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

A switching detection method based on projected subspace classification

José Borges, Vincent Verdult, Michel Verhaegen and Miguel Ayala Botto

Abstract— **In this paper an innovative switching detection method for piecewise linear systems is presented. The principle used for switching detection is based on finding projected subspaces from batches of input-output data, which are taken from the full data set. The method runs off-line, incrementally over all the data and, at each time, a different batch is used to compute the projected subspace. In this way, the segmentation and classification of data are entirely based on the information retrieved from the projected subspace, i.e. the subspace dimension and basis. The output of the method is a matrix of weights that assigns each pair of input-output measured data to the respective local system. Simulation experiments show the effectiveness of the proposed approach.**

I. INTRODUCTION

Piecewise linear (PWL) systems, together with piecewise affine systems, form important subclasses of hybrid systems [1], therefore, they have become highly attractive research fields in the past recent years.

The identification of both types of systems is a topic of major interest and several approaches were proposed, e.g. [2], [3] in the input-output form and [4] in the state-space form. These approaches normally assume that the data classification, i.e. the partitioning of data according to the respective local linear system, is known. This paper contributes to the solution of this last problem by defining a suitable framework and analyzing the intrinsic mechanisms.

The approach proposed in this paper stems from the classical subspace identification of LTI systems [5]. The principle is to use the order detection mechanisms of subspace identification, but applied to smaller batches of data with size $N_w \ll N$, to detect the switching between local linear systems. A moving window is used to select a small batch of data. The outputs of this batch are projected onto the orthogonal complement of the inputs, as in the LTI subspace methods. If no switching occurs in this batch, the dimension of the projected subspace equals the order of the local linear system. On the other hand, if a switching occurs within this batch, then the order will increase. By comparing the successive subspaces obtained from this moving window it is possible to classify the data according to which local model is active. The method allows the detection of switchings between local systems when the dynamics changes and as well when only the system zeros change.

This work was partially supported by the FCT grant - SFRH/BD/16370/2004, and by the FCT project POCTI/EME/44336/2002.

J. Borges and M. Ayala Botto are with GCAR, Department of Mechanical Engineering, Instituto Superior Tcnico, Technical University of Lisbon, Portugal. jborges@ist.utl.pt

Vincent Verdult and Michel Verhaegen are with the Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands v.verdult@dcsc.tudelft.nl

At the end, all batches with similar attributes are collected and associated to a local system using a weighting vector, which is then assigned to the respective input-output measured data. Assuming the measured data has enough information about all local systems, then these local systems can be identified using both the measured data and the weights that result from the switching detection method.

In [6] a concise overview on identification methods, including the data partitioning, for PWA systems is presented. Very recently two contributions appeared, [7] and [8], that tackle the problem of data partitioning using also a subspace framework. Though both also use the principle of iteratively computing subspaces from smaller batches of data, several features make a clear distinction with respect to the present work: [7] uses a PWL framework that results from running in parallel a set of LTI systems, while here it is adopted a representation where the state space is common to all models. Furthermore their test for switching detection requires the transient part of the output to vanish, while in this paper the initial system response after the switching is precisely used for detection. The authors of [8] use the same model structure as [7]. They present two methods for the embedding of inputs and outputs in a given subspace, and then a deterministic generalized principal component analysis algorithm is used to segment the data. The approach that proposed in this paper is based on projecting the outputs onto the orthogonal complement of the inputs for data classification and appears to be less complex, since it is simply based in grouping subspaces with the same order and then, among these, the ones with the same subspaces.

This paper is organized as follows. Section II describes the structure of PWL systems used in the paper and formulates the identification problem for this systems. In Section III the framework for switching detection is proposed. The main parts are the description of the projected subspace for each batch of data, the rank detection mechanism and the data classification. The simulation results are presented in Section IV.

II. PIECEWISE LINEAR SYSTEMS

A single local linear system can be described by the state-space model Σ_i,

$$\Sigma_i : \begin{cases} x(k+1) = A_i x(k) + B_i u(k) \\ y(k) = C_i x(k) + D_i u(k) \end{cases}, \tag{1}$$

where $u(k) \in \mathbb{R}^m$ is the input, $y(k) \in \mathbb{R}^\ell$ the output and $x(k) \in \mathbb{R}^n$ the embedding state. In the present paper the case when process noise $w(k)$ is added to the state equation is not treated. This problem can be addressed by formulating the model in innovations form, with $w(k) = Kv(k)$ and where K is the Kalman gain, and then by using the instrumental variables approach to deal with the noise term. Indeed, this procedure resumes to the generalization of the PO-MOESP method [5] to the current framework. The PWL system results from combining M linear models using a switching signal $p_i(k)$,

$$x(k+1) = \sum_{i=1}^{M} p_i(k)\Big(A_i x(k) + B_i u(k)\Big), \tag{2}$$

$$y(k) = \sum_{i=1}^{M} p_i(k)\Big(C_i x(k) + D_i u(k)\Big) + v(k), \tag{3}$$

where $v_i(k)$ is zero-mean white noise. The switching signal $p_i(k)$ determines which local model Σ_i is active at a given time instant k, with the restriction that only one local system can be active at a time, and should satisfy for all k:

$$p_i(k) \in \{0,1\}, \quad \sum_{i=1}^{M} p_i(k) = 1. \tag{4}$$

this formulation of the weights implies hard switchings between the local systems of (2)–(3).

The aim of the identification problem for the PWL system defined in (2)–(3) is twofold: the determination of the switching signal $p_i(k)$ and the estimation of state-space models Σ_i from a finite number N of measurements of the inputs $u(k)$ and outputs $y(k)$. This paper focus is on the detection of switching between local systems. The identification of state-space models Σ_i with known switching in the state-space framework was addressed before, e.g. by [4] using subspace identification techniques. Each local systems is assumed to be stable, observable and controllable.

III. A FRAMEWORK FOR THE SWITCHING DETECTION

In this section the switching detection method is formulated using projected subspaces. The data is classified according to the subspace dimension and basis. Several issues, e.g. dealing with noisy data, are treated and a discussion on the requirements of the method, as well as its limitations, are also presented.

A. Notation and Global Assumptions

A general form of the noisy output (3), with respect to a given initial time k_0, can be written as,

$$y(k) = \Gamma(k_0,k)x(k_0) + \sum_{j=k_0}^{k} \Phi(k,j)u(j) + \sum_{i=1}^{M} p_i(k) D_i u(k) + v_i(k), \tag{5}$$

where,

$$\Gamma(k_0,k) = \sum_{i=1}^{M} p_i(k) C_i \left(\prod_{\sigma=k_0}^{k-1} \sum_{\tau=1}^{M} p_\tau(\sigma) A_\tau \right), \tag{6}$$

$$\Phi(k,j) = \sum_{i=1}^{M} p_i(k) C_i \left(\prod_{\sigma=j+1}^{k-1} \sum_{\tau=1}^{M} p_\tau(\sigma) A_\tau \right) p_i(j) B_i. \tag{7}$$

Given a batch of outputs, which is defined as discussed in the introduction, produced by a single local system Σ_i, the terms of equation (5) can be combined into a compact form,

$$Y_{k_0,s,N_w} = \Gamma_{k_0,s,N_w} X_{k_0,N_w} + \Phi_{k_0,s} U_{k_0,s,N_w} + V_{k_0,s,N_w}. \tag{8}$$

Matrices using the notation $(\cdot)_{k_0,s,N_w}$ have a common structure and are usually called block Hankel matrices. The subscript k_0 denotes the time for the upper-left entry of the matrix, $N_w \ll N$ the number of columns and $s \ll N_w$ the number of rows. The Hankel matrices are defined in the usual way:

$$Y_{k_0,s,N_w} := \begin{pmatrix} y(k_0) & y(k_0+1) & \cdots & y(k_0+N_w-1) \\ y(k_0+1) & y(k_0+2) & \cdots & y(k_0+N_w) \\ \vdots & \vdots & \ddots & \vdots \\ y(k_0+s-1) & y(k_0+s) & \cdots & y(k_0+N_w+s-2) \end{pmatrix}$$

with U_{k_0,s,N_w} and V_{k_0,s,N_w} defined following the same pattern. Given $1 \le i \le s$ and $1 \le j \le N_w$, the columns of Γ_{k_0,s,N_w} are shortly defined as follows:

$$\Gamma_{k_0,s,N_w}(i\ell, (j-1)n+1 : jn) := \Gamma(k_0+j-1, k_0+i-1).$$

The matrix with the state vectors is block diagonal:

$$X_{k_0,N_w} := \operatorname{diag}\left(x(k_0), x(k_0+1), \ldots, x(k_0+N_w-1)\right). \tag{9}$$

Now, with $1 \le i \le s$ and $1 \le j \le s$, matrix $\Phi_{k_0,s}$ is defined as:

$$\Phi_{k_0,s} := \begin{cases} 0, & i > j \\ \sum_{\tau=1}^{M} p_\tau(k_0) D_\tau, & i = j \\ \Phi(k_0+j-1, k_0+i-1). & i < j \end{cases}. \tag{10}$$

This notation can be sometimes relaxed to lighten the presentation along the paper. This means that the subscripts k_0 and s can be sometimes omitted. It is assumed that the input $u(k)$ is persistently exciting of order s, i.e. $PE(s)$, such that

$$\operatorname{rank}\left(U_{k_0,s,N_w} U_{k_0,s,N_w}^T\right) = sm. \tag{11}$$

Furthermore, it is also assumed that input $u(k)$ and the measurement white noise $v(k)$ are ergodic uncorrelated data sequences. Adopting a notation similar to [9], the expected value operator is defined as,

$$\mathbb{E}_N\left[\sum_{j=0}^{N-1} u(j)v(j)^T\right] = \lim_{N\to\infty}\left[\frac{1}{N}\sum_{j=0}^{N-1} u(j)v(j)^T\right]. \tag{12}$$

Due to the ergodicity of data, $\mathbb{E}_N[\cdot]$ applies to the average over one *infinitely* long experiment, while the common operator $\mathbb{E}[\cdot]$ denotes the average over an infinite number of experiments. The use of $\mathbb{E}_{N_w}[\cdot]$ implies a tradeoff for the choice of the window size N_w: it should be small enough to capture fast switching systems and large enough to keep the statistic properties.

B. Projected Subspaces

When the window of data contains a batch of outputs exclusively from the same local system - the case when there is a switching is treated in section III-C - and using the assumption that $u(k)$ is $PE(s)$, the influence of the inputs can be removed from the data equation (8),

$$Y_{N_w}\Pi^{\perp}_{U_{N_w}} = \Gamma_{N_w}X_{N_w}\Pi^{\perp}_{U_{N_w}} + V_{N_w}\Pi^{\perp}_{U_{N_w}}. \tag{13}$$

The operator $\Pi^{\perp}_{U_{N_w}}$ is an orthogonal projector onto the column-space of U_{N_w}, which is defined as,

$$\Pi^{\perp}_{U_{N_w}} := I_{N_w} - U^T_{N_w}\left(U_{N_w}U^T_{N_w}\right)^{-1}U_{N_w}, \tag{14}$$

and, therefore, it satisfies the property $\Phi_{k_0,s}U_{N_w}\Pi^{\perp}_{U_{N_w}} = 0$. By projecting out the influence of the inputs from the data equation (8) only two components remain in the output: the dynamical behavior of the system, and the projected additive measurement noise. The next pair of Lemmas provide effective means to decouple the influence of each of these components in the projected outputs.

Lemma 1: Assuming $u(k)$ is $PE(s)$ and all $x(k) \neq 0$, then,

$$\text{rank}\left(\Gamma_{N_w}X_{N_w}\Pi^{\perp}_{U_{N_w}}\right) = \text{rank}\left(\Gamma_{N_w}\right).$$

The Proof of Lemma 1 is a generalization from the LTI case to the current framework, and is not presented here due to space constraints. Next Lemma also generalizes to PWL systems a result for the LTI case, which is due to [9], on the use of singular value decomposition (SVD) to retrieve the column- and row-spaces of noisy matrices.

Lemma 2: Given the minimal system (2)–(3). If the input $u(k)$ is $PE(s)$ and uncorrelated with $v(k)$, which is an ergodic zero mean white noise sequence whose variance σ_v^2 is, possibly, unknown,

$$\mathbb{E}_{N_w}\left[V_{i,s,N_w}V^T_{i,s,N_w}\right] = \sigma_v^2 I_{ls},$$
$$\mathbb{E}_{N_w}\left[U_{N_w}V^T_{i,s,N_w}\right] = 0.$$

Then the following relation holds,

$$\text{range}\left(Y_{N_w}\Pi^{\perp}_{U_{N_w}}\right) = \text{range}\left(\Gamma_{N_w}\right).$$

The Proof of Lemma 2 is also not presented here due to space constraints.

1) Noise removal: In Lemma 1 it is assumed that the SVD of $\Gamma_{N_w}X_{N_w}\Pi^{\perp}_{U_{N_w}}$ is available, though in practical applications this situation never happens because only the outputs are accessible. Therefore the SVD of the noisy matrix $Y_{N_w}\Pi^{\perp}_{U_{N_w}}$ has to be computed, resulting also in a "noisy" singular values decomposition. In this work it is assumed that the measurement noise is white, therefore its effect in the matrix of singular values can be filtered out by left-multiplying this matrix with a weighting matrix W.

Actually, the SVD is often used as a robust tool for additive noise filtering, e.g. [10] shows that the overall procedure of order reduction using SVD is equivalent to a FIR-filtering operation on the noisy signal $y(k)$.

The weighting matrix W can be obtained as,

$$W = \text{diag}\left(1 - \frac{\sigma_v^2}{\sigma_1^2}, \ldots, 1 - \frac{\sigma_v^2}{\sigma^2_{s\ell+n_w}}\right). \tag{15}$$

Assuming the Signal to Noise Ratio (SNR) allows to distinguish the measurement noise from the output signal, i.e. there is a clear gap between the singular values related to the system and the ones related with the noise, the variance σ_v^2 can be computed from the singular values using a minimum variance estimator [9], though other approaches could be used [11]. The dimension n_w of the projected subspace is then revealed by checking the rank of the filtered singular values matrix.

Lemma 2 shows that the computation of the column space for $Y_{N_w}\Pi^{\perp}_{U_{N_w}}$ does not change in the presence of additive white noise in the output signal. This is not the case when the measurement noise is coloured. The treatment of the switching detection for nonwhite additive noise is a topic of current research.

2) Efficient computation of the projected output: Similarly to the procedure described in [5], an effective way to compute the projected output results from using *RQ*-factorization.

C. Switching Detection

The complete projection of the inputs is possible only if all the outputs in the batch of data are produced by the same local system. Let assume now that the batch of output data contains outputs collected from two successive local systems Σ_p and Σ_q, which were activated as defined in Section II for the weights, then the outputs can be collected as follows,

$$Y_{k_0,s,N_w} = \begin{bmatrix} Y^- \\ Y^+ \end{bmatrix}, \tag{16}$$

where the superscript $(\cdot)^-$ stands for a row-wise partition that contains data related with the first local system Σ_p, with size $[r \times N_w]$, and superscript $(\cdot)^+$ stands for a nonempty row-wise partition that contains data both from Σ_p and Σ_q, with size $[s - r \times N_w]$. Next Lemma states the mechanism for the subspace projection when the window of data reaches a switching between successively activated local systems.

Lemma 3: Given a batch of data with size N_w. If the initial time is k_0 and the batch of data takes outputs from two successively activated systems Σ_p and Σ_q, then the projection of the mixed part Y^+ satisfies:

$$Y^+\Pi^{\perp}_{U_{N_w}} = \Gamma^+X_{N_w}\Pi^{\perp}_{U_{N_w}} + \sum_{i=0}^{N_w-1}\Phi^+_{k_0+i,s-r}\mathscr{U}(k_0+i)\Pi^{\perp}_{U_{N_w}} + V^+\Pi^{\perp}_{U_{N_w}}$$

Proof: In the case when the window has both output from Σ_i and Σ_{i+1}, the data equation (8) is replaced by,

$$\begin{bmatrix} Y^- \\ Y^+ \end{bmatrix} = \begin{bmatrix} \Gamma^- \\ \Gamma^+ \end{bmatrix} X_{N_w} + \begin{bmatrix} \Phi^- \\ 0 \end{bmatrix} U_{k_0,s,N_w} + \sum_{i=0}^{N_w-1}\begin{bmatrix} 0 \\ \Phi^+_{k_0+i,s-r} \end{bmatrix}\mathscr{U}(k_0+i) + \begin{bmatrix} V^- \\ V^+ \end{bmatrix}, \tag{17}$$

$$
Y_{0,3,3}\Pi^{\perp}_{U_{0,3,3}} = \left[\begin{array}{c} Y^- \\ \hline Y^+ \end{array}\right] \Pi^{\perp}_{U_{0,3,3}}
$$

$$
= \left[\begin{array}{ccc} C_1x(0) & C_1x(1) & C_1x(2) \\ \hline C_1A_1x(0)+C_1B_1u(0)+D_1u(1) & C_1A_1x(1)+C_1B_1u(0)+D_1u(1) & \mathbf{C_2A_1x(2)+C_2B_1u(0)+D_2u(1)} \\ C_1A_1^2x(0)+C_1A_1B_1u(0)+C_1B_1u(1)+D_1u(2) & \mathbf{C_2A_1^2x(1)+C_2A_1B_1u(0)+C_2B_1u(1)+D_2u(2)} & \mathbf{C_2A_2A_1x(2)+C_2A_2B_1u(0)+C_2B_2u(1)+D_2u(2)} \end{array}\right] \Pi^{\perp}_{U_{0,3,3}}
$$

Fig. 1. Illustrative example of the partitioning for the projected output matrix $Y_{k_0,3,3}\Pi^{\perp}_{U_{k_0,3,3}}$ according to definition (16). The entries in bold belong to the local linear system that is active after the switching. The initial time instant is $k_0 = 0$, the number of rows $s = 3$ and the window dimension $N_w = 3$.

where $U(k_0+i)$ is defined as a matrix of zeros, except for the (k_0+i)-th column that is equal to same column in the input Hankel matrix U_{k_0,s,N_w},

$$
\mathscr{U}(k_0+i) := \left[\begin{array}{ccc} 0_{s\times k_0+i-1} & U_{k_0+i,s,1} & 0_{s\times N_w-k_0+i} \end{array}\right]. \quad (18)
$$

By applying the projector $\Pi^{\perp}_{U_{N_w}}$ to equation (17), and after some straightforward manipulations results,

$$
\left[\begin{array}{c} Y^-\Pi^{\perp}_{U_{N_w}} \\ Y^+\Pi^{\perp}_{U_{N_w}} \end{array}\right] = \left[\begin{array}{c} \Gamma^- X_{N_w}\Pi^{\perp}_{U_{N_w}} \\ \Gamma^+ X_{N_w}\Pi^{\perp}_{U_{N_w}} \end{array}\right] + \left[\begin{array}{c} 0 \\ \sum_{i=0}^{N_w-1} \Phi^{+}_{k_0+i,s-r}\mathscr{U}(k_0+i)\Pi^{\perp}_{U_{N_w}} \end{array}\right] + \left[\begin{array}{c} V^-\Pi^{\perp}_{U_{N_w}} \\ V^+\Pi^{\perp}_{U_{N_w}} \end{array}\right], \quad (19)
$$

which means that also part of the inputs will show up in the output at the switching time. ■

Lemma 3 is illustrated in Figure 1 for the "noiseless" case,

$$
\left[\begin{array}{c} Y^-\Pi^{\perp}_{U_{N_w}} \\ Y^+\Pi^{\perp}_{U_{N_w}} \end{array}\right] - \left[\begin{array}{c} V^-\Pi^{\perp}_{U_{N_w}} \\ V^+\Pi^{\perp}_{U_{N_w}} \end{array}\right] = \left[\begin{array}{c} \Gamma^- X_{N_w}\Pi^{\perp}_{U_{N_w}} \\ \Gamma^+ X_{N_w}\Pi^{\perp}_{U_{N_w}} \end{array}\right] + \left[\begin{array}{c} 0 \\ \sum_{i=0}^{N_w-1} \Phi^{+}_{k_0+i,s-r}\mathscr{U}(k_0+i)\Pi^{\perp}_{U_{N_w}} \end{array}\right]. \quad (20)
$$

Next Lemma characterizes the relation between the rank of the projected outputs for batches of data taken before and during the switching times.

Lemma 4: Under the assumptions that:

i) All local systems Σ_i, with $i = 1,\dots,M$, are both observable and controllable.

ii) The input is $PE(s)$.

Then,

$$
\operatorname{rank}\left(Y_{j_1,N_w}\Pi^{\perp}_{U_{j_1,N_w}}\right) < \operatorname{rank}\left(Y_{j_2,N_w}\Pi^{\perp}_{U_{j_2,N_w}}\right) \leq s, \quad (21)
$$

where $Y_{j_1,N_w}\Pi^{\perp}_{U_{j_1,N_w}}$ and $Y_{j_2,N_w}\Pi^{\perp}_{U_{j_2,N_w}}$ are the projected outputs for two successive batches of data starting at $k_0 = j_1$ and $k_0 = j_2$, respectively, and with $j_2 > j_1$. The first batch takes data from only one local system Σ_p, while the second batch takes data from two local systems Σ_p and Σ_q.

Proof:

i. Naturally, $\operatorname{rank}(Y^-\Pi^{\perp}_{U_{N_w}}) \leq \operatorname{rank}(Y_{j_1,N_w}\Pi^{\perp}_{U_{j_2,N_w}})$

ii. Consider the SVD of the left term in equation (20),

$$
\left[\begin{array}{c} Y^-\Pi^{\perp}_{U_{N_w}} - V^-\Pi^{\perp}_{U_{N_w}} \\ Y^+\Pi^{\perp}_{U_{N_w}} - V^+\Pi^{\perp}_{U_{N_w}} \end{array}\right] = USV^T, \quad (22)
$$

and as well the following SVD,

$$
\left[\begin{array}{c} \Gamma^- X_{N_w}\Pi^{\perp}_{U_{N_w}} \\ \Gamma^+ X_{N_w}\Pi^{\perp}_{U_{N_w}} \end{array}\right] = U_\Gamma S_\Gamma V_\Gamma^T, \quad (23)
$$

$$
\left[\begin{array}{c} 0 \\ \sum_{i=0}^{N_w-1} \Phi^{+}_{k_0+i,s-r}\mathscr{U}(k_0+i)\Pi^{\perp}_{U_{N_w}} \end{array}\right] = U_\Phi S_\Phi V_\Phi^T. \quad (24)
$$

By replacing the SVD results of (22)–(23) into equation (20), and then after some straightforward manipulations to isolate the singular values of (20) results,

$$
\underbrace{U^T\, U_\Gamma}_{\mathbf{U}_\Gamma}\, S_\Gamma\, \underbrace{V_\Gamma^T\, V}_{\mathbf{V}_\Gamma{}^T} + \underbrace{U^T\, U_\Phi}_{\mathbf{U}_\Phi}\, S_\Phi\, \underbrace{V_\Phi^T\, V}_{\mathbf{V}_\Phi{}^T} = S, \quad (25)
$$

which can be combined into a valid SVD,

$$
\left[\begin{array}{cc} \mathbf{U}_\Gamma & \mathbf{U}_\Phi \end{array}\right] \left[\begin{array}{cc} S_\Gamma & 0 \\ 0 & S_\Phi \end{array}\right] \left[\begin{array}{c} \mathbf{V}_\Gamma^T \\ \mathbf{V}_\Phi^T \end{array}\right] = S. \quad (26)
$$

Expression (26) decouples the contribution of each term to the dimension of the projected outputs (20):

$$
\operatorname{rank}(S) = \operatorname{rank}(S_\Gamma) + \operatorname{rank}(S_\Phi). \quad (27)
$$

iii. Using the controllability and observability assumptions then the term $\sum_{i=0}^{N_w-1} \Phi^{+}_{k_0+i,s-r}\mathscr{U}(k_0+i)\Pi^{\perp}_{U_{N_w}}$ is full rank. Furthermore, if at time $k_0 = \tau$ the part $Y^-\Pi^{\perp}_{U_{N_w}} = 0$, then $Y^+\Pi^{\perp}_{U_{N_w}} = Y_{\tau,N_w}\Pi^{\perp}_{U_{\tau,N_w}}$ and,

$$
\operatorname{rank}\left(Y_{\tau,N_w}\Pi^{\perp}_{U_{\tau,N_w}}\right) = s. \quad (28)
$$

When the window of data incrementally enters the switching time the pure part $Y^-\Pi^{\perp}_{U_{N_w}}$ will be successively replaced by mixed $Y^+\Pi^{\perp}_{U_{N_w}}$ and the rank will increase, therefore proving inequality (21). ■

1) Reverse case: When the batch of data "leaves" the switching period the notation $(\cdot)^-$ and $(\cdot)^+$ is permuted and, therefore, the results of previous lemmas still hold, as long as inequality (21) in Lemma 4 is reversed, which means that the order decreases in this case.

2) Types of changes in the system detected by the method: The key term for the rank accumulation feature used in switching detection is $\sum_{i=0}^{N_w-1} \Phi^+_{k_0+i,s-r} \mathscr{U}(k_0+i)\Pi^\perp_{U_{N_w}}$. Due to the structure of matrix $\Phi^+_{k_0+i,s-r}$, which takes matrices A_i, B_i, C_i and D_i, the switching detection method used in this framework is able to detect changes when the successive local systems:

- are completely different.
- change only in the dynamics, i.e. poles of the system.
- change only the system zeros, while keeping constant the dynamics.

D. Data classification

The batches of data are grouped if the respective subspaces have equal dimension and basis, which is equivalent to say that these data samples are related to the same local system. The test to compare the subspace dimension is straightforward. The test to to compare the subspace basis relies on the distance between the respective projected subspaces. According to [12], a measure for the distance between two projected subspaces can be efficiently computed using,

$$d_{jp} = 1 - \sigma_{\min}\left(U^T_{\Sigma_j} U_{\Sigma_p}\right)^2, \tag{29}$$

where the matrices U_{Σ_j} and U_{Σ_p} are the left singular vectors for the systems Σ_j and Σ_p, respectively. The noise properties are assumed to be constant. If the projected distance satisfies $d_{jp} \leq \delta \simeq 0$, then the subspaces are coincident and, therefore, the batches of data should be assigned to the same local system Σ_i. The constant δ is introduced as a statistical threshold. The test to the system zeros is performed only when the systems have the same dynamics. The parameters $\widehat{\theta}_i := [\mathrm{vec}(B_i)^T \mathrm{vec}(D_i)^T]^T$ are estimated using a a least squares setting, like in [5], for each local systems $i = p$ and $i = q$, and then compared using a suitable metric, e.g. the Mahalanobis distance,

$$d_{R_\theta} = \sqrt{\left(\widehat{\theta}_p - \widehat{\theta}_q\right)^T R_\theta^{-1} \left(\widehat{\theta}_p - \widehat{\theta}_q\right)}. \tag{30}$$

Initially the batches are grouped according to the respective subspace dimension. When consecutive batches have equal dimension and projected distance $d_{jp} \leq \delta$, these are immediately concatenated and assigned to the same local system.

The general procedure for the assignment of data can be summarized as follows:

Algorithm: *data classification*,

(I) $k \leftarrow k_0$, $p \leftarrow 0$, $M \leftarrow 1$, $i \leftarrow M$
(II) compute $Y_{k,s,N_w}\Pi^\perp_{U_{k,s,N_w}} = U_{[k]}S_{[k]}V^T_{[k]}$
(III) model set $\mathscr{M} \leftarrow \mathscr{M}_i$
(IV) $p_i(k : k+N_w-1) \leftarrow 1$
(V) **while** $k < N - N_w + 1$
  a) $k \leftarrow k+1$
  b) update $Y_{k,s,N_w}\Pi^\perp_{U_{k,s,N_w}}$
  c) compute $U_{[k]}$ and $S_{[k]}$
  d) **if** $\rho_{[k]} \neq \rho_{[k-1]}$
    i) **if** $\{U_{[k]}, S_{[k]}\} \supset \mathscr{M}$
      A) select i
      else if $k - k_0 > N_w + s - 2$
      B) model set $\mathscr{M} \leftarrow \mathscr{M} \cup \mathscr{M}_i$
      C) $M \leftarrow M+1$
      D) $i \leftarrow M$
    ii) **end**
    iii) $k_0 \leftarrow k$
    else if $k - k_0 > N_w + s - 2$
    iv) $p_i(k_0 : k+N_w-2) \leftarrow 1$
  e) **end**
(VI) **end**

The set of all models is denoted by $\mathscr{M}$ and consists of local models characterized as $\mathscr{M}_i := \{U_i, S_i\}$, $i = 1, \ldots, M$. The rank of the subspace whose batch of data starts at time k is denoted by $\rho_{[k]}$. A unit value is assigned to the weights indexed to data belonging to the same local system, as defined in (4). The output of the method is, therefore, a matrix of weights that assigns each input-output sample pair to a given local system Σ_i.

In the conditions of Algorithm 1 the window moves one step at a time and if the subspace dimension changes, then the subsequent batch is only assigned to a new local system when the dimension stays constant for more $N_w + s - 2$ iterations. This algorithm is presented here only to illustrative the switching detection method. More elaborated algorithms, based e.g. in tree search or recursive SVD computation, are currently under research.

IV. SIMULATION EXAMPLES

To illustrate the effectiveness of the detection method, two different cases are presented in this section:

1) A PWL system where 1st order local system Σ_1 commutes with a 2nd order local system Σ_2:

$$\Sigma_1 : A_1 = 0.45, B_1 = 0.4, C_1 = 0.27, D_1 = 0.32,$$
$$\Sigma_2 : A_2 = \begin{bmatrix} 0 & 0.8 \\ -0.4 & 0.01 \end{bmatrix}, B_2 = \begin{bmatrix} 0.24 \\ 0.3 \end{bmatrix},$$
$$C_2 = \begin{bmatrix} 0.19 & 0.45 \end{bmatrix}, D_2 = 0.$$

2) A PWL system where 2nd order local system Σ_3 and Σ_4 commute between each other:

$$\Sigma_3 : A_3 = A_2, B_3 = B_2, \quad \Sigma_4 : A_4 = A_3, B_4 = \begin{bmatrix} 0.4 \\ 0 \end{bmatrix},$$
$$C_3 = C_2, D_3 = D_2, \qquad C_4 = C_3, D_4 = D_3.$$

100 Monte Carlo experiments were performed, where at each run the input sequences $u(k)$ are pseudo-random numbers and the measurement noise sequences $v(k)$ are white noise with variance designed such that the SNR of 37.95 dB holds. This value was selected because higher values will result in a series of "false alarms" due to the approach used to compute the weighting matrix W in (15). The switching signal, i.e. the weights $p(k)$ for the simulations, is also randomly generated in each case taking into account the restrictions of (4). The

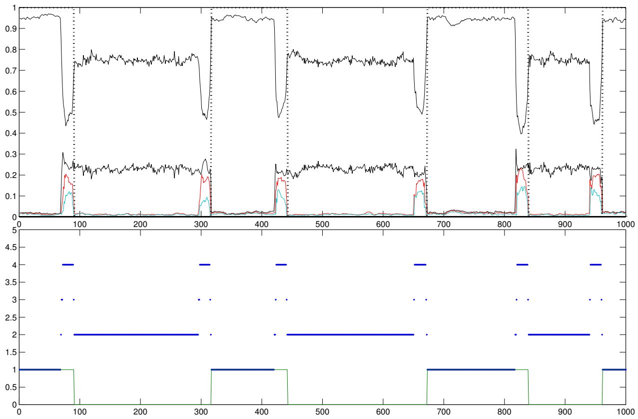

Fig. 2. Results for switching detection using Σ_1 and Σ_2. Top: normalized singular values (system and noise), the vertical lines represent the switching times. Bottom: dimension of the projected subspaces (dots) and estimated vector of weights p_1 for local system Σ_1 (solid line).

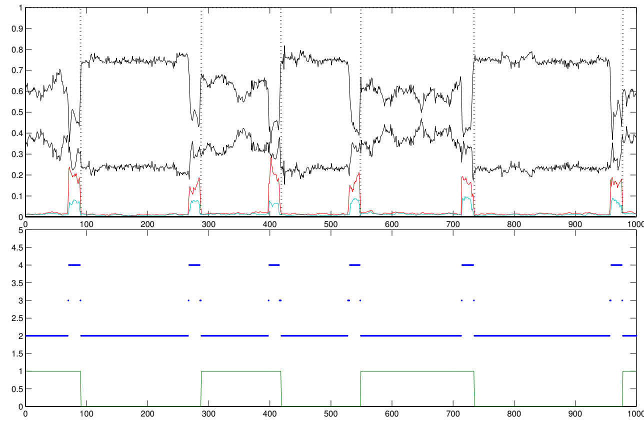

Fig. 3. Results for switching detection using Σ_3 and Σ_4. Top: normalized singular values (system and noise), the vertical lines represent the switching times. Bottom: dimension of the projected subspaces (dots) and estimated vector of weights p_3 for local system Σ_3 (solid line).

constant parameters along the 100 experiments are the batch size $N_w = 20$ and the number of rows for each Hankel matrix is $s = 4$.

A typical plot for one of the Monte Carlo experiments is presented in figures 2, in the case of PWL system $\{\Sigma_1, \Sigma_2\}$, and in figures 3, in the case of PWL system $\{\Sigma_3, \Sigma_4\}$. The top plot for both figures 2 and 3 present the normalized singular values, and the bottom plot present subspace dimension and weights for one of the local linear systems.

Figure 2 shows the case when the 1st order local system Σ_1 switches with the 2nd order local system Σ_2. The initial 90 data samples correspond to system Σ_1 active, which means that only one singular value is related with the system and the remaining with measurement noise. When the batch of data enters the switching, in the conditions of Lemma 4, the number of singular values will increase, which reveals the signature of switching. As the window leaves the switching, the number of energetic singular values will decrease again until its number is steady, which means that the data is related only with one local system, in this case Σ_2. The process repeats until the end of the data sequence. Similar behavior is also verified for local systems of figure 3, i.e. when the batches of data enters the switching times the “less energetic” singular values related with noise start to increase, until the time that the batch contains only mixed data from the two local systems, meaning that the subspace dimension is maximum and equal to $s = 4$. When the batch leaves the switching the dimension reduces until a steady values that reflects the dimension of the respective local system. When the batch takes data from only one local system the singular values related to noise keep approximately a constant low value that reflects the noise variance.

V. CONCLUSIONS

The switching between local linear systems of PWL systems can be successfully detected by applying the switching detection framework, which was presented in this paper, to the available system input-output measurements. This method is based on the detection of rank variations for projected subspaces that are computed from successive batches of data. The data is classified by assigning a unitary weight to the data whose respective subspace dimension and projected distance are equal. Simulation results, which were obtained using different types of systems, confirm the effectiveness of the method in the conditions defined in the paper. Future work include the treatments of other types of noise and development of efficient algorithms for data assignment.

REFERENCES

[1] W. P. M. H. Heemels, B. De Schuter, and A. Bemporad. Equivalence of hybrid dynamical models. *Automatica*, 37(7):1085–1091, July 2001.

[2] A. Bemporad, A.and Garulli, S. Paoletti, and A. Vicino. A greedy approach to identification of piecewise affine models. In O. Maler and A. Pnuelli, editors, *Hybrid Systems: Computation and Control*, number 2623 in LNCS, pages 97–112. Springer-Verlag, 2003.

[3] R. Vidal, S. Soatto, Y. Ma, and S. Sastry. An algebraic geometric approach to the identification of a class of linear hybrid systems. *42rd IEEE Conf. Dec. Cont. - Hawaii*, December 2003.

[4] V. Verdult and M. Verhaegen. Subspace identification of piecewise linear systems. *43rd IEEE Conf. Dec. Cont. - Bahamas*, pages 3838–3843, December 2004.

[5] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.

[6] Jacob Roll. *Local and Piecewise Affine Approaches to System Identification*. PhD thesis, Apr 2003.

[7] M. Pekpe, G. Mourot, K. Gasso, and J. Ragot. Identification of switching systems using change detection technique in the subspace framework. *43rd IEEE Conf. Dec. Cont. - Bahamas*, pages 3720–3725, December 2004.

[8] K. Huang, A. Wagner, and Y. Ma. Identification of hybrid linear time-invariant systems via subspace embedding and segmentation (ses). *43rd IEEE Conf. Dec. Cont. - Bahamas*, pages 3227–3234, December 2004.

[9] B. Moor. The singular value decomposition and long and short spaces of noisy matrices. *IEEE Trans. Signal Proces.*, 41(9):2826–2838, December 1993.

[10] P. C. Hansen and S. H. Jensen. Fir filter representation of reduced-rank noise reduction. *IEEE Trans. Signal Proces.*, 46(6):1737–1741, June 1998.

[11] K. Konstantinides. Statistical analysis of effective singular values in matrix rank determination. *IEEE Trans. Acous. Speech Sig. Proces.*, 36(5):757–763, May 1988.

[12] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins, Baltimore, 3rd edition, 1996.