Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Nolinear System Identification Based on Local Sub-Model Networks

Lianming Sun * Akira Sano **

* The University of Kitakyushu, Kitakyushu, 808-0135 Japan
(e-mail: sun@env.kitakyu-u.ac.jp)
** Keio University, Yokohama, 223-8522 Japan
(e-mail: sano@sd.keio.ac.jp)

Abstract: Local sub-model networks based nonlinear approximated model and the identification algorithm are developed for nonlinear system. The structure of local sub-model is selected through a criterion with respect to both the approximation accuracy and model simplicity for the further applications. The computional load of identification does not change so much with the increaing of number of sub-models. The sub-model can be a linear model, or a simple block-oriented nonlinear model to improve the model accuracy and convergence performance of the identification algortihm.

Keywords: Nonlinear systems; system identification; model structures; networks; parameter estimation.

1. INTRODUCTION

Nonlinear system analysis and system design are very urgent and challenging issues in a wide range of application areas, and the key step is construction of an effective model to satisfy practical requirement, see Billings [1980], Sjöberg et al. [1995] and Nelles [2001]. A system model may be obtained through system identification, nevertheless, nonlinear system identification is often not an easy task due to the complex model structure, complicated numerical nonlinear optimization, etc.

Many works have been considered to construct a global model for nonlinear system directly. For example, block-oriented models such as Hammerstein model, Wiener model, Hammerstein-Wiener model are applied in control applications, see Janczak [2005]. Although their structures are very simple, and the identification procedures can be performed easily, it seems that they are only effective to the systems whose nonlinearity can be separated from linear dynamics. Wiener or Volterra series based approximation is independent of the system structure as shown in Schetzen [1980], however, it often requires a large number of kernels and coefficients for high accuracy, so it often causes numerical problems easily. NARMAX model allows the approximation to use linear combination of parameters and regression containing the nonlinear components, i.e., linear in parameters, see Chen et al. [1989], and the regression structure is required to be chosen appropriately before parameter estimation from some prior system information. GMDH in Ivakhnenko [1970] is the method to choose the nonlinear components involved in the regression from a short data record, but its efficiency will decrease when handling a long data record. On the other hand, neural networks, genetic algorithms map from the input space into output space even though the physical structure of the nonlinear system is unknown, see Hunt et al. [1992], however, they usually need a large number of samples to optimize the weight coefficients of the nonlinear model. The wavelet based neural networks using orthogonal nonlinear basis functions have also been considered in Sjöberg et al. [1995].

Instead of construction of global models, some methods have constructed the nonlinear model by networks of a set of local models assigned by weighting functions to indicate their respective contribution in the nonlinear model. For example, local linear model tree (LOLIMOT) using simple linear models in Nelles et al. [1996], or Takagi-Sugeno fuzzy models using neuro-fuzzy model in Takagi et al. [1985], Pekpe et al. [2007], divides the operating space into several partitions corresponding to the operating regime of local models. The partition can be performed through measuring the deviation of the operating state from the partition center, or the self-organizing map, see Jeongho et al. [2003]. Furthermore, the comparison of global model and local model networks shows that the local model networks has some advantages, e.g., it does not require the structure information of nonlinear system, is very effective to dynamic systems, and can be implemented easily, see Brown et al. [1999], Koga et al. [2006]. However, the number of local models, convergence performance and computational complexity of identification depend on the structure of the local model largely.

In this paper, a nonlinear identification scheme based on local sub-model networks is developed. The model has the similar structure as LOLIMOT since it has an explicit interpretability, and the parameters of local sub-models can be estimated easily. In the new identification, the local model structure is selected through a specified criterion function to adjust trade-off between the accuracy and simplicity of the nonlinear model. Only the parameters of sub-models corresponding to new partitions are estimated per iteration, so the computational load does not change so much with increasing of sub-models number compared

978-1-1234-7890-2/08/$20.00 © 2008 IFAC 10.3182/20080706-5-KR-1001.2669

with other networks based identification methods. Furthermore, the structure of sub-model is extended to simple block-oriented nonlinear model for the system with severe nonlinearity to improve the model accuracy and convergence performance of parameter estimation.

2. PROBLEM STATEMENT

Assume the nonlinear system can be represented by

$$y(k)=f\big(y(k-1),\cdots,y(k-n_a),u(k-1),\cdots,\\ u(k-n_b)\big)+e(k) \tag{1}$$

where $u(k)$, $y(k)$ and $e(k)$ are the system input, output and noise at sampling instant k respectively. $f(\cdot)$ is an unknown nonlinear mapping function which will be approximated by a local sub-model networks illustrated in Fig. 1. The networks is composed of M sub-models followed by weight functions w_j, $j=1,2,\cdots,M$. w_j is determined by the regression of sub-model deviation from the partition centers, i.e., the closer the regression to the j-th partition center, the larger w_j will be. Each sub-model has the same structure and model order, and the parameters are estimated form the input and output data iteratively. The sub-model can be chosen as a linear dynamic model similarly as LOLIMOT, or the simple block-oriented model which is easy to be estimated, e.g., a Hammerstein model, if the application requires more accurate nonlinear model.

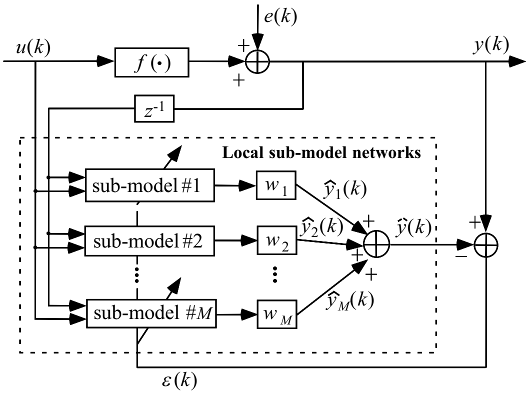

Fig. 1. Illustration of local sub-model networks

3. IDENTIFICATION ALGORITHM

In the new identification algorithm, the structure of the local sub-model will be determined first from the input and output data by using a specified criterion function with consideration of both approximation accuracy and model simplicity.

3.1 Selection of Sub-Model Structure

The selection of model structure is an essential step in system identification, see Haber et al. [1990], Sjöberg et al. [1995]. It influences the result of system identification greatly.

Two aspects of model quality are often considered. One is the model accuracy to approximate the main characteristics of the system, the other one is the model simplicity for applications. In the local sub-model networks, the structure of sub-model should be considered carefully. The structure of the local sub-models as well as the number of sub-models can be determined by using some criteria such as Mallow's C_{p} statistics, Akaike's information criterion (AIC), or Bayesian information criterion (BIC), etc., as illustrated in Haber et al. [1990]. However, the computational load may be heavy since not only the estimation of sub-models, but also the partitions have to be re-calculated when using different structure of local models. In the proposed identification algorithm, a simple scheme to select local model orders just using one local linear model is proposed.

Consider the case where the sub-model is a linear ARX model. Define the mean squares error $V_1(\boldsymbol{\theta}_1)$ by

$$V_1(\boldsymbol{\theta}_1)=\frac{1}{N}\sum_{k=1}^{N}\varepsilon^2(k) \tag{2}$$

where the nonlinear system is approximated by only a single linear model and $\boldsymbol{\theta}_1$ is its parameter vector obtained by minimizing $V_1(\boldsymbol{\theta}_1)$ through least squares method, and

$$\varepsilon(k)=y(k)-\hat{y}(k), \tag{3}$$

$$\hat{y}(k)=\boldsymbol{\phi}^T(k)\boldsymbol{\theta}_1, \tag{4}$$

$$\boldsymbol{\phi}(k)=\big[\,y(k-1)\quad\cdots\quad y(k-n_a)\\ u(k-1)\quad\cdots\quad u(k-n_b)\,\big]^T, \tag{5}$$

$$\boldsymbol{\theta}_1=\big[\,a_1\quad\cdots\quad a_{n_a}\quad b_1\quad\cdots\quad b_{n_b}\,\big]^T. \tag{6}$$

Then the model denominator and numerator orders n_a, n_b can be selected by minimizing the criterion function

$$n_a,n_b=\arg\min_{n_a,n_b}J(n_a,n_b) \tag{7}$$

where $J(n_a,n_b)$ is defined by

$$J(n_a,n_b)=V_1(\boldsymbol{\theta}_1)+\lambda V_2(n_a,n_b), \tag{8}$$

$$V_2(n_a,n_b)=\frac{n_a+n_b}{\alpha+\prod\limits_{i,j}|p_i-z_j|}. \tag{9}$$

Here p_i and z_j are poles and zeros of the transfer function of the estimated model, α is a small regularization positive number. $J(n_a,n_b)$ in (8) can be interpreted as follows. The first term $V_1(\boldsymbol{\theta}_1)$ is the accuracy indicator of local model networks, whereas the second term $V_2(n_a,n_b)$ is a penalty function corresponding to the simplicity of model structure, and the constant λ is used as a weight adjustment between these two terms. When the locations of some poles p_i are close to some zeros z_j, the penalty term becomes large to prevent the excessive model complexity. In the applications, λ can be chosen as a proportional factor of the variance of output $y(k)$ to adjust the weights of $V_1(\boldsymbol{\theta}_1)$ and $V_2(n_a,n_b)$.

As initial values, let the mean and standard deviation vectors of the regression of the linear model with parameter vector $\boldsymbol{\theta}_1$ be denoted as $\boldsymbol{c}_1$ and $\boldsymbol{\sigma}_1$, where

$$\boldsymbol{c}_1=\big[\,c_{1,1}\quad\cdots\quad c_{1,n_a+n_b}\,\big]^T$$

$$\boldsymbol{\sigma}_1=\big[\,\sigma_{1,1}\quad\cdots\quad \sigma_{1,n_a+n_b}\,\big]^T$$

and $c_{1,i}$, $\sigma_{1,i}$ are the mean, standard deviation of i-th entry of the regression $\boldsymbol{\phi}(k)$.

For the simplicity of calculations, let the structure of all the local linear models have same orders selected by (8), then the regressions of each sub-model have n_a+n_b entries.

3.2 Partition of Regressions

In the M-th iteration, the regressions $\boldsymbol{\phi}(k)$ for $k = 1, \cdots, N$ are divided into M partitions. The output of the approximated model is given by

$$\hat{y}(k) = \sum_{j=1}^{M} \boldsymbol{\phi}^T(k)\boldsymbol{\theta}_j w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j) \tag{10}$$

where $\boldsymbol{\theta}_j$ is the $(n_a + n_b) \times 1$ parameter vector of the sub-model corresponding to the j-th partition. The weight functions $w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j)$ take important roles in local sub-model networks. They indicate which sub-model will dominate the model output at instant k, and make the operating regime of the networks be adaptive to the current state of nonlinearity. Following the methods used in LOLIMOT Nelles et al. [1996] and neuro-fuzzy model Takagi et al. [1985], $w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j)$ can be given by

$$w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j) = \frac{q_j(k)}{\sum\limits_{l=1}^{M} q_l(k)}, \tag{11}$$

$$q_l(k) = e^{-\frac{1}{2}\left(\frac{\left(y(k-1)-c_{l,1}\right)^2}{\sigma_{l,1}^2}+\cdots+\frac{\left(u(k-n_b)-c_{l,n_a+n_b}\right)^2}{\sigma_{l,n_a+n_b}^2}\right)}. \tag{12}$$

Then the prediction error $\varepsilon(k)$ becomes to

$$\varepsilon(k) = y(k) - \hat{y}(k). \tag{13}$$

Similarly as the regression partition, the prediction error $\varepsilon(k)$ can also be divided into M partitions and the corresponding local mean squares error denoted as MSE_j, $j = 1, \cdots, M$, can be calculated by

$$\text{MSE}_j = \frac{1}{N_j} \sum_{\substack{\boldsymbol{\phi}(k) \in j-\text{th} \\ \text{partition}}} \varepsilon^2(k) \tag{14}$$

where N_j is the total regression number of the j-th partition. In order to reduce the global error in the $(M + 1)$-th iteration, the partition to be divided into two new partitions is determined by choosing the partition whose local error MSE_j is the largest one among the M partitions. Let such partition number be denoted as m, then m is given by

$$m = \arg\max_j \text{MSE}_j. \tag{15}$$

There are $n_a + n_b$ candidates to divide the m-th partition with respect to the regression size. Consider the m-th partition in the last iteration. Following the i-th entry of regression $\boldsymbol{\phi}(k)$, which is denoted as $x_i(k)$, the m-th partition can be divided into two new partitions

$$\begin{aligned} &\text{If } x_i(k) < c_{m,i} \\ &\quad \boldsymbol{\phi}(k) \in m-\text{th partition;} \\ &\text{else} \\ &\quad \boldsymbol{\phi}(k) \in (M+1)-\text{th partition.} \end{aligned} \tag{16}$$

Then the mean and standard deviation denoted as $\boldsymbol{c}_m$, σ_m of the new m-th partition, $\boldsymbol{c}_{M+1}$ and $\boldsymbol{\sigma}_{M+1}$ of $(M + 1)$-th partition, are calculated from the regressions in m-th and $(M + 1)$-th partitions respectively. Mean of the other partitions $\boldsymbol{c}_1, \cdots, \boldsymbol{c}_{m-1}, \boldsymbol{c}_{m+1}, \cdots, \boldsymbol{c}_M$ as well as the standard deviation $\boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_{m-1}, \cdots, \boldsymbol{\sigma}_{m+1}, \cdots, \boldsymbol{\sigma}_M$ are kept the same values as those in the last iteration for the simplicity of computation.

3.3 Parameter Estimation of Local Models

Following Fig. 1, the nonlinear system is approximated by $M + 1$ sub-models

$$\hat{y}(k) = \sum_{j=1}^{M+1} \boldsymbol{\phi}^T(k)\boldsymbol{\theta}_j w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j) \tag{17}$$

in the $(M+1)$-th iteration, where $\boldsymbol{\theta}_j$ is the parameter vector of the sub-model corresponding to the j-th partition. Corresponding to the new partitions, the weight functions $w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j)$ are updated by

$$w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j) = \frac{q_j(k)}{\sum\limits_{l=1}^{M+1} q_l(k)} \tag{18}$$

Notice that in (17) only the m and $(M + 1)$-th partitions are the new ones, then by letting the parameter vectors of $1, \cdots, m-1, m+1, \cdots, M$-th sub-models be the same as those in the last iteration, only the parameter vectors $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_{M+1}$ remain unknown. Define the MSE cost function $V(\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M+1})$ by

$$\begin{aligned} &V(\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M+1}) = \\ &\frac{1}{N}\sum_{k=1}^{N}\Big(y_{\text{res}}(k) - \boldsymbol{\phi}^T(k)\boldsymbol{\theta}_m w_m(\boldsymbol{\phi}(k), \boldsymbol{c}_m, \boldsymbol{\sigma}_m) \\ &\qquad -\boldsymbol{\phi}^T(k)\boldsymbol{\theta}_{M+1} w_{M+1}(\boldsymbol{\phi}(k), \boldsymbol{c}_{M+1}, \boldsymbol{\sigma}_{M+1})\Big)^2 \end{aligned} \tag{19}$$

where the signal $y_{\text{res}}(k)$ is given by

$$\begin{aligned} y_{\text{res}}(k) = y(k) &- \sum_{j=1}^{m-1} \boldsymbol{\phi}^T(k)\boldsymbol{\theta}_j w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j) \\ &- \sum_{j=m+1}^{M} \boldsymbol{\phi}^T(k)\boldsymbol{\theta}_j w_j(\boldsymbol{\phi}(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j). \end{aligned} \tag{20}$$

Then the parameter vectors $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_{M+1}$ corresponding to this partition candidate are estimated by some algorithms such as least squares through minimizing the cost function $V(\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M+1})$. Notice that only $2(n_a + n_b)$ parameters are required to be estimated for one partition candidate, the computational complexity is almost the same in every iteration and does not increase too much even though the sub-model number increases.

Following $(n_a + n_b)$ entries of the regression $\boldsymbol{\phi}(k)$, there are $(n_a + n_b)$ candidates to divide the m-th partition, consequently $(n_a + n_b)$ sets of the estimated parameters and the corresponding global MSE functions $V(\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M+1})$ are obtained. Choosing the partition candidate corresponding to the smallest global MSE function $V(\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M+1})$ yields the optimal partition of regressions $\boldsymbol{\phi}(k)$ in this iteration, correspondingly the parameters of m, $(M+1)$-th sub-models, the mean vector as well as the standard deviation vector can be updated.

It is noticed that the number of sub-models increases with the increasing of iteration number. If some sub-models with the similar mapping property are merged into one sub-model, the complexity of the networks can be decreased.

3.4 Identification Procedures

Assume that the order range of sub-model is $n_1 \leq n_a, n_b \leq n_2$, where n_1 and n_2 are determined by prior information. If no prior information is available, n_1 can be chosen as 1, and n_2 can be given by an enough large integer such that mean squares error does not decrease too much for $n_a, n_b > n_2$. The identification using local sub-model networks can be performed by the following procedures.

Step 1(a): Let $n_a = n_b = n_1$, and the regressions be constructed by (5) from the observation data.

Step 1(b): Estimate parameter vector $\boldsymbol{\theta}_1$ in (6).

Step 1(c): Calculate criterion function $J(n_a, n_b)$.

Step 1(d): If both n_a and n_b are larger than n_2, go to Step 2. Otherwise let $n_a = n_a + 1$ if $n_a < n_2$; otherwise let $n_a = n_1$, $n_b = n_b + 1$ then return to Step 1(b).

Step 2: Choose the orders n_a, n_b such that $J(n_a, n_b)$ has the smallest value, and the corresponding estimation of parameter vector $\boldsymbol{\theta}_1$. Let $M = 1$, calculate $\boldsymbol{c}_1$ and $\boldsymbol{\sigma}_1$ by using the all the regressions, and weight function $w_1(\boldsymbol{\phi}(k), \boldsymbol{c}_1, \boldsymbol{\sigma}_1) = 1$. Choose the partition number to be divided in the next iteration as $m = 1$, and the division entry number $i = 1$, then start the identification iteration.

Step 3(a): Divide the regressions $\boldsymbol{\phi}(k)$ in m-th partition into two new partitions following (16) with respect to i-th entry of $\boldsymbol{\phi}(k)$.

Step 3(b): Calculate weight functions w_1, $\cdots$, w_{M+1} for the current division candidate.

Step 3(c): Calculate $y_{\text{res}}(k)$ in (20).

Step 3(d): Estimate $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_{M+1}$ by minimizing $V(\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M+1})$ in (19).

Step 3(e): Let $i = i + 1$. If $i \leq n_a + n_b$, return to Step 3(a), otherwise go to Step 4.

Step 4: Choose the regression partition such that $V(\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M+1})$ has the smallest value from $(n_a + n_b)$ candidates, then update the corresponding parameter vectors $\boldsymbol{\theta}_m$, $\boldsymbol{\theta}_{M+1}$, mean and standard deviation $\boldsymbol{c}_m$, $\boldsymbol{\sigma}_m$, $\boldsymbol{c}_{M+1}$, $\boldsymbol{\sigma}_{M+1}$ of the new partitions.

Step 5: Calculate the local errors MSE_j for $j = 1, \cdots, M+1$ in (14).

Step 6: Let m be number of the partition that has largest local error, and let $M = M + 1$. Return to Step 3(a) to continue the next iteration.

4. EXTENSION TO LOCAL NONLINEAR MODEL NETWORKS

A simple nonlinear local sub-model can also be used in the networks. For example, by introducing some nonlinear components of input signals, the performance to deal with nonlinearity can be improved significantly. In this section, the networks using a Hammerstein model as local sub-models are considered.

Let the model output of the networks model with M sub-models be represented by

$$\hat{y}(k) = \sum_{j=1}^{M} \left(\boldsymbol{\phi}_L^T(k)\boldsymbol{\theta}_{j,L} + \boldsymbol{\phi}_{NL}^T(k)\boldsymbol{\theta}_{j,NL} \right) w_j(\boldsymbol{\phi}_L(k), \boldsymbol{c}_j, \boldsymbol{\sigma}_j) \tag{21}$$

where $\boldsymbol{\phi}_L(k)$ is the linear part of regression, $\boldsymbol{\phi}_{NL}(k)$ is the nonlinear part that contains the nonlinear terms of input signal, e.g.,

$$\boldsymbol{\phi}_{NL}(k) = \left[\, g\big(u(k-1)\big) \quad \cdots \quad g\big(u(k-n_b)\big) \,\right]^T \tag{22}$$

Here $g(\cdot)$ is a nonlinear basis function, $\boldsymbol{\theta}_{j,L}$ and $\boldsymbol{\theta}_{j,NL}$ are the linear regression, nonlinear regression corresponding parameter vectors respectively. The weight function w_j is determined by the linear regression $\boldsymbol{\phi}_L(k)$ and its mean vector $\boldsymbol{c}_j$, standard deviation vector $\boldsymbol{\sigma}_j$. When the order selection of sub-model and partition of regression are performed similarly as the linear sub-model case, where only the linear part $\boldsymbol{\phi}_L(k)$ is considered for order selection in (8) and regression partition in (16), the parameter estimation can be performed similarly as the linear sub-model case.

Moreover, it is possible to decrease the computation load if the nonlinear regression part is also considered in (8) to reduce the model orders.

5. NUMERICAL EXAMPLES

Two simulation examples are considered in this section. The first one is the example to deal with local linear model networks.

5.1 Example 1: Case of Local Linear Model

Consider a true nonlinear system which is represented by

$$\begin{aligned} & s(k) - 1.5s(k-1) + 0.7s(k-2) \\ =\; & r(k-1) + 0.5r(k-2) \\ & r(k) = u(k) + 0.1u^2(k) \\ & y(k) = 0.1s(k) + 0.025s^3(k) + 0.0064s^5(k) \\ & \quad + e(k) \end{aligned} \tag{23}$$

where $e(k)$ is an i.i.d white Gaussian noise with $\mathcal{N}(0, 0.05^2)$. Provide that 4096 input and output data $u(k)$, $y(k)$ are collected for system identification. Here $u(k)$ is chosen as a uniformly distributed random signal on [0, 1]. The system is a black-box system, where the information of nonlinear mapping function is unknown, and the intermediate signals $r(k)$ and $s(k)$ cannot be observed either. The system will be identified to construct a model of local linear model networks from $u(k)$ and $y(k)$.

Let $n_a = n_b = n$ for the simplicity of notation to select the local model order. The function $J(n_a, n_b)$ in (8) for $n = 1, 2, \cdots, 9$ is shown in Fig. 2. λ is chosen as $\sum_{k=1}^{N}(y(k) - \bar{y})^2/(6 \cdot 10^4 N)$, where $\bar{y}$ is the mean of $y(k)$.

It illustrates that in this example $J(4,4)$ has the smallest value, so the model orders are chosen as $n_a = n_b = 4$ for system identification.

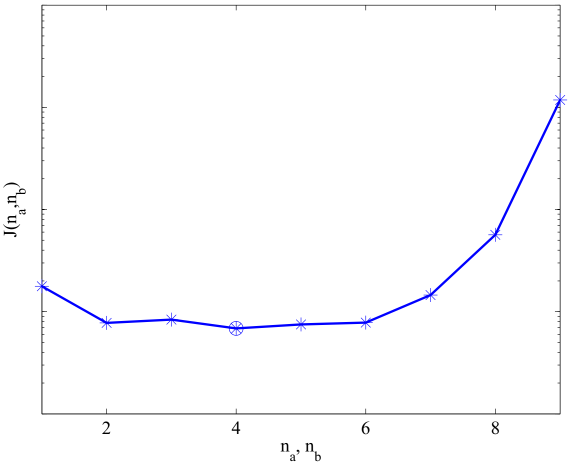

Fig. 2. $J(n_a, n_b)$ vs n_a and n_b in local model structure selection. ◯: selected model order

As a comparison, the function $V_1(\boldsymbol{\theta}_1)$ and $V_2(n_a, n_b)$ are plotted in Fig. 3, where $V_1(\boldsymbol{\theta}_1)$ decreases slightly with increasing of model orders, while the penalty function $V_2(n_a, n_b)$ increases quickly for high model orders. In order to verify the effectiveness of the selection approach of local model orders, the global MSE of the obtained model with respect to various model orders is shown in Fig. 4. It can be seen that the global MSE in the proposed algorithm where $n = 4$ is much lower than the conventional LOLIMOT model in Nelles et al. [1996] where $n = 2$. Though the values of MSE for $n \geq 5$ are slight smaller than that of $n = 4$, the computational load increases largely to deal with high dimension of matrix computation and too many candidates of regression partition. So it implies that the model structure selection by (8) is effective in the proposed identification algorithm.

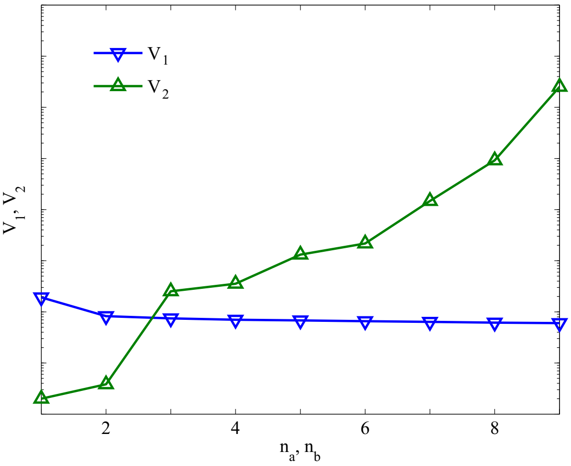

Fig. 3. $V_1(\boldsymbol{\theta}_1)$ and $V_2(n_a, n_b)$ vs model orders n_a, n_b

The observed system output $y(k)$ and the output of networks $\hat{y}(k)$ in the 21-th iteration for $k = 3000, \cdots, 3100$ are illustrated in Fig. 5. It can be seen that $\hat{y}(k)$ is very close to the system output observations so the identified model can be used to represent the characteristics of the plant system.

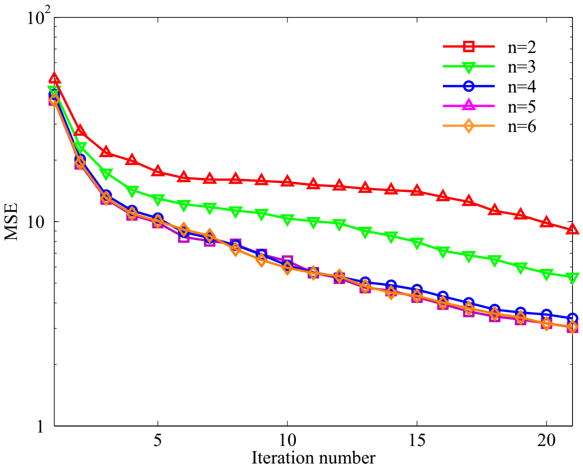

Fig. 4. Global MSE vs local model orders $n_a = n_b = 2, 3, 4, 5, 6$. (Average of 20 simulation runs)

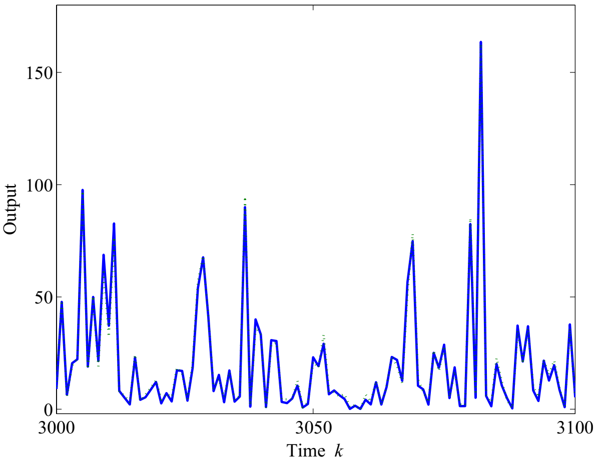

Fig. 5. System output $y(k)$ and model output $\hat{y}(k)$ in Example 1. Solid line: $y(k)$; dotted line: $\hat{y}(k)$

5.2 Example 2: Case of Local Nonlinear Model

Consider a nonlinear system which given by

$$\begin{aligned} & s(k) - 1.5s(k-1) + 0.7s(k-2) \\ = & \, r(k-1) + 0.5r(k-2) \\ & r(k) = u(k) + 0.1\sqrt{u(k)} \\ & y(k) = 0.1s(k) + 0.025s^3(k) + 0.0064s^5(k) \\ & \quad + e(k) \end{aligned} \tag{24}$$

where $e(k)$ is an i.i.d white Gaussian noise with $\mathcal{N}(0, 7.5^2)$. 4096 input and output data are used for identification. The sub-model is chosen as a Hammerstein model whose nonlinear regression part is given by a square function, i.e., $g(x) = x^2$ in (22). The other simulation conditions are the same as those in Example 1.

The orders of linear dynamics in Hammerstein sub-model, weight function w_j are determined just from $\boldsymbol{\phi}_L(k)$. So the computational load increases slightly to estimate $\boldsymbol{\theta}_{m,NL}$ and $\boldsymbol{\theta}_{M+1,NL}$. The orders of linear dynamics of sub-model is selected as $n_a = n_b = 4$ based on $J(n_a, n_b)$ in (8). As a comparison, the global MSE of the linear sub-model networks and Hammerstein sub-model networks is plotted in Fig. 6.

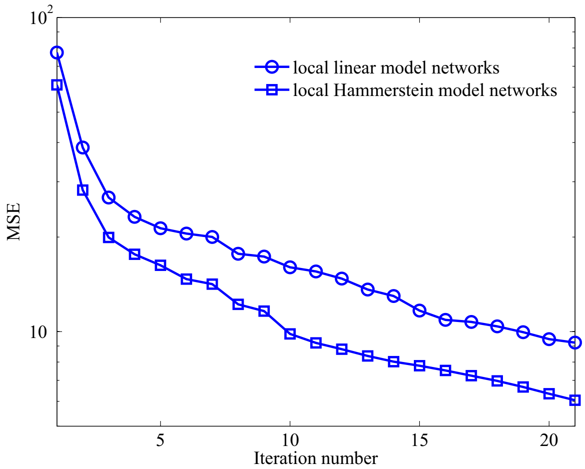

Fig. 6. Global MSE of local linear model networks and local Hammerstein model networks (Average of 20 simulation runs)

It can be seen that MSE obtained by using the Hammerstein sub-model networks is about 1 times faster to reach low level than that by using linear local model networks. Though the nonlinear component of $\phi_{NL}(k)$ is a square function, which is quite different from the true nonlinearity of square root, the convergence performance of the networks using nonlinear local model is superior to that of the linear local model networks. The comparison of true observed system output $y(k)$ and the output $\hat{y}(k)$ of obtained networks of local Hammerstein models in the 21-th iteration is also illustrated in Fig. 7, where the model output $\hat{y}(k)$ is close to the true one so the obtained networks model is valid for approximation of the true nonlinear system.

6. CONCLUSIONS

The local sub-model networks based identification algorithm has been developed for nonlinear system identification. By selecting appropriate orders of the local sub-model, the networks can approximate the characteristics of the nonlinear system effectively. Furthermore, by using local nonlinear models, the convergence performance of the networks can be improved significantly. The selection of nonlinear regressions, more effective criterion function to select model order for the local nonlinear model networks, recursive identification algorithm and merging of sub-models will be considered in the future research work.

REFERENCES

S.A. Billings. Identification of non-linear systems–A survey. *IEE Proc.*, volume 127, Pt. D, number 6, pages 272–285. 1980.

M.D. Brown, and G.W. Irwin. Nonlinear identification and control of turbogenerators using local model networks. *Proc. American Control Conf.*, volume 6, pages 4213–4217. 1999.

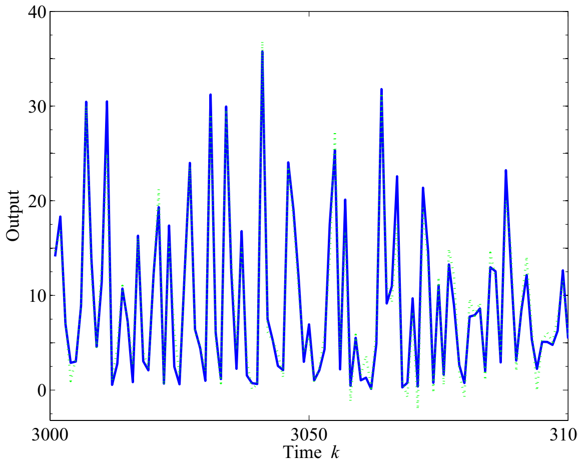

Fig. 7. System output $y(k)$ and model output $\hat{y}(k)$ in Example 2. Solid line: $y(k)$; dotted line: $\hat{y}(k)$

S. Chen and S.A. Billings. Representation of non-linear systems: The NARMAX model. *Int. J. Control*, volume 49, number 3, pages 1013–1032. 1989.

R. Haber and H. Unbehauen. Structure identification of nonlinear dynamic system–A survey on input output approaches. *Automatica*, volume 26, number 4, pages 651–677. 1990.

K.J. Hunt, et al. Neural networks for control systems–A survey. *Automatica*, volume 28, number 6, pages 1083–1112. 1992.

A.G. Ivakhnenko. heuristic self-organization in problems of engineering cybernetics. *Automatica*, volume 6, pages 207–219. 1970.

A. Janczak. *Identification of Nonlinear Systems using Neural Networks and Polynomial Models: A Block-Oriented Approach.* Springer. 2005.

C. Jeongho, J.C. Principe, and M.A. Motter. Local Hammerstein modeling based on self-organizing map. *IEEE Workshop on Neural Networks for Signal Processing*, pages 809–818, Sept., 2003.

K. Koga, and A. Sano. Query-based approach to prediction of MR damper force with application to vibration control. *Proc. American Control Conf.*, pages 3259–3265. 2006.

O. Nelles. *Nonlinear System Identification: from Classical Approaches to Neural Networks and Fuzzy Models.* Springer. 2001.

O. Nelles, and R. Isermann. Basis function networks for interpolation of local linear models. In *35th IEEE Conference on Decision and Control*, volume 1, pages 470–475. 1996.

K.M. Kekpe, J.P. Cassar, and S. Chenikher. Identification of MIMO Takagi-Sugeno model of a bioreactor. *IEEE Int. Fuzzy Systems Conf.*, pages 1–6, July, 2007.

M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems.* John Wiley and Sons, New York. 1980.

J. Sjöberg, et al. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, volume 31, pages 1691–1724. 1995.

T. Takagi and M. Sugeno. Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. on Systems, Man, and Cybernetics*, volume 15, number 1, pages 116–132. 1985.