A Local Information Criterion for Order Identification of Nonlinear ARX Systems

Wenxiao Zhao[1,2], Han-Fu Chen[1,2], Er-Wei Bai[3,4], and Kang Li[4]

[1]*Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science,*
[2]*National Center for Mathematics and Interdisciplinary Sciences,*
Chinese Academy of Sciences, Beijing 100190, China.
[3]*Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA.*
[4]*School of Electronics, Electrical Engineering and Computer Science, Queen's University, Belfast, UK.*

***Abstract*— We consider the local order estimation of nonparametric nonlinear autoregressive systems with exogenous inputs (NARX), which may have different local dimensions at different points of interest. By minimizing the kernel-based local information criterion introduced in the paper, the strongly consistent estimates for the local orders of the NARX system at of interest points are obtained. The theoretical result derived here is tested by a simulation example.**

***Index Terms*— Nonlinear ARX system, recursive local linear estimator, order estimation, strong consistency.**

I. Introduction

Consider a single-input single-output (SISO) nonparametric nonlinear autoregressive system with exogenous input (NARX),

$$y_{k+1} = f(y_k, \cdots, y_{k+1-M}, u_k, \cdots, u_{k+1-M}) + \varepsilon_{k+1}, \quad (1)$$

where u_k and y_k are the system input and output, respectively, ε_k is the driven noise, M is the known upper bound of the true system order and $f(\cdot)$ is the unknown function representing the system dynamics.

In recent years identification of system (1) has been an active research topic, estimating not only the nonlinear function $f(\cdot)$ itself ([3], [5], [14], [19], [20], [21], [22]) but also the system orders ([2], [4], [11], [12], [17], [18]). As far as the estimation of the nonlinear function $f(\cdot)$ is of concern, the approaches can roughly be divided into two categories, the parametric approach ([14], [20]) and the nonparametric approach ([3], [5], [19], [21], [22]), according to the description of $f(\cdot)$. In the former, it is usually assumed that $f(\cdot) = f(\cdot, \theta)$ with a known structure of $f(\cdot)$ and unknown parameter θ, and consequently identification of $f(\cdot)$ is transformed into a parametric optimization problem for θ. While in the latter approach, it is often to estimate the values of $f(\cdot)$ at the points of interest referred to as *Model on Demand* in the literature. This paper is on the estimation of the orders in (1), where $f(\cdot)$ is a nonparametric nonlinearity.

It is natural to ask what is the necessity and advantage of order estimation for (1) in a nonparametric setting of $f(\cdot)$? Let us first consider system (1) and assume the function is defined in the domain $\Gamma = \{x = [y(1), \cdots, y(M), u(1), \cdots, u(M)]^T \mid y(i) \in [0,1],\ u(i) \in [0,1],\ i = 1, \cdots, M\}$. By uniformly dividing the interval $[0,1]$ into ten equal subintervals, the domain Γ is equally divided into 10^{2M} small cubes. Assume that the function value at a fixed point in each cube is to be estimated and ten measurements in each cube are adequate to get reliable estimates. Then the total number of the required measurements is up to 10^{2M+1}, which is impossible in practice even for moderate M. This phenomenon is called the *Curse of Dimensionality*. If the actual order $M_0 \ll M$ can somehow be calculated, the total measurements required is 10^{2M_0+1} which is much smaller than 10^{2M+1}. Thus, order estimation of nonlinear dynamic systems is of both theoretical and practical significance.

Over the last few decades considerable progress has been made on the order estimation of linear stochastic systems, for example, the Akaike's information criterion (AIC) [1], Bayesian information criterion (BIC) and their generalizations [8], the recursive algorithms [9], being just a few among many others. But these approaches are not applicable to system (1) due to its nonparametric and nonlinear nature. The order estimation of nonlinear systems has also been studied in recent years, e.g., [2], [6], [11], [13], [15], [17], [18]. In [2] an approach to estimate the orders of the linearized nonlinear system is introduced. The so-called Lipschitz number approach and false nearest neighbors approach are proposed in [11] and [13], respectively, and successive research appeared in [6], [16], [17], etc. These two approaches do not identify the nonlinearity $f(\cdot)$ itself, while estimating the orders. The methods in [2], [11], and [13] are however sensitive to the system noises, and to the authors' knowledge, their convergence and consistency are unclear. The stepwise approach and the analysis of variance (ANOVA) approach are suggested in [15] and [18] based on hypothesis tests, and

Email: wxzhao@amss.ac.cn; hfchen@iss.ac.cn; er-wei-bai@uiowa.edu; k.li@qub.ac.uk.

The research of Wenxiao Zhao was supported by NSF of China under Grants No. 61104052, 61273193, and 91130008. The research of Han-Fu Chen was supported by NSF of China under Grants No. 61273193, 61120106011, and 61134013. The research of Er-Wei Bai was supported in part by NSF, DoE, IEC and NIH grants.

978-1-4673-5769-2/13/$31.00 ©2013 IEEE

the nonlinear system under investigation is parameterized. For these approaches a review is given in [12]. Note that the order estimation in the above papers is in a global sense, i.e., the true order is unique over the whole regression space. In contrast to this, sometimes the true orders of a nonlinear system are not unique and may vary from point to point. To this end, let us consider an example given below.

Example (i): The finite impulse response system is given by

$$y_{k+1} = f_2(u_k, u_{k-1}, u_{k-2}) + \varepsilon_{k+1}, \tag{2}$$

where

$$f_2(u_k, u_{k-1}, u_{k-2}) = \begin{cases} u_k u_{k-1} u_{k-2}, & \text{if} \quad u_k > 1, \\ u_k u_{k-1}, & \text{if} \quad -1 \le u_k \le 1, \\ u_k, & \text{if} \quad u_k < -1. \end{cases}$$

The above example demonstrates a need for the local order estimation at points of interests. To the authors' knowledge, there has not much been done on this topic, though in reference [4] a forward/backward approach was proposed. Although numerical simulations seem to suggest that the forward/backward approach works well in terms of variable selection, determination of the system order and its theoretical study remain open.

The contribution of the paper is as follows. First, a kernel-based local information criterion, for simplicity of reference, named as the local information criterion (LIC), is proposed for the local order estimation of system (1). Then under moderate conditions, the estimates generated from LIC are proved to converge almost surely to the true local orders of system (1) at the points of interest.

The rest of the paper is arranged as follows. The LIC and the strong consistency of the estimates are given in Section II. A simulation example is given in Section III and some concluding remarks are addressed in Section IV.

Notations. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be the basic probability space. Let $\mathscr{B}^m$ denote the Borel σ-algebra on $\mathbb{R}^m$. For a vector $x = [x_1 \cdots x_m]^T \in \mathbb{R}^m$, denote its Euclidean norm by $\|x\|$ and its sub-vector by $x(i:j) \triangleq [x_i \cdots x_j]^T \in \mathbb{R}^{j-i+1}$. Denote by $\|\nu(\cdot)\|_{\text{var}}$ the total variation norm of a signed measure $\nu(\cdot)$. For two positive sequences $\{a_N\}_{N\ge1}$ and $\{b_N\}_{N\ge1}$, by $a_N \sim b_N$ it means that $c_1 b_N \le a_N \le c_2 b_N,\ \forall\ N \ge 1$, for some positive constants c_1 and c_2.

II. Local Order Estimation

A. Local Information Criterion for Order Estimation

Denote the regressor and the point of interest in the regression space $\mathbb{R}^{2M}$ by $\varphi_k(M,M)$ and $\varphi^*(M,M)$, respectively,

$$\varphi_k(M,M) \triangleq [y_k \cdots y_{k+1-M}\ u_k \cdots u_{k+1-M}]^T, \tag{3}$$
$$\varphi^*(M,M) \triangleq [y_1^* \cdots y_M^*\ u_1^* \cdots u_M^*]^T. \tag{4}$$

Similar to (3) and (4), for any fixed $1 \le p \le M$ and $1 \le q \le M$, define

$$\varphi_k(p,q) \triangleq [y_k \cdots y_{k+1-p}\ u_k \cdots u_{k+1-q}]^T, \tag{5}$$
$$\varphi^*(p,q) \triangleq [y_1^* \cdots y_p^*\ u_1^* \cdots u_q^*]^T. \tag{6}$$

We make the following assumption.

A1) The upper bound M for order is finite and known, and the true order of the NARX system at $\varphi^*(M,M)$ is $(p_0, q_0),\ 1 \le p_0 \le M,\ 1 \le q_0 \le M$, i.e.,

$$\begin{aligned} f(\varphi^*(M,M)) =& f(y_1^*, \cdots, y_{p_0}^*, \alpha_{p_0+1}, \cdots, \alpha_M, \\ & u_1^*, \cdots, u_{q_0}^*, \beta_{q_0+1}, \cdots, \beta_M) \end{aligned} \tag{7}$$

for all $\alpha_i \in \mathbb{R},\ i = p_0+1, \cdots, M$ and $\beta_j \in \mathbb{R},\ j = q_0+1, \cdots, M$.

At a given $\varphi^*(M,M)$, the task is to define and estimate the local order of $f(\cdot)$ based on the observations $\{u_k, y_k\}_{k\ge1}$. Since $f(\cdot)$ is nonlinear and nonparametric, it is difficult to choose the quantitative information based on which the algorithms estimating the local order can be designed. Most of the nonparametric identification algorithms only give estimates for the function value $f(\varphi^*(M,M))$ rather than the order. Denote the gradient of $f(\cdot)$ at $\varphi^*(M,M)$ by $\nabla f(\varphi^*(M,M)) \triangleq \left[\frac{\partial f}{\partial y_1^*} \ \cdots \ \frac{\partial f}{\partial y_M^*} \ \frac{\partial f}{\partial u_1^*} \ \cdots \ \frac{\partial f}{\partial u_M^*}\right]^T \in \mathbb{R}^{2M}$ if it exists. It is clear that if the true order of $f(\cdot)$ is (p_0, q_0), then $\frac{\partial f}{\partial y_i^*} = 0,\ i = p_0+1, \cdots, M$ and $\frac{\partial f}{\partial u_j^*} = 0,\ j = q_0+1, \cdots, M$, i.e.,

$$\nabla f(\varphi^*(M,M)) = \Bigg[\frac{\partial f}{\partial y_1^*} \cdots \frac{\partial f}{\partial y_{p_0}^*} \ \underbrace{0 \ \cdots \ 0}_{M-p_0} \ \frac{\partial f}{\partial u_1^*} \cdots \frac{\partial f}{\partial u_{q_0}^*} \ \underbrace{0 \ \cdots \ 0}_{M-q_0}\Bigg]^T, \tag{8}$$

and by the Taylor's expansion,

$$f(\varphi) = f(\varphi^*(M,M)) + \nabla f(\varphi^*(M,M))^T (\varphi - \varphi^*(M,M)) \tag{9}$$

for $\varphi \in \mathbb{R}^{2M}$ being close to $\varphi^*(M,M)$. From (8) and (9) it is seen that that if we can find a local linear model of $f(\cdot)$ at $\varphi^*(M,M)$, then we can estimate the local order by determining the biggest p and q such that $\frac{\partial f}{\partial y_p^*} \neq 0$ and $\frac{\partial f}{\partial u_q^*} \neq 0$.

To this end, we further impose the following assumption.

A2) $|f(x)| \le c_1\|x\|^r + c_2,\ x \in \mathbb{R}^{2M}$ for some positive constants $c_1,\ c_2$, and r and $f(\cdot)$ is twice differentiable at $\varphi^*(M,M)$. Further, $\frac{\partial f}{\partial y_p^*} \neq 0$ and $\frac{\partial f}{\partial u_q^*} \neq 0$ for some $p = 1, \cdots, M$ and $q = 1, \cdots, M$.

Definition 1: At the point $\varphi^*(M,M)$, the local order of $f(\cdot)$ is defined as (s_0, t_0), where

$s_0 \triangleq \max\left\{p = 1,\cdots,M \mid \frac{\partial f}{\partial y_p^*} \neq 0\right\}$ and $t_0 \triangleq \max\left\{q = 1,\cdots,M \mid \frac{\partial f}{\partial u_q^*} \neq 0\right\}$.

It is natural to ask why (s_0,t_0) rather than (p_0,q_0) is defined as the local order of $f(\cdot)$ at $\varphi^*(M,M)$? It is clear that if $\frac{\partial f}{\partial y_{p_0}^*} \neq 0$ and $\frac{\partial f}{\partial u_{q_0}^*} \neq 0$, then $(s_0,t_0) = (p_0,q_0)$. But sometimes the local orders given by Definition 1 are smaller than (p_0,q_0). Next we provide two examples to illustrate Definition 1.

Example (ii): For the linear system

$$y_{k+1} = a_1 y_k + \cdots + a_{p_0} y_{k+1-p_0} + b_1 u_k + \cdots + b_{q_0} u_{k+1-q_0} + \varepsilon_{k+1}$$

with $a_{p_0} \neq 0$, $b_{q_0} \neq 0$ and $f(\varphi^*(M,M)) = a_1 y_1^* + \cdots + a_{p_0} y_{p_0}^* + b_1 u_1^* + \cdots + b_{q_0} u_{q_0}^*$, it is clear that $\frac{\partial f}{\partial y_{p_0}^*} \neq 0$, $\frac{\partial f}{\partial u_{q_0}^*} \neq 0$, $\frac{\partial f}{\partial y_i^*} = 0,\ i = p_0+1,\cdots,M$, and $\frac{\partial f}{\partial u_j^*} = 0,\ j = q_0+1,\cdots,M$. Thus for this example assumption A2) takes place and the system order (s_0,t_0) derived by Definition 1 equals (p_0,q_0), which is consistent with the linear system theory.

Example (iii): For the nonlinear system

$$y_{k+1} = a y_k y_{k-1} + b u_k u_{k-1} + \varepsilon_{k+1}$$

with $a \neq 0$, $b \neq 0$, $f(\varphi(2,2)) = f(y_1,y_2,u_1,u_2) = a y_1 y_2 + b u_1 u_2$ at the fixed point $\varphi^*(2,2) = [y_1^*\ y_2^*\ u_1^*\ u_2^*]^T = [0\ 1\ 0\ 1]^T \in \mathbb{R}^4$, it is clear that $\nabla f(\varphi^*(2,2)) = [a\ 0\ b\ 0]^T$, and by Taylor's expansion $f(\varphi(2,2)) = f(\varphi^*(2,2)) + a\frac{\partial f}{\partial y_1^*}(y_1 - y_1^*) + b\frac{\partial f}{\partial u_1^*}(u_1 - u_1^*)$ for all $\varphi(2,2)$ close to $\varphi^*(2,2)$. This implies that it is reasonable to define the local order at the given point by $(s_0,t_0) = (1,1)$ rather than $(p_0,q_0) = (2,2)$.

Based on the above discussion the key step in our approach to estimating the local order is to find the local linear model of $f(\cdot)$ at $\varphi^*(M,M)$. In [3] and [22], the kernel function-based local linear estimator (LLE) and its recursive version (RLLE) are considered, which estimate the values of the nonlinear function at fixed points together with their gradients. Based on RLLE, we now introduce the order estimate for the NARX system (1) at a given point $\varphi^*(M,M)$.

Let us first introduce assumption A3):

A3) Select $b_k = \frac{1}{k^\delta}$ for some $\delta \in \left(0, \frac{1}{2(2M+1)}\right)$; $w(\cdot)$ is chosen as a symmetric probability density function (pdf) and $w(x) = O\left(\rho^{\|x\|}\right)$ for some $0 < \rho < 1$ as $\|x\| \to \infty$, and the matrix $\begin{bmatrix} 1 & 0 \\ 0 & \int_{\mathbb{R}^{2M}} w(x)xx^T \mathrm{d}x \end{bmatrix} > 0$.

With the given order (p,q) and measurements $\{\varphi_k(M,M), y_{k+1}\}_{k=1}^N$ the RLLE estimate of $f(\cdot)$ at time $N+1$ is given by

$$\begin{aligned} \theta_{N+1}(p,q) &= \left[\theta_{0,N+1}(p,q) \quad \theta_{1,N+1}^T(p,q)\right]^T \\ &\triangleq \operatorname*{argmin}_{\substack{\theta_0(p,q)\in\mathbb{R} \\ \theta_1(p,q)\in\mathbb{R}^{p+q}}} \sum_{k=1}^N w_k(\varphi^*(M,M)) \\ &\cdot \left(y_{k+1} - \theta_0(p,q) - \theta_1(p,q)^T(\varphi_k(p,q) - \varphi^*(p,q))\right)^2, \end{aligned} \tag{10}$$

where the kernel function $w_k(\varphi^*(M,M))$ is given by

$$w_k(\varphi^*(M,M)) = \frac{1}{b_k^{2M}} w\left(\frac{1}{b_k}(\varphi_k(M,M) - \varphi^*(M,M))\right). \tag{11}$$

Denote

$$X_k(p,q) \triangleq \begin{bmatrix} 1 \\ \varphi_k(p,q) - \varphi^*(p,q) \end{bmatrix}. \tag{12}$$

By some simple manipulations, RLLE in (10) can be expressed by

$$\begin{aligned} \theta_{N+1}(p,q) = &\left(\sum_{k=1}^N w_k(\varphi^*(M,M)) X_k(p,q) X_k(p,q)^T\right)^{-1} \\ &\cdot \left(\sum_{k=1}^N w_k(\varphi^*(M,M)) X_k(p,q) y_{k+1}\right), \end{aligned} \tag{13}$$

if the matrices $\sum_{k=1}^N w_k(\varphi^*(M,M)) X_k(p,q) X_k(p,q)^T,\ N \geq 1$ are nonsingular. Notice that by the matrix inverse lemma, $\theta_{N+1}(p,q)$ given by (13) can be computed in a recursive way.

A widely used kernel is the Gaussian pdf. Other important kernels include the rectangle kernel, triangle kernel, Epanechnikov kernel, etc.

For estimating the local order (s_0,t_0), we introduce the following local information criterion (LIC) $L_{N+1}(p,q)$:

$$L_{N+1}(p,q) \triangleq \sigma_{N+1}(p,q) + a_N \cdot (p+q), \tag{14}$$

where

$$\begin{aligned} \sigma_{N+1}(p,q) &\triangleq \sum_{k=1}^N w_k(\varphi^*(M,M)) \\ &\cdot \left(y_{k+1} - \theta_{0,N+1}(p,q) - \theta_{1,N+1}(p,q)^T(\varphi_k(p,q) - \varphi^*(p,q))\right)^2, \end{aligned} \tag{15}$$

$\{a_N\}_{N\geq 1}$ is a positive sequence tending to infinity as $N \to \infty$, and $\theta_{0,N+1}(p,q)$ and $\theta_{1,N+1}(p,q)$ are RLLE generated by (10) with the given order (p,q).

The order estimate (p_{N+1}, q_{N+1}) of (s_0,t_0) is defined by minimizing $L_{N+1}(p,q)$, i.e.,

$$(p_{N+1}, q_{N+1}) \triangleq \operatorname*{argmin}_{\substack{1\leq p\leq M \\ 1\leq q\leq M}} L_{N+1}(p,q). \tag{16}$$

Remark 1: RLLE $\theta_{0,N+1}(p,q)$ and $\theta_{1,N+1}(p,q)$ generated by (10) serve as the estimates for $f(\varphi^*(M,M))$ and its gradient, respectively. Notice that $\frac{\partial f}{\partial y_i^*}=0,\ i=s_0+1,\cdots,M$ and $\frac{\partial f}{\partial u_j^*}=0,\ j=t_0+1,\cdots,M$. Thus if RLLE approximates the true values well, then the function $\sigma_{N+1}(p,q)$ decreases as p and q increase but the performance does not change much for $p\geq s_0$ and $q\geq t_0$. On the other hand, as p and q increase, $(p+q)$ increases and this compensates decreasing of $\sigma_{N+1}(p,q)$. This indicates that (16) with appropriately chosen $\{a_N\}_{N\geq 1}$ defines a reasonable estimate for (s_0,t_0).

We list some further conditions used for convergence analysis of the order estimates. Note that (1) is an infinite impulse response nonlinear system, so the second order statistics may not contain adequate information for its identification, while ergodicity and mixing properties are often required. See, e.g., [10] in statistics literature. These conditions, in fact, are on the asymptotical independency and stationarity of the sequence $\{\varphi_k(M,M)\}_{k\geq 1}$, and they can be guaranteed by assuming stability of the system with input excited in a certain sense as shown in [21] and [22]. The conditions given in [21] and [22] cover a large class of systems, including the ARX system, the Hammerstein systems and the Wiener system, etc. So for ease of presentation, in this paper we assume that $\{\varphi_k(M,M)\}_{k\geq 1}$ is a mixing process with an asymptotically stationary distribution.

A4) $\{\varepsilon_k\}_{k\geq 0}$ is an independent and identically distributed (iid) sequence with $E\varepsilon_k=0,\ 0<E|\varepsilon_k|^{2+\eta}<\infty$ for some $\eta\in(0,2]$; $\varphi_k(M,M)$ and ε_{k+1} are mutually independent for each $k\geq 1$.

A5) The sequence $\{\varphi_k(M,M)\}_{k\geq 1}$ is geometrically ergodic, i.e., there exists an invariant probability measure $P_{\mathrm{IV}}(\cdot)$ on $(\mathbb{R}^{2M},\mathscr{B}^{2M})$ and some constants $c_1>0$ and $0<\rho_1<1$ such that $\|P_k(\cdot)-P_{\mathrm{IV}}(\cdot)\|_{\mathrm{var}}\leq c_1\rho_1^k$, where $P_k(\cdot)$ is the marginal distribution of $\varphi_k(M,M)$. $P_{\mathrm{IV}}(\cdot)$ is with a bounded pdf, denoted by $f_{\mathrm{IV}}(\cdot)$, which is with a continuous second order derivative at $\varphi^*(M,M)$.

A6) $\{\varphi_k(M,M)\}_{k\geq 1}$ is an α-mixing with mixing coefficients $\{\alpha(k)\}_{k\geq 1}$ satisfying $\alpha(k)\leq c_2\rho_2^k$ for some $c_2>0$ and $0<\rho_2<1$ and $E\|\varphi_k(M,M)\|^r<\infty$ for $k\geq 1$.

A7) The sequence $\{a_N\}_{N\geq 1}$ satisfies

$$\frac{N^{1-4\delta}}{a_N}\underset{N\to\infty}{\longrightarrow}0,\quad \frac{a_N}{N^{1-2\delta}}\underset{N\to\infty}{\longrightarrow}0, \tag{17}$$

where $\delta>0$ is given in A3).

The convergence of (16) is considered in the next section.

B. Strong Consistency of Estimates

For any fixed $1\leq p\leq M$ and $1\leq q\leq M$, define

$$\nabla f(\varphi^*(p,q))\triangleq\left[\frac{\partial f}{\partial y_1^*}\quad\cdots\quad\frac{\partial f}{\partial y_p^*}\quad\frac{\partial f}{\partial u_1^*}\quad\cdots\quad\frac{\partial f}{\partial u_q^*}\right]^T, \tag{18}$$

and

$$\bar{\theta}_{1,N+1}(p,q)\triangleq\Big[\theta_{1,N+1}(p,q)(1:p)^T\ \underbrace{0\cdots0}_{M-p}\quad \theta_{1,N+1}(p,q)(p+1:p+q)^T\ \underbrace{0\cdots0}_{M-q}\Big]^T\in\mathbb{R}^{2M}, \tag{19}$$

$$\widetilde{\theta}_{N+1}(p,q)\triangleq\big[f(\varphi^*(M,M))-\theta_{0,N+1}(p,q)\quad \nabla f(\varphi^*(M,M))^T-\bar{\theta}_{1,N+1}(p,q)^T\big]^T\in\mathbb{R}^{1+2M}, \tag{20}$$

and the maximal and minimal eigenvalues of $\sum_{i=1}^N w_i(\varphi^*(M,M))X_i(p,q)X_i(p,q)^T$ by $\lambda_{\max}^{(p,q)}(N)$ and $\lambda_{\min}^{(p,q)}(N)$, respectively.

Theorem 1: Under conditions A1)-A7), the order estimate (p_N,q_N) given by (16) is strongly consistent,

$$(p_N,q_N)\underset{N\to\infty}{\longrightarrow}(s_0,t_0)\quad \text{a.s.} \tag{21}$$

Proof: Due to space limitation, here we only present the key idea of the proof. Interested readers can find the details in [23].

Because all $p,\ q,\ s_0$, and t_0 are positive integers between 1 and M, for (21) it suffices to show that any limit point of $\{(p_N,q_N)\}_{N\geq 1}$ coincides with (s_0,t_0). Assume that (p',q') is a limit point of $\{(p_N,q_N)\}_{N\geq 1}$, i.e., there exists a subsequence of $\{(p_N,q_N)\}_{N\geq 1}$ denoted by $\{(p_{N_k},q_{N_k})\}_{k\geq 1}$, such that $(p_{N_k},q_{N_k})\underset{k\to\infty}{\longrightarrow}(p',q')$. Since $\{(p_N,q_N)\}_{N\geq 1}$ and (p',q') are nonnegative integers, there exists $K>0$ such that

$$(p_{N_k},q_{N_k})=(p',q'),\quad \forall\ k\geq K. \tag{22}$$

For (21) we need to prove the impossibility of the following cases: (i) $p'<s_0$; (ii) $q'<t_0$; (iii) $p'+q'>s_0+t_0$, based on which we will obtain (21). ■

Remark 2: LIC considered in the paper look similar to the well known AIC, BIC, and their generalizations. However, AIC, BIC, and others are in a global sense and thus they are inapplicable to the local order estimation. While for LIC the kernel function $w_k(\varphi^*(M,M))$ plays a bandwidth like role to stress those measurements which are close to the given point and to take their average.

III. Simulation

Consider the FIR system (2) where the inputs $\{u_k\}_{k\geq 1}$ and the noises $\{\varepsilon_k\}_{k\geq 1}$ are mutually independent iid Gaussian variables with distributions $\mathscr{N}(0,1)$ and $\mathscr{N}(0,0.1^2)$, respectively. Assume the upper bound M of system orders is 4. It is noticed that the right-hand side of (2) is free of the system output, so $\varphi^*(M,M)$ defined by (4) changes to $\varphi^*(M)\triangleq[u_1^*,\cdots,u_M^*]^T$, and $L_{N+1}(p,q)$ and $\sigma_{N+1}(p,q)$ defined by (14) and (15) correspondingly change to functions $L_{N+1}(q)$ and $\sigma_{N+1}(q)$, respectively. We choose three points for test,

$\varphi_1^* = [1\ 1\ 0\ 0]^T,\ \varphi_2^* = [0.5\ 0\ 0\ 0]^T$, and $\varphi_3^* = [-1\ 0\ 0\ 0]^T$. With the data set $\{u_k, y_{k+1}\}_{k=1}^N,\ \ N = 2000$ and the parameter $\delta = 0.01$, Figures 1, 2, and 3 show the performance of the proposed estimator with $a_N = 0.001N^{1-3\delta}$. The top diagrams of these figures show the performance of $\sigma_N(q)$ while the bottom diagrams show the performance of $L_N(q)$, $q = 1,\ 2,\ 3$, and 4. The corresponding values of $\sigma_N(q)$ and $L_N(q)$ are listed in Table 1. It can be found that the LIC gives the correct estimates for the local orders of the system.

TABLE I

VALUES OF $\sigma_N(q)$ AND $L_N(q)$, $q = 1,\ 2,\ 3,\ 4$

q	1	2	3	4
$\sigma(q)$ at φ_1^*	36.5239	35.0562	34.8976	34.9033
$L_N(q)$ at φ_1^*	36.5829	35.1742	35.0746	35.1393
$\sigma(q)$ at φ_2^*	29.1722	28.4396	28.4417	28.4506
$L_N(q)$ at φ_2^*	29.3314	28.7580	28.9193	29.0875
$\sigma(q)$ at φ_3^*	33.1448	34.5534	34.5532	34.5502
$L_N(q)$ at φ_3^*	34.7370	37.7378	39.3298	40.9190

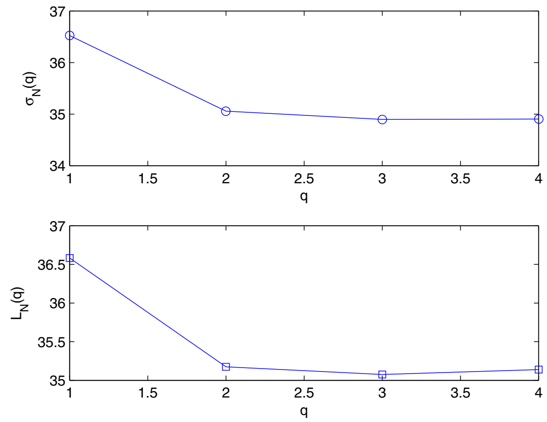

Fig.1 Performance of estimator at φ_1^*

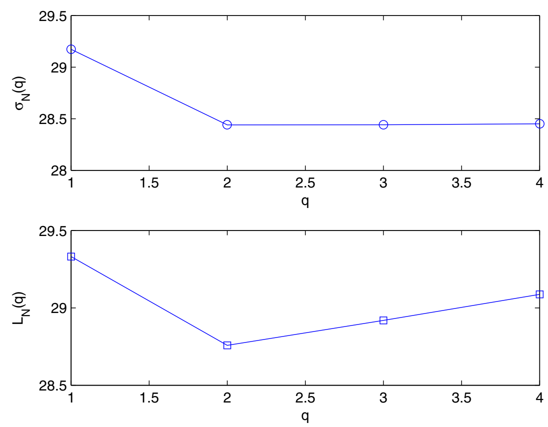

Fig.1 Performance of estimator at φ_2^*

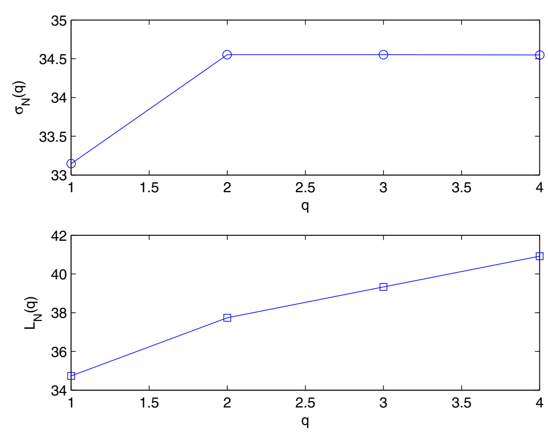

Fig.1 Performance of estimator at φ_3^*

IV. CONCLUDING REMARKS

In the paper LIC is suggested for the local order estimation of NARX systems and the consistency of the estimates is established under moderate conditions. Some important issues connected with LIC are summarized as follows.

1. This paper together with the results in [22] propose a framework for the local parameter and order estimation of nonparametric nonlinear systems. In applications, we may first perform the proposed algorithm to find the local order estimates and the local linear models at points chosen in the domain of interest, and then interpolate the whole function.
2. LIC is based on the recursive local linear estimator introduced in [22]. We can also use its nonrecursive version investigated in [3] to construct LIC and to carry out corresponding convergence analysis.
3. The results in the paper can be easily extended to the case $1 \le s_0 \le M_1$ and $1 \le t_0 \le M_2$ for some known but different M_1 and M_2. For future research, it is of interest to remove the upper bound assumption on the true system orders.
4. The order estimation algorithms in the paper are nonrecursive, i.e., for each $N \ge 1$ we need to calculate the function $L_{N+1}(p, q),\ \ 1 \le p \le M,\ \ 1 \le q \le M$ and then to find the minimum to serve as the estimate. It is interesting to consider the recursive way to obtain the order estimates.

5 The close-loop order estimation of NARX systems also deserves further research.

REFERENCES

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control* , Vol. 19, pp. 716–723, 1974.

[2] M. Autinl, M. Biey, and M. Haslerl, "Order of discrete time nonlinear systems determined from input-output signals", *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 296–299, 1992.

[3] E. W. Bai, "Non-parametric nonlinear system identification: An asymptotical minimum mean squared error estimator," *IEEE Trans. Automatic Control*, Vol. 55, No. 7, pp. 1615–1626, 2010.

[4] E. W. Bai, L. Kang, W. X. Zhao, and W. Y. Xu, "A kernel based forward/backward stepwise approach to local nonlinear non-parametric variable selection", *Automatica*, submitted for a possible publication.

[5] E. W. Bai, R. Tempo and Y. Liu, "Identification of IIR nonlinear systems without prior structural information," *IEEE Trans. Automatic Control*, Vol. 52, No. 3, pp. 442–453, 2007.

[6] J. D. Bomberger and D. Seborg, "Determination of model order for NARX models directly from input-output data", *J. Process Control*, Vol. 8, pp. 459–468, 1998.

[7] H. F. Chen and L. Guo, "Consistent estimation of the order of stochastic control systems", *IEEE Trans. Automatic Control*, Vol. 32, pp. 531–535, 1987.

[8] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Boston, MA: Birkhäuser, 1991.

[9] H. F. Chen and W. X. Zhao, "New method of order estimation for ARMA/ARMAX processes", *SIAM J. Control and Optimization*, Vol. 48, pp. 4157–4176, 2010.

[10] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Applications*. London, U.K.: Chapman & Hall/CRC, 1996.

[11] X. He and H. Asada, "A new method for identifying orders of input-output models for nonlinear dynamic systems", *Proceedings of the American Control Conference* , San Francisco, USA, pp. 2520–2523, 1993.

[12] X. Hong, R. J. Mitchell, S. Chen, et al, "Model selection approaches for nonlinear system identification:a review", *Int. J. of System Science*, Vol. 39, pp.925–949, 2008.

[13] M. B. Kennel, R. Brown, and H. Abarbanel, "Determining embedding dimension for phase-space reconstruction using geometrical construction", *Physical Review*, Vol. 45, pp. 3403–3411, 1992.

[14] B. Ninness and S. J. Henriksen, "Bayesian system identification via Markov chain Monte Carlo techniques," *Automatica*, Vol. 46, No. 1, pp. 40–51, 2010.

[15] P. Peduzzi, "A stepwise variable selection procedure for nonlinear regression methods", *Biometrics*, Vol. 36, pp. 510–516, 1980.

[16] S. Ramdania, J. F. Castiesa, F. Boucharab, et al, "Influence of noise on the averaged false neighbors method for analyzing time series", *Physica D*, Vol. 223, pp. 229–241, 2006.

[17] C. Rhodes and M. Morari, "Determining the model input/output order of nonlinear systems", *AIChE Journal*, Vol. 44, No. 1, pp. 151–163, 1998.

[18] J. Roll, I. Lind, and L. Ljung, "Connections between optimisation-based regressor selection and analysis of variance", *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 4907–4914, San Diego, USA, 2006.

[19] J. Roll, A. Nazin and L. Ljung, "Nonlinear system identification via direct weight optimation," *Automatica*, Vol. 41, No. 3, pp. 475–490, 2005.

[20] T. Schon, A. Wills and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, Vol. 47, No. 1, pp. 39–49, 2011.

[21] W. X. Zhao, H. F. Chen, and W. X. Zheng, "Recursive identification for nonlinear ARX systems based on stochastic approximation," *IEEE Trans. Automatic Control*, Vol. 55, No. 6, pp.1287–1299, 2010.

[22] W. X. Zhao, W. X. Zheng, and E. W. Bai, "A recursive local linear estimator for identification of nonlinear ARX systems: asymptotical convergence and applications," *IEEE Trans. Automatic Control*, provisionally accepted as a paper.

[23] W. X. Zhao, H. F. Chen, E. W. Bai, and K. Li, "Kernel-based local rrder estimation of nonlinear nonparametric systems", Technical Report in Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 2013.