A Note on the Convergence of Distributed RLS

Zhaobo Liu and Chanying Li *Member, IEEE*

***Abstract*—This note focuses on the convergence problem of a distributed recursive least squares (RLS) estimator, which converges fast in many circumstances. We find that the convergence behavior of this distributed RLS is quite different for the scalar-parameter and high-dimensional-parameter cases.**

***Index Terms*—distributed estimation, strong consistency, mean square convergence, recursive least-squares, noises**

I. Introduction

An adaptive network is built up from a set of nodes which could communicate with their neighbors through interlinks. Each node observes partial information related to an unknown parameter of common interest and performs local estimation separately [8], [10], [12], [18]. Such network enjoys a certain advantages in robustness and privacy. Specifically, consider a network consisting of n nodes. The corresponding stochastic linear regression models are described by

$$y_{k,i} = \theta^\tau \phi_{k,i} + \varepsilon_{k,i}, \quad k \geq 0, \ i = 1, \ldots, n, \tag{1}$$

where $\varepsilon_{k,i}$ is a scalar noise sequence and $\theta \in \mathbb{R}^m$ is an unknown deterministic parameter vector. At time k, each node i observes a noisy signal $y_{k,i} \in \mathbb{R}$ and a data signal $\phi_{k,i} \in \mathbb{R}^m$. Let the network topology be depicted by a directed weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of the nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of the edges that any $(i, j) \in \mathcal{E}$ means $\mathcal{G}$ contains a directed path from j to i. The structure of the graph $\mathcal{G}$ is described by the weighted adjacency matrix $\mathcal{A} = \{a_{ij}\}_{n\times n}$, where $a_{ij} > 0$ for $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. Moreover, $\sum_{j=1}^n a_{ij} = 1$ for each $i \in [1, n]$.

At each step k, analogous to the distributed optimization problem [16], we hope to optimize $\sum_{i=1}^n f_{k,i}(x)$, where

$$f_{k,i}(x) = \sum_{j=1}^k (y_{j,i} - \phi_{j,i}^T x)^2, \quad i = 1, \ldots, n$$

are the local loss functions. For the fixed k, [16] achieves it literally by minimizing $f_{k,i}$ at each node and then weighting the minimizers from neighbor nodes. However, our problem is a dynamic optimization process with the optimization function varying at each time. This process consumes a lot of computing resources and hence a recursive heuristic algorithm is preferable. Since the recursive least squares (RLS) is one of the most frequently used algorithms in practice due to its good performance [1], [7], [17], a natural estimation strategy for the network is that the nodes update their estimates according to the RLS rule with the local data and then combine the estimates of their neighbours to produce the final estimates. Mathematically, given initial estimate $\theta_{0,i} \in \mathbb{R}^m$ at node i, the strategy is defined as follows for time k.

1) Adaption:

$$\begin{cases} \beta_{k+1,i} = \theta_{k,i} + L_{k,i}(y_{k,i} - \theta_{k,i}^\tau \phi_{k,i}) \\ L_{k,i} = P_{k+1,i}\phi_{k,i} = \frac{P_{k,i}\phi_{k,i}}{1+\phi_{k,i}^\tau P_{k,i}\phi_{k,i}} \\ P_{k+1,i}^{-1} = I_m + \sum_{j=0}^k \phi_{j,i}\phi_{j,i}^\tau \end{cases}. \tag{2}$$

2) Combination:

$$\theta_{k+1,i} = \sum_{j=1}^n a_{ij}\beta_{k+1,j}. \tag{3}$$

Denote $\tilde{\theta}_{k,i} \triangleq \theta_{k,i} - \theta$ and $\widetilde{\Theta}_k \triangleq \mathrm{col}\{\tilde{\theta}_{k,1}, \ldots, \tilde{\theta}_{k,n}\}$. Then

$$\widetilde{\Theta}_{k+1} = (\mathcal{A} \otimes I_m)(I_{mn} - F_k)\widetilde{\Theta}_k + (\mathcal{A} \otimes I_m)L_k V_k,$$

where $\otimes$ is the Kronecker product, $F_k \triangleq L_k\Phi_k^\tau$, $L_k \triangleq \mathrm{diag}\{L_{k,1}, \ldots, L_{k,n}\}$, $\Phi_k \triangleq \mathrm{diag}\{\phi_{k,1}, \ldots, \phi_{k,n}\}$ and $V_k \triangleq \mathrm{col}\{\varepsilon_{k,1}, \ldots, \varepsilon_{k,n}\}$. Compared with other network strategies, this estimator converges fast in many circumstances.

Now, given the connectivity of the network, one might guess that if the data at one single node is sufficient to guarantee the strong consistency (mean-square convergence) of the standard RLS, then the distributed RLS (2)–(3) converges to the true parameter accordingly. This guess is comfirmed for the scalar-parameter case, where we establish a necessary and sufficient condition on the regression data to guarantee the convergence of the distributed RLS to the true parameter, in the sense of the strong consistency and mean-square convergence. But we cannot take the guess for granted when the unknown parameter is of high dimension. Interestingly, for some "defective" data, the cooperation of the nodes in the connected network causes a diverging error, even if the standard RLS at each node can identify the true parameter by using the local data. According to the "defective" data construction provided later, we find that the distributed RLS (2)–(3) is not suitable for the regression data which vary sharply in the high-dimensional-parameter case. As to the other types of distributed strategies using the RLS, the readers are referred to [3], [9], [14], [19].

The rest of the note is organized as follows. The next section presents the main results, while Sections III–IV address the proofs of the theorems.

II. Main Results

At the beginning of this section, we provide some basic definitions as follows.

This work was supported by the National Natural Science Foundation of China under Grants 11925109 and 11688101.

L. Liu and C. Li are with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P. R. China. They are also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, P. R. China. Corresponding author: Chanying Li (cyli@amss.ac.cn).

Definition 1. *[13] A $n \times n$ non-negative matrix T is irreducible if for every pair i, j of its index set, there exists a positive integer $p = p(i,j)$ such that $T^p[i,j] > 0$, where $A[i,j]$ is the (i,j)th entry of matrix A. An irreducible matrix is said to be aperiodic, if the period of any one (and so of each one) of its indices is equal to 1. T is called stochastic (or more precisely row stochastic) if $\sum_{j=1}^n T[i,j] = 1$, $\forall i \in [1,n]$. It is called doubly stochastic if also $\sum_{i=1}^n T[i,j] = 1$, $\forall j \in [1,n]$.*

Definition 2. *[2] A sequence of random variables $\{x_k\}_{k\ge1}$ converges to x in probability if*

$$\lim_{k\to+\infty} P(\|x_k - x\| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Moreover, $\{x_k\}$ converges to x almost surely if

$$P(\lim_{k\to+\infty} \|x_k - x\| = 0) = 1.$$

We use $\|x_k\| \xrightarrow{a.s.} x$ to represent it.

Definition 3. *[2] A σ-algebra generated by random variables $x_1, \dots, x_k$ is denoted by $\sigma\{x_1, \dots, x_k\}$. $E(y|\,\sigma\{x_1, \dots, x_k\})$ is the conditional expectation of y under $x_1, \dots, x_k$.*

We discuss the convergence of the distributed RLS (2)–(3) regarding network (1) in the following sense

$$\|\widetilde{\Theta}_k\| \xrightarrow{a.s.} 0 \text{ and } E\|\widetilde{\Theta}_k\|^2 \to 0 \quad \text{as } k \to +\infty. \tag{4}$$

A. Scalar-Parameter Case

First, consider the scalar-parameter case under assumptions

A1 $\mathcal{A}$ is an irreducible and aperiodic doubly stochastic matrix with $\mathcal{A}^\tau \mathcal{A}$ being irreducible.

A2 There is a constant $M > 0$ such that

$$\begin{cases} E(\varepsilon_{k,i}|\mathcal{G}_k) = 0 \\ \sup\limits_{k\ge0} E(\varepsilon_{k,i}^2|\mathcal{G}_k) \le M \end{cases} \text{a.s.} \quad i = 1, \dots, n,$$

where $\mathcal{G}_k \triangleq \sigma\{V_l, 0 \le l \le k-1\}$.

A3 $\phi_{k,i}, i = 1, \dots, n, k \ge 0$ are non-random constants.

Remark 1. *If graph $\mathcal{G}$ is undirected, connected and containing a self-loop at each node, then it corresponds to a special case of Assumption A1. See the network topology of [18].*

Now, we derive a necessary and sufficient condition for the convergence of the distributed RLS in a collaborative manner when the unknown parameter is a scalar.

Theorem 1. *Let $m = 1$. Under Assumptions A1–A3, (4) holding for all initial $\Theta_0 \in \mathbb{R}^n$ is equivalent to*

$$\lim_{k\to+\infty} \sum_{i=1}^n P_{k+1,i}^{-1} = +\infty. \tag{5}$$

Remark 2. *(i) If (5) fails, as proved in Section III, any initial values $\{\theta_{0,i}, 1 \le i \le n\}$ except the ones satisfying $\sum_{i=1}^n \mu_i(\theta_{0,i} - \theta) = 0$ will lead to $\liminf_{k\to+\infty} E\|\widetilde{\Theta}_k\|^2 > 0$ and $\|\widetilde{\Theta}_k\| \overset{P}{\nrightarrow} 0$, where $\mu_1, \dots, \mu_n > 0$ are some constants determined by data $\{\phi_{k,i}\}$ and matrix $\mathcal{A}$.*
(ii) When the noises and data further satisfy

$$\begin{cases} E\varepsilon_{k,i}^2 > 0, \text{ for all } k \ge 0, 1 \le i \le n \\ E\varepsilon_{k,i}\varepsilon_{k,j} = 0, \text{ for all } k \ge 0, 1 \le i < j \le n \\ \sum_{k=0}^{+\infty} \sum_{i=1}^n \phi_{k,i}^2 \ne 0 \end{cases}, \tag{6}$$

then given an initial $\Theta_0 \in \mathbb{R}^n$ (including $\theta_{0,i} = \theta, i \in [1,n]$), (4) is equivalent to (5). See Appendix A.

We come to an analogous conclusion of Theorem 1 when data $\Phi_k = \text{diag}\{\phi_{k,1}, \dots, \phi_{k,n}\}, k \ge 0$ is a random sequence.

A2' There is a constant $M > 0$ such that

$$\begin{cases} E(\varepsilon_{k,i}|\mathcal{G}'_k) = 0 \\ \sup\limits_{k\ge0} E(\varepsilon_{k,i}^2|\mathcal{G}'_k) \le M \end{cases} \text{a.s.} \quad i = 1, \dots, n,$$

where $\mathcal{G}'_k \triangleq \sigma\{\Phi_i, V_l, 0 \le i \le k, 0 \le l \le k-1\}$.

The proof of the following theorem is given in Appendix A.

Theorem 2. *Under Assumptions A1 and A2', for any initial $\Theta_0 \in \mathbb{R}^n$, $\lim_{k\to+\infty} \|\widetilde{\Theta}_k\| = 0$ holds almost surely on set $\{\lim_{k\to+\infty} \sum_{i=1}^n P_{k+1,i}^{-1} = +\infty\}$.*

B. High-dimensional-parameter Case

In this part, let the network topology be strongly connected:

A1' $\mathcal{A}$ is irreducible.

Recalling the well-known results [5, Theorem 1] and [11, Theorem 3.1], we know that under Assumptions A2–A3, for each single node i, if $\inf_{k\ge0} E\varepsilon_{k,i}^2 > 0$, then $\hat{\theta}_{k,i} \xrightarrow{a.s.} \theta$ and $E(\hat{\theta}_{k,i} - \theta)^2 \to 0$ are both equivalent to

$$\lambda_{\min}(P_{k+1,i}^{-1}) \to +\infty, \tag{7}$$

where $\hat{\theta}_{k,i}$ is the standard RLS estimates at node i. So, the next two theorems in fact show that despite the connectivity, the distributed RLS (2)–(3) can fail, even if the data at each node are sufficient to identify the parameter by using the standard RLS. As proved later, the "defective" data in Theorem 3 are those that vary sharply.

Theorem 3. *Let $m > 1$ and Assumptions A1'and A2 hold. If $\|E\widetilde{\Theta}_0\| \ne 0$, then there is a series of data $\{\Phi_k\}_{k=0}^{+\infty}$ satisfying*

$$\lim_{k\to+\infty} \lambda_{\min}(P_{k+1,i}^{-1}) = +\infty, \quad i = 1, \dots, n, \tag{8}$$

such that $\sup_{k\ge0} E\|\widetilde{\Theta}_k\|^2 = +\infty$.

Let the noises in Assumption A2 be further specified by

A2" $\{(\varepsilon_{k,1}, \dots, \varepsilon_{k,n})^\tau\}_{k\ge0}$ is an i.i.d random sequence with a multivariate normal distribution $N(0, \Sigma)$.

Theorem 4. *Let $m > 1$ and Assumptions A1' and A2" hold. If $\|E\widetilde{\Theta}_0\| \ne 0$, then there is a series of data $\{\Phi_k\}_{k=0}^{+\infty}$ satisfying (8) such that*
(i) for some set D_0 with $P(D_0) > 0$,

$$\sup_{k\ge0} \|\widetilde{\Theta}_k\| = +\infty, \quad \text{a.s. on } D_0; \tag{9}$$

(ii) for any $\varepsilon > 0$, $\limsup_{k\to+\infty} P(\|\widetilde{\Theta}_k\| > \varepsilon) > 0$.

Remark 3. *Although parameter θ is modeled as a deterministic vector, it is also applicable to the stochastic framework.*
(i) Theorems 3 holds for any random parameter θ and Theorem 4 can be derived as well if θ has a normal distribution. See the proofs in Section IV.
(ii) Let parameter θ, data $\{\phi_{k,i}\}$ and noises $\{\varepsilon_{k,i}\}$ in model (1) all be random such that $\theta, \{\phi_{0,i}\}$ and $\{\varepsilon_{k,i}\}$ are mutually

independent for each $i \in [1,n]$. Moreover, $\phi_{k,i} \in \sigma\{y_{j,i}, j = 0,\ldots,k-1\}, k \geq 0$. If θ is Gaussian distributed and $\{\varepsilon_{k,i}\}$ is an i.i.d sequence with the standard normal distribution, then in view of [15], for each node i and any initial value $\hat{\theta}_{0,i}$, $\left\{\lambda_{\min}(P_{k+1,i}^{-1}\right) \to +\infty\} \subset \{\lim_{t\to+\infty} \hat{\theta}_{k,i} = \theta\}$, where $\hat{\theta}_{k,i}$ is the standard RLS at node i.
So, combing (i) and (ii), we draw an analogous conclusion in the stochastic framework.

III. Proof of Theorem 1

We preface the proof with a simple lemma below.

Lemma 1. *Let $\{e_k\}$ be a series of nonnegative real numbers. (i) If for some $d_k \geq 0$ and $\sum_{k=0}^{+\infty} d_k < +\infty$, $e_{k+1} \leq e_k + d_k$, for $\forall k \geq 0$, then $\lim_{k\to+\infty} e_k$ exists.*
(ii) If there exist two nonnegative sequences $\{a_k\}$,$\{b_k\}$ satisfying $\sum_{k=0}^{+\infty} a_k = +\infty$ and $\sum_{k=0}^{+\infty} b_k < +\infty$ such that $e_{k+1} \leq (1-a_k)e_k + b_k$, for $\forall k \geq 0$, then $\lim_{k\to+\infty} e_k = 0$.

Proof. (i) Fix an integer $k > 0$. Then, for any $l \geq k$, $e_l \leq e_k + \sum_{i=k}^{l-1} d_i \leq e_k + \xi_k$, where $\xi_k = \sum_{i=k}^{+\infty} d_i$. So, $e_k \geq \limsup_{l\to+\infty} e_l - \xi_k$, which together with $\lim_{k\to+\infty}\xi_k = 0$ yields $\liminf_{k\to+\infty} e_k \geq \limsup_{l\to+\infty} e_l - \lim_{k\to+\infty}\xi_k = \limsup_{l\to+\infty} e_l$. Then, $\lim_{k\to+\infty} e_k$ exists.

To prove (ii), note that $e_{k+1} \leq e_k + b_k$, where $\sum_{k=0}^{+\infty} b_k < +\infty$. Therefore, $\lim_{k\to+\infty} e_k$ exists by (i). Suppose $e \triangleq \lim_{k\to+\infty} e_k > 0$, so there is a $N > 0$ such that $e_k > \frac{e}{2}$ for all $k > N$. Consequently, $e_{N+i} - e_{N+1} = \sum_{k=N+1}^{N+i-1}(e_{k+1} - e_k) \leq -\sum_{k=N+1}^{N+i-1} a_k e_k + \sum_{k=N+1}^{N+i-1} b_k \leq -\frac{a}{2}\sum_{k=N+1}^{N+i-1} a_k + \sum_{k=N+1}^{+\infty} b_k$, which shows $e_{N+i} \to -\infty$ by letting $i \to +\infty$. This leads to a contradiction and hence $e = 0$. ∎

Lemma 2. *Let $\{e_k, k \geq 0\}$ and $\{d_k, k \geq 0\}$ be two nonnegative processes adapted to a filtration $\{\mathcal{H}_k, k \geq 0\}$. If $E[e_{k+1}|\mathcal{H}_k] \leq e_k + b_k - d_k, k \geq 0$, for some $b_k \geq 0$ with $\sum_{k=0}^{+\infty} b_k < +\infty$, then $\sum_{k=0}^{+\infty} d_k < +\infty$, a.s.. In addition, if $\lim_{k\to+\infty} Ee_k = 0$, then $\lim_{k\to+\infty} e_k = 0$, a.s..*

Proof. As a matter of fact, $\sum_{k=0}^{+\infty} d_k < +\infty$ almost surely is a direct result of [4, Lemma 1.2.2] and this lemma further shows that there exists a random variable e_∞ such that $E|e_\infty| < +\infty$ and $\lim_{k\to+\infty} e_k = e_\infty$, a.s.. Since $e_k \geq 0$, by *Fatou's lemma*, $0 = \liminf_{k\to+\infty} Ee_k \geq Ee_\infty$, which indicates $e_\infty = 0$ almost surely. ∎

Let $\{A_i; i = 1,\ldots,n\}$ be a sequence of $m\times m$ symmetric random matrices satisfying $0 \leq A_i \leq I_m$. Denote $I_1(A) \triangleq \mathrm{diag}\{A_1,\ldots,A_n\}$ and $\psi_1 \triangleq (\mathcal{A}\otimes I_m)I_1(A)$. The following lemma shows

Lemma 3. *Under Assumption A1, for any σ-algebra $\mathcal{F}$, there is a constant $s \in (0,1)$ determined by h and $\mathcal{A}$ such that $\lambda_{\min}(E[I_{mn} - \psi_h^\tau\psi_h|\mathcal{F}]) \geq s\lambda_{\min}(E[\sum_{k=1}^{h}\sum_{i=1}^{n}(I_m - A_{k,i}^2)|\mathcal{F}])$.*

Proof. Denote $\mathcal{B} \triangleq \mathcal{A}^\tau\mathcal{A}$. Since $\mathcal{B}$ is irreducible, for any $i \in [1, n-1]$, there is an integer $d_i \geq 2$ and some distinct $c_1^i,\ldots,c_{d_i}^i \in [1,n]$ such that $c_1^i = i$, $c_{d_i}^i = i+1$, $\mathcal{B}[c_j^i, c_{j+1}^i] > 0$, $j \in [1, d_i - 1]$. Let $q \triangleq \sum_{i=1}^{n-1} d_i - (n-2)$ and define a sequence of $b_j, j = 1,\ldots,q$ with $b_1 = c_1^1$ and $b_j = c^{l+1}_{j-\sum_{i=1}^{l}(d_i-1)}$, where $l \in [0, n-2]$ and $1 + \sum_{i=1}^{l}(d_i - 1) < j \leq 1 + \sum_{i=1}^{l+1}(d_i - 1)$. Hence $\mathcal{B}[b_j, b_{j+1}] > 0$ for all $j \in [1, q-1]$, $\varpi \triangleq \min_{j\in[1,q-1]} \mathcal{B}[b_j, b_{j+1}] > 0$.

Denote $\rho \triangleq \lambda_{\min}(E[\sum_{i=1}^{n}(I_m - A_i^2)|\mathcal{F}])$, and select $0 < s < \frac{\varpi}{64n^3q(1+n^2)}$. Now, suppose for a constant vector $x \in \mathbb{R}^{mn}$ with $\|x\| = 1$,

$$x^\tau E\left[I_{mn} - \psi_1^\tau\psi_1\middle|\mathcal{F}\right] x < s\rho. \tag{10}$$

To this end, write $x = \mathrm{col}\{x_1,\ldots,x_n\} \in \mathbb{R}^{mn}$. A direct calculation yields

$$\begin{aligned} & x^\tau E\left[I_{mn} - \psi_1^\tau\psi_1\middle|\mathcal{F}\right] x \\ = & \sum_{1\leq i<j\leq n} \mathcal{B}[i,j]E[\|A_ix_i - A_jx_j\|^2|\mathcal{F}] \\ & + x^\tau E[I_{mn} - I_1^2(A)|\mathcal{F}]x, \end{aligned} \tag{11}$$

which, together with (10), implies that for any $i \in [1, n-1]$, $\sum_{j=1}^{d_i-1} E[\|A_{c_j^i}x_{c_j^i} - A_{c_{j+1}^i}x_{c_{j+1}^i}\|^2|\mathcal{F}] < \varpi^{-1}s\rho$, hence $\sum_{j=1}^{q-1} E[\|A_{b_j}x_{b_j} - A_{b_{j+1}}x_{b_{j+1}}\|^2|\mathcal{F}] < \varpi^{-1}ns\rho$, and by *Cauchy-Schwarz inequality*,

$$E\left[\|A_ix_i - A_jx_j\|^2\middle|\mathcal{F}\right] < \varpi^{-1}qns\rho, \quad \forall i \neq j. \tag{12}$$

Furthermore, since $x_i^\tau(I_m - A_i^2)x_i + x_j^\tau(I_m - A_j^2)x_j \geq \frac{1}{2}\|(I_m - A_i)x_i - (I_m - A_j)x_j\|^2$, (10) and (11) imply

$$\begin{aligned} & \frac{1}{2}\max_{i,j} E[\|(I_m - A_i)x_i - (I_m - A_j)x_j\|^2|\mathcal{F}] \\ \leq & \sum_{i=1}^{n} E[x_i^\tau(I_m - A_i^2)x_i|\mathcal{F}] < s\rho. \end{aligned}$$

So

$$\begin{aligned} & \sum_{j=1}^{q-1}\mathcal{B}[b_j, b_{j+1}]E[\|x_{b_j} - x_{b_{j+1}}\|^2|\mathcal{F}] \\ \leq & \sum_{1\leq i<j\leq n} \mathcal{B}[i,j]E[\|x_i - x_j\|^2|\mathcal{F}] \\ \leq & \; 2\sum_{1\leq i<j\leq n} \mathcal{B}[i,j](E[\|A_ix_i - A_jx_j\|^2 \\ & + \|(I_m - A_i)x_i - (I_m - A_j)x_j\|^2|\mathcal{F}]) \\ \leq & \; 2s\rho + 2n^2s\rho. \end{aligned} \tag{13}$$

By *Cauchy-Schwarz inequality* and (13), $E[\|x_i - x_j\|^2|\mathcal{F}] \leq q\sum_{j=1}^{q-1} E[\|x_{b_j} - x_{b_{j+1}}\|^2|\mathcal{F}] < 2\varpi^{-1}qns\rho(1+n^2) < \frac{\rho}{16n^2}$, $i < j$.

Now, it yields that $\|x_1 - x_i\|^2 < \frac{\rho}{16n^2}$ for all $i > 1$. Since $\sum_{i=1}^{n}\|x_i\|^2 = 1$, $\|x_1\|^2 \geq \frac{1}{2n-1} - \frac{1}{16n^2}\rho > \frac{1}{4n}$. Therefore,

$$\begin{aligned} x^\tau E\left[I_{mn} - \psi_1^\tau\psi_1\middle|\mathcal{F}\right] x \geq & \sum_{i=1}^{n} x_i^\tau E\left[I_m - A_i^2\middle|\mathcal{F}\right] x_i \\ \geq & \; \frac{1}{2}x_1^\tau E\left[\sum_{i=1}^{n}(I_m - A_i^2)\middle|\mathcal{F}\right] x_1 - \sum_{i=2}^{n}\|x_1 - x_i\|^2 \\ \geq & \; \frac{\rho}{8n} - \frac{\rho}{16n} > s\rho, \end{aligned}$$

which contradicts to (10). So, on every trajectory, $x^\tau E\left[I_{mn} - \psi_1^\tau\psi_1|\mathcal{F}\right] x \geq s\rho$ holds for all unit vector $x \in \mathbb{R}^{mn}$ and Lemma 3 follows. ∎

Taking $m=1$ in Lemma 3 gives

Corollary 1. *Let $c=(c_1,\dots,c_n)^\tau\in\mathbb{R}^n$ be a sequence of random variables satisfying $c_i\in[0,1]$ for all $i\in[1,n]$. Denote $I(c)\triangleq diag\{c_1,\dots,c_n\}$, then there is a constant $s\in(0,1)$ depending on $\mathcal{A}$ such that $\lambda_{max}(I(c)\mathcal{A}^\tau\mathcal{A}I(c))\le 1-s\sum_{i=1}^n(1-c_i^2)$.*

Proof of Theorem 1. First, we show the sufficiency. Without loss of generality, assume $\lim\limits_{k\to+\infty}P_{k+1,1}^{-1}=+\infty$. Since $m=1$,

$$\widetilde{\Theta}_{k+1}=\mathcal{A}(I_n-F_k)\widetilde{\Theta}_k+\mathcal{A}L_kV_k. \tag{14}$$

Denoting $\Lambda_k=E\left[\widetilde{\Theta}_k\widetilde{\Theta}_k^\tau\right]$, Assumption A3 shows

$$\begin{aligned}\Lambda_{k+1} \;=\; &\mathcal{A}(I_n-F_k)\Lambda_k(I_n-F_k)\mathcal{A}^\tau\\ &+\mathcal{A}L_kE[V_kV_k^\tau]L_k\mathcal{A}^\tau.\end{aligned} \tag{15}$$

In view of Assumption A2, applying *Neumann inequality* and Corollary 1 leads to

$$\begin{aligned}\mathrm{tr}(\Lambda_{k+1}) \;\le\; &\left(1-s\sum_{i=1}^n(1-(P_{k+1,i}P_{k,i}^{-1})^2)\right)\mathrm{tr}(\Lambda_k)\\ &+nM\sum_{i=1}^nP_{k+1,i}^2\phi_{k,i}^2.\end{aligned} \tag{16}$$

Because $\lim_{k\to+\infty}P_{k+1,1}^{-1}=+\infty$, we have $\prod_{k=0}^{+\infty}(1-(1-(P_{k+1,1}P_{k,1}^{-1})^2))=0$, which infers

$$\sum_{k=0}^{+\infty}(1-(P_{k+1,1}P_{k,1}^{-1})^2)=+\infty. \tag{17}$$

Furthermore,

$$\begin{aligned}&\sum_{i=1}^n\sum_{k=0}^{+\infty}P_{k+1,i}^2\phi_{k,i}^2=\sum_{i=1}^n\sum_{k=0}^{+\infty}(1-P_{k+1,i}P_{k,i}^{-1})P_{k+1,i}\\ &<\;\sum_{i=1}^n\sum_{k=0}^{+\infty}\left(P_{k+1,i}^{-1}P_{k,i}-1\right)P_{k+1,i}\le\sum_{i=1}^nP_{0,i}\\ &<\;+\infty,\end{aligned} \tag{18}$$

we thus conclude $\lim_{k\to+\infty}\mathrm{tr}(\Lambda_k)=0$ from Lemma 1(ii).

To prove the strong consistency, (14) and Corollary 1 yield

$$\begin{aligned}&E[\widetilde{\Theta}_{k+1}^\tau\widetilde{\Theta}_{k+1}|\mathcal{G}_k]\le\widetilde{\Theta}_k^\tau\widetilde{\Theta}_k\\ &-s\sum_{j=1}^n(1-(P_{k+1,j}P_{k,j}^{-1})^2)\|\widetilde{\Theta}_k\|^2+nM\sum_{i=1}^nP_{k+1,i}^2\phi_{k,i}^2.\end{aligned}$$

Since $\widetilde{\Theta}_k^\tau\widetilde{\Theta}_k\in\mathcal{G}_k$, by (18) and Lemma 2, $\widetilde{\Theta}_{k+1}^\tau\widetilde{\Theta}_{k+1}\to0$ as $k\to+\infty$ almost surely with the convergence rate $\sum_{k=0}^{+\infty}\sum_{i=1}^n(1-(P_{k+1,i}P_{k,i}^{-1})^2)\|\widetilde{\Theta}_k\|^2<+\infty$, a.s..

Now, we prove the necessity under $\lim_{k\to+\infty}\sum_{i=1}^nP_{k+1,i}^{-1}<+\infty$. In this case,

$$\sum_{k=0}^{+\infty}\sum_{i=1}^nP_{k+1,i}\phi_{k,i}^2<+\infty. \tag{19}$$

Denote $\Pi_k\triangleq\prod_{i=k}^0\mathcal{A}(I_n-F_i)$, we first prove $\lim_{k\to+\infty}\Pi_k$ exists. In fact, since $\mathcal{A}$ is an irreducible and aperiodic doubly stochastic matrix, we have $\lim_{k\to}\mathcal{A}^k=\frac{1}{n}\cdot\mathbf{1}\mathbf{1}^\tau$. Then, by (19), given any $\varepsilon>0$, there is a $k_1>0$ such that

$$\begin{cases}\sum_{k=k_1}^{+\infty}\sum_{i=1}^nP_{k+1,i}\phi_{k,i}^2<\frac{\varepsilon}{3}\\ \|\mathcal{A}^k-\mathcal{A}^l\|_1<\frac{\varepsilon}{3},\quad\forall k,l>k_1\end{cases}. \tag{20}$$

Therefore, for every $k>2k_1$,

$$\begin{aligned}&\left\|\Pi_k-\mathcal{A}^{k-k_1}\Pi_{k_1}\right\|_1=\left\|\sum_{j=k_1}^{k-1}\mathcal{A}^{k-j-1}(\Pi_{j+1}-\mathcal{A}\Pi_j)\right\|_1\\ &=\;\|\sum_{j=k_1}^{k-1}\mathcal{A}^{k-j}F_{j+1}\Pi_j\|_1\le\sum_{j=k_1}^{k-1}\|\mathcal{A}^{k-j}F_{j+1}\Pi_j\|_1\\ &\le\;\sum_{j=k_1}^{k-1}\sum_{i=1}^nP_{j+2,i}\phi_{j+1,i}^2<\frac{\varepsilon}{3}.\end{aligned}$$

Combining this and (20) infers that for all $k,l>2k_1$,

$$\begin{aligned}\|\Pi_k-\Pi_l\|_1\;\le\;&\|\Pi_k-\mathcal{A}^{k-k_1}\Pi_{k_1}\|_1+\|\Pi_l-\mathcal{A}^{l-k_1}\Pi_{k_1}\|_1\\ &+\|(\mathcal{A}^{l-k_1}-\mathcal{A}^{k-k_1})\Pi_{k_1}\|_1\\ <\;&\frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\varepsilon,\end{aligned} \tag{21}$$

which means $\lim_{k\to+\infty}\Pi_k$ exists.

Now, denote $\Pi\triangleq\lim_{k\to+\infty}\Pi_k$. Observe that $\Pi_{k+1}=\mathcal{A}(I_n-F_{k+1})\Pi_k$ and $\lim_{k\to+\infty}\mathcal{A}(I_n-F_k)=\mathcal{A}$ then $\Pi=\mathcal{A}\Pi$. Consequently, $\Pi=\mathbf{1}\cdot(\mu_1,\dots,\mu_n)$ for some $\mu_i\ge0, i=1,\dots,n$. We now prove $\mu_i>0$ for all $i=1,\dots,n$. First, (19) infers that there is a $k_2>k_1$ such that $\sum_{k=k_2}^{+\infty}\sum_{i=1}^nP_{k+1,i}\phi_{k,i}^2<1$. Furthermore, $1-P_{k+1,i}\phi_{k,i}^2>0$ for all $k\ge0$, $i=1,\dots,n$ and $\lim_{k\to}\mathcal{A}^k=\frac{1}{n}\cdot\mathbf{1}\mathbf{1}^\tau$, we then conclude that as long as k_2 is sufficiently large, $\min_{i,j}\Pi_{k_2}[i,j]>0$. Further, since $\mathcal{A}$ is a doubly stochastic matrix, for all $k\ge0$, $\min_{i,j}\Pi_{k+1}[i,j]\ge(1-\max_jP_{k+1,j}\phi_{k,j}^2)\min_{i,j}\Pi_k[i,j]$. As a result, by (20) and $\min_{i,j}\Pi_{k_2}[i,j]>0$, $\min\{\mu_1,\dots,\mu_n\}=\liminf_{k\to+\infty}\min_{i,j}\Pi_{k+1}[i,j]\ge\min_{i,j}\Pi_{k_2}[i,j]\prod_{k=k_2}^{+\infty}(1-\sum_{i=1}^nP_{k+1,i}\phi_{k,i}^2)>0$. So, in view of (15),

$$\begin{aligned}\liminf_{k\to+\infty}\mathrm{tr}(\Lambda_{k+1})\;\ge\;&\liminf_{k\to+\infty}\mathrm{tr}(\Pi_k\widetilde{\Theta}_0\widetilde{\Theta}_0^\tau\Pi_k^\tau)\\ =\;&n\left(\sum_{j=1}^n\mu_j(\theta_{0,j}-\theta)\right)^2,\end{aligned} \tag{22}$$

which infers $\liminf\limits_{k\to+\infty}E\widetilde{\Theta}_k^\tau\widetilde{\Theta}_k>0$ if $\sum_{j=1}^n\mu_j(\theta_{0,j}-\theta)\neq0$.

The last part is addressed to proving $\widetilde{\Theta}_k\overset{P}{\nrightarrow}0$. By (16) and (18), Lemma 1(i) shows that $\lim_{k\to+\infty}E\|\widetilde{\Theta}_k\|^2$ exists. Denote $Q\triangleq\sup_{k\ge0}E\|\widetilde{\Theta}_k\|^2$, $\Pi(k,i)\triangleq\prod_{j=k}^{k-i+1}\mathcal{A}(I_n-F_k)$. By (18), for any fixed $\varepsilon>0$, there is a $k_3>0$ such that $\sum_{j=k_3}^{+\infty}\sum_{i=1}^nP_{j+1,i}^2\phi_{j,i}^2<\frac{\varepsilon}{4M}$. In addition, similar to (21), there is a $k_4>k_3$ such that for any $k,l>k_4$, $\|\Pi(k,k-k_3-1)-\Pi(l,l-k_3-1)\|<\frac{\varepsilon}{2Q}$.

So, as long as $k,l>k_4$,

$$\begin{aligned}&\widetilde{\Theta}_{k+1}-\widetilde{\Theta}_{l+1}\\ &=\;\Pi(k,k-k_3+1)\widetilde{\Theta}_{k_3}+\sum_{j=k_3}^k\Pi(k,k-j)\mathcal{A}L_jV_j\\ &\quad-\Pi(l,l-k_3+1)\widetilde{\Theta}_{k_3}-\sum_{j=k_3}^l\Pi(l,l-j)\mathcal{A}L_jV_j,\end{aligned}$$

which infers

$$
\begin{aligned}
& E\|\widetilde{\Theta}_{k+1}-\widetilde{\Theta}_{l+1}\|^2 \\
\leq\ & E\|(\Pi(k,k-k_3+1)-\Pi(l,l-k_3+1))\widetilde{\Theta}_{k_3}\|^2 \\
& +2\sum_{j=k_3}^{\max\{k,l\}} E\|L_jV_j\|^2 \\
<\ & \frac{\varepsilon}{2Q}\cdot Q+2M\sum_{j=k_3}^{+\infty}\sum_{i=1}^{n}P_{j+1,i}^2\phi_{j,i}^2<\varepsilon.
\end{aligned}
$$

This means $\{\widetilde{\Theta}_k\}_{k\geq 0}$ is a Cauchy sequence in $L^2(dP)$, and hence there exists a random vector $Z\in L^2(dP)$ such that $\lim_{k\to+\infty}E\|\widetilde{\Theta}_k-Z\|^2=0$. So, $\widetilde{\Theta}_k\xrightarrow{P}Z$. Note that $Z\neq 0$ due to $\lim_{k\to+\infty}E\|\widetilde{\Theta}_k\|^2\neq 0$. ■

IV. Proofs of Theorems 3

Since a deterministic parameter can be viewed as a random variable having a degenerate Gaussian distribution with zero variance, it suffices to prove Remark 3 by assuming that θ in Theorems 3–4 is random. In addition, let θ in Theorem 4 be Gaussian distributed.

We first prove a technical lemma. Let $A[i,j]$ be the (i,j)th entry of matrix A. Fix a $j^*\in\{1,\ldots,n\}$ and let d be the smallest integer that $\mathcal{A}^{d+1}[j^*,j^*]>0$. define a sequence of vectors in $\mathbb{R}^{mn}$:

$$
\begin{cases}
\mathcal{P}_0\triangleq\{C:C[m(j^*-1)+1,1]=0\} \\
\mathcal{P}_l\triangleq\{C:\sum_{k=1}^n b_{l,k}C[m(k-1)+1,1]=0\},l\in[1,d]
\end{cases},
$$

where for $l\geq 1$ and $k=1,\ldots,n$,

$$
b_{l,k}=\sum_{i\neq j^*}\sum_{i_1,\ldots,i_{l-1}}a_{j^*i}a_{ii_1}\ldots a_{i_{l-1}k}.
$$

Lemma 4. *Given $k\geq 1$, let $f=(f_1,\ldots,f_{mn})^\tau:\mathbb{R}^k\to\mathbb{R}^{mn}$ be a map that each $f_i(z)$ is a polynomial of $z\in\mathbb{R}^k, 1\leq i\leq mn$. If $f(\mathbb{R}^k)\not\subset\mathcal{P}_l$ for some $l\in[0,d]$, then for any nonempty open set $U\subset\mathbb{R}^k$, there is a $z\in U$ such that $f(z)\notin\mathcal{P}_l$.*

Proof. Since $\sum_{k=1}^n b_{l,k}f_{m(k-1)+1}(z)$ is a polynomial, if for some nonempty open set $U\subset\mathbb{R}^k$, $\sum_{k=1}^n b_{l,k}f_{m(k-1)+1}(z)\equiv 0$ for all $z\in U$, then the polynomial must be identically zero on $\mathbb{R}^k$. This contradicts to $f(\mathbb{R}^k)\not\subset\mathcal{P}_l, l\in[0,d]$. ■

Lemma 5. *Let $C\in\mathbb{R}^{mn}$ be a vector and $B_i\in\mathbb{R}^{m\times m}$, $i=1,\ldots,n$ be a sequence of positive definite matrices. Define a map $Q_0:\mathbb{R}^{mn}\to\mathbb{R}^{mn\times mn}$ by $Q_0(z)\triangleq diag\{(B_i^{-1}+v_iv_i^\tau)^{-1}B_i^{-1},i=1,\ldots,n\}$, where $z=col\{v_1,\ldots,v_n\}$ and $v_i\in\mathbb{R}^m$, $1\leq i\leq n$. For each $l\in[0,d]$,*
(i) if $C\notin\mathcal{P}_{l+1}$, then for any nonempty open set $U\in\mathbb{R}^{mn}$, there is a $z\in U$ such that $(\mathcal{A}\otimes I_m)(Q_0(z)C)\notin\mathcal{P}_l$; (ii) if $C\notin\mathcal{P}_0$, then for any nonempty open set $U\in\mathbb{R}^{mn}$, there is a $z\in U$ such that $(\mathcal{A}\otimes I_m)(Q_0(z)C)\notin\mathcal{P}_d$.

Proof. (i) Let $D(z)\triangleq(\mathcal{A}\otimes I_m)Q_0(z)C$, then

$$
\begin{aligned}
D[m(i-1)+1,1](z) \ =\ & -\sum_{k=1}^n a_{ik}\frac{v_k^\tau C_1^{(k)}B_kv_k}{1+v_k^\tau B_kv_k} \\
& +\sum_{k=1}^n a_{ik}C[m(k-1)+1,1],
\end{aligned}
$$

where $C_1^{(k)}\in\mathbb{R}^{m\times m}$ satisfies

$$
C_1^{(k)}=\begin{bmatrix} C[m(k-1)+1,1] & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C[mk,1] & 0 & \ldots & 0\end{bmatrix},\quad k=1,\ldots,n.
$$

Becasue each component of $D(z)\prod_{k=1}^n(1+v_k^\tau B_kv_k)$ is a polynomial and $\prod_{k=1}^n(1+v_k^\tau B_kv_k)>0$, in view of Lemma 4, it is sufficient to prove $D(\mathbb{R}^{mn})\prod_{k=1}^n(1+v_k^\tau B_kv_k)\not\subset\mathcal{P}_l$.

Suppose $(\mathcal{A}\otimes I_m)(Q_0(\mathbb{R}^{mn})C)\subset\mathcal{P}_l$ and let $v_j=(x_{1j},\ldots,x_{mj})^\tau$, $j\in[1,n]$. If $l\geq 1$, the constant term of $\prod_{k=1}^n(1+v_k^\tau B_kv_k)\sum_{k=1}^n b_{l,k}D[m(k-1)+1,1]$ is

$$
\begin{aligned}
& \sum_{k=1}^n\sum_{i=1}^n b_{l,i}a_{i,k}C[m(k-1)+1,1] \\
=\ & \sum_{k=1}^n b_{l+1,k}C[m(k-1)+1,1]=0,
\end{aligned}
$$

which implies $C\in\mathcal{P}_{l+1}$. It contradicts to $C\notin\mathcal{P}_{l+1}$. If $l=0$, the coefficient of $x_{1j^*}^2$ of $\prod_{k=1}^n(1+v_k^\tau B_kv_k)D[m(j^*-1)+1,1]$ is $B_{j^*}[1,1]\sum_{k\neq j^*}a_{j^*k}C[m(k-1)+1,1]=0$, which implies $C\in\mathcal{P}_1$ since B_j is positive definite. Hence, it leads to a contradiction again.

(ii) If $(\mathcal{A}\otimes I_m)(Q_0(\mathbb{R}^{mn})C)\subset\mathcal{P}_d$, then the coefficient of $x_{1j^*}^2$ and the constant term of

$$
\prod_{k=1}^n(1+v_k^\tau B_kv_k)\sum_{k=1}^n b_{d,k}D[m(k-1)+1,1]
$$

are $B_{j^*}[1,1]\sum_{k\neq j^*}\sum_{i=1}^n b_{d,i}a_{i,k}C[m(k-1)+1,1]=0$, and $\sum_{k=1}^n\sum_{i=1}^n b_{d,i}a_{i,k}C[m(k-1)+1,1]=0$, respectively. As a result, $\sum_{i=1}^n b_{d,i}a_{i,j^*}C[m(j^*-1)+1,1]=\sum_{i\neq j^*}\sum_{i_1,\ldots,i_d}a_{j^*i}a_{ii_1}\ldots a_{i_dj^*}C[m(j^*-1)+1,1]=0$. So, by $C\notin\mathcal{P}_0$, $\sum_{i\neq j^*}\sum_{i_1,\ldots,i_d}a_{j^*i}a_{ii_1}\ldots a_{i_dj^*}=0$. Hence $\mathcal{A}^{d+1}[j^*,j^*]=\sum_{i=1}^n\sum_{i_1,\ldots,i_d}a_{j^*i}a_{ii_1}\ldots a_{i_dj^*}=a_{j^*j^*}(\sum_{i_1,\ldots,i_d}a_{j^*i_1}\ldots a_{i_dj^*})$, which together with $\mathcal{A}^{d+1}[j,j]>0$ implies $\mathcal{A}^d[j^*,j^*]=\sum_{i_1,\ldots,i_d}a_{j^*i_1}\ldots a_{i_dj^*}>0$. This contradicts to the definition of d. ■

Letting

$$
z=\mathrm{col}\{\underbrace{\mathbf{0},\ldots,\mathbf{0}}_{j^*-1},v,\underbrace{\mathbf{0},\ldots,\mathbf{0}}_{n-j^*}\}
$$

in Lemma 5 shows

Corollary 2. *Let $C\in\mathbb{R}^{mn}$ and $B\in\mathbb{R}^{m\times m}$ be a vector and a positive definite matrix. Denote $Q_0^*:\mathbb{R}^m\to\mathbb{R}^{mn\times mn}$ by*

$$
Q_0^*(v)\triangleq diag\{\underbrace{I_m,\ldots,I_m}_{j^*-1},(B^{-1}+vv^\tau)^{-1}B^{-1},\underbrace{I_m,\ldots,I_m}_{n-j^*}\},
$$

(i) If $C\notin\mathcal{P}_{l+1}$, then for any nonempty open set $U\in\mathbb{R}^m$, there is $z\in U$ such that $(\mathcal{A}\otimes I_m)(Q_0^*(z)C)\not\subset\mathcal{P}_l$.
(ii) If $C\notin\mathcal{P}_0$, then for any nonempty open set $U\in\mathbb{R}^m$, there is $z\in U$ such that $(\mathcal{A}\otimes I_m)(Q_0^*(z)C)\not\subset\mathcal{P}_d$.

The next lemma with the proof given in Appendix B is the main reason for the failure of the diffusion RLS in Remark 3. We introduce some necessary notations. For $C\in\mathbb{R}^{mn}$ and $B\in\mathbb{R}^{m\times m}$ defined in Corollary 2, denote maps $Q_1:\mathbb{R}^m\to$

$\mathbb{R}^{mn\times mn}$, $Q_2:\mathbb{R}^m\times\mathbb{R}^m\to\mathbb{R}^{mn\times mn}$ and $Q_3:\mathbb{R}^m\times\mathbb{R}^m\to\mathbb{R}^{mn\times mn}$ by

$$\begin{cases} Q_1(v_1)\triangleq \mathrm{diag}\{\underbrace{I_m,\ldots,I_m}_{j^*-1},B_1B^{-1},\underbrace{I_m,\ldots,I_m}_{n-j^*}\} \\ Q_2(v_1,v_2)\triangleq \mathrm{diag}\{\underbrace{I_m,\ldots,I_m}_{j^*-1},B_2B_1^{-1},\underbrace{I_m,\ldots,I_m}_{n-j^*}\} \\ Q_3(v_1,v_2)\triangleq(\mathcal{A}\otimes I_m)Q_2(v_1,v_2)(\mathcal{A}\otimes I_m)Q_1(v_1)C \end{cases},$$

where $B_1\triangleq(B^{-1}+v_1v_1^\tau)^{-1}$, $B_2\triangleq(B^{-1}+v_1v_1^\tau+v_2v_2^\tau)^{-1}$ and $v_1,v_2\in\mathbb{R}^m$.

Lemma 6. *Let $a_{lj^*}>0$ for some $l\in[1,n]$, where j^* is the fixed index defined before. If $C\notin\mathcal{P}_1$, then for any $L>0$, there exist some $v_1,v_2\in\mathbb{R}^m$ such that $(\mathcal{A}\otimes I_m)(Q_1(v_1))C\notin\mathcal{P}_0$, $Q_3(v_1,v_2)\notin\mathcal{P}_d$, $|Q_3(v_1,v_2)[m(l-1)+1,1]|>L$.*

Lemma 7. *Let $C\in\mathbb{R}^{mn}$ and $\{B_i\in\mathbb{R}^{m\times m}\}$ be defined in Lemma 5. For any $K>0$, if $C\notin\mathcal{P}_d$, then there exist some $z_j=col\{v_{j,1},\ldots,v_{j,n}\}\in\mathbb{R}^{mn}$, $j\in[1,m]$, such that*

$$\begin{cases} \inf_{i\in[1,n]}\lambda_{\min}\left(B_i^{-1}+\sum_{j=1}^m v_{j,i}v_{j,i}^\tau\right)>K \\ \prod_{k=j}^1(\mathcal{A}\otimes I_m)G_k(z_1,\ldots,z_k)C\notin\mathcal{P}_{d-j}, \quad j\in[1,m] \end{cases},$$

where $\mathcal{P}_{-l}\triangleq\mathcal{P}_{d-l+1}$, $l\geq1$ and

$$\begin{cases} G_j(z_1,\ldots,z_j)\triangleq diag\left\{B_{1,j}B_{1,j-1}^{-1},\ldots,B_{n,j}B_{n,j-1}^{-1}\right\} \\ B_{i,j}\triangleq(B_i^{-1}+\sum_{k=1}^j v_{k,i}v_{k,i}^\tau)^{-1}, \quad 1\leq i\leq n \end{cases}$$

for $1\leq j\leq m$ and $B_{i,0}\triangleq B_i$, $1\leq i\leq n$.

Proof. Let e_j denote the jth column of the identity matrix I_m, $j\in[1,m]$ and $z_j^*=\mathrm{col}\{v_{j,1}^*,\ldots,v_{j,n}^*\}=\mathrm{col}\{\sqrt{nK}\cdot e_j,\ldots,\sqrt{nK}\cdot e_j\}$. Then, for $i\in[1,n]$, $\lambda_{\min}(B_i^{-1}+\sum_{j=1}^m v_{j,i}^*(v_{j,i}^*)^\tau)\geq\lambda_{\min}(\sum_{j=1}^m v_{j,i}^*(v_{j,i}^*)^\tau)=nK>K$. Since $\lambda(z_1,z_2,\ldots,z_m)\triangleq\inf_{i\in[1,n]}\lambda_{\min}(B_i^{-1}+\sum_{j=1}^m v_{j,i}v_{j,i}^\tau)$ is continuous in $z_1,\ldots,z_m$, there exists a neighbourhood U_1 of z_1^* such that $\lambda(s,z_2^*,\ldots,z_m^*)>K$ for all $s\in U_1$. By Lemma 5, there is a $z_1\in U_1$ such that $(\mathcal{A}\otimes I_m)G_1(z_1)C\notin\mathcal{P}_{d-1}$. An analogous argument shows that we can select a series of $z_1,\ldots,z_m$ satisfying

$$\begin{cases} \inf_{i\in[1,n]}\lambda_{\min}\left(B_i^{-1}+\sum_{j=1}^m v_{j,i}v_{j,i}^\tau\right)>K \\ \prod_{k=j}^1(\mathcal{A}\otimes I_m)G_k(z_1,\ldots,z_k)C\notin\mathcal{P}_{d-j} \end{cases}, \quad j\in[1,m],$$

which is exactly the result as desired. ∎

Lemma 8. *Let $E\widetilde{\Theta}_0[m(j^*-1)+1,1]\neq0$ and $a_{lj^*}>0$ for some $l\in[1,n]$. Then, under Assumption A1', there is a sequence of deterministic matrices $\{\Phi_i\}_{i=0}^{+\infty}$ such that $\lim_{t\to+\infty}\lambda_{\min}(\sum_{i=0}^t\Phi_i\Phi_i^\tau)=+\infty$ and for $R_t\triangleq\prod_{i=t}^0(\mathcal{A}\otimes I_m)(I_{mn}-F_i)E\widetilde{\Theta}_0$, $\limsup_{t\to+\infty}\frac{|R_t[m(l-1)+1,1]|}{16(t+1)^4}>1$.*

Proof. It suffices to construct a series of deterministic $\{\Phi_i\}_{i=0}^{+\infty}$ such that for any $k\geq0$, $s\in[0,d]$ and $t_k=k(m+3)(d+1)$,

$$\begin{cases} R_{t_k+j}\notin\mathcal{P}_{d-j} \\ \lambda_{\min}\left(\sum_{i=0}^{t_k+m}\Phi_i\Phi_i^\tau\right)>t_k+m \\ |R_{t_{k+1}}[m(l-1)+1,1]|>20(t_{k+1}+1)^4 \end{cases}. \tag{23}$$

First, since $E\widetilde{\Theta}_0\notin\mathcal{P}_0$, by Lemma 5, there is a Φ_0 such that $R_0\notin\mathcal{P}_d$. Let $k=0$. In view of Lemma 7, we can find some $\Phi_{t_k+j}, j=1\ldots,m$, such that for all $j\in[1,m]$,

$$\begin{cases} \lambda_{\min}\left(\sum_{i=0}^{t_k+m}\Phi_i\Phi_i^\tau\right)>t_k+m \\ R_{t_k+j}=\prod_{i=j}^1(\mathcal{A}\otimes I_m)(I_{mn}-F_{t_k+i})R_{t_k}\notin\mathcal{P}_{d-j} \end{cases}. \tag{24}$$

Moreover, by Lemma 5, there are some $\Phi_{t_k+j}, j=m+1,\ldots,(m+3)(d+1)-2$ such that for all $j\in[m+1,(m+3)(d+1)-2]$, $R_{t_k+j}=\prod_{i=j}^{m+1}(\mathcal{A}\otimes I_m)(I_{mn}-F_{t_k+i})R_{t_k+m}\notin\mathcal{P}_{d-j}$. Finally, by noting that $R_{t_k+(m+3)(d+1)-2}\notin\mathcal{P}_1$, Lemma 6 indicates that for some

$$\Phi_{t_{k+1}-i}=\mathrm{diag}\{\underbrace{\mathbf{0},\ldots,\mathbf{0}}_{j^*-1},v_{i,j^*},\underbrace{\mathbf{0},\ldots,\mathbf{0}}_{n-j^*}\}, \quad i=0,1,$$

one has

$$\begin{cases} R_{t_{k+1}-1}\notin\mathcal{P}_0, \quad R_{t_{k+1}}\notin\mathcal{P}_d \\ |R_{t_{k+1}}[m(l-1)+1,1]|>20(t_{k+1}+1)^4 \end{cases}. \tag{25}$$

So, we obtain a series of $\{\Phi_j, j=0,\ldots,t_1\}$ fulfilling (23). By repeating (24) to (25) for all $k\geq1$, (23) is proved immediately based on the mathematical induction. ∎

Proof of Theorem 3. Considering $\|E\widetilde{\Theta}_0\|\neq0$ and Assumption A1', we suppose, without loss of generality, there are some $j^*,l\in\{1,\ldots,n\}$ such that $E\widetilde{\Theta}_0[m(j^*-1)+1,1]\neq0$ and $a_{lj^*}>0$. Let $\{\Phi_k\}_{k=0}^{+\infty}$ be the deterministic sequence constructed in Lemma 8. Then, by virtue of Assumption A2 and (14), $E(\widetilde{\Theta}_k[m(l-1)+1,1])=R_k[m(l-1)+1,1]$, and hence $\sup_{k\geq0}E\|\widetilde{\Theta}_k\|^2\geq\sup_{k\geq0}\|E\widetilde{\Theta}_k\|^2\geq\sup_{k\geq0}(E(\widetilde{\Theta}_k[m(l-1)+1,1]))^2=\sup_{k\geq0}(R_k[m(l-1)+1,1])^2=+\infty$, where R_k is define in Lemma 8. ∎

Proof of Theorem 4. Let $\{\Phi_k\}_{k=0}^{+\infty}$ be defined in the proof of Theorem 3. Since θ is Gaussian distributed, $\widetilde{\Theta}_k[m(l-1)+1,1]$ possesses a normal distribution by Assumption A2". Note that for any random variable $\xi\sim N(E\xi,\sigma^2)$ and $k\geq1$,

$$\begin{aligned} & P\left(16(k+1)^3|\xi|<|E\xi|\right) \\ =\ & I_{\{\sigma\neq0\}}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\frac{1}{16(k+1)^3}\frac{|E\xi|}{|\sigma|}-\frac{E\xi}{\sigma}}^{\frac{1}{16(k+1)^3}\frac{|E\xi|}{|\sigma|}-\frac{E\xi}{\sigma}}e^{-\frac{x^2}{2}}dx \\ \leq\ & I_{\left\{\frac{|E\xi|}{|\sigma|}<8k\right\}}\cdot\frac{1}{\sqrt{2\pi}}\frac{1}{(k+1)^2} \\ & +I_{\left\{\frac{|E\xi|}{|\sigma|}\geq8k\right\}}\cdot\frac{1}{\sqrt{2\pi}}\int_{4k}^{+\infty}\frac{2}{x^3}dx\leq\frac{1}{\sqrt{2\pi}}\frac{1}{k^2}, \end{aligned}$$

and

$$\begin{aligned} & P(|\xi|\leq\varepsilon) \\ =\ & I_{\{\sigma\neq0\}}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\frac{\varepsilon}{|\sigma|}-\frac{E\xi}{\sigma}}^{\frac{\varepsilon}{|\sigma|}-\frac{E\xi}{\sigma}}e^{-\frac{x^2}{2}}dx+I_{\{\sigma=0,|E\xi|\leq\varepsilon\}} \\ \leq\ & 2\cdot I_{\{|E\xi|\leq\varepsilon\}}+I_{\{\sigma>\varepsilon\}}\cdot\frac{2}{\sqrt{2\pi}} \\ & +I_{\{\sigma\leq\varepsilon,|E\xi|>\varepsilon\}}\cdot\frac{1}{\sqrt{2\pi}}\int_{\frac{1}{\varepsilon}(|E\xi|-\varepsilon)}^{+\infty}e^{-\frac{x^2}{2}}dx. \end{aligned} \tag{26}$$

Define

$$D_0\triangleq\bigcap_{k=1}^{+\infty}\left\{|\widetilde{\Theta}_k[m[l-1]+1,1]|\geq\frac{|R_k[m(l-1)+1,1]|}{16(k+1)^3}\right\},$$

then $P(D_0) \geq 1-\sum_{k=1}^{+\infty} P(\{16(k+1)^3|\widetilde{\Theta}_k[m(l-1)+1,1]| < |R_k[m(l-1)+1,1]|\}) \geq 1-\frac{1}{\sqrt{2\pi}}\sum_{k=1}^{+\infty}\frac{1}{k^2} > 0$. According to Lemma 8, $\|\widetilde{\Theta}_k\| \geq |\widetilde{\Theta}_k[m(l-1)+1,1]| \geq \frac{1}{16(k+1)^3}R_k[m(l-1)+1,1] > k+1$, i.o. on D_0. So, (9) holds. Moreover, by Lemma 8 again, $\limsup_{k\to+\infty}|E\widetilde{\Theta}_k[m(l-1)+1,1]| = +\infty$, which together with (26) yields $\liminf_{k\to+\infty} P(|\widetilde{\Theta}_k[m(l-1)+1,1]| \leq \varepsilon) \leq \frac{2}{\sqrt{2\pi}}$, and hence $\limsup_{k\to+\infty} P(\|\widetilde{\Theta}_k\| > \varepsilon) \geq \limsup_{k\to+\infty} P(|\widetilde{\Theta}_k[m(l-1)+1,1]| > \varepsilon) \geq 1-\frac{2}{\sqrt{2\pi}}$. The proof is completed. ∎

Appendix A

Proof of Remark 2(i)(b). The argument is based on the proof of Theorem 1 from (15)–(22). Considering (6), let l be the smallest integer such that $\sum_{i=1}^n \phi_{l,i}^2 \neq 0$. An analogous proof of Theorem 1 shows that for some $\mu_i' > 0, i = 1,\ldots,n$, $\lim_{k\to+\infty}\Pi(k,k-l) = \mathbf{1}\cdot(\mu_1',\ldots,\mu_n')$. As a result, $\liminf_{k\to+\infty}\mathrm{tr}(\Lambda_{k+1}) \geq \liminf_{k\to+\infty}\mathrm{tr}(\Pi(k,k-l)\mathcal{A}L_l E[V_l V_l^\tau]L_l\mathcal{A}^\tau\Pi(k,k-l)^\tau) = n\sum_{i=1}^n(\sum_{j=1}^n a_{ij}\mu_i')^2 P_{l+1,i}^2\phi_{l,i}^2 E\varepsilon_{l,i}^2 > 0$, and $\widetilde{\Theta}_k \overset{P}{\nrightarrow} 0$ follows as proved in Theorem 1. ∎

Proof of Theorem 2. By (14) and Corollary 1, $E[\widetilde{\Theta}_{k+1}^\tau\widetilde{\Theta}_{k+1}|\mathcal{G}_k'] \leq \widetilde{\Theta}_k^\tau\widetilde{\Theta}_k - s\sum_{i=1}^n(1-(P_{k+1,i}P_{k,i}^{-1})^2)\|\widetilde{\Theta}_k\|^2 + nM\sum_{i=1}^n P_{k+1,i}^2\phi_{k,i}^2$. Since $\widetilde{\Theta}_k^\tau\widetilde{\Theta}_k \in \mathcal{G}_k'$, according to (18) and [4, Lemma 1.2.2], $\lim_{k\to+\infty}\widetilde{\Theta}_{k+1}^\tau\widetilde{\Theta}_{k+1}$ exists almost surely and

$$\sum_{k=0}^{+\infty}\sum_{i=1}^n(1-(P_{k+1,i}P_{k,i}^{-1})^2)\|\widetilde{\Theta}_k\|^2 < +\infty, \quad \text{a.s..} \tag{27}$$

Denote $\Theta_\infty \triangleq \lim_{k\to+\infty}\widetilde{\Theta}_k^\tau\widetilde{\Theta}_k$ and

$$\begin{aligned} S &\triangleq \{\Theta_\infty \neq 0\} \cap \left\{\lim_{k\to+\infty}\textstyle\sum_{i=1}^n P_{k+1,i}^{-1} = +\infty\right\} \\ S' &\triangleq \{\Theta_\infty \neq 0\} \cap \left\{\sum_{k=0}^{+\infty}\sum_{i=1}^n(1-(P_{k+1,i}P_{k,i}^{-1})^2) = +\infty\right\}. \end{aligned}$$

Note that by (17), $\{\lim_{k\to+\infty}\sum_{i=1}^n P_{k+1,i}^{-1} = +\infty\} \subset \{\sum_{k=0}^{+\infty}\sum_{i=1}^n(1-(P_{k+1,i}P_{k,i}^{-1})^2) = +\infty\}$, then $S \subset S'$. Moreover,

$$\sum_{k=0}^{+\infty}\sum_{i=1}^n(1-(P_{k+1,i}P_{k,i}^{-1})^2)\|\widetilde{\Theta}_k\|^2 = +\infty \quad \text{on } S',$$

which implies $P(S) \leq P(S') = 0$ by (27). ∎

Appendix B

Proof of Lemma 6. The first step is to seek a pair (v_1,v_2) that

$$|Q_3(v_1,v_2)[m(l-1)+1,1]| > L. \tag{28}$$

To this end, denote $D(v_1) \triangleq (\mathcal{A}\otimes I_m)Q_1(v_1)C$. In the later discussion, we suppress v_1 in $D(v_1)$ for brevity. Calculate

$$\begin{aligned} & Q_3(v_1,v_2)[m(l-1)+1,1] \\ =\ & ((\mathcal{A}\otimes I_m)(Q_2(v_1,v_2))D)[m(l-1)+1,1] \\ =\ & a_{lj^*}(1,0,\ldots,0)B_2B_1^{-1} \\ & \cdot(D[m(j^*-1)+1,1],\ldots,D[mj^*,1])^\tau \\ & +\sum_{i\neq j^*} a_{li}D[m(i-1)+1,1] \\ =\ & a_{lj^*}(1,0,\ldots,0)\left(I_m - \frac{B_1v_2v_2^\tau}{1+v_2^\tau B_1v_2}\right) \\ & \cdot(D[m(j^*-1)+1,1],\ldots,D[mj^*,1])^\tau \\ & +\sum_{i\neq j^*} a_{li}D[m(i-1),1] \\ =\ & -a_{lj^*}\frac{v_2^\tau D_1B_1v_2}{1+v_2^\tau B_1v_2} + \sum_{i=1}^n a_{li}D[m(i-1)+1,1], \end{aligned}$$

where $D_1 \in \mathbb{R}^{m\times m}$ is defined by

$$D_1 \triangleq \begin{bmatrix} D[m(j^*-1)+1,1] & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ D[mj^*,1] & 0 & \ldots & 0 \end{bmatrix}.$$

Similarly, for all $i = 1,\ldots,n$,

$$\begin{aligned} D[m(i-1)+1,1] &= -a_{ij^*}\frac{v_1^\tau C_1Bv_1}{1+v_1^\tau Bv_1} \\ &\quad + \sum_{k=1}^n a_{ik}C[m(k-1)+1,1], \end{aligned} \tag{29}$$

where $C_1 \in \mathbb{R}^{m\times m}$ is defined by

$$C_1 \triangleq \begin{bmatrix} C[m(j^*-1)+1,1] & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C[mj^*,1] & 0 & \ldots & 0 \end{bmatrix}.$$

Now, write $v_i = r_iz_i$, where $r_i > 0$ and $|z_i| = 1$, $i = 1,2$. Since for any $r_1 > 0$,

$$a_{ij^*}\frac{|v_1^\tau C_1Bv_1|}{1+v_1^\tau Bv_1} = a_{ij^*}\frac{|z_1^\tau C_1Bz_1|}{r_1^{-2}+z_1^\tau Bz_1} < a_{ij^*}\frac{|z_1^\tau C_1Bz_1|}{z_1^\tau Bz_1},$$

it is trivial that $a_{ij^*}\frac{|z_1^\tau C_1Bz_1|}{z_1^\tau Bz_1} \leq a_{ij^*}\lambda_{max}(B^{-1})\|B\|\|C\|_1$. Then, by (29), for all $i = 1,\ldots,n$, $|D[m(i-1)+1,1]| \leq (1+\lambda_{max}(B^{-1})\|B\|)\|C\|_1$, which infers $Q_3(v_1,v_2)[m(l-1)+1,1] < -a_{lj^*}\frac{v_2^\tau D_1B_1v_2}{1+v_2^\tau B_1v_2} + (1+\lambda_{max}(B^{-1})\|B\|)\|C\|_1$.

Next, for any $L > 0$, denote $c \triangleq L\cdot a_{lj^*}^{-1} + a_{lj^*}^{-1}(1+\lambda_{max}(B^{-1})\|B\|)\|C\|_1$. If we could find a v_1 such that $D \notin \mathcal{P}_0$ and $K \triangleq 2cB_1 - (D_1B_1 + B_1D_1^\tau)$ is not semi-positive definite, then there is a v_2' such that for any v_2 in some sufficiently small neighbourhood of v_2', $z_2^\tau(D_1-cI_m)B_1z_2 > \frac{c}{r_2^2}$, which can deduce (28). So, according to Corollary 2, there exists a v_2 in this neighbourhood fulfilling both (28) and $Q_3(v_1,v_2) = (\mathcal{A}\otimes I_m)Q_2(v_1,v_2)D \notin \mathcal{P}_d$.

To construct the desired v_1, compute the leading principal minor of order 2 of K by

$$\begin{aligned} & K[1,1]K[2,2] - K^2[1,2] \\ =\ & 4(cB_1[1,1] - D_1[1,1]B_1[1,1])(cB_1[2,2] \\ & -D_1[2,1]B_1[1,2]) - (2cB[1,2] - D_1[2,1]B_1[1,1] \\ & -D_1[1,1]B_1[1,2])^2 \\ =\ & 4c(c-D_1[1,1])(B_1[1,1]B_1[2,2] \\ & -B_1^2[1,2]) - (D_1[1,1]B_1[1,2] - D_1[2,1]B_1[1,1])^2. \end{aligned}$$

Let $z_1 = (q_1,q_2,0,\ldots,0)^\tau$, where q_1,q_2 are two real numbers satisfying $q_1^2+q_2^2 = 1$ and $q_2 \neq 0$. Then,

$$K[1,1]K[2,2] - K^2[1,2] < 0 \tag{30}$$

is equivalent to

$$4c(c-D_1[1,1])\left((B_1^{-1})^*[2,2]-\frac{((B_1^{-1})^*[1,2])^2}{(B_1^{-1})^*[1,1]}\right) < \frac{(D_1[1,1](B_1^{-1})^*[1,2]-D_1[2,1](B_1^{-1})^*[1,1])^2}{(B_1^{-1})^*[1,1]}. \quad (31)$$

where $(B_1^{-1})^*$ is the adjugate of matrix B_1^{-1}.

Calculating the adjugate matrix of B_1^{-1} shows that there exist two constants $M_1, M_2 > 0$ depending on B such that $|l_i| < M_2$ for $i = 1, 2, 3$, where $l_1 \triangleq (B_1^{-1})^*[1,1] - r_1^2 q_2^2 M_1$, $l_2 \triangleq (B_1^{-1})^*[1,2] + r_1^2 q_1 q_2 M_1$, $l_3 \triangleq (B_1^{-1})^*[2,2] - r_1^2 q_1^2 M_1$. Therefore,

$$\begin{aligned}
&\left|(B_1^{-1})^*[2,2]-\frac{((B_1^{-1})^*[1,2])^2}{(B_1^{-1})^*[1,1]}\right| \\
=\ &\left|\frac{r_1^2(l_3q_2^2M_1+l_1q_1^2M_1+2q_1q_2M_1l_2)+l_1l_3-l_2^2}{r_1^2q_2^2M_1+l_1}\right| \\
\leq\ &\frac{2r_1^2M_1M_2(q_1^2+q_2^2)+2M_2^2}{|r_1^2q_2^2M_1+l_1|},
\end{aligned}$$

which yields

$$\limsup_{r_1\to+\infty}\left|(B_1^{-1})^*[2,2]-\frac{((B_1^{-1})^*[1,2])^2}{(B_1^{-1})^*[1,1]}\right| \leq 1+\frac{2M_2}{q_2^2}. \quad (32)$$

In order to estimate the right hand side of (31), we define two functions $H_1(\cdot)$ and $H_2(\cdot)$ by $H_1(q_1/q_2) \triangleq \lim_{r_1\to+\infty} D_1[1,1] = \sum_{k=1}^n a_{j^*k}C[m(k-1)+1,1] - a_{j^*j^*}\frac{z_1^\tau C_1 B z_1}{z_1^\tau B z_1}$ and $H_2(q_1/q_2) \triangleq \lim_{r_1\to+\infty} D_1[2,1] = \sum_{k=1}^n a_{j^*k}C[m(k-1)+2,1] - a_{j^*j^*}\frac{z_1^\tau C_2 B z_1}{z_1^\tau B z_1}$, where $C_2 \in \mathbb{R}^{m\times m}$ satisfies

$$C_2 = \begin{bmatrix} 0 & C[m(j^*-1)+1,1] & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & C[mj^*,1] & 0 & \dots & 0 \end{bmatrix}.$$

As a result,

$$\begin{aligned}
&\frac{1}{M_1r_1^2}\frac{(D_1[1,1](B_1^{-1})^*[1,2]-D_1[2,1](B_1^{-1})^*[1,1])^2}{(B_1^{-1})^*[1,1]} \\
=\ &\frac{(D_1[1,1](l_2-r_1^2q_1q_2M_1)-D_1[2,1](r_1^2q_2^2M_1+l_1))^2}{M_1r_1^2(r_1^2q_2^2M_1+l_1)}, \\
\to\ &q_2^2\left(H_1(q_1/q_2)\,q_1/q_2+H_2(q_1/q_2)\right)^2 \qquad (33)
\end{aligned}$$

as $r_1 \to +\infty$. Therefore, if

$$H_1(q_1/q_2)\,q_1/q_2 + H_2(q_1/q_2) \neq 0, \quad (34)$$

then (30) will follow directly from (32) and (33) by letting $r_1 > N(q_1, q_2)$ for some sufficiently large number $N(z_1)$.

So, the remainder is to show that there is a $x \in \mathbb{R}$ such that

$$H_1(x)x + H_2(x) \neq 0, \quad (35)$$

which is equivalent to

$$\begin{aligned}
&-a_{j^*j^*}\frac{(x,1,0,\dots,0)(xC_1+C_2)B(x,1,0,\dots,0)^\tau}{(x,1,0,\dots,0)B(x,1,0,\dots,0)^\tau} \\
&+x\sum_{k=1}^n a_{j^*k}C[m(k-1)+1,1] \\
&+\sum_{k=1}^n a_{j^*k}C[m(k-1)+2,1] \neq 0.
\end{aligned}$$

If (35) fails, then the coefficient of x^3 of

$$(x,1,0,\dots,0)B(x,1,0,\dots,0)^\tau(H_1(x)x+H_2(x))$$

is

$$B[1,1]\cdot\sum_{k\neq j^*} a_{j^*k}C[m(k-1)+1,1] = 0,$$

which contradicts to $C \notin \mathcal{P}_1$. So, (34) holds if $q_1/q_2 = x$.

We now can conclude that all $v_1 = r_1(q_1, q_2, 0, \dots, 0)^\tau$ with $q_1^2 + q_2^2 = 1$, $q_2 \neq 0$, $q_1/q_2 = x$ and $r_1 > N(q_1, q_2)$ will result in (30). Note that $C \notin \mathcal{P}_1$, by Corollary 2 again, there always exists some v_1 fulfilling both $D \notin \mathcal{P}_0$ and (30), which means K cannot be a semi-positive definite matrix. ∎

REFERENCES

[1] P. D. Achlerkar and B. K. Panigrahi, "Recursive Least Squares-Based Adaptive Parameter Estimation Scheme for Signal Transformation and Grid Synchronization," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 2, pp. 2427–2439, 2021

[2] P. Billingsley, *Probability and Measure*, John Wiley & Sons, 2008.

[3] F. S. Cattivelli, C. G. Lopes and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, 2008.

[4] H. Chen, *Stochastic Approximation and Its Applications*, Kluwer Academic Publishers, 2002.

[5] H. Drygas, "Weak and strong consistency of the least squares estimators in regression models," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 34, pp. 119–127, 1976.

[6] F. Eicker, "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *The Annals of Mathematical Statistics* vol. 34, pp. 447–456, 1963.

[7] K. Jiang, A. C. Victorino, A. Charara , "Adaptive Estimation of Vehicle Dynamics Through RLS and Kalman Filter Approaches," *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, IEEE, 2015.

[8] A. Khalili, M. A. Tinati, A. Rastegarnia and J. A. Chanbers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. on Signal Processing*, vol. 60, no. 2, pp. 974–979, 2012.

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

[10] C. G. Lopes and A. H. Sayed, "Dffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, 2008.

[11] T. L. Lai, H. Robbins and C. Z. Wei, "Strong consistency of least squares estimates in multiple regression II*," *Journal of Multivariate Analysis*, vol. 9, pp. 343–361, 1979.

[12] M. J. Piggott and V. Solo, "Stability of distributed adaptive algorithms II: Diffusion algorithms," *54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 7428–7433.

[13] E. Seneta, *Non-negative matrices and Markov chains*, Springer Science & Business Media, 2006.

[14] A. K. Sahu, S. Kar, J. M. F. Moura and H. V. Poor, "Distributed constrained recursive nonlinear least-squares estimation: Algorithms and asymptotics", *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 426–441, 2016.

[15] J. Sternby, "On consistency for the method of least squares using martingale theory," *IEEE Trans. Autom. Control* vol. 22, pp. 346–352, 1977.

[16] C. A. Uribe, S. Lee, A. Gasnikov, and A. Nedić, "A dual approach for optimal algorithms in distributed optimization over networks", *In 2020 Information Theory and Applications Workshop (ITA)*, pp. 1–37, IEEE.

[17] Z. Wei, K. J. Tseng, N. Wai, et al, "Adaptive estimation of state of charge and capacity with online identified battery model for vanadium redox flow battery," *Journal of Power Sources*, vol.332, pp. 389–398, 2016.

[18] S. Y. Xie and L. Guo, "Analysis of distributed adaptive filters based on diffusion strategies over sensor networks," *IEEE Trans. Autom. Control*, vol. 63, pp. 3643–3658, 2018.

[19] S. Y. Xie, Y. Q. Zhang and L. Guo, "Learning and prediction theory of distributed least squares," *IEEE Trans. Autom. Control*, accepted.