2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

LPV subspace identification using a novel nuclear norm regularization method

P.M.O. Gebraad, J.W. van Wingerden, G.J. van der Veen and M. Verhaegen

Index Terms— **Subspace identification, LPV systems, System identification, Regularization**

Abstract— **It is well-known that recently proposed Linear Parameter-Varying (LPV) subspace identification techniques suffer from a curse of dimensionality leading to an ill-posed parameter estimation problem. In this paper we will focus on regularization methods to solve the parameter estimation problem. Tikhonov and TSVD regularization are conventional general-purpose regularization methods. These general-purpose regularization methods give preference to a solution with a small 2-norm. In principle many other types of additional information about the desired solution can be incorporated in order to stabilize the ill-posed problem. The main contribution of this paper is that we propose a novel regularization strategy for LPV subspace methods: the nuclear norm regularization method. By applying state-of-the-art convex optimization techniques, the method stabilizes the parameter estimation problem by including information on the desired solution that is specific to the (LPV) subspace identification scheme. We will conclude the paper with a summarizing comparison between the different regularization techniques.**

I. Introduction

In system identification, measured input and output signals of a system are used to calculate a model describing its dynamics. The identification methods for Linear Time Invariant (LTI) systems are well-established, but the resulting models are only valid in one operating point. To obtain a model that is valid throughout a certain operating region, identification methods for Linear Parameter-Varying (LPV) systems can be used. LPV systems are linear time-varying systems where the time variation is governed by a known scheduling signal parameterizing the 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 e.g. [1], [2]). This paper is concerned with an LPV subspace identification technique with a global approach, which means that in the experiments the scheduling and input are excited simultaneously, so that the dynamic dependence of the input-output behavior on the scheduling can be found from the collected data. Current LPV subspace identification methods can obtain models with affine dependence of the state-space system matrices on arbitrary scheduling sequences. They have the advantage over nonlinear programming approaches given in e.g. [3] that the model parameters are found through convex optimization, for which in principle no initial estimate is needed.

A major drawback of early subspace approaches is that they are computationally demanding, as the dimensions of the data matrix involved grows quickly with the system order and the number of scheduling parameters. Therefore, the state-of-the-art subspace technique in [4] uses a kernel method to reduce the data matrix to a square matrix with dimensions equal to the number of samples. The computational complexity then depends on the number of samples included in the data. However, the estimation of a large number of model parameters using a small number of samples can lead to non-unique solutions, or to large variance error in the estimated parameters when we have noise in the data [5]. Solving the parameter estimation problem therefore requires regularization techniques. These techniques introduce additional requirements on the estimated parameters, thereby allowing a small bias, in order to reduce the variance error. Using newly available convex optimization techniques for nuclear norm minimization, this paper presents a new regularization technique incorporating requirements in the parameter estimation problem that are useful further on in the subspace identification scheme.

The outline of this paper is as follows; we start in Section II with a brief review of the LPV subspace identification scheme. In Section III we explain the conventional, and the new regularization techniques. In Section IV a simulation example is used to give a proof of concept for the new technique. We end with our conclusions on the applicability of the new regularization technique.

II. LPV subspace identification

In this section we present a brief summary of the LPV subspace identification scheme given in [4].

A. Problem formulation

State-space representations of LPV systems have system matrices that are a known function of the scheduling. In this paper, we consider LPV systems with a parameter-independent output equation. This is not only for simplicity, but also because many practical LPV systems have a parameter independent output equations. However, the method presented here can be extended to model structures with an LPV output equation [6]. In this paper, we consider the

The authors are with the Delft Center for Systems and Control (DCSC), Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands. P.M.O.Gebraad@student.tudelft.nl (P.M.O. Gebraad) G.J.vanderVeen@tudelft.nl (G.J. van der Veen) J.W.vanWingerden@tudelft.nl (J.W. van Wingerden) m.verhaegen@tudelft.nl (M. Verhaegen)

following LPV model:

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

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

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^r$, $y_k \in \mathbb{R}^\ell$, are the state, input and output vectors. The vector $e_k \in \mathbb{R}^\ell$ denotes the zero mean white innovation process. The matrices $A^{(i)} \in \mathbb{R}^{n\times n}$, $B^{(i)} \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{\ell\times n}$, $D \in \mathbb{R}^{\ell\times r}$, $K^{(i)} \in \mathbb{R}^{n\times \ell}$ are the local system, input, output, direct feedthrough, and observer gain matrices; and $\mu_k^{(i)} \in \mathbb{R}$ the local weights. The index m is referred to as the number of local models or scheduling parameters. Note that the system, input, and observer matrices depend linearly on the time-varying scheduling vector. The time-varying system matrix is now given by:

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

This can be similarly done for the other system matrices. We assume that we have an affine dependence and the scheduling is given by:

$$\mu_k = \begin{bmatrix} 1, & \mu_k^{(2)}, & \cdots, & \mu_k^{(m)} \end{bmatrix}^T. \tag{3}$$

Assume $\mu_k \in \mathscr{P}_c$, where $\mathscr{P}_c$ defines a parameter polytope. We can rewrite (1)-(2) in the predictor form as:

$$x_{k+1} = \sum_{i=1}^{m} \mu_k^{(i)} \left(\tilde{A}^{(i)} x_k + \tilde{B}^{(i)} u_k + K^{(i)} y_k \right), \tag{4}$$

$$y_k = C x_k + D u_k + e_k, \tag{5}$$

with

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

It is well-known that an invertible linear transformation of the state does not change the input-output behavior of a state-space system. Therefore, we can only determine the system matrices up to a similarity transformation $T \in \mathbb{R}^{n\times n}$: $T^{-1}A^{(i)}T$, $T^{-1}B^{(i)}$, $T^{-1}K^{(i)}$, CT, and D.

The identification problem can now be formulated as: given the input sequence u_k, the output sequence y_k, and the scheduling sequence μ_k over a time $k = \{1,\ldots,N\}$; find, if they exist, the LPV system matrices $A^{(i)}$, $B^{(i)}$, $K^{(i)}$, C, and D for all $i \in \{1,2,\cdots,m\}$ up to a global similarity transformation.

B. Assumptions and notation

First we define the transition matrix for discrete-time time-varying systems [7]; i.e.

$$\phi_{j,k} = \tilde{A}_{k+j-1} \cdots \tilde{A}_{k+1} \tilde{A}_k. \tag{6}$$

For ease of notation we define: $z_k = \begin{bmatrix} u_k^T, & y_k^T \end{bmatrix}^T$ and $\overline{B}^{(i)} = \begin{bmatrix} \tilde{B}^{(i)}, & K^{(i)} \end{bmatrix}$. We define a past window length by p. This window is used to define the following stacked vector:

$$\overline{z}_k^p = \begin{bmatrix} z_k \\ z_{k+1} \\ \vdots \\ z_{k+p-1} \end{bmatrix}.$$

We assume that the state sequence:

$$X = \begin{bmatrix} x_{p+1}, & \cdots, & x_N \end{bmatrix}, \tag{7}$$

has full row rank; and that the matrix,

$$\Gamma^p = \begin{bmatrix} C \\ C\tilde{A}^{(1)} \\ \vdots \\ C\left(\tilde{A}^{(1)}\right)^{p-1} \end{bmatrix}, \tag{8}$$

has full column rank. This last matrix can be interpreted as the extended observability matrix of the first local model. For persistency of excitation it is also required that the scheduling sequence satisfies the following relation:

$$\operatorname{rank}\left(\begin{bmatrix} \mu_0, & \mu_1, & \cdots, & \mu_{N-p} \end{bmatrix} \right) = m,$$

and $N - p + 1 > m$.
Further, we define the matrix:

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

with:

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

This operator $\mathscr{L}_j$ is used to define a time-invariant LPV controllability matrix $\mathscr{K}^p$, by:

$$\mathscr{K}^p = \begin{bmatrix} \mathscr{L}_p, & \mathscr{L}_{p-1}, & \cdots & \mathscr{L}_1 \end{bmatrix} \in \mathbb{R}^{n\times \tilde{q}}, \tag{9}$$

with the size $\tilde{q}$ given by:

$$\tilde{q} = (r+l) \textstyle\sum_{j=1}^{p} m^j. \tag{10}$$

C. Regression problem

The first objective of the algorithm is to reconstruct the state sequence up to a similarity transformation. The state x_{k+p} is given by:

$$x_{k+p} = \phi_{p,k} x_k + \mathscr{K}^p N_k^p \overline{z}_k^p,$$

where $\phi_{p,k}$ is the transition matrix given in (6); $\mathscr{K}^p$ is the time-invariant LPV controllability matrix given in (9); and the matrix N_k^p is a matrix solely composed of the scheduling sequence; i.e.

$$N_k^p = \begin{bmatrix} P_{p|k} & & & \\ & P_{p|k+1} & & \\ & & \ddots & \\ & & & P_{p|k+p-1} \end{bmatrix},$$

$$P_{p|k} = \mu_{k+p-1} \otimes \cdots \otimes \mu_k \otimes I_{r+l}.$$

where $\otimes$ denoted the Kronecker product defined in [8].
The key approximation in this algorithm is that we assume that $\phi_{j,k} \approx 0$ for all $j \geq p$. This approximation is commonly used in the LTI identification literature (*e.g.* N4SID [9], SSARX [10], PBSID [11]). For finite p, this approximation might result in biased estimates. However, it can be shown that, if the system in (4)-(5) is uniformly exponentially stable, the approximation error can be made arbitrarily small by

choosing p large enough [12]. With this approximation, the state x_{k+p} is approximated by:

$$x_{k+p} \approx \mathscr{K}^p N_k^p \bar{z}_k^p. \tag{11}$$

The input-output behavior is now approximately given by:

$$y_{k+p} \approx C\mathscr{K}^p N_k^p \bar{z}_k^p + Du_{k+p} + e_{k+p} := y_{k+p}^{(p)}. \tag{12}$$

Now we define the stacked matrices U, Y, and Z:

$$U = \begin{bmatrix} u_{p+1}, & \cdots, & u_N \end{bmatrix}, \tag{13}$$

$$Y = \begin{bmatrix} y_{p+1}, & \cdots, & y_N \end{bmatrix}, \tag{14}$$

$$Z = \begin{bmatrix} N_1^p \bar{z}_1^p, & \cdots, & N_{N-p}^p \bar{z}_{N-p}^p \end{bmatrix}. \tag{15}$$

If $\begin{bmatrix} Z^T, & U^T \end{bmatrix}^T$ has full row rank, $C\mathscr{K}^p$ and D can be estimated by solving the following least squares problem:

$$\min_{C\mathscr{K}^p, D} ||Y - C\mathscr{K}^p Z - DU||_F^2, \tag{16}$$

where $||\cdots||_F$ represents the Frobenius norm [13]. For finite p this linear problem will be biased due to the approximation made in (11). In LTI literature a number of papers appeared that study the effect of the window size; although they prove one of the asymptotic properties of the algorithms (if $p \to \infty$ the bias disappears), it is hard to quantify the effect for finite p [12], [14], [11].

D. Observability matrix times controllability matrix

The algorithm we described, can be seen as the LPV counterpart of the PBSID$_{\text{opt}}$ algorithm [11], [14]. In the PBSID$_{\text{opt}}$ algorithm, the LTI equivalent of $C\mathscr{K}^p$ is estimated to construct, approximately, the extended observability matrix times the extended controllability matrix (see (4.11) in [11]). In the LPV case, a similar approach can be followed. However, in this case we construct, approximately, the product between the extended observability matrix of the first local model, given in (8), and the extended LPV controllability matrix, given in (9). This matrix product can then be written as follows:

$$\Gamma^p \mathscr{K}^p \approx \begin{bmatrix} C\mathscr{L}_p & C\mathscr{L}_{p-1} & \cdots & C\mathscr{L}_1 \\ 0 & C\tilde{A}^{(1)}\mathscr{L}_{p-1} & \cdots & C\tilde{A}^{(1)}\mathscr{L}_1 \\ & & \ddots & \\ 0 & & & C\left(\tilde{A}^{(1)}\right)^{p-1}\mathscr{L}_1 \end{bmatrix}. \tag{17}$$

The zeros appear in this equation based on the approximation that $\phi_{j,k} \approx 0$ for all $j \geq p$ (similarly as in the LTI PBSID$_{\text{opt}}$ [11])[1]. Equation (17) can therefore be constructed from:

$$C\mathscr{K}^p = \begin{bmatrix} C\mathscr{L}_p, & C\mathscr{L}_{p-1}, & \cdots & C\mathscr{L}_1 \end{bmatrix},$$

estimated from (16).

[1]Without this approximation the algorithm becomes more complex and computationally intensive, see the LPV-PBSID algorithm in [6].

E. Estimation of the state sequence

Now we can compute $\Gamma^p \mathscr{K}^p Z$, which equals by definition the extended observability matrix times the state sequence, $\Gamma^p X$. Under the assumptions stated in Section II-B that X and Γ^p both have full rank and that $p\ell > n$, we can estimate the state sequence and the order of the system based on a rank revealing Singular Value Decomposition (SVD). We will use the following SVD:

$$\widehat{\Gamma^p \mathscr{K}^p} Z = \begin{bmatrix} \mathscr{U} & \mathscr{U}_\perp \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} V \\ V_\perp \end{bmatrix}, \tag{18}$$

where Σ_n is the diagonal matrix containing the n largest singular values; and V is the corresponding row space. Note that we can find the system order by detecting a gap between the singular values [15]. The state is now estimated by:

$$\widehat{X} = \Sigma_n V. \tag{19}$$

It is well known that once the state, input, output, and scheduling sequence are known, the system matrices can be estimated [16]. First, we use (2), which is now a linear relation in C and D, and where e_k represents white noise. From this equation an estimate can be found of the C and D matrix, as well as the noise sequence. The estimated noise sequence is used to transform (1) into a linear expression depending on $A^{(i)}$, $B^{(i)}$, and $K^{(i)}$. Consequently, all system matrices can be estimated.

F. Kernel method

The method described above suffers from a curse of dimensionality, as $\tilde{q}$, the number of rows in data matrix Z, grows exponentially with the size of the past window, see (10). An effective way to reduce the dimensionality of the parameter estimation problem is to use the kernel method for the LPV PBSID$_{\text{opt}}$ scheme, presented in [4],[6]. It assumes that the solution to the estimation problem (16) is of the form:

$$\begin{bmatrix} C\mathscr{K}^p, & D \end{bmatrix} = \alpha \begin{bmatrix} Z^T & U^T \end{bmatrix}. \tag{20}$$

This results into a dual to the estimation problem (16), given by:

$$\min_{\alpha} \left(\|Y - \alpha\Phi\|_F^2 \right) \tag{21}$$

where $\Phi = Z^T Z + U^T U \in \mathbb{R}^{(N-p)\times(N-p)}$. From the solution of the above problem, the matrix $\Gamma^p K^p Z$ can be found through a linear mapping $L(\alpha)$, defined as:

$$\Gamma^p K^p Z = \begin{bmatrix} \alpha \sum_{j=1}^{p} Z_{1,j}^T Z_{1,j} \\ \alpha \sum_{j=2}^{p} Z_{2,j}^T Z_{1,j} \\ \vdots \\ \alpha \sum_{j=p}^{p} Z_{p,j}^T Z_{1,j} \end{bmatrix} \triangleq L(\alpha), \tag{22}$$

with:

$$Z_{i,j} = \begin{bmatrix} P_{p-j+1|j-i+1} z_{j-i+1}, & \cdots, & P_{p-j+1|N+j-i} z_{N+j-i}. \end{bmatrix}$$

When $\tilde{q} > N - p$, using the kernel method improves the numerical efficiency of the LPV subspace identification algorithm, as it reduces the size of the data matrices. In order to fully exploit the improved numerical efficiency of the dual problem, the kernel matrices $Z_{1,j}^T Z_{i,j}$ are to be constructed directly from data without first constructing $Z_{i,j}$. Before giving the expressions that show how to do this, first we introduce a notation $\odot$ for the Hadamard (or entry-wise) product, [8]:

$$(A \odot B)^{(i.j)} = A^{(i.j)} \cdot B^{(i.j,)}.$$

where indices $(i.j)$ denote the different elements in the A and B matrices. The Hadamard product of a sequence of matrices is denoted by:

$$\overset{j}{\underset{v=i}{\mathrm{H}}} (M_v) = M_i \odot M_{i+1} \odot \ldots \odot M_j. \tag{23}$$

Also, we define the row vector $\tilde{N} = \begin{bmatrix} 1, & \cdots, & N-p \end{bmatrix}$, and the notation:

$$\mu_{\tilde{N}+i} = \begin{bmatrix} \mu_{i+1}, & \ldots, & \mu_{i+N-p} \end{bmatrix}$$

and similarly for $z_{\tilde{N}+i}$. It is then then found that the kernel matrices needed to formulate problem (21) and construct (22), can be calculated from data through:

$$Z_{i,j}^T Z_{1,j} = \overset{p-j}{\underset{v=0}{\mathrm{H}}} \left(\mu_{\tilde{N}+v+j-i}^T \mu_{\tilde{N}+v+j-1} \right) \odot \left(z_{\tilde{N}+j-i}^T z_{\tilde{N}+j-1} \right), \tag{24}$$

$$Z^T Z = \sum_{j=1}^{p} Z_{1,j}^T Z_{1,j}. \tag{25}$$

For derivations of the kernel method we refer to [4],[6].

III. Regularization methods

The parameter estimation problem (21) in the kernel method for LPV subspace identification is often ill-posed, as it estimates a large number of parameters from a reduced set of data points in order to reduce memory requirements. This makes the solution α sensitive to measurement error in the data, or to numerical error in the calculation of the matrix Φ. Regularization techniques 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.

A. Conventional 2-norm regularization

Conventionally, Tikhonov or TSVD regularization techniques are employed in LPV subspace identification [17]. These techniques add to the estimation problem the requirement that the 2-norm of the solution is small. The underlying assumption is that a solution with small elements is less sensitive to noise.

1) Tikhonov regularization: In Tikhonov regularization, a term is added to the parameter estimation problem, which yields preference to a solution with a small 2-norm:

$$\min_{\alpha} \left(\|Y - \alpha\Phi\|_F^2 + \lambda^2 \|\alpha\|_F^2 \right), \tag{26}$$

with regularization parameter $\lambda > 0$ defining the trade-off between the fit of the model and the variation of the solution. The basic idea is that λ can be chosen such that a well-conditioned matrix $\Phi + \lambda^2 I$ is formed, that replaces the ill-conditioned matrix Φ in the inversion problem (21).

2) TSVD regularization: Using a Truncated Singular Value Decomposition (TSVD), one can try to obtain a well-conditioned full-rank matrix Φ_λ replacing Φ in (21), which only includes the information in Φ corresponding to the largest singular values:

$$\Phi = \begin{bmatrix} \mathscr{U}_\lambda & \mathscr{U}_\perp \end{bmatrix} \begin{bmatrix} \Sigma_\lambda & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} V_\lambda \\ V_\perp \end{bmatrix}, \; \Phi_\lambda = \mathscr{U}_\lambda \Sigma_\lambda V_\lambda, \tag{27}$$

where Σ_λ is a diagonal matrix containing the $\lambda \in \mathbb{N}$ largest singular values of Φ. The minimum 2-norm regularized solution is given by:

$$\hat{\alpha} = Y V_\lambda^T \Sigma_\lambda^{-1} \mathscr{U}_\lambda^T.$$

As in the Tikhonov regularization, regularization parameter λ balances the bias error and variance error of the identified model.

B. Nuclear norm regularization

The nuclear norm regularization method tries to incorporate requirements in the estimation problem (21) that will result in a minimum order model later on in the LPV PBSID$_{\text{opt}}$ scheme. As the state sequence is found from an SVD of $\Gamma^p \mathscr{K}^p Z$, this translates into a minimum rank requirement on this matrix (see section II-E), resulting in the following optimization problem:

$$\min_{\alpha} \mathrm{Rank}\left(L(\alpha)\right) \text{ s.t. } \min \|Y - \alpha\Phi\|_F^2. \tag{28}$$

The above problem is a difficult nonconvex problem in general. Therefore, a convex heuristic for problem (28) is used:

$$\min_{\alpha} \left(\lambda \|Y - \alpha\Phi\|_F^2 + \|L(\alpha)\|_* \right), \tag{29}$$

where $\|\cdot\|_*$ denotes the nuclear norm, the sum of singular values of a matrix. Regularization parameter λ is used to change the balance between the minimum order requirement and the prediction error norm. The above problem is solved using the interior-point convex optimization techniques presented in [18], which solves problems of the form:

$$\min_{\chi} \left(\|L(\chi)\|_* + \frac{1}{2} \chi^T \mathcal{C} \chi + \mathcal{d}^T \chi \right), \tag{30}$$

with vector $\chi \in \mathbb{R}^n$ as optimization variable, matrix $\mathcal{C} \in \mathbb{R}^{n \times n}$ and vector $\mathcal{d} \in \mathbb{R}^n$ being weighting factors, and $L(\chi) \in \mathbb{R}^{n \times n}$ being a linear matrix valued function:

$$L(\chi) = \begin{bmatrix} L_1\chi, & L_2\chi, & \cdots, & L_q\chi \end{bmatrix} \text{ with } L_i \in \mathbb{R}^{p \times n},$$

with $p > q$. Vectorization of the optimization variable α is used to rewrite problem (29) in the form (30):

$$\chi = \mathrm{vec}\left(\alpha^T\right), \tag{31}$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operation which stacks the columns of a matrix as a vector. This results in weighting factors:

$$\mathcal{C} = 2\lambda\left(I_l \otimes \Phi^T\Phi\right), \; d = -2\lambda\,\mathrm{vec}\left(\Phi^T Y^T\right), \tag{32}$$

where I_l denotes the $l \times l$ identity matrix. From (22) follows the definition of the matrices $\{L_i\}_{i=1}^{q}$:

$$\begin{bmatrix} L_{lk+1} \\ \hline L_{lk+2} \\ \hline \vdots \\ \hline L_{lk+1} \end{bmatrix} = \begin{bmatrix} Z_k & 0 & \cdots & 0 \\ \hline 0 & Z_k & & 0 \\ \hline \vdots & & \ddots & \vdots \\ \hline 0 & \cdots & & Z_k \end{bmatrix}, \tag{33}$$

for $k = 0, \ldots, p-1$ (hence $q = lp$), with $Z_k = \sum_{j=k}^{p} Z_{k,j}^T Z_{1,j}$.

IV. Simulation example

This section presents a proof of concept for the nuclear norm regularization method, by showing that it can outperform the conventional regularization methods in the identification of a second order LPV system with $l = r = 2$, with system matrices:

$$[A_1|A_2] = \left[\begin{array}{cc|cc} 0 & -.07 & 0.002 & 0 \\ -6.5 & -0.5 & -0.01 & 0.5 \end{array}\right],$$

$$[B_1|B_2] = \left[\begin{array}{cc|cc} -2.2 & -1 & 0 & 0 \\ -6.6 & -2 & 0 & 0 \end{array}\right], C = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix},$$

$$D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \; [K_1|K_2] = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right].$$

To identify the system, the system is excited with the following scheduling signal:

$$\mu_k = \left[1, \quad \cos(2\pi k/5) + \sin(\pi k/5)\right]^T,$$

and a stochastic Gaussian input signal u_k with variance $\mathrm{var}(u_k) = 1$. A Gaussian noise signal e_k is used with a Signal to Noise Ratio (SNR)[2] of 75dB. The data u_k, y_k, and μ_k is collected and used in the LPV PBSID$_{\mathrm{opt}}$ algorithm to find the system matrices. We limit the data set to $N = 80$ samples. It is found that the best results are obtained using $p = 5$. Since $\tilde{q}$ exceeds $N - p$, the kernel method is used. Regularization is used to limit the effect of the noise on the estimated parameters. As a measure for performance of the algorithm, we use the value of the Variance-Accounted-For (VAF) on a data set different from the identification data set. In this validation data set, we add a Gaussian stochastic component with variance 0.25 to the scheduling signal, use a different realization u_k, and use no noise e_k. The VAF value is defined as:

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

[2] $\mathrm{SNR} = 20\log_{10}\frac{\mathrm{var}(y_k)}{\mathrm{var}(e_k)}$

where $\hat{y}_k$ is the estimated system output, and y_k is the true system output. To show the effect of noise on the result, Monte-Carlo simulations with 100 runs are carried out. In each run, different realizations of the input and noise are used. Of the conventional 2-norm regularization techniques, the best results are obtained with TSVD regularization with $\lambda = 37$. A significant improvement is made by using the nuclear norm regularization with $\lambda = 0.0501$, both in terms of a higher mean, and a smaller variance of the VAF, see Figure 1. Also, nuclear norm regularization results in a model with a closer fit to the true system in terms of the eigenvalues of system matrices A_1 and A_2, see Figure 2.

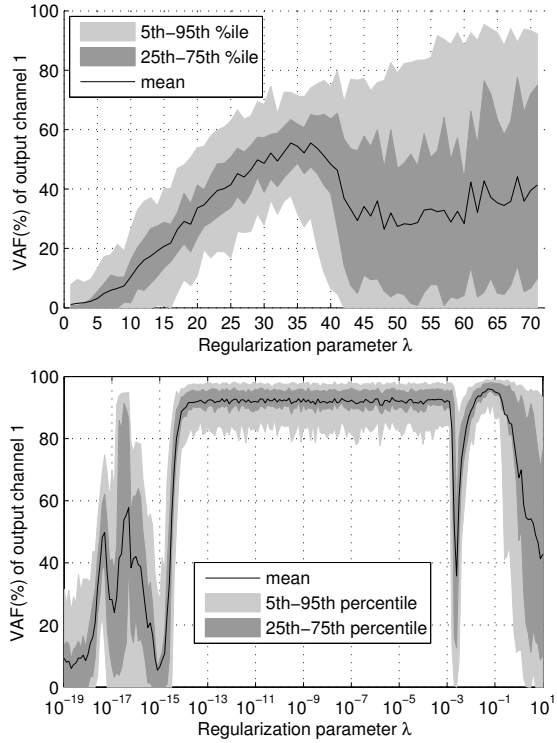

Fig. 1. Performance of LPV PBSID$_{\mathrm{opt}}$ using TSVD regularization (upper) and nuclear norm regularization (lower) for a large range of values of the parameter λ. Performance is expressed as the distribution of the VAF of a validation data set, resulting from a Monte-Carlo simulation with 100 runs. The results are shown for the first output channel, the second output gives similar results.

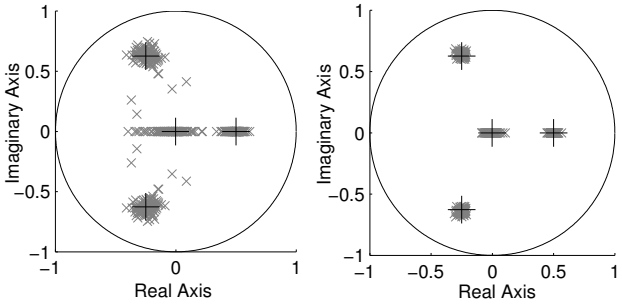

Fig. 2. Eigenvalues of A_1 and A_2 ($\times$) as estimated by LPV PBSID$_{\mathrm{opt}}$ with TSVD regularization (left) and nuclear norm regularization (right). The true eigenvalues are indicated with +.

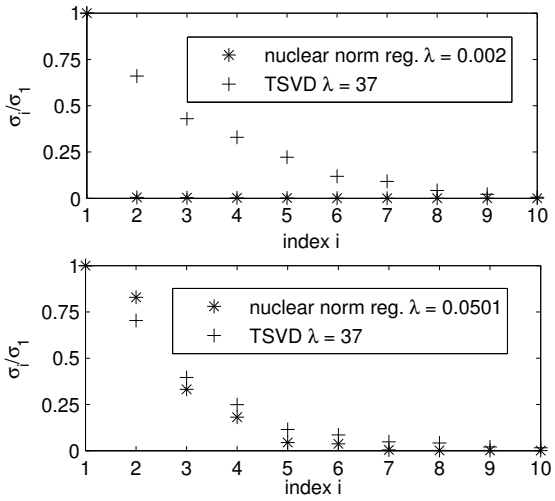

Fig. 3. Normalized singular values of the matrix $\Gamma_p \mathscr{K}_p Z$, resulting from nuclear norm regularization ($*$) for $\lambda = 0.002$ (upper), $\lambda = 0.0501$ (lower), and from TSVD regularization (+) for $\lambda = 37$.

Note in Figure 1 that in nuclear norm regularization the dependence of the performance on the regularization parameter λ is quite complicated. Figure 3 provides more insight by showing the singular values $\{\sigma_i\}_{i=1}^{pl}$ of the matrix $\Gamma_p \mathscr{K}_p Z$ for different values of λ. In general, if we want to accurately fit the data with an n^{th} order model, the gap between the singular values $\{\sigma_1, \ldots, \sigma_n\}$ corresponding to the system dynamics and the other singular values due to noise, needs to be large. Comparing Figures 1 and 3, it can be seen that decreased performance indeed occurs when there is only a small gap between σ_2 and σ_3 (e.g. for $\lambda = 0.002$). By changing λ we are only directly affecting the balance between the sum of singular values and the prediction error. It therefore takes some effort to find a value of λ that results in a matrix $\Gamma_p \mathscr{K}_p Z$ with appropriate rank, e.g. in this case $\lambda = 0.0501$ resulting in a well-defined rank 2 for $\Gamma_p \mathscr{K}_p Z$. For higher order systems, it becomes more difficult to use the nuclear norm regularization to manipulate the parameter estimation problem so that a minimum rank matrix $\Gamma_p \mathscr{K}_p Z$ results.

V. Conclusions

The main contribution of this paper is a novel nuclear norm regularization method for LPV identification with the $\text{PBSID}_{\text{opt}}$ method. The nuclear norm regularization stabilizes the ill-posed parameter estimation problem in the identification scheme by adding additional information on the desired solution to the optimization problem, based on a minimum order requirement for the resulting LPV model. This information consists of a rank requirement on the matrix from which the state sequence is found. In our understanding, this requirement is more meaningful than the minimum 2-norm requirement on the solution of the parameter estimation problem, that are used by the conventional Tikhonov or TSVD regularization methods. By using the nuclear norm as a heuristic for the rank of a matrix in the estimation problem, the problem remains convex. We have shown in a simulation example of the identification of a second order system that the nuclear norm regularization in combination with the kernel method for the LPV $\text{PBSID}_{\text{opt}}$ scheme can outperform Tikhonov and TSVD regularization. The nuclear norm regularization method is useful when a relatively small amount of experimental data is used to identify the model from. In this case the order reducing capabilities of the nuclear norm method can prevent overfitting to noise.

For higher order systems, it becomes more difficult to use the nuclear norm criterion to obtain a state matrix that is both low rank, and results in a well-conditioned system matrix estimation problem later on in the scheme. So far, the nuclear norm method has been proven useful in simulation examples with models up to order 2 or 3, with up to two scheduling parameters.

Note that the subspace identification and nuclear norm regularization scheme presented in this paper, can also be employed for identifying LTI systems, using $\mu_k = 1$.

References

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

[2] P. Apkarian and P. Gahinet. A convex characterization of gain-scheduled H_∞ controllers. *IEEE Transactions on Automatic Control*, 40(5):853–864, 1995.

[3] L. H. Lee and K. Poolla. Identification of linear parameter-varying systems using non-linear programming. *Journal of Dynamic Systems, Measurements and control*, 121(1):71–78, 1999.

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

[5] J. Sjöberg and L. Ljung. Overtraining, regularization and searching for a minimum, with application to neural networks. *International Journal of Control*, 62(6):1391–1407, 1995.

[6] J. W. van Wingerden. *Control of Wind Turbines with 'Smart' Rotors: Proof of Concept & LPV Subspace Identification*. PhD thesis, Delft University of Technology, 2008.

[7] W.J. Rugh. *Linear System Theory*. Upper Sadle River, 1996.

[8] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1994.

[9] P. Van Overschee and B. De Moor. *Subspace Identification for linear systems*. Kluwer Academic Publishers, 1996.

[10] M. Jansson. A new subspace identification method for open and closed loop data. In *Proceedings of the* 16^{th} *IFAC World Congress*, Prague, Czech Republic, 2005.

[11] A. Chiuso. The role of vector auto regressive modeling in predictor based subspace identification. *Automatica*, 43(6):1034–1048, 2007.

[12] T. Knudsen. Consistency analysis of subspace identification methods based on a linear regression approach. *Automatica*, 37(1):81–89, 2001.

[13] G. H. Golub and C. F. van Loan. *Matrix computations*. The John Hopkins University Press, 1996.

[14] A. Chiuso and G. Picci. Consistency analysis of certain closed-loop subspace identification methods. *Automatica*, 41(3):377–391, 2005.

[15] M. Verhaegen and P. Dewilde. Subspace model identification part 1: The output-error state-space model identification class of algorithms. *International Journal of Control*, 56(5):1187–1210, 1992.

[16] M. Lovera. *Subspace identification methods: theory and applications*. PhD thesis, Politecnico Di Milano, 1997.

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

[18] Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256, 2009.