2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

An Improved Algebraic Geometric Solution to the Identification of Switched ARX Models with Noise

Sohail Nazari[1], Qing Zhao[1*], Biao Huang[2]

Abstract — **In this paper, we present an improved algebraic geometry solution for the identification of switched ARX models in the presence of measurement noise. The procedure utilizes the highest order of sub-models, which is estimated by using statistical analysis of effective singular values in matrix rank determination. After embedding sub-models into a large continuous-time model for omitting the necessity of switching sequence, an analytical solution for the two-mode system is obtained using matrix differential calculus. The improvements made to the previous method are verified by simulations on two linear systems. Also the effectiveness of the proposed method is shown by using a two mode experimental pilot plant.**

I. Introduction

The switching Auto Regressive eXogenous (SARX) model, also known as the Jump ARX model, is used to describe the behavior of certain hybrid systems. In this modeling, the hybrid system has different continuous time operational modes (sub-systems), and a discrete-valued variable switches the operating mode under certain conditions. Therefore the hybrid system is defined by interaction of continuous-time and discrete-valued dynamics. The discrete switching variable depends on other variables such as the time or the input characteristics. Existence of both continuous dynamics and discrete switching increases the complexity of the system control and identification problem.

In the literature, the hybrid system identification using SARX models can be categorized into four major groups: the clustering-based approach [1], the Bayesian approach [2], the bounded error approach [3], and the algebraic geometry approach [4]. The latter approach will be discussed in detail. Several reviews of these four different approaches can be seen in [5, 6].

One of the main advantages of the algebraic approach is that only an upper bound for the order of the sub-models needs to be assumed. Then a procedure estimating the actual maximum order of the sub-models is given prior to identification. After the order estimation, all sub-models are embedded into a large continuous time model, called the hybrid decoupling polynomial (HDP) by using geometric mapping methods. The next step is to estimate all parameters of this embedded model (finding coefficients of the HDP). Finally, a procedure is introduced to recover the parameters of each sub-model from the embedded mapped model. This method has been shown to work effectively in the absence of noise, but has some major drawbacks, such as difficulty in estimating the highest order of the sub-models in the presence of noise. In [4], an additional step is proposed to ensure that the method provides acceptable results even in the presence of noise. Another disadvantage of the AG algorithm is in recovering the parameter of each sub-model when the noise variance is high. In Vidal and his co-workers work [4, 7], where there is measurement noise, both the variance of the estimation error and the mean square error are high. .

In this paper, for estimating the highest order of the sub-models, a new method is introduced that depends on the signal to noise ratio (SNR) of the data rather than the system itself and is easy to tune. Furthermore, the utilization of total least square (TLS) regression in parameter estimation of the embedded model simplifies the original algorithm presented in [4]. After estimating the parameters for HDP, a new method is developed to find the parameters of each sub-model from the identified embedded model.

In this new method, the derivatives of the HDP with respect to the regressors are calculated, and the equations are formulated so they can be used in the framework of a linear regression problem. For a two-mode model a unique closed-form solution to this equation is derived.

This paper is organized as follows. Section II presents the modeling and system description, the problem statement and a new procedure for estimating the maximum order of the sub-models. Section III reports the proposed identification method of the SARX model as well as the new algorithm, while Section IV presents the simulation results, which illustrate the improvement made by the proposed algorithm. Section V concludes the paper.

II. Model Descriptions

A. SARX Model

The Switched ARX (SARX) model has the following representation;

Manuscript received September 22, 2010.

1 S. Nazari is with the Electrical and Computer Engineering Department, University of Alberta, Alberta, Canada. (e-mail: snazari@ ualberta.ca).

1 Q. Zhao is with the Electrical and Computer Engineering Department, University of Alberta, Alberta, Canada. (phone: +1 780 492-5792 fax: +1 780 492-1811 e-mail: qingz@ ualberta.ca), *: The corresponding author.

2 B. Huang is with the Chemical Engineering Department, University of Alberta, Alberta, Canada. (e-mail: biao.huang@ualberta.ca)

$$y_t = \sum_{j=1}^{n_\alpha(\lambda_t)} \alpha_j(\lambda_t) y_{t-j} + \sum_{j=1}^{n_\beta(\lambda_t)} \beta_j(\lambda_t) u_{t-j} + C(\lambda_t) + \omega_t \,, \quad (1)$$

where $u_t \in \mathbb{R}$ is the input, and $y_t \in \mathbb{R}$ is the output of the system. λ_t is the discrete state, also known as the mode of operation. In this paper, λ_t is considered as an unknown arbitrary sequence from the integer set, $\lambda_t \in \mathbb{Z} : \{1,2,\ldots,n\}$, and n is the total number of modes. ω_t is the additive white noise that has a Gaussian distribution with zero mean and δ^2 variance. Another representation of the SARX system is shown as follows:

$$SARX\ system : \begin{cases} b_1^T x_t = \varepsilon_1 \\ \vdots \\ b_n^T x_t = \varepsilon_n \end{cases}, \quad (2)$$

where $b_i = [1\, \alpha_1^i \alpha_2^i \cdots \alpha_{n_\alpha}^i \, \beta_1^i \cdots \beta_{n_c}^i \, c^i]^T$, and correspondingly, $x_t = [y_t \cdots y_{t-n_\alpha} \, u_{t-1} \cdots u_{t-n_\beta} \, 1]^T \in \mathbb{R}^K$. i is the discrete-valued state and ε_i is the noise. The identification problem is defined as the following.

Problem 1. Identification of the switched ARX model.

Given the input/output data $\{u_t, y_t\}_{t=0}^T$ from a hybrid ARX system (1), identify the maximum order of the ARX model ($\max n_\alpha(\lambda_t)$ *and* $\max n_\beta(\lambda_t)$), the model parameters $\{\alpha_j(i)\}_{j=1}^{n_\alpha(i)}$, $\{\beta_j(i)\}_{j=1}^{n_\beta(i)}$, c_i and the discrete-valued states $\{\lambda_t\}$.

The number of modes can be estimated as in [4], however, in this paper n is assumed to be known.

B. Embedded SARX Model

The main difference between the SARX model and the ARX model is the discrete switching sequence (λ_t), which significantly increases the complexity of the identification procedure. Since each data point belongs to a different mode, the linear regression cannot be used to estimate the parameters. This problem motivates us to eliminate the role of the switching sequence and to find a representation of (2) that is independent from i (modes of operation).The embedding of all sub-equations into one equation can be done, by taking the product of all sub-equations in (2). The outcome is the following polynomial equation:

$$P_n(x_t) = \prod_{i=1}^{n} b_i^T x_t = \varepsilon \quad (3)$$

where P_n is called the hybrid decoupling polynomial (HDP) [7]. Although taking the product is not the only way to eliminate the switching sequence, the use of the HDP leads to certain advantages in algebraic structure. The HDP is a multivariate polynomial of degree n with $K = \max n_\alpha(\lambda_t) + \max n_\beta(\lambda_t) + 1$ variables, which can be linearly written in terms of its coefficients as

$$P_n(x) = \prod_{i=1}^{n} b_i^T x = \sum_{I=1}^{M_n(K)} h_{n1,\ldots,nk} X_I = H_n^T v_n(x) = \varepsilon \,, \quad (4)$$

where, $h_{n1,\ldots,nk} \in \mathbb{R}$ is the coefficient of the monomial $X_I = x_1^{n1} \ldots x_k^{nk}$, where $0 \le n_j \le n \;\; j = 1 \ldots K$; $n_1 + \ldots + n_k = n$. $v_n : \mathbb{R}^K \to \mathbb{R}^{M_n(K)}$ is a Veronese map of degree n, which is defined as [9] $v_n : [x_1, \ldots, x_K]^T \to [\ldots, X_I, \ldots]^T$ with I chosen in the degree-lexicographic order, e.g. $v_2([x_1 x_2 x_3]) = \left[x_1^2 \;\; x_1x_2 \;\; x_1x_3 \;\; x_2^2 \;\; x_2x_3 \;\; x_3^2\right] = \left[X_1 X_2 X_3 X_4 X_5 X_6\right]$; and $M_n(K) = \binom{n+k-1}{n}$ is the total number of independent monomials in (4).

Equations (3) and (4) should hold for all the data points and therefore, the following equation becomes a linear ARX model with the parameter vector H_n:

$$L_n(n_\alpha, n_\beta) H_n = [v_n(x_{\max(n_\alpha, n_\beta)}) \ldots v_n(x_{\max(n_\alpha, n_\beta)+T-1})]^T H_n = \varepsilon_{T\times 1}, \quad (5)$$

where T is the number of data points, and $L_n(n_\alpha, n_\beta) \in \mathbb{R}^{T \times M_n(K)}$ is the matrix of the embedded and mapped input/output data via the Veronese variety. Note that in this linear model, the switching sequence is eliminated completely.

C. Determination of highest order of sub-models

The main advantage of the AG approach is that, the orders of sub-models are not needed for the identification process, only the highest order of sub-models ($\max n_\alpha(\lambda_t)$ and $\max n_\beta(\lambda_t)$) need to be estimated.

The following theorem developed in [7] estimates the maximum order of the ARX sub-models.

Theorem 1. (Identifying the maximum orders)

By using the input/output data from the SARX system and $L_n(i,j) \in \mathbb{R}^{T \times M_n(i+j+2)}$, if T is sufficiently large and the input along with the switching sequences are adequately excited, then the maximum order of SARX models can be calculated by:

$$(n_\alpha, n_\beta) = \underset{(i,j): M_n(i+j+1) < T}{\arg\min} \{(i+j\} : rank(L_n(i,j) < M_n(i+j+1)\} \quad (6)$$

The results obtained by Theorem 1 are precise in the deterministic case, but when the data is noisy, the rank of the L_n matrix cannot be determined easily and uniquely. More likely $L_n(i,j)$ has full rank with the noisy data. Therefore, instead of using the rank of L_n, the effective rank should be considered. In [7], the rank of a noisy matrix is obtained through the following minimization problem:

$$Rank(L) = \underset{j=1,\ldots,k}{\arg\min} \left\{ \frac{\sigma_j^2(L)}{\sum_{i=1}^{j-1} \sigma_i^2(L)} + \kappa . j \right\} \quad (7)$$

where σ is a singular value of L, and $\kappa \in \mathbb{R}$ is a weighting parameter used to penalize the choice of higher ranks. This method for the rank estimation method has some problems, especially in selecting the parameter κ. No rules for choosing κ are provided in [4]. The authors only suggest a low limit for κ without giving any justification and as a result the choice of κ involves trial and error. It will be shown in section IV that the suggested low limit works only

for the suggested system used in [4]. For other systems presented in this paper the range for correct κ is extremely narrow. Since this method is not reliable due to its trial and error nature, we are motivated to find an alternative method whose results are independent from the specific system.

The above problem can be approached by alternative method for calculating the rank of a noisy matrix. In [11], a method is developed by using statistical analysis for finding the effective singular values in the matrix rank determination. The definition of the problem given is as follows.

Problem 2. Finding an effective rank for a noisy matrix.

For the equation $B = A + E$ in which all the matrices have the dimension of $m \times n$, the problem is to find the effective rank of B (say r) so that the r[th] singular value of B is greater than the 2-norm of E.

$\beta_r > \varepsilon > \beta_{r+1}$, $\varepsilon = \| E \|_2$,

where each element in the matrix E is an i.i.d. random variables with a Gaussian distribution with zero mean and a variance of σ^2 , and β_r is r[th] singular value of the matrix B. In [11], several different bounds for ε are introduced. One of these bounds uses chi-square (χ^2) distribution properties as follow: $S^2 = \| e_j \|_2^2 = \sum_{i=1}^{m} | e_{ij} |^2$, so that S^2/σ^2 has a χ^2 distribution with m degrees of freedom. Following that, for a given level of significance α, a constant *c* can be found such that $prob\left(S^2/\sigma^2 > c\right) \le \alpha$. The bounds for ε are:

$$\sqrt{c}\sigma \le \varepsilon \le \sqrt{nc}\sigma \tag{8}$$

By using the right hand side of the above inequality, the effective rank of matrix B with respect to the noise matrix E can be found. Notice that *c* is determined from the table of the chi-square distribution with *m* degrees of freedom and the given level of significance. The choice of the level of significance depends on the SNR of the signal [11].

The rank of matrix $L_n(i,j)$ used in Theorom 1 can then be found by using (8) with $n = T$, $m = M_n(K)$. As will be shown in the simulation section, the tuning factor *c* does not depend on the system anymore if the above method is used. Also, it will be shown that this method is more robust to the noise than the method used by Ma and Vidal [4].

III. Estimation of Model Parameters

As described in Problem 1, the parameters for each sub-model in (2) should be estimated. In the algebraic approach, two major steps are taken to reach this goal. In the first step, the coefficients (H_n) for each monomial in the HDP constructed in (4) are estimated. In the next step, the estimated coefficients are used to recover the parameters of the sub-models (b_i).

First, the estimations of the parameters (H_n) for the embedded linear model (5) is discussed. This estimation is a standard regression problem where both the input/output data are corrupted with noise. Therefore instead of using the Least Square method, the Total Least Square method [8] should be used. TLS solution is the minimization solution to the following cost function:

$$X_{TLS} = \arg\min_X \| [\Delta A \ \Delta B] \| \text{ subject to } \left(A^{m\times n} + \Delta A\right) X = B^{m\times d} + \Delta B.$$

Where ΔA and ΔB are error in variables for the system and represent the noise in our problem set. In this paper singular value decomposition (SVD) method for deriving TLS solution is used. Also, the first column of L_n matrix is used as *B* matrix and the rest of L_n is used as *A* matrix.

A. Parameter estimation of the SARX model (b_i) – deterministic case

The main difficulty of the algebraic geometry method is to recover the model parameters (b_i) from the estimated coefficients of HDP (H_n). Vidal et al. [4] suggested a method for solving this problem by using the derivatives of HDP in the deterministic situation:

$$DP_n(x) = \frac{\partial P_n(x)}{\partial x} = \frac{\partial \prod_{i=1}^{n} b_i^T x}{\partial x} = \sum_{i=0}^{n} b_i \prod_{l\neq i}^{n} b_l^T x. \tag{9}$$

If x belongs to i[th] sub-model, then $b_i^T x = 0$, and since the first element of each b_i is equal to 1, each parameter set can be found uniquely through the following equation:

$$b_j = \frac{DP_n(x_j)}{(e_1^T DP_n(x_j))} . \tag{10}$$

where x_j is the data point belonging the j^{th} sub-model. In this method, at least one data-point is needed for each sub-model. The procedure involves finding the closest data-point to sub-models in presence of noise. In [7], by using the first-order approximation to smallest distance of data points and hyperplanes representing the sub-models the following lemma is presented:

Lemma 1. The closest point from the data set to one of the sub-spaces can be chosen by

$$x_n = \arg\min_{x\in\{x_i\}} \frac{|P_n(x)|}{\left\| \left(I - e_1 e_1^T\right) DP_n(x) \right\|}. \tag{11}$$

Now that one point near one sub-model has been found, the parameters corresponding to that sub-space can be calculated by using (10). In order to find another point (x_{n-1}) in the remaining sub-spaces, (11) should be penalized for choosing a point from n[th] hyperplane:

$$x_{n-1} = \arg\min_{x\in\{x_i\}} \frac{|P_n(x)|}{\left| \left\| \left(I - e_1 e_1^T\right) DP_n(x) \right\| \right| \left| b_n^T x \right|} \tag{12}$$

Repeating this procedure will result in finding all the sub-model parameters. This procedure works well for data with small noise variances, but when the variance of the noise increases, this algorithm does not provide accurate results due to the sub-optimality of (12) and can result in a the local minimum. Even when the appropriate data point is found, equation (10) holds only in the absence of the noise.

B. Parameter estimation of SARX model (b_i) via Element-

wise approach

In order to overcome the problems addressed in the previous section, we propose an approach that avoids finding the data point for each sub-model. In fact, in this approach, all data points are utilized without being clustered and results in a closed-form solution for the two-mode models. We name this as element-wise approach, in the sense that the parameters of the original sub-models are solved one by one in an analytical form. The derivation is as follows.

Consider the two-mode model to be $\begin{cases} b_1^T x_t = \varepsilon_1 \\ b_2^T x_t = \varepsilon_2 \end{cases}$, and the corresponding HDP is: $p_n(x_t) = (b_1^T x_t) \times (b_2^T x_t)$.

The derivative with respect to x_t is

$$DP_n(x_t) = \frac{\partial P_n(x_t)}{\partial x_t} = b_1(b_2^T x_t) + b_2(b_1^T x_t). \quad (13)$$

$DP_n(x_t)$ is a vector since x_t is a vector. In order to fit the problem into a linear regression problem, one can consider each element of the above equation:

$$D_i P_n(x_t) = b_{1i}(b_2^T x_t) + b_{2i}(b_1^T x_t),$$

in which the i^{th} element of $DP_n(x_t)$ is $D_i P_n(x_t)$, and the j^{th} element of b_i is b_{ij}. Now, starting with the first element, and since previously assumed $b_{11} = b_{21} = 1$, we have

$$D_1 P_n(x_t) = (b_2^T x_t) + (b_1^T x_t) = (b_1^T + b_2^T) x_t = x_t^T (b_1 + b_2).$$

Now, using all the data points in the above equation results in

$$\begin{bmatrix} D_1 P_n(x_1) \\ \vdots \\ D_1 P_n(x_T) \end{bmatrix} = \begin{bmatrix} x_1^T \\ \vdots \\ x_T^T \end{bmatrix} (b_1 + b_2) = L(b_1 + b_2),$$

which is a linear regression problem with regression vector L, and the parameter vector to be estimated is $(b_1 + b_2)$. Since $D_1 P_n(x_t)$ is noisy, and some elements of x_t also have noise, the TLS method will be used for parameter estimation and $(b_1 + b_2)$ can be estimated. Let

$$(b_1 + b_2) = A^1. \quad (14)$$

Using (13) in (14) we have

$$DP_n(x_t) = b_1((A^1 - b_1)^T x_t) + (A^1 - b_1)(b_1^T x_t).$$

Using the second element of the above equation, we obtain

$$D_2 P_n(x_t) = b_{12}((A^1 - b_1)^T x_t) + (A_2^1 - b_{12})(b_1^T x_t)$$

$$= \left(b_{12}(A^1 - b_1)^T + (A_2^1 - b_{12}) b_1^T\right) x_t = x_t^T \left(b_{12}(A^1 - b_1) + (A_2^1 - b_{12}) b_1\right).$$

As done previously, all the data points are used in the above equation and by using the SVD and TLS, we can estimate the following:

$$\left(b_{12}(A^1 - b_1) + (A_2^1 - b_{12}) b_1\right) = A^2, \quad (15)$$

The goal is to find b_1. Based on the second element of each vector in (15), we have

$$A_2^2 = \left(b_{12}(A_2^1 - b_{12}) + (A_2^1 - b_{12}) b_{12}\right) = 2b_{12}A_2^1 - 2{b_{12}}^2.$$

This ends to ${b_{12}}^2 - A_2^1 b_{12} + \frac{1}{2} A_2^2 = 0$, which is a second-order equation with respect to b_{12}. The solution for b_{12} results in

$$b_{12} = \frac{A_2^1 \pm \sqrt{(A_2^1)^2 - 2A_2^2}}{2}. \quad (16)$$

Now that b_{12} is found, by using the third element of each vector in (15), we have $A_3^2 = \left(b_{12}(A_3^1 - b_{13}) + (A_2^1 - b_{12}) b_{13}\right)$. Then by solving for b_{13}, the above equation gives $b_{13} = \frac{A_3^2 - b_{12}A_3^1}{A_2^1 - 2b_{12}}$, and with the similar procedure the other parameters can be solved.

Remark 1. From the above procedure, b_1 is found, and b_2 can be found from (14). This method illustrates that for a two-mode model, the parameter estimates can be found without clustering any data point.

Remark 2. The proposed method does not increase the complexity with respect to the previous method discussed in section IIIB. In the previous method, for a two-mode model, two minimization problems should be solved (equations (11, 12)), while in our method, two TLS regression problems should be solved instead.

Remark 3. Equation (16) has two solutions. Selecting each of them and continuing the procedure will lead to either one of the parameter sets. Since only one quadratic equation needs to be solved, this method results in a unique solution for the two-mode models. However, in the three-mode models, using the same approach and with some straightforward mathematical manipulations, one ends up solving *K* different polynomial equations with order three. Since one cannot determine which solution belongs to which parameter vector, the unique solution cannot be found. Therefore, further studies are needed for generalizing this new method.

Remark 4. After finding the parameter estimates, one can cluster the data points via the following minimization. Clustering the data points will re-construct the switching sequence.

$$\lambda_t = \arg\min_{i=1,\ldots,n} (b_i^T x_t)^2 \quad (17)$$

IV. Simulation Results

A. Estimating the highest order of sub-models

The simulation performed in this sub-section is for demonstrating the effectiveness of the proposed method in estimating the maximum order of sub-models compared to the existing results in [4]. For this reason, two systems are considered. The first system was given in [4], and the second system is chosen to be more complex than the first one. These systems are:

$$System1: \begin{cases} y_t = 2y_{t-1} + 10 + w(t) & y_{t-1} < 0 \\ y_t = -1.5y_{t-1} + 10 + w(t) & y_{t-1} \geq 0 \end{cases}.$$

$$System\,2:\begin{cases} y_t = -.5y_{t-1} + .3y_{t-2} + u_{t-1} + 5 + w(t) & y_{t-1} \geq 0 \\ y_t = .7y_{t-1} - .4y_{t-2} + 2u_{t-1} - 3 + w(t) & y_{t-1} < 0 \end{cases}.$$

In both systems, u has a uniform distribution in [-10 10], and the noise has a Gaussian distribution with zero mean. Ma and Vidal [4] suggested a low bound of κ suitable for the system. However, the simulations below show that the desired κ for these two systems fall into very narrow intervals implying difficulties in selecting a suitable κ.

In each simulation run, the noise and input sequences are assigned randomly with random initialization in order to prevent any replication. For each system, an interval is defined for κ. That interval is divided into 1000 points, and for each κ, the system is simulated 20 times. By using Theorem 1 and the selected κ, the correctness of the order estimation is measured for each run, and the mean is recorded for each κ. This procedure is repeated for different noise variances.

As Fig.1 reveals, the desired interval for κ in system 1 is $\left[10^{-6}, 1.2\times10^{-5}\right]$; this results challenges the claim made by the authors in [4]. Fig.2 reveals that the desired interval for κ in system 2 is $\left[0.01, 0.03\right]$.

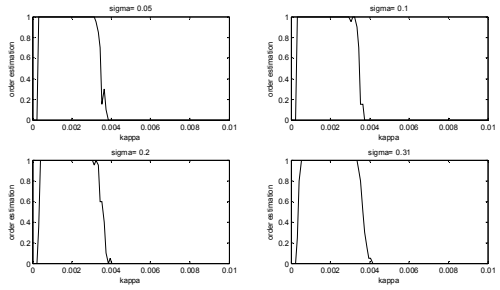

Fig.1 Finding the interval for κ for system 1 in different noise standard deviations

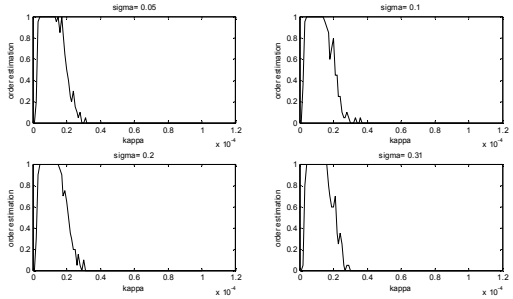

Fig.2 Finding the interval for κ for system 2 in different noise standard deviation

After finding the correct range for κ in each system, the procedure proposed by Ma and Vidal [4] finds the order of the system well. Thus, for each system, a different interval for κ should be found. Notice that this identification method is not supervised; therefore it is difficult to apply this method in practice.

On the other hand, the alternative method proposed in this article shows that for both systems, the same value for α will result in the correct estimation of the maximum order of sub-systems in the same range of noise variances (Table 1 and 2). Thus, the result of the proposed method does not depend on the system, but only depends on the signal to noise ratio of the system or the variance of the noise [11]. The variance of the noise should be estimated first for this method. Hereby, the variance of the noise or at least its bound is assumed to be known.

TABLE 1. CORRECT DETECTION OF SYSTEM ORDER FOR SYSTEM 1 USING PROPOSED APPROACH WITH $\alpha = 10^{-5}$

Noise Variance	.0025	.01	.04	0.1
Efficiency percentage of correct detection of system order	94.35	95.2	95.4	95.5

TABLE 2. CORRECT DETECTION OF SYSTEM ORDER FOR SYSTEM 2 USING PROPOSED APPROACH WITH $\alpha = 10^{-5}$

Noise Variance	.0025	.01	.04	0.1
Efficiency percentage of correct detection of system order	98.9	99.85	100	100

B. Parameter estimation of SARX models

Two 2-mode systems are used in the simulations. The first one, which is less complex, was introduced in the work of Vidal et al. This system has also been used in other works [12] and has become a benchmark problem.

$$System\,1:\begin{cases} y_t = 2u_{t-1} + 10 + w(t) & u_{t-1} < 0 \\ y_t = -1.5u_{t-1} + 10 + w(t) & u_{t-1} \geq 0 \end{cases}$$

The second system is already introduced as system 2 in section IV A. In these two-mode systems, the element-wise (EW) approach is applied and the results are compared to that obtained from the previous approach in [4].

In both systems, u has uniform distribution in the range of [-10 10], and y(0) is -10. $\omega(t)$ is Gaussian noise with zero mean and different variances. The noise variances used in the following simulations are $\sigma^2 = [.01\ .04\ 0.1\ 0.25]$. Simulations have been run 100 times for each noise variance. In each run, the generated noise is initialized with a random number in order to prevent any replication. The error between the estimated parameter $\hat{b}$ and the true parameters b is computed as $error = \max_{i=1,\ldots,n} \min_{j=1,\ldots,n} \frac{\hat{b}_i - b_j}{b_j}$.

The accuracy of the clustering is computed by using (17). Simulation results are presented in Fig. 3. This figure demonstrates that the overall performance of the new approach is better than the original method. The most important observation from Fig.3 is that the original approach fails to have satisfactory results when the system becomes more complex. In system 2, the difference between the performances of the new approaches and the original method is significant. Another important point is that the variance of the estimation error decreases dramatically for both systems 1 and 2 when applying the proposed method. This is due to the utilization of the TLS regression. Moreover, in the original approach, parameter κ, which varies for each system, should be assigned correctly in order to get the correct results. This requirement gives our new

approach a significant advantage over the original one.

C. Experimental switched control system

In this section an experimental setup introduced in [13] is used to show the effectiveness of the proposed method on the real system. The pilot plant is a tank system with two different level controllers. Two controllers have different transient responses in set-point tracking. One of them results in a fast response with high overshoot and the other one provides slower responses with much less overshoot. Switching between two controllers is performed following a random sequence. In this experimental setup the goal is to identify the closed-loop model for the two mode hybrid system.

The orders of the sub-models are selected as $n_\alpha = 2, n_\beta = 2$, which is same as the ones used in [13]. By applying the proposed EW approach to the de-trended data collected from the plant, the following parameters are estimated for the two mode system:

$$\begin{cases} y_k = 1.3547 y_{k-1} - 0.4304 y_{k-2} + 0.0406 u_{k-1} - 0.0344 u_{k-2} \\ y_k = 2.4864 y_{k-1} - 1.7158 y_{k-2} + 0.1422 u_{k-1} - 0.1552 u_{k-2} \end{cases}.$$

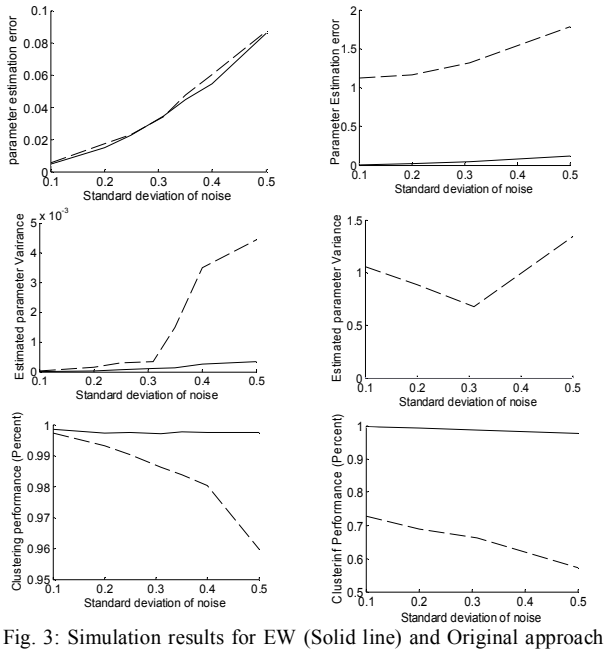

Fig. 3: Simulation results for EW (Solid line) and Original approach (Dashed Line) for both systems. System 1 is in the left and system 2 in the right.

The data set available from this experiment consist of 1000 data points. The first 700 data points are used for training and estimating purposes and the rest are used for model validation. The sampling time is 3 second. The self-validation and cross-validation of the model are shown in fig.4. As it can be seen in this figure the model estimated by the proposed method shows a good performance in both self-validation and cross-validation. The results obtained by this method are significantly better than the results reported in [13], which are obtained by using expectation maximization (EM) algorithm. The MSE of the self-validation results is 9.86×10^{-6} and the MSE for cross-validation results is 1.04×10^{-5}. Unfortunately these MSEs for the illustrated algorithm in [13] are not reported. In overall this experiment shows the applicability of our proposed algorithm.

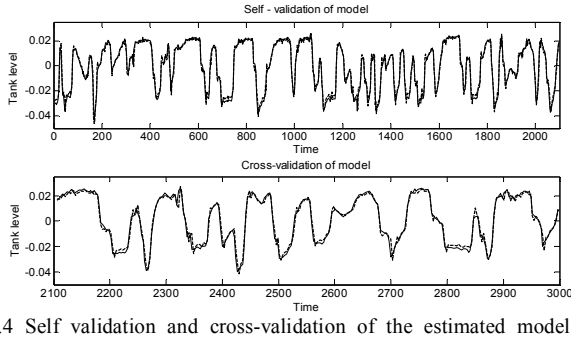

Fig.4 Self validation and cross-validation of the estimated model using the proposed method. The solid line is the actual output and

V. Conclusion

We have improved the algebraic geometry method for identifying the switched ARX models. The proposed method shows improved results when the data is corrupted with noise. The results in experimental example show that the developed approach has a great performance in the practical systems. The procedure of finding the closed-form solution for the parameters can be extended to the systems with higher number of modes in future work.

VI. References

[1] G. Ferrari, M. Muselli, D. Liberati, M. Morari, "A clustering technique for the identification of piecewise affine systems." *Automatica*, Vol. 39(2), pp. 205–217, 2003.

[2] A.L. Juloski, S. Weiland, W. Heemels, "A Bayesian approach to identification of hybrid systems.", *IEEE Trans. on Automatic Contro*l Vol. 50(10),pp. 1520–1533,2005.

[3] A. Bemporad, A. Garulli, S. Paoletti, A. Vicino, "A bounded-error approach to piecewise affine system identification." *IEEE Trans. on Automatic Control,* Vol. 50(10), pp. 1567–1580, 2005.

[4] Y. Ma, R. Vidal, "Identification of deterministic switched ARX systems via identification of algebraic varieties." In: *Hybrid Systems: Computation and Control*. vol. 3414 of Lecture Notes in Computer Science, Springer, Heidelberg, 2005, pp. 449–465.

[5] S. Paoletti, A.Lj. Juloski, G. Ferrari and R. Vidal, " Identification of Hybrid systems: a tutorial", *European Journal of Control*, Vol. 13 (2-3), pp. 242-260, 2007

[6] A. L. Juloski, W. Heemels, G. Ferrari, R. Vidal, S. Paoletti, J. Niessen,"Comparison of four procedures for the identification of hybrid systems ",*Lecture Notes in Computer Science*, Vol. 3414, pp. 354-369, 2005.

[7] R. Vidal,"Identification of PWARX hybrid models with unknown and possibly different orders." In: *Proceedings of the IEEE American Control Conference*, Boston, MA. 2004, pp 547–552

[8] I. Markovsky, S. Van Huffel, "Overview of total least-squares methods", *Signal Processing*, Vol. 87 (10), pp. 2283-2302, 2007.

[9] J. Harris, *Algebraic Geometry: A first course*, Springer- verlog,1992.

[10] K. Kanatani and C. Matsunaga."Estimating the number of independent motions for multibody motion segmentation". In *Asian Conf. on Computer Vision*, 2002.

[11] Konestnantinites and Yao," Statistical Analysis of Effective Singular Values in Matrix Rank Determination", *IEEE Transactions On Acoustics, Speech, And Signal Processing*, Vol. 36, No. 5, May 1988.

[12] H. Niessen and A. Jouleski. "Comparison of three precodeures for identification of hybrid systems". In conference on *control applications*,2004.

[13] X. Jing and B. Huang, "Robust Identification of Piecewise/Switching Autoregressive Exogenous Process". *AICHE Journal*, Vol. 56, No. 7, July 2010.