Proceedings of the 40th Chinese Control Conference
July 26-28, 2021, Shanghai, China

On the Convergence of Least Squares Estimator for Nonlinear Autoregressive Models*

Zhaobo Liu[1], Chanying Li[2,3]

1. College of Computer Science and Software Engineering of Shenzhen University, Shenzhen 518060, P. R. China
E-mail: liuzhaobo@szu.edu.cn

2. Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P. R. China
E-mail: cyli@amss.ac.cn

3. School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, P. R. China

Abstract: This paper is concerned with the least squares estimator for a basic class of nonlinear autoregressive models, whose outputs are not necessarily to be ergodic. Several asymptotic properties of the least squares estimator have been established under mild conditions. These properties suggest the strong consistency of the least squares estimates in nonlinear autoregressive models which are not divergent.

Key Words: Nonlinear Autoregressive Models, Strong Consistency, Least Squares, Harris Recurrent

1 Introduction

System identification is a mature field that has gained a notable attraction in the research community ([1],[4],[12],[13],[15]), however our knowledge of parameter estimation for nonlinear autoregressive models is still developing (e.g. [2],[9],[16]). Consider the following general nonlinear autoregressive (AR) model,

$$y_{t+1} = \theta^\tau \phi(y_t, \dots, y_{t-n+1}) + w_{t+1}, \quad t \ge 0, \tag{1}$$

where θ is the $m \times 1$ unknown parameter vector, y_t, w_t are the scalar observations and random noise signals, and $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is a known Lebesgue measurable vector function, respectively. In the literature, asymptotic theory has been established for estimating nonlinear AR models based on an important assumption that the underlying systems satisfy certain ergodic or stationary conditions. Since the least squares (LS) estimator is one of the most basic and effective methods in parameter estimation, the goal of this article is to estimate parameter θ in model (1) by using LS estimator, no matter the outputs are ergodic or not.

Recently, [11] established some asymptotic properties of the LS estimator for model (1) by restricting ϕ as

$$\phi(z_1, \dots, z_n) = \mathrm{col}\{\phi^{(1)}(z_1), \dots, \phi^{(n)}(z_n)\}, \tag{2}$$

where $\phi^{(i)} = (f_{i1}, \dots, f_{im_i})^\tau : \mathbb{R} \to \mathbb{R}^{m_i}, i = 1, \dots, n$ are some known Lebesgue measurable vector functions and $m_i \ge 1$ are n integers satisfying $\sum_{i=1}^n m_i = m$. [11] shows that the LS estimates converge to the true parameter almost surely on the set where vectors $(y_t, \dots, y_{t-n+1})^\tau$ do not diverge to infinity. This means the LS estimator is very likely to be strong consistency when applied to model (1)–(2) in practice owing to that most real systems are not divergent. It is worth nothing that [7] has completely solved the convergence problem of the LS estimator for the linear AR model

$$y_{t+1} = \sum_{i=1}^n \theta_i y_{t-i+1} + w_{t+1},$$

which is a special case of model (1)–(2).

As to nonlinear model (1) without restriction (2), we naturally wonder if the LS estimator still has a similar asymptotic behavior. Indeed, there are some situations that (2) does not hold, such as exponential autoregressive model (EAR) or threshold autoregressive model (TAR) (see [2],[3]). Now, define

$$P_{t+1}^{-1} = I_m + \sum_{i=0}^t \phi(y_t, \dots, y_{t-n+1})\phi(y_t, \dots, y_{t-n+1}),$$

which is refered to as "information matrix". The key of studying the strong consistency of the LS for model (1) is to estimate the growth rate of the minimal eigenvalue of P_{t+1}^{-1} ([5],[6],[8],[14]). In Section 2, under some mild assumptions, the minimal eigenvalue of P_{t+1}^{-1} are estimated in both the Gaussian case and non-Gaussian cases, respectively. The proofs of the main results are included in Section 3.

2 Main Results

We shall discuss the parameter estimation of model (1) by treating θ as a random variable and a constant vector, respectively. Firstly, consider model (1) under the Gaussian case. Assume

A1 The noise $\{w_t\}$ is an i.i.d random sequence with $w_1 \sim N(0, 1)$ and parameter $\theta \sim N(\theta_0, I_m)$ is independent of $\{w_t\}$.

A2 There is a bounded open set $E \subset \mathbb{R}^n$ and a number $\delta^* > 0$ such that
(i) $f_i \in C(\mathbb{R}^n)$, $1 \le i \le m$;
(ii) for every unit vector $x \in \mathbb{R}^n$,

$$J\left(\{y \in \overline{E} : |\phi^\tau(y)x| = \delta^*\}\right) = 0, \tag{3}$$

and

$$\inf_{\|x\|=1} \ell\left(\{y \in E : |\phi^\tau(y)x| > \delta^*\}\right) > 0, \tag{4}$$

*This work is supported by the National Natural Science Foundation (NNSF) of China under Grant 11925109 and the National Key R&D Program of China (2018YFA0703800).
*Corresponding author: Chanying Li.

where $J(\cdot)$ and $\ell(\cdot)$ denote the Jordan measure and Lebesgue measure, respectively.

Recall that the LS estimate $\hat{\theta}_t$ for parameter θ is recursively defined by

$$\left\{\begin{array}{l}\hat{\theta}_{t+1}=\hat{\theta}_t+P_{t+1}\phi_t(y_{t+1}-\phi_t^\tau\hat{\theta}_t)\\ P_{t+1}=P_t-(1+\phi_t^\tau P_t\phi_t)^{-1}P_t\phi_t\phi_t^\tau P_t,\ \ P_0=I_m\\ \phi_t=\phi(y_t,\ldots,y_{t-n+1}),\ \ t\geq 0\end{array}\right.,\quad (5)$$

where $\hat{\theta}_0$ is the deterministic initial condition of the algorithm and ϕ_0 the random initial vector of system (1). Clearly, by (1) and (5),

$$P_{t+1}^{-1}=I_m+\sum_{i=0}^{t}\phi_i\phi_i^\tau. \quad (6)$$

Denote $\lambda_{\min}(t+1)$ as the minimal eigenvalue of P_{t+1}^{-1} in (6) and let

$$N_t(M)\triangleq\sum_{i=1}^{t}I_{\{\|Y_i\|\leq M\}}, \quad (7)$$

where $Y_t\triangleq(y_{t+n-1},\ldots,y_t)^\tau$ and $M>0$ is a constant. Since in the Gaussian case, [14] showed that

$$\left\{\lim_{t\to+\infty}\lambda_{\min}(t+1)=+\infty\right\}=\left\{\lim_{t\to+\infty}\hat{\theta}_t=\theta\right\}, \quad (8)$$

we now estimate $\lambda_{\min}(t+1)$ in the following theorem.

Theorem 1. *Under Assumptions A1–A2, for any constant $M>0$,*

$$\liminf_{t\to+\infty}\frac{\lambda_{\min}(t+1)}{N_t(M)}>0\quad \textit{a.s. on } \Omega(M), \quad (9)$$

where $\Omega(M)\triangleq\{\lim_{t\to+\infty}N_t(M)=+\infty\}$.

Corollary 1. *Let Assumptions A1–A2 hold. Then,*

$$\lim_{t\to+\infty}\hat{\theta}_t=\theta\quad \textit{a.s. on } \left\{\liminf_{t\to+\infty}\|Y_t\|<+\infty\right\}. \quad (10)$$

Next, let θ be a non-random parameter and assume

A1' $\{w_t\}$ is an i.i.d random sequence with $Ew_1=0$ and $E|w_1|^\beta<+\infty$ for some $\beta>2$. Moreover, w_1 has a density $\rho(x)$ such that for every proper interval $I\subset\mathbb{R}$,

$$\inf_{x\in I}\rho(x)>0\quad\text{and}\quad\sup_{x\in\mathbb{R}}\rho(x)<+\infty.$$

We construct the LS estimator from partial data in this case. That is, for some constant $C_\phi>0$, ϕ_t in (5) is modified as

$$\phi_t\triangleq I_{\{\|Y_{t-n+1}\|\leq C_\phi\}}\phi(y_t,\ldots,y_{t-n+1}).$$

Here, we remark that by [6, Theorem 1] and [5, Lemma 3.1],

$$\|\hat{\theta}_{t+1}-\theta\|^2=O\left(\frac{\log\left(\lambda_{\max}(t+1)\right)}{\lambda_{\min}(t+1)}\right),\quad\text{a.s.,} \quad (11)$$

where $\lambda_{\max}(t+1)$ denotes the maximal eigenvalue of P_{t+1}^{-1}. Analogous to Theorem 1, we deduce

Theorem 2. *Under Assumptions A1' and A2, there is a constant $M_\phi>0$ depending only on ϕ such that for any $C_\phi>M_\phi$ and $M>0$,*

$$\liminf_{t\to+\infty}\frac{\lambda_{\min}(t+1)}{N_t(M)}>0\quad \textit{a.s. on } \Omega(M).$$

In addition, if $M\geq C_\phi$, then

$$\|\hat{\theta}_t-\theta\|^2=O\left(\frac{\log N_t(M)}{N_t(M)}\right)\quad \textit{a.s. on } \Omega(M).$$

Finally, we propose another assumption that is stronger than Assumption A2 but easier to check. To this end, we introduce a series of operators. Firstly, rewrite

$$\phi(z)=\mathrm{col}\{f_1(z),\ldots,f_m(z)\},$$

where $z=(z_1,\ldots,z_n)^\tau$ and $f_i:\mathbb{R}^n\to\mathbb{R}, i=1,\ldots,n$ are some known Lebesgue measurable vector functions. Denote D_u as the partial differential operator regarding to variable u, then for any sufficiently smooth functions $\{g_l\}_{l\geq 1}:\mathbb{R}^n\to\mathbb{R}$, and variables u_i, $i\geq 1$, recursively define $\Lambda_1^{(0)}(g_1)\triangleq g_1$ and

$$\begin{aligned}&\Lambda_{l+1}^{(u_1,\ldots,u_l)}(g_1,\cdots,g_{l+1})\\ \triangleq\ &\Lambda_l^{(u_2,\ldots,u_l)}\left(\frac{D_{u_1}g_1}{D_{u_1}g_{l+1}},\cdots,\frac{D_{u_1}g_l}{D_{u_1}g_{l+1}}\right),\quad l\geq 1.\end{aligned}$$

Now, let $l_1<\cdots<l_s$ be s positive integers. For each $k\in[1,s]$, denote $\mathcal{H}_k^{(l_1,\ldots,l_s)}$ as the k-permutations of $\{l_1,\ldots,l_s\}$. Moreover, for variable set $Z=\{z_1,\ldots,z_n\}$, denote $Z^0=\{(0)\}$,

$$Z^k\triangleq\{(z_{i_1},\ldots,z_{i_k}):z_{i_j}\in Z,1\leq j\leq k\},\quad k\geq 1.$$

For $(i_1,\ldots,i_k)\in\mathcal{H}_k^{(1,\ldots,m)}$, $b\in Z^{k-1}$, $k\in[1,m]$, define

$$\Gamma_{(i_1,\ldots,i_k)}^{(b)}\triangleq\Lambda_k^{(b)}\left(f_{i_1},\ldots,f_{i_k}\right). \quad (12)$$

In addition, let $b'\in Z^{k-1}$ and $z_i\in Z$, then define

$$\bar{\Gamma}_s^{(b)}\triangleq D_{z_i}\Gamma_{(i_1,\ldots,i_k)}^{(b')}\quad\text{with}\quad b=\mathrm{col}\{b',z_i\}=\begin{pmatrix}b'\\ z_i\end{pmatrix}.$$

Given function g, denote $A(g)\triangleq\{x:g(x)=0\}$. We classify the sets $A(\bar{\Gamma}_s^{(b)}), s=(i_1,\ldots,i_k)\in\mathcal{H}, b\in Z^k$ into three types:

$$\left\{\begin{array}{l}Z_s^1(b)=\mathrm{int}(A(\bar{\Gamma}_s^{(b)}))\\ Z_s^2(b)=d(A(\bar{\Gamma}_s^{(b)}))\backslash Z_s^1(b)\\ Z_s^3(b)=A(\bar{\Gamma}_s^{(b)})\backslash d(A(\bar{\Gamma}_s^{(b)}))\end{array}\right., \quad (13)$$

where $d(A)$ denotes the derived set of A. Based on above notations, we provide our assumption as follows.

A2' There is an open sets E belonging to $\mathbb{R}^n$ such that
(i) $f_i\in C(\mathbb{R}^n)$ and $f_i\in C^m(E)$, $1\leq i\leq n$;
(ii) for every unit vector $x\in\mathbb{R}^m$, there is a point $y\in E$ such that $|\phi^\tau(y)x|\neq 0$;
(iii) for any $j\in[1,m]$, $s\in\mathcal{H}_j^{(1,\ldots,m)}$, $b\in Z^j$,

$$Z_s^1(b)\cap E=\emptyset\text{ or }E. \quad (14)$$

We assert that

Theorem 3. *Theorems 1 and 2 hold for model (1) if Assumption A2 is replaced by A2'.*

3 Proof of Theorems

Apparently, Theorems 1–2 are the direct results of following proposition, whose proof is contained in Appendix B.

Proposition 1. *Under Assumptions A1' and A2, let θ be a random variable independent of $\{w_t\}_{t\geq 1}$.Then, there is a constant $M_\phi > 0$ depending only on ϕ such that for any $C_\phi > M_\phi$ and $M, K > 0$,*

$$\liminf_{t\to+\infty} \frac{\lambda_{\min}(t+1)}{N_t(M)} > 0 \quad \textit{a.s. on } \Omega(M) \cap \{\|\theta\| \leq K\}.$$

To prove Theorem 3, it suffices to show Assumption A2' is stronger than A2. Denote

$$S(\alpha) = \{y \in E : \mathrm{dist}(y, \partial(E)) > \alpha\}, \quad \alpha > 0. \tag{15}$$

Clearly, $\lim_{\alpha\to 0} S(\alpha) = E$. It is not hard to prove that there is a point $\alpha > 0$ such that for any unit $x \in \mathbb{R}^m$,

$$\ell\left(\{y \in S(\alpha) : |\phi^\tau(y)x| > 0\}\right) > 0 \tag{16}$$

(Readers can also refer to [11, Lemmas 3.1 and 3.4]). Given (16), the following lemma is natural.

Lemma 1. *If (16) holds, then there exist some open boxes $S_1(\alpha), \ldots, S_q(\alpha)$ such that for any unit $x \in \mathbb{R}^m$,*

$$\ell\left(\{y \in \mathcal{S} : |\phi^\tau(y)x| > 0\}\right) > 0, \tag{17}$$

where $\mathcal{S} \triangleq \bigcup_{j=1}^q S_j(\alpha)$ and $\overline{S(\alpha)} \subset \overline{\mathcal{S}} \subset E$.

Proof. For each point $y \in \overline{S(\alpha)}$, there is an open box $\chi_y \subset S$ containing y, then Lemma 1 follows by *Heine-Borel theorem.* □

Lemma 2. *Under Assumption A2', let $\mathcal{S}$ be the set defined in Lemma 1. Then, for any non-zero $x = (x_1, \ldots, x_m)^\tau \in \mathbb{R}^m$ and $\delta \in \mathbb{R}$,*

$$J\left(\{y \in \overline{\mathcal{S}} : |\phi^\tau(y)x| = \delta\}\right) = 0, \tag{18}$$

or

$$\{y \in \overline{\mathcal{S}} : |\phi^\tau(y)x| = \delta\} = \overline{\mathcal{S}}. \tag{19}$$

Proof. Since $\overline{\mathcal{S}} \subset E$, for any $j \in [1, m]$, $s \in \mathcal{H}_j^{(1,\ldots,m)}$, $b \in Z^j$, we have

$$Z_s^1(b) \cap \overline{\mathcal{S}} = \emptyset \text{ or } \overline{\mathcal{S}}. \tag{20}$$

For $s = (i_1, \ldots, i_m) \in \mathcal{H}_m^{(1,2,\ldots,m)}$, $k \in [1, m]$, $b \in Z^{m-k+1}$, denote

$$\varrho_{k,s}^{(b)}(y) = \sum_{j=1}^{k} \bar{\Gamma}^{(b)}_{(i_j, i_{k+1}, \ldots, i_m)}(y) x_{i_j}.$$

By using an induction method, we shall show that for any $k \in [1, m]$, $b \in Z^{m-k+1}$, $s = (i_1, \ldots, i_m) \in \mathcal{H}_m^{(1,2,\ldots,m)}$,

$$J\left(\{y \in \overline{\mathcal{S}} : \varrho_{k,s}^{(b)}(y) = 0\}\right) = 0, \tag{21}$$

or

$$\{y \in \overline{\mathcal{S}} : \varrho_{k,s}^{(b)}(y) = 0\} = \overline{\mathcal{S}}.$$

In fact, for $k = 1$, $s = (i_1, \ldots, i_m) \in \mathcal{H}_m^{(1,2,\ldots,m)}$ and $b \in Z^m$, if $x_{i_1} = 0$, we have $\varrho_{1,s}^{(b)}(y) \equiv 0$ on $\overline{\mathcal{S}}$. Otherwise,

$$\begin{aligned} &\{y \in \overline{\mathcal{S}} : \varrho_{1,s}^{(b)}(y) = 0\} = \overline{\mathcal{S}} \cap A(\bar{\Gamma}^{(b)}_{(i_1,\ldots,i_m)}) \\ = \;& \overline{\mathcal{S}} \cap (Z^1_{(i_1,\ldots,i_m)}(b) \cup Z^3_{(i_1,\ldots,i_m)}(b)) \\ = \;& \overline{\mathcal{S}} \text{ or } \overline{\mathcal{S}} \cap (Z^3_{(i_1,\ldots,i_m)}(b)). \end{aligned}$$

Note that $Z^3_{(i_1,\ldots,i_m)}(b) \cap \overline{\mathcal{S}}$ then is a finite set, its Jordan measure must be zero. The assertion thus follows for $k = 1$. Now, assume that for some $h > 1$, this assertion still holds when $k \in [1, h)$. Next, consider $k = h$. For any $b^* = (b_1, \ldots, b_{m-h+1}) \in Z^{m-h+1}$, $s = (i_1, \ldots, i_m) \in \mathcal{H}_m^{(1,2,\ldots,m)}$, if for any $j \in [1, h]$,

$$J(A(\bar{\Gamma}^{(b^*)}_{(i_j, i_{h+1}, \ldots, i_m)})) > 0,$$

then

$$\overline{\mathcal{S}} \subset A(\bar{\Gamma}^{(b^*)}_{(i_j, i_{h+1}, \ldots, i_m)}) \text{ and } \{y \in \overline{\mathcal{S}} : \varrho_{h,s}^{(b^*)}(y) = 0\} = \overline{\mathcal{S}}.$$

We further assume that there is an index $j^* \in [1, h]$ such that

$$J(A(\bar{\Gamma}^{(b^*)}_{(i_{j^*}, i_{h+1}, \ldots, i_m)})) = 0.$$

let

$$s' = (i_1, \ldots, i_{j^*-1}, i_{j^*+1}, \ldots, i_h, i_{j^*}, i_{h+1}, \ldots, i_m).$$

If for all $b \in b^* \times Z$,

$$J(\{y \in \overline{\mathcal{S}} : \varrho_{h-1,s'}^{(b)}(y) = 0\}) \neq 0,$$

then by our assumption,

$$\bigcap_{b \in b^* \times Z} \{y \in \overline{\mathcal{S}} : \varrho_{h-1,s'}^{(b)}(y) = 0\} = \overline{\mathcal{S}}, \tag{22}$$

and hence

$$\begin{aligned} &\{y \in \overline{\mathcal{S}} : \varrho_{h,s}^{(b^*)}(y) = \delta\} \\ = \;& \{y \in \overline{\mathcal{S}} : \sum_{j \in [1,h] \setminus \{j^*} \Gamma^{(b^*)}_{(i_j, i_{j^*}, i_{h+1}, \ldots, i_m)}(y) x_{i_j} \\ & = -x_{i_{j^*}}\} = \overline{\mathcal{S}}. \end{aligned}$$

So it suffices to consider the set

$$\begin{aligned} T_{b^*} \;=\; & \{b \in b^* \times Z : J(\{y \in \overline{\mathcal{S}} : \varrho_{h-1,s'}^{(b)}(y) = 0\}) = 0\} \\ \neq \; & \emptyset. \end{aligned}$$

Fix an $\varepsilon > 0$. There are finite some disjoint open boxes $O(1), \ldots, O(N)$ covering

$$A(\bar{\Gamma}^{(b^*)}_{(i_{j^*}, i_{h+1}, \ldots, i_m)}) \cup \bigcup_{b \in T_{b^*}} \{y \in \overline{\mathcal{S}} : \varrho_{h-1,s'}^{(b)}(y) = 0\},$$

and satisfying $\sum_{l=1}^N O(l) < \varepsilon$. Note that

$$\begin{aligned} &\{y \in \overline{\mathcal{S}} : \varrho_{h,s}^{(b^*)}(y) = 0\} \\ \subset \;& \{y \in \overline{\mathcal{S}} \setminus \bigcup_{l=1}^N O(l) : \varrho_{h,s}^{(b^*)}(y) = 0\} \cup \bigcup_{l=1}^N O(l) \\ = \;& \{y \in \overline{\mathcal{S}} \setminus \bigcup_{l=1}^N O(l) : \sum_{j \in [1,h] \setminus \{j^*\}} \\ & \Gamma^{(b^*)}_{(i_j, i_{j^*}, i_{h+1}, \ldots, i_m)}(y) x_{i_j} = -x_{i_{j^*}}\} \cup \bigcup_{l=1}^N O(l), \end{aligned} \tag{23}$$

we then divide set

$$\{y \in \overline{\mathcal{S}} \setminus \bigcup_{l=1}^{N} O(l) : \sum_{j\in[1,h]\setminus\{j^*\}} \Gamma^{(b^*)}_{(i_j,i_{j^*},i_{h+1},\ldots,i_m)}(y)x_{i_j} = -x_{i_{j^*}}\}$$

into some connected components. In words,

$$\{y \in \overline{\mathcal{S}} \setminus \bigcup_{l=1}^{N} O(l) : \sum_{j\in[1,h]\setminus\{j^*\}} \Gamma^{(b^*)}_{(i_j,i_{j^*},i_{h+1},\ldots,i_m)}(y)x_{i_j} = -x_{i_{j^*}}\} = \bigcup_{a\in A} X_a,$$

where A is some index set and each connected component X_a is closed. By our assumption, for any $b \in b^* \times Z \setminus T_{b^*}$,

$$\{y \in \overline{\mathcal{S}} : \varrho^{(b)}_{h-1,s'}(y) = 0\} = \overline{\mathcal{S}}.$$

Without loss of generality, suppose

$$T_{b^*} = \{(b^*, z_i) : i = 1, \ldots, |T_{b^*}|\},$$

then X_a has the form $X_a = X'_a \times \{v \in \mathbb{R}^{n-|T_{b^*}|} : (u, v) \in \overline{\mathcal{S}} \setminus \bigcup_{l=1}^{N} O(l)$ for some $u \in \mathbb{R}^{|T_{b^*}|}\}$. By the theorem of implicit function and the Riemann's integrability of continuous function, we know that for each fixed $a \in A$, surface X'_a has zero Jordan measure in $\mathbb{R}^{|T_{b^*}|}$, and thus $J(X_a) = 0$. Moreover, for a box $O = \prod_{i=1}^{n} I_i$, let

$$E(O) = \bigcup_{l=1}^{n} \prod_{i=1}^{l} \partial(I_i) \times I_l \times \prod_{i=l+1}^{n} \partial(I_i)$$

be the set of the edges of O. Denote

$$\{S_j(\alpha) : j = 1, \ldots, q\} \cup \{O(l) : l = 1, \ldots, N\} \triangleq \mathcal{O},$$

and

$$\mathcal{O}_0 = \{O \in \mathcal{O} : \partial(O) \cap X_a \neq \emptyset\}, \tag{24}$$

it is then easy to see $\mathcal{O}_0 \neq \emptyset$. Fix $O \in \mathcal{O}_0$ and suppose

$$\inf_{e\in E(O)} \operatorname{dist}(X_a \cap \partial(O), e) > 0. \tag{25}$$

Let

$$e^* = \arg \min_{e\in E(O)} \operatorname{dist}(X_a \cap \partial(O), e)$$

and

$$y^* = \arg \min_{y\in X_a\cap\partial(O)} \operatorname{dist}(y, e^*).$$

Now, write $O = \prod_{i=1}^{n}[c_i, d_i]$, $e^* = \prod_{i=1}^{l-1}\{e_i\} \times [c_l, d_l] \times \prod_{i=l+1}^{n} e_i,\ e_i \in \{c_i, d_i\},\ i \neq l$, and $y^* = (y_1, \ldots, y_n)^\tau$.

If $y_l \in (c_i, d_i)$ and $(b^*, z_l) \in T_{b^*}$, then by the theorem of implicit function and the definition of X_a, there is a $y' \in X_a \cap \partial(O)$ such that $\operatorname{dist}(y', e^*) < \operatorname{dist}(y, e^*)$. This is a contradiction. Hence, $y_l \in \{c_l, d_l\}$ or $(b^*, z_l) \notin T_{b^*}$. Note that $(b^*, z_l) \notin T_{b^*}$ infers that

$$(y_1, \ldots, y_{l-1}, c_l, y_{l+1}, \ldots, y_n)^\tau \in X_a,$$

and the distance between $(y_1, \ldots, y_{l-1}, c_l, y_{l+1}, \ldots, y_n)^\tau$ and e^* is equal to $\operatorname{dist}(y, e^*)$, so we can assume $y_l \in \{c_l, d_l\}$ directly. Since $\operatorname{dist}(y, e^*) > 0$, there is a $l' \neq l$ such that $e_{l'} \neq y_{l'}$. Without loss of generality, suppose $l > l'$ and consider

$$e' = \prod_{i=1}^{l'-1} e_i \times [c_{l'}, d_{l'}] \times \prod_{i=l'+1}^{l-1} e_i \times y_l \times \prod_{i=l+1}^{n} e_i.$$

As a result,

$$\begin{aligned}(\operatorname{dist}(y, e'))^2 &= \sum_{i\neq l',l} (e_i - y_i)^2 \\ &< \sum_{i\neq l} (e_i - y_i)^2 = (\operatorname{dist}(y, e^*))^2,\end{aligned}$$

which contradicts to the definition of e^*. So, for any $O \in \mathcal{O}_0$,

$$\inf_{e\in E(O)} \operatorname{dist}(X_a \cap \partial(O), e) = 0. \tag{26}$$

This implies there is an edge e of O such that $X_a \cap e \neq \emptyset$.

If $|A| = +\infty$, then there is a $O = \prod_{i=1}^{n}[c_i, d_i] \in \mathcal{O}_0$ and $e \in E(O)$ such that there are infinite many $a \in A$ satisfying $X_a \cap e \neq \emptyset$. Let $e = \prod_{i=1}^{l-1}\{e_i\} \times [c_l, d_l] \times \prod_{i=l+1}^{n} e_i,\ e_i \in \{c_i, d_i\},\ i \neq l$, and given $g \in [c_l, d_l]$ and $\varepsilon > 0$, denote

$$\mathbf{G}(g, \varepsilon) = \prod_{i=1}^{l-1}\{e_i\} \times [g - \varepsilon, g + \varepsilon] \times \prod_{i=l+1}^{n} e_i. \tag{27}$$

Then, there is a $g^* \in [c_l, d_l]$ such that for any $\zeta > 0$,

$$|\{y \in \overline{\mathcal{S}} : \sum_{j\in[1,h]\setminus\{j^*\}} \Gamma^{(b^*)}_{(i_j,i_{j^*},i_{h+1},\ldots,i_m)}(y)x_{i_j} = -x_{i_{j^*}}\} \cap \mathbf{G}(g^*, \zeta)| = +\infty.$$

By *Rolle's theorem*,

$$|\{y \in \overline{\mathcal{S}} : \sum_{j\in[1,h]\setminus\{j^*\}} D_{z_l}\Gamma^{(b^*)}_{(i_j,i_{j^*},i_{h+1},\ldots,i_m)}(y)x_{i_j} = 0\} \cap \mathbf{G}(g^*, \zeta)| = +\infty.$$

Note that $Z^2_s(b) \cap e = \emptyset$ for all $s \in \mathcal{H}^{(1,\ldots,m)}_j$, $b \in Z^j$, $j = 1, \ldots, m$, then there is an η such that

$$\mathbf{G}(g^*, \eta) \cap \bigcup_{j=1}^{m} \bigcup_{b\in Z^j} \bigcup_{s\in\mathcal{H}^{(1,\ldots,m)}_j} Z^3_s(b) = \emptyset,$$

and hence for all $s \in \mathcal{H}^{(1,\ldots,m)}_j$, $b \in Z^j$, $j = 1, \ldots, m$,

$$\begin{aligned}A(\bar{\Gamma}^{(b)}_s) \cap \mathbf{G}(g^*, \eta) &= Z^1_s(b) \cap \mathbf{G}(g^*, \eta) \\ &= \emptyset \text{ or } \mathbf{G}(g^*, \eta).\end{aligned}$$

By applying Lemma 3(i) in Appendix A, $\mathbf{G}(g^*, \eta) \subset X_a$ for some $a \in A$, which yields a contradiction.

So, $|A| < +\infty$. Observe that $J(X_a) = 0$, $a \in A$, then we obtain $J(\bigcup_{a\in A} X_a) = 0$. It follows that there exist some finite boxes with volume less than ε, covering

$$\{y \in \overline{\mathcal{S}} \setminus \bigcup_{l=1}^{N} O(l) : \sum_{j\in[1,h]\setminus\{j^*\}} \Gamma^{(b^*)}_{(i_j,i_{j^*},i_{h+1},\ldots,i_m)}(y)x_{i_j} = -x_{i_{j^*}}\}.$$

This, together with (23) and the arbitrariness of ε, leads to

$$J(\{y \in \overline{\mathcal{S}} : \varrho_{h,s}^{(b^*)}(y) = 0\}) = 0.$$

The induction completes.

Now, let $k = m$ in (21). For $i = 1, \ldots, n$, we have

$$J(\{y \in \overline{\mathcal{S}} : \sum_{j=1}^{m} D_{z_i} f_j(y) = 0\}) = 0$$

or

$$\{y \in \overline{\mathcal{S}} : \sum_{j=1}^{m} D_{z_i} f_j(y) = 0\} = \overline{\mathcal{S}}. \tag{28}$$

If for all $i \in [1, n]$, (28) holds, then (19) follows. Otherwise, the set

$$T' = \{i \in [1, n] : J(\{y \in \overline{\mathcal{S}} : \sum_{j=1}^{m} D_{z_i} f_j(y)(y) = 0\}) = 0\}$$

is not empty. For any $\varepsilon > 0$, we can choose some finite disjoint open boxes $O'(1), \ldots, O'(N)$ covering

$$\bigcup_{i \in T'} \{y \in \overline{\mathcal{S}} : \sum_{j=1}^{m} D_{z_i} f_j(y)(y) = 0\}$$

and satisfying $\sum_{l=1}^{N} O'(l) < \varepsilon$. By a similar discussion, we can also show that

$$\{y \in \overline{\mathcal{S}} \setminus \bigcup_{l=1}^{N} O'(l) : \sum_{j=1}^{m} f_j(y) x_j = \delta\} \tag{29}$$

can be divided into finite connected components with zero Jordan measure. This implies (18). $\square$

Now, by the definition of $\mathcal{S}$, for any non-zero $x \in \mathbb{R}^m$,

$$\ell(\{y \in \overline{\mathcal{S}} : |\phi^\tau(y)x| > \delta^*\}) > 0, \tag{30}$$

so $\{y \in \overline{\mathcal{S}} : |\phi^\tau(y)x| > \delta^*\} \neq \overline{\mathcal{S}}$. Thus by Lemma 2, we know

$$J\left(\{y \in \overline{\mathcal{S}} : |\phi^\tau(y)x| = \delta^*\}\right) = 0. \tag{31}$$

Then we deduce Assumption A2 under Assumption A2'.

A Property of Operators $\{\Lambda_l^{(u_1,\ldots,u_{l-1})}\}_{l\geq 1}$

We now show that operators $\{\Lambda_l^{(u_1,\ldots,u_{l-1})}\}_{l\geq 1}$ have the following property:

Lemma 3. *Fix $i' \in [0, m-1]$, $i \in [1, n]$ and vector*

$$(c_1, \ldots, c_n)^\tau \in \mathbb{R}^n.$$

For some open set $E \subset \mathbb{R}^n$, let functions $g_1, \ldots, g_{m-i'} \in C^{m-i'}(E)$. Let $x = (x_1, \ldots, x_m)^\tau \in \mathbb{R}^m$ be a non-zero vector and $d > c$ be two numbers satisfying

$$\Xi_i \triangleq (c_1, \ldots, c_{i-1})^\tau \times [c, d] \times (c_{i+1}, \ldots, c_n)^\tau \subset E.$$

Also, let $\{r_l\}_{l=1}^{2^{m-i'-1}}$ be a sequence of numbers that $d \geq r_1 > r_2 > \ldots > r_{2^{m-i'-1}} \geq c$ and for $1 \leq l \leq 2^{m-i'-1}$,

$$\sum_{j=1}^{m-i'} D_{z_i} g_j(c_1, \ldots, c_{i-1}, r_i, c_{i+1}, \ldots, c_n) x_j = 0.$$

Then, the following two statements hold:
(i) there exist some $j' \geq 1$ and array $s = (i_1, \ldots, i_{j'}) \in \mathcal{H}_{j'}^{(1,2,\ldots,m-i')}$ such that

$$A(D_{z_i} \Lambda_{j'}^{(\{z_i\}^{j'-1})}(g_{i_1}, \ldots, g_{i_{j'}})) \cap \Xi_i \neq \emptyset;$$

(ii) if for every $j' \geq 1$ and $s = (i_1, \ldots, i_{j'}) \in \mathcal{H}_{j'}^{(1,2,\ldots,m-i')}$, set

$$A(D_{z_i} \Lambda_{j'}^{(\{z_i\}^{j'-1})}(g_{i_1}, \ldots, g_{i_{j'}})) \cap \Xi_i$$

is either $\emptyset$ or Ξ_i, then

$$\sum_{j=1}^{m-i'} D_{z_i} g_j(y) x_{ij} = 0, \quad \forall y \in \Xi_i.$$

Proof. The proof of Lemma 3 can be referred to [10] and [11, Lemma 3.7]. $\square$

B Proof of Proposition 1

The proof is similar as that of [10, Proposition 1] but more concise. As a matter of fact, taking δ^* from (4) in Assumption A2, [10, Lemma 1] follows with $\mathcal{S}$ replaced by E. So, for every unit vector $x \in \mathbb{R}^m$, we can directly define

$$U_x \triangleq \{y : |\phi^\tau(y)x| > \delta^*\} \cap E.$$

Next, with random process g_x and set $\mathcal{T}(O, r)$ defined in [10, Appendix A], we proceed to show [10, Lemma 3]. For this, select a box O containing $\overline{E}$ and define

$$\mathcal{T}(x, O, r) \triangleq \{U \in \mathcal{T}(O, r) : \partial U_x \cap U \neq \emptyset\}.$$

The remainder is mainly devoted to proving

$$\lim_{r\to+\infty} \sup_{\|x\|=1} \sum_{U \in \mathcal{T}(x,O,r)} \ell(U) = 0. \tag{32}$$

To show (32), note that

$$\partial(U_x) \subset V_x \triangleq \{y \in \overline{E} : |\phi^\tau(y)x| = \delta^*\}.$$

Denote

$$W(x, r) \triangleq \bigcup_{U \in \mathcal{T}'(x,O,r)} U,$$

where $\mathcal{T}'(x, O, r) \triangleq \{U \in \mathcal{T}(O, r) : V_x \cap U \neq \emptyset\}$. So, it suffices to show

$$\lim_{r\to+\infty} \sup_{\|x\|=1} \ell(W(x, r)) = 0. \tag{33}$$

If (33) is false, then there is a number $\varepsilon > 0$ and a unit vector sequence $\{x(k)\}_{k=1}^{+\infty}$ such that $\lim_{k\to+\infty} x(k) = x^*$ for some unit vector x^* and

$$\ell(W(x(k), 2^k)) > \varepsilon, \quad \forall k \geq 1. \tag{34}$$

Now, according to the definition of the Jordan measure, (3) in Assumption A3(ii) indicates that

$$\lim_{r\to+\infty} \ell(W(x^*, r)) = 0.$$

Moreover, since

$$\lim_{k\to+\infty} \sup_{y\in V_{x(k)}} \mathrm{dist}(y, V_{x^*}) = 0,$$

for any $\varepsilon' > 0$ and all sufficiently large integers k', k with $k' < k$,

$$|\mathcal{T}'(x^*, O, 2^k)| < \frac{\varepsilon' 2^{kn}}{\ell(O)}$$

and

$$|\mathcal{T}'(x(k), O, 2^k)| < (1 + 2^{k-k'+1})^n |\mathcal{T}'(x^*, O, 2^k)|.$$

The above two inequalities immediately lead to

$$\begin{aligned} \ell(W(x(k), 2^k)) &= \frac{\ell(O)}{2^{kn}} \cdot |\mathcal{T}'(x(k), O, 2^k)| \\ &< (1 + 2^{k-k'+1})^n \varepsilon', \end{aligned}$$

which contradicts to (34) by selecting $k' = k - 1$ and $\varepsilon' < 5^{-n}\varepsilon$.

Finally, [11, Equation (3.65)] follows from (32) and hence [10, Lemma 3] holds. The rest of the procedures thus keep the same as those for [10, Proposition 1].

References

[1] K. J. Åström, P. Eykhoff, System identification: A survey, *Automatica*, 7: 123–162, 1970.

[2] K. S. Chan, R. S. Tsay, Limiting properties of the least squares estimator of a continuous threshold autoregressive model, *Biometrika*, 85: 413–426, 1998.

[3] G. Chen, M. Gan, G. Chen, Generalized exponential autoregressive models for nonlinear time series: stationarity, estimation and applications, *Information Sciences*, 438: 46–57, 2018.

[4] G. C. Goodwin, M. Gevers, B. Ninness, Quantifying the error in estimated transfer functions with application to model order selection, *IEEE Transactions on Automatic Control*, 37: 913–929, 1992.

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

[6] T. L. Lai, C. Z. Wei, Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems, *Annals of Statistics*, 10: 154–166, 1982.

[7] T. L. Lai, C. Z. Wei, Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters, *Journal of Multivariate Analysis*, 13: 1–23, 1983.

[8] C. Li, J. Lam, Stabilization of discrete-time nonlinear uncertain systems by feedback based on LS algorithm, *SIAM Journal on Control and Optimization*, 51: 1128–1151, 2013.

[9] D. Li, S Ling, On the least squares estimation of multiple-regime threshold autoregressive models, *Journal of Econometrics*, 167: 240–253, 2012.

[10] Z. Liu, C. Li, Asymptotic Behavior of Least Squares Estimator for Nonlinear Autoregressive Models, *Science China-Information Sciences*, provisionally accepted.

[11] Z. Liu, C. Li, Asymptotic Behavior of Least Squares Estimator for Nonlinear Autoregressive Models. arXiv preprint arXiv:1909.06773, 2019.

[12] L. Ljung, Perspectives on system identification, *Annual Reviews in Control*, 34: 1–12, 2010.

[13] M. Yu, J. Liu, L. Zhao, Nuclear Norm Subspace System Identification and Its Application on a Stochastic Model of Plague, *Journal of Systems Science and Complexity*, 33: 43–60, 2020.

[14] J. Sternby, On consistency for the method of least squares using martingale theory, *IEEE Transactions on Automatic and Control*, 22: 346–352, 1977.

[15] X. M. Wang, M. Hu, Y. Zhao, et al, Credit Scoring Based on the Set-Valued Identification Method, *Journal of Systems Science and Complexity*, 33: 1297–1309, 2020.

[16] W. X. Zhao, H. F. Chen, W. X. Zheng, Recursive identification for nonlinear ARX systems based on stochastic approximation algorithm, *IEEE Transactions on Automatic and Control*, 55: 1287–1299, 2010.