Sparse Estimation for Predictor-Based Subspace Identification of LPV Systems

P.M.O. Gebraad * **J.W. van Wingerden** * **M. Verhaegen** *

* *Delft Center of Systems and Control (DCSC), Delft University of Technology, Mekelweg 2, 2628CD Delft, The Netherlands. E-mail: {P.M.O.Gebraad, J.W.vanWingerden, M.Verhaegen}@tudelft.nl*

Abstract: This paper presents a Basis Pursuit DeNoising (BPDN) sparse estimation approach as a regularization technique in a predictor-based subspace method for the identification of Linear Parameter-Varying (LPV) state-space systems. It is known that in this identification method, the choice of the past window of a state predictor factorization will influence the conditioning of the main parameter estimation problem. Therefore, prior knowledge of the system order may be needed to choose the past window in such a way that this problem is well-conditioned. It will be demonstrated that sparse estimation through BPDN can reduce the sensitivity of the conditioning with respect to the past window parameter. In this way, we can simplify the task of choosing the past window to an extend that the need for prior knowledge of the system order is eliminated. Also, this paper will pay attention to the synthesis of stabilizing observer gain matrices in the identified LPV innovation-type state-space model.

Keywords: LPV systems, System identification, Subspace methods, Sparse estimation

1. INTRODUCTION

Linear Parameter-Varying (LPV) systems are linear time-varying systems for which the time variation is governed by a known scheduling signal which parameterizes operating conditions varying throughout a certain operating region. Of particular interest are LPV models with a state-space representation, as they are convenient to use for systems with multiple inputs and outputs, and can be used in optimal control synthesis, see for example Apkarian and Gahinet (1995), de Souza and Trofino (2005).

LPV identification methods with a global approach calculate LPV models from data collected in experiments in which the scheduling and input are excited simultaneously. They can be used to find the dynamic dependence of the system's input-output behavior on the scheduling. This paper is concerned with the state-of-the-art identification method of van Wingerden and Verhaegen (2009), called LPV Predictor-Based Subspace IDentification (LPV PBSID$_{\text{opt}}$). This method finds, through a global approach, discrete-time innovation-type LPV state-space models with affine dependence of the system matrices on the scheduling.

In the PBSID$_{\text{opt}}$ method, the Markov parameters of the model are found through solving a linear least-squares problem, based on a factorization of state predictors which are a function of the input, output and scheduling data in a certain past time window. The main issue addressed in this paper is that the conditioning of this least-squares problem is known to be sensitive to the size of the past window. This paper will show how sparse estimation can be used as a means of regularization of this problem.

Once the parameter estimation problem is solved, the LPV Markov parameters are used to find the state sequence, and subsequently the system matrices of the LPV innovation-type state-space model. A second issue that is addressed in this paper is that with the currently used least-squares techniques of finding the parameter-varying observer gain matrix, the observer form of the model is not guaranteed to be stable. We will show how to optimally choose the observer gain such that this stability is guaranteed throughout the operating region of the scheduling.

The outline of this paper is as follows; in Section 2 we present the model structure used in the subspace identification scheme. Section 3 gives a brief review of the LPV PSBID$_{\text{opt}}$ scheme, and in Section 3.4 an addition to this scheme is proposed which ensures stabilizing observer gains. In Section 4 we give a brief explanation of conventional regularization techniques for the parameter estimation step in LPV PSBID$_{\text{opt}}$, and we introduce the new sparse estimation approach. In Section 5 two simulation examples are used to give a proof of concept for the new techniques. We end with our conclusions in Section 6.

2. THE AFFINE LPV STATE-SPACE MODEL

The identification algorithm considers LPV systems in the discrete-time deterministic-stochastic state-space form:

$$x_{k+1} = \sum_{i=1}^{m} \mu_k^{(i)} \left(A^{(i)} x_k + B^{(i)} u_k \right) + w_k, \tag{1}$$

$$y_k = C x_k + D u_k + v_k, \tag{2}$$

where k is the time index, and $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^r$, $y_k \in \mathbb{R}^\ell$, are the state, input and output vectors. The matrices $A^{(i)} \in \mathbb{R}^{n\times n}$, $B^{(i)} \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{\ell\times n}$, $D \in \mathbb{R}^{\ell\times r}$, are local state, input, output and feedthrough matrices. The LPV system described above has a parameter-independent output equation, but if needed the method can be extended to model structures with an LPV output equation, see van Wingerden (2008). The scalars $\mu_k^{(i)}$ are the scheduling parameters, which can be interpreted as weighting factors that are used to interpolate local models. They are assumed to be measurable in real time. The system matrices depend linearly on the time-varying scheduling parameters; the time-varying state matrix is given by:

$$A_k = \sum_{i=1}^{m} \mu_k^{(i)} A^{(i)},$$

and similarly for the input matrix B. We assume that we have an affine dependence and we collect the scheduling parameters in a vector μ_k of the form:

$$\mu_k = \left[\, 1,\ \mu_k^{(2)},\ \cdots,\ \mu_k^{(m)} \,\right]^T. \tag{3}$$

Assume $\mu_k \in \mathscr{P}_c$, where $\mathscr{P}_c$ defines the operating region of the scheduling as a convex parameter polytope with vertices $\left\{\mu_{(i)} \in \mathbb{R}^m\right\}_{i=1}^h$, i.e. $\mathscr{P}_c = \mathrm{co}\left(\mu_{(1)}, \dots, \mu_{(h)}\right)$. The process noise $w_k \in \mathbb{R}^n$ and the measurement noise $v_k \in \mathbb{R}^\ell$ are assumed to be zero-mean white noise sequences, assumed to have constant joint covariance matrices:

$$\mathrm{E}\left(\begin{bmatrix} w_k \\ v_k \end{bmatrix} \begin{bmatrix} w_j^T & v_j^T \end{bmatrix}\right) = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \Delta(k-j) \tag{4}$$

with expectation operator $\mathrm{E}(\cdot)$ and Kronecker delta function $\Delta(k)$:

$$\Delta(k) = \begin{cases} 1 & \text{for } k = 0, \\ 0 & \text{for } k \neq 0. \end{cases} \tag{5}$$

3. LPV PREDICTOR-BASED SUBSPACE IDENTIFICATION

This section summarizes the identification scheme first presented in van Wingerden and Verhaegen (2009), called LPV Predictor-Based Subspace IDentification (LPV PBSID$_{\mathrm{opt}}$).

3.1 Problem formulation

The LPV PBSID$_{\mathrm{opt}}$ algorithm considers the model (1)-(2) in the innovation form:

$$x_{k+1} = \textstyle\sum_{i=1}^m \mu_k^{(i)} \left(A^{(i)} x_k + B^{(i)} u_k + K^{(i)} e_k\right), \tag{6}$$

$$y_k = C x_k + D u_k + e_k, \tag{7}$$

with $e_k \in \mathbb{R}^\ell$ denoting a zero mean white innovation process, and with observer gain matrices $K^{(i)} \in \mathbb{R}^{n\times\ell}$. The identification problem is to determine from input, output and scheduling sequences u_k, y_k, μ_k measured over a time $k = \{1, \dots, N\}$, the LPV system matrices $\left\{A^{(i)}, B^{(i)}, K^{(i)}\right\}_{i=1}^m$, C, and D. As an invertible linear state transformation $T \in \mathbb{R}^{n\times n}$ does not change the input-output behavior of the system, we determine the system matrices up to a similarity transformation: $T^{-1}A^{(i)}T$, $T^{-1}B^{(i)}$, $T^{-1}K^{(i)}$, CT and D.

3.2 Regression problem

We can rewrite system (6)-(7) in the predictor form:

$$x_{k+1} = \textstyle\sum_{i=1}^m \mu_k^{(i)} \left(\tilde{A}^{(i)} x_k + \tilde{B}^{(i)} u_k + K^{(i)} y_k\right), \tag{8}$$

$$y_k = C x_k + D u_k + e_k, \tag{9}$$

with:

$$\tilde{A}^{(i)} = A^{(i)} - K^{(i)} C, \quad \tilde{B}^{(i)} = B^{(i)} - K^{(i)} D.$$

Note that eq. (8) can be rewritten using the Kronecker product $\otimes$, as:

$$x_{k+1} = \left[\tilde{A}^{(1)}, \cdots, \tilde{A}^{(m)}\right](\mu_k \otimes x_k) + \left[\tilde{B}^{(1)}, \cdots, \tilde{B}^{(m)}\right](\mu_k \otimes u_k) + \\ + \left[K^{(1)}, \cdots, K^{(m)}\right](\mu_k \otimes y_k).$$

We can choose a past window p, and relate the state at time $k+p$ to the state at time k and the input, output and scheduling data in the window in between those time instants:

$$x_{k+p} = \phi_{p,k} x_k + \mathscr{K}_p z_k^p. \tag{10}$$

In this state predictor, $\phi_{p,k}$ is a transition matrix which is the product of the past state matrices:

$$\phi_{p,k} = \tilde{A}_{k+p-1} \cdots \tilde{A}_{k+1} \tilde{A}_k, \text{ with } \tilde{A}_k = \textstyle\sum_{i=1}^m \mu_k^{(i)} \tilde{A}^{(i)}. \tag{11}$$

The extended LPV controllability matrix $\mathscr{K}^p$ contains products of the state matrices with the input and observer matrices:

$$\mathscr{K}^p = \left[\, \mathscr{L}_p,\ \dots,\ \mathscr{L}_1 \,\right] \in \mathbb{R}^{n\times\tilde{q}}, \text{ with} \tag{12}$$

$$\mathscr{L}_1 = \left[\, \tilde{B}^{(1)},\ K^{(1)},\ \cdots,\ \tilde{B}^{(m)},\ K^{(m)} \,\right],$$

$$\mathscr{L}_p = \left[\, \tilde{A}^{(1)} \mathscr{L}_{p-1},\ \cdots,\ \tilde{A}^{(m)} \mathscr{L}_{p-1} \,\right].$$

The regressor vector z_k^p contains the corresponding Kronecker products of the past input, output data with the scheduling:

$$z_k^p = \begin{bmatrix} \mu_{k+p-1} \otimes \ldots \otimes \mu_{k+1} \otimes \mu_k \otimes \begin{bmatrix} u_k \\ y_k \end{bmatrix} \\ \mu_{k+p-1} \otimes \ldots \otimes \mu_{k+1} \otimes \begin{bmatrix} u_{k+1} \\ y_{k+1} \end{bmatrix} \\ \vdots \\ \mu_{k+p-1} \otimes \begin{bmatrix} u_{k+p-1} \\ y_{k+p-1} \end{bmatrix} \end{bmatrix} \in \mathbb{R}^{\tilde{q}}. \tag{13}$$

The size of this regressor is given by:

$$\tilde{q} = (r+l) \textstyle\sum_{j=1}^p m^j. \tag{14}$$

For a stable predictor form (8)-(9), we can choose the past window p large enough such that:

$$\phi_{j,k} \approx 0 \; \forall j \geq p, \tag{15}$$

which implies that $x_{k+p} \approx \mathscr{K}^p z_k^p$, and that:

$$y_{k+p} \approx C \mathscr{K}_p z_k^p + D u_{k+p} + e_{k+p}. \tag{16}$$

Now we define the stacked matrices U_p^N, Y_p^N, and Z:

$$U_p^N = \left[\, u_{p+1},\ \cdots,\ u_N \,\right], \quad Y_p^N = \left[\, y_{p+1},\ \cdots,\ y_N \,\right], \tag{17}$$

$$Z = \left[\, z_1^p,\ \cdots,\ z_{N-p}^p \,\right]. \tag{18}$$

Based on approximation (16), the parameters $C\mathscr{K}^p$ and D can be estimated by solving a linear least-squares problem in which the prediction error is minimized:

$$\min_{C\mathscr{K}^p, D} ||Y_p^N - C\mathscr{K}^p Z - D U_p^N||_F, \tag{19}$$

where $||\cdots||_F$ represents the Frobenius norm. The above problem may be ill-posed, in which case regularization or sparse estimation techniques are needed to find a reliable estimate of $C\mathscr{K}_p$, see Section 4.

3.3 Estimation of the state sequence

We define the extended observability matrix Γ_p of the first local model as:

$$\Gamma_p = \left[\, C^T,\ \left(CA^{(1)}\right)^T,\ \dots,\ \left(C\left(A^{(1)}\right)^{p-1}\right)^T \,\right]^T, \tag{20}$$

From $C\mathscr{K}^p$, we can approximate the product of Γ^p and $\mathscr{K}^p$, by constructing:

$$\Gamma^p \mathscr{K}^p \approx \begin{bmatrix} C\mathscr{L}_p & C\mathscr{L}_{p-1} & \cdots & C\mathscr{L}_1 \\ 0 & C\tilde{A}^{(1)}\mathscr{L}_{p-1} & \cdots & C\tilde{A}^{(1)}\mathscr{L}_1 \\ & & \ddots & \\ 0 & & 0 & C\left(\tilde{A}^{(1)}\right)^{p-1}\mathscr{L}_1 \end{bmatrix}. \tag{21}$$

The zeros appear in the lower triangular part of this matrix as a consequence of approximation (15).

Now we can compute $\Gamma^p \mathscr{K}^p Z$, for which it holds that:

$$\Gamma^p \mathscr{K}^p Z \approx \Gamma^p X_p^N, \text{ where } X_p^N = [x_{p+1}, \ldots, x_N]. \quad (22)$$

Under the assumptions that X_p^N and Γ^p both have full rank and that $p\ell > n$, we estimate the state sequence (up to a similarity transformation) and the order of the system based on a rank revealing Singular Value Decomposition (SVD):

$$\widehat{\Gamma^p \mathscr{K}^p} Z = [\,\mathscr{U} \;\; \mathscr{U}_\perp\,] \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} \mathscr{V} \\ \mathscr{V}_\perp \end{bmatrix}, \quad \widehat{X_p^N} = \Sigma_n \mathscr{V}, \quad (23)$$

where Σ_n is the diagonal matrix containing the n largest singular values, and $\mathscr{U}$ and $\mathscr{V}$ are the corresponding column and row spaces. The system order n is thus found by detecting a gap between the singular values.

3.4 Estimation of the system matrices, guaranteeing a stable observer form

Once the state, input, output, and scheduling sequence are known, the system matrices can be estimated by solving the linear relations (6)-(7) in a least-squares sense. We extend this approach with H_∞ techniques to find observer gain matrices $K^{(i)}$ that result in a stable H_∞-optimal predictor form of the identified model. First, the matrices $A^{(i)}$, $B^{(i)}$, C and D are estimated by solving the following linear least-squares problems:

$$\min_{\breve{A},\breve{B}} \left\| W_p^{N-1}(\breve{A},\breve{B}) \right\|_F^2, \qquad \min_{C,D} \left\| V_p^N(C,D) \right\|_F^2, \quad (24)$$

where $\breve{A} = \left[A^{(1)}, \cdots, A^{(m)}\right]$, $\breve{B} = \left[B^{(1)}, \cdots, B^{(m)}\right]$, and

$$W_p^{N-1} = X_{p+1}^N - [\breve{A},\breve{B}] \begin{bmatrix} \mu_{p+1} \otimes x_{p+1}, & \cdots, & \mu_{N-1} \otimes x_{N-1} \\ \mu_{p+1} \otimes u_{p+1}, & \cdots, & \mu_{N-1} \otimes u_{N-1} \end{bmatrix},$$

$$V_p^N = Y_p^N - [C,D] \begin{bmatrix} X_p^N \\ U_p^N \end{bmatrix}.$$

Then from the residuals W_p^{N-1}, V_p^N the covariance matrices of the process and measurement noise, defined in Eq. (4), can be estimated:

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \approx \frac{1}{N-p-2} \begin{bmatrix} W_p^{N-1} \\ V_p^{N-1} \end{bmatrix} \begin{bmatrix} W_p^{N-1} \\ V_p^{N-1} \end{bmatrix}^T. \quad (25)$$

Using the square root formulation of Verhaegen and Verdult (2007), §5.5.1, we rewrite the system (1)-(2) as:

$$x_{k+1} = \textstyle\sum_{i=1}^m \mu_k^{(i)} \left(A^{(i)} x_k + B^{(i)} u_k\right) + B_\varepsilon \varepsilon_k, \quad (26)$$

$$y_k = C x_k + D u_k + D_\varepsilon \varepsilon_k, \quad (27)$$

with $\varepsilon_k \sim (0, I_{n+l})$, and:

$$B_\varepsilon = \left[\, S R^{-T/2}, \; Q_x^{1/2} \,\right], \quad D_\varepsilon = \left[\, R^{1/2}, \; 0_{l\times n} \,\right],$$

where $Q_x = Q - S R^{-1} S^T$. For this system, an observer of the form:

$$\hat{x}_{k+1} = \textstyle\sum_{i=1}^m \mu_k^{(i)} \left(A^{(i)} \hat{x}_k + B^{(i)} u_k + K^{(i)} e_k\right), \quad (28)$$

can be designed, with state estimate $\hat{x}_k \in \mathbb{R}^n$ and output error $e_k = y_k - C\hat{x}_k - D u_k$. The error system $\mathscr{E}$ of this observer is:

$$\tilde{x}_{k+1} = \textstyle\sum_{i=1}^m \mu_k^{(i)} \left[\left(A^{(i)} - K^{(i)} C\right) \tilde{x}_k + \left(B_\varepsilon - K^{(i)} D_\varepsilon\right) \varepsilon_k\right], \quad (29)$$

with state error $\tilde{x}_k = x_k - \hat{x}_k$. From Calafiore and Dabbene (2009) it follows that we can minimize $\|\mathscr{E}\|_\infty$, the least upper bound of the error system gain, and guarantee the stability of $\mathscr{E}$ in the scheduling space $\mathscr{P}_c$, by solving the following optimization problem with linear matrix inequality constraints:

$$\min \gamma^2 \text{ s.t. } \exists P \succ I_n,$$
$$\begin{bmatrix} P & 0_{n,n} & P\left(\sum_{i=1}^m \left(\mu^{(i)} A^{(i)}\right) - S R^{-1} C\right) & P Q_x^{1/2} \\ * & I_n & I_n & 0_{n,n} \\ * & * & P + \gamma^2 C^T R^{-1} C & 0_{n,n} \\ * & * & * & \gamma^2 I_n \end{bmatrix} \succ 0 \quad (30)$$
$$\forall \mu \in \left\{\mu_{(1)}, \ldots, \mu_{(h)}\right\},$$

and calculating the observers gains as:

$$K^{(i)} = \left(\gamma^2 A^{(i)} C_P + S\right) \left(\gamma^2 C C_P + R\right)^{-1} \quad (31)$$

with $C_P = (P - I_n)^{-1} C^T$. See Appendix A for the derivation. With these observer gains, it is guaranteed that $\|\mathscr{E}\|_\infty < \gamma$, which yields an upper bound on the root mean square of the state error $\tilde{x}_k$.

4. REGULARIZATION METHODS FOR LPV PBSID$_{\text{OPT}}$

If the data matrix $\Phi = \left[\, Z^T, \; \left(U_p^N\right)^T \,\right]$ is full rank, the unique solution to the estimation problem (19) is given by:

$$[\, C\mathscr{K}_p, \; D \,] = Y\Phi\left(\Phi^T \Phi\right)^{-1}. \quad (32)$$

The problem is ill-posed if $\Phi^T\Phi$ is either non-invertible, or if Φ is (nearly) rank-deficient with no significant gap in its singular values, rendering the problem underdetermined or the estimates of $C\mathscr{K}_p$ and D sensitive to noise in the data. Regularization techniques are computational techniques that aim at modifying the ill-posed problem in such a way that its solution is unique and less sensitive to error in the data, thereby preventing overfitting.

4.1 Conventional regularization techniques

Conventionally, Tikhonov or Truncated Singular Value Decomposition (TSVD) regularization techniques are employed in LPV PBSID$_{\text{opt}}$. The trade-off parameters of these techniques can be chosen using L-curve or Generalized Cross-Validation (GCV) criteria. See Verdult and Verhaegen (2005) for a more detailed explanation. Tikhonov regularization, used in the simulation example of Section 5.2, adds the requirement that the 2-norm of the solution is small. The underlying assumption is that a solution with a small value elements is less sensitive to noise.

4.2 Regularization through sparse estimation

The past window p is an important factor influencing the conditioning of the parameter estimation problem in the LPV identification scheme. Although no theoretically well-founded rules for choosing p are available, as a rule of thumb, the past window can be chosen within a small factor (2 to 4) of the system order n to come to a well-posed problem without using more samples than strictly necessary. However, in general we cannot assume the system order to be known beforehand.

For a stable system (8)-(9) the parameter $C\mathscr{K}_p$ has a decaying structure, since the blocks $\mathscr{L}^i \to 0$ as $i \to \infty$. Hence, if p is chosen (too) large, insignificant elements in the left part of the parameter matrix can be set to zero in order to overcome the problem of overparameterization. This sparsity is imposed by posing the parameter estimation as a Basis Pursuit DeNoising (BPDN) problem:

$$\min \left\| [C\mathscr{K}_p, \; D] \right\|_{2,1} \text{ s.t. } \left\| Y_p^N - C\mathscr{K}_p Z - D U_p^N \right\|_F \le \sigma \quad (33)$$

where $\|\cdot\|_{2,1}$ denotes the $\ell_{2,1}$-norm, defined as the sum of the two-norms of columns of a matrix. The scalar trade-off parameter $\sigma > 0$ balances the 2-norm of the residual against the sparsity of the solution.

BPDN with AIC parameter selection In Rojas and Hjalmarsson (2011) it is suggested, based on the Akaike Information Criterion (AIC), to use the trade-off parameter choice:

$$\sigma = (1+2\tilde{q}/N)\,\bar{\varphi}, \text{ with } \bar{\varphi} = \left\|Y_p^N - \overline{C\mathscr{K}_p}Z - \bar{D}U_p^N\right\|_F, \quad (34)$$

where $\{\overline{C\mathscr{K}_p}, \bar{D}\}$ is the solution of the parameter estimation problem (19) without regularization. This parameter choice, denoted by BPDN-AIC, was found to yield an improvement of the resulting model quality in some LPV identification experiments, but since the sparsity of the solution is only ensured for $N \to \infty$, for higher order systems and small N the results were found not to be satisfactory.

BPDN with a stopping criterion based on validation data prediction error In Algorithm 1, we introduce a scheme to automatically find a trade-off point using a criterion based on the prediction error on validation data, denoted by BPDN-SV. This scheme is based on SPGL1, a solver for BPDN problems by van den Berg and Friedlander (2008). Different iterative steps of the scheme are shown in Figure 1. Line 4 of Algorithm 1 calculates the gradient $\nabla\varphi_{id}$ [1] of the non-increasing convex Pareto trade-off curve of φ_{id}, the 2-norm of the prediction error on the identification data (defined in line 9), against τ, the upper bound on the $\ell_{2,1}$-norm of the solution. In line 5, $\nabla\varphi_{id}$ is used to perform Newton root-finding iterations on this curve. In each iteration, σ, the upper bound on the 2-norm of the residual in the BPDN problem, is decreased at the cost of an increased τ in the equivalent least-squares problem with an $\ell_{2,1}$-constraint. The latter problem, also known as a Lasso problem, is solved in line 6. The Newton root-finding approach converges to small φ_{id} in a small number of iterations, even if the Lasso problems are solved only approximately, and it provides an additional stopping criterion when $\nabla\varphi_{id}$ is small.

The Lasso problems are approximately solved using Spectral Projected Gradient (SPG) techniques. Compared to the interior-point method of Candès and Romberg (2011), SPG turns out to be better suited to solve these problems if they involve an ill-conditioned data matrix Z_{id}, making it more suited to the example regularization problems of Section 5.

In this adapted scheme, an increase of φ_{val}, the prediction error norm for a validation dataset, is used as stopping criterion for the root-finding iterations. The final solution $C\mathscr{K}_p$ is selected as the iterative solution $\widetilde{C\mathscr{K}_p}$ for which the smallest φ_{val} was found. After the stop, φ_{val} can be minimized further by performing a line search between the last two or three iterations of τ, but to reduce the computational burden this is left out in Algorithm 1 and the simulation examples.

[1] For the derivation of the explicit formulation of $\nabla\varphi_{id}$, we refer to van den Berg and Friedlander (2008).

Algorithm 1 BPDN-SV algorithm for parameter estimation in LPV PBSID$_{\text{opt}}$. For brevity of the notation, we hold $D = 0$. We define $\|\cdot\|_{\infty,2}$ as the largest 2-norm of the rows of a matrix. The data matrices Y_{id}, Z_{id} are Y_p^N, Z for an identification dataset $\{u_k, y_k, \mu_k\}_{k=1}^N$, constructed by eq. (17)-(18). Similarly, Y_{val}, Z_{val} are constructed from a different validation dataset. The small scalar $\nabla_{tol} > 0$ is used to define an additional stopping point when the trade-off curve gradient $\nabla\varphi_{id}$ is small.

1: **given** $Y_{id}, Z_{id}, Y_{val}, Z_{val}$
2: $\widetilde{C\mathscr{K}_p}, \tau \leftarrow 0, \quad \varphi_{val}, \varphi_{val}^{best} \leftarrow \|Y_{val}\|_F, \quad \varphi_{id} \leftarrow \|Y_{id}\|_F$
3: **repeat**
4: $\quad \nabla\varphi_{id} \leftarrow -\left\|Z_{id}\left(Y_{id} - \widetilde{C\mathscr{K}_p}Z_{id}\right)^T / \varphi_{id}\right\|_{\infty,2}$
5: $\quad \tau \leftarrow \tau - \varphi_{id}/\nabla\varphi_{id}$
6: $\quad \widetilde{C\mathscr{K}_p} \leftarrow \arg\min_{\widetilde{C\mathscr{K}_p}} \left\|Y_{id} - \widetilde{C\mathscr{K}_p}Z_{id}\right\|_F$ s.t. $\left\|\widetilde{C\mathscr{K}_p}\right\|_{2,1} \leq \tau$
7: $\quad \varphi_{val}^{old} \leftarrow \varphi_{val}$
8: $\quad \varphi_{val} \leftarrow \left\|Y_{val} - \widetilde{C\mathscr{K}_p}Z_{val}\right\|_F$
9: $\quad \varphi_{id} \leftarrow \left\|Y_{id} - \widetilde{C\mathscr{K}_p}Z_{id}\right\|_F$
10: $\quad$ **if** $\varphi_{val} < \varphi_{val}^{best}$ **then**
11: $\quad\quad \varphi_{val}^{best} \leftarrow \varphi_{val}, \quad C\mathscr{K}_p \leftarrow \widetilde{C\mathscr{K}_p}$
12: $\quad$ **end if**
13: **until** $\varphi_{val} > \varphi_{val}^{old}$ **or** $\nabla\varphi_{id} < \nabla_{tol}$
14: **return** $C\mathscr{K}_p$

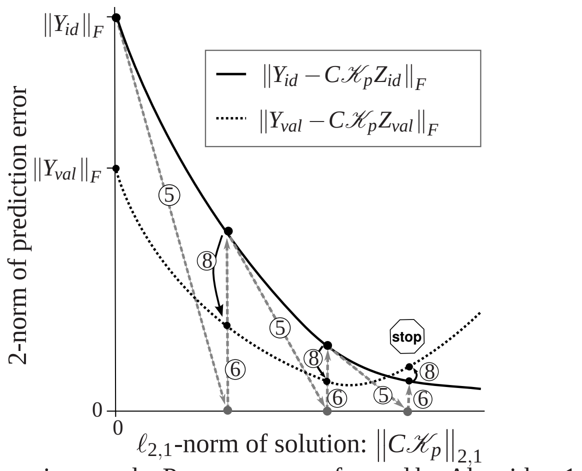


Fig. 1. Iterations on the Pareto curve performed by Algorithm 1. Encircled numbers correspond to lines in this algorithm.

5. SIMULATION EXAMPLES

5.1 Example I: Identification of a second-order LPV system

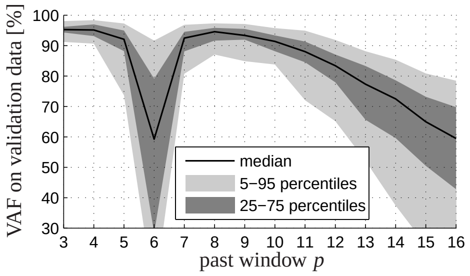
System description In §3.8.1 of van Wingerden (2008), an LPV identification experiment is presented on a second order LPV model of the form (1)-(2) that might represent the simplified out-of-plane dynamics of a flexible rotor blade of a wind turbine, using the blade rotation angle as scheduling. The model has the system matrices:

$$\left[A^{(1)}|A^{(2)}\right] = \left[\begin{array}{cc|cc} 0 & 0.073 & -0.0021 & 0 \\ -6.52 & -0.50 & -0.014 & 0.52 \end{array}\right],$$

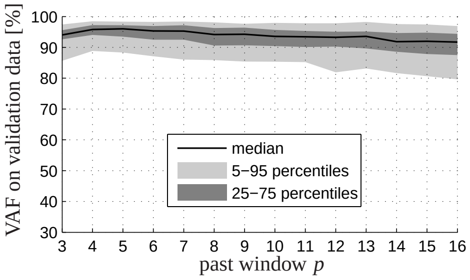
$$\left[B^{(1)}|B^{(2)}\right] = \left[\begin{array}{c|c} -0.72 & 0 \\ -9.63 & 0 \end{array}\right], \; C = \left[\, 1 \;\; 0 \,\right], \; D = 0\,.$$

Identification experiment In the identification experiment of van Wingerden (2008), the order was assumed to be known beforehand, and accordingly, the past window was tuned with the rule of thumb mentioned in Section 4.2 as $p = 8$. We redo the identification for a range of p. We use a small data set, $N = 75$, excite the system with a zero-mean Gaussian input with $\text{var}(u_k) = 1$, and a scheduling $\mu_k^{(2)} = \cos(2\pi k/10) + 0.2$, and add zero-mean Gaussian process and measurement noise with a Signal-To-Noise Ratio (SNR) of 25dB, i.e.:

$$\text{SNR}(y_k, w_k) = \text{SNR}(y_k, v_k) = 25\text{dB},$$



(a) Without regularization



(b) With sparse estimation

Fig. 2. Distribution of VAF values on validation data for 2nd order models estimated in 80 Monte-Carlo experiments with different realizations of the noise and input sequences, for a range of past windows p.

with the SNR of a noise signal s_k defined as:

$$\mathrm{SNR}(y_k, s_k) = 20\log_{10}\frac{\mathrm{var}\,(y_k)}{\mathrm{var}\,(s_k)}.$$

The quality of the identified models is expressed in the Variance Accounted For (VAF) for an input sequence different from the one used in identification. In the validation data, no noise is added. The VAF value is defined as:

$$\mathrm{VAF} = 100\% \cdot \max\left\{0,\ 1 - \frac{\mathrm{var}\,(\hat{y}_k - y_k)}{\mathrm{var}\,(y_k)}\right\},$$

where $\hat{y}_k$ is the estimated system output, and y_k is the output from the true model. To show the effect of noise on the model estimate, Monte-Carlo simulations with 80 runs are carried out with different realizations of the input and noise in each run. Figure 2(a) shows that without regularization, the quality of the identified model may decrease for other choices of the past window than $p = 8$. In this identification experiment, we found that the conventional regularization methods mentioned in Section 4.1 did not improve the results.

When performing the same identification experiment again using the BPDN approach of Section 4.2 the quality of the estimated model remains at approximately the same level for a large range of p, see Figure 2(b). This demonstrates that sparse estimation enables estimating the model without prior knowledge of the system order, or tuning of p. We used the BPDN-SV approach of Section 4.2.2, using the first three quarters of the sequences in the identification dataset to build Y_{id} and Z_{id}, and the remainder for Y_{val} and Z_{val}. The BPDN-AIC approach of Section 4.2.1 yielded similar results (not shown for brevity), but in Example II we will show that for higher order systems, it does not perform that well. A disadvantage of the BPDN approach is that it takes more calculation time, see Figure 3; BPDN-SV takes 2 to 7 Newton root-finding steps to obtain the regularized $C\mathscr{K}_p$ parameter, and in each step a Lasso problem is solved with 15 to 60 SPG iterations, while in the case without regularization a single problem (19) is solved by a direct method based on a Cholesky decomposition.

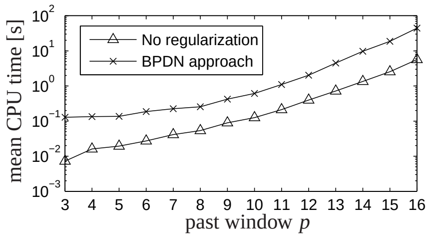


Fig. 3. Mean CPU time needed to estimate the state sequence in the experiments of Example I. The algorithms were run using MATLAB R2010b on an Intel Q8400 2.7 GHz PC.

Table 1. VAF on validation data and stability region of the predictor form of the identified models.

	VAF		Stability region	
$p = 5$	LS	H_∞	LS	H_∞
SNR=15dB	71.7	61.7	$\mu_k^{(2)} \in [-2.7,\ 3.7]$	$\mu_k^{(2)} \in [-2.8,\ 3.6]$
SNR=25dB	87.2	77.3	$\mu_k^{(2)} \in [-2.0,\ 4.9]$	$\mu_k^{(2)} \in [-3.2,\ 3.7]$
SNR=50dB	97.9	96.4	$\mu_k^{(2)} \in [-2.1,\ 3.3]$	$\mu_k^{(2)} \in [-2.3,\ 2.7]$
SNR=75dB	0	99.5	$\mu_k^{(2)} \in [-1.3,\ 0.6]$	$\mu_k^{(2)} \in [-2.8,\ 2.3]$
SNR=100dB	0	99.7	$\mu_k^{(2)} \in [-1.8,\ 0.4]$	$\mu_k^{(2)} \in [-1.8,\ 2.0]$
$p = 10$	LS	H_∞	LS	H_∞
SNR=15dB	63.8	64.5	$\mu_k^{(2)} \in [-0.5,\ 3.4]$	$\mu_k^{(2)} \in [-4.8,\ 5.3]$
SNR=25dB	83.2	79.1	$\mu_k^{(2)} \in [-2.5,\ 6.9]$	$\mu_k^{(2)} \in [-2.8,\ 3.3]$
SNR=50dB	97.7	96.9	$\mu_k^{(2)} \in [-1.6,\ 5.2]$	$\mu_k^{(2)} \in [-2.2,\ 2.6]$
SNR=75dB	98.4	98.8	$\mu_k^{(2)} \in [-1.5,\ 6.9]$	$\mu_k^{(2)} \in [-3.0,\ 3.6]$
SNR=100dB	0	98.2	$\mu_k^{(2)} \in [-1.7,\ 0.6]$	$\mu_k^{(2)} \in [-2.2,\ 2.7]$

Stabilization of the predictor In Table 1, we see the VAF values for a validation dataset resulting from simulating the predictor form of the models identified with the BPDN method as described above, with noise of different SNRs. The observer gains $K^{(i)}$ are generated by solving the linear relations (6)-(7) in a least-squares sense (LS), or by solving the H_∞ design problem described in Section 3.4. For the high SNR values, the LS method may produce a predictor form which is unstable in the operating range of the scheduling $\mu_k^{(2)} \in [-0.8, 1.2]$, resulting in a bad fit to the data. With the H_∞ design, resulting in a minimization of the upper bound on the gain of the error system, the stability of the predictor is guaranteed. This design may however be conservative for the particular scheduling sequence in the validation dataset for lower SNR values, resulting in a lower VAF than with the LS method.

5.2 *Example II: Identification of a fourth-order LPV system*

System description To demonstrate the method for a higher order system, we identify a fourth order system of the form (6)-(7) with $r = l = m = 2$ and varying poles:

$$\left[A^{(1)}|A^{(2)}\right] = \left[\begin{array}{cccc|cccc} 0.67 & 0.67 & 0 & 0 & -0.2 & -0.2 & 0 & 0 \\ -0.67 & 0.67 & 0 & 0 & 0.2 & -0.2 & 0 & 0 \\ 0 & 0 & -0.67 & -0.67 & 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0.67 & -0.67 & 0 & 0 & -0.2 & 0.2 \end{array}\right],$$

$$B^{(1)} = \begin{bmatrix} 0.66 & 1.97 & 4.32 & -2.64 \\ -0.53 & 0.48 & -0.49 & -0.34 \end{bmatrix}^T,\ B^{(2)} = 0,$$

$$K^{(1)} = \begin{bmatrix} -0.70 & 0.17 & 0.65 & -0.94 \\ -0.15 & 0.56 & -0.47 & 0.10 \end{bmatrix}^T,\ K^{(2)} = 0,$$

$$C = \begin{bmatrix} -0.37 & 0.075 & -0.52 & 0.58 \\ -0.90 & 0.75 & 0.12 & 0.098 \end{bmatrix},\ D = 0.$$

Identification experiment We excite the system with a zero-mean Gaussian input signal with $\mathrm{var}\,(u_k) = 1$, and scheduling $\mu_k^{(2)} = 0.5 + \sin(k)$, add zero-mean Gaussian noise e_k with

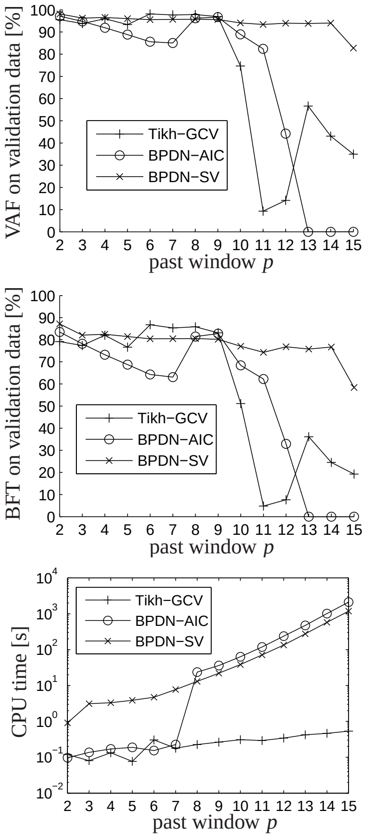


Fig. 4. VAF and BFT on validation data of estimated 4$^{\text{th}}$ order models, averaged over the two outputs. The lower figure gives the calculation time for estimating the state sequence using MATLAB R2010b on an Intel Q8400 2.7GHz PC.

variance 0.25, and collect $N = 250$ samples. We perform the identification with this data using three different regularization methods: Tikhonov with GCV parameter selection (Tikh-GCV), BPDN-AIC, and BPDN-SV. In BPDN-SV, we use the last one-third of the identification dataset to build Y_{val}, Z_{val}. In Tikh-GCV, we use a kernel method explained in van Wingerden and Verhaegen (2009) to reduce the size of the data matrices. We again evaluate the quality of the identified models by the VAF for a validation dataset with a different realization of the input. Further, we calculate the Best Fit percentage (BFT) on this validation data, showing possible bias of the models:

$$\text{BFT} = 100\% \cdot \max\left\{0, 1 - \frac{\|y_k - \hat{y}_k\|_2}{\|y_k - \bar{y}\|_2}\right\},$$

where $\bar{y}$ is the mean of output signal y_k. The results in Figure 4 show that by using BPDN-SV, we extend the range of past windows for which we get good results from $p = 9$ to $p = 14$. As in Example I, the performance improvement through BPDN-SV regularization goes at the cost of an increase in calculation time due the use of iterative methods. In terms of calculation time and estimated model quality, BPDN-SV outperforms BPDN-AIC for larger p. Other conventional regularization methods mentioned in Section 4.1 did not outperform Tikh-GCV.

6. CONCLUSIONS

This paper presents a basis pursuit denoising sparse estimation approach as a regularization method in a predictor-based subspace identification scheme for linear parameter-varying state-space systems. It uses a criterion based on the prediction error on validation data to find the trade-off parameters of the regularization method. In two simulation examples, we demonstrate that this regularization reduces the sensitivity of the performance of the LPV PBSID$_{\text{opt}}$ scheme to the choice of the past window parameter p, eliminating the need for prior knowledge of the system order in choosing the p parameter. Also, we presented a scheme to find H_∞ optimal observer gains stabilizing the observer form of the identified models. Simulation Example I demonstrates that this prevents instability of the observer form that may occur when using the conventional least-squares approach to find the system matrices.

Appendix A. DERIVATION OF THE H_∞-OPTIMAL LPV OBSERVER GAIN DESIGN PROBLEM

From Corollary 1 in Calafiore and Dabbene (2009) it follows that if we define $\mathscr{R} = D_\varepsilon {D_\varepsilon}^T$, and an orthonormal basis for the null space of D_ε as $\mathscr{N} = [\,0_{n\times l},\ I_n\,]^T$ then

$$\mathscr{E} \text{ is stable and } \|\mathscr{E}\|_\infty < \gamma \text{ for } \mu \in \mathscr{P}_c \text{ if } \exists P \succ I_n,$$

$$\begin{bmatrix} P & 0_{n\times n} & P\left(\sum_{i=1}^{m}\left(\mu^{(i)}A^{(i)}\right) - B_\varepsilon {D_\varepsilon}^T \mathscr{R}^{-1} C\right) & PB_\varepsilon \mathscr{N} \\ * & I_n & I_n & 0_{n\times n} \\ * & * & P + \gamma^2 C^T \mathscr{R}^{-1} C & 0_{n\times n} \\ * & * & * & \gamma^2 I_n \end{bmatrix} \succ 0$$

$$\forall \mu \in \left\{\mu_{(1)}, \ldots, \mu_{(v)}\right\}$$

$$\text{and } K^{(i)} = \left(A^{(i)} C_P + \frac{1}{\gamma^2} B_\varepsilon {D_\varepsilon}^T\right)\left(CC_P + \frac{1}{\gamma^2}\mathscr{R}\right)^{-1}.$$

Since $\mathscr{R} = R$, $B_\varepsilon {D_\varepsilon}^T = S$ and $B_\varepsilon \mathscr{N} = Q_x^{1/2}$, the above simplifies to (30).

REFERENCES

Apkarian, P. and Gahinet, P. (1995). A convex characterization of gain-scheduled H_∞ controllers. *IEEE Trans. Automat. Control*, 40(5), 853–864.

Calafiore, G. and Dabbene, F. (2009). Observer design with guaranteed RMS gain for discrete-time LPV systems with Markovian jumps. *Internat. J. Robust Nonlinear Control*, 19, 676–691.

Candès, E. and Romberg, J. (2011). ℓ_1-magic : Recovery of sparse signals via convex programming. URL `http://www.l1-magic.org/`.

de Souza, C.E. and Trofino, A. (2005). Gain-scheduled H_2 controller synthesis for linear parameter varying systems via parameter-dependent Lyapunov functions. *Internat. J. Robust Nonlinear Control*, 16(5), 243–257.

Rojas, C. and Hjalmarsson, H. (2011). Sparse estimation based on a validation criterion. In *Proc. of the* 50th *IEEE Conf. on Decision and Control (CDC)*. Orlando, Florida, USA.

van den Berg, E. and Friedlander, M.P. (2008). Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2), 890–912.

van Wingerden, J.W. (2008). *Control of wind turbines with 'smart' rotors: proof of concept & LPV subspace identification*. Ph.D. thesis, Delft University of Technology.

van Wingerden, J.W. and Verhaegen, M. (2009). Subspace identification of bilinear and LPV systems for open- and closed-loop data. *Automatica*, 45(2), 372–381.

Verdult, V. and Verhaegen, M. (2005). Kernel methods for subspace identification of multivariable LPV and bilinear systems. *Automatica*, 41(9), 1557–1565.

Verhaegen, M. and Verdult, V. (2007). *Filtering and system identification*. Cambridge University Press.