2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

# Informative Data and Identifiability in LPV-ARX Prediction-Error Identification

Arne G. Dankers, Roland Tóth, Peter S. C. Heuberger, Xavier Bombois and Paul M. J. Van den Hof

*Abstract*— **In system identification, the concepts of informative data and identifiable model structures are important for addressing the statistical properties of estimated models. In this paper, these two concepts are generalized from the classical LTI prediction-error identification framework to the situation of LPV model structures and appropriate definitions are introduced. For two particular cases (piecewise constant and periodic scheduling trajectories) conditions are derived for the data sets to be informative w.r.t. the LPV-ARX model structure. Moreover, conditions are derived under which the LPV-ARX model structure is globally identifiable.**

## I. Introduction

Efficient control of high-tech systems such as precision mechatronic devices, aircrafts, and chemical plants, requires accurate but simple models of the nonlinear and/or time-varying behavior of these applications. For many nonlinear systems, the linear parameter-varying (LPV) framework offers a nice trade-off between accuracy and parsimony. Moreover, it offers convex control synthesis for nonlinear systems in a computationally attractive setting [1].

In the LPV framework, signal relations are considered to be linear just as in the LTI case, but these relations are assumed to be varying as a function of a measurable signal, the so-called scheduling variable. Recently an LPV prediction-error identification framework has been developed in [1] providing a theoretical basis which can be used for the estimation of LPV predictor models. In this framework LPV-ARX, LPV-ARMAX, etc. model structures are defined which are generalizations of the LTI model structures.

In the LTI prediction-error identification theory, it is well known that the data set must be informative w.r.t. the model structure in order to obtain a consistent estimate of the dynamic relations, and that a model structure must be identifiable in order to obtain a consistent estimate of the parameter vector [2], [3], [4].

The concepts of informative data and identifiability are also important in the LPV prediction-error identification framework, however the current LTI definitions are not directly transferable to the LPV setting. This is due to the lack of a transfer function representation in the LPV framework, and the inclusion of the scheduling variable.

The first half of this paper focuses on investigating and defining informative data and identifiability for LPV systems.

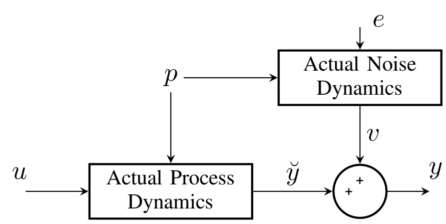


Fig. 1. LPV data generating system

In the current literature, conditions on the LPV data set are derived such that the identification problem is well conditioned [5], [6]. While this is related to informative data and identifiability, neither paper defines these concepts. Getting a clear understanding of these definitions is important in the analysis of the LPV prediction-error framework, for instance when determining consistency and convergence of the estimates of the dynamic relations of the signals and the estimates of the parameters.

In the second half of the paper the new definitions of informative data and identifiability are used to derive conditions on the data set with respect to the LPV-ARX model structure. In particular conditions are derived for two common trajectories of the scheduling variable - periodic and piecewise constant - and, conditions are derived such that the LPV-ARX model structure is identifiable.

In Sec. II the LPV prediction-error framework is briefly summarized. In Sec. III and IV the concepts of identifiability and informative data for LTI and LPV models respectively are investigated. In Sec. V the established definitions are applied to the LPV-ARX model structure.

## II. LPV prediction-error framework

In this section the LPV prediction-error framework will be briefly introduced. For a detailed presentation and analysis see [1]. Throughout the remainder of the paper, let $u$, $p$, $y$ denote the input, scheduling, and output variables respectively. In Fig. 1 the LPV data generating system is shown. The actual process dynamics have the form:

$$A_0(q,p,k)\breve{y}(k) = B_0(q,p,k)u(k)$$

where $A_0(q,p,k)$ and $B_0(q,p,k)$ are polynomials in the shift operator $q^{-1}$, where $q^{-1}u(k) = u(k-1)$:

$$A_0(\cdot) = 1+\sum_{i=1}^{n_a^0} a_i^0(p,k)q^{-i},\ B_0(\cdot) = \sum_{i=0}^{n_b^0} b_i^0(p,k)q^{-i}, \quad (1)$$

with $n_a^0, n_b^0 \geq 0$ and with $p$-dependent coefficients $a_i^0$ and $b_i^0$ defined as

$$a_i^0(p,k) = \sum_{j=0}^{n_\alpha^0} \alpha_{i,j}^0 f_j(p,k),\ b_i^0(p,k) = \sum_{j=0}^{n_\beta^0} \beta_{i,j}^0 g_j(p,k),$$

This work is supported in part by NSERC of Canada and the Netherlands Organization for Scientific Research (grant no. 680-50-0927)

The authors are with the Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands {a.g.dankers, r.toth, p.s.c.heuberger, x.j.a.bombois, p.m.j.vandenhof}@tudelft.nl

$f_j$ and $g_j$ are functions that are bounded on the range of $p$

*Remark 1:* In this paper the coefficients $a_i$ and $b_i$ will be functions of the instantaneous value of $p$, i.e. $f_j(p,k) = f_j\big(p(k)\big)$ and $g_j(p,k) = g_j\big(p(k)\big)$. In general however, it is possible that they also depend on shifted values of $p$.

The actual noise dynamics have the form:

$$D_0(q,p,k)v(k) = C_0(q,p,k)e(k),$$

where $e$ is a white noise process, and the $p$-dependent polynomials $D_0$ and $C_0$ with order $n_d^0 \geq n_c^0 \geq 0$ are monic and defined similarly to $A_0$ with a linear parameterization in terms of the nonlinear functions $h_j(\cdot)$ and $r_j(\cdot)$.

The LPV data generating system can now be defined as:

$$\begin{aligned} A_0(q,p,k)\breve{y}(k) &= B_0(q,p,k)u(k), \\ D_0(q,p,k)v(k) &= C_0(q,p,k)e(k), \\ y(k) &= \breve{y}(k) + v(k). \end{aligned} \tag{2}$$

A model, denoted $\mathcal{M}(\theta)$ where $\theta$ is the parameter vector, will be used to approximate the actual process dynamics and noise dynamics. A model set $\mathcal{M}$ results when $\theta$ is assumed to vary over a set, $\theta \in D_{\mathcal{M}} \subset \mathbb{R}^d$, where $d$ is the number of parameters. In this paper, the one-step-ahead prediction error will be minimized to find the optimal parameter vector (so called prediction-error identification).

The LPV-ARX model structure is a generalization of the LTI-ARX model structure, i.e. let $C(\cdot) = 1$ and $D(\cdot) = A(\cdot)$. From [1], the set of predictors with LPV-ARX structure can be formulated (under the assumption that noise-free observations of $p(k)$, $p(k-1)$, $\ldots$ are available) as:

$$\hat{y}(k|\theta) = B(q,p,k,\theta)u(k) + \big(1 - A(q,p,k,\theta)\big)y(k), \tag{3}$$

where $A$ and $B$ are defined analogously to (1) and have 'orders' $n_a, n_\alpha, n_b, n_\beta$ respectively, and

$$\theta = [\beta_{0,0} \ \cdots \ \beta_{n_b,n_\beta} \ \alpha_{1,0} \ \cdots \ \alpha_{n_a,n_\alpha}] \in D_{\mathcal{M}} \subset \mathbb{R}^d.$$

The LPV-ARX predictor model, (3), is a function of the data $\{u(k), p(k), y(k)\}_{k=0,\ldots,N-1}$, and the initial conditions $\{u(k)\}_{k=-n_b,\ldots,-1}$ and $\{y(k)\}_{k=-n_a,\ldots,-1}$. Denote the data set as:

$$Z^N = \left\{ \begin{array}{l} \{u(k), p(k), y(k)\}_{k=0,\ldots,N-1}, \\ \{u(k)\}_{k=-n_b,\ldots,-1}, \\ \{y(k)\}_{k=-n_a,\ldots,-1}. \end{array} \right\}. \tag{4}$$

Throughout the paper, the predictor model will be denoted $\hat{y}(k|\theta)$ or $\hat{y}(k|Z^N,\theta)$ or $\hat{y}(k|Z^N)$ depending on whether certain dependencies are to be emphasized.

One difference between the LPV prediction-error framework and the LTI prediction-error framework is that in the LPV case there is no transfer function representation - the dynamic map between $u$ and $y$ is not constant over time, but depends on the time-varying signal $p$ [1].

## III. LTI Informative Data and Identifiability

In this section the concepts of informative data and identifiability in the LTI prediction-error framework will be briefly recalled. A predictor model in the LTI case is defined as [2]:

$$\hat{y}(k|\theta) = W_u(q)u(k) + W_y(q)y(k) = W(q)z(k).$$

where $z(k) = [u(k) \ y(k)]^T$. When studying informative data and identifiability in the classical LTI framework $u$ and $y$ are assumed to be quasistationary and the data set is assumed to be $\{u(k), y(k)\}_{k=0,\ldots,\infty}$ where the initial conditions become negligible due to the infinite size of the data set.

*Statement 1:* The fundamental reason that conditions must be put on the data set is to ensure that "different" models result in "different" predicted outputs. □

A data set that meets these requirements is said to be informative w.r.t. the model structure under consideration. The reason why different is in quotations is because it has been used in an ambiguous manner. The following definition formalizes model equality.

*Definition 1:* LTI predictor models $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{M}$ are *equal* if $W_1(e^{j\omega}) - W_2(e^{j\omega}) = 0$ for almost all $\omega \in [0,\pi]$. □

Using Definition 1, Statement 1 can be formalized.

*Definition 2:* A data set $Z^\infty$ is informative w.r.t. an LTI model structure $\mathcal{M}$ if for any $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{M}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathrm{E}[\big((\hat{y}_1(k|Z^N) - \hat{y}_2(k|Z^N)\big)^2] = 0 \Rightarrow \mathcal{M}_1 = \mathcal{M}_2$$

where $\mathrm{E}[\cdot]$ is the expected value operator, and equality of models is defined in Definition 1. □

Next, the concept of identifiability will be investigated.

*Statement 2:* A model structure is said to be *globally identifiable at* $\theta_1$ if any model that is "different" from the model represented by $\theta_1$ is represented by a "different" parameter vector in $D_{\mathcal{M}}$. The map from $D_{\mathcal{M}}$ to $\mathcal{M}$ is one to one and onto (bijective). □

Again, using Definition 1, Statement 2 can be formalized.

*Definition 3:* A model structure $\mathcal{M}$ is *globally identifiable at* $\theta_1$ if, for any $\theta_2 \in D_{\mathcal{M}}$,

$$\mathcal{M}(\theta_1) = \mathcal{M}(\theta_2) \Longrightarrow \theta_1 = \theta_2,$$

where equality of models is defined in Definition 1. □

One last concept which will be useful later on is the concept of persistence of excitation.

*Definition 4:* Let $R_u(k) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \mathrm{E}[u(i)u(i-k)]$. A signal $u$ is *persistently exciting of order* $n$ if the matrix

$$\bar{R} = \begin{bmatrix} R_u(0) & \cdots & R_u(n-1) \\ \vdots & & \vdots \\ R_u(n-1) & \cdots & R_u(0) \end{bmatrix} \tag{5}$$

is nonsingular [2]. □

*Remark 2:* Informativity is a property of a data set (and dependent on a model structure), identifiability is a property of a model structure, and persistence of excitation is a property of a signal (and is independent of the model structure).

Informative data is important to ensure that the identification criterion can discriminate between models. A common identification criterion is:

$$V_N(\theta) = \frac{1}{N} \sum_{k=0}^{N-1} \big(y(k) - \hat{y}(k|\theta, Z^N)\big)^2. \tag{6}$$

If the data set is not informative then the identification criterion cannot discriminate between different models.

If a model structure is not identifiable it *could* happen that the minimizer of the identification criterion will be a set of parameter vectors. It is clear that identifiability is a less crucial property than informative data.

Next, the concepts of informative data and identifiability will be generalized such that they can be applied to the LPV prediction-error framework.

## IV. LPV Informative Data and Identifiability

When formulating the definitions of informative data and identifiability in the LPV framework, the following points will be taken into consideration:

- LPV models cannot be described by a transfer function, therefore model equality cannot be defined as in Definition 1 and,
- the scheduling variable $p$ is a function that is given to the user for only a finite time period with no obvious extension to an infinite time signal.

Due to the second reason all the results in the sequel will be stated in a finite time framework.

The equivalence of two LPV predictor models will be defined as follows:

*Definition 5:* Two models $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{M}$ are *equal* if

$$\hat{y}_1(k|Z^N) - \hat{y}_2(k|Z^N) = 0,\ k = 0, \dots, N-1$$

for *all* data sets $Z^N$ of the form (4) for all $N > 0$. □

Statement 1 can now be formalized in the LPV setting.

*Definition 6:* A data set $Z^N$ of the form (4) is informative w.r.t. an LPV model structure $\mathcal{M}$ if for any $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{M}$,

$$\hat{y}_1(k|Z^N) - \hat{y}_2(k|Z^N) = 0, k = 0, ..., N-1 \Rightarrow \mathcal{M}_1 = \mathcal{M}_2$$

where equality of models is defined in Definition 5. □

Formalizing Statement 2 in the LPV setting results in:

*Definition 7:* An LPV model structure $\mathcal{M}$ is *globally identifiable at* $\theta_1$ if, for any $\theta_2 \in D_{\mathcal{M}}$,

$$\mathcal{M}(\theta_1) = \mathcal{M}(\theta_2) \Longrightarrow \theta_1 = \theta_2,$$

where equality of models is defined in Definition 5. □

By Definition 7, $\mathcal{M}$ is globally identifiable at $\theta_1$ if for any $\theta_2 \in D_{\mathcal{M}}$, $\theta_2 \neq \theta_1$ there exists a data set $Z^N$ such that

$$\sum_{k=0}^{N-1} \left(\hat{y}(k|\theta_1, Z^N) - \hat{y}(k|\theta_2, Z^N)\right)^2 \neq 0. \tag{7}$$

Thus identifiability is a property of the model structure, not the data set (it only depends on the existence of a data set).

*Remark 3:* For a model structure that is identifiable, and a data set that is informative w.r.t. the model structure the following implication holds: if two models have the same predicted outputs then $\theta_1 = \theta_2$. □

The concept of persistence of excitation must also be adapted to a finite time framework.

*Definition 8:* Let $R_u^N(k) = \frac{1}{N}\sum_{i=0}^{N-1} u(i)u(i-k)$. A finite length signal $\{u(k)\}_{k=0,\dots,N-1}$ is *persistently exciting of order* $n$ if the matrix

$$\bar{R}_N = \begin{bmatrix} R_u^N(0) & \cdots & R_u^N(n-1) \\ \vdots & & \vdots \\ R_u^N(n-1) & \cdots & R_u^N(0) \end{bmatrix} \tag{8}$$

is nonsingular. □

*Remark 4:* The definitions in Sec. IV are formulated in a setting where only one realization of a system is known (due to the reasons mentioned at the beginning of this section), whereas in Sec. III they are formulated in a stochastic setting where the probability distributions of the signals are known.

In the next section, Definitions 6, 7, 8 are applied to the LPV-ARX model structure.

## V. Informative data and Identifiablilty of the LPV-ARX model structure

Given a LPV-ARX model structure, conditions on the data set will be derived for two specific trajectories of the scheduling variable: periodic and piecewise constant. The proofs of the theorems will use a notation based on the proofs in [5] but will not be restricted to LPV-ARX models with $n_\alpha = n_\beta$, $n_a = n_b$, and $f_i = g_i$ as is the case in [5]. First, a lemma, a definition and an assumption will be presented.

*Lemma 1 ([7]):* Let $A$ and $B$ be an $m \times n$ and $n \times p$ matrices. Then $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$. Moreover, if $B$ is such that $\text{rank}(B) = n$, then $\text{rank}(AB) = \text{rank}(A)$.

*Definition 9 ([8]):* A set of $n$ functions $\{f_1, \dots, f_n\}$ defined on a domain $\Omega$ is called *unisolvent* on $\Omega$ if

$$\begin{bmatrix} f_1(x_1) \\ \vdots \\ f_1(x_n) \end{bmatrix}, \dots, \begin{bmatrix} f_n(x_1) \\ \vdots \\ f_n(x_n) \end{bmatrix}$$

are linearly independent for any $x_i \in \Omega$, $x_i \neq x_j, i \neq j$.

Some examples of unisolvent sets of functions are: $\{1, x, \dots, x^{n-1}\}$ on any interval $[a, b]$; and $\{1, \cos(x), \sin(x), \dots, \cos(nx), \sin(nx)\}$, on interval $[-\pi, \pi]$ [8]. An example of a set that is not unisolvent is $\{1, x^2\}$ defined on $[-a, a]$.

The noise will be characterized in the following way:

*Assumption 1:* Let $W \in \mathbb{R}^{N \times n_b}$ be a matrix with columns containing delayed versions of $u$. Let $\vec{y}_{-i}$ denote a vector containing a delayed version of y: $\vec{y}_{-i}^{\,T} = [y(-i)\ \dots\ y(N{-}1{-}i)]$. Recall from (2) that $y = \breve{y} + v$. Assume $v$ is such that the matrix $[W\ \vec{y}_{-1}\ \cdots\ \vec{y}_{-n_a}]$ has rank equal to $\text{rank}(W) + n_a$, i.e. the noise is such that the vectors $\vec{y}_{-1}, \dots \vec{y}_{-n_a}$, are linearly independent of each other and $W$. □

Note that vectors containing samples from an independent identically distributed random variable satisfy Assumption 1 with probability one.

Finally, before stating the main theorem of this section, consider a representation of a signal that reveals its persistence of excitation:

$$u(k) = \sum_{\ell=1}^{n_u} \sum_{m=0}^{n_\ell - 1} \zeta_{\ell,m} k^m \xi_\ell^k,\ \xi_i, \zeta_i \in \mathbb{C}, \tag{9}$$

where $\zeta_{\ell,n_\ell-1} \neq 0, \xi_i \neq \xi_j, i \neq j$. Any finite length signal can be representation in this way, and $u$ is persistently exciting of order $\sum_{\ell=1}^{n_u} n_\ell$ for sufficiently large $N$ [9]. Consequently, the order of excitation of $u$ can be increased either by adding another basis function $\xi_i^k$ where $\xi_i \neq \xi_j, i \neq j$, or by increasing $n_\ell$ of one of the existing basis functions (again, for sufficiently large $N$).

To keep the proofs of the theorems clear, the following sub-class of (9) will be considered in the sequel:

$$\mathcal{U}: \{u(k) = \sum_{\ell=1}^{n_u} \zeta_\ell \xi_\ell^k,\ \xi_i, \zeta_i \in \mathbb{C}, \zeta_i \neq 0, \xi_i \neq \xi_j, i \neq j\}. \quad (10)$$

The initial conditions of $u$ are $u(-n_i), \ldots, u(-1)$. The class $\mathcal{U}$ is comprised of sums of decaying exponentials and sinusoids. For example $u(k) = 0.5e^{-j\omega k} + 0.5e^{j\omega k}$ and $u(t) = a^k + b^k$, $a \neq b$ are in the set. The theorems of this section could be derived completely analogously using the class of signals (9).

*Remark 5:* Investigate the persistency of excitation for finite length signals $u \in \mathcal{U}$. The matrix $\bar{R}_N$ in Definition 8 can be factored as $\bar{R}_N = \tilde{R}_N^T \tilde{R}_N$, where:

$$\begin{aligned}\tilde{R}_N &= \begin{bmatrix} u(0) & \ldots & u(-n+1) \\ \vdots & & \vdots \\ u(N-1) & \ldots & u(N-n) \end{bmatrix} \\ &= \begin{bmatrix} 1 & \cdots & 1 \\ \xi_1 & \cdots & \xi_{n_u} \\ \vdots & & \vdots \\ \xi_1^{N-1} & \cdots & \xi_{n_u}^{N-1} \end{bmatrix} B_i \begin{bmatrix} 1 & \xi_1^{-1} & \cdots & \xi_1^{-n+1} \\ 1 & \xi_2^{-1} & \cdots & \xi_2^{-n+1} \\ \vdots & \vdots & & \vdots \\ 1 & \xi_{n_u}^{-1} & \cdots & \xi_{n_u}^{-n+1} \end{bmatrix}\end{aligned}$$

where $B_i = \mathrm{diag}(\zeta_1, \ldots, \zeta_{n_u})$. There are two cases:

- $(n \leq n_u \leq N)$. By Lemma 1, $\mathrm{rank}(\tilde{R}_N) = n$.
- $(n \leq N \leq n_u)$. By Lemma 1, $\mathrm{rank}(\tilde{R}_N) \leq n$.

In the second case the rank can be less than $n$ if $\xi_1, \ldots, \xi_{n_u}$ are chosen properly. However, the probability of this reduction in rank is zero with $\xi_1, \ldots, \xi_{n_u}$ chosen arbitrarily. Therefore in both cases the signal $u$ can be said to be generically persistently exciting of order $n$. □

In the following theorem conditions are derived such that the data set is informative w.r.t. the LPV-ARX model structure. The class of permissible functions in the parameterization of $a_i$ and $b_i$ is restricted by unisolvency; the class of possible input signals is restricted by persistency of excitation (in the sense of Definition 8); and the scheduling trajectory is restricted to be piecewise constant.

*Theorem 1:* Consider an LPV-ARX model structure with a parameterization in terms of the set of functions $\mathcal{F} = \{f_1, \ldots, f_{n_\alpha}\}$ and $\mathcal{G} = \{g_1, \ldots, g_{n_\beta}\}$. Let $u \in \mathcal{U}$ with the number of initial conditions $n_i = n_b$ (see (10)). Let $p$ be piecewise constant with $\ell$ levels, each of length $m_1, \ldots, m_\ell$ respectively. Let $m_i \geq n_a + n_b + 1$, $i = 1, \ldots, \ell$. Let $\mathcal{F}$ and $\mathcal{G}$ be unisolvent on the interval $[\min(p), \max(p)]$. Then, generically (due to Remark 5), the data set $Z^N$ is informative if and only if

(a) $n_u \geq n_b + 1$,
(b) $\ell \geq \max\{n_\alpha + 1, n_\beta + 1\}$.

Note that conditions (a)-(b) imply that $N \geq d$.

*Proof:* From the definition of informative data, it must be shown that for this data set, if the predicted outputs of two models are the same, then the models are the same. Using (3), the difference between two LPV-ARX predictors is

$$\begin{aligned}\hat{y}(k|\theta_1) - \hat{y}(k|\theta_2) = \big(B(q, p(k), \theta_1) - B(q, p(k), \theta_2)\big)u(k) \\ - \big(A(q, p(k), \theta_1) - A(q, p(k), \theta_2)\big)y(k)\end{aligned}$$

which can be grouped together into a matrix expression: $[\phi_u\ \phi_y](\theta_1 - \theta_2)$, where

$$\begin{aligned}\theta_i^T = [&\beta_{0,0}^i \ \ldots\ \beta_{n_b,0}^i \ \ldots\ \beta_{0,n_\beta}^i \ \ldots\ \beta_{n_b,n_\beta}^i \ \ldots \\ &\ldots\ \alpha_{1,0}^i \ \ldots\ \alpha_{n_a,0}^i \ \ldots\ \alpha_{1,n_\alpha}^i \ \ldots\ \alpha_{n_a,n_\alpha}^i] \in \mathbb{R}^d\end{aligned}$$

$$\phi_u^T = \begin{bmatrix} u(0)g_0\big(p(0)\big) & \cdots & u(N-1)g_0\big(p(N-1)\big) \\ \vdots & & \vdots \\ u(-n_b)g_0\big(p(0)\big) & \cdots & u(N-n_b-1)g_0\big(p(N-1)\big) \\ \vdots & & \vdots \\ u(0)g_{n_\beta}\big(p(0)\big) & \cdots & u(N-1)g_{n_\beta}\big(p(N-1)\big) \\ \vdots & & \vdots \\ u(-n_b)g_{n_\beta}\big(p(0)\big) & \cdots & u(N-n_b-1)g_{n_\beta}\big(p(N-1)\big) \end{bmatrix}$$

where $i = 1, 2$, and $\phi_y$ is defined similarly, but with shifted versions of $y$. Let $\phi(Z^N) = [\phi_u\ \phi_y] \in \mathbb{R}^{N \times d}$. For a particular data set $Z_{\text{part}}^N$ to be informative, the implication in the definition of informative data can then be written as:

$$\phi(Z_{\text{part}}^N)(\theta_1 - \theta_2) = 0_N \Rightarrow \phi(Z^N)(\theta_1 - \theta_2) = 0\ \forall Z^N \quad (11)$$

$N \geq d$. The proof will proceed as follows. First it will be shown that $\phi(Z_{\text{part}}^N)$ is full rank iff the conditions listed in the theorem hold (part 1). Then it will be shown that the data is informative iff $\phi(Z_{\text{part}}^N)$ is full rank (part 2).

(Part 1 sufficiency). Assume the conditions hold. It must be shown that $\phi(Z_{\text{part}}^N)$ is full rank. The matrix $\phi(Z_{\text{part}}^N)$ will be factored which will allow for analysis of its rank.

By assumption, $p$ has $\ell$ unique levels, which are each $m_1, \ldots, m_\ell$ samples long. Within the sequence $p$, it is possible to find $\ell$ sequences such that:

$$\begin{aligned} p_1 = p(k_1) = \cdots &= p(k_1 + m_1 - 1) \\ &\vdots \\ p_\ell = p(k_\ell) = \cdots &= p(k_\ell + m_\ell - 1)\end{aligned}$$

where $p_i \neq p_j, i \neq j$, $0 = k_1 < k_2 < \cdots < k_\ell$. Using this notation, it is possible to factor $\phi(Z_{\text{part}}^N)$ such that:

$$\begin{aligned}&\begin{bmatrix} U_1 & & & Y_1 & & \\ & \ddots & & & \ddots & \\ & & U_\ell & & & Y_\ell \end{bmatrix} \begin{bmatrix} G \otimes I_{n_b+1} & \\ & F \otimes I_{n_a} \end{bmatrix} \\ &= \begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_\ell \end{bmatrix} P \begin{bmatrix} G \otimes I_{n_b+1} & \\ & F \otimes I_{n_a} \end{bmatrix} \\ &= WPH \in \mathbb{R}^{N \times d} \end{aligned} \quad (12)$$

where

$$G = \begin{bmatrix} g_0(p_1) & \cdots & g_{n_\beta}(p_1) \\ \vdots & & \vdots \\ g_0(p_\ell) & \cdots & g_{n_\beta}(p_\ell) \end{bmatrix}, F = \begin{bmatrix} f_0(p_1) & \cdots & f_{n_\alpha}(p_1) \\ \vdots & & \vdots \\ f_0(p_\ell) & \cdots & f_{n_\alpha}(p_\ell) \end{bmatrix},$$

$$U_i = \begin{bmatrix} u(k_i) & \cdots & u(k_i - n_b) \\ \vdots & & \vdots \\ u(k_i+m_i-1) & \cdots & u(k_i+m_i-n_b-1) \end{bmatrix},$$

$$Y_i = \begin{bmatrix} y(k_i-1) & \cdots & y(k_i - n_a) \\ \vdots & & \vdots \\ y(k_i+m_i-2) & \cdots & y(k_i+m_i-n_a-1) \end{bmatrix},$$

$P$ is a permutation matrix, $W_i = [U_i \; Y_i] \in \mathbb{R}^{m_i \times (n_a+n_b+1)}$, and $\sum_{i=1}^{\ell} m_i = N$. Note that the initial conditions of $u$ appear in $U_1$. Since $u \in \mathcal{U}$, each matrix $U_i$ can be written as

$$\begin{aligned} U_i &= \begin{bmatrix} 1 & \cdots & 1 \\ \xi_1 & \cdots & \xi_{n_u} \\ \vdots & & \vdots \\ \xi_1^{m_i-1} & \cdots & \xi_{n_u}^{m_i-1} \end{bmatrix} B_i \begin{bmatrix} 1 & \xi_1^{-1} & \cdots & \xi_1^{-n_b} \\ 1 & \xi_2^{-1} & \cdots & \xi_2^{-n_b} \\ \vdots & \vdots & & \vdots \\ 1 & \xi_{n_u}^{-1} & \cdots & \xi_{n_u}^{-n_b} \end{bmatrix} \\ &= U_{1,i} B_i U_{2,i} \end{aligned} \tag{13}$$

where $B_i = \mathrm{diag}(\zeta_1 \xi_1^{k_i}, \ldots, \zeta_{n_u} \xi_{n_u}^{k_i})$.

To conclude this section of the proof, Lemma 1 will be used to show $\phi(Z^N_{\text{part}})$ is full rank. Generically,

$$\text{rank}(W) = \ell(n_a + n_b + 1)$$

by Assumption 1 and Remark 5. And,

$$\text{rank}(H) = (n_b+1)\text{rank}(G) + n_a \text{rank}(F) = d$$

by unisolvency of $\mathcal{F}$ and $\mathcal{G}$. Moreover,

$$\begin{aligned} \dim(W) &= N \times \ell(n_a+n_b+1), \quad N \ge \ell(n_a+n_b+1) \\ \dim(H) &= \ell(n_a+n_b+1) \times d, \quad \ell(n_a+n_b+1) \ge d \end{aligned}$$

where the inequalities hold by condition (b). Then by Lemma 1, $\text{rank}(WPH) = d$, i.e. $\phi(Z^N_{\text{part}})$ is full rank.

(Part 1 necessity). Assume that $\phi(Z^N_{\text{part}})$ is full rank. It must be shown that the conditions (a) and (b) hold.

(a). Proof by contradiction. Suppose $n_u < n_b + 1$. Then

$$\text{rank}(U_{2,i}) = \min\{n_u, n_b+1\} = n_u,$$

and by Lemma 1, $\text{rank}(U_i) = \min\{m_i, n_u\} < n_b + 1$. Since the $\dim(U_i) = m_i \times (n_b + 1)$, at least one column can be reduced to zeros by elementary column operations, i.e. there exists a permutation matrix $P_u$ such that $U_i P_u = [\tilde{U}_i \; 0]$. Since each $U_{2,i}$ is the same for every $i$, the same $P_u$ will put a column of zeros at the end of every $U_i$. This means that there exists a permutation matrix such that:

$$\begin{aligned} \phi_u(Z^N_{\text{part}}) &= \begin{bmatrix} g_0(p_1)U_1 & \cdots & g_{n_\beta}(p_1)U_1 \\ g_0(p_2)U_2 & \cdots & g_{n_\beta}(p_2)U_2 \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} P_u & & \\ & P_u & \\ & & \ddots \end{bmatrix} \\ &= \begin{bmatrix} g_0(p_1)[\tilde{U}_1 \; 0] & \cdots & g_{n_\beta}(p_1)[\tilde{U}_1 \; 0] \\ g_0(p_2)[\tilde{U}_2 \; 0] & \cdots & g_{n_\beta}(p_2)[\tilde{U}_2 \; 0] \\ \vdots & & \vdots \end{bmatrix} \end{aligned}$$

which is clearly not full rank since it has columns of zeros. This is a contradiction to the original assumption that $\phi(Z^N_{\text{part}})$ is full rank, so conclude that $n_u \ge (n_b + 1)$.

(b). Proof by contradiction. Suppose $\ell < \max\{n_\alpha+1, n_\beta+1\}$. By assumption, and by Lemma 1

$$d = \text{rank}(\phi(Z^N_{\text{part}})) \le \min\{\text{rank}(W), \text{rank}(H)\}$$

so $\text{rank}(H)$ must be at least $d$. However,

$$\text{rank}(H) = \min\{\ell, n_\beta\}(n_b+1) + \min\{\ell, n_\alpha+1\}n_a < d$$

Which is a contradiction; conclude $\ell \ge \max\{n_\alpha+1, n_\beta+1\}$.

Finally, it will be shown that the data is informative iff $\phi(Z^N_{\text{part}})$ is full rank.

(Part 2 sufficiency). Assume $\phi(Z^N_{\text{part}})$ is full rank. It must be shown that implication (11) holds. Since $\phi(Z^N_{\text{part}})$ is full rank, the left hand side equation implies that $\theta_1 - \theta_2 = 0$. If $\theta_1 - \theta_2 = 0$, then the right hand side equation will also equal zero, which means the implication holds.

(Part 2 necessity). Assume implication (11) holds. It must be shown that $\phi(Z^N_{\text{part}})$ is full rank. Proof by contradiction: suppose $\phi(Z^N_{\text{part}})$ is not full rank. Let $\theta_1$, $\theta_2$ be such that $\theta_1 \neq \theta_2$, and $\phi(Z^N_{\text{part}})(\theta_1 - \theta_2) = 0$. Since $\mathcal{F}$ and $\mathcal{G}$ are unisolvent, and the noise is non zero, then by Part 1 of the proof, there exists a $Z^N$ such that $\phi(Z^N)$ is full rank so that $\phi(Z^N)(\theta_1 - \theta_2) \neq 0$. Therefore the implication does not hold which is a contradiction. Conclude that $\phi(Z^N_{\text{part}})$ must be full rank. ■

*Theorem 2:* Consider an LPV-ARX model structure with a parameterization in terms of the set of functions $\mathcal{F} = \{f_1, \ldots, f_{n_\alpha}\}$ and $\mathcal{G} = \{g_1, \ldots, g_{n_{beta}}\}$. Let $u \in \mathcal{U}$ with the number of initial conditions $n_i = n_b$ (see (10)). Let $p$ be periodic with period $T_p$, $\ell$ unique values per period, and $m$ periods. Let $m \ge n_a + n_b + 1$. Let $\mathcal{F}$ and $\mathcal{G}$ be unisolvent on the interval $[\min(p), \max(p)]$. Then, generically (due to Remark 5) the data set $Z^N$ is informative if and only if

(a) $n_n \ge n_b + 1$,
(b) $\ell \ge \max\{n_\alpha + 1, n_\beta + 1\}$.

with the exception of the special case that if $n_u = n_b + 1$, then $u$ cannot contain a sinusoid of the same period as $p$.

*Proof:* The proof is the same as the proof of Theorem 1 except for a few small changes. Compared to the proof of sufficiency of Part 1 of Theorem 1 the elements of $U_i$ and $Y_i$ in this case are different. Since $p$ is periodic with period $T_p$, with $m$ periods in the data set, and $\ell$ unique values per period, $p_1, \ldots, p_\ell$ are defined as:

$$\begin{aligned} p_1 &= p(0) = \cdots = p((m-1)T_p) \\ &\vdots \\ p_\ell &= p(\ell-1) = \cdots = p(\ell - 1 + (m-1)T_p) \end{aligned}$$

where $p_i \neq p_j, i \neq j$. Using this notation, $\phi(Z^N_{\text{part}})$ can be factored in the form (12), where $G$, $F$, $H$, $W$ and $P$ are defined the same as in Theorem 1, but

$$U_i^T = \begin{bmatrix} u(i) & u(i+T_p) & \cdots & u(i+mT_p) \\ \vdots & \vdots & & \vdots \\ u(i-n_b) & u(i+T_p-n_b) & \cdots & u(i+mT_p-n_b) \end{bmatrix}$$

$$Y_i^T = \begin{bmatrix} y(i-1) & y(i+T_p-1) & \cdots & y(i+mT_p-1) \\ \vdots & \vdots & & \vdots \\ y(i-n_a) & y(i+T_p-n_a) & \cdots & y(i+mT_p-n_a) \end{bmatrix}$$

TABLE I

DATA SETS

| Data Set | Input | Period of $p$ |
|---|---|---|
| $Z_1^N$ | $u(k) = 0.9^k + (-0.8)^k$ | [1 2 3] |
| | $u(-1) = 0.9^{-1} - 0.8^{-1}$ | $p(-1) = 3$ |
| $Z_2^N$ | $u(k) = 0.9^k + (-0.8)^k$ | [1 2 2] |
| | $u(-1) = 0.9^{-1} - 0.8^{-1}$ | $p(-1) = 2$ |
| $Z_3^N$ | $u(k) = 0.9^k$ | [1 2 3] |
| | $u(-1) = 0.9^{-1}$ | $p(-1) = 3$ |
| $Z_4^N$ | $u(k) = \sin(2\pi k/6)$ | [1 2 3] |
| | $u(-1) = \sin(-\pi/3)$ | $p(-1) = 3$ |
| $Z_{\text{val}}^N$ | $u(k)$ = white noise | 0.5[1 2 3 4 5 6] |
| | $u(-1)$ = 0 | $p(-1) = 3$ |

Each matrix $U_i$ can be factored the same way as in (13) but with $\xi_k$ replaced by $\xi_k^{T_p}$ and $B_i = \text{diag}(\zeta_1 \xi_1^i, \ldots, \zeta_{n_u} \xi_{n_u}^i)$. The Vandermonde matrices will be full rank as long as $\xi_i \neq \xi_j$, $i \neq j$, with one exception. If $\xi_k = e^{i\pi/T_p}$ and $\xi_\ell = e^{-i\pi/T_p}$, then $\xi_k^{T_p} = \xi_\ell^{T_p} = 1$. This corresponds to $u$ having a sinusoidal component with the same period as the scheduling parameter. Therefore if this is the case, $n_u$ must be equal to at least $n_b + 2$ to ensure that $U_i$ has rank $n_b + 1$.

The rest of the proof is analogous to Theorem 1. ■

*Theorem 3:* The LPV-ARX model structure is globally identifiable at any $\theta \in D_{\mathcal{M}}$, if and only if the non-linear functions $\{f_j\}$, $j = 1, \ldots, n_\alpha$ are linearly independent and the functions $\{g_j\}$, $j = 1, \ldots, n_\beta$ are linearly independent.

*Proof:* By Definition 7 it must be shown that for any $\theta_2 \in D_{\mathcal{M}}$, $\theta_2 \neq \theta_1$, there exists a data set such that the predicted outputs are different. Using the same notation as Theorem 1, the difference of two predicted outputs can be written as a matrix expression, $\phi(Z^N)(\theta_1 - \theta_2)$. Therefore, the LPV-ARX model structure is identifiable iff there exists a data set $\tilde{Z}^N$ such that $\phi(\tilde{Z}^N)$ is full rank.

(Necessary). If $g_i$ can be written as a linear combination of the other nonlinear functions $g_j, j \neq i$, then the rows of $\phi_u^T$ (see notation in Theorem 1) that are functions of $g_i$ can written as a linear combination of the other nonlinear functions $g_j, j \neq i$, and so the matrix $\phi_u^T$ will have less than full row rank for all $\tilde{Z}^N$. The same argument holds for $\phi_y$.

(Sufficient). Since the functions $\{f_1, \ldots, f_{n_\alpha}\}$ and $\{g_1, \ldots, g_{n_\beta}\}$ are linearly independent, there always exists a scheduling variable trajectory such that the sets $\mathcal{F}$ and $\mathcal{G}$ are unisolvent. Then by Theorem 1 a data set $\tilde{Z}^N$ exists such that $\phi$ is full rank. ■

*Example 1:* Choose an LPV-ARX model structure with $n_a = n_b = 1$ and $n_\alpha = n_\beta = 2$, and $f_j(x) = g_j(x) = x^j$. Let $N = 90$, with SNR of 45dB. Four different data sets (as shown in Table I) were used to estimate the actual parameter vector by minimizing (6). Only $Z_1^N$ is informative.

The actual and estimated parameter vectors are tabulated in Table II. Noise free simulations of the estimated systems using a validation data set $Z_{\text{val}}^N$ are plotted in Fig. 2. From Table II and Fig. 2 it can be seen that the models estimated using non-informative data are quite inaccurate, whereas the prediction of the output $y$ using the model estimated via $Z_1^N$ is barely distinguishable from the true system. □

TABLE II

ESTIMATED PARAMETER VECTORS

| True Parameter Vector | |
|---|---|
| [0.20 0.40 0.30 0.50 0.60 0.60 0.20 −0.60 0.20] | |
| **Data Set** | **Estimated Parameter Vector** |
| $Z_1^N$ | [0.20 0.41 0.29 0.47 0.62 0.59 0.21 −0.61 0.20] |
| $Z_2^N$ | [0.46 0.00 0.43 0.89 0.00 0.80 −0.20 0.00 0.00] |
| $Z_3^N$ | [0.00 0.00 0.00 0.72 0.90 0.89 0.21 −0.62 0.20] |
| $Z_4^N$ | [0.00 0.00 0.46 0.61 0.00 0.76 −0.64 0.71 −0.16] |

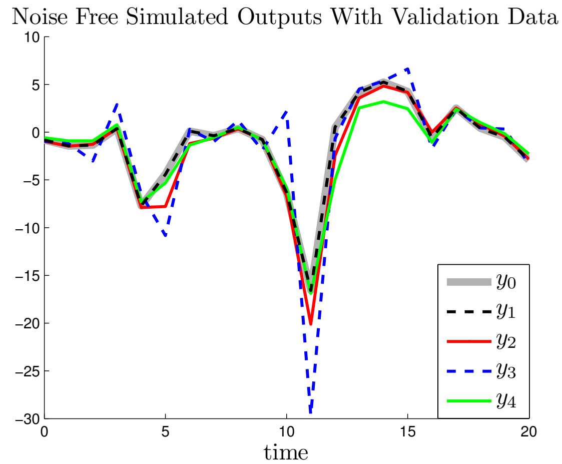


Fig. 2. Plots of the noise-free outputs of the 4 estimated systems

## VI. CONCLUSION

The notions of informative data and identifiability were investigated for LPV models. Specific conditions were derived to ensure informative data and to ensure identifiability for the LPV-ARX model structure. It was assumed that $a_i$ and $b_i$ depended on instantaneous values of $p$, however the theorems can be extended to include dynamic dependencies. The definitions presented give a framework within which it is possible to investigate informative data and identifiability for LPV-ARMAX, LPV-OE, and LPV-BJ model structures.

## REFERENCES


[1] R. Tóth, *Modeling and Identification of Linear Parameter-Varying Systems*, ser. Lecture Notes in Control And Information Sciences. Heidelberg: Springer-Verlag, 2010.

[2] L. Ljung, *System Identification. Theory for the User*, 2nd ed. Prentice Hall, 1999.

[3] M. Gevers, A. Bazanella, X. Bombois, and L. Mišković, "Identification and the information matrix: How to get just sufficiently rich?" *IEEE Transactions on Automatic Control*, vol. 54, pp. 2828–2840, 2009.

[4] M. Gevers, A. Brazanella, and X. Bombois, "Connecting informative experiments, information matrix and the minima of a prediction error identification criterion," in *Proceedings of 15th IFAC symposium on System Identification (SYSID2009)*, 2009.

[5] X. Wei and L. Del Re, "On persistent excitation for parameter estimation of quasi-LPV systems and its application in modelling of diesel engine torque," in *14th IFAC Symposium on System Identification*, Newcastle, Australia, 2006, pp. 517–522.

[6] B. Bamieh and L. Giarré, "Identification of linear parameter varying models," in *Proceedings of the 38th Conference on Decision and Control*, Phoenix, Arizona, 1999, pp. 1505–1510.

[7] G. Strang, *Linear Algebra and its Applications*, 4th ed. Thomson Brooks/Cole, 2006.

[8] P. J. Davis, *Interpolation and Approximation*. New York: Dover Publications Inc., 1975.

[9] A. Dankers, "Using laguerre filters for system modelling and identification," Master's thesis, University of Calgary, Feb. 2010.