Master's Thesis

Analysis and Enhancement of an Iterative Identification Approach for Nonlinear Systems

Author: Patrick Schrangl

Supervisor: Luigi del Re
Harald Waschl

Finished: March 2015

Abstract

Mathematical models for real systems are required for different tasks in engineering, such as predicting or controlling a system's behavior. Many systems are too complex for describing them using physical knowledge, so data-based modeling techniques provide a sensible alternative if few information is given. However, these methods require suitable identification data. Additionally, many real-world systems show nonlinear behavior, therefore this work concentrates on the identification of nonlinear dynamic systems. Previous research shows that the class of polynomial models is suitable to describe many nonlinear problems, such as engine emissions. This work uses an iterative identification approach which allows to identify these models. A D-optimal design of experiments approach is used to excite the system in order to achieve high-quality models while minimizing the effort for acquiring data. Regarding model complexity, it is beneficial to select only significant model regressors in order to limit complexity and avoid overfitting. An enhanced iterative identification approach that is able to identify systems online, i. e. in real-time whilst capturing measurement data, is developed in this work. The pursued approach allows different regressor selection methods to reduce model complexity and improve model quality. Moreover, in this work several commonly used model quality assessment criteria are compared against each other. The algorithm is analyzed by means of simulation studies on generic polynomial systems. The proposed methods are combined in a Matlab toolbox which is evaluated on real-world automotive applications and compared to previously identified models in order to show its effectiveness.

Introduction

The proposed approach uses polynomial nonlinear ARX (PNARX) models, which are ARX (auto-regressive with exogenous input) models that include polynomial combinations of inputs and outputs in the data matrix $\boldsymbol{\Phi}$. This model class is a discrete-time model of the form

$$y_k = \boldsymbol{\varphi}_k^\mathsf{T}\boldsymbol{\theta} + e_k,$$

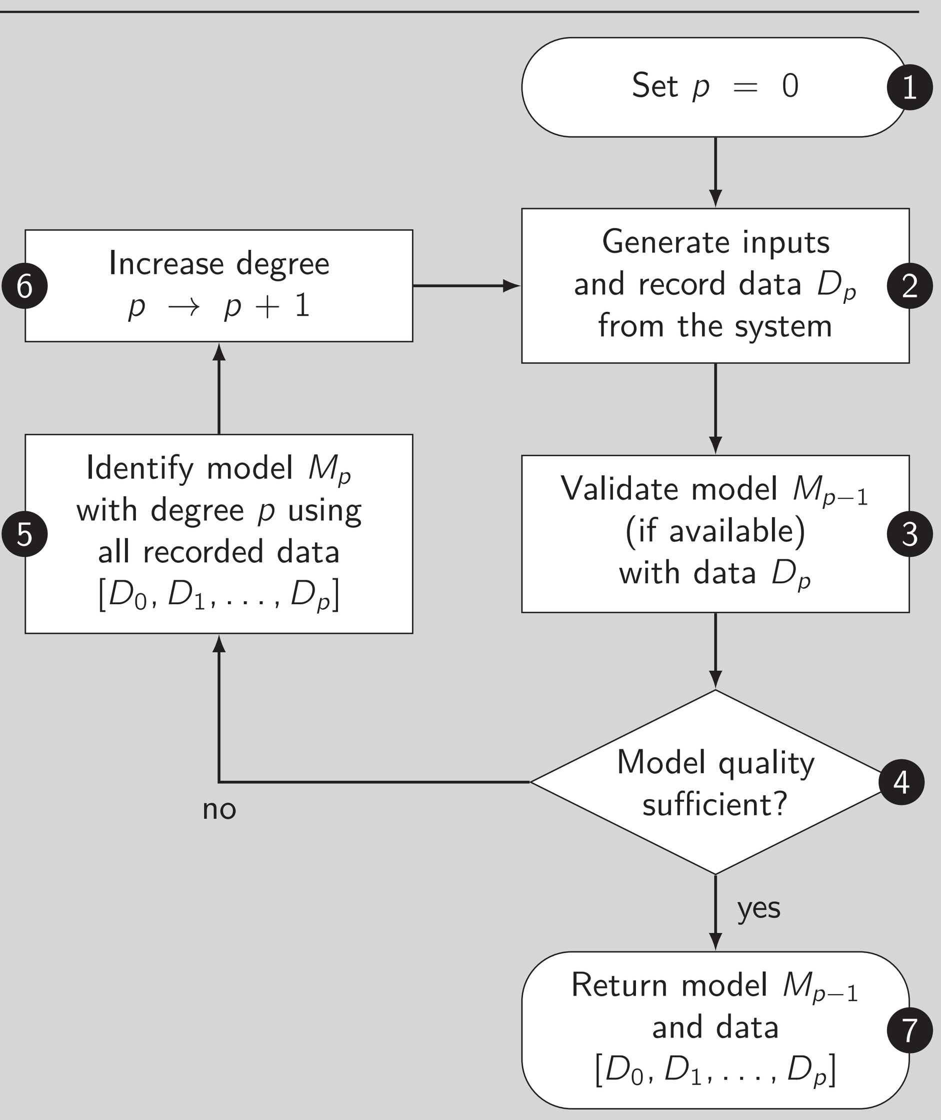
where y denotes a system output, φ denotes the regressors (polynomial combinations of degree p of inputs and outputs), $\boldsymbol{\theta}$ denotes the unknown parameter vector and e denotes the model error. Assuming N recorded data values, the resulting matrix equation, $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{e}$, is *linear-in-parameters*, thus it can be solved using the *least squares* (LS) method, yielding

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{y}.$$

Especially for the identification of nonlinear models suitable identification data is needed both to reduce measurement time and costs and to obtain high-quality models. Therefore the system should be excited in an *optimal* way. In this approach, an iterative static *D*-optimal *design of experiments* (DOE) strategy is applied to generate sufficiently exciting input data. This approach maximizes the determinant of the information matrix $\boldsymbol{H} = \boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}$, which leads to a well-invertible matrix $\boldsymbol{H}$ and therefore to a precise estimation of the model parameters $\boldsymbol{\theta}$.

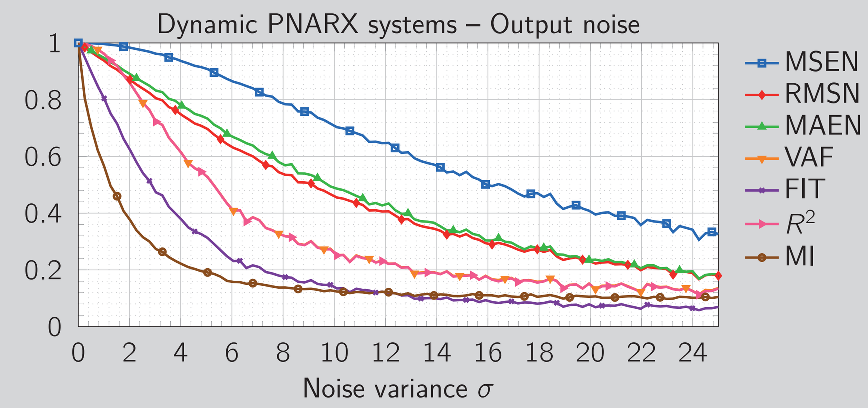
Identification Algorithm

The iterative identification approach, where the degree p of the polynomial model is increased iteratively until the model quality is sufficient, is depicted to the right. The algorithm consists of two main parts: 1. Generate input data (DOE) and record input and output data from the system, 2. Model identification and model reduction (regressor selection). The algorithm combines identification and validation. Note that the termination depends on the model quality. Therefore different model assessment criteria need to be evaluated and tested. Prior to model identification (regressor selection) is performed to address the problem of overfitting, which occurs especially for higher polynomial degrees due to a huge number of possible regressors.



Assessment Criteria

The properties of different model assessment criteria have been evaluated. The considered criteria are five error-based criteria: normalized mean squared error (MSEN), normalized root mean squared error (RMSN), normalized mean absolute error, (MAEN), fit-value (FIT) and coefficient of determination (R^2). Additionally, two stochastic criteria, variance accounted for (VAF) and mutual information (MI), have been assessed using different systems (linear, static polynomial and dynamic polynomial) through a simulation study. The criteria have been evaluated with respect to their sensitivity on: **noise** (normally distributed), **offset** errors, **scaling** errors and **nonlinear** transformations. The simulation results are briefly summarized in the plot and table below.



Criterion	Sensitive on		
	Offset	Scaling	Nonlinearity
Error-based[1]	yes	yes	yes
VAF	**no**	yes	yes
MI	**no**	**no**	**slightly**

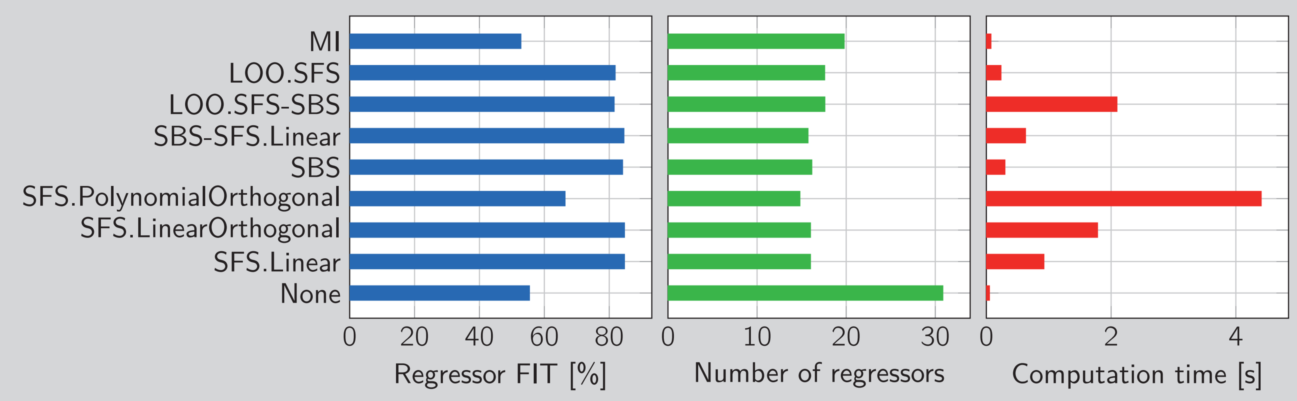
[1] MSEN, RMSN, MAEN, FIT and R^2

Regressor Selection

When dealing with polynomial regression models the number of possible regressors increases very fast with the polynomial degree, especially when the dynamic system orders are high. This immediately leads to overfitting problems. Therefore different systematic methods for selecting the most significant regressors, based on the captured data, were implemented and evaluated by a simulation study:

- Sequential forward selection (SFS) algorithms
- Sequential backward selection (SBS) algorithms
- Mutual information (MI) based regressor selection

The plot shows the mean results of the simulation on static PNARX models. It could be shown that regressor selection plays an important role when identifying PNARX models, however, a single *best* method could not be unveiled in this study. Generally, the MI method is not suitable for this problem, but could be used for different tasks such as input pre-selection.

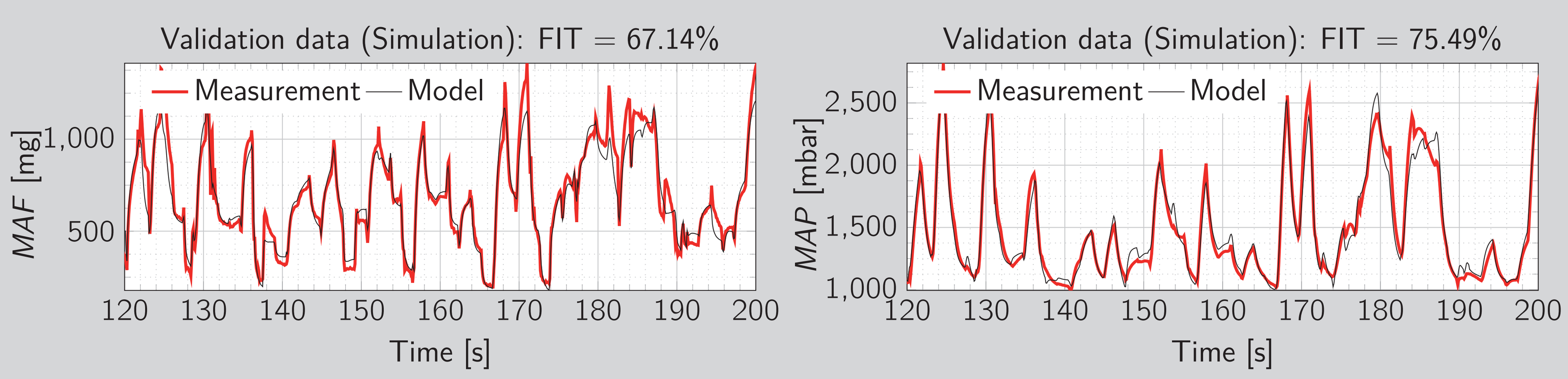


Applications

The implemented toolbox and the presented algorithms have been applied to two real-world systems in the field of automotive engineering

- the air path of a BMW N47 Diesel engine (dynamic model)
- the stationary exhaust gas emissions of an Otto engine (static model)

For the air path model four inputs, i. e. engine speed, exhaust gas recirculation (EGR) valve position, variable geometry turbine guide vane position and fuel injection amount have been used to create a model of the two outputs: intake manifold air mass flow (MAF) and intake manifold air pressure (MAP). A polynomial degree of $p = 2$ turns out to be sufficient. Exemplary validation plots of the outputs are depicted in the following figures.



Conclusion and Outlook

- **Implementation** of the approaches in the form of a generic Matlab toolbox
- **Simulation** by means of generic static and dynamic PNARX systems
- Implementation and comparison of different **model assessment criteria** and different **regressor selection methods**
- Testing the approach on **real automotive systems** (engine air path and engine emissions)
- **Outlook:** dynamic DOE approach; automatic approaches for estimating system orders and pre-selecting inputs (e. g. MI); toolbox extensions (e. g. GUI)