Systems & Control Letters 164 (2022) 105228

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Performance analysis of the compressed distributed least squares algorithm[☆]

Die Gan, Zhixin Liu [*]

Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 22 September 2021
Received in revised form 21 February 2022
Accepted 6 April 2022
Available online 3 May 2022

Keywords:
Sparse signal
Compressive sensing
Distributed least squares
Performance analysis

ABSTRACT

In this paper, we consider the distributed estimation problem of unknown high-dimensional sparse signals for a random dynamic system. We propose a compressed distributed algorithm by using the compressive sensing theory and the distributed least squares (LS) algorithm. Under a compressed cooperative persistent excitation condition, the upper bound of the estimation error is established which is positively related to the restricted isometry constant. Our results are obtained without relying on some stringent conditions such as independency or stationarity of the regression vectors. Finally, we provide a simulation example to show that the compressed distributed least squares algorithm has better performance than the regularized distributed LS algorithm with l_1 penalty for the estimation of high-dimensional sparse signals.

1. Introduction

Over the past decades, wireless sensor networks (WSNs) have attracted increasing research attention due to wide practical applications including target tracking, surveillance, and biomedical health monitoring [1,2]. A large amount of data from WSNs may help improve the performance of the estimation and filtering problems by designing suitable algorithms [3]. Compared with the centralized algorithms, the distributed ones which only depend on local information exchange have the advantages of robustness and scalability, as well as reducing communication load and calculation pressure. Some distributed adaptive estimation and filtering algorithms are proposed based on the incremental, consensus and diffusion strategies, and the theoretical analysis on the performance analysis of the algorithms are also established (cf., [4–9]).

Sparsity is one of the important characteristics of high-dimensional signals like audio signals [10], image signals [11] and biomedical data [12]. The prior knowledge about the sparsity of the signals can be exploited to design appropriate algorithms to improve the estimation performance [13]. Many estimation algorithms based on sparse signals have been proposed in existing literature, see e.g., [14–17]. Most of them are designed using the sparse optimization method where a penalty term is added into the cost function to avoid overfitting. Theoretical results of the distributed adaptive filtering algorithms for sparse random systems are established under some signal conditions. For example, Liu et al. in [17] proposed distributed sparse recursive LS algorithms by using l_1 and l_0 norms as the penalties and the mean stability and mean-square convergence were analyzed with the independent and identically distributed regressors. Lorenzo and Sayed in [15] provided the convergence and mean-square performance analysis of the distributed least mean squares (LMS) algorithm regularized by convex penalties where the assumption of independent regressors is required. However, the independency assumption of regression vectors is too strong to be applied to practical feedback control systems. Moreover, the high-dimensional data are used in the design and analysis of the algorithms, which may lead to high computational complexity, slow convergence rate and degraded mean square error (MSE) performance.

Another branch to deal with sparse signals is the compressive sensing (CS) theory. The CS theory is a novel sampling theory where fewer measurements are required to get a higher accuracy estimation of unknown sparse signals than those in Shannon sample principle (cf., [18,19]). The CS theory can be applied to deal with the estimation problem of high-dimensional sparse signals. Xu et al. in [20] introduced a distributed compressed estimation scheme to estimate the compressed parameter.

☆ This work was supported by the National Key R&D Program of China under Grant 2018YFA0703800, the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDA27000000 and Natural Science Foundation of China under grant U21B6001, and National Science Foundation of Shandong Province, China (ZR2020ZD26).

* Corresponding author at: Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.
E-mail addresses: gandie@amss.ac.cn (D. Gan), lzx@amss.ac.cn (Z. Liu).

https://doi.org/10.1016/j.sysconle.2022.105228
0167-6911/

Numerical simulations show that the distributed estimation algorithm based on the CS theory can improve performance and reduce bandwidth. However, the comprehensive theoretical analysis for the compressed estimation algorithms is still lacking. Li and Li in [21] provided the stability analysis of the compressed–combine–reconstruct–adaptive algorithm for the independent regression vectors. We remark that an elegant theoretical result for the compressed distributed LMS was established by Xie and Guo in [22] under a compressed cooperative information condition.

The distributed LS algorithm has been widely investigated due to its faster convergence rate and practical applications in the area of cloud technologies [23]. The rigorous theoretical analysis for the convergence of the distributed LS algorithm was given in [24], where a "weakest" condition in the existing literature, i.e., cooperative excitation condition was introduced to guarantee the convergence of the distributed LS. However, the cooperative excitation condition may not be satisfied for the high-dimensional sparse regression signals. In order to solve the estimation problem of the sparse random dynamic system, we propose a compressed distributed LS algorithm using the noisy observations and sparse regression vectors from its neighbors. Different from the sparse optimization method in [15–17], we use the compressed regression vectors to estimate the unknown signals in a low-dimensional space by the distributed least squares algorithm. Then, the signal reconstruction algorithm is used to obtain the estimate of the original high-dimensional sparse signal. We introduce a compressed cooperative persistent excitation condition under which the comprehensive analysis for the performance of the proposed algorithm is established. We show that the upper bound for the estimation error is positively related to the restricted isometry constant. We note that compared with [21], our theoretical results are established without relying on the assumptions of the independency and stationarity of regression signals. A numerical example is given to show that the compressed distributed LS algorithm can estimate the unknown high-dimensional sparse signal, while the classical distributed LS algorithm (cf., [24]) cannot fulfill the estimation task due to lack of adequate excitation condition.

The rest of this paper is organized as follows. We first introduce some preliminaries including matrix analysis, the CS theory and graph theory in Section 2. Section 3 presents the compressed distributed LS algorithm. In Section 4, we provide the theoretical analysis for the compressed distributed LS algorithm. A simulation example is presented in Section 5, and concluding remarks are made in Section 6.

2. Some preliminaries

In this paper, we will construct the distributed algorithm to estimate unknown high-dimensional sparse signals and provide the performance analysis of the algorithm. For this purpose, we need to introduce some notations and basic results on the matrix analysis, compressive sensing theory and graph theory.

2.1. Notations

For an m-dimensional vector $\boldsymbol{x}$, the p-norm of $\boldsymbol{x}$ is defined as $\|\boldsymbol{x}\|_p = (\sum_{j=1}^{m} |x_j|^p)^{1/p}$ $(1 \leq p < \infty)$, where x_j denotes the jth element of $\boldsymbol{x}$. For $p = 1$, $\|\boldsymbol{x}\|_1$ is the sum of absolute values of all the elements in $\boldsymbol{x}$; and for $p = 2$, $\|\boldsymbol{x}\|_2$ is the Euclidean norm, we simply write $\|\cdot\|_2$ as $\|\cdot\|$. We also use the notation $\|\boldsymbol{x}\|_0$ which represents the number of nonzero elements of $\boldsymbol{x}$; For an $m \times n$-dimensional real matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\|\boldsymbol{A}\|$ denotes the operator norm induced by the Euclidean norm, i.e., $\|\boldsymbol{A}\| = (\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^T))^{\frac{1}{2}}$, where the notation T denotes the transpose operator and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix. Correspondingly, $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of the matrix. For two real symmetric matrices $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{Y} \in \mathbb{R}^{n\times n}$, $\boldsymbol{X} \geq \boldsymbol{Y}$ ($\boldsymbol{X} > \boldsymbol{Y}$) means that $\boldsymbol{X} - \boldsymbol{Y}$ is a positive semi-definite (definite) matrix. For matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ with suitable dimensions, we have the following matrix inversion formula provided that the relevant matrices are invertible (see [25]),

$$(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}\boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}^{-1} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}. \tag{1}$$

Let $\{\boldsymbol{A}_t\}$ be a matrix sequence and $\{b_t\}$ be a positive scalar sequence. Then by $\boldsymbol{A}_t = O(b_t)$ we mean that there exists a positive constant C independent of t and b_t such that $\|\boldsymbol{A}_t\| \leq Cb_t$ holds for all $t \geq 0$, and by $\boldsymbol{A}_t = o(b_t)$ we mean that $\lim_{t\to\infty} \frac{\|\boldsymbol{A}_t\|}{b_t} = 0$.

2.2. Compressive sensing theory

The CS theory has emerged as a new framework for sampling theory which has many attractive properties, such as robustness to noise, fault tolerance, and bandwidth saving. The vector $\boldsymbol{x}$ is called s-sparse if $\|\boldsymbol{x}\|_0 \leq s$, that is, $\boldsymbol{x}$ has at most s nonzero elements. Assume that the sparse signal $\boldsymbol{x}$ obeys the following equation,

$$\boldsymbol{z} = \boldsymbol{D}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{z}$ is the measurement, $\boldsymbol{D} \in \mathbb{R}^{d\times m}$ is the sensing matrix whose number of rows is much smaller than the number of columns, and $\boldsymbol{\epsilon} \in \mathbb{R}^d$ is the measurement perturbation bounded by $\|\boldsymbol{\epsilon}\| \leq C$. We are interested in how to recover the signal $\boldsymbol{x}$ from the noisy measurement $\boldsymbol{z}$. We see that for a general signal $\boldsymbol{x}$, it is hard or even impossible to recover $\boldsymbol{x}$. However, the recovery problem becomes possible when the signal $\boldsymbol{x}$ is sparse. Candès et al. introduced the definition of the restricted isometry property (RIP) of sensing matrix $\boldsymbol{D}$ in CS theory to study the reconstruction problem of sparse signals, and they proved that the sparse signal $\boldsymbol{x}$ can be recovered with high accuracy if $\boldsymbol{D}$ satisfies the RIP and the noise is small (see e.g., [19,26]).

Definition 1 (*[26] (RIP)*)**.** Let $\boldsymbol{D} \in \mathbb{R}^{d\times m}$ be the sensing matrix, and $\boldsymbol{D}_L$ ($L \subseteq \{1, \dots, m\}$) be the sub-matrix obtained by extracting the columns of $\boldsymbol{D}$ corresponding to the indices in the set L. For given integer s ($1 \leq s \leq m$), we define the s-restricted isometry constant $\delta_s \in [0, 1)$ to be the smallest quantity such that the following inequality

$$(1 - \delta_s)\|\boldsymbol{b}\|^2 \leq \|\boldsymbol{D}_L\boldsymbol{b}\|^2 \leq (1 + \delta_s)\|\boldsymbol{b}\|^2 \tag{3}$$

holds for all real vector $\boldsymbol{b}$ and all subsets L with cardinality at most s. Then we say that $\boldsymbol{D}$ satisfies the RIP with order s.

Remark 1. From (3), the restricted isometry constant δ_s reflects the degree of preservation of the signal's 2-norm, with $\delta_s = 0$ being exactly preserved. The condition (3) is equivalent to that all eigenvalues of the matrix $\boldsymbol{D}_L^T\boldsymbol{D}_L$ lie in $[1 - \delta_s, 1 + \delta_s]$. Moreover, from Definition 1, we can see that the s-restricted isometry constant δ_s increases with s.

How to construct the matrix $\boldsymbol{D}$ satisfying the RIP attracts much attention of researchers in the fields of information theory and signal processing. For example, DeVore in [27] presented a deterministic construction method of sensing matrix using the mutual incoherence. Li et al. in [28] proposed a construction method of the binary sensing matrix via algebraic curves over finite fields. Xu et al. constructed a class of special sensing matrices consisting of partial Fourier matrices in [29]. There is also some literature focusing on the construction of random sensing matrices (see e.g., [26,30,31]). In the simulation example given in Section 5, we use Gaussian matrix $\boldsymbol{D} \in \mathbb{R}^{d\times m}$ whose entries are independent realizations of Gaussian random variables with zero mean and variance $1/d$. We have the following result about the matrix $\boldsymbol{D}$.

Lemma 1 (*[30]*). *For given d, m, and $0 < \delta < 1$, if the sensing matrix $\boldsymbol{D} \in \mathbb{R}^{d\times m}$ is a Gaussian or Bernoulli random matrix, then there exist positive constants c_1, c_2 depending only on δ such that the RIP (3) holds for $\boldsymbol{D}$ with the prescribed δ and any $s \le c_1 d/\log(m/s)$ with probability no less than $1 - e^{-c_2 d}$.*

Remark 2. In [30], Baraniuk et al. show that the constants c_1 and c_2 satisfies the inequality $c_2 \le \frac{\delta^2}{16} - \frac{\delta^3}{48} - c_1\left[1 + \frac{1+\log(12/\delta)}{\log(m/s)}\right]$ and the constant c_1 can be small enough to ensure $c_2 > 0$. For example, if $c_1 = \frac{\delta^3}{120}$, then we can obtain that the sensing $\boldsymbol{D}$ satisfies the RIP with probability no less than $1 - e^{-c_2 d}$ if $d \ge 120 s \log(m/s)/\delta^3$.

After constructing the sensing matrix, the recovery problem of the sparse signal $\boldsymbol{x}$ can be transformed into solving the following convex optimization problem,

$$\min_{\boldsymbol{x}'\in\mathbb{R}^m} \|\boldsymbol{x}'\|_1, \quad s.t. \quad \|\boldsymbol{D}\boldsymbol{x}' - \boldsymbol{z}\| \le C. \tag{4}$$

For the signals obtained by solving the above convex optimization problem, Candès, Romberg and Tao in [19] established the following lemma on the upper bound of the error between the recovered signals and original signals.

Lemma 2 (*[19]*). *Let s satisfy $\delta_{3s} + 3\delta_{4s} < 2$. Then for any s-sparse signal $\boldsymbol{x}$ and any perturbation $\boldsymbol{\epsilon}$ with $\|\boldsymbol{\epsilon}\| \le C$, the recovered signal $\boldsymbol{x}^*$ obtained by solving the optimization problem (4) obeys*

$$\|\boldsymbol{x}^* - \boldsymbol{x}\| \le C_s C,$$

where the positive constant C_s can be taken as

$$C_s \triangleq \frac{4}{\sqrt{3(1-\delta_{4s})} - \sqrt{1+\delta_{3s}}}.$$

Remark 3. Note that for s satisfying the condition of Lemma 2, we can get the true value of the sparse signal by the reconstruction process if there are no measurement perturbations in (2).

2.3. Graph theory

We consider a sensor network with n sensors. The communication between sensors are usually modeled as an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, 3, \dots, n\}$ is the set of sensors (or nodes), $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set, and $\mathcal{A} = \{a_{ij}\} \in \mathbb{R}^{n\times n}$ is the weighted adjacency matrix. The adjacency matrix $\mathcal{A} = \{a_{ij}\}$ is defined as: $a_{ij} > 0$ if $(i,j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. For the sensor i, the set of its neighbors is denoted as $N_i = \{j \in \mathcal{V} | (i,j) \in \mathcal{E}\}$, and we assume that sensor i belongs to N_i. Each sensor can only exchange information with its neighbors. A path of length ℓ is a sequence of nodes $\{i_1, \dots, i_\ell, i_{\ell+1}\}$ such that $(i_h, i_{h+1}) \in \mathcal{E}$ with $1 \le h \le \ell$. The graph $\mathcal{G}$ is called connected if there is a path between any two nodes. The diameter $D_{\mathcal{G}}$ of the graph $\mathcal{G}$ is defined as the maximum shortest path length between any two sensors. If all the elements of a matrix $\mathcal{A} = \{a_{ij}\} \in \mathbb{R}^{n\times n}$ are nonnegative, then it is called a nonnegative matrix, and furthermore if $\sum_{j=1}^n a_{ij} = 1$ holds for all $i \in \{1, \dots, n\}$, then it is a stochastic matrix. For simplicity of analysis, the performance analysis of the distributed algorithm proposed in this paper is considered under the condition that the weighted adjacency matrix $\mathcal{A}$ is symmetric and stochastic.

3. Compressed distributed LS algorithm

3.1. Problem statement

We consider a network consisting of n sensors labeled $1, \dots, n$. For each sensor i, the observation is assumed to obey the following discrete-time random regression model,

$$y_{t+1,i} = \boldsymbol{\varphi}_{t,i}^T \boldsymbol{\theta} + w_{t+1,i}, \quad t = 0, 1, 2, \dots, \tag{5}$$

where $y_{t,i}$ is the scalar observation of the sensor i at time t, $\boldsymbol{\theta} \in \mathbb{R}^m$ is an unknown s-sparse parameter vector that needs to be estimated (i.e., $\boldsymbol{\theta}$ has at most s non-zero elements), $\boldsymbol{\varphi}_{t,i} \in \mathbb{R}^m$ is $3s$-sparse random regression vector, and $\{w_{t,i}\}$ is a measurement noise sequence. The above system model (5) includes many parameterized systems, such as ARX system and Hammerstein system.

This paper aims at designing a distributed algorithm to identify the sparse parameter $\boldsymbol{\theta}$ by using the observation signals and sparse regression vectors from its neighbors, and establishing the performance analysis of the proposed algorithm.

3.2. Design of the algorithm

We know that for the general random regression vectors $\{\boldsymbol{\varphi}_{t,i}\}$ without sparsity, the distributed LS algorithm has been widely used to estimate the unknown parameter vector $\boldsymbol{\theta}$ due to the fast convergence rate. Xie et al. in [24] introduced a cooperative excitation condition to guarantee the convergence of the distributed LS algorithm. However, for some cases such as high dimensional data classification (e.g., [32]), the regression vectors $\{\boldsymbol{\varphi}_{t,i}\}$ may be sparse (i.e., many elements of $\boldsymbol{\varphi}_{t,i}$ are zero). Under such a situation, the cooperative excitation condition mentioned in [24] is hard to be satisfied. Hence, the classical distributed LS algorithm is not able to accurately estimate the high-dimensional sparse signal $\boldsymbol{\theta}$. In addition, the diffusion of high-dimensional regression vectors $\{\boldsymbol{\varphi}_{t,i}\}$ over sensor networks is required in the classical distributed LS algorithm, which may cause high computation cost and communication pressure.

Now, we propose the compressed distributed LS algorithm based on the CS theory and distributed LS algorithm. To be specific, each sensor i at the time instant t can receive m-dimensional sparse regression vectors $\{\boldsymbol{\varphi}_{t,j}, j \in N_i\}$. Using the sensing matrix $\boldsymbol{D} \in \mathbb{R}^{d\times m} (d \ll m)$,[1] we can obtain the compressed d-dimensional regression vectors $\{\boldsymbol{\phi}_{t,j}, j \in N_i\}$ by

$$\boldsymbol{\phi}_{t,j} = \boldsymbol{D}\boldsymbol{\varphi}_{t,j}. \tag{6}$$

Then, the distributed LS algorithm is used to estimate the d-dimensional parameter

$$\boldsymbol{\vartheta} = \boldsymbol{D}\boldsymbol{\theta} \tag{7}$$

in the compressed space. Finally, the estimate $\boldsymbol{\theta}_{t+1,i}$ for the unknown parameter vector $\boldsymbol{\theta}$ is recovered by solving the convex optimization problem (4). The details of the compressed distributed LS algorithm are illustrated in Algorithm 1.

The positive constant C in Step 3 can be actually taken as the upper bound of the estimation error $\|\widetilde{\boldsymbol{\vartheta}}_{t,i}\|$, and the explicit value of C can be found in Theorem 1 of Section 4.

Remark 4. For the reconstruction process (Step 3) in the above Algorithm 1, there are some algorithms, e.g., orthogonal matching pursuit (OMP), compressive sampling matching pursuit (CoSaMP), and interior-point (IP) algorithms to solve the convex optimization problem (11) in the literature, see e.g., [33,34].

3.3. Assumptions

In order to analyze the performance of the compressed distributed LS algorithm, we need to introduce some assumptions on the observation noise, the network topology, the sensing matrix and the regression vectors.

[1] $d \ll m$ means that d is far smaller than m.

Algorithm 1 Compressed distributed LS algorithm

Input: $\{\boldsymbol{\varphi}_{t,i}, y_{t+1,i}\}_{i=1}^n$, $t=0,1,2,\cdots$
Output: $\{\boldsymbol{\theta}_{t+1,i}\}_{i=1}^n$, $t=0,1,2,\cdots$
for every sensor $i=1,\cdots,n$ **do**
Initialization: Begin with an initial vector $\boldsymbol{\vartheta}_{0,i}$ and an initial positive definite matrix $\boldsymbol{P}_{0,i}>0$.
for each time $t=0,1,2,\cdots$ **do**
Step 1. Compression: $\boldsymbol{\phi}_{t,i}=\boldsymbol{D}\boldsymbol{\varphi}_{t,i}$.
Step 2. Estimation in a low-dimensional dimension.
(i) Adaption.

$$\begin{aligned}\bar{\boldsymbol{\vartheta}}_{t+1,i} &= \boldsymbol{\vartheta}_{t,i}+d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\phi}_{t,i}(y_{t+1,i}-\boldsymbol{\phi}_{t,i}^T\boldsymbol{\vartheta}_{t,i}),\\ \bar{\boldsymbol{P}}_{t+1,i} &= \boldsymbol{P}_{t,i}-d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\phi}_{t,i}\boldsymbol{\phi}_{t,i}^T\boldsymbol{P}_{t,i},\\ d_{t,i} &= (1+\boldsymbol{\phi}_{t,i}^T\boldsymbol{P}_{t,i}\boldsymbol{\phi}_{t,i})^{-1},\end{aligned} \tag{8}$$

(ii) Diffusion.

$$\boldsymbol{P}_{t+1,i}^{-1}=\sum_{j\in N_i}a_{ij}\bar{\boldsymbol{P}}_{t+1,j}^{-1}, \tag{9}$$

$$\boldsymbol{\vartheta}_{t+1,i}=\boldsymbol{P}_{t+1,i}\sum_{j\in N_i}a_{ij}\bar{\boldsymbol{P}}_{t+1,j}^{-1}\bar{\boldsymbol{\vartheta}}_{t+1,j}. \tag{10}$$

Step 3. Restruction:

$$\boldsymbol{\theta}_{t+1,i}=\operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{B}}\|\boldsymbol{\beta}\|_1 \tag{11}$$

where $\mathbb{B}=\left\{\boldsymbol{\beta}\in\mathbb{R}^m \,\middle|\, \|\boldsymbol{D}\boldsymbol{\beta}-\boldsymbol{\vartheta}_{t+1,i}\|\le C\right\}$

Assumption 1. For any $i\in\{1,\dots,n\}$, the noise sequence $\{w_{t,i},\mathscr{F}_t\}$ is a martingale difference sequence, and there exists a constant $\gamma>2$, such that

$$\sup_{t\ge 0}E(|w_{t+1,i}|^\gamma|\mathscr{F}_t)<\infty,\quad \text{a.s.}$$

where $\mathscr{F}_t=\sigma\{\boldsymbol{\phi}_{k,i},w_{k,i},k\le t,i=1,\dots,n\}$ is a sequence of non-decreasing σ-algebras and $E[\cdot|\cdot]$ denotes the conditional expectation operator.

We can verify that the i.i.d. (independent and identically distributed) zero-mean bounded or Gaussian noise $\{w_{t,i}\}$ which are independent of the regressors can satisfy Assumption 1.

Assumption 2. The undirected communication graph $\mathscr{G}$ is connected and the weighted adjacency matrix $\mathcal{A}$ is symmetric and stochastic.

Remark 5. In Assumption 2, we require that the matrix $\mathcal{A}$ is symmetric and stochastic which is used to study the nonnegative definite properties of $\boldsymbol{\Delta}_{t+1}$ (see Lemma 3). For the directed graph case, if $\mathscr{G}$ is strongly connected and balanced, then Lemma 3 still holds by revising $\boldsymbol{\Delta}_{t+1}$ as $\boldsymbol{\Delta}'_{t+1}\triangleq\bar{\boldsymbol{P}}_{t+1}-(\mathcal{A}^T\otimes\boldsymbol{I}_d)\boldsymbol{P}_{t+1}(\mathcal{A}\otimes\boldsymbol{I}_d)$. Thus, we can obtain similar theoretical results by following the proof line of Theorems 1 and 2.

Assumption 3. The sensing matrix $\boldsymbol{D}\in\mathbb{R}^{d\times m}$ satisfies the RIP with order $4s$ where the $3s$- and $4s$-restricted isometry constants denoted as δ_{3s} and δ_{4s} (see Definition 1) satisfy $\delta_{3s}+3\delta_{4s}<2$.

Remark 6. The above properties (RIP) of the sensing matrix $\boldsymbol{D}$ which is often used in the compressive sensing theory can guarantee that the sparse signals can be recovered with high accuracy (see Lemma 2).

Assumption 4 (*Compressed Cooperative Persistent Excitation Condition*)**.** There exists a positive constant M such that for the adapted sequence $\{\boldsymbol{\phi}_{t,i},\mathscr{F}_t,t\ge 0\}$,

$$\lambda_{\min}^{n,t}\xrightarrow[t\to\infty]{}\infty,\quad \sup_{t\ge 0}\frac{r_t}{\lambda_{\min}^{n,t}}\le M,$$

where $r_t\triangleq\lambda_{\max}(\sum_{i=1}^n\boldsymbol{P}_{0,i}^{-1})+\sum_{i=1}^n\sum_{k=0}^t\|\boldsymbol{\phi}_{k,i}\|^2$, and

$$\lambda_{\min}^{n,t}\triangleq\lambda_{\min}\left(\sum_{i=1}^n\boldsymbol{P}_{0,i}^{-1}+\sum_{i=1}^n\sum_{k=0}^{t-D_{\mathscr{G}}+1}\boldsymbol{\phi}_{k,i}\boldsymbol{\phi}_{k,i}^T\right).$$

Remark 7. We provide an intuitive illustration for the above excitation condition of the compressed signals. Consider an extreme case where all regression vectors $\{\boldsymbol{\varphi}_{k,i}\}$ are equal to zero, it is clear that the unknown sparse parameter $\boldsymbol{\theta}$ cannot be estimated since the measurement signal $y_{t,i}$ does not contain any information about $\boldsymbol{\theta}$. In order to estimate $\boldsymbol{\theta}$, we need to impose some excitation conditions on the regression vectors. The following persistent excitation condition for the single sensor case is commonly used for the convergence analysis of LS algorithm in the literature (see e.g., [35,36]),

$$\sup_{t\ge 0}\frac{\lambda_{\max}\left(\sum_{k=1}^t\boldsymbol{\varphi}_{k,i}\boldsymbol{\varphi}_{k,i}^T\right)}{\lambda_{\min}\left(\sum_{k=1}^t\boldsymbol{\varphi}_{k,i}\boldsymbol{\varphi}_{k,i}^T\right)}<\infty. \tag{12}$$

However, it is difficult for the sparse regression vectors $\{\boldsymbol{\varphi}_{k,i}\}$ to satisfy the above excitation condition (see the simulation example in Section 5). Hence, we propose the compressed cooperative persistent excitation condition (Assumption 4) where $\boldsymbol{\varphi}_{k,i}$ is replaced by the compressed signal $\boldsymbol{\phi}_{k,i}$.

4. Performance analysis of the algorithm

Now, we will provide the performance analysis for the compressed distributed LS algorithm. By (5), we have

$$\begin{aligned}y_{t+1,i} &= \boldsymbol{\phi}_{t,i}^T\boldsymbol{\vartheta}+\boldsymbol{\varphi}_{t,i}^T\boldsymbol{\theta}-\boldsymbol{\phi}_{t,i}^T\boldsymbol{\vartheta}+w_{t+1,i}\\ &= \boldsymbol{\phi}_{t,i}^T\boldsymbol{\vartheta}+\boldsymbol{\varphi}_{t,i}^T(\boldsymbol{I}_m-\boldsymbol{D}^T\boldsymbol{D})\boldsymbol{\theta}+w_{t+1,i}.\end{aligned} \tag{13}$$

Set

$$\bar{w}_{t+1,i}\triangleq\xi_{t,i}+w_{t+1,i},\quad \xi_{t,i}=\boldsymbol{\varphi}_{t,i}^T(\boldsymbol{I}_m-\boldsymbol{D}^T\boldsymbol{D})\boldsymbol{\theta}, \tag{14}$$

where $\bar{w}_{t+1,i}$ can be regarded as the“new" noise, and $\xi_{t,i}$ is called the sensing deviation. Then by (14), the dynamical system (13) can be rewritten as

$$y_{t+1,i}=\boldsymbol{\phi}_{t,i}^T\boldsymbol{\vartheta}+\bar{w}_{t+1,i}. \tag{15}$$

Denote the compressed estimation error as

$$\widetilde{\boldsymbol{\vartheta}}_{t,i}=\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_{t,i}. \tag{16}$$

Then by (8) and (15), we have

$$\begin{aligned}\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}}_{t+1,i} &= \boldsymbol{\vartheta}-\boldsymbol{\vartheta}_{t,i}-d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\phi}_{t,i}(y_{t+1,i}-\boldsymbol{\phi}_{t,i}^T\boldsymbol{\vartheta}_{t,i})\\ &= (\boldsymbol{I}_d-d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\phi}_{t,i}\boldsymbol{\phi}_{t,i}^T)\widetilde{\boldsymbol{\vartheta}}_{t,i}-d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\phi}_{t,i}\bar{w}_{t+1,i}\\ &= \bar{\boldsymbol{P}}_{t+1,i}\boldsymbol{P}_{t,i}^{-1}\widetilde{\boldsymbol{\vartheta}}_{t,i}-d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\phi}_{t,i}\bar{w}_{t+1,i}.\end{aligned}$$

Hence by (9) and 10, the error $\widetilde{\boldsymbol{\vartheta}}_{t,i}$ in the compressed space evolves according to the following equation

$$\begin{aligned}\widetilde{\boldsymbol{\vartheta}}_{t+1,i} &= \boldsymbol{P}_{t+1,i}\sum_{j\in N_i}a_{ij}\boldsymbol{P}_{t,j}^{-1}\widetilde{\boldsymbol{\vartheta}}_{t,j}\\ &\quad -\boldsymbol{P}_{t+1,i}\sum_{j\in N_i}a_{ij}\bar{\boldsymbol{P}}_{t+1,j}^{-1}d_{t,j}\boldsymbol{P}_{t,j}\boldsymbol{\phi}_{t,j}\bar{w}_{t+1,j}.\end{aligned} \tag{17}$$

Denote $\boldsymbol{Y}_t = col\{y_{t,1}, \ldots, y_{t,n}\}$ and

$$\begin{aligned}&\boldsymbol{W}_t = col\{w_{t,1}, \ldots, w_{t,n}\}, \quad \boldsymbol{\Xi}_t = col\{\xi_{t,1}, \ldots, \xi_{t,n}\},\\ &\overline{\boldsymbol{W}}_t = col\{\bar{w}_{t,1}, \ldots, \bar{w}_{t,n}\}, \quad \widetilde{\boldsymbol{Z}}_t = col\{\widetilde{\boldsymbol{\vartheta}}_{t,1}, \ldots, \widetilde{\boldsymbol{\vartheta}}_{t,n}\},\\ &\boldsymbol{d}_t = diag\{d_{t,1}, \ldots, d_{t,n}\}, \quad \boldsymbol{\Psi}_t = diag\{\boldsymbol{\phi}_{t,1}, \ldots, \boldsymbol{\phi}_{t,n}\},\\ &\bar{\boldsymbol{P}}_t = diag\{\bar{\boldsymbol{P}}_{t,1}, \ldots, \bar{\boldsymbol{P}}_{t,n}\}, \quad \boldsymbol{P}_t = diag\{\boldsymbol{P}_{t,1}, \ldots, \boldsymbol{P}_{t,n}\},\end{aligned}$$

where $col(\cdot, \ldots, \cdot)$ denotes a vector stacked by the specified vectors, and $diag(\cdot, \ldots, \cdot)$ denotes a block matrix formed in a diagonal manner of the corresponding vectors or matrices. Then dynamical system (15) and error Eq. (17) can be written as the following matrix forms,

$$\begin{aligned}\boldsymbol{Y}_{t+1} &= \boldsymbol{\Psi}_t^T \boldsymbol{Z} + \overline{\boldsymbol{W}}_{t+1},\\ \widetilde{\boldsymbol{Z}}_{t+1} &= \boldsymbol{P}_{t+1}(\mathcal{A}\otimes \boldsymbol{I}_d)\boldsymbol{P}_t^{-1}\widetilde{\boldsymbol{Z}}_t\\ &\quad - \boldsymbol{P}_{t+1}(\mathcal{A}\otimes \boldsymbol{I}_d)\bar{\boldsymbol{P}}_{t+1}^{-1}(\boldsymbol{d}_t\otimes \boldsymbol{I}_d)\boldsymbol{P}_t\boldsymbol{\Psi}_t\overline{\boldsymbol{W}}_{t+1}.\end{aligned} \tag{18}$$

We will first consider the upper bound of the compressed estimation error $\widetilde{\boldsymbol{Z}}_{t+1}$. For this, we introduce function $V_t = \widetilde{\boldsymbol{Z}}_t^T\boldsymbol{P}_t^{-1}\widetilde{\boldsymbol{Z}}_t$. By following the proof line of [24], we can obtain a preliminary result of V_t.

Lemma 3. *Under Assumption 2, we have the following inequality,*

$$\begin{aligned}&V_{t+1} + \sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{\Psi}_k\boldsymbol{d}_k\boldsymbol{\Psi}_k^T\widetilde{\boldsymbol{Z}}_k + \sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{P}_k^{-1}\boldsymbol{\Delta}_{k+1}\boldsymbol{P}_k^{-1}\widetilde{\boldsymbol{Z}}_k\\ &\le V_0 - 2\sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{P}_k^{-1}\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Psi}_k\overline{\boldsymbol{W}}_{k+1}\\ &\quad + 2\sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{P}_k^{-1}\boldsymbol{\Delta}_{k+1}\boldsymbol{\Psi}_k\overline{\boldsymbol{W}}_{k+1}\\ &\quad + \sum_{k=0}^t \overline{\boldsymbol{W}}_{k+1}^T\boldsymbol{d}_k\boldsymbol{\Psi}_k^T\boldsymbol{P}_k\boldsymbol{\Psi}_k\overline{\boldsymbol{W}}_{k+1},\end{aligned} \tag{19}$$

where $\boldsymbol{\Delta}_{t+1} \triangleq \bar{\boldsymbol{P}}_{t+1} - (\mathcal{A}\otimes \boldsymbol{I}_d)\boldsymbol{P}_{t+1}(\mathcal{A}\otimes \boldsymbol{I}_d)$ is a nonnegative definite matrix satisfying $\boldsymbol{\Delta}_{t+1} \le \bar{\boldsymbol{P}}_{t+1}$.

In the following, we will analyze the last three terms on the right hand side (RHS) of (19). We note that the estimation step (i.e., Step 2 in Algorithm 1) is obtained by using the compressed regression vector $\boldsymbol{\phi}_{t,i} = \boldsymbol{D}\boldsymbol{\varphi}_{t,i}$ not the original regression vector $\boldsymbol{\varphi}_{t,i}$, which leads to the sensing deviation term $\xi_{t,i}$ defined in (14). Unlike the measurement noise $w_{t+1,i}$, $\xi_{t,i}$ generally has no good statistical properties, such as independency, martingale difference sequence, which makes it hard to use the theoretical results in probability theory and stochastic process in our analysis. We will use the properties of the sensing matrix $\boldsymbol{D}$ to deal with the cumulative effect of sensing deviation $\xi_{t,i}$.

Lemma 4. *Under Assumption 3, the following inequality for the sensing deviation holds*

$$\|\boldsymbol{\Xi}_t\|^2 \le \frac{9\delta_{4s}^2\|\boldsymbol{\theta}\|^2}{1-\delta_{4s}}\sum_{i=1}^n \|\boldsymbol{\phi}_{t,i}\|^2. \tag{20}$$

Proof. Denote the set

$$\Lambda_{t,i} = \left\{l_{t,i}^{(1)}, \ldots, l_{t,i}^{(3s)}, j^{(1)}, \ldots, j^{(s)}\right\},$$

where $l_{t,i}^{(1)}, \ldots, l_{t,i}^{(3s)}$ are the indices of $3s$ nonzero elements of $\boldsymbol{\varphi}_{t,i}$ and $j^{(1)}, \ldots, j^{(s)}$ are the indices of s nonzero elements of $\boldsymbol{\theta}$. The analysis for the case where the cardinality of $\Lambda_{t,i}$ is less than $4s$ (i.e., part of the nonzero elements of the vectors $\boldsymbol{\varphi}_{t,i}$ and $\boldsymbol{\theta}$ are in the same position) is almost the same as that for the case where the set $\Lambda_{t,i}$ has $4s$ elements. Thus, we just consider the latter. The vectors obtained by extracting the $4s$ nonzero elements from $\boldsymbol{\varphi}_{t,i}$ and $\boldsymbol{\theta}$ are denoted as $\boldsymbol{\varphi}_{t,i}^{(4s)}$ and $\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}$, and the indices of these $4s$ elements come from the set $\Lambda_{t,i}$. Correspondingly, we extract the $4s$ columns from the matrix $\boldsymbol{D}$, and denote the new matrix as $\boldsymbol{D}_{t,i}^{(4s)}$.

By Assumption 3 and Remark 1, we see that all eigenvalues of the matrix $(\boldsymbol{D}_{t,i}^{(4s)})^T\boldsymbol{D}_{t,i}^{(4s)}$ lie in the interval $[1-\delta_{4s}, 1+\delta_{4s}]$. Thus, we have

$$\begin{aligned}&|\xi_{t,i}| = \left\|\boldsymbol{\varphi}_{t,i}^T(\boldsymbol{I}_m - \boldsymbol{D}^T\boldsymbol{D})\boldsymbol{\theta}\right\|\\ &= \left\|\left(\boldsymbol{\varphi}_{t,i}^{(4s)}\right)^T\left[\boldsymbol{I}_{4s} - \left(\boldsymbol{D}_{t,i}^{(4s)}\right)^T\boldsymbol{D}_{t,i}^{(4s)}\right]\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}\right\|\\ &\le \left\|\left(\boldsymbol{\varphi}_{t,i}^{(4s)}\right)^T\left[(1+\delta_{4s})\boldsymbol{I}_{4s} - \left(\boldsymbol{D}_{t,i}^{(4s)}\right)^T\boldsymbol{D}_{t,i}^{(4s)}\right]\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}\right\|\\ &\quad + \delta_{4s}\left\|\left(\boldsymbol{\varphi}_{t,i}^{(4s)}\right)^T\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}\right\|\\ &\le \left\|(\boldsymbol{\varphi}_{t,i}^{(4s)})^T\right\|\cdot\left\|(1+\delta_{4s})\boldsymbol{I}_{4s} - \left(\boldsymbol{D}_{t,i}^{(4s)}\right)^T\boldsymbol{D}_{t,i}^{(4s)}\right\|\cdot\left\|\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}\right\|\\ &\quad + \delta_{4s}\left\|\left(\boldsymbol{\varphi}_{t,i}^{(4s)}\right)^T\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}\right\|\\ &\le 2\delta_{4s}\left\|\boldsymbol{\varphi}_{t,i}^{(4s)}\right\|\cdot\left\|\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}\right\| + \delta_{4s}\left\|\boldsymbol{\varphi}_{t,i}^{(4s)}\right\|\cdot\left\|\breve{\boldsymbol{\theta}}_{t,i}^{(4s)}\right\|\\ &= 3\delta_{4s}\left\|\boldsymbol{\varphi}_{t,i}\right\|\cdot\|\boldsymbol{\theta}\| \le \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\boldsymbol{D}\boldsymbol{\varphi}_{t,i}\|\cdot\|\boldsymbol{\theta}\|\\ &= \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\boldsymbol{\phi}_{t,i}\|\|\boldsymbol{\theta}\|.\end{aligned} \tag{21}$$

By $\boldsymbol{\Xi}_t = col\{\xi_{t,1}, \ldots, \xi_{t,n}\}$, we see that the result of the lemma holds. ■

Lemma 5. *Under Assumptions 1–3, the following result holds almost surely,*

$$\widetilde{\boldsymbol{Z}}_{t+1}^T\boldsymbol{P}_{t+1}^{-1}\widetilde{\boldsymbol{Z}}_{t+1} \le O(\log r_t) + \frac{54\delta_{4s}^2}{1-\delta_{4s}}\|\boldsymbol{\theta}\|^2 r_t.$$

Proof. By the definition of $\overline{\boldsymbol{W}}_{t+1}$ and $\boldsymbol{\Xi}_t$, we have $\overline{\boldsymbol{W}}_{t+1} = \boldsymbol{\Xi}_t + \boldsymbol{W}_{t+1}$. Therefore, we have

$$\begin{aligned}&\sum_{k=0}^t \overline{\boldsymbol{W}}_{k+1}^T\boldsymbol{d}_k\boldsymbol{\Psi}_k^T\boldsymbol{P}_k\boldsymbol{\Psi}_k\overline{\boldsymbol{W}}_{k+1}\\ &\le \sum_{k=0}^t \lambda_{\max}(\boldsymbol{d}_k\boldsymbol{\Psi}_k^T\boldsymbol{P}_k\boldsymbol{\Psi}_k)\|\boldsymbol{\Xi}_k + \boldsymbol{W}_{k+1}\|^2\\ &\le 2\sum_{k=0}^t \lambda_{\max}(\boldsymbol{d}_k\boldsymbol{\Psi}_k^T\boldsymbol{P}_k\boldsymbol{\Psi}_k)(\|\boldsymbol{\Xi}_k\|^2 + \|\boldsymbol{W}_{k+1}\|^2).\end{aligned} \tag{22}$$

Substituting (22) into (19), we have

$$\begin{aligned}&V_{t+1} + \sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{\Psi}_k\boldsymbol{d}_k\boldsymbol{\Psi}_k^T\widetilde{\boldsymbol{Z}}_k + \sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{P}_k^{-1}\boldsymbol{\Delta}_{k+1}\boldsymbol{P}_k^{-1}\widetilde{\boldsymbol{Z}}_k\\ &\le V_0 - 2\sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{P}_k^{-1}\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Psi}_k\boldsymbol{W}_{k+1}\\ &\quad + 2\sum_{k=0}^t \widetilde{\boldsymbol{Z}}_k^T\boldsymbol{P}_k^{-1}\boldsymbol{\Delta}_{k+1}\boldsymbol{\Psi}_k\boldsymbol{W}_{k+1}\\ &\quad + 2\sum_{k=0}^t \lambda_{\max}(\boldsymbol{d}_k\boldsymbol{\Psi}_k^T\boldsymbol{P}_k\boldsymbol{\Psi}_k)\|\boldsymbol{W}_{k+1}\|^2\end{aligned}$$

$$
\begin{aligned}
&-2\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\bar{P}_{k+1}\Psi_k\Xi_k \\
&+2\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}\Psi_k\Xi_k \\
&+2\sum_{k=0}^{t}\lambda_{\max}(d_k\Psi_k^T P_k\Psi_k)\|\Xi_k\|^2. \qquad (23)
\end{aligned}
$$

Using the martingale convergence theorem and martingale estimation theorem, we have the following results for the second to the fourth term on the RHS of (23) (see Lemma 4.4 in [24] for details),

$$
\begin{aligned}
&\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\bar{P}_{k+1}\Psi_k W_{k+1} \\
&\quad = O(1)+o\Big(\sum_{k=0}^{t}\widetilde{Z}_k^T\Psi_k d_k\Psi_k^T\widetilde{Z}_k\Big), \qquad (24)\\
&\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}\Psi_k W_{k+1} \\
&\quad = O(1)+o\Big(\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}P_k^{-1}\widetilde{Z}_k\Big), \qquad (25)\\
&\sum_{k=0}^{t}\lambda_{\max}(d_k\Psi_k^T P_k\Psi_k)\|W_{k+1}\|^2 = O(\log r_t). \qquad (26)
\end{aligned}
$$

In the following, we estimate the last three terms on the RHS of (23). By the definition of $d_{k,i}$, we have

$$
d_{k,i}\phi_{k,i}^T P_{k,i}\phi_{k,i} = 1-d_{k,i}. \qquad (27)
$$

Combining this equation with (8), we can deduce that

$$
\begin{aligned}
P_{k,i}^{-1}\bar{P}_{k+1,i}\phi_{k,i} &= \phi_{k,i}-d_{k,i}\phi_{k,i}\phi_{k,i}^T P_{k,i}\phi_{k,i}\\
&= \phi_{k,i}-\phi_{k,i}(1-d_{k,i}) = \phi_{k,i}d_{k,i}.
\end{aligned}
$$

Hence we have

$$
P_k^{-1}\bar{P}_{k+1}\Psi_k = \Psi_k d_k. \qquad (28)
$$

By (28) and Hölder inequality, we see that the fifth term on the RHS of (23) satisfies the following inequality

$$
\begin{aligned}
&\Big\|2\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\bar{P}_{k+1}\Psi_k\Xi_k\Big\| = 2\Big\|\sum_{k=0}^{t}\widetilde{Z}_k^T\Psi_k d_k\Xi_k\Big\|\\
&\le 2\left(\sum_{k=0}^{t}\widetilde{Z}_k^T\Psi_k d_k\Psi_k^T\widetilde{Z}_k\right)^{\frac12}\left(\sum_{k=0}^{t}\|\Xi_k\|^2\right)^{\frac12},
\end{aligned}
$$

where the fact $d_k \le I_n$ is used in the last inequality. Hence by Lemma 4, we have

$$
\begin{aligned}
&\Big\|2\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\bar{P}_{k+1}\Psi_k\Xi_k\Big\|\\
&\le \left(\sum_{k=0}^{t}\widetilde{Z}_k^T\Psi_k d_k\Psi_k^T\widetilde{Z}_k\right)^{\frac12}\\
&\quad\cdot\frac{6\delta_{4s}\|\theta\|}{\sqrt{1-\delta_{4s}}}\left(\sum_{k=0}^{t}\sum_{i=1}^{n}\|\phi_{k,i}\|^2\right)^{\frac12}\\
&\le \frac{\left(\sum_{k=0}^{t}\widetilde{Z}_k^T\Psi_k d_k\Psi_k^T\widetilde{Z}_k\right)}{2}\\
&\quad+\frac{18\delta_{4s}^2\|\theta\|^2}{1-\delta_{4s}}\left(\sum_{k=0}^{t}\sum_{i=1}^{n}\|\phi_{k,i}\|^2\right), \qquad (29)
\end{aligned}
$$

where the inequality $xy \le \frac{x^2+y^2}{2}$ is used.

From (28), we have

$$
\begin{aligned}
&\Delta_{k+1}\Psi_k\Psi_k^T\Delta_{k+1}\\
&\le (\lambda_{\max}(\Psi_k^T\Delta_{k+1}\Psi_k))\Delta_{k+1}\\
&\le \lambda_{\max}(\Psi_k^T\bar{P}_{k+1}\Psi_k)\Delta_{k+1}\\
&= \lambda_{\max}(\Psi_k^T P_k\Psi_k d_k)\Delta_{k+1} \le \Delta_{k+1}. \qquad (30)
\end{aligned}
$$

Similar to the analysis of (29), we can deduce that the sixth term on the RHS of (23) satisfies the following inequality,

$$
\begin{aligned}
&\Big\|2\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}\Psi_k\Xi_k\Big\|\\
&\le \Big\|\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}\Psi_k\Psi_k^T\Delta_{k+1}P_k^{-1}\widetilde{Z}_k\Big\|^{\frac12}\\
&\quad\cdot 2\left(\sum_{k=0}^{t}\Big\|\Xi_k\Big\|^2\right)^{\frac12}\\
&\le \Big\|\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}P_k^{-1}\widetilde{Z}_k\Big\|^{\frac12}\\
&\quad\cdot\frac{6\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\theta\|\left(\sum_{k=0}^{t}\sum_{i=1}^{n}\|\phi_{k,i}\|^2\right)^{\frac12}\\
&\le \frac{\Big\|\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}P_k^{-1}\widetilde{Z}_k\Big\|}{2}\\
&+\frac{18\delta_{4s}^2}{1-\delta_{4s}}\|\theta\|^2\left(\sum_{k=0}^{t}\sum_{i=1}^{n}\|\phi_{k,i}\|^2\right), \qquad (31)
\end{aligned}
$$

where the inequality (30) is used in the second inequality.

For the last term on the RHS of (23), we have the following inequalities

$$
\begin{aligned}
&2\sum_{k=0}^{t}\lambda_{\max}(d_k\Psi_k^T P_k\Psi_k)\|\Xi_k\|^2 \le 2\left(\sum_{k=0}^{t}\|\Xi_k\|^2\right)\\
&\le \frac{18\delta_{4s}^2}{1-\delta_{4s}}\|\theta\|^2\left(\sum_{k=0}^{t}\sum_{i=1}^{n}\|\phi_{k,i}\|^2\right), \qquad (32)
\end{aligned}
$$

where (20) and the fact $\lambda_{\max}(d_k\Psi_k^T P_k\Psi_k) \le 1$ are used.

Substituting (24)–(26), (29), (31) and (32) into (23) yields

$$
\begin{aligned}
&V_{t+1}+\left(\frac12+o(1)\right)\sum_{k=0}^{t}\widetilde{Z}_k^T\Psi_k d_k\Psi_k^T\widetilde{Z}_k\\
&+\left(\frac12+o(1)\right)\sum_{k=0}^{t}\widetilde{Z}_k^T P_k^{-1}\Delta_{k+1}P_k^{-1}\widetilde{Z}_k\\
&\le O(\log r_t)+\frac{54\delta_{4s}^2}{1-\delta_{4s}}\|\theta\|^2\left(\sum_{k=0}^{t}\sum_{i=1}^{n}\|\phi_{k,i}\|^2\right)\\
&\le O(\log r_t)+\frac{54\delta_{4s}^2}{1-\delta_{4s}}\|\theta\|^2 r_t, \qquad (33)
\end{aligned}
$$

which completes the proof of the lemma. ■

In fact, if the regression vectors $\varphi_{t,i}$ are bounded, then by following the proof line of the above lemma, we can obtain a certain non-asymptotic performance of the estimation error as shown for the classical LS algorithm in [37] and [38].

According to the above lemma, we can easily get the following result on the upper bound of the estimation error $\widetilde{\boldsymbol{Z}}_{t+1}$.

Theorem 1. *Suppose that Assumptions 1–4 hold. The compressed estimation error $\widetilde{\boldsymbol{Z}}_{t+1}$ has the following upper bound,*

$$\|\widetilde{\boldsymbol{Z}}_{t+1}\|^2 = O\left(\frac{C_s^2\delta_{4s}^2}{1-\delta_{4s}}\right). \quad \text{a.s.} \tag{34}$$

Proof. By the matrix inversion formula (1) and (8), we have $\bar{\boldsymbol{P}}_{t+1,i}^{-1} = \boldsymbol{P}_{t,i}^{-1} + \boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T$. Then by (9), we have the following equation,

$$\boldsymbol{P}_{t+1,i}^{-1} = \sum_{j=1}^{n}\sum_{k=0}^{t} a_{ij}^{(t+1-k)}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T + \sum_{j=1}^{n} a_{ij}^{(t+1)}\boldsymbol{P}_{0,j}^{-1},$$

where $a_{ij}^{(t)}$ is the ith row, jth column entry of the weight matrix $\mathcal{A}^t$, $t \geq 1$. Note that by Assumption 2, we see that $a_{\min} \triangleq \min_{i,j\in\mathcal{V}} a_{ij}^{(D_\mathcal{G})} > 0$ holds for $t \geq D_\mathcal{G}$. Thus, we have

$$\lambda_{\min}(\boldsymbol{P}_{t+1,i}^{-1}) \geq a_{\min}\lambda_{\min}^{n,t}. \tag{35}$$

By (35) and Lemma 5, we have

$$\|\widetilde{\boldsymbol{Z}}_{t+1}\|^2 \leq O\left(\frac{\log r_t}{\lambda_{\min}^{n,t}}\right) + \frac{54\delta_{4s}^2\|\boldsymbol{\theta}\|^2}{1-\delta_{4s}}\frac{r_t}{\lambda_{\min}^{n,t}}. \quad \text{a.s.} \tag{36}$$

By (36) and Assumption 4, the result of the theorem can be obtained. ■

Remark 8. By (36), we see that the upper bound of the estimation error $\widetilde{\boldsymbol{Z}}_{t+1}$ consists of two parts: the first part is mainly caused by the measurement noise $w_{t,i}$, and the second part is mainly concerned with the sensing deviation $\xi_{t,i}$ defined in (14). Under Compressed Cooperative Persistent Excitation Condition (Assumption 4), the first part tends to zero as $t \to \infty$, and the second part tends to zero as the 4s-restricted isometry constant δ_{4s} goes to zero. Furthermore, by Remark 2, we can see that the δ_{4s} can be arbitrarily small when the dimension of the sensing matrix $\boldsymbol{D}$ satisfies the inequality $d \geq 480s\log(m/4s)/\delta_{4s}^3$. How to relax Assumption 4 to non-persistent excitation condition (e.g., [39]) requires new techniques to deal with the sensing deviation, which falls into our future work.

Theorem 2. *Under Assumptions 1–4, we have the following upper bound for the estimation error of the compressed distributed LS algorithm for any $i \in \{1,\ldots,n\}$,*

$$\|\widetilde{\boldsymbol{\theta}}_{t+1,i}\|^2 = O\left(\frac{C_s^2\delta_{4s}^2}{1-\delta_{4s}}\right), \quad \text{a.s.,}$$

where $\widetilde{\boldsymbol{\theta}}_{t+1,i} = \boldsymbol{\theta} - \boldsymbol{\theta}_{t+1,i}$ and $C_s = \frac{4}{\sqrt{3(1-\delta_{4s})}-\sqrt{1+\delta_{3s}}}$ defined in Lemma 2.

Proof. Note that $\widetilde{\boldsymbol{\vartheta}}_{t+1,i} = \boldsymbol{D}\boldsymbol{\theta} - \boldsymbol{\vartheta}_{t+1,i}$. According to Theorem 1, the constant C in Algorithm 1 can be taken as $C = \sqrt{\frac{55M\delta_{4s}^2\|\boldsymbol{\theta}\|^2}{1-\delta_{4s}}}$ with M being defined in Assumption 4. Furthermore, by Assumption 4 and Lemma 2, we have for large t

$$\|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}\| \leq C_s\sqrt{\frac{55M\delta_{4s}^2\|\boldsymbol{\theta}\|^2}{1-\delta_{4s}}} \quad \text{a.s.} \tag{37}$$

This completes the proof of the theorem. ■

Remark 9. By the definition of RIP (Definition 1), we have $\delta_{3s} \leq \delta_{4s}$. It is clear that by (37), the estimation error $\widetilde{\boldsymbol{\theta}}_{t,i}$ of the compressed DLS algorithm goes to zero as δ_{4s} tends to zero.

Remark 10. Compared with [21], our results for the performance analysis of the compressed distributed LS algorithm in this paper (see Theorems 1 and 2) are derived without using any independency or stationarity assumptions on the regression vectors. It is clear that our results are more applicable to practical feedback systems.

5. A simulation example

In this section, we provide an example to illustrate the performance of the compressed distributed LS algorithm proposed in this paper.

Example 1. Consider a network composed of $n = 20$ sensors whose dynamics obey the following model

$$y_{t+1,i} = \boldsymbol{\varphi}_{t,i}^T\boldsymbol{\theta} + w_{t+1,i} \tag{38}$$

with the dimension $m = 300$. The noise sequence $\{w_{t,i}, t \geq 1, i = 1,\ldots,n\}$ in (5) is independent and identically distributed with $w_{t,i} \sim \mathcal{N}(0, 0.04)$ (Gaussian distribution with zero mean and variance 0.04). The regression vectors $\{\boldsymbol{\varphi}_{t,i} \in \mathbb{R}^{300}, i = 1,\ldots,20, t \geq 1\}$ are generated according to the following expression,

$$\boldsymbol{\varphi}_{t,i} = \Big[0,\ldots,0, \underbrace{1.1^t + \sum_{k=0}^{t-1} 1.1^k \varepsilon_{t-k,i}}_{i^{th}}, 0,\ldots,0\Big]^T, \tag{39}$$

where the noise sequences $\{\varepsilon_{t,i}, i = 1,\ldots,20, t \geq 1\}$ in (39) are independent and uniformly distributed in $(0, 0.1)$. All sensors will estimate an unknown 3-sparse parameter $\boldsymbol{\theta} = [\underbrace{0,\ldots,0}_{297}, 2.4, 3.5, 4.6]^T$.

The initial estimate is taken as $\boldsymbol{\theta}_{0,i} = [\underbrace{0.8,\ldots,0.8}_{300}]^T$ for $i = 1, 2,\ldots,20$. The network structure is shown in Fig. 1. Here we use the Metropolis rule in [40] to construct the weights, i.e.,

$$a_{li} = \begin{cases} 1 - \sum_{j\neq i} a_{ij} & \text{if } l = i \\ 1/(\max\{n_i, n_l\}) & \text{if } l \in N_i \setminus \{i\} \end{cases}, \tag{40}$$

where n_i is the degree of the node i.

From the definition of the noise sequence $\{w_{t,i}\}$, we see that Assumption 1 holds. By (40) and the structure of the network topology in Fig. 1, Assumption 2 can be satisfied. We estimate the unknown sparse parameter $\boldsymbol{\theta}$ by using the compressed distributed LS algorithm, the distributed LS algorithm in [24] and regularized distributed LS algorithm with l_1 penalty (where the construction of the regularized parameter is chosen similar to that in [41]), respectively. In the compressed distributed LS algorithm, the sensing matrix $\boldsymbol{D}$ is selected as a 15×300-dimensional random matrix whose elements are Gaussian random variables with zero mean and variance $1/15$. Baraniuk et al. in [30] showed that such a sensing matrix can satisfy RIP condition (3) with probability no less than $1 - 2e^{-d\bar{c}}$ for some positive constant $\bar{c}$ (see Lemma 1). For the decompression procedure (i.e., Step 3 in Algorithm 1), the OMP algorithm is used to solve the optimization problem (11). It can be shown that the compressed regression vector $\boldsymbol{\phi}_{t,i} = \boldsymbol{D}\boldsymbol{\varphi}_{t,i}$ can satisfy the compressed cooperative persistent excitation condition (i.e., Assumption 4), while the original

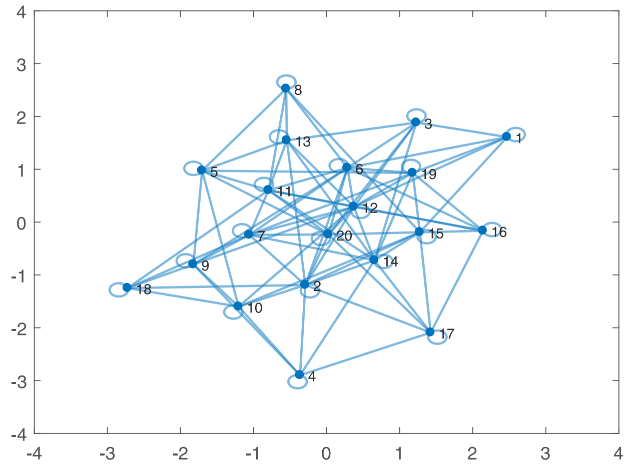

Fig. 1. Network topology of 20 sensors.

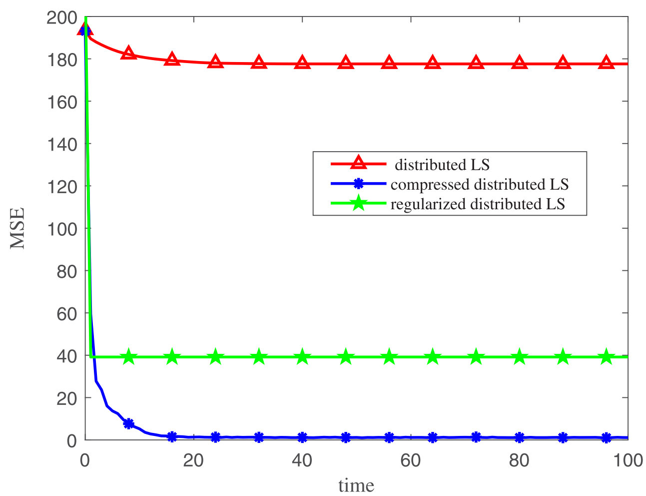

Fig. 2. The MSEs of the compressed distributed LS algorithm, the distributed LS algorithm in [24], and the regularized distributed LS with l_1 penalty.

regression vector $\varphi_{t,i}$ cannot satisfy the cooperative excitation condition used in [24].

We repeat the simulation for 200 times with the same initial value, and the simulation results are shown in Fig. 2. We can see that the mean square error (MSE) of our compressed distributed LS algorithm is much smaller than that of the distributed LS algorithm in [24], and is also smaller than that of the regularized distributed LS algorithm with l_1 penalty, which means that for the sparse regression vectors, the compressed distributed LS algorithm has better estimation performance than the non-compressed distributed LS algorithms.

6. Concluding remarks

This paper proposes a compressed distributed algorithm to estimate unknown high-dimensional sparse signals based on the distributed LS algorithm and the CS theory. We provide an upper bound for the estimation error of compressed signals under the compressed cooperative persistent excitation condition, and further establish the upper bound for the estimation error of the original high-dimensional sparse signals. We give a simulation result to show that the compressed distributed LS algorithm proposed in this paper can estimate the high-dimensional sparse signals while the non-compressed distributed algorithms may not. Many interesting problems deserve to be further investigated, for example, the relaxation of the compressed cooperative persistent excitation condition, the optimization of the sensing matrix and the combination of distributed adaptive identification and control (e.g., [39,42]).

CRediT authorship contribution statement

Die Gan: Methodology, Writing – original draft. **Zhixin Liu:** Supervision, Conceptualization, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to express our sincere thanks to Prof. Lei Guo for his discussions and valuable suggestions.

References

[1] A.H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, Z.J. Towfic, Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior, IEEE Signal Process. Mag. 30 (3) (2013) 155–171.
[2] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Comput. Netw. 52 (12) (2008) 2292–2330.
[3] I.D. Schizas, G.B. Giannakis, Z. Luo, Distributed estimation using reduced-dimensionality sensor observations, IEEE Trans. Signal Process. 55 (8) (2007) 4284–4299.
[4] S. Xie, L. Guo, Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition, SIAM J. Control Optim. 56 (5) (2018) 3404–3431.
[5] A. Khalili, M.A. Tinati, A. Rastegarnia, J.A. Chanbers, Steady-state analysis of diffusion LMS adaptive networks with noisy links, Signal Process. 60 (2012) 974–979.
[6] D. Gan, S. Xie, Z. Liu, Stability of the distributed Kalman filter using general random coefficients, Sci. China Inf. Sci. 64 (2021) 172204.
[7] Y. Yu, H. Zhao, R.C. de Lamare, Y. Zakharov, L. Lu, Robust distributed diffusion recursive least squares algorithms with side information for adaptive networks, IEEE Trans. Signal Process. 67 (6) (2019) 1566–1581.
[8] D. Gan, Z. Liu, Strong consistency of the distributed stochastic gradient algorithm, in: Proceedings of the 58th IEEE Conference on Decision and Control, Nice, France, 2019, pp. 5082–5087.
[9] S. Battilotti, F. Cacace, M. d'Angelo, A. Germani, Asymptotically optimal consensus-based distributed filtering of continuous-time linear systems, Automatica 122 (2020) 109189.
[10] A. Griffin, T. Hirvonen, C. Tzagkarakis, A. Mouchtaris, P. Tsakalides, Single-channel and multi-channel sinusoidal audio coding using compressed sensing, IEEE Trans. Audio, Speech, Lang. Process. 19 (5) (2011) 1382–1395.
[11] Y. Bai, M. Tang, Object tracking via robust multitask sparse representation, IEEE Signal Process. Lett. 21 (8) (2014) 909–913.
[12] S. Vinga, Structured sparsity regularization for analyzing high-dimensional omics data, Brief Bioinf. 22 (1) (2021) 77–87.
[13] J.A. Bazerque, G.B. Giannakis, Distributed spectrum sensing for cognitive radio networks by exploiting sparsity, IEEE Trans. Signal Process. 58 (3) (2010) 1847–1862.
[14] J. Li, X. Li, Online sparse identification for regression models, Systems Control Lett. 141 (2020) 104710.
[15] P. Di Lorenzo, A.H. Sayed, Sparse distributed learning based on diffusion adaptation, IEEE Trans. Signal Process. 61 (6) (2013) 1419–1433.
[16] Y. Liu, C. Li, Z. Zhang, Diffusion sparse least-mean squares over networks, IEEE Trans. Signal Process. 60 (8) (2012) 4480–4485.
[17] Z. Liu, Y. Liu, C. Li, Distributed sparse recursive least-squares over networks, IEEE Trans. Signal Process. 62 (6) (2014) 1386–1395.
[18] R.G. Baraniuk, Compressive sensing [Lecture notes], IEEE Signal Process. Mag. 24 (4) (2007) 118–121.
[19] E.J. Candès, J.K. Romberg, T. Tao, Stable singal recovery from incomplete and inaccurate measurements, IEEE Trans. Inform. Theory 59 (8) (2006) 1207–1223.

[20] S. Xu, R.C. de Lamare, H.V. Poor, Distributed compressed estimation based on compressive sensing, IEEE Signal Process. Lett. 22 (9) (2015) 1311–1315.
[21] L. Li, D.H. Li, A distributed estimation method over network based on compressed sensing, Int. J. Distrib. Sens. Netw. 15 (4) (2019) 1–10.
[22] S. Xie, L. Guo, Analysis of compressed distributed adaptive filters, Automatica 112 (2020) 108707.
[23] V. Breschi, A. Bemporad, I.V. Kolmanovsky, Cooperative constrained parameter estimation by ADMM-RLS, Automatica 121 (2020) 109175.
[24] S. Xie, Y. Zhang, L. Guo, Convergence of a distributed least squares, IEEE Trans. Automat. Control 66 (10) (2021) 4952–4959.
[25] G. Zielke, Inversion of modified symmetric matrices, J. Assoc. Comput. Mach. 15 (3) (1968) 402–408.
[26] E.J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (12) (2005) 4203–4215.
[27] R.A. DeVore, Deterministic constructions of compressed sensing matrices, J. Complexity 23 (2007) 918–925.
[28] S. Li, F. Gao, G. Ge, S. Zhang, Deterministic construction of compressed sensing matrices via algebraic curves, IEEE Trans. Inform. Theory 58 (8) (2012) 5035–5041.
[29] G. Xu, Z. Xu, Compressed sensing matrices from Fourier matrices, IEEE Trans. Inform. Theory 61 (1) (2015) 469–478.
[30] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, Constr. Approx. 28 (3) (2006) 253–263.
[31] M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Comm. Pure Appl. Math. 61 (8) (2008) 1025–1045.
[32] B. Ghaddar, J. Naoum-Sawaya, High dimensional data classification and feature selection using support vector machines, European J. Oper. Res. 265 (3) (2018) 993–1004.
[33] J. Tropp, A. Gilbert, Signal recovery from partial information via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (2) (2007) 4655–4666.
[34] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, An interior-point method for large-scale ℓ_1-regularized least squares, IEEE J. Sel. Top. Sign. Proces. 1 (4) (2007) 606–617.
[35] J.B. Moore, On strong consistency of least squares identification algorithms, Automatica 14 (5) (1978) 505–509.
[36] H.F. Chen, Strong consistency and convergence rate of the least squares identification, Sci.nce China (Series A) 25 (7) (1982) 771–784.
[37] Y. Abbasi-Yadkori, D. Pál, C. Szepesvári, Improved algorithms for linear stochastic bandits, in: Proceedings of the 24th International Conference on Neural Information Processing Systems, Granada, Spain, 2011, pp. 2312–2320.
[38] M.K.S. Faradonbeh, A. Tewari, G. Michailidis, Finite time identification in unstable linear systems, Automatica 96 (2018) 342–353.
[39] M.K.S. Faradonbeh, A. Tewari, G. Michailidis, Input perturbations for adaptive control and learning, Automatica 117 (2020) 108950.
[40] L. Xiao, S. Boyd, S. Lall, A scheme for robust distributed sensor fusion based on average consensus, in: Proceedings of the 4th Fourth International Symposium on Information Processing in Sensor Networks, Boise, ID, USA, 2005, pp. 63–70.
[41] W. Zhao, G. Yin, E.W. Bai, Sparse system identification for stochastic systems with general observation sequences, Automatica 121 (2020) 109162.
[42] M.K.S. Faradonbeh, A. Tewari, G. Michailidis, On adaptive linear-quadratic regulators, Automatica 117 (2020) 108982.