J Syst Sci Complex (2024) 37(4): 1488–1506

Compressed Least Squares Algorithm of Continuous-Time Linear Stochastic Regression Model Using Sampling Data*

XIE Siyu · ZHANG Shujun · WANG Ziming · GAN Die

DOI: 10.1007/s11424-024-3297-7
Received: 1 August 2023 / Revised: 13 October 2023

Abstract In this paper, the authors consider a sparse parameter estimation problem in continuous-time linear stochastic regression models using sampling data. Based on the compressed sensing (CS) method, the authors propose a compressed least squares (LS) algorithm to deal with the challenges of parameter sparsity. At each sampling time instant, the proposed compressed LS algorithm first compresses the original high-dimensional regressor using a sensing matrix and obtains a low-dimensional LS estimate for the compressed unknown parameter. Then, the original high-dimensional sparse unknown parameter is recovered by a reconstruction method. By introducing a compressed excitation assumption and employing stochastic Lyapunov function and martingale estimate methods, the authors establish the performance analysis of the compressed LS algorithm under the condition on the sampling time interval without using independence or stationarity conditions on the system signals. At last, a simulation example is provided to verify the theoretical results by comparing the standard and the compressed LS algorithms for estimating a high-dimensional sparse unknown parameter.

Keywords Compressed excitation condition, compressed sensing, continuous-time model, least squares, linear stochastic regression, parameter identification, sampling data.

1 Introduction

Parameter estimation problem has attracted much attention in many research areas, e.g., identification, signal processing, adaptive control, statistical learning, and so on, see [1–3] for

XIE Siyu · ZHANG Shujun
School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731*, China.* Email: syxie@uestc.edu.cn; zhangsj@uestc.edu.cn.
WANG Ziming
China Academy of Aerospace Science and Innovation, Beijing 100176*, China.* Email: wangziming@lsec.cc.ac.cn.
GAN Die (Corresponding author)
Zhongguancun Laboratory, Beijing 100094*, China.* Email: gandie@amss.ac.cn.
*This research was supported by the Major Key Project of Peng Cheng Laboratory under Grant No. PCL 2023AS1-2, and Project funded by China Postdoctoral Science Foundation under Grant Nos. 2022M722926 and 2023T160605.
◇*This paper was recommended for publication by Editor GUO Jin.*
†These authors contributed equally: XIE Siyu and ZHANG Shujun.

some design and theoretical analyses on estimation and filtering algorithms. Over the past few decades, considerable progress has been made in estimating unknown parameters for discrete-time linear regression models. For example, the stochastic gradient (SG) algorithm, LS algorithm, maximum likelihood algorithm, least mean squares (LMS) algorithm, and so on[4–8]. Note that the LS, which can be traced back to Gauss, is a most basic, widely used and comprehensively studied estimation algorithm in many fields of science and engineering. Moreover, when the unknown parameter is time-invariant, the LS algorithm may generate more accurate estimates in the transient phase and have faster convergence speed compared with SG and LMS algorithms. So the LS appears to be more suitable for applications that require fast speed and accurate estimates for unknown constant parameters. This is one of the main motivations for us to consider the LS-based estimation algorithm in this paper.

Most of the existing techniques on system identification are oriented to discrete-time systems. However, practical dynamic systems in physics and engineering are naturally described by differential equations according to physical laws[9]. The continuous-time case tends to be addressed by modifying these discrete-time techniques. For example, [10] considered adaptive estimation and control problems for continuous-time stochastic systems containing both modeled and unmodeled dynamics, [11] presented some options to deal with the problem of parameter estimation in continuous-time stochastic systems under white and colored noise perturbations using classical methods, [12] considered time-varying matrix estimation problems in continuous-time stochastic models under a colored noise based on the LMS with forgetting factor, [13] compared the instrumental variables and the least squares methods applied to parameter estimation in continuous-time systems and so on. See [14, 15] for more information about other estimation algorithms for continuous-time systems.

Moreover, 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[16]. The authors in [17] analyzed LS estimation of continuous-time autoregressive model with exogenous inputs (ARX) models using discrete-time approximations of the derivatives, [18] identified a continuous-time Hammerstein system driven by a random signal from observations sampled in time, [19] established an upper bound for the estimation error of a standard LS algorithm used to identify a continuous-time model from filtered and sampling input-output data, and [20] extended the deterministic learning theory to sampling data nonlinear systems. Also, we proposed the LS algorithm based on the sampling data to identify the unknown parameter vector and established the almost sure convergence results of the proposed algorithm[21].

Although estimation algorithms are able to estimate unknown parameters in continuous-time stochastic systems based on sampling data, some prior knowledge about the system can be helpful for performance improvement. A common prior information is sparsity, i.e., the unknown high-dimensional signals can be sparse on some basis, which means that only a few elements of the signals are non-zero in the domain. For example, some practical situations like radar systems, multi propagation, channel estimation in Ultra Wideband communication systems, and so on. There are many algorithms that have been developed in the literature for sparse signal estimation problems in discrete-time[22, 23]. One technique for sparse system identification is to regularize the error function by adding another item which takes into account the sparsity of the unknown parameter. For example, [24] proposed an LMS algorithm with ℓ_0-norm constraint in order to accelerate the sparse system identification, and [25] presented a theoretical performance analysis of ℓ_0-LMS for white Gaussian input data. Also, [26] used an ℓ_1-norm penalty in the standard LMS cost function and performed convergence analysis of the zero-attracting LMS algorithm based on white input signals, and [27] presented the ℓ_1-norm regularized versions of the recursive least squares algorithm. Moreover, there are some other alternatives to induce sparse solutions, for example, ℓ_r $(0 < r < 1)$-norm[28].

Another technique is inspired by the CS theory[29], which is a new type of sampling theory that appeared in the beginning of the 21st century and makes the additional assumption that the signal is sparse or can be sparse on some orthonormal basis compared with Nyqvist sampling theory. Reference [30] applied CS in sensor networks and used the CS technique in the transit layer, [31] identified the sparse system in the compressed domain and applied the CS method in estimation layer, [32] proposed a novel diffusion compressed estimation scheme for sparse signals in compressed state, [33] proposed a tail iteratively reweighted least squares algorithm to solve the ℓ_q $(1 \leq q \leq 2)$ minimization problem for the CS sparse signal recovery issue, and so on. Also, we have incorporated the CS technique into the LMS and LS algorithms with theoretical analyses on the stability and error bounds[34, 35]. Note that most related work focus on discrete-time models, and a rigorous theoretical analysis on continuous-time models and sampling data for sparse signal estimation problems is still lacking.

Inspired by the CS theory, this paper proposes a compressed LS algorithm for estimating unknown sparse high-dimensional parameter of the continuous-time linear stochastic regression models using the sampling data. At each sampling time instant, we first use the LS algorithm to construct a low-dimensional estimate for the compressed unknown signal by using the compressed regressor data. Then we use a suitable signal reconstruction algorithm to obtain a high-dimensional sparse estimate for the original unknown parameter. The main contributions of this paper are summarized as follows:

1) Based on the CS method, we propose a compressed LS algorithm to deal with the challenges of parameter sparsity for the continuous-time linear stochastic regression models using the sampling data. In our theoretical analysis, we first provide an upper bound for the compressed estimation error of the compressed algorithm under some conditions on the sampling time interval and the stochastic compressed regression signals, and then give an upper bound for the desired high-dimensional estimation error.

2) The theoretical analysis of the compressed LS algorithm is provided under a compressed excitation condition, which does not require the independence or stationarity of the system signal as commonly used in existing work[36], and is much weaker than the excitation conditions imposed on the original regression vectors.

3) We show in the simulation part that even in the case where the traditional uncompressed algorithm cannot estimate the unknown high-dimensional sparse signal due to lack of sufficient information, the proposed compressed LS algorithm can accomplish the estimation task under

a compressed information condition, because it is much weaker than the information condition for the uncompressed algorithms.

The rest of the paper is organized as follows. Section 2 introduces some preliminaries on matrix theory, stochastic process, compressed sensing theory, and gives the continuous-time stochastic linear regression model studied in this paper. Section 3 proposes the compressed LS algorithm based on sampling data and introduces the assumptions used to fulfill the theoretical analysis. The convergence results of the proposed compressed LS algorithm are established in Section 4, the performance evaluation case studies are presented in Section 5, and the main conclusions of this paper are summarized in Section 6, respectively.

2 Problem Formulation

2.1 Some Preliminaries

2.1.1 Matrix Theory

For an m-dimensional vector x, the p-norm of x is defined as $\|x\|_p = (\sum_{j=1}^m |x_j|^p)^{1/p}$ $(1 \leq p < \infty)$, where x_j denotes the j-th element of x. The notation $\|x\|_0$ refers to the number of nonzero elements of x. For an $m \times m$-dimensional matrix A, the Euclidean norm, denoted by $\|A\|$, is defined as $\|A\| \stackrel{\Delta}{=} (\lambda_{\max}\{AA^{\mathrm{T}}\})^{\frac{1}{2}}$, where $\lambda_{\max}(\cdot)$ represents the largest eigenvalue of the matrix and T denotes the transpose operator. The smallest eigenvalue of the matrix is denoted as $\lambda_{\min}(\cdot)$, and the determinant of the matrix is denoted by $\det(\cdot)$. Consider 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 $\lim_{k\to\infty} \|A_k\|/a_k = 0$, then we say $A_k = o(a_k)$. For matrices A, B, C and D with suitable dimensions, the matrix inverse formula (cf., [3]) is given by:

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

where all relevant matrices are assumed to be invertible.

2.1.2 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 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 2.1 (see [3]) *Suppose that $\{m_k, \mathcal{F}_k\}$ is an adapted process (m_k can be a number* *or a matrix), and* $\{\nu_k, \mathcal{F}_k\}$ *is a martingale difference sequence* (ν_k *can be a number or a matrix) satisfying* $\sup_n \mathbb{E}[\|\nu_{n+1}\|^\beta|\mathcal{F}_n] < \infty$ *with* $\beta \in (0, 2]$*. Then for any* $\eta > 0$*, we have*

$$\sum_{k=0}^{n} m_k \nu_{k+1} = O\Big(M_n(\beta) \log^{\frac{1}{\beta}+\eta}(M_n(\beta) + e)\Big) \quad \text{a.s.},$$

where

$$M_n(\beta) = \bigg(\sum_{k=0}^{n} \|m_k\|^\beta\bigg)^{\frac{1}{\beta}}.$$

2.1.3 Compressed Sensing Theory

The CS theory has emerged as a new framework for sampling theory, offering several attractive properties such as robustness to noise, fault tolerance, and bandwidth saving. In CS theory, the vector x is called p-sparse if $\|x\|_0 \le p$, that is, x has at most p nonzero elements. Assume that the sparse signal x obeys the following equation:

$$z = Dx + \varepsilon, \tag{2}$$

where z is the measurement, $D \in \mathbb{R}^{d\times m}$ is the sensing matrix whose number of rows is much smaller than the number of columns, and $\varepsilon \in \mathbb{R}^d$ is the measurement perturbation bounded by $\|\varepsilon\| \le C$. We are interested in how to recover the signal x from the noisy measurement z. Recovering a general signal x accurately is typically challenging or even impossible. However, when the signal x is sparse, the recovery problem becomes feasible. Candès, et al. introduced the concept of the restricted isometry property (RIP) of sensing matrix D in CS theory to study the reconstruction problem of sparse signals, and they proved that the sparse signal x can be recovered with high accuracy if D satisfies the RIP and the noise is small (see e.g., [29, 37]).

Definition 2.2 (see [29]) Let $D \in \mathbb{R}^{d\times m}$ be the sensing matrix, and D_L ($L \subseteq \{1, \cdots, m\}$) be the sub-matrix obtained by extracting the columns of D corresponding to the indices in the set L. For given integer p ($1 \le p \le m$), we define the p-restricted isometry constant $\gamma_p \in [0, 1)$ to be the smallest quantity such that the following inequality

$$(1 - \gamma_p)\|b\|^2 \le \|D_L b\|^2 \le (1 + \gamma_p)\|b\|^2 \tag{3}$$

holds for all real vector b and all subsets L with cardinality at most p. Then we say that D satisfies the RIP with order p.

Remark 2.3 From (3), the restricted isometry constant γ_p reflects the degree of preservation of the signal's 2-norm, with $\gamma_p = 0$ being exactly preserved. The condition (3) is equivalent to that all eigenvalues of the matrix $D_L^{\mathrm{T}} D_L$ lie in $[1-\gamma_p, 1+\gamma_p]$. Moreover, from Definition 2.2, we can see that the p-restricted isometry constant γ_p increases with p.

The construction of a sensing matrix D satisfying the RIP has attracted significant attention in the fields of information theory and signal processing. Researchers have proposed various methods for constructing such matrices. For example, DeVore in [38] introduced a deterministic construction method based on the concept of mutual incoherence, and Xu and Xu in [39] constructed a class of special sensing matrices called partial Fourier matrices. In addition to deterministic constructions, there is also a body of literature focusing on random sensing matrices. Random matrices can be generated using various probability distributions, and under certain conditions, they are known to satisfy the RIP with high probability. For example, Gaussian random matrices are commonly used as random sensing matrices. In this case, each entry of $D \in \mathbb{R}^{d\times m}$ is an independent realization of a Gaussian random variable with zero mean and variance $1/d$. Gaussian matrices exhibit good RIP properties and have been widely studied in theoretical analysis and practical applications. We have the following result about random sensing matrices.

Lemma 2.4 (see [40]) *For given d, m, and $0 < \delta < 1$, if the sensing matrix $D \in \mathbb{R}^{d\times m}$ is a Gaussian or Bernoulli random matrix, then there exist positive constants c_1, c_2 depending only on δ such that the RIP* (3) *holds for D with the prescribed δ and any $p \le c_1 d/\log(m/p)$ with probability no less than $1 - \mathrm{e}^{-c_2 d}$.*

Remark 2.5 In [40], Baraniuk, et al. showed that the constants c_1 and c_2 satisfy the inequality $c_2 \le \frac{\delta^2}{16} - \frac{\delta^3}{48} - c_1\left[1 + \frac{1+\log(12/\delta)}{\log(m/p)}\right]$ and the constant c_1 can be small enough to ensure $c_2 > 0$. For example, if $c_1 = \frac{\delta^3}{120}$, then we can obtain that the sensing D satisfies the RIP with probability no less than $1 - \mathrm{e}^{-c_2 d}$ if $d \ge 120 s \log(m/p)/\delta^3$.

After constructing the sensing matrix, the recovery problem of the sparse signal x can be formulated as the following convex optimization problem:

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

For the signals obtained by solving the above convex optimization problem, Candès and Tao in [37] established the following lemma on the upper bound of the error between the recovered signals and original signals, which provides a theoretical guarantee for the accuracy of the recovery of sparse signals using the convex optimization approach. It has been widely used in the analysis and design of compressed sensing systems and has significantly contributed to the development of signal processing and related fields.

Lemma 2.6 (see [37]) *Let p satisfy $\gamma_{3p} + 3\gamma_{4p} < 2$ where γ_{3p} and γ_{4p} are defined in Definition* 2.2. *Then for any p-sparse signal x and any perturbation ε with $\|\varepsilon\| \le C$, the recovered signal x^* obtained by solving the optimization problem* (4) *obeys $\|x^* - x\| \le C_p C$, where the positive constant C_p can be taken as $C_p \triangleq \frac{4}{\sqrt{3(1-\gamma_{4p})} - \sqrt{1+\gamma_{3p}}}$.*

Remark 2.7 Note that for p satisfying the condition of Lemma 2.6, we can get the true value of the sparse signal by the reconstruction process if there are no measurement perturbations in (2).

2.2 Continuous-Time Stochastic Regression Model

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

$$o_t = S\theta^{\mathrm{T}}\varphi_t + n_t, \quad t \ge 0, \tag{5}$$

where S is the integral operator, i.e.,

$$S\varphi_t = \int_0^t \varphi_s ds,$$

$o_t \in \mathbb{R}$ is the scalar observation at time t, $\theta \in \mathbb{R}^m$ is an unknown p-sparse parameter vector that needs to be estimated (i.e., θ has at most p non-zero elements), $\varphi_s \in \mathbb{R}^m$ is a stochastic regression vector, and the system noise n_t is a standard Wiener process. We can see that many models such as the continuous-time ARX model and autoregressive moving average system with exogenous inputs (ARMAX) model can be included by (5).

With the application of computer and communication technology in many practical scenarios, we can only receive the discrete-time sampling data $\{o_{t_k}, \varphi_{t_k}\}_{k=0}^{\infty}$ rather than continuous-time signals, where t_k is the k-th sampling time instant. Besides, the sampling stochastic regression vector is always sparse. Here we assume that the sampling stochastic regression vector φ_{t_k} is $3p$-sparse for all k. In this paper, we aim at designing an algorithm to estimate the unknown sparse parameter vector θ by using the sampling data based on compressed sensing theory, and providing the asymptotic results for the proposed algorithm.

3 Compressed LS Algorithm Based on Sampling Data

Now, we propose the compressed LS algorithm based on sampling data (see Algorithm 1), where $\delta_k = t_{k+1} - t_k$ and t_k represents the k-th sampling time instant.

Algorithm 1 Compressed LS algorithm based on sampling data

Input: $\{\varphi_{t_k} \in \mathbb{R}^m, o_{t_{k+1}} \in \mathbb{R}\}$, $k = 0, 1, \cdots$

Output: $\{\theta_{t_{k+1}} \in \mathbb{R}^m\}$, $k = 0, 1, \cdots$

Initialization: Begin with any initial vector $\xi_0 \in \mathbb{R}^d$ and any initial positive definite matrix $P_0 \in \mathbb{R}^{d\times d}$.

for each $k = 0, 1, \cdots$ **do**

Step 1 Compression: $\phi_{t_k} = D\varphi_{t_k} \in \mathbb{R}^d$, where $D \in \mathbb{R}^{d\times m}$ is the sensing matrix.

Step 2 Estimation in a low-dimensional space:

$$\xi_{t_{k+1}} = \xi_{t_k} + \delta_k a_{t_k} P_{t_k} \phi_{t_k} (o_{t_{k+1}} - o_{t_k} - \phi_{t_k}^{\mathrm{T}} \xi_{t_k} \delta_k), \tag{6}$$

$$P_{t_{k+1}} = P_{t_k} - \delta_k^2 a_{t_k} P_{t_k} \phi_{t_k} \phi_{t_k}^{\mathrm{T}} P_{t_k}, \tag{7}$$

$$a_{t_k} = \frac{1}{1 + \delta_k^2 \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k}}, \tag{8}$$

$$\delta_k = t_{k+1} - t_k. \tag{9}$$

Step 3 Reconstruction:

$$\theta_{t_{k+1}} = \arg\min_{\theta\in\Xi} \|\theta\|_1, \tag{10}$$

$$\text{where} \quad \Xi = \left\{\theta \in \mathbb{R}^m \,\middle|\, \|D\theta - \xi_{t_{k+1}}\| \leq C\right\}. \tag{11}$$

Remark 3.1 In Step 3 of Algorithm 1, the positive constant C can indeed be chosen as an upper bound on the estimation error $\|\widetilde{\xi}_{t_k}\|$, and the explicit value of C can be found in Theorem 4.4 of Section 4. To solve the convex optimization problem (10) in the reconstruction process (i.e., Step 3), various algorithms including orthogonal matching pursuit (OMP), compressive sampling matching pursuit (CoSaMP), and interior-point (IP) algorithms can be employed, see e.g., [41, 42].

To proceed with the theoretical analysis of the proposed algorithm, we introduce some assumptions on the sensing matrix, the sampling interval, and regression vectors.

Assumption 3.2 The sensing matrix $D \in \mathbb{R}^{d\times m}$ satisfies the RIP with order $4p$ where the $3p$- and $4p$-restricted isometry constants denoted as γ_{3p} and γ_{4p} (see Definition 2.2) satisfy $\gamma_{3p} + 3\gamma_{4p} < 2$.

Remark 3.3 The above properties (RIP) of the sensing matrix D which is often used in the compressive sensing theory can guarantee that the sparse signals can be recovered with high accuracy (see Lemma 2.6). By Lemma 2.4, we know that for any given restricted isometry constant, if the sensing matrix D is taken as a Gaussian or Bernoulli random matrix, then it satisfies RIP condition with a high probability.

Assumption 3.4 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 3.5 The stochastic regression signal ϕ_t in the model (5) 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 3.6 (Compressed persistent excitation condition) There exists a positive constant M such that

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

where

$$r_n \triangleq 1 + \sum_{k=0}^{n} \delta_k^2 \|\phi_{t_k}\|^2, \quad \lambda_{\min}^{n} \triangleq \lambda_{\min}\Big(P_0^{-1} + \sum_{k=0}^{n} \delta_k^2 \phi_{t_k} \phi_{t_k}^{\mathrm{T}}\Big).$$

Remark 3.7 Assumption 3.5 is often used to deal with the effect of the approximation deviation caused by the sampling data, see Theorem 4.2. Assumption 3.6 is called compressed persistent excitation condition since it is imposed on the compressed regression vectors ϕ_{t_k}, rather than the original high-dimensional sparse regression vector φ_{t_k}, which is different from the existing literature, see e.g., [21, 22]. Moreover, it can be verified that Assumption 3.6 is much weaker than the traditional persistent excitation condition imposed on the original high dimensional regression vector[4], i.e., $\sup_n \frac{\lambda_{\max}\left(\sum_{k=0}^{n}\varphi_{t_k}\varphi_{t_k}^{\mathrm{T}}\right)}{\lambda_{\min}\left(\sum_{k=0}^{n}\varphi_{t_k}\varphi_{t_k}^{\mathrm{T}}\right)} < \infty$.

4 Performance Results of Algorithm 1

In this section, we will investigate the asymptotic convergence properties of the proposed Algorithm 1. Denote $\widetilde{\xi}_{t_k}$ as the compressed estimate error at the time instant t_k in the low-dimensional space, i.e., $\widetilde{\xi}_{t_k} = \xi - \xi_{t_k}$, where $\xi = D\theta$. Substituting (5) into (6), we can obtain the following error equation:

$$\begin{aligned}\widetilde{\xi}_{t_{k+1}} =& \xi - \xi_{t_{k+1}}\\ =&\xi - \xi_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\left\{\int_{t_k}^{t_{k+1}}\varphi_s^{\mathrm{T}} ds\theta - \phi_{t_k}^{\mathrm{T}}\xi_{t_k}\delta_k + n_{t_{k+1}} - n_{t_k}\right\}\\ =&\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\left\{\int_{t_k}^{t_{k+1}}\varphi_s^{\mathrm{T}} ds\theta - \varphi_{t_k}^{\mathrm{T}}\theta\delta_k + \varphi_{t_k}^{\mathrm{T}}\theta\delta_k - \phi_{t_k}^{\mathrm{T}}\xi\delta_k\right.\\ &\left. + \phi_{t_k}^{\mathrm{T}}\xi\delta_k - \phi_{t_k}^{\mathrm{T}}\xi_{t_k}\delta_k + n_{t_{k+1}} - n_{t_k}\right\}\\ =&\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\left\{\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\delta_k + \mathit{\Delta}_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right\},\end{aligned} \tag{12}$$

where

$$\mathit{\Delta}_{t_{k+1}} = \int_{t_k}^{t_{k+1}}\phi_s ds - \phi_{t_k}\delta_k, \tag{13}$$

$$w_{t_k} = \varphi_{t_k}^{\mathrm{T}}\theta\delta_k - \phi_{t_k}^{\mathrm{T}}\xi\delta_k, \tag{14}$$

$$\overline{n}_{t_{k+1}} = n_{t_{k+1}} - n_{t_k}. \tag{15}$$

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

1) For $k \geq 0$, $\overline{n}_{t_k}$ obeys a normal distribution with mean 0 and variance δ_k;

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

In many existing results on the estimation of the unknown parameter vector of continuous-time systems based on the sampling data (cf., [17, 20]), continuous-time systems are often discretized, and the approximation error $\mathit{\Delta}_{t_k}$ caused by discretization is seldom considered. Different from these studies, we will consider the effect of the approximation error $\mathit{\Delta}_{t_k}$ and the sensing error w_{t_k} on the estimation performance of algorithm, see the following theorem.

Theorem 4.1 *Under Assumption* 3.2*, for the dynamical system* (5)*, the compressed estimation error* $\widetilde{\xi}_{t_{n+1}}$ *satisfies the following relationship:*

$$\begin{aligned}&\widetilde{\xi}_{t_{n+1}}^{\mathrm{T}} P_{t_{n+1}}^{-1}\widetilde{\xi}_{t_{n+1}} + \left(\frac{1}{2} + o(1)\right)\sum_{k=0}^{n} a_{t_k}\delta_k^2\left(\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\right)^2\\ \leq& 7\sum_{k=0}^{n}\left(\mathit{\Delta}_{t_{k+1}}^{\mathrm{T}}\theta\right)^2 + \frac{63\gamma_{4p}^2\|\theta\|^2}{(1-\gamma_{4p})}\sum_{k=0}^{n}\left(\|\phi_{t_k}\|^2\delta_k^2\right) + O\left(\log\det(P_{t_{n+1}}^{-1})\right) \quad \text{a.s.}\end{aligned}$$

Proof By (12), we have

$$\widetilde{\xi}_{t_{k+1}} = \left(I - a_{t_k}\delta_k^2 P_{t_k}\phi_{t_k}\phi_{t_k}^{\mathrm{T}}\right)\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right). \tag{16}$$

Multiply both sides of (7) by $P_{t_k}^{-1}$, we have

$$P_{t_{k+1}}P_{t_k}^{-1} = I - a_{t_k}\delta_k^2 P_{t_k}\phi_{t_k}\phi_{t_k}^{\mathrm{T}}. \tag{17}$$

Substituting (17) into (16), we can obtain that

$$\widetilde{\xi}_{t_{k+1}} = P_{t_{k+1}}P_{t_k}^{-1}\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}),$$

which is equivalent to the following equation

$$P_{t_{k+1}}^{-1}\widetilde{\xi}_{t_{k+1}} = P_{t_k}^{-1}\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_{k+1}}^{-1}P_{t_k}\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right). \tag{18}$$

Now, we consider the Lyapunov candidate function $W_{t_k} = \widetilde{\xi}_{t_k}^{\mathrm{T}}P_{t_k}^{-1}\widetilde{\xi}_{t_k}$. From (12) and (18), we have

$$\begin{aligned} W_{t_{k+1}} =& \widetilde{\xi}_{t_{k+1}}^{\mathrm{T}}P_{t_{k+1}}^{-1}\widetilde{\xi}_{t_{k+1}} \\ =& \left[\widetilde{\xi}_{t_k} - a_{t_k}\delta_k^2 P_{t_k}\phi_{t_k}\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right)\right]^{\mathrm{T}} \\ & \cdot\left[P_{t_k}^{-1}\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_{k+1}}^{-1}P_{t_k}\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right)\right]. \end{aligned}$$

By matrix inverse formulas (1) and (7), we have $P_{t_{k+1}}^{-1} = P_{t_k}^{-1} + \delta_k^2\phi_{t_k}\phi_{t_k}^{\mathrm{T}}$. By this equation and the definition of a_{t_k} in (8), hence we obtain $P_{t_{k+1}}^{-1}P_{t_k} = I + \delta_k^2\phi_{t_k}\phi_{t_k}^{\mathrm{T}}P_{t_k}$ and $1 - a_{t_k} = a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}$. Then we have

$$\begin{aligned} W_{t_{k+1}} =& \widetilde{\xi}_{t_{k+1}}^{\mathrm{T}}P_{t_{k+1}}^{-1}\widetilde{\xi}_{t_{k+1}} \\ =& \left[\widetilde{\xi}_{t_k} - a_{t_k}\delta_k^2 P_{t_k}\phi_{t_k}\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k} - a_{t_k}\delta_k P_{t_k}\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right)\right]^{\mathrm{T}} \\ & \cdot\left[P_{t_k}^{-1}\widetilde{\xi}_{t_k} - \delta_k\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right)\right]. \end{aligned}$$

Hence we can get the following equation

$$\begin{aligned} W_{t_{k+1}} =& \widetilde{\xi}_{t_k}^{\mathrm{T}}P_{t_k}^{-1}\widetilde{\xi}_{t_k} - a_{t_k}\delta_k^2\left(\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\right)^2 - 2a_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right) \\ & + a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta + w_{t_k} + \overline{n}_{t_{k+1}}\right)^2 \\ \le& W_{t_k} - a_{t_k}\delta_k^2\left(\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\right)^2 - 2a_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\varDelta_{t_{k+1}}^{\mathrm{T}}\theta - 2a_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}w_{t_k} \\ & - 2a_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\overline{n}_{t_{k+1}} + 3a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta\right)^2 + 3a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}w_{t_k}^2 \\ & + 3a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}\overline{n}_{t_{k+1}}^2. \end{aligned} \tag{19}$$

By summing both sides of (19) from $k = 0$ to n, we have

$$W_{t_{n+1}} - W_0 + \sum_{k=0}^{n} a_{t_k}\delta_k^2(\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k})^2$$

$$\leq \underbrace{-2\sum_{k=0}^{n} a_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\varDelta_{t_{k+1}}^{\mathrm{T}}\theta}_{I_1} \underbrace{-2\sum_{k=0}^{n} a_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}w_{t_k}}_{I_2} \underbrace{-2\sum_{k=0}^{n} a_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\overline{n}_{t_{k+1}}}_{I_3} \tag{20}$$
$$+\underbrace{3\sum_{k=0}^{n} a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta)^2}_{I_4} + \underbrace{3\sum_{k=0}^{n} a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}w_{t_k}^2}_{I_5} + \underbrace{3\sum_{k=0}^{n} a_{t_k}\delta_k^2\phi_{t_k}^{\mathrm{T}}P_{t_k}\phi_{t_k}\overline{n}_{t_{k+1}}^2}_{I6}.$$

In the following, we estimate the six terms $I_1, I_2, \cdots, I_6$ on the right hand side of (20). By the Hölder inequality and mean inequality, we have for any $h > 0$,

$$\begin{aligned}|I_1| &\leq \left\{\sum_{k=0}^{n}\left(ha_{t_k}\delta_k\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\right)^2\right\}^{\frac{1}{2}} \cdot \left\{\sum_{k=0}^{n}\left(\frac{2}{h}\varDelta_{t_{k+1}}^{\mathrm{T}}\theta\right)^2\right\}^{\frac{1}{2}} \\ &\leq \frac{h^2}{2}\sum_{k=0}^{n} a_{t_k}\delta_k^2\left(\phi_{t_k}^{\mathrm{T}}\widetilde{\xi}_{t_k}\right)^2 + \frac{2}{h^2}\sum_{k=0}^{n}\left(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta\right)^2.\end{aligned} \tag{21}$$

Denote the set $\varLambda_t = \left\{i_{t_k}^{(1)}, \cdots, i_{t_k}^{(3p)}, j^{(1)}, \cdots, j^{(p)}\right\}$, where $i_{t_k}^{(1)}, \cdots, i_{t_k}^{(3p)}$ are the indices of $3p$ nonzero elements of φ_{t_k} and $j^{(1)}, \cdots, j^{(p)}$ are the indices of p nonzero elements of θ. The analysis for the case where the cardinality of $\varLambda_{t_k}$ is less than $4p$ (i.e., part of the nonzero elements of the vectors φ_{t_k} and θ are in the same position) is almost the same as that for the case where the set $\varLambda_{t_k}$ has $4p$ elements. Thus, we just consider the latter. The vectors obtained by extracting the $4p$ nonzero elements from φ_{t_k} and θ are denoted as $\varphi_{t_k}^{(4p)}$ and $\breve{\theta}_{t_k}^{(4p)}$, and the indices of these $4p$ elements come from the set $\varLambda_{t_k}$. Correspondingly, we extract the $4p$ columns from the matrix D, and denote the new matrix as $D_{t_k}^{(4p)}$.

By Assumption 3.2 and Remark 2.3, we see that all eigenvalues of the matrix $\left(D_{t_k}^{(4p)}\right)^{\mathrm{T}} D_{t_k}^{(4p)}$ lie in the interval $[1-\gamma_{4p}, 1+\gamma_{4p}]$. Thus, we have

$$\begin{aligned}|w_{t_k}| &= \left\|\varphi_{t_k}^{\mathrm{T}}(I_m - D^{\mathrm{T}}D)\theta\right\|\delta_k \\ &= \left\|\left(\varphi_{t_k}^{(4p)}\right)^{\mathrm{T}}\left[I_{4p} - \left(D_{t_k}^{(4p)}\right)^{\mathrm{T}} D_{t_k}^{(4p)}\right]\breve{\theta}_{t_k}^{(4p)}\right\|\delta_k \\ &\leq \left\|\left(\varphi_{t_k}^{(4p)}\right)^{\mathrm{T}}\left[(1+\gamma_{4p})I_{4p} - \left(D_{t_k}^{(4p)}\right)^{\mathrm{T}} D_{t_k}^{(4p)}\right]\breve{\theta}_{t_k}^{(4p)}\right\|\delta_k + \gamma_{4p}\left\|\left(\varphi_{t_k}^{(4p)}\right)^{\mathrm{T}}\breve{\theta}_{t_k}^{(4p)}\right\|\delta_k \\ &\leq \left\|(\varphi_{t_k}^{(4p)})^{\mathrm{T}}\right\| \cdot \left\|(1+\gamma_{4p})I_{4p} - \left(D_{t_k}^{(4p)}\right)^{\mathrm{T}} D_{t_k}^{(4p)}\right\| \cdot \left\|\breve{\theta}_{t_k}^{(4p)}\right\|\delta_k + \gamma_{4p}\left\|\left(\varphi_{t_k}^{(4p)}\right)^{\mathrm{T}}\breve{\theta}_{t_k}^{(4p)}\right\|\delta_k \\ &\leq 2\gamma_{4p}\left\|\varphi_{t_k}^{(4p)}\right\| \cdot \left\|\breve{\theta}_{t_k}^{(4p)}\right\|\delta_k + \gamma_{4p}\left\|\varphi_{t_k}^{(4p)}\right\| \cdot \left\|\breve{\theta}_{t_k}^{(4p)}\right\|\delta_k \\ &= 3\gamma_{4p}\left\|\varphi_{t_k}\right\| \cdot \|\theta\|\delta_k \\ &\leq \frac{3\gamma_{4p}}{\sqrt{1-\gamma_{4p}}}\|D\varphi_{t_k}\|\|\theta\|\delta_k \\ &= \frac{3\gamma_{4p}}{\sqrt{1-\gamma_{4p}}}\|\phi_{t_k}\|\|\theta\|\delta_k.\end{aligned} \tag{22}$$

By (22), we can obtain the following inequality for any $h > 0$,

$$
\begin{aligned}
|I_2| \le & \left\{ \sum_{k=0}^{n} (h a_{t_k} \delta_k \phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k})^2 \right\}^{\frac{1}{2}} \cdot \left\{ \sum_{k=0}^{n} \left(\frac{2}{h} w_{t_k} \right)^2 \right\}^{\frac{1}{2}} \\
\le & \frac{h^2}{2} \sum_{k=0}^{n} a_{t_k} \delta_k^2 (\phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k})^2 + \frac{2}{h^2} \sum_{k=0}^{n} w_{t_k}^2 \\
\le & \frac{h^2}{2} \sum_{k=0}^{n} a_{t_k} \delta_k^2 (\phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k})^2 + \frac{18 \gamma_{4p}^2 \|\theta\|^2}{h^2 (1-\gamma_{4p})} \sum_{k=0}^{n} \left(\|\phi_{t_k}\|^2 \delta_k^2 \right).
\end{aligned} \tag{23}
$$

Note that $\delta_k a_{t_k} \phi_{t_k}^{\mathrm{T}} \widetilde{\theta}_{t_k} \in \mathcal{F}_{t_k}$, by Lemma 2.1, we can derive the following estimate for I_3,

$$
|I_3| = O(1) + o\left(\sum_{k=0}^{n} \delta_k^2 a_{t_k} (\phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k})^2 \right). \tag{24}
$$

By (8), it is clear that $\delta_k^2 a_{t_k} \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} \le 1$. Thus, we have

$$
I_4 \le 3 \sum_{k=0}^{n} \left(\varDelta_{t_{k+1}}^{\mathrm{T}} \theta \right)^2. \tag{25}
$$

Similarly, by (22), we have

$$
I_5 \le 3 \sum_{k=0}^{n} w_k^2 \le \frac{27 \gamma_{4p}^2 \|\theta\|^2}{1-\gamma_{4p}} \sum_{k=0}^{n} \left(\|\phi_{t_k}\|^2 \delta_k^2 \right). \tag{26}
$$

For I_6, it follows that

$$
I_6 = \sum_{k=0}^{n} \delta_k^2 a_{t_k} \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} \left\{ \overline{n}_{t_{k+1}}^2 - \mathbb{E}\left[\overline{n}_{t_{k+1}}^2 | \mathcal{F}_{t_k} \right] \right\} + \sum_{k=0}^{n} \delta_k^2 a_{t_k} \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} \mathbb{E}\left[\overline{n}_{t_{k+1}}^2 | \mathcal{F}_{t_k} \right]. \tag{27}
$$

Note that $\{\overline{n}_{t_{k+1}}^2 - \mathbb{E}[\overline{n}_{t_{k+1}}^2 | \mathcal{F}_{t_k}], \mathcal{F}_{t_k}\}$ is a martingale difference sequence, by the C_r-inequality and the Lyapunov inequality, we see that for any $\alpha \in (1,2]$,

$$
\begin{aligned}
& \sup_k \mathbb{E}\left\{ \left[\overline{n}_{t_{k+1}}^2 - \mathbb{E}(\overline{n}_{t_{k+1}}^2 | \mathcal{F}_{t_k}) \right]^{\alpha} | \mathcal{F}_{t_k} \right\} \\
\le & 2 \sup_k \mathbb{E}\left[|\overline{n}_{t_{k+1}}|^{2\alpha} | \mathcal{F}_{t_k} \right] + 2 \sup_k \mathbb{E}\left[\left(\mathbb{E}\left[|\overline{n}_{t_{k+1}}|^2 | \mathcal{F}_{t_k} \right] \right)^{\alpha} | \mathcal{F}_{t_k} \right] \\
\le & 4 \sup_k \mathbb{E}\left[|\overline{n}_{t_{k+1}}|^{2\alpha} | \mathcal{F}_{t_k} \right] \\
< & \infty
\end{aligned}
$$

holds almost surely. From Lemma 2.1, we can derive that for any $\eta > 0$,

$$
\sum_{k=0}^{n} \delta_k^2 a_{t_k} \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} \left\{ \overline{n}_{t_{k+1}}^2 - \mathbb{E}\left[\overline{n}_{t_{k+1}}^2 | \mathcal{F}_{t_k} \right] \right\} = O\left(M_n(\alpha) \log^{\frac{1}{\alpha}+\eta} (M_n(\alpha) + e) \right) \quad \text{a.s.},
$$

where

$$
M_n(\alpha) \triangleq \left\{ \sum_{k=0}^{n} \left(\delta_k^2 a_{t_k} \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} \right)^{\alpha} \right\}^{\frac{1}{\alpha}}.
$$

In the following, we need to analyze $M_n(\alpha)$. Since

$$P_{t_{k+1}}^{-1} = P_{t_k}^{-1} + \delta_k^2 \phi_{t_k} \phi_{t_k}^{\mathrm{T}} = P_{t_k}^{-1} \left(I + \delta_k^2 P_{t_k} \phi_{t_k} \phi_{t_k}^{\mathrm{T}} \right), \tag{28}$$

then we can derive that

$$\det(P_{t_{k+1}}^{-1}) = \det(P_{t_k}^{-1}) \left(1 + \delta_k^2 \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} \right).$$

Thus, it follows that

$$\delta_k^2 \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} = \frac{\det(P_{t_{k+1}}^{-1}) - \det(P_{t_k}^{-1})}{\det(P_{t_k}^{-1})}, \quad a_{t_k} \delta_k^2 \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} = \frac{\det(P_{t_{k+1}}^{-1}) - \det(P_{t_k}^{-1})}{\det(P_{t_{k+1}}^{-1})}.$$

Then, it follows that

$$\begin{aligned} \sum_{k=0}^{n} \delta_k^2 a_{t_k} \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} &= \sum_{k=0}^{n} \frac{\det(P_{t_{k+1}}^{-1}) - \det(P_{t_k}^{-1})}{\det(P_{t_{k+1}}^{-1})} \\ &\le \sum_{k=0}^{n} \int_{\det(P_{t_k}^{-1})}^{\det(P_{t_{k+1}}^{-1})} \frac{dx}{x} \\ &= \log\det(P_{t_{n+1}}^{-1}) + \log\det(P_0^{-1}). \end{aligned} \tag{29}$$

According to the inequality $a_{t_k} \delta_k^2 \phi_{t_k}^{\mathrm{T}} P_{t_k} \phi_{t_k} < 1$ and (29), we can deduce the following equation

$$M_n(\alpha) = O(1) + o\left(\log\det\left(P_{t_{n+1}}^{-1} \right) \right), \tag{30}$$

where $\alpha > 1$ is used. Thus, from (27), (29) and (30), it can be deduced that

$$I_6 = O(1) + o\left(\log\det\left(P_{t_{n+1}}^{-1} \right) \right) + O\left(\log\det\left(P_{t_{n+1}}^{-1} \right) \right) = O\left(\log\det\left(P_{t_{n+1}}^{-1} \right) \right). \tag{31}$$

Hence by (20), (21), (23)–(26), (31), taking $h = \frac{1}{\sqrt{2}}$, we have the following inequality

$$\begin{aligned} & W_{t_{n+1}} - W_0 + \sum_{k=0}^{n} a_{t_k} \delta_k^2 (\phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k})^2 \\ \le & \frac{1}{2} \sum_{k=0}^{n} a_{t_k} \delta_k^2 (\phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k})^2 + 7 \sum_{k=0}^{n} (\varDelta_{t_{k+1}}^{\mathrm{T}} \theta)^2 + \frac{63\gamma_{4p}^2 \|\theta\|^2}{1-\gamma_{4p}} \sum_{k=0}^{n} (\|\phi_{t_k}\|^2 \delta_k^2) + o\bigg(\sum_{k=0}^{n} \delta_k^2 a_{t_k} (\phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k})^2 \bigg) \\ & + O\left(\log\det\left(P_{t_{n+1}}^{-1} \right) \right), \end{aligned}$$

which implies that

$$\begin{aligned} & \widetilde{\xi}_{t_{n+1}}^{\mathrm{T}} P_{t_{n+1}}^{-1} \widetilde{\xi}_{t_{n+1}} + \left(\frac{1}{2} + o(1) \right) \sum_{k=0}^{n} a_{t_k} \delta_k^2 \left(\phi_{t_k}^{\mathrm{T}} \widetilde{\xi}_{t_k} \right)^2 \\ \le & 7 \sum_{k=0}^{n} \left(\varDelta_{t_{k+1}}^{\mathrm{T}} \theta \right)^2 + \frac{63\gamma_{4p}^2 \|\theta\|^2}{1-\gamma_{4p}} \sum_{k=0}^{n} \left(\|\phi_{t_k}\|^2 \delta_k^2 \right) + O\left(\log\det\left(P_{t_{n+1}}^{-1} \right) \right). \end{aligned}$$

This completes the proof of Theorem 4.1. ∎

Therefore, an upper bound for the compressed estimation error of the proposed compressed LS algorithm is established as follows.

Theorem 4.2 *Under Assumptions* 3.2 *and* 3.5*, the compressed estimation error* $\widetilde{\xi}_{t_{n+1}}$ *has the following upper bound:*

$$\|\widetilde{\xi}_{t_{n+1}}\|^2 \leq \frac{7L^2\|\theta\|^2}{4}\frac{\sum_{k=0}^{n}\delta_k^4}{\lambda_{\min}^n} + \frac{63\gamma_{4p}^2\|\theta\|^2}{1-\gamma_{4p}} \cdot \frac{r_n}{\lambda_{\min}^n} + O\left(\frac{\log r_n}{\lambda_{\min}^n}\right). \tag{32}$$

Proof By Theorem 4.1, we have

$$\widetilde{\xi}_{t_{n+1}}^{\mathrm{T}} P_{t_{n+1}}^{-1} \widetilde{\xi}_{t_{n+1}} \leq 7\sum_{k=0}^{n}(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta)^2 + \frac{63\gamma_{4p}^2\|\theta\|^2}{1-\gamma_{4p}}\sum_{k=0}^{n}\left(\|\phi_{t_k}\|^2\delta_k^2\right) + O\left(\log\det\left(P_{t_{n+1}}^{-1}\right)\right),$$

Then we have the following equation

$$\|\widetilde{\xi}_{t_{n+1}}\|^2 \leq \frac{7\sum_{k=0}^{n}(\varDelta_{t_{k+1}}^{\mathrm{T}}\theta)^2}{\lambda_{\min}^n} + \frac{63\gamma_{4p}^2\|\theta\|^2}{1-\gamma_{4p}} \cdot \frac{r_n}{\lambda_{\min}^n} + O\left(\frac{\log\det(P_{t_{n+1}}^{-1})}{\lambda_{\min}^n}\right) \quad \text{a.s.}, \tag{33}$$

where $\lambda_{\min}^n$ and r_n are defined in Assumption 3.6. Using (13) and Assumption 3.5, we obtain the following inequality

$$\|\varDelta_{t_{k+1}}\| = \left\|\int_{t_k}^{t_{k+1}}(\phi_s - \phi_{t_k})ds\right\| \leq \int_{t_k}^{t_{k+1}} L(s-t_k)ds = \frac{L}{2}\delta_k^2. \tag{34}$$

Note that

$$\log\det(P_{t_{n+1}}^{-1}) = \sum_{i=0}^{d}\log\lambda_i(P_{t_{n+1}}^{-1}) \leq d\log\lambda_{\max}(P_{t_{n+1}}^{-1}) = O\left(\log r_n\right), \tag{35}$$

where d is the dimension of the vector ϕ_{t_k}. Substituting (34) and (35) into (33) yields

$$\|\widetilde{\xi}_{t_{n+1}}\|^2 \leq \frac{7L^2\|\theta\|^2}{4}\frac{\sum_{k=0}^{n}\delta_k^4}{\lambda_{\min}^n} + \frac{63\gamma_{4p}^2\|\theta\|^2}{1-\gamma_{4p}} \cdot \frac{r_n}{\lambda_{\min}^n} + O\left(\frac{\log r_n}{\lambda_{\min}^n}\right), \tag{36}$$

which completes the proof of the theorem. ∎

Remark 4.3 By (36), we see that the upper bound of the estimation error $\widetilde{\xi}_{t_{n+1}}$ consists of three parts: The first part is mainly caused by the approximation error $\varDelta_{t_{k+1}}$; the second part is mainly concerned with the sensing error w_{t_k}; the third part is mainly caused by the system noise $\overline{n}_{t_k}$. Under Assumptions 3.4 and 3.6, the first part and the third part tend to zero as $n \to \infty$, and the second part tends to zero as the $4p$-restricted isometry constant γ_{4p} goes to zero. Furthermore, by Remark 2.5, we can see that the γ_{4p} can be arbitrarily small when the dimension of the sensing matrix D satisfies the inequality $d \geq 480p\log(m/4p)/\gamma_{4p}^3$.

Theorem 4.4 *Under Assumptions* 3.2–3.6*, we have the following upper bound for the estimation error of the compressed LS algorithm:*

$$\|\widetilde{\theta}_{t_{n+1}}\|^2 = O\left(\frac{C_p^2\gamma_{4p}^2}{1-\gamma_{4p}}\right) \quad \text{a.s.},$$

where $\widetilde{\theta}_{t_{n+1}} = \theta - \theta_{t_{n+1}}$ *and* $C_p = \frac{4}{\sqrt{3(1-\gamma_{4p})}-\sqrt{1+\gamma_{3p}}}$ *defined in Lemma* 2.6*.*

Proof By Assumptions 3.4–3.6 and Theorem 4.2, we have

$$\sup_n \|\widetilde{\theta}_{t_{n+1}}\|^2 \leq \frac{64M\gamma_{4p}^2\|\theta\|^2}{1-\gamma_{4p}} \quad \text{a.s.,} \tag{37}$$

where M is defined in Assumption 3.6. Note that $\widetilde{\xi}_{t_{n+1}} = D\theta - \xi_{t_{n+1}}$, by (37), the constant C in Algorithm 1 can be taken as $C = \sqrt{\frac{64M\gamma_{4p}^2\|\theta\|^2}{1-\gamma_{4p}}}$. Furthermore, by Assumption 3.6 and Lemma 2.6, we have for large n,

$$\|\theta_{t_{n+1}} - \theta\| \leq C_p\sqrt{\frac{64M\gamma_{4p}^2\|\theta\|^2}{1-\gamma_{4p}}} \quad \text{a.s.} \tag{38}$$

This completes the proof of the theorem. ∎

Remark 4.5 Theorem 4.4 gives an upper bound of the original high-dimensional estimation error, which is positively related to the RIP constant γ_{4p}. The estimation error goes to zero when γ_{4p} tends to zero.

5 A Simulation Example

In this section, we provide a simulation example to verify the performance of the compressed LS algorithm based on sampling data. Consider the following dynamic system:

$$o_t = S\varphi_t^{\mathrm{T}}\theta + n_t,$$

with the dimension $m = 100$. Set the regression vector $\varphi_t = [\sin(t), \cos(t), \sqrt{t}, 1+\sin(2t), 2-\cos(2t), \underbrace{0,\cdots,0}_{95}]^{\mathrm{T}} \in \mathbb{R}^{100}$, the system noise $\{n_t \geq 0\}$ is a standard Wiener process, and the unknown sparse parameter $\theta = [0.3, 1.2, 0.8, \underbrace{0,\cdots,0}_{94}, 0.1, 0, 0.2]^{\mathrm{T}} \in \mathbb{R}^{100}$. For the settings of Algorithm 1, we take the initial estimate as $\theta_0 = [\underbrace{0.1,\cdots,0.1}_{100}]^{\mathrm{T}}$ and the initial covariance matrix as $P_0 = \mathrm{diag}(1,1,1,1,1)$. The sampling interval δ_k is same as that in [21], i.e.,

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

where $\lceil\cdot\rceil$ is a round up operator. According to [21], we have for any given $a_0 > 0, b_0 > 0, c_0 > 1$, $\sum_{k=0}^n \delta_k^2 = \infty$ and $\sum_{k=0}^n \delta_k^4 = O\left(\sum_{k=0}^n \frac{1}{b_0^{4k}} b_0^{2k}\right) < \infty$. Thus, the time interval δ_k satisfies Assumption 3.4. Here we choose $a_0 = 0.24, b_0 = 1.2, c_0 = 200$.

The sensing matrix D is selected as a 5×100-dimensional matrix whose elements are Gaussian random variables with zero mean and variance $1/5$. Reference [40] showed that such a sensing matrix can satisfy RIP condition (3) with probability no less than $1-\mathrm{e}^{-5c}$ for some positive constant c (see Lemma 2.4). For the reconstruction procedure (i.e., Step 3 in Algorithm 1), the OMP algorithm is used to solve the optimization problem. It can be shown that the compressed regression vector $\phi_{t_k} = D\varphi_{t_k} \in \mathbb{R}^5$ can satisfy Assumptions 3.5 and 3.6 (i.e., compressed persistent excitation condition) while the original high-dimensional regression vector φ_{t_k} cannot satisfy the excitation condition in [21]. We compare our algorithm with the traditional uncompressed LS algorithm based on sampling data (cf., [21]) in Figure 1. From Figure 1, we can see that the mean square error (MSE) (averaged over 200 runs) of our compressed LS algorithm is much smaller than that of the uncompressed LS algorithm in [21], which means that for the sparse regression vectors, the compressed distributed LS algorithm has better estimation performance than the uncompressed LS algorithm based on sampling data. Moreover, in Figure 2, 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. Then we find 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 performs better than cases of fixed constant sampling intervals.

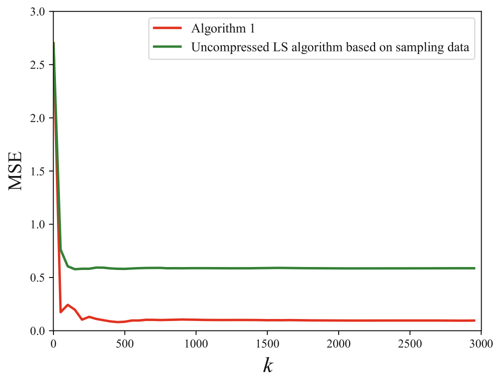

Figure 1 MSEs of the compressed and uncompressed LS algorithms based on sampling data

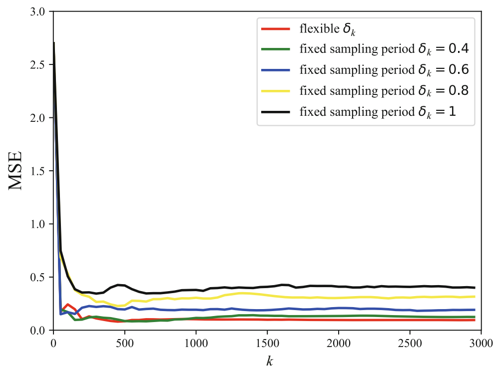

Figure 2 MSEs of the compressed LS algorithm under flexible sampling interval and constant sampling intervals

6 Concluding Remarks

In this paper, we proposed the compressed LS algorithm to estimate an unknown high-dimensional sparse parameter of continuous-time linear stochastic regression models using the sampling data based on compressed sensing methods. Under a compressed excitation condition and suitably choosing sampling time interval, we established the performance result and upper bound for the compressed parameter vector and the original high-dimensional parameter vector of the proposed compressed LS algorithm. Also, the theoretical results do not require independence and stationarity conditions on the system signals, which indicates that our theory is available for feedback systems, which plays a key role in the system and control fields. There remains some interesting open problems in the direction of this paper. For example, to consider the networked systems and the compressed distributed LS algorithm.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Widrow B and Stearns S, *Adaptive Signal Processing*, Prentice-Hall Englewood Cliffs NJ, 1985.

[2] Solo V and Kong X, *Adaptive Signal Processing Algorithms*, Prentice Hall, Upper Saddle River, 1995.

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

[4] Moore J B, On strong consistency of least squares identification algorithms, *Automatica*, 1978, **14**(5): 505–509.

[5] Lai T L and Wei C Z, Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Annals of Statistics*, 1982, **10**(1): 154–166.

[6] Guo L, Stability of recursive stochastic tracking algorithms, *SIAM Journal on Control and Optimization*, 1994, **32**(5): 1195–1225.

[7] Guo L, Convergence and logarithm laws of self-tuning regulators, *Automatica*, 1995, **31**(3): 435–450.

[8] Guo J, Jia R, Su R, et al., Identification of FIR systems with binary-valued observations against data tampering attacks, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, **53**(9): 5861–5873.

[9] Khalil H K, *Nonlinear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 2002.

[10] Chen H F and Guo L, Continuous-time stochastic adaptive tracking-robustness and asymptotic properties, *SIAM Journal on Control and Optimization*, 2006, **28**(3): 513–527.

[11] Escobar J A, Gallardo-Hernández A G, and Gonzalez-Olvera M A, How to deal with parameter estimation in continuous-time stochastic systems, *Circuits, Systems, and Signal Processing*, 2021, **41**: 2338–2357.

[12] Escobar J A and Poznyak A S, Time-varying matrix estimation in stochastic continuous-time

models under coloured noise using LSM with forgetting factor, *International Journal of Systems Science*, 2011, **42**: 2009–2020.

[13] Escobar J A and Enqvist M, Instrumental variables and LSM in continuous-time parameter estimation, *ESAIM: Control, Optimisation and Calculus of Variations*, 2017, **23**: 427–442.

[14] Poyton A A, Varziri M S, Mcauley K B, et al., Parameter estimation in continuous-time dynamic models using principal differential analysis, *Computers and Chemical Engineering*, 2006, **30**(4): 698–708.

[15] Wang W and Zhang H, H_∞ filtering for continuous-time systems with pointwise time-varying delay, *Journal of Systems Science & Complexity*, 2012, **25**(1): 90–104.

[16] Garnier H and Wang L P, *Identification of Continuous-Time Models from Sampled Data*, Springer, London, 2008.

[17] Soderstrom T, Fan H, Carlsson B, et al., Least squares parameter estimation of continuous-time ARX models from discrete-time data, *IEEE Transactions on Automatic Control*, 1997, **42**(5): 659–673.

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

[19] 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.

[20] 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.

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

[22] Zhao W, Yin G, and Bai E W, Sparse system identification for stochastic systems with general observation sequences, *Automatica*, 2020, **121**: 109162.

[23] Gan D and Liu Z, Distributed sparse identification for stochastic dynamic systems under cooperative non-persistent excitation condition, *Automatica*, 2023, **151**: 110958.

[24] Gu Y, Jin J, and Mei S, ℓ_0 norm constraint lms algorithm for sparse system identification, *IEEE Signal Processing Letters*, 2009, **16**(9): 774–777.

[25] Su G, Jin J, Gu Y, et al., Performance analysis of ℓ_0 norm constraint least mean square algorithm, *IEEE Transactions on Signal Processing*, 2012, **60**(5): 2223–2235.

[26] Shi K and Shi P, Convergence analysis of sparse LMS algorithms with ℓ_1-norm penalty based on white input signal, *Signal Processing*, 2010, **90**(12): 3289–3293.

[27] Angelosante D, Bazerque J A, and Giannakis G B, Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1-norm, *IEEE Transactions on Signal Processing*, 2010, **58**(7): 3436–3447.

[28] Xu Z, Chang X, Xu F, et al., $l_{1/2}$ regularization: A thresholding representation theory and a fast solver, *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(7): 1013–1027.

[29] Candès E J, Romberg J K, and Tao T, Stable singal recovery from incomplete and inaccurate measurements, *IEEE Transactions on Information Theory*, 2006, **59**(8): 1207–1223.

[30] Bajwa W, Haupt J, Sayeed A, et al., Compressive wireless sensing, *International Conference on Information Processing in Sensor Networks*, Nashville, TN, USA, 2006, 134–142.

[31] Hosseini S H and Shayesteh M G, Compressed sensing for denoising in adaptive system identification, 20*th Iranian Conference on Electrical Engineering*, Tehran, Iran, 2012, 1238–1242.

[32] Xu S, de Lamare R C, and Poor H V, Distributed compressed estimation based on compressive sensing, *IEEE Signal Processing Letters*, 2015, **22**: 1311–1315.

[33] Luo Y, Li S, Ma C, et al., Tail-iteratively reweighted least squares technique for compressed sensing, *China Automation Congress* (*CAC*), Xiamen, 2022, 3455–3459.

[34] Xie S and Guo L, Analysis of compressed distributed adaptive filters, *Automatica*, 2020, **112**: 108707.

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

[36] Liu Z, Liu Y, and Li C, Distributed sparse recursive least-squares over networks, *IEEE Transactions on Signal Processing*, 2014, **62**(6): 1386–1395.

[37] Candès E J and Tao T, Decoding by linear programming, *IEEE Transactions on Information Theory*, 2005, **51**(12): 4203–4215.

[38] DeVore R A, Deterministic constructions of compressed sensing matrices, *Journal of Complexity*, 2007, **23**: 918–925.

[39] Xu G and Xu Z, Compressed sensing matrices from Fourier matrices, *IEEE Transactions on Information Theory*, 2015, **61**(1): 469–478.

[40] Baraniuk R, Davenport M, DeVore R, et al., A simple proof of the restricted isometry property for random matrices, *Constructive Approximation*, 2006, **28**(3): 253–263.

[41] Tropp J and Gilbert A, Signal recovery from partial information via orthogonal matching pursuit, *IEEE Transactions on Information Theory*, 2007, **53**(2): 4655–4666.

[42] Kim S J, Koh K, Lustig M, et al., An interior-point method for large-scale ℓ_1-regularized least squares, *IEEE Journal of Selected Topics in Signal Processing*, 2007, **1**(4): 606–617.