51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Order and Structural Dependence Selection of LPV-ARX Models Revisited

Roland Tóth, Håkan Hjalmarsson and Cristian R. Rojas

***Abstract*— Accurate parametric identification of Linear Parameter-Varying (LPV) systems requires an optimal prior selection of model order and a set of functional dependencies for the parameterization of the model coefficients. In order to address this problem for linear regression models, a regressor shrinkage method, the Non-Negative Garrote (NNG) approach, has been proposed recently. This approach achieves statistically efficient order and structural coefficient dependence selection using only measured data of the system. However, particular drawbacks of the NNG are that it is not applicable for large-scale over-parameterized problems due to computational limitations and that adequate performance of the estimator requires a relatively large data set compared to the size of the parameterization used in the model. To overcome these limitations, a recently introduced L_1 sparse estimator approach, the so-called SPARSEVA method, is extended to the LPV case and its performance is compared to the NNG.**

***Index Terms*— Linear parameter-varying systems; ARX model; identification, order selection; sparse estimators; compressive system identification.**

I. Introduction

Modeling and identification of nonlinear systems by *Linear Parameter-Varying* (LPV) models has recently received extensive attention in the literature. This has resulted in significant developments in the estimation of LPV polynomial models in the *Input-Output* (IO) setting both in discrete and continuous time [1]–[7]. A general mark of LPV models is that the signal relations are considered to be linear just as in the *Linear Time-Invariant* (LTI) case, but the model parameters are functions of a measurable time-varying signal $p : \mathbb{Z} \to \mathbb{P} \subseteq \mathbb{R}^{n_\mathrm{p}}$ (in discrete time), called the *scheduling variable*. In principle, using scheduling variables as changing operating conditions, endogenous/free signals or even latent variables of the plant, the LPV system class can describe both nonlinear and time-varying phenomena accurately.

To successfully find adequate LPV models of a given system via identification, prior knowledge about the original system is of paramount importance. One of the most important priors is the selected model structure and the corresponding model set within which the identification method should find an LPV model estimate of the plant. Such a selection is rather complicated as it is outmost desired to estimate an accurate LPV model of the underlaying system using as few parameters as possible.

Besides the classical questions of model structure selection in terms of model order (or relative orders in the *multiple-input multiple-output* (MIMO) case), input delay and noise structure, in the LPV setting a rather important, but relatively rarely investigated problem is the selection of the functional dependencies which explain how the model coefficients can change due to the variation of the scheduling variable p. To be able to describe the dynamic behavior of a given system by an LPV model, this relation of the coefficient functions to p must be precisely captured. However, in practice, prior information on such a relation is hardly available and in most cases only a class of functions, like polynomials, trigonometrical functions, etc., can be pointed out. To be able to capture these unknown relations, the lack of reliable prior knowledge often leads to the situation of choosing a large set of dependencies, like monomials $\{\psi_{i,j}\}_{j=1}^{s_i}$ with $\psi_{i,j} : \mathbb{P} \to \mathbb{R}$, for the parametrization of the model coefficients in the form

$$a_i(p(k)) = \theta_{i,0} + \theta_{i,1}\psi_{i,1}(p(k)) + \ldots + \theta_{i,1}\psi_{i,s_i}(p(k)), \quad (1)$$

with $\theta_{i,j} \in \mathbb{R}$ being the unknown parameters. Often from this large set of functions only a few are needed to capture the behavior of the system adequately. Furthermore, as quality of a model is clearly related to the performance of the application in which the model will be used, there is a desire for a minimal parametrization based on the parsimony principle (Occam's razor) and utilization complexity in terms of control synthesis, prediction, etc.

The resulting model structure selection problem can be resolved by proposing a model structure capable of explaining a rich set of possible dynamics, and, by imposing a sparsity pattern on the parameters, letting the data decide which substructure is appropriate to use. In the LTI literature, typical model structure selection tools such as AIC, BIC/MDL, cross-validation, etc. are used for this purpose. However, recently computationally more attractive sparse estimation or statistical regularization (shrinkage) methods have been developed like the *Non-Negative Garrote* (NNG) or the *Least Absolute Shrinkage and Selection Operator* (LASSO) [8]–[10], the *Ridge Regression*, the *Elastic Net* methods [11] and *L_1-estimators* like the SPARSEVA [12], [13].

The NNG approach has been extended to LPV-ARX polynomial models in [14] to tackle the order and dependency selection problem in the LPV global identification setting (identification w.r.t. varying scheduling trajectory) based on

R. Tóth is with the Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands, email: R.Toth@tue.nl.

H. Hjalmarsson and C. R. Rojas are with ACCESS Linnaeus Center, Electrical Engineering, KTH–Royal Institute of Technology, S-100 44 Stockholm, Sweden, emails: hjalmars@kth.se and crro@kth.se. This work was supported by the European Research Council under the advanced grant LEARN, contract 267381.

convex optimization. In this method, a natural ordering of model complexity is imposed on the parameters, which provides the possibility to estimate the order of the input and output side polynomials simultaneously together with of the required structural dependence of the coefficients. However, a particular drawback of the NNG is that it requires a relatively large data set compared to the size of the parameter vector to be estimated (sensitivity for the underdetermined regression case). As in the LPV setting, over-parametrization might result in an overly large number of parameters, the resulting estimation problem can fall close to the underdetermined case which might rise difficulties in the application of the NNG. Furthermore, the NNG requires the optimization of a regularization parameter which has a non-trivial relationship between the expected prediction error and sparsity level of the estimate.

In this paper, we study the application of the recently developed L_1 sparse estimator approach called SPARSEVA (which stands for *SPARSe Estimator based on a VAlidation criterion*) for the joint model order and dependency selection problem of discrete-time LPV-ARX models, and compare the resulting approach to the LPV-NNG. The motivation of this study is that the SPARSEVA approach has been recently reported to have favorable properties in the LTI setting in solving underdetermined sparse linear-regression problems [13] and requires no regularization parameter selection [12]. Therefore, it provides a competitive alternative of the NNG and it is an exciting question to explore if these advantages can be fruitfully used in handling the challenging problems of LPV model structure selection.

For the sake of generality, it must also be noted that an alternative of over-parametrization based model structure selection in the LPV context is the use of "non-parametric" estimation approaches like [4], [15]. These methods have been recently developed to learn model structures directly without prior assumptions on the dependencies. However, in this paper, we focus our comparison only on approaches based on the over-parametrization principle.

The paper is organized as follows: In Section II, a short review of the LPV-ARX model structure and its linear-regression based identification method is given, defining the problem setting for model structure selection in this context. Section III gives an introduction to the SPARSEVA approach and presents how its modified form can be used to solve the LPV-ARX model structure selection problem. In Section IV, the NNG approach is briefly summarized and it is compared to the developed LPV-SPARSEVA method in Section V. In Section VI, performance of both algorithms are compared on simulated data. Finally, in Section VII, conclusions are drawn and perspectives on future work are given.

II. LPV IDENTIFICATION VIA ARX MODELS

For the simplicity of the notation, assume that the original data-generating system is *single-input single-output* (SISO) and, in discrete-time, it can be written in the form of

$$y(k)+\sum_{i=1}^{n_\mathrm{a}} a_i^\mathrm{o}(p(k))y(k-i)=\sum_{j=0}^{n_\mathrm{b}} b_j^\mathrm{o}(p(k))u(k-j)+e(k) \quad (2)$$

where $k \in \mathbb{Z}$ is the time, $u : \mathbb{Z} \to \mathbb{R}$ and $y : \mathbb{Z} \to \mathbb{R}$ denote the input and the output signals respectively, $p : \mathbb{Z} \to \mathbb{P}$ is the *scheduling variable* with range $\mathbb{P} \subseteq \mathbb{R}^{n_\mathrm{p}}$ and e is a white Gaussian noise process with zero mean. Furthermore, the coefficient functions $a_i^\mathrm{o}, b_j^\mathrm{o} : \mathbb{P} \to \mathbb{R}$ have *static dependence* on p, i.e., they only depend on $p(k)$. Based on measurements $\mathcal{D}_N = \{(u(k), p(k), y(k))\}_{k=1}^N$, our goal is to estimate a model of this system in the form of

$$y(k)+\sum_{i=1}^{n_\mathrm{a}} a_i(p(k))y(k-i)=\sum_{j=0}^{n_\mathrm{b}} b_j(p(k))u(k-j)+\epsilon_\theta(k) \quad (3)$$

where the functions

$$\begin{bmatrix} \phi_1 & \cdots & \phi_{n_\mathrm{g}} \end{bmatrix}^\top \triangleq \begin{bmatrix} a_1 & \cdots & a_{n_\mathrm{a}} & b_0 & \cdots & b_{n_\mathrm{b}} \end{bmatrix}^\top,$$

with $n_\mathrm{g} = n_\mathrm{a} + n_\mathrm{b} + 1$ are considered – as it is almost exclusively done in the LPV identification literature (see, e.g., [1]–[3], [5], [16]–[18]) – to be parameterized as

$$\phi_i(\centerdot) = \theta_{i0} + \sum_{j=1}^{s_i} \theta_{i,j}\psi_{i,j}(\centerdot), \quad (4)$$

where $\{\theta_{i,j}\}_{i=1,j=1}^{n_\mathrm{g},s_i}$ are unknown parameters and $\{\psi_{i,j}\}_{i=1,j=1}^{n_\mathrm{g},s_i}$ are functions chosen by the user. In this case, (3) can be rewritten as

$$y(k) = \varphi^\top(k)\theta + \epsilon_\theta(k), \quad (5)$$

where

$$\begin{aligned} \theta &= \begin{bmatrix} \theta_{1,0} & \cdots & \theta_{1,s_1} & \theta_{2,0} & \cdots & \theta_{n_\mathrm{g},s_{n_\mathrm{g}}} \end{bmatrix}^\top \\ \varphi(k) &= \begin{bmatrix} -y(k-1) & -\psi_{1,1}(p(k))y(k-1) & \cdots \\ & -\psi_{1,s_1}(p(k))y(k-1) & -y(k-2) & \cdots \\ & -\psi_{n_\mathrm{a},s_{n_\mathrm{a}}}(p(k))y(k-n_\mathrm{a}) & u(k) & \cdots \end{bmatrix}^\top. \end{aligned}$$

It is assumed that model structure (3) contains the true system (2), i.e., there exists a θ_o, s.t. $y(k) = \varphi^\top(k)\theta_\mathrm{o}+e(k)$.

Given a data set $\mathcal{D}_N$, the *least-squares* (LS) parameter estimate for the linear regression model (5) is

$$\hat{\theta}_N = \arg\min_{\theta\in\mathbb{R}^n} V_N(\theta, \mathcal{D}_N), \quad (6)$$

where $n = \sum_{i=1}^{n_\mathrm{g}}(1 + s_i)$ (according to (4)), and

$$V_N(\theta, \mathcal{D}_N) \triangleq \frac{1}{N}\|\epsilon_\theta(k)\|_{\ell_2}^2. \quad (7)$$

To guarantee a unique solution of (6), it is assumed that $\{\psi_{ij,}\}_{i=1,j=1}^{n_\mathrm{g},s_i}$ are chosen such that (3) is globally identifiable (there exist no θ and θ', such that the 1-step ahead predictor resulting from (3) is not distinguishable for θ and θ') and that $\mathcal{D}_N$ provides a *persistently exciting* regressor in (5) (see [5], [19], [20]). Note that identifiability in particular holds for (5) with $e \neq 0$ iff for each $i \in \{1, \ldots, n_\mathrm{g}\}$, $\{\psi_{i,j}\}_{j=1}^{s_i}$ corresponds to a set of linearly independent functions on $\mathbb{P}$. By organizing the data as

$$Y_N = \begin{bmatrix} y(1) & y(2) & \ldots & y(N) \end{bmatrix}^\top, \quad (8a)$$

$$\Phi_N = \begin{bmatrix} \varphi(1) & \varphi(2) & \ldots & \varphi(N) \end{bmatrix}^\top, \quad (8b)$$

the optimal solution to (6) can be written as

$$\hat{\theta}_N^\mathrm{LS} = \left(\Phi_N^\top\Phi_N\right)^{-1}\Phi_N^\top Y_N \triangleq \Phi_N^+ Y_N. \quad (9)$$

III. The LPV-SPARSEVA approach

In this section, an extension of the SPARSEVA approach (developed in [12]) is introduced for the estimation of the model structure (3) by taking into account the sparsity in the parameter vector. The SPARSEVA estimator is a variant of the LASSO estimator [9] (an L_1 penalized least squares estimator), and w.r.t. the considered linear regression form (5), it corresponds to the minimizer of the convex program:

$$\min_{\theta\in\mathbb{R}^{n_\mathrm{g}}} \quad \|\theta\|_1, \tag{10a}$$

$$\text{s.t.} \quad V_N(\theta,\mathcal{D}_N) \leq V_N(\hat{\theta}_N^{\mathrm{LS}},\mathcal{D}_N)(1+\varepsilon_N). \tag{10b}$$

Here $\varepsilon_N > 0$ is a quantity defining a particular level set of V_N and can be typically chosen as:

- $\varepsilon_N = 2\left(\sum_{i=1}^{n_\mathrm{g}}(1+s_i)\right)/N$.
- $\varepsilon_N = \left(\sum_{i=1}^{n_\mathrm{g}}(1+s_i)\right)\ln(N)/N$.

Note that the $\hat{\theta}_N^{\mathrm{LS}}$ solution corresponds to the minimum of $V_N(\theta,\mathcal{D}_N)$. However, in case of an over-parameterization, which happens frequently in the LPV case, this minimum is achieved by using the extra degrees of freedom in θ to fit the noise in the data and hence decrease the ℓ_2-loss. Based on the sparsity prejudice, it is thus required to increase the ℓ_2-loss in terms of the ε_N defined level sets of $V_N(\theta,\mathcal{D}_N)$ to find the sparse solution we are looking for. The first choice of ε_N is motivated by the AIC criterion, while the second one is related to the BIC/MDL criterion [21]. In Section V, it is going to be shown how these choices of ε_N relate to consistency/sparsity of the SPARSEVA scheme. Note that the problem (10a-b) can be efficiently solved by any convex optimization algorithms even in case of underdetermined situations, i.e., $N \ll n_\mathrm{g}$ (under minor conditions on Φ_N, see [13]).

Even though SPARSEVA can be considered as a variant of the LASSO, it has the advantage of not requiring the tuning of regularization parameters via techniques such as *cross-validation*, which involve solving multiple times a convex program over a grid of values of the regularization parameters. This tuning is automatically taken into account by choosing the value of ε_N, as explained in detail in [12].

An "adaptive" version of (10a), called A-SPARSEVA, has better sparsity properties than SPARSEVA [12], and w.r.t. (5), it is defined as the minimizer of the following convex program:

$$\min_{\theta\in\mathbb{R}^{n_\mathrm{g}}} \quad \|w\odot\theta\|_1 \tag{11a}$$

$$\text{s.t.} \quad V_N(\theta,\mathcal{D}_N) \leq V_N(\hat{\theta}_N^{\mathrm{LS}},\mathcal{D}_N)(1+\varepsilon_N), \tag{11b}$$

where $w := \left[\ [\hat{\theta}_N^{\mathrm{LS}}]_1^{-1} \quad \cdots \quad [\hat{\theta}_N^{\mathrm{LS}}]_{n_\mathrm{a}+n_\mathrm{b}}^{-1}\ \right]^\top$ and $\odot$ is the Hadamard product. The better sparsity property of this scheme follows by the weighting w based re-orientation of the $V_N(\theta,\mathcal{D}_N)$ level set associated feasible θ set in the parameter space such that the relaxation gap between the θ solution with minimal support and the ℓ_1 solution is decreased. The estimation properties of A-SPARSEVA can be further improved by removing the columns of Φ_N corresponding to the zero entries of the A-SPARSEVA estimate $\hat{\theta}_N^{\mathrm{A}}$, and re-estimating θ by least squares on the reduced Φ_N. Note that in case the assumed model structure is not ARX, the prediction error is not equal to the equation error in (3). This means that minimizing the ℓ_2-loss w.r.t. the prediction will not lead to a convex problem in terms of (11a-b) and the introduced selection criterion for ε_N will be invalid due to the bias of the LS estimate.

IV. The LPV-NNG approach

The *Nonnegative Garrote* (NNG) method was first presented in [8] as a coefficient shrinkage method for linear regression models in statistics. This approach is based on a penalization of the least-squares solution of (9) by attaching weights to it, which in turn are regularized. The NNG problem w.r.t. (5) can be also written as a convex program

$$\min_{w\in\mathbb{R}^{n_\mathrm{g}}} \quad V_N(w\odot\hat{\theta}_N^{\mathrm{LS}},\mathcal{D}_N) + \lambda\|w\|_1, \tag{12a}$$

$$\text{s.t.} \quad w \succeq 0, \tag{12b}$$

where λ is the model complexity, i.e., regularization parameter. For a given λ, (12a-b) as a a convex optimization problem can be solved in the decision variable w, providing the final estimate as $\hat{\theta}_N^{\mathrm{NNG}} = w\odot\hat{\theta}_N^{\mathrm{LS}}$. As λ increases, the weights of the less important regressors will shrink, and finally end up at zero. Thus, as λ increases, the model becomes less complex.

By this approach, it is also possible to take into account the natural ordering of time lags by penalizing higher model order in the NNG estimate as reported in [14]. This is achieved by adding some linear constraints on the weights w. For LPV-ARX models, these constraints can be

$$1 \geq \sum_{j=0}^{s_1} w_{1,j} \geq \sum_{j=0}^{s_2} w_{2,j} \geq \ldots \geq \sum_{j=0}^{s_{n_\mathrm{a}}} w_{n_\mathrm{a},j}, \tag{12c}$$

$$1 \geq \sum_{j=0}^{s_{n_\mathrm{a}+1}} w_{(n_\mathrm{a}+1),j} \geq \ldots \geq \sum_{j=0}^{s_{n_\mathrm{g}}} w_{n_\mathrm{g},j}. \tag{12d}$$

This is a natural[1] extension of the NNG method, for order selection of LPV-ARX models in system identification.

An efficient way to implement this strategy is to use a path following parametric estimation. For this purpose a Lagrangian multipliers based method has been proposed in [22]. Starting from $\lambda = 0$, this method calculates a piecewise affine solution path for λ. In this way it efficiently explores the change in the model fit as a function of λ. For more details see [22]. Then, based on the resulting (finite) set of λ values where the solution changes, the final parameter estimate $\hat{\theta}_N^{\mathrm{NNG}}$ and its associated λ value is selected by cross validation with a BIC or AIC criterion.

V. Comparison of the SPARSEVA and the NGG

Under the assumptions of Section II, it is possible to show (see [12], [23]) that the considered A-SPARSEVA L_1 estimator (11a-b) enjoys the following properties:

i) $\hat{\theta}_N$ is consistent in probability iff $\varepsilon_N \to 0$ as $N \to \infty$.

[1] Note that other choices for the ordering of the parameters, e.g., the maximum instead of the sum, are also possible.

ii) Under condition (i), $\hat{\theta}_N$ has the *sparseness property* (i.e., $P\{[\hat{\theta}_N]_i = 0\} \to 1$ as $N \to \infty$ for every index i such that $[\theta_o]_i = 0$) if and only if $N\varepsilon_N \to \infty$.

iii) Under conditions (i) and (ii), $\hat{\theta}_N$ (with least squares re-estimation) has the *oracle property*. This means that

$$\sqrt{N}(\hat{\theta}_N - \theta_o) \in \text{As } \mathcal{N}(0, M^\dagger), \tag{13}$$

where M is the information matrix when the support of θ_o is known.

Notice that the A-SPARSEVA enjoys consistency, sparseness and the oracle property if we choose $\varepsilon_N = \left(\sum_{i=1}^{n_g}(1+s_i)\right)\ln(N)/N$, resembling the BIC criterion. For the non-asymptotic properties of the method, especially in case of relatively small data records, see [13].

The available results (see [10], [24]) on the NNG, under the assumptions of Section II, are more modest:

i) $\hat{\theta}_N$ is consistent in probability iff $\lambda \to 0$ as $N \to \infty$.

ii) if $\lambda \to 0$ in a fashion that $\|\theta_o - \hat{\theta}_N^{\text{LS}}\|_\infty = o(\lambda)$ as $N \to \infty$, then $\hat{\theta}_N$ has the *sparseness property*.

As the regularization parameter λ has a non-trivial relationship with the level sets of V_N, therefore it is generally not possible to give a simple choice of it, like in the A-SPARSEVA case, which can ensure both consistency and sparseness. Therefore, calculation of a piece-wise affine solution path for λ and cross validation based selection of the right λ value is a more reliable path to follow. However, application of such an algorithm renders asymptotic analysis to be impractical, even though decreasing tendency of the selected λ can be observed in practice. This clearly signifies the importance of the A-SPARSEVA scheme with the BIC type of selection of ε_N. Furthermore, depending on the size of the regression problem, the signal to noise ratio and the sparsity level of θ_o, the calculation of the piece-wise affine solution path has a varying computational time ranging from a few seconds to hours. For the sake of fairness, it is important to highlight that it might be also possible to give a selection rule which could bring to the NNG the same properties as the A-SPARSEVA enjoys, but such a selection rule is difficult to obtain due to the non-trivial relationship of λ with the level sets of V_N. In conclusion, these properties suggest that the A-SPARSEVA is better tuned to the large-scale linear-regression problems we face in case of LPV-ARX model selection problem. To test this hypothesis, a serious Monte-Carlo study is conducted in the next section.

VI. Simulations

In order to test the applicability of the proposed LPV-SPARSEVA method and to compare its performance to the NNG, a representative simulation example is considered. This example is taken from [14].

The data-generating system is defined as an LPV-ARX$(9,3)$ model:

$$A(q,p)y = B(q,p)u + e, \tag{14}$$

where the noise e is white with a Gaussian distribution $\mathcal{N}(0,\sigma_e^2)$, $p(k) \in \mathbb{P}$ with $\mathbb{P} = [-2\pi, 0]$ and

$$\begin{aligned} A(q,p) = 1 &+ (0.24 + 0.1p)q^{-1} + (0.6 - 0.1\sqrt{-p})q^{-2} \\ &+ 0.3\sin(p)q^{-3} + (0.17 + 0.1p)q^{-4} \\ &+ 0.3\cos(p)q^{-5} - 0.27q^{-6} + (0.01p)q^{-7} \\ &- 0.07q^{-8} + 0.01\cos(p)q^{-9}, \\ B(q,p) = 1 &+ (1.25 - p)q^{-1} - (0.2 + \sqrt{-p})q^{-2}, \end{aligned}$$

are polynomials in q with static coefficient dependence on p. This LPV-ARX$(9,3)$ model is stable for all constant trajectories of p (*uniform frozen stability*) and has fast and slow modes which change rapidly with the variation of p (see [14] for details). Note that this model is a particularly difficult one to correctly estimate as it is possible to approximate it with an LPV-ARX$(8,3)$ model without a significant loss of accuracy.

For estimation purposes, 100 estimation and 100 validation data records have been generated by system (14) for each data length $N \in \{200 + 50k\}_{k=1}^{37}$, resulting in 37×100 estimation and validation data records with length in the interval $[200, 2000]$. During each computation, u, p and e have been considered as independent realizations of three white noise sequences with distributions $u(k) \in \mathcal{N}(0,1)$, $p(k) \in \mathcal{U}(-2\pi, 0)$ and $e(k) \in \mathcal{N}(0,\sigma_e^2)$ respectively. To study the effect of a change in the power of the noise, this generation of the data sequences has been repeated for various noise variances $\sigma_e^2 \in \{6.25{\cdot}10^{-8}, 6.25{\cdot}10^{-2}, 6.5, 56.25\}$ corresponding to average *Signal to Noise Ratios*[2] (SNR's): 100dB, 40dB, 20dB, 12.5dB respectively. This has resulted in a total of $4 \times 37 \times 100 = 14800$ estimation and validation data sets defining a serious Monte-Carlo study under various conditions.

Using these data sets, the A-SPARSEVA approach described in Section III with LS re-optimization, the NNG approach described in Section IV with BIC based cross validation and the ordinary LS approach are used to estimate the system. In order to fairly assess the quality of the resulting model estimates a base-line estimator or so called `oracle` estimator in terms of an LS method has been applied with the priori knowledge of which elements of θ_o are zero. Note that the latter approach cannot be applied in practice as the optimal model structure is unknown (part of the identification problem itself). The results are compared in terms of

- The *Mean Squared Error* (MSE) of the prediction on the validation data:

$$\text{MSE} = \mathbb{E}\{\|y(k) - \hat{y}_{\hat{\theta}_N}(k)\|_2^2\}. \tag{15}$$

computed as an average over each 100 runs for a given N and σ_e^2.

- The average of the *fit score* or the *Best Fit Rate* (BFR) [25]:

$$\text{BFR} = 100\%{\cdot}\max\left(1 - \frac{\|y(k) - \hat{y}_{\hat{\theta}_N}(k)\|_2}{\|y(k) - \bar{y}\|_2}, 0\right), \tag{16}$$

[2]The SNR is defined as $\text{SNR} \triangleq 10 \cdot \log_{10}\left(\frac{\|y(k)-v(k)\|_2^2}{\|v(k)\|_2^2}\right)$ where $v(k) = -\sum_{i=1}^{n_a} a_i(p(k))v(k-i) + e(k)$.

where $\bar{y}$ is the mean of y and $\hat{y}_{\hat{\theta}_N}$ is the simulated model output based on the validation data.

- The L_1 parameter estimation error: $\|\hat{\theta}_N - \theta_o\|_1$.
- The percentage of correctly estimated zero elements.

By the previously described approaches, an LPV-ARX$(9, 3)$ model is estimated based on the collected data and using the coefficient parametrization (4) with

$$\psi_{i,1}(p) = p, \qquad \psi_{i,3}(p) = \sin(p),$$
$$\psi_{i2,}(p) = \sqrt{-p}, \qquad \psi_{i,4}(p) = \cos(p),$$

for all i, i.e. $s_1 = \ldots = s_{12} = 4$. This parametrization corresponds to $5 \cdot 12 = 60$ unknown θ_{ij}'s to be estimated. Note that (14) is in the model class and the model order is correct, but the coefficients are overparameterized, as only a subset of $\{\psi_{ij}\}$ is required for the estimation of each ϕ_i.

The average results of the 100 Monte-Carlo runs in each cases are given in Figure 1. From these results it follows that in the low noise cases (SNR$=100$dB, 40dB) the proposed LPV-A-SPARSEVA scheme correctly estimates the true support of θ_o, i.e., it correctly identifies the underlying model structure of the system and hence it achieves the same results as the `oracle` approach. For SNR$=100$dB, the NNG with BIC shows a worse performance in the MSE of the prediction and BFR of the simulation error than the ordinary LS, but it starts to recover quickly as N increases. This phenomenon, is due to some numerical errors in the calculation of the piece-wise affine solution path and has no real relation to the performance of the method. This is clearly shown by the fact that in average almost 100% of the zeros in θ_0 are correctly estimated by the NNG-BIC. For SNR$=40$dB, the NNG-BIC shows its true performance and clearly improves the estimation w.r.t. the LS approach. However, the proposed A-SPARSEVA scheme is able to explore better the sparsity structure of θ_0 and provides the correct selection of the regressor terms with a small error on the estimated number of zeros. For the sake of fairness, it is important to note that a significant amount of parameters identified by the NNG-BIC to be non-zero are actually estimated with a magnitude less than 10^{-4}. The real difference in the performance of the A-SPARSEVA and the NNG-BIC becomes apparent in the medium (SNR$=20$dB) and the high noise cases (SNR$=12.5$dB), where the A-SPARSEVA clearly out-performs the NNG-BIC. For increasing data lengths beyond $N = 2000$, the performance of all approaches start to converge to the `oracle` base-line estimator with the same relative rate which can be observed in Figure 1.

VII. Conclusions and future work

In this paper, the order and structural dependence selection problem of LPV-ARX models, addressed in [14], has been revisited. Extension of two sparse estimator approaches, the *Non-Negative Garrote* (NNG) also studied in [14] and an L_1 sparse estimator presented in [12] have been analyzed and compared in this context. It has been shown, that the L_1 sparse estimator called SPARSEVA, has favorable properties over the NNG approach as it does not require optimization of the regularization parameter, it is applicable to solve underdetermined sparse linear-regression problems and shows better performance in the recovery of the sparsity structure. These advantages are important as in the LPV case usually serious over-parameterization is needed to adequately capture the dynamics of the data-generating system. Hence, the LPV extension of the SPARSEVA provides a reliable approach to estimate the required order and dependency structure in LPV-ARX models, giving a practical tool for the support of LPV-IO identification approaches. The proposed method is extendable to the multivariable case and for more difficult noise settings, providing important objectives for further research.

References

[1] B. Bamieh and L. Giarré, "Identification of linear parameter varying models," *Int. Journal of Robust and Nonlinear Control*, vol. 12, pp. 841–853, 2002.

[2] X. Wei, "Advanced LPV techniques for diesel engines," Ph.D. dissertation, Johannes Kepler University, Linz, 2006.

[3] V. Laurain, M. Gilson, R. Tóth, and H. Garnier, "Refined instrumental variable methods for identification of LPV Box-Jenkins models," *Automatica*, vol. 46, no. 6, pp. 959–967, 2010.

[4] K. Hsu, T. L. Vincent, and K. Poolla, "Nonparametric methods for the identification of linear parameter varying systems," in *Proc. of the Int. Symposium on Computer-Aided Control System Design*, San Antonio, Texas, USA, Sept. 2008, pp. 846–851.

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

[6] S. K. Wattamwar, S. Weiland, and T. Backx, "Identification of low-order parameter-varying models for large-scale systems," *Journal of Process Control*, vol. 20, no. 2, pp. 158–172, 2010.

[7] X. Jin, B. Huang, and D. S. Shook, "Multiple model LPV approach to nonlinear process identification with EM algorithm," *Journal of Process Control*, vol. 21, no. 1, pp. 182–193, 2011.

[8] L. Breiman, "Better subset regression using the nonnegative garotte," *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.

[9] R. Tibshirani, "Regression shrinkage and selection with the Lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[10] M. Yuan and Y. Lin, "On the non-negative garrotte estimator," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 143–161, 2007.

[11] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[12] C. R. Rojas and H. Hjalmarsson, "SPARSEVA: Sparse estimation based on a validation criterion," in *Proc. of the 50th IEEE Conf. on Decision and Control*, Orlando, Florida, USA, Dec. 2011, pp. 2825–2830.

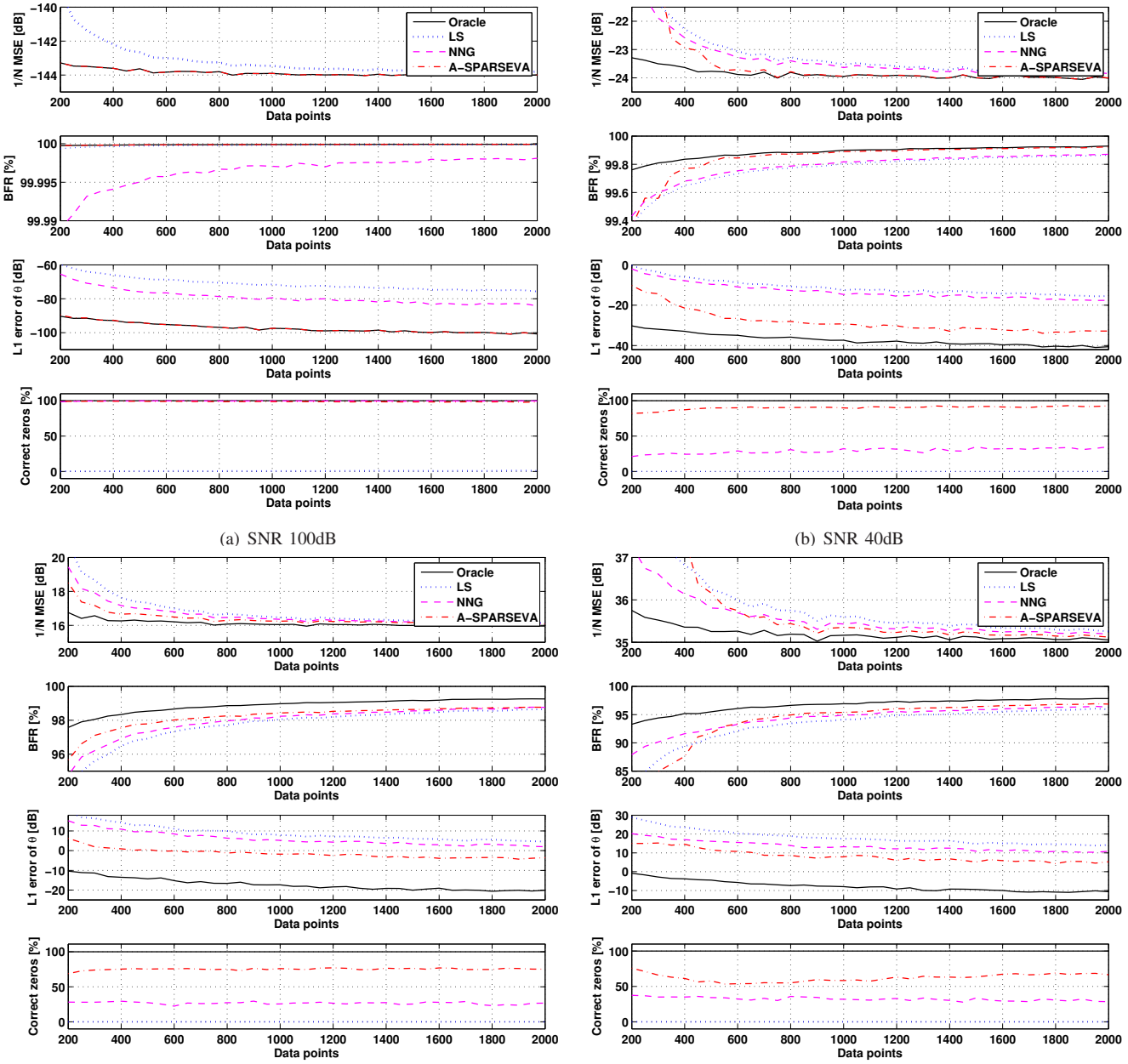
[13] R. Tóth, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Compressive system identification in the linear time-invariant framework," in *Proc. of the 50th IEEE Conf. on Decision and Control*, Orlando, Florida, USA, Dec. 2011, pp. 783–790.

[14] R. Tóth, C. Lyzell, M. Enqvist, P. S. C. Heuberger, and P. M. J. Van den Hof, "Order and structural dependence selection of LPV-ARX models using a nonnegative garrote approach," in *Proc. of the 48th IEEE Conf. on Decision and Control*, Shanghai, China, Dec. 2009, pp. 7406–7411.

[15] R. Tóth, V. Laurain, W. Zheng, and K. Poolla, "Model structure learning: A support vector machine approach for LPV linear-regression models," in *Proc. of the 50th IEEE Conf. on Decision and Control*, Orlando, Florida, USA, Dec. 2011, pp. 3192–3197.

[16] 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.

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



(a) SNR 100dB

(b) SNR 40dB

(c) SNR 20dB

(d) SNR 12.5dB

Fig. 1. Monte Carlo simulation results with various SNR and data length N.

[18] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof, "Asymptotically optimal orthonormal basis functions for LPV system identification," *Automatica*, vol. 45, no. 6, pp. 1359–1370, 2009.

[19] A. G. Dankers, R. Tóth, P. S. C. Heuberger, X. Bombois, and P. M. J. Van den Hof, "Identifiability and the informativity of data sets for LPV-ARX identification," in *Proc. of the 50th IEEE Conf. on Decision and Control*, Orlando, Florida, USA, Dec. 2011, pp. 799–804.

[20] M. Gevers, A. Bazanella, and L. Mišković, "Informative data: how to get just sufficiently rich?" in *Proc. of the 47th IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 4522–4527.

[21] L. Ljung, *System Identification, theory for the user*, 2nd ed. Prentice-Hall, 1999.

[22] C. Lyzell, J. Roll, and L. Ljung, "The use of nonnegative garrote for order selection of ARX models," in *Proc. of the 45th IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 1974–1979.

[23] R. Tóth, H. Hjalmarsson, and C. R. Rojas, "Sparse estimation of polynomial dynamical models," in *Proc. of the the 16th IFAC Symposium on System Identification*, Brussels, Belgium, 2012, pp. 983–988.

[24] L. Mohammadi and S. van De Geer, "On nonnegative garrote estimator in a linear regression model," Tech. Rep., 2006.

[25] L. Ljung, *System Identification Toolbox, for use with Matlab*. The Mathworks Inc., 2006.