

Sequential Identification of Engine Subsystems by Optimal Input Design

Markus Hirsch and Luigi del Re

Institute for Design and Control of Mechatronical Systems, Johannes Kepler University Linz, Austria

Copyright © 2009 SAE International

ABSTRACT

Complexity and nonlinearity of engines makes precise first principle engine models often difficult to obtain, as for instance for emissions. System identification is a well known possible alternative, successfully used in several automotive applications. In most cases system identification is concerned with the estimation of the unknown parameters of a known set of equations. Unfortunately, for many engine subsystems, there is no sufficiently precise or real time suitable model. This paper presents a sequential algorithm which allows to derive real time suitable models on line by a combination of model structure hypothesis of increasing complexity and an associated optimal input design and selection process. This paper introduces the method and shows its use both for a rather simple and a very difficult engine identification task, a dynamical model of the airpath of a Diesel engine and a dynamical model of nitrogen oxides and particulate matter.

INTRODUCTION

The increasing number of possible control actions in modern engines allows a much better optimization of the combustion, but complicates the search for the optimal set points in terms of fuel consumption, emissions, efficiency etc. Control oriented models of these highly nonlinear processes would be very helpful, and for some subsystems, for instance the air path, there is a wide choice. Others, like emission models, are still underdeveloped. In all cases, it would be very important to be able to derive them in a short time and with a limited effort.

This paper proposes an answer to this question. Like many others, it concentrates on nonlinear data based models, but starts from the consideration that the quality of such models depends as much on the specific data used as on the specific model assumptions. Indeed, in some cases, like air path, sensible assumptions on a control oriented nonlinear dynamic model can be found, but in others, like emissions, it is hardly possible. Especially for these cases, practical experience shows that taking a different measurement leads not only to different parameter estimates but also to choose a different model structure. This leads to the natural question: is there any "perfect" choice of the

measurement to be used? And can we obtain a model which works everywhere? In the linear case and for the known structure, the question is essentially solved, but the same does not hold for nonlinear systems, besides the trivial and impossible answer to "measure all cases" as some approaches, as the artificial neural networks, essentially do.

This paper proposes an iterative combination of optimal as well as robust input design and model parameter estimation to achieve both high precision models and limited measurement time. The main idea is to start with a simple model, to use a modified D-optimal design of experiments and optimal parameter estimation to determine the best representation of the system for this specific model class, to perform the corresponding measurements. If the quality of the resulting model is not sufficient, then the procedure can be repeated iteratively for model classes of increased complexity.

The paper is structured as follows: the next chapter presents the structure of the polynomial NARX model as well as the basic parameter estimation algorithm; chapter three describes the input design procedure, followed by the description of the iterative combination of parameter estimation and input design. Chapter five presents the application of theory on a test bench engine by identifying firstly a model for the airpath states, manifold absolute pressure (MAP) and mass air flow (MAF) as a result of engine speed, total amount of injected fuel mass and levels of the exhaust-gas recirculation valve (EGR) and the variable geometry turbine actuator (VGT). Secondly also the identification of models for engine emissions of particulate matter (PM) and nitrogen oxides (NO_x) are presented. All measurements are performed and validated on a production 2 liter 4 cylinder CRDI Diesel engine on a dynamical engine test bench.

DATA BASED MODEL IDENTIFICATION

System identification is usually understood as the optimal estimation of parameters of a candidate model using a given data set. If no a priori information is available, many model classes offering sufficient approximation properties are suitable candidates. In our case, we concentrate on NARX models with polynomial nonlinearities (cp. :[1])

$$\begin{aligned}
y[k] &= \theta_0 + \sum_{m=1}^l \sum_{i_1 \dots i_m=1}^n \theta_{i_1 \dots i_m} \cdot \prod_{p=1}^m x_{i_p}[k] + \varepsilon[k] \\
x[k] &= [y[k-1] \dots y[k-n_a] \ u_1[k] \dots u_1[k-n_{b_1}] \dots \\
&\quad u_2[k] \dots u_2[k-n_{b_2}] \dots u_r[k] \dots u_r[k-n_{b_r}]]' \\
n &= n_a + \sum_{h=1}^r n_{b_h}
\end{aligned} \quad (1)$$

Eq. (1) presents the structure of the full factorial polynomial NARX model of maximum order l for a system with r inputs, autoregressive terms up to an order n_a and input delays up to an order n_b . This structure indicates that the actual output $y[k]$ is defined by polynomial combinations of past model outputs and measured values of past model inputs $u_i[k-j]$ given by the model and an actual model error $\varepsilon[k]$. Notice that this structure is linear in parameters; therefore it can be written in the form

$$\begin{aligned}
y[k] &= \Theta' \cdot X[k] + \varepsilon[k] \\
\Theta &= [\theta_0 \ \theta_1 \dots \theta_n \ \theta_{n_1} \ \theta_{n_2} \dots \theta_{nn} \dots \theta_{n'}]' \\
X[k] &= [1 \ x_1[k] \dots x_2[k] \ x_1[k]x_1[k] \ x_1[k]x_2[k] \dots \\
&\quad x_n[k]x_n[k] \dots x_{n'}[k]]'
\end{aligned} \quad (2)$$

and in matrix notation for a whole time series respectively

$$Y = \Theta' X + E \quad (3).$$

Here X is the data matrix containing the elements of the polynomial and Y , the output vector, while E is the unknown error vector. Notice also that the dimension of the model with increasing order l of a full polynomial expansion will typically increase extremely fast, with dire effects both on the condition of the associated optimal estimation problem and/or on the measurement effort. Therefore a reduction of regressors due to physical knowledge on the system can be very helpful (an example for that is given at the section applications).

Once a model structure has been defined corresponding to special choices of the parameter set in (1), the values of the parameters can be estimated by optimization. Since the model is linear in the parameters these estimates can be computed using standard least squares, if standard conditions on the error hold, otherwise numerical methods can be used to reduce bias. In the best case, estimation of Θ can be done explicitly by inversion of the Information matrix $M = X'X$.

$$\hat{\Theta} = M^{-1} \cdot X' \cdot Y \quad (4)$$

In general eq. (1) for a polynomial system will contain many elements of the regression vector X of little relevance. Using parameter selection procedures, (e.g. AIC or BIC criteria – cp. [2-3]), the set of remaining terms can be simplified. This can strongly improve the condition of M and so the quality of the estimate of Θ .

INPUT DESIGN

The idea of input design is to have an input sequence $u(t)$ that excites the system preferably with the minimal length and makes identification of each terms of Θ separable in some sense. Random signals (e.g. white noise or pseudo random binary signals e.g. [4]) are known as good candidates to this end, especially for linear systems. Indeed, studies like [5] have confirmed that simultaneous and independent random excitations present the best input signals. Unfortunately, this occurs only after a rather long time (in theory infinity), because it is essentially a blind excitation – all possible components are excited.

However, if something more is known on the system to be identified, this information can be used to tailor the signal to it. In our case, we do not have any information on the real model structure but if we assume the indices l , n_a and n_b to be sufficiently large, in general we may expect our system to be represented sufficiently well by the polynomial NARX approach given in (1). The task of our algorithm is to enforce a consistent estimation of the minimum complexity model belonging to the class of eq. (1) up to a given precision level.

To this end, let us recall that an essential requirement for the explicit least square identification is the invertibility of the information matrix M , where according to (1) and (2) correlations of the inputs u and past outputs y are included:

$$M = X'X = \begin{bmatrix} \sum_{k=1}^N X_1[k]^2 & \sum_{k=1}^N X_1[k]X_2[k] & \dots \\ \sum_{k=1}^N X_1[k]X_2[k] & \sum_{k=1}^N X_2[k]^2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad (5)$$

Optimal input design algorithms define input signals so that inputs operate in a pre-defined valid region Ω and thereby maximize a given criterion $J(M)$, usually related to the condition of M , a common one being the D-criterion, defined as $J(M) = \det(M)/N$ where N is the number of measurement points (for some alternative proposals see [6]). The input sequence can be computed by solving the problem

$$u(t) = \underset{u(t)}{\operatorname{argmax}}(\det(M)), \quad u \in \Omega \quad (6)$$

The solution of this problem is not trivial for several reasons:

The problem is a non convex one, as there exists in general not only one input sequence for optimal excitation of a regressor set. There exist different ways to solve this problem: One is that Ω consists only of a discrete set of measurement candidates from which points can be selected according to (6) (see [7] and

references therein). Another possibility of finding the global maximum of $J(M)$ within a continuous range Ω is doing some multi-shoot optimizations where optimizations starts on many different random points and best results are assumed to be close to the global optimum. We found desired input points by numerically solving (6) starting on a random point within Ω , though we knew that probably only a local maximum was found which presents a good (but not the best) candidate for input design too.

Another problem is that in general autoregressive terms are considered in (1), which means that the matrix M includes past values of the output measurements which cannot be considered during input design directly. To solve this problem some knowledge of the system has already been given: By this knowledge an estimation of the future outputs can be done and values therefore influence the input design. Another possibility is to neglect autoregressive terms in the input design but to excite these effects by adapting the periodic time of input points. If this time is similar to the rising time of the outputs variations, excitation due to autoregressive terms should be sufficient.

SEQUENTIAL PROCEDURE

Standard proposals for input design are based on a given model structure. As this is not possible here, we start with a simple model structure and define inputs for optimal parameter estimation according to (6). After measurements are done, model parameters are identified and non-significant parameters are removed by parameter selection algorithms. Model validation on a data set not used for identification allows deciding whether the model accuracy is sufficient or not. If not, the model structure can be extended. This is done by increasing the polynomial order in (1). Then new measurements have to be defined in such a way they complement old measurements in terms of parameter estimation for the new structure optimally. The information matrix for optimization then contains entries due to the new regressors (of a higher degree polynomial) but considers also old already measured data. This means that the input design for new points results in maximizing $\det(M)$ by adding new measurements.

The algorithm starts with the estimation of an initial model of simple (e.g. linear) structure and validation. If the model quality is not sufficient, the iterative identification procedure is started. Figure 1 shows the sequential identification loop.

The procedure is done sequentially until a sufficient approximation of the system is found or a predefined maximum model complexity is reached. In order to minimize measurement time, measurements necessary for validation can be done so that that they are D-optimal for the model of next higher complexity. This means that if model performance is not sufficient, a more complex

model can instantly be identified just adding the validation data set to the old identification data set.

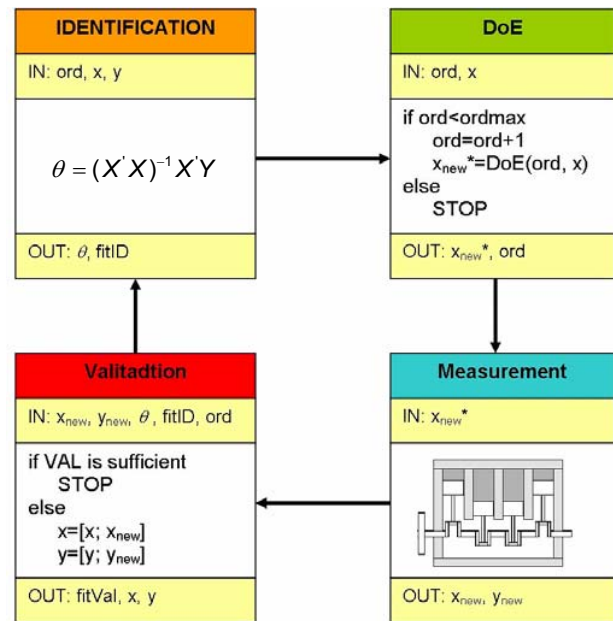


Figure 1: Iterative System Identification loop

ENGINE APPLICATIONS

A production 2 liter EU4 common rail Diesel engine was used as test object for the current work. Figure 2 depicts a scheme of the engine with a variable geometry turbine turbocharger, high pressure exhaust gas recirculation, variable intake swirl and multiple injections per combustion cycle. By the combination of compressor and intercooler the density of the fresh charge gas is increased. The variable geometry turbine (VGT) is used to exploit the available exhaust energy, and a part of the exhaust gas is recirculated (EGR) mainly for the reduction of nitrogen oxides. The swirl valve (VSA) is used to control the swirl in the combustion chamber. Table 1 summarizes the measurement quantities and actuators relevant for this work. The analyses were done on an AVL dynamic engine test bench, whereas a development engine control unit (ECU) was connected with a real-time hardware in the loop system for data acquisition and control.