J Syst Sci Complex (2024) 37(2): 609–628

Distributed Least Squares Algorithm of Continuous-Time Stochastic Regression Model Based on Sampled Data*

ZHU Xinghua · GAN Die · LIU Zhixin

DOI: 10.1007/s11424-024-3016-4
Received: 15 January 2023 / Revised: 5 March 2023

Abstract In this paper, the authors consider the distributed adaptive identification problem over sensor networks using sampled data, where the dynamics of each sensor is described by a stochastic differential equation. By minimizing a local objective function at sampling time instants, the authors propose an online distributed least squares algorithm based on sampled data. A cooperative non-persistent excitation condition is introduced, under which the convergence results of the proposed algorithm are established by properly choosing the sampling time interval. The upper bound on the accumulative regret of the adaptive predictor can also be provided. Finally, the authors demonstrate the cooperative effect of multiple sensors in the estimation of unknown parameters by computer simulations.

Keywords Cooperative excitation condition, distributed least squares, regret, sampled data, stochastic differential equation.

1 Introduction

With the rapid development of information technology in recent decades, wireless sensor networks are widely applied in various fields such as pollution control, ocean sampling, UAV formation control, and machine health monitoring (cf., [1, 2]). Sensor networks can gather a

ZHU Xinghua
The Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, P. R. China. Email: zxh@amss.ac.cn.
GAN Die
Zhongguancun Laboratory, Beijing 100094, *China.* Email: gandie@amss.ac.cn.
LIU Zhixin (Corresponding author)
The Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, P. R. China. Email: lzx@amss.ac.cn.
*This work was supported by the Natural Science Foundation of China under Grant No. T2293772, the National Key R&D Program of China under Grant No. 2018YFA0703800, the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDA27000000, and the National Science Foundation of Shandong Province under Grant No. ZR2020ZD26.
◇ *This paper was recommended for publication by Editor GUO Jin.*

large amount of data, and how to use these data to design appropriate algorithms to estimate unknown parameters is a meaningful research direction. The centralized and distributed methods are commonly used in the design of estimation algorithms. In the centralized method, a fusion center is needed to process the data from all sensors, while the sensors in the distributed scheme can cooperatively estimate unknown parameters of interest by using local measurements. It is clear that the distributed algorithms have the advantages of robustness to node or link failures, reducing calculation pressure and communication load, see [3–8] among many references.

For the discrete-time dynamical systems, incremental-, consensus-, diffusion-based distributed estimation or filtering algorithms have been proposed and considerable progress has been made in the corresponding analysis of the algorithms. For the time-invariant unknown parameter case, a bias-compensated distributed least squares (LS) algorithm based on diffusion strategy was proposed by Bertrand, et al. in [9] and the mean-square stability with independent and ergodic regressors was obtained. Cattivelli, et al. in [10] designed a diffusion-based distributed LS algorithm and studied the mean-square performance of the proposed algorithm under similar conditions of regressors. Arablouei, et al. in [11] presented a partial diffusion-based distributed LS algorithm and established the mean-square stability results of the proposed algorithm under independent and ergodic inputs. We remark that the independency and ergodicity assumptions of regressors or inputs using in the above literature are too stringent to be satisfied for dynamical feedback systems. In order to overcome this issue, Xie, et al. in [4] proposed a cooperative non-persistent excitation (non-PE) condition, under which the convergence of the diffusion-based distributed LS algorithm was established. Other algorithms such as distributed stochastic gradient algorithm, distributed stochastic approximation algorithm, and distributed weighted LS algorithm were also investigated, see [12–14] for more references. For the case where the unknown parameters are time-varying, the stability and performance analysis for the classical distributed least mean squares algorithms with independent and stationary signals were investigated (cf., [15]). For non-independent or non-stationary signals, Xie and Guo in [5, 16] introduced a joint information condition, and then established stability and performance analysis theory of the proposed algorithms under this condition. The distributed Kalman filter and distributed forgetting factor least squares algorithm were also studied in [17, 18].

It is known that the dynamics in physics and engineering systems such as advection-diffusion, oil spill, and electromagnetic induction are naturally modeled by (stochastic) differential equations (cf., [3, 19]). For the distributed identification problem of continuous-time systems, some theoretical results based on continuous-time signals were obtained (cf., [3, 20–24]). For example, Chen, et al. in [22] established the convergence result of the consensus-based identification algorithm for a group of continuous-time subsystems with uniformly bounded regressors satisfying a cooperative PE condition. You and Wu in [3] proposed a consensus-based identification algorithm to estimate unknown parameters of the advection-diffusion partial differential equations and proved the convergence results of the proposed algorithm under a similar cooperative PE condition. Papusha, et al. in [23] introduced a consensus-based gradient algorithm and investigated the asymptotic convergence of the algorithm with cooperative PE inputs. Furthermore, Javed, et al. in [24] analyzed the uniform exponential stability of the consensus-based gradient algorithm with cooperative PE regressors. More related results about distributed estimation algorithms for continuous-time systems can be found in [25–27]. In summary, for continuous-time distributed estimation algorithms, the regression signals are often required to be deterministic and satisfy PE conditions.

We note that the data collected by sensors are often digital (or discrete-time) due to the communication and computer technologies. Therefore, the identification problem of continuous-time (stochastic) systems based on discrete-time signals is more practical, but few results are obtained in this direction. For single sensor case, Marelli and Fu in [28] converted the identification problem of continuous-time systems based on sampled data into the identification problem of sampling systems by applying constant sampling intervals and signal reconstruction techniques, and illustrated the effectiveness of the proposed method by simulation examples. Pan, et al. in [29] proposed a refined instrumental variable method for continuous-time systems by using similar techniques, and established consistency results of the estimator under PE inputs. More related analysis can be found in [30–32]. In these references, the PE conditions of regression or input signals are required, which is hard to be satisfied in general. In order to relax the PE condition, we put forward a least squares algorithm for continuous-time system based on sampled data in [39], and provided the convergence analysis of the algorithm under a non-PE condition. However, the investigation of distributed identification problem of continuous-time systems over sensor networks is still lack.

In this paper, we design the distributed algorithm to estimate unknown parameters of continuous-time systems based on sampled data, and give theoretical analysis for the convergence of the algorithm. The main contributions of this paper can be summarized as follows.

1) We introduce a local objective function to characterize the accumulative sampling prediction error. By minimizing the objective function, we put forward the online distributed LS algorithm based on sampled data.

2) By employing the martingale estimation theory and stochastic Lyapunov function method, we provide upper bounds on the estimation error and the accumulative regret of the adaptive predictor of the distributed LS algorithm. Furthermore, under a cooperative non-PE excitation condition of regression signals, the convergence can be obtained by properly choosing flexible sampling time intervals.

The rest of this paper is arranged as follows. The problem formulation is given in Section 2. In Section 3, asymptotic analysis of the proposed algorithm are established. Simulation results are given in Section 4, and concluding remarks are made in Section 5.

2 Problem Formulation

2.1 Some Preliminaries

For an n-dimensional square matrix $A \in \mathbb{R}^{n\times n}$, we use $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ and $|A|$ to denote its minimum eigenvalue, maximum eigenvalue and determinant, respectively. For an $m \times n$-

dimensional matrix $A \in \mathbb{R}^{m\times n}$, we use $\|A\|$ and A^{T} to denote its Euclidean norm and transpose operator. A square matrix is said to be stochastic if all its elements are nonnegative and the sum of all rows are equal to 1. For a positive scalar sequence $\{a_k, k \geq 0\}$ and a matrix sequence $\{A_k, k \geq 0\}$, the notation $A_k = O(a_k)$ means that $\|A_k\| \leq Ca_k$ holds for all $k \geq 0$ where C is a positive constant independent of t, and $A_k = o(a_k)$ means that $\lim_{k\to\infty} \frac{\|A_k\|}{a_k} = 0$. The Kronecker product $A \otimes B$ of two matrices $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $B = [b_{ij}] \in \mathbb{R}^{p\times q}$ is defined as the following block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{mp\times nq}.$$

In this paper, we consider the distributed estimation problem of stochastic regression models. Thus, we need to introduce some related concepts and lemmas in probability theory. For a probability space $(\Omega, \mathcal{F}, P)$, we use $\{\mathcal{F}_k, k \geq 0\}$ to denote a family of nondecreasing σ-algebras of $\mathcal{F}$, i.e., $\mathcal{F}_{k_1} \subseteq \mathcal{F}_{k_2}$ for $k_1 \leq k_2$. If a stochastic sequence ν_k is $\mathcal{F}_k$-measurable, then ν_k is called $\mathcal{F}_k$ adapted. An adapted stochastic sequence $\{\nu_k, \mathcal{F}_k\}$ is called a martingale difference sequence if $\mathbb{E}(\nu_{k+1}|\mathcal{F}_k) = 0$ where $\mathbb{E}(\cdot|\cdot)$ represents the conditional expectation operator. The following martingale estimation theorem plays an important role in estimating the summation of stochastic sequences.

Lemma 2.1 ([33]) *Suppose that m_k is $\mathcal{F}_k$-measurable and $\{\xi_k, \mathcal{F}_k\}$ is a martingale difference sequence satisfying $\sup_k \mathbb{E}[\|\xi_{k+1}\|^\beta|\mathcal{F}_k] < \infty$ with $\beta \in (0, 2]$. Then for any $\eta > 0$, the following equation holds almost surely,*

$$\sum_{k=0}^{n} m_k\xi_{k+1} = O\Big(U_n(\beta)\log^{\frac{1}{\beta}+\eta}(U_n(\beta) + e)\Big),$$

where $U_n(\beta)$ is defined by $U_n(\beta) = \Big(\sum_{k=0}^{n} \|m_k\|^\beta\Big)^{\frac{1}{\beta}}$.

An adapted continuous stochastic process $\{X_t, \mathcal{F}_t; t \geq 0\}$ is called a Wiener process if $X_0 = 0$ holds almost surely (a.s.), and for any $0 \leq s_1 < s_2 < \infty$, $X_{s_1} - X_{s_2}$ is independent of $\mathcal{F}_{s_1}$, and obeys a normal distribution with mean zero, variance $s_2 - s_1$.

For the convenience of analysis, we model the communication between sensors as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. $\mathcal{N}_i = \{j \in \mathcal{V}|(j, i) \in \mathcal{E}\}$ is used to denote the neighbor set of the sensor i. A path with length h in the graph $\mathcal{G}$ is defined as a sequence of sensors $\{i_1, \cdots, i_h\}$ satisfying $(i_j, i_{j+1}) \in \mathcal{E}$ for all $1 \leq j \leq h-1$. The diameter of the graph $\mathcal{G}$, denoted as $D_{\mathcal{G}}$, is defined as the maximum shortest length of the path between any two sensors. The elements of the weighted adjacency matrix $\mathcal{A} = [a_{ij}]_{n\times n}$ are non-negative, and $a_{ij} > 0$ if and only if $(i, j) \in \mathcal{E}$. More details on the graph theory can be found in [34].

2.2 Distributed LS Algorithm Based on Sampled Data

We consider a network consisting of N sensors. For each sensor $i \in \{1, \cdots, N\}$, its dynamics is described by the following continuous-time stochastic differential equation

$$y_{t,i} = S\phi_{t,i}^{\mathrm{T}}\theta + v_{t,i}, \tag{1}$$

where S is the integral operator (i.e., $S\phi_{t,i}^{\mathrm{T}} = \int_0^t \phi_{s,i}^{\mathrm{T}} ds$), $y_{t,i}$ is the scalar output of the sensor i at time t, $\phi_{t,i}$ is an l-dimensional stochastic regressor, $v_{t,i}$ is the system noise, and $\theta \in \mathbb{R}^l$ is the unknown parameter vector to be estimated. It is clear that many models can be written in the form of the model (1), such as the continuous-time ARX model and ARMAX model (cf., [35]).

Limited by communication and computer technologies, sensors often collect discrete-time or sampled data. In this paper, we design the distributed algorithm to cooperatively estimate the unknown parameter vector θ in (1) by using sampled data from neighbors. The following matrix inversion formula is useful in deriving the recursive form of the algorithm.

Lemma 2.2 ([33]) *Let A, B, C and D be matrices with appropriate dimensions, and the related matrices be invertible. Then we have the matrix inversion formula as follows*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}. \tag{2}$$

We recursively define the following local objective function $J_{t_{k+1},i}(\beta)$,

$$J_{t_{k+1},i}(\beta) = \sum_{j\in N_i} a_{ij} \left\{ J_{t_k,i}(\beta) + \left(y_{t_{k+1},j} - y_{t_k,j} - \beta^{\mathrm{T}}\delta_k\phi_{t_k,j}\right)^2 \right\} \tag{3}$$

with $J_{0,i}(\beta) = 0$. By simple calculation, we can derive the following equation from (3),

$$J_{t_{k+1},i}(\beta) = \sum_{j=1}^{N}\sum_{s=0}^{k} a_{ij}^{(k+1-s)} \left(y_{t_{s+1},j} - y_{t_s,j} - \beta^{\mathrm{T}}\delta_s\phi_{t_s,j}\right)^2,$$

where $a_{ij}^{(k+1-s)}$ is the (i,j)-th element of the matrix $\mathcal{A}^{k+1-s}$, t_s represents the s-th sampling time instant and $\delta_s = t_{s+1} - t_s$ is the sampling period. By minimizing the above local objective function $J_{t_{k+1},i}(\beta)$, we obtain the following distributed LS algorithm based on sampled data,

$$\begin{aligned} &\theta_{t_{k+1},i} \triangleq \arg\min_{\beta} J_{t_{k+1},i}(\beta) \\ &= \left(\sum_{j=1}^{N}\sum_{s=0}^{k} a_{ij}^{(k+1-s)}\delta_s^2\phi_{t_s,j}\phi_{t_s,j}^{\mathrm{T}}\right)^{-1} \left(\sum_{j=1}^{N}\sum_{s=0}^{k} a_{ij}^{(k+1-s)}\delta_s\phi_{t_s,j}\left(y_{t_{s+1},j} - y_{t_s,j}\right)\right). \end{aligned} \tag{4}$$

Denote

$$P_{t_{k+1},i} = \left(\sum_{j=1}^{N}\sum_{s=0}^{k} a_{ij}^{(k+1-s)}\delta_s^2\phi_{t_s,j}\phi_{t_s,j}^{\mathrm{T}}\right)^{-1} \tag{5}$$

and

$$\overline{P}_{t_{k+1},i}^{-1} = P_{t_k,j}^{-1} + \delta_k^2\phi_{t_k,j}\phi_{t_k,j}^{\mathrm{T}}.$$

It is clear that the information matrix $P_{t_{k+1},i}$ can be written into the following recursive form:

$$P_{t_{k+1},i}^{-1} = \sum_{j\in N_i} a_{ij}\left(P_{t_k,j}^{-1} + \delta_k^2 \phi_{t_k,j}\phi_{t_k,j}^{\mathrm{T}}\right).$$

By Lemma 2.2, we can derive the following equation for $\overline{P}_{t_{k+1},i}$,

$$\overline{P}_{t_{k+1},i} = P_{t_k,i} - \frac{\delta_k^2 P_{t_k,i}\phi_{t_k,i}\phi_{t_k,i}^{\mathrm{T}} P_{t_k,i}}{1+\delta_k^2\phi_{t_k,i}^{\mathrm{T}} P_{t_k,i}\phi_{t_k,i}}. \tag{6}$$

Substituting (5) and (6) into (4), we can derive the recursive expression of the distributed LS algorithm as follows:

$$\begin{aligned}\theta_{t_{k+1},i} =& P_{t_{k+1},i}\sum_{j\in N_i} a_{ij}\left[P_{t_k,j}^{-1}\theta_{t_k,j} + \delta_k\phi_{t_k,j}\left(y_{t_{k+1},j} - y_{t_k,j}\right)\right]\\ =& P_{t_{k+1},i}\sum_{j\in N_i} a_{ij}\overline{P}_{t_{k+1},j}^{-1}\overline{P}_{t_{k+1},j}\left[P_{t_k,j}^{-1}\theta_{t_k,j} + \delta_k\phi_{t_k,j}\left(y_{t_{k+1},j} - y_{t_k,j}\right)\right]\\ =& P_{t_{k+1},i}\sum_{j\in N_i} a_{ij}\overline{P}_{t_{k+1},j}^{-1}\left[\theta_{t_k,j} + \frac{\delta_k P_{t_k,j}\phi_{t_k,j}}{1+\delta_k^2\phi_{t_k,j}^{\mathrm{T}}P_{t_k,j}\phi_{t_k,j}}\left(y_{t_{k+1},j} - y_{t_k,j} - \delta_k\phi_{t_k,j}^{\mathrm{T}}\theta_{t_k,j}\right)\right].\end{aligned} \tag{7}$$

Denote $\overline{\theta}_{t_{k+1},i} \triangleq \theta_{t_k,i} + \frac{\delta_k P_{t_k,i}\phi_{t_k,i}}{1+\delta_k^2\phi_{t_k,i}^{\mathrm{T}}P_{t_k,i}\phi_{t_k,i}}\left(y_{t_{k+1},i} - y_{t_k,i} - \delta_k\phi_{t_k,i}^{\mathrm{T}}\theta_{t_k,i}\right)$. Then, the equation (7) becomes

$$\theta_{t_{k+1},i} = P_{t_{k+1},i}\sum_{j\in N_i} a_{ij}\overline{P}_{t_{k+1},j}^{-1}\overline{\theta}_{t_{k+1},j}.$$

From the above analysis, the distributed least squares algorithm based on sampled data for the continuous-time linear regression model (1) can be summarized by Algorithm 1.

Denote

$$\begin{aligned}&Y_t \triangleq (y_{t,1},\cdots,y_{t,N})^{\mathrm{T}},\\ &\varPsi_t \triangleq \mathrm{diag}\{\phi_{t,1},\cdots,\phi_{t,N}\},\\ &V_t \triangleq (v_{t,1},\cdots,v_{t,N})^{\mathrm{T}},\\ &\vartheta \triangleq \mathrm{col}\{\theta,\cdots,\theta\},\end{aligned}$$

where diag$\{\cdots\}$ denotes a block diagonal matrix in which each block is the corresponding vector or matrix, and col$\{\cdots\}$ denotes a column vector whose elements are stacked by specified vectors. Then the system (1) can be written into the following compact form:

$$Y_t = S\varPsi_t^{\mathrm{T}}\vartheta + V_t. \tag{13}$$

Algorithm 1 Distributed LS Algorithm Based on Sampled Data

For any $i \in \{1, \cdots, n\}$, begin with an initial estimate $\theta_{0,i} \in \mathbb{R}^l$ and a positive definite matrix $P_{0,i} \in \mathbb{R}^{l\times l}$. The distributed LS algorithm based on sampled data is recursively designed as follows.

Step 1: Adaptation. $\overline{\theta}_{t_{k+1},i}$ and $\overline{P}_{t_{k+1},i}$ are recursively updated according to the following equations,

$$\overline{\theta}_{t_{k+1},i} = \theta_{t_k,i} + \delta_k a_{t_k,i} P_{t_k,i} \phi_{t_k,i} \left(y_{t_{k+1},i} - y_{t_k,i} - \delta_k \phi^{\mathrm{T}}_{t_k,i} \theta_{t_k,i} \right), \tag{8}$$

$$\overline{P}_{t_{k+1},i} = P_{t_k,i} - \delta_k^2 a_{t_k,i} P_{t_k,i} \phi_{t_k,i} \phi^{\mathrm{T}}_{t_k,i} P_{t_k,i}, \tag{9}$$

$$a_{t_k,i} = \frac{1}{1 + \delta_k^2 \phi^{\mathrm{T}}_{t_k,i} P_{t_k,i} \phi_{t_k,i}}. \tag{10}$$

Step 2: Combination. Generate $\theta_{t_{k+1},i}$ and $P_{t_{k+1},i}$ according to the following combination manner,

$$P_{t_{k+1},i} = \left\{ \sum_{j\in N_i} a_{ij} \overline{P}^{-1}_{t_{k+1},j} \right\}^{-1}, \tag{11}$$

$$\theta_{t_{k+1},i} = P_{t_{k+1},i} \sum_{j\in N_i} a_{ij} \overline{P}^{-1}_{t_{k+1},j} \overline{\theta}_{t_{k+1},j}. \tag{12}$$

Similarly, we introduce the following notations:

$$\begin{aligned}
&\vartheta_{t_k} \triangleq \mathrm{col}\{\theta_{t_k,1}, \cdots, \theta_{t_k,N}\},\\
&\widetilde{\vartheta}_{t_k} \triangleq \mathrm{col}\{\widetilde{\theta}_{t_k,1}, \cdots, \widetilde{\theta}_{t_k,N}\}, \quad \widetilde{\theta}_{\mathrm{t_k},i} = \theta_{\mathrm{t_k,i}} - \theta,\\
&\widetilde{\overline{\vartheta}}_{t_k} \triangleq \mathrm{col}\{\widetilde{\overline{\theta}}_{t_k,1}, \cdots, \widetilde{\overline{\theta}}_{t_k,N}\}, \quad \widetilde{\overline{\theta}}_{\mathrm{t_k},i} = \overline{\theta}_{\mathrm{t_k,i}} - \theta,\\
&a_{t_k} \triangleq \mathrm{diag}\{a_{t_k,1}, \cdots, a_{t_k,N}\},\\
&P_{t_k} \triangleq \mathrm{diag}\{P_{t_k,1}, \cdots, P_{t_k,N}\},\\
&\overline{P}_{t_k} \triangleq \mathrm{diag}\{\overline{P}_{t_k,1}, \cdots, \overline{P}_{t_k,N}\},\\
&\mathscr{A} \triangleq \mathcal{A} \otimes I_l,
\end{aligned}$$

where $\mathcal{A}$ is the weighted adjacency matrix and $\otimes$ is the Kroneker product. Then the distributed LS algorithm (8)–(12) can be equivalently written into the following equations:

$$\begin{cases}
\overline{\vartheta}_{t_{k+1}} = \vartheta_{t_k} + \delta_k b_{t_k} P_{t_k} \varPsi_{t_k} \left(Y_{t_{k+1}} - Y_{t_k} - \delta_k \varPsi^{\mathrm{T}}_{t_k} \vartheta_{t_k} \right),\\
\overline{P}_{t_{k+1}} = P_{t_k} - \delta_k^2 b_{t_k} P_{t_k} \varPsi_{t_k} \varPsi^{\mathrm{T}}_{t_k} P_{t_k},\\
b_{t_k} = a_{t_k} \otimes I_l,\\
a_{t_k} = \left(I_N + \delta_k^2 \varPsi^{\mathrm{T}}_{t_k} P_{t_k} \varPsi_{t_k} \right)^{-1},
\end{cases} \tag{14}$$

and

$$\begin{cases} \mathrm{vec}\{P_{t_{k+1}}^{-1}\} = \mathscr{A}\mathrm{vec}\{\overline{P}_{t_{k+1}}^{-1}\}, \\ \vartheta_{t_{k+1}} = P_{t_{k+1}}\mathscr{A}\overline{P}_{t_{k+1}}^{-1}\overline{\vartheta}_{t_{k+1}}, \end{cases} \tag{15}$$

where vec$\{\cdot\}$ represents the matrix stacked by the blocks of a block-diagonal matrix on top of each other.

3 Asymptotic Results of the Distributed LS Algorithm.

3.1 Convergence Analysis of the Algorithm

In order to investigate the asymptotic properties of the distributed LS algorithm proposed in Subsection 2.2, we first introduce some assumptions.

Assumption 3.1 (Noise) The noise $\{v_{t,i}, \mathcal{F}_t\}$ is a Wiener process, where $\{\mathcal{F}_t\}$ is a family of nondecreasing σ-algebras with $\mathcal{F}_t = \sigma\{\phi_{s,i}, v_{s,i}, i = 1, \cdots, N, s \leq t\}$.

Remark 3.2 Assumption 3.1 is a basic assumption for continuous-time noises. Many noises in the investigation of control and filtering problems of continuous-time systems are modeled by Wiener process, such as the population growth model, circuit system (cf., [36]). From Assumption 3.1, we know that for $k \geq 0$, $\overline{v}_{t_k,i} \triangleq v_{t_k,i} - v_{t_{k-1},i}$ follows a normal distribution with mean 0 and variance δ_k, and $\{\overline{v}_{t_k,i}, \mathcal{F}_{t_k}\}$ is a martingale difference sequence satisfying $0 < \sup_k \mathbb{E}[|\overline{v}_{t_k,i}|^\gamma|\mathcal{F}_{t_k}] < \infty$ a.s. for any constant $\gamma \geq 2$.

Assumption 3.3 (Network topology) The graph $\mathcal{G}$ is connected, and the weighted adjacency matrix $\mathcal{A}$ is symmetric and doubly stochastic.

Remark 3.4 For the undirected graph case, connectivity is often used in theoretical investigations of multi-agent systems and distributed algorithms. In fact, the convergence results obtained in this paper can be also extended to the directed graph case where $\mathcal{G}$ is balanced and strongly connected. Denote $\mathcal{A}^h = [a_{ij}^{(h)}]_{N\times N}$, where $a_{ij}^{(h)}$ represents the (i,j)-th element of the matrix $\mathcal{A}^h$. By Lemma 8.1.2 in [34] and Assumption 3.3, we can easily derive that $a_{ij}^{(D_\mathcal{G})} \geq a_{\min} > 0$, where $a_{\min} = \min_{i,j\in\mathcal{V}} a_{ij}^{(D_\mathcal{G})}$. By the property of stochastic matrices, we have for any $t > D_\mathcal{G}, a_{ij}^{(t)} \geq a_{\min} > 0$.

Assumption 3.5 The stochastic regressor $\phi_{t,i}$ is Lipschitz continuous for almost all sample paths, i.e., $\|\phi_{t,i} - \phi_{s,i}\| \leq L|t-s|$, a.s. holds for all $t > 0$, $s > 0$, and $i \in \{1, \cdots, N\}$, where $L > 0$ is a constant.

Remark 3.6 Intuitively speaking, if the nonlinear regression vectors $\phi_{t,i}$ grows too fast with t, then it is difficult to deal with the approximation derivation from the sampled data. Thus, we need to add some assumptions on the growth rate of $\phi_{t,i}$, and Assumption 3.5 is often used in the analysis and control of nonlinear dynamical systems (cf., [37, 38]). We can verify that many signals can be included in Assumption 3.5, such as sine signals, cosine signals, linear signals.

Assumption 3.7 The sampling interval sequence $\{\delta_k, k \geq 0\}$ satisfies $\sum_{k=0}^t \delta_k^2 = \infty$ and $\sum_{k=0}^t \delta_k^4 < \infty$.

Remark 3.8 From the simulation example in Section 4, we see that the constant sampling intervals can not guarantee convergence of the proposed algorithm. Thus, it is necessary to introduce the flexible sampling intervals to investigate the performance of the algorithm.

Assumption 3.9 (Cooperative non-persistent information condition) The growth rate of $\log\left(\lambda_{\max}\left\{P_{t_{n+1}}^{-1}\right\}\right)$ and $\lambda_{\min}\left\{P_{t_{n+1}}^{-1}\right\}$ satisfies the following relationship:

$$\lim_{n\to\infty}\frac{\log R_n}{\lambda_{\min}^n}=0\quad \text{a.s.},$$

where $\lambda_{\min}^n=\lambda_{\min}\left\{\sum_{j=1}^N P_{0,j}^{-1}+\sum_{j=1}^N\sum_{k=0}^{n-D_{\mathcal{G}}+1}\delta_k^2\phi_{t_k,j}\phi_{t_k,j}^{\rm T}\right\}$ and

$$R_n=\lambda_{\max}\left\{P_0^{-1}\right\}+\sum_{j=1}^N\sum_{k=0}^n\delta_k^2\|\phi_{t_k,j}\|^2. \tag{16}$$

Remark 3.10 For the single sensor case, we proved that convergence of the LS algorithm based on sampled data can be achieved under the following information condition (see [39]):

$$\lim_{n\to\infty}\frac{\log\left(\lambda_{\max}\left\{P_{0,i}^{-1}\right\}+\sum_{k=0}^n\delta_k^2\|\phi_{t_k,i}\|^2\right)}{\lambda_{\min}\left\{P_{0,i}^{-1}+\sum_{k=0}^n\delta_k^2\phi_{t_k,i}\phi_{t_k,i}^{\rm T}\right\}}=0,\quad \text{a.s.}. \tag{17}$$

The cooperative non-persistent information condition (Assumption 3.9) can be degenerated to (17) when $N=1$. Moreover, Assumption 3.9 can reveal the cooperative effect of multiple sensors in a sense that multiple sensors can cooperate to accomplish the estimation task through information exchange, even if none of them can do it alone due to lack of sufficient excitations. We will illustrate this point by a simulation example in Section 4.

Denote

$$\begin{aligned}
&\varXi_{t_{k+1}}\triangleq{\rm diag}\{\varXi_{t_{k+1},1},\cdots,\varXi_{t_{k+1},N}\}\ \text{with}\ \ \varXi_{t_{k+1},i}=\int_{t_k}^{t_{k+1}}\phi_{s,i}ds-\delta_k\phi_{t_k,i}, \qquad (18)\\
&\overline{V}_{t_{k+1}}\triangleq V_{t_{k+1}}-V_{t_k}=(\overline{v}_{t_{k+1},1},\cdots,\overline{v}_{t_{k+1},N})^{\rm T}.
\end{aligned}$$

By the above notations and definition of $\varPsi_t$, we have $\varXi_{t_{k+1}}=\int_{t_k}^{t_{k+1}}\varPsi_s ds-\delta_k\varPsi_{t_k}$. From (13) and (14), we have

$$\begin{aligned}
\widetilde{\overline{\vartheta}}_{t_{k+1}}=&\overline{\vartheta}_{t_{k+1}}-\vartheta\\
=&\vartheta_{t_k}-\vartheta+\delta_k b_{t_k}P_{t_k}\varPsi_{t_k}\left(\int_{t_k}^{t_{k+1}}\varPsi_s^{\rm T}ds\vartheta+\overline{V}_{t_{k+1}}-\delta_k\varPsi_{t_k}^{\rm T}\vartheta_{t_k}\right)\\
=&\widetilde{\vartheta}_{t_k}+\delta_k b_{t_k}P_{t_k}\varPsi_{t_k}\left(\varXi_{t_{k+1}}^{\rm T}\vartheta+\overline{V}_{t_{k+1}}-\delta_k\varPsi_{t_k}^{\rm T}\widetilde{\vartheta}_{t_k}\right)\\
=&\left(I_{Nl}-\delta_k^2 b_{t_k}P_{t_k}\varPsi_{t_k}\varPsi_{t_k}^{\rm T}\right)\widetilde{\vartheta}_{t_k}+\delta_k b_{t_k}P_{t_k}\varPsi_{t_k}\left(\varXi_{t_{k+1}}^{\rm T}\vartheta+\overline{V}_{t_{k+1}}\right)\\
=&\overline{P}_{t_{k+1}}P_{t_k}^{-1}\widetilde{\vartheta}_{t_k}+\delta_k b_{t_k}P_{t_k}\varPsi_{t_k}\left(\varXi_{t_{k+1}}^{\rm T}\vartheta+\overline{V}_{t_{k+1}}\right),
\end{aligned} \tag{19}$$

where l is the dimension of unknown parameter θ. By (15) and the fact that $P_{t_{k+1}}\mathscr{A}\overline{P}^{-1}_{t_{k+1}}\vartheta=\vartheta$, we can derive the following error equation:

$$\begin{aligned}\widetilde{\vartheta}_{t_{k+1}}=&\vartheta_{t_{k+1}}-\vartheta\\=&P_{t_{k+1}}\mathscr{A}\overline{P}^{-1}_{t_{k+1}}\overline{\vartheta}_{t_{k+1}}-P_{t_{k+1}}\mathscr{A}\overline{P}^{-1}_{t_{k+1}}\vartheta\\=&P_{t_{k+1}}\mathscr{A}\overline{P}^{-1}_{t_{k+1}}\widetilde{\overline{\vartheta}}_{t_{k+1}}.\end{aligned}\tag{20}$$

Before analyzing the above equation, we need to introduce a lemma which reveals the relationship between the matrix $\overline{P}_{t_{k+1}}$ and $P_{t_{k+1}}$ in the algorithm (14)–(15). This lemma will be used to deal with the influence of neighbor relationship on the convergence of the proposed algorithm.

Lemma 3.11 (see [4]) *By* (11), *we have the following inequalities:*

$$\begin{aligned}&\mathscr{A}P_{t_{k+1}}\mathscr{A}\leq\overline{P}_{t_{k+1}},\\&|\overline{P}^{-1}_{t_{k+1}}|\leq|P^{-1}_{t_{k+1}}|,\end{aligned}$$

where $\overline{P}_{t_{k+1}}$ and $P_{t_{k+1}}$ are defined by (14) *and* (15).

As mentioned in [38, 40], the approximation deviation $\varXi_{t_k}$ defined by (18) should not be ignored, and we take the influence of $\varXi_{t_k}$ on the performance analysis of the algorithm into account in this paper. Based on the lemmas introduced above, we can first obtain the following theorem for the estimation error without any information excitation condition on the regression vector $\phi_{t,i}$.

Theorem 3.12 *Under Assumption* 3.1, *the following equation holds almost surely,*

$$\widetilde{\vartheta}^{\mathrm{T}}_{t_{n+1}}P^{-1}_{t_{n+1}}\widetilde{\vartheta}_{t_{n+1}}+\left(\frac{1}{2}+o(1)\right)\sum_{k=0}^{n}\delta_k^2\widetilde{\vartheta}^{\mathrm{T}}_{t_k}\varPsi_{t_k}a_{t_k}\varPsi^{\mathrm{T}}_{t_k}\widetilde{\vartheta}_{t_k}=O\left(\sum_{k=0}^{n}(\varXi^{\mathrm{T}}_{t_{k+1}}\vartheta)^2+\log|P^{-1}_{t_{n+1}}|\right).$$

Proof Consider the Lyapunov candidate function $W_{t_k}=\widetilde{\vartheta}^{\mathrm{T}}_{t_k}P^{-1}_{t_k}\widetilde{\vartheta}_{t_k}$. By (20) and the lemma 3.11, we have

$$\begin{aligned}\widetilde{\vartheta}^{\mathrm{T}}_{t_{k+1}}P^{-1}_{t_{k+1}}\widetilde{\vartheta}_{t_{k+1}}&=\widetilde{\overline{\vartheta}}^{\mathrm{T}}_{t_{k+1}}\overline{P}^{-1}_{t_{k+1}}\mathscr{A}P_{t_{k+1}}P^{-1}_{t_{k+1}}P_{t_{k+1}}\mathscr{A}\overline{P}^{-1}_{t_{k+1}}\widetilde{\overline{\vartheta}}_{t_{k+1}}\\&\leq\widetilde{\overline{\vartheta}}^{\mathrm{T}}_{t_{k+1}}\overline{P}^{-1}_{t_{k+1}}\widetilde{\overline{\vartheta}}_{t_{k+1}}.\end{aligned}\tag{21}$$

Substituting (19) into the equation (21), we obtain the following equation:

$$\begin{aligned}&\widetilde{\overline{\vartheta}}^{\mathrm{T}}_{t_{k+1}}\overline{P}^{-1}_{t_{k+1}}\widetilde{\overline{\vartheta}}_{t_{k+1}}\\=&\left\{\overline{P}_{t_{k+1}}P^{-1}_{t_k}\widetilde{\vartheta}_{t_k}+\delta_kb_{t_k}P_{t_k}\varPsi_{t_k}\left(\varXi^{\mathrm{T}}_{t_{k+1}}\vartheta+\overline{V}_{t_{k+1}}\right)\right\}^{\mathrm{T}}\overline{P}^{-1}_{t_{k+1}}\\&\cdot\left\{\overline{P}_{t_{k+1}}P^{-1}_{t_k}\widetilde{\vartheta}_{t_k}+\delta_kb_{t_k}P_{t_k}\varPsi_{t_k}\left(\varXi^{\mathrm{T}}_{t_{k+1}}\vartheta+\overline{V}_{t_{k+1}}\right)\right\}\\=&\widetilde{\vartheta}^{\mathrm{T}}_{t_k}P^{-1}_{t_k}\overline{P}_{t_{k+1}}P^{-1}_{t_k}\widetilde{\vartheta}_{t_k}+2\delta_k\left(\varXi^{\mathrm{T}}_{t_{k+1}}\vartheta+\overline{V}_{t_{k+1}}\right)^{\mathrm{T}}\varPsi^{\mathrm{T}}_{t_k}b_{t_k}\widetilde{\vartheta}_{t_k}\\&+\delta_k^2\left(\varXi^{\mathrm{T}}_{t_{k+1}}\vartheta+\overline{V}_{t_{k+1}}\right)^{\mathrm{T}}\varPsi^{\mathrm{T}}_{t_k}P_{t_k}b_{t_k}\overline{P}^{-1}_{t_{k+1}}b_{t_k}P_{t_k}\varPsi_{t_k}\left(\varXi^{\mathrm{T}}_{t_{k+1}}\vartheta+\overline{V}_{t_{k+1}}\right).\end{aligned}\tag{22}$$

Noticing that matrices a_{t_k}, b_{t_k}, Ψ_{t_k} and P_{t_k} are all block diagonal, we can derive the following equations:

$$b_{t_k}\Psi_{t_k}=\Psi_{t_k}a_{t_k},\quad \Psi_{t_k}^{\mathrm{T}}b_{t_k}=a_{t_k}\Psi_{t_k}^{\mathrm{T}},\quad P_{t_k}b_{t_k}=b_{t_k}P_{t_k}. \tag{23}$$

By a simple calculation and (14), the first term on the right-hand-side (RHS) of (22) satisfies the following equation:

$$\begin{aligned}\widetilde{\vartheta}_{t_k}^{\mathrm{T}}P_{t_k}^{-1}\overline{P}_{t_{k+1}}P_{t_k}^{-1}\widetilde{\vartheta}_{t_k}=&\widetilde{\vartheta}_{t_k}^{\mathrm{T}}P_{t_k}^{-1}\left(P_{t_k}-\delta_k^2P_{t_k}\Psi_{t_k}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\right)P_{t_k}^{-1}\widetilde{\vartheta}_{t_k}\\ =&\widetilde{\vartheta}_{t_k}^{\mathrm{T}}P_{t_k}^{-1}\widetilde{\vartheta}_{t_k}-\delta_k^2\widetilde{\vartheta}_{t_k}^{\mathrm{T}}\Psi_{t_k}a_{t_k}\Psi_{t_k}^{\mathrm{T}}\widetilde{\vartheta}_{t_k}.\end{aligned} \tag{24}$$

For the second term on the RHS of (22), by (23) we can derive the following equation:

$$2\delta_k\left(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+\overline{V}_{t_{k+1}}\right)^{\mathrm{T}}\Psi_{t_k}^{\mathrm{T}}b_{t_k}\widetilde{\vartheta}_{t_k}=2\delta_k\vartheta^{\mathrm{T}}\Xi_{t_{k+1}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}\widetilde{\vartheta}_{t_k}+2\delta_k\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}\widetilde{\vartheta}_{t_k}. \tag{25}$$

For the last term on the RHS of (22), we have

$$\begin{aligned}&\delta_k^2\left(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+\overline{V}_{t_{k+1}}\right)^{\mathrm{T}}\Psi_{t_k}^{\mathrm{T}}P_{t_k}b_{t_k}\overline{P}_{t_{k+1}}^{-1}b_{t_k}P_{t_k}\Psi_{t_k}\left(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+\overline{V}_{t_{k+1}}\right)\\ =&\delta_k^2\vartheta^{\mathrm{T}}\Xi_{t_{k+1}}\Psi_{t_k}^{\mathrm{T}}P_{t_k}b_{t_k}\overline{P}_{t_{k+1}}^{-1}b_{t_k}P_{t_k}\Psi_{t_k}\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+2\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}\Psi_{t_k}^{\mathrm{T}}P_{t_k}b_{t_k}\overline{P}_{t_{k+1}}^{-1}b_{t_k}P_{t_k}\Psi_{t_k}\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta\\ &+\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}\Psi_{t_k}^{\mathrm{T}}P_{t_k}b_{t_k}\overline{P}_{t_{k+1}}^{-1}b_{t_k}P_{t_k}\Psi_{t_k}\overline{V}_{t_{k+1}}.\end{aligned} \tag{26}$$

By (23) and $\delta_k^2a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}=I_N-a_{t_k}$, we can deduce that

$$\begin{aligned}&\delta_k^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}b_{t_k}\overline{P}_{t_{k+1}}^{-1}b_{t_k}P_{t_k}\Psi_{t_k}\\ =&\delta_k^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}b_{t_k}\left(P_{t_k}^{-1}+\delta_k^2\Psi_{t_k}\Psi_{t_k}^{\mathrm{T}}\right)b_{t_k}P_{t_k}\Psi_{t_k}\\ =&\delta_k^2a_{t_k}^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}+\left(I_N-a_{t_k}\right)\delta_k^2a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\\ =&\delta_k^2a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}.\end{aligned} \tag{27}$$

Substituting (27) into (26) yields

$$\begin{aligned}&\delta_k^2\left(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+\overline{V}_{t_{k+1}}\right)^{\mathrm{T}}\Psi_{t_k}^{\mathrm{T}}P_{t_k}b_{t_k}\overline{P}_{t_{k+1}}^{-1}b_{t_k}P_{t_k}\Psi_{t_k}\left(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+\overline{V}_{t_{k+1}}\right)\\ =&\delta_k^2\vartheta^{\mathrm{T}}\Xi_{t_{k+1}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+2\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta\\ &+\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\overline{V}_{t_{k+1}}.\end{aligned} \tag{28}$$

Substituting (22), (24), (25) and (28) into (21), we have the following inequality:

$$\begin{aligned}&\widetilde{\vartheta}_{t_{k+1}}^{\mathrm{T}}P_{t_{k+1}}^{-1}\widetilde{\vartheta}_{t_{k+1}}\le\widetilde{\widetilde{\vartheta}}_{t_{k+1}}^{\mathrm{T}}\overline{P}_{t_{k+1}}^{-1}\widetilde{\widetilde{\vartheta}}_{t_{k+1}}\\ =&\widetilde{\vartheta}_{t_k}^{\mathrm{T}}P_{t_k}^{-1}\widetilde{\vartheta}_{t_k}-\delta_k^2\widetilde{\vartheta}_{t_k}^{\mathrm{T}}\Psi_{t_k}a_{t_k}\Psi_{t_k}^{\mathrm{T}}\widetilde{\vartheta}_{t_k}+2\delta_k\vartheta^{\mathrm{T}}\Xi_{t_{k+1}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}\widetilde{\vartheta}_{t_k}\\ &+2\delta_k\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}\widetilde{\vartheta}_{t_k}+\delta_k^2\vartheta^{\mathrm{T}}\Xi_{t_{k+1}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta\\ &+2\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta+\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\overline{V}_{t_{k+1}}.\end{aligned} \tag{29}$$

Summing both sides of (29) from $k=0$ to n leads to

$$
\begin{aligned}
&W_{t_{n+1}} - W_{t_0} + \sum_{k=0}^{n} \delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} \\
\leq& 2\sum_{k=0}^{n} \delta_k \vartheta^{\mathrm{T}} \Xi_{t_{k+1}} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} \\
&+ 2\sum_{k=0}^{n} \delta_k \overline{V}_{t_{k+1}}^{\mathrm{T}} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} + \sum_{k=0}^{n} \vartheta^{\mathrm{T}} \Xi_{t_{k+1}} a_{t_k} \delta_k^2 \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k} \Xi_{t_{k+1}}^{\mathrm{T}} \vartheta \\
&+ 2\sum_{k=0}^{n} \delta_k^2 \overline{V}_{t_{k+1}}^{\mathrm{T}} a_{t_k} \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k} \Xi_{t_{k+1}}^{\mathrm{T}} \vartheta + \sum_{k=0}^{n} \delta_k^2 \overline{V}_{t_{k+1}}^{\mathrm{T}} a_{t_k} \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k} \overline{V}_{t_{k+1}}.
\end{aligned} \tag{30}
$$

In the following, we proceed to estimate the RHS of (30) term by term. By Hölder inequality, we can obtain the following inequality for the first term on the RHS of (30),

$$
\begin{aligned}
2\sum_{k=0}^{n} \delta_k \vartheta^{\mathrm{T}} \Xi_{t_{k+1}} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} &\leq 2\left(\sum_{k=0}^{n} \|\vartheta^{\mathrm{T}} \Xi_{t_{k+1}}\|^2\right)^{\frac{1}{2}} \cdot \left(\sum_{k=0}^{n} \delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k}\right)^{\frac{1}{2}} \\
&\leq 2\sum_{k=0}^{n} \|\vartheta^{\mathrm{T}} \Xi_{t_{k+1}}\|^2 + \frac{1}{2}\sum_{k=0}^{n} \delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k}.
\end{aligned} \tag{31}
$$

By Assumption 3.1 and Remark 3.2, we see that $\{\overline{V}_{t_{k+1}}, \mathcal{F}_{t_k}, t_k \geq 0\}$ is a martingale difference sequence, and satisfies

$$
0 < \sup_k \mathbb{E}[|\overline{V}_{t_k}|^{\gamma} | \mathcal{F}_{t_k}] < \infty \tag{32}
$$

almost surely for any $\gamma \geq 2$. Thus, there exists a set Ω_0 with $\mathbb{P}\{\Omega_0\} = 1$ such that (32) holds for $\omega \in \Omega_0$. In the following, we will consider the estimate error $\widetilde{\vartheta}_{t_k}$ on the set Ω_0. By $\delta_k a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} \in \mathcal{F}_{t_k}$ and Lemma 2.1, we get the following estimation for the second term on the RHS of (30),

$$
\begin{aligned}
2\sum_{k=0}^{n} \delta_k \overline{V}_{t_{k+1}}^{\mathrm{T}} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} &= O\left(\left\{\sum_{k=0}^{n} \delta_k^2 (a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k})^2\right\}^{\frac{1}{2}+\eta}\right) \\
&= O(1) + o\left(\sum_{k=0}^{n} \delta_k^2 (a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k})^2\right) \quad \text{a.s.},
\end{aligned} \tag{33}
$$

where η is a positive constant. Note that

$$
a_{t_k} \delta_k^2 \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k} \leq I_N. \tag{34}
$$

The third term on the RHS of (30) can be estimated by

$$
\sum_{k=0}^{n} \vartheta^{\mathrm{T}} \Xi_{t_{k+1}} a_{t_k} \delta_k^2 \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k} \Xi_{t_{k+1}}^{\mathrm{T}} \vartheta \leq \sum_{k=0}^{n} \|\vartheta^{\mathrm{T}} \Xi_{t_{k+1}}\|^2. \tag{35}
$$

By Hölder inequality and (34), we can derive the estimation for the fourth term on the RHS of (30) as follows:

$$\begin{aligned} &2\sum_{k=0}^{n}\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta \\ \le&\sum_{k=0}^{n}\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\overline{V}_{t_{k+1}}+\sum_{k=0}^{n}(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta)^2 \\ \le&\sum_{k=0}^{n}\|a_{t_k}\delta_k^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\|\cdot\|\overline{V}_{t_{k+1}}\|^2+\sum_{k=0}^{n}(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta)^2 \\ =&\sum_{k=0}^{n}\lambda_{\max}\left\{a_{t_k}\delta_k^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\right\}\cdot\left\{\sum_{i=1}^{N}\overline{v}_{t_{k+1},i}^2\right\}+\sum_{k=0}^{n}(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta)^2. \end{aligned} \tag{36}$$

We derive the estimation for the last term on the RHS of (30) as follows:

$$\sum_{k=0}^{n}\delta_k^2\overline{V}_{t_{k+1}}^{\mathrm{T}}a_{t_k}\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\overline{V}_{t_{k+1}}\le\sum_{k=0}^{n}\lambda_{\max}\left\{a_{t_k}\delta_k^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\right\}\cdot\left\{\sum_{i=1}^{N}\overline{v}_{t_{k+1},i}^2\right\}. \tag{37}$$

Substituting (31)–(37) into (30) yields

$$\begin{aligned} &W_{t_{n+1}}+\left(\frac{1}{2}+o(1)\right)\sum_{k=0}^{n}\delta_k^2\widetilde{\vartheta}_{t_k}^{\mathrm{T}}\Psi_{t_k}a_{t_k}\Psi_{t_k}^{\mathrm{T}}\widetilde{\vartheta}_{t_k} \\ =&O(1)+4\sum_{k=0}^{n}(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta)^2+2\sum_{k=0}^{n}\lambda_{\max}\left\{a_{t_k}\delta_k^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\right\}\cdot\left\{\sum_{i=1}^{N}\overline{v}_{t_{k+1},i}^2\right\}\quad \text{a.s..} \end{aligned} \tag{38}$$

By following the proof line of Lemma 4.4 in [4], we can derive that

$$\sum_{k=0}^{n}\lambda_{\max}\left\{a_{t_k}\delta_k^2\Psi_{t_k}^{\mathrm{T}}P_{t_k}\Psi_{t_k}\right\}\cdot\left\{\sum_{i=1}^{N}\overline{v}_{t_{k+1},i}^2\right\}=O(\log|P_{t_{n+1}}^{-1}|). \tag{39}$$

From (38) and (39), we can deduce the results of the theorem. ∎

Based on the above analysis, we can obtain the upper bound on the estimation error in the following theorem.

Theorem 3.13 *Under Assumptions* 3.1–3.9*, the distributed LS estimator* $\vartheta_{t_{n+1}}$ *defined by* (14)–(15) *converges to the true parameter* ϑ *almost surely.*

Proof By Theorem 3.12, we can derive that $\widetilde{\vartheta}_{t_{n+1}}^{\mathrm{T}}P_{t_{n+1}}^{-1}\widetilde{\vartheta}_{t_{n+1}}=O\big(\sum_{k=0}^{n}(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta)^2+\log|P_{t_{n+1}}^{-1}|\big)$ holds almost surely. Thus,

$$\|\widetilde{\vartheta}_{t_{n+1}}\|^2=O\left(\frac{\sum_{k=0}^{n}(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta)^2+\log|P_{t_{n+1}}^{-1}|}{\lambda_{\min}\left\{P_{t_{n+1}}^{-1}\right\}}\right),\quad \text{a.s..} \tag{40}$$

From Assumption 3.5 and (18), we can derive the inequality as follows:

$$\|\Xi_{t_{k+1}}\|=\left\|\int_{t_k}^{t_{k+1}}(\Psi_s-\Psi_{t_k})ds\right\|\le\int_{t_k}^{t_{k+1}}L(s-t_k)ds=\frac{L}{2}\delta_k^2, \tag{41}$$

where L is defined in Assumption 3.5. Now, we estimate the term $\log|P_{t_{n+1}}^{-1}|$ in (40). By (11), we have the following equation:

$$P_{t_{n+1},i}^{-1} = \sum_{j\in N_i} a_{ij}\overline{P}_{t_{n+1},j}^{-1} = \sum_{j\in N_i} a_{ij}\left(P_{t_n,j}^{-1} + \delta_k^2 \Psi_{t_n,j}\Psi_{t_n,j}^{\mathrm{T}}\right). \tag{42}$$

By Theorem 3.1 in [4], we can get the following inequality:

$$\max_{1\le i\le N} \lambda_{\max}\{P_{t_{n+1},i}^{-1}\} \le \lambda_{\max}\Big\{P_0^{-1}\Big\} + \sum_{j=1}^{N}\sum_{k=0}^{n} \delta_k^2\|\phi_{t_k,j}\|^2 ds.$$

Consequently, we have the following equation:

$$\log|P_{t_{n+1}}^{-1}| \le Nl\log\Big(\max_{1\le i\le N}\lambda_{\max}\{P_{t_{n+1},i}^{-1}\}\Big) = O(\log R_n), \tag{43}$$

where Nl is the dimension of the matrix $P_{t_{n+1}}$. By (15) and (42), we have for any $n\ge 0$,

$$\begin{aligned}\mathrm{vec}\{P_{t_{n+1}}^{-1}\} =& \mathscr{A}\mathrm{vec}\{P_{t_n}^{-1}\} + \mathscr{A}\mathrm{vec}\left\{\delta_n^2\Psi_{t_n}\Psi_{t_n}^{\mathrm{T}}\right\}\\ =&\mathscr{A}^{n+1}\mathrm{vec}\{P_0^{-1}\} + \sum_{k=0}^{n}\mathscr{A}^{n+1-k}\mathrm{vec}\Big\{\delta_k^2\Psi_{t_k}\Psi_{t_k}^{\mathrm{T}}\Big\}.\end{aligned}$$

By the proof of Theorem 3.3 in [4] and Assumption 3.3 as well as Remark 3.4, we obtain the following result:

$$\lambda_{\min}\left\{P_{t_{n+1}}^{-1}\right\} = \min_{1\le i\le N}\lambda_{\min}\left\{P_{t_{n+1},i}^{-1}\right\} \ge a_{\min}\lambda_{\min}^n,$$

where $\lambda_{\min}^n = \big\{\sum_{j=1}^N P_{0,j}^{-1} + \sum_{j=1}^N\sum_{k=0}^{n-D_{\mathcal{G}}+1}\delta_k^2\phi_{t_k,j}\phi_{t_k,j}^{\mathrm{T}}\big\}$. Substituting this inequality, (41) and (43) into (40), we have the following result for the estimation error:

$$\|\widetilde{\vartheta}_{t_{n+1}}\|^2 = O\left(\frac{\sum_{k=0}^n\delta_k^4 + \log R_n}{\lambda_{\min}^n}\right) \quad \text{a.s..} \tag{44}$$

By Assumptions 3.7 and 3.9, the results of the theorem can be obtained. ∎

Comparing with [38, 40], we take the approximation deviation into account in the performance analysis of the algorithm. Thus, the upper bound on the estimation error of the distributed LS algorithm based on sampled data consists of two parts: The approximation deviation and the error caused by stochastic noises. We remark that convergence results are obtained in this paper for stochastic regression vectors without using PE information conditions or relying on independency assumptions of regressors or inputs.

3.2 Analysis of the Regret

In this subsection, we will study the adaptive prediction ability of the distributed LS algorithm based on sampled data. For this purpose, we introduce the accumulative regret of the proposed algorithm, which is one of important metrics to evaluate the performance of adaptive learning and estimation algorithms.

For $i \in \{1, \cdots, N\}$, we denote $\overline{y}_{t_k,i} = y_{t_{k+1},i} - y_{t_k,i}$. By the fact $\mathbb{E}[(v_{t_{k+1},i} - v_{t_k,i})|\mathcal{F}_{t_k}] = 0$, we see that the best predictor to $\overline{y}_{t_k,i}$ is $\mathbb{E}[\overline{y}_{t_k,i}|\mathcal{F}_{t_k}] = \mathbb{E}[(\int_{t_k}^{t_{k+1}} \phi_{s,i}^{\mathrm{T}} ds \cdot \theta)|\mathcal{F}_{t_k}]$. Since the parameter θ is unknown, we replace θ by the estimate $\theta_{t_k,i}$ and obtain an adaptive predictor $\overline{\overline{y}}_{t_k,i}$ as follows:

$$\overline{\overline{y}}_{t_k,i} = \delta_k \phi_{t_k,i}^{\mathrm{T}} \theta_{t_k,i}, \tag{45}$$

The regret $\mathcal{R}_{t_k,i}$ of the sensor i can be defined as the difference between the best predictor and the adaptive predictor. By $\phi_{t_k,i} \in \mathcal{F}_{t_k}$, $\mathcal{R}_{t_k,i}$ satisfies the following equation:

$$\mathcal{R}_{t_k,i} = \left\{\mathbb{E}[\overline{y}_{t_k,i}|\mathcal{F}_{t_k}] - \overline{\overline{y}}_{t_k,i}\right\}^2 = \left\{\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}} \phi_{s,i}^{\mathrm{T}} ds \cdot \theta - \delta_k \phi_{t_k,i}^{\mathrm{T}} \theta_{t_k,i}\right)\middle|\mathcal{F}_{t_k}\right]\right\}^2, \tag{46}$$

The following theorem provides the upper bound on the accumulative regret of the adaptive predictor (45).

Theorem 3.14 *Under Assumption* 3.1 *and Assumptions* 3.5, 3.7, *the accumulative regret satisfies*

$$\sum_{i=1}^{N}\sum_{k=0}^{n} \mathcal{R}_{t_k,i} = O(\log R_n) \quad \text{a.s.},$$

provided that $\delta_k^2 \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k} = O(1)$.

Proof By using Jensen inequality and (46), we have the following inequality:

$$\begin{aligned}\sum_{i=1}^{N}\sum_{k=0}^{n} \mathcal{R}_{t_k,i} &= \sum_{i=1}^{N}\sum_{k=0}^{n} \left\{\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}} \phi_{s,i}^{\mathrm{T}} ds \cdot \theta - \delta_k \phi_{t_k,i}^{\mathrm{T}} \theta_{t_k,i}\right)\middle|\mathcal{F}_{t_k}\right]\right\}^2 \\ &\le \sum_{i=1}^{N}\sum_{k=0}^{n} \mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}} \phi_{s,i}^{\mathrm{T}} ds \cdot \theta - \delta_k \phi_{t_k,i}^{\mathrm{T}} \theta_{t_k,i}\right)^2\middle|\mathcal{F}_{t_k}\right].\end{aligned} \tag{47}$$

By the fact that $a_{t_k}\left(I_N + \delta_k^2 \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k}\right) = I_N$, we have

$$\Psi_{t_k} \Psi_{t_k}^{\mathrm{T}} = \Psi_{t_k} a_{t_k} \Psi_{t_k}^{\mathrm{T}} + \Psi_{t_k} (\delta_k^2 a_{t_k} \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k}) \Psi_{t_k}^{\mathrm{T}}. \tag{48}$$

By (18), (41) and (48), we can obtain that

$$\begin{aligned}\sum_{i=1}^{N}\sum_{k=0}^{n} \mathcal{R}_{t_k,i} \le& \sum_{i=1}^{N}\sum_{k=0}^{n} \mathbb{E}\left[\left(\Xi_{t_{k+1},i}^{\mathrm{T}}\theta + \delta_k \phi_{t_k}^{\mathrm{T}} \widetilde{\theta}_{t_k}\right)^2\middle|\mathcal{F}_{t_k}\right] \\ \le& \sum_{i=1}^{N}\sum_{k=0}^{n} \mathbb{E}\left[2(\Xi_{t_{k+1},i}^{\mathrm{T}}\theta)^2 + 2\delta_k^2 \widetilde{\theta}_{t_k,i}^{\mathrm{T}} \phi_{t_k,i} \phi_{t_k,i}^{\mathrm{T}} \widetilde{\theta}_{t_k,i}\middle|\mathcal{F}_{t_k}\right] \\ =& \sum_{k=0}^{n} \mathbb{E}\left[2(\Xi_{t_{k+1}}^{\mathrm{T}}\vartheta)^2 + 2\delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k}\middle|\mathcal{F}_{t_k}\right] \\ \le& \frac{L^2}{2}\|\vartheta\|^2 \sum_{k=0}^{n} \delta_k^4 + 2\sum_{k=0}^{n} \delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} + 2\sum_{k=0}^{n} \delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} (\delta_k^2 a_{t_k} \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k}) \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} \\ =& O\left(\sum_{k=0}^{n} \delta_k^4 + \sum_{k=0}^{n} \delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k}\right),\end{aligned} \tag{49}$$

where $\phi_{t_k} \in \mathcal{F}_{t_k}$ and the condition $\delta_k^2 \Psi_{t_k}^{\mathrm{T}} P_{t_k} \Psi_{t_k} = O(1)$ are used. From Theorem 3.12, (41) and (43), we know that

$$\sum_{k=0}^{n} \delta_k^2 \widetilde{\vartheta}_{t_k}^{\mathrm{T}} \Psi_{t_k} a_{t_k} \Psi_{t_k}^{\mathrm{T}} \widetilde{\vartheta}_{t_k} = O\bigg(\sum_{k=0}^{n} (\Xi_{t_{k+1}}^{\mathrm{T}} \vartheta)^2 + \log |P_{t_{n+1}}^{-1}| \bigg) = O\bigg(\sum_{k=0}^{n} \delta_k^4 + \log R_n \bigg), \quad \text{a.s..}$$

Substituting the above equation into (49), we can obtain the desired result of the theorem. ∎

Remark 3.15 For a typical case where the regression vectors are bounded, we have $R_n = O(n)$. By Theorem 3.14, the averaged accumulative regret satisfies $\frac{1}{N \sum_{k=0}^{n} \delta_k} \sum_{i=1}^{N} \sum_{k=0}^{n} \mathcal{R}_{t_k,i} \to 0$.

4 Simulation Results

In this section, we conduct computer simulations to verify theoretical results obtained in this paper.

Consider a network consisting of three sensors whose dynamics obey the equation (1). We set the 3-dimensional regression vectors $\phi_{t,i}$ $(i = 1, 2, 3)$ as follows:

$$\phi_{t,1} = \left[1 + \sin\left(\frac{2\pi}{3} t \right), 0, 0 \right]^{\mathrm{T}}, \quad \phi_{t,2} = \left[0, 2 + \cos\left(\frac{2\pi}{3} t \right), 0 \right]^{\mathrm{T}}, \quad \phi_{t,3} = \left[0, 0, 1 + \sin\left(\frac{4\pi}{3} t \right) \right]^{\mathrm{T}}.$$

It can be easily verified that Assumption 3.5 holds. The noises $\{v_{t,i},\ t \geq 0\}$, $i = 1, 2, 3$ in (1) are Wiener processes. We take the sampling interval δ_k as follows:

$$\delta_k = \frac{0.2}{2^{\left\lceil \log_4^{3\lceil \frac{k}{500} \rceil + 4} \right\rceil - 1}}, \tag{50}$$

where $\lceil \cdot \rceil$ is a round-up operator. It is clear that $\sum_{k=0}^{n} \delta_k^2 = \infty$ and $\sum_{k=0}^{n} \delta_k^4 = O\left(\sum_{k=0}^{n} \frac{1}{2^{2k}} \right) < \infty$. Thus, Assumption 3.7 holds. Besides, we see that the three sensors can cooperate to satisfy Assumption 3.9 even though none of the regression signals $\{\phi_{t,i}, i = 1, 2, 3\}$ satisfy the condition (17). The parameter vector to be estimated is taken as $\theta = (3, 4, 5)^{\mathrm{T}}$, and the initial values of the estimate are taken as

$$\theta_{0,1} = (-1, 0, 0)^{\mathrm{T}}, \quad \theta_{0,2} = (0, -1, 0)^{\mathrm{T}}, \quad \theta_{0,3} = (0, 0, -1)^{\mathrm{T}}.$$

The initial covariance matrix is taken as

$$P_{0,1} = P_{0,2} = P_{0,3} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

First, we consider the cooperative distributed LS algorithm proposed in this paper and the non-cooperative LS algorithm based on sampled data in [39], and the mean square errors (MSEs) (averaged over 150 runs) of these two algorithms are shown in Figure 1. From Figure 1, we see that MSEs of the distributed LS algorithm based on sampled data quickly converge to 0, while

MSEs of the non-cooperative LS algorithm can not. To some extent, this simulation example can demonstrate the cooperative effect of sensor networks: Multiple sensors can cooperate to finish the estimation task by information exchange even though none of them can due to the lack of adequate excitation.

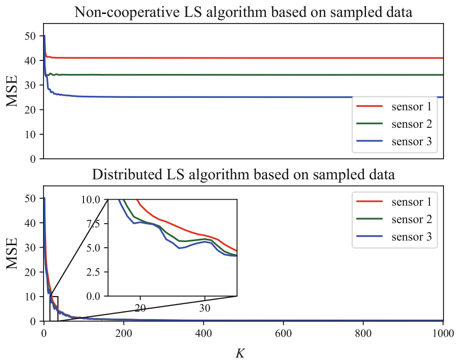

Figure 1 MSEs of the non-cooperative LS algorithm based on sampled data and distributed LS algorithm in this paper (Algorithm 1)

Then, we compare the performance of the distributed LS algorithm with constant and time-varying sampling intervals. The time-varying sampling interval is taken according to (50), and the constant sampling intervals are taken as $\delta_k = 0.4$, 0.6, 0.8, and 1. Figure 2 shows MSEs (averaged over 150 runs) of the algorithm in these two settings (constant and time-varying sampling intervals). It is clear that MSE of the algorithm with time-varying sampling interval gradually converges to 0 as the iteration step k goes to infinity. However, MSEs of the algorithm with constant sampling intervals can not converge to 0. Figure 3 shows the average accumulative regret $\frac{1}{Nn}\sum_{i=1}^{N}\sum_{k=0}^{n}\mathcal{R}_{t_k,i}$ (averaged over 150 runs) in the above two settings (constant and flexible sampling intervals). Compared with the average accumulative regret under fixed constant sampling intervals, the average accumulative regret under flexible sampling intervals has better performance.

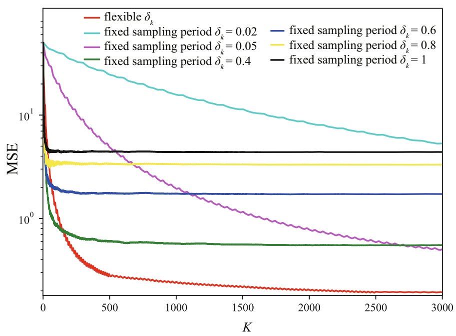

Figure 2 MSEs of distributed LS algorithm with constant and time-varying sampling intervals

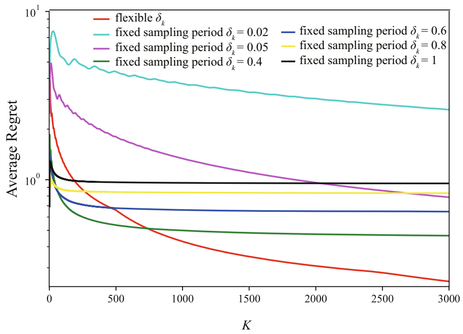

Figure 3 The average accumulative regrets of distributed LS algorithm with constant and time-varying sampling intervals

5 Concluding Remarks

For the continuous-time stochastic regression model described by stochastic differential equations, we proposed a distributed LS algorithm based on the sampled data over sensor networks to cooperatively estimate the unknown parameter vector. Under the cooperative non-persistent information condition, we established almost sure convergence of the proposed algorithm with properly sampling time interval. We also provided the upper bound on the accumulative regret of the adaptive predictor based on the proposed distributed LS algorithm without any excitation condition. Simulation results were conducted to illustrate the cooperative effect of multiple sensors. Some interesting problems deserve further investigation, e.g., how to design a suitable distributed algorithm based on sampled data to estimate time-varying parameters, and how to integrate the distributed control with distributed estimation algorithms.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Akyildiz I F and Vuran M C, *Wireless Sensor Networks*, John Wiley & Sons Inc, New York, 2010.

[2] Fiorelli E, Leonard N E, Bhatta P, et al., Multi-AUV control and adaptive sampling in Monterey Bay, *IEEE Journal of Oceanic Engineering*, 2006, **31**(4): 935–948.

[3] You J and Wu W, Online passive identifier for spatially distributed systems using mobile sensor networks, *IEEE Transactions on Control Systems Technology*, 2017, **25**(6): 2151–2159.

[4] Xie S, Zhang Y, and Guo L, Convergence of a distributed least squares, *IEEE Transactions on Automatic Control*, 2020, **66**(10): 4952–4959.

[5] Xie S and Guo L, Analysis of distributed adaptive filters based on diffusion strategies over sensor networks, *IEEE Transactions on Automatic Control*, 2018, **63**(11): 3643–3658.

[6] Demetriou M A, Guidance of mobile actuator-plus-sensor networks for improved control and estimation of distributed parameter systems, *IEEE Transactions on Automatic Control*, 2010, **55**(7): 1570–1584.

[7] Gan D and Liu Z, Performance analysis of the compressed distributed least squares algorithm, *Systems & Control Letters*, 2022, **164**: 105228.

[8] Geng N, Meng Q, Gong D, et al., How good are distributed allocation algorithms for solving urban search and rescue problems? A comparative study with centralized algorithms, *IEEE Transactions on Automation Science and Engineering*, 2018, **16**(1): 478–485.

[9] Bertrand A, Moonen M, and Sayed A H, Diffusion bias-compensated RLS estimation over adaptive networks, *IEEE Transactions on Signal Processing*, 2011, **59**(11): 5212–5224.

[10] Cattivelli F S, Lopes C G, and Sayed A H, Diffusion recursive least-squares for distributed estimation over adaptive networks, *IEEE Transactions on Signal Processing*, 2008, **56**(5): 1865–1877.

[11] Arablouei R, Dogancay K, Werner S, et al., Adaptive distributed estimation based on recursive least-squares and partial diffusion, *IEEE Transactions on Signal Processing*, 2014, **62**(14): 3510–3522.

[12] Yang Q, Zhang Z, and Fu M, Distributed weighted least-squares estimation for networked systems with edge measurements, *Automatica*, 2020, **120**: 91–109.

[13] Stankovic S S, Stankovic M S, and Stipanovic D M, Decentralized parameter estimation by consensus based stochastic approximation, *IEEE Transactions on Automatic Control*, 2010, **56**(3): 531–543.

[14] Gan D and Liu Z, Strong consistency of the distributed stochastic gradient algorithm, *Proceedings of the* 2019 *IEEE* 58*th Conference on Decision and Control*, Nice, 2019, 5082–5087.

[15] Abdolee R, Vakilian V, and Champagne B, Tracking performance and optimal adaptation step-sizes of diffusion-LMS networks, *IEEE Transactions on Control of Network Systems*, 2016, **5**(1): 67–78.

[16] Xie S and Guo L, Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition, *SIAM Journal on Control and Optimization*, 2018, **56**(5): 3404–3431.

[17] Gan D, Xie S, and Liu Z, Stability of the distributed Kalman filter using general random coefficients, *Science China Information Sciences*, 2021, **64**(7): 1–14.

[18] Sayed A H and Lopes C G, Distributed recursive least-squares strategies over adaptive networks, *Proceedings of the* 2006 *Fortieth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, 2006, 233–237.

[19] Orlov Y and Bentsman J, Adaptive distributed parameter systems identification with enforceable identifiability conditions and reduced-order spatial differentiation, *IEEE Transactions on Automatic Control*, 2000, **45**(2): 203–216.

[20] Li R G and Wu H N, Iterative approach with optimization-based execution scheme for parameter identification of distributed parameter systems and its application in secure communication, *IEEE Transactions on Circuits and Systems* I*: Regular Papers*, 2020, **67**(9): 3113–3126.

[21] Maghenem M, Lora A, and Panteley E, Formation-tracking control of autonomous vehicles under relaxed persistency of information conditions, *IEEE Transactions on Control Systems Technology*, 2017, **26**(5): 1860–1865.

[22] Chen W, Wen C, Hua S, et al., Distributed cooperative adaptive identification and control for a

group of continuous-time systems with a cooperative PE condition via consensus, *IEEE Transactions on Automatic Control*, 2013, **59**(1): 91–106.

[23] Papusha I, Lavretsky E, and Murray R M, Collaborative system identification via parameter consensus, *Proceedings of IEEE American Control Conference*, Portland, 2014, 13–19.

[24] Javed M U, Poveda J I, and Chen X, Information conditions for uniform exponential stability of the cooperative gradient algorithm over weakly connected digraphs, *IEEE Control Systems Letters*, 2021, **6**: 67–72.

[25] Manatakis D V and Manolakos E S, Predictive tracking of continuous object boundaries using sparse local estimates, *IEEE Access*, 2020, **8**: 152881–152899.

[26] Battilotti S, Cacace F, d'Angelo M, et al., Asymptotically optimal consensus-based distributed filtering of continuous-time linear systems, *Automatica*, 2020, **122**: 109189.

[27] Liang S, Zeng X, and Hong Y, Distributed sub-optimal resource allocation over weight-balanced graph via singular perturbation, *Automatica*, 2018, **95**: 222–228.

[28] Marelli D and Fu M, A continuous-time linear system identification method for slowly sampled data, *IEEE Transactions on Signal Processing*, 2010, **58**(5): 2521–2533.

[29] Pan S, González R A, Welsh J S, et al., Consistency analysis of the simplified refined instrumental variable method for continuous-time systems, *Automatica*, 2020, **113**: 108767.

[30] Greblicki W, Continuous-time hammerstein system identification from sampled data, *IEEE Transactions on Automatic Control*, 2006, **51**(7): 1195–1200.

[31] Yuan C and Wang C, Design and performance analysis of deterministic learning of sampled-data nonlinear systems, *Science China Information Sciences*, 2014, **57**(3): 1–18.

[32] Ortega R, Bobtsov A, and Nikolaev N, Parameter identification with finite-convergence time alertness preservation, *IEEE Control Systems Letters*, 2021, **6**: 205–210.

[33] Guo L, *Time-Varying Stochastic Systems, Stability and Adaptive Theory,* 2nd Edition, Science Press, Beijing, 2020.

[34] Godsil C and Royle G F, *Algebraic Graph Theory*, Springer-Verlag, Berlin, 2014.

[35] Johansson R, Identification of continuous-time models, *IEEE Transactions on Signal Processing*, 1994, **42**(4): 887–897.

[36] Oksendal B, *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin, 2013.

[37] Zhang J, Zhao C, and Guo L, On PID control theory for nonaffine uncertain stochastic systems, *Journal of Systems Science & Complexity*, 2023, **36**(1): 165–186.

[38] Hu X L and Welsh J S, Continuous-time model identification from filtered sampled data: Error analysis, *IEEE Transactions on Automatic Control*, 2020, **65**(10): 4005–4015.

[39] Zhu X, Gan D, and Liu Z, Performance analysis of least squares of continuous-time model based on sampled data, *IEEE Control Systems Letters*, 2022, **6**: 3086–3091.

[40] Hu X L and Welsh J S, Bias analysis of continuous-time model identification from filtered sample output data, *Proceedings of the IEEE 51st IEEE Conference on Decision and Control*, Maui, 2012, 208–213.