{k+1}^T). \tag{5.4}$$

By (5.4), we have

$$\begin{aligned} &\mathrm{Tr}(\widetilde{\boldsymbol{\Theta}}_{k+1}^T \widetilde{\boldsymbol{\Theta}}_{k+1}) + b_{k+1} \\ &\quad = \mathrm{Tr}\Big(\widetilde{\boldsymbol{\Theta}}_k^T \big(\boldsymbol{I}_{mn} - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\big)^2 \widetilde{\boldsymbol{\Theta}}_k \\ &\qquad - 2\mu \widetilde{\boldsymbol{\Theta}}_k^T \big(\boldsymbol{I}_{mn} - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\big) \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} (\boldsymbol{\zeta}_{k+1}^T + \boldsymbol{W}_{k+1}^T) \\ &\qquad + \mu^2 (\boldsymbol{\zeta}_{k+1} + \boldsymbol{W}_{k+1}) \boldsymbol{R}_k^{-1} \boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} (\boldsymbol{\zeta}_{k+1}^T + \boldsymbol{W}_{k+1}^T) \Big) + b_{k+1} \\ &\quad = \mathrm{Tr}\Big(\widetilde{\boldsymbol{\Theta}}_k^T \big(\boldsymbol{I}_{mn} - \mu\nu \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\big)^2 \widetilde{\boldsymbol{\Theta}}_k \Big) \\ &\qquad - \mathrm{Tr}\Big(2\mu \left[\widetilde{\boldsymbol{\Theta}}_k^T \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} - \frac{\mu(1+4\nu)(1+\bar{c})}{2} \boldsymbol{\zeta}_{k+1} \boldsymbol{R}_k^{-1} \right] \boldsymbol{\zeta}_{k+1}^T \Big) \\ &\qquad - \mathrm{Tr}\Big(\mu^2 (1+4\nu)(1+\bar{c}) \boldsymbol{\zeta}_{k+1} \boldsymbol{R}_k^{-1} \boldsymbol{\zeta}_{k+1}^T \Big) - \mathrm{Tr}\Big(2\mu \widetilde{\boldsymbol{\Theta}}_k^T \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{W}_{k+1}^T \Big) \\ &\qquad + \mathrm{Tr}\Big(2\mu^2\nu \widetilde{\boldsymbol{\Theta}}_k^T \mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L} \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} (\boldsymbol{\zeta}_{k+1}^T + \boldsymbol{W}_{k+1}^T) \Big) \\ &\qquad + \mathrm{Tr}\Big(\mu^2 \boldsymbol{\zeta}_{k+1} \boldsymbol{R}_k^{-1} \boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{\zeta}_{k+1}^T \Big) + \mathrm{Tr}\Big(\mu^2 \boldsymbol{W}_{k+1} \boldsymbol{R}_k^{-1} \boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{W}_{k+1}^T \Big) \\ &\qquad + \mathrm{Tr}\Big(2\mu^2 \boldsymbol{\zeta}_{k+1} \boldsymbol{R}_k^{-1} \boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k \boldsymbol{R}_k^{-1} \boldsymbol{W}_{k+1}^T \Big) + b_{k+1} \\ &\quad \triangleq J_1 - J_2 - J_3 - J_4 + J_5 + J_6 + J_7 + J_8 + b_{k+1}. \end{aligned} \tag{5.5}$$

In the following, we analyze the right-hand side of (5.5) term by term.

By Lemma 2.1 and the definition of $\boldsymbol{X}_k(Q)$, we have

$$\|\mathscr{L}(\boldsymbol{X}_k(Q) \otimes \boldsymbol{I}_m)\mathscr{L}\| \leq 4 \max_{1 \leq i \leq n} \frac{\|\varphi_k^i\|^2}{r_k^i} \leq 4. \tag{5.6}$$

Since for any $A \geq 0$ and any X, Y with proper dimensions, the inequalities $2\mathrm{Tr}(X^T AY) \leq \mathrm{Tr}(X^T AX) + \mathrm{Tr}(Y^T AY)$ and $\mathrm{Tr}(X^T AX) \leq \|A\|\mathrm{Tr}(X^T X)$ hold. By these two inequalities and (5.6), we have

$$\begin{aligned} J_5 \leq\, & \mu^2\nu\,\mathrm{Tr}\Big(2\widetilde{\boldsymbol{\Theta}}_k^T \mathscr{L}(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m)\mathscr{L}\widetilde{\boldsymbol{\Theta}}_k \\ & + \boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T \mathscr{L}(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m)\mathscr{L}\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T \\ & + \boldsymbol{W}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T \mathscr{L}(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m)\mathscr{L}\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{W}_{k+1}^T\Big) \\ \leq\, & \mu^2\nu\,\mathrm{Tr}\Big(2\widetilde{\boldsymbol{\Theta}}_k^T \mathscr{L}(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m)\mathscr{L}\widetilde{\boldsymbol{\Theta}}_k \\ & + 4\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T + 4\boldsymbol{W}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{W}_{k+1}^T\Big). \end{aligned}$$

Then we can obtain the following inequality:

$$\begin{aligned} J_1 + J_5 + J_6 + J_7 \leq\, & \mathrm{Tr}\Big(\widetilde{\boldsymbol{\Theta}}_k^T\big[\big(\boldsymbol{I}_{mn} - \mu\nu\mathscr{L}(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m)\mathscr{L}\big)^2 \\ & + 2\mu^2\nu\mathscr{L}(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m)\mathscr{L}\big]\widetilde{\boldsymbol{\Theta}}_k \\ & + \mu^2(1+4\nu)\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T \\ & + \mu^2(1+4\nu)\boldsymbol{W}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{W}_{k+1}^T\Big). \end{aligned} \tag{5.7}$$

From $\mu(1+4\nu) \leq 1$ and (5.6), we have

$$\big(\boldsymbol{I}_{mn} - \mu\nu\mathscr{L}\left(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m\right)\mathscr{L}\big)^2 + 2\mu^2\nu\mathscr{L}(\boldsymbol{X}_k(Q)\otimes \boldsymbol{I}_m)\mathscr{L} \leq \boldsymbol{I}_{mn}.$$

Then

$$\begin{aligned} & \mathrm{Tr}\Big(\mu^2(1+4\nu)\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T\Big) - J_3 \\ & = \mathrm{Tr}\Big(-\mu^2(1+4\nu)\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-\frac{1}{2}}(\boldsymbol{I}_n - \boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-\frac{1}{2}})\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{\zeta}_{k+1}^T \\ & \qquad - \mu^2(1+4\nu)\bar{c}\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T\Big). \end{aligned} \tag{5.8}$$

On the other hand, it is clear that

$$\big\|\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-\frac{1}{2}}\big\| = \left\|\mathrm{diag}\left\{\frac{\|\boldsymbol{\varphi}_k^1\|^2}{r_k^1}, \ldots, \frac{\|\boldsymbol{\varphi}_k^n\|^2}{r_k^n}\right\}\right\| \leq 1. \tag{5.9}$$

Hence $\boldsymbol{I}_n - \boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-\frac{1}{2}} \geq 0$. Then according to (5.8), we have

$$\begin{aligned} & \mathrm{Tr}\Big(\mu^2(1+4\nu)\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T\Big) - J_3 \\ & \quad \leq -\mu^2(1+4\nu)\bar{c}\,\mathrm{Tr}\Big(\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T\Big). \end{aligned} \tag{5.10}$$

Combining (5.3), (5.5), (5.7) with (5.10), we obtain the following inequality:

$$\begin{aligned} & \mathrm{Tr}\left(\widetilde{\boldsymbol{\Theta}}_{k+1}^T\widetilde{\boldsymbol{\Theta}}_{k+1}\right) + b_{k+1} \\ & \leq \mathrm{Tr}\Big(\widetilde{\boldsymbol{\Theta}}_k^T\widetilde{\boldsymbol{\Theta}}_k - \mu^2(1+4\nu)\bar{c}\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\zeta}_{k+1}^T - 2\mu\widetilde{\boldsymbol{\Theta}}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{W}_{k+1}^T \\ & \quad + \mu^2(1+4\nu)\boldsymbol{W}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{W}_{k+1}^T \\ & \quad + 2\mu^2\boldsymbol{\zeta}_{k+1}\boldsymbol{R}_k^{-1}\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\boldsymbol{R}_k^{-1}\boldsymbol{W}_{k+1}^T\Big) + b_k. \end{aligned}$$

By the summation of both sides of the above inequality, it can be derived that

$$
\begin{aligned}
&\mathrm{Tr}\left(\widetilde{\mathbf{\Theta}}_{k+1}^T\widetilde{\mathbf{\Theta}}_{k+1}\right)+b_{k+1}+\mu^2(1+4\nu)\bar{c}\cdot\mathrm{Tr}\left(\sum_{j=0}^{k}\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\boldsymbol{\zeta}_{j+1}^T\right)\\
&\quad\le \mathrm{Tr}\left(\widetilde{\mathbf{\Theta}}_0^T\widetilde{\mathbf{\Theta}}_0\right)+b_0+\mu^2(1+4\nu)\mathrm{Tr}\left(\sum_{j=0}^{k}\boldsymbol{W}_{j+1}\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right)\\
&\qquad-2\mu\mathrm{Tr}\left(\sum_{j=0}^{k}\widetilde{\mathbf{\Theta}}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right)+2\mu^2\mathrm{Tr}\left(\sum_{j=0}^{k}\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right).
\end{aligned}
\tag{5.11}
$$

Now let us estimate the last three terms of (5.11). By Lemma 2.2 and Assumptions 4.2 and 4.3, we can choose a small constant $\delta_0\in(0,\frac{1}{2})$ such that

$$
\begin{aligned}
&\left|\mathrm{Tr}\left(\sum_{j=0}^{k}\widetilde{\mathbf{\Theta}}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right)\right|\le m_1\left\|\sum_{j=0}^{k}\widetilde{\mathbf{\Theta}}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right\|\\
&\quad=O\left(\left(\sum_{j=0}^{k}\left\|\widetilde{\mathbf{\Theta}}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\right\|^2\right)^{\frac{1}{2}+\delta_0}\right)=O\left(\left(\sum_{j=0}^{k}\left\|\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\right\|^2\right)^{\frac{1}{2}+\delta_0}\right),
\end{aligned}
\tag{5.12}
$$

where the last equality holds by (5.1).

Notice $\boldsymbol{\zeta}_{k+1}\in\mathscr{F}_k$ and $\left\|\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\right\|\le 1$. Applying Lemma 2.2 to the last term of (5.11) yields

$$
\begin{aligned}
&\left|\mathrm{Tr}\Big(\sum_{j=0}^{k}\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\Big)\right|\le O\left(\left\|\sum_{j=0}^{k}\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right\|\right)\\
&\quad=O\left(\left(\sum_{j=0}^{k}\left\|\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\right\|^2\right)^{\frac{1}{2}+\delta_0}\right)\\
&\quad=O\left(\left(\sum_{j=0}^{k}\left\|\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\right\|^2\left\|\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\right\|^2\right)^{\frac{1}{2}+\delta_0}\right)=O\left(\left(\sum_{j=0}^{k}\left\|\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\right\|^2\right)^{\frac{1}{2}+\delta_0}\right).
\end{aligned}
\tag{5.13}
$$

Finally we estimate $\mathrm{Tr}(\sum_{j=0}^{k}\boldsymbol{W}_{j+1}\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T)$. Note that

$$
\begin{aligned}
&\sum_{j=0}^{\infty}\mathbb{E}\left(\left\|\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\right\|^2\cdot\left|\|\boldsymbol{W}_{j+1}\|^2-\mathbb{E}\left(\|\boldsymbol{W}_{j+1}\|^2|\mathscr{F}_j\right)\right|\Big|\mathscr{F}_j\right)\le 2c_0\sum_{j=0}^{\infty}\left\|\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\right\|^2\\
&\quad\le 2c_0\sum_{j=0}^{\infty}\mathrm{Tr}\left(\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\right)=2c_0\sum_{j=0}^{\infty}\sum_{i=1}^{n}\frac{\|\boldsymbol{\varphi}_j^i\|^2}{(r_j^i)^2}<\infty\ \text{ a.s.},
\end{aligned}
$$

where Lemma 4.4 is used in the last inequality and c_0 is defined in Assumption 4.3. Then by Lemma 2.3, we obtain the almost sure convergence of

$$\sum_{j=0}^{k} \|\boldsymbol{\Phi}_j \boldsymbol{R}_j^{-1}\|^2 \left(\|\boldsymbol{W}_{j+1}\|^2 - \mathbb{E}\left(\|\boldsymbol{W}_{j+1}\|^2 | \mathscr{F}_j\right)\right).$$

Hence, we have

$$\begin{aligned}
&\left|\mathrm{Tr}\left(\sum_{j=0}^{k} \boldsymbol{W}_{j+1}\boldsymbol{R}_j^{-1}\boldsymbol{\Phi}_j^T\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right)\right| \le m_1\left(\sum_{j=0}^{k}\left\|\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\boldsymbol{W}_{j+1}^T\right\|^2\right)\\
&\quad\le m_1\sum_{j=0}^{k}\left\|\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\right\|^2\left(\|\boldsymbol{W}_{j+1}\|^2 - \mathbb{E}\left(\|\boldsymbol{W}_{j+1}\|^2 \,|\mathscr{F}_j\right)\right)\\
&\qquad + m_1\sum_{j=0}^{k}\|\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\|^2\mathbb{E}\left(\|\boldsymbol{W}_{j+1}\|^2\,|\mathscr{F}_j\right)\\
&\quad\le m_1\left(\sum_{j=0}^{k}\|\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\|^2\left(\|\boldsymbol{W}_{j+1}\|^2 - \mathbb{E}\left(\|\boldsymbol{W}_{j+1}\|^2\,|\mathscr{F}_j\right)\right) + c_0\sum_{j=0}^{\infty}\|\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-1}\|^2\right) < \infty.
\end{aligned}$$

By (5.11), (5.12), (5.13), we see that there exists constant c' such that for any instant k, the inequality

$$\begin{aligned}
&\mathrm{Tr}\left(\widetilde{\boldsymbol{\Theta}}_{k+1}^T\widetilde{\boldsymbol{\Theta}}_{k+1}\right) + b_{k+1} + \mu^2(1+4\nu)\bar{c}\,\mathrm{Tr}\left(\sum_{j=0}^{k}\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\boldsymbol{\zeta}_{j+1}^T\right)\\
&\quad\le c' + O\left(\left(\sum_{j=0}^{k}\|\boldsymbol{\zeta}_{j+1}\boldsymbol{R}_j^{-1}\|^2\right)^{\frac{1}{2}+\delta_0}\right)
\end{aligned} \tag{5.14}$$

holds. By $\delta_0 < \frac{1}{2}$ and $b_k \ge 0$ for all $k \ge 0$, we complete the proof of this lemma by (5.14). ∎

5.2. Proof of Theorem 4.9.

Proof. We first prove the necessity part of the theorem. By Lemma 23 (2) in [11], we can see that $\boldsymbol{\Pi}(k,0) \xrightarrow[k\to\infty]{} 0$ implies $\|\boldsymbol{R}_k\| \xrightarrow[k\to\infty]{} \infty$. By (4.12) and $\|\boldsymbol{R}_k\| \xrightarrow[k\to\infty]{} \infty$, we have $\|\boldsymbol{R}_k^0\| \xrightarrow[k\to\infty]{} \infty$. Then by the proof of Lemma 4.8, we have for any i, $r_k^i \xrightarrow[k\to\infty]{} \infty$. Thus from (4.11), it is clear that

$$\limsup_{k\to\infty}\left\|\boldsymbol{R}_k(\boldsymbol{R}_k^0)^{-1}\right\| \le \lim_{k\to\infty}\mathrm{Tr}\left(\boldsymbol{R}_k\left(\boldsymbol{R}_k^0\right)^{-1}\right) = \lim_{k\to\infty}\sum_{i=1}^{n}\frac{r_k^i}{r_k^{0,i}} = n \tag{5.15}$$

and, similarly,

$$\limsup_{k\to\infty}\|\boldsymbol{R}_k^{-1}(\boldsymbol{R}_k^0)\| \le n. \tag{5.16}$$

Then by (4.5) and (5.15), we have

$$\sum_{k=0}^{\infty}\|\mathbf{\Psi}_k^{\zeta}(\boldsymbol{R}_k^0)^{-\frac{1}{2}}\|^2=\sum_{k=0}^{\infty}\|\mathbf{\Psi}_k^{\zeta}\boldsymbol{R}_k^{-\frac{1}{2}}\boldsymbol{R}_k^{\frac{1}{2}}(\boldsymbol{R}_k^0)^{-\frac{1}{2}}\|^2<\infty. \tag{5.17}$$

By $\mathbf{\Pi}^0(k+1,0)=(\boldsymbol{I}_{mn}-\mu\boldsymbol{G}_k)\mathbf{\Pi}^0(k,0)+(\mu\boldsymbol{G}_k-\mu\boldsymbol{G}_k^0)\mathbf{\Pi}^0(k,0)$, we can obtain the following equation:

$$\begin{aligned}
\mathbf{\Pi}^0(k+1,0)&=\mathbf{\Pi}(k+1,0)+\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\left(\mu\boldsymbol{G}_j-\mu\boldsymbol{G}_j^0\right)\mathbf{\Pi}^0(j,0)\\
&=\mathbf{\Pi}(k+1,0)+\mu\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\Big(\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^T-\mathbf{\Phi}_j^0(\boldsymbol{R}_j^0)^{-1}\mathbf{\Phi}_j^{0^T}\Big)\mathbf{\Pi}^0(j,0)\\
&\quad+\mu\nu\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\mathscr{L}\big(\big[\boldsymbol{X}_j(Q)-\boldsymbol{X}_j^0(Q)\big]\otimes\boldsymbol{I}_m\big)\mathscr{L}\mathbf{\Pi}^0(j,0)\\
&=\mathbf{\Pi}(k+1,0)+\mu\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\mathbf{\Psi}_j^{\zeta^T}\mathbf{\Pi}^0(j,0)\\
&\quad-\mu\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\mathbf{\Psi}_j^{\zeta}(\boldsymbol{R}_j^0)^{-1}\mathbf{\Phi}_j^{0^T}\mathbf{\Pi}^0(j,0)\\
&\quad+\mu\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\left(\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\mathbf{\Phi}_j^{0^T}-\mathbf{\Phi}_j(\boldsymbol{R}_j^0)^{-1}\mathbf{\Phi}_j^{0^T}\right)\mathbf{\Pi}^0(j,0)\\
&\quad+\mu\nu\sum_{j=0}^{k}\mathbf{\Pi}(k+1,j+1)\mathscr{L}\left(\big[\boldsymbol{X}_j(Q)-\boldsymbol{X}_j^0(Q)\big]\otimes\boldsymbol{I}_m\right)\mathscr{L}\mathbf{\Pi}^0(j,0)\\
&\triangleq\mathbf{\Pi}(k+1,0)+\mu H_{k,1}-\mu H_{k,2}+\mu H_{k,3}+\mu\nu H_{k,4}.
\end{aligned}$$

In the following we prove that all the terms $H_{k,1}$, $H_{k,2}$,$H_{k,3}$, $H_{k,4}$ converge to zero as $k\to\infty$. From (4.5), (4.7), $\|\mathbf{\Pi}^0(k,j)\|\le 1$, and $\mathbf{\Pi}(k,0)\xrightarrow[k\to\infty]{}0$, we have

$$\begin{aligned}
\|H_{k,1}\|\le&\left\|\sum_{j=0}^{M}\mathbf{\Pi}(k+1,j+1)\mathbf{\Phi}_j\boldsymbol{R}_j^{-1}\mathbf{\Psi}_j^{\zeta^T}\mathbf{\Pi}^0(j,0)\right\|\\
&+\left(\sum_{j=M+1}^{k}\left\|\mathbf{\Pi}(k,j+1)\mathbf{\Phi}_j\boldsymbol{R}_j^{-\frac{1}{2}}\right\|^2\right)^{\frac{1}{2}}\left(\sum_{j=M+1}^{k}\left\|\boldsymbol{R}_j^{-\frac{1}{2}}\mathbf{\Psi}_j^{\zeta^T}\right\|^2\right)^{\frac{1}{2}}\\
\to&\,0,\quad\text{as } k\to\infty \text{ and then } M\to\infty.
\end{aligned}$$

For $H_{k,2}$, by $\|\mathbf{\Pi}^0(k,j)\|\le 1$ we have for any $M>0$

$$\begin{aligned}
mn &> \mathrm{Tr}\left[\boldsymbol{\Pi}^{0^T}(M,0)\boldsymbol{\Pi}^0(M,0)\right] \\
&\geq \sum_{j=M}^{\infty} \mathrm{Tr}\left[\boldsymbol{\Pi}^{0^T}(j,0)\boldsymbol{\Pi}^0(j,0) - \boldsymbol{\Pi}^{0^T}(j+1,0)\boldsymbol{\Pi}^0(j+1,0)\right] \\
&= \sum_{j=M}^{\infty} \mathrm{Tr}\left[\boldsymbol{\Pi}^{0^T}(j,0)\left(\boldsymbol{I}_{mn} - \boldsymbol{\Pi}^{0^T}(j+1,j)\boldsymbol{\Pi}^0(j+1,j)\right)\boldsymbol{\Pi}^0(j,0)\right] \\
&= \sum_{j=M}^{\infty} \mathrm{Tr}\left[\boldsymbol{\Pi}^{0^T}(j,0)\left(\boldsymbol{I}_{mn} - (\boldsymbol{I} - \mu\boldsymbol{G}_j^0)(\boldsymbol{I}_{mn} - \mu\boldsymbol{G}_j^0)\right)\boldsymbol{\Pi}^0(j,0)\right] \\
&= \sum_{j=M}^{\infty} \mathrm{Tr}\left[\boldsymbol{\Pi}^{0^T}(j,0)\left(\mu\boldsymbol{G}_j^0 + \mu\boldsymbol{G}_j^0(\boldsymbol{I} - \mu\boldsymbol{G}_j^0)\right)\boldsymbol{\Pi}^0(j,0)\right] \\
&\geq \sum_{j=M}^{\infty} \mathrm{Tr}\left[\boldsymbol{\Pi}^{0^T}(j,0)\left(\mu\boldsymbol{G}_j^0\right)\boldsymbol{\Pi}^0(j,0)\right] \qquad (5.18)\\
&\geq \mu \sum_{j=M}^{\infty} \left\|\boldsymbol{\Pi}^{0^T}(j,0)\boldsymbol{\Phi}_j^0(\boldsymbol{R}_j^0)^{-\frac{1}{2}}\right\|^2. \qquad (5.19)
\end{aligned}$$

Then by (5.17), (5.19), $\|\boldsymbol{\Pi}(k,j)\| \leq 1$, and $\boldsymbol{\Pi}(k,0) \xrightarrow[k\to\infty]{} 0$, we can further obtain that

$$\begin{aligned}
\|H_{k,2}\| \leq& \left\|\sum_{j=0}^{M} \boldsymbol{\Pi}(k+1,j+1)\boldsymbol{\Psi}_j^{\zeta}(\boldsymbol{R}_j^0)^{-1}\boldsymbol{\Phi}_j^{0^T}\boldsymbol{\Pi}^0(j,0)\right\| \\
&+ \left(\sum_{j=M+1}^{k}\left\|\boldsymbol{\Psi}_k^{\zeta}(\boldsymbol{R}_k^0)^{-\frac{1}{2}}\right\|^2\right)^{\frac{1}{2}} \left(\sum_{j=M+1}^{k}\left\|\boldsymbol{\Pi}^{0^T}(j,0)\boldsymbol{\Phi}_j^0(\boldsymbol{R}_j^0)^{-\frac{1}{2}}\right\|^2\right)^{\frac{1}{2}} \\
&\longrightarrow 0, \quad \text{as } k\to\infty \text{ and then } M\to\infty.
\end{aligned}$$

From (4.7) and (5.19), it is clear that

$$\begin{aligned}
\|H_{k,3}\| =& \left\|\sum_{j=0}^{k} \boldsymbol{\Pi}(k+1,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-\frac{1}{2}}\left(\boldsymbol{R}_j^{-\frac{1}{2}}\left(\boldsymbol{R}_j^0\right)^{\frac{1}{2}} - \boldsymbol{R}_j^{\frac{1}{2}}\left(\boldsymbol{R}_j^0\right)^{-\frac{1}{2}}\right)\left(\boldsymbol{R}_j^0\right)^{-\frac{1}{2}}\boldsymbol{\Phi}_j^{0^T}\boldsymbol{\Pi}^0(j,0)\right\| \\
=& \left\|\sum_{j=0}^{M} \boldsymbol{\Pi}(k+1,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-\frac{1}{2}}\left(\boldsymbol{R}_j^{-\frac{1}{2}}\left(\boldsymbol{R}_j^0\right)^{\frac{1}{2}} - \boldsymbol{R}_j^{\frac{1}{2}}\left(\boldsymbol{R}_j^0\right)^{-\frac{1}{2}}\right)\left(\boldsymbol{R}_j^0\right)^{-\frac{1}{2}}\boldsymbol{\Phi}_j^{0^T}\boldsymbol{\Pi}^0(j,0)\right\| \\
&+ 2\sqrt{n}\left(\sum_{j=M+1}^{k}\left\|\boldsymbol{\Pi}(k+1,j+1)\boldsymbol{\Phi}_j\boldsymbol{R}_j^{-\frac{1}{2}}\right\|^2\right)^{\frac{1}{2}} \\
&\times \left(\sum_{j=M+1}^{k}\left\|\boldsymbol{\Pi}^{0^T}(j,0)\boldsymbol{\Phi}_j^0(\boldsymbol{R}_j^0)^{-\frac{1}{2}}\right\|^2\right)^{\frac{1}{2}} \longrightarrow 0, \quad \text{as } k\to\infty \text{ and then } M\to\infty,
\end{aligned}$$

where (5.15) and (5.16) are used in the above inequality.

Finally, we will show $H_{k,4} \xrightarrow[k\to\infty]{} 0$. By Step 1 in Algorithm 3.1, it is clear that

$$x_j^i(Q) = \sum_{l=1}^{n} a_{il}^{(Q)} \frac{\|\boldsymbol{\varphi}_j^l\|^2}{r_j^l}. \tag{5.20}$$

By (3.10), we have

$$\begin{aligned}
&a_{il}^{(Q)}\left(\frac{\|\varphi_j^l\|^2}{r_j^l}-\frac{\|\varphi_j^{0,l}\|^2}{r_j^{0,l}}\right)\\
&\quad = a_{il}^{(Q)}\left(\frac{{\varphi_j^l}^T\psi_j^{\zeta,l}}{r_j^l}+\frac{{\psi_j^{\zeta,l}}^T\varphi_j^{0,l}}{r_j^{0,l}}+\frac{{\varphi_j^l}^T}{\sqrt{r_j^l}}\left(\sqrt{\frac{r_j^{0,l}}{r_j^l}}-\sqrt{\frac{r_j^l}{r_j^{0,l}}}\right)\frac{\varphi_j^{0,l}}{\sqrt{r_j^{0,l}}}\right)\\
&\quad \triangleq a_{il}^{(Q)}P_{l1,j}+a_{il}^{(Q)}P_{l2,j}+a_{il}^{(Q)}P_{l3,j}.
\end{aligned}$$

Then by (5.20), we have

$$\begin{aligned}
&\boldsymbol{X}_j(Q)-\boldsymbol{X}_j^0(Q)\\
&\quad=\sum_{l=1}^n \operatorname{diag}\left\{a_{1l}^{(Q)}\left(\frac{\|\varphi_j^l\|^2}{r_j^l}-\frac{\|\varphi_j^{0,l}\|^2}{r_j^{0,l}}\right),\dots,a_{nl}^{(Q)}\left(\frac{\|\varphi_j^l\|^2}{r_j^l}-\frac{\|\varphi_j^{0,l}\|^2}{r_j^{0,l}}\right)\right\}\\
&\quad=\sum_{l=1}^n \operatorname{diag}\left\{a_{1l}^{(Q)}P_{l1,j}+a_{1l}^{(Q)}P_{l2,j}+a_{1l}^{(Q)}P_{l3,j},\dots,a_{nl}^{(Q)}P_{l1,j}+a_{nl}^{(Q)}P_{l2,j}+a_{nl}^{(Q)}P_{l3,j}\right\}.
\end{aligned}$$

Hence

$$\begin{aligned}
H_{k,4}=&\sum_{l=1}^n\sum_{j=0}^k \boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\operatorname{diag}\left\{a_{1l}^{(Q)}P_{l1,j},\dots,a_{nl}^{(Q)}P_{l1,j}\right\}\right]\otimes\boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\\
&+\sum_{l=1}^n\sum_{j=0}^k \boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\operatorname{diag}\left\{a_{1l}^{(Q)}P_{l2,j},\dots,a_{nl}^{(Q)}P_{l2,j}\right\}\right]\otimes\boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\\
&+\sum_{l=1}^n\sum_{j=0}^k \boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\operatorname{diag}\left\{a_{1l}^{(Q)}P_{l3,j},\dots,a_{nl}^{(Q)}P_{l3,j}\right\}\right]\otimes\boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\\
\triangleq& H_{k,41}+H_{k,42}+H_{k,43}.
\end{aligned}$$

We will estimate $H_{k,41}$, $H_{k,42}$, $H_{k,43}$, respectively. For $H_{k,41}$, by the definition of $P_{l1,j}$, we have

$$\begin{aligned}
&\|H_{k,41}\|\\
&\quad\le\sum_{l=1}^n\sum_{j=0}^M\left\|\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\operatorname{diag}\left\{a_{1l}^{(Q)}P_{l1,j},\dots,a_{nl}^{(Q)}P_{l1,j}\right\}\right]\otimes\boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|\\
&\qquad+\left(\sum_{l=1}^n\sum_{j=M+1}^k\left\|\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left[\operatorname{diag}\left\{\sqrt{\frac{a_{1l}^{(Q)}}{r_j^l}}{\varphi_j^l}^T,\dots,\sqrt{\frac{a_{nl}^{(Q)}}{r_j^l}}{\varphi_j^l}^T\right\}\otimes\boldsymbol{I}_m\right]\right\|^2\right)^{\frac{1}{2}}\\
&\qquad\cdot\left(\sum_{l=1}^n\sum_{j=M+1}^k\left\|\left[\operatorname{diag}\left\{\sqrt{\frac{a_{1l}^{(Q)}}{r_j^l}}\psi_j^{\zeta,l},\dots,\sqrt{\frac{a_{nl}^{(Q)}}{r_j^l}}\psi_j^{\zeta,l}\right\}\otimes\boldsymbol{I}_m\right]\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|^2\right)^{\frac{1}{2}}
\end{aligned}$$

$$\begin{aligned}&\le \sum_{l=1}^{n}\sum_{j=0}^{M}\left\|\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\mathrm{diag}\left\{a_{1l}^{(Q)}P_{l1,j},\dots,a_{nl}^{(Q)}P_{l1,j}\right\}\right]\otimes \boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|\\&\quad+\left(\sum_{j=M+1}^{k}\mathrm{Tr}\left(\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left[\boldsymbol{X}_j(Q)\otimes \boldsymbol{I}_m\right]\mathscr{L}\boldsymbol{\Pi}^T(k+1,j+1)\right)\right)^{\frac{1}{2}}\\&\quad\cdot 2\left(\sum_{l=1}^{n}\sum_{j=M+1}^{k}\frac{\|\psi_j^{\zeta,l}\|^2}{r_j^l}\right)^{\frac{1}{2}}.\end{aligned}$$

Similarly to (5.18), we have

$$\sum_{j=M}^{\infty}\mathrm{Tr}\left(\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left[\boldsymbol{X}_j(Q)\otimes \boldsymbol{I}_m\right]\mathscr{L}\boldsymbol{\Pi}^T(k+1,j+1)\right)<\infty. \tag{5.21}$$

Then by (4.8), we obtain $\|H_{k,41}\|\to 0$ as $k\to\infty$ and then $M\to\infty$.

For $H_{k,42}$, by the definition of $P_{l2,j}$, we have

$$\begin{aligned}&\|H_{k,42}\|\\&\le \sum_{l=1}^{n}\sum_{j=0}^{M}\left\|\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}([\mathrm{diag}\left\{a_{1l}^{(Q)}P_{l2,j},\dots,a_{nl}^{(Q)}P_{l2,j}\right\}]\otimes \boldsymbol{I}_m)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|\\&\quad+2\left(\sum_{l=1}^{n}\sum_{j=M+1}^{k}\left\|\left[\mathrm{diag}\left\{\sqrt{\frac{a_{1l}^{(Q)}}{r_j^{0,l}}}\psi_j^{\zeta,l^T},\dots,\sqrt{\frac{a_{nl}^{(Q)}}{r_j^{0,l}}}\psi_j^{\zeta,l^T}\right\}\otimes \boldsymbol{I}_m\right]\right\|^2\right)^{\frac{1}{2}}\\&\quad\cdot\left(\sum_{l=1}^{n}\sum_{j=M+1}^{k}\left\|\left[\mathrm{diag}\left\{\sqrt{\frac{a_{1l}^{(Q)}}{r_j^{0,l}}}\varphi_j^{l,0},\dots,\sqrt{\frac{a_{nl}^{(Q)}}{r_j^{0,l}}}\varphi_j^{l,0}\right\}\otimes \boldsymbol{I}_m\right]\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|^2\right)^{\frac{1}{2}}\\&\le \sum_{l=1}^{n}\sum_{j=0}^{M}\left\|\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\mathrm{diag}\left\{a_{1l}^{(Q)}P_{l2,j},\dots,a_{nl}^{(Q)}P_{l2,j}\right\}\right]\otimes \boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|\\&\quad+2\left(\sum_{l=1}^{n}\sum_{j=M+1}^{k}\frac{\|\psi_j^{\zeta,l}\|^2}{r_j^{0,l}}\right)^{\frac{1}{2}}\left(\sum_{j=M+1}^{k}\mathrm{Tr}\left(\boldsymbol{\Pi}^{0^T}(j,0)\mathscr{L}[\boldsymbol{X}_j(Q)\otimes\boldsymbol{I}_m]\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right)\right)^{\frac{1}{2}}.\end{aligned}$$

Combining (5.17) with (5.18), we obtain $\|H_{k,42}\|\to 0$ as $k\to\infty$ and then $M\to\infty$.

For $H_{k,43}$, by (5.15)–(5.16) and the definition of $P_{l3,j}$, we have

$$\begin{aligned}&\|H_{k,43}\|\\&\quad=\sum_{l=1}^{n}\sum_{j=0}^{k}\left\|\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\mathrm{diag}\left\{a_{1l}^{(Q)}P_{l3,j},\dots,a_{nl}^{(Q)}P_{l3,j}\right\}\right]\otimes\boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|\\&\quad\le\sum_{l=1}^{n}\sum_{j=0}^{M}\left\|\boldsymbol{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\mathrm{diag}\left\{a_{1l}^{(Q)}P_{l3,j},\dots,a_{nl}^{(Q)}P_{l3,j}\right\}\right]\otimes\boldsymbol{I}_m\right)\mathscr{L}\boldsymbol{\Pi}^0(j,0)\right\|\end{aligned}$$

$$+2\sqrt{n}\left(\sum_{l=1}^{n}\sum_{j=M+1}^{k}\left\|\mathbf{\Pi}(k+1,j+1)\mathscr{L}\left[\operatorname{diag}\left\{\sqrt{\frac{a_{1l}^{(Q)}}{r_j^l}}\varphi_j^{l\,T},\ldots,\sqrt{\frac{a_{nl}^{(Q)}}{r_j^l}}\varphi_j^{l\,T}\right\}\otimes \boldsymbol{I}_m\right]\right\|^2\right)^{\frac{1}{2}}$$

$$\cdot\left(\sum_{l=1}^{n}\sum_{j=M+1}^{k}\left\|\left[\operatorname{diag}\left\{\sqrt{\frac{a_{1l}^{(Q)}}{r_j^{0,l}}}\varphi_j^{0,l},\ldots,\sqrt{\frac{a_{nl}^{(Q)}}{r_j^{0,l}}}\varphi_j^{0,l}\right\}\otimes \boldsymbol{I}_m\right]\mathscr{L}\mathbf{\Pi}^0(j,0)\right\|^2\right)^{\frac{1}{2}}$$

$$\leq \sum_{l=1}^{n}\sum_{j=0}^{M}\left\|\mathbf{\Pi}(k+1,j+1)\mathscr{L}\left(\left[\operatorname{diag}\left\{a_{1l}^{(Q)}P_{l3,j},\ldots,a_{nl}^{(Q)}P_{l3,j}\right\}\right]\otimes \boldsymbol{I}_m\right)\mathscr{L}\mathbf{\Pi}^0(j,0)\right\|$$

$$+2\sqrt{n}\left(\sum_{j=M+1}^{k}\operatorname{Tr}\left(\mathbf{\Pi}(k+1,j+1)\mathscr{L}\left[\boldsymbol{X}_j(Q)\otimes \boldsymbol{I}_m\right]\mathscr{L}\mathbf{\Pi}^T(k+1,j+1)\right)\right)^{\frac{1}{2}}$$

$$\cdot\left(\sum_{j=M+1}^{k}\operatorname{Tr}\left(\mathbf{\Pi}^{0^T}(j,0)\mathscr{L}[\boldsymbol{X}_j^0(Q)\otimes \boldsymbol{I}_m]\mathscr{L}\mathbf{\Pi}^0(j,0)\right)\right)^{\frac{1}{2}}.$$

Hence by (5.18) and (5.21), we have $\|H_{k,43}\|\to 0$ as $k\to\infty$ and then $M\to\infty$. In summary, we show that $\mathbf{\Pi}(k,0)\to 0$, $k\to\infty$ implies $\mathbf{\Pi}^0(k,0)\to 0$, $k\to\infty$.

Let us move on the sufficiency. Conversely, if $\mathbf{\Pi}^0(k,0)\to 0$, then a similar argument leads to $\|\boldsymbol{R}_k^0\|\to\infty$. From the proof of the Lemma 4.8, we see that $r_k^{0,i}\to\infty$ is satisfied for any $i\in\{1,\ldots,n\}$. Hence for any $i\in\{1,\ldots,n\}$, we have $r_k^i\to\infty$. Thus, (5.15)–(5.17) still hold. Then by the expression

$$\mathbf{\Pi}(k+1,0)=\mathbf{\Pi}^0(k+1,0)+\sum_{j=0}^{k}\mathbf{\Pi}^0(k+1,j+1)(\mu\boldsymbol{G}_j^0-\mu\boldsymbol{G}_j)\mathbf{\Pi}(j,0),$$

we can prove $\mathbf{\Pi}(k,0)\to 0$, $k\to\infty$ by a similar argument as that discussed above. This completes the proof of this theorem. ∎

6. Simulation results. In this section, we provide an example to illustrate the performance of the distributed extended SG algorithm proposed in this paper.

Consider a network composed of $n=12$ sensors whose dynamics obey the following observation model:

$$y_{k+1}^i=\theta_0^T\phi_k^{0,i}+\omega_{k+1}^i+0.5\,\omega_k^i=\theta^T\varphi_k^{0,i}+\omega_{k+1}^i, \tag{6.1}$$

where $\theta^T=[\,\theta_0^T,0.5\boldsymbol{I}_3]\in\mathbb{R}^{3\times 6}$ and $\varphi_k^{0,i}=[(\phi_k^{0,i})^T,(\omega_k^i)^T]^T\in\mathbb{R}^6$. Here we assume that the unknown parameter matrix

$$\theta_0=\begin{bmatrix}-1 & 0 & 1\\ 0 & 1 & 2\\ 1 & 2 & 3\end{bmatrix}.$$

Let the regression vectors $\{\phi_k^{0,i}\in\mathbb{R}^3, k\geq 0, i=1,\ldots,12\}$ be generated according to the following expression:

$$\phi_k^{0,i}=\left[0,1.2^k+\sum_{t=0}^{k-1}1.2^t\cos\left(\frac{i\pi}{n}\right)\varepsilon_t^i,0\right]^T,\ i=1,4,7,10,$$

$$\phi_k^{0,i} = \left[0, 0, 1.2^k + \sum_{t=0}^{k-1} 1.2^t \cos\left(\frac{i\pi}{n}\right)\varepsilon_t^i\right]^T, \quad i = 2, 5, 8, 11,$$

$$\phi_k^{0,i} = \left[1.2^k + \sum_{t=0}^{k-1} 1.2^t \cos\left(\frac{i\pi}{n}\right)\varepsilon_t^i, 0, 0\right]^T, \quad i = 3, 6, 9, 12,$$

where the $\{\varepsilon_k^i\}$ are i.i.d. with $\varepsilon_k^i \sim \mathcal{N}(0, 0.16)$ (Gaussian distribution with zero mean and variance 0.16). Moreover, by Remark 4.4, we set the noise $\omega_{k+1}^i = U_{k+1}^i \phi_k^{0,i}$ with $\{U_{k+1}^i, k \geq 0, i = 1, \ldots, 12\}$ being independent and uniformly distributed in $[-0.5, 0.5]$, and also independent of $\mathscr{F}_k$.

The network structure is shown in Figure 1. Here we use the Metropolis rule to construct the weights. i.e.,

$$a_{li} = \begin{cases} 1 - \sum_{j \neq i} a_{ij} & \text{if } l = i, \\ 1/(\max\{n_i, n_l\}) & \text{if } l \in \mathcal{N}_i \setminus \{i\}, \end{cases}$$

where n_i is the degree of the node i. The initial estimate is taken as $\hat{\theta}_0^i = \mathbf{0} \in \mathbb{R}^{6\times 3}$, and the step sizes are taken as $\mu = 0.25$ and $\nu = 0.6$.

From the structure of the network topology in Figure 1 and the observation model, it is clear that Assumptions 4.1 and 4.2 hold. By the settings of the regression vector $\phi_k^{0,i}$ and the noise ω_{k+1}^i, we can verify that for each sensor i the extended regression vector $\varphi_k^{0,i}$ lacks sufficient excitation. However, they can cooperate to satisfy Assumption 4.5. We repeat the simulation for $s = 100$ times with the same initial states. Thus, for each sensor i, we can get the following 100 sequences:

$$\left\{\mathrm{Tr}\left(\widetilde{\theta}_{k,p}^{i,T}\widetilde{\theta}_{k,p}^i\right), k = 1, \ldots, 1000\right\}, \; i = 1, \ldots, 12, \; p = 1, \ldots, 100,$$

where p denotes the pth simulation result. Then

$$\frac{1}{100}\sum_{p=1}^{100} \mathrm{Tr}\left(\widetilde{\theta}_{k,p}^{i,T}\widetilde{\theta}_{k,p}^i\right), \; k = 1, \ldots, 1000, \; i = 1, \ldots, 12,$$

is used to approximate the estimation error of sensor i in Figure 2, which shows the estimation errors of the 12 sensors by using the noncooperative SG algorithm in [6] and the proposed distributed SG algorithm with colored noise. From Figure 2, we

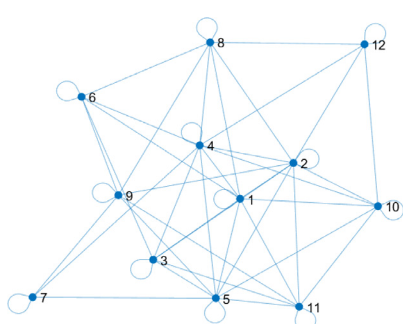

FIG. 1. *Network topology of 12 sensors.*

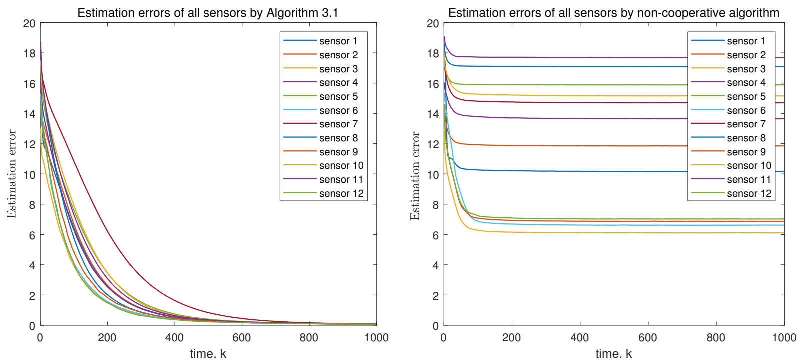

FIG. 2. *Comparison between Algorithm* 3.1 *and noncooperative extended SG algorithm.*

can see that if we use the noncooperative algorithm to estimate θ, the errors of all 12 sensors will not tend to zero since all sensors do not satisfy the excitation condition in [6], while the estimation errors of all 12 sensors in Algorithm 3.1 converge to zero because all sensors cooperatively satisfy Assumption 4.5, which shows the cooperative effect of multiple sensors that the estimation task can be fulfilled through exchanging information between sensors even if any individual sensor cannot.

7. Concluding remarks. This paper proposed a distributed extended SG algorithm to estimate unknown parameter matrices of dynamic stochastic systems with colored noise. The extended regression vectors are introduced by integrating the observed signals and estimates of noises. Then the algorithm is developed by combining the diffusion strategy of extended regression vectors with the consensus strategy of neighbors' estimates. Under the cooperative non-PE conditions on regressors, the almost sure convergence of the proposed algorithm was established. The convergence results for the distributed algorithm are obtained without resorting to the independence or stationarity conditions of stochastic regression vectors, which makes our theory applicable to stochastic feedback systems. By a simulation example, we reveal the cooperative effect of multiple sensors in accomplishing the task. For further research, it is of interest to optimize the diffusion process between sensors and analyze the convergence rate of the distributed extended SG algorithm Moreover, how to establish the performance analysis of distributed algorithms for the dynamic systems with complex noise models, e.g., the noise model with infinite unknown parameters, the time-varying colored noise model, is another interesting research topic.

REFERENCES

[1] R. Abdolee, V. Vakilian, and B. Champagne, *Tracking performance and optimal adaptation step-sizes of diffusion-LMS networks*, IEEE Trans. Control Netw. Syst., 5 (2018), pp. 67–78, https://doi.org/10.1109/TCNS.2016.2578044.

[2] F. Barani, A. Savadi, and H. S. Yazdi, *Convergence behavior of diffusion stochastic gradient descent algorithm*, Signal Process., 183 (2021), 108014, https://doi.org/10.1016/j.sigpro.2021.108014.

[3] G. Battistelli and L. Chisci, *Stability of consensus extended Kalman filter for distributed state estimation*, Automatica J. IFAC, 68 (2016), pp. 169–178, https://doi.org/10.1016/j.automatica.2016.01.071.

[4] A. Bertrand, M. Moonen, and A. H. Sayed, *Diffusion bias-compensated RLS estimation over adaptive networks*, IEEE Trans. Signal Process., 59 (2011), pp. 5212–5224, https://doi.org/10.1109/TSP.2011.2163631.

[5] A. S. Bijral, A. D. Sarwate, and N. Srebro, *Data-dependent convergence for consensus stochastic optimization*, IEEE Trans. Automat. Control, 62 (2017), pp. 4483–4498, https://doi.org/10.1109/TAC.2017.2671377.

[6] H.-F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Birkhäuser, Boston, MA, 1991.

[7] W. Chen, C. Wen, S. Hua, and C. Sun, *Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative PE condition via consensus*, IEEE Trans. Automat. Control, 59 (2014), pp. 91–106, https://doi.org/10.1109/TAC.2013.2278135.

[8] Y. Fu and W. Zhao, *Support recovery and parameter identification of multivariate ARMA systems with exogenous inputs*, SIAM J. Control Optim., 61 (2023), pp. 1835–1860, https://doi.org/10.1137/21M146346X.

[9] D. Gan and Z. Liu, *Distributed order estimation of ARX model under cooperative excitation condition*, SIAM J. Control Optim., 60 (2022), pp. 1519–1545, https://doi.org/10.1137/21M1421362.

[10] D. Gan and Z. Liu, *Distributed sparse identification for stochastic dynamic systems under cooperative non-persistent excitation condition*, Automatica J. IFAC, 151 (2023), 110958, https://doi.org/10.1016/j.automatica.2023.110958.

[11] D. Gan and Z. Liu, *Convergence of the distributed SG algorithm under cooperative excitation condition*, IEEE Trans. Neural Netw. Learn. Syst., 35 (2024), pp. 7087–7101, https://doi.org/10.1109/TNNLS.2022.3213715.

[12] D. Gan, S. Xie, and Z. Liu, *Stability of the distributed Kalman filter using general random coefficients*, Sci. China Inf. Sci., 64 (2021), 172204, https://doi.org/10.1007/s11432-020-2962-9.

[13] J. Gao, X.-W. Liu, Y.-H. Dai, Y. Huang, and J. Gu, *Distributed stochastic gradient tracking methods with momentum acceleration for non-convex optimization*, Comput. Optim. Appl., 84 (2022), pp. 531–572, https://doi.org/10.1007/s10589-022-00432-5.

[14] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934.

[15] X. He, X. Ren, H. Sandberg, and K. H. Johansson, *How to secure distributed filters under sensor attacks*, IEEE Trans. Automat. Control, 67 (2022), pp. 2843–2856, https://doi.org/10.1109/TAC.2021.3092603.

[16] X. Huang, J. Liang, X. Shen, and Q. Liang, *A multimodal data harness approach of mobile sensors trajectory planning for target tracking*, IEEE Internet Things J., 10 (2023), pp. 9252–9261, https://doi.org/10.1109/JIOT.2022.3222665.

[17] M. U. Javed, J. I. Poveda, and X. Chen, *Excitation conditions for uniform exponential stability of the cooperative gradient algorithm over weakly connected digraphs*, IEEE Control Syst. Lett., 6 (2022), pp. 67–72, https://doi.org/10.1109/LCSYS.2021.3049153.

[18] J. Li, J. Guo, and W. Shiu, *Bounds on normalized Laplacian eigenvalues of graphs*, J. Inequal. Appl., 2014 (2014), 316, https://doi.org/10.1186/1029-242X-2014-316.

[19] W. Li, Z. Wang, D. W. C. Ho, and G. Wei, *On boundedness of error covariances for Kalman consensus filtering problems*, IEEE Trans. Automat. Control, 65 (2020), pp. 2654–2661, https://doi.org/10.1109/TAC.2019.2942826.

[20] Y. Luo, J. Zhou, and W. Yang, *Distributed state estimation with colored noises*, IEEE Trans. Circuits Syst. II Express Briefs, 69 (2022), pp. 2807–2811, https://doi.org/10.1109/TCSII.2021.3136184.

[21] R. Olfati-Saber, *Distributed Kalman filtering for sensor networks*, in Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, 2007, IEEE, Piscataway, NJ, pp. 5492–5498.

[22] R. Olfati-Saber, J. A. Fax, and R. M. Murray, *Consensus and cooperation in networked multi-agent systems*, Proc. IEEE, 95 (2007), pp. 215–233.

[23] I. D. Schizas, G. Mateos, and G. B. Giannakis, *Distributed LMS for consensus-based in-network adaptive processing*, IEEE Trans. Signal Process., 57 (2009), pp. 2365–2382, https://doi.org/10.1109/TSP.2009.2016226.

[24] V. Vahidpour, A. Rastegarnia, M. Latifi, A. Khalili, and S. Sanei, *Performance analysis of distributed Kalman filtering with partial diffusion over noisy network*, IEEE Trans. Aerosp. Electron. Syst., 56 (2020), pp. 1767–1782, https://doi.org/10.1109/TAES.2019.2933961.

[25] S. Xie and L. Guo, *Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition*, SIAM J. Control Optim., 56 (2018), pp. 3404–3431, https://doi.org/10.1137/16M1106791.

[26] S. Xie, Y. Zhang, and L. Guo, *Convergence of a distributed least squares*, IEEE Trans. Automat. Control, 66 (2021), pp. 4952–4959, https://doi.org/10.1109/TAC.2020.3047989.

[27] Y. Yang, Y. Qin, Q. Pan, Y. Yang, and Z. Li, *Distributed information filter for linear systems with colored measurement noise*, in 2019 22th International Conference on Information Fusion (FUSION), IEEE, Piscataway, NJ, 2019, pp. 1–7.

[28] P. Yi, Y. Hong, and F. Liu, *Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems*, Automatica J. IFAC, 74 (2016), pp. 259–269, https://doi.org/10.1016/j.automatica.2016.08.007.

[29] J. Zeng, W. Mi, and W. X. Zheng, *A diffusion bias-compensated LMS algorithm for distributed estimation with ARMAX models*, Digital Signal Process., 142 (2023), 104202, https://doi.org/10.1016/j.dsp.2023.104202.

[30] J. Zhou, Y. Xu, J.-Y. Li, and Y. Li, *Robust state estimation for discrete time systems with colored noises and communication constraints*, J. Franklin Inst., 355 (2018), pp. 5790–5810, https://doi.org/10.1016/j.jfranklin.2018.05.043.