Frequency Domain Model Selection for Servo Systems ensuring Practical Identifiability

Mathias Tantau, Mark Wielitzka, and Tobias Ortmaier
Institute of Mechatronic Systems
Leibniz University Hannover
An der Universität 1, 30823 Garbsen, Germany
{*mathias.tantau, mark.wielitzka, tobias.ortmaier*}*@imes.uni-hannover.de*

Lars Perner
Lenze Automation GmbH
Am Alten Bahnhof 11,
38122 Braunschweig, Germany
lars.perner@lenze.com

***Abstract*— Physically motivated models of servo control systems with coupled mechanics are required for control design, simulation etc. Often, however, the effort of modelling prohibits these model-based methods in industrial applications. Therefore, all approaches of automatic modelling / model selection are naturally appealing. In this paper a procedure for model selection in frequency domain is proposed that minimizes the Kullback-Leibler distance between model and measurement while considering only those models that are practically identifiable. It aims at mechanical models of servo systems including multiple-mass resonators. Criteria for practical identifiability are derived locally from the sensitivity matrix which is calculated for different formulations of the equation error. In experiments with two industry-like testbeds the methodology proves to reveal the characteristic mechanical properties of the two setups.**

***Index Terms*— Model Selection; Frequency Domain; Servo Systems; Practical Identifiability; Sensitivity**

I. INTRODUCTION

Physically motivated models, mostly called bright grey-box models or first-principle models of servo control systems with coupled mechanics are required for control design [1], simulation, feed-forward, model-based fault diagnosis [2] etc. However, the effort of modelling prohibits these model-based methods in industrial applications. Especially, in the case of small series products up to individual solutions for specialized customers there may not always be the time for thorough model development and validation.

Therefore, all approaches of automatic modelling are naturally appealing. Automatic model selection is also called model structure identification [3] or structure and parameter identification. As a first step in this direction one could search for the model structure that best describes a measured input-output behaviour in the sense of minimizing the Kullback-Leibler distance, see [4], while considering criteria of practical identifiability as described below.

While many previous works on servo system identification exist, in the sense of identifying the parameters of a given model, e.g. [5], [6], works on model selection are mainly restricted to data-driven, static models [7] of other disciplines such as biology [8] and finance [9]. The difficulty with the transfer function models used to describe servo systems is that linearity in the physical parameters is generally not maintained. Extensions of identifiability tests to nonlinear models exist [10], [11], but they have rarely been applied to frequency domain. Frequency domain identification is of interest, since it allows an intuitive evaluation of transfer function models and it avoids time domain simulations, which are time-consuming and less suited for automation because of frequent instability problems. The presented model selection strategy that includes practical identifiability into the model composition decisions is innovative, even if the frequency domain aspect is left aside. An analysis of practical identifiability in frequency domain can be found in [12], where contour lines of the cost function are interpreted. A limited model selection in the field of dynamic models for servo control applications is performed in [13], but the process is not fully automatic and leaves the final decision for the commissioning engineer.

In this paper an automated model selection strategy is proposed for dynamic transfer function models with physical interpretation of the parameters. It optimizes the accuracy of the model in frequency domain but also ensures practical identifiability of all parameters, respectively avoids inclusion of superficial parameters. Three different options for formulating the cost function are considered together with the calculation of the sensitivity matrix for practical identifiability analysis.

The resulting model is optimal for the given excitation, but the excitation is not optimal for the selected model. Optimal design of experiments, see for example [14], is seen as an alternative approach to the problem of unsatisfactory identifiability. It is not followed here because for the given nonlinear models the necessity of accurate prior knowledge about model parameters would be prohibitive [15].

II. MODEL SELECTION AND PRACTICAL IDENTIFIABILITY

In this section the methodology for optimizing the model structure is explained along with the set of candidate submodels. Finally, criteria for practical identifiability are stated.

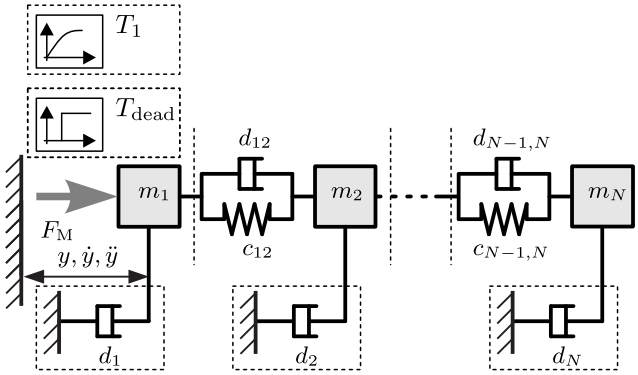

Fig. 1. Class of candidate models

A. Candidate models

Only a limited variety of multiple-mass models is chosen as candidate models as shown in Fig. 1. In general, the candidate models are composed of submodels as indicated by the dashed lines with 0 or more estimation parameters. Only the purely translational case is shown in the figure but the purely rotary case would work equivalently. All models consist of a chain of N elastically coupled masses $m_1, m_2, \cdots, m_N$ of which the first element is actuated. The force of the driving motor F_{M} may be subject to input delay time T_{dead}. In addition, the torque control dynamics are either neglected or modelled as a P_{T1} element. As the number of masses in the chain varies the spring-damper elements between adjacent masses will also appear / disappear, but there is never a spring-damper element in front of the first mass. In order to incorporate the effect of bearing and guideway friction dampers between masses and the environment can be included for each existing mass.

This set of candidate models could be extended easily, e.g. by springs between distant masses, but including a large set of candidate models without justification through experience and prior knowledge could be seen critically as data dredging / mining) [4], [16]. What results is a single input single output (SISO) transfer function model so that in the following only the SISO case is considered.

B. Model selection

In the process of model selection an exhaustive search over all combinations of submodels is carried out. More elegant procedures such as genetic programming [17] are avoided here in order to reduce the complexity and the effect of coincidence. For each model the parameters are optimized by matching the calculated and measured frequency response in an equation error formulation. The exact cost function is given in the next section. Then, the best model is chosen based on its merit as explained next.

Depending on the intended purpose of a model different ways to determine its merit can be defined. Reasonable criteria are the minimization of the Kullback-Leibler distance to the 'true model' by means of the Akaike information criterion or cross validation, a test for whiteness of the residuals, for cross-correlation between the inputs and the residuals, a χ^2 significance test of the cost function and others [15]. However, the problem with all these approaches is that they are based on the stochastic nature of measurements and they seek for the true model. If the repeatability is high, which can always be achieved by averaging several measurements under similar conditions [16], very complex models would result with possibly superfluous parameters. Since physical interpretability of model parameters is the prerequisite of this paper, it is reasonable to check *practical identifiability* of all model parameters. Practical identifiability means that all model parameters can be identified accurately from experimental data with real sparce, noisy data for a given excitation [18]. Criteria for checking practical identifiability are given below. Only models with all parameters practically identifiable are kept in the set of candidate models, see next sections. Among the models that fulfill these criteria the one with the best fit on an independent test data set is chosen. The separation into training and test data ensures minimization of the Kullback-Leibler distance [19].

C. Sensitivity calculation

The notion of practical identifiability comes from the parameter-linear model

$$\boldsymbol{S}_{\mathrm{lin}}\boldsymbol{p} = \boldsymbol{y} \tag{1}$$

with sensitivity matrix $\boldsymbol{S}_{\mathrm{lin}} \in \mathbb{R}^{N_y \times N_p}$, parameter vector $\boldsymbol{p} \in \mathbb{R}^{N_p \times 1}$, and system outputs $\boldsymbol{y} \in \mathbb{R}^{N_y \times 1}$. In the least-squares sense the cost function $J_{\mathrm{lin}} = (\boldsymbol{S}_{\mathrm{lin}}\boldsymbol{p} - \boldsymbol{y}_{\mathrm{m}})^{\mathrm{T}}(\boldsymbol{S}_{\mathrm{lin}}\boldsymbol{p} - \boldsymbol{y}_{\mathrm{m}})$ is minimized in order to determine optimal parameters, where $\boldsymbol{y}_{\mathrm{m}}$. In the study of practical identifiability the sensitivity matrix $\boldsymbol{S}_{\mathrm{lin}}$ is checked for (multi-)collinearity and sensitivity of the parameters as explained later. In the nonlinear case the model can be written as a nonlinear function with multiple outputs $\boldsymbol{g} : \mathbb{R}^{N_{\mathrm{p}}} \to \mathbb{R}^{N_{\mathrm{y}}}$ and the parameters $\boldsymbol{p}$ as inputs:

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{p}). \tag{2}$$

The cost function $J \in \mathbb{R}$ is of the form:

$$J = (\boldsymbol{y} - \boldsymbol{y}_{\mathrm{m}})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{y}_{\mathrm{m}}) = ||\boldsymbol{\Lambda}(\boldsymbol{y} - \boldsymbol{y}_{\mathrm{m}})||_2^2 . \tag{3}$$

$\boldsymbol{W} \in \mathbb{R}^{N_y \times N_y}$ is a weighting matrix, often chosen as the inverse covariance matrix of the measurements assuming zero mean, Gaussian noise. This gives the maximum likelihood estimator. The decomposition $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda} = \boldsymbol{W}$ exists if $\boldsymbol{W}$ is positive definite [15]. Motivated by the Taylor series expansion [10]:

$$\boldsymbol{g}(\boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{p}_0) + \left.\frac{\partial \boldsymbol{g}(\boldsymbol{p})}{\partial \boldsymbol{p}^{\mathrm{T}}}\right|_{\boldsymbol{p}=\boldsymbol{p}_0} (\boldsymbol{p} - \boldsymbol{p}_0) + \cdots \tag{4}$$

a substitute for the $\boldsymbol{S}_{\mathrm{lin}}$ of the linear case can be found:

$$\boldsymbol{S}_{\mathrm{nl}} = \boldsymbol{\Lambda}\left.\frac{\partial \boldsymbol{g}(\boldsymbol{p})}{\partial \boldsymbol{p}^{\mathrm{T}}}\right|_{\boldsymbol{p}=\boldsymbol{p}_0}. \tag{5}$$

The weighting $\boldsymbol{\Lambda}$ should be considered in the analysis of practical identifiability in agreement with the penalty function (3) as the weights also influence the result. $\boldsymbol{S}_{\mathrm{nl}} \in \mathbb{R}^{N_y \times N_y}$ characterizes the effect of changing a parameter locally on the output.

In frequency domain identification the observations and the noise are complex, which gives several options to formulate the penalty function. In the sequel, different cases are considered together with the implications for calculating the appropriate sensitivity matrix.

Case 1: The cost function is the distance in the complex plane between model $\boldsymbol{G}(\boldsymbol{p}) \in \mathbb{R}^{N_{\mathrm{f}} \times 1}$ and measurement $\boldsymbol{G}_{\mathrm{M}} \in \mathbb{R}^{N_{\mathrm{f}} \times 1}$ for each of the N_{f} frequency components [15], [20]. This can be written in complex notation:

$$J = (\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}})^{\mathrm{H}}\, \boldsymbol{W}\, (\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}}), \tag{6}$$

with $\boldsymbol{W} \in \mathbb{R}^{N_y \times N_y}$ or in real notation:

$$J = \left|\left| \boldsymbol{\Lambda}_{\mathrm{r/i}} \begin{bmatrix} \mathrm{Re}\,\{\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}}\} \\ \mathrm{Im}\,\{\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}}\} \end{bmatrix} \right|\right|_2^2. \tag{7}$$

$\boldsymbol{\Lambda}_{\mathrm{r/i}} \in \mathbb{R}^{2N_{\mathrm{f}} \times 2N_{\mathrm{f}}}$ can be defined by the square root of the inverse sample covariance matrix of the real measurement vector or if not available e.g. as the unity matrix. Assuming that real and imaginary part at each spectral line of the observation are uncorrelated having equal variances, matrix $\boldsymbol{\Lambda}_{\mathrm{r/i}}$ has the form

$$\begin{pmatrix} \boldsymbol{\Lambda}_{\mathrm{r}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_{\mathrm{i}} \end{pmatrix} \quad \text{with} \quad \boldsymbol{\Lambda}_{\mathrm{r}} = \boldsymbol{\Lambda}_{\mathrm{i}} \in \mathbb{R}^{N_{\mathrm{f}} \times N_{\mathrm{f}}} \tag{8}$$

and matrix $\boldsymbol{W}$ is given by $2\boldsymbol{\Lambda}_{\mathrm{r}}^{\mathrm{T}}\boldsymbol{\Lambda}_{\mathrm{r}}$ due to variance summation of uncorrelated variables [15].

The latter cost function formulation is favoured here because the resulting sensitivity matrix $\boldsymbol{S}_1 \in \mathbb{R}^{2N_{\mathrm{f}} \times N_p}$ is real and all criteria of practical identifiability can readily be applied:

$$\boldsymbol{S}_1 = \boldsymbol{\Lambda}_{\mathrm{r/i}} \begin{bmatrix} \boldsymbol{S}_{\mathrm{r}} \\ \boldsymbol{S}_{\mathrm{i}} \end{bmatrix}, \tag{9}$$

$$\boldsymbol{S}_{\mathrm{r}} = \frac{\mathrm{d}\mathrm{Re}\,\{\boldsymbol{G}(\boldsymbol{p})\}}{\mathrm{d}\boldsymbol{p}} = \mathrm{Re}\left\{\frac{\mathrm{d}\boldsymbol{G}(\boldsymbol{p})}{\mathrm{d}\boldsymbol{p}}\right\}, \tag{10}$$

$$\boldsymbol{S}_{\mathrm{i}} = \frac{\mathrm{d}\mathrm{Im}\,\{\boldsymbol{G}(\boldsymbol{p})\}}{\mathrm{d}\boldsymbol{p}} = \mathrm{Im}\left\{\frac{\mathrm{d}\boldsymbol{G}(\boldsymbol{p})}{\mathrm{d}\boldsymbol{p}}\right\}. \tag{11}$$

It can easily be verified that exchanging the decomposition into real and imaginary part, and the derivative operator in (10,11) is possible, because the complex unit can be treated as a prefactor in the derivative. The right side of (10,11) allows to calculate the real and imaginary part of the sensitivity numerically after the complex derivative has been calculated symbolically by the help of computer algebra, which simplifies the calculations. Still, for very complex models even $\mathrm{d}\boldsymbol{G}/\mathrm{d}\boldsymbol{p}$ cannot be calculated symbolically.

Case 2: Unfortunately, the distance in the complex plane is not very robust for identifying multiple mass models. In [1] the followeing cost function is used instead:

$$J = \left|\left| \boldsymbol{\Lambda}_{\mathrm{A/P}} \begin{bmatrix} |\boldsymbol{G}(\boldsymbol{p})| - |\boldsymbol{G}_{\mathrm{M}}| \\ \angle\{\boldsymbol{G}(\boldsymbol{p})\} - \angle\{\boldsymbol{G}_{\mathrm{M}}\} \end{bmatrix} \right|\right|_2^2. \tag{12}$$

In our experience this cost function facilitates a more safe convergence to the global optimum in parameter identification, but it is more difficult to adjust the weights. Assuming normal distributions for phase and amplitude the sample covariance matrix could be used again. Often, however, the sample covariance would be too expensive to measure and the variance is small. Then, maximizing the likelihood is not the major concern but rather the convergence to the global optimum. Thus, all frequencies can be weighted equally by a diagonal $\boldsymbol{\Lambda}_{\mathrm{A/P}}$, but it is still difficult to choose the weighting for amplitude and phase because different units are compared.

The sensitivity matrix is given by:

$$\boldsymbol{S}_2 = \boldsymbol{\Lambda}_{\mathrm{A/P}} \begin{bmatrix} \boldsymbol{S}_{\mathrm{A2}} \\ \boldsymbol{S}_{\mathrm{P2}} \end{bmatrix}, \tag{13}$$

$$\boldsymbol{S}_{\mathrm{A2},kj} = \frac{1}{|\boldsymbol{G}_k|}\left[\mathrm{Im}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{i},kj} + \mathrm{Re}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{r},kj}\right], \tag{14}$$

$$\boldsymbol{S}_{\mathrm{P2},kj} = \frac{1}{|\boldsymbol{G}_k|^2}\left[\mathrm{Re}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{i},kj} - \mathrm{Im}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{r},kj}\right]. \tag{15}$$

$\boldsymbol{G}_k$ denotes the k-th spectral component of the modelled transfer function. As it turns out, it is possible to write this Jacobian matrix in terms of $\boldsymbol{S}_{\mathrm{r}}, \boldsymbol{S}_{\mathrm{i}}$. So, whenever these two can be calculated from analytic expressions, numeric differentiation is not necessary for $\boldsymbol{S}_2$ either.

Case 3: In the third case to consider the logarithm of the amplitude is evaluated instead of the amplitude itself:

$$J = \left|\left| \boldsymbol{\Lambda}_{\mathrm{A/P}} \begin{bmatrix} \log_{10}\{|\boldsymbol{G}(\boldsymbol{p})|\} - \log_{10}\{|\boldsymbol{G}_{\mathrm{M}}|\} \\ \angle\{\boldsymbol{G}(\boldsymbol{p})\} - \angle\{\boldsymbol{G}_{\mathrm{M}}\} \end{bmatrix} \right|\right|_2^2. \tag{16}$$

The advantage is that in the log scale a reasonable weighting of amplitude and phase can be defined. The penalty depends less on the absolute size of the amplitudes, which eliminates the dependence on testbed properties. For example π and 1 can be chosen as weightings for amplitude and phase, respectively. Then $20\,\mathrm{dB}$ difference in the amplitude has an equivalent effect as $180°$ in the phase.

The sensitivity matrix can be calculated symbolically, if analytic expressions exist for $\boldsymbol{S}_{\mathrm{r}}, \boldsymbol{S}_{\mathrm{i}}$:

$$\boldsymbol{S}_3 = \boldsymbol{\Lambda}_{\mathrm{A/P}} \begin{bmatrix} \boldsymbol{S}_{\mathrm{A3}} \\ \boldsymbol{S}_{\mathrm{P3}} \end{bmatrix}, \tag{17}$$

$$\boldsymbol{S}_{\mathrm{A3},kj} = \frac{1/\ln(10)}{|\boldsymbol{G}_k|^2}\left[\mathrm{Im}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{i},kj} + \mathrm{Re}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{r},kj}\right], \tag{18}$$

$$\boldsymbol{S}_{\mathrm{P3},kj} = \frac{1}{|\boldsymbol{G}_k|^2}\left[\mathrm{Re}\left\{\boldsymbol{G}_k\right\}\boldsymbol{S}_{\mathrm{i},kj} - \mathrm{Im}\left\{\boldsymbol{G}_k\right\}\boldsymbol{S}_{\mathrm{r},kj}\right]. \quad (19)$$

As said before, cases 2 and 3 tend to find the correct model parameters for multiple mass models easier and weighting is more robust in the third case. Therefore, in the following only the third case, (16) is used. A potential problem is, however, that the measured and simulated phase must be interpolated continuous beyond $\pm 2\pi$, which fails sometimes, especially for coarse frequency resolutions, high measurement noise or low damping.

D. Assessing practical identifiability

In the process of model selection models that are not practically identifiable are excluded. Criteria for assessing practical identifiability are not reviewed here but can be found in [21], [22], [23], [24], [4], [10], [25], [26], [11]. Only the criteria used in this paper are introduced briefly.

For a model structure to be valid all parameters must have a certain minimal importance according to the **msqr parameter importance index (PII)** given by [10]:

$$\delta_j^{\mathrm{msqr}} = \left|\left|\boldsymbol{S}_j \Delta p_j\right|\right|_2 \quad (20)$$

with $\boldsymbol{S}_j$ the column of the sensitivity matrix corresponding to parameter j. If this criterion is violated, at least one of the parameters is of little importance and cannot be identified precisely. Also, the model is not parsimonious. The parameter Δp_j is supposed to be a normalization constant in the unit of parameter p_j, for example the nominal value or a quarter of the range of reasonable values [10]. In [11] normalization is omitted. Here, the previously identified parameter value is chosen for Δp_j. Defining the threshold for this parameter is somewhat arbitrary. Gabor et.al [11] set it to four orders of magnitude below the maximum PII.

Furthermore, the **collinearity index** γ_k:

$$\gamma_k = \frac{1}{\sqrt{\lambda_k}}, \quad (21)$$

reveals (multi-)collinearity among parameters, which is critical if γ_k exceeds $5...20$ [10]. λ_k is the smallest singular value of $\tilde{\boldsymbol{S}}$, that is the matrix $\boldsymbol{S}$ with all columns normalized to unit length. This normalization ensures a clear separation of collinearity and the PIIs.

Alternatively, collinearity can be measured by the **scaled condition indices** [24]:

$$\tilde{\eta}_k = \frac{\mu_{\max}}{\mu_k}. \quad (22)$$

μ_k and $\mu_{\max}$ are the k-th and the maximal singular value of $\tilde{\boldsymbol{S}}$, respectively. Critical values are $10...30$ [24].

In the experiments of the next section a combination of these three exclusion criteria is used for model selection. The exact thresholds will be given together with the results.

Fig. 2. Testbed 1: Stacker crane

III. EXPERIMENTAL RESULTS

As an experimental validation the structure and parameter identification is demonstrated for two testbeds. Since these results are highly dependent on the chosen thresholds, this dependency is investigated more thoroughly afterwards.

A. Structure and parameter optimization

The structure optimization is applied to the two testbeds shown in Fig. 2 and 3. Testbed 1 is a stacker crane with $5.6\,\mathrm{m}$ mast height and $5\,\mathrm{m}$ length of the horizontal axis (x) along the shelf. All experiments are carried out on x with the vertical axis in a position of $2\,\mathrm{m}$. Testbed 2 has only one axis which is driven in direct drive. In Figs. 4 and 5 the FRFs of training and test are shown. They have both been recorded with stepped sine excitation, but different amplitudes.

Parameter ranges have been chosen $0...1000$ for stiffnesses, and $0...0.1$ for dampers and moments of inertias (physical units). If more prior knowledge is available, the ranges can be narrowed. Parameters are identified with particle swarm optimization which is parameterized with a number of particles that equals 200 times the number of parameters of the current model. In the structure identification a maximum of 4 masses are considered (120 different models) for which the calculation takes approx. seven hours while for a maximum of 3 masses it takes only $35\,\mathrm{min.}$, implemented in Matlab on an i7 4-core computer running at $3.7\,\mathrm{GHz}$ with $16\,\mathrm{GB}$ DDR4 RAM. The overproportionally long calculation time for 4-mass models can be explained by the fact that they often require numerical sensitivity matrix calculation due to TF complexity.

Fig. 3. Testbed 2: Linear positioning system

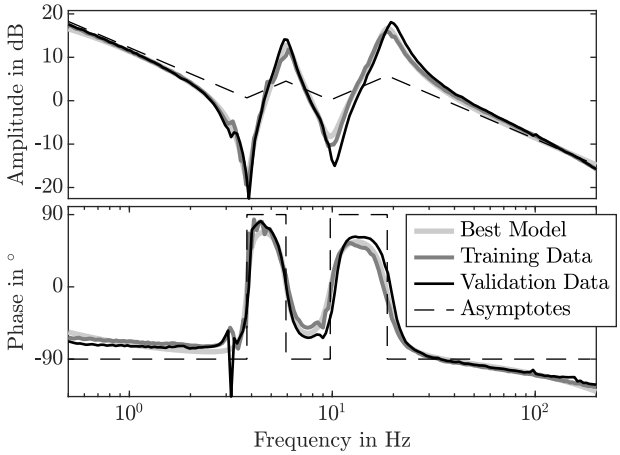

Fig. 4. Frequency response functions of testbed 1, measurement and best model

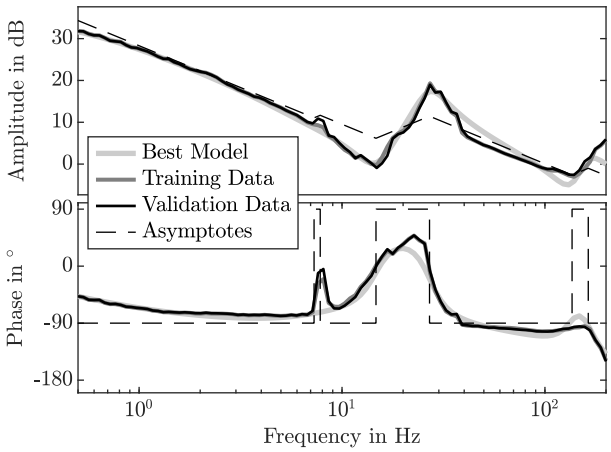

Fig. 5. Frequency response functions of testbed 2, measurement and best model

TABLE I

THREE CHOSEN SETTINGS FOR THE IDENTIFIABILITY CRITERIA AND CORRESPONDING BEST MODEL FOR BOTH TESTBEDS. THE FIRST ROW IN EACH SECTION CORRESPONDS TO THE FIRST SET OF SETTINGS AND SO ON.

Criterion	Threshold	Testbed 1	Testbed 2
Max. scaled	30	6.4	11.3
condition	10	6.4	10.3
index	10	6.4	2.6
$\frac{\text{Max. PII}}{\text{Min. PII}}$	1000	10.2	12.0
	1000	10.2	7.7
	1000	10.2	7.5
Max.	20	4.0	7.0
collinearity	10	4.0	7.0
index	5	4.0	1.9

In Figs. 4 and 5 the best model is also shown, including asymptotes. The chosen thresholds for the three criteria are given in the first row of each section in Tab. I. For testbed 1 a 3-mass system with additional damper to the base at mass 2 and delay time results, 9 estimation parameters. For testbed 2 it is a 4-mass system with additional damper at mass 1 and delay time, 12 estimation parameters.

The effect of changing the thresholds has been studied for two alternative settings, see rows two and three in Tab. I. Testbed 1 is always assigned the same model, while for testbed 2 different models result (3-mass system with additional damper at mass 3 and delay time, 9 estimation parameters, for the second set of thresholds and a 2-mass system with additional damper at mass 2 and delay time, 6 estimation parameters, for the third set). The corresponding FRFs are not shown. So especially for testbed 2, which cannot be categorized visually very clearly, the algorithmic result also depends strongly on the thresholds.

B. Threshold dependencies

To further investigate the dependence on thresholds of the identifiability criteria each of the three criteria is applied once exclusively in Figs. 6 and 7. Its threshold is varied in steps from 1 to 10000. For each step the model with the lowest J is selected among all models that fulfil this criterion. The number of estimation parameters of the best model for this threshold is plotted. For reference, the thresholds of the first setting in Tab. I are indicated by dashed lines.

Clearly, a strong dependence on the thresholds exists and mostly the number of parameters increases as the threshold increases. Exceptions exist where models with fewer parameters are more critical regarding one of the criteria than models with one or two more parameters.

What is disguised by this figure is the interplay between the three criteria. E.g. a model may be highly collinear, while all parameters have approx. the same sensitivity or vice versa. Thus, it can happen that the resulting model is very simple although all criteria applied exclusively allow for more complex models. The combined result would be hard to visualize in a 2D plane. Also, it is not known how many models are ruled out by a certain criterion and a given threshold. In order to investigate this Figs. 8 and 9 show the

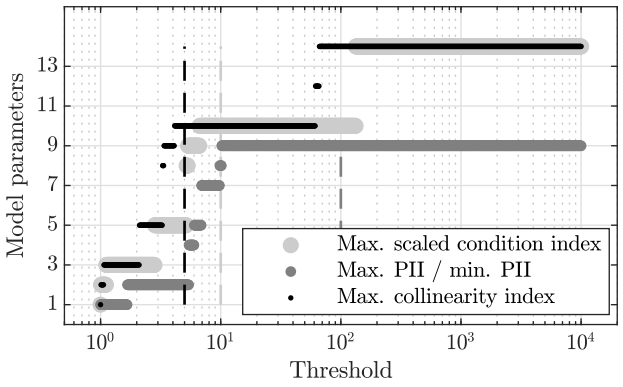

Fig. 6. Number of estimation parameters of the best model for a given threshold of the one current criterion, testbed 1

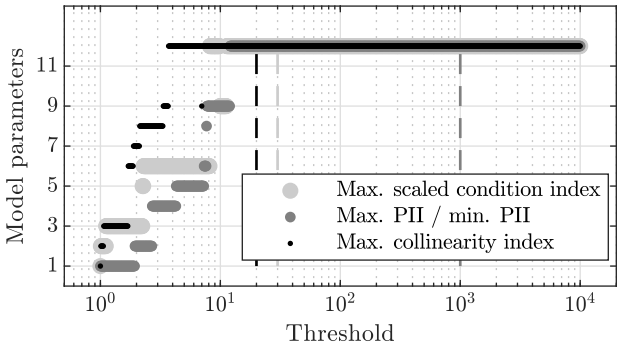

Fig. 7. Number of estimation parameters of the best model for a given threshold of the one current criterion, testbed 2

number of models that fulfil a certain criterion depending on the chosen threshold out of all 120 models.

The graphs of testbed 1 show a clear edge at approx. 60 models. Only if the thresholds are chosen below that, the number of valid models diminishes steeply. As expectable from the above results, testbed 2 shows a more gradual slope.

IV. DISCUSSION

This approach towards structure and parameter identification in frequency domain minimizes the difference between model and test data while considering criteria of practical identifiability based on local sensitivity analysis. It can handle different formulations of the cost function. The chosen model seems intuitively plausible, but it depends strongly on the chosen thresholds. Especially when different interpretations of the data are admitted by visual inspection the algorithm will also suggest different models depending on the thresholds. It is questionable if it is possible to choose the thresholds in advance without carefully analysing all criteria for a given model, also all scaled condition indices, not only

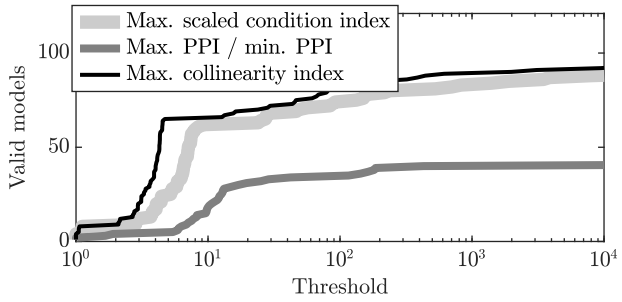

Fig. 8. Number of models that suffice the considered criterion depending on the threshold, testbed 1

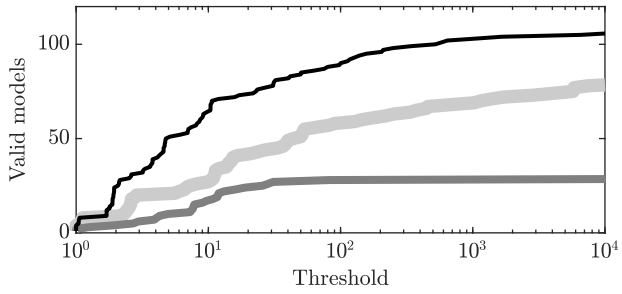

Fig. 9. Number of models that suffice the considered criterion depending on the threshold, testbed 2, legend as in Fig. 8

the maximum.

The selected model is not necessarily the best choice for a certain intended purpose such as feedforward, control design, but it can be used with confidence as a practically identifiable model with good fit. Any results obtained by this methods should be reviewed carefully regarding integrity.

What is disregarded is the model uncertainty, for example it is questionable if a model with the additional damper at a different mass may or may not lead to an equally good model. This aspect could further be investigated by calculating the Akaike weights [4] and structural distinguishability.

It was chosen to include only those submodels that have exclusively identifiable parameters. Although this is sound, it sometimes leads to unintuitive conclusions, e.g. when an elastically coupled mass is not included only because the damping parameter cannot be identified securely. Here, different strategies could be further investigated for the transition from parameter investigations to submodel inclusion/exclusion decisions.

V. CONCLUSIONS

A procedure for model selection in frequency domain has been proposed that minimizes the Kullback-Leibler distance while also maintaining practical identifiability of all estimation parameters. It aims at mechanical models of servo

systems including multiple-mass resonators. Criteria for practical identifiability are derived locally from the sensitivity matrix which is calculated for different formulations of the equation error. In most cases an analytic expression can be found and only for complex 4-mass models numeric differentiation is necessary.

Because of the tests for practical identifiability the dependence on knowledge of measurement noise is less critical. This is an advantage because often this knowledge is not available and by averaging the noise level can be reduced arbitrarily.

In experiments with two industry-like testbeds the method proves to reveal the characteristic mechanical properties of the two setups. Due to normalization thresholds for the criteria of practical identifiability can be chosen almost independently of the testbed properties but the results are still depended on the exact choice.

ACKNOWLEDGMENT

This work was sponsored by the German Forschungsvereinigung Antriebstechnik e.V. (FVA) and the AiF Arbeitsgemeinschaft industrieller Forschungsvereinigungen "Otto von Guericke" e.V.

REFERENCES

[1] F. Schütte, *Automatisierte Reglerinbetriebnahme für elektrische Antriebe mit schwingungsfähiger Mechanik*. Shaker, 2003.

[2] M. Witczak, A. Obuchowicz, and J. Korbicz, "Genetic programming based approaches to identification and fault diagnosis of non-linear dynamic systems," *International Journal of Control*, vol. 75, no. 13, pp. 1012–1031, 2002.

[3] J. D. Stigter and M. B. Beck, "A new approach to the identification of model structure," *Environmetrics*, vol. 5, no. 3, pp. 315–333, 1994.

[4] K. P. Burnham and D. R. Anderson, *Model selection and inference - a practical use of the information-theoretic approach*. New York: Springer, 1998.

[5] Y.-Y. Chen, P.-Y. Huang, and J.-Y. Yen, "Frequency-domain identification algorithms for servo systems with friction," *IEEE transactions on control systems technology*, vol. 10, no. 5, pp. 654–665, 2002.

[6] S. Villwock, "Identifikationsmethoden für die automatisierte Inbetriebnahme und Zustandsüberwachung elektrischer Antriebe," Ph.D. dissertation, Universität Siegen, 2007.

[7] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Statistical Science*, vol. 44, no. 4, pp. 382–417, 1999.

[8] C. T. Volinsky, D. Madigan, A. E. Raftery, and R. A. Kronmal, "Bayesian model averaging in proportional hazard models: Assessing stroke risk," Department of Statistics, University of Washington, Tech. Rep. 302, 1996.

[9] D. Draper, "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 45–70, 1995.

[10] R. Brun, P. Reichert, and H. R. Künsch, "Practical identifiability analysis of large environmental simulation models," *Water Resources Research*, vol. 37, no. 4, pp. 1015–1030, 2001.

[11] A. Gábor, A. F. Villaverde, and J. R. Banga, "Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems," *BMC systems biology*, vol. 11, no. 1, p. 54, 2017.

[12] A. M. Bizeray, J.-H. Kim, S. R. Duncan, and D. A. Howey, "Identifiability and parameter estimation of the single particle lithium-ion battery model," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 5, pp. 1862–1877, 2018.

[13] F. Schütte, S. Beineke, H. Grotstollen, N. Fröhleke, U. Witkowski, U. Rückert, and S. Rüping, "Structure-and parameter identification for a two-mass-system with backlash and friction using a self-organizing map," in *European Conference on Power Electronics and Applications*, vol. 3, 1997, pp. 3–358.

[14] L. Pronzato and A. Pázman, "Design of experiments in nonlinear models," *Lecture notes in statistics*, vol. 212, 2013.

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

[16] C. Chatfield, "Model uncertainty, data mining and statistical inference," *Journal of the Royal Statistical Society: Series A*, vol. 158, no. 3, pp. 419–444, 1995.

[17] M. Tantau, M. Wielitzka, and T. Ortmaier, "Structure and parameter identification of process models with hard non-linearities for industrial drive trains by means of degenerate genetic programming," in *ICINCO*, vol. 2, Praque, Czech Republic, 2019, pp. 368–376.

[18] Q. D. Vu, "Parameter estimation in complex nonlinear dynamical systems," PhDThesis, Technische Universität Ilmenau, 2015.

[19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[20] R. J. Galarza, J. H. Chow, and R. C. Degeneff, "Transformer model reduction using time and frequency domain sensitivity techniques," *IEEE transactions on power delivery*, vol. 10, no. 2, pp. 1052–1059, 1995.

[21] D. E. Farrar and R. R. Glauber, "Multicollinearity in regression analysis: the problem revisited," *The Review of Economic and Statistics*, pp. 92–107, 1967.

[22] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, vol. 12, no. 3, pp. 591–612, 1970.

[23] G. W. Stewart *et al.*, "Collinearity and least squares regression," *Statistical Science*, vol. 2, no. 1, pp. 68–84, 1987.

[24] D. A. Belsley, "A guide to using the collinearity diagnostics," *Computer Science in Economics and Management*, vol. 4, no. 1, pp. 33–50, 1991.

[25] P. Kovács, T. Petres, and L. Tóth, "A new measure of multicollinearity in linear regression models," *International Statistical Review*, vol. 73, no. 3, pp. 405–412, 2005.

[26] K. Nienałtowski, M. Włodarczyk, T. Lipniacki, and M. Komorowski, "Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics," *BMC systems biology*, vol. 9, no. 1, p. 65, 2015.