Modeling and Model Reduction in Automotive Systems

Nilsson, Oskar

2006

Document Version:
Publisher's PDF, also known as Version of record

Modeling and Model Reduction in Automotive Systems

Oskar Nilsson

Department of Automatic Control
Lund University
Lund, December 2006

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

ISSN 0280–5316
ISRN LUTFD2/TFRT--3242--SE

Printed in Sweden,
Lund University, Lund 2006

Acknowledgments

First of all, I would like to thank my supervisor Anders Rantzer for his support, encourage and fruitful discussions throughout my work.

My thanks goes to Klaus Papadakis and Ingemar Odenbrand at the Department of Chemical Engineering, Lund University, for their helpful suggestions concerning the Lambda sensor chemistry. Also, Per Tunestål, Jan-Erik Everitt with others deserve acknowledgments for their assistance in the Combustion Engines Lab.

I owe my gratitude to Jonathan Chauvin at École Nationale Supérieure des Mines de Paris, for contributing with the heuristic model reduction part in Chapter 4. I am also grateful for the proofreading performed by Anders Rantzer, Rolf Johansson, Henrik Sandberg and Martin Ohlin.

My roommate Martin Ohlin is greatly appreciated for his cheerful welcome in the mornings. Our discussions have not always been work-related but essential for keeping a pleasant atmosphere. Others who I would like to thank for their technical and, in particular, moral support are Brad Schofield, Mathias Persson and Martin Kjær.

Leif Andersson and Anders Blomdell have greatly facilitated all contact with computers. Leif is also acknowledged for typesetting support of this thesis. The secretaries Agneta Tuszynski, Britt-Marie Mårtensson and Eva Schildt have solved all my administrative problems and are important cogs in the social machinery of the department.

This research was partially done in the framework of the HYCON Network of Excellence, contract number FP6-IST-511368. Partial financial support was also received from Toyota Motor Corporation and our industrial contact Akira Ohata is acknowledged for providing industrial problem settings.

Finally I would like to thank my friends and family for your encouragement and support.

Oskar

Contents

1

Introduction

1.1 Motivation

Automotive industry development

The current control design development process in automotive industry involves many expensive experiments and hand-tuning by experienced personnel. This process is time-consuming and even if only small changes have been done between two models, many tuning tasks have to be made over and over again.

Model based development is a promising approach to reduce costs, development time and dependency of the undocumented knowledge possessed by experienced personnel. The idea is to replace expensive experiments with simulation of mathematical models.

Complexity versus fidelity

The modeling process is highly dependent of the model purpose. Depending of the model usage different effects in the physical plant should either be taken into consideration or be neglected. However, what effects to be included can be very hard to know and requires experience and understanding of the real process.

A model is always approximate and the level of accuracy is typically a function of its complexity. A very detailed model is more likely to cover the most important dynamics but large complexity has many down-

sides. In general simulation time and memory requirements scales badly with complexity, also model analysis is made harder. Yet another inconvenience is that larger models generally contain more model parameters. In control design, this could yield that the hand-tuning saved by model based control design is replaced by time consuming calibration of model parameters.

Why model reduction?

A systematic method to reduce model complexity would be very useful in many situations. If a detailed component model has been developed, the modeling effort could later be reused for other purposes. For example, component models could be combined to model a more overall behaviour. The model complexity could then be reduced with a model reduction method to match the required detail level of the actual purpose.

In model based control design, simple models are highly preferred. Some methods, e.g. Linear Quadratic Gaussian control or Model Predictive Control, yield controllers with complexity comparable to the model.

A control design approach where model reduction plays a central role is illustrated in Figure 1.1. A complex physical model with a large number of uncertain parameters could be reduced by a model reduction method. Ideally, the resulting model should not only be of low complexity but should also contain few parameters, facilitating calibration. The small model is then calibrated with experiment data and used for model based control design.

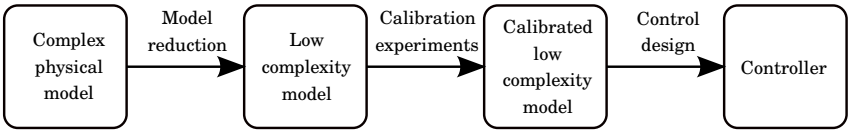

Figure 1.1 Example of alternative controller development process

In some cases fast simulation models for real-time purposes are essential, e.g. in on-board fault diagnostics where computing power is not abundant.

1.2 Outline of the thesis

The thesis combines the areas of modeling and model reduction of automotive systems. First, in Chapter 2 a brief introduction to model reduction methods of linear and nonlinear systems is presented. In Chapter 3 a model of an exhaust gas oxygen sensor, also called Lambda sensor, is derived. This sensor is a core component in the emission control in modern spark ignition combustion engines. In Chapter 4, two model reduction methodologies are applied on a detailed engine air path model. One of the methodologies is systematic with mathematical founding while the other is heuristic and based on intuition and experience. A general model reduction method for nonlinear systems is presented in Chapter 5 together with numerical examples. Finally, Chapter 6 contains concluding remarks together with the direction of further research after this thesis.

2

Background

Model reduction methods currently used in automotive industry are mostly ad hoc and heuristic. This chapter presents this approach together with other model reduction methods for linear and nonlinear systems. For further reading, a broad overview of model reduction methods is presented in [Antoulas and Sorensen, 2001].

2.1 The goal of model reduction

The goal of model reduction is defined depending on the application. In general the reduced model should be easier to handle, which in most cases implies facilitated analysis and simulation, without too much loss in quality. A model developed for control design purposes is usually equipped with input and output signals and in that case, the quality demand could be a bound on the input-output relationship deviation.

2.2 Model reduction of linear systems

A linear time-invariant system can be represented in many different ways. A common description is the state-space form

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2.1}$$

where $x(t) \in \mathbf{R}^n$, $u(t) \in \mathbf{R}^l$ and $y(t) \in \mathbf{R}^m$. If the model order n is much larger than the number of inputs and outputs($n \gg l$, $n \gg m$) it can be suspected that the model contains redundant states. The model reduction problem is how to find and remove such redundancy.

Gramians

Gramians is a central concept in model reduction, they give a measure on how strong a connection is between states and input resp. output signals.

The controllability function, as defined in [Scherpen, 1993], is the minimum amount of input energy required to drive the system from the zero state to x_0.

$$L_c(x_0) = \min_{\substack{u \in L_2(-\infty,0) \\ x(-\infty)=0 \\ x(0)=x_0}} \frac{1}{2} \int_{-\infty}^{0} ||u(t)||^2 dt \tag{2.2}$$

Further, the observability function is the amount of energy the initial state x_0 generates in the output signal while the input signal is zero.

$$L_o(x_0) = \frac{1}{2} \int_{0}^{\infty} ||y(t)||^2 dt, \quad x(0) = x_0, \quad u \equiv 0 \tag{2.3}$$

For linear systems, as defined in (2.1), these functions become the quadratic expressions

$$L_c(x_0) = \frac{1}{2} x_0^T P^{-1} x_0 \qquad\qquad L_o(x_0) = \frac{1}{2} x_0^T Q x_0$$

where P and Q are called the controllability gramian resp. the observability gramian. It can be shown, see [Moore, 1981], that these gramians are

$$P = \int_0^\infty e^{At} B B^T e^{A^T t} dt \qquad\qquad Q = \int_0^\infty e^{A^T t} C^T C e^{At} dt$$

Often stability is assumed and these integrals are defined. A more numerically feasible way to compute the gramians is to determine the

unique solutions to the Lyapunov equations

$$\begin{aligned} AQ + QA^T + BB^T &= 0 \\ A^T P + PA + C^T C &= 0 \end{aligned} \tag{2.4}$$

Moreover, the column vectors of P span the controllable subspace in $\mathbf{R}^n$ and correspondingly the null space of Q is the unobservable subspace.

Balanced truncation

Balanced truncation is a popular model reduction technique introduced in [Moore, 1981]. The method guarantees preserved stability and comes with an a priori error bound.

The idea of the method is to apply a coordinate change so that each state is equally controllable and observable. The model is then reduced by truncating states with relatively weak input-output dependency. Applying a linear coordinate change, $T\tilde{x} = x$, to the state-space form given in (2.1) yields the system

$$\begin{aligned} \dot{\tilde{x}}(t) &= TAT^{-1}\tilde{x}(t) + TBu(t) \\ y(t) &= CT^{-1}\tilde{x}(t) + Du(t) \end{aligned}$$

How controllable and observable these new states are is determined by the previously mentioned gramians. From (2.4) it can be derived that the new gramians become $\tilde{P} = TPT^T$ and $\tilde{Q} = T^{-T}QT^{-1}$. A balanced realization is achieved if the coordinate change makes the gramians diagonal and equal.

$$\tilde{P} = \tilde{Q} = \tilde{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \tag{2.5}$$

Methods for computing this coordinate change T can be found in [Li, 2000]. The diagonal elements $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ are the Hankel singular values that indicate how important a state is for the input-output relationship. Consequently, the reduced model is derived by

truncating states in the balanced realization corresponding to small singular values. Moreover, the approximation error is bounded as

$$\max_u \frac{\|\tilde{y}(t) - y(t)\|_2}{\|u(t)\|_2} \leq 2 \sum_{k=n-r+1}^{n} \sigma_k \tag{2.6}$$

if the r least important states are truncated.

This model reduction method is applied on a combustion engine model in Chapter 4.

Balanced truncation of linear time-varying systems

Balanced truncation has been extended to also cover the linear time-varying case, see [Verriest and Kailath, 1983; Shokoohi *et al.*, 1983]. For this class of linear systems, the matrices A, B, C and D are time-varying.

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)x(t) + D(t)u(t) \end{aligned} \tag{2.7}$$

The method follows the ideas of balanced truncation of time-invariant systems. For time-varying systems one can use the notion of controllability or observability over a time interval, say $t \in [0, T]$. The energy functions in (2.2) and (2.3) are then slightly modified. The controllability function is here the minimum required input energy to reach x_0 at time t starting from the zero state at $t = 0$.

$$L_c(x_0, t) = \min_{\substack{u \in L_2(0,t) \\ x(0)=0 \\ x(t)=x_0}} \frac{1}{2} \int_0^t ||u(\tau)||^2 d\tau$$

The observability function is the energy induced by the initial state $x(t) = x_0$ in the output-signal over the time interval $[t, T]$, while the input-signal is zero.

$$L_o(x_0, t) = \frac{1}{2} \int_t^T ||y(\tau)||^2 d\tau, \quad x(t) = x_0, \quad u \equiv 0$$

As in the time-invariant case, these functions are quadratic but the gramians $P(t)$ and $Q(t)$ are now time-dependant.

$$L_c(x_0, t) = \frac{1}{2} x_0^T P(t)^{-1} x_0 \qquad L_o(x_0, t) = \frac{1}{2} x_0^T Q(t) x_0$$

Furthermore, the time-varying generalization of the Lyapunov equations in (2.4) becomes

$$\frac{dP}{dt}(t) = A(t)P(t) + P(t)A(t)^T + B(t)B(t)^T \qquad P(0) = 0$$
$$\frac{dQ}{dt}(t) = -Q(t)A(t) - A(t)^T Q(t) - C(t)^T C(t) \qquad Q(T) = 0$$

Once more, a balanced realization is achieved if a time-varying coordinate change yields diagonal and equal gramians

$$\tilde{P}(t) = \tilde{Q}(t) = \tilde{\Sigma}(t) = \begin{bmatrix} \sigma_1(t) & & \\ & \ddots & \\ & & \sigma_n(t) \end{bmatrix}$$

The low order model is derived by truncating states corresponding to small singular values $\sigma_i(t)$. It is reasonable to let the reduced model order vary with time as $\sigma_i(t)$ is time-varying. A priori error bounds, similar to (2.6), are available for the time-varying case, see [Lall and Beck, 2003; Sandberg and Rantzer, 2004].

This theory will be revisited in Chapter 5, where the time-varying gramians are used as tools to reduce nonlinear systems.

Descriptor form

Another linear state-space representation is the descriptor form

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2.8}$$

This representation can be transformed to the standard state-space form in (2.1) if E is invertible. However, if E and A are sparse $E^{-1}A$ can be dense and it might therefore be beneficial to keep the form in (2.8).

If E is singular, the system is a set of algebraic-differential equations and the problem becomes more involved. A generalization of the gramians defined for the standard state-space form is presented in [Stykel, 2004]. Theory together with numerical methods are defined and also in this case an a priori error bound is available.

2.3 Model reduction of nonlinear systems

Model reduction of nonlinear systems is a research area under heavy development. Currently there is no method that generally provides guaranteed preserved stability or error bounds.

Ad hoc methods

Indirect model reduction is performed in all modeling work when complexity is chosen to match the intended model purpose. There are three common ways to reduce complexity:

- To discard effects that by intuition or experience have a relatively weak impression on the dynamics.
- Separation of time-scales and replacing relatively fast dynamics with static gains.
- Averaging effects into one pseudo-effect.

All three approaches require great knowledge and intuition of the modeled object. The second mentioned method is more formally called the singular perturbation method. The differential equations of $\dot{x} = f(x, u)$ are divided into two parts, one relatively faster than the other

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2, u) \\ \dot{x}_2 &= f_2(x_1, x_2, u) \end{aligned}$$

If x_2 corresponds to the fast dynamics, one introduces a factor ϵ according to

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2, u) \\ \epsilon\dot{x}_2 &= f_2(x_1, x_2, u) \end{aligned}$$

and then set $\epsilon = 0$. The original system is now replaced with a set of differential algebraic equations with fewer states, for more details see [Khalil, 2002].

Linearization around equilibrium point or trajectory

In some applications the intended model usage is in the neighborhood of a certain operating point in state-space. Then the detailed nonlinear model could be linearized at this point, giving rise to a linear model. This model can then be reduced with a linear reduction method. The reduced model will however also be linear and only be a valid approximation close to the operating point.

Sometimes a nominal input-signal is available and the effect a deviation from this signal would give is of interest. A linearization around a trajectory in state-space is then a valid approximation. This yields a time-varying system as in (2.7) and the theory of balanced truncation of linear time-varying systems can be applied. This procedure has been done successfully, see [Sandberg, 2006].

Balancing nonlinear systems

Balancing using energy functions An extension to nonlinear systems of the mentioned balanced truncation method is proposed in [Scherpen, 1993]. Here nonlinear systems of the form

$$\begin{aligned} \dot{x} &= f(x) + g(x)u \\ y &= h(x) \end{aligned} \tag{2.9}$$

are considered. Again, the controllability and observability functions in (2.2) and (2.3) are used. For the given nonlinear system it can be shown that, under some conditions, $L_c(x)$ and $L_o(x)$ are the unique smooth solutions of

$$\frac{\partial L_c}{\partial x}(x) f(x) + \frac{1}{2}\frac{\partial L_c}{\partial x}(x) g(x) g^T(x) \frac{\partial^T L_c}{\partial x}(x) = 0, \quad L_c(0) = 0$$

and

$$\frac{\partial L_o}{\partial x}(x) f(x) + \frac{1}{2} h^T(x) h(x) = 0, \quad L_o(0) = 0$$

After a coordinate transformation, $x = \psi(z)$, the functions can be writ-

ten

$$\tilde{L}_c(z) = \frac{1}{2} z^T z \qquad \tilde{L}_o(z) = \frac{1}{2} z^T \begin{bmatrix} \tau_1(z) & & \\ & \ddots & \\ & & \tau_n(z) \end{bmatrix} z$$

This form is not balanced, in the linear case it is sometimes called "input normalized". However, an additional coordinate change can balance L_c and L_o. For more details see [Scherpen, 1993]. In analogy with the linear case, the functions $\tau_1(z) \geq \cdots \geq \tau_n(z)$ are called the singular value functions of the system. Model reduction is performed by truncating states in the balanced form corresponding to small singular functions.

A linearized version of the method applied to a linear system yields the same result as the standard balanced truncation method would. Further, the singular value functions become constant and $\tau_i(z) = \sigma_i^2$ as given in (2.5).

The main disadvantage with this method is that it requires great numerical effort and it is therefore only applicable to very small systems, see [Newman and Krishnaprasad, 1998].

A resent contribution to this problem setting is among others [Fujimoto and Tsubakino, 2006].

Balancing using empirical gramians A model reduction method for nonlinear systems of the form

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x) \end{aligned}$$

is proposed in [Lall *et al.*, 2002]. The approach applies ideas concerning linear systems, introduced in [Moore, 1981], to nonlinear systems.

Here state-space data is collected while impulse input-signals in different directions are injected. The data is then used to estimate a constant controllability gramian matrix. Similarly, a constant observability gramian matrix is constructed from simulation data generated by different initial values distributed on the unit sphere.

When the gramians have been computed they are balanced using linear theory, see Section 2.2. The reduced nonlinear model is then

derived by applying the corresponding linear coordinate change $Tz = x$

$$\dot{z} = T^{-1} f(Tz, u)$$
$$y = h(Tz)$$

followed by truncation of states, as in the linear case. This method also yields the same reduced system as standard linear balanced truncation if applied to a linear system. In [Liu and Wagner, 2002] the method is applied on an automotive model.

The method is much less computationally intensive than the method using energy functions. However, the heuristic use of simulations does not leave much room for proofs and analysis.

Proper orthogonal decomposition

Karhunen-Loéve expansion [Karhunen, 1946; Loève, 1945], or proper orthogonal decomposition (POD), is a model reduction method for state-space models based on principal component analysis. The method was pioneered for applications in turbulence models in [Lumley, 1967] and is one of the most commonly used tools for model reduction of nonlinear systems. It uses simulation data to find a low-dimensional subspace that captures most of the state dynamics. Figure 2.3 illustrates a possible truncation of state-space, (x_1, x_2) to $(\hat{x}_1)$.

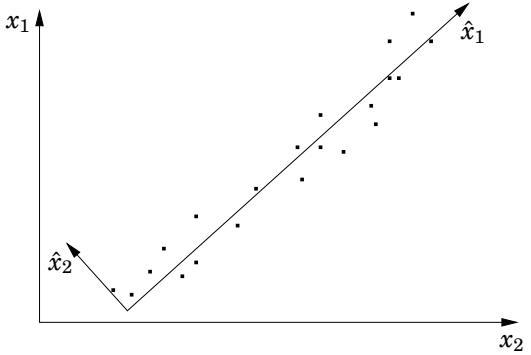

Figure 2.1 $\hat{x}_2$ is a dominant state-subspace in (x_1, x_2)

The method can briefly be described in three steps.

1. Simulate the nonlinear system

$$\dot{x} = f(x,u)$$

and collect snapshots of the state vector in a matrix X.

$$X = [\, x(t_0) \quad x(t_1) \quad \dots \quad x(t_N) \,],\ x(t) \in \mathbf{R}^n$$

2. Factorize X with the singular value decomposition

$$U\Sigma V^T = X$$

3. Choose truncation level after size of singular values in Σ. Truncate $U \in \mathbf{R}^{nxn}$ to $\hat{U} \in \mathbf{R}^{nx\hat{n}}$ so that $x \approx \hat{U}\hat{x}$ where $\hat{x} \in \mathbf{R}^{\hat{n}}$. Then the reduced model becomes

$$\dot{\hat{x}} = \hat{U}^T f(\hat{U}\hat{x}, u) \tag{2.10}$$

This method lacks general error bounds, which can easily be demonstrated. Put short, a state can be important even though it is small. For example, scaling of states by a diagonal coordinate change

$$x^* = \mathrm{diag}(c_1, c_2, ..., c_n)x \quad c_i > 0$$

does not change the dynamic behaviour of the system but can make the method choose an arbitrary subspace. Further, all states are usually not interesting for control purposes and as this method does not take any output-signal into consideration one is forced to use a larger subspace than might be necessary.

A common source of large models is discretization of Partial Differential Equations (PDE's), where the states share the same physical units. In this case size comparison might be feasible. More details and numerous examples can be found in [Astrid, 2004].

Trajectory piecewise-linear model reduction

A novel approach to nonlinear model reduction is presented in [Rewieński, 2003]. The method is based on linearizations distributed over one or many training trajectories. The method can be divided into three steps.

First step: Linearizations Simulate the nonlinear system

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= g(x,u) \end{aligned}$$

with a training input $u_0(t)$. Then choose a set of linearization points (x_0^i, u_0^i) along the training trajectory, see Figure 2.2. Observe that the points are in general not stationary.

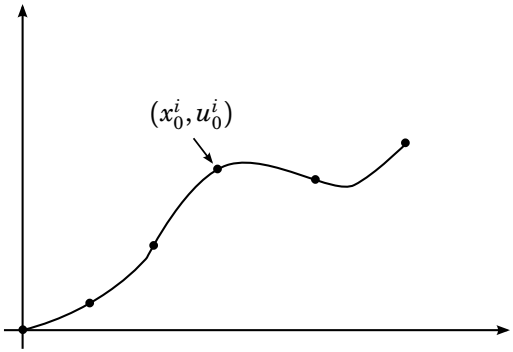

Figure 2.2 Linearizations distributed over a training trajectory

Close to linearization i, the linearization

$$\begin{aligned} \dot{x} &\simeq f(x_0^i, u_0^i) + A_i(x - x_0^i) + B_i(u - u_0^i) \\ y &\simeq g(x_0^i, u_0^i) + C_i(x - x_0^i) + D_i(u - u_0^i) \end{aligned} \tag{2.11}$$

approximates the nonlinear system. The matrices A_i, B_i, C_i and D_i are the partial derivatives

$$\begin{aligned} A_i &= \frac{\partial f}{\partial x}(x_0^i, u_0^i) & B_i &= \frac{\partial f}{\partial u}(x_0^i, u_0^i) \\ C_i &= \frac{\partial g}{\partial x}(x_0^i, u_0^i) & D_i &= \frac{\partial g}{\partial u}(x_0^i, u_0^i) \end{aligned}$$

Second step: Piece-wise linear approximation The local linear approximation (2.11) can be rewritten as

$$\begin{aligned} \dot{x} &\simeq f_i(x,u) \\ y &\simeq g_i(x,u) \end{aligned}$$

Now let the original nonlinear system be approximated in a less local way by a weighted sum of the local linearizations.

$$\begin{aligned} \dot{x} &\simeq \sum_i w_i(x,u)\, f_i(x,u) \\ y &\simeq \sum_i w_i(x,u) g_i(x,u) \end{aligned} \tag{2.12}$$

The weighting function $w_i(x,u)$ is close to one in the neighborhood of linearization i and zero otherwise. Additionally, $w_i(x,u) \geq 0$ and $\sum_i w_i(x,u) = 1$ for all x and u.

Third step: Model reduction So far the original nonlinear system has been approximated but no gain in terms of number of states or simulation time has been achieved. In this step linear model reduction theory is used to reduce the local linear models.

In [Rewieński, 2003] the Krylov subspace method was used. This method uses the Arnoldi algorithm, which is numerically effective even for very large systems. However, it has the drawback of not generally provide guaranteed preserved stability or error bounds, see [Grimme, 1997]. The use of balanced truncation has also been investigated, see [Vasilyev *et al.*, 2006].

The Krylov subspace method generates an orthonormal projection $z = Wx, W \in \mathbf{R}^{\hat{n}xn}$, which is used globally to reduce all the local models. For further details see [Rewieński, 2003]. Introducing the new coordinates in (2.12) yields

$$\begin{aligned} \dot{z} &= \sum_i w_i(W^T z,u) W f_i(W^T z,u) \\ y &\simeq \sum_i w_i(W^T z,u) g_i(W^T z,u) \end{aligned}$$

This reduced model has fewer states, $\hat{n} < n$, and improved simulation time compared to the original model.

2.4 Summary

Model reduction of linear systems is a well developed research area. Methods as balanced truncation provide error bounds and guaranteed preserved stability.

How to reduce nonlinear systems is however still a quite open problem and there is a large room for improvement of existing methods. Theorems concerning preserved stability or error bounds are sparse.

Common for all mentioned nonlinear methods in this chapter, except the piece-wise linear approach, is that even though the order is reduced, simulation time is in general not shorter. This is due to the fact that in practice most models are sparse and this sparsity is lost using the substitution shown in (2.10). Many states and small right-hand-side functions are replaced with few states but large right-hand-side functions.

3

Modeling the exhaust gas oxygen sensor

3.1 Introduction

Lambda sensors, or exhaust gas oxygen(EGO) sensors, are core components in the emission control in modern spark ignition combustion engines. The sensor, shown in Figure 3.1, is typically placed in the exhaust gas manifold between the engine and the catalyst. The performance of catalysts is highly dependent on exhaust gas composition and, e.g., the air-fuel ratio needs to be precisely controlled. A common air-fuel ratio control setup is illustrated in Figure 3.2.

Figure 3.1 An exhaust gas oxygen sensor

There exists many different kinds of oxygen sensors but the most commonly used is the zirconia switch-type sensor, this chapter is focused on this type. The sensor generates a voltage of roughly one Volt

if the air-fuel ratio is rich and zero Volt otherwise, see Figure 3.3. The lambda value is another name for air-fuel ratio and in this chapter, a normalized lambda value is considered, where the value one implies stoichiometric conditions.

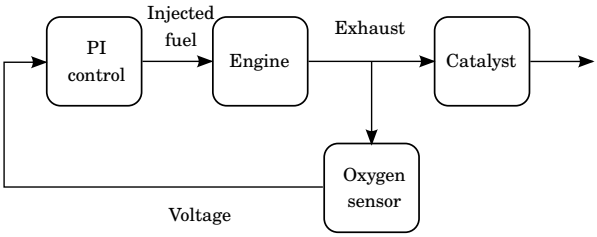

Figure 3.2 An air-fuel ratio control scheme

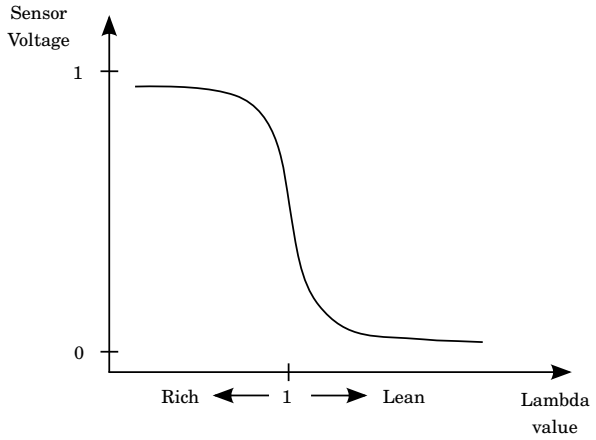

Figure 3.3 Lambda sensor characteristics

To meet future emission legislations, it is required to refine and extend current lambda control strategies. Good understanding of catalyst operation is essential to improve emission performance. It is necessary to understand the interaction of the catalyst and the lambda control system, including the lambda sensors, to optimize the exhaust gas treatment. Physically based simulation models are then vital tools

to analyze and evaluate new control strategies. An important part in this is the sensor models, and their ability to correctly reproduce effects of significance to catalyst operation. Of particular interest is the shift in voltage characteristics with respect to lambda value that is observed when the exhaust gas is diluted with hydrogen or carbon monoxide.

3.2 Modeling the exhaust gas oxygen sensor

A model with moderate complexity, which captures the lambda characteristics and it's dependency of hydrogen and carbon monoxide is sought for. A model of reasonable complexity level is developed in [Fleming, 1977] but with the drawback of not being able to model hydrogen dependency. The model presented in [Auckenthaler *et al.*, 2002] is a very detailed and complex model based on state of the art methods in literature. It models the hydrogen dependency along with many other effects. Unfortunately it suffers from numerical ill-conditioning. Due to the great time-scale difference between the electrode dynamics and the diffusion, the model becomes very stiff and is therefore difficult to use, e.g. in simulation. Possible model extensions could be

- Extend Fleming's model with hydrogen dependency.
- Derive an equilibrium approximation to Auckenthaler's model in order to avoid the numerical stiffness problem.

Both alternatives have been investigated, but more progress was obtained by following a model description found in [Barrick *et al.*, 1996]. This model is compact, considers hydrogen dependency and is static, so no numerical stiffness problem arises. Whether or not the dynamics of the sensor can be neglected depends on the control scheme and sensor placement. Here it is assumed that the dynamics can be disregarded.

The physics of the sensor

The sensor mainly consists of four manifold layers between the exhaust and reference gas(outside air), as can be seen in the cross-section in Figure 3.4.

A close-up of the four layers is shown in Figure 3.5. As the figure indicates, the model by Barrick takes into account the species H_2, CO,

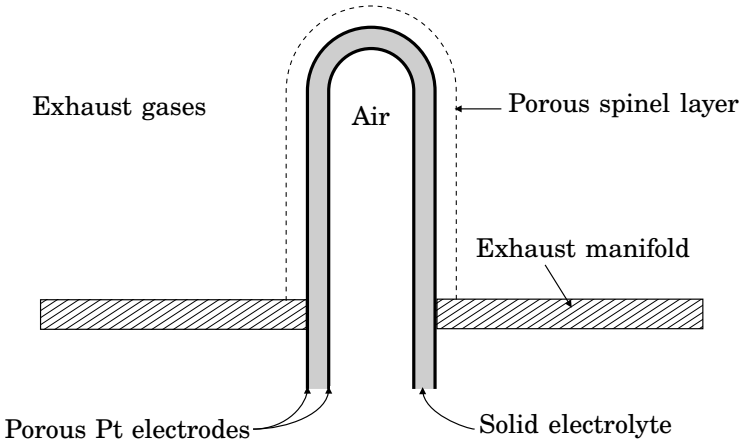

Figure 3.4 Sensor cross-section

O_2, CO_2 and H_2O, other species are assumed not to affect the sensor voltage.

Firstly, the exhaust gases have to diffuse through the porous spinel layer to affect the sensor voltage. Different species have different mass, and therefore different diffusion velocities, so the concentrations at the platinum surface are different compared to the ones in the exhaust gas close to the sensor. Secondly, at the cathode surface the platinum acts as a catalyst for the chemical reactions bringing the modified gas mixture to chemical equilibrium. And finally, the difference in gas concentrations at the electrodes yields the sensor voltage. The sensor model can thus be divided into three parts

- Diffusion through the porous spinel layer
- Platinum surface reactions
- The resulting sensor voltage

Here these three parts will be studied in more detail.

Diffusion N_2 is assumed to be abundant in the exhaust gas and is viewed as a media for the other species. Binary diffusion (also called

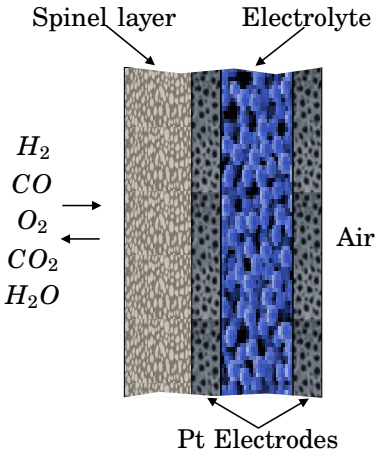

Figure 3.5 Sensor layer close-up

Fick's law diffusion) is then a valid approximation. In [Barrick *et al.*, 1996] a more detailed transport model with Stefan-Maxwell diffusion was investigated without gaining much accuracy. Thus, in principle the species diffuse through the porous layer without interacting with each other. When the platinum surface is reached they will combine according to the reactions in Equation 3.2 until chemical equilibrium is reached and diffuse out of the sensor again. Balancing the steady state flux of the three kinds of atoms gives rise to three linear equations

$$\begin{aligned} D_{CO_2}(X_{CO_2} - X^{exh}_{CO_2}) &= -D_{CO}(X_{CO} - X^{exh}_{CO}) \\ D_{H_2O}(X_{H_2O} - X^{exh}_{H_2O}) &= -D_{H_2}(X_{H_2} - X^{exh}_{H_2}) \\ D_{O_2}(X_{O_2} - X^{exh}_{O_2}) &= \frac{1}{2}D_{CO}(X_{CO} - X^{exh}_{CO}) + \frac{1}{2}D_{H_2}(X_{H_2} - X^{exh}_{H_2}) \end{aligned} \tag{3.1}$$

where

- X_i, X^{exh}_i are the molar fractions of gas i at the platinum surface resp. in the exhaust.
- D_i is the diffusion coefficient of gas i in N_2, which is dependent of temperature, pressure and tortuosity of the material.

Platinum surface reactions The model contains five active species interacting through two simplified reactions

$$\begin{aligned} CO + \frac{1}{2}O_2 &\rightleftharpoons CO_2 \\ H_2 + \frac{1}{2}O_2 &\rightleftharpoons H_2O \end{aligned} \tag{3.2}$$

A key simplification is to assume chemical equilibrium at the platinum surface, which induces two nonlinear algebraic equations

$$\begin{aligned} X_{CO}\sqrt{X_{O_2}} &= k_C X_{CO_2} \\ X_{H_2}\sqrt{X_{O_2}} &= k_H X_{H_2O} \end{aligned} \tag{3.3}$$

where k_C and k_H are temperature dependent constants arising from reaction velocities.

Sensor voltage In order to keep down model complexity, the three phase boundary cites (where species can be adsorbed) are assumed to be abundant. There is no competition for vacant sites, so the voltage model only depends on O_2 concentration, see [Barrick *et al.*, 1996]. The sensor voltage is then obtained by

$$V = -\frac{RT}{4F}\ln\frac{X_{O_2}}{0.21} \tag{3.4}$$

where R is the universal gas constant, T temperature in Kelvin, F Faraday's constant and 0.21 is the molar fraction of oxygen in the reference air.

An extension to this voltage model was described in [Fleming, 1977] where carbon monoxide's chemical effect on the voltage was included. It is however equivalent to (3.4) when assuming chemical equilibrium at the platinum surface so the simpler version was chosen in favour of low complexity.

Parameter estimation

For this model no calibration experiments are needed since all parameters are physical constants. Some are very well known, e.g. Faraday's constant, and others can be estimated using semi-empirical formulas.

Diffusion velocities The diffusion velocities have been estimated using the method by Chapman and Enskog, see [Reid *et al.*, 1977]. This method has an accuracy of about 6% error margin. As mentioned, the velocity is temperature and pressure dependant. However, all velocities have approximately the same dependencies, so by dividing with a nominal velocity in equation 3.1, the temperature and pressure dependencies can be omitted.

The method by Chapman and Enskog estimates the diffusion velocity in an open geometry. In the sensor however, gases diffuse through a porous media and this has to be taken into account. The standard way to deal with this, see [Smith, 1981], is to multiply the nominal velocity with a material dependent factor

$$D_i^* = \frac{\epsilon}{\tau} D_i$$

The tortuosity, τ, and the porosity, ϵ, are material specific and the same for all species i. Inserted in (3.1), the factor does not have an impact and can therefore also be disregarded.

Chemical equilibrium constants The program *HSC Chemistry*^{TM} has been used to estimate the chemical equilibrium constants k_C and k_H.

Lambda characteristics perturbation

This effect is analyzed in [Saji *et al.*, 1988], where a lambda sensor was exposed to different test gas mixtures. Some results presented in the article are illustrated in Figure 3.6, where it can be seen that deuterium has a different impact than H_2 although they have the same chemical properties. This displays that the perturbation effect is not due to chemical reactions but to physical properties, i.e. diffusion. The article also claims that this effect depends on non-equilibrium gas concentrations in combination with diffusion.

For example, for lambda equal to 1.1 a gas mixture in chemical equilibrium would contain O_2 and almost no H_2. The O_2 would be able to undisturbed diffuse to the cathode and as the sensor mainly produces voltage in function of difference in O_2 concentration, no(or low) voltage would be obtained. However, if the gas is not in chemical equilibrium there is O_2 and H_2 present in the exhaust. H_2 diffuses faster

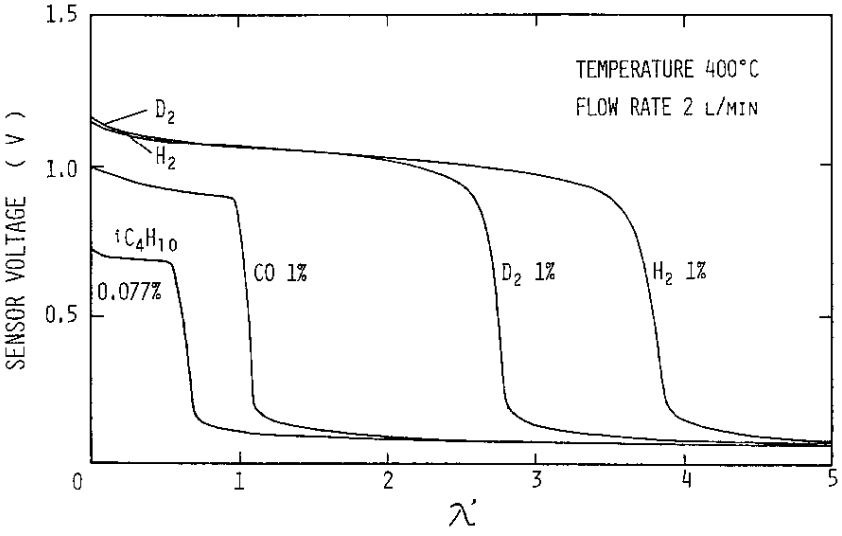

Figure 3.6 Lambda characteristics for non-equilibrium gas concentrations from [Saji *et al.*, 1988]. Reproduced by permission of ECS - The Electrochemical Society

than oxygen and the cathode would be exposed to a disproportionally large amount of H_2. The cathode acts as a catalyst for the reaction

$$H_2 + \frac{1}{2}O_2 \rightleftharpoons H_2O$$

and the oxygen is depleted inducing a high sensor voltage. Similar reasoning can be applied to other disturbing gases, the deviation depends on the degree of non-equilibrium and the mass difference between the species. This explains the perturbations in Figure 3.6, taken from [Saji *et al.*, 1988].

If the exhaust gas is heated to a higher temperature the gas would have a higher probability to reach equilibrium before exposing the sensor as can be seen in Figure 3.7, also taken from [Saji *et al.*, 1988].

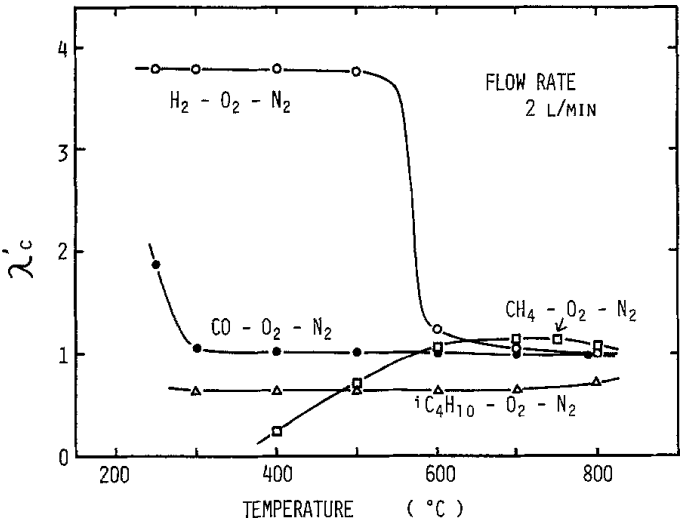

Figure 3.7 Switch point sensibility to temperature from [Saji *et al.*, 1988]. Reproduced by permission of ECS - The Electrochemical Society

Possible model extensions

- Sensor dynamics can probably be neglected, but should preferably be modeled in case it has importance.

  One way to introduce dynamics could be by adding, to the current static model, a linear first order filter with a time constant corresponding to the diffusion time of oxygen.

- Consider, and if necessary include in model, effects of NO_x and methane in the exhaust gas.

  A first attempt could be to model the effect NO_x and methane has on gas outside the sensor and in that way perturb the voltage. This can be motivated since NO_x and methane are both heavy species, and therefore diffuse slowly. For this reason they probably do not have an active role at the cathode layer.

3.3 Implementation

The implementation has been done in the Modelica language, see [Fritzson, 2004]. The model sums up into a set of nonlinear algebraic equations

$$
\begin{aligned}
D_{CO_2}(X_{CO_2} - X^{exh}_{CO_2}) &= -D_{CO}(X_{CO} - X^{exh}_{CO}) \\
D_{H_2O}(X_{H_2O} - X^{exh}_{H_2O}) &= -D_{H_2}(X_{H_2} - X^{exh}_{H_2}) \\
D_{O_2}(X_{O_2} - X^{exh}_{O_2}) &= \frac{1}{2}D_{CO}(X_{CO} - X^{exh}_{CO}) + \frac{1}{2}D_{H_2}(X_{H_2} - X^{exh}_{H_2}) \\
X_{CO}\sqrt{X_{O_2}} &= k_C X_{CO_2} \\
X_{H_2}\sqrt{X_{O_2}} &= k_H X_{H_2O} \\
V &= -\frac{RT}{4F}\ln\frac{X_{O_2}}{0.21}
\end{aligned}
$$

where the exhaust molar fractions, X_i^{exh}, are considered as model inputs and the sensor voltage, V, as model output. The sole alteration that has to be done to get a working Modelica code is the coordinate change

$$Y_i = \log(X_i)$$

for molar fractions inside the sensor. The new coordinates give far better numerical results for calculating small concentrations. The Modelica code has been simulated with the software tool Dymola, see [Dynasim AB, 2006]. The top view diagram is shown in Figure 3.8.

3.4 Simulations

The top plot in Figure 3.9 shows a 6 species (N_2 is omitted) varying gas configuration, this gas is at all lambda values in chemical equilibrium. The lower plot of the same figure show the same gas mixtures but the species has been manually modified to deviate from chemical equilibrium, by pushing the reactions in (3.2) from the equilibrium points.

Exposing the model to these gas mixtures yields the voltage in Figure 3.10. As can be seen, the model shows promising and reasonable results, the voltage shift due to non-equilibrium H_2 is clearly visible.

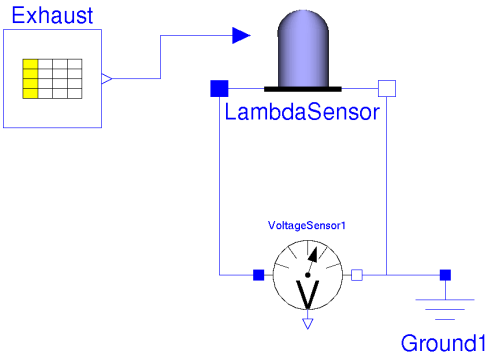

Figure 3.8 Layout of Dymola model

3.5 Model validation

Data from test-gas experiments has been used in validation purposes. In-house experiments were conducted but the equipment for oxygen concentration measurement turned out to have inadequate resolution. Instead, other data was used that, unfortunately, are proprietary information and not publishable.

Two types of experiments were used for validation, a constant flow of hydrogen or carbon monoxide were mixed with oxygen and nitrogen. The gas mixture was then heated to 500°C and exposed to the sensor, see Figure 3.11. During the tests the sensor voltage together with the lambda value were measured while the gas composition was changed according to Figure 3.12.

The lambda value for the hydrogen experiment is defined by the combustion reaction

$$2H_2 + \lambda O_2 \rightarrow n_p(X_{H_2}H_2 + X_{O_2}O_2 + X_{H_2O}H_2O)$$

where n_p is the total mole amount of the gas. An expression for the lambda value is achieved by balancing the amount of hydrogen and

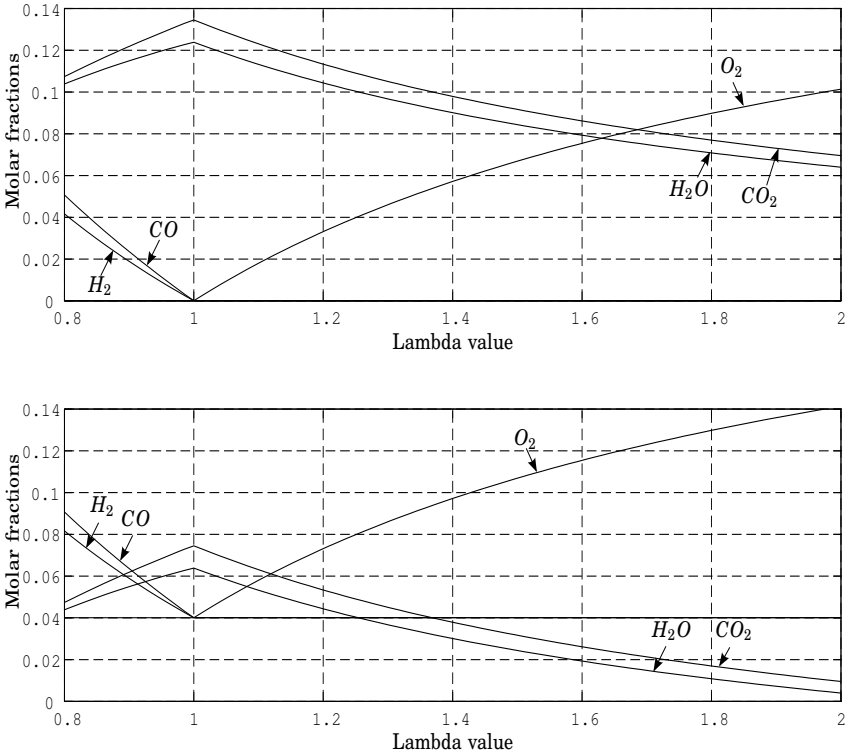

Figure 3.9 Gas mixtures with the corresponding lambda value

oxygen atoms.

$$\lambda_H = \frac{X_{H_2O} + 2X_{O_2}}{X_{H_2} + X_{H_2O}} \tag{3.5}$$

The same procedure for carbon monoxide yields

$$\lambda_C = \frac{X_{CO_2} + 2X_{O_2}}{X_{CO} + X_{CO_2}} \tag{3.6}$$

As can be seen, the gas configuration is not uniquely defined even though the hydrogen (resp. carbon monoxide) concentration is known.

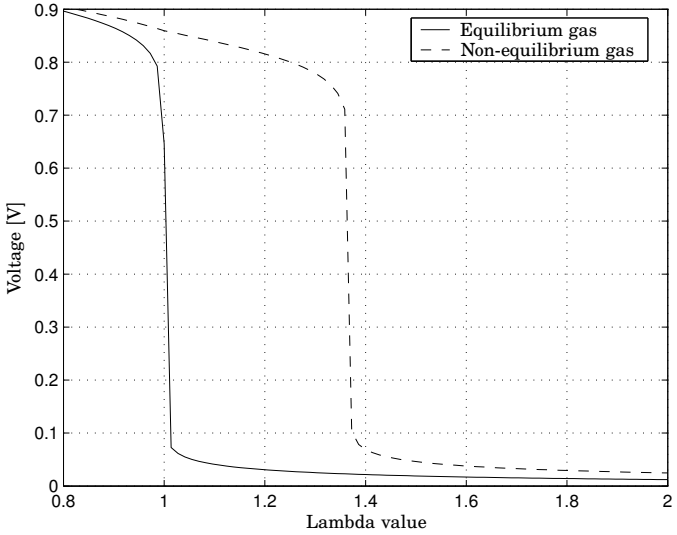

Figure 3.10 Lambda characteristics simulation

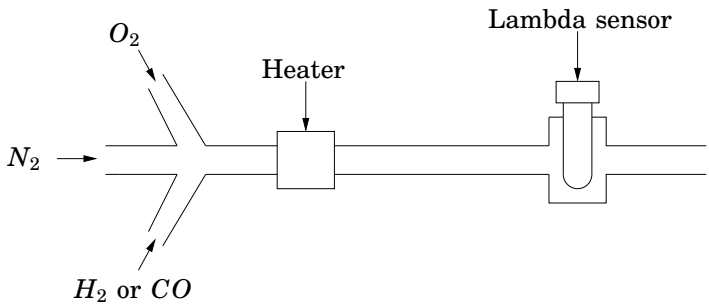

Figure 3.11 Experiment setup

To analyse the completeness of combustion reactions at the sensor position in Figure 3.11, the software *Cantera* [Cantera, 2006] was used to simulate the reaction dynamics. It turned out that a temperature above 700°C is needed to activate the reactions in the mixed gas. The

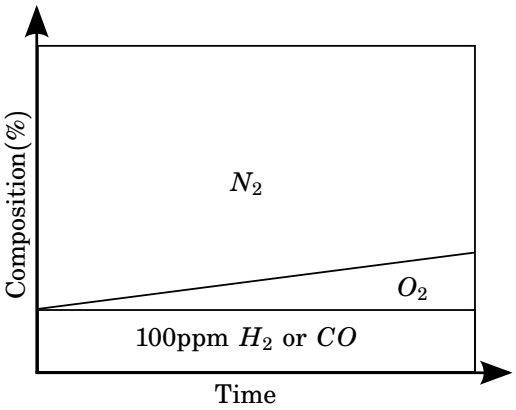

Figure 3.12 Experiment gas composition

reaction speeds can be seen in Figure 3.13 and 3.14. For lower temperatures it's a good approximation to assume that the gases do not react and the concentrations remain unchanged until they reach the sensor.

The absence of H_2O and CO_2 in the mixed gas change (3.5) and (3.6) into

$$\lambda_H = \frac{2X_{O_2}}{X_{H_2}}$$

$$\lambda_C = \frac{2X_{O_2}}{X_{CO}}$$

Now the gas composition is uniquely defined by the lambda value and the corresponding experiment can be simulated with the model derived in Section 3.2.

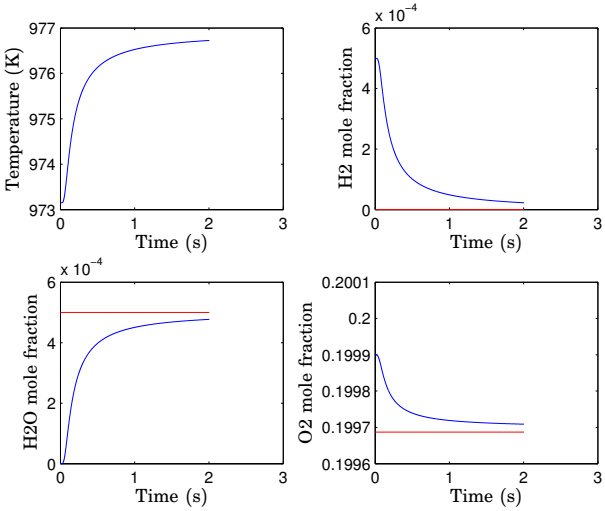

Figure 3.13 Reaction speed with hydrogen

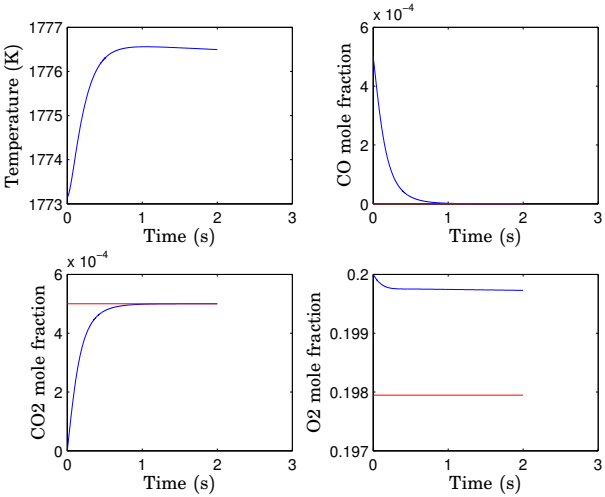

Figure 3.14 Reaction speed with carbon monoxide

Validation results The simulated model has been compared with the experiment data, with 100ppm of H_2 resp. CO at a temperature of 500°C. As mentioned, the experiment data are proprietary information and is not publishable. However, the model output, shown in Figure 3.15, captures the sensor behaviour well. The curves have approximately the same switching point as the measured data and the voltage, at the rich and lean sides, does not deviate much. The mean absolute error compared to experiment data is 0.14V for CO and 0.057V for H_2.

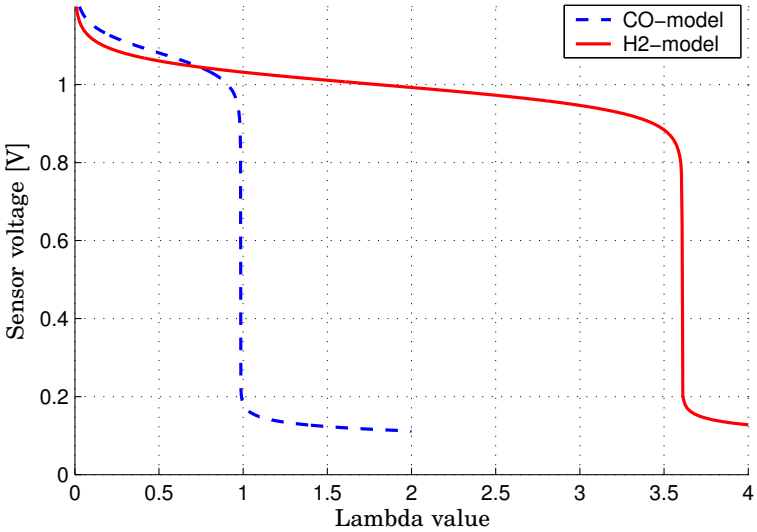

Figure 3.15 Model output for a gas mixture with 100ppm CO or H_2 at 500°C

Calibration Most of the model parameters described in Section 3.2 are estimated with a level of uncertainty. If they are treated more as non fixed parameters than natural constants, then higher accuracy to the experiment data can be achieved. In Figure 3.16 the calibrated model's output is shown. Now the mean absolute error is reduced to 0.033V for CO and 0.0261V for H_2. Here the diffusion velocities are modified but kept inside the Chapman and Enskog method's error

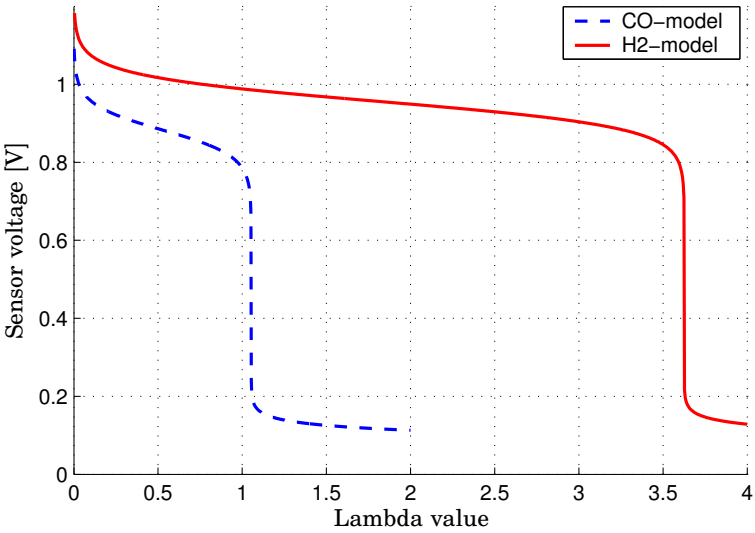

Figure 3.16 Calibrated model output for a gas mixture with 100ppm CO or H_2 at 500°C

margin. The reaction constant was modified corresponding to a 100°C change to match the voltage level for rich mixtures of carbon monoxide.

3.6 Conclusions

A simple and static model with H_2 dependency has been developed and implemented in the Modelica language. Simulations show reasonable results where the effects of H_2 in a non-equilibrium gas can be observed.

The model has successfully been validated with test gas experiment data. By adapting parameters within reasonable physical limits, higher fidelity to experiment measurements was achieved. The mean error in sensor output voltage did not exceed 3% of the maximum output, when the model was compared to experiment data.

4

A model reduction case study

The contents of this chapter is based on the article [Nilsson *et al.*, 2006].

Low complexity plant models are essential for model based control design. Often a detailed high order model is available and simplification to a low order approximate model is needed. This chapter presents a case study of two model reduction methodologies applied on the automotive engine air path. The first methodology is based on balanced truncation of models obtained by linearization around equilibria and trajectories. Under appropriate assumptions, this technique yields strict bounds on the approximation error. The second is a heuristic methodology, based on intuition commonly used in modeling of engine dynamics. Although it is successfully used in practice, the approximation error is seldom known. The two methodologies are used to derive simple models for the required fuel charge in a spark ignition engine, given throttle and swirl flap positions and engine speed. Performance, complexity and similarities of the two resulting low order models are compared.

4.1 Introduction

The air path dynamics is a major problem in automotive engine control. The main problem for spark ignition (SI) engines is to regulate

the Air/Fuel Ratio. The electronic fuel control system of a modern SI automobile engine employs individual fuel injectors located in the inlet manifold runners close to the intake valves to deliver precisely timed and accurately metered fuel to all cylinders. This fuel management system acts in concert with the three-way catalytic converter (TWC) to control HC, CO, and NO_x emissions. Figure 4.1 illustrates

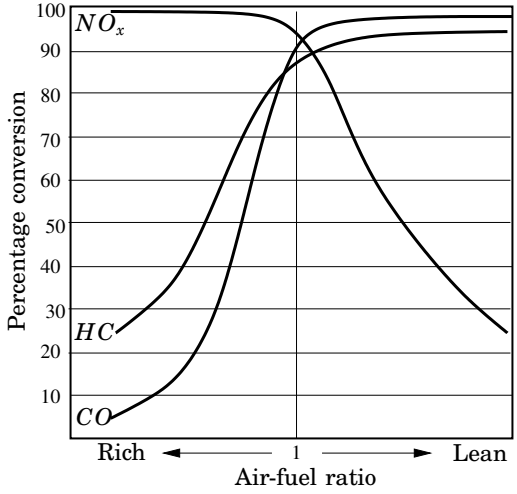

Figure 4.1 Typical conversion efficiency of a three-way catalyst

the conversion efficiencies provided by a typical TWC as a function of exhaust air-fuel ratio (A/F) for the three constituents. It can be seen that there is only a very narrow range of A/F near the stoichiometric value (14.64) over which high simultaneous conversion efficiencies may be attained, see [Heywood, 1988]. In order to utilize the TWC effectively, feedback from an exhaust gas oxygen (EGO) sensor in the vehicle exhaust system is used to regulate the A/F operating point, see Figure 4.2. If the operating point is changed by, for example, an increased torque demand, the injected fuel amount has to increase. The EGO sensor will not instantly detect the unbalance in the A/F ratio and a good feed-forward control is needed to adapt the injected fuel amount before the deviation is detected by the sensor. To this purpose,

a good low complexity model of the air entering the cylinder is essential. Given throttle position, swirl flap position and engine speed the required fuel charge (Fc) has to be estimated to achieve stoichiometric conditions.

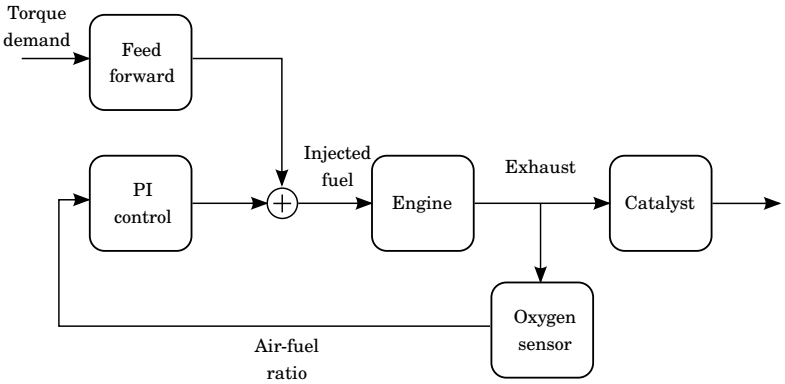

Figure 4.2 A standard air-fuel ratio control scheme

4.2 Model properties

The article is focused on the setup shown in Figure 4.3, which shows an illustration of the air path. The throttle is used to get the desired airflow and the swirl flap is used for inducing turbulence and thereby achieving better mixing in the cylinder. Volume 1 and 2 in the figure represent the connecting pipes between the elements.

The base for the model reduction is a detailed one-cylinder model provided by Toyota Motor Corporation. It is written in the Modelica language, see [Fritzson, 2004], and managed with the software tool Dymola, a multi-domain modeling and simulation tool, see [Dynasim AB, 2006]. The model is based on conservation laws such as mass balances. The top view of the Dymola model can be seen in Figure 4.4, which shows the same physical layout as Figure 4.3.

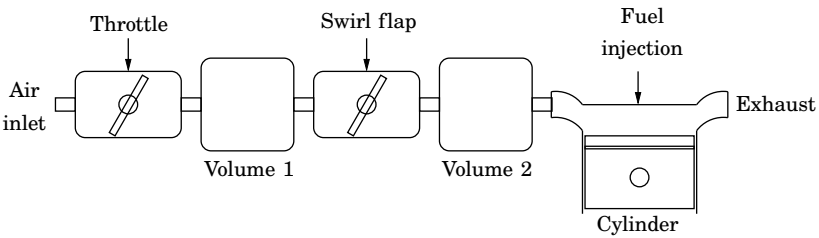

Figure 4.3 Schematic of the engine air path

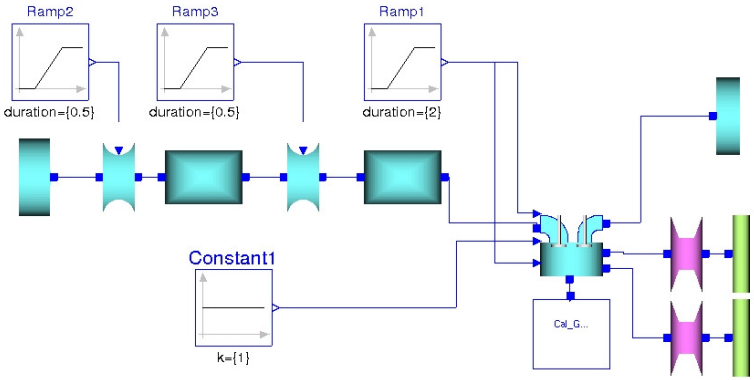

Figure 4.4 Top view of the one cylinder Dymola model

Translating the model Dymola induces a nonlinear differential algebraic equation (DAE) with 37 continuous time states, distributed as

- 11 states in volume 1
- 11 states in volume 2
- 15 states in the cylinder

The states in the three objects are among other things mass, energy, momentum and concentrations of the seven species gas mixture. The model is ideal for simulation but is too complex for model based control

design. By applying model reduction techniques a low-order approximate model can be derived.

4.3 Model reduction by balanced truncation

In this section the balanced truncation method is applied to obtain a low order approximate model. The theory of balanced truncation, as described in section 2.2, is clearly not directly applicable on the Dymola model as it is non-linear, hybrid and is defined both by equations and algorithms. However, with the below described methodology an approximate low order model is obtained. The methodology can be separated into three steps.

Obtain a linear time-varying system by repeated linearization The Dymola model was simulated with constant input signals, i.e. throttle position, swirl flap position and engine speed. The simulation gave rise to a state-trajectory around which the non-linear model can be linearized. Since the model is nonlinear the linearized model becomes time-varying, i.e. the A,B,C and D matrices are time dependent.

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)x(t) + D(t)u(t) \end{aligned} \tag{4.1}$$

The state vector x is the deviation from the nominal state trajectories and all 37 states have known physical interpretations. The fuel charge is defined once per engine cycle (when the inlet valve closes) and is proportional to the amount of oxygen in the cylinder. To be able to linearize the system at any time point a continuous version of the signal was defined that coincides with the discrete when the inlet valve closes.

Dymola has the functionality of derivation of linearizations of the nonlinear DAE. By using Dymola's scripting capabilities this can be done repeatedly at times t_k and snapshots, with $50\mu s$ intervals, of the continuous linear time-varying system in (4.1) is obtained. With the assumption that the $A(t)$,$B(t)$,$C(t)$ and $D(t)$ matrices are constant between the times t_k a discrete linear time-varying (LTV) system can be derived by zero order hold sampling. The discrete system (4.2) only

captures the state vector at the snapshot times, $x(t_k) = x_k$.

$$\begin{aligned} x_{k+1} &= \Phi_k x_k + \Gamma_k u_k \\ y_k &= C_k x_k + D_k u_k \end{aligned} \tag{4.2}$$

Resample the discrete LTV system once per engine cycle The required fuel charge is defined once per engine-cycle and the system is therefore sampled. Resampling, with the sampling periods defined by the closing of the inlet valve, gives rise to a cycle-to-cycle model for the fuel charge. Here the input signals are assumed to be constant during the cycle and n denotes the number of sampling intervals per cycle.

$$x_{k+n} = \tilde{\Phi} x_k + \tilde{\Gamma} u_k$$

Letting $k = 0$ represent the first sample time in the cycle the matrices $\tilde{\Phi}$ and $\tilde{\Gamma}$ can be computed according to

$$\begin{aligned} \tilde{\Phi} &= \prod_{i=1}^{n} \Phi_{n-i} \\ \tilde{\Gamma} &= \sum_{i=0}^{n-2} \left(\prod_{j=1}^{n-i-1} \Phi_{n-j} \right) \Gamma_i + \Gamma_{n-1} \end{aligned}$$

Linearization around trajectories captures hybrid phenomena such as dynamics changing depending of time and position in state-space, but not instantaneous changes such as reset maps. The Dymola model contains reset maps, for example the cylinder mass is instantaneously increased by the amount of fuel injected, similarly the oxygen amount is reset after combustion. These two and others all occur when the inlet valve closes and can all be represented as an instantaneous linear transformation of the state vector

$$x_k^* = H x_k$$

This can be included in the LTV system by introducing the transformation in the beginning of the cycle, i.e. when the inlet valve closes.

$$x_1 = \Phi_0 H x_0 + \Gamma_0 u_0$$

$\tilde{\Gamma}$ is not affected and the only alteration is in the calculation of $\tilde{\Phi}$, which becomes

$$\tilde{\Phi} = \left(\prod_{i=1}^{n} \Phi_{k+n-i} \right) H$$

Only slight cycle-to-cycle variations can be seen in $\tilde{\Phi}$ and $\tilde{\Gamma}$ but to improve numerical precision the matrices are calculated by averaging over several cycles.

Apply balanced truncation to obtain low-order model By linearization and resampling a linear time-invariant model for the required fuel charge has been derived. And now the balanced truncation method, as described in Section 2.2, can be applied.

The model has 37 states, a number that can significantly be reduced without much loss of accuracy. The five largest Hankel singular values described by (2.5) are shown in Figure 4.5. The plot indicates

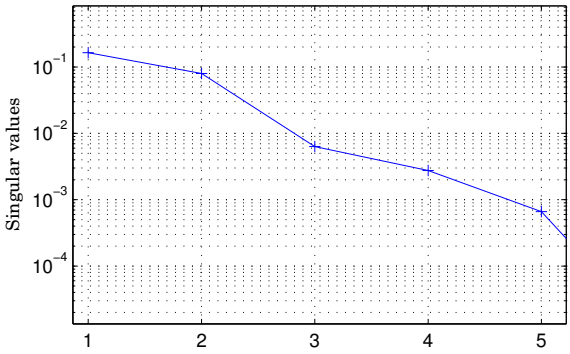

Figure 4.5 The five largest singular values of the balanced realization

how well the model can be represented with a lower order approximation. How many states the low order model should have is a trade off between approximation error and model complexity. In this case simulation shows that it is reasonable to truncate all but two states. The new low order state vector will be a linear combination of the physical states and reducing to a second order system yields the coordinate

change

$$\bar{x} = Tx = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} x$$

where $T \in \mathbf{R}^{2x37}$. The absolute value of the elements in T_1 and T_2 can be seen in Figure 4.6, which illustrates the relative importance of the physical states in the reduced model. The most important states are listed in Table 4.1.

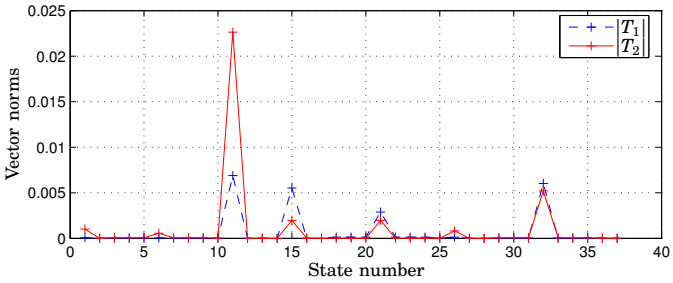

Figure 4.6 Relative importance of physical states in the reduced model

Table 4.1 The most important physical states

State number	Physical interpretation
11	Amount of oxygen in cylinder
15	Mass in volume 1
21	Amount of oxygen in volume 1
32	Amount of oxygen in volume 2

Results

The nonlinear Dymola model with 37 states has been approximated with a 2 state linear time-invariant system. Figure 4.7 shows the required fuel charge computed by the original Dymola model, the linearized model and the reduced model as a response to the illustrated

change of input signals. As can be seen, both the linearized and reduced model approximates well the Dymola model's result. For this trajectory the approximation error is dominated by the linearization and not the truncation.

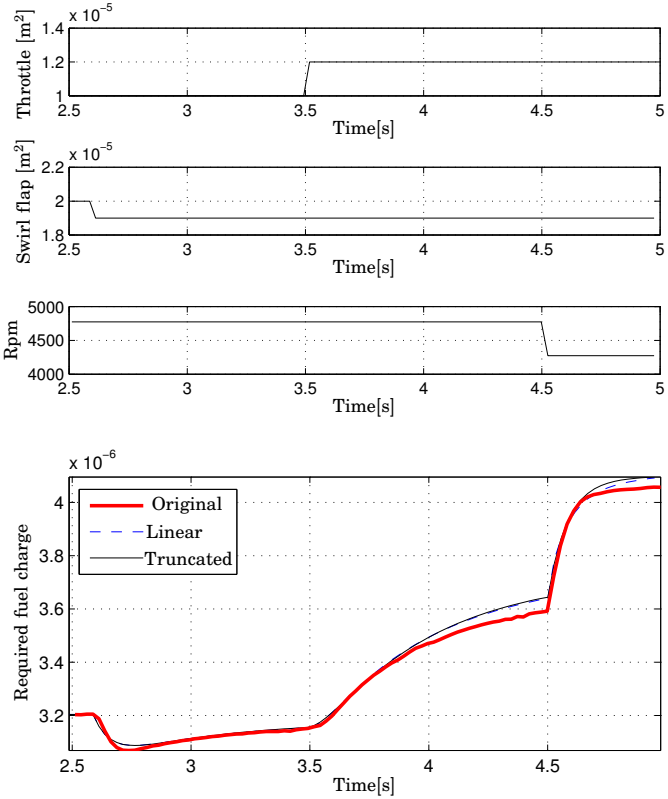

Figure 4.7 Simulation results for the linearized and truncated model

Equation (2.6) gives a bound on the approximation error between the two linear models and keeping two states yields

$$\max_u \frac{\|\tilde{y}(t) - y(t)\|_2}{\|u(t)\|_2} \leq 2 \sum_{k=3}^{37} \sigma_k \tag{4.3}$$

where $\tilde{y}$ is the reduced model output. In this case the input signal $u(t)$ consists of three scalar signals, throttle position, swirl flap position and engine speed. The norm is calculated according to

$$\|u(t)\|_2 = \sqrt{\int_0^\infty u^T(t)u(t)dt}$$

The throttle and swirl flap positions are given in effective area $[m^2] \sim 10^{-5}$ and engine speed in [rpm]$\sim 10^3$, as can be seen the engine speed will greatly dominate the input signal norm. To achieve more reasonable results the inputs are balanced by including a 10^{-8} gain before the throttle and swirl flap positions in the model, which could correspond to a unit change. Now the input signals all have the approximate magnitude of 10^3 and if all but two states are truncated, according to (4.3), the following holds

$$\|\tilde{y}(t) - y(t)\|_2 \leq 2.6784 \cdot 10^{-10} \|u(t)\|_2$$

$y(t)$ and $\tilde{y}(t)$ denotes the outputs of the 37- and 2-state linear models. For the trajectory in Figure 4.7 this implies that the output error is bounded as

$$\|\tilde{y}(t) - y(t)\|_2 \leq 5.0 \cdot 10^{-6}$$

while the actual approximation error is $6.9 \cdot 10^{-7}$. This is not uncommon when applying balanced truncation, the result is often much better than the error bound indicates. There are two reasons for the difference, the norm is worst-case and that the bound is often conservative.

4.4 Heuristic model reduction

A common way to accomplish model reduction is to use experience and insight into the physics to omit dynamics with little importance. In this case the problem is split into two parts

1. The cylinder dynamics
2. The air path including the two volumes

where the first part has far faster dynamics than the second. Therefore the common technique of time scale dynamics reduction can be performed on the first part. The result is a static mapping from pressure in volume 2 and engine speed to required fuel charge. Simplification of more detailed physical modeling yields that the dynamics of the second part can be modeled as two first order dynamics. The states represent the pressure in each volume and the throttle and swirl flap positions act as input signals in a similar configuration as in Figure 4.3. The effect of varying engine speed is introduced in the model by letting the gains and time constants of the first order dynamics be dependent of engine speed, N_e. The complete model is the combination of the two parts and is illustrated in Figure 4.8.

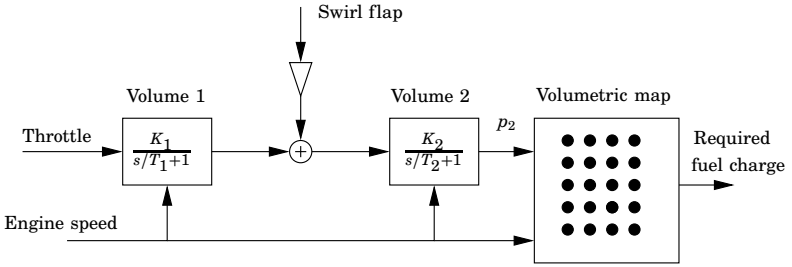

Figure 4.8 Structure of the heuristically derived model

A more detailed description of how the model structure was obtained for the two parts is presented below. For further details and background see [Chevalier *et al.*, 2000; Hendricks *et al.*, 1996] and [Føns *et al.*, 1999].

Time scale dynamics reduction

One simple model of the air flow in an intake manifold is the filling and emptying model. The air flow enters the manifold through the throttle and is pumped out of the manifold into the cylinder. Assuming no leaks, the intake mass air flow D_{air}, into the manifold and the flow entering in the cylinder, M_{asp} are identical only in steady state.

Pumping fluctuations Pumping fluctuations are caused by any disturbance initiated at the boundary of inlet manifold such as moving piston, moving valve and moving throttle plate. These disturbances travel along the pipe experiencing many reflections. When the engine is operated in the steady-state, they finally settle down into a standing wave. The source of pumping noise is periodic, so the pumping fluctuations are frequency locked to the engine event frequency.

Mean model of the aspirated flow In spite of the complexity of the fluid dynamic phenomena occurring during a transient (due to fast opening or closing of the throttle), the conventional volumetric efficiency η (function of the engine working point), identified during steady-state conditions, is used to describe the inlet air mass flow rate. So the speed-density gives an accurate description of the air mass flow rate through the inlet valve

$$D_{\text{asp,map}} = \eta_{\text{map}}(\bar{P}_2, N_e)\frac{V_{\text{cyl}}\bar{P}_2}{R\bar{T}_2}\frac{N_e}{120}$$

where:

- $\bar{P}_2$ is the mean pressure over a TDC in volume 2
- η_{map} is the volumetric efficiency
- $\bar{T}_2$ is the mean temperature over a TDC in volume 2
- N_e is the engine speed
- V_{cyl} is the cylinder volume
- R is the universal gas constant

The volumetric efficiency η_{map} is highly nonlinear function of the engine speed (N_e) and manifold pressure ($\bar{P}_2$). It can only be estimated via experimentation. Figure 4.9 shows the volumetric efficiency of a commercial gasoline engine. The used volumetric efficiency should preferably been generated from the Dymola model. The map derived by experiment data is considered to be accurate enough for the purposes of this article.

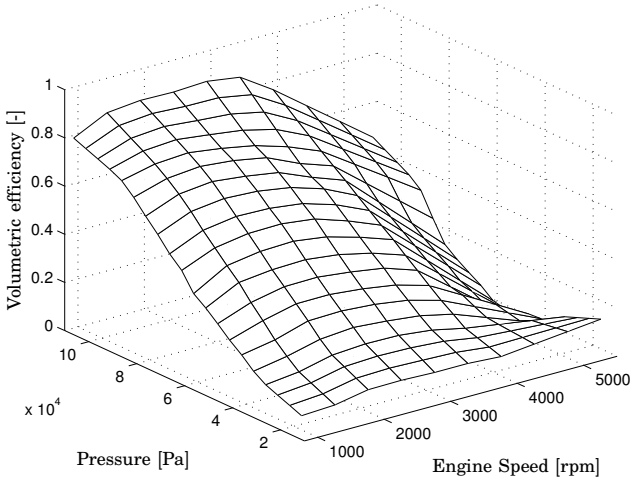

Figure 4.9 The volumetric efficiency map

Air path dynamics reduction

During throttle transients, the difference between these two flows equal the rate of change of the air mass in the manifold plenum. Assuming that the manifold pressure is uniform and the intake manifold temperature is constant, the continuity equation and ideal gas law can be applied to the manifold plenum. Usually, the flow is defined as a function of the total mass M_T and can be factorized as

$$d(M_T, O_{valve}) = p(M_T)M_T$$

where O_{valve} is the effective area of the valve and p is a positive increasing (concave) function with respect to the total mass M_T as proposed in [Heywood, 1988],

$$p(z, O_{valve}) = p_0(O_{valve})\sqrt{2\frac{\gamma}{\gamma-1}((\frac{z}{z_0})^{-\frac{2}{\gamma}} - (\frac{z}{z_0})^{-\frac{\gamma+1}{\gamma}})}$$

Here z_0 is the mass in atmospheric conditions, γ is the heat ratio and p_0 is a function transforming the valve opening into the effective area

of the valve. By linearization of the flow equation, the model of the two volumes writes as two first order dynamics with the states p_1 and p_2 (the pressure in the volume 1 and 2 respectively) as described in Figure 4.10. Although the parameters depend on engine speed, the calibration task is easier than for the original nonlinear model. Moreover, first-order dynamics (with the parameters varying with respect to the operating conditions) is well representative of the filling-emptying dynamics.

The combination of the time-scale reduction and the volume's dynamics yields the total model showed in Figure 4.8.

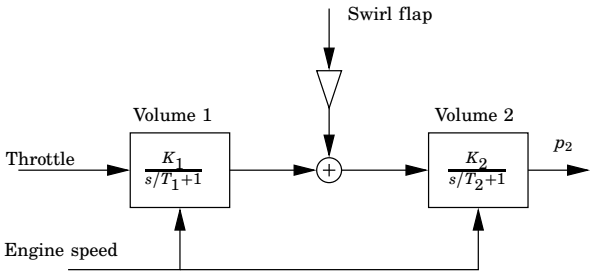

Figure 4.10 Air path dynamics approximation of the two volumes

Results

The nonlinear Dymola model with 37 states has been approximated with a second order linear parameter varying system combined with a look-up table. Figure 4.11 shows the required fuel charge computed by this model as a response to the illustrated change of input signals. The model is able to approximately describe the variation of the requested fuel charge. Nevertheless, the qualitative response is not so good. Indeed, without observers and correction mapping, a precise estimation of the aspirated flow is not available. It means that the air-fuel ratio controller-action will be necessary and predominant as the requested fuel charge is not well predicted.

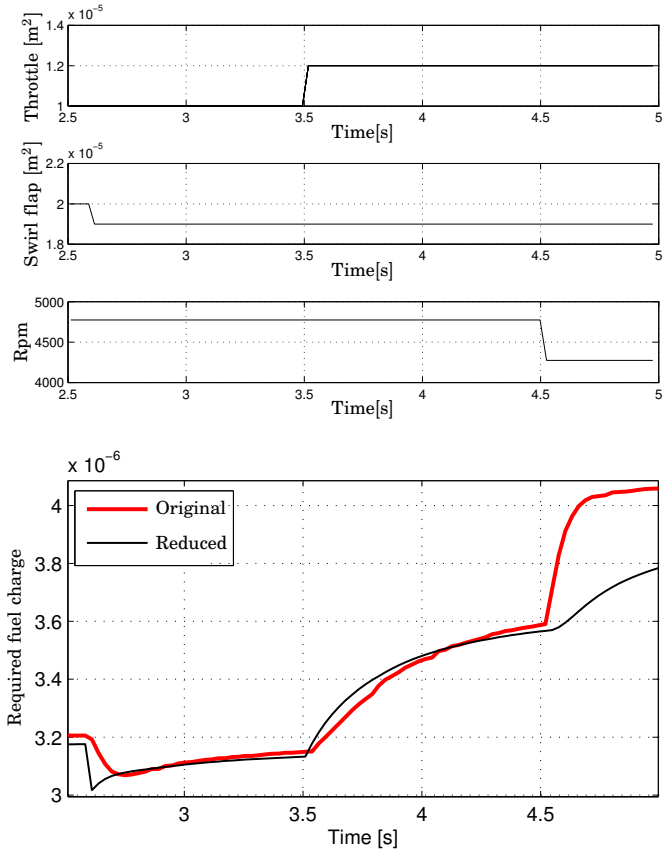

Figure 4.11 Simulation results for the heuristically derived model

4.5 Methodology comparison and conclusions

Two low order models have been derived, one using balanced truncation and one using heuristic methods. The complexity of the two resulting models is almost the same but the methodology is quite different.

Methodology comparison

In the case of the heuristic procedure the methodology is based on modeling and simplifications are made based on intuition and experience. When an appropriate model complexity is chosen, parameters are determined by physical properties or, as in this case, by tuning to fit simulation data. The tuning could also been done to fit experimental data, that is not the case for balanced truncation which needs a detailed model.

The resulting model from the balanced truncation based technique is always a linear time-invariant system, the heuristic procedure does not have this restriction and has therefore greater potential. On the other hand, it can be very hard to tune the parameters, especially for larger nonlinear systems. This time-consuming tuning can be compared to the rather heavy computations needed to derive the linearizations required for balanced truncation, which can be carried out by a computer without human supervision.

The balanced truncation methodology is very systematic and does not need physical knowledge of the model, neither is any parameter fitting necessary. It also delivers a bound on the approximation error compared to the linearized model.

In this example better performance was achieved with balanced truncation, this is however more a question of how well you can fit the parameters in the heuristics based model. In cases when nonlinear dynamics is essential the balanced truncation technique used here will not be sufficient.

Similarities in choice of states can be seen in the two methods. For example, both methods neglect the fast dynamics occurring in the cylinder and other gas species than oxygen are ignored. The heuristic method chose the volume pressures as states, which are (at constant temperature) proportional to the amount of oxygen. For the model generated by balanced truncation, the oxygen concentrations are present as components in the two states.

Conclusions

Both methodologies have their advantages and disadvantages. If a detailed model is available and linear behaviour is expected then the balanced truncation technique could be preferred. This technique can

require a large computation time but needs very little manual attention. Using the heuristic method requires more experience and knowledge, it may also involve extensive parameter fitting, but renders more insight to the simplifications made.

5

A new approach to balanced truncation of nonlinear systems

In this chapter, a new method for simplification of nonlinear input-output models is outlined. The method is based on state transformation followed by truncation of some states. Consider the following example.

EXAMPLE 5.1
The nonlinear system

$$\begin{aligned}\dot{x}_1 &= -3x_1^3 + x_1^2x_2 + 2x_1x_2^2 - x_2^3\\ \dot{x}_2 &= 2x_1^3 - 10x_1^2x_2 + 10x_1x_2^2 - 3x_2^3 - u\\ y &= 2x_1 - x_2\end{aligned}$$

has exactly the same input-output relationship as the system

$$\dot{y} = -y^3 + u \tag{5.1}$$

It is a challenge for any model reduction procedure to detect that a reduction like this is possible. A general methodology for such problems will be developed, but first a simple proof of the equivalence in this particular case is given.

Hence, note that the system can be rewritten as

$$\begin{aligned}\dot{x}_1 &= -(2x_1 - x_2)^2 x_1 + (x_1 - x_2)^3 \\ \dot{x}_2 &= -(2x_1 - x_2)^2 x_2 + 2(x_1 - x_2)^3 - u \\ y &= 2x_1 - x_2\end{aligned}$$

With the new variables $z_1 = 2x_1 - x_2$, $z_2 = x_2 - x_1$, this means that

$$\begin{aligned}\dot{z}_1 &= -z_1^3 + u \\ \dot{z}_2 &= -z_1^2 z_2 - z_2^3 - u \\ y &= z_1\end{aligned}$$

In particular, the state z_2 does not appear in the output. Hence, it can be truncated and (5.1) holds. □

In the example, a linear coordinate transformation followed by state truncation gave a simplified model without approximation error. The goal of the method described in this chapter is to provide a systematic way to find such transformations whenever they exist and otherwise to find good approximations.

5.1 Method description

Let the system to be reduced have the form

$$\begin{aligned}\frac{dx}{dt}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t))\end{aligned}$$

To find states that are redundant or that have small importance for the input-output relationship, linearizations of the system dynamics will be used. Local importance of states would be revealed if one would linearize the system around a stationary point. A combination of several linearization points could then indicate which the important states are in the nonlinear system. However, some states may only have an active role during transient behaviour, which will be later commented in

Example 5.4. Instead, linearization around a trajectory will be used as a tool to find an approximate low-order model.

Recall the theory concerning balanced truncation of linear time-varying systems presented in Section 2.2. The time-varying gramians give information about state importance even in transient regions of state-space. These gramians can be computed in the neighborhood of simulated trajectories using linearization of the system dynamics. When there exists a linear coordinate transformation that disconnects some states from the input-output relationship, this will be revealed in those localized gramians.

The choice of training trajectory, around which linearization is made, is an important aspect of the reduction procedure. The corresponding training input should be chosen as a typical input-signal, which is rich enough to excite all dynamics important to the intended model use.

The first step of the procedure is to simulate the system for $t = [0, t_f]$ with the training input-signal. Along the simulated trajectory, define

$$A(t) = \frac{\partial f}{\partial x}(x(t), u(t)) \quad B(t) = \frac{\partial f}{\partial u}(x(t), u(t)) \quad C(t) = \frac{\partial h}{\partial x}(x(t), u(t))$$

and compute the time-varying gramians $P(t)$, $Q(t)$ by simulation of the Lyapunov equations

$$\frac{dP}{dt}(t) = A(t)P(t) + P(t)A(t)^T + B(t)B(t)^T \qquad P(0) = 0 \qquad (5.2)$$

$$\frac{dQ}{dt}(t) = -Q(t)A(t) - A(t)^T Q(t) - C(t)^T C(t) \qquad Q(t_f) = 0 \qquad (5.3)$$

The controllability gramian $P(t)$ reveals how large deviation in input-signal is needed to perturb $x(t)$. If a certain state component is hard to perturb for all times, one can suspect that this state is in general hard to affect in the nonlinear system. Similarly, $Q(t)$ shows how much the output-signal is affected if $x(t)$ is perturbed. If the output-signal is weakly influenced by a certain state-perturbation, independently of when the perturbation is made, it can be suspected that this state needs a high variation to affect the output-signal in the nonlinear system.

In order to isolate the overall important states with a constant state-transformation, one could use the *average gramians*

$$\bar{P} = \frac{1}{t_f}\int_0^{t_f} P(\tau)d\tau \qquad\qquad \bar{Q} = \frac{1}{t_f}\int_0^{t_f} Q(\tau)d\tau$$

and then treat $\bar{P}$ and $\bar{Q}$ as if they belonged to a linear time-invariant system. Then rank deficiency of the matrix $\bar{P}\bar{Q}$ indicates that some states are obsolete and can be truncated from the model without changing the input-output relationship.

Define

$$T = [\,T_1 \quad \dots \quad T_n\,] \qquad\qquad T^{-1} = \begin{bmatrix} S_1 \\ \vdots \\ S_n \end{bmatrix}$$

to diagonalize $T^{-1}\bar{P}\bar{Q}T$ with diagonal elements in decreasing order and introduce new coordinates according to the formula $Tz = x$. The coordinate change T is obtained in the same manner as for balanced truncation of linear time-invariant systems. Then the truncated model

$$\begin{aligned} \frac{dz}{dt}(t) &= \begin{bmatrix} S_1 \\ \vdots \\ S_m \end{bmatrix} f\big([\,T_1 \quad \dots \quad T_m\,]\,z(t), u(t)\big) \\ \widehat{y}(t) &= h\big([\,T_1 \quad \dots \quad T_m\,]\,z(t), u(t)\big) \end{aligned} \tag{5.4}$$

approximates the input-output behavior of the original system near the simulated trajectory provided that the m largest eigenvalues of $T^{-1}\bar{P}\bar{Q}T$ are dominating.

For further illustration, the method is applied in the three following examples.

EXAMPLE 5.2

Consider again the system defined in Example 5.1. Simulating this system along various trajectories and computing the observability gramian according to the differential equation

$$\frac{dQ}{dt}(t) = -Q(t)A(t) - A(t)^T Q(t) - C(t)^T C(t) \qquad Q(t_f) = 0$$

one observes that the rank of $Q(t)$ never exceeds one for any trajectory. The reason is that $q(t) = [1 \quad 2]\,Q(t)\,[1 \quad 2]^T$ obeys the differential equation

$$\frac{dq}{dt}(t) = -[1 \quad 2]\left[Q(t)A(t) + A(t)^T Q(t) + C(t)^T C(t)\right][1 \quad 2]^T$$

with $q(t_f) = 0$. In this case

$$\begin{aligned} A(t)\begin{bmatrix}1\\2\end{bmatrix} &= \begin{bmatrix} -9x_1^2 + 2x_1x_2 + 2x_2^2 & x_1^2 + 4x_1x_2 - 3x_2^2 \\ 6x_1^2 - 20x_1x_2 + 10x_2^2 & -10x_1^2 + 20x_1x_2 - 9x_2^2 \end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} \\ &= \begin{bmatrix} -7x_1^2 + 10x_1x_2 - 4x_2^2 \\ -14x_1^2 - 20x_1x_2 - 8x_2^2 \end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix}(-7x_1^2 + 10x_1x_2 - 4x_2^2) \\ C(t)\begin{bmatrix}1\\2\end{bmatrix} &= [2 \quad -1]\begin{bmatrix}1\\2\end{bmatrix} = 0 \end{aligned}$$

so

$$\frac{dq}{dt}(t) = -2q(t)[-7x_1(t)^2 + 10x_1(t)x_2(t) - 4x_2(t)^2], \qquad q(t_f) = 0$$

and $q(t) = 0$ for all t. In particular, $\bar{Q}$ and $\bar{P}\bar{Q}$ are singular and one state can be truncated without affecting the input-output relationship. By simulation of the differential equations (5.2) and (5.3) one obtains $\bar{P}$ and $\bar{Q}$. For a certain training input-signal $u_0(t)$ the average gramians are computed to

$$\bar{P} = \begin{bmatrix} 0.0018 & -0.0145 \\ -0.0145 & 0.2936 \end{bmatrix} \qquad \bar{Q} = \begin{bmatrix} 1.3058 & -0.6529 \\ -0.6529 & 0.3264 \end{bmatrix}$$

and the matrix $\bar{Q}$ is, as expected, singular. The corresponding coordinate change $Tz = x$ is determined so that $T^{-1}\bar{P}\bar{Q}T$ is diagonal with decreasing order, in this case

$$T^{-1}\bar{P}\bar{Q}T = \begin{bmatrix} 0.1171 & 0 \\ 0 & 0 \end{bmatrix}.$$

Truncation of z_2 and substitution according to (5.4) yields the nonlinear system

$$\begin{aligned} \dot{z_1} &= -1.23z_1^3 - 0.901u \\ y &= -1.11z_1 \end{aligned}$$

which is equivalent to $\dot{y} = -y^3 + u$ □

EXAMPLE 5.3—A SEVEN-STATE SYSTEM

The procedure can be applied to larger examples and also when lossless truncation is not possible. Consider the seven-state system

$$\begin{aligned} \dot{x}_1 &= -x_1^3 + u \\ \dot{x}_2 &= -x_2^3 - x_1^2x_2 + 3x_1x_2^2 - u \\ \dot{x}_3 &= -x_3^3 + x_5 + u \\ \dot{x}_4 &= -x_4^3 + x_1 - x_2 + x_3 + 2u \\ \dot{x}_5 &= x_1x_2x_3 - x_5^3 + u \\ \dot{x}_6 &= x_5 - x_6^3 - x_5^3 + 2u \\ \dot{x}_7 &= -2x_6^3 + 2x_5 - x_7 - x_5^3 + 4u \\ y &= x_1 - x_2^2 + x_3 + x_4x_3 + x_5 - 2x_6 + 2x_7 \end{aligned}$$

Following the described procedure, the system is linearized along a simulated training trajectory generated by a a 10Hz square wave signal with amplitude one. Further, the gramians are calculated according to Equation (5.2) and (5.3). Again, the balancing coordinate change T is computed so that $T^{-1}\bar{P}\bar{Q}T$ is diagonal with decreasing order. The diagonal elements computed for this system and choice of training trajectory are shown in Figure 5.1. The size of these diagonal elements indicate the importance of the new states for the input-output relationship. If the nonlinear system is truncated to one state, using the substitution in (5.4), the reduced system becomes

$$\begin{aligned} \dot{z}_1 &= -0.492z_1 - 0.0879z_1^3 + 5.08u \\ \tilde{y} &= 1.34z_1 + 0.0792z_1^2 \end{aligned}$$

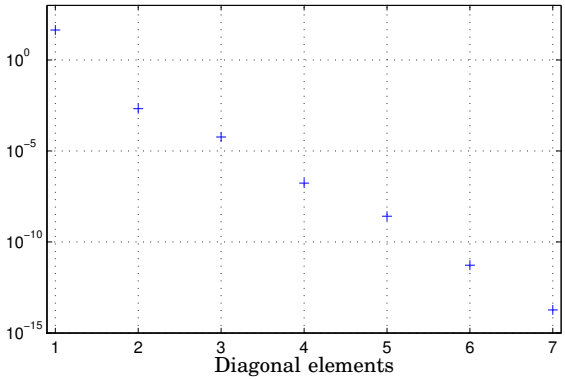

Figure 5.1 Diagonal elements for the system in Example 5.3

A comparison achieved by simulating the original and reduced system with the same input signal $u(t)$ can be seen in Figure 5.2. The system lack physical interpretation and the input-signal used for comparison can therefore be chosen arbitrarily. In this case it has been chosen to be the sum of a sinusoidal and a square-wave signal. □

EXAMPLE 5.4—MASS-SPRING-DAMPER SYSTEM

In this example, the method is applied to a two-dimensional mass-spring-damper system. Figure 5.3 shows six masses connected with springs and dampers. The input-signal is an external force, in horizontal and vertical direction, on the leftmost mass. The output-signal is the position coordinates of the top middle mass.

A thin line in the figure represents a linear spring-damper with an unforced length l_0 according to the figure. The masses, except the two rightmost ones, are also connected to the ground with linear spring-dampers.

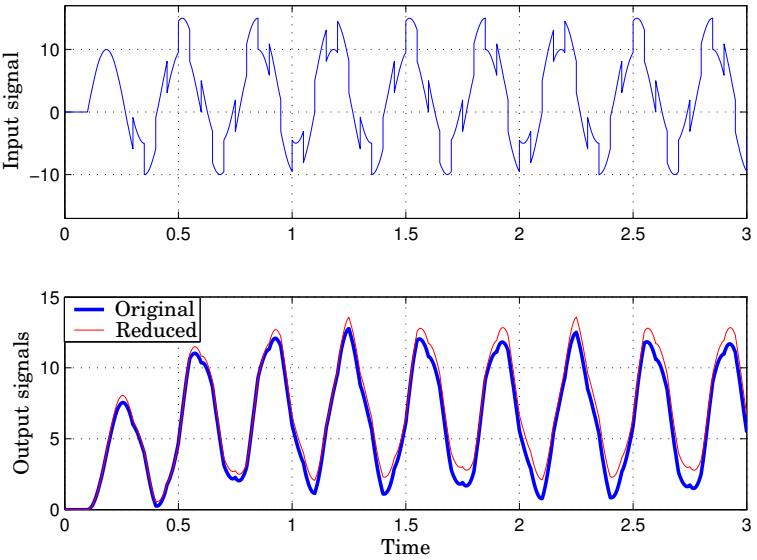

Figure 5.2 Simulation results for the system in example 5.3

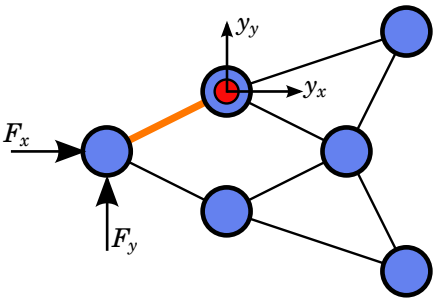

Figure 5.3 Mass-spring-damper system in Example 5.4. Force on the left mass as input signals and position of the marked mass as output.

The motion equations for each mass consist of four differential equations

$$\dot{p}_x = v_x \qquad \dot{p}_y = v_y$$
$$\dot{v}_x = \frac{1}{M}\sum_i F_{x,i} \qquad \dot{v}_y = \frac{1}{M}\sum_i F_{y,i}$$

where p_x and p_y are the position coordinates with the corresponding velocities v_x and v_y. The mass is denoted M and the forces $F_{x,i}$ and $F_{y,i}$ are the forces in horizontal and vertical direction inflicted by spring-damper i

$$F_{x,i} = \left(K(l_i - l_{0i}) + D\frac{d}{dt}(l_i)\right)\cos\theta_i$$
$$F_{y,i} = \left(K(l_i - l_{0i}) + D\frac{d}{dt}(l_i)\right)\sin\theta_i$$

Here l_i is the length of spring-damper i, D the damping coefficient and K the spring coefficient. In this example all coefficients have been set to one, $M = K = D = 1$. Further, the angle θ_i is the angle of the spring-damper. Here, only small angle perturbations are considered and θ_i is therefore assumed to be constant.

The thick line is a nonlinear damper that gives a force proportional to the deformation rate to the power of three,

$$F_{x,i} = D\left(\frac{d}{dt}(l_i)\right)^3\cos\theta_i \qquad F_{y,i} = D\left(\frac{d}{dt}(l_i)\right)^3\sin\theta_i$$

Linearization of the model around any stationary point would neglect this nonlinear damper, it only affects the linearization during transient behaviour. In the case of the leftmost mass, the external forces also contribute to the equations.

The model has four states per mass, yielding a total of 24 states, and can be written on the form

$$\dot{x} = f(x,u)$$
$$y = h(x,u)$$

In this example the described method is compared to the Proper Orthogonal Decomposition method as defined in Section 2.3. Reduction to 8 states is performed with both methods using the same training trajectory. A simulation result can be seen in Figure 5.4. A better result is obtained with the described method, which partly is due to the fact that the Proper Orthogonal Decomposition method does not take the output function into consideration.

□

5.2 Summary

A method for simplification of nonlinear input-output models has been outlined. The given procedure is focused on reducing the number of states using information obtained by linearization around trajectories.

The number of states is one factor contributing to simulation time and in the presented examples computation time has been reduced. Although, in the procedure many short right-hand-side functions can be replaced by fewer more complicated ones. Therefore, simulation time does not necessarily diminish in the general case.

No proofs concerning preserved stability or error bounds are presented. However, the methodology is closely tied to existing theory on error bounds and good results are shown in form of examples and simulation data.

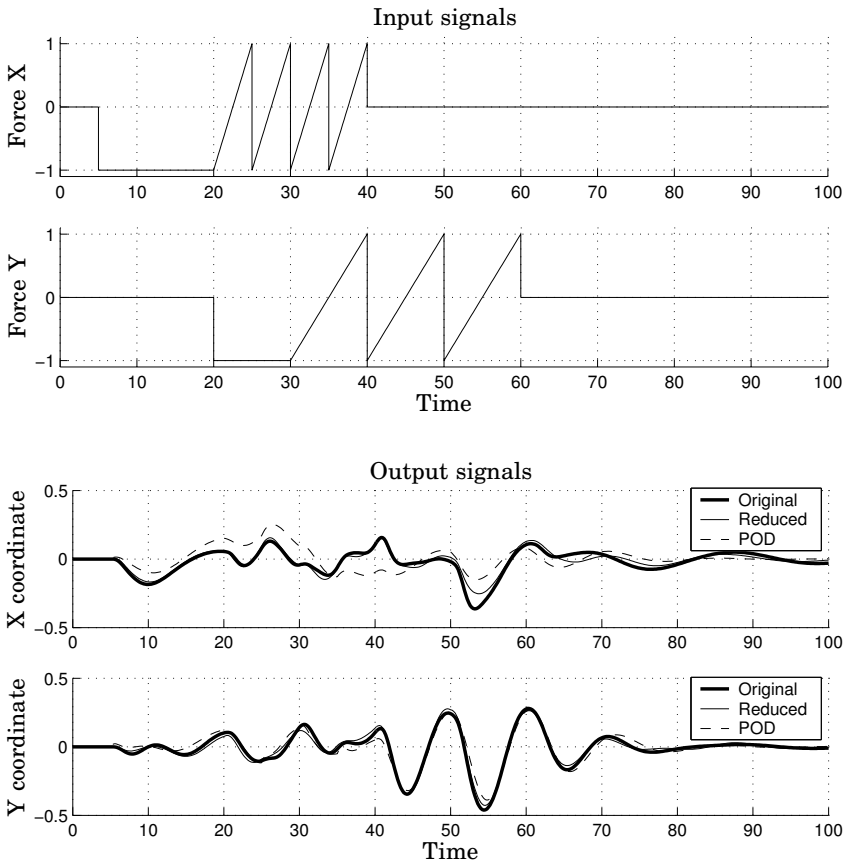

Figure 5.4 Simulation results for example 5.4. Method comparison for model reduction from 24 to 8 states.

6

Conclusions and future work

The current control design development process in automotive industry involves many expensive experiments and hand-tuning of control parameters. Model based control design is a promising approach to reduce costs and development time, in this process low complexity models are essential. This thesis combines the areas of modeling and model reduction of automotive systems.

A model of the exhaust gas oxygen sensor has been developed. The end result is a simple and static model that is sensitive to non-equilibrium concentrations of H_2. The model has been implemented in the Modelica language and successfully been validated with test gas experiment data. By adapting parameters within reasonable physical limits, higher fidelity to experiment measurements was achieved. The mean error in sensor output voltage did not exceed 3% of the maximum output, when the model was compared to experiment data.

Also, a model reduction method comparison has been conducted on an engine air path model. The heuristic method commonly used when modeling engine dynamics is compared with a more systematic method based on balanced truncation. Both methodologies have their advantages and disadvantages. If a detailed model is available and linear behaviour is expected then the balanced truncation methodology could be preferred. This technique may require a large computation time but needs very little manual attention. Using the heuristic method requires more experience and knowledge, it may also involve exten-

sive parameter fitting, but renders more insight to the simplifications made.

Finally, a method for model reduction of nonlinear systems has been presented. The given procedure is focused on reducing the number of states using information obtained by linearization around trajectories. No proofs concerning preserved stability or error bounds are presented. However, the methodology is closely tied to existing theory on error bounds and good results are shown in form of examples and simulation data.

6.1 Future work

The continued work of this thesis will be in the area of model reduction of nonlinear systems. So far, a method for order reduction has been developed but an attractive method should also diminish simulation time and the number of model parameters. Further, nonlinear model reduction of more realistic examples has to be considered.

7

Bibliography

Antoulas, A. and D. Sorensen (2001): "Approximation of large-scale dynamical systems: An overview." *International Journal of Applied Mathematics and Computational Science*, **11**, pp. 1093–1121.

Astrid, P. (2004): *Reduction of process simulation models: a proper orthogonal decomposition approach*. PhD thesis, Technische Universiteit Eindhoven, Netherlands.

Auckenthaler, T. S., C. H. Onder, and H. P. Geering (2002): "Modelling of a solid-electrolyte oxygen sensor." SAE Technical Paper 2002-01-1293.

Barrick, G., D. Calhoun, X. Chen, B. von Dohlen, X. Huang, T. Leise, H. Lomeli, R. Michler, S. Othman, P. Worfolk, and D. Baker (1996): "A mathematical model for an exhaust oxygen sensor." Technical Report IMA Pre-print Series #1422-6. Institute for Mathematics & its Applications, University of Minnesota – Minneapolis, Minnesota 55455.

Cantera (2006): "Object-Oriented Software for Reacting Flows." Home page: http://cantera.sourceforge.net.

Chevalier, A., C. Vigild, and E. Hendricks (2000): "Predicting the port air mass flow of SI engine in Air/Fuel ratio control applications." Number 2000-01-0260. SAE Technical Paper.

Dynasim AB (2006): "Dymola-Dynamic Modeling Laboratory, Users Manual." http://www.dynasim.se.

Fleming, W. J. (1977): "Physical principles governing nonideal behavior of the zirconia oxygen sensor." *Journal of the Electrochemical Society*, **124:1**, pp. 21–28.

Føns, M., C. Vigild, A. Chevalier, E. Hendricks, S. Sorenson, and M. Müller (1999): "Mean value engine modelling of an SI engine with EGR." Number 1999-01-0909. SAE Technical Paper.

Fritzson, P. (2004): *Principles of object-oriented modeling and simulation with Modelica 2.1*. Wiley-IEEE Press.

Fujimoto, K. and D. Tsubakino (2006): "On computation of nonlinear balanced realization and model reduction." In *Proceedings of American Control Conference*, pp. 460–465. IEEE, Minneapolis, USA.

Grimme, E. J. (1997): *Krylov Projection Methods for Model Reduction*. PhD thesis, University of Illinois at Urbana-Champaign.

Hendricks, E., A. Chevalier, M. Jensen, S. Sorenson, D. Trumpy, and J. Asik (1996): "Modelling of the intake manifold filling dynamics." Number 960037. SAE Technical Paper.

Heywood, J. (1988): *Internal Combustion Engine Fundamentals*. McGraw-Hill, New York.

Karhunen, K. (1946): "Zur spektraltheorie stochastischer prozesse." *Ann. Acad. Sci. Fennicae*, **Ser.A1:34**.

Khalil, H. K. (2002): *Nonlinear systems*, third edition. Prentice Hall, Upper Saddle River, New Jersey.

Lall, S. and C. Beck (2003): "Error-bounds for balanced model-reduction of linear time-varying systems." *IEEE Transactions on Automatic Control*, **48:6**, pp. 946–956.

Lall, S., J. Marsden, and S. Glavaski (2002): "A subspace approach to balanced truncation for model reduction of nonlinear control systems." *International Journal of Robust and Nonlinear Control*, **12**, pp. 519–535.

Li, J.-R. (2000): *Model Reduction of Large Linear Systems via Low Rank System Grammians*. PhD thesis, Massachusetts Institute of Technology.

Liu, Z. and J. Wagner (2002): "Nonlinear model reduction for dynamic and automotive system descriptions." *Journal of Dynamic Systems, Measurement, and Control*, **124:4**, pp. 637–647.

Loève, M. (1945): "Fonctions aléatoires de second ordre." *Comptes Rendus Acad. Sci. Paris*, pp. 220–469.

Lumley, J. (1967): "The structure of inhomogeneous turbulence." *Atmospheric turbulence and wave propagation*, **4**, pp. 166–178.

Moore, B. (1981): "Principal component analysis in linear systems: controllability, observability, and model reduction." *IEEE Transactions on Automatic Control*, **26:1**, pp. 17–32.

Newman, A. J. and P. S. Krishnaprasad (1998): "Computation for nonlinear balancing." In *Proceedings of the 37th IEEE Conference on Decision & Control*, vol. 4, pp. 4103–4104. IEEE, Tampa, Florida, USA.

Nilsson, O., A. Rantzer, and J. Chauvin (2006): "A model reduction case study: Automotive engine air path." In *Proceedings of the IEEE International Conference on Control Applications*. Munich, Germany.

Reid, R., J. Prausnitz, and T. Sherwood (1977): *The Properties of Gases and Liquids*. McGraw-Hill, New York.

Rewieński, M. J. (2003): *A Trajectory Piecewise-Linear Approach to Model Order Reduction of Nonlinear Dynamical Systems*. PhD thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Saji, K., H. Kondo, T. Takeuchi, and I. Igarashi (1988): "Voltage step characteristics of oxygen concentration cell sensors for nonequilibrium gas mixtures." *Journal of the Electrochemical Society: Electrochemical Science and Technology*, **135:7**, pp. 1686–1691.

Sandberg, H. (2006): "A case study in model reduction of linear time-varying systems." *Automatica*, **42:3**, pp. 467–472.

Sandberg, H. and A. Rantzer (2004): "Balanced truncation of linear time-varying systems." *IEEE Transactions on Automatic Control*, **49:2**, pp. 217–229.

Scherpen, J. (1993): "Balancing for nonlinear systems." *Systems and Control Letters*, **21:2**, pp. 143–153.

Shokoohi, S., L. Silverman, and P. van Dooren (1983): "Linear time-variable systems: Balancing and model reduction." *IEEE Transactions on Automatic Control*, **28:8**, pp. 810–822.

Smith, J. M. (1981): *Chemical Engineering Kinetics*. McGraw-Hill, New York.

Stykel, T. (2004): "Gramian based model reduction for descriptor systems." *Mathematics of Control, Signals and Systems*, **16**, pp. 297–319.

Vasilyev, D., M. J. Rewieński, and J. White (2006): "Macromodel generation for biomems components using a stabilized balanced truncation plus trajectory piecewise-linear approach." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, **25:2**.

Verriest, E. I. and T. Kailath (1983): "On generalized balanced realizations." *IEEE Transactions on Automatic Control*, **28:8**, pp. 833–844.

Department of Automatic Control
Lund University
Box 118
SE-221 00 Lund Sweden

Document name
LICENTIATE THESIS

Date of issue
December 2006

Document Number
ISRN LUTFD2/TFRT--3242--SE

Author(s)
Oskar Nilsson

Supervisor
Anders Rantzer
Rolf Johansson

Sponsoring organisation
HYCON, Toyota Motor Corporation

Title and subtitle
Modeling and model reduction in automotive systems

Abstract

The current control design development process in automotive industry and elsewhere involves many expensive experiments and hand-tuning of control parameters. Model based control design is a promising approach to reduce costs and development time. In this process low complexity models are essential.

This thesis combines the areas of modeling and model reduction in automotive systems. A model of the exhaust gas oxygen sensor, used for air-fuel ratio control in automotive spark ignition engines, is developed and successfully validated.

A model reduction case study is also performed on an engine air path. The heuristic method commonly used when modeling engine dynamics is compared with a more systematic approach based on the balanced truncation method.

Finally, a method for model reduction of nonlinear systems has been derived. The procedure is focused on reducing the number of states using information obtained by linearization around trajectories. The methodology is closely tied to existing theory on error bounds and good results are shown in form of examples and simulation data.

Key words
Modeling, Model reduction, Automotive systems, Lambda sensor, Engine air path

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280–5316

ISBN

Language
English

Number of pages
76

Recipient's notes

Security classification