2015 American Control Conference
Palmer House Hilton
July 1-3, 2015. Chicago, IL, USA

CVA Identification of Nonlinear Systems with LPV State-Space Models of Affine Dependence

Wallace E. Larimore*, Pepijn B. Cox† and Roland Tóth†

*Adaptics, Inc, 1717 Briar Ridge Road, McLean, VA 22101 USA
Phone: 703 532-0062, Fax: 703 536-3319, Email: larimore@adaptics.com
†Department of Electrical Engineering, Eindhoven University of Technology,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: {p.b.cox; r.toth}@tue.nl

***Abstract*— This paper discusses an improvement on the extension of linear subspace methods (originally developed in the Linear Time-Invariant (LTI) context) to the identification of Linear Parameter-Varying (LPV) and state-affine nonlinear system models. This includes the fitting of a special polynomial shifted form based LPV Autoregressive with eXogenous input (ARX) model to the observed input-output data. The estimated ARX model is used for filtering away the effects of future inputs on future outputs to obtain the so called "corrected future" analogous to the LTI case. The generality of the applied LPV-ARX parametrization now permits the estimation of the input-output map of a rather general class of LPV state-space models with matrices depending affinely on the scheduling. This is achieved by a canonical variate analysis (CVA) between the past and the corrected future which provides an estimate of a relevant set of state variables and their trajectories for the system, necessary for the construction of the minimal order state equations.**

I. Background and model structures

Over the past decade, considerable progress has been made in identification of parameterized dynamic systems using *linear parameter-varying* (LPV) and *state-affine* (SA) nonlinear models. Such methods, however, generally (i) involve model representations (*e.g.*, LPV *input-output* (IO) or nonlinear IO forms) that are relatively "easy" to identify, but require the use of difficult or computationally demanding realization/model reduction approaches or restrictive parameterizations to obtain state-space representation of the models, useful for the main stream of control synthesis methods [1], (ii) they rely on iterative nonlinear parameter optimization that may have computational or convergence difficulties [2], or (iii) involve subspace identification methods with over-restrictive approximation of the signal behaviors and/or computational requirements that grow exponentially in the number of states, inputs, outputs, and scheduling variables used (see *e.g.*, [3]–[5]). This can result in heavy computational requirements, lack of convergence of computations, and/or poor resulting accuracy in the computed solution corresponding to an uncertain overall performance of the whole identification task.

This work was partly supported by the Netherlands Organization for Scientific Research (NWO, grant. no.: 639.021.127).

In this paper, recent results [6] are reviewed in extending *linear time-invariant* (LTI) *subspace identification methods* (SIM) to the estimation of discrete-time LPV models of nonlinear systems. SIMs are well developed and understood in the context of LTI systems with optimal properties for particular cases [7] and conjectured to hold in the general LTI case [8]. However it is somewhat surprising that linear SIM can be extended and successfully used to capture LPV and SA nonlinear models that are often fundamentally bilinear.

A. LPV state-space models with affine dependence

Consider a discrete-time LPV system whose signal relations, *i.e.*, the corresponding IO map between the input signals $u : \mathbb{Z} \to \mathbb{R}^{n_\mathrm{u}}$ and output signals $y : \mathbb{Z} \to \mathbb{R}^{n_\mathrm{y}}$, can be described by a *state-space* (SS) representation

$$\left[\begin{array}{c|c} \mathcal{A}(\rho) & \mathcal{B}(\rho) \\ \hline \mathcal{C}(\rho) & \mathcal{D}(\rho) \end{array}\right] \begin{cases} x_{t+1} = \mathcal{A}(\rho_t)x_t + \mathcal{B}(\rho_t)u_t + w_t \\ \quad y_t = \mathcal{C}(\rho_t)x_t + \mathcal{D}(\rho_t)u_t \ + v_t \end{cases} \tag{1}$$

where $x_t : \mathbb{Z} \to \mathbb{R}^{n_\mathrm{x}}$ is the state vector and the matrices are bounded time-varying functions (also called *parameter-varying* (PV) functions) of a vector of *scheduling variables* $\rho_t : \mathbb{Z} \to \mathbb{R}^s$ with $\rho_t = [\ \rho_t^{(1)} = 1 \quad \rho_t^{(2)} \quad \dots \quad \rho_t^{(s)}\ ]^\top$ and s being finite and w_t and v_t are zero-mean colored noise processes independent of u_t and ρ_t, where the latter are assumed to be known or measured with no error in real time.

In this paper, as usual in much of the literature (*e.g.*, [3]–[5]), only LPV-SS models are considered which have *affine* dependence on the scheduling variables of the form $\mathcal{A}(\rho_t) = \rho_t^{(1)}A_1 + \cdots + \rho_t^{(s)}A_s$ and similarly for $\mathcal{B}(\rho_t)$, $\mathcal{C}(\rho_t)$, and $\mathcal{D}(\rho_t)$. Here, the parameter-varying matrices, like $\mathcal{A}(\rho_t)$, are expressed as a linear combination of constant matrices $A = [A_1 \ \cdots \ A_s]$ with time-varying weighting $\rho_t^{(i)}$. Other classes of dependencies are assumed to be captured already in ρ, *i.e.*, ρ can depend nonlinearly on other measurable variables that describe the operating point of the system.

The case of particular interest in system identification is the input-output LPV-ARX (autoregressive with external inputs) shifted form. This form is known to result in a constant parameter matrix $\mathcal{C}$ for the corresponding LPV state-space representation (1) and static dependence of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{D}$ on the respective scheduling variables (see Section I-C).

In much of the literature, LPV models are often restricted to the *strict-LPV* case where the scheduling functions ρ_t are *not* functions of the system inputs u_t, outputs y_t, and/or states x_t, *i.e.*, are truly exogenous variables. Note that, in the strict-LPV case, (1) is a linear system with the time-variation parameterized by ρ_t. The more general case including ρ_t as functions of u_t, y_t, and x_t is often called the *quasi-LPV* case. In previous papers [6], [9], [10], results were obtained for the strict-LPV case. In this paper, the *effect of the quasi-LPV case w.r.t. polynomial nonlinear systems* is considered, i.e., what happens if such nonlinear systems are identified using a CVA approach with an LPV model structure. It will be shown that the same results hold as for the strict-LPV case so that analogous but much more general methods are available for identifying polynomial nonlinear systems using linear model structures based subspace identification methods.

In general, the LPV state-space equations (1) can be considerably simplified by introducing the Kronecker product variables $\rho_t \otimes x_t$ and $\rho_t \otimes u_t$ in the form

$$\begin{bmatrix} x_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \rho_t \otimes x_t \\ \rho_t \otimes u_t \end{bmatrix} + \begin{bmatrix} w_t \\ v_t \end{bmatrix}, \qquad (2)$$

where $\otimes$ denotes the Kronecker product and (A, B, C, D) are the row collection of the sub-matrices, *e.g.*, $A = [A_1 \ \cdots \ A_s]$. Furthermore, introduce the notation $[M; N] = \begin{bmatrix} M^\top & N^\top \end{bmatrix}^\top$ which corresponds to stacking the vectors or matrices M and N. In later subsections, the full exploitation of this structure will result in an LTI subspace like formulation of data-driven LPV state-space model realization.

B. The LPV-ARX equivalent of the IO map

The first step in a *canonical variate analysis* (CVA) based procedure to estimate the random process (2) is to characterize its IO map and estimate it by a high order LPV-ARX model. As a generalization of the results of [1], in [11], it has been shown that any system with LPV-SS representation (2) with affine dependence also admits an IO representation in the form of

$$\alpha_0(\bar{\rho}_t) y_t = \sum_{i=1}^{\ell} \alpha_i(\bar{\rho}_t) y_{t-i} + \sum_{i=0}^{\ell} \beta_i(\bar{\rho}_t) u_{t-i} + e_t, \qquad (3)$$

where $\ell = n_\mathrm{x}$, $\bar{\rho}_t = [\ \rho_t^\top \ \ \ldots \ \ \rho_{t-n_\mathrm{x}}^\top \ ]^\top$, each α_i and β_i is a matrix of multivariate polynomials of degree at most ℓ and e_t is a quasi-stationary zero mean noise process corresponding to the filtered sum of w_t and v_t by a corresponding LPV filter. Furthermore, if the equivalence class of (2) under constant state transformation $\breve{x} = Tx$ with T nonsingular, has a companion-observability canonical form, which in the SISO case is

$$\left[\begin{array}{ccccc|c} -\alpha_1(\rho_t) & 1 & 0 & \ldots & 0 & \gamma_1(\rho_t) \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ -\alpha_{\ell-1}(p) & 0 & \ldots & 0 & 1 & \gamma_{\ell-1}(\rho_t) \\ -\alpha_\ell(\rho_t) & 0 & \ldots & \ldots & 0 & \gamma_\ell(\rho_t) \\ \hline 1 & 0 & \ldots & \ldots & 0 & \beta_0(\rho_t) \end{array}\right] \qquad (4)$$

then each α_i and β_i reduces to an affine function in ρ_{t-i} and $\alpha_0 \equiv 1$, *i.e.*, (4) is equivalent with

$$y_t = \sum_{i=1}^{\ell} \alpha_i(\rho_{t-i}) y_{t-i} + \sum_{i=0}^{\ell} \beta_i(\rho_{t-i}) u_{t-i} + e_t \qquad (5)$$

where $\beta_i(\rho_{t-i}) = \gamma_i(\rho_{t-i}) + \alpha_i(\rho_{t-i})\beta_0(\rho_{t-i})$. If representation (3) is seen as two polynomials in the time-shift operator q^{-i} (with scheduling dependent matrices in ρ_t and its finite many time-shifted versions) acting on y_t and u_t, then according to the realization theory provided in [12], it is possible to left multiply this form with such polynomials, *i.e.*, increasing ℓ, till e_t becomes (almost) a white noise process. In fact, if $\ell \to \infty$, then it is possible to choose the corresponding sequence of coefficient function such that e_t is white. This means that for appropriately large choice of ℓ, it is possible to approximate (2) with (3) such that e_t is a white noise process with covariance matrix Σ_{ee}. This corresponds to the generalization of the *high-order* ARX (HO-ARX) approximation of LTI systems under general noise conditions [13] and resembles to *nonlinear* (N)ARX model structures applied in nonlinear system identification.

C. State-space realization of the shifted IO form

As it was discussed, the LPV-IO model structure (5) with a shifted form of dependency guarantees the existence of an equivalent state-space model with no dynamic dependence (*e.g.*, (4)), meaning that the IO maps of both models are exactly the same. Note that this holds true in the MIMO case as well. This, together with the HO-ARX principle (*i.e.*, for high enough order, the ARX structure is capable to capture a larger generality of noise scenarios) allows the fitting of an IO model using linear regression methods up to some usually high order with small residual modeling error as measured by *Akaike's information criterion* (AIC). Then, the existence of an exactly corresponding SS model (2), although of possible high state order, is guaranteed.

Assume that each coefficient function α_i and β_i is an affine matrix function with coefficients $\{a_{i,j}\}_{j=1}^s$ and $\{b_{i,j}\}_{j=1}^s$ respectively which can be collected as $a_i = [\ a_{i,1} \ \ \cdots \ \ a_{i,s}\ ]$. According to (4) in the SISO case, the state variables, with the simplification that $\beta_0 \equiv 0$, are related as

$$q x_t^{(1)} = x_t^{(2)} - a_1(\rho_t \otimes x_t^{(1)}) + b_1(\rho_t \otimes u_t), \qquad (6a)$$

$$\vdots$$

$$q x_t^{(\ell-1)} = x_t^{(\ell)} - a_{\ell-1}(\rho_t \otimes x_t^{(\ell-1)}) + b_{\ell-1}(\rho_t \otimes u_t), \qquad (6b)$$

$$q x_t^{(\ell)} = -a_\ell(\rho_t \otimes x_t^{(\ell)}) + b_\ell(\rho_t \otimes u_t), \qquad (6c)$$

$$y_t = x_t^{(1)}, \qquad (6d)$$

where q is the *forward time-shift operator*, *i.e.*, $qu_t = u_{t+1}$. Based on [14], the above equations are a special case of

$$y = f_1(q^1 y, q^1 u) + \cdots + f_\ell(q^\ell y, q^\ell u), \qquad (7)$$

where f_i are analytic functions for $i = 1, \ldots, \ell$ known as *additive nonlinear auto regressive with exogenous inputs*

(ANARX) models. Such systems are always realizable in the classical state space form:

$$qx_t^{(1)} = x_t^{(2)} + f_1(x_t^{(1)}, u_t), \tag{8a}$$

$$\vdots$$

$$qx_t^{(\ell-1)} = x_t^{(\ell)} + f_{\ell-1}(x_t^{(1)}, u_t), \tag{8b}$$

$$qx_t^{(\ell)} = f_\ell(x_t^{(1)}, u_t), \tag{8c}$$

$$y_t = x_t^{(1)}. \tag{8d}$$

ANARX models have been widely used in modeling and control applications including IO linearization of nonlinear systems by dynamic output feedback [15]. If the nonlinear vector field $f(\cdot)$, expressing (8) as $qx_t = f(x_t, u_t)$, is continuously differentiable and if the origin is an equilibrium (i.e., $f(0) = 0$), then f can always be factorized as

$$f(y_t, u_t) = \mathcal{A}(\rho_t)y_t + \mathcal{B}(\rho_t)u_t, \tag{9}$$

where $\rho_t = \mu(y_t, u_t)$ with $\mu : \mathbb{R}^{(n_\mathrm{u}+n_\mathrm{y})} \to \mathbb{P}^s$ is an affine function. Hence, (6) becomes a special case of (8). If $f(0) \neq 0$, then by appropriate state and input transformation this property can be achieved. The price to be paid in the latter case is that it is generally only possible to eliminate x_t in f by allowing dynamic dependence of ρ_t on u_t and y_t (see [16]). It is also important to note, that instead of an affine mapping characterized by μ, we can also absorb the nonlinearities of $\mathcal{A}$ and $\mathcal{B}$ into ρ_t by allowing μ to be a continuous function and hence transforming $\mathcal{A}$ and $\mathcal{B}$ to an affine mapping resulting in an LPV-SS representation with affine dependence. In fact, this means that by assuming affine dependence of (1) in LPV system identification, we assume that, up to an affine combination, all nonlinearities in the factorization (9) are exactly captured in μ.

II. CVA SUBSPACE ID OF AFFINE LPV-SS MODELS

The most important question in the identification of (1) is how the state trajectory corresponding to the recorded data set can be efficiently estimated. For that purpose, the CVA method described in the sequel is applied while the exact selection of the model order is performed via the AIC measure of model fit. As we will see, many aspects of the CVA subspace identification scheme for LTI systems directly extend to strict-LPV systems (see [6], [9]) under the assumption of an affine dependence on the scheduling and that the scheduling variable is a free independent variable in the system. This is mainly due to the same linear relationship of the correlation structure, which, based on the above stated assumption, is only needed to be augmented with the extra ρ signal. However, in practice, ρ depends in many cases on u and y and hence, in this quasi-LPV case, the underlying dynamical structure of the representation is potentially highly nonlinear - a fundamental difference - which is apparent in comparing the LPV-SS model with affine dependence (2) and the state-affine NL-SS model (8). So the interesting question is that by not neglecting that ρ does depend on y and u can we still use the CVA to identify an LPV-SS model of the system and if not what remedies can we have?

To answer this question, let us first discuss the steps of the CVA scheme for LPV-SS identification [8]:

Algorithm 1 CVA scheme for LPV-SS identification

1: Estimate an LPV-ARX model (5) of sufficiently high order via least-squares and AIC based order selection (See Sections III-A and III-B).
2: Compute the *corrected future* (13) by removing the effects of future u on future y using the estimated LPV-ARX model (5).
3: Perform CVA (see Section III-D) between the past data and the corrected future of y for optimal candidate state estimates based on various state orders.
4: Estimate the SS model parameters by least-squares using (2).
5: Compute the AIC for each state order for optimal order selection.
6: **return** the realization of the LPV-SS model corresponding to the estimated state evolution and optimal state order.

A. The corrected future

Let us discuss first the intended procedure in the LTI case. Consider a state-space description of an LTI process. A k^{th}-order linear Markov process has been shown in [17] to have a representation in the following general state space form

$$x_{t+1} = Ax_t + Bu_t + w_t \tag{10a}$$

$$y_t = Cx_t + Du_t + Fw_t + v_t \tag{10b}$$

where x_t is a k^{th}-order *Markov state* and w_t and v_t are white noise processes that are independent with covariance matrices Σ_{ww} and Σ_{vv} respectively. These state equations are more general than typically used since the noise $Fw_t + v_t$ in (10b) is correlated with w_t in (10a). This is a consequence of requiring that the state in (10a) is a k^{th}-order Markov state. Requiring w_t and v_t to be uncorrelated may result in a SS model where the state is higher dimensional than the Markov order k resulting in a non-minimal realization.

The focus in this paper is on the restricted identification task of modeling the open-loop dynamic behavior from u_t to y_t. Assume that u_t can have arbitrary autocorrelation and possibly involve feedback from y_t. The discussion below summarizes the procedure described in detail in [8] for removing effects of such possible autocorrelation.

The ℓ-length future $\mathcal{F}_t(y) = [\ y_{t+\ell}^\top\ \dots\ y_t^\top\ ]^\top$ of the process is related to the ℓ-length past $\mathcal{P}_t(y) = [\ y_{t-1}^\top\ \dots\ y_{t-\ell}^\top\ ]^\top$ and $\mathcal{P}_t(u) = [\ u_{t-1}^\top\ \dots\ u_{t-\ell}^\top\ ]^\top$ through the state x_t and the future inputs $\mathcal{F}_t(u) = [\ u_{t+\ell}^\top\ \dots\ u_t^\top\ ]^\top$ in the form

$$\mathcal{F}_t(y) = \Psi^\top x_t + \Omega^\top \mathcal{F}_t(u) + \mathcal{F}_t(e), \tag{11}$$

where x_t lies in some fixed subspace of $\mathcal{P}_t(y, u)$, $\Psi^\top = [CA^{\ell-1}; \dots; CA; C]$ and if $i = j$, then the (i, j)-th block of the upper-block-triangular Ω is D, else it is $CA^{j-i}B$. The presence of the future inputs $\mathcal{F}_t(u)$ creates a major problem in determining the state subspace from the observed past and future. If the term $\Omega^\top \mathcal{F}_t(u)$ could be removed from the above equation, then the state subspace could be estimated accurately. The approach used in the CVA method is to fit an ARX model and compute an estimate $\hat{\Psi}$ of Ψ based on the estimated ARX parameters. Note that an ARX process can be expressed in state-space form with state $x_t = \mathcal{P}_t(y, u)$ and hence it satisfies the state relation (11). Then, the state-space realization of the ARX model $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ and $\hat{\Psi}$ and $\hat{\Omega}$ are themselves functions of the ARX model parameters $\hat{\Theta}_A$.

Now, since the effect of the future inputs $\mathcal{F}_t(u)$ on future outputs $\mathcal{F}_t(y)$ can be accurately predicted by the ARX model with moderate sample size, the term $\Omega^\top \mathcal{F}_t(u)$ can thus be predicted (in a one to ℓ step-ahead sense) and subtracted from both sides of (11). Then a CVA can be done between the *corrected future* $\mathcal{F}_t(y) - \Omega^\top \mathcal{F}_t(u)$ and the past $\mathcal{P}_t(y, u)$ to determine the state x_t as discussed next. A variety of procedures to deal with autocorrelation and feedback in subspace system identification for LTI systems have been developed in [8], [18]–[22]. The statistical justification for the whole CVA procedure is discussed in detail in [8].

In case of an LPV-SS model (2) with affine dependence, note that the same procedure holds except in

$$\mathcal{F}_t(y) = \Psi^\top x_t + \Omega^\top \mathcal{F}_t(u) + \mathcal{F}_t(e), \tag{12}$$

it holds that

$$\Psi^\top = \left[\sum_{i=1}^{s} C_i \rho_{t+\ell}^{(i)} \prod_{j=1}^{\ell-1} \sum_{k=1}^{s} A_k \rho_{t+j}^{(k)}; \ldots; \sum_{i=1}^{s} C_i \rho_t^{(i)} \right]$$

and if $i = j$, then the (i, j)-th block of the upper-block-triangular Ω is $\sum_{k=1}^{s} D_k \rho_{t+\ell-i+1}^{(k)}$, else it is

$$\sum_{k=1}^{s} C_k \rho_{t+\ell-i+1}^{(k)} \prod_{l=i}^{j-2} \left(\sum_{m=1}^{s} A_m \rho_{t+\ell-l}^{(m)} \right) \sum_{k=1}^{s} B_k \rho_{t+\ell-j}^{(k)}.$$

B. Correlation of the past and the corrected future

The procedure we intend to apply in this section for the CVA of the past and the corrected future is the same as in the strict-LPV case, which has been proven to result in a maximum likelihood estimation of the state-trajectory in [6]. Our contribution here is to extend this result to the quasi-LPV case.

Definition 1 (Corrected future): Let the LPV-ARX process of order ℓ be given by (5). Denote by $\tilde{y}$ *future outputs due to future inputs*, *i.e.*, the simulated response of (5) for u on the time interval $[t, \ell]$ starting with zero initial conditions:

$$\tilde{y}_{t+j} = \sum_{i=1}^{j} a_i (\rho_{t+j-i} \otimes \tilde{y}_{t+j-i}) + \sum_{i=0}^{j} b_i (\rho_{t+j-i} \otimes u_{t+j-i}). \tag{13}$$

Denote the *corrected future outputs* as $\mathcal{F}_t(\bar{y})$ computed as

$$\mathcal{F}_t(\bar{y}) = \mathcal{F}_t(y) - \mathcal{F}_t(\tilde{y}). \tag{14}$$

□

The terms in the above definition are justified in the following Theorem.

Theorem 1 (Correlation structure, quasi-LPV case): Consider the LPV-ARX process of order ℓ be given by (5) where ρ_t is not independent of u_t and y_t (quasi-LPV process). For every $t \in [\ell + 1, N - \ell]$, the corrected future outputs $\mathcal{F}_t(\bar{y}) = [\ \bar{y}_{t+\ell}^\top \ \ldots \ \bar{y}_t^\top\ ]^\top$, as in Def. 1, are linear (time-invariant) functions of the corrected augmented future $\mathcal{F}_t(\rho \otimes \bar{y})$ and the augmented past $\mathcal{P}_t(\rho \otimes y)$ and $\mathcal{P}_t(\rho \otimes u)$. This relation can be expressed recursively as

$$\begin{aligned}
&\underbrace{\begin{bmatrix} \bar{y}_{t+\ell}^\top & \bar{y}_{t+\ell-1}^\top & \cdots & \bar{y}_{t+1}^\top & \bar{y}_t^\top \end{bmatrix}^\top}_{\mathcal{F}_t(\bar{y})} = \\
&\qquad \begin{bmatrix} e_{t+\ell}^\top & e_{t+\ell-1}^\top & \cdots & e_{t+1}^\top & e_t^\top \end{bmatrix}^\top + \\
&\underbrace{\begin{bmatrix} 0 & a_1 & \cdots & a_{\ell-1} & a_\ell \\ & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & a_1 & a_2 \\ & & & \ddots & a_1 \\ 0 & & \cdots & & 0 \end{bmatrix}}_{a_{\mathcal{F}}} \underbrace{\begin{bmatrix} \rho_{t+\ell} \otimes \bar{y}_{t+\ell} \\ \rho_{t+\ell-1} \otimes \bar{y}_{t+\ell-1} \\ \vdots \\ \rho_{t+1} \otimes \bar{y}_{t+1} \\ \rho_t \otimes \bar{y}_t \end{bmatrix}}_{\mathcal{F}_t(\rho \otimes \bar{y})} \\
&+ \underbrace{\begin{bmatrix} 0 & \cdots & & 0 \\ a_\ell & \ddots & & \\ \vdots & \ddots & \ddots & \vdots \\ a_2 & \cdots & a_\ell & 0 \\ a_1 & \cdots & a_{\ell-1} & a_\ell \end{bmatrix}}_{a_{\mathcal{P}}} \underbrace{\begin{bmatrix} \rho_{t-1} \otimes y_{t-1} \\ \rho_{t-2} \otimes y_{t-2} \\ \vdots \\ \rho_{t-\ell+1} \otimes y_{t-\ell+1} \\ \rho_{t-\ell} \otimes y_{t-\ell} \end{bmatrix}}_{\mathcal{P}_t(\rho \otimes y)} \\
&+ \underbrace{\begin{bmatrix} 0 & \cdots & & 0 \\ b_\ell & \ddots & & \\ \vdots & \ddots & \ddots & \vdots \\ b_2 & \cdots & b_\ell & 0 \\ b_1 & \cdots & b_{\ell-1} & b_\ell \end{bmatrix}}_{b_{\mathcal{P}}} \underbrace{\begin{bmatrix} \rho_{t-1} \otimes u_{t-1} \\ \rho_{t-2} \otimes u_{t-2} \\ \vdots \\ \rho_{t-\ell+1} \otimes u_{t-\ell+1} \\ \rho_{t-\ell} \otimes u_{t-\ell} \end{bmatrix}}_{\mathcal{P}_t(\rho \otimes u)}
\end{aligned} \tag{15}$$

For a proof see Appendix A. □

Equation (15) can be rewritten as

$$(\bar{I} - a_{\mathcal{F}}) \mathcal{F}_t(\rho \otimes \bar{y}) = \begin{bmatrix} a_{\mathcal{P}} & b_{\mathcal{P}} \end{bmatrix} \begin{bmatrix} \mathcal{P}_t(\rho \otimes y) \\ \mathcal{P}_t(\rho \otimes u) \end{bmatrix} + \mathcal{F}_t(e), \tag{16}$$

where $\bar{I}$ is the flattened identity matrix according to the left hand side of (15). This follows since the first component of the scheduling vector ρ_t is the constant 1 so each of the terms $\bar{y}_{t+i}$ for $i = 1, \ldots, \ell$ composing $\mathcal{F}_t(\bar{y})$ are included as a sub-vector in the corrected augmented future vector $\mathcal{F}_t(\rho \otimes \bar{y})$. Furthermore, if the order of the estimated ARX model is high enough, *i.e.*, ℓ is taken appropriately large, then e_t can be considered as a white noise process.

Note that in (16), $(\bar{I} - a_{\mathcal{F}})$ and $[\ a_{\mathcal{P}} \ \ b_{\mathcal{P}}\ ]$ are constant matrices, hence the information from the past is projected by a linear time-invariant map to the augmented corrected future and this mapping is explicitly dependent on the scheduling ρ_t. Furthermore, in this mapping, $\mathcal{P}_t(\rho \otimes y) = \mathcal{P}_t(\mu(y, u) \otimes y)$ and $\mathcal{P}_t(\rho \otimes u) = \mathcal{P}_t(\mu(y, u) \otimes u)$ are not correlated with $\mathcal{F}_t(e)$ if ℓ is chosen appropriately large. Hence, a linear time-invariant CVA can be performed between $\mathcal{F}_t(\rho \otimes \bar{y})$ and $[\ \mathcal{P}_t^\top(\rho \otimes y) \ \ \mathcal{P}_t^\top(\rho \otimes u)\ ]^\top$. This justifies the use of a CVA to synthesize a state vector for the LPV-ARX process that can be used for state order selection and estimation of the constant matrices (A, B, C, D) in (2).

Note, that each step of the above described procedure uses an orthogonal projection, where the errors in parameters removed from earlier models are orthogonal to the resulting projected model. This is a fundamental statistical concept applied in sequence to the projection from the LPV-ARX model into the CVA model and then into the state-space model. As long as asymptotically in terms of the sample size, each model is estimated using a maximum likelihood procedure, the overall algorithm is guaranteed to converge to a maximum likelihood procedure [23].

III. Computational aspects and properties

There are a number of sophisticated computational, statistical, and numerical methods that have been developed for CVA methods over the years and are well-suited to the estimation of LPV-SS models with affine dependence. These are briefly mentioned here along with references where they are described in detail. The ADAPTx algorithm for subspace identification of LTI systems, and later, LPV and nonlinear systems, is described in the last subsection as a sequence of nested models that are successively projected from one into the next. The key to refining the model estimate with each projection is that residual error of the earlier models is orthogonal to the resulting projected model.

The overall strategy is to first identify the terms of an LPV-ARX model using linear regression that only involves solving systems of linear equations to obtain coefficients of these models. This is followed by a realization step to construct an LPV state-space model of the system dynamics from which, by re-substitution of the scheduling variable relation, a nonlinear model results. Note that this model is guaranteed to describe the captured signal relations, but not necessarily qualify as a nonlinear state-space model [24].

A. Order-recursive fitting of LPV-ARX models

Estimated models with various orders and scheduling maps often have partial nesting in that some contain others as special cases where some coefficients are set to zero. When the additional terms are additive, as in LPV-ARX and state-affine models, there can be considerable efficiency in starting with a simple parametrization based model structure (low ARX order and simple scheduling map) and computing the change in the AIC model fit using more complex models with additional parameters. For LPV-ARX models, the computational load to estimate a single model is proportional to n^3 where $n = (sn_{\mathrm{y}}^2 + sn_{\mathrm{y}}n_{\mathrm{u}})^{\ell+1}$ is the number of parameters, whereas in updating such a model from n_1 parameters to n_2 parameters is proportional to $(n_2 - n_1)^2$. So fitting a single model with n_2 parameters requires the same order of computation as fitting all orders from 1 to n_2. These computations are accurate to double precision using SVD based CVA methods (see [25], [26]).

B. Use of AIC for comparison of multiple models

Fitting of LPV-ARX models involves a multitude of models with increasing numbers of terms (lag order and monomial degree). The problem of structure selection requires the solution of a statistical multiple comparison problem among the various fitted models. A very general approach to solving this problem using the AIC [27] is developed in [28]. In the ADAPTx software, this is applied to the choice of ARX order and monomial degree, and state order in state-space model construction.

C. Calculation of the corrected future

In dealing with feedback in dynamic systems, a way to avoid problems of bias in parameter estimation is to "remove the effects of future inputs on future outputs" to obtain the "corrected future". This strategy [8] of the ADAPTx algorithm is shown theoretically to be asymptotically equivalent in large samples to two other methods in [19].

D. Canonical variate analysis using a generalized SVD

The CVA is computed using a *generalized* (G)SVD where the variables are orthogonal with respect to the covariance matrix of the observations, while other subspace methods use a "CVA weighting". The latter choice does not provide the optimal statistical selection of the model state order provided by using the generalized SVD and the AIC. For illustration, the main idea is to calculate

$$M = \mathbb{E}\left\{ \begin{bmatrix} \mathcal{P}_t(\rho \otimes y) \\ \mathcal{P}_t(\rho \otimes u) \end{bmatrix}^\top \begin{bmatrix} \mathcal{P}_t(\rho \otimes y) \\ \mathcal{P}_t(\rho \otimes u) \end{bmatrix} \right\}$$

$$N = \mathbb{E}\left\{ \mathcal{F}_t^\top(\rho \otimes \bar{y}) \mathcal{F}_t(\rho \otimes \bar{y}) \right\}$$

$$X = (\bar{I} - a_{\mathcal{F}})^{-1} \begin{bmatrix} a_{\mathcal{P}} & b_{\mathcal{P}} \end{bmatrix}$$

and then find U and V such that $I = U^\top M U$ and $I = V^\top N V$ and $X = U \Sigma V^\top$ with Σ diagonal, containing the squared canonical correlations, to arrive at the state estimate

$$\hat{x}_t = [I \quad 0] U^\top \begin{bmatrix} \mathcal{P}_t(\rho \otimes y) \\ \mathcal{P}_t(\rho \otimes u) \end{bmatrix}. \tag{17}$$

Then, compute Φ as the auto-correlation of $[\hat{x}_{t+1}; y_t]$ and Ψ the cross-correlation of $[\rho_t \otimes \hat{x}_t; \rho_t \otimes u_t]$ with $[\hat{x}_{t+1}; y_t]$ to obtain the estimate of the state-space matrices as

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} = \Phi \Psi^\dagger, \tag{18}$$

where $\Psi^\dagger$ denotes the right pseudo inverse of Ψ. Detailed derivation with GSVD is given in [29], with some additional accuracy issues discussed in [30].

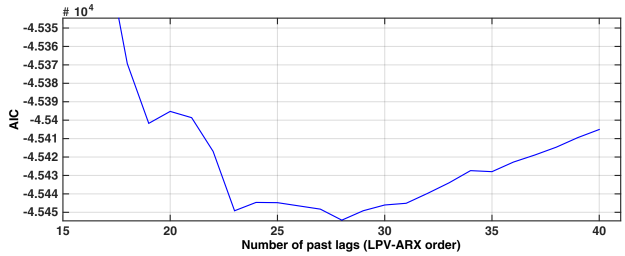
IV. Example: Forced Lorenz Attractor

To demonstrate the effectiveness of the proposed procedure, identification of a nonlinear Lorenz attractor is considered. The Lorenz attractor can be expressed as a parameter-varying system of the form of (1) with sub-matrices

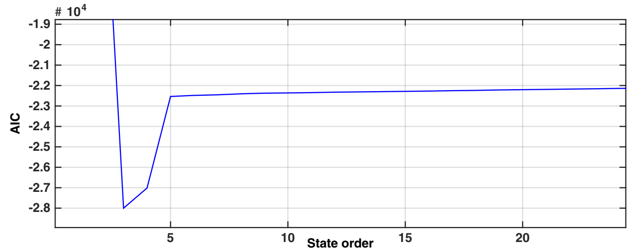
$$\left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right] = \left[\begin{array}{ccc|c} 1-\sigma T & \sigma T & 0 & T \\ \gamma T & 1-T & 0 & 0 \\ 0 & 0 & 1-\beta T & 0 \\ \hline 1 & 0 & 0 & 0 \end{array}\right],$$

$$\left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right] = \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right],$$

scheduling map $\rho_t = \begin{bmatrix} 1 & y_t \end{bmatrix}^\top$, sampling time $T > 0$, and parameters (σ, γ, β) (where γ is usually denoted with ρ



(a)



(b)

Fig. 1: Results of AIC based order selection: (a) order of the LPV-ARX model; (b) order of the resulting state-space model by CVA.

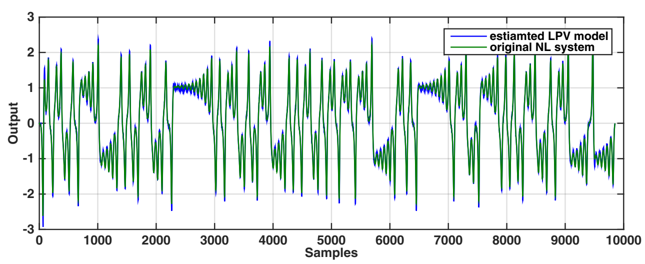


Fig. 2: 5-step ahead prediction of the estimated 4^{th}-order LPV-SS model compared to the output of the data generating system using noise free validation data.

that is used in this paper for the scheduling parameter). Note that since one of the scheduling variables is y_t, this example corresponds to a quasi-LPV model. If u_t is set to zero for all t, then the system is "self-exciting," *i.e.*, it has sustained behavior that does not damp out. Also note that the term $\rho_t \otimes x_t$ involves the product $y_t x_t$ that is a bilinear term.

To obtain an estimation data set, simulation of the Lorenz attractor with parameter values of $\sigma = 10$, $\gamma = 28$, $\beta = 8/3$ and with independent white noise u_t w_t and v_t having standard deviations 10^{-1}, 10^{-6}, and 10^{-2}, respectively, is performed. As a next step, recursive estimation of LPV-ARX models with scheduling map $\rho_t = \begin{bmatrix} 1 & y_t \end{bmatrix}^\top$ is accomplished which has resulted in a plot of AIC verses LPV-ARX order given in Figure 1a. The minimum AIC occurs at an ARX order of 28. An LPV-ARX model of this order was used in calculating the corrected future. A CVA between the augmented past and corrected augmented future in terms of Theorem 1 produced squared canonical correlations with values of 0.999999, 0.99999, 0.999, 0.999, 0.95, 0.94, 0.57 for the first seven elements. The CVA provided state estimates of various orders with the AIC of model fit is plotted as a function of state order in Figure 1b. The first 4 states appear to provide the best model fit. They were used to estimate an LPV-SS model (2) using (18) to compute $[\hat{A}\ \hat{B}; \hat{C}\ \hat{D}]$. Comparison of the trajectories of the observed verses the identified models are shown in Figure 2 using validation data.

There is a strong qualitative resemblance between these two models. The presence of noise in the simulation of the true process perturbs the trajectory, so a precise comparison is difficult to make since in such nonlinear systems small differences result in divergence between the trajectories. Remarkably, however, the identified state affine model is qualitatively very similar and follows the major transitions of the observed output data.

V. Summary and Conclusions

In this paper, the validity of the previously developed CVA subspace method for the identification of LPV models has been investigated in the quasi-LPV scenario where the scheduling variable is not assumed to be independent of the system inputs and outputs. As a central result, it is proven that still a linear and time invariant correlation structure between the past and the corrected future holds under the assumption that the scheduling variable accurately captures the underlying nonlinearities of the system modulo an affine combination. From this structure, it is shown that the state of the system can be determined by a linear CVA between the past and corrected future. The procedure has been successfully demonstrated on the identification of the Lorenz attractor in a quasi-LPV state-space form.

Appendix A. Proof of Theorem 1

First, in (5), replace t by $t+j$, and then consider t as the present time t dividing the past and the present-future for recursive computation of future outputs y_{t+j} with j considered as the number of steps ahead of the present time t with $j = 0, 1, \ldots, \ell$.

Second, for each j, the computation of terms in (3) are partitioned into present and future terms (with sums from $i = 0$ or 1 to j as in (19b)) and into past terms (with sums from $i = j+1$ to ℓ as in (19c))

$$y_{t+j} = e_{t+j} + \tag{19a}$$

$$+\sum_{i=1}^{j} a_i(\rho_{t+j-i} \otimes y_{t+j-i}) + \sum_{i=0}^{j} b_i(\rho_{t+j-i} \otimes u_{t+j-i}) \tag{19b}$$

$$+\sum_{i=j+1}^{\ell} a_i(\rho_{t+j-i} \otimes y_{t+j-i}) + \sum_{i=j+1}^{\ell} b_i(\rho_{t+j-i} \otimes u_{t+j-i}) \tag{19c}$$

Now consider a fixed j steps ahead in predicting y_{t+j}. It is apparent that the second term of (19b) has contributions only from present-future inputs and all of (19c) has contributions only from past inputs and outputs. The first term of (19b) has contributions from both the past and present-future since the future outputs y_{t+j-i} defined for $i = 1$ to j are previously and recursively defined by (19b) and (19c). So the strategy is to separately split out the past and present-future contributions for each recursion in the computation of y_{t+j-i} in the first term of (19b).

Lemma 1 (Past and Present-Future Effects on Outputs): Let time t split the past and present-future, and let y_{t+j} be a present-future output with $0 \leq j \leq \ell$. Then, the contribution of present-future inputs to y_{t+j} is given by $\tilde{y}_{t+j}$ as in (13), and the contribution of past inputs and outputs to y_{t+j} is given by $\bar{y}_{t+j}$ as in (14).

Proof: Proceeding by induction, (i) first Lemma 1 is demonstrated to be true for $j = 0$, and then (ii) second it is shown for any choice of j^* satisfying $0 < j^* \leq \ell$ that if Lemma 1 is true for all j satisfying $0 \leq j < j^*$, then Lemma 1 is also true for $j = j^*$.

It is only necessary to keep track of the contribution from the present-future since doing this necessarily determines the correct contribution from the past as $y_{t+j-i} = \tilde{y}_{t+j-i} + \bar{y}_{t+j-i}$ with y_{t+j-i} a particular observed output. To show (i), the contribution from the present-future given by both (13) and (19b) agree since the first term of each are zero and the second terms are identical. To show (ii), assume that the splits $\tilde{y}_{t+j}$ and $\bar{y}_{t+j}$ are correct for $0 \leq j < j^*$, then it is required to show that this is also true for $j = j^*$. If the substitution $y_{t+j-i} = \tilde{y}_{t+j-i} + \bar{y}_{t+j-i}$ is made in the first term in (19b), then this produces a sum of two terms, respectively associated with the present-future and the past. Then associating $\tilde{y}_{t+j-i}$ with the present-future gives

$$\tilde{y}_{t+j} = \sum_{i=1}^{j} a_i(\rho_{t+j-i} \otimes \tilde{y}_{t+j-i}) + \sum_{i=0}^{j} b_i(\rho_{t+j-i} \otimes u_{t+j-i}) \tag{20}$$

while associating the $\bar{y}_{t+j-i}$ term with past via (19c) gives the expression below for computing $\bar{y}_{t+j}$ in (15).

$$\bar{y}_{t+j} = e_{t+j} + \sum_{i=1}^{j} a_i(\rho_{t+j-i} \otimes \bar{y}_{t+j-i}) \tag{21a}$$

$$+ \sum_{i=j+1}^{\ell} \alpha_i(\rho_{t+j-i} \otimes y_{t+j-i}) + \sum_{i=j+1}^{\ell} b_i(\rho_{t+j-i} \otimes u_{t+j-i}) \tag{21b}$$

The expression for $\tilde{y}_{t+j}$ agrees with (13), which proves Lemma 1, and (21a) and (21b) is precisely the recursive form of the matrix equation (15) that proves Theorem 1. ■

References

[1] R. Tóth, H. Abbas, and W. Werner, "On the state-space realization of LPV input-output models: Practical approaches," *IEEE Trans. on Control Systems Technology*, vol. 20, pp. 139–153, 2012.

[2] V. Verdult, L. Ljung, and M. Verhaegen, "Identification of composite local linear state-space models using a projected gradient search," *Int. Journal of Control*, vol. 75, no. 16-17, pp. 1125–1153, 2002.

[3] J. W. van Wingerden and M. Verhaegen, "Subspace identification of bilinear and LPV systems for open- and closed-loop data," *Automatica*, vol. 45, no. 2, pp. 372–381, 2009.

[4] V. Verdult and M. Verhaegen, "Subspace identification of multivariable linear parameter-varying systems," *Automatica*, vol. 38, no. 5, pp. 805–814, 2002.

[5] P. L. dos Santos, J. A. Ramos, and J. L. M. de Carvalho, "Identification of LPV systems using successive approximations," in *Proc. of the 47th IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 4509–4515.

[6] W. E. Larimore, "Identification of nonlinear parameter-varying systems via canonical variate analysis," in *Proc. of the American Control Conf.*, Washington, DC, USA, June 2013, pp. 2250–2265.

[7] D. Bauer, "Comparing the CCA subspace method to pseudo maximum likelihood methods in the case of no exogenous inputs," *Journal Time Series Analysis*, vol. 26, no. 5, pp. 631–668, 2005.

[8] W. Larimore, "Large sample efficiency for ADAPTx subspace system identification with unknown feedback," in *Proc. of the IFAC Dynamics and Control of Process Systems Conference*, Cambridge, MA, USA, July 2004, pp. 293–298.

[9] W. E. Larimore and M. Buchholz, "ADAPT-LPV software for identification of Nonlinear Parameter-Varying systems," in *Proc. of the 16th IFAC Symposium on System Identification*, Brussels, Belgium, July 2012, pp. 1820–1825.

[10] M. Buchholz and W. E. Larimore, "Subspace identification of an aircraft linear parameter-varying flutter model," in *Proc. of the American Control Conference*, Washington, DC, USA, June 2013, pp. 2263–2267.

[11] M. Petreczky and G. Mercere, "Affine LPV systems: realization theory, input-output equations and relationship with linear switched systems,," in *Proc. of the 51th IEEE Conf. on Decision and Control*, Maui, Hawaii, USA, Dec. 2012, pp. 4511–4516.

[12] R. Tóth, *Modeling and Identification of Linear Parameter-Varying Systems*, ser. Lecture Notes in Control and Information Sciences, Vol. 403. Heidelberg: Springer, 2010.

[13] Y. C. Zhu, "Black-box identification of MIMO transfer functions: Asymptotic properties of prediction error models," *Int. J. Adap. Contr. and Signal Proc.*, vol. 3, pp. 357–373, 1989.

[14] Ü. Kotta and N. Sadegh, "Two approaches for state space realization of NARMA models: Bridging the gap," *Mathematical and Computer Modeling of Dynamical Systems*, vol. 8, no. 1, pp. 21–32, 2002.

[15] A. Kaldmae and Ü. Kotta, "Input-output linearization by dynamic output feedback," in *Proc. of the European Control Conf.*, Zurich, Switzerland, July 2013, pp. 1728–1733.

[16] H. Abbas, R. Tóth, M. Petreczky, N. Meskin, and J. Mohammadpour, "Embedding of nonlinear systems in a linear parameter-varying representation," in *Proc. of the 19th IFAC World Congress*, Cape Town, South Africa, Aug. 2014, pp. 6907–6913.

[17] A. Lindquist and M. Pavon, "Markovian representation of discrete-time stationary stochastic vector processes," in *Proc. of the 20th IEEE Conference on Decision and Control*, 1981, pp. 1345–1356.

[18] M. Jansson, "Subspace identification and ARX modelling," in *Proc. of the the 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, 2003.

[19] A. Chiuso, "On the asymptotic properties of closed-loop CCA-type subspace algorithms: Equivalence results and role of the future horizon,," *IEEE Trans. on Automatic Control*, vol. 55, no. 3, pp. 634–649, 2010.

[20] R. Shi and J. F. MacGregor, "A framework for subspace identification methods," in *Proc. of the American Cont. Conf.*, 2001, pp. 3678–3683.

[21] S. J. Qin, "An overview of subspace identification," *Computers and Chemical Engineering*, vol. 30, no. 10-12, pp. 1502–1513, 2006.

[22] S. J. Qin and L. Ljung, "On the role of future horizon in closed-loop subspace identification," in *Proc. of the 14th IFAC Symposium on System Identification*, 2006.

[23] D. Cox and D. Hinkley, *Theoretical Statistics*. New York: Chapman and Hall, 1974.

[24] J. Belikov, Ü. Kotta, and M. Tõnso, "Comparison of LPV and nonlinear system theory: A realization problem," *Systems & Control Letters*, vol. 64, pp. 72–78, 2014.

[25] W. Larimore, "Order-recursive factorization of the pseudoinverse of a covariance matrix," *IEEE Trans. on Automatic Control*, vol. 35, pp. 1299–1303, 1990.

[26] ——, "Reply to "comment on order-recursive factorization of the pseudoinverse of a covariance matrix"," *IEEE Trans. on Automatic Control*, vol. 47, pp. 1953–1957, 2002.

[27] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. of the 2nd International Symposium on Information Theory*, Budapest, Hungary, 1973, pp. 267–281.

[28] W. Larimore, "Predictive inference, sufficiency, entropy, and an asymptotic likelihood principle," *Biometrika*, vol. 70, pp. 175–181, 1983.

[29] ——, "System identification, reduced-order filtering and modeling via canonical variate analysis," in *Proc. of the American Control Conference*, San Francisco, CA, USA, 1983, pp. 175–181.

[30] W. Larimore and H. Javaherian, "Identification of linear parameter-varying engine models," in *Proc. of the American Control Conf.*, Washington, DC, USA, June 2013, pp. 2250–2265.