Identification of Linear Parameter-Varying Systems: A Reweighted $\ell_{2,1}$-Norm Regularization Approach

D. Turk [a,*] J. Gillis [a] G. Pipeleers [a] J. Swevers [a]

[a] *KU Leuven, Department of Mechanical Engineering, Member of Flanders Make*
Celestijnenlaan 300 - box 2420, 3001 Leuven, Belgium

Abstract

This paper presents a regularized nonlinear least-squares identification approach for linear parameter-varying (LPV) systems. The objective of the method is, on the one hand, to obtain an LPV model of which the response fits the system measurements as accurately as possible and, on the other hand, to favor models with an as simple as possible dependency on the scheduling parameter. This is accomplished by introducing $\ell_{2,1}$-norm regularization into the nonlinear least-squares problem. The resulting nonsmooth optimization problem is reformulated into a nonlinear second-order cone program and solved using a sequential convex programming approach. Through an iterative reweighting of the regularization, the parameters that do not substantially contribute to the system response are penalized heavily, while the significant parameters remain unaffected or are penalized only slightly. Numerical and experimental validations of the proposed method show a substantial model simplification in comparison with the nonregularized solution, without significantly sacrificing model accuracy.

Key words: Linear Parameter-Varying (LPV) Systems, System Identification, State-space Models, Basis Selection, Regularization

1 Introduction

System identification is a well-established subfield of automatic control that is concerned with deriving mathematical models of dynamic systems based on available input-output data. The dynamic systems of a particular interest in this paper are linear parameter-varying (LPV) systems: nonlinear systems described by a linear

* Corresponding author.
Email address: dora.turk@kuleuven.be (D. Turk).

model with coefficients varying as a function of one or more scheduling parameters. Literature study shows that LPV modeling and control have a clear track record in a wide variety of application areas such as: thermal control systems with different heat sources and temperature probes, active control of vibro-acoustic systems with multiple accelerometers and actuators, control of mechatronic motion systems, control of magnetic bearings, adaptive active vehicle suspension control, (turbofan) engine control, and vehicle drivetrain control [1].

The literature on LPV system identification distinguishes between global and local approaches. The global techniques (e.g. [2], [3]) directly identify an LPV model based on data obtained from experiments where both the input signal and the scheduling parameters are continuously changing. Experiments of this kind are referred to as global experiments. Systems for which the scheduling parameter cannot be fixed during the identification experiment and hence only global approach can be applied are for example: LPV identification of the rotor dynamics of a helicopter [4] and wind turbine [5]. Local identification techniques identify an LPV model based on data obtained from several so-called local experiments, that is, experiments during which the scheduling parameters are fixed. The local identification techniques (e.g. [6], [7]) typically first identify linear time-invariant (LTI) models based on these local experiments, and then interpolate them into a parameter-dependent model. Systems that can be identified with this approach are for example: an overhead crane with a varying cable length and a temperature-dependent vibro-acoustic panel [8], and am industrial pick and place unit (gantry system) [7]. In [9], these two seemingly exclusive approaches are combined.

An important issue in LPV system identification is the determination of an adequate dependency of the model on the scheduling parameter(s), called model selection. A complex dependency complicates consequent LPV control synthesis and analysis. Model selection is recognized as being the most critical step in any system identification procedure (see [10], Chapter 11), and as especially challenging in the LPV case if no prior knowledge on the scheduling parameter dependency of the system model is available.

The traditional approach to model selection in system identification is to implement an iterative procedure where in each iteration a candidate model is selected, the model parameters are estimated and the model is evaluated using criteria that try to find a good balance between model accuracy, model parameter uncertainty and model complexity, e.g. Akaike's criterion (AIC) [11] or minimum description complexity (MDC) [12].

A more recent approach to model selection is to combine it with parameter estimation by augmenting the parameter estimation objective with a penalty such that simpler models are favored. This well-known strategy in system identification and machine learning is called regularization. The most common penalty functions are ℓ_2 norm (Tikhonov regularization or ridge regression [13]), ℓ_1 norm (LASSO [14] or basis pursuit), and their combination (elastic net regularization [15]). The advantage of the ℓ_2 norm is that it is differentiable, while the ℓ_1 norm induces sparsity.

The survey given in [16] explores connections between system identification and machine learning, also brought up in [17] and [18].

Regularization is also a common approach in LPV system identification. In [19], an ℓ_1 sparse estimator approach called SPARSEVA (SPARSe Estimator based on VAlidation criterion) method [20] for ARX models is extended to the LPV case and compared to the Non-Negative Garrote approach (NNG), [21], [22], and [23]. The SPARSEVA estimator is a variant of the LASSO estimator (ℓ_1 penalized least-squares) which does not require tuning of regularization parameters, while the NNG is based on a penalization of the least-squares solution by attaching weights to it. In [24], a Basis Pursuit DeNoising (BPDN) sparse estimation approach is presented, which is a regularization technique in a predictor-based subspace state-space model identification framework.

This paper presents a nonlinear least-squares LPV system identification method with $\ell_{2,1}$-norm regularization. The method directly identifies state-space models, as they are preferred in control synthesis over the input-output models, and as the transformation from input-output models to their state-space equivalents is complex and sometimes limiting [25]. Through a reweighted $\ell_{2,1}$-norm regularization, an automatic reduction of model structure complexity is accomplished by discarding redundant basis function dependencies from the state-space matrices. The choice of the norm is motivated by the appealing properties the $\ell_{2,1}$-norm regularization has when it comes to selecting a subset of relevant features in machine learning [26], [27]. Furthermore, applying an iterative reweighting scheme, as shown in [28], [29], [30], improves the effectiveness of the sparsity inducing property of the ℓ_1 norm. Due to its noonsmooth nature, the obtained optimization problem is hard to solve, and the way proposed in this paper is to reformulate it into a nonlinear second-order cone program (NSOCP) and solve it by applying the sequential convex programming approach. To keep the focus mainly on the effects of the proposed regularization, only local frequency domain data are considered in this paper. Adaptation to account for any combination of global and local, time and frequency domain data is straightforward.

The paper is organized as follows. Section 2 introduces the chosen LPV model structure. In section 3, the reweighted $\ell_{2,1}$-norm regularization is developed and reformulated into conic constraints by introducing auxiliary variables. The same section discusses the process of solving the presented optimization problem and gives instructions for its implementation in the form of a pseudo code. In section 4 the proposed approach is numerically validated on a flexible gantry system, whereas its experimental validation on an XY-motion system is given in section 5. The obtained validation results form the bottom line for the conclusions conveyed in section 6.

2 LPV model structure

In this paper we focus on the following fully parameterized discrete-time state-space model:

$$\begin{cases} x(t+1) = \mathscr{A}(p(t))x(t) + \mathscr{B}(p(t))u(t) \\ y(t) = \mathscr{C}(p(t))x(t) + \mathscr{D}(p(t))u(t), \end{cases} \tag{1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^r$, $y(t) \in \mathbb{R}^l$, and $p(t) \in \mathbb{R}^{N_p}$, are respectively the state vector, the input vector, the output vector, and the scheduling parameter vector, at time instance t. The state-space matrices of the introduced model are parameter-dependent:

$$\mathscr{X}(p(t)) = X^{(0)} + \sum_{i=1}^{N_b} X^{(i)} \psi_i(p(t)), \tag{2}$$

$\forall (\mathscr{X}, X) \in \{(\mathscr{A}, A), (\mathscr{B}, B), (\mathscr{C}, C), (\mathscr{D}, D)\}$; $A^{(i)} \in \mathbb{R}^{n \times n}$, $B^{(i)} \in \mathbb{R}^{n \times r}$, $C^{(i)} \in \mathbb{R}^{l \times n}$, $D^{(i)} \in \mathbb{R}^{l \times r}$, $\forall i = 0, 1, ..., N_b$, where N_b is the number of basis functions ψ_i of the scheduling parameter $p(t)$ that are employed for the parameterization. No particular choice of basis functions is here assumed. Since the identification method proposed in this paper considers only local data, the basis functions ψ_i are considered to depend only on the instantaneous time values of the scheduling parameter $p(t)$, which excludes dynamic dependency. For simplicity of notation all model matrices $\{\mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D}\}$ are assumed to have the same number of basis functions. This assumption can, however, be relaxed without loss of generality.

Let the model parameters be stacked in a vector θ as follows:

$$\theta_{X^{(i)}} = vec\left(X^{(i)}\right), \quad \forall X \in \{A, B, C, D\}, \quad \forall i = 0, ..., N_b, \tag{3}$$

$$\theta_X = [\theta_{X^{(0)}}^T \quad \theta_{X^{(1)}}^T \quad ... \quad \theta_{X^{(N_b)}}^T]^T, \quad \forall X \in \{A, B, C, D\}, \tag{4}$$

$$\theta = [\theta_A^T \quad \theta_B^T \quad \theta_C^T \quad \theta_D^T]^T. \tag{5}$$

3 LPV model identification

This section discusses the reweighted $\ell_{2,1}$-norm regularization approach for LPV system identification. It starts with the description of the frequency domain data available for identification in section 3.1, introduces the nonlinear least-squares approach in section 3.2, which provides foundations for the regularization algorithm disclosed in section 3.3.

3.1 Identification data

Assume that N local experiments have been performed, each providing a frequency domain data set and corresponding scheduling parameter value. The frequency domain data set is in this paper assumed to be the frequency response function (FRF) matrix of the system measured at a specified set of frequencies $F^{[q]}$, for a fixed value of the scheduling parameter $p^{[q]}$, that is:

$$\{G_{\mathrm{m}}^{[q]}(\omega_k)\},\ k = 1,\ ...,\ F^{[q]},\ q = 1,\ ...,\ N. \tag{6}$$

In general, the measured FRFs contain stochastic components originating from various noise sources and system nonlinearities, of which the characteristics are discussed in [31] and [32] respectively. Stochastic frequency domain identification methods can take these characteristics into account [33].

3.2 The nonlinear least-squares approach

In parameter estimation, a commonly used fitness-criterion is the weighted squared error:

$$V_{\mathrm{NLS}}(\theta) = \varepsilon(\theta)^{\mathrm{H}} W \varepsilon(\theta), \tag{7}$$

where $\varepsilon(\theta)$ denotes the difference between the model response and the corresponding measured system response, $\varepsilon(\theta)^{\mathrm{H}}$ is the conjugate transpose of $\varepsilon(\theta)$, and W is a positive definite weighting matrix. Stochastic identification approaches, e.g. maximum likelihood and generalized total least squares, use information about the characteristics of the stochastic components to determine this weighting matrix [33]. Deterministic identification approaches use other arguments, e.g. common sense, to select these weights. For the particular cases considered in section 4 and section 5, estimates of the sample total variance of the measured FRFs are used as weights, yielding that the nonlinear least-squares estimate corresponds to the maximum likelihood estimate [31].

The frequency domain model error vector for each $q = 1,\ ...,\ N$ equals

$$\varepsilon^{[q]}(\theta) = [\varepsilon^{[q]}(\theta,\omega_1)^T \quad \varepsilon^{[q]}(\theta,\omega_2)^T \quad ... \quad \varepsilon^{[q]}(\theta,\omega_{F^{[q]}})^T]^T \tag{8}$$

and consists of:

$$\varepsilon^{[q]}(\theta,\omega_k) = vec\left(G^{[q]}(\theta,\omega_k) - G_{\mathrm{m}}^{[q]}(\omega_k)\right),\quad k = 1,\ ...,\ F^{[q]}, \tag{9}$$

where $G^{[q]}(\theta,\omega_k)$ is the model frequency response function matrix. Eventually, all

data are stacked in one vector, i.e.:

$$\varepsilon(\theta) = [\varepsilon^{[1]}(\theta)^T \quad \varepsilon^{[2]}(\theta)^T \quad \dots \quad \varepsilon^{[N]}(\theta)^T]^T. \tag{10}$$

The nonlinear least-squares parameter estimate is given by

$$\hat{\theta} = \arg\min_{\theta} V_{\mathrm{NLS}} \tag{11}$$

and can be calculated via the Levenberg-Marquardt algorithm.

3.3 A reweighted $\ell_{2,1}$-norm regularization approach

As mentioned in the introduction, this paper follows a regularization approach to select the most appropriate scheduling parameter dependency. This approach assumes a predefined set of basis functions—based on physical insights or guess—to describe the scheduling parameter dependency of the initial, purposely too complex model. Throughout the identification algorithm, the model complexity is iteratively reduced by discarding basis function dependencies from the state-space matrices where they are assessed as redundant. This strategy is achieved via the reweighted $\ell_{2,1}$-norm regularization approach, thoroughly explained in what follows.

3.3.1 $\ell_{2,1}$-norm regularization

The $\ell_{2,1}$-norm of an arbitrary matrix $M \in \mathbb{R}^{m\times n}$ is defined as

$$\|M\|_{2,1} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{m} M_{i,j}^2} = \sum_{j=1}^{n} \|M_{1:m,\,j}\|_2, \tag{12}$$

and has a desirable "grouping" property [26], [27]: in case (12) is added to an optimization problem where all elements of M are optimization variables, the optimization favors solutions M with as many zero columns as possible. In order to group model parameters associated with the same basis functions, the concept of columns in (12) is in this paper extended to matrices of model parameters $\left(X^{(i)}\right)$; these matrices are vectorized and the values of the resulting Euclidean norm are summed up into the regularization objective

$$V_{\mathrm{reg}}^{\theta_X} = \sum_{i=1}^{N_b} \left\|\theta_{X^{(i)}}\right\|_2, \quad \forall X \in \{A,B,C,D\}, \tag{13}$$

that is added to the model-fitness criterion (7).

To effectively reduce the model complexity without discarding parameter sets that significantly contribute to the model response, we chose to add penalization weights,

of which the explicit definition is given at the end of this section. Combining the nonlinear least-squares approach (section 3.2), with the concept of weighted $\ell_{2,1}$-norm regularization yields the following optimization problem:

$$\underset{\theta}{\text{minimize}}\ V_{\text{NLS}}(\theta)+\gamma\sum_{i=1}^{N_b}\Bigg(\phi\left(\hat{\theta}_{A^{(i)}}\right)\left\|\theta_{A^{(i)}}\right\|_2+\phi\left(\hat{\theta}_{B^{(i)}}\right)\left\|\theta_{B^{(i)}}\right\|_2+ \\ +\phi\left(\hat{\theta}_{C^{(i)}}\right)\left\|\theta_{C^{(i)}}\right\|_2+\phi\left(\hat{\theta}_{D^{(i)}}\right)\left\|\theta_{D^{(i)}}\right\|_2\Bigg), \tag{14}$$

where γ is a scalar, the value of which determines the rigorousness of the penalization, that is, the importance of the model simplicity with regard to the accuracy $V_{\text{NLS}}(\theta)$, and ϕ is an expression for the penalization weights, defined in section 3.3.3. It is important to notice that the arguments $\hat{\theta}$ are not optimization variables. The choice of $\hat{\theta}$ is discussed in section 3.3.3.

The objective function (14) is nonsmooth, which makes the optimization problem challenging to solve. The way to tackle it chosen in this paper is to reformulate the $\ell_{2,1}$ regularization objective into second-order cone constraints by including auxiliary optimization variables (s), which results in a nonlinear second-order cone programming (NSOCP) problem, [34]:

$$\begin{aligned}\underset{\theta,\,s}{\text{minimize}}\ V_{\text{NLS}}(\theta)+\gamma\sum_{i=1}^{N_b}\Bigg(&\phi\left(\hat{\theta}_{A^{(i)}}\right)s_{A^{(i)}}+\phi\left(\hat{\theta}_{B^{(i)}}\right)s_{B^{(i)}}+ \\ &+\phi\left(\hat{\theta}_{C^{(i)}}\right)s_{C^{(i)}}+\phi\left(\hat{\theta}_{D^{(i)}}\right)s_{D^{(i)}}\Bigg), \\ \text{subject to}\quad &\left\|\theta_{A^{(i)}}\right\|_2\le s_{A^{(i)}}, \\ &\left\|\theta_{B^{(i)}}\right\|_2\le s_{B^{(i)}}, \\ &\left\|\theta_{C^{(i)}}\right\|_2\le s_{C^{(i)}}, \\ &\left\|\theta_{D^{(i)}}\right\|_2\le s_{D^{(i)}}. \\ &i=1,\ldots,N_b.\end{aligned} \tag{15}$$

In that way not only is it possible to handle the nonsmoothness, but also the Hessian of the Lagrangian coincides with the one of the nonlinear least-squares problem (7), which is handy. Motivated by the Levenberg-Marquardt algorithm, we decided to solve (15) by "cutting" it into convex SOCP subproblems. The objective function of the SOCP subproblem is a convex quadratic function, while the second-order

cone constraints remain as linear cone constraints, that is:

$$\begin{aligned}
\underset{\Delta\theta,\,\Delta s}{\text{minimize}}\ \ & \nabla V_{\text{NLS}}\left(\theta^{(k)}\right)^{T}\Delta\theta+\frac{1}{2}\Delta\theta^{T}M^{(k)}\Delta\theta+\gamma\sum_{i=1}^{N_b}\Bigg(\phi\left(\hat{\theta}_{A^{(i)}}\right)\Delta s_{A^{(i)}}+ \\
& +\phi\left(\hat{\theta}_{B^{(i)}}\right)\Delta s_{B^{(i)}}+\phi\left(\hat{\theta}_{C^{(i)}}\right)\Delta s_{C^{(i)}}+\phi\left(\hat{\theta}_{D^{(i)}}\right)\Delta s_{D^{(i)}}\Bigg) \\
\text{subject to}\quad & \left\|\theta^{(k)}_{A^{(i)}}+\Delta\theta_{A^{(i)}}\right\|_2\leq s^{(k)}_{A^{(i)}}+\Delta s_{A^{(i)}}, \\
& \left\|\theta^{(k)}_{B^{(i)}}+\Delta\theta_{B^{(i)}}\right\|_2\leq s^{(k)}_{B^{(i)}}+\Delta s_{B^{(i)}}, \\
& \left\|\theta^{(k)}_{C^{(i)}}+\Delta\theta_{C^{(i)}}\right\|_2\leq s^{(k)}_{C^{(i)}}+\Delta s_{C^{(i)}}, \\
& \left\|\theta^{(k)}_{D^{(i)}}+\Delta\theta_{D^{(i)}}\right\|_2\leq s^{(k)}_{D^{(k)}}+\Delta s_{D^{(i)}}, \\
& i=1,...,N_b. \qquad (16)
\end{aligned}$$

In (16),

$$\Delta s=[\Delta s_A^T\ \ \Delta s_B^T\ \ \Delta s_C^T\ \ \Delta s_D^T]^T, \qquad (17)$$

and $M^{(k)}$ is an approximation of the Hessian defined as in the original Levenberg-Marquardt algorithm:

$$M^{(k)}=2\left(\nabla V_{\text{NLS}}\left(\theta^{(k)}\right)^{T}\nabla V_{\text{NLS}}\left(\theta^{(k)}\right)+\lambda^{2}\text{diag}\left(\nabla V_{\text{NLS}}\left(\theta^{(k)}\right)^{T}\nabla V_{\text{NLS}}\left(\theta^{(k)}\right)\right)\right), \qquad (18)$$

where $\nabla V_{\text{NLS}}\left(\theta^{(k)}\right)$ is the Jacobian matrix of V_{NLS}, and λ is the damping parameter. For solving (16), we used YALMIP [35] with the Embedded Conic Solver (ECOS) [36]. What remains is to update the estimates of the model parameters, at the end of each iteration, which is done as follows:

$$\theta^{(k+1)}_{X^{(i)}}=\theta^{(k)}_{X^{(i)}}+\Delta\theta_{X^{(i)}},\quad \forall X\in\{A,B,C,D\},\quad \forall i=0,1,...,N_b, \qquad (19)$$

where $\Delta\theta=[\Delta\theta_A^T\ \ \Delta\theta_B^T\ \ \Delta\theta_C^T\ \ \Delta\theta_D^T]^T$.

The algorithm is stopped once the step size $\|\Delta\theta\|_2$ is smaller than a specified threshold or when the obtained model accuracy is sufficient. A somewhat similar approach for solving nonlinear second-order cone programs is proposed in [37].

3.3.2 Initial estimates

Just like the nonlinear least-squares approach, the proposed reweighted $\ell_{2,1}$ minimization approach requires estimates of the model parameters to start from. In both

the numerical (section 4) and experimental (section 5) case of validation, these initial estimates θ_0 are generated using the State-space Model Interpolation of Local Estimates (SMILE) technique, [38].

3.3.3 Selection of the penalization weights

This section discusses several alternative formulations of the penalization weights. They are selected such that model parameters not contributing substantially to the system response are penalized.

The importance of a subset of model parameters $\theta_{X^{(i)}}$ can be ranked according to its magnitude. Consequently, penalization weights inversely proportional to the magnitude of some valid estimates of these model parameters $\hat{\theta}_{X^{(i)}}$—the process of obtaining them is discussed below in the text—arise as a natural choice:

$$\phi\left(\hat{\theta}_{X^{(i)}}\right) = \begin{cases} \left(\left\|\hat{\theta}_{X^{(i)}}\right\|_2\right)^{-1} & \text{if } \left\|\hat{\theta}_{X^{(i)}}\right\|_2 \neq 0 \\ \infty & \text{otherwise} \end{cases}, \tag{20}$$

$\forall X \in \{A,B,C,D\}$, $\forall i = 1,...,N_b$. For numerical stability a small tuning parameter ε is added, which yields:

$$\phi\left(\hat{\theta}_{X^{(i)}}\right) = \left(\left\|\hat{\theta}_{X^{(i)}}\right\|_2 + \varepsilon\right)^{-1}, \tag{21}$$

$\forall X \in \{A,B,C,D\}$, $\forall i = 1,...,N_b$. When $\left\|\hat{\theta}_{X^{(i)}}\right\|_2$ is large, ε is negligible and (21) is small. When $\left\|\hat{\theta}_{X^{(i)}}\right\|_2$ has a value close to zero, ε starts playing a role since its inverse will determine the value of (21), which will in this case be large. By tuning the value of ε, one tunes the intensity of the penalization strategy.

However, by revisiting the expressions for the state-space matrices of the chosen LPV model structure (1) - (2), one can notice that the contribution of a subset of parameters to a state-space matrix and consequently to the overall system response might be better assessed if expressed through the product of these parameters with the basis function they are assigned to. Since the value $\psi_i(p(t))$ varies with the scheduling parameter $p(t)$, the infinity norm, that is, the maximal absolute value of ψ_i within the given scheduling parameter range, is here considered:

$$\phi\left(\hat{\theta}_{X^{(i)}}\right) = \left(\left\|\hat{\theta}_{X^{(i)}}\right\|_2 \left\|\psi_i\right\|_\infty + \varepsilon\right)^{-1}, \tag{22}$$

$\forall X \in \{A,B,C,D\}$, $\forall i = 1,...,N_b$.

A step further in assessing the importance of a particular subset of model parameters is comparing their contribution to a state-space matrix against the LTI contri-

bution, which yields:

$$\phi\left(\hat{\theta}_{X^{(i)}}\right) = \left(\frac{\left\|\hat{\theta}_{X^{(i)}}\right\|_2 \left\|\psi_i\right\|_\infty}{\left\|\hat{\theta}_{X^{(0)}}\right\|_2} + \varepsilon\right)^{-1}, \tag{23}$$

$\forall X \in \{A,B,C,D\},\ \forall i = 1,...,N_b.$

Moreover, by rescaling and restructuring the optimization problem into

$$\begin{aligned}
\underset{\theta,\, s}{\text{minimize}}\ V_{\text{NLS}}(\theta) + \gamma \sum_{i=1}^{N_b} \Bigg(& \phi\left(\hat{\theta}_{A^{(i)}}\right) s_{A^{(i)}} + \phi\left(\hat{\theta}_{B^{(i)}}\right) s_{B^{(i)}} + \\
& + \phi\left(\hat{\theta}_{C^{(i)}}\right) s_{C^{(i)}} + \phi\left(\hat{\theta}_{D^{(i)}}\right) s_{D^{(i)}} \Bigg), \\
\text{subject to} \quad & \varphi\left(\hat{\theta}_{A^{(i)}}\right) \left\|\theta_{A^{(i)}}\right\|_2 \leq s_{A^{(i)}}, \\
& \varphi\left(\hat{\theta}_{B^{(i)}}\right) \left\|\theta_{B^{(i)}}\right\|_2 \leq s_{B^{(i)}}, \\
& \varphi\left(\hat{\theta}_{C^{(i)}}\right) \left\|\theta_{C^{(i)}}\right\|_2 \leq s_{C^{(i)}}, \\
& \varphi\left(\hat{\theta}_{D^{(i)}}\right) \left\|\theta_{D^{(i)}}\right\|_2 \leq s_{D^{(i)}}, \\
& i = 1,...,N_b,
\end{aligned} \tag{24}$$

with

$$\varphi\left(\hat{\theta}_{X^{(i)}}\right) = \frac{\left\|\psi_i\right\|_\infty}{\left\|\hat{\theta}_{X^{(0)}}\right\|_2}, \tag{25}$$

the value of the regularization part of the optimization criterion will converge to the number of nonzero subsets $\theta_{X^{(i)}}$, namely

$$\begin{aligned}
\phi\left(\hat{\theta}_{X^{(i)}}\right) s_{X^{(i)}} \ &\approx\ \frac{\dfrac{\left\|\psi_i\right\|_\infty}{\left\|\hat{\theta}_{X^{(0)}}\right\|_2} \left\|\theta_{X^{(i)}}\right\|_2}{\dfrac{\left\|\hat{\theta}_{X^{(i)}}\right\|_2 \left\|\psi_i\right\|_\infty}{\left\|\hat{\theta}_{X^{(0)}}\right\|_2} + \varepsilon} \\
&\approx\ \begin{cases}
1 & \text{if } \dfrac{\left\|\psi_i\right\|_\infty \left\|\theta_{X^{(i)}}\right\|_2}{\left\|\hat{\theta}_{X^{(0)}}\right\|_2} \gg \varepsilon \\
0 & \text{if } \dfrac{\left\|\psi_i\right\|_\infty \left\|\hat{\theta}_{X^{(i)}}\right\|_2}{\left\|\hat{\theta}_{X^{(0)}}\right\|_2} \ll \varepsilon
\end{cases}
\end{aligned} \tag{26}$$

$\forall X \in \{A,B,C,D\},\ \forall i = 1,...,N_b$, which is convenient for keeping track of the algorithm progress. In case there exist $\theta_{X^{(i)}}$ for which $\left\|\hat{\theta}_{X^{(0)}}\right\|_2 = 0$, expression (21) or

(22) should be used instead.

The new optimization problem (24) gets solved sequentially as before:

$$\begin{aligned}
\underset{\Delta\theta,\,\Delta s}{\text{minimize}}\ & \nabla V_{\text{NLS}}\left(\theta^{(k)}\right)^{T}\Delta\theta+\frac{1}{2}\Delta\theta^{T}M^{(k)}\Delta\theta+\gamma\sum_{i=1}^{N_b}\Bigg(\phi\left(\hat{\theta}_{A^{(i)}}\right)\Delta s_{A^{(i)}}+ \\
& +\phi\left(\hat{\theta}_{B^{(i)}}\right)\Delta s_{B^{(i)}}+\phi\left(\hat{\theta}_{C^{(i)}}\right)\Delta s_{C^{(i)}}+\phi\left(\hat{\theta}_{D^{(i)}}\right)\Delta s_{D^{(i)}}\Bigg) \\
\text{subject to}\quad & \varphi\left(\hat{\theta}_{A^{(i)}}\right)\left\|\theta_{A^{(i)}}^{(k)}+\Delta\theta_{A^{(i)}}\right\|_2\leq s_{A^{(i)}}^{(k)}+\Delta s_{A^{(i)}} \\
& \varphi\left(\hat{\theta}_{A^{(i)}}\right)\left\|\theta_{B^{(i)}}^{(k)}+\Delta\theta_{B^{(i)}}\right\|_2\leq s_{B^{(i)}}^{(k)}+\Delta s_{B^{(i)}} \\
& \varphi\left(\hat{\theta}_{A^{(i)}}\right)\left\|\theta_{C^{(i)}}^{(k)}+\Delta\theta_{C^{(i)}}\right\|_2\leq s_{C^{(i)}}^{(k)}+\Delta s_{C^{(i)}} \\
& \phi\left(\hat{\theta}_{A^{(i)}}\right)\left\|\theta_{D^{(i)}}^{(k)}+\Delta\theta_{D^{(i)}}\right\|_2\leq s_{D^{(i)}}^{(k)}+\Delta s_{D^{(i)}} \\
& i=1,\ldots,N_b. \qquad (27)
\end{aligned}$$

The described principle makes sense though only under assumption that the weights are based upon a valid set of parameter estimates $\hat{\theta}$. The algorithm should, therefore, alternate between estimating θ by solving (24) and redefining the penalization weights, as in [28]. This functionality is realized in Algorithm 1. The inner loop is iteratively solving (24) by updating θ via (27), as long as the step size is larger than a specified tolerance $\Delta\theta_{\min}$. Once $\Delta\theta_{\min}$ is reached, the outer loop takes the latest estimate of θ to calculate new weights ϕ, redefines the optimization problem, and runs the inner loop anew. The procedure is being repeated as long as the relative change of the weights between the cycles is above a certain threshold $\Delta\phi_{\min}$.

As already mentioned, the value of the regularization factor γ determines the importance of the model simplicity with regard to the accuracy. This implies an inevitable trade-off between the two objectives for a fixed γ. Nevertheless, following the reasoning of Algorithm 1, where better estimates of the model parameters are obtained over and over thanks to updated penalization weights (and vise versa), one could use the updated estimates of the model parameters and run again the entire algorithm with a reduced γ to possibly improve model accuracy. Too small γ results in regularization having no influence on the model identification.

Data: θ_0, G_m, $\Delta\theta_\mathrm{min}$, $\Delta\phi_\mathrm{min}$

Result: $\hat{\theta}$

$\ell \leftarrow 0, \quad \hat{\theta}^{[\ell]} \leftarrow \theta_0, \quad \left|\phi\left(\hat{\theta}_{X^{(i)}}^{[\ell-1]}\right)\right| \leftarrow 0, \ \ \forall X \in \{A,B,C,D\}, \ \ \forall i = 1,...,N_b;$

while $\exists X, \exists i \ \ \dfrac{\left|\phi\left(\hat{\theta}_{X^{(i)}}^{[\ell]}\right) - \phi\left(\hat{\theta}_{X^{(i)}}^{[\ell-1]}\right)\right|}{\left|\phi\left(\hat{\theta}_{X^{(i)}}^{[\ell-1]}\right)\right|} > \Delta\phi_\mathrm{min}$ **do**

  $\theta \leftarrow \hat{\theta}^{[\ell]}$;

  while $\|\Delta\theta\|_2 > \Delta\theta_\mathrm{min}$ **do**

    $\Delta\theta \leftarrow$ solve (27);

    $\theta \leftarrow \theta + \Delta\theta$;

  end

  $\ell \leftarrow \ell + 1$;

  $\hat{\theta}^{[\ell]} \leftarrow \theta$;

end

$\hat{\theta} = \hat{\theta}^{[\ell]}$;

Algorithm 1: Reweighted $\ell_{2,1}$-norm regularization procedure

4 Numerical validation

This section discusses the application of the developed reweighted $\ell_{2,1}$-norm regularization approach to a flexible gantry system. This case study is part of a research project which develops methods to concurrently design and optimize systems and their controller. To facilitate this concurrent design, low-order system models are required that explicitly depend on a selected set of design parameters.

4.1 System description and identification data

The gantry system shown in Fig. 1 consists of a bridge that is moving along two linear guides and is actuated at each side by a motor via a rack and pinion. The bridge is a mechanical truss structure constructed of square tubes of fixed thickness $t = 2\,\mathrm{mm}$ and side length p. The side length of the tubes is the design parameter in the above mentioned study and hence the scheduling parameter for the LPV model identification. This parameter influences the mass and the stiffness of the gantry. The head of the gantry system is fixed in the middle of the bridge. In this case study, the input to the system is the motor torque, which is the same for both motors. The

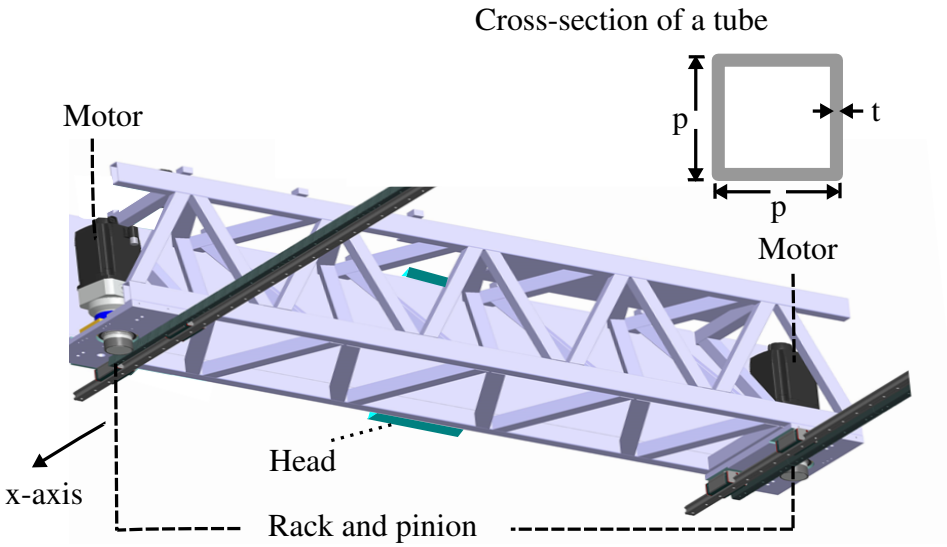

Fig. 1. Gantry crane

outputs are the acceleration of the head in the x-direction and the acceleration of the gantry on the rack, also in the x-direction, which is the same at both sides because of system symmetry and because both motors apply the same torques.

The models used to generate the data for the LPV identification are five continuous-time 6^{th} order LTI *Single-Input Multiple-Output* state-space models for the side lengths of the tubes equal to

$$p = \{24,\ 30,\ 36,\ 42,\ 48\}\ \text{mm}. \tag{28}$$

These models have been derived from a high-order flexible multi-body model of this system by applying the model reduction techniques described in [39]. The 6^{th} order models were then discretized and the corresponding FRFs, evaluated at 300 equally distributed frequency lines within the band $f \in [50, 300]$ Hz and corrupted with complex Gaussian noise yielding a SNR of 40 dB, form the identification data set. The square root of the reciprocal value of the known noise variance is used to build the weights in the fitness criterion (7) of all performed LTI and LPV identifications. By doing this, the weighted error criterion (7) corresponds to the maximum likelihood criterion [33].

4.2 Initial LPV model

An initial LPV model is obtained using the SMILE technique, as suggested in section 3.3.2. The SMILE technique starts from LTI models identified via the nonlinear least-squares frequency domain system identification method [10] based on the aforementioned FRFs. The next step is the interpolation of the LTI models using

basis functions of the scheduling parameter suggested by the user. Since there is no prior knowledge available on the scheduling parameter dependency of the model, and since there are five LTI models, a fourth-order polynomial scheduling parameter dependency and hence following set of basis functions:

$$\psi_1 = p(t), \quad \psi_2 = p(t)^2, \quad \psi_3 = p(t)^3, \quad \psi_4 = p(t)^4, \tag{29}$$

was chosen. With this choice, the interpolation can be performed without introducing errors; the LTI models match exactly to the LPV model for the corresponding fixed values of the scheduling parameter. Table 1 shows for the obtained SMILE model the magnitude of the parameters of the matrices linked to the different basis functions. The left side of the table displays their ℓ_2 norm, whereas the right side shows their relative magnitude indicating the contribution to the system response compared to the linear counterpart. It can be seen that all basis functions (ψ_1, ... , ψ_4) are being used in each model matrix ($\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, $\mathscr{D}$), although the parameters building matrix $\mathscr{B}$ are relatively small.

Employing all four basis functions for each state-space model matrix (1) - (2) may not be necessary. This assessment is often a matter of compromise one is willing to make when trading off model accuracy for model simplicity. The technique presented in this paper implements this compromise and delivers a model the most accordant with the requirements. Nevertheless, the trade-off curve is in most cases a function more complex than a straight line, which means there might exist points in which the lost model accuracy is negligible in comparison with the obtained model simplification. Finding these points is the main challenge of the presented technique, taken up in what follows.

4.3 Results

Algorithm 1 solving (24) with the weights (23) and (25) was firstly run with $\gamma = 10^{-3}$. The resulting model matrices $\mathscr{B}$ and $\mathscr{D}$ are independent of every basis function and consequently, independent of the scheduling parameter. Fig. 2 shows the magnitude of the difference between the FRF data used for the identification and the FRF of the model (red crosses). This error is further referred to as error magnitude. Comparing this error magnitude with the standard deviation of the used FRF data (blue crosses), one notices a bias, that is a systematic modeling error due to oversimplification of the scheduling parameter dependency. This systematic modeling error can be reduced by giving less importance to the regularization, i.e. decreasing γ.

With the same settings, Algorithm 1 was run again, but with $\gamma = 5 \cdot 10^{-5}$. The magnitude measures of the obtained model parameters are given in Table 2. It can be seen that the parameters building components of $\mathscr{B}$ remained negligible, and so did the parameters building $\mathscr{D}$, except for those associated with basis function

Table 1
Parameter magnitude of the SMILE model.

i	$\|\theta_{A^{(i)}}\|_2$	$\|\theta_{B^{(i)}}\|_2$	$\|\theta_{C^{(i)}}\|_2$	$\|\theta_{D^{(i)}}\|_2$	$\frac{\|\psi_i\|_\infty \|\theta_{A^{(i)}}\|_2}{\|\hat{\theta}_{A^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{B^{(i)}}\|_2}{\|\hat{\theta}_{B^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{C^{(i)}}\|_2}{\|\hat{\theta}_{C^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{D^{(i)}}\|_2}{\|\hat{\theta}_{D^{(0)}}\|_2}$
0	25.965	1.000	0.470	1.580				
1	82.659	$1.908 \cdot 10^{-4}$	57.619	49.620	0.152	$9.156 \cdot 10^{-6}$	5.878	1.508
2	$2.833 \cdot 10^{3}$	0.004	$2.425 \cdot 10^{3}$	$2.126 \cdot 10^{3}$	0.250	$8.420 \cdot 10^{-6}$	11.875	3.100
3	$3.689 \cdot 10^{4}$	0.118	$4.507 \cdot 10^{4}$	$4.078 \cdot 10^{4}$	0.157	$1.301 \cdot 10^{-5}$	10.593	2.855
4	$1.782 \cdot 10^{5}$	1.670	$3.095 \cdot 10^{5}$	$2.886 \cdot 10^{5}$	0.036	$8.866 \cdot 10^{-6}$	3.491	0.970

Table 2
Parameter magnitude of the regularized model.

i	$\|\theta_{A^{(i)}}\|_2$	$\|\theta_{B^{(i)}}\|_2$	$\|\theta_{C^{(i)}}\|_2$	$\|\theta_{D^{(i)}}\|_2$	$\frac{\|\psi_i\|_\infty \|\theta_{A^{(i)}}\|_2}{\|\hat{\theta}_{A^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{B^{(i)}}\|_2}{\|\hat{\theta}_{B^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{C^{(i)}}\|_2}{\|\hat{\theta}_{C^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{D^{(i)}}\|_2}{\|\hat{\theta}_{D^{(0)}}\|_2}$
0	25.965	2.999	0.471	1.783				
1	82.659	$7.858 \cdot 10^{-10}$	57.034	$4.604 \cdot 10^{-10}$	0.152	$1.2579 \cdot 10^{-11}$	5.810	$1.240 \cdot 10^{-11}$
2	$2.833 \cdot 10^{3}$	$2.362 \cdot 10^{-8}$	$2.422 \cdot 10^{3}$	33.466	0.250	$1.815 \cdot 10^{-11}$	11.844	0.043
3	$3.689 \cdot 10^{4}$	$5.790 \cdot 10^{-7}$	$4.438 \cdot 10^{4}$	$3.312 \cdot 10^{-8}$	0.157	$2.135 \cdot 10^{-11}$	10.415	$2.055 \cdot 10^{-12}$
4	$1.781 \cdot 10^{5}$	$1.311 \cdot 10^{-5}$	$2.900 \cdot 10^{5}$	$1.612 \cdot 10^{-6}$	0.036	$2.321 \cdot 10^{-11}$	3.268	$4.801 \cdot 10^{-12}$

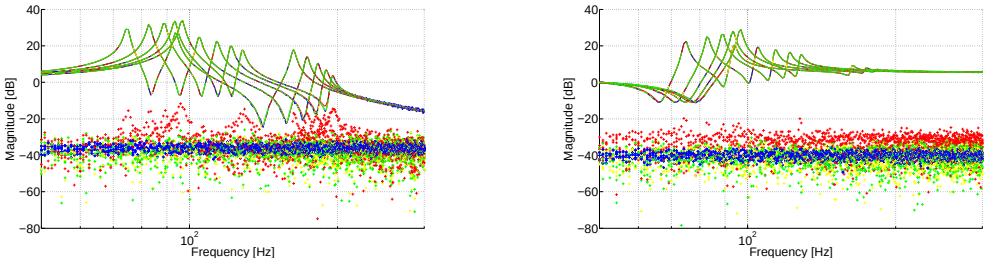

Fig. 2. Error magnitude of the initial SMILE model (+), the regularized model obtained with $\gamma = 10^{-3}$ (+), and the regularized model obtained with $\gamma = 5 \cdot 10^{-5}$ (+), compared against the standard deviation of the FRF data (+). The solid lines show the corresponding FRFs. Left subfigure concerns the first model output, right the second.

ψ_2. Judging by Fig. 2, the error magnitude of this model (green crosses) is comparable with the error magnitude of the SMILE model (yellow crosses) and the standard deviation of the measured FRFs (blue crosses). Hence, by allowing for one more basis function (yielding only two more nonzero parameters), the model error and accordingly the value of the weighted nonlinear least-squares criterion (7) is significantly smaller, more precisely $V_{\mathrm{NLS}} = 3.652 \cdot 10^{3}$, compared to the one of the model with $\mathscr{B}$ and $\mathscr{D}$ completely independent of the scheduling parameter ($V_{\mathrm{NLS}} = 1.876 \cdot 10^{4}$). For the reference, the SMILE model error equals $V_{\mathrm{NLS}} = 3.405 \cdot 10^{3}$. Taking into account the substantial model simplification with respect to the SMILE model, we conclude that the model obtained with $\gamma = 5 \cdot 10^{-5}$ got the best of both criteria—accuracy and simplicity—and represents a good compromise between the two extremes. Fig 3 portrays its magnitude and phase surface as a function of the scheduling parameter. The transitions of the system resonances and antiresonances are smooth, which confirms the validity of the model over the considered operating range.

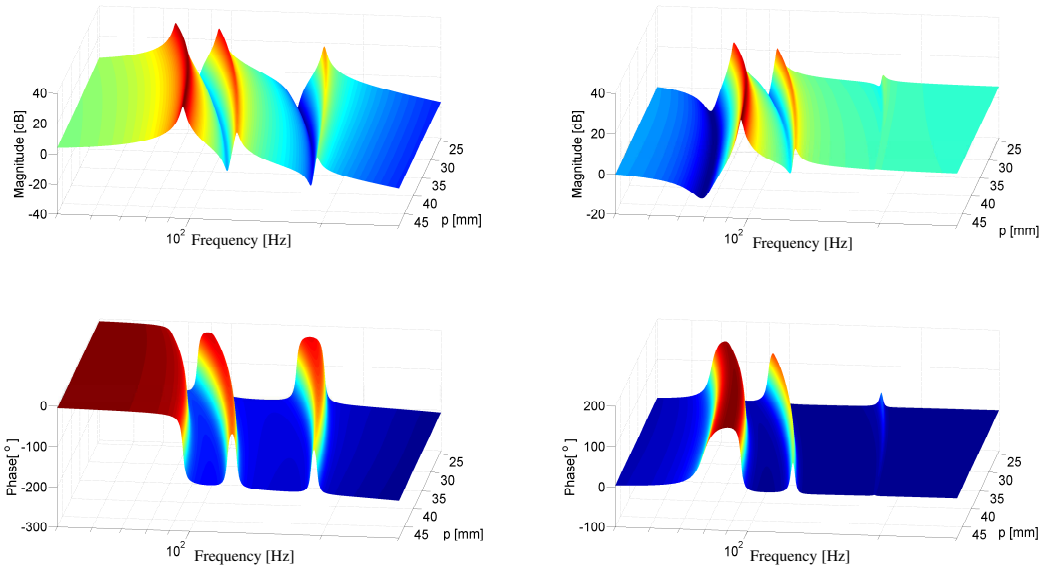

Fig. 3. The regularized LPV model of the gantry crane system: magnitude (top) and phase (bottom) surface of the first (left) and second (right) output, as a function of the scheduling parameter.

Remark on computational complexity. It is difficult to formulate a general statement on the comparison of computational complexity of the non-regularized and regularized optimization problem. This is because the number of computations required to find a solution depends on many things, e.g. the number of data samples involved, the selected basis functions and on the order of the model being identified. The total number of iterations needed to find an optimum may be different as well. However, to get a rough idea on this issue, a comparison of the average computation time of one iteration of the two optimization problems is given for this specific case. One iteration of the Levenberg-Marquardt algorithm corresponds to solving a linear least squares problem, one iteration of the regularized problem (NSOCP) corresponds to solving one SOCP. For this specific case, the SOCP on the average takes 0.16 s, whereas one iteration of the Levenberg-Marquardt algorithm takes on the average 0.02 s, yielding a ratio of 8. The code of both methods is implemented in MATLAB 8.3.0.532 (R2014a) and executed on the 64-bit Operating System with Intel® Core™ i5-3210M CPU @ 2.50 GHz and 8 GB of RAM.

5 Experimental validation

5.1 Setup description

The experimental setup is the XY-motion system shown in Fig. 4. The system consists of two perpendicularly mounted linear stages (X and Y) and a flexible can-

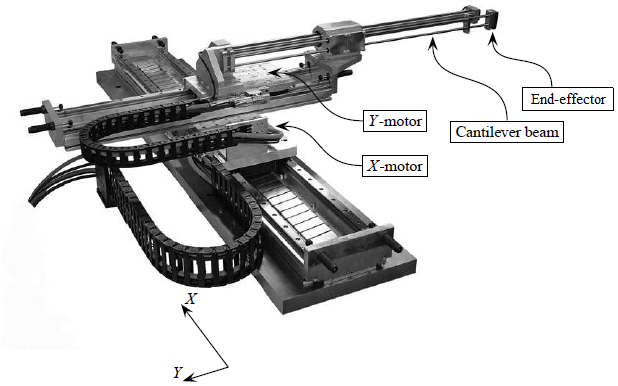

Fig. 4. XY-motion system

tilever beam. The length of this beam is changed by the position of the Y-motor, such that the cantilever beam resonances and hence the dynamics in the X-direction depend on the position of the Y-motor [38]. The reference velocity for the velocity controlled X-motor is the system input, while the acceleration of the end-effector in the same direction represents the system output. The acceleration is measured by an inductive accelerometer having a flat frequency spectrum within $f \in [0, 250]$ Hz. The reference position for the position controlled Y-motor is the scheduling parameter of the system we aim to identify.

5.2 Experiment design and initial LPV model

Given the operating range of the setup, the following positions of the Y-motor were chosen as scheduling points at which local experiments were performed:

$$p = \{-0.151,\ -0.093,\ p = -0.0639,\ -0.035,\ 0.023\}\ \text{m}, \tag{30}$$

where 0 m corresponds to the middle position of the Y-stage.

Both measurement noise and nonlinear system distortions corrupt the FRF measurements, and have to be accounted for in the estimation of the variance of the FRF. In order to estimate this variance and average out both types of distortions and hence improve the quality of the FRF measurement, ten experiments with different realizations of a random-phase multisine signal were performed, and this for each of the four chosen local scheduling points. These mesurements were performed in a zero-order-hold setting, also meaning that the input is noise free. The multisine signals are composed of frequencies in the range $f \in [3, 50]$ Hz, with an as high as possible amplitude without causing motor current saturation, and a duration of 8.192 s resulting in a frequency resolution of 0.1221 Hz. During each experiment, 10 periods were measured. Consequently, for each of the five local scheduling points, there are 10×10 FRFs. Their average and sample total variance are then calculated using the procedure described in [40]. The measurement taken

in the middle of the operating range, i.e. at $p = -0.0639\,\mathrm{m}$, is used for validation purposes only, whereas the remaining four are used for the identification. The reciprocal value of the square root of their sample total variance is used to build the weights in the fitness criterion (7) of all performed LTI and LPV identifications. Through this choice of weighting, and because of noise-free input data, the weighted error criterion corresponds to the maximum likelihood criterion [33].

To obtain initial estimates of the LPV model parameters using again the SMILE technique, LTI models describing the system behavior at the aforementioned scheduling points must be identified first. This LTI model identification was performed with the nonlinear least-squares frequency domain identification technique [10] using the weights as described above. The obtained LTI models are of order 4. Since there is no direct feedthrough between the input and the output due to data sampling, the state space model matrix $\mathscr{D}$ and hence parameter vector θ_D is fixed to zero.

No prior information about suitable basis functions was available, so a third-order polynomial scheduling parameter dependency i.e. the following set of basis functions:

$$\psi_1 = p(t), \quad \psi_2 = p(t)^2, \quad \psi_3 = p(t)^3, \tag{31}$$

was chosen to interpolate the four LTI models. In that way, no interpolation error is introduced and the LPV model evaluated at the values of the scheduling parameter $p = \{-0.151,\ -0.093,\ -0.035,\ 0.023\}$ m matches exactly to the corresponding LTI models.

5.3 Results

Table 3 and Table 4 show for the SMILE model and the regularized model the magnitude of the parameters of the matrices linked to the different basis functions, as in Table 1. The regularized model is obtained using Algorithm 1 within a loop that gradually decreases γ to improve model accuracy. The magnitude of the parameters in B_i and C_i, $\forall i = 1,2,3$ of the regularized model is significantly smaller than the magnitude of the parameters in B_0 and C_0, respectively. The proposed $\ell_{2,1}$-norm regularization enabled us to select a simpler scheduling parameter dependency, in the sense that the input matrix ($\mathscr{B}$) and the complete output equation (matrices $\mathscr{C}$ and $\mathscr{D}$) are independent of the scheduling parameter.

When it comes to accuracy, the regularized model results in a 7.33% larger V_{NLS} (7) value compared to the SMILE model, indicating a trade-off between the model accuracy and simplicity. However, by looking at Fig. 5 - Fig. 7, which portray the fit of the initial SMILE model and the regularized model to the measured FRFs, one can see no significant difference between the two. Their overlap with the measured FRFs evidently implies that both models capture the local LPV behavior of the

Table 3
Parameter magnitude of the SMILE model.

i	$\|\theta_{A^{(i)}}\|_2$	$\|\theta_{B^{(i)}}\|_2$	$\|\theta_{C^{(i)}}\|_2$	$\|\theta_{D^{(i)}}\|_2$	$\frac{\|\psi_i\|_\infty \|\theta_{A^{(i)}}\|_2}{\|\hat{\theta}_{A^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{B^{(i)}}\|_2}{\|\hat{\theta}_{B^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{C^{(i)}}\|_2}{\|\hat{\theta}_{C^{(0)}}\|_2}$
0	2.617	1.312	$3.865 \cdot 10^4$	0			
1	0.549	0.365	$6.631 \cdot 10^4$	0	0.025	0.042	0.259
2	35.976	2.069	$5.056 \cdot 10^5$	0	0.249	0.036	0.297
3	197.322	6.765	$2.781 \cdot 10^6$	0	0.206	0.018	0.247

Table 4
Parameter magnitude of the regularized model.

i	$\|\theta_{A^{(i)}}\|_2$	$\|\theta_{B^{(i)}}\|_2$	$\|\theta_{C^{(i)}}\|_2$	$\|\theta_{D^{(i)}}\|_2$	$\frac{\|\psi_i\|_\infty \|\theta_{A^{(i)}}\|_2}{\|\hat{\theta}_{A^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{B^{(i)}}\|_2}{\|\hat{\theta}_{B^{(0)}}\|_2}$	$\frac{\|\psi_i\|_\infty \|\theta_{C^{(i)}}\|_2}{\|\hat{\theta}_{C^{(0)}}\|_2}$
0	2.611	1.358	$3.585 \cdot 10^4$	0			
1	0.169	$8.491 \cdot 10^{-16}$	$1.842 \cdot 10^{-11}$	0	0.008	$9.425 \cdot 10^{-17}$	$7.746 \cdot 10^{-17}$
2	31.732	$5.085 \cdot 10^{-15}$	$1.222 \cdot 10^{-10}$	0	0.219	$8.510 \cdot 10^{-17}$	$7.746 \cdot 10^{-17}$
3	174.820	$3.921 \cdot 10^{-14}$	$8.933 \cdot 10^{-10}$	0	0.182	$9.891 \cdot 10^{-17}$	$8.536 \cdot 10^{-17}$

XY-motion system sufficiently well. The fact that in both cases the model error magnitude is at the level of the standard deviation of the measured FRFs $\sigma(G_\mathrm{m})$ confirms that the obtained accuracy is very close to the maximum one can achieve.

Fig. 8 shows the accuracy of the obtained models for the validation FRF measurement at $p = -0.0639\,\mathrm{m}$. The error magnitude obtained with both the SMILE and regularized model are similar as for the FRF measurements that were used in the identification, that is, it is at the same level as the total sample variance of the FRF measurement. This shows that both models have excellent interpolating capabilities.

With respect to the computation requirements, one SOCP iteration needs 0.027 s, while the Levenberg-Marquardt iteration needs 0.003 s, which makes it 9 times faster than the SOCP.

6 Conclusion

With the aim of reducing the complexity of LPV models in terms of dependency on the scheduling parameters, and consequently avoiding data overfitting, we explored the reweighted $\ell_{2,1}$-norm regularization in a nonlinear least-squares system identification setting. The reweighted $\ell_{2,1}$-norm regularization is the outcome of combining the idea of applying the $\ell_{2,1}$-norm regularization to the parameter estimation problem, and the concept of reweighted ℓ_1 minimization. Reformulation of the optimization problem with added regularization into a nonlinear second-order cone programming problem resulted in an approach successfully validated numerically on a flexible gantry system, and experimentally on an XY-motion system.

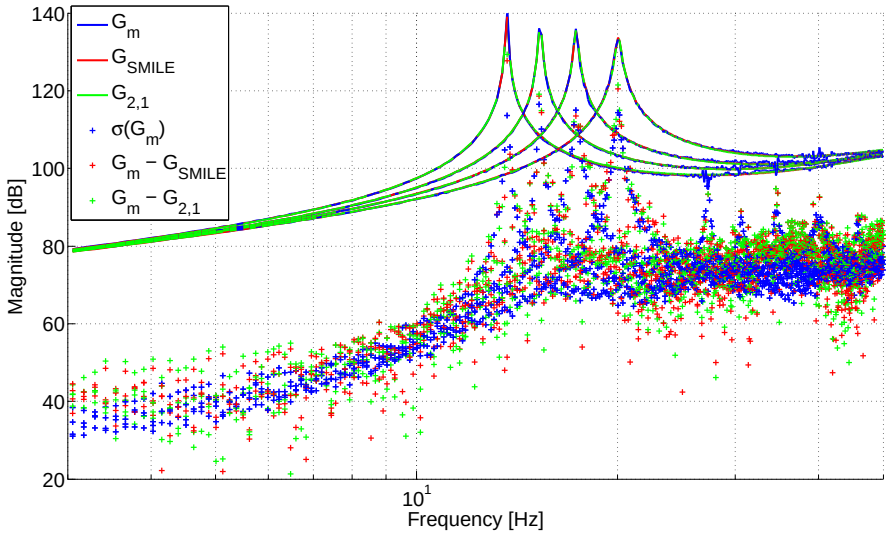

Fig. 5. Local identification fit - magnitude. G_{m}, G_{SMILE}, and $G_{2,1}$ indicate the measured FRF, the FRF of the SMILE model and the FRF of the regularized model, respectively. $\sigma(G_{\mathrm{m}})$ indicates the estimated standard deviation of the FRF data.

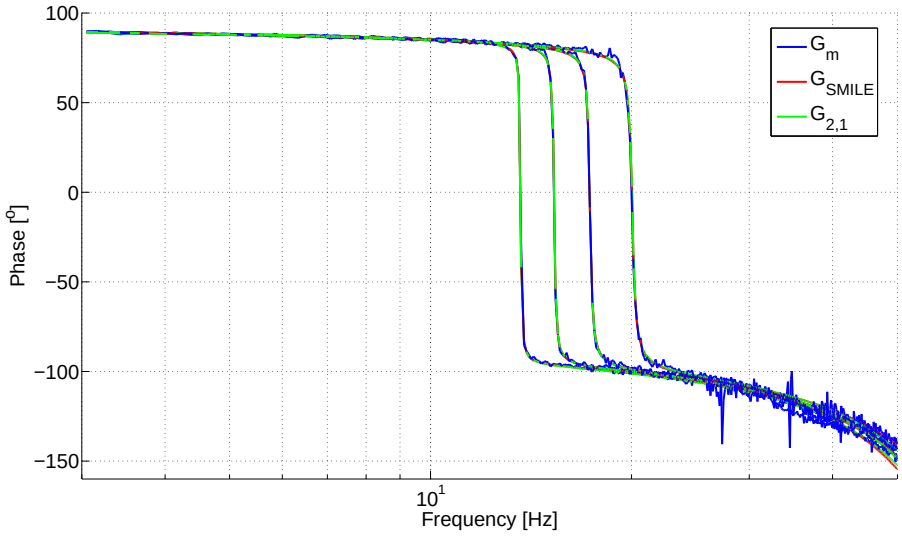

Fig. 6. Local identification fit - phase. G_{m}, G_{SMILE}, and $G_{2,1}$ indicate the measured FRF, the FRF of the SMILE model and the FRF of the regularized model, respectively. $\sigma(G_{\mathrm{m}})$ indicates the estimated standard deviation of the FRF data.

The obtained LPV models have significantly simpler scheduling parameter dependency than the nonregularized version, while still being a good approximation of the system behavior.

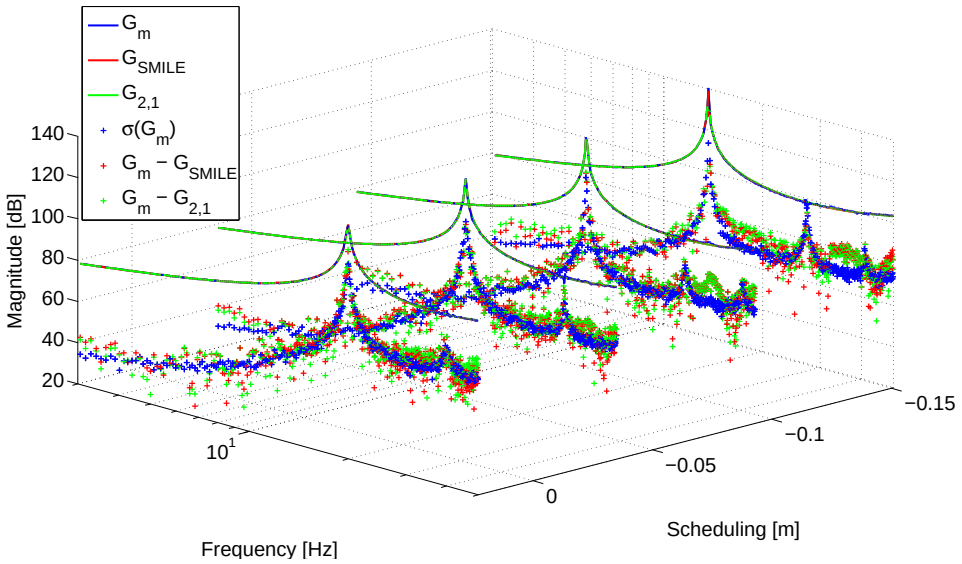

Fig. 7. Local identification fit in 3D - magnitude. G_{m}, G_{SMILE}, and $G_{2,1}$ indicate the measured FRF, the FRF of the SMILE model and the FRF of the regularized model, respectively. $\sigma(G_{\mathrm{m}})$ indicates the estimated standard deviation of the FRF data.

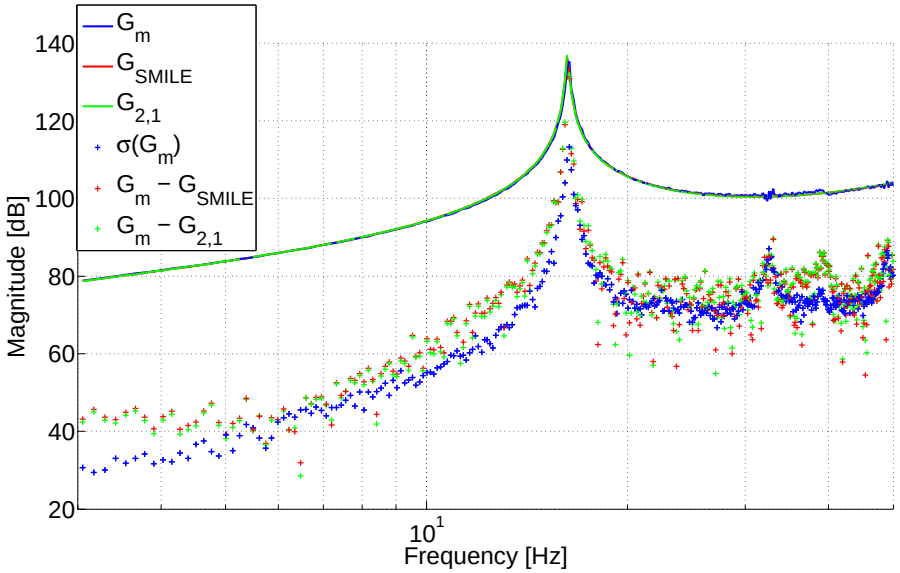

Fig. 8. Local validation fit at $p = -0.0639\,\mathrm{m}$ - magnitude. G_{m}, G_{SMILE}, and $G_{2,1}$ indicate the measured FRF, the FRF of the SMILE model and the FRF of the regularized model, respectively. $\sigma(G_{\mathrm{m}})$ indicates the estimated standard deviation of the FRF data.

Acknowledgement

This research is sponsored by the Fund for Scientific Research, by the KU Leuven-BOF PFV/10/002 Center-of-Excellence Optimization in Engineering (OPTEC), by Flanders Make SBO project ROCSIS: Robot and Optimal Control of System of Interacting Subsystems, IWT SBO project MBSE4Mechatronics (Model-Based Sys-

tem Engineering for Mechatronics), and by the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Ministers Office, Science Policy programming (IAP VII, DYSCO).

References

[1] C. Hoffmann and H. Werner, "A survey of linear parameter-varying control applications validated by experiments or high-fidelity simulations," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 2, pp. 416–433, 2015.

[2] B. Bamieh and L. Giarre, "Identification of linear parameter varying models," *International Journal of Robust and Nonlinear Control*, vol. 12, no. 9, pp. 841–853, 2002.

[3] F. Felici, J.-W. Van Wingerden, and M. Verhaegen, "Subspace identification of MIMO LPV systems using a periodic scheduling sequence," *Automatica*, vol. 43, no. 10, pp. 1684–1697, 2007.

[4] V. Verdult, M. Lovera, and M. Verhaegen, "Identification of linear parameter-varying state-space models with application to helicopter rotor dynamics," *International Journal of Control*, vol. 77, no. 13, pp. 1149–1159, 2004.

[5] J. Van Wingerden, I. Houtzager, F. Felici, and M. Verhaegen, "Closed-loop identification of the time-varying dynamics of variable-speed wind turbines," *International Journal of Robust and Nonlinear Control*, vol. 19, no. 1, pp. 4–21, 2009.

[6] J. De Caigny, R. Pintelon, J. Camino, and J. Swevers, "Interpolated modeling of LPV systems based on observability and controllability," in *Proc. of the 16th IFAC Symposium on System Identification, Brussels, Belgium*, pp. 1773–1778, 2013.

[7] M. Steinbuch, R. Van De Molengraft, and A. van der Voort, "Experimental modelling and LPV control of a motion system," in *Proceedings of the American Control Conference, 2003*, vol. 2, pp. 1374–1379, IEEE, 2003.

[8] J. De Caigny, R. Pintelon, J. F. Camino, and J. Swevers, "Interpolated modeling of LPV systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 6, pp. 2232–2246, 2014.

[9] D. Turk, G. Pipeleers, and J. Swevers, "A combined global and local identification approach for LPV systems," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 184–189, 2015.

[10] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach, Second Edition*. John Wiley & Sons, 2012.

[11] H. Akaike, "Stochastic theory of minimal realization," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 667–674, 1974.

[12] S. Beheshti and M. A. Dahleh, "A new information theoretic approach to order estimation problem," in *13th IFAC Symposium on System Identification*, 2003.

[13] T. A. Johansen, "On Tikhonov regularization, bias and variance in nonlinear system identification," *Automatica*, vol. 33, no. 3, pp. 441–446, 1997.

[14] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[15] 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.

[16] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[17] L. Ljung and T. Chen, "Convexity issues in system identification," in *Control and Automation (ICCA), 2013 10th IEEE International Conference on*, pp. 1–9, IEEE, 2013.

[18] T. Chen, M. S. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto, "System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 2933–2945, 2014.

[19] R. Tóth, H. Hjalmarsson, and C. R. Rojas, "Order and structural dependence selection of LPV-ARX models revisited," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 6271–6276, IEEE, 2012.

[20] C. R. Rojas and H. Hjalmarsson, "Sparse estimation based on a validation criterion," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pp. 2825–2830, IEEE, 2011.

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

[22] 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.

[23] R. Tóth, C. Lyzell, M. Enqvist, P. S. Heuberger, and P. M. Van den Hof, "Order and structural dependence selection of LPV-ARX models using a nonnegative garrote approach," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pp. 7406–7411, IEEE, 2009.

[24] P. Gebraad, J.-W. van Wingerden, and M. Verhaegen, "Sparse estimation for predictor-based subspace identification of LPV systems," in *16th IFAC Symposium on System Identification*, pp. 1749–1754, 2012.

[25] R. Tóth, H. S. Abbas, and H. Werner, "On the state-space realization of LPV input-output models: Practical approaches," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 1, pp. 139–153, 2012.

[26] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou, "$\ell_{2,1}$-norm regularized discriminative feature selection for unsupervised learning," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, p. 1589, 2011.

[27] J. Liu, S. Ji, and J. Ye, "Multi-task feature learning via efficient $l_{2,1}$-norm minimization," in *Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial Intelligence*, pp. 339–348, AUAI Press, 2009.

[28] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[29] H. Zou, "The adaptive LASSO and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[30] D. P. Wipf and S. S. Nagarajan, "A new view of automatic relevance determination," in *Advances in Neural Information Processing Systems*, pp. 1625–1632, 2008.

[31] R. Pintelon and J. Schoukens, "Measurement of frequency response functions using periodic excitations, corrupted by correlated input/output errors," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 6, pp. 1753–1760, 2001.

[32] J. Schoukens, T. Dobrowiecki, and R. Pintelon, "Parametric and nonparametric identification of linear systems in the presence of nonlinear distortions-a frequency domain approach," *IEEE Transactions on Automatic Control*, vol. 43, no. 2, pp. 176–190, 1998.

[33] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van Hamme, "Parametric identification of transfer functions in the frequency domain-a survey," *IEEE Transactions on Automatic Control*, vol. 39, no. 11, pp. 2245–2260, 1994.

[34] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, no. 1, pp. 193–228, 1998.

[35] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in matlab," in *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pp. 284–289, IEEE, 2005.

[36] A. Domahidi, E. Chu, and S. Boyd, "ECOS: An SOCP solver for embedded systems," in *European Control Conference (ECC)*, pp. 3071–3076, 2013.

[37] H. Kato and M. Fukushima, "An SQP-type algorithm for nonlinear second-order cone programs," *Optimization Letters*, vol. 1, no. 2, pp. 129–144, 2007.

[38] J. De Caigny, J. F. Camino, and J. Swevers, "Interpolating model identification for SISO linear parameter-varying systems," *Mechanical Systems and Signal Processing*, vol. 23, no. 8, pp. 2395–2417, 2009.

[39] E. Lappano, E. Nijman, F. Naets, W. Desmet, and D. Mundo, "A parametric model order reduction for simulations of beam-based structures," *eLiQuiD Best Engineering Training in Electric, Lightweight and Quiet Driving*, p. 45, 2016.

[40] R. Pintelon, J. Schoukens, W. Van Moer, and Y. Rolain, "Identification of linear systems in the presence of nonlinear distortions," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 4, pp. 855–863, 2001.