Order Reduction for a Realtime Engine Model Using Flat and Nonlinear Galerkin Methods

Georg Fuchs, Alois Steindl, and Stefan Jakubek

Abstract—In this paper a methodology for the order reduction of large scale nonlinear dynamic systems is proposed and discussed. It is based on the known concepts of linear and nonlinear Galerkin projection. For the actual mode of operation the projection space is determined by snapshot decomposition. Nonlinear Galerkin methods require the existence of an attracting manifold of the system. As an approximation an iterative scheme is proposed, which is augmented and stabilized by making use of the local Jacobian of the system. The main idea pursued in this context is based on the notion that many technical systems can be easily decomposed into operating regimes which are characterized by locally constant Jacobians. The effectiveness of the proposed methodology is demonstrated by a large scale dynamic model of a turbocharged heavy-duty diesel engine.

Keywords—Model order reduction, Singular value decomposition, Snapshot method, Galerkin methods, Local model network

I. INTRODUCTION

Modern internal combustion engines are very complex systems with a variety of different calibration parameters and actuators. Especially regarding today’s rigorous legal emissions regulations, conflicting goals between fuel efficiency, driveability, performance and emissions must be handled, which is a very challenging task for control engineers.

The mathematical models used in the automotive industry mostly comprise nonlinear ordinary differential equations. Online simulations require real-time capable models which are computationally very demanding due to their extensive, nonlinear structure and require very powerful and fast numerical integration algorithms. Therefore, a reduction of the complex nonlinear model to a simplified version only containing the main dynamic characteristics seems to be a very promising approach for subsequent hardware-in-the-loop operation as well as controller design procedures.

The main idea behind model order reduction techniques is to simplify the original system to its dominating dynamic modes. There exists a multitude of different techniques in literature, some of which are specified in section I-A. In this work the linear (flat) and nonlinear Galerkin methods will be applied to a real-time heavy-duty diesel engine model. The engine features a single-stage turbocharger and exhaust gas recirculation (EGR), and measurements can directly be acquired at the test bench for the parameter identification. It offers highly nonlinear dynamics through a wide operating range in engine speed and torque and constitutes a well-suited application example for the performance evaluation of order reduction approaches. In the application of the nonlinear Galerkin method, a novel approach for the iterative solution of the nonlinear invariant manifold is adopted. It is based on the idea of decomposing the operating range into subdomains where the local Jacobians are obtained by model linearization. They are assumed to be constant within each subdomain and then utilized for the iterative solution in the nonlinear Galerkin manifold. This approach provides substantial advantages concerning computing time, since the local Jacobians can be calculated from the original model in an offline a priori linearization. Direct simulation outputs from the original nonlinear model for all states are used in the snapshot method for the assembly of the projection matrices.

A. Model Reduction Schemes

When dealing with large, complex models, consisting of systems of nonlinear ordinary differential equation systems, there exist different schemes of model order reduction. One main group of such methods are singular value decomposition (SVD)-based approximation methods. They contain balanced approximations, Hankel-norm approximations, proper orthogonal decomposition (POD, comprising the Galerkin projection) and modal approximation methods.

The basic idea of POD is the assumption that a state trajectory in the original state manifold of dimension $d$ can be approximated by a projection of the trajectory onto a lower-dimensional state space (a sub-manifold) of dimension $m < d$. The projection is obtained using a Galerkin projection which will be discussed in this paper in section II. The eigenfunctions used for the projection are obtained from an empirical approach, called the snapshot method (see section II-D), which takes advantage of system outputs collected in measurements or simulation [1]. A very well-arranged compilation of POD and other model order reduction methods can be found in [2] and [3].

The application of these methods has already been focus of several publications. In [4], the dimension reduction of the dynamics of a fluid conveying tube is presented, using linear and nonlinear Galerkin methods and center manifold reduction. In [5] an approach for the identification of the temporal coefficients of an empirical approximator of a process is shown. It uses experimental data gained from the process for a POD. [6] introduces a model order reduction using a nonlinear Galerkin projection for a finite element model of a horizontal axis wind turbine which serves for material fatigue
assessments in long-time simulations. The application of the method of model reduction is controlled by error estimation. Further approaches can be found in [7] and [8].

B. Operating Regime Decomposition

Modern diesel engines, may it be in the automotive area or for heavy-duty applications, offer a variety of inputs used for control purposes. As mentioned before, the models consist of generally extensive systems of coupled nonlinear ordinary differential equations. For a large class of these nonlinear dynamic systems, there exist methods that are based on the identification of subdomains of the system that can sufficiently accurately be described by local models [9]. These subdomains can be assembled in a so-called local model network (LMN) which provides multilateral characterization of the overall system, see Fig. 1. Some of the related methodologies completely leave the partitioning to the user so that a solid idea about the nature of the nonlinearity of the system is required, cf. [10]. Knowledge bounded least squares methods use both linguistic information and numerical data to identify so-called fuzzy models [11]. Others make use of the input/output data of the system to identify linear subdomains [12], [13].

In the current work the approach of local linear networks is pursued. The local Jacobian matrix is achieved from a numerical linearization within the respective network subdomain. It is then adopted in the iteration scheme of the nonlinear invariant Galerkin manifold (see II-C) for the computation of the vector of the inessential states parallel to the integration of the main states of the reduced system. The realization of this approach is shown in section III.

II. MODEL REDUCTION BY GALERKIN METHODS

The Galerkin reduction methods were originally introduced for the approximation of dissipative partial differential equation problems. In a geometrical interpretation, they can be regarded as an approximation of the system dynamics on the phase manifold by projection onto a sub-space, which is able to capture the essential dynamical behavior of the original system. The two methods shown in this paper are the linear (also called flat) and the nonlinear Galerkin method (see sections II-B and II-C).

In the general case a nonlinear model of the real process is obtained from a mathematical-physical modeling approach. Such a model can be written in state space representation as

\[ \dot{x}(t) = f(x(t), u(t)), \]  

with \( x \in \mathbb{R}^d \) the state vector of dimension \( d \) and \( u \in \mathbb{R}^r \) the \( r \)-dimensional input vector.

A. System Projection

For the reduction of the system order a projection of the complete differential equation from the original state space onto a sub-space is performed. Galerkin’s method assumes that only the essential modes on this sub-space of the underlying model are important for the main dynamical behavior, whereas the remaining modes can either be neglected completely (flat Galerkin method) or are governed by the “main” modes in some algebraic relation (nonlinear Galerkin method).

For the solution of equation (1) the following approach is chosen:

\[ x = \Phi_1 \xi + \Phi_2 \eta, \]  

where \( \xi \in \mathbb{R}^m \) is the vector containing the essential modes and \( \eta \in \mathbb{R}^{d-m} \) contains the less important remaining modes in the reduced system. The matrices \( \Phi_1 \) and \( \Phi_2 \) span \( m \)- and \( d-m \)-dimensional sub-spaces \( \mathcal{X} \) and \( \mathcal{Y} \). The composition of these matrices will be the topic of chapter II-D.

Substituting (2) in (1) yields

\[ \dot{\Phi}_1 \xi + \dot{\Phi}_2 \eta = F(\xi, \eta, u). \]  

A projection onto the sub-spaces \( \mathcal{X} \) and \( \mathcal{Y} \) is accomplished using the matrices \( \Phi_1^T \) and \( \Phi_2^T \). Assuming \( \Phi_1^T \Phi_1 = I \), \( \Phi_1^T \Phi_2 = 0 \) and \( \Phi_2^T \Phi_1 = 0 \), \( \Phi_2^T \Phi_2 = I \), this leads to two coupled systems of differential equations:

\[ \dot{\xi} = \Phi_1^T F(\xi, \eta, u) \]  
\[ \dot{\eta} = \Phi_2^T F(\xi, \eta, u). \]  

B. Linear Galerkin Method

The linear, or also called flat Galerkin method finds an approximation to the system (4),(5) by neglecting the states \( \eta = 0 \). For the solution only the first equation

\[ \dot{\xi} = \Phi_1^T F(\xi, u) \]  

is considered, yielding the approximation \( x \approx \Phi_1 \xi \). Geometrically, the linear Galerkin approach represents a projection of the original differential equation from the state space of order \( d \) onto a sub-space of order \( m \) without accounting for the remaining \( d-m \) states.

C. Nonlinear Galerkin Method

The nonlinear Galerkin method assumes that an algebraic relationship

\[ \eta = \Theta(\xi) \]  

can be found which means that the behavior of the modes \( \eta \) is directly determined by the dynamic behavior of the \( \xi \) modes. Such a relation is called invariant manifold. The problem in this case is that there is no a priori information about an invariant manifold and the calculation can be very complex. Instead, one can come up with an approximate invariant manifold (AIM), which can be found without too
much computational effort. The AIM can be described as the approximate solution of the equation
\[ \dot{\eta} = \Phi_2^T F(\Phi_1 \xi + \Phi_2 \eta, u). \] (8)

There exist different methods of finding these approximations (see [14], [15] and [3] for more detailed information about the AIM calculation). In this work the approach of Titi [16] was used. Here, the dynamics of \( \eta \) are disregarded in order to obtain a quasi-stationary AIM. (8) then becomes an algebraic relation:
\[ \Phi_2^T F(\xi, \eta, u) = 0. \] (9)

This approach leads to a coupled system of differential-algebraic equations (4) and (9). The algebraic part can be solved by a fixed-point iteration, which is shown in section III.

D. Snapshot Methodology for POD

Measurements results and analysis of the complete nonlinear model show that for typical operating conditions of an engine only certain modes show significant dynamic behavior. The idea behind the snapshot method is to excite the system with the inputs \( u \) in a way that is typical for the real engine operation. It is based on an empirical concept using the outputs generated from measurements or directly from the model simulation. It is a well-acknowledged technique, which has proved to be very efficient in several previous works, see e.g. [8] and [5]. For a defined input signal the states of the system are recorded and analyzed according to the dominant dynamic behavior. The \((n \times d)\) matrix \( X \) is called the snapshot record matrix, where \( d \) is the number of states in the reduced system and \( n \) the number of recorded snapshots. For the assembly of \( \mathcal{X} \) and \( \mathcal{Y} \) the principal eigenmodes are needed, extracted from the \((d \times d)\) covariance matrix
\[ C = \frac{1}{d}((X - \mu)^T (X - \mu)), \] (10)
with \( \mu \) containing the mean values over all samples for each state. The eigenvalues \( \lambda_j \) and the eigenvectors \( s_j \) of the covariance matrix are computed according to
\[ CS_j = \lambda_j s_j. \] (11)

For the POD, the eigenvectors of \( C \) are assembled in the matrices \( \Phi_1 \) and \( \Phi_2 \) according to the relative magnitude of the eigenvalues compared to each other:
\[ \Phi_1 = [s_1 \ldots s_m], \quad \Phi_2 = [s_{m+1} \ldots s_d]. \] (12)

For the diesel engine system, appropriate input signal excitation sequences have to be found. Modern internal combustion engines undergo strongly dynamic exposures, such as abrupt load alterations or rapid acceleration maneuvers. These cycles are accompanied by partly rough actuator position variations. For the composition of the snapshot matrix in order to obtain preferably much information from the original system, the input signals should contain as much dynamical portions as possible. Fig. 2 shows one example of the input excitation of the throttle actuator signal \( u_{thr} \), actuated around its halfway opened position \( (u_{thr} = 50\%) \) in a rectangular distribution with an amplitude of ±20% and superposed with additional noise. Corresponding signal sequences have been chosen for the other system inputs.

Of course the range of validity for the snapshot method is limited. For a certain excitation of the system with certain input variables, very sufficient results can be obtained in the reduced model using the same input variables. The method reaches its limits when additional inputs are actuated that are not accounted for in the snapshot acquisition process. Then the reduced model is not able to correctly map the dynamic response of the original system to these additional inputs.

III. ITERATIVE NONLINEAR GALERKIN PROJECTION

In section II-C the nonlinear equation (9) emerged. For reasons of computational speed and accuracy, an efficient numerical procedure has to be introduced to find approximate solutions on the invariant Galerkin manifold. On this account a local linear model structure, described in section I-B, is adopted, using the local Jacobian for the iteration scheme.

A. Linearized System

The nonlinear system equation (1) can be linearized in the current state, yielding the Jacobian matrices \( A \) and \( B \) according to
\[ A = \frac{\partial f(x, u)}{\partial x}, \quad B = \frac{\partial f(x, u)}{\partial u}. \] (13)

Using an LMN approach, the operating region of the nonlinear model is separated into single subdomains (compare Fig. 1). Within these subdomains, the local Jacobians are computed and assumed to be constant for each of these partitions. The local Jacobians are now applied for the AIM iteration of (9) (see the following subsection).

B. Equation Error Minimization Problem

The solution of (9) is obtained recursively in the following way: the equation error is calculated and minimized in order to compute the optimal step width in the \( \eta \) direction. According to a Taylor series expansion approximation, the change of \( f \) in the \( \eta \) direction is
\[ \Delta f = \frac{\partial f}{\partial \eta}(x, u) \Delta \eta, \] (14)

\[ \text{wt: 90-70-50-30-10} \]

\[ \text{time [s]} \]

[Fig. 2. Input signal sequence of the throttle actuator position]
with
\[ \frac{\partial f}{\partial \eta}(x(\xi, \eta), u) = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \eta} = \frac{\partial f}{\partial \Phi_2} . \] (15)

The Jacobian matrix \( \frac{\partial f}{\partial x} = A \) is computed offline for the different operating regimes and is chosen online according to the regime the model is currently running in. Combining the equations from above, the equation error \( r_k = \Phi^T_2 f_k \) for the next step in \( \eta \) can be computed:
\[ r_{k+1} = \Phi^T_2 \left( f_k + \frac{\partial f}{\partial x} \Phi_2 \Delta \eta \right) . \] (16)

The error is used for optimizing the performance function
\[ J = \frac{1}{2} r^T_{k+1} r_{k+1} \rightarrow \min, \] (17)
from which the optimal step size \( \Delta \eta \) follows to
\[ \Delta \eta = - \left( \frac{\partial f}{\partial x} \right)^{-1} \Phi^T_2 f_k . \] (18)

The described iteration scheme is performed in parallel to the numerical integration of the reduced system states. The advantage of this method is that due to the fact that the local Jacobians, which are calculated offline for the subdomains of the LMN, can be switched online during simulation, a constitutive increase in performance with a low additional computing expense is allowed for. The results obtained are discussed in the following section.

IV. RESULTS

The goal of this work was the adaption of the model order reduction methods described in sections II and III to a real-time heavy-duty diesel engine model. The model description and the results of the reduction methods are the topics of this section.

A. System description

The model considered was obtained from a physical-mathematical modeling approach using the concept of mean value modeling (MVM) [17]. MVM means the replacement of the discontinuous operation of the pistons by continuous processes for mass transportation through the cylinders and production of power. The thermodynamic and chemical processes inside the cylinders during the combustion cycles are considered as mean values over a cycle. Such a simplifying modeling approach is well-suited for realtime test bench applications because of low computational demand and thus higher simulation speed. The model is built up by modularly connectable zero-dimensional tank components, which reproduce the thermodynamic characteristics inside the pipe connections, coolers, etc. These storage elements are connected to each other through coupling elements (valves, throttles). The exhaust gas turbocharger is modeled by a quasi-stationary, parameter based approach. The complete system has a state order of 16. The parameters of the model were identified using measurements from the test bench.

B. Results

The linear (flat) and nonlinear Galerkin methods were applied to the engine model and the results compared to the solution of the original system, which was computed using a fixed-step Runge-Kutta numerical integration method. The original system was also used for the offline calculation of the local Jacobians \( \frac{\partial f}{\partial x} \). The main focus of this work lies in the examination of the general functionality of the two presented Galerkin methods and the sensitivity of the performance with respect to the reduced model order.

Different input signal and operating condition test cases were designed, with three input variables: the throttle actuator position \( u_{thr} \), the EGR valve actuator position \( u_{egr} \) and the engine speed \( n_{eng} \). For the snapshot recording procedure and the subsequent proper orthogonal decomposition, these input variables were varied in the range deployed for the respective use case (compare II-D). Consequently, the model order was reduced and the differences in performance of the order reduction methods are discussed.

Case 1: \( n_{eng} = 600 \text{ U/min} \), \( u_{egr} = 100 \% \), \( u_{thr} \) varied. The throttle actuator input signal was varied according to the signal shown in Fig. 2. Fig. 3 shows the results for the exhaust turbocharger speed for a reduced model order of 10. It can be seen that the system dynamics are reproduced sufficiently by both the flat and the nonlinear Galerkin methods. Obviously, the flat method shows a clear offset from the original solution, which is increasing with time.

![Fig. 3. Exhaust turbocharger speed, reduced system order 10](image)

Fig. 3. Exhaust turbocharger speed, reduced system order 10

Next, the reduced system order was lowered to nine. For the same input signal the results were compared again, see Fig. 4. An order reduction by one already shows considerably more significant differences between the results of the flat and the nonlinear methods. The nonlinear method traces the solution of the original system very well, whereas the flat method already indicates strong errors in the amplitude. The performance discrepancy is even more obvious, when the order is dramatically decreased to 4 states, see Fig. 5. Here, the solution of the flat method can not reproduce the dynamics of the original system any more (outside the plot range), but the nonlinear method can still perform well, showing decent accord with the reference solution.

Case 2: \( u_{egr} = 100 \% \), \( u_{thr} \) and \( n_{eng} \) varied simultaneously. In this test case two input variables were varied at the same time: the throttle actuator position and the engine speed. The according pattern is shown in Fig. 6.
The results for exhaust turbocharger speed, intake and exhaust manifold pressure (Fig. 7) already show much better performance of the nonlinear method for a system order of 9, especially at the beginning of the simulation.

The difference becomes more distinct again for a further reduction of the system order, in this case from 9 to 8 states (Fig. 8). The nonlinear Galerkin approximation fits the original solution very well, whereas the flat Galerkin method result shows very poor behavior.

Case 3: $u_{thr} = 20\%$, $u_{egr}$ and $n_{eng}$ varied. Again, two input variables - the EGR valve actuator signal and the engine speed - were varied according to the pattern given in Fig. 9. Fig. 10 shows the results for intake and exhaust manifold pressure and EGR mass flow for the reduced system order 10. When the order is decreased - here from 10 to 8 states - again
one can see the considerably better results of the nonlinear method (see Fig. 11). They seem even more remarkable when the wide range of passed nonlinear system behavior (1000 - 2500 rpm) is envisioned.

![Fig. 10. Intake manifold pressure, EGR mass flow and exhaust manifold pressure, reduced system order 10](image)

![Fig. 11. Intake manifold pressure, EGR mass flow and exhaust manifold pressure, reduced system order 8](image)

The results of this section allow for some important conclusions: The linear and nonlinear Galerkin methods are both able to reproduce the main system dynamics very well for high orders of the reduced model. In the process of decreasing the model order, the omitting of relevant dynamic states is clearly better compensated by the nonlinear method due to the fact that it can account for these omitted states using the AIM. The use of the local Jacobian achieves impressive results even for running through strongly nonlinear operating regions. The linear method, however, completely neglecting the omitted states, shows considerable performance declines for reduced system orders.

V. CONCLUSION

The application of flat and nonlinear Galerkin model order reduction methods to a mean value diesel engine model was presented. The existing approach of the nonlinear Galerkin method was combined with an iterative solution scheme using local Jacobians, obtained from a local model network approach and assumed constant within each operating subdomain. Subsequent reduction of the model order revealed significant performance advantages of the nonlinear method using the local Jacobian-based iteration scheme, compared to the flat method. The low computational expense increase is outweighed by remarkable benefits in dynamic accuracy. The obtained results provide an outlook to further applications in other technical disciplines.

REFERENCES


Georg Fuchs received the M.S. degree in mechanical engineering in 2008 from the Vienna University of Technology (VUT). He is presently research assistant at the Institute of Mechanics and Mechatronics at the VUT, pursuing his Ph.D. degree in automatic control engineering. His current research interests are internal combustion engine modeling, simulation and control, numerical linearization and integration methods, and real-time hardware applications.