Identification of linear hybrid systems: a geometric approach

van Luong Le, Fabien Lauer, Gérard Bloch

Identification of linear hybrid systems: a geometric approach

Van Luong Le, Fabien Lauer and Gérard Bloch

Abstract— This paper deals with the identification of linear hybrid systems switching between multiple linear subsystems. We propose a new approach based on the geometric properties of hybrid systems in parameter space. More precisely, the data are mapped in that space such that each submodel is represented by a hypersphere. Then, we show how these hyperspheres can be easily separated by Principal Component Analysis (PCA) and derive a condition under which this separation is optimal for systems with two modes. Finally, classical (robust) regression is applied to estimate the system parameters from the classified data set. A simple procedure is also proposed to extend the method to the identification of switched systems with multiple modes. Experiments show that the final algorithm can accurately estimate both the parameters and the number of modes while being simple to apply and far more robust to noise than other methods.

I. INTRODUCTION

This paper deals with the identification of linear hybrid systems, i.e., systems switching between linear or affine subsystems. More precisely, we consider SISO systems in Switched AutoRegressive with eXogenous input (SARX) form as

$$y_i = \boldsymbol{\theta}_{\lambda_i}^\top \boldsymbol{\varphi}_i + e_i, \tag{1}$$

where $\lambda_i \in \{1, \ldots, n\}$ is the discrete state or mode with n the number of submodels, $\{\boldsymbol{\theta}_j\}_{j=1}^n$ are the parameter vectors of the submodels, $e_i \in \mathbb{R}$ is a noise term and $\boldsymbol{\varphi}_i = [y_{i-1}, \ldots, y_{i-n_a}, \ u_{i-1}, \ldots, u_{i-n_b}]^\top \in \mathbb{R}^p$, $p = n_a + n_b$, is the regression vector with n_a and n_b being the model orders. In the case of affine subsystems, the regression vector $\boldsymbol{\varphi}_i$ is simply replaced by $\tilde{\boldsymbol{\varphi}}_i = [\boldsymbol{\varphi}_i^\top, 1]^\top$. As in most works, the following identification problem is considered.

Problem 1. *Given a collection $\mathcal{S}$ of N input-output pairs $(\boldsymbol{\varphi}_i, y_i)$, $i = 1, \ldots, N$, generated by a hybrid system* (1), *estimate:*

- *the submodels $f_j(\boldsymbol{\varphi}) = \boldsymbol{\theta}_j^\top \boldsymbol{\varphi}$,*
- *the switching sequence $\{\lambda_i\}_{i=1}^N$,*
- *the number of submodels n.*

Related work. Many approaches have been proposed over the last decade for the identification of hybrid systems [1]. These include the algebraic approach [2] and various convex [3], [4], [5], [6] and nonconvex [7], [8], [9], [10], [11], [12] optimization-based approaches. The algebraic approach [2] gives the exact parameter vectors but only in the noiseless

V.L. Le and G. Bloch are with the Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, France, van-luong.le@univ-lorraine.fr, gerard.bloch@esstin.uhp-nancy.fr
F. Lauer is with the LORIA, Université de Lorraine, CNRS, Inria, France, fabien.lauer@loria.fr

case and is rather sensitive to noise otherwise. While the nonconvex optimization-based approaches do not guarantee a global solution, the convex ones typically use a convex relaxation of a non-convex problem and require some data conditions [4] to guarantee the equivalence. Note that the global optimization approach of [8] is only applicable to hinging hyperplane systems and with a small data set due to a prohibitive computing time. Besides these particular limitations, all these methods also intrinsically include some important hyperparameters to tune.

Contribution. This paper presents a new approach for the identification of linear hybrid systems. We consider their properties directly in the parameter space instead of in the data space where other methods typically minimize an error criterion. In the parameter space, we derive a mapping of the data to construct hyperspheres which represent each of the submodels. Then, for hybrid systems with two modes, we propose a simple method based on Principal Component Analysis (PCA) to separate the two hyperspheres. After this data classification step, the submodel parameter vectors can be estimated by any robust linear regression method. Contrary to other methods in the literature, this approach does not include any hyperparameter to tune and does not rely on non-convex optimization. Finally, a simple iterative procedure is also proposed for the identification of switched systems with multiple modes.

Paper organization. We start in Sect. II by emphasizing main properties of hybrid systems in parameter space. Then, Sect. III is dedicated to solving the identification problem of hybrid systems with two modes, while Sect. IV extends the proposed method to systems with multiple modes. The paper ends with numerical examples in Sect. V and conclusions in Sect. VI.

II. HYPERSPHERES OF HYBRID SYSTEMS

In this section, the identification problem is considered from the viewpoint of the parameter space, first for the noiseless case in Sect. II-A, then for the noisy one in Sect. II-B.

A. Noiseless case

In the parameter space $\mathbb{R}^p$ of $\boldsymbol{\theta}$, $p = n_a + n_b$, a data pair $(\boldsymbol{\varphi}_i, y_i)$ defines a subspace of dimension $p - 1$ (a hyperplane) $\mathbb{P}_i$ as

$$\mathbb{P}_i: \quad \boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i = 0. \tag{2}$$

Property 1. *Let $\mathcal{S}_j$ be the data subset of $\mathcal{S}$ generated by the jth subsystem, if there are at least p linearly independent*

data points in $\mathcal{S}_j$ then the corresponding parameter vector $\boldsymbol{\theta}_j$ is at the intersection of all $\mathbb{P}_i$ with $(\boldsymbol{\varphi}_i, y_i) \in \mathcal{S}_j$, i.e.,

$$\boldsymbol{\theta}_j = \bigcap_{i \in I_j} \mathbb{P}_i, \tag{3}$$

where $I_j = \{i : (\boldsymbol{\varphi}_i, y_i) \in \mathcal{S}_j\}$.

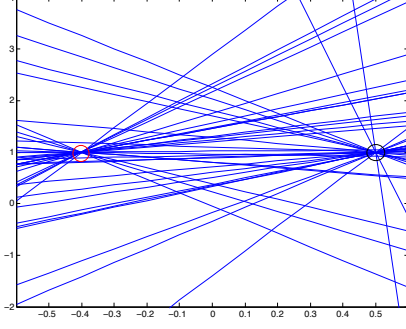Figure 1 illustrates Property 1 for a noiseless data set generated by a SARX system with two modes.

Fig. 1. Parameter space $\mathbb{R}^2$ of a SARX system with 2 modes. The hyperplanes (lines) $\mathbb{P}_i$ intersect either at $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$.

As all hyperplanes $\{\mathbb{P}_i\}_{i \in I_j}$ intersect at the true parameter points $\boldsymbol{\theta}_j$ in parameter space, the identification problem reduces to finding these points. However, directly searching for them as the solution to a system of equations is known as a hard combinatorial problem, which is intractable in practice for a large N. This is due to the fact that the discrete state λ_i is unknown. If the data were well classified into each submodel, many classical regression methods could be applied to estimate the submodel parameters from the data subsets.

Property 1 is now exploited to transform the data classification problem in the data space to a problem of separating hyperspheres representing submodels in the parameter space. More precisely, each hyperplane $\{\mathbb{P}_i\}_{i \in I_j}$ will be mapped to a point $\boldsymbol{z}_i$ lying on a particular hypersphere for the mode j.

Let $\boldsymbol{z}_i$ be the orthogonal projection of an arbitrary point $\boldsymbol{\theta}$ on $\mathbb{P}_i$,

$$\boldsymbol{z}_i = \boldsymbol{z}(\boldsymbol{\theta}, \mathbb{P}_i) = \boldsymbol{\theta} + \boldsymbol{g}_i, \tag{4}$$

with

$$\boldsymbol{g}_i = \boldsymbol{g}(\boldsymbol{\theta}, \mathbb{P}_i) \triangleq -\frac{\boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i}{\|\boldsymbol{\varphi}_i\|_2^2} \boldsymbol{\varphi}_i. \tag{5}$$

We note the following properties:
- $\forall i \in I_j$, $\boldsymbol{\theta}_j \in \mathbb{P}_i$ and $\boldsymbol{g}_i \perp (\boldsymbol{\theta}_j - \boldsymbol{z}_i)$,
- $\boldsymbol{g}_i = \boldsymbol{0}$ iff $\boldsymbol{\theta} \in \mathbb{P}_i$,
- $\|\boldsymbol{g}_i\|_2 = |d_i|$,

where d_i is the algebraic distance from $\boldsymbol{\theta}$ to $\mathbb{P}_i$,

$$d_i = d(\boldsymbol{\theta}, \mathbb{P}_i) = \frac{\boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i}{\|\boldsymbol{\varphi}_i\|_2}. \tag{6}$$

By fixing an arbitrary point $\boldsymbol{\theta} \neq \boldsymbol{\theta}_j$, each point triplet $\{\boldsymbol{\theta}_j, \boldsymbol{\theta}, \boldsymbol{z}_i\}_{i \in I_j}$ forms a right triangle with the right angle at $\boldsymbol{z}_i$ and the hypotenuse $\boldsymbol{\theta}\boldsymbol{\theta}_j$. We know from basic geometry that

for fixed points $\boldsymbol{\theta}$, $\boldsymbol{\theta}_j$, all the points $\boldsymbol{z}_i$ lie on a hypersphere whose diameter is the line segment $\boldsymbol{\theta}\boldsymbol{\theta}_j$, as illustrated in Fig. 2 (left) for $p = 2$. Keeping $\boldsymbol{\theta}$ fixed and switching $\boldsymbol{\theta}_j$ for a different j, we obtain a different hypersphere. Therefore, each hypersphere represents a submodel and all these hyperspheres intersect at the chosen point $\boldsymbol{\theta}$. Figure 2 (right) shows an example with 2 modes and $p = 2$.
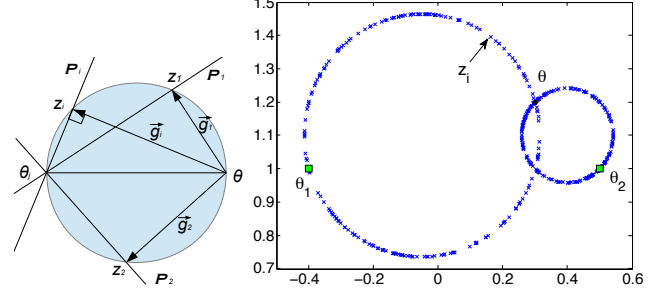

Fig. 2. Vectors $\boldsymbol{g}_i$, points $\boldsymbol{z}_i$ and hyperplanes $\mathbb{P}_i$ of a submodel (left) and an illustration of two modes (right): the point $\boldsymbol{\theta}$ (dot), two submodel parameter points $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ (squares) and points $\boldsymbol{z}_i$ (crosses).

Hereafter, the data classification into each mode is considered as a hypersphere separation in the parameter space. The difficulty of this separation problem is intimately related to the choice of $\boldsymbol{\theta}$, which will be discussed in Sect. III for systems with two modes.

B. Noisy case

We now consider the effect of noise in the parameter space. According to model (1), $(\boldsymbol{\varphi}_i, y_i, e_i)$ defines a hyperplane $\mathbb{P}_{i,e_i}$ as

$$\mathbb{P}_{i,e_i}: \quad \boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i + e_i = 0. \tag{7}$$

Thus, $\mathbb{P}_i$ as defined by (2) in the noiseless case and $\mathbb{P}_{i,e_i}$ are parallel and at a distance $|d_{e_i}| = |\frac{e_i}{\|\boldsymbol{\varphi}_i\|_2}|$. When the point $\boldsymbol{\theta}_j$ is the intersection of all $\{\mathbb{P}_i\}_{i \in I_j}$, the distance from $\boldsymbol{\theta}_j$ to $\mathbb{P}_{i,e_i}$ is also $|d_{e_i}|$. Then, all $\{\mathbb{P}_{i,e_i}\}_{i \in I_j}$ intersect the hypersphere of center $\boldsymbol{\theta}_j$ and radius $r_e = \max_{i \in I_j} |d_{e_i}|$.

Property 2. *In the case where the data are corrupted by a bounded noise with $\|\boldsymbol{e}\|_\infty = \max_{i \in \{1,\dots,N\}} |e_i| \leq \epsilon$, one has*

$$\boldsymbol{\theta}_j \in \bigcap_{i \in I_j} \mathbb{P}_i^\epsilon, \tag{8}$$

where $\mathbb{P}_i^\epsilon$ is a slab with a thickness 2ϵ defined by

$$\mathbb{P}_i^\epsilon: \quad -\epsilon \leq \boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i \leq \epsilon. \tag{9}$$

We note that $|d_{e_i}|$ is large for a small $\|\boldsymbol{\varphi}_i\|_2$. This implies that for linear submodels, the effect on the parameter vector of the noise is more serious with regressors close to the origin, whereas for affine models, this effect is reduced since $\|\tilde{\boldsymbol{\varphi}}_i\|_2 \geq 1$.

The distance from the orthogonal projection $\boldsymbol{z}_i'$ of $\boldsymbol{\theta}$ on $\mathbb{P}_{i,e_i}$ with $i \in I_j$ to the jth mode hypersphere (obtained for the noiseless case by (4) from the projections $\boldsymbol{z}_i$ of $\boldsymbol{\theta}$ on $\mathbb{P}_i$) is $|\Delta_{i,e_i}|$ with

$$\Delta_{i,e_i} = \sqrt{R_j^2 + d_{e_i}^2 + 2R_j d_{e_i} \cos(\overrightarrow{\boldsymbol{O}_j \boldsymbol{z}_i}, \overrightarrow{\boldsymbol{g}_i})} - R_j, \tag{10}$$

where $\boldsymbol{O}_j, R_j$ are the center and the radius of the jth hypersphere. Since $|\Delta_{i,e_i}|$ is the smallest distance from $\boldsymbol{z}'_i$ to the hypersphere, $|\Delta_{i,e_i}| \le |d_{e_i}|$.

We see that when $\boldsymbol{z}_i \to \boldsymbol{\theta}$, $\cos(\overrightarrow{\boldsymbol{O}_j \boldsymbol{z}_i}, \overrightarrow{\boldsymbol{g}_i}) \to 0$ and the points $\boldsymbol{z}'_i$ close to $\boldsymbol{\theta}$ are less affected by the noise, in which case,

$$|\Delta_{i,e_i}| \to \sqrt{R_j^2 + d_{e_i}^2} - R_j = \frac{d_{e_i}^2}{\sqrt{R_j^2 + d_{e_i}^2} + R_j} \le \frac{d_{e_i}^2}{2R_j}.$$

This property implies also a decreased effect of the noise in data classification since the area around $\boldsymbol{\theta}$ is the most difficult one to separate. This property is illustrated in Fig. II-B. With a uniform noise, points $\boldsymbol{z}'_i$ close to $\boldsymbol{\theta}$ almost lie on the hyperspheres.
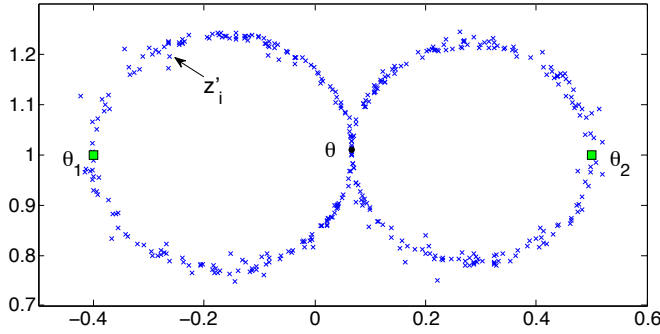

Fig. 3. Illustration of 2 modes with a uniform noise ($p = 2$): the point $\boldsymbol{\theta}$ (dot), two submodel parameter points $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ (squares) and points $\boldsymbol{z}'_i$ (crosses).

III. IDENTIFICATION OF LINEAR HYBRID SYSTEMS WITH TWO MODES

In this section, we concentrate on the identification of two-mode SARX systems which have the two hypersphere structure in parameter space. We propose a procedure including the four following steps.

1) Choose a point $\boldsymbol{\theta}$.
2) Map the data points $(\boldsymbol{\varphi}_i, y_i)$ to points $\boldsymbol{z}_i$ on hyperspheres as in Sect. II-A, i.e.,

$$\boldsymbol{z}_i = \boldsymbol{\theta} - \frac{\boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i}{\|\boldsymbol{\varphi}_i\|_2^2} \boldsymbol{\varphi_i}. \quad (11)$$

3) Separate the two hyperspheres to classify the data into two groups.
4) Estimate the two submodels from the data in each group.

A. Choice of the point $\boldsymbol{\theta}$

We now search for a $\boldsymbol{\theta}$ which will ease the separation of the hyperspheres. If $\boldsymbol{\theta}$ belongs to the segment $\boldsymbol{\theta}_1 \boldsymbol{\theta}_2$, i.e.,

$$\boldsymbol{\theta} = \boldsymbol{\theta}_1 + \mu(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1), \quad \mu \in (0,1), \quad (12)$$

the two hyperspheres are tangent at $\boldsymbol{\theta}$. With this structure, a linear classification is sufficient to separate the two hyperspheres. However, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ are not known yet, and we

must find a $\boldsymbol{\theta}$ satisfying (12) from the data only. The condition (12) is equivalent to

$$f_{\boldsymbol{\theta}}(\boldsymbol{\varphi}) \doteq \boldsymbol{\theta}^\top \boldsymbol{\varphi} = (1-\mu) f_1(\boldsymbol{\varphi}) + \mu f_2(\boldsymbol{\varphi}), \ \forall \boldsymbol{\varphi} \in \mathbb{R}^p. \quad (13)$$

In particular, for $\mu = \frac{1}{2}$, $f_{\boldsymbol{\theta}}(\boldsymbol{\varphi})$ is the average function of the two submodels. This observation leads us to consider the least squares solution as a candidate for $\boldsymbol{\theta}$.

Proposition 1. *Given a data set $\mathcal{S}$ generated by a switched system with two modes such that the matrix $[\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_N]$ is of full rank and the following condition holds,*

$$\sum_{i=1}^N \frac{\boldsymbol{\varphi}_i^T \boldsymbol{\theta}_1 - y_i}{\|\boldsymbol{\varphi}_i\|_2^2} \boldsymbol{\varphi}_i = \gamma \sum_{i=1}^N \frac{\boldsymbol{\varphi}_i^T \boldsymbol{\theta}_2 - y_i}{\|\boldsymbol{\varphi}_i\|_2^2} \boldsymbol{\varphi}_i, \quad (14)$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the subsystem parameter vectors and $\gamma \in \mathbb{R}$, the least squares solution $\boldsymbol{\theta}^* \in \mathbb{R}^p$ minimizing the cost function

$$L(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^N \frac{(\boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i)^2}{\|\boldsymbol{\varphi}_i\|_2^2}, \quad (15)$$

is on the segment $\boldsymbol{\theta}_1 \boldsymbol{\theta}_2$, i.e., $\boldsymbol{\theta}^* = \boldsymbol{\theta}_1 + \mu(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$ with $\mu = \frac{-\gamma}{1-\gamma} \in [0,1]$.

Proof: Since, with $\boldsymbol{g}_i$ given by (5), $\boldsymbol{g}_i(\boldsymbol{\theta}_j) = \boldsymbol{0}$, $\forall i \in I_j, j \in \{1,2\}$, condition (14) is rewritten as

$$\sum_{i \in I_2} \boldsymbol{g}_i(\boldsymbol{\theta}_1) = \gamma \sum_{i \in I_1} \boldsymbol{g}_i(\boldsymbol{\theta}_2). \quad (16)$$

Thus, the vectors $\overline{\boldsymbol{g}}(\boldsymbol{\theta}_1) = \sum_{i \in I_2} \boldsymbol{g}_i(\boldsymbol{\theta}_1)$ and $\overline{\boldsymbol{g}}(\boldsymbol{\theta}_2) = \sum_{i \in I_1} \boldsymbol{g}_i(\boldsymbol{\theta}_2)$ are collinear. Moreover, since, $\forall i \in I_2$, $\boldsymbol{g}_i(\boldsymbol{\theta}_1)$ points towards $\boldsymbol{\theta}_2$, and similarly, $\forall i \in I_1$, $\boldsymbol{g}_i(\boldsymbol{\theta}_2)$ points towards $\boldsymbol{\theta}_1$, $\overline{\boldsymbol{g}}(\boldsymbol{\theta}_1)$ and $\overline{\boldsymbol{g}}(\boldsymbol{\theta}_2)$ are in opposite directions, which implies $\gamma \le 0$.

For any $\boldsymbol{\theta}$ in the segment $\boldsymbol{\theta}_1 \boldsymbol{\theta}_2$ such that $\boldsymbol{\theta} - \boldsymbol{\theta}_1 = \mu(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$, we have

$$\boldsymbol{g}_i(\boldsymbol{\theta}) = \mu \boldsymbol{g}_i(\boldsymbol{\theta}_2), \ \forall i \in I_1,$$

$$\boldsymbol{g}_i(\boldsymbol{\theta}) = (1-\mu) \boldsymbol{g}_i(\boldsymbol{\theta}_1), \ \forall i \in I_2.$$

Thus, (16) is equivalent to

$$\sum_{i \in I_2} \boldsymbol{g}_i(\boldsymbol{\theta}) = \frac{(1-\mu)}{\mu} \gamma \sum_{i \in I_1} \boldsymbol{g}_i(\boldsymbol{\theta}).$$

Then, if we take $\boldsymbol{\theta}$ with $0 \le \mu = \frac{-\gamma}{1-\gamma} \le 1$, we get

$$\sum_{i=1}^N \boldsymbol{g}_i(\boldsymbol{\theta}) = \sum_{i \in I_1} \boldsymbol{g}_i(\boldsymbol{\theta}) + \sum_{i \in I_2} \boldsymbol{g}_i(\boldsymbol{\theta})$$

$$= \left(1 + \frac{1-\mu}{\mu} \gamma\right) \sum_{i \in I_1} \boldsymbol{g}_i(\boldsymbol{\theta}) = \boldsymbol{0}.$$

On the other hand, the gradient of $L(\boldsymbol{\theta})$ is zero at $\boldsymbol{\theta}^*$, i.e., $\nabla L(\boldsymbol{\theta})|_{\boldsymbol{\theta}^*} = \sum_{i=1}^N \boldsymbol{g}_i(\boldsymbol{\theta}^*) = \boldsymbol{0}$. Since $\boldsymbol{\theta}^*$ is unique if $\text{rank}[\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_N] = p$, we conclude that $\boldsymbol{\theta} \equiv \boldsymbol{\theta}^*$. $\square$

Proposition 2. *Given a data set $\mathcal{S}$ generated by a switched system with two modes such that for each value $\boldsymbol{\varphi}$ in the data set, the data set includes both of the outputs $f_1(\boldsymbol{\varphi})$ and*

$f_2(\varphi)$ and the matrix $[\varphi_1, \ldots, \varphi_N]$ is of full rank, the least squares solution $\boldsymbol{\theta}^* \in \mathbb{R}^p$ minimizing (15) is the midpoint of the segment $\boldsymbol{\theta}_1 \boldsymbol{\theta}_2$.

Proof: With such a data set, it is easy to see that $\gamma = -1$ in (14). Then, according to the proof of Proposition 1, $\mu = 0.5$ and the statement follows. □

For data sets not satisfying (14), the point $\boldsymbol{\theta}^*$ may be out of alignment with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Then the two hyperspheres are no longer tangent and the separation task is more difficult. In particular, points in the intersection of the two corresponding hyper-balls will be badly classified by a linear classifier. However, a good submodel can be estimated from such a data set with few outliers by a robust regression method. As shown in [4], the ℓ_1-norm optimization based method can handle a data set including about 50% of outliers, which is enough for systems with two modes.

B. Data classification and submodel estimation

We propose a linear classification method based on the Principal Component Analysis (PCA) for the matrix $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N, \boldsymbol{\theta}]$. As a pre-processing step in PCA, the matrix $\boldsymbol{Z}$ is centered to the matrix $\tilde{\boldsymbol{Z}} = [\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_N, \tilde{\boldsymbol{z}}_{\boldsymbol{\theta}}]$ such that $\sum_{i=1}^N \tilde{\boldsymbol{z}}_i + \tilde{\boldsymbol{z}}_{\boldsymbol{\theta}} = \boldsymbol{0}$. Note that with the choice $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ as in Sect. III-A, the centered matrix is simply $\tilde{\boldsymbol{Z}} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N, \boldsymbol{0}]$. The eigenvector corresponding to the largest eigenvalue of $\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}^\top$ represents a direction on which the data of $\tilde{\boldsymbol{Z}}$ has the largest variance. This eigenvector has the same direction as the vector $\overrightarrow{\boldsymbol{\theta}_1 \boldsymbol{\theta}_2}$ in the case of two tangent hyperspheres. Then, a projection of vectors $\tilde{\boldsymbol{z}}_i$ on this vector can separate data into two groups corresponding to the submodels.

Let $\lambda_1, \boldsymbol{v}_1$ be the largest eigenvalue and the corresponding eigenvector of $\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}^\top$. Then, two data groups are found by

$$\hat{I}_1 = \left\{ i \in \{1, \ldots, N\} : \tilde{\boldsymbol{z}}_i^\top \boldsymbol{v}_1 \geq 0 \right\}, \qquad (17)$$
$$\hat{I}_2 = \left\{ i \in \{1, \ldots, N\} : \tilde{\boldsymbol{z}}_i^\top \boldsymbol{v}_1 < 0 \right\}.$$

Now, we can use any robust regression method to estimate the submodel parameter vectors. A simple robust convex method minimizes the ℓ_1-norm of the error as

$$\hat{\boldsymbol{\theta}}_j = \arg\min_{\boldsymbol{\theta}} \sum_{i \in \hat{I}_j} \|\boldsymbol{\varphi}_i^\top \boldsymbol{\theta} - y_i\|_1, \quad j = 1, 2. \qquad (18)$$

Finally, one can re-estimate the mode $\hat{\lambda}_i$ with

$$\hat{\lambda}_i = \arg\min_{j \in \{1,2\}} |y_i - \boldsymbol{\theta}_j^\top \boldsymbol{\varphi}_i|, \ i = 1, \ldots, N. \qquad (19)$$

For a refinement, the submodels can be re-estimated by (18) with the new classification (19).

IV. SYSTEMS WITH MORE THAN TWO MODES

Inspired by the methods in [4], [5], which estimate the submodels one by one, we propose a simple procedure to identify hybrid systems with more than two modes. For these methods, after estimating a parameter vector $\hat{\boldsymbol{\theta}}$, the data points verifying the error condition, $|y_i - \boldsymbol{\varphi}_i^\top \hat{\boldsymbol{\theta}}| \leq \delta$ where δ is a

fixed threshold, are associated to the estimated submodel and removed from the data set to estimate the next parameter vector. The number of modes is finally estimated as the number of submodels required to satisfy the error condition for all data points.

Algorithm 1

Require: a data set $\mathcal{S} = \{(\boldsymbol{\varphi}_i, y_i)\}_{i=1}^N$ and a threshold δ.
Initialize $\mathcal{S}_0 = \mathcal{S}$, $j = 0$.
while $|\mathcal{S}_0| \neq 0$ **do**
  Set $j = j + 1$.
  Estimate $\boldsymbol{\theta}^*$ by least squares with the data in $\mathcal{S}_0$.
  Map the data $(\boldsymbol{\varphi}_i, y_i)$ to points $\boldsymbol{z}_i$ by (11) with $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.
  Classify the data set $\mathcal{S}_0$ into two groups via the points $\boldsymbol{z}_i$ and (17).
  Estimate two parameter vectors $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ corresponding to the groups as in (18).
  Identify the jth submodel parameter vector by

$$\hat{\boldsymbol{\theta}}_j = \arg\max_{\boldsymbol{\theta} \in \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}} |\mathcal{S}(\boldsymbol{\theta})|,$$

  where $\mathcal{S}(\boldsymbol{\theta}) = \left\{ (\boldsymbol{\varphi}_i, y_i) \in \mathcal{S}_0 : |y_i - \boldsymbol{\varphi}_i^\top \boldsymbol{\theta}| \leq \delta \right\}$.
  Set $\mathcal{S}_0 = \mathcal{S}_0 \backslash \mathcal{S}(\hat{\boldsymbol{\theta}}_j)$.
end while
return the estimated number of modes $\hat{n} = j$ and the estimated parameter matrix $\hat{\Theta} = [\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{\hat{n}}]$.

Algorithm 1 extends the proposed method of Sect. III in a similar manner. This procedure includes only one hyperparameter to tune, the threshold δ. As for the methods [4], [5], when the threshold δ is appropriately chosen, the number of modes n is recovered. Nevertheless, experiments in the next section will show that the proposed method is less sensitive to choice of δ. For a bounded noise, δ can be chosen such that $\delta \geq \|e\|_\infty$. If the noise is a white Gaussian noise with a standard deviation σ_e, the threshold δ may be chosen in $[\sigma_e, 3\sigma_e]$. In this case, to avoid the effect of outliers, the algorithm should be stopped before $|\mathcal{S}_0| = 0$.

V. NUMERICAL EXPERIMENTS

In this section, we show the efficiency of the proposed method via some numerical examples. To evaluate the quality of the results, we compute the Normalized Parametric Error (NPE), $NPE = \frac{\|\hat{\Theta} - \Theta\|_F}{\|\Theta\|_F}$, where $\|.\|_F$ is the Frobenius norm and $\Theta, \hat{\Theta}$ are the true and estimated parameter matrices in which each column is a subsystem parameter vector. The proposed method (GEO) is compared with the algebraic approach [2] (ALG), the continuous optimization based approach [11] (PE-MCS) and the sparse optimization based approaches [4] (SO1) and [5] (SO2). Over 100 trials with different input, switching and noise sequences, we report the NPE mean and standard deviation. The sparse optimization based methods [4], [5] estimate the parameter vectors one by one until the data set is empty. Thus the number of modes cannot be fixed. In this case, we retain the best estimated parameter vectors as the ones closest to the true ones to

compute the NPE. In all experiments, the data are generated with a uniform input distribution $\mathcal{U}(-4, 4)$ and an additive Gaussian noise e_i of standard deviation σ_e, except for Sect. V-B.2 which uses a normally distributed input of unit variance. In order to observe the sensitivity to noise of the methods, σ_e is varied from 0 to 0.9. The hyperparameters in the method PE-MCS [11] are left to their default values. The threshold δ used to assign data points to a submodel in methods SO1 [4] and SO2 [5] is chosen in accordance with σ_e. The other hyperparameters are tuned to get the best results. The stopping criterion in Algorithm 1 is $|\mathcal{S}_0| \leq 0.01N$.

For the examples with two modes, we are only interested in the parametric error (NPE), whereas for the examples with three modes, the estimated number of modes is also important and its sensitivity to δ will be studied.

A. Static examples

1) Toy examples: We take two one-dimensional switched static systems, intersecting or parallel, as toy examples,

$$y_i = \begin{cases} u_{i-1} + 2 + e_i & \text{if } \lambda_i = 1, \\ -u_{i-1} + e_i & \text{if } \lambda_i = 2, \end{cases} \quad (20)$$

$$y_i = \begin{cases} u_{i-1} + 2 + e_i & \text{if } \lambda_i = 1, \\ u_{i-1} + e_i & \text{if } \lambda_i = 2. \end{cases} \quad (21)$$

For (20) or (21), training sets of $N = 100$ points are generated with a uniformly distributed random sequence of $\lambda_i \in \{1, 2\}$. Figure 4 shows the submodel estimates with the noisy data for each of these systems for $\sigma_e = 0.5$. Figures 5 and 6 report the comparisons with the other methods for a varying noise standard deviation. They show that the proposed method can handle high noise regimes more efficiently. The other methods do not work well in these regimes even though these examples are very simple.
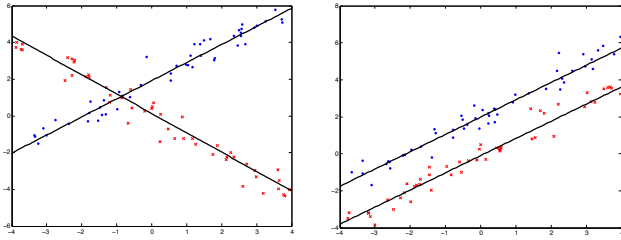
Fig. 4. System (20) (left), system (21) (right) : noisy data ($\bullet$, $\times$) and estimated submodels ($-$).

2) Three mode example: We consider the following switched hybrid system with 3 modes

$$y_i = \begin{cases} u_{i-1} + 0.5u_{i-2} + 2 + e_i & \text{if } \lambda_i = 1, \\ -0.5u_{i-1} - 1.3u_{i-2} + e_i & \text{if } \lambda_i = 2, \\ -u_{i-1} + u_{i-2} - 1 + e_i & \text{if } \lambda_i = 3. \end{cases} \quad (22)$$

Training sets of $N = 900$ points are generated by (22) with a fixed $\sigma_e = 0.3$ and with two scenarios for the random switching sequence $\{\lambda_i\}_{i=1}^{900}$. In scenario 1, $|\{i : \lambda_i = 1\}| = 100$, $|\{i : \lambda_i = 2\}| = 300$, $|\{i : \lambda_i = 3\}| = 500$, while in
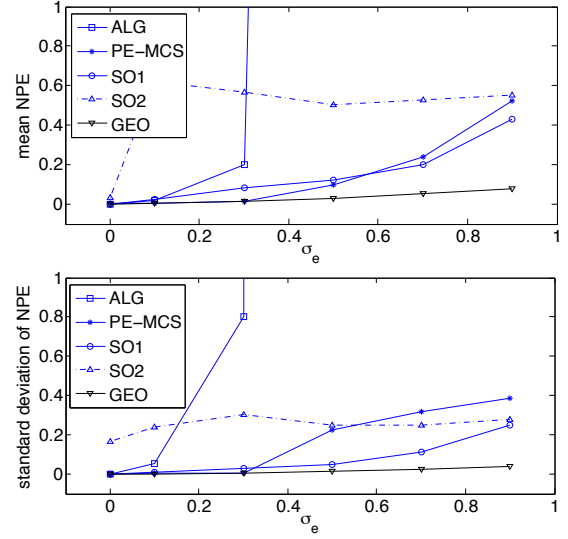
Fig. 5. Static system (20): NPE mean (top) and standard deviation (bottom) versus Gaussian noise standard deviation.
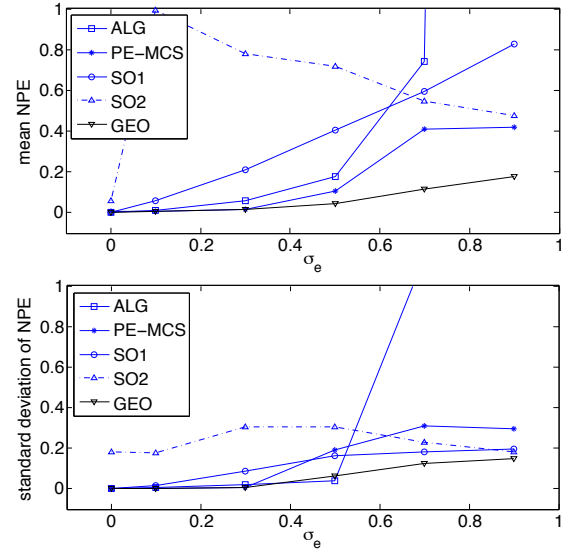
Fig. 6. Static system (21): NPE mean (top) and standard deviation (bottom) versus Gaussian noise standard deviation.

scenario 2, λ_i is uniformly distributed in $\{1, 2, 3\}$. Figure 7 (left) presents the first scenario which satisfies the sufficient condition on the data for SO1 [4]. The number of modes is well estimated with $\delta \geq 3\sigma_e$ while Fig. 7 (right) shows the second scenario. The proposed method still correctly estimates the number of modes, while the SO1 method breaks down in these unfavorable conditions.

B. Dynamical Examples

1) Two mode example: Consider the linear hybrid system used in [5]:

$$y_i = \begin{cases} 0.2y_{i-1} + 0.24y_{i-2} + 2u_{i-1} + e_i & \text{if } \lambda_i = 1, \\ -1.4y_{i-1} - 0.53y_{i-2} + u_{i-1} + e_i & \text{if } \lambda_i = 2. \end{cases} \quad (23)$$

Training sets of $N = 400$ points are generated by (23) with $\lambda_i = 1$ for $i \in [1, 100] \cup [201, 300]$, and $\lambda_i = 2$ for $i \in$
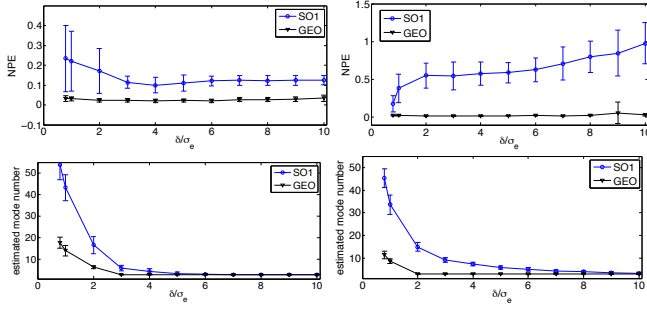
Fig. 7. System (22) with scenario 1 (left) and scenario 2 (right): NPE (top) and estimated number of modes (bottom) versus the threshold δ.

Fig. 9. Switched linear system in [4] (left) and switched affine system (right): NPE mean (top) and estimated number of modes (bottom) versus the threshold δ.

$[101, 200] \cup [301, 400]$. Figure 8 shows that the proposed method gives the model with the smallest error and which is the less affected by noise.
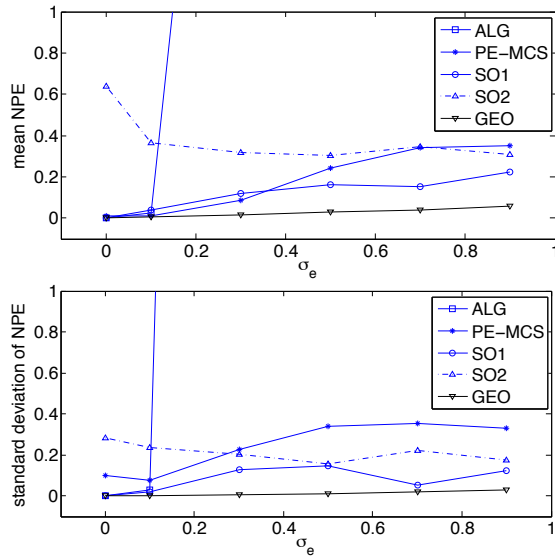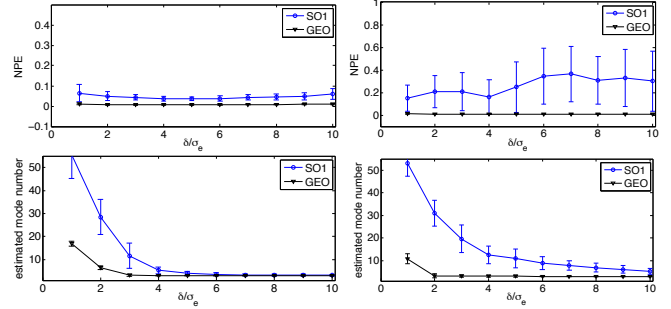

Fig. 8. System (23): NPE mean (top) and standard deviation (bottom) versus Gaussian noise standard deviation.

2) Three mode examples: Consider the switched linear system with the three following subsystem parameters:

$$\boldsymbol{\theta}_1 = [-0.40 \quad 0.25 \ -0.15 \quad 0.08]^\top,$$
$$\boldsymbol{\theta}_2 = [\ 1.55 \ -0.58 \ -2.10 \quad 0.96]^\top,$$
$$\boldsymbol{\theta}_3 = [\ 1.00 \ -0.24 \ -0.65 \quad 0.30]^\top,$$

used in the experiments of [4] and an affine one obtained by adding constants $2, -1, 1$ respectively to its three subsystems.

Training sets of $N = 600$ points are generated by these systems with a uniformly distributed random sequence of $\lambda_i \in \{1, 2, 3\}$ and $\sigma_e = 0.1$. Figure 9 shows that the proposed method (GEO) can estimate switched affine systems as accurately as switched linear ones, whereas the SO1 method [4] has difficulties to deal with such systems.

VI. CONCLUSIONS

This paper presented a geometric approach for hybrid system identification. We first focused on systems switching between two modes and then extended the proposed method to deal with an arbitrary number of modes. The proposed procedures proved their simplicity, accuracy and noise robustness in numerical experiments. For hybrid systems with two modes, there is no hyperparameter to tune and no (non-convex) optimization problem to solve, contrary to most of the other methods. For hybrid systems with multiple modes, there is only one classical hyperparameter, δ, to which the proposed method is less sensitive than others.

In future work, although a simple iterative algorithm was proposed to deal with more than two modes, the multiple hypersphere separation problem should be investigated.

REFERENCES

[1] A. Garulli, S. Paoletti, and V. A., "A survey on switched and piecewise affine system identification," in *Proc. of the 16th IFAC Symposium on System Identification, Brussels, Belgium*, 2012, pp. 344–355.

[2] R. Vidal, S. Soatto, Y. Ma, and S. Sastry, "An algebraic geometric approach to the identification of a class of linear hybrid systems," in *Proc. of the 42nd IEEE Conf. on Decision and Control*, 2004, pp. 167–172.

[3] H. Ohlsson and L. Ljung, "Piecewise affine system identification using sum-of-norms regularization," in *Proc. of the 18th IFAC World Congress, Milano, Italy*, 2011.

[4] L. Bako, "Identification of switched linear systems via sparse optimization," *Automatica*, vol. 47, no. 4, pp. 668–677, 2011.

[5] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps, "A sparsification approach to set membership identification of switched affine systems," *IEEE Trans. on Automatic Control*, vol. 57, no. 3, pp. 634–648, 2012.

[6] F. Lauer, V. L. Le, and G. Bloch, "Learning smooth models of nonsmooth functions via convex optimization," in *Proc. of the IEEE Int. Workshop on Machine Learning for Signal Processing, Santander, Spain*, 2012.

[7] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, vol. 39, no. 2, pp. 205–217, 2003.

[8] J. Roll, A. Bemporad, and L. Ljung, "Identification of piecewise affine systems via mixed-integer programming," *Automatica*, vol. 40, no. 1, pp. 37–50, 2004.

[9] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A bounded-error approach to piecewise affine system identification," *IEEE Trans. on Automatic Control*, vol. 50, no. 10, pp. 1567–1580, 2005.

[10] A. Juloski, S. Weiland, and W. Heemels, "A Bayesian approach to identification of hybrid systems," *IEEE Trans. on Automatic Control*, vol. 50, no. 10, pp. 1520–1533, 2005.

[11] F. Lauer, G. Bloch, and R. Vidal, "A continuous optimization framework for hybrid system identification," *Automatica*, vol. 47, no. 3, pp. 608–613, 2011.

[12] F. Lauer, "Estimating the probability of success of a simple algorithm for switched linear regression," *Nonlinear Analysis: Hybrid Systems*, vol. 8, pp. 31–47, 2013.