Performance Analysis of Least Squares of Continuous-Time Model Based on Sampling Data

Xinghua Zhu, Die Gan, and Zhixin Liu

***Abstract*—In this letter, we consider the parameter estimation problem of continuous-time linear stochastic regression models described by stochastic differential equations using the sampling data. An online least squares (LS) algorithm is proposed by minimizing the accumulative prediction error at discrete sampling time instants. By employing both the stochastic Lyapunov function and martingale estimate methods, we establish the convergence analysis of the proposed algorithm under conditions of the excitation of the sampling data and the sampling time interval. We also provide the upper bound of the accumulative regret for the adaptive predictor. A simulation example is given to verify our theoretical results.**

***Index Terms*—Continuous-time linear stochastic regression models, sampling data, parameter identification, convergence.**

I. Introduction

PARAMETER estimation or filtering is an important issue in areas of identification, adaptive control, and statistical learning, which enjoys practical applications in many engineering systems such as radar system [1] and power system [2]. A large number of works arise on the design and theoretical analysis of estimation and filtering algorithms, see [3]–[11] for some references.

In the past half century, considerable progresses have been made for the parameter estimation problem of discrete-time models. For time-invariant parameters, Moore Considered the well-known least squares in [3], and established the convergence analysis under the persistent excitation (PE) condition. Furthermore, the PE condition was generalized to the weakest possible excitation condition in [4]. There were also some investigations on other algorithms such as the stochastic gradient algorithm, the weighted LS algorithm, the maximum likelihood algorithm and Bayes methods [5], [11]. For time-varying parameters, the stability and performance analysis of some algorithms like the least mean square (LMS) algorithm, the forgetting factor least squares and the Kalman filter were established in [6]–[8], [12].

We note that dynamic systems in physics and engineering are naturally described by (stochastic) differential equations according to physical laws (cf., [13], [14]). There were some results for estimation algorithms of continuous-time systems based on the continuous-time data (cf., [9], [10], [15], [16]). For example, [15] considered the extended least squares of continuous-time estimation problems without noise or disturbance; The convergence of the continuous-time least squares algorithm was investigated in [16]. See [17], [18] for more information about other estimation algorithms for continuous-time systems.

However, with applications of the communication and computer technology, we can only measure the digital or discrete-time data, which inspires the research of the parameter estimation problem of continuous-time models based on sampling data. Some asymptotic analyses for the estimation algorithms were obtained under assumptions on regression vectors. For example, Yuan and Wang [19] designed an adaptive neural network weight estimator, and analyzed the convergence of the algorithm with uniformly bounded regressors satisfying PE condition. Greblicki in [20] investigated the convergence of the estimation algorithm using stationary signals. Ortega *et al.* in [21] presented a discrete-time gradient descent algorithm, and they established the convergence of the algorithm under the PE condition. The above estimation algorithms are designed based on the discretized model where the approximation errors are not taken into consideration in the convergence analysis. Observing this fact, Hu and Welsh in [22] studied the LS algorithm and the convergence analysis was established under the PE condition of regressors. How to relax the PE condition is an unresolved issue.

Manuscript received March 21, 2022; revised May 29, 2022; accepted June 1, 2022. Date of publication June 10, 2022; date of current version June 22, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0703800; in part by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant XDA27000000; in part by the ŞNatural Science Foundation of China under Grant U21B6001; and in part by the National Science Foundation of Shandong Province under Grant ZR2020ZD26. Recommended by Senior Editor V. Ugrinovskii. *(Corresponding author: Zhixin Liu.)*

The authors are with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: zxh@amss.ac.cn; gandie@amss.ac.cn; lzx@amss.ac.cn).

Digital Object Identifier 10.1109/LCSYS.2022.3182040

In this letter, we estimate the problem of a continuous-time linear stochastic regression model based on the sampling data. The LS algorithm is proposed by minimizing the accumulative sampling prediction error. By employing the stochastic Lyapunov function and martingale estimate methods, we establish the upper bound of the estimation error which is caused by system noise and the approximation error of the sampling time interval. The consistency result of the algorithm is obtained under conditions of the flexible sampling time interval and non-persistent excitation of regression signals. Finally, we discuss the upper bound of accumulative regret of the adaptive predictor without any excitation condition on regressors. We remark that our results are obtained without relying on assumptions of independency or stationarity of regression signals, which makes our results applicable to feedback systems.

The rest of this letter is organized as follows. Section II gives the problem formulation. The asymptotic results of the proposed algorithm are provided in Section III. A simulation example is given in Section IV and the concluding remarks are made in Section V.

II. Problem Formulation

A. Some Preliminaries

1) Matrix Theory: For an $m \times m$-dimensional matrix A, we use $\|A\|$ to represent the Euclidean norm, i.e., $\|A\| \triangleq (\lambda_{\max}\{AA^{\tau}\})^{\frac{1}{2}}$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix, and τ denotes the transpose of the matrix. Correspondingly, the smallest eigenvalue of the matrix is denoted as $\lambda_{\min}(\cdot)$. The determinant of the matrix A is denoted as $det(A)$. For a matrix sequence $\{A_k, k \geq 0\}$ and a positive scalar sequence $\{a_k, k \geq 0\}$, if there exists a positive constant C, such that $\|A_k\| \leq Ca_k$ holds for all $k \geq 0$, then we say $A_k = O(a_k)$; and if $\lim_{k\to\infty} \frac{\|A_k\|}{a_k} = 0$, then we say $A_k = o(a_k)$. For matrices A, B, C and D are with suitable dimensions, we have the following matrix inverse formula (cf., [9])

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \quad (1)$$

provided that the relevant matrices are invertible.

2) Stochastic Processes: Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\{\mathcal{F}_t, t \geq 0\}$ be a nondecreasing family of sub-σ-algebras of $\mathcal{F}$. A continuous, adapted, stochastic process $\{W_t, \mathcal{F}_t; t \geq 0\}$ is called a standard Wiener process if: (1) $W_0 = 0$ holds almost surely (a.s.); (2) $\{W_t, \mathcal{F}_t; t \geq 0\}$ is an independent incremental process, and for any $0 \leq s < t < \infty$, the increment $W_t - W_s$ is independent of $\mathcal{F}_s$ and obeys the normal distribution with mean zero and variance $t - s$.

An adapted stochastic sequence $\{v_k, \mathcal{F}_k\}$ is called a martingale difference sequence if $\mathbb{E}(v_{k+1}|\mathcal{F}_k) = 0$, where $\mathbb{E}(\cdot|\cdot)$ denotes the conditional expectation operator. For a martingale difference sequence, we have the following martingale estimation theorem.

Lemma 1 [10, Th. 2.8]: Suppose that $\{a_k, \mathcal{F}_k\}$ is an adapted process (a_k can be a number or a matrix), and $\{v_k, \mathcal{F}_k\}$ is a martingale difference sequence (v_k can be a number or a matrix) satisfying $\sup_n \mathbb{E}[\|v_{n+1}\|^{\beta}|\mathcal{F}_n] < \infty$ with $\beta \in (0, 2]$. Then for any $\eta > 0$, we have

$$\sum_{k=0}^{n} a_k v_{k+1} = O\Big(Q_n(\beta) \log^{\frac{1}{\beta}+\eta}(Q_n(\beta) + e)\Big) \ a.s.,$$

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

B. Continuous-Time Stochastic Regression Model

Let us consider the continuous-time stochastic linear regression model described by the following stochastic differential equation,

$$y_t = S\theta^{\tau}\phi_t + v_t, \ \ t \geq 0, \quad (2)$$

where S is the integral operator (i.e., $S\phi_t = \int_0^t \phi_s ds$), y_t is the scalar observation at time t, θ is an unknown l-dimensional parameter vector to be estimated, and ϕ_s is an l-dimensional stochastic regressor, and the system noise v_t is a standard Wiener process. We can see that many models such as the continuous-time ARX model and ARMAX model (cf., [23]) can be included by (2).

With the application of computer and communication technology in many practical scenarios, we can only receive the discrete-time sampling data $\{y_{t_k}, \phi_{t_k}\}_{k=0}^{\infty}$ rather than continuous-time signals, where t_k is the k-th sampling time instant. In this letter, we aim at designing an algorithm to estimate the unknown parameter vector θ by using the sampling data, and providing the asymptotic results for the proposed algorithm.

C. The LS Algorithm Based on Sampling Data

In order to construct the algorithm to estimate the unknown parameter vector θ based on the sampling data, we introduce the accumulative sampling prediction error $\sum_{j=0}^{k}(y_{t_{j+1}} - y_{t_j} - \theta^{\tau}\delta_j\phi_{t_j})^2$, where $\delta_j = t_{j+1} - t_j$, and t_j represents the j-th sampling time instant. By minimizing it, we can obtain the following LS algorithm to estimate θ at time t_k,

$$\theta_{t_{k+1}} = \Big(\sum_{j=0}^{k} \delta_j^2 \phi_{t_j}\phi_{t_j}^{\tau}\Big)^{-1} \sum_{j=0}^{k} \delta_j \phi_{t_j}(y_{t_{j+1}} - y_{t_j}) \quad (3)$$

Denote

$$P_{t_{k+1}} = \Big(\sum_{j=0}^{k} \delta_j^2 \phi_{t_j}\phi_{t_j}^{\tau}\Big)^{-1}. \quad (4)$$

By the matrix inverse formula (1), we have

$$P_{t_{k+1}} = P_{t_k} - \delta_k^2 a_{t_k} P_{t_k}\phi_{t_k}\phi_{t_k}^{\tau}P_{t_k}, \quad (5)$$

$$a_{t_k} = \frac{1}{1 + \delta_k^2 \phi_{t_k}^{\tau} P_{t_k}\phi_{t_k}}. \quad (6)$$

Substituting (5) into (3) yields the following recursive least squares algorithm,

$$\theta_{t_{k+1}} = \theta_{t_k} + \delta_k a_{t_k} P_{t_k}\phi_{t_k}\big\{y_{t_{k+1}} - y_{t_k} - \phi_{t_k}^{\tau}\theta_{t_k}\delta_k\big\}. \quad (7)$$

III. Asymptotic Results of the Algorithm

A. Convergence Analysis

In this section, we will investigate the asymptotic properties of the proposed algorithm (5)-(7). Denote $\widetilde{\theta}_{t_k}$ as the estimate error at the time instant t_k, i.e., $\widetilde{\theta}_{t_k} = \theta - \theta_{t_k}$. Substituting (2) into (7), we can obtain the following error equation,

$$\begin{aligned}\widetilde{\theta}_{t_{k+1}} &= \theta - \theta_{t_{k+1}} \\ &= \theta - \theta_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\left\{\int_{t_k}^{t_{k+1}}\phi_s^\tau ds\theta \right.\\ &\qquad\left. - \phi_{t_k}^\tau\theta_{t_k}\delta_k + v_{t_{k+1}} - v_{t_k}\right\} \\ &= \widetilde{\theta}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\{\phi_{t_k}^\tau\widetilde{\theta}_{t_k}\delta_k + \Delta_{t_{k+1}}^\tau\theta + \bar{v}_{t_{k+1}}\},\end{aligned} \tag{8}$$

where $\bar{v}_{t_{k+1}} = v_{t_{k+1}} - v_{t_k}$ and

$$\Delta_{t_{k+1}} = \textstyle\int_{t_k}^{t_{k+1}}\phi_s ds - \phi_{t_k}\delta_k. \tag{9}$$

By the fact that $\{v_t, \mathcal{F}_t\}$ is a standard Wiener process, we have the following properties:

(1) for $k \geq 0$, $\bar{v}_{t_k}$ obeys a normal distribution with mean 0 and variance δ_k;

(2) $\{\bar{v}_{t_{k+1}}, \mathcal{F}_{t_k}, k \geq 0\}$ is a martingale difference sequence, and $0 < \mathbb{E}[|\bar{v}_{t_{k+1}}|^\beta|\mathcal{F}_{t_k}] < \infty$ holds a.s. for any constant $\beta \geq 2$, where $F_{t_k} = \sigma\{\phi_s, v_s, s \leq t_k\}$.

In many existing results on the estimation of the unknown parameter vector of continuous-time systems based on the sampling data (cf., [19], [24]), continuous-time systems are often discretized, and the approximation error Δ_{t_k} caused by discretization is seldom considered. Different from these studies, we will consider the effect of the approximation error Δ_{t_k} on the estimation performance of algorithm, see the following proposition.

Proposition 1: For the dynamical system (2), the LS algorithm (5)-(7) satisfies the following relationship,

$$\begin{aligned}&\tilde{\theta}_{t_{n+1}}^\tau P_{t_{n+1}}^{-1}\tilde{\theta}_{t_{n+1}} + (\frac{1}{2} + o(1))\sum_{k=0}^n a_{t_k}\delta_k^2(\phi_{t_k}^\tau\tilde{\theta}_{t_k})^2 \\ &= 4\sum_{k=0}^n(\Delta_{t_{k+1}}^\tau\theta)^2 + O(\log|P_{t_{n+1}}^{-1}|) \quad a.s.\end{aligned}$$

Proof: By (8), we have

$$\widetilde{\theta}_{t_{k+1}} = (I - a_{t_k}\delta_k^2 P_{t_k}\phi_{t_k}\phi_{t_k}^\tau)\widetilde{\theta}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta + \bar{v}_{t_{k+1}}). \tag{10}$$

Multiply both sides of (5) by $P_{t_k}^{-1}$, we have

$$P_{t_{k+1}}P_{t_k}^{-1} = I - a_{t_k}\delta_k^2 P_{t_k}\phi_{t_k}\phi_{t_k}^\tau. \tag{11}$$

Substituting (11) into (10), we can obtain that $\widetilde{\theta}_{t_{k+1}} = P_{t_{k+1}}P_{t_k}^{-1}\widetilde{\theta}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta + \bar{v}_{t_{k+1}})$, which is equivalent to the following equation,

$$P_{t_{k+1}}^{-1}\tilde{\theta}_{t_{k+1}} = P_{t_k}^{-1}\tilde{\theta}_{t_k} - a_{t_k}\delta_k P_{t_{k+1}}^{-1}P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta + \bar{v}_{t_{k+1}}). \tag{12}$$

Now, we consider the Lyapunov candidate function $W_{t_k} = \tilde{\theta}_{t_k}^\tau P_{t_k}^{-1}\tilde{\theta}_{t_k}$. From (8) and (12), we have

$$\begin{aligned}W_{t_{k+1}} &= \tilde{\theta}_{t_{k+1}}^\tau P_{t_{k+1}}^{-1}\tilde{\theta}_{t_{k+1}} \\ &= \left[\widetilde{\theta}_{t_k} - a_{t_k}\delta_k^2 P_{t_k}\phi_{t_k}\phi_{t_k}^\tau\widetilde{\theta}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta + \bar{v}_{t_{k+1}})\right]^\tau \\ &\quad\cdot\left[P_{t_k}^{-1}\tilde{\theta}_{t_k} - a_{t_k}\delta_k P_{t_{k+1}}^{-1}P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta + \bar{v}_{t_{k+1}})\right].\end{aligned}$$

Note that $P_{t_{k+1}}^{-1} = P_{t_k}^{-1} + \delta_k^2\phi_{t_k}\phi_{t_k}^\tau$. By this equation and the definition of a_{t_k} in (6), we have $P_{t_{k+1}}^{-1}P_{t_k} = I + \delta_k^2\phi_{t_k}\phi_{t_k}^\tau P_{t_k}$ and $1 - a_{t_k} = a_{t_k}\delta_k^2\phi_{t_k}^\tau P_{t_k}\phi_{t_k}$. From these two equations, we can get the following equation,

$$\begin{aligned}W_{t_{k+1}} &= W_{t_k} - a_{t_k}\delta_k^2(\phi_{t_k}^\tau\tilde{\theta}_{t_k})^2 - 2a_{t_k}\delta_k\phi_{t_k}^\tau\tilde{\theta}_{t_k}\Delta_{t_{k+1}}^\tau\theta \\ &\quad - 2a_{t_k}\delta_k\phi_{t_k}^\tau\tilde{\theta}_{t_k}\bar{v}_{t_{k+1}} + a_{t_k}\delta_k^2\phi_{t_k}^\tau P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta)^2 \\ &\quad + a_{t_k}\delta_k^2\phi_{t_k}^\tau P_{t_k}\phi_{t_k}\bar{v}_{t_{k+1}}^2 + 2a_{t_k}\delta_k^2\phi_{t_k}^\tau P_{t_k}\phi_{t_k}\Delta_{t_{k+1}}^\tau\theta\bar{v}_{t_{k+1}}.\end{aligned} \tag{13}$$

By summing both sides of (13) from $k = 0$ to n, we have

$$\begin{aligned}&W_{t_{n+1}} - W_0 + \sum_{k=0}^n a_{t_k}\delta_k^2(\phi_{t_k}^\tau\tilde{\theta}_{t_k})^2 \\ &= -2\sum_{k=0}^n a_{t_k}\delta_k\phi_{t_k}^\tau\tilde{\theta}_{t_k}\bar{v}_{t_{k+1}} - 2\sum_{k=0}^n a_{t_k}\delta_k\phi_{t_k}^\tau\tilde{\theta}_{t_k}\Delta_{t_{k+1}}^\tau\theta \\ &\quad + \sum_{k=0}^n a_{t_k}\delta_k^2\phi_{t_k}^\tau P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta)^2 + \sum_{k=0}^n a_{t_k}\delta_k^2\phi_{t_k}^\tau P_{t_k}\phi_{t_k}\bar{v}_{t_{k+1}}^2 \\ &\quad + 2\sum_{k=0}^n a_{t_k}\delta_k^2\phi_{t_k}^\tau P_{t_k}\phi_{t_k}\Delta_{t_{k+1}}^\tau\theta\bar{v}_{t_{k+1}}.\end{aligned} \tag{14}$$

For the first term on the right hand side (RHD) of (14), noticing $\delta_k a_{t_k}\phi_{t_k}^\tau\tilde{\theta}_{t_k} \in \mathcal{F}_{t_k}$ and Lemma 1, we can derive that for all $\eta > 0$,

$$\sum_{k=0}^n\delta_k a_{t_k}\phi_{t_k}^\tau\tilde{\theta}_{t_k}\bar{v}_{t_{k+1}} = O(1) + o\left(\sum_{k=0}^n\delta_k^2 a_{t_k}(\phi_{t_k}^\tau\tilde{\theta}_{t_k})^2\right). \tag{15}$$

Applying the Hölder inequality, we see that the second term on the RHD of (14) satisfies the following inequality,

$$\begin{aligned}&2\sum_{k=0}^n a_{t_k}\delta_k\phi_{t_k}^\tau\tilde{\theta}_{t_k}\Delta_{t_{k+1}}^\tau\theta \leq \{\sum_{k=0}^n(a_{t_k}\delta_k\phi_{t_k}^\tau\tilde{\theta}_{t_k})^2\}^{\frac{1}{2}}\cdot\{\sum_{k=0}^n(2\Delta_{t_{k+1}}^\tau\theta)^2\}^{\frac{1}{2}} \\ &\leq \frac{1}{2}\sum_{k=0}^n a_{t_k}\delta_k^2(\phi_{t_k}^\tau\tilde{\theta}_{t_k})^2 + 2\sum_{k=0}^n(\Delta_{t_{k+1}}^\tau\theta)^2,\end{aligned} \tag{16}$$

where $a_{t_k} \leq 1$ is used in the last inequality. By (6), we have $\delta_k^2 a_{t_k}\phi_{t_k}^\tau P_{t_k}\phi_{t_k} \leq 1$. Thus, the third term on the RHD of (14) satisfies

$$\sum_{k=0}^n\delta_k^2 a_{t_k}\phi_{t_k}^\tau P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta)^2 \leq \sum_{k=0}^n(\Delta_{t_{k+1}}^\tau\theta)^2. \tag{17}$$

Using the mean inequality and (17), we have for the last term on the RHD of (14),

$$\begin{aligned}&2\sum_{k=0}^n\delta_k^2 a_{t_k}\phi_{t_k}^\tau P_{t_k}\phi_{t_k}\Delta_{t_{k+1}}^\tau\theta\bar{v}_{t_{k+1}} \\ &\leq \sum_{k=0}^n\delta_k^2 a_{t_k}\phi_{t_k}^\tau P_{t_k}\phi_{t_k}(\Delta_{t_{k+1}}^\tau\theta)^2 + \sum_{k=0}^n\delta_k^2 a_{t_k}\phi_{t_k}^\tau P_{t_k}\phi_{t_k}\bar{v}_{t_{k+1}}^2 \\ &\leq \sum_{k=0}^n(\Delta_{t_{k+1}}^\tau\theta)^2 + \sum_{k=0}^n\delta_k^2 a_{t_k}\phi_{t_k}^\tau P_{t_k}\phi_{t_k}\bar{v}_{t_{k+1}}^2.\end{aligned} \tag{18}$$

Substituting (16)-(18) into (14) yields

$$W_{t_{n+1}}+(\frac{1}{2}+o(1))\sum_{k=0}^{n}a_{t_k}\delta_k^2(\phi_{t_k}^{\tau}\tilde{\theta}_{t_k})^2 = O(1)+4\sum_{k=0}^{n}(\Delta_{t_{k+1}}^{\tau}\theta)^2+\sum_{k=0}^{n}\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}\bar{v}_{t_{k+1}}^2. \quad (19)$$

For the last term on the RHD of (19), it follows that

$$\begin{aligned}\sum_{k=0}^{n}\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}\bar{v}_{t_{k+1}}^2 &= \sum_{k=0}^{n}\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}\{\bar{v}_{t_{k+1}}^2-\mathbb{E}[\bar{v}_{t_{k+1}}^2|\mathcal{F}_{t_k}]\}\\ &+\sum_{k=0}^{n}\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}\mathbb{E}[\bar{v}_{t_{k+1}}^2|\mathcal{F}_{t_k}]. \quad (20)\end{aligned}$$

Note that $\{\bar{v}_{t_{k+1}}^2-\mathbb{E}[\bar{v}_{t_{k+1}}^2|\mathcal{F}_{t_k}],\mathcal{F}_{t_k}\}$ is a martingale difference sequence, by the C_r-inequality and the Lyapunov inequality, we see that for any $\alpha\in(1,2]$, $\sup_k\mathbb{E}\{[\bar{v}_{t_{k+1}}^2-\mathbb{E}(\bar{v}_{t_{k+1}}^2|\mathcal{F}_{t_k})]^{\alpha}|\mathcal{F}_{t_k}\}\le 2\sup_k\mathbb{E}[|\bar{v}_{t_{k+1}}|^{2\alpha}|\mathcal{F}_{t_k}]+2\sup_k\mathbb{E}[(\mathbb{E}[|\bar{v}_{t_{k+1}}|^2|\mathcal{F}_{t_k}])^{\alpha}|\mathcal{F}_{t_k}]\le 4\sup_k\mathbb{E}[|\bar{v}_{t_{k+1}}|^{2\alpha}|\mathcal{F}_{t_k}]<\infty$ holds almost surely. From Lemma 1, we can derive that for any $\eta>0$,

$$\sum_{k=0}^{n}\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}\{\bar{v}_{t_{k+1}}^2-\mathbb{E}[\bar{v}_{t_{k+1}}^2|\mathcal{F}_{t_k}]\} = O\Big(M_n(\alpha)\log^{\frac{1}{\alpha}+\eta}(M_n(\alpha)+e)\Big),\quad a.s., \quad (21)$$

where $M_n(\alpha)\triangleq\{\sum_{k=0}^{n}(\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k})^{\alpha}\}^{\frac{1}{\alpha}}$. Thus, by Lemma 1 again, we have for any $\eta>0$,

$$\sum_{k=0}^{n}a_{t_k}\delta_k^2\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}\{\bar{v}_{t_{k+1}}^2-\mathbb{E}[\bar{v}_{t_{k+1}}^2|\mathcal{F}_{t_k}]\} = O\Big(M_n(\alpha)\log^{\frac{1}{\alpha}+\eta}(M_n(\alpha)+e)\Big),\quad a.s., \quad (22)$$

where $M_n(\alpha)\triangleq\{\sum_{k=0}^{n}(\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k})^{\alpha}\}^{\frac{1}{\alpha}}$. In the following, we need to analyze $M_n(\alpha)$. According to (4), we have

$$P_{t_{k+1}}^{-1}=P_{t_k}^{-1}+\delta_k^2\phi_{t_k}\phi_{t_k}^{\tau}=P_{t_k}^{-1}(I+\delta_k^2P_{t_k}\phi_{t_k}\phi_{t_k}^{\tau}). \quad (23)$$

Taking the determinant from both sides of (23), we have

$$|P_{t_{k+1}}^{-1}|=|P_{t_k}^{-1}|(1+\delta_k^2\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}).$$

Thus, it follows that $\delta_k^2\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}=\frac{|P_{t_{k+1}}^{-1}|-|P_{t_k}^{-1}|}{|P_{t_k}^{-1}|}$ and $a_{t_k}\delta_k^2\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}=\frac{|P_{t_{k+1}}^{-1}|-|P_{t_k}^{-1}|}{|P_{t_{k+1}}^{-1}|}$. Then, it follows that

$$\begin{aligned}\sum_{k=0}^{n}\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k} &= \sum_{k=0}^{n}\frac{|P_{t_{k+1}}^{-1}|-|P_{t_k}^{-1}|}{|P_{t_{k+1}}^{-1}|}\\ &\le\sum_{k=0}^{n}\int_{|P_{t_k}^{-1}|}^{|P_{t_{k+1}}^{-1}|}\frac{dx}{x}=\log|P_{t_{n+1}}^{-1}|+\log|P_0^{-1}|. \quad (24)\end{aligned}$$

According to the inequality $a_{t_k}\delta_k^2\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}<1$ and (24), we can deduce the following equation

$$M_n(\alpha)=O(1)+o(\log|P_{t_{n+1}}^{-1}|), \quad (25)$$

where $\alpha>1$ is used. Thus, from (20), (24) and (25), it can be deduced that

$$\begin{aligned}&\sum_{k=0}^{n}a_{t_k}\delta_k^2\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}\bar{v}_{t_{k+1}}^2=O(1)+o(\log|P_{t_{n+1}}^{-1}|)\\ &+\sum_{k=0}^{n}\delta_{k+1}^2\delta_k^2 a_{t_k}\phi_{t_k}^{\tau}P_{t_k}\phi_{t_k}=O(\log|P_{t_{n+1}}^{-1}|).\end{aligned}$$

This completes the proof of the proposition. ■

To proceed with our analysis for strong consistency of the proposed algorithm, we introduce some assumptions on the sampling interval and regression vectors.

Assumption 1: The sampling interval δ_k is designed to satisfy $\sum_{k=0}^{n}\delta_k^2=\infty$ and $\sum_{k=0}^{n}\delta_k^4<\infty$.

Assumption 2: The stochastic regression signal ϕ_t in model (2) is $\mathcal{F}_t$-measurable, where $\mathcal{F}_t=\sigma\{\phi_s,v_s,s\le t\}$ is a family of nondecreasing σ-algebras. And ϕ_t is also Lipshitz continuous for almost all sample paths, i.e., there exists a positive constant L such that for all $t>0$ and $s>0$, $\|\phi_t-\phi_s\|\le L|t-s|$ holds almost surely.

Assumption 3 (Non-Persistent Excitation Condition): The growth rate of $\log(\lambda_{\max}\{P_{t_{n+1}}^{-1}\})$ is slower than $\lambda_{\min}\{P_{t_{n+1}}^{-1}\}$, that is

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

holds, where $R_n=1+\sum_{k=0}^{n}\delta_k^2\|\phi_{t_k}\|^2$, and $\lambda_{\min}^n=\lambda_{\min}\{P_0^{-1}+\sum_{k=0}^{n}\delta_k^2\phi_{t_k}\phi_{t_k}^{\tau}\}$.

Remark 1: Chen and Guo in [16] has proved the convergence of the standard continuous-time LS algorithm of the model (2) under the excitation condition

$$\lim_{t\to\infty}\frac{\log\left(e+\int_0^t\|\phi_s\|^2ds\right)}{\lambda_{\min}\left\{P_0^{-1}+\int_0^t\phi_s\phi_s^{\tau}ds\right\}}=0. \quad (26)$$

We see that Assumption 3 can be degenerated to (26) as $\delta_k\to 0$.

Based on the above proposition and assumptions, we can obtain the following theorem on the upper bound of the estimation error.

Theorem 1: Under Assumption 2, we have

$$\|\widetilde{\theta}_{t_{n+1}}\|^2=O\Big(\frac{\log R_n}{\lambda_{\min}^n}\Big)+L^2\|\theta\|^2\frac{\sum_{k=0}^{n}\delta_k^4}{\lambda_{\min}^n}\quad a.s.$$

Proof: By Proposition 1, we have

$$\tilde{\theta}_{t_{n+1}}^{\tau}P_{t_{n+1}}^{-1}\tilde{\theta}_{t_{n+1}}=4\sum_{k=0}^{n}(\Delta_{t_{k+1}}^{\tau}\theta)^2+O(\log|P_{t_{n+1}}^{-1}|),$$

by which we can deduce the following equation,

$$\|\widetilde{\theta}_{t_{n+1}}\|^2=O\Big(\frac{\log|P_{t_{n+1}}^{-1}|}{\lambda_{\min}^n}\Big)+\frac{4\sum_{k=0}^{n}(\Delta_{t_{k+1}}^{\tau}\theta)^2}{\lambda_{\min}^n}\quad a.s., \quad (27)$$

where $\lambda_{\min}^n$ is defined in Assumption 3. Using (9) and Assumption 2, we obtain the following inequality

$$\begin{aligned}\|\Delta_{t_{k+1}}\| &= \left\|\int_{t_k}^{t_{k+1}}(\phi_s-\phi_{t_k})ds\right\|\\ &\le\int_{t_k}^{t_{k+1}}L(s-t_k)ds=\frac{L}{2}\delta_k^2. \quad (28)\end{aligned}$$

Noticing (4), we have,

$$\log|P_{t_{n+1}}^{-1}| = \sum_{i=0}^{l} \log \lambda_i(P_{t_{n+1}}^{-1}) \\ \leq l \log \lambda_{\max}(P_{t_{n+1}}^{-1}) = O(\log R_n), \quad (29)$$

where l is the dimension of the vector ϕ_{t_k} and R_n is defined in Assumption 3. Substituting (28) and (29) into (27) yields

$$\|\widetilde{\theta}_{t_{n+1}}\|^2 = O\left(\frac{\log R_n}{\lambda_{\min}^n}\right) + L^2\|\theta\|^2 \frac{\sum_{k=0}^n \delta_k^4}{\lambda_{\min}^n}. \quad (30)$$

This completes the proof of the theorem. ■

Remark 2: Under the condition of Theorem 1, if Assumptions 3 and 1 are satisfied, we can derive that $\|\widetilde{\theta}_{t_{n+1}}\|^2 \to 0$ almost surely as $n \to +\infty$, which implies that the estimation $\theta_{t_{n+1}}$ converges to the true parameter vector θ almost surely.

B. Analysis of the Regret

Denote $\Delta y_{t_k} = y_{t_{k+1}} - y_{t_k}$. At the sampling instant $t_k \geq 0$, the best prediction to the change of y_t between the time instants t_k and t_{k+1} is $\mathbb{E}[\Delta y_{t_k}|\mathcal{F}_{t_k}] = \mathbb{E}[(\int_{t_k}^{t_{k+1}} \phi_s^\tau ds \cdot \theta)|\mathcal{F}_{t_k}]$ since $\mathbb{E}[(v_{t_{k+1}} - v_{t_k})|\mathcal{F}_{t_k}] = 0$. The difference between the best prediction and the adaptive prediction can be regarded as regret, denoted by $\mathcal{R}_{t_k}$, i.e.,

$$\mathcal{R}_{t_k} = \{\mathbb{E}[\Delta y_{t_k}|\mathcal{F}_{t_k}] - \Delta\hat{y}_{t_k}\}^2 \\ = \left\{\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}} \phi_s^\tau ds \cdot \theta - \delta_k \phi_{t_k}^\tau \theta_{t_k}\right)\middle|\mathcal{F}_{t_k}\right]\right\}^2, \quad (31)$$

where $\phi_{t_k} \in \mathcal{F}_{t_k}$ is used in the last inequality. We develop the following theorem concerning the upper bound of the accumulative regret.

Theorem 2: Under Assumption 2, the accumulative regret has the following upper bound,

$$\sum_{k=0}^n \mathcal{R}_{t_k} = \frac{9L^2}{2}\|\theta\|^2 \sum_{k=0}^n \delta_k^4 + O(\log R_n) \quad a.s.,$$

provided that $\delta_k^2 \phi_{t_k}^\tau P_{t_k} \phi_{t_k} = O(1)$.

Proof: From (31) and Jensen inequality, we have

$$\sum_{k=0}^n \mathcal{R}_{t_k} = \sum_{k=0}^n \left\{\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}} \phi_s^\tau ds \cdot \theta - \delta_k \phi_{t_k}^\tau \theta_{t_k}\right)|\mathcal{F}_{t_k}\right]\right\}^2 \\ \leq \sum_{k=0}^n \mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}} \phi_s^\tau ds \cdot \theta - \delta_k \phi_{t_k}^\tau \theta_{t_k}\right)^2 \middle|\mathcal{F}_{t_k}\right]. \quad (32)$$

Noticing (6), we can derive the following equation,

$$\phi_{t_k}\phi_{t_k}^\tau = a_{t_k}\phi_{t_k}\phi_{t_k}^\tau + \phi_{t_k}(\delta_k^2 a_{t_k}\phi_{t_k}^\tau P_{t_k}\phi_{t_k})\phi_{t_k}^\tau. \quad (33)$$

Substituting (28) and (33) into (32), we can obtain the following inequality,

$$\sum_{k=0}^n \mathcal{R}_{t_k} \leq \sum_{k=0}^n \mathbb{E}\left[\left(\Delta_{t_{k+1}}^\tau \theta + \delta_k \phi_{t_k}^\tau \tilde{\theta}_{t_k}\right)^2 \middle|\mathcal{F}_{t_k}\right] \\ \leq \sum_{k=0}^n \mathbb{E}\left[2(\Delta_{t_{k+1}}^\tau \theta)^2 + 2\delta_k^2 \tilde{\theta}_{t_k}^\tau \phi_{t_k}\phi_{t_k}^\tau \tilde{\theta}_{t_k} \middle|\mathcal{F}_{t_k}\right] \\ \leq \frac{L^2}{2}\|\theta\|^2 \sum_{k=0}^n \delta_k^4 + 2\sum_{k=0}^n a_{t_k}\delta_k^2(\phi_{t_k}^\tau \tilde{\theta}_{t_k})^2 \\ +2\sum_{k=0}^n \delta_k^2 \tilde{\theta}_{t_k}^\tau \phi_{t_k}(a_{t_k}\delta_k^2 \phi_{t_k}^\tau P_{t_k}\phi_{t_k})\phi_{t_k}^\tau \tilde{\theta}_{t_k}, \quad (34)$$

where $\phi_{t_k} \in \mathcal{F}_{t_k}$ is used. From Proposition 1, (28) and (29), the following equation can be obtained,

$$\sum_{k=0}^n a_{t_k}\delta_k^2 \tilde{\theta}_{t_k}^\tau \phi_{t_k}\phi_{t_k}^\tau \tilde{\theta}_{t_k} = 4\sum_{k=0}^n (\Delta_{t_{k+1}}^\tau \theta)^2 + O(\log|P_{t_{n+1}}^{-1}|) \\ = L^2\|\theta\|^2 \sum_{k=0}^n \delta_k^4 + O\Big(\log R_n\Big) \quad a.s. (35)$$

Noticing $\delta_k^2 \phi_{t_k}^\tau P_{t_k}\phi_{t_k} = O(1)$ and substituting this and (35) into (34), we have

$$\sum_{k=0}^n \mathcal{R}_{t_k} = \frac{9L^2}{2}\|\theta\|^2 \sum_{k=0}^n \delta_k^4 + O(\log R_n).$$

This completes the proof of the theorem. ■

Remark 3: We remark that under the condition of Theorem 2, if Assumption 1 is further satisfied, we can obtain that $\sum_{k=0}^n \mathcal{R}_{t_k} = O(\log R_n)$, which is the minimum order of magnitude that one may at most expect to achieve for the accumulative regret (cf. [25]).

IV. Simulation Results

In this section, we provide a simulation example to verify the effectiveness of the LS algorithm based on sampling data.

Consider the dynamic system (2). We set the 3-dimensional regression vector ϕ_t as $\phi_t = [1 + \sin(\frac{2\pi}{3}t), 2 + \cos(\frac{2\pi}{3}t), 1 + \sin(\frac{4\pi}{3}t)]^\tau$. It can be verified that Assumptions 2 and 3 hold. The system noise $\{v_t \geq 0\}$ is a standard, one-dimensional Wiener process. The exact value and the initial value of parameter vector are given as $\theta = [3, 4, 5]$ and $\theta_0 = [0, 0, 0]$. The initial covariance matrix is $P_0 = diag(3, 3, 3)$. The sampling interval δ_k is taken according to the following manner

$$\delta_k = \frac{a_0}{b_0^{\lceil \log_{b_0^2} (b_0^2-1)\lceil \frac{k}{c_0}\rceil + b_0^2 \rceil - 1}},$$

where $\lceil\cdot\rceil$ is a round up operator. It is clear that for any given $a_0 > 0, b_0 > 0, c_0 > 1$, $\sum_{k=0}^n \delta_k^2 = \infty$ and $\sum_{k=0}^n \delta_k^4 = O(\sum_{k=0}^n \frac{1}{b_0^{4k}} b_0^{2k}) < \infty$ hold. Thus, the time interval δ_k satisfies Assumption 1. Here we choose $a_0 = 0.12, b_0 = 1.2, c_0 = 200$. We compare our algorithm with cases where the sampling intervals are constants and taken as $\delta_k = 0.4$, $\delta_k = 0.6$, $\delta_k = 0.8$ and $\delta_k = 1$, respectively.

The mean square errors (MSEs) (averaged over 100 runs) of these two scenarios are shown in Figure 1, from which we see that the MSE of the algorithm with fixed constant sampling interval increases as δ_k increases, while the MSE of the proposed algorithm based on flexible sampling interval gradually converges to 0 as the iteration step k tends to infinity. The average accumulative regret $\frac{1}{n}\sum_{k=0}^n \mathcal{R}_{t_k}$ (averaged over 100 runs) of the above two scenarios (flexible and fixed constant

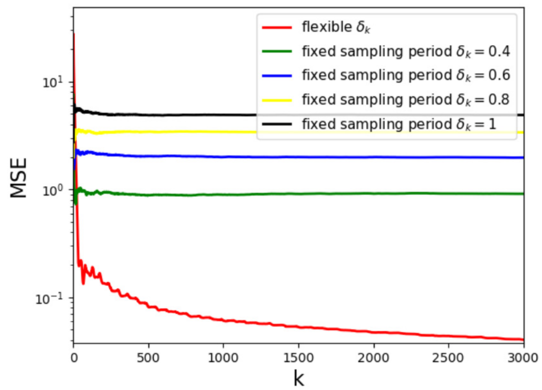

Fig. 1. MSEs of the algorithm under flexible sampling interval and constant sampling intervals.

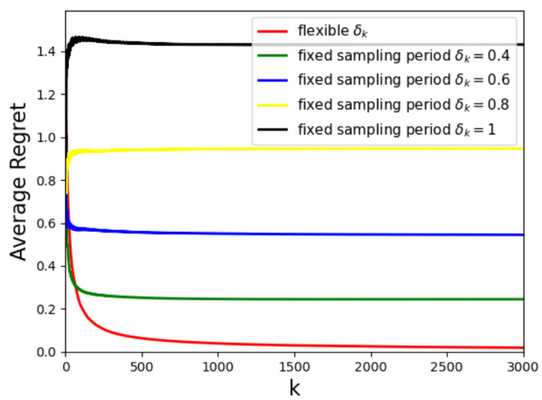

Fig. 2. Average accumulative regret of the algorithm under flexible sampling interval and constant sampling intervals.

sampling intervals) can be seen from Figure 2. We can find that the average accumulative regret with flexible sampling intervals performs better than cases of fixed constant sampling intervals.

V. Conclusion

For a continuous-time stochastic regression model, we proposed the LS algorithm based on the sampling data to identify the unknown parameter vector. By properly choosing the sampling time interval and introducing the non-persistent excitation condition, we established the almost sure convergence results of the proposed algorithm. We also provided the accumulative regret analysis without any excitation condition. Some interesting problems deserve to be further investigated, e.g., the design of the distributed algorithm to estimate the unknown parameter using local measurement, the adaptive control by using the estimation algorithm based on the sampling data.

References

[1] S. Haykin, "Radar signal processing," *IEEE ASSP Mag.*, vol. 2, no. 2, pp. 2–18, Apr. 1985.
[2] M. Vaezi and A. Izadian, "Piecewise affine system identification of a hydraulic wind power transfer system," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 6, pp. 2077–2086, Nov. 2015.
[3] J. B. Moore, "On strong consistency of least squares identification algorithms," *Automatica*, vol. 14, no. 5, pp. 505–509, 1978.
[4] T. L. Lai and C. Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *Ann. Stat.*, vol. 10, no. 1, pp. 154–166, 1982.
[5] H. Chen and L. Guo, "Strong consistency of recursive identification by no use of persistent excitation condition," *Acta Mathematicae Applicatae Sinica*, vol. 2, no. 2, pp. 133–145, 1985.
[6] L. Guo, L. Ljung, and P. Priouret, "Performance analysis of the forgetting factor RLS algorithm," *Int. J. Adapt. Control Signal Process.*, vol. 7, no. 6, pp. 525–537, 1993.
[7] L. Guo, "Stability of recursive stochastic tracking algorithms," *SIAM J. Control Optim.*, vol. 32, no. 5, pp. 1195–1225, 1994.
[8] D. Gan and Z. Liu, "On the stability of Kalman filter with random coefficients," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 2397–2402, 2020.
[9] L. Guo, *Time-Varying Stochastic Systems, Stability and Adaptive Theory*, 2nd ed. Beijing, China: Sci. Press, 2020.
[10] L. Guo and H. Chen, *Identification and Stochastic Adaptive Control*. Cambridge, MA, USA: Birkhaüser, 1991.
[11] F. S. Richards, "A method of maximum-likelihood estimation," *J. Roy. Stat. Soc. B, Methodol.*, vol. 23, no. 2, pp. 469–475, 1961.
[12] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.
[13] H. K. Khalil, *Nonlinear Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
[14] P. A. Gnoffo, *Conservation Equations and Physical Models for Hypersonic Air Flows in Thermal and Chemical Nonequilibrium*. Washington, DC, USA: Nat. Aeronaut. Space Admin., 1989.
[15] D. G. DeWolf and D. M. Wiberg, "An ordinary differential equation technique for continuous-time parameter estimation," *IEEE Trans. Autom. Control*, vol. 38, no. 4, pp. 514–528, Apr. 1993.
[16] H. Chen and L. Guo, "Continuous-time stochastic adaptive tracking—Robustness and asymptotic properties," *SIAM J. Control Optim.*, vol. 28, no. 3, pp. 513–527, 1990.
[17] A. Poyton, M. S. Varziri, K. B. McAuley, P. J. McLellan, and J. O. Ramsay, "Parameter estimation in continuous-time dynamic models using principal differential analysis," *Comput. Chem. Eng.*, vol. 30, no. 4, pp. 698–708, 2006.
[18] J. Wang, Y. Wei, T. Liu, A. Li, and Y. Wang, "Fully parametric identification for continuous time fractional order Hammerstein systems," *J. Franklin Inst.*, vol. 357, no. 1, pp. 651–666, 2020.
[19] C. Yuan and C. Wang, "Design and performance analysis of deterministic learning of sampled-data nonlinear systems," *Sci. China Inf. Sci.*, vol. 57, no. 3, pp. 1–18, 2014.
[20] W. Greblicki, "Continuous-time Hammerstein system identification from sampled data," *IEEE Trans. Autom. Control*, vol. 51, no. 7, pp. 1195–1200, Jul. 2006.
[21] R. Ortega, A. Bobtsov, and N. Nikolaev, "Parameter identification with finite-convergence time alertness preservation," *IEEE Control Syst. Lett.*, vol. 6, pp. 205–210, 2021.
[22] X.-L. Hu and J. S. Welsh, "Continuous-time model identification from filtered sampled data: Error analysis," *IEEE Trans. Autom. Control*, vol. 65, no. 10, pp. 4005–4015, Oct. 2020.
[23] R. Johansson, "Identification of continuous-time models," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 887–897, Apr. 1994.
[24] T. Soderstrom, H. Fan, B. Carlsson, and S. Bigi, "Least squares parameter estimation of continuous-time ARX models from discrete-time data," *IEEE Trans. Autom. Control*, vol. 42, no. 5, pp. 659–673, May 1997.
[25] T. L. Lai, "Asymptotically efficient adaptive control in stochastic regression models," *Adv. Appl. Math.*, vol. 7, no. 1, pp. 23–45, 1986.