Identification of Linear Switching System with Unknown Dimensions

Amir. H. Shirdel
Chemical Engineering Department
Åbo Akademi University
Amir.shirdel@abo.fi

Kaj-Mikael Björk
Information Technology Department
Åbo Akademi University
Kaj-mikael.bjork@abo.fi

Hannu. T. Toivonen
Information Technology Department
Åbo Akademi University
Hannu.toivonen@abo.fi

Abstract

Switching systems have the property that their dynamical behavior may switch between a number of different modes. Identification of switching systems consists of identifying both the individual models which describe the system in the various modes, as well as the time instants when the mode changes have occurred. This soft computing problem therefore consists of numerically demanding coupled parameter identification and clustering problem. In this contribution a method based on sparse optimization techniques is proposed for identification of switching systems. In the two proposed methods, the modes and the associated models are determined one by one by solving linear problems which are sparse with respect to the number of violated epsilon-insensitive constraints and also solving by extending the models with an alternative LP-formulation of the optimization problem. The performance of the identification procedure is demonstrated by simulated examples.

1. Introduction

Switching systems are dynamical systems with the property that they can switch between a number of modes with different dynamical properties [1], [2]. An important class of switching systems consists of hybrid systems, whose continuous dynamics depend on discrete-valued logical variables.

Systems with several operating regimes can also be modeled as switching systems, where the modes are associated with the various operating conditions. In this case the mode is usually known or is a function of known variables. In more general cases, the mode switches may be random, or they may depend on variables which are unknown.

In practice, an important problem consists of black-box identification of switching systems using input-output data. A special problem in switching system identification is the fact that the times of the mode switches may not be known. In these cases the switching times between the various modes should be identified simultaneously with the individual models, which make the identification of switching systems significantly more demanding than standard system identification. In order to cope with the difficulties, a number of emerging framework soft computing have been developed [3], [4]. Therefore, in most studies of switching system identification various simplifying assumptions have been made.

A special class of switching systems consists of systems where the modes depend on the state. Identification of such systems can be roughly decomposed into two parts. First, clustering techniques are used to determine regions where the various system modes are active, followed by identification of the individual models using standard techniques.

Piecewise affine models are commonly used to describe switching system. In [5] statistical clustering of the input data has been applied to identification of piecewise affine systems. The fact that the clustering and model identification steps interact and cannot be solved independently of each other has been addressed in [6], [7], where piecewise affine models were identified using combinations of clustering, classification and linear identification methods. In [8] a Bayesian approach was applied to identification of hybrid systems, where the model parameters are described as random variables.

In [9] identification of switching system was applied by minimizing the sum of the product of the squared prediction errors of individual models. By construction, in the noise-free case this objective function is minimized by the correct models associated with the various modes. In other words, hybrid decoupling constraint establishes a connection between linear hybrid system identification, polynomial factorization and hyperplane clustering. An important feature of this approach is that it can be applied to on-line identification [10], [11]. The procedure was further modified in [12] for robustness to outliers in the data. The method involves, however, the solution of a computationally demanding nonconvex optimization problem, for which convergence to the globally optimal solution can in
general not be ensured. An alternative approach was proposed in [13], where subspace identification was applied to the identification of switching systems with a minimum dwell time in each mode.

In [14], two classes of estimators for hybrid systems have been proposed: the ME estimator, based on the minimum of the sub-model errors, and the PE estimator, based on the product of the sub model errors. In terms of efficiency, the proposed estimators have been tailored to tackle large-scale problems and have been shown to yield fast and accurate results in experiments with numerous data.

In [15], segmentation of time-varying systems and signals with piecewise model parameters which are constant in time was formulated as a least squares problem with sum of norms regularization of the parameter jumps, a generalization of \( l_1 \)-regularization. A nice characteristic of this method is that the number of parameter jumps can be controlled using a single tuning parameter.

In [16], a sparse optimization method was applied to identification of switched linear systems. The approach is based on the fact if the input-output data are described using a single linear model, then the vector of prediction errors will contain many entries which are (close to) zero if the model is selected as one associated with any of the system modes. This leads to a sparse optimization problem, which was relaxed to an \( l_1 \)-norm minimization problem.

Recently in [21], an approach to identify hybrid systems with unknown nonlinearities in the sub-models using combination of sparse optimization and support vector machines was presented.

For many demanding problems which a single theory has difficulty to manage it is often useful to combine the strengths of two or more sets of theories. In the soft computing realm, this has been common in many years. For example, [19] solved a heat exchanger network system with a combined genetic algorithm and Mixed Integer Nonlinear Programming (MINLP) technique. In [20] a mixed Tabu search and MILP (Mixed Integer Linear Programming) method was used to solve a very complex production scheduling problem. In this paper, a hybrid method combining the strengths of sparse optimization and Linear Programming (LP) and Quadratic Programming (QP) will be used to solve the identification problem for switching systems.

In this paper, an alternative sparse optimization approach to switching system identification is presented. In the first proposed method, the identification problem is posed as a support vector regression (SVR) problem. Support vector machines and SVR are a powerful technique for data classification and regression [17], which has been successfully applied to robust black box system identification. In support vector regression, a regularization term, which is taken as a quadratic function of the parameters, is minimized subject to linear constraints on the prediction error magnitudes. Here, we use the fact that when the model parameters are equal to the parameters associated with one of the system modes, a significant proportion of the constraints on the prediction errors will be satisfied, corresponding to the time instants when that particular mode has been active. Such a solution can be found using sparse optimization techniques, which makes it possible to find the model parameters of the various modes one by one.

The second proposed method will also extend the models with an alternative LP-formulation of the optimization problem. The LP-formulation has the advantage of being even faster to solve than the QP (in general). The LP version along with a quite simple algorithm will also used to illustrate how to solve problems, where the number of subsystems is not known. We will show that a simpler identification algorithm structure than used in [16] will, and that the linear formulation identifies the systems very accurately, even if the number of subsystems is unknown. In the end this hybrid system can be used as a benchmark for other soft computing techniques used to identify unknown systems. This will be elaborated on in the last section.

### 2. PROBLEM FORMULATION

We consider time-varying linear dynamical systems which can be described by

\[
y(k) = \varphi(k)^T \theta(k) + e(k) \tag{1}
\]

where \( y(k) \) is the output, \( e(k) \) is a disturbance, \( \varphi(k) \) is a state vector, and \( \theta(k) \) is a parameter vector. For example, in an ARX model, we have

\[
\varphi(k)^T = [y(k-1), \ldots, y(k-r), u(k-1), \ldots, u(k-r)]
\]

It is assumed that the system dynamics switch between a numbers of modes, so that

\[
\theta(k) \in \{\theta_1, \theta_2, \ldots, \theta_n\}
\]

where \( \theta_i \) is the vector of system parameters in the \( i \)th mode.

The problem studied in this paper is to identify the mode-wise parameter vectors \( q_i \) from input-output data \( \{\varphi(k), y(k), k = 1, 2, \ldots, N\} \). It is not
assumed that the time instant when the various modes have been active is known, and these should be identified as well.

3. A SPARSE SVR APPROACH

The approach proposed in this paper is based on the fact that if mode \( i \) has been active at \( N_i \) time instants, then, assuming that \( \varepsilon(k) \leq \varepsilon \), the inequality

\[
|y(k) - \varphi(k)^T \hat{\theta}| \leq \varepsilon
\]

holds at these \( N_i \) time instants. The most frequent mode can then be found by solving the optimization problem

\[
\min \| \xi \|_0
\]

subject to

\[
(y(k) - \varphi(k)^T \hat{\theta} - \xi_k)^2 \leq \varepsilon^2, k = 1, 2, ..., N
\]

where \( \| \cdot \|_0 \) denotes the number of nonzero entries in \( \xi \). This is, however, a hard combinatorial optimization problem, whose solution is in general intractable. Therefore, these kinds of problems are usually replaced by the convex problem

\[
\min \| \xi \|
\]

subject to

\[
(y(k) - \varphi(k)^T \hat{\theta} - \xi_k)^2 \leq \varepsilon^2, k = 1, 2, ..., N
\]

Defining the variables

\[
\xi_k = \xi_k^+ - \xi_k^-, \quad \xi_k^+, \xi_k^- \geq 0
\]

we can write (4) as

\[
\min_{\hat{\theta}, \xi^+, \xi^-} \sum_{k=1}^{N} (\xi_k^+ + \xi_k^-)
\]

subject to

\[
(y(k) - \varphi(k)^T \hat{\theta} - \xi_k) \leq \varepsilon + \xi_k^+, \quad \xi_k^+ \geq 0
\]

\[
y(k) + \varphi(k)^T \hat{\theta} \leq \varepsilon + \xi_k^-, \quad \xi_k^- \geq 0
\]

Here, we will modify the optimization problem further by using reweighting of the slack variables \( \xi_k^+, \xi_k^- \) for improved accuracy, as well as a regularization term which penalizes excessively large parameter vector. This gives the optimization problem

\[
\min_{\hat{\theta}, \xi^+, \xi^-} \frac{1}{2} \| \hat{\theta} \|^2 + \sum_{k=1}^{N} C_k (\xi_k^+ + \xi_k^-)
\]

subject to the \( \varepsilon \)-insensitive constraints

\[
(y(k) - \varphi(k)^T \hat{\theta}) \leq \varepsilon + \xi_k^+, \quad \xi_k^+ \geq 0
\]

\[
y(k) + \varphi(k)^T \hat{\theta} \leq \varepsilon + \xi_k^-, \quad \xi_k^- \geq 0
\]

The problem defined by (7) and (8) is the one minimized in support vector regression (SVR) [18], [19], and is usually solved via a dual maximization problem. The constraint \( \varepsilon \) in (8) is a design parameter, which defines the accuracy of the model predictions, and should be selected to reflect the disturbance magnitudes affecting the system (1).

In order to find and identify the parameters associated with one mode, the following sparse SVR problem with reweighting can be applied.

**Algorithm 3.1.** Sparse SVR with reweighting.

**Step 0.** Initialization: set iteration index \( m = 0 \) and initial weights \( C_k = C_{\text{svm}} \).

**Step 1.** Set \( m \leftarrow m + 1 \) and solve the SVR problem defined by (7), (8) to give the parameter estimate \( \hat{\theta}^{(m)} \).

**Step 2.** Check for convergence. If convergence has been achieved, determine the time instants at which

\[
|y(k) - \varphi(k)^T \hat{\theta}^{(m)}| \leq \delta
\]
where $\delta > \varepsilon$ is a constant. Determine the parameter estimate $\hat{\theta}_i$ by SVR using these time instants and stop. Otherwise, continue from step 3.

Step 3. Let
$$e^{(m)}(k) = y(k) - \varphi(k)^T \hat{\theta}^{(m)}$$
and define the weights
$$C^{(m)}_k = \left\{ \begin{array}{ll} \frac{C_{\text{sym}}}{1 + C_{\text{sym}} (|e^{(m)}(k)| - \varepsilon)} & \text{if} |e^{(m)}(k)| \geq \varepsilon \\ 1 & \text{otherwise} \end{array} \right.$$ 

Continue from step 1.

The above algorithm returns a parameter vector which, optimally, is an estimate of one of the parameter vectors $\theta_i$ of the system modes. In order to find the models describing all the modes, the algorithm has to be applied recursively as follows, cf. [16].

Algorithm 3.2. Identification of switching system.

Step 0. Initialization: set $i = 1$.

Step 1. Apply Algorithm 3.1 to find a parameter estimate $\hat{\theta}_i$, and remove the data pairs $\varphi(k), y(k)$ at which mode $i$ has been active, cf. step 2 of Algorithm 3.1.

Step 2. Check the reduced data set for convergence: if all data pairs have been accounted for, stop. Otherwise, set $i \leftarrow i + 1$ and continue from step 1 using the reduced data set.

4. A LINEARLY REFORMULATED MODEL

In [16], Bako showed a very efficient way to identify switching linear systems with the help of sparse optimization. Previously in this paper, we showed that it is sufficient to work with the primal problem in order to get similar results (utilizing quadratic programming in the sparse optimization framework). In this section, we will show that it is sufficient to use a linear reformulation of the primal problem if we use an $l_1$-norm instead of an $l_2$-norm in Eq. 7. This formulation can be described as follows:

$$\min_{\hat{\theta}, \xi^+, \xi^-} \frac{1}{2} \|\hat{\theta}\| + \sum_{k=1}^{N} C_k (\xi^+_k + \xi^-_k)$$
subject to
$$y(k) - \varphi(k)^T \hat{\theta} \leq \varepsilon + \xi^+_k$$
$$y(k) + \varphi(k)^T \hat{\theta} \leq \varepsilon + \xi^-_k$$
$$\xi^+_k, \xi^-_k \geq 0$$

which can be rewritten as

$$\min_{\hat{\theta}, \xi^+, \xi^-} \frac{1}{2} \sum_{i=1}^{I} (\hat{\theta}^+_i + \hat{\theta}^-_i) + \sum_{k=1}^{N} C_k (\xi^+_k + \xi^-_k)$$
subject to
$$\hat{\theta}^+_i \geq \hat{\theta}_i \quad \forall i \in \{1,2,\ldots,I\}$$
$$\hat{\theta}^-_i \geq \hat{\theta}_i \quad \forall i \in \{1,2,\ldots,I\}$$
$$y(k) - \varphi(k)^T \hat{\theta} \leq \varepsilon + \xi^+_k$$
$$y(k) + \varphi(k)^T \hat{\theta} \leq \varepsilon + \xi^-_k$$
$$\xi^+_k, \xi^-_k \geq 0$$

Here $\hat{\theta}_i = \hat{\theta}^+_i - \hat{\theta}^-_i$, where $\hat{\theta}_i$ is the $i$th component of parameter vector $\hat{\theta}$. The optimization problem described in Eq. 11 is now entirely linear, and standard Linear Programming (LP) pieces of software can be used to solve the problem. The advantage of using LP-problem formulations is found in the fact that LP is usually faster to solve than the Quadratic Programming (QP) problem in Eqs. (7) and (8). In this way it is likely that we can solve larger identification problems with the linear reformulation than the original formulation by Bako [16] or the formulation in Eq. 7-8.

5. NUMERICAL EXAMPLES

In this section the numerical examples are shown in order to illustrate the performance of the proposed identification methods.
Example 5.1

In this example, we consider the system defined by

\[
y(k) = a_1(k)y(k-1) + a_2(k)y(k-2) + b_1(k)u(k-1) + b_2(k)u(k-2) + e(k) \tag{12}
\]

where the parameter vector

\[
\theta(k) = [a_1(k) \ a_2(k) \ b_1(k) \ b_2(k)]^T
\]

belongs to the set \( \theta(k) \in \{\theta_1, \theta_2, \theta_3\} \), corresponding to the three system modes, with the parameter values

\[
\begin{align*}
\theta_1 &= [-0.4 \ 0.25 \ -0.15 \ 0.08]^T \\
\theta_2 &= [1.55 \ -0.58 \ -2.1 \ 0.96]^T \\
\theta_3 &= [1.0 \ -0.25 \ -0.65 \ 0.3]^T
\end{align*}
\]

The three modes were taken to occur randomly with probabilities \( p_i; i = 1, 2, 3 \). Two different cases were considered:

**Case 1**: \( p_1 = 0.6, p_2 = 0.3, p_3 = 0.1 \)

**Case 2**: \( p_1 = 1/3, p_2 = 1/3, p_3 = 1/3 \)

| \( \hat{\theta}_1 \) | -0.396±0.01 | 0.248±0.004 | -0.153±0.012 | 0.083±0.017 |
| \( \hat{\theta}_2 \) | 1.545±0.009 | -0.577±0.013 | -2.092±0.03 | 0.953±0.03 |
| \( \hat{\theta}_3 \) | 1.002±0.025 | -0.239±0.02 | -0.650±0.05 | 0.294±0.06 |

| Table 1 | AVERAGE OF ESTIMATED PARAMETERS IN CASE 1. |

| MSE | 0.1002±0.005 |
| MSPE | 0.1372±0.01 |

The system was excited using a normally distributed white noise input sequence \( \{u(k)\} \) with zero mean and unit variance. The disturbance \( \{e(k)\} \) was normally distributed zero mean white noise with variance 0.1, corresponding to the signal to noise ratio 30dB. The identification method presented in section 3 was applied using SVR design parameters \( C_{svm} = 100 \) and \( \varepsilon = 0.15 \) in the reweighting scheme of Algorithm 3.1. Training data sequences generated from the switching system and used for identification for the Case 1 are shown in Fig. 1. Figures 2 and 3 show the prediction error after applying Algorithm 3.1 to find the first mode in Case 1. The mode found in the first stage was the most frequent one, i.e., mode 1.
Figures 4 and 5 show the prediction errors at the second and third applications of Algorithm 3.1, after removing the data points associated with modes already found. In the three stages of Algorithm 3.2, the number of iterations required for convergence in Algorithm 3.1 was four, two and one, respectively. After three stages of Algorithm 3.2, there may be some data points which have not been assigned to any mode. However, by computing the prediction errors obtained using the models found, these data points were found to be described by either mode 1 or mode 2. All three system modes were found by the algorithms in Case 1. The estimated parameters after 20 independent runs are given in Table 1.

Table 2 shows the mean square values of the disturbance sequence \( \{ e(k) \} \), which is the minimum prediction error which can be achieved if the system parameters are known, and the prediction error obtained with the identified model. The prediction errors were computed by using the modes found by the algorithm at the various time instants. This is an optimistic assumption if the identified model is applied to new data, as the model does not give any prediction of which mode will be active. Such information should be determined separately.

Table 3

<table>
<thead>
<tr>
<th>Table 3</th>
<th>AVERAGE OF ESTIMATED PARAMETER IN CASE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta}_1 )</td>
<td>-0.395±0.015</td>
</tr>
<tr>
<td>( \hat{\theta}_2 )</td>
<td>1.549±0.017</td>
</tr>
<tr>
<td>( \hat{\theta}_3 )</td>
<td>0.985±0.015</td>
</tr>
</tbody>
</table>

Figure 6 shows the predicted output and prediction errors for the training data achieved with the identified model in Case 1. The identification method assigned 767 data points of 800 points to the correct mode.

In Case 2, the algorithm found all three system modes. The estimated parameters after 20 independent runs are given in Table 3, and the mean square errors are given in Table 4 (cf. Table 2). In this case, the algorithm assigned 751 data points to the correct mode. The result is somewhat worse than in Case 1, as it is harder to identify modes which occur with the same frequencies.

Table 4

<table>
<thead>
<tr>
<th>Table 4</th>
<th>AVERAGE OF MEAN SQUARE PREDICTION ERROR IN CASE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.0995±0.005</td>
</tr>
<tr>
<td>MSPE</td>
<td>0.1423±0.025</td>
</tr>
</tbody>
</table>
Example 5.2

In this example, the approach described in Section 3 is applied to system (12) using the $l_1$-relaxation presented in Section 4. While the sparse optimization routine was applied to identify the modes, one at a time, the final model for each mode was identified afterwards with only the points assigned to the mode present. In this example 769 out of 800 points were correctly, and the standard deviation of the prediction error was 0.34, whereas the standard deviation of the simulated system noise was 0.31. In Table 5, the identified mode-wise system parameters are listed.

<table>
<thead>
<tr>
<th>System</th>
<th>a1</th>
<th>a2</th>
<th>b1</th>
<th>b2</th>
</tr>
</thead>
<tbody>
<tr>
<td>System 1</td>
<td>-0.4</td>
<td>0.25</td>
<td>-0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>System 2</td>
<td>1.55</td>
<td>-0.58</td>
<td>-2.1</td>
<td>0.96</td>
</tr>
<tr>
<td>System 3</td>
<td>1</td>
<td>-0.24</td>
<td>-0.65</td>
<td>0.3</td>
</tr>
</tbody>
</table>

6. IDENTIFICATION OF SYSTEMS OF UNKNOWN SIZE

Above, it has been assumed the number of subsystems as well as the sizes of the subsystems is known. In [16], Bako introduced an iterative approach to find these. Our experience is that since the linear optimization model is not as sensitive to problem size as the quadratic models can be, the system can simply be allowed to be “over-sized” in both parameters and in the number of subsystems. Let’s assume that we have a similar system as in Section 5, but with different subsystems sizes, cf. Table 6.

<table>
<thead>
<tr>
<th>System</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
<th>b4</th>
</tr>
</thead>
<tbody>
<tr>
<td>System 1</td>
<td>-0.4</td>
<td>0.25</td>
<td>0.5</td>
<td>0</td>
<td>-0.15</td>
<td>0.08</td>
<td>1.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>System 2</td>
<td>1.55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-2.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>System 3</td>
<td>1</td>
<td>-0.24</td>
<td>0</td>
<td>0</td>
<td>-0.65</td>
<td>0.3</td>
<td>-0.4</td>
<td>0</td>
</tr>
</tbody>
</table>

Also, the first subsystem represents roughly 50% of the data points, the second about 30% and the third about 20%. We assume not to know the number of subsystems, but we assume that the number is not greater than seven. In this case we will start with identifying seven subsystems using the sparse optimization algorithm solving iteratively the LP-problem in Eq. 11 (in the same manner as described in Section 4). Again, when these subsystems are identified, the system parameters are fine-tuned by finding the best fit, only with the assigned data-points for each subsystem. This step is usually only performed for fine-tuning the parameters as in the example in the previous section. Then a step that was not performed in the previous example is performed; i.e. to re-assign each of the points to the subsystem that is the best fit. This step is important not only for assigning the correct subsystem to each data point but also helps the upcoming steps. The next step is to merge subsystems that are close to each other (in case of not having for instance seven subsystems). This procedure is simply by comparing each system parameter for every pair of subsystems (to check for similarity). If close enough, merge the subsystems until no pair of similar subsystems are found. Then the final step is to again fine-tune the parameters with the assigned data-points (for each subsystem) in order to find the best possible system parameters. See Figure 7 for a schematic flow of the algorithm.

The system in Table 6 was simulated and identified with the above proposed algorithm. The results were very interesting. First of all, in the end 779 out of 800 data points were correctly assigned to subsystems. Also only 3 subsystems were present at the end and that the system parameters were very close to the simulated system, see Table 7. The variance of the prediction error obtained with the identified system was 0.104, whereas the variance of the disturbance sequence was 0.107. See Figure 8 for the error plot for the data.
7. CONCLUSIONS AND FUTURE WORK

In this paper a class of methods for identification of switching dynamical systems has been considered. The approach is based on the fact that a support vector regression problem with an \( \epsilon \) - insensitive cost is equivalent to an \( l_1 \) -relaxed sparse optimization problem. This connection is used to identify the various modes of a switching system successively by solving a sequence of reweighted support vector regression problems. In a closely related approach, the quadratic objective function of the support vector method is replaced by a linear cost. Simulated examples demonstrate that the proposed algorithms correctly identify the various modes of a switching system.

The main contribution has been to present an alternative version to [16], which is simple and efficient in identifying the correct modes of a switching system. Also, using an LP-formulation makes it likely possible to solve larger identification problems.

We also consider cases where the number of subsystems and/or their sizes is unknown. In the former case, the number of subsystems can simply be overestimated and in a final stage merged together (if similar enough). For systems with unknown sizes, the number of parameters can likewise be overestimated, and the minimum number of parameters required to describe the system is determined by sparse optimization.

In the recent reference [21], a class of methods closely related to the one presented in this paper was studied. We believe, however, that our reweighting scheme (Algorithm 3.1), which is tailored for the support vector formulation (7), (8) with \( \epsilon \) - insensitive constraints, and the generalization to handle switching systems of unknown sizes (Section 6) enlarges the scope of the proposed class of methods.

Finally, the technique can be used for comparisons to other problems using soft computing techniques; for example, a blast furnace identification work with both ANFIS and the method proposed in this paper is undertaken at the moment of preparing this paper. This problem has several thousands of data points, many inputs with models with more than ten time steps backwards in time. For further research is also left the track to combine soft computing techniques, such as genetic algorithms to be combined with sparse optimization for identifying more complex switching systems.

### Table 7
THE SIMULATED SYSTEM VS. THE IDENTIFIED SYSTEM.

<table>
<thead>
<tr>
<th>The simulated system</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
<th>b4</th>
</tr>
</thead>
<tbody>
<tr>
<td>system 1</td>
<td>-0.4</td>
<td>0.25</td>
<td>0.5</td>
<td>0</td>
<td>-0.13</td>
<td>0.08</td>
<td>1.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>system 2</td>
<td>1.55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-2.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>system 3</td>
<td>1</td>
<td>-0.24</td>
<td>0</td>
<td>0</td>
<td>-0.03</td>
<td>0.3</td>
<td>-0.4</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The identified system</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>b1</th>
<th>b2</th>
<th>b3</th>
<th>b4</th>
</tr>
</thead>
<tbody>
<tr>
<td>system 1</td>
<td>-0.402</td>
<td>0.25</td>
<td>0.502</td>
<td>0.001</td>
<td>-0.138</td>
<td>0.083</td>
<td>1.201</td>
<td>-0.494</td>
</tr>
<tr>
<td>system 2</td>
<td>1.551</td>
<td>0</td>
<td>0</td>
<td>0.002</td>
<td>-2.104</td>
<td>0.003</td>
<td>-7E-04</td>
<td>-0.005</td>
</tr>
<tr>
<td>system 3</td>
<td>0.999</td>
<td>-0.241</td>
<td>0.004</td>
<td>-7E-04</td>
<td>-0.639</td>
<td>0.283</td>
<td>-0.363</td>
<td>0.005</td>
</tr>
</tbody>
</table>
8. ACKNOWLEDGMENTS

This work has been funded by the Foundation of Åbo Akademi University and the center for optimization and Systems Engineering at Åbo Akademi as well as IAMSR at Åbo Akademi and the TUF foundation at Arcada University of Applied Sciences. We are also thankful to the anonymous reviewers for informing us about the recent reference [21].

9. REFERENCES