Ziming Wang; Yiming Xing; Xinghua Zhu
Performance analysis of least squares algorithm for multivariable stochastic systems

Kybernetika, Vol. 59 (2023), No. 1, 28–44

Persistent URL: http://dml.cz/dmlcz/151582

KYBERNETIKA — VOLUME *59* (2023), NUMBER 1, PAGES 28–44

PERFORMANCE ANALYSIS OF LEAST SQUARES ALGORITHM FOR MULTIVARIABLE STOCHASTIC SYSTEMS

Ziming Wang, Yiming Xing and Xinghua Zhu

In this paper, we consider the parameter estimation problem for the multivariable system. A recursive least squares algorithm is studied by minimizing the accumulative prediction error. By employing the stochastic Lyapunov function and the martingale estimate methods, we provide the weakest possible data conditions for convergence analysis. The upper bound of accumulative regret is also provided. Various simulation examples are given, and the results demonstrate that the convergence rate of the algorithm depends on the parameter dimension and output dimension.

Keywords: least squares, martingale theory, non-persistent excitation

Classification: 93A10, 93E12, 93E24

1. INTRODUCTION

As a famous and significant issue, parameter estimation or filtering plays an important role in the areas of identification, adaptive control, and statistical learning, which appears in diverse scientific and engineering applications, including radar system [11], power system [23] and twin-rotor system [22]. The theoretical analysis of the estimation and filtering algorithms has attracted the great interest of scientists, see [2, 3, 17] for some references.

Great progress has been made in parameter estimation of discrete-time models in the past half-century. The unknown parameters to be estimated in the model include time-variant and time-invariant parameters. For time-variant parameters, the existing literature mainly focuses on the stability and performance analysis of the identification algorithm. For example, [6, 7, 10, 26] respectively studied the stability and performance of the forgetting factor least squares, the least mean square (LMS) algorithm, and the Kalman filter. For time-invariant parameters, Moore established the convergence analysis for the well-known recursive least square (RLS) algorithm under the persistent excitation (PE) condition [18]. Furthermore, the PE condition was generalized to the weakest possible excitation condition by Lai and Wei [13]. Other theoretical analyses of identification algorithms for time-invariant parameters have also been widely studied,

DOI: 10.14736/kyb-2023-1-0028

including the bias compensation least squares methods [27, 28], the stochastic gradient algorithm [1], the weighted LS algorithm, the maximum likelihood algorithm [14, 24] and Bayes methods [19, 20]. However, these theoretical analyses only consider the case where the output vector y is one-dimensional, which may fail to hold in some practical scenarios, such as Multiple Input Multiple Output (MIMO) radar systems [15] and neural networks [25].

In this paper, we consider the multivariable cases of the linear stochastic regression model, whose dynamics can be written as follows:

$$\boldsymbol{y}(k+1) = \Phi^T(k)\boldsymbol{\theta} + \boldsymbol{v}(k+1), \tag{1.1}$$

where $\boldsymbol{y}(k) = [y_1(k), y_2(k), \ldots, y_m(k)]^T \in \mathbb{R}^m$ is the output vector containing all the system output variables, $\Phi(k) \in \mathbb{R}^{n\times m}$ is the regression matrix containing the input and output data, $\boldsymbol{\theta} \in \mathbb{R}^{\boldsymbol{n}}$ is the parameter vector consisting of all the system parameters to be estimated, and $\boldsymbol{v}(k) = [v_1(k), v_2(k), \ldots, v_m(k)]^T \in \mathbb{R}^m$ is a random noise vector.

The multivariate linear regression form (1.1) for multivariable systems was early studied by Sen and Sinha [21]. They described the original model as a transfer function matrix and proposed a recursive pseudo-inverse algorithm based on least squares to avoid large matrix inverse calculation in the offline least squares method. In [16], the RLS algorithm was presented for the multivariable systems with the linear form in (1.1), and its convergence properties were theoretically analyzed under the PE condition. However, the PE condition is often difficult to verify in the commonly used feedback control systems [4]. In this paper, we devote to generalizing the PE condition used in [16] to the weakest possible Lai-Wei excitation condition [13], in the convergence analysis of the RLS algorithm for the multivariable systems. Moreover, an upper bound is provided on the accumulative regret.

The rest of this paper is organized as follows. In Section 2, we introduce the RLS algorithm for the multivariable system. After that, our main theoretical results are shown in Section 3, and the corresponding proof processes are listed in Section 4. Section 5 provides simulation examples to show the effectiveness of the RLS algorithm. Finally, the concluding remarks are made in Section 6.

2. RECURSIVE LEAST SQUARE ALGORITHM FOR MULTIVARIABLE SYSTEMS

In this section, in order to identify the unknown parameter vector $\boldsymbol{\theta}$, in (1.1), we introduce the RLS algorithm for multivariable systems. We first recall some basic notations, then introduce the RLS algorithm for the estimation problem (1.1).

2.1. Basic notations

In the sequel, $X \in \mathbb{R}^n$ is viewed as an n-dimensional column vector, and $A \in \mathbb{R}^{m\times n}$ is viewed as an $m \times n$ dimensional matrix. Let $\lambda_{\max}(\cdot)$ and $\lambda_{\min})(\cdot)$ denote the largest and the smallest eigenvalues of the corresponding matrix, respectively. Throughout this paper, we use $|\cdot|$ to denote the determinant of the corresponding matrix, which should not be confused with the absolute value of a scalar from the context.

For any matrix $A \in \mathbb{R}^{m\times n}$, $\|A\|$ denotes the operator norm induced by the Euclidean norm, i. e., $\left(\lambda_{\max}\left(AA^T\right)\right)^{1/2}$, where $(\cdot)^T$ denotes the transpose operator. For two symmetric matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$, $A > B (A \geq B)$ means $A - B$ is a positive definite (semidefinite) matrix. The symbol $\mathrm{E}[\cdot]$ is used to denote the mathematical expectation operator, and $\mathrm{E}\left[\cdot \mid \mathcal{F}_k\right]$ denote the conditional mathematical expectation operator, where $\{\mathcal{F}_k\}$ is a sequence of nondecreasing σ-algebras [5]. We also use $\log(\cdot)$ to denote the natural logarithm function, and $\mathrm{tr}(\cdot)$ to denote the trace of a symmetric matrix.

Let $\{A_k, k \geq 0\}$ be a matrix sequence and $\{b_k, k \geq 0\}$ be a positive scalar sequence. Then $A_k = O\left(b_k\right)$ means that there exists a constant $M > 0$ such that $\|A_k\| \leq Mb_k, \forall k \geq 0$. Also, $A_k = o\left(b_k\right)$ means that $\lim_{k\to\infty} \|A_k\| / b_k = 0$.

2.2. Recursive least squares algorithm for multivariable systems

To estimate the unknown parameter vector in model (1.1), we introduce the following accumulative prediction error:

$$J(\boldsymbol{\theta}) := \sum_{j=0}^{t} \|\boldsymbol{y}(j+1) - \Phi^T(j)\boldsymbol{\theta}\|^2. \tag{2.1}$$

We can obtain the following recursive least squares (RLS) algorithm by minimizing the above prediction error,

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \mathbf{P}(k)\Phi(k)A(k)\big[\boldsymbol{y}(k+1) - \Phi^T(k)\hat{\boldsymbol{\theta}}(k)\big], \tag{2.2}$$

$$A(k) = \left(\mathbf{I}_m + \Phi^T(k)\mathbf{P}(k)\Phi(k)\right)^{-1}, \tag{2.3}$$

$$\mathbf{P}(k+1) = \left(\mathbf{I}_n - \mathbf{P}(k)\Phi(k)A(k)\Phi^T(k)\right)\mathbf{P}(k), \tag{2.4}$$

where $\mathbf{P}(k)$ is the covariance matrix of dimension $n \times n$. The initial estimate of $\hat{\boldsymbol{\theta}}(k)$ is generally set as $\hat{\boldsymbol{\theta}}(\mathbf{0}) = C_1\mathbf{1}_n$, and the initial positive definite matrix $\mathbf{P}(0)$ can be taken as $\mathbf{P}(0) = C_2\mathbf{I}_n$, where $C_1 \in \mathbb{R}$ and $C_2 > 0$ are two constants. We remark that for the RLS algorithm (2.2)–(2.4), it does not require computing the matrix inverse $\mathbf{P}^{-1}(k)$ at each recursion. For the convenience of subsequent theoretical analysis, we can obtain the following equivalent recursive form of RLS algorithm [16] by applying the matrix inversion lemma [9] $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$:

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \mathbf{P}(k+1)\Phi(k)[\boldsymbol{y}(k+1) - \Phi^T(k)\hat{\boldsymbol{\theta}}(k)], \tag{2.5}$$

$$\mathbf{P}^{-1}(k+1) = \mathbf{P}^{-1}(k) + \Phi(k)\Phi^T(k), \quad \mathbf{P}(0) = p_0\mathbf{I}_n. \tag{2.6}$$

3. ASYMPTOTIC RESULTS OF THE ALGORITHM

3.1. Analysis of parameter convergence

In this section, we give our main convergence results. We start by introducing some basic assumptions to be used throughout this paper. For the system in (1.1) and the RLS

algorithm in (2.5) – (2.6), we assume that $\{\boldsymbol{v}(k), \mathcal{F}_k\}$ is a martingale difference vector sequence defined on a probability space $\{\Omega, \mathcal{F}, P\}$, i. e.,

$$\mathrm{E}\left[\boldsymbol{v}(k+1) \mid \mathcal{F}_k\right] = \mathbf{0}, \quad \text{a.s.} \tag{3.1}$$

where $\mathcal{F}_k$ is the σ-algebra generated by the observations up to and including time k. Moreover, we assume that

$$\mathrm{E}\left[\|\boldsymbol{v}(k+1)\|^2 \mid \mathcal{F}_k\right] = \sigma^2 < \infty, \quad \text{a.s.} \tag{3.2}$$

We denote by $\mathbf{I}$ the identity matrix with appropriate dimensions. Define the parameter estimation error vector as

$$\tilde{\boldsymbol{\theta}}(k) := \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(k). \tag{3.3}$$

Inserting (1.1) into (2.2) gives the following recursive form of the error vector:

$$\tilde{\boldsymbol{\theta}}(k+1) = \left(\mathbf{I} - \mathbf{P}(k)\Phi(k)A(k)\Phi^T(k)\right)\tilde{\boldsymbol{\theta}}(k) - \mathbf{P}(k)\Phi(k)A(k)\boldsymbol{v}(k+1). \tag{3.4}$$

The following theorem provides some properties of the error vector, which is crucial in the convergence analysis.

Theorem 3.1. Assume (3.1) – (3.2) hold. As $t \to \infty$, we have

$$1) \quad \tilde{\boldsymbol{\theta}}^T(t+1)\mathbf{P}^{-1}(t+1)\tilde{\boldsymbol{\theta}}(t+1) = O(\log r_t), \quad \text{a.s.} \tag{3.5}$$

$$2) \quad \sum_{k=0}^{t} \|A^{\frac{1}{2}}\Phi^T(k)\tilde{\boldsymbol{\theta}}(k)\|^2 = O(\log r_t), \quad \text{a.s.} \tag{3.6}$$

where r_t is defined by $r_t = 1 + \sum\limits_{k=0}^{t} \|\Phi(k)\|^2$.

By using result 1) of Theorem 3.1, we can obtain the following theorem on the convergence rate of the RLS algorithm:

Theorem 3.2. Assume (3.1) – (3.2) hold. As $t \to \infty$, we have

$$\|\tilde{\boldsymbol{\theta}}_{t+1}\|^2 = O\left(\frac{\log r_t}{\lambda_{\min}\left(\mathbf{P}^{-1}(t+1)\right)}\right), \quad \text{a.s.}$$

Remark 3.1. Theorem 3.2 shows that if

$$\lim_{t\to\infty} \frac{\log\left(1 + \sum\limits_{k=0}^{t} \|\Phi(k)\|^2\right)}{\lambda_{\min}\left(\mathbf{P}^{-1}(0) + \sum\limits_{k=0}^{t} \Phi(k)\Phi^T(k)\right)} = 0, \quad \text{a.s.} \tag{3.7}$$

then the RLS estimate $\hat{\boldsymbol{\theta}}(t)$ will converge to the true unknown parameter. In the traditional single variable case (where $m = 1$), (3.7) reduces to the well-known Lai-Wei excitation condition, which is known to be the weakest possible data condition for the

convergence of the classical LS estimates [13], and is much weaker than the well-known persistence of excitation (PE) condition or sufficient excitation (SE) condition usually used in the parameter estimation of finite-dimensional linear control systems. The PE or SE condition is defined as follows:

$$0 < a\mathbf{I}_n < \sum_{k=t}^{t+h} \Phi(k)\Phi^T(k) \leq b\mathbf{I}_n < \infty, \quad \forall\, t \geq 0, \tag{3.8}$$

where h is a positive integer, a and b are two positive constants. By this inequality, we can easily derive that (3.7) holds. But the reverse does not.

Remark 3.2. We remark that Theorem 3.2 gives the rate of convergence of the estimation error as $O\left(\frac{\log r_t}{\lambda_{\min}(\mathbf{P}^{-1}(t+1))}\right)$, whereas Theorem 1 in [16] gives the rate of convergence as $O\left(\frac{r_t^c}{\lambda_{\min}(\mathbf{P}^{-1}(t+1))}\right)$ with $c \in (0,1)$. It is clear that the convergence result in Theorem 3.2 is much better than Theorem 1 in [16].

3.2. Regret analysis

We all know that regret is an important performance metric for measuring online learning algorithms. In this section, we will analyze the adaptive prediction ability of the proposed algorithm.

The result 2) of Theorem 3.1 is important for the analysis of the regret of the RLS algorithm, where the regret of RLS algorithm is defined as

$$R_k = \left\{ \mathrm{E}\left[\boldsymbol{y}(k+1)|\mathcal{F}_k\right] - \Phi^T(k)\hat{\boldsymbol{\theta}}(k) \right\}^2 = \left\{ \Phi^T(k)\boldsymbol{\theta} - \Phi^T(k)\hat{\boldsymbol{\theta}}(k) \right\}^2, \tag{3.9}$$

in which $\boldsymbol{\theta}$ is the true parameter and $\hat{\boldsymbol{\theta}}(k)$ is the estimation of the parameter at kth step.

Remark 3.3. Since the noise sequence $\{\boldsymbol{v}(k), \mathcal{F}_k\}$ is a martingale difference vector sequence, the best prediction to the future observation $\boldsymbol{y}(k+1)$ at any time instant $k \geq 1$ is $\mathrm{E}\left[\boldsymbol{y}(k+1)|\mathcal{F}_k\right] = \Phi^T(k)\boldsymbol{\theta}$. Its adaptive predictor is constructed as $\Phi^T(k)\hat{\boldsymbol{\theta}}(k)$. The squared difference between the best prediction and the adaptive prediction can be regarded as the regret, i. e.,

$$R_k = \left\{ \Phi^T(k)\boldsymbol{\theta} - \Phi^T(k)\hat{\boldsymbol{\theta}}(k) \right\}^2 = \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k),$$

where $\tilde{\boldsymbol{\theta}}(k) = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(k)$ is the parameter estimation error vector.

We develop the following theorem concerning the upper bound on the accumulative regret for the above adaptive predictor.

Theorem 3.3. Assume (3.1) – (3.2) hold. Then, the sample path of the accumulated regrets has the following bound as $t \to \infty$:

$$\sum_{k=0}^{t} R_k = O\left(\log r_t\right), \quad \text{a.s.}$$

provided that $\|\Phi^T(t)\boldsymbol{P}(t)\Phi(t)\| = O(1)$, a.s.

Remark 3.4. We remark that the order $O\left(\log r_t\right)$ for the accumulated regrets may be shown to be the best possible among all adaptive predictors in a certain sense, as is already known in the traditional single variable case. And the proposed upper bound on the accumulative regret, $O(\log r_t)$, is the minimum order of magnitude that one may expect to achieve (cf. [12]). Moreover, the precise constant in $O(\cdot)$ may be determined if we have further conditions on the regressors (see [8] in the single variable case).

4. PROOF

In this section, we give proof of our main results. We start by proving the following elementary lemma.

Lemma 4.1. (Theorem 2.8 in [3]) Suppose that $\{a_k, \mathcal{F}_k\}$ is an adapted process and $\{v_k, \mathcal{F}_k\}$ is a martingale difference sequence (a_k and v_k can be a random number or a random matrix) satisfying $\sup_t \mathbb{E}[\|v_{t+1}\|^\beta | \mathcal{F}_t] < \infty$ for some $\beta \in (0, 2]$. Then for any $\eta > 0$, we have

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

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

The following lemma gives a property about the trace of positive semidefinite matrices, which is useful in the following analysis.

Lemma 4.2. If $A > 0$, $0 \le B \le C$ are symmetric matrices of the same dimension n, we have

$$\operatorname{tr}\left((A+B)^{-1}B\right) \le \operatorname{tr}\left((A+C)^{-1}C\right).$$

P r o o f. By $A > 0$, there exists an invertible matrix P_A such that $P_A A P_A^T = \mathbf{I}$. Since $P_A B P_A^T$ and $P_A C P_A^T$ are also positive semidefinite, there exist orthogonal matrices Q_B and Q_C such that

$$Q_B P_A B P_A^T Q_B^T = \Lambda_B, \quad Q_C P_A B P_A^T Q_C^T = \Lambda_C,$$

where Λ_B and Λ_C are diagonal matrices. We denote by $0 \le \lambda_1(\Lambda_B) \le \cdots \le \lambda_n(\Lambda_B)$ and $0 \le \lambda_1(\Lambda_C) \le \cdots \le \lambda_n(\Lambda_C)$ the eigenvalues of Λ_B and Λ_C, respectively. Meanwhile, it is obvious that

$$Q_B P_A A P_A^T Q_B^T = \mathbf{I}, \quad Q_C P_A A P_A^T Q_C^T = \mathbf{I}.$$

Thus,

$$\begin{aligned}\operatorname{tr}\left((A+B)^{-1}B\right) &= \operatorname{tr}\left((P_A^{-1}Q_B^{-1}(\mathbf{I}+\Lambda_B)Q_B^{-T}P_A^{-T})^{-1}P_A^{-1}Q_B^{-1}\Lambda_B Q_B^{-T}P_A^{-T}\right)\\ &= \operatorname{tr}\left(P_A^T Q_B^T(\mathbf{I}+\Lambda_B)^{-1}\Lambda_B Q_B^{-T}P_A^{-T}\right)\\ &= \sum_{i=1}^{n}\frac{\lambda_i(\Lambda_B)}{1+\lambda_i(\Lambda_B)}.\end{aligned} \tag{4.1}$$

Similarly,

$$\operatorname{tr}\left((A+C)^{-1}C\right) = \sum_{i=1}^{n}\frac{\lambda_i(\Lambda_C)}{1+\lambda_i(\Lambda_C)}. \tag{4.2}$$

By $B \le C$ we have $\Lambda_B \le \Lambda_C$, i. e. $\lambda_i(\Lambda_B) \le \lambda_i(\Lambda_C)$ for $i = 1, \dots, n$. Hence, by (4.1) and (4.2), we can finish the proof. □

Lemma 4.3. We have

$$\operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right) \le m^2\frac{|\mathbf{P}^{-1}(k+1)| - |\mathbf{P}^{-1}(k)|}{|\mathbf{P}^{-1}(k+1)|}. \tag{4.3}$$

P r o o f . From the definition of $\mathbf{P}(k)$, we have

$$\begin{aligned}\mathbf{P}^{-1}(k+1) &= \mathbf{P}^{-1}(k) + \Phi(k)\Phi^T(k)\\ &\ge \mathbf{P}^{-1}(k) + \phi_i(k)\phi_i^T(k) = \mathbf{P}^{-1}(k)\left[\mathbf{I} + \mathbf{P}(k)\phi_i(k)\phi_i^T(k)\right].\end{aligned} \tag{4.4}$$

Taking determinant on both sides and using the fact that $|\mathbf{I}+BC| = |\mathbf{I}+CB|$, we have

$$|\mathbf{P}^{-1}(k+1)| \ge |\mathbf{P}^{-1}(k)||\mathbf{I}+\mathbf{P}(k)\phi_i(k)\phi_i^T(k)| = |\mathbf{P}^{-1}(k)|\left(1+\phi_i^T(k)\mathbf{P}(k)\phi_i(k)\right). \tag{4.5}$$

Therefore, we have

$$\phi_i^T(k)\mathbf{P}(k)\phi_i(k) \le \frac{\left|\mathbf{P}^{-1}(k+1)\right| - \left|\mathbf{P}^{-1}(k)\right|}{|\mathbf{P}^{-1}(k)|}. \tag{4.6}$$

Thus,

$$\operatorname{tr}\left(\Phi^T(k)\mathbf{P}(k)\Phi(k)\right) \le \sum_{i=1}^{m}\phi_i^T(k)\mathbf{P}(k)\phi_i(k) \le m \times \frac{\left|\mathbf{P}^{-1}(k+1)\right| - \left|\mathbf{P}^{-1}(k)\right|}{|\mathbf{P}^{-1}(k)|}. \tag{4.7}$$

We denote

$$t_k = \frac{\left|\mathbf{P}^{-1}(k+1)\right| - \left|\mathbf{P}^{-1}(k)\right|}{|\mathbf{P}^{-1}(k)|},$$

then (4.7) leads to

$$\Phi^T(k)\mathbf{P}(k)\Phi(k) \le mt_k\mathbf{I}. \tag{4.8}$$

Noticing that $A(k) = \left(\mathbf{I} + \Phi^T(k)\mathbf{P}(k)\Phi(k)\right)^{-1}$, by Lemma 4.2 and (4.8) we have

$$\begin{aligned} &\operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right) \\ &\leq \operatorname{tr}\left(\left(\mathbf{I} + mt_k\mathbf{I}\right)^{-1} mt_k\mathbf{I}\right) = m^2 \times \frac{t_k}{1+mt_k} \\ &\leq m^2 \times \frac{t_k}{1+t_k} = m^2 \frac{|\mathbf{P}^{-1}(k+1)| - |\mathbf{P}^{-1}(k)|}{|\mathbf{P}^{-1}(k+1)|}, \end{aligned} \tag{4.9}$$

which completes the proof. □

We are now in the position to complete the proof of Theorem 3.1.

P r o o f . of Theorem 3.1. Multiplying $\mathbf{P}^{-1}(k)$ to both sides of (2.4), we have

$$\mathbf{P}(k+1)\mathbf{P}^{-1}(k) = \mathbf{I} - \mathbf{P}(k)\Phi(k)A(k)\Phi^T(k). \tag{4.10}$$

Similarly, multiplying $\mathbf{P}(k)$ to both side of (2.6), we have

$$\mathbf{P}^{-1}(k+1)\mathbf{P}(k) = \mathbf{I} + \Phi(k)\Phi^T(k)\mathbf{P}(k). \tag{4.11}$$

Now (3.4) can be written as

$$\tilde{\boldsymbol{\theta}}(k+1) = \mathbf{P}(k+1)\mathbf{P}^{-1}(k)\tilde{\boldsymbol{\theta}}(k) - \mathbf{P}(k)\Phi(k)A(k)\boldsymbol{v}(k+1). \tag{4.12}$$

Consider the Lyapunov function $\boldsymbol{V}(k) = \tilde{\boldsymbol{\theta}}^T(k)\mathbf{P}^{-1}(k)\tilde{\boldsymbol{\theta}}(k)$. Using (4.10) – (4.12), we have

$$\begin{aligned} \boldsymbol{V}(k+1) &= \tilde{\boldsymbol{\theta}}^T(k+1)\mathbf{P}^{-1}(k+1)\tilde{\boldsymbol{\theta}}(k+1) \\ &= \left[\tilde{\boldsymbol{\theta}}^T(k)(\mathbf{I} - \Phi(k)A(k)\Phi^T(k)\mathbf{P}(k)) - \boldsymbol{v}^T(k+1)A(k)\Phi^T(k)\mathbf{P}(k)\right] \\ &\quad \times \left[\mathbf{P}^{-1}(k)\tilde{\boldsymbol{\theta}}(k) - \mathbf{P}^{-1}(k+1)\mathbf{P}(k)\Phi(k)A(k)\boldsymbol{v}(k+1)\right] \\ &= \tilde{\boldsymbol{\theta}}^T(k)\mathbf{P}^{-1}(k)\tilde{\boldsymbol{\theta}}(k) - \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k) - 2\tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\boldsymbol{v}(k+1) \\ &\quad + \boldsymbol{v}^T(k+1)A(k)\Phi^T(k)\mathbf{P}(k)\mathbf{P}^{-1}(k+1)\mathbf{P}(k)\Phi(k)A(k)\boldsymbol{v}(k+1) \\ &= \boldsymbol{V}(k) - \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k) - 2\tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\boldsymbol{v}(k+1) \\ &\quad + \boldsymbol{v}^T(k+1)A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\boldsymbol{v}(k+1). \end{aligned} \tag{4.13}$$

Taking sum of (4.13) from 0 to t, we have

$$\begin{aligned} &\boldsymbol{V}(t+1) + \sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k) \\ &= \boldsymbol{V}(0) - 2\sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\boldsymbol{v}(k+1) + \sum_{k=0}^{t} \boldsymbol{v}^T(k+1)A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\boldsymbol{v}(k+1) \\ &\leq \boldsymbol{V}(0) - 2\sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\boldsymbol{v}(k+1) + \sum_{k=0}^{t} \operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right)\|\boldsymbol{v}(k+1)\|^2. \end{aligned} \tag{4.14}$$

Next, we evaluate the last two terms of (4.14). Noticing that $0 < A(k) \leq \mathbf{I}$ and $\tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k) \in \mathcal{F}_k$, for any $\delta > 0$ we can use Lemma 4.1 to get the following estimation of the second last term of (4.14):

$$\sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\boldsymbol{v}(k+1) = O\left(\left\{\sum_{k=0}^{t}\left\|A^{\frac{1}{2}}(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k)\right\|^2\right\}^{\frac{1}{2}+\delta}\right). \tag{4.15}$$

Taking $0 < \delta < 1/2$, we have

$$\begin{aligned}\sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\boldsymbol{v}(k+1) &= O(1) + o\left(\sum_{k=0}^{t}\left\|A^{\frac{1}{2}}(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k)\right\|^2\right)\\ &= O(1) + o\left(\sum_{k=0}^{t}\tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k)\right).\end{aligned} \tag{4.16}$$

To evaluate the last term of (4.14), we first use Lemma 4.3 to get

$$\begin{aligned}&\sum_{k=0}^{t}\operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right) \leq m^2\sum_{k=0}^{t}\frac{|\mathbf{P}^{-1}(k+1)| - |\mathbf{P}^{-1}(k)|}{|\mathbf{P}^{-1}(k+1)|}\\ &\leq m^2\sum_{k=0}^{t}\int_{|\mathbf{P}^{-1}(k)|}^{\mathbf{P}^{-1}(k+1)}\frac{1}{x}\,\mathrm{d}x = m^2(\log|\mathbf{P}^{-1}(t+1)| + \log|\mathbf{P}(0)|).\end{aligned} \tag{4.17}$$

Since $\left\{\|\boldsymbol{v}(k+1)\|^2 - \mathrm{E}[\|\boldsymbol{v}(\mathrm{k}+1)\|^2|\mathcal{F}_\mathrm{k}], \mathcal{F}_\mathrm{k}\right\}$ is a martingale difference sequence and

$$\begin{aligned}&\sup_k \mathrm{E}\left[\left|\|\boldsymbol{v}(\mathrm{k}+1)\|^2 - \mathrm{E}[\|\boldsymbol{v}(\mathrm{k}+1)\|^2|\mathcal{F}_\mathrm{k}]\right||\mathcal{F}_\mathrm{k}\right]\\ &\leq 2\sup_k \mathrm{E}\left[\|\boldsymbol{v}(\mathrm{k}+1)\|^2|\mathcal{F}_\mathrm{k}\right] < \infty,\end{aligned}$$

we can use Lemma 4.2 and (4.17) to get

$$\begin{aligned}&\sum_{k=0}^{t}\operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right)\left\{\|\boldsymbol{v}(k+1)\|^2 - \mathrm{E}[\|\boldsymbol{v}(\mathrm{k}+1)\|^2|\mathcal{F}_\mathrm{k}]\right\}\\ &= O\left(\sum_{k=0}^{t}\operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right)\cdot\log^{1+\eta}\left(\sum_{k=0}^{t}\operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right) + e\right)\right)\\ &= o(\log|\mathbf{P}^{-1}(t+1)|) + O(1) \quad \text{a.s.}\end{aligned}$$

Thus by (4.17), the last term of (4.14) satisfies

$$\begin{aligned}
&\sum_{k=0}^{t} \operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right)\|\boldsymbol{v}(k+1)\|^2 \\
&\le \sigma^2 \sum_{k=0}^{t} \operatorname{tr}\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right) + o(\log|\mathbf{P}^{-1}(n+1)|) + O(1) \\
&\le m^2 \log|\mathbf{P}^{-1}(t+1)| + o(\log|\mathbf{P}^{-1}(t+1)|) + O(1) \\
&\le m^2 n \log\lambda_{max}\left(\mathbf{P}^{-1}(t+1)\right) + o(\log|\mathbf{P}^{-1}(t+1)|) + O(1) \\
&\le m^2 n \log r_t + o(\log|\mathbf{P}^{-1}(t+1)|) + O(1) = O(\log r_t).
\end{aligned} \tag{4.18}$$

Inserting (4.16) and (4.18) into (4.14), we have

$$\boldsymbol{V}(t+1) + \sum_{k=0}^{t} \|A^{\frac{1}{2}}(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k)\|^2 = O(\log r_t),$$

thus we conclude that (3.5) – (3.6) hold. □

Based on the results of Theorem 3.1, the rest of the theorems can be proved.

P r o o f . of Theorem 3.2. By Theorem 3.1 and the fact that $\mathbf{P}(t+1) > 0$, we have

$$\|\tilde{\boldsymbol{\theta}}_{t+1}\|^2 \le \frac{\tilde{\boldsymbol{\theta}}^T(t+1)\mathbf{P}^{-1}(t+1)\tilde{\boldsymbol{\theta}}(t+1)}{\lambda_{\min}\left(\mathbf{P}^{-1}(t+1)\right)} = O\left(\frac{\log r_t}{\lambda_{\min}\left(\mathbf{P}^{-1}(t+1)\right)}\right) \quad \text{a.s.}$$

□

P r o o f . of Theorem 3.3. By the definition of $A(k)$ in (2.3), we have

$$\Phi(k)\Phi^T(k) = \Phi(k)A(k)\Phi^T(k) + \Phi(k)\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right)\Phi^T(k).$$

Then by $\|\Phi^T(k)\boldsymbol{P}(k)\Phi(k)\| = O(1)$ a.s. and Theorem 3.1, we know that

$$\begin{aligned}
&\sum_{k=0}^{t} R_k = \sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k) \\
&= \sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)A(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k) + \sum_{k=0}^{t} \tilde{\boldsymbol{\theta}}^T(k)\Phi(k)\left(A(k)\Phi^T(k)\mathbf{P}(k)\Phi(k)\right)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k) \\
&= O\left(\sum_{k=0}^{t} \|A^{\frac{1}{2}}(k)\Phi^T(k)\tilde{\boldsymbol{\theta}}(k)\|^2\right) = O(\log r_t),
\end{aligned}$$

which completes the proof. □

5. SIMULATION EXAMPLES

In this section, we present numerical examples to illustrate the performance of the proposed algorithm. Specifically, we test the convergence property of the proposed algorithm in two cases, where in the first, PE and condition (3.7) are both satisfied, while in the second, condition (3.7) is satisfied but the PE condition is not.

5.1. Convergence property when PE and condition (3.7) are both satisfied

Consider a multivariate linear regression system with $m = 3$ and $n = 4$,

$$\boldsymbol{y}(t) = \Phi^T(t)\boldsymbol{\theta} + \boldsymbol{v}(t), \tag{5.1}$$

where $\boldsymbol{y}(t) = (y_1(t), y_2(t), y_3(t))^T \in \mathbb{R}^3$ is the system output, $\Phi(t) \in \mathbb{R}^{4\times 3}$ is the regression matrix of input and output data, $\boldsymbol{v}(t) = (v_1(t), v_2(t), v_3(t))^T \in \mathbb{R}^3$ is a white noise vector sequence with zero mean and variances $\sigma_1^2 = 0.10^2$ for $v_1(t)$, $\sigma_2^2 = 0.20^2$ for $v_2(t)$ and $\sigma_3^2 = 0.30^2$ for $v_3(t)$, respectively, and $\boldsymbol{\theta} \in \mathbb{R}^4$ is the parameter vector to be determined. Here, the true $\boldsymbol{\theta}$ is taken as

$$\boldsymbol{\theta} = [1.78, 2.75, 0.97, 4.26]^T.$$

The initial value is selected as $\hat{\boldsymbol{\theta}}(0) = [0.01, 0.01, 0.01, 0.01]^T$, $P(0) = \mathbf{I}_4$. In the simulation, the regression matrix $\{\Phi(t)\}$ is designed to be a 4×3 matrix sequence containing pseudo-random values generated from a normal distribution. One may check that $\{\Phi(t)\}$ satisfies the exciting condition (3.7). Using the true parameters, we calculate the response values $\boldsymbol{y}(t), t = 1, 2, \ldots$ by the linear regression form (5.1). Then RLS algorithm is used to get the parameter estimation $\hat{\boldsymbol{\theta}}(t)$ from the input-output data $\{\boldsymbol{y}(t), \Phi(t)\}$. The estimation accuracy is evaluated by the relative error, which is defined by

$$\epsilon := \frac{\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|}{\|\boldsymbol{\theta}\|}.$$

The averaged accumulative regrets is defined by

$$\text{Regret}_t = \frac{1}{t}\sum_{k=0}^{t} R_k.$$

In this simulation study, we collect 1000 data $\{\Phi(t), \boldsymbol{y}(t), t = 1, 2, \ldots, 1000\}$, where the first 500 data is used to estimate the parameter of the linear regression model, and the rest 500 data is used for model validation. The predicted output can be obtained by

$$\hat{y}(t) = \Phi(t)\hat{\boldsymbol{\theta}}(500), \quad t = 501, 502, \ldots, 1000.$$

The estimated parameters are shown in Table 5.1 and Figure 5.1, which shows that all the estimations converge fast to their true values as t increases. The relative error and averaged accumulative regrets are shown in Figure 5.2. It is clear that the relative errors ϵ and averaged accumulative regrets approach zero very fast. The predicted outputs and the true outputs are illustrated in Figure 5.3. It shows that the estimated model can fit the validation data well.

5.2. Convergence property when Condition (3.7) is satisfied but the PE/SE condition is not

In this subsection, we consider a case where the condition (3.7) is satisfied but the PE/SE condition is not satisfied. The results show that both the parameter estimates and the

k	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\epsilon(\%)$
5	1.01205	2.36697	0.77317	3.36507	22.10651
10	1.44907	2.51318	0.71106	3.60785	13.13711
20	1.67966	2.59640	0.87289	3.96180	6.62003
50	1.70344	2.68361	0.92971	4.12768	3.00724
100	1.74821	2.73247	0.96986	4.15996	1.92715
200	1.75337	2.72663	0.96776	4.21953	1.00738
500	1.77032	2.74553	0.97514	4.25187	0.26291
True values	1.78000	2.75000	0.97000	4.26000	

Tab. 5.1. The parameter estimates and their errors against time k.

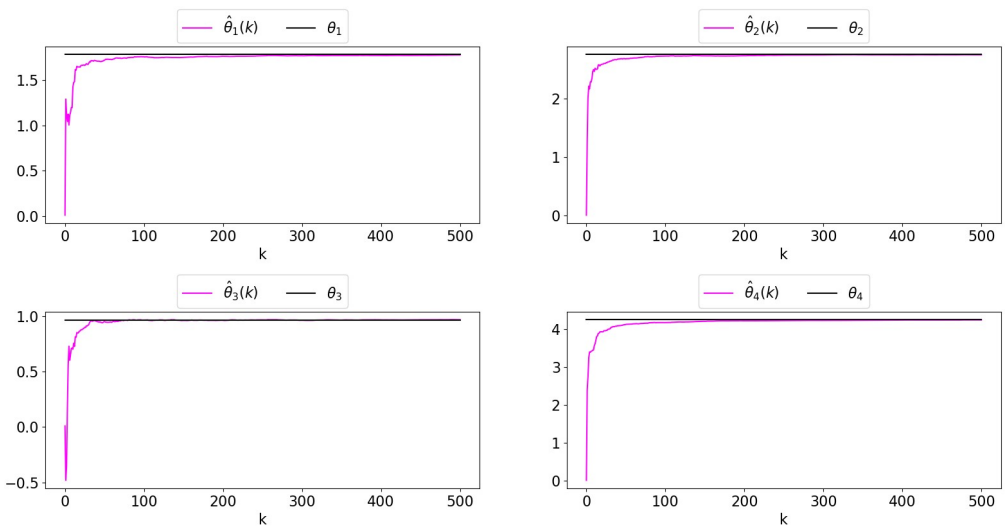
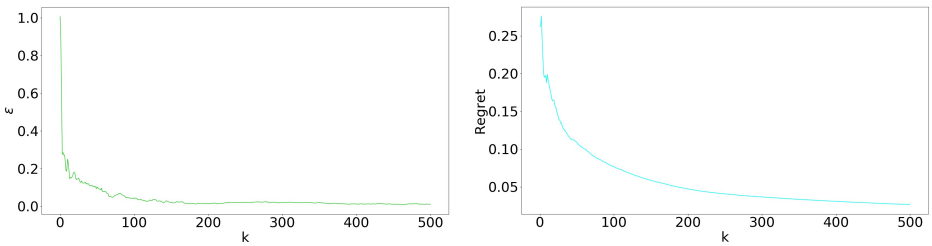


Fig. 5.1. The parameter estimates against time k.



(a) Relative error (b) Regret

Fig. 5.2. The relative errors and the averaged accumulative regrets against time k.

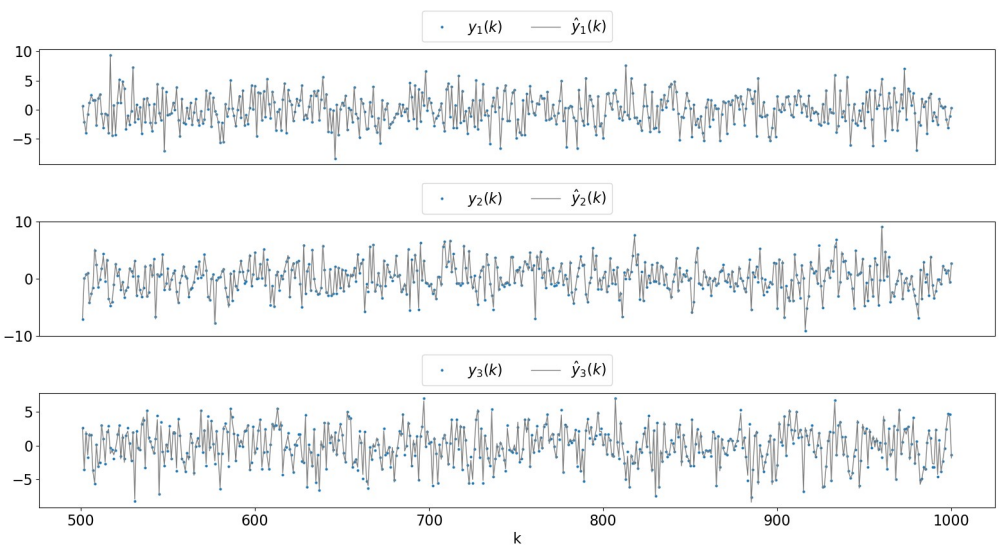


Fig. 5.3. The prediction outputs and the true outputs against time k.

regrets can converge even if the PE/SE condition is not satisfied, which fully validates our theoretical results.

Consider a multivariate linear regression system with $m = 5$ and $n = 5$,

$$\boldsymbol{y}(t) = \Phi^T(t)\boldsymbol{\theta} + \boldsymbol{v}(t),$$

where $\boldsymbol{v}(t) = (v_1(t), v_2(t), \ldots, v_5(t))^T \in \mathbb{R}^5$ is a white noise vector sequence with zero mean and variances $\sigma_i^2 = 0.10^2$ for $v_i(t)$, $i = 1, 2, \ldots, 5$. The true $\boldsymbol{\theta}$ is taken as

$$\boldsymbol{\theta} = [1.13, 0.75, 3.73, 0.18, 0.94]^T.$$

Let the regression matrix $\Phi(t) \in R^{5\times 5}$ be given by the following state space model:

$$\begin{aligned}\mathbf{u}(t) &= A(t)\mathbf{u}(t-1) + B(t)\\ \Phi(t) &= C(t)\mathbf{u}(t),\end{aligned}$$

and the state coefficient matrices are taken as follows:

$$\begin{aligned}A(t) &= \operatorname{diag}\left\{\sqrt{\frac{t}{t+1}}, \sqrt{\frac{t}{t+1}}, \cdots, \sqrt{\frac{t}{t+1}}\right\} \in \mathbb{R}^{5\times 5},\\ B(t) &= \operatorname{diag}\left\{\frac{0.7}{2^t}, \frac{0.4}{2^t}, \frac{1.1}{2^t}, \frac{0.9}{2^t}, \frac{1.3}{2^t}\right\} \in \mathbb{R}^{5\times 5},\\ C(t) &= \begin{bmatrix} 1.0 & 0.5 & & 0.8 & \\ 0.5 & 1.7 & 0.2 & & 0.1 \\ & 0.2 & 1.6 & & 0.6 \\ 0.8 & & & 1.1 & \\ & 0.1 & 0.6 & & 1.3 \end{bmatrix} \in \mathbb{R}^{5\times 5}.\end{aligned}$$

Let $\mathbf{u}(0) = \operatorname{diag}\{1.0, 1.0, \cdots, 1.0\} \in \mathbb{R}^{5\times 5}$ be the initial state.

By the definition of $\mathbf{u}(t)$, we have

$$\begin{aligned}\|\mathbf{u}(t)\| &= \left\| \frac{1}{\sqrt{t+1}}\mathbf{u}(0) + \sum_{i=1}^{t}\left(B(i)\prod_{j=i+1}^{t}A(j)\right)\right\|\\ &= \left\| \frac{1}{\sqrt{t+1}}\mathbf{u}(0) + \sum_{i=1}^{t}\left(\frac{B(0)}{2^i}\cdot\sqrt{\frac{i+1}{t+1}}\right)\right\|\\ &< \left\| \frac{1}{\sqrt{t+1}}\left(\mathbf{u}(0) + 3B(0)\right)\right\|\\ &= \frac{4.9}{\sqrt{t+1}}.\end{aligned}$$

Thus, we have

$$\|\Phi(t)\| \le \|C(t)\|\|\mathbf{u}(t)\| < 2.28 \cdot \frac{4.9}{\sqrt{t+1}} = \frac{11.172}{\sqrt{t+1}}. \tag{5.2}$$

Based on this, we can see that the regression vector $\Phi(t)$ does not satisfy the PE/SE condition, since for any fixed h,

$$\left\|\sum_{k=t}^{t+h}\Phi(k)\Phi^T(k)\right\| \le \sum_{k=t}^{t+h}\|\Phi(k)\|^2 < \sum_{k=t}^{t+h}\frac{125}{k+1} < \frac{125h}{t+1} \to 0, \quad \text{as } t\to\infty,$$

which implies (3.8) is not satisfied because the positive constant a in (3.8) does not exist.

On the other hand, condition (3.7) is satisfied because $\lambda_{\min}\left(\sum_{k=0}^{t}\Phi(k)\Phi^T(k)\right)$ goes to infinity as t increases. Specifically,

$$\lambda_{\min}\left(\Phi(t)\Phi^T(t)\right) = \lambda_{\min}\left(C(t)\mathbf{u}(t)\mathbf{u}^T(t)C^T(t)\right) \ge \frac{\lambda_{\min}\left(C(t)C^T(t)\right)}{t+1} > \frac{0.02}{t+1}.$$

By (5.2) and the fact that

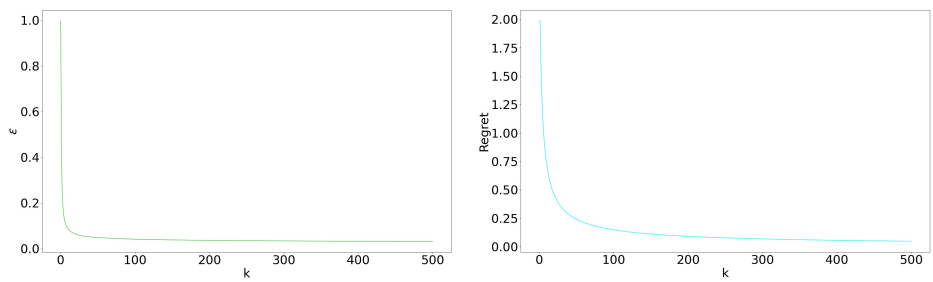
$$\lambda_{\min}\left(\sum_{k=0}^{t}\Phi(k)\Phi^T(k)\right) \ge \sum_{k=0}^{t}\lambda_{\min}\left(\Phi(k)\Phi^T(k)\right) > \sum_{k=0}^{t}\frac{0.02}{k+1},$$

we have

$$\frac{\log\left(1+\sum_{k=0}^{t}\|\Phi(k)\|^2\right)}{\lambda_{\min}\left(\mathbf{P}^{-1}(0)+\sum_{k=0}^{t}\Phi(k)\Phi^T(k)\right)} < \frac{\log\left(1+\sum_{k=0}^{t}\frac{125}{k+1}\right)}{\sum_{k=0}^{t}\frac{0.02}{k+1}} \to 0, \quad \text{as } t\to\infty,$$

which validates condition (3.7).

With initial value selected as $\hat{\boldsymbol{\theta}}(0) = [0.01, 0.01, 0.01, 0.01, 0.01]^T$ and $P(0) = \mathbf{I}_5$, we perform 500 iterative steps to estimate the parameters. The resulting relative errors and average accumulative regrets are shown in Figure 5.4.



(a) Relative error (b) Regret

Fig. 5.4. The relative errors and the averaged accumulative regrets against time k. Here, the PE condition is not satisfied.

Clearly, the relative errors ϵ and the averaged accumulative regrets approach zero even if the PE/SE condition is not satisfied, which fully demonstrates that our theory is effective and the PE/SE condition is not necessary.

6. CONCLUSION

For a stochastic regression model, we studied the RLS algorithm to identify the unknown parameter vector for multivariable systems. We established the almost sure convergence results of the RLS algorithm under the weakest possible condition. The accumulated regret analysis is also provided. Simulations are included, which illustrate the convergence results. Some interesting problems deserve to be further investigated, e. g., the design of the distributed algorithm to estimate the unknown parameter using local measurement.

(Received May 9, 2022)

REFERENCES

[1] H.-F. Chen and L. Guo: Strong consistency of recursive identification by no use of persistent excitation condition. Acta Math. Appl. Sinica *2* (1985), 2, 133–145. DOI:10.1007/BF01539484

[2] H.-F. Chen and L. Guo: Continuous-time stochastic adaptive trackingrobustness and asymptotic properties. SIAM J. Control Optim. *28* (1990), 3, 513–527. DOI:10.1137/0328030

[3] H.-F. Chen and L. Guo: Identification and Stochastic Adaptive Control. Volume 5. Springer Science Business Media, 1991.

[4] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum: Feedback Control Theory. Courier Corporation, 2013.

[5] R. Durrett: Probability: Theory and Examples. Vol. 49. Cambridge University Press, 2019.

[6] D. Gan and Z. X. Liu: On the stability of Kalman filter with random coefficients. IFAC-PapersOnLine *53* (2020), 2, 2397–2402. DOI:10.1016/j.ifacol.2020.12.2514

[7] L. Guo: Stability of recursive stochastic tracking algorithms. SIAM J. Control Optim. *32* (1994), 5, 1195–1225, 1994. DOI:10.1137/S0363012992225606

[8] L. Guo: Convergence and logarithm laws of self-tuning regulators. Automatica *31* (1995), 3, 435–450. DOI:10.1016/0005-1098(94)00127-5

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

[10] L. Guo, L. Ljung, and P. Priouret: Performance analysis of the forgetting factor rls algorithm. Int. J. Adaptive Control Signal Process. *7* (1993), 6, 525–537. DOI:doi.org/10.1002/acs.4480070604

[11] S. Haykin: Radar signal processing. IEEE Signal Process. Magazine *2* (1993), 2, 2–18.

[12] T. L. Lai: Asymptotically efficient adaptive control in stochastic regression models. Advances Appl. Math. *7* (1986), 1, 23–45.

[13] T. L. Lai and Ch. Z. Wei: Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Ann. Statist. *10* (1982), 1, 154–166. DOI:10.1214/aos/1176345697

[14] J. Li, F. Ding, and G. Yang: Maximum likelihood least squares identification method for input nonlinear finite impulse response moving average systems. Math. Comput. Modell. *55* (2012), 3–4, :442–450.

[15] J. Li and P. Stoica: MIMO Radar Signal Processing. John Wiley and Sons, 2008.

[16] Y. Liu and F. Ding: Convergence properties of the least squares estimation algorithm for multivariable systems. Applied Math. Modell. *37* (2013), 1–2, 476–483. DOI:10.1016/j.apm.2012.03.007

[17] D. Marelli and M. Fu: A continuous-time linear system identification method for slowly sampled data. IEEE Trans. Signal Process. *58* (2010), 5, 2521–2533. DOI:10.1109/TSP.2009.2040017

[18] J. B. Moore: On strong consistency of least squares identification algorithms. Automatica *14* (1978), 5, 505–509. DOI:10.1016/0005-1098(78)90010-9

[19] M. Niedzwiecki and L. Guo: Nonasymptotic results for finite-memory wls filters. In: Proc. 28th IEEE Conference on Decision and Control, Vol. 2, 1989, pp. 1785–1790.

[20] F. S. G. Richards: A method of maximum-likelihood estimation. J. Roy. Statist. Soc,: Series B (Methodological) *23* (1961), 2, 469–475. DOI:10.1111/j.2517-6161.1961.tb00430.x

[21] A. Sen and N. K. Sinha: On-line estimation of the parameters of a multivariable system using matrix pseudo-inverse. Int. J. Systems Sci. *7* (1976), 4, 461–471. DOI:10.1080/00207727608941930

[22] B. Subudhi and D. Jena: Nonlinear system identification of a twin rotor mimo system. In: IEEE Region 10 Conference 2009, pp. 1–6.

[23] M. Vaezi and A. Izadian: Piecewise affine system identification of a hydraulic wind power transfer system. IEEE Trans. Control Systems Technol. *23* (2015), 6, :2077–2086. DOI:10.1109/TCST.2015.2398311

[24] W. Wang, F. Ding, and J. Dai: Maximum likelihood least squares identification for systems with autoregressive moving average noise. Appl. Math. Modell. *36* (2012), 5, 1842–1853. DOI:10.1016/j.apm.2011.07.083

[25] T. P. Vogels, K. Rajan, and L. F. Abbott: Neural network dynamics. Ann. Rev. Neurosci. *28* (2005), 357–376. DOI:10.1146/annurev.neuro.28.061604.135637

[26] B. Widrow and S. D. Stearns: Adaptive Signal Processing. Prentice-Hall Englewood Cliffs, NJ 1985.

[27] Y. Zhang: Unbiased identification of a class of multi-input single-output systems with correlated disturbances using bias compensation methods. Math. Computer Modell. *53* (2011), 9–10, 1810–1819. DOI:10.1016/j.mcm.2010.12.059

[28] Y. Zhang and G. Cui: Bias compensation methods for stochastic systems with colored noise. Appl. Math- Modell. *35* (2011), 4, 1709–1716. DOI:10.1016/j.apm.2010.10.003

Ziming Wang, NCMIS, LSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190. P. R. China.
e-mail: wangziming@lsec.cc.ac.cn

Yiming Xing, Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, Illinois 61822. U. S. A.
e-mail: yimingx4@illinois.edu

Xinghua Zhu, Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190. P. R. China.
e-mail: zxh@amss.ac.cn