Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Identification of an Additive NFIR System and its applications in generalized Hammerstein models

Er-Wei Bai
Dept. of Electrical and Computer Engineering
University of Iowa, Iowa City, Iowa 52242
er-wei-bai@uiowa.edu

***Abstract*— Identification of a nonlinear additive system is considered. An input signal is designed in such a way that the problem of identification of nonlinear additive systems is reduced to a problem of identification of static nonlinear functions. Then, three approaches are established to estimate the order of the system. The methods exploit the structure of the nonlinear additive model so that their implementations are easy.**

I. Introduction

In this paper, we consider identification of a nonlinear FIR additive model,

$$y(t) = f_1[u(t)]+f_2[u(t-1)]+...+f_p[u(t-(p-1))]+\mu+v(t) \tag{I.1}$$

where $u(t)$ and $y(t)$ are the system input and output at time $t = 1, 2, ...$, respectively. The unknown constant μ represents the possible DC offset and the noise sequence $v(t)$ is assumed to be independent identically distributed random variables with zero mean, independent of the input and $\mathbf{E}|v(t)|^{4+\rho} < \infty$ for some $\rho > 0$. In particular, $\mathbf{E}|v(t)|^2 = \sigma_v^2$. Here, $\mathbf{E}$ stands for the expectation operator. No a priori information on the structures of f_i's is assumed. Also, only an upper bound $(p-1)$ on the order is assumed. The exact order is unknown.

The model (I.1) is often referred to as the nonlinear FIR additive system in the literature [3], [6]. It is one of the widely used nonlinear and non-parametric techniques to describe nonlinear behaviors. A generalized Hammerstein model is also a special case of the additive model. Let

$$b_i(z) = b_{1i} + b_{2i}z^{-1} + ... + b_{q_i i}z^{-(q_i-1)}, \quad i = 1, ..., q$$

where z^{-1} represents the unit delay. The non-parametric generalized FIR Hammerstein model [5] is in the form of

$$y(t) = b_1(z)\bar{f}_1(u(t))+b_2(z)\bar{f}_2(u(t))+...+b_q(z)\bar{f}_q(u(t))+v(t) \tag{I.2}$$

for some unknown $b_i(z)$'s and nonlinear functions $\bar{f}(\cdot)$'s. In particular, if $b_1(z) = b_2(z) = ... = b_q(z)$ and $\bar{f}_1 = \bar{f}_2 = ... = \bar{f}_q$, the generalized Hammerstein model becomes a standard FIR Hammerstein model. The generalized Hammerstein model has extensive applications in controls, e.g., in thermal power plant, heat exchanger, stream flow process, cement kiln and closed circuit cement ball grinding mill [5]. Now, let

$$p = \max_i q_i, \quad b_{ji} = 0 \ if \ j > q_i$$

We can re-write the generalized Hammerstein model (I.2) as

$$y(t) = \sum_{i=1}^{q_1} b_{i1}\bar{f}_1(u(t-i+1))+...+\sum_{i=1}^{q_q} b_{iq}\bar{f}_q(u(t-i+1))+v(t)$$

$$= \sum_{l=1}^{q} b_{1l}\bar{f}_l(u(t)) + ... + \sum_{l=1}^{q} b_{pl}\bar{f}_l(u(t-p+1)) + v(t)$$

$$= f_1(u(t) + f_2(u(t-1)) + ... + f_p(u(t-p+1)) + v(t)$$

which is exactly an additive model.

Additive models can also be viewed as a generalization of the well known linear FIR models [11], [15]

$$y(t) = \alpha_1 u(t) + \alpha_2 u(t-1) + ... + \alpha_p u(t-p+1) + v(t)$$

where the linear terms $\alpha_j u(t-j+1)$'s are replaced by the nonlinear terms $f_j[u(t-j+1)]$'s which provide capabilities to describe nonlinear behavior that linear systems are inadequate to model.

Though widely used in many fields, to the best of our knowledge, identification of the nonlinear additive model (I.1) has not received much attention in the control/identification community. Most studies on the additive models are found in the regression literature [3], [6] where the setting can be quite different. Three main difficulties associated with identification of the additive model (I.1) are

- The couplings of f_i's. Note that the output $y(t)$ depends on all the nonlinearities $f_i[u(t-i+1)]$, $i = 1, ..., p$, and this coupling obviously makes identification of the additive model non-trivial.
- The order estimation. Since no information on the exact order is available, order estimation has to be part of identification.
- The lack of a priori information on the structures of the unknown nonlinearities. In fact, f_i's can be discontinuous.

We make a comment on the last difficulty. Though identification of nonlinear systems without structural information remains an intractable task, considerable advancements have been made for identification of a static nonlinearity $y = f(u)$ without a priori structural information on f. For example, the kernel methods [8], the orthonormal basis and series expansion methods [1], [13] and the smooth spline method [2], [16] have been proposed and analyzed in details. Therefore, it is our intention to focus on the first two difficulties in the paper. Clearly, if the decoupling problem can be solved, identification of the additive model is reduced to identification of static nonlinearities f_i's separately. We will elaborate on this issue more later.

The main contributions of the paper are some novel ideas that make the order determination and decoupling possible without a priori knowledge on the unknown nonlinearities.

II. INPUT DESIGNS

Given the range of interest $I = [-a, a]$ for the input, we divide I into $(m-1)$ partitions

$$-a = a_1 < a_2 < ... < a_m = a \tag{II.1}$$

The symmetry of I is not necessary and is for simplicity only. First, for $t \in [1, mp]$, we define $u(t)$ as

$$u(t) = \begin{cases} a_{i+1} & t = ip+1,\ i = 0, ..., m-1 \\ a_1 & otherwise \end{cases} \tag{II.2}$$

To average out the effects of the noise, we repeat $u(t)$ n times, i.e., $u(t)$ is periodic with period mp and

$$\{u(t)\}_{t=1}^{mp} = \{u(t)\}_{t=mp+1}^{2mp} = ... = \{u(t)\}_{t=(n-1)mp+1}^{N}$$

where $N = mpn$. With proper initial conditions,

$$u(0) = u(-1) = ... = u(-p+2) = a_1 \tag{II.3}$$

it can be easily verified that

$$y(1) = f_1(a_1) + ... + f_p(a_1) + \mu + v(1)$$
$$y(p+1) = f_1(a_2) + ... + f_p(a_1) + \mu + v(p+1)$$
$$y((m-1)p+1) = f_1(a_m) + ... + f_p(a_1) + \mu + v((m-1)p+1)$$
$$y(p) = f_1(a_1) + ... + f_p(a_1) + \mu + v(p)$$
$$y(p+p) = f_1(a_1) + ... + f_p(a_2) + \mu + v(p+p)$$
$$y((m-1)p+p) = f_1(a_1) + ... + f_{p-1}(a_1) + f_p(a_m) + \mu + v((m-1)p+p)$$

To further simplify notation, let

$$\mu_j = \sum_{l=1, l\neq j}^{p} f_l(a_1) + \mu,\ j = 1, 2, ..., p$$

and

$$y_{ijk} = y((k-1)mp + (i-1)p + j),$$
$$v_{ijk} = v((k-1)mp + (i-1)p + j) \tag{II.4}$$

The outputs $y(t)$ at $t = (k-1)mp + (i-1)p + j$ can be expressed by

$$y_{ijk} = f_j(a_i) + \mu_j + v_{ijk} \tag{II.5}$$

for $k = 1, 2, ..., n$, $i = 1, 2, ..., m$ and $j = 1, 2, ..., p$.

Clearly, the index $j = 1, ..., p$ separates the contributions of $f_j(\cdot)$'s, the index $i = 1, ..., m$ covers input partitions a_i's and $k = 1, ..., n$ is to average out the effect of the noise. Note $u(t)$ is periodic with period mp and the total length of $u(t)$ is nmp, i.e., n determines how many periods are used for identification. To have a better resolution for the input range I, m has to be increased and to reduce the effect of the noise, n needs to be increased. The input length $N = npm$ is linear in m, n and p. This linear property is a property of the additive model and is quite different from a general nonlinear system where N grows exponentially as pnm^p [9]. This difference is crucial because the order is not known in advance and we have to assume a large upper bound p to begin with.

It should be remarked that the initial condition assumption (II.3) does not impose any restriction to the system at all. If the actual initial condition is different or unknown, we may assign

$$u(1) = u(2) = ... = u(p-1) = a_1$$

and then, reset the time index $\tilde{t} = t - p + 1$.

III. REGRESSOR SELECTION

The order determination problem amounts whether to include one more delayed input variable in the system. Since no structural information is available, a method for order determination without requiring the knowledge of f_j's has to be developed for the nonlinear additive model (I.1). Our idea in this paper is not to tackle the order determination problem directly but to investigate which f_j's contribute to the output. This is actually a regressor selection problem. The term f_j that contributes to the output with the maximum j gives rise to the order. We discuss three approaches from an intuitive visual inspection method to sophisticated statistical approaches.

At this point, we comment that identification of the nonlinear additive model (I.1) is actually ill-defined. Note that if for any $l \neq j$, we replace $f_j[u(t-j+1)]$ and $f_l[u(t-l+1)]$ by $f_j[u(t-j+1)] + \bar{\mu}$ and $f_l[u(t-l+1)] - \bar{\mu}$ respectively for any constant $\bar{\mu}$, the input-output measurements of the system remain the same. In other words, the DC offset in each f_j is not identifiable without normalizations. For this purpose, we make the following assumption.

Assumption 3.1: Consider the nonlinear additive model (I.1) and the partition (II.1). Assume that for each $j = 1, 2, ..., p$,

$$\frac{1}{m}\sum_{i=1}^{m} f_j(a_i) = 0$$

Now, we define 4 averages: the grand average $\bar{y}_{...}$ of all observations, the average $\bar{y}_{ij.}$ of the ith level of the jth function f_j, the average $\bar{y}_{.j.}$ of the jth function f_j, and the average $\bar{y}_{i..}$ of the ith level,

$$\begin{array}{lcl} \bar{y}_{...} & = & \frac{1}{nmp}\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{p} y_{ijk} \\ \bar{y}_{ij.} & = & \frac{1}{n}\sum_{k=1}^{n} y_{ijk} = f_j(a_i) + \mu_j + \frac{1}{n}\sum_{k=1}^{n} v_{ijk} \\ \bar{y}_{.j.} & = & \frac{1}{m}\sum_{i=1}^{m} \bar{y}_{ij.} = \frac{1}{nm}\sum_{i=1}^{m}\sum_{k=1}^{n} y_{ijk} \\ \bar{y}_{i..} & = & \frac{1}{p}\sum_{j=1}^{p} \bar{y}_{ij.} = \frac{1}{pn}\sum_{j=1}^{p}\sum_{k=1}^{n} y_{ijk} \end{array} \tag{III.1}$$

where the dot "." subscript implies the average with respect to the subscript it replaces.

A. Visual inspection

μ_j is constant when i is varied. Whether f_j contributes to the output can be visually inspected if the graph of $f_j(a_i)$ vs a_i is available. Since $f_j(a_i)$ is unknown, we need to have an estimate of $f_j(a_i)$. To this end, let

$$z_{ij} = \bar{y}_{ij.} - \bar{y}_{.j.} = f_j(a_i) + \frac{1}{n}\sum_{k=1}^{n}(v_{ijk} - \frac{1}{m}\sum_{i=1}^{m} v_{ijk}) \quad \text{(III.2)}$$

Theorem 3.1: Consider the above equation.

1)
$$\frac{1}{n}\sum_{k=1}^{n}(v_{ijk} - \frac{1}{m}\sum_{i=1}^{m} v_{ijk}) \text{ and } \frac{1}{n}\sum_{k=1}^{n}(v_{ilk} - \frac{1}{m}\sum_{i=1}^{m} v_{ilk})$$
are independent if $j \neq l$.

2)
$$\mathbf{E}\frac{1}{n}\sum_{k=1}^{n}(v_{ijk} - \frac{1}{m}\sum_{i=1}^{m} v_{ijk}) = 0,$$
$$\mathbf{E}[\frac{1}{n}\sum_{k=1}^{n}(v_{ijk} - \frac{1}{m}\sum_{i=1}^{m} v_{ijk})]^2 = \frac{1}{n}(1 - \frac{1}{m})\sigma_v^2$$

3) $z_{ij} \to f_j(a_i)$ for every i and j in probability as $n \to \infty$.

Proof: The first one is trivial because v_{ijk} and v_{ilk} are independent. The second part shows the convergence in the mean squares which implies the convergence in probability of the third part. Thus, what we have to show is the second part. Note

$$\mathbf{E}\frac{1}{n^2}\sum_{k=1}^{n}(v_{ijk} - \frac{1}{m}\sum_{i=1}^{m} v_{ijk})\sum_{l=1}^{n}(v_{ijl} - \frac{1}{m}\sum_{d=1}^{m} v_{djl})$$
$$= \sigma_v^2(n/n^2 - 2/(mn) + 1/(nm)) = \frac{1}{n}(1 - \frac{1}{m})\sigma_v^2$$

This completes the proof.

The implication of the above result is that z_{ij} is computable based on the input-output measurements and converges to $f_j(a_i)$. Therefore, an estimate of the graph of $f_j(a_i)$ vs a_i is obtained by the graph of z_{ij} vs a_i by varying i for large n. Accordingly, the contribution of $f_j(a_i)$ can be visually inspected by the graph of z_{ij} vs a_i.

B. Relative contributions

The above idea can be made precise. Again, note that the average of $f_j(a_i)$ with respect to a_i is zero and hence, the magnitude $\sum_{i=1}^m f_j^2(a_i)$ is an indication of how much the term $f_j(\cdot)$ contributes to the output. If $f_j(a_i)$ does not contribute, i.e., $f_j(a_i) = 0$, $i = 1, ..., m$, $\sum_{i=1}^m f_j^2(a_i) = 0$. On the other hand, if $\sum_{i=1}^m f_j^2(a_i) \neq 0$, then some of $f_j(a_i)$'s must be non-zero. Keep in mind however what we are really interested is not whether $f_j(\cdot)$ contributes or not, but whether the contribution is significant or not. In other words, a relative contribution is more appropriate for order determination. This can be measured by calculating the ratio

$$\frac{\sum_{i=1}^m f_j^2(a_i)}{\sum_{l=1,l\neq j}^p \sum_{i=1}^m f_l^2(a_i)} \quad \text{(III.3)}$$

The numerator is the contribution of f_j and the denominator is the contribution of the rest terms. This ratio quantifies the relative contribution of f_j compared to the sum of the rest terms. Now the question is how to find this ratio which is not available. To this end, we define

$$\begin{aligned} SS_j &= n\sum_{i=1}^m(\bar{y}_{ij.} - \bar{y}_{.j.})^2, \ j = 1, 2, ..., p \\ SS_E &= \sum_{k=1}^n \sum_{i=1}^m \sum_{j=1}^p (y_{ijk} - \bar{y}_{ij.})^2 \end{aligned} \quad \text{(III.4)}$$

and based on SS_j, we further define the relative contribution index

$$R_j = \frac{\frac{1}{n(m-1)}SS_j}{\frac{1}{n(m-1)}\sum_{l=1,l\neq j}^p SS_l}, \ j = 1, ..., p \quad \text{(III.5)}$$

Theorem 3.2: Consider the relative contribution index R_j. If $\sum_{l=1,l\neq j}^p \sum_{i=1}^m f_l^2(a_i) \neq 0$, then

$$R_j \to \frac{\sum_{i=1}^m f_j^2(a_i)}{\sum_{l=1,l\neq j}^p \sum_{i=1}^m f_l^2(a_i)}$$

in probability as $n \to \infty$.

Proof: SS_j and SS_l are independent for $j \neq l$. Now, from

$$\begin{aligned} \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.}) &= \sum_{k=1}^n y_{ijk} - n\bar{y}_{ij.} = 0 \\ \sum_{i=1}^m (\bar{y}_{ij.} - \bar{y}_{.j.}) &= \sum_{i=1}^m \bar{y}_{ij.} - m\bar{y}_{.j.} = 0 \\ \sum_{j=1}^p (\bar{y}_{ij.} - \bar{y}_{i..}) &= \sum_{j=1}^p \bar{y}_{ij.} - p\bar{y}_{i..} = 0 \end{aligned}$$

it is easily verified that

$$\begin{aligned} &\mathbf{E}\frac{SS_E}{mp(n-1)} = \mathbf{E}\frac{\sum_{k=1}^n \sum_{i=1}^m \sum_{j=1}^p (y_{ijk} - \bar{y}_{ij.})^2}{mp(n-1)} = \sigma_v^2 \\ &\mathbf{E}[\frac{SS_E}{mp(n-1)} - \sigma_v^2]^2 \sim O(\frac{1}{n}) \ as \ n \to \infty \\ &\mathbf{E}\frac{SS_j}{m-1} = \mathbf{E}\frac{n\sum_{i=1}^m (\bar{y}_{ij.} - \bar{y}_{.j.})^2}{m-1} = \sigma_v^2 + \frac{n\sum_{i=1}^m f_j^2(a_i)}{m-1} \\ &\mathbf{E}[\frac{SS_j}{n(m-1)} - (\sigma_v^2/n + \frac{\sum_{i=1}^m f_j^2(a_i)}{m-1})]^2 \sim O(\frac{1}{n}) \ as \ n \to \infty \end{aligned} \quad \text{(III.6)}$$

Thus, $\frac{SS_j}{n(m-1)} \to \frac{1}{(m-1)}\sum_{i=1}^m f_j^2(a_i)$ in probability. This implies that the numerator of R_j converges to $\frac{1}{(m-1)}\sum_{i=1}^m f_j^2(a_i)$ and the denominator of R_j converges to $\frac{1}{(m-1)}\sum_{l=1,l\neq j}^p \sum_{i=1}^m f_l^2(a_i)$. Further, $\sum_{l=1,l\neq j}^p \sum_{i=1}^m f_l^2(a_i) > 0$ implies that R_j is a continuous function of the numerator and denominator. Now, from Slutsky's theorem [4] and the fact that convergence in probability and convergence in distribution are equivalent if the limit is a constant, we have

$$R_j \to \frac{\frac{1}{(m-1)}\sum_{i=1}^m f_j^2(a_i)}{\frac{1}{(m-1)}\sum_{l=1,l\neq j}^p \sum_{i=1}^m f_l^2(a_i)}$$

Though some equations in (III.6) were not used in the convergence proof of R_j, they are needed later. Also, as an by-product, the result tells us that $\frac{SS_E}{mp(n-1)}$ actually provides an unbiased and consistent estimate of the unknown noise variance σ_v^2.

To determine whether the term f_j contributes, we compute R_j, $j = 1, 2, ..., p$. Let the threshold d, for example $d = 0.05$ or 5%, be chosen. If $R_j \geq d$, we say the term f_j contributes and otherwise, f_j does not contribute. Since R_j converges to the quantity in (III.3), the test is very reliable for large n. For small n, an improvement can be made. Note that

$SS_j/n(m-1)$ is a biased estimate of $\frac{1}{m-1}\sum_{i=1}^m f_j^2(a_i)$ and the bias term $\frac{\sigma_v^2}{n}$ goes to zero as n gets large. To compensate this bias term for small n, we may subtract an unbiased and consistent estimate of σ_v^2/n from $SS_j/n(m-1)$ and redefine R_j as

$$\frac{SS_j/[n(m-1)] - SS_E/[nmp(n-1)]}{\sum_{l=1,l\neq j}^p \{SS_l/[n(m-1)] - SS_E/[nmp(n-1)]\}} \quad \text{(III.7)}$$

So far no assumption on the probability distribution of the noise is assumed and the obtained results are asymptotic in nature. To be able to test R_j statistically for not so large n, additional information on the unknown noise is needed. For instance, if the noise is assumed to be Gaussian, $SS_j/n(m-1)$ is (non-central) chi-square distributed and then, hypotheses test or the confidence internal calculation becomes feasible. This is close to the idea of the analysis of variance that will be discussed in the next section.

C. The analysis of variance

The analysis of variance (ANOVA) is a powerful tool in statistics [3], [12] and was first introduced to system identification in [9], [10] where a nonlinear FIR system was examined. The ANOVA enumerates all possible combinations of the input lags and compares their contributions by some statistical measures. Thus, the input length and the computational complexity are high and grow exponentially as the order increases referred to as the curse of dimensionality [6], [9] that limits the use of the ANOVA to very low order systems. By taking advantages of the form of the additive model (I.1) and carefully designing the input signals as in (II.2), however, the curse of the dimensionality is no longer a problem. In fact, the input length and the computational complexity are linear functions of n, m and p for the nonlinear additive model (I.1) as discussed before.

Similar to the previous section, the idea of the ANOVA is to study the variance of y_{ijk} for fixed j by varying i. If f_j has no contribution, the variance is the same as the noise variance σ_v^2. If the variance is significantly larger than σ_v^2, f_j likely contributes to the output.

Strictly speaking, to carry out the ANOVA, the noise has to be assumed to be Gaussian. It has been reported in the literature [7], [9], [10] that, however, the ANOVA is robust even when this assumption is not satisfied. Before presenting the results, we define two variables.

$$\begin{aligned} SS_T &= \textstyle\sum_{k=1}^n \sum_{i=1}^m \sum_{j=1}^p (y_{ijk} - \bar{y}_{...})^2, \\ SS_R &= mn \textstyle\sum_{j=1}^p (\bar{y}_{.j.} - \bar{y}_{...})^2 \end{aligned}$$

Theorem 3.3: Consider the nonlinear additive model (I.1) and the input (II.2). Assume the noise is i.i.d. Gaussian with zero mean and variance σ_v^2. Then, the random variables SS_T, SS_E, SS_R and SS_j, $j = 1, 2, ..., p$ defined in (III.4) satisfy

$$SS_T = SS_E + SS_1 + ... + SS_p + SS_R \quad \text{(III.8)}$$

and SS_E, SS_R and SS_j, $j = 1, 2, ..., p$ are statistically independent. Moreover, $SS_E \sim \chi^2(mp(n-1))$ and if $f_j(a_i) = 0$, $i = 1, 2, ..., m$, for some j, $SS_j \sim \chi^2(m-1)$ and consequently

$$F_j = \frac{SS_j/(m-1)}{SS_E/[mp(n-1)]} \quad \sim F(m-1, mp(n-1))$$

i.e., when $f_j(a_i) = 0$, $i = 1, 2, ..., m$, F_j is F-distributed with $(m-1)$ and $mp(n-1)$ degrees of freedom.

Proof: Equation (III.8) is from (III.6) and the independence is from the Cochran theorem, see page 69 of [12]. The rest part comes from the definitions of χ^2 and F distributions [12].

Since $SS_j/(m-1)$ is an estimate of $\frac{n}{m-1}\sum_{i=1}^m f_j^2(a_i) + \sigma_v^2$ and $SS_E/[mp(n-1)]$ is an estimate of σ_v^2, an implication of the results is that if f_j does not contribute to the output, i.e., $f_j(a_i) = 0$, $i = 1, 2, ..., m$, $F_j \approx \sigma_v^2/\sigma_v^2 = 1$. If f_j does contribute, $F_j > 1$. Mathematically, this can be put into a form of hypotheses test for each $j = 1, 2..., p$,

$$\begin{array}{rll} & H_{0,j}: & f_j(a_i) = 0, \ \forall i \\ \text{against} & H_{1,j}: & \text{at least one of } f_j(a_i) \neq 0 \end{array} \quad \text{(III.9)}$$

If $H_{0,j}$ is rejected, the output does depend on f_j. To test the hypotheses, we calculate F_j, $j = 1, ..., p$ based on the measured input-output data. Let the threshold d be taken from $F_\alpha(m-1, mp(n-1))$-distribution table, where α denotes the level of significance, i.e., the probability to reject $H_{0,j}$ though $H_{0,j}$ is true. The hypothesis is rejected if $F_j > d$ and we conclude that f_j does contribute.

Once the tests are carried out for F_j, $j = 1, 2, ..., p$, we have determined which f_j contributes to the output and the maximum j so that f_j contributes to the output is the system order.

We comment that the hypotheses test is done in terms of the probability

$$Prob\{\text{reject } H_{0,j} \ : \ H_{0,j} \text{ is true}\}$$

This is often referred to as the type I error. The test can also be done in terms of the type II error

$$Prob\{\text{reject } H_{0,j} : H_{0,j} \text{ is false}\} = 1 -$$

$$Prob\{\text{fail to reject } H_{0,j} \ : \ H_{0,j} \text{ is false}\}$$

This test is based on non-central F distributions and is more involved. Interested readers can find details from [12].

D. Discussion and simulation

Three order determination methods are proposed and each one has its own advantage. The visual inspection approach is very intuitive and simple. It may also reveal the structures of the unknown nonlinearities graphically. The relative contribution index R_j accurately measures the relative contribution that is often more important than the individual contribution in order determination. The ANOVA approach is a robust approach but does not measure relative contributions.

We now consider a numerical example of the form of a generalized Hammerstein model (I.2) with $b_1 = b_{11} + b_{21}z^{-1}$, $b_2 = b_{12} + b_{22}z^{-1}$, $\bar{f}_1 = u(t)$ and $\bar{f}_2 = u(t)^2$ or equivalently

$$y(t) = \underbrace{b_{11}u(t) + b_{12}u(t)^2 - \mu_1}_{f_1(u(t))}$$

$$+\underbrace{b_{21}u(t-1)+b_{22}u(t-1)^2-\mu_2}_{f_2(u(t-1))}$$
$$+\underbrace{0}_{f_3(u(t-2))+f_4(u(t-3))+f_5(u(t-4))}+\underbrace{\mu_1+\mu_2}_{\mu}+v(t)$$

The purpose of μ_i is to make the average of $f_i(\cdot)$ zero. The above system is frequently encountered in communications. For instance, it describes nonlinear distortions due to amplifiers in satellite link communication [14] and in ADSL [17].

In the simulation, the structures of f_i's as well as the exact order are unknown and only an upper bound $p = 5$ is assumed. The input range is $I = [-a, a] = [-2, 2]$ with $m = 11$ and

$$a_i = (i-1)\cdot 0.4 - 1, \quad i = 1, 2, ..., 11(= m)$$

The noise $v(t)$ is i.i.d. Gaussian with zero mean and $\sigma_v^2 = 0.25$, and $n = 100$ that implies $mp(n-1) = 5445$. For $b_{11} = 1$, $b_{12} = 0.7$, $b_{21} = 0.8$ and $b_{22} = -0.6$, Figure 1 shows the simulation results of z_{ij} $(= f_j(a_i) + \frac{1}{n}\sum_{k=1}^{n}(v_{ijk} - \frac{1}{m}\sum_{i=1}^{m} v_{ijk}))$ vs a_i, for $j = 1, 2, 3, 4, 5$. R_j and F_j, $j = 1, 2, 3, 4, 5$, are listed in Table 1. Further, the significance level or the probability to reject H_{0j} though H_{0j} is true is taken to be 0.01 which results in $F_{0.01}(10, 5445) = 2.3242$.

The results shown in the figure are the averages of 100 Monte Carlo simulations. The ranges of R_j and F_j for 100 Monte Carlo simulations are in the table. By visually inspecting Figure 1, it is clear that f_1 and f_2 contribute while f_3, f_4 and f_5 are almost identically zero and do not contribute. We conclude that the order is $2 - 1 = 1$. If the relative contribution index R_j is used, we notice that $R_3,\ R_4$ and $R_5 \approx 0$ and this implies that f_3, f_4 and f_5 do not contribute while f_1 and f_2 do. Therefore, the order is $2 - 1 = 1$. Finally, $F_1,\ F_2 > F_{0.01}(10, 5445) = 2.3242$ and $F_3,\ F_4,\ F_5 < 2.3242$ so the ANOVA method also concludes that f_1 and f_2 contribute and f_3, f_4 and f_5 do not. For this example, all three methods reach the same conclusion.

IV. Concluding remarks

In this paper, we have studied identification of nonlinear additive models. We feel however some ideas presented in the paper can be extended to identification of other nonlinear systems. In particular, the partition of the input range and the adoption of the relative contribution index may provide a fresh way of thinking in order estimation for nonlinear systems.

References

[1] Chen, B. S.A. Billings and W. Luo (1989), "Orthonormal least squares methods and their application to nonlinear system identification", *Int. J. Control*, **50**, pp.1873-1896

[2] Eubank, R. (1999), Nonparametric Regression and Spline Smoothing, 2nd Edition, Marcel Dekker, New York, NY

[3] Fan, J. and Q. Yao (2003), Nonlinear Time Series: Nonparametric and Parametric Models, Springer, New York, NY

[4] A. Gut, An Intermedia Course in Probability, Springer-Verlag, New York, 1995

[5] Haber, R. and L. Keviczky (1999), Nonlinear System Identification: Input-Output Modeling Approach, Vols. 1 and 2, Kluwer Academic Publisher, Dordrecht, The Netherlands

[6] Hastie, T. and R. Tibshirani (1997), Generalized Additive Models, Chapman & Hall, London

[7] Krishnaiah, P. (1980), HANDBOOK OF STATISTICS, **1**, North-Holland, Amsterdam

[8] Krzyzak, A. (1992), "Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation", *IEEE Trans. on Info. Theory*, **38**, pp.1323-1338

[9] Lind, I. (2000), "Model order selection of NFIR models by the analysis of variance method", *Proc. IFAC Symposium SYSID 2000*, pp.367-372

[10] Lind, I. and L. Ljung (2003), "Structure selection with ANOVA: local linear models", *Proc. IFAC Symposium SYSID 2003*, pp.51-56

[11] Ljung, L. (1999) System Identification: Theory for the Users, 2nd Ed. Prentice-Hall, Upper Saddle River

[12] Montgomery, D.G. (2001), Design and Analysis of Experiments, 5th Edition, John Wiley and Sons, New York, NY.

[13] Pawlak, M. (1991), "On the series expansion approach to the identification of Hammerstein systems", *IEEE Trans. on Automatic Control*, **36**, pp763-767

[14] S. Prakriya, D. Hatzinakos, "Blind identification of linear subsystems of LTI-ZMNL-LTI models with cyclostationary inputs", *IEEE Transactions on Signal Processing*, Vol.45, pp.2023 -2036, 1997

[15] Soderstrom, T. and P. Stoica (1989), System Identification, Prentice-Hall, New York, NY

[16] Wahba, G. (1990), Spline Models for Observational Data, SIAM, Philadelphia, PA

[17] Zojer, B, R. Kolan, J. Pichler and G. Pailo (2000), "A broadband high-voltage slic for a splitter and transformless combined ADSL-Lite-POTS linecard", *IEEE JSSC*, **35**, pp1976-1987

j	1	2	3	4	5
R_j	$[1.3703, 1.5850]$	$[0.6268, 0.7249]$	$[6 * 10^{-5}, 0.0012]$	$[9 * 10^{-5}, 0.0012]$	$[0.0001, 0.0015]$
F_j	$[1067, 1211]$	$[713, 814]$	$[0.2688, 2.3029]$	$[0.2418, 2.2649]$	$[0.2282, 1.9707]$
$F_{0.01}(10, 5445) = 2.3242$					

TABLE I
THE RANGES OF R_j, F_j AND $F_{0.01}(10, 5445)$

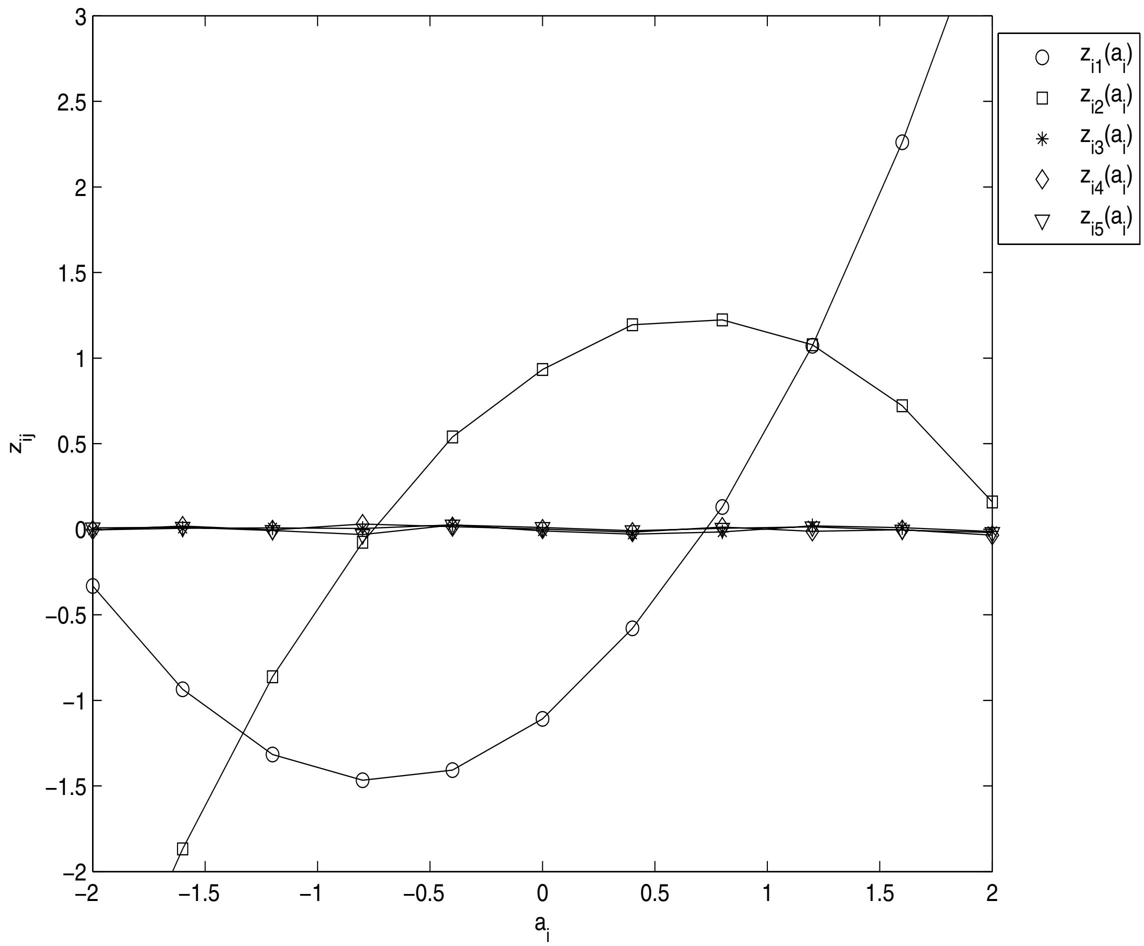

Fig. 1. z_{ij} vs a_i, $j = 1, 2, 3, 4, 5$.