DOI: 10.58895/ksp/1000162754-6 erschienen in:
Proceedings. 33. Workshop Computational Intelligence - Berlin, 23. – 24. November 2023
DOI: 10.5445/KSP/1000162754 | https://www.ksp.kit.edu/site/books/e/10.5445/KSP/1000162754/

Physical Interpretability of Data-Driven State Space Models

Hermann Klein, Max Schüssler, Oliver Nelles

Universität Siegen, Department Maschinenbau,
Institut für Mechanik und Regelungstechnik – Mechatronik
Paul-Bonatz-Str. 9-11, 57068, Siegen, Germany
E-Mail: {hermann.klein,oliver.nelles}@uni-siegen.de
research@maxschuessler.de

1 Introduction

Nonlinear state space models are powerful model architectures for system identification. They provide the necessary flexibility for the description of nonlinear dynamic processes while still maintaining in a quite compact respresentation. Typically, the data-driven state space models are black-box models, a fact that causes shortcomings regarding interpretability [9]. Since the modeling performance is satisfying and competitive to recurrent neural networks [5], we strive for an increase in interpretability. Interpretability in terms of physical insights can be achieved by the incorporation of prior process information. This leads to a gray-box identification strategy. Eventually, the goal of this contribution is to combine both modeling power and the interpretability of its parameters.

Gray-box methods require prior process knowledge besides measurement data. Prior knowledge can be available in different forms. This work studies the case in which the process structure is a priori known in form of a physical equation. If a structured identification is carried out, the model is forced into an explainable form. Then, both the modeling task and the interpretability problem are solved. The central idea of the proposed method is the reduction of the model parameters down to the physically required number.

An advantageous architecture for the structured identification is the Local Model State Space Network (LMSSN), developed by Schüssler [6]. Its local linear behavior supports the desire for interpretability, because the well-known foundations of linear system theory can be applied.

2 Gray-Box Modeling with the Local Model State Space Network

LMSSN is an extension of the linear time-discrete state space for modeling nonlinear processes. Detailed information about LMSSN can be found in [6]. On the one hand, it can be seen as a state space framework in which the multidimensional nonlinear functions in the state and output equations are implemented with two Local Model Networks (LMN) [1]. On the other hand, it can be seen as a deep neural network consisting out of one recurrent layer and a dense layer. Expressed with the above mentioned LMNs, a single-input single-output LMSSN of order n_x with the input $u(k)$, the state $\hat{x}(k)$ and the output $\hat{y}(k)$ is defined by

$$\begin{aligned} \underline{\hat{x}}(k+1) &= \sum_{j=1}^{n_{\text{state}}} \left(\underline{o}_j^{[\text{d}]} + \underline{A}_j^{[\text{d}]}\underline{\hat{x}}(k) + \underline{b}_j^{[\text{d}]}u(k)\right) \cdot \Phi_{\text{state},j}(k) \\ \hat{y}(k) &= \sum_{j=1}^{n_{\text{out}}} \left(p_j + \underline{c}_j^{\text{T}}\underline{\hat{x}}(k) + du(k)\right) \cdot \Phi_{\text{out},j}(k). \end{aligned} \tag{1}$$

Here, $\underline{o}_j^{[\text{d}]}$ is the offset of the state equation, $\underline{A}_j^{[\text{d}]}$ and $\underline{b}_j^{[\text{d}]}$ can be interpreted as the slopes of the j-th model of the state equation. Accordingly, $\underline{p}_j^{[\text{d}]}$ is the offset of the output euqation and $\underline{c}_j^{[\text{d}]}$ and $\underline{d}_j^{[\text{d}]}$ are the slopes of the j-th model of the output equation. (The superscript $(\cdot)^{[\text{d}]}$ marks the discrete-time description.) There are altogether n_{state} superposed affine models in the state equation network and n_{out} in the output equation network. The basis functions $\Phi_j(k)$ express the j-th local validity. They are realized with normalized Gaussians and generate a global nonlinear function by superposition of the local affine models. Due to the fact that these local state space models are fully-parameterized, LMSSN is a black-box model.

In the following, the steps of structured identification are presented. The procedure requires a novel initialization technique for LMSSN models. Normally, LMSSN is initialized as a linear state space model, before it is divided into serveral local models with the help of the Local Linear Model Tree (LOLIMOT) or the Hierachical Local Model Tree (HILOMOT) algorithm [1]. The initial model for this tree-construction algorithm is the Best Linear Approximation (BLA) of the process estimated from input-output data $\{u(kT_0), y(kT_0)\}$. It is generated via a subspace-based system identification method [2] and results in a linear fully-parameterized model. If the model is restructured and restricted, it leads to a linear gray-box model. With the help of LOLIMOT, a nonlinear gray-box model is then derived by splitting the extended input space $\underline{\tilde{u}} = [\underline{\hat{x}}, u]^{\mathrm{T}}$ and adding local models. Note that the structure of the initial model is kept as splitting progresses and it is passed to the local models generated by LOLIMOT. The restructuring step will be described more closely in the following. Figure 1 shows the workflow for gray-box structured identification.

Canonical state space forms like the Canonical Controllable Form (CCF) are easy to achieve via similiarity transformations [8]. An arbitrary gray-box structure requires a more sophisticated restructuring strategy because the transformation can not be calculated directly. Therefore, three possible methods are stated. A nonlinear unconstraint optimization with an additional penalty term, called Penalty Method (PM) [7] can force the free parameters to their desired values. Alternatively, a classical gray-box technique is the Prediction Error Method (PEM) [6]. It uses hard constraints for implementation of the known parameters $\underline{\theta}_{\mathrm{con}}$ by placing them in the model. Here, the fully-parameterized black-box model based on BLA is only necessary for initialization. Another alternative is to estimate a specific transformation matrix that leads to the desired state space form [4].

After the restructuring step, only the free parameters will be optimized while the constrained parameters are kept "frozen". In the following step, LOLIMOT generates a nonlinear global model by partitioning $\underline{\tilde{u}}$. Finally, the described procedure yields a nonlinear model containing the desired gray-box structure. As a post-processing step, the physical parameters can be extracted, which is useful for analysis and gives insights into process.

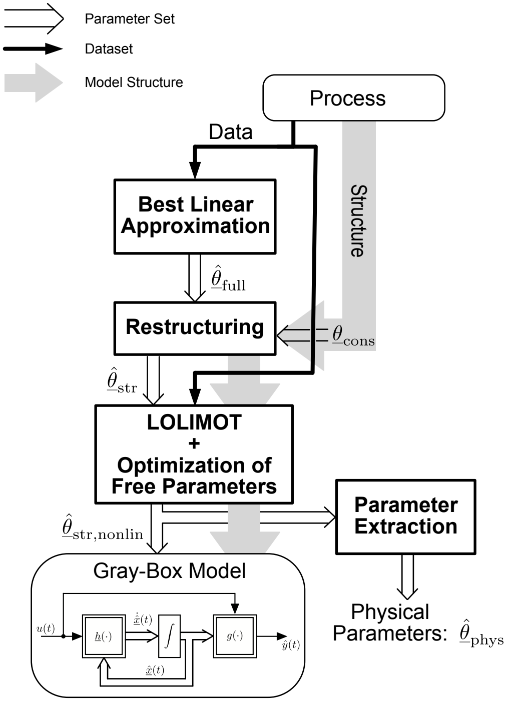
Parameter Set
Dataset
Model Structure
Process
Data
Structure
Best Linear Approximation
$\hat{\underline{\theta}}_{\mathrm{full}}$
Restructuring
$\underline{\theta}_{\mathrm{cons}}$
$\hat{\underline{\theta}}_{\mathrm{str}}$
LOLIMOT + Optimization of Free Parameters
Parameter Extraction
$\hat{\underline{\theta}}_{\mathrm{str,nonlin}}$
Gray-Box Model
$u(t)$
$\underline{h}(\cdot)$
$\dot{\hat{x}}(t)$
$\hat{x}(t)$
$g(\cdot)$
$\hat{y}(t)$
Physical Parameters: $\hat{\underline{\theta}}_{\mathrm{phys}}$

Figure 1: Structured identification procedure. The BLA yields the parameter vector $\hat{\underline{\theta}}_{\text{full}}$. For restructuring the constraint parameters $\underline{\theta}_{\text{cons}}$ and their indexes are used. The structured parameters are in the vector $\hat{\underline{\theta}}_{\text{str}}$ while the parameters of the nonlinear model are in $\hat{\underline{\theta}}_{\text{str,nonlin}}$. The parameter extraction decodes $\hat{\underline{\theta}}_{\text{str,nonlin}}$ to the interpretable parameters $\hat{\underline{\theta}}_{\text{phys}}$.

3 Test Process

The proposed structured LMSSN is demonstrated on simulated data from a mechanical system which is a moving body with a single degree of freedom. The system's input $u(t)$ is the excitation force acting on the center of gravity while the output $y(t)$ is its position:

$$M\ddot{y}(t) + D\dot{y}(t) + \underbrace{F_{\mathrm{S}}(y)}_{=F_{\mathrm{off}}(e^{\gamma y}-1)} = u(t). \tag{2}$$

Here, the body's mass is M and the linear damping ratio is D. A static progressive spring curve $F_{\mathrm{S}}(y)$ leads to a nonlinear equation of motion. The stiffness characteristic is parameterized with the curve offset F_{off} and the exponential stiffness rate γ. For the excitation of the system, a step-like signal containing 96 events with random levels in the interval $[0\,\mathrm{N}; 1\,\mathrm{N}]$ is chosen. The numerical integration of the equation of motion, necessary for output calculation, is carried out with the Euler-forward method. After data generation, the output signal was artifically disturbed with additive white Gaussian noise with an signal-to-noise ratio of 49 dB.

Next, we state the specific prior knowledge of the process (see (2)) required for gray-box identification. In the present case, the gray-box knowledge is the information that the process can be approximated by a second order system whose numerator equals one (PT$_2$ system[1]). This knowledge is applied to the model. A favorable structure for the above mentioned task is a Nonlinear Controllable Form (NCF). Compared to CCF, the initial NCF has an additional offset in the state equation. The NCF is written as

$$\begin{aligned}\underline{\dot{x}}(t) &= \underbrace{\begin{bmatrix} 0 & 1 \\ \theta_{\mathrm{free},1} & \theta_{\mathrm{free},2} \end{bmatrix}}_{\underline{A}(\underline{\theta}_{\mathrm{free}})} \underline{x}(t) + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\underline{b}} u(t) + \underbrace{\begin{bmatrix} 0 \\ \theta_{\mathrm{free},3} \end{bmatrix}}_{\underline{o}(\underline{\theta}_{\mathrm{free}})}, \\ y(t) &= \underbrace{\begin{bmatrix} \theta_{\mathrm{free},4} & 0 \end{bmatrix}}_{\underline{c}^{\mathrm{T}}(\underline{\theta}_{\mathrm{free}})} \underline{x}(t) + \underbrace{\begin{bmatrix} 0 \end{bmatrix}}_{d} u(t) + \underbrace{\begin{bmatrix} 0 \end{bmatrix}}_{p}.\end{aligned} \tag{3}$$

For the sake of completeness, the linear case relations are stated as

$$\begin{aligned}&\theta_{\mathrm{free},1} = -\frac{C}{M}, \quad \theta_{\mathrm{free},2} = -\frac{D}{M}, \\ &\theta_{\mathrm{free},3} = 0, \quad \theta_{\mathrm{free},4} = -\frac{1}{M}.\end{aligned} \tag{4}$$

Here, $C = \text{const.}$ is the stiffness of a hypothetical linear spring.

[1] A PT$_2$ system is a second order transfer function without zeros.

4 Parameter Extraction

The final step is the interpretation of the model parameters in a physical manner, compare Fig. 1. Regarding the nonlinear equation, these are the body's mass, damping ratio and stiffness. The parameter extraction is able to deliver the spring force $\hat{F}_\mathrm{S}(y)$ as a function dependent on the position y. Since we modeled with local affine functions, we find $\hat{F}_\mathrm{S}(y)$ as the weighted sum of n_LM local affine stiffness functions $\hat{F}_{\mathrm{affine},i}(y)$, as

$$\begin{aligned}\hat{F}_\mathrm{S}(y) &= \sum_{i=1}^{n_\mathrm{LM}} \hat{F}_{\mathrm{affine},i}(y)\Phi_i(y) \\ &= \sum_{i=1}^{n_\mathrm{LM}} (\hat{C}_{\mathrm{lin},i}y + \hat{C}_{\mathrm{off},i})\Phi_i(y).\end{aligned} \tag{5}$$

Here, $\hat{C}_{\mathrm{lin},i}$ is the linear stiffness and $\hat{C}_{\mathrm{off},i}$ the offset of the i-th local stiffness model. With (3), we can extract $\hat{C}_{\mathrm{lin},i}$ and $\hat{C}_{\mathrm{off},i}$ from the estimated model parameters

$$\hat{C}_{\mathrm{lin},i} = -\frac{\hat{\theta}_{1,i}}{\hat{\theta}_4}, \quad \hat{C}_{\mathrm{off},i} = -\hat{\theta}_{3,i}. \tag{6}$$

Alternatively, the function $\hat{\theta}_1(y)$ which describes the variation of the parameter $\hat{\theta}_1$ with $y(t)$ can be constructed from (5) and (6) as weighted sum,

$$\hat{\theta}_1(y) = \sum_{i=1}^{n_\mathrm{LM}} \hat{\theta}_{1,i}\Phi_i(y). \tag{7}$$

The left side of Fig. 2 visualizes the extracted validity functions and $\hat{\theta}_1(y)$. Additionally, on the right is the true spring curve plotted, which has been fitted accurately by the LMN.

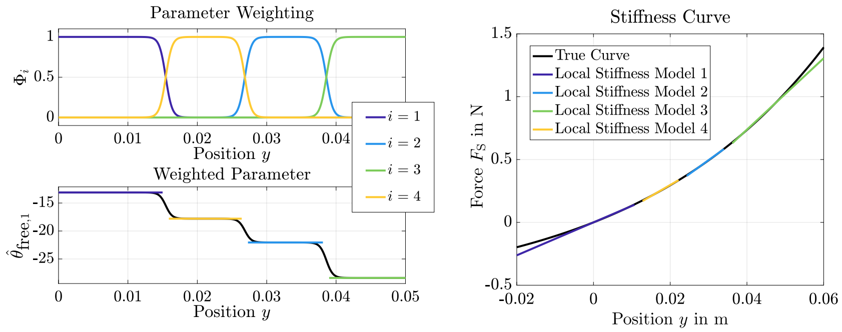

Figure 2: Local Linear Stiffness

5 Conclusion

This contribution emphasized that a nonlinear data-driven state space model with physically inspired structure is able to combine interpretability with high performance. Furthermore, computational resources can be preserved compared to an unstructured black-box model because the gray-box model contains less parameters that have to be optimized.

We are able to demonstrate our apporach on a widely known but simple example process. Next, our gray-box method shall be expanded on more complex processes like the Bouc-Wen hysteresis benchmark.

References

[1] O. Nelles. “Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes”. Springer International Publishing. 2020.

[2] T. McKelvey, H. Akcay, L. Ljung. “Subspace-based multivariable system identification from frequency response data”. In: *IEEE Transactions On Automatic Control* (**41**, 960-979). 1996.

[3] M. Schüssler. “Machine learning with Nonlinear State Space Models”. In: *Schriftenreihe Der Arbeitsgruppe Mess- Und Regelungstechnik - Mechatronik, Department Maschinenbau*. 2022.

[4] G. Mercère, O. Prot, J. Ramos. “Identification of Parameterized Gray-Box State-Space Systems: From a Black-Box Linear Time-Invariant Representation to a Structured One”. In: *IEEE Transactions On Automatic Control*. (**59**, 2873-2885). 2014.

[5] M. Schüssler, T. Münker, O. Nelles. “Deep Recurrent Neural Networks for Nonlinear System Identification”. 2019.

[6] L. Ljung. “System Identification: Theory for the User.” Prentice-Hall. 1986.

[7] J. Nocedal, S. Wright. “Numerical Optimization”. Springer. 2006.

[8] B. Friedland “Control System Design: An Introduction to State-Space Methods”. Dover Publications. 2012.

[9] P. Parrilo, L. Ljung. “Initialization of Physical Parameter Estimates”. In: *IFAC Proceedings Volumes*. (**36**, 1483-1488). 2003.