When is a pole spurious?

Ivan Markovsky, Jeroen Boets, Bart Vanluyten, Katrien De Cock, and Bart De Moor
K.U.Leuven, ESAT/SISTA, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
e-mail: **ivan.markovsky@esat.kuleuven.be**

Abstract

The stabilization diagrams, used in modal analysis, rely on the intuitive notion of a spurious pole. In this paper, we give a definition of a spurious pole, based on the most powerful unfalsified model (MPUM) of the data, i.e., on an exact model for the data. The poles of the MPUM are by definition physical and a pole that is not physical is by definition spurious. Our definition does not make assumptions about the data, apart from the basic postulation of the linear time-invariant model class. In this sense it is unprejudiced. Since the MPUM can be constructed from the data, one can compute the physical poles and thus answer the question in the title. If, however, one knows a priori that the data correspond to a noise corrupted trajectory of a true data generating system or that there is an unobserved process noise acting on the true system, one should use this knowledge. In this case, the MPUM concept has to be modified to allow for approximation. Methods for approximate system identification and model reduction are reviewed and applied for spurious pole detection of simulated data.

1 Introduction

In modal analysis, it is a common practice to draw what is called a stabilization diagram and identify a model for the observed data by visually selecting certain poles from the diagram. Empirical observations suggest that the stabilization diagram gives a good indication for deciding which of the model poles correspond to poles of the true system. These poles, called physical poles, tend to stabilize on the diagram. The extra poles, resulting from the fact that the model order is selected higher than the true system order are called spurious poles and do not stabilize on the diagram. Techniques for automatic detection of the physical and spurious poles are of high interest in modal analysis.

The separation of the identified poles into physical and spurious ones, however, is rather speculative. First, the observed data might not be generated by a linear time-invariant (LTI) system, so that one might not be able to properly talk about physical poles and as a consequence about spurious poles. Second, if the data happen to be (or are very close of being) generated by an LTI system of a low (compared to the data length) order, it is the total behavior of the model (with contribution of all poles) that approximates the given data, so it is not justified to keep some of the poles and discard the others. For poorly identified physical poles, the spurious poles might carry equally important information about the true system. This is manifested in model reduction techniques, such as the balanced model reduction method [6] that use the full order model to arrive at a good reduced order model and do not simply select some of the poles.

Despite the loose connection between physical poles and poles of a true data generating system, the stabilization diagrams are a successful and very much used technique. One possible explanation for this is that they are good heuristics for model order selection. The problem of deciding when a pole is spurious is related to *order selection* and *model reduction*. All subspace identification methods, see, e.g., [7], implicitly do order selection and subsequently model reduction by computing the singular value decomposition of certain matrices derived from data and counting the number of singular values that are larger than a given tolerance.

Another method that is related to the detection of spurious poles appeared in the signal processing literature. In [2], Kumaresan and Tufts argue that identifying a large order model from the extended Yule-Walker

equations and reducing that model gives better result than identifying directly a low order model from the extended Yule-Walker equations. Choosing a high model order includes spurious poles in the model, which need to be eliminated on the reduction step. The method of Kumaresan and Tufts is based on the fact that when the least norm solution of the extended Yule-Walker equations is chosen, the spurious poles are outside the unit circle while the physical poles (for a stable true system) tend to be inside the unit circle. Thus the separation problem reduces to the one of picking the stable poles of the identified high order model.

Another commonly used method for order selection is to plot the misfit versus model complexity trade-off curve and (visually) select the order corresponding to the corner of the curve. The order is a measure of the model complexity and the misfit refers to the lack of fit (in some sense) between the data and the model. When used in combination with a maximum likelihood identification method, the trade-off curve has a natural candidate for the misfit measure: the objective function being minimized by the maximum likelihood method.

In this paper, we put forward a definition for the notions of physical and spurious poles and apply

- the method of Kung [3], which is based on balanced model reduction,
- the stochastic subspace identification method of [7, Chapter 3],
- the method of Kumaresan and Tufts [2], and
- the maximum likelihood method combined with order selection from the misfit–complexity trade-off curve,

as alternatives to the stabilization diagram method for detection of spurious poles.

2 Model class: autonomous LTI systems

Let $\mathscr{L}$ be the discrete-time LTI model class and let σ be the shift operator $(\sigma w)(t) := w(t+1)$. If we need to specify the number of variables $\mathtt{w} := \dim(w)$ of a trajectory w of the model $\mathscr{B}$, we write $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$. Two common representations of a model $\mathscr{B} \in \mathscr{L}$ are the constant coefficients difference equation

$$R_0\sigma^0 w + R_1\sigma^1 w + \cdots + R_{\mathtt{l}}\sigma^{\mathtt{l}} w = 0 \tag{1}$$

and an input/state/output representation

$$w = \Pi \operatorname{col}(u,y), \quad \sigma x = Ax + Bu, \quad y = Cx + Du, \tag{2}$$

where Π is a permutation matrix. The polynomial matrix

$$R(\xi) := R_0\xi^0 + R_1\xi^1 + \cdots + R_{\mathtt{l}}\xi^{\mathtt{l}} \in \mathbb{R}^{g\times \mathtt{w}}[\xi]$$

and the tuple of matrices (A,B,C,D,Π) are parameters of the model. The notations $\mathscr{B}(R)$ and $\mathscr{B}_{\text{i/s/o}}(A,B,C,D)$ stand for the systems defined by (1) and (2), respectively, i.e.,

$$\mathscr{B}(R) = \{\, w \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \mid (1) \text{ holds} \,\} \quad \text{and} \quad \mathscr{B}_{\text{i/s/o}}(A,B,C,D,\Pi) = \{\, w \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \mid (2) \text{ holds} \,\}.$$

The integer $\mathtt{l}$ is called the *lag* of the difference equation (1) and the integer $\mathtt{n}$ is called the *order* of the state space representation (2). For a given $\mathscr{B} \in \mathscr{L}$, the smallest lag $\mathbf{l}(\mathscr{B})$ and order $\mathbf{n}(\mathscr{B})$ of a difference equation and state space representation are by definition independent of the representations and are called the lag and the order of the system $\mathscr{B}$. The difference equation representation $\mathscr{B}(R) = \mathscr{B}$ is called minimal if the degree of the representation is equal to the lag of the system $\mathscr{B}$, i.e., $\operatorname{degree}(R) = \mathbf{l}(\mathscr{B})$. The state space representation $\mathscr{B}_{\text{i/s/o}}(A,B,C,D,\Pi) = \mathscr{B}$ is called minimal if $\dim(A) = \mathbf{n}(\mathscr{B})$.

The representations $\mathscr{B}(R) = \mathscr{B}$ and $\mathscr{B}_{\text{i/s/o}}(A,B,C,D,\Pi) = \mathscr{B}$ of a system $\mathscr{B} \in \mathscr{L}$ are not unique due to

- equivalence transformations:

$$\mathscr{B}(R) = \mathscr{B}(UR), \text{ for any unimodular matrix } U$$

and

$$\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D,\Pi) = \mathscr{B}_{\mathrm{i/s/o}}(VAV^{-1},VB,CV^{-1},D,\Pi), \text{ for any nonsingular matrix } V.$$

- non-minimality of the representations.

In addition, the state space representation $\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D,\Pi)$ is not unique due to nonuniqueness in the choice of the input/output partition, i.e., the choice of the permutation matrix Π. In this paper, however, we consider the subclass $\mathscr{L}_0^{\mathtt{p}}$ of the LTI model class $\mathscr{L}^{\mathtt{p}}$ consisting of autonomous systems (the subscript 0 stands for "no inputs"), so that the input/output partitioning issue does not occur.

The difference equation and state space representations for an autonomous system $\mathscr{B} \in \mathscr{L}_0$ specialize to, respectively,

$$P_0\sigma^0 y + P_1\sigma^1 y + \cdots + P_{\mathtt{l}}\sigma^{\mathtt{l}} y = 0, \quad \text{where} \quad P(\xi) := \sum_{i=0}^{\mathtt{l}} R_i \xi^i \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}}[\xi],\ \det\big(P(\xi)\big) \neq 0 \tag{3}$$

and

$$\sigma x = Ax, \quad y = Cx. \tag{4}$$

$\mathscr{B}(P)$ denotes the system defined by (3) and $\mathscr{B}(A,C)$ denotes the system defined by (4).

The zeros of the characteristic equation $\det(P(\xi)) = 0$ of (3) determine the asymptotic behavior (growth, oscillation, or decay) of the model $\mathscr{B}(P)$ and are called *poles* of the system. Note that although the parameter P is not unique, the poles of $\mathscr{B}(P)$ are invariant of the representation. We denote by $\lambda(\mathscr{B})$ the set of the poles of $\mathscr{B}$, i.e.,

$$\lambda(\mathscr{B}) := \{\, z \in \mathbb{C} \mid \det\big(P(z)\big) = 0 \,\}, \quad \text{where} \quad P \text{ is any polynomial matrix, such that } \mathscr{B}(P) = \mathscr{B}.$$

If a state space representation $\mathscr{B}(A,C)$ is minimal, then the set of eigenvalues of A coincides with the set of poles of $\mathscr{B}(A,C)$. If, however, the order of the representation $\mathscr{B} = \mathscr{B}(A,C)$ is higher than the order of the system $\mathscr{B}$, then there are eigenvalues of A that do not correspond to poles. Such eigenvalues do not have effect on $\mathscr{B}$ and can be arbitrary complex numbers.

The complexity of a general LTI system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$, can be measured by the number of inputs $\mathbf{m}(\mathscr{B})$, the lag $\mathbf{l}(\mathscr{B})$, and the order $\mathbf{n}(\mathscr{B})$. The model class of LTI systems (with $\mathtt{w}$ variables) of *bounded complexity* — $\mathbf{m}(\mathscr{B}) \leq \mathtt{m}$, $\mathbf{l}(\mathscr{B}) \leq \mathtt{l}$, $\mathbf{n}(\mathscr{B}) \leq \mathtt{n}$, where $\mathtt{m}$, $\mathtt{l}$, and $\mathtt{n}$ are given integers — is denoted by $\mathscr{L}^{\mathtt{w},\mathtt{n}}_{\mathtt{m},\mathtt{l}}$. For an autonomous system $\mathscr{B} \in \mathscr{L}_0^{\mathtt{p}}$, $\mathbf{m}(\mathscr{B}) = 0$, so that the complexity specification is given only by the lag and the order. Note that, $\mathbf{l}(\mathscr{B}) \leq \mathbf{n}(\mathscr{B})$ for any $\mathscr{B} \in \mathscr{L}$, and $\mathbf{l}(\mathscr{B}) = \mathbf{n}(\mathscr{B})$, for any single-output system.

A trajectory $w \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}}$ of the system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}} \subseteq (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}}$ is an infinite time series $\big(w(1),\ldots,w(t),\ldots\big)$. The restriction of w to the interval $[t_1,t_2]$ is denoted by $w|_{[t_1,t_2]} := \big(w(t_1),\ldots,w(t_2)\big)$ and the restriction of the system behavior $\mathscr{B}$ to the interval $[t_1,t_2]$ is denoted by $\mathscr{B}|_{[t_1,t_2]} := \{\, w|_{[t_1,t_2]} \mid w \in \mathscr{B} \,\}$.

3 Spurious poles: definition using exact modeling

By definition, a pole is physical if and only if it is not spurious. Thus, instead of talking about spurious poles, we can equivalently talk about physical poles. This turns out to be more convenient.

Intuitively, a pole is physical with respect to data y_{d} if it is a pole of a *true data generating system* $\bar{\mathscr{B}} \in \mathscr{L}_0$ for y_{d}. The difficulty in making this intuitive idea precise is in specifying the meaning of the statement

"$\bar{\mathscr{B}} \in \mathscr{L}_0$ is a data generating system for y_d". The classical approaches to define this statement are to introduce process or measurement noise in the description of the model class and adopt a stochastic setting. Using measurement noise, the true data generating system is the output error model

$$y_\mathrm{d} = \bar{y} + \tilde{y}, \quad \text{where} \quad \bar{y} \in \bar{\mathscr{B}} \in \mathscr{L}_0 \quad \text{and} \quad \tilde{y} \sim \mathrm{N}(0, \sigma^2 I), \tag{5}$$

i.e., the data are assumed to be a noise corrupted trajectory of $\bar{\mathscr{B}}$, where the noise is zero mean, Gaussian, with covariance matrix known up to a scaling factor. Using the process noise, the true data generating system is an ARMA model. In this case, $\bar{\mathscr{B}}$ is augmented with an input e that is a white Gaussian process, and it is assumed that $\mathrm{col}(y_\mathrm{d}, e_\mathrm{d}) \in \bar{\mathscr{B}}$ for some noise realization e_d.

There are two basic problems with the classical approach.

1. One has to *assume* a priori that a true data generating system $\bar{\mathscr{B}}$ (i.e., an output error or ARMA model) exists. In practice, such assumptions need not hold, so a notion of an approximated data generating system is needed.
2. Even when $\bar{\mathscr{B}}$ exists, typically it is not computable from the data. The theory ensures that the result of a "good" (i.e., consistent) identification method converges to $\bar{\mathscr{B}}$ asymptotically as the number of data points increases. The model $\bar{\mathscr{B}}$, however, is attained exactly only when an infinite amount of data are available.

We use a deterministic approach. The system $\bar{\mathscr{B}} \in \mathscr{L}_0$ is defined as the most powerful unfalsified model (MPUM) for y_d [8], which makes no assumptions about the data and allows to construct a model $\bar{\mathscr{B}}$ for any finite time series $y_\mathrm{d} \in (\mathbb{R}^\mathsf{p})^T$.

Definition 1 (Most powerful unfalsified model)**.** Consider a time series $y_\mathrm{d} \in (\mathbb{R}^\mathsf{p})^T$. The system $\mathscr{B}_\mathrm{mpum}(y_\mathrm{d})$ is called the MPUM of y_d in the model class $\mathscr{L}_0^\mathsf{p}$ if

1. $\mathscr{B}_\mathrm{mpum}(y_\mathrm{d})$ is unfalsified by y_d, i.e., $y_\mathrm{d} \in \mathscr{B}_\mathrm{mpum}(y_\mathrm{d})$,
2. $\mathscr{B}_\mathrm{mpum}(y_\mathrm{d})$ is in the model class, i.e., $\mathscr{B}_\mathrm{mpum}(y_\mathrm{d}) \in \mathscr{L}_0^\mathsf{p}$, and
3. any other unfalsified model in the model class is less powerful, i.e.,

$$y_\mathrm{d} \in \mathscr{B} \in \mathscr{L}_0^\mathsf{p} \implies \mathscr{B}_\mathrm{mpum}(y_\mathrm{d}) \subseteq \mathscr{B}.$$

It can be shown that the MPUM exists and is unique. Moreover, algorithms for computing it have been developed; see [8, 7, 5, 4].

With a "true data generating system $\bar{\mathscr{B}}$ for y_d" defined to be the MPUM $\mathscr{B}_\mathrm{mpum}(y_\mathrm{d})$ of y_d, the intuitive notion of a spurious pole leads to the following definition.

Definition 2 (Physical poles)**.** Consider a time series $y_\mathrm{d} \in (\mathbb{R}^\mathsf{p})^T$. The poles $\lambda\big(\mathscr{B}_\mathrm{mpum}(y_\mathrm{d})\big)$ of the MPUM of y_d are called physical with respect to the data y_d. Any $z \in \mathbb{C}$, such that $z \notin \lambda\big(\mathscr{B}_\mathrm{mpum}(y_\mathrm{d})\big)$, is called a spurious pole.

Clearly, by computing the MPUM, we can answer the basic question "When is a pole spurious?" Next we discuss the question:

> Assuming that there is a true data generating system $\bar{\mathscr{B}}$ for y_d, i.e., $y_\mathrm{d} \in \bar{\mathscr{B}}$, under want conditions the MPUM $\mathscr{B}_\mathrm{mpum}(y_\mathrm{d})$ of y_d coincides with $\bar{\mathscr{B}}$?

The conditions turn out to be restrictive, so the notion of an MPUM is modified in Section 5 to allow for an approximation.

4 Recovering the data generating system from exact data

In the general case when the true data generating system $\bar{\mathscr{B}}$ is in $\mathscr{L}^{\mathsf{w}}$ (rather than in $\mathscr{L}_0^{\mathsf{p}}$ as considered in the previous section), a key condition for $\bar{\mathscr{B}} = \mathscr{B}_{\mathrm{mpum}}(w_{\mathrm{d}})$ turns out to be the persistency of excitation of an input component of the data w_{d}. The time series $u_{\mathrm{d}} = \big(u_{\mathrm{d}}(1), \ldots, u_{\mathrm{d}}(T)\big)$ is persistently exciting of order L if the Hankel matrix

$$\mathscr{H}_L(u_{\mathrm{d}}) := \begin{bmatrix} u_{\mathrm{d}}(1) & u_{\mathrm{d}}(2) & u_{\mathrm{d}}(3) & \cdots & u_{\mathrm{d}}(T-L+1) \\ u_{\mathrm{d}}(2) & u_{\mathrm{d}}(3) & u_{\mathrm{d}}(4) & \cdots & u_{\mathrm{d}}(T-L+2) \\ u_{\mathrm{d}}(3) & u_{\mathrm{d}}(4) & u_{\mathrm{d}}(5) & \cdots & u_{\mathrm{d}}(T-L+3) \\ \vdots & \vdots & \vdots & & \vdots \\ u_{\mathrm{d}}(L) & u_{\mathrm{d}}(L+1) & u_{\mathrm{d}}(L+2) & \cdots & u_{\mathrm{d}}(T) \end{bmatrix}.$$

is of full row rank.

Theorem 1 ([9])**.** *If the system $\mathscr{B} \in \mathscr{L}^{\mathsf{w},\mathsf{n}}_{\mathsf{m},\mathsf{l}}$ and the time series $w_{\mathrm{d}} \in (\mathbb{R}^{\mathsf{w}})^T$ satisfy the following identifiability conditions:*

1. *$w_{\mathrm{d}} \in \mathscr{B}|_{[1,T]}$,*
2. *$\mathscr{B}$ is controllable, and*
3. *an input component u_{d} of w_{d} is persistently exciting of order $\mathbf{l}(\mathscr{B}) + 1 + \mathbf{n}(\mathscr{B})$,*

then $\mathscr{B} = \mathscr{B}_{\mathrm{mpum}}(w_{\mathrm{d}})$.

The first condition is the most restrictive one: it requires the data to be exact. In practice, the data are almost never exact, because

- it is likely to be generated by a more complex system than any LTI systems of bounded complexity (cf., the model class $\mathscr{L}^{\mathsf{w},\mathsf{n}}_{\mathsf{m},\mathsf{l}}$),
- there are unobserved inputs that act on the system (cf., the process noise in the ARMA model),
- there are measurement noises (cf., the measurement noise in the output error model (5)).

The second condition of the theorem is more restrictive than it might look at first. For example, autonomous models do not satisfy it. The third condition is the least restrictive but its verification requires prior knowledge ($\mathbf{l}(\mathscr{B})$ and $\mathbf{n}(\mathscr{B})$) about the true data generating system.

Next we modify the result for the case at hand: autonomous model class.

Theorem 2. *If the system $\mathscr{B} \in \mathscr{L}^{\mathsf{p},\mathsf{n}}_{0,\mathsf{l}}$ and the time series $y_{\mathrm{d}} \in (\mathbb{R}^{\mathsf{p}})^T$ satisfy the following identifiability conditions:*

1. *$y_{\mathrm{d}} \in \mathscr{B}|_{[1,T]}$, and*
2. *y_{d} is persistently exciting of order $\mathbf{n}(\mathscr{B})$,*

then $\mathscr{B} = \mathscr{B}_{\mathrm{mpum}}(y_{\mathrm{d}})$.

Proof. In the infinite time case ($T = \infty$) $\mathscr{B}_{\text{mpum}}(y_\text{d}) = \text{span}(y_\text{d}, \sigma y_\text{d}, \ldots, \sigma^t y_\text{d}, \ldots)$. On one hand, since $\mathscr{B}$ is LTI and by assumption 1, we have

$$y_\text{d} \in \mathscr{B} \implies \text{span}(y_\text{d}, \sigma y_\text{d}, \ldots, \sigma^t y_\text{d}, \ldots) = \mathscr{B}_{\text{mpum}}(y_\text{d}) \subseteq \mathscr{B}.$$

On the other hand, by assumption 2, we have

$$\dim\big(y_\text{d}, \sigma y_\text{d}, \ldots, \sigma^t y_\text{d},)\ldots\big) \geq \mathbf{n}(\mathscr{B}) \implies \mathscr{B}_{\text{mpum}}(y_\text{d}) \supseteq \mathscr{B}.$$

Therefore, $\mathscr{B} = \mathscr{B}_{\text{mpum}}(y_\text{d})$.

In the finite horizon case, $\mathscr{B}_{\text{mpum}}(y_\text{d}) = \text{colspan}\big(\mathscr{H}_L(y_\text{d})\big)$, for all $L \geq \mathsf{l}$. By the LTI assumption and by assumption 1,

$$y_\text{d} \in \mathscr{B}|_{[1,T]} \implies \text{colspan}\big(\mathscr{H}_\mathsf{n}(y_\text{d})\big) = \mathscr{B}_{\text{mpum}}(y_\text{d})|_{[1,\mathsf{n}]} \subseteq \mathscr{B}|_{[1,\mathsf{n}]}.$$

By assumption 2,

$$\dim\big(\mathscr{H}_\mathsf{n}(y_\text{d})\big) \geq \mathbf{n}(\mathscr{B})|_{[1,\mathsf{n}]} \implies \mathscr{B}_{\text{mpum}}(y_\text{d})|_{[1,\mathsf{n}]} \supseteq \mathscr{B}|_{[1,\mathsf{n}]}.$$

Therefore, $\mathscr{B}|_{[1,\mathsf{n}]} = \mathscr{B}_{\text{mpum}}(y_\text{d})|_{[1,\mathsf{n}]}$. But $\mathscr{B}$ is an autonomous system of order n, so that $\dim(\mathscr{B}) = \mathsf{n}$. Therefore, $\mathscr{B} = \mathscr{B}_{\text{mpum}}(y_\text{d})$. □

In Theorem 2, we again assume that the data are exact so the comments made for Theorem 1 hold also for the modified result. Next, we show a numerical illustration of the result in the exact case and the implication of the presence of a (small) perturbation of the data. Then we consider methods for approximate identification and show their performance on the data in the example.

5 Simulation example and the approximation issue

Consider an exact data generating system $\bar{\mathscr{B}} \in \mathscr{L}^1_{0,4}$ and a trajectory $y_\text{d} \in \mathbb{R}^{250}$ of that system that is persistently exciting of order 4. The poles of $\bar{\mathscr{B}}$ are $0.8556 \pm 0.4674\,j$ and $0.8980 \pm 0.3797\,j$. Figure 1 shows the corresponding modes. The MPUM of y_d, computed by Algorithm 8.1 of [4], is the system $\mathscr{B}(P)$, where

$$P(\xi) = 0.1278\xi^0 - 0.4716\xi^1 + 0.7037\xi^2 - 0.4961\xi^3 + 0.1414\xi^4.$$

The assumptions of Theorem 2 are satisfied and indeed the poles of the MPUM $\mathscr{B}(P)$ are the poles $\{\, 0.8556 \pm 0.4674\,j,\ 0.8980 \pm 0.3797\,j \,\}$ of $\bar{\mathscr{B}}$.

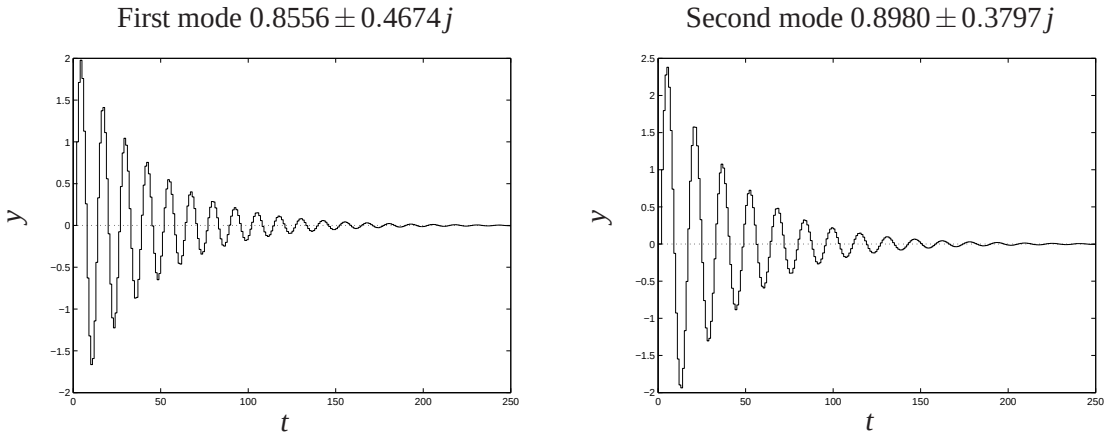

Figure 1: Physical modes in the example.

Next, we perturb the exact data y_d by a zero mean white Gaussian noise with standard deviation 0.05. The perturbed data are no longer exact for the true data generating system but the signal-to-noise (SNR) ratio

is high. The MPUM for the perturbed data fit the signal *and* the noise and is not equal to the true data generating system. Due to the high SNR ratio, however, it is relatively close to it.

Note that for noisy data $y_{\underline{d}}$, the order of the MPUM $\mathscr{B}_{\mathrm{mpum}}(y_{\mathrm{d}})$ is high ($\mathbf{n}(\mathscr{B}_{\mathrm{mpum}}(y_{\mathrm{d}})) = T$), while the true data generating system $\mathscr{B}$ is by assumption in a model class $\mathscr{L}_{0,\mathtt{l}}$ of bounded complexity. If we know the parameter $\mathtt{l}$, we could find an approximation $\hat{\mathscr{B}}$ of $\mathscr{B}_{\mathrm{mpum}}(y_{\mathrm{d}})$ in $\mathscr{L}_{0,\mathtt{l}}$. The optimal approximating system

$$\hat{\mathscr{B}}^* = \arg \min_{\hat{\mathscr{B}} \in \mathscr{L}_{0,\mathtt{l}}} \|\mathscr{B}_{\mathrm{mpum}}(y_{\mathrm{d}}) - \hat{\mathscr{B}}\|$$

satisfies the inequality

$$\|\hat{\mathscr{B}}^* - \bar{\mathscr{B}}\| \leq \|\mathscr{B}_{\mathrm{mpum}}(y_{\mathrm{d}}) - \bar{\mathscr{B}}\|,$$

i.e., in the $\|\cdot\|$-distance sense, $\hat{\mathscr{B}}^*$ is a better approximation of $\bar{\mathscr{B}}$ than $\mathscr{B}_{\mathrm{mpum}}(y_{\mathrm{d}})$.

The question occurs of how to infer from the data only what is the parameter $\mathtt{l}$. Equivalently, how to infer from the data what is the right model class. This question has been addressed in different communities under different names: it is called order selection in system identification, hyper parameter estimation in machine learning, and is closely related to the choice of the regularization parameter in numerical linear algebra. Correspondingly, different algorithms for model selection are derived: the Akaike information criterion, the cross-validation, and the L-curve, are a few examples.

Surprisingly in modal analysis, it is generally not recognized that the spurious pole detection problem addresses precisely the same issue and the stabilization diagrams are actually a method for selection of the model class. Next, we illustrate on the simulation example, described above, the following techniques for model selection:

1. a stabilization diagram using a covariance-driven stochastic subspace identification method,
2. the pole separation method of Kumaresan and Tufts,
3. the singular value analysis in Kung's algorithm,
4. the principal angle analysis in output-only subspace identification, and
5. the trade-off curve, used with a maximum likelihood method.

Our aim is to show the performance of a few alternatives to the stabilization diagrams. We are not exhaustive in covering the various available methods for model selection, e.g., we do not consider the well known Akaike information criterion and the cross-validation method. A more complete comparison of model selection methods on real-life data is a topic for further research.

5.1 A stabilization diagram

Given an identification method and data y_{d}, a set of models of orders 1, 2, 3, ... are identified and the modal frequencies (which are related to the poles' angles), are plotted against the model order. Typically, the horizontal axis represents the frequency and the vertical axis represents the model order. Thus, going upwards the diagram, one observes an increasing number of poles, corresponding to the increasing number of modes of the identified systems. Some of the poles "stabilize" in the sense that their frequencies do not change much, while some of the poles fluctuate. According to a heuristic rule behind the stabilization diagram, the former are physical and the latter are spurious poles.

Figure 2 shows the stabilization diagram, obtain using a covariance-driven stochastic subspace identification method[1], for the perturbed data in the simulation example. The vertical dotted lines correspond to the modes

[1]The authors are thankful to Bart Peeters from LMS, Leuven for providing the necessary software for plotting the stabilization diagram.

that stabilize. The method detects 6 stabilizing modes, corresponding to frequencies 0.0225Hz, 0.0638Hz, 0.0797Hz, 0.0866Hz, 0.1116Hz, and 0.1358Hz. The ones at 0.0638Hz and 0.0797Hz correspond to the physical modes, however, the other four are spurious.

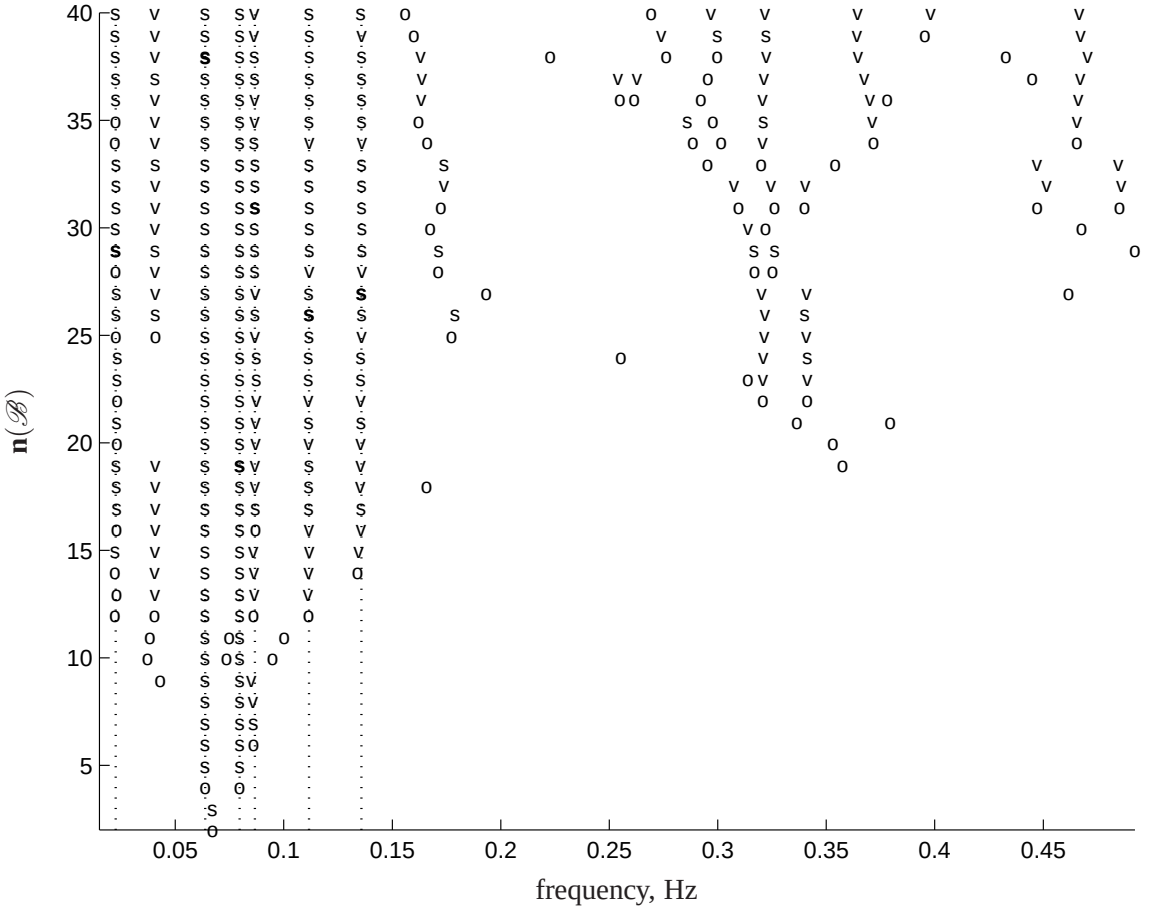

Figure 2: Stabilization diagram for the example. "○" pole that does not stabilize, "v" pole that stabilize in frequency only, "s" pole that stabilize in frequency and mode shape (however note that in this example the mode shape is a scalar because the system is single output).

5.2 The pole separation method of Kumaresan and Tufts

The model class $\mathscr{L}^1_{0,\mathsf{l}}$ with distinct poles corresponds to the "sum of damped exponentials" model class, considered by Kumaresan and Tufts [2]. The aim of the method, presented in [2], is to recover the data generating system $\bar{\mathscr{B}}$ from the data y_{d}, without knowing its lag l or equivalently its order n.

If the data y_{d} were exact, then rank $\big(\mathscr{H}_L(y_{\mathrm{d}})\big) \leq \mathsf{n}$, for all L, because by assumption y_{d} satisfies an autonomous LTI system of order n. Moreover, assuming that y_{d} is persistently exciting of order l, rank $\big(\mathscr{H}_L(y_{\mathrm{d}})\big) = \mathsf{n}$, for all $L \geq \mathsf{l}$. Consider the first time $L = \mathsf{l}+1$ when $\mathscr{H}_L(y_{\mathrm{d}})$ becomes rank deficient and let the rows of $\hat{P}$ form a basis for its left null space, i.e.,

$$\hat{P}\mathscr{H}_{\mathsf{l}+1}(y_{\mathrm{d}}) = 0.$$

Let $\mathscr{B}(P) = \bar{\mathscr{B}}$ be a difference equation representation of the true data generating system $\bar{\mathscr{B}}$ and let

$$P = \begin{bmatrix} P_0 & P_1 & \cdots & P_{\mathsf{l}} \end{bmatrix}, \quad \text{where} \quad P(\xi) =: \sum_{i=0}^{\mathsf{l}} P_i \xi^i.$$

Due to the persistency of excitation assumption

$$\operatorname{row\,span}(\hat{P}) = \operatorname{row\,span}(P).$$

Therefore, with

$$\hat{P}(\xi) := \sum_{i=0}^{\mathsf{l}} \hat{P}_i \xi^i, \quad \text{where} \quad \hat{P} =: \begin{bmatrix} \hat{P}_0 & \hat{P}_1 & \cdots & \hat{P}_{\mathsf{l}} \end{bmatrix}, \ \hat{P}_i \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$$

we have that $\mathscr{B}(\hat{P}) = \mathscr{B}(P) = \bar{\mathscr{B}}$ and

$$\lambda\big(\mathscr{B}(\hat{P})\big) = \lambda\big(\mathscr{B}(P)\big).$$

In the scalar case $\mathsf{p} = 1$, if P and $\hat{P}$ are normalized $P_0 = 1$ and $\hat{P}_0 = 1$, moreover we have that $P = \hat{P}$.

With $L > \mathsf{l}+1$, the Hankel matrix $\mathscr{H}_L(y_{\mathrm{d}})$ has a left kernel of dimension $L\mathsf{p} - \mathsf{n}$. Let the rows of $\hat{P}$ form a basis for that space. Assume for simplicity that $\mathsf{n} = \mathsf{lp}$. (This assumption holds always in the scalar case and generically in the multivariable case.) Then there is a change of basis matrix T, such that

$$T\hat{P} = \begin{bmatrix} P_0 & P_1 & \cdots & P_{\mathsf{l}} & 0 & \cdots & 0 \\ 0 & P_0 & P_1 & \cdots & P_{\mathsf{l}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & P_0 & P_1 & \cdots & P_{\mathsf{l}} \end{bmatrix}.$$

In a polynomial notation

$$\hat{P}(\xi) = S(\xi)P(\xi),$$

where $\operatorname{degree}(S(\xi)) = L - \mathsf{l} - 1$, so that

$$\lambda\big(\mathscr{B}(\hat{P})\big) = \lambda\big(\mathscr{B}(P)\big) \cup \lambda\big(\mathscr{B}(S)\big).$$

Define $\hat{\mathscr{B}} = \mathscr{B}(\hat{P})$ to be the identified model. It contains the poles of $\mathscr{B}$ (physical poles) and the $L - \mathsf{l} - 1$ zeros of S (spurious poles).

The choice of S (and therefore the choice of the spurious poles) depends on the identification method being used. For example, in the scalar case we are looking for a nontrivial solution $\hat{P} \in \mathbb{R}^{1\times L}$ of the system $\hat{P}\mathscr{H}_L(y_{\mathrm{d}}) = 0$. In [2], the normalization $\hat{P}_0 = 1$ is used and the *least norm* solution of the resulting system of equations is chosen. The least norm solution is shown to have the property that the spurious poles are outside the unit circle. Moreover, their exact locations depend only on L and T and not on y_{d}.

The fact that the spurious poles corresponding to the least norm solution $\hat{P}$ are outside the unit circle is used to separate them from the physical poles. The physical poles of a stable system $\bar{\mathscr{B}}$ are inside the unit circle and can be distinguished from the spurious poles. Figure 3 shows the poles plot of a 20th order system identified by the method of Kumaresan and Tufts. The poles are marked by crosses, the zeros by circles, and the four poles of $\bar{\mathscr{B}}$ by squares. The only identified poles that are stable are the physical poles, so that they can be detected by the method of Kumaresan and Tufts.

5.3 The singular value analysis in Kung's algorithm

Kung's algorithm [3] is a realization algorithm: it computes a state space representation of a system from its impulse response. Algorithms for impulse response realization, however, can be used for autonomous system identification by a trivial substitution, so Kung's algorithm is applicable in the setting of our simulation example.

A feature that makes Kung's algorithm extremely popular in system identification and signal processing is that it allows to compute very good approximate models in the case when the data are not exact. The reason for this is that in the infinite time case ($T = \infty$) Kung's algorithm corresponds to the balanced model reduction method, which is known to be a very effective heuristic for model reduction. In the finite time case, Kung's algorithm performs *finite-time* balanced model reduction. In many cases, finite-time balanced model reduction is virtually as good as its infinite-time counterpart.

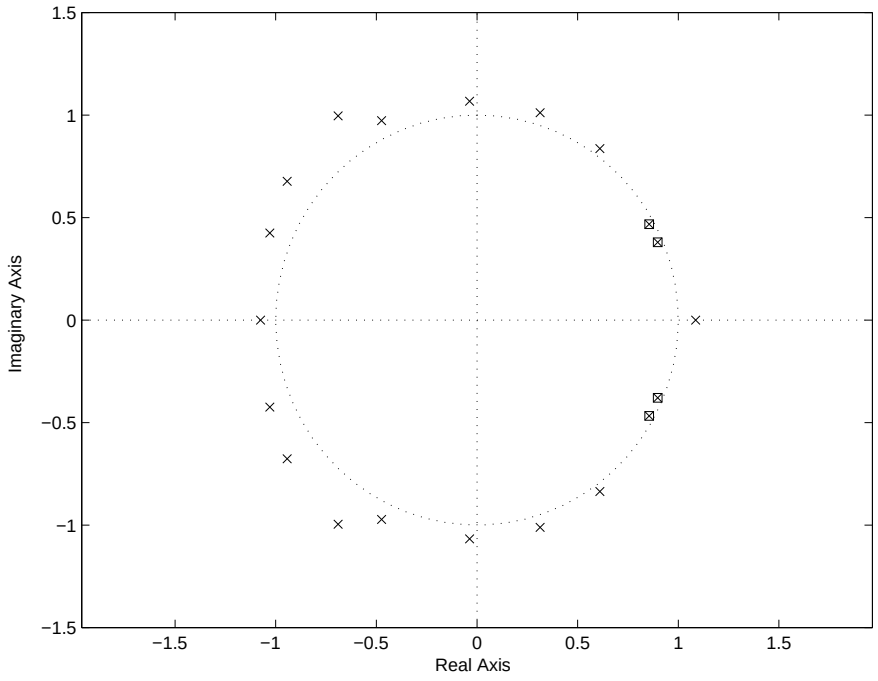

Figure 3: Poles plot for the system identified by the method of Kumaresan and Tufts.
$\times$ — poles of the identified system, $\square$ — physical poles.

In Kung's algorithm, the order selection is done on the basis of the singular values of the Hankel matrix $\mathscr{H}_L(y_\mathrm{d})$, where the parameter L is chosen large enough compared to the lag of the true data generating system. (For example, L can be chosen, so that $\mathscr{H}_L(y_\mathrm{d})$ is nearly square.) Figure 4, right, shows the singular values plot of $\mathscr{H}_{20}(y_\mathrm{d})$. The first four singular values are much larger than the remaining singular values, which indicates that there is a very good approximate model of order $\mathsf{n} = 4$. Computing an approximate model in the model class $\mathscr{L}^1_{0,4}$ by Kung's method, we obtain an approximation of the true poles, which is correct up to the 4th digit.

Balanced model reduction has been used for spurious poles detection in [1].

5.4 The principal angle analysis in output-only subspace identification

Similar to the order selection in Kung's algorithm, in output-only subspace identification the system order is chosen from what are called principal angles (between the subspaces spanned by the past and the future data). Figure 4, left, shows the principal angles computed by the N4SID method when evoked with the data in the example. Four of the principal angles are close to 0°, while the others are close to 90°. This is again an indication that there is a very good model of order $\mathsf{n} = 4$.

5.5 The misfit–complexity trade-off in a maximum likelihood identification

Finally, we consider an optimization-based identification method that is the maximum likelihood method in the output error model (5). In this case, an indication for a relevant model class is given by the misfit–complexity trade-off curve. This curve shows the fitting error (misfit) of the identified model and the data as a function of the model order and typically has an "L" shape. A good model class is indicated by the corner: it indicates a simple model that fits still well the data. Figure 5 shows the misfit–complexity trade-off curve for the data in the example. The order of the true data generating system is easy to infer from the curve: models of order 4 or higher achieve virtually the same fit and models of order less than 4 have significantly

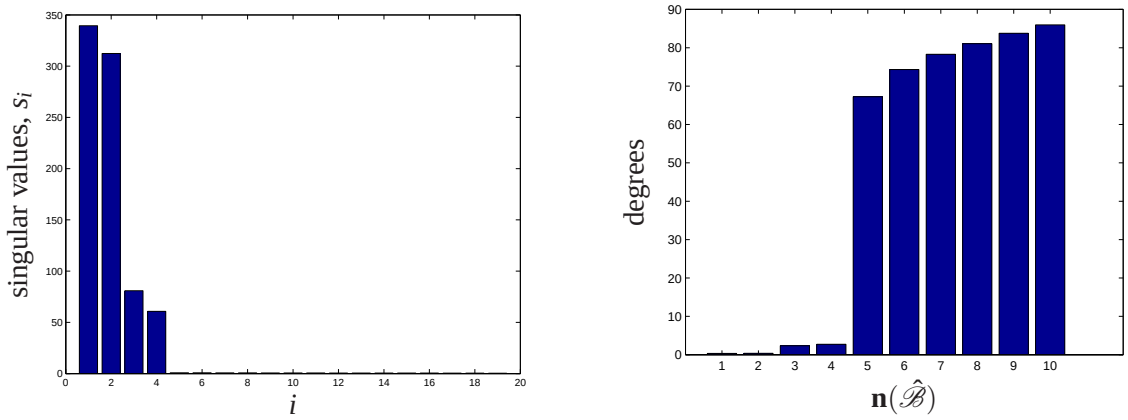

Figure 4: Left: Singular values of $\mathscr{H}_{20}(y_\mathrm{d})$. Right: Principal angles in output-only subspace identification.

higher misfit with the data. Therefore, the simplest model that fits well the data is of order 4, which is the true system's order.

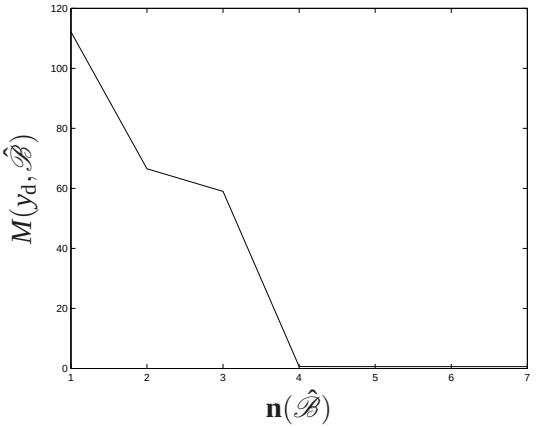

Figure 5: Order selection from the misfit–complexity trade-off curve.

6 Conclusions

A basic presumption used in spurious poles detection methods, based on stabilization diagrams, is that as the model order increases, the estimation accuracy improves. This presumption is not universally valid. It depends on

1. the identification method being used, and
2. how much the model order is increased.

For example, for the Yule-Walker estimation method it is a well known rule-of-thumb that the approximation accuracy is highest for model order about 2/3 the data length, which is typically much more than the intended model order. Therefore, a subsequent model reduction step is needed. For the maximum likelihood methods, however, the increased number of parameters is likely to deteriorate the results due to the increased number of tunable parameters and the resulting increased danger of convergence to a local minimum.

Once a model of complexity higher than intended is estimated, the next step is to perform model reduction in order to obtain a model of certain desired lower complexity. Model reduction is as hard as the identification step itself; optimal model reduction as optimal approximate system identification, in general, requires

a nonconvex optimization. In addition, it is not a priori clear why the two step procedure (first identify a high order model and then do model reduction) should be superior over the direct procedure. In fact, from an optimization point of view, the direct identification of an optimal low complexity model can not be outperformed. An advantage of the two step procedure, however, is the existence of effective heuristic methods for performing the two steps separately. This is exploited in the methods of Kumaresan and Tufts, Kung, and the subspace identification methods.

Acknowledgments

We thankful to Bart Peeters from LMS, Leuven for providing his software for plotting the stabilization diagram.

Research supported by: **Research Council KUL:** GOA-Mefisto 666, GOA-Ambiorics, IDO/99/003 and IDO/02/009 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc & fellow grants; **Flemish Government: FWO:** PhD/postdoc grants, projects, G.0078.01 (structured matrices), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), G.0240.99 (multilinear algebra), G.0407.02 (support vector machines), G.0197.02 (power islands), G.0141.03 (Identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), G.0452.04 (QC), G.0499.04 (robust SVM), research communities (ICCoS, ANMMM, MLDM); AWI: Bil. Int. Collaboration Hungary/ Poland; **IWT:** PhD Grants; GBOU (McKnow) **Belgian Federal Government:** DWTC (IUAP IV-02 (1996-2001) and Belgian Federal Science Policy Office IUAP V-22 (2002-2006) (Dynamical Systems and Control: Computation, Identification & Modelling)); PODO-II (CP/01/40: TMS and Sustainibility); **EU:** PDT-COIL, BIOPATTERN, eTUMOUR, FP5-Quprodis; ERNSI; Eureka 2063-IMPACT; Eureka 2419-FliTE; Contract Research/agreements: ISMC/IPCOS, Data4s, TML, Elia, LMS, IPCOS, Mastercard.

References

[1] I. Goethals and B. De Moor. Model reduction and energy analysis as a tool to detect spurious modes. In *Proc. of the International Conference on Noise and Vibration Engineering*, Leuven, Belgium, 2002.

[2] R. Kumaresan and D. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans. Acust., Speech, Signal Process.*, 30(6):833–840, 1982.

[3] S. Kung. A new identification method and model reduction algorithm via singular value decomposition. In *Proceedings of the 12th Asilomar Conference on Circuits, Systems, and Computers*, pages 705–714, Pacific Grove, CA, 1978.

[4] I. Markovsky, J. C. Willems, B. De Moor, and S. Van Huffel. *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*. Number 11 in Monographs on Mathematical Modeling and Computation. SIAM, March 2006.

[5] I. Markovsky, J. C. Willems, P. Rapisarda, and B. De Moor. Algorithms for deterministic balanced subspace identification. *Automatica*, 41(5):755–766, 2005.

[6] B. Moore. Principal component analysis in linear systems: Controllability, observability and model reduction. *IEEE Trans. Automat. Control*, 26(1):17–31, 1981.

[7] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer, Boston, 1996.

[8] J. C. Willems. From time series to linear system—Part II. Exact modelling. *Automatica*, 22(6):675–694, 1986.

[9] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor. A note on persistency of excitation. *Control Lett.*, 54(4):325–329, 2005.