Hybrid system identification using a structural approach and its model based control: An experimental validation

Naresh N. Nandola, Sharad Bhartiya

Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai 400 076, India

Keywords:
Identification of hybrid system
Control of hybrid system
Experimental validation
Three tank benchmark system

ABSTRACT

For implementation of various model based techniques such as in control and fault diagnosis, data-driven identification is key for enabling cheap and rapid development of models of hybrid systems of industrial interest. In the present work, a novel identification method is proposed for a class of hybrid systems which are linear and separable in the discrete variables (that is discrete states and discrete inputs). The method takes cognizance of the fact that the separable structure of the hybrid system constrains the evolution of system dynamics. In particular, the proposed method identifies models corresponding to a certain number of modes, far fewer than the total possible modes of the system. It then generates the models for the remaining modes without any further requirement for input–output data by exploiting the separable structure of the hybrid system. We experimentally validate the method by identifying the model for a three tank benchmark hybrid system followed by model predictive control using the identified model.

1. Introduction

The requirement for enhanced operational flexibility in the process industry has led to greater focus on modeling of hybrid systems which incorporate discrete decisions/variables along with continuous process states in the model. Such models play an important role in various model based tools such as in control, soft sensing and diagnosis. While use of first principles models in such applications is feasible, these would be necessarily limited to small and well understood systems. On the other hand, for practical implementation, data-driven identification is key for enabling cheap and rapid development of models of hybrid system of industrial interest. The challenge in the identification of hybrid systems lies in the fact that the model parameters depend on the mode or location [1]. Thus, if the correspondence between the identification data and the mode of the system is not known a priori, the hybrid system identification problem becomes one of simultaneous identification and estimation of the model parameters and modes of the system, respectively. In fact, substantial literature on identification of hybrid systems pays great emphasis on the ability of identification algorithms to estimate the mode as well as the continuous state model. Roll et al. [1] have classified the strategies for identification of the piecewise affine (PWA) model into the following categories: (i) identifying both, models for the different locations as well as the partition (mode) of the hybrid systems simultaneously; (ii) identifying both, models for the location and the partitions simultaneously but adding only one partition at time; (iii) the partition and modes are identified in several steps, and (iv) the partition is known a priori. While the first three approaches estimate the mode of the hybrid system, the last approach assumes that the mode is known. The mode of the system changes upon occurrence of an event. From a control viewpoint, these events may be characterized as state event, control event or time event. Control and time events are triggered by the user and hence do not need to be identified. State events also would not require identification as long as the identification data contains noise-free
information about the event related states. If information about the event related states is absent, then the mode would require estimation along with model parameters as in the first three approaches.

In the current work, we assume that the mode of the system can be inferred from the input–output data and is therefore known a priori. Although this simplifies the identification procedure, we still need to identify models for a large number of locations. Note that if ns binary variables are used to describe the occurrence of events, then the total number of possible modes is 2ns.

A straightforward approach would involve identification of linear models for all feasible modes of the system and use these models in control. However, for practical implementation such an effort is non-trivial since: (i) the number of modes are usually very large, and (ii) the identification signal must be persistently exciting to ensure that sufficient length of rich input–output data exists for each of the feasible modes.

We propose a novel identification method for a class of hybrid systems which are linear and separable in the discrete variables (that is, discrete states and discrete inputs). The method takes cognizance of the fact that the separable structure of the hybrid system constrains the evolution of system dynamics. Thus, the identified models are consistent with the class of hybrid system that is characterized by separability in discrete variables. In particular, the proposed method identifies models corresponding to certain number of modes, far fewer than the 2ns total possible modes of the system. It then generates the models for the remaining modes without any further requirement for input–output data by exploiting the separable structure of the hybrid system. Since only certain modes need identification for modeling of the composite hybrid system, the proposed method is referred to as a reduced mode approach. The identified models can be used for various model-based tools discussed previously. In this work, the novel identification method and model predictive control using the identified models have been experimentally validated using a 3-tank benchmark setup.

The paper has been organized as follows: Section 2 briefly reviews the literature on identification of hybrid systems. Section 3 presents the reduced mode strategy for identification of hybrid system. Section 4 presents an experimental validation of the identification procedure as well as experimental implementation of MPC using identified models for control of a three-tank benchmark hybrid system setup. Finally, the paper is summarized in Section 5.

2. A brief review on hybrid system identification

The identification theory for continuous state systems is well developed in the literature (see for example Ref. [2]). However, hybrid systems add extra complexity due to the presence of discrete states or discrete inputs. Here, the identification involves estimating the partition-discriminating modes of the hybrid system followed by identification of the parameters of the continuous state models in each mode. Description of hybrid systems using PWA models is the dominant modeling approach discussed in literature.

In case of known partitions, the identification problem is one of identifying the continuous state dynamic models for each mode, [1] which can be performed by any available standard identification techniques. On the other hand, for unknown partitions, the identification problem is more complex. Julian et al. [3], Pucar & Sjoberg [4], Messai et al. [5] and Bartuni [6] proposed hybrid system identification approaches assuming both partition and models parameters as unknowns. They assumed a structure of the model and formulated an optimization problem which minimizes an appropriately defined criterion function in order to obtain all partition and parameters of the location models simultaneously. In contrast, Breiman [7], Heredia and Arce [8], Ernst [9] and Hush & Horne [10] avoided solving a single complex optimization problem as above but solved several simple optimization problems in order to achieve the same goal of identifying partition and parameters of all locations. They initially assumed a single location with simple structure and added more locations by fitting the residual of previous models. Temporad et al. [11] and Ferrari-Trecate et al. [12] used a clustering based approach to identify the partitions of the hybrid system. Roll et al. [1] presented a mixed integer linear or quadratic programming based identification approach that is guaranteed to converge to its global optima, a feature not guaranteed in the previously reported methods.

Approaches that assume knowledge of the partition a priori have also been presented in the literature. Choi and Choi [13] divided the state space into several regions from the distribution information of the regression vectors such that each region contains sufficient data. This was followed by identifying affine models corresponding to each of these regions. Egbunonu and Guay [14] presented a method for identification of state space models for switched linear hybrid systems with the assumption of known partition. They considered a deterministic state space model framework for each of the sub models. The switched linear system was converted to the well-known mixed logical dynamical framework and model parameters were obtained using mixed integer program. On the other hand, Billings and Voon [15] and Simani et al. [16] partitioned the state-space independent of the identification data. Billings and Voon [15] used a rectangular partitioning parallel to coordinate axes while Simani et al. [16] used a simplicial partition. A Bayesian approach for identifying model parameters that considers model parameters as random variables has been proposed by Juloski et al. [17]. In yet another approach, model parameters for the various modes are obtained using roots of an appropriately defined polynomial which is independent of evolution of the discrete states of the system [18,19].

3. Identification of hybrid systems using a reduced mode approach

We motivate our method by considering the following switched nonlinear hybrid system [20].

$$\dot{x}^c = f_c(x^c, u^c, \delta)$$ (1)
Finally, the continuous and discrete states can be represented by an augmented state vector \( \mathbf{x} \) as follows,

\[
\begin{bmatrix}
\dot{x}^c_{k+1} \\
\dot{x}^d_{k+1}
\end{bmatrix} = 
\begin{bmatrix}
\Phi_i & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
x^c_k \\
0
\end{bmatrix} + 
\begin{bmatrix}
\Gamma_i & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
x^d_k \\
0
\end{bmatrix} + 
\begin{bmatrix}
u^c_k \\
b^d_i
\end{bmatrix}, \quad i = 1, 2, \ldots, 2^{\text{ns}}.
\]

(4)

Note that the inequality in (3) can be rewritten in discrete-time form by merely replacing the continuous-time variable \( t \) by a sampling time index \( k \) as follows.

\[
E_1 u^c_k + E_2 \delta_k + E_3 x_k \leq E_4.
\]

(5)

Finally, the continuous and discrete states can be represented by an augmented state vector \( \mathbf{x} \) as follows,

\[
\begin{align*}
\mathbf{x}_{k+1} &= \Phi_i \mathbf{x}_k + \Gamma_i u^c_k + f^i_d. \\
E_1 u^c_k + E_2 \delta_k + E_3 x_k &\leq E_4.
\end{align*}
\]

(6)

(7)

The output of the PWA model may be written as,

\[
y_{k+1} = C \mathbf{x}_{k+1}.
\]

(8)

Remark. The binary valued vector \( \delta \in \{0, 1\}^{\text{ns}} \) can assume \( 2^{\text{ns}} \) values. Each of the \( 2^{\text{ns}} \) values correspond to distinct locations or modes of the system. Some of these modes may be infeasible. For example, let \( \delta_1 = 1 \) if the level of water \( \geq 20 \) cm in a tank of height 60 cm and 0 otherwise. Also let \( \delta_2 = 1 \) if level of water is \( \geq 30 \) cm and 0 otherwise. Then the mode \((0, 1)\) is infeasible. For simplicity and without loss of generality, we will assume that all \( 2^{\text{ns}} \) modes are feasible.

If the partition of the hybrid system is predetermined, as considered here, the identification problem reduces to estimating the \( 2^{\text{ns}} \) models of the PWA system. We assume that a linear affine model adequately describes the continuous state dynamics in each location. Each of the \( 2^{\text{ns}} \) models may be represented in discrete-time as follows,

\[
\begin{align*}
\dot{x}^c_{k+1} &= \Phi_i x^c_k + \Gamma_i u^c_k + f^i_d. \\
E_1 u^c_k + E_2 \delta_k + E_3 x_k &\leq E_4.
\end{align*}
\]

(9)

Here the variable \( \delta_i \) is the \( i \)th element of the binary vector \( \delta \in \{0, 1\}^{\text{ns}} \) which serves to describe whether a particular phenomenon modeled by \( f_i(x^c, u^c) \) is active or not. For such systems, the proposed method identifies a set of models for uniquely defined modes called the basis modes using the input–output data. For a hybrid system with \( \text{ns} \) binary variables, the total number of basis modes is \( \text{ns} + 1 \). The method then reconstitutes models for all non-basis modes using the identified models for the \( \text{ns} + 1 \) basis modes alone. Thus, once the models for the basis modes have been obtained, no further data is needed for obtaining the models for the remaining \( 2^{\text{ns}} - \text{ns} - 1 \) modes. As we show below this is possible since the dynamics of the hybrid system are constrained by the structure in Eq. (9).

We begin by defining the basis modes as the \( \text{ns} + 1 \) columns of the matrix,

\[
\beta = 
\begin{bmatrix}
I_{\text{ns} \times \text{ns}} & 0
\end{bmatrix}
= 
\begin{bmatrix}
e_1 & e_2 & \ldots & e_{\text{ns}} & e_0
\end{bmatrix}
\]

(10)

where \( I_{\text{ns} \times \text{ns}} \) is the identity matrix of size \( \text{ns} \) and 0 is the column vector consisting of \( \text{ns} \) zeros, \( e_i \)'s correspond to unit vectors and the \( e_0 \) the zero vector. As seen from Eq. (9), the hybrid system in the \( i \)th basis mode will evolve according to the following dynamic equation,

\[
\dot{x}^c = f_i + f_{e_i}, \quad \text{for } e_i \text{ basis mode}, i = 1, 2, \ldots, \text{ns}.
\]

(11)
This is because in the $i$th basis mode only the $i$th element of $\delta \in \{0, 1\}^{ns}$ is non-zero and all remaining elements are zero. On the other hand, the hybrid system, in the basis mode corresponding to $e_0$ will evolve as follows,

$$\dot{x}^{c} = f_0$$

for the origin mode.

The basis modes can thus be considered as the $ns + 1$ corners of a unit hypercube including the origin. The corresponding linear discrete-time models for the $ns + 1$ basis modes corresponding to the $ns$ unit vector and the zero vector may be represented as follows,

$$x_{k+1} = \Phi (e_i) x_k + \Gamma (e_i) u_k^{c} + f_d (e_i) \quad i = 0, 1, 2, \ldots, ns. \tag{13}$$

Note that the system matrices are location dependent. It is assumed that these $ns + 1$ basis models have been identified using input–output data. We claim that the remaining $2^{ns} - ns - 1$ models can be obtained from the $ns + 1$ basis models without further need for input–output data. In case of an infeasible basis mode, a dummy model, with all system matrices of appropriate dimension set to 0, is defined. In order to motivate the key idea behind reconstituting of the models for the remaining $2^{ns} - ns - 1$ non-basis locations, consider the linear representation of the basis models in continuous-time domain as follows,

$$\dot{x} = A (e_i) x^{c} + B (e_i) u^{c} + f (e_i) \tag{14}$$

where the system matrices of the basis model in Eq. (14) can be related to the nonlinear system of Eq. (9) as follows,

$$A (e_i) = \left. \frac{\partial f^{c}}{\partial x^{c}} \right|_{x_{i}^{c}, u_{i}^{c}, e_i} \tag{15}$$

$$B (e_i) = \left. \frac{\partial f^{c}}{\partial u^{c}} \right|_{x_{i}^{c}, u_{i}^{c}, e_i} \tag{16}$$

$$f (e_i) = f_{i} (x_{i}^{c}, u_{i}^{c}, e_i). \tag{17}$$

The discrete-time and continuous-time matrices for a system with time-invariant coefficients are related as follows,

$$\Phi (e_i) = e^{A(e_i)T} \tag{18}$$

$$\Gamma (e_i) = \int_{0}^{T} e^{A(e_i)[T(t)-t]} \, dt \tag{19}$$

$$f_d (e_i) = \int_{0}^{T} e^{A(e_i)[T(t)-t]} \, df (e_i) \tag{20}$$

where $T$ is the sampling period. At this point it is assumed that the $ns + 1$ basis models have been identified using input–output data. Then the models corresponding to the non-basis locations can be easily obtained by a binary combination of the basis models. Let $\delta_i$ be a non-basis location whose continuous state model is desired. Then $\delta_i$ can be expressed as the following binary combination of the basis vectors $e_1, e_2, \ldots, e_{ns}$ as follows,

$$\delta_i = \delta_{i,1} e_1 + \cdots + \delta_{i,ns} e_{ns} \tag{21}$$

where $\delta_{i,j}$ represent the coefficients of expansion of $\delta_i$ with respect to the basis set. Then the system matrices of the non-basis model in the continuous-time domain for any non-basis location $\delta_i$ can be obtain as follows,

$$A (\delta_i) = A (e_0) + \sum_{j=1}^{ns} \delta_{i,j} A (e_j) \tag{22}$$

$$B (\delta_i) = B (e_0) + \sum_{j=1}^{ns} \delta_{i,j} B (e_j) \tag{23}$$

$$f (\delta_i) = f (e_0) + \sum_{j=1}^{ns} \delta_{i,j} f (e_j). \tag{24}$$

Subsequently, the continuous-time system matrices (22)–(24) can be easily converted to discrete-time using appropriate transformations of Eqs. (18)–(20). Repeating the procedure for all non-basis $\delta_i$, will yield the discrete-time models for all non-basis locations from the $ns + 1$ basis models. These models constitute the identified models for a hybrid system described by Eq. (9).

An advantage of the reduced mode method is that the models for all locations are consistent with the a priori knowledge of the structure of the hybrid system embodied in Eq. (9). Thus, this approach makes lesser demands on the data set used for identification by exploiting the knowledge of the structure of the system. We demonstrate the rationale behind our approach using an example.
Example 1. Consider the model for the mode characterized by the binary variables $\delta_{1,1} = [1 \ 1 \ 0 \ 0 \ldots 0]$. From Eq. (9), the hybrid systems for this case is described by,

$$\dot{x}^c = f_0 + f_1 + f_2 \tag{25}$$

whose linearization yields

$$\dot{x}^c = \frac{\partial (f_0 + f_1 + f_2)}{\partial x^c} \bigg|_{x_0, u_0} \left[ x^c - x_0^c \right] + \frac{\partial (f_0 + f_1 + f_2)}{\partial u^c} \bigg|_{x_0, u_0} \left[ u^c - u_0^c \right]$$

$$+ f_0 \left( x_0^c, u_0^c \right) + f_1 \left( x_0^c, u_0^c \right) + f_2 \left( x_0^c, u_0^c \right). \tag{26}$$

The system matrices of the location described by Eq. (26) can be written in terms of system matrices of basis models as follows,

$$A (\delta_{1,1}) = A (e_0) + A (e_1) + A (e_2)$$

$$B (\delta_{1,1}) = B (e_0) + B (e_1) + B (e_2)$$

$$f (\delta_{1,1}) = f (e_0) + f (e_1) + f (e_2). \quad \blacksquare$$

While the reduced mode approach appears to be advantageous for hybrid systems which are linear and separable in the binary variables as described by Eq. (9), it offers no advantage for hybrid systems that are not separable in binary variables. For such non-separable hybrid systems, all $2^n$ models must be identified using input–output data. However, a majority of practical applications lie in between these two extremes, that is, part of the binary variables are linear and separable while others are non-separable. Thus, a broader class of hybrid systems which are partially separable in $\delta$ may be represented as follows,

$$\dot{x}^c = f_0 (x^c, u^c, \delta) = f_0 (x^c, u^c) + \sum_{j=1}^{n_d} \delta_{ij} f_j (x^c, u^c, \delta_n). \tag{27}$$

Here, the original binary vector is partitioned into two parts as follows, $\delta = [\delta_n \ \delta_s]^T$ where $\delta_n \in \{0, 1\}^{n_d}$ represents the part of the vector $\delta$ which is separable and $\delta_s \in \{0, 1\}^{ns - nd}$ represents the non-separable part of $\delta$. In order to use the proposed reduced mode identification method, Eq. (27) can be converted into a separable form through enumeration of the locations corresponding to the non-separable $\delta_s$ followed by their re-aggregation. This can be achieved by substituting the values for $\delta_n$ into flow-field $f_j(.)$ and hence obtaining $2^{nr}$ (where, $nr = ns - nd$) sub-flow-fields. Thus, Eq. (27) can be rewritten as follows,

$$\dot{x}^c = f_0 (x^c, u^c) + \sum_{j=1}^{n_d} \delta_{n,j} \left[ \sum_{n=1}^{2^{nr}} \ell_n (\delta_n) f_{j,n} (x^c, u^c) \right]$$

$$= f_0 (x^c, u^c) + \sum_{j=1}^{n_d} \sum_{n=1}^{2^{nr}} (\delta_{n,j} \ell_n (\delta_n)) f_{j,n} (x^c, u^c) \tag{28}$$

where $\ell_n (\delta_n)$ is a logical multiplier, which is a function of binary variables $\delta_n$ and is designed such that it takes on a value 1 when a particular combination of $\delta_n$ is encountered and 0 otherwise (see Ref. [22,23]). Note that Eq. (28) is of a separable-in-$\delta_n$ form, albeit nonlinear. New binary variables for the multiplication of $\delta_{n,j}$ and $\ell_n (\delta_n)$ along with corresponding constraints can be defined to convert the partially separable model to a separable form. Thus, the system reduces to $n_d 2^{nr} + 1$ separable functions of continuous variables which can be treated like basis models and the remaining $2^{ns} - n_d 2^{nr} - 1$ models can be obtained from these models. Thus, the partially separable model in Eq. (27) can be converted to a model that is separable and linear in the binary variables as in Eq. (9). We demonstrate this idea using an example.

Example 2. Consider a hybrid system with $ns = 5$ binary variables, of which 3 are separable and linear ($nd = 3$). Thus,

$$\dot{x}^c = f_0 (x^c, u^c) + \sum_{j=1}^{3} \delta_j f_j (x^c, u^c, \delta_4, \delta_5). \tag{29}$$

In order to obtain a separable form of the above system, the sub-flow-fields $f_j$ can be expanded as follows,

$$\dot{x}^c = f_0 (x^c, u^c) + \sum_{j=1}^{3} \delta_j \left[ (1 - \delta_4)(1 - \delta_5)f_{j,00} (x^c, u^c) \right.$$

$$\left. + (1 - \delta_5)(\delta_4)f_{j,01} (x^c, u^c) + (\delta_4)(1 - \delta_5)f_{j,10} (x^c, u^c) \right] \tag{30}$$
where \( f_{1,\alpha_1\alpha_2} (\cdot) \) indicates the function form of \( f_i (\cdot) \) when \( \delta_4, \delta_5 \in \{0, 1\} \) take on the value \( \alpha_1, \alpha_2 \). Further expansion leads to the following form

\[
\dot{x}^c = f_0 (x^c, u^c) + \left[ \ell_{1,0} f_{1,00} (x^c, u^c) + \ell_{2,0} f_{2,00} (x^c, u^c) + \ell_{3,0} f_{3,00} (x^c, u^c) + \ell_{1,0} f_{1,01} (x^c, u^c) + \ell_{2,0} f_{2,01} (x^c, u^c) + \ell_{3,0} f_{3,01} (x^c, u^c) + \ell_{1,0} f_{1,10} (x^c, u^c) + \ell_{2,0} f_{2,10} (x^c, u^c) + \ell_{3,0} f_{3,10} (x^c, u^c) + \ell_{1,1} f_{1,11} (x^c, u^c) + \ell_{2,1} f_{2,11} (x^c, u^c) + \ell_{3,1} f_{3,11} (x^c, u^c) \right]
\]

(31)

where \( \ell_i \)'s are binary variables defined by multiplication of binary variables \( \delta_1, \delta_2 \) and \( \delta_3 \) with \( \delta_4 \) and \( \delta_5 \). The corresponding constraints are merged with Eq. (7). Thus, Eq. (29) is reduced to 12 new binary variables which are linear and separable. Thus, using the reduced mode method would entail identification of models corresponding to these 13 basis locations while models for the remaining 19 (that is \( 2^5 - 13 \) ) non-basis locations can be generated based on the structure of the hybrid system of Eq. (31).

In the above example, if all 5 binary variables were in separable form, then only 6 models would require identification using input–output data to obtain the PWA representation of the hybrid system. On the other hand, if all binary variables were in a non-separable form, then all 32 models would require identification using input–output data. Thus, for the intermediate cases, the number of models which need to be identified using the input–output data varies between the two extremes of 6 and 32.

### 3.1. Identification of basis models

As discussed above the basis models correspond to the basis modes \( \delta_i \), described in Eq. (10). A standard practice in obtaining MIMO models, involves identification of individual MISO models in time series or state space form and stacking them together to obtain the MIMO model [24]. Common choices for time series models include auto-regressive exogenous (ARX), auto-regressive moving average exogenous (ARMAX), output error (OE) among others. The parameters for such models can be estimated using the prediction-error identification method (PEM) framework by solving an optimization problem [2] as follows,

\[
\hat{\theta}_N = \hat{\theta}_N (Z^N) = \min_{\theta} \sum_{k=1}^{N} \| y_k - \hat{y}_k^c (\theta) \|_n
\]

(32)

where \( \| \cdot \|_n \) denotes the \( n \)-norm, \( Z^N \) is the data set of length \( N \), \( \hat{\theta}_N \) is the vector of parameter estimates, \( y_k \) represents the measurement vector and \( \hat{y}_k^c \) the prediction. A common check for the quality of the estimates is that the residual sequence should be a white noise process [2]. This in turn, necessitates an appropriate choice of the structure and order of time-series model. State-space realizations of time-series models (input–output models) of different order typically result in different numbers of states. Also in using subspace identification methods, the choice of the state-space dimension is critical to obtaining adequate models. Thus, the \( n+1 \) basis models in the state space form may be of different state dimensions. In case the basis models are of different dimension, the addition of matrices in (22)–(24) needed in the reduced mode method is not defined and the reduced mode approach would fail. To overcome this, we have considered padding of the smaller sized matrices appropriately so that they can be combined to produce models for the non-basis locations, as demonstrated using the following example.

**Example 3.** Let us assume a hybrid system with \( q \)-inputs and two basis locations with the following state-space system matrices,

| Basis Model 1: | \([A_1]_{p \times p}, [B_1]_{p \times q} \) and \([f_1]_{p \times 1}\) |
| Basis Model 2: | \([A_2]_{n \times n}, [B_2]_{n \times q} \) and \([f_2]_{n \times 1}\) |

Let us also assume that Model-1 consists of \( p \) states and Model 2 consist of \( n \) states with \( p > n \). Here, Basis Model 2 can be padded with zeros in order to match with the size of Model 1. Thus, the augmented Basis Model 2 can be represented as follows,

\[
[A_2]_{p \times p} = \begin{bmatrix} [A_2]_{n \times n} & [0]_{n \times (p-n)} \\ [0]_{(p-n) \times n} & [0]_{(p-n) \times (p-n)} \end{bmatrix},
\]

\[
[B_2]_{p \times q} = \begin{bmatrix} [B_2]_{n \times m} \\ [0]_{(p-n) \times m} \end{bmatrix}
\]

and

\[
[f_2]_{p \times 1} = \begin{bmatrix} [f_2]_{n \times 1} \\ [0]_{(p-n) \times 1} \end{bmatrix}
\]

Now, the models can be combined as in (22)–(24) to produce the models for non-basis locations.
3.2. A composite model for all locations of hybrid system

The $ns + 1$ identified models and the remaining reconstituted models describe the $2^{ns}$ sub-models of hybrid system. The $2^{ns}$ models obtained above may then be combined using a corresponding scalar logical multiplier $\ell_i$ [22] resulting in a composite discrete-time identified model of \( (1)-(3) \), which accounts for all locations of the hybrid system as follows,

\[
x_{k+1} = \left( \sum_{i=1}^{2^{ns}} \ell_{i,k}\Phi_1 \right) x_k + \left( \sum_{i=1}^{2^{ns}} \ell_{i,k}\Gamma_1 \right) u_k + \left( \sum_{i=1}^{2^{ns}} \ell_{i,k}f_i \right)
\]

\[
y_k = \left( \sum_{i=1}^{2^{ns}} \ell_{i,k}C_1 \right) x_k
\]

\[
E_1 u_k^c + E_2 \delta_k + E_3 x_k \leq E_4.
\]

The logical multiplier $\ell_i$ is based on the variables $\delta_k$, and is designed to take on a value 1 if and only if the $i$th combination of the binary variables is encountered and zero, otherwise [22,23]. Thus, the different $\ell_i$ correspond to the modes of the system. The primary use of this identified model in MPC lies in obtaining future predictions. As described in [23], Eqs. (33)-(35) may be put in a readily usable form for MPC by defining the following terms,

\[
\tilde{L}_k = \left[ (\ell_{1,k})I \ (\ell_{2,k})I \cdots (\ell_{2^{ns},k})I \right]
\]

\[
\tilde{\Phi} = \left[ \Phi_1^T \ \Phi_2^T \cdots \Phi_{2^{ns}}^T \right]^T
\]

\[
\tilde{\Gamma} = \left[ \Gamma_1^T \ \Gamma_2^T \cdots \Gamma_{2^{ns}}^T \right]^T
\]

\[
\tilde{f} = \left[ f_{d1}^T \ f_{d2}^T \cdots \ f_{d_{2^{ns}}}^T \right]^T
\]

\[
\tilde{c} = \left[ c_1^T \ c_2^T \cdots \ c_{2^{ns}}^T \right]^T
\]

where $I$ is an identity matrix of appropriate size. Thus, the composite model of the hybrid system in Eqs. (33)-(35) takes the form,

\[
x_{k+1} = (\tilde{L}_k \tilde{\Phi}) x_k + (\tilde{L}_k \tilde{\Gamma}) u_k + \tilde{L}_k \tilde{f}
\]

\[
E_1 u_k^c + E_2 \delta_k + E_3 x_k \leq E_4
\]

\[
y_k = (\tilde{L}_k \tilde{c}) x_k
\]

where the augmented matrices $\tilde{L}_k$, $\tilde{\Phi}$, $\tilde{\Gamma}$ and $\tilde{f}$ are constituted from $\ell_{i,k}$, $\Phi_i$, $\Gamma_i$ and $f_{di}$, respectively. $\tilde{L}_k$ may be interpreted as an operator that retrieves the system matrices corresponding to the current location as identified by $\delta_k$ from the composite system matrices $\tilde{\Phi}$, $\tilde{\Gamma}$ and $\tilde{f}$. Note that the $2^{ns}$ system models in Eqs. (41)-(43) include all locations, basis as well as non-basis.

Eqs. (41)-(43) represent the final form of the identified model for all locations of the nonlinear hybrid system in the vicinity of a single operating point. One may obtain multiple such models around different operating points. These models can be combined using various weighting techniques such as Bayes’ rule to obtain prediction of the nonlinear hybrid system [25,26]. The final form of the weighted multiple models for all location is identified to that shown in Eqs. (41)-(43) with the system matrices $\tilde{\Phi}$, $\tilde{\Gamma}$, $\tilde{f}$ replaced by corresponding weighted average matrices. For details on the multiple model approach for MPC, see Ref. [23]. The multiple models may finally be used for forecasting in MPC.

In the next section we demonstrate the applicability of the reduced model identification method on an experimental three-tank system fabricated in Automation laboratory IIT Bombay. This experimental hybrid system has 6 binary variables corresponding to control events resulting in $2^6$ or 64 locations. The proposed reduced mode method is used for identification of the 7 basis models from input–output data and then reconstituting the models for the remaining 57 locations. Subsequently the identification and model predictive control is validated on the experimental setup.

4. Application

We evaluate the identification strategy on the experimental three-tank hybrid system. We briefly describe the setup along with a first principles model whose parameters were determined using experimental data followed by identification and control of experimental setup.
4.1. The three-tank benchmark hybrid system

The three-tank benchmark system has been used by a number of researchers as a benchmark to test and validate various issues arising out of the hybrid system behavior [27,28]. A similar experimental setup has been designed and fabricated at the Automation Laboratory, IIT Bombay. The setup has three cylindrical tanks each of height 60 cm and internal diameter of 15 cm. These tanks are interconnected by pipes of 0.5” diameter. The Piping & Instrumentation Diagram is shown in Fig. 1. Various details of the setup are summarized in Table 1.

As shown in Fig. 1, the system consists of two independent pumps that deliver the liquid flows $Q_1$ and $Q_2$ to Tank-1 and Tank-2, respectively through the two control valves CV-01 and CV-02. Six independent solenoid (on/off) valves ($S_1$, $S_2$, $S_{13}$, $S_{23}$, $S_{11}$ and $S_{13}$) can be manipulated to interrupt the flows into or out of the three tanks. An additional solenoid valve at the bottom of Tank-2 has been provided to simulate a leak in Tank-2. In addition, ball valves are connected in series to each of the seven solenoid valves that can be used to vary the resistances to flow. Tank-1 and Tank-3 as well as Tank-2 and Tank-3 are connected through upper and lower pipes. The lower pipes are located at the bottom of the tanks while the upper pipes are located at a height $h_v$ (0.3 m). This system exhibits typical characteristics of a hybrid dynamical system. The system transits between its locations due to the logic inputs (the solenoid valves) and continuous variables (for example, if $h_1 > h_v$ then the outflow dynamics in Tank-1 changes). The nonlinearity in the system results from the constitutive relationship between the exit flows and the level in each tank.
The flow rate through the control valves is measured by flow transmitters. The levels in the three tanks are measured by differential pressure (DP) transmitters. The details of sensors are given in Table 2. All the sensors and control valves (actuators) are interfaced with a desktop computer using a data acquisition module NI USB 6112 from National Instruments®. All solenoid valves are interfaced to the desktop computer using a separate data acquisition module, PCI-9112 from Ad-Link®.

4.2. Identification and validation of the first principles model of the setup

The six solenoid valves may be assigned binary indicator variables, whose values equal 1 when the corresponding valve is open and 0 otherwise. Thus, the opening and closing of the valves may be classified as control events. A liquid volume balance for each of the three tanks yields the dynamic model as follows [27].

\[
\begin{align*}
\frac{d h_1}{d t} &= (Q_1 - V_{13}Q_{13v_{13}} - V_1Q_{13v_1} - V_{11}Q_{11}) \\
\frac{d h_2}{d t} &= (Q_2 - V_{23}Q_{23v_{23}} - V_2Q_{23v_2}) \\
\frac{d h_3}{d t} &= (V_{13}Q_{13v_{13}} + V_{23}Q_{23v_{23}} + V_1Q_{13v_1} + V_2Q_{23v_2} - V_{N3}Q_N)
\end{align*}
\]

where \( V_1, V_2, V_{13}, V_{23}, V_{11} \) and \( V_{N3} \) represents binary indicator variables for solenoid valves \( S_1, S_2, S_{13}, S_{23}, S_{11} \) and \( S_{13} \), respectively. Variables \( Q_1 \) and \( Q_2 \) represent flow rates through control valves CV-01 and CV-02, respectively and \( A_i \) represents the cross sectional area of tank. Variables \( Q_{13}, Q_{23}, Q_{11}, Q_{21}, Q_N \) represent flow rates through the respective solenoid valves \( V_i \) and may be evaluated using the following constitutive equations [27].

\[
\begin{align*}
Q_{13v_i} &= k_{i3} \text{sign}(h_i - h_3) \sqrt{|2gh_i(h_i - h_3)|}, \quad i = 1, 2 \\
Q_{11} &= k_{11} \sqrt{2gh_1} \\
Q_N &= k_{N3} \sqrt{2gh_3}
\end{align*}
\]

The expressions for \( Q_{j3v_i} \) depends on whether the heights \( h_j \) (\( j = 1, 2, 3 \)) are greater or less than \( h_i = 0.3 \) m. Thus,

\[
Q_{j3v_i} = k_i \text{sign} \left( \max \{ h_i, h_j \} - \min \{ h_3, h_i \} \right) \sqrt{2gh_j \left( \max \{ h_i, h_j \} - \min \{ h_3, h_i \} \right)}, \quad i = 1, 2 
\]

where \( k_1, k_{13}, k_{11} \) and \( k_{N3} \) are the discharge coefficients for flow through the pair of solenoid and ball valves. These were obtained by fixing the position of the ball valves and obtaining data for steady state operation. The value of the coefficients are tabulated in Table 3.

Since \( \delta = \{ V_{13}, V_{23}, V_1, V_2, V_{11}, V_{N3} \} \in (0, 1)^6 \), a total of \( 2^6 = 64 \) locations of the hybrid system are considered. Eqs. (44)-(50) represent the first principles model (function \( f_0 \) of (9)). Fig. 2 validates the first principles model (44)-(50) (dashed line) with the plant data (solid line) in response to arbitrary changes in discrete as well as continuous inputs. From the figure it can be seen that the predictions of the first principles model match with measurements of the process setup.

4.3. Model identification for three-tank process setup

Based on the multiple model approach, two models of the form (41)-(43) are obtained using two different data sets. Model-I is based on a data set where all levels were below 0.3 m (or a low level operating point). Model-II is based on a data
set where all levels were above 0.3 m (or a high level operating point). These models were then used to predict the plant behavior using the multiple model scheme that uses Bayes’ rule for the online calculation of weights of each of the two models [25]. Since we have a validated first principles model available, three different cases were considered. In Case-I, all $2^6 = 64$ models are identified using data generated from the first principles model. In Case-II, the proposed reduced mode method is used to identify models for basis locations only using data generated from first principles model. Case-II gives us sufficient confidence in the reduced mode method, and in Case-III we identify the basis models using measurements from the experimental setup. In each of the three cases, models at two operating points as discussed above are obtained. We have used ARX models as the model structure. These ARX models are then converted to state space form to obtain the identified models of the form (41)–(43).

Case-I. Identification of models using simulation data for all location of the system.

The three-tank process setup consists of 6 binary variables (on/off valves) which results in 64 sub-models of hybrid systems. Since obtaining 64 models on the experimental setup is cumbersome, we have used the validated first principles model to generate the necessary data. To ensure that we have persistently exciting data of sufficient length corresponding to each location, we generated separated data sets corresponding to each location. Finally these models were combined as shown in Eqs. (41)–(43). This exercise was performed for both the operating points (Model-I and Model-II) and the two models were combined using Bayes’ rule. The model was then validated on the experimental setup. Fig. 3 shows the validation of the identified models using the multiple model scheme. The dashed line represents the predictions using the identified models for the three-tank while the solid line represents the experimental setup measurements for arbitrary changes in the input. Note that while the 64 models were identified using simulation data from first principles model, the validation in Fig. 3 is based on measurements from the experimental setup. From the figure, it is clear that the multiple identified model performs satisfactorily.

Case-II. Identification of models using simulation data for the $ns + 1$ basis locations.

In this case also, the first principles model is simulated to obtain $ns + 1 (=7)$ data sets corresponding to the basis locations. These data sets are subsequently used to identify models corresponding to basis locations. The non-basis models are then obtained by the proposed reduced mode identification scheme. Finally these models were combined as shown in Eqs. (41)–(43). This exercise was performed for both the operating points (Model-I and Model-II) and the two models were combined using the Bayes’ rule. The model was then validated on the experimental setup. As shown in the Fig. 4, predictions using identified models (dashed line) closely match with plant measurements (solid line). Note that while the 7 models were identified using simulation data from the first principles model, the validation in Fig. 4 is based on measurements from the experimental setup. The identified model in this case given nearly identical results as in Case-I (Compared Figs. 3 and 4) since the proposed identification method is equivalent to enumeration of all models.

Case-III. Identification of models using experimental data for the $ns + 1$ basis locations.

This case is same as Case-II with the difference that the data sets used to obtain the identified models represent measurements from the experimental setup. Like in Case-II we identify only 7 models corresponding to the basis locations. The non-basis models are then obtained by the proposed reduced mode identification scheme. Finally these models were combined as shown in Eqs. (41)–(43). This exercise was performed for both the operating points (Model-I and Model-II)
Fig. 3. Case-I: Identification of models using simulation data for all locations of the system: Experimental validation of multiple identified model of three-tank process setup. Measurements (solid line), model prediction (dashed line).

Fig. 4. Case-II: Identification of models using simulation data for the $n_t + 1$ basis locations: Experimental validation of multiple identified model of three-tank process setup. Measurements (solid line), model prediction (dashed line).

and the two models were combined using the Bayes’ rule. The model was then validated on the experimental setup. Fig. 5 documents the validation results for this case. The dashed line represents the identified models while the solid line represents measurements from the experimental setup. Here too, the identified model predicts the behavior of the setup. On comparison with Fig. 3 (Case-I) and Fig. 4 (Case-II), it is clear that the performance of the Case-III model is marginally poorer than those for Case-I and Case-II. This behavior is attributed to the fact that in this case the measurements from the setup are noisy and the ARX model structure has a poor noise model (additive white noise). This suggests the need for better noise models such as in ARMA, OE, etc (or a very high order ARX). The data used in the Case-I and Case-II was based on simulation of the first principles model, and hence this problem was not encountered. The issue of noise modeling is treated well in the literature and the current work can easily be extended. However, we do not pursue this point further.
4.4. Control of three-tank experimental setup

We have used the three different identified models discussed in Case I, II and III to achieve level control of the three-tank setup using MPC. The detailed formulation of the MPC using the multiple model approach has been discussed elsewhere [23]. Implementation of MPC requires an online solution of MINLP. The standard branch and bound (BB) approach is computationally expensive and thus impractical. In our previous work [29], we have made use of the generalized outer approximation method to achieve practical implementation of the control strategy. Figs. 6–8 document the implementation of MPC on the experimental setup for level control using the multiple identified models of Case-I, Case-II and Case-III, respectively. From these figures, it can be seen that the control using multiple identified model of Case-I and Case-II produce identical performances. However, control using multiple identified model of Case-III produces slightly different results than the first two cases. In fact, control using the multiple identified model of Case-III performs relatively poorer than the first two cases. As discussed above, this is due to the lack of appropriate noise modeling during identification from the experimental
Fig. 7. Case-II: Identification of models using simulation data for the \( n_s + 1 (= 7) \) basis locations: Experimental implementation of model predictive control of levels \( h_1, h_2, \) and \( h_3 \) in the three-tank process setup using the multiple identified models (solid line).

Fig. 8. Case-III: Identification of models using experimental data for the \( n_s + 1 (= 7) \) basis locations: Experimental implementation of model predictive control of levels \( h_1, h_2, \) and \( h_3 \) in the three-tank process setup using the multiple identified models (solid line).

The growing number of applications of hybrid dynamical systems demands efficient and rapid modeling of hybrid systems. For complex processes, a good first principles model is difficult to obtain. In this work, a novel data driven modeling scheme for the nonlinear hybrid dynamical system is proposed. The proposed scheme assumes that the locations of the data. However, the overall control performance is satisfactory. In particular, the poor quality of the model of Case-III results in a slight offset in Tank-2 (see between 1000 s and 2000 s and between 3200 s and 4500 s), which reflects in the quality of the prediction using the models. Note that the peak in Tank-3 level at 4500 s is an evidence of interaction between Tank-2 and Tank-3. A negative step change in Tank-2 causes a rise in the level of Tank-3 as the outlet is in Tank-3 only. Corresponding control moves are not provided for brevity.

5. Summary
In conclusion, the reduced mode method of identification exploits the structure of hybrid systems that are linear and separable in the binary variables. The structure acts as a constraint for the evolution of the system dynamics and the model identified by the proposed method for the different modes are consistent with this structural constraint.

References