

Linköping Studies in Science and Technology
Thesis No. 1390

Observer Design and Model Augmentation for Bias Compensation with Engine Applications

Erik Höckerdal



Linköpings universitet
INSTITUTE OF TECHNOLOGY

Department of Electrical Engineering
Linköpings universitet, SE-581 83 Linköping, Sweden

Linköping 2008

**Observer Design and Model Augmentation for Bias Compensation
with Engine Applications**

© 2008 Erik Höckerdal

hockerdal@isy.liu.se
<http://www.vehicular.isy.liu.se>
Department of Electrical Engineering,
Linköpings universitet,
SE-581 83 Linköping,
Sweden.

ISBN 978-91-7393-734-4

ISSN 0280-7971

LIU-TEK-LIC-2008:48

Printed by LiU-Tryck, Linköping, Sweden 2008

Abstract

Control and diagnosis of complex systems demand accurate knowledge of certain quantities to be able to control the system efficiently and also to detect small errors. Physical sensors are expensive and some quantities are hard or even impossible to measure with physical sensors. This has made model-based estimation an attractive alternative.

Model-based estimators are sensitive to errors in the model and since the model complexity needs to be kept low, the accuracy of the models becomes limited. Further, modeling is hard and time consuming and it is desirable to design robust estimators based on existing models. An experimental investigation shows that the model deficiencies in engine applications often are stationary errors while the dynamics of the engine is well described by the model equations. This together with fairly frequent appearance of sensor offsets have led to a demand for systematic ways of handling stationary errors, also called bias, in both models and sensors.

In the thesis systematic design methods for reducing bias in estimators are developed. The methods utilize a default model and measurement data. In the first method, a low order description of the model deficiencies is estimated from the default model and measurement data, resulting in an automatic model augmentation. The idea is then to use the augmented model for estimator design, yielding reduced stationary estimation errors compared to an estimator based on the default model. Three main results are: a characterization of possible model augmentations from observability perspectives, an analysis of what augmentations that are possible to estimate from measurement data, and a robustness analysis with respect to noise and model uncertainty.

An important step is how the bias is modeled, and two ways of describing the bias are introduced. The first is a random walk and the second is a parameterization of the bias. The latter can be viewed as an extension of the first and utilizes a parameterized function that describes the bias as a function of the operating point of the system. The parameters, rather than the bias, are now modeled as random walks, which eliminates the trade-off between noise suppression in the parameter convergence and rapid change of the offset in transients. This is achieved by storing information about the bias in different operating points. A direct application for the parameterized bias is the adaptation algorithms that are commonly used in engine control systems.

The methods are applied to measurement data from a heavy duty diesel engine. A first order model augmentation is found for a third order model and by modeling the bias as a random walk, an estimation error reduction of 50 % is achieved for a European transient cycle. By instead letting a parameterized function describe the bias, simulation results indicate similar, or better, improvements and increased robustness.

Acknowledgments

First, I would like to thank my supervisors Dr. Erik Frisk and Dr. Lars Eriksson for their guidance and many interesting discussions. I would also like to thank Professor Lars Nielsen for letting me join his research group of Vehicular Systems at the Department of Electrical Engineering.

All colleagues at Vehicular systems also deserve a place in this acknowledgment for creating such a nice research atmosphere. Special thanks goes to Dr. Jan Åslund for his guidance in the theory of matrices. Dr. David Törnqvist at the division of automatic control also deserves a special thanks for the discussions about this works relation to SLAM.

I would also like to thank my colleagues at Scania for showing great interest in the project and for coming with valuable input. Special thanks goes to Mats Jennische, David Elfvik, Björn Völeker, and Erik Geijer Lundin for always taking time to discuss my thoughts and help collecting data.

This work has been supported by SCANIA CV AB and Swedish Governmental Agency for Innovation Systems VINNOVA through the research program GRÖNA BILEN 2.

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

Erik Höckerdal
Linköping, 2008

CONTENTS

1	Introduction	1
1.1	Problem Statement and Solution	3
1.2	System Overview	3
1.3	Contributions	5
1.4	Thesis Outline	5
1.5	Publications	6
2	Air Mass-flow Estimation in Heavy Duty Diesel Engines	7
2.1	Air Mass-flow Sensor Variations	8
2.1.1	Calibration Curve Evaluation	9
2.2	Estimators	10
2.2.1	Model Output	12
2.2.2	Air Mass-flow Sensor	12
2.2.3	Extended Kalman Filter - EKF	13
2.3	Quality Measures	14
2.3.1	Reference Signal Based Measures	14
2.3.2	Signal Noise Based Measure	15
2.4	Experimental Evaluation	16
2.4.1	Model Errors	16
2.4.2	Comparison of Estimators	17
2.5	Conclusions	19

3	Model Augmentation for Bias Compensation	21
3.1	Problem Formulation	21
3.1.1	Problem and Chapter Outline	23
3.2	Discretization and Linearization	23
3.3	Possible Augmentations	24
3.4	Augmentation Estimation	27
3.4.1	Bias Estimation	27
3.4.2	Augmentation Computation	29
3.4.3	Properties of the Estimated Augmentation	29
3.4.4	Evaluation of Augmentation Estimation Approaches	32
3.4.5	Method Summary	33
3.5	Experimental Evaluation	34
3.5.1	Evaluation Using Simulated Data	34
3.5.2	Two Experimental Evaluations	39
3.6	Conclusions	42
4	Parameterizing the Bias	45
4.1	Method Outline	45
4.2	Observability	47
4.2.1	Relations to SLAM	47
4.2.2	Growing Estimation Error Covariance	50
4.3	Noise Sensitivity	54
4.4	Experimental Evaluation	55
4.4.1	Observers	57
4.4.2	Convergence Speed	57
4.4.3	Adaptation Map Development	58
4.5	Conclusions	62
5	Conclusions and Future Work	65
5.1	Conclusions	65
5.2	Future work	66
	References	69
A	Plant Model	73
A.1	Basic Structure	73
A.2	Model Details	75
B	Experimental Set-up and Data	81
C	State Marginalization	83

INTRODUCTION

Stricter emission legislations and customer demands on low fuel consumption drive the technical development of engines and force new solutions to be introduced. On diesel engines are for example exhaust gas recirculation (EGR) and variable geometry turbo (VGT) systems introduced, see Figure 1.1. The technical development, with increased system complexity and tightened requirements from customers and legislators, increases the demands on the control and diagnosis systems. Two examples of important quantities that significantly affect the emissions of diesel engines are air to fuel ratio and EGR-fraction. As a consequence of the increased demands on the control and diagnosis system, the information quality of these quantities have to be increased, or new quantities that are hard, or even impossible, to measure with physical sensors have to be introduced. Furthermore, there is a desire to keep the costs down, that is, it is desirable to have as few, cheap, and reliable sensors in the system as possible. This has made estimation an important and active research area, see (Lino et al., 2008; García-Nieto et al., 2008; Andersson and Eriksson, 2004) for some examples.

To achieve cost-effective estimation with high accuracy, model based estimators are often used. This has driven the development of new models that are suitable for estimator design. These models have to be simple enough to be evaluated in real time, in for example an engine control unit, and at the same time describe the system behavior accurately enough for the estimation task. Development of models like this is a delicate balance between computational complexity of the model and how well it manages to describe the true system. Typically a large engineering effort is spent on modeling, which often is based on

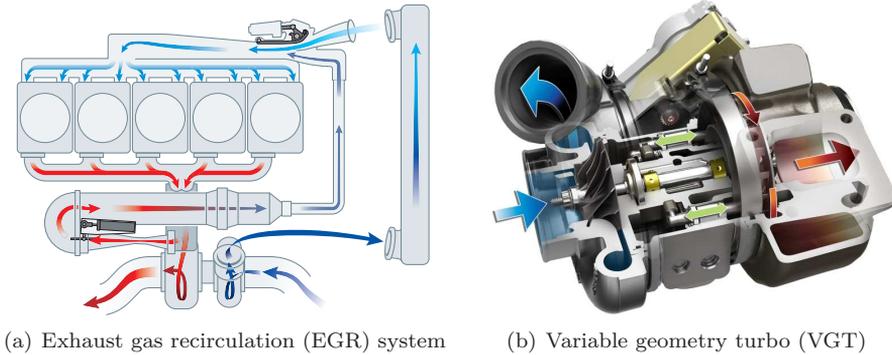


Figure 1.1: Technical solutions introduced on modern diesel engines to be able to fulfill the stricter emission legislations.

first law physics. In spite of the significant amount of work devoted to modeling there will always be errors in the model.

In all model-based control or diagnosis systems, the performance of the system is directly dependent on the accuracy of the model. In addition, as stated above, modeling is time consuming and even if much time is spent on physical modeling, there will always be errors in the model. This is especially true if there are constraints on the model complexity, as is the case in most real time applications. Another scenario is that a model developed for some purpose, for example control, exists but needs improvement before it can be used for other purposes, for example diagnosis. That is, there exist a lot of models, on which much modeling time is spent, that needs improvement before they can be used in an application. Hereafter these already available models will be called default models. There is a desire for a systematic method for improving these default models without involving further system or component modeling efforts.

In engine control and engine diagnosis it is crucial to have unbiased estimates. In model-based diagnosis, the true system is often monitored by comparing measured signals to estimated signals. If the magnitude of the difference, the residual, is above a certain limit a decision that something is wrong is made. In engine control, one objective is to maximize torque output while keeping the emissions below legislated levels and the fuel consumption as low as possible. Here, stationary estimation errors are crucial since fuel consumption and emissions often are in direct conflict with each other. If the stationary estimation errors can be reduced then the control system can balance closer to the emission limits without risking crossing them. For diesel engines this is especially hard since the control system does not normally have any feedback information from a λ - or NO_x -sensor and have to rely on estimated signals instead (Wang, 2008). In both cases, biased estimates impair the performance, and it is obvious that

there is a desire for methods that reduce the stationary model errors.

During the development of engines and engine control systems a lot of test are performed in engine test cells and measurement data are collected. This means that there are a lot of measurements from the engine and that it is fairly easy to gain new measurements.

1.1 Problem Statement and Solution

Based on the discussion above about model quality and the normally very good availability to measurement data from the real system. The objective of this thesis is to develop systematic methods for reducing stationary estimation errors when a default model and measurement data is given, this without involving further modeling efforts.

The starting point is a default model and measurement data from the true system. From this it can be determined if the model describes the system sufficiently well or if it has to be modified to be applicable to the current estimation application. In this thesis only modifications with respect to stationary estimation errors are considered.

If it is concluded that the model suffer from too large stationary errors and cannot be used in its current state. Then the methods for reducing stationary estimation errors developed in this thesis can be applied.

Basically, the ideas in the methods are to augment the default model with bias states to compensate for operating point dependent stationary errors. Then this augmented model can be used in any suitable estimator design to get an estimator that has reduced stationary errors compared to using the default model directly.

1.2 System Overview

This section serves as an overview of the system and the default model that are used for evaluation of the developed methods throughout this thesis. It also introduces the nomenclature used, and presents two important control quantities used in the control of diesel engines. Even though the methods developed in this thesis are not specially devoted to engine applications they are all applied and evaluated on the gas flow system of a Scania heavy duty diesel engine presented in Figure 1.2.

The default model used in the evaluations of the methods throughout this thesis is developed in (Wahlström and Eriksson, 2006), and presented in Appendix A. A schematic of the model is presented in Figure 1.3, where most of the modeled variables are presented. Control inputs to the model are injected amount of fuel u_δ , and EGR and VGT positions u_{egr} and u_{vgt} . Besides the control inputs, the engine speed n_e is used as a parameterization input, and

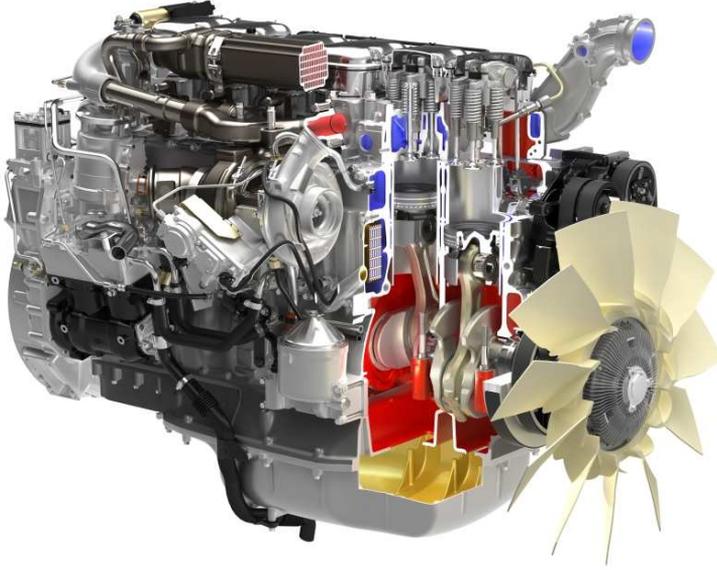


Figure 1.2: Illustration of the Scania inline six cylinder engine with VGT and EGR used for evaluation throughout the thesis.

thus the engine model can be expressed in state space form as

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

That the engine speed is used as an input to the model is due to the fact that the modeling is focused on the gas flows and does not include modeling of the produced torque. States are intake and exhaust manifold pressures p_{im} and p_{em} , and turbine speed n_{trb} . Also presented in Figure 1.3 are the, compressor mass-flow W_c , EGR mass-flow W_{egr} , mass-flow into the engine W_{ei} , mass-flow out of the engine W_{eo} , and turbine mass-flow W_t . Outputs from the model are the states, p_{im} , p_{em} , and n_{trb} , and the compressor mass-flow W_c .

Besides these variables the air to fuel ratio

$$\lambda = \frac{W_{air}}{W_{fuel} (A/F)_s}$$

and EGR-fraction

$$EGR_{frac} = \frac{W_{egr}}{W_{ei}},$$

are used, where W_{air} and W_{fuel} are the mass-flow of fresh air and fuel into the engine and $(A/F)_s$ is the stoichiometric mixture of air and fuel. The reasons for using these are their impact on the emissions and that they both are dependent on the air mass-flow into the cylinders.

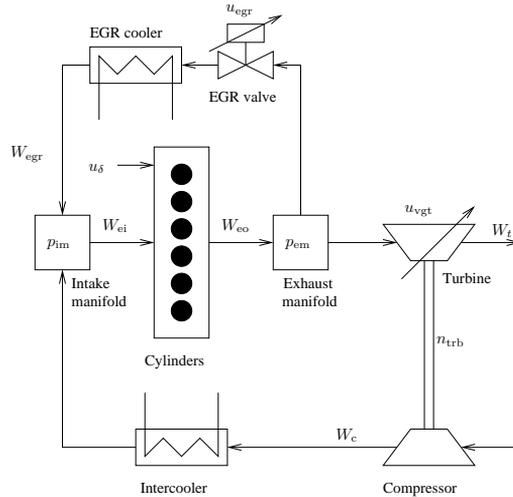


Figure 1.3: The model structure of the diesel engine. It has three control inputs, and three states related to the engine. In addition, it has one external parameterization input, n_e .

1.3 Contributions

The main contributions in the thesis are:

- An empirical analysis of model and sensor errors in heavy duty diesel engines.
- The methods for estimating a bias reducing model augmentation using a default model and measurements from the true system, summarized in Section 3.4.5.
- A parametrization of all model augmentations that maintains system observability in Theorem 3.2.
- The parametrization of all model augmentations that are possible to find with the proposed estimation algorithms in Theorem 3.3.
- A new algorithm for engine map adaptation with variable parameter update rate, Chapter 4.

1.4 Thesis Outline

The thesis is organized as follows. Chapter 2 is based on (Höckerdal et al., 2008a) and it describes an important estimation problem from the automotive industry. This particular example is used to analyze how the quality of a sensor

can be improved as well as how the quality can be assessed. This example also illustrates the effect a model with stationary errors has on the estimates when used in estimator design.

Chapter 3 is based on (Höckerdal et al., 2008b) and (Höckerdal et al., 2008c) and it presents a systematic method for bias reduction in model based estimator design. The method applies the idea of introducing extra states, modeled as random walk processes, for estimating the stationary errors. An automatized method for estimating appropriate augmentations from measurement data is developed.

In engine applications the bias is typically operating point dependent. If such biases are modeled as random walks the information about the bias in each operating point is discarded as soon as the system changes operating point. Chapter 4 addresses this problem by modeling the bias as function of known states and/or inputs instead of as a random walk. By introducing unknown parameters in the function and apply joint parameter and state estimation, the observer becomes adaptive and can handle operating point dependent biases that change over time.

Chapter 5 contains the main conclusions and a discussion of possible future work.

1.5 Publications

The thesis is based on the following publications where the author contributed with the majority of the work.

- Erik Höckerdal, Lars Eriksson, and Erik Frisk. Air Mass-Flow Measurement and Estimation in Diesel Engines Equipped with EGR and VGT. *SAE World Congress*, Detroit, 2008.
- Erik Höckerdal, Erik Frisk, and Lars Eriksson. Observer Design and Model Augmentation for Bias Compensation Applied to an Engine. *IFAC World Congress*, Seoul, 2008.
- Erik Höckerdal, Erik Frisk, and Lars Eriksson. Observer Design and Model Augmentation for Bias Compensation with a Truck Engine Application. *Control Engineering Practice*, 2008, <http://dx.doi.org/10.1016/j.conengprac.2008.09.004>.

AIR MASS-FLOW ESTIMATION IN HEAVY DUTY DIESEL ENGINES

A central quantity used in the engine control systems is the air mass-flow into the intake system. It is used for many purposes and influences both the engine performance and emissions and it is essential to have an air mass-flow measurement of good quality. This sensor signal, on a diesel engine with EGR and VGT, is studied in detail in this chapter and there are two issues that are addressed. The first deals with variations in the sensor characteristics, i.e., how accurate is the air mass-flow sensor, while the second studies how the quality of a sensor can be improved, by for example an estimator, and also how the quality can be assessed.

Air Mass-flow Sensor Variations

The first problem that is addressed is the air mass-flow sensor quality. One important issue is the sensor's long term stability and variation. Two questions are asked: how does the sensor characteristic vary with time, and how does it vary between engine configurations?

To answer these questions, systematic engine test cell measurements have been conducted on a limited range of air mass-flow sensors over the span of several weeks. A central piece of information is a calibration curve that has been recorded and stored for all days and all tests. The data is stored and analyzed with respect to day-to-day variations, aging, changes between configurations etc, and all these changes are quantified using experimental data.

Methods for Measuring and Improving Sensor Signals

The second issue that is addressed concerns methods for improving sensor signals, for example by applying an estimator. Due to factors like system aging and different operating conditions caused by geographical location, for example pressure, temperature, humidity etc. of the surrounding air, the air mass-flow sensor signal is in need of continuous adaptation. Several approaches to cope with this adaptation are investigated including ad hoc mapping and Kalman filtering. The investigation analyzes the effect model quality has on the estimates from a model based estimator. The quality of a physical sensor is determined by, for example accuracy, drift and aging, while the determination of the quality of an estimator is a more subtle task. An estimator is the result of a design work and hence connected to factors like application, model, control error and robustness.

The air mass-flow is used to estimate EGR-fraction and λ , two different applications that, in some sense, demand different quality properties of the signal. Therefore several quality measures are presented in Section 2.3, and used in the evaluation of the methods used for improving air mass-flow sensor signal. For example, if the signal is used to estimate the amount of EGR fed through the engine it is suitable to use an absolute measure, but if it is used to estimate a concentration, for example λ , it might be better to choose a relative measure. In some applications a bias is not crucial and it is better to use a signal noise based measure.

2.1 Air Mass-flow Sensor Variations

The air mass-flow signal is needed for computations of λ and EGR-fraction. Both are important quantities that greatly affects the emissions. In a diesel engine λ , defined as

$$\lambda = \frac{W_{\text{air}}}{W_{\text{fuel}}(A/F)_s},$$

can not be allowed below a certain limit, $\lambda_{\text{smoke lim}}$, to avoid generating smoke. Normally in a diesel engine, when λ is greater than $\lambda_{\text{smoke lim}}$, W_{fuel} is decided by the desired torque. However when the desired torque forces λ toward $\lambda_{\text{smoke lim}}$, and when λ reaches $\lambda_{\text{smoke lim}}$, the control law enters a mode where W_{fuel} is proportional to \hat{W}_{air} , (Wahlström, 2006). This is particularly important during transients where the torque demand is high, i.e., during acceleration. In these cases, an error in the air mass-flow signal results in either creation of smoke or reduced torque output. The other important quantity, the EGR-fraction

$$EGR_{\text{frac}} = \frac{W_{\text{tot}} - W_{\text{air}}}{W_{\text{tot}}},$$

where W_{tot} is the total gas mass-flow into the engine and W_{air} is the air mass-flow into the engine. The EGR-fraction is used to control the NO_x emissions.

Both λ and EGR-fraction are important for the emissions and they rely heavily on the air mass-flow W_{air} . Hence, it is important to have a high quality measurement or estimation of the air mass-flow. Note that the EGR-fraction also depends on W_{tot} , which is computed using the volumetric efficiency of the engine and is, by empirical experience, considered to be accurate.

A calibration curve,

$$W_{\text{ref}} = (1 + r(W_{\text{raw}})) W_{\text{raw}},$$

is used for the air mass-flow sensor variation analysis. The calibration curve is found by comparing the production air mass-flow sensor W_{raw} to a reference W_{ref} from a long series of engine measurements. W_{ref} is a sensor mounted in the engine test cell for the purpose of accurately being able to measure the air mass-flow into the engine. The difference between W_{raw} and W_{ref} has been stored in a calibration curve that is implemented as a lookup-table consisting of 12 grid points. These calibration curves have been recorded over several weeks to enable the study of day-to-day variations. The calibration curve is computed according to

$$r(W_{\text{raw}}) = \frac{W_{\text{ref}}}{W_{\text{raw}}} - 1, \quad (2.1)$$

and has the typical appearance presented in Figure 2.1.

The reference sensor is mounted on a straight pipe in the test cell, where the air mass-flow over the cross section of the pipe is orthogonal to the sensor and cylindrically symmetric, and is considered to give very accurate measurements of the air mass-flow. More details concerning the reference sensor W_{ref} is given in Appendix B.

2.1.1 Calibration Curve Evaluation

Calibration curves from two diesel engines, one inline six cylinder and one V8, are gathered from test runs in an engine test cell. 13 calibration curves are collected over a total time of about two weeks for the six cylinder engine and 21 calibration curves over four weeks for the V8 engine. Figure 2.1 presents the typical appearance of a calibration curve, the upper for a 6 cylinder engine and the lower for a V8 engine. These calibration curves are used to analyze the quality of the air mass-flow sensor.

The difference between engine configurations can be seen by comparing the upper and lower plots in Figure 2.1 and Figure 2.2, where Figure 2.1 presents the day-to-day variations and Figure 2.2 presents the trend of four grid points in the calibration curve. These four grid points are spread out over the operating region of the air mass-flow into the engine. Figure 2.1 shows that the day-to-day variations are quite large, especially for the V8 engine where the standard deviation varies between 2–3 %-units. For the 6 cylinder engine the variations are smaller. Further, the difference between the minimum and maximum values for each parameter in the calibration curve varies between approximately 1,5–4 %-units for the inline six cylinder engine and 3–12 %-units for the V8 engine.

Another difference between the two engine configurations is the appearance of the calibration curve. For the 6 cylinder engine the line starts at approximately 5 %, has a slightly positive slope, and ends at approximately 10 %. For the V8 engine the line is quite different, it starts at about 1 %, varies quite a bit, and ends at -1 %.

The 6 cylinder engine data indicates that the air mass-flow into the engine has to be continuously monitored. The following small example gives a rough estimate of the consequence of an incorrect air mass-flow measurement.

Example 2.1

Assume that the engine control system controls the EGR-fraction to 30 % based on the air mass-flow sensor and that the air mass-flow sensor signal is incorrect and reads $W_{\text{air}}/1.1$, which is approximately the worst case according to Figure 2.1. That is,

$$EGR_{\text{frac}} = \frac{W_{\text{tot}} - W_{\text{air}}/1.1}{W_{\text{tot}}} = 30\%.$$

Then the true fresh air-fraction would become

$$(1 - 0.3)W_{\text{tot}} = \frac{W_{\text{air}}}{1.1} \Rightarrow W_{\text{air}} = 1.1(1 - 0.3)W_{\text{tot}} = 0.77W_{\text{tot}},$$

and thereby the true EGR-fraction 23 %, which would have a significant effect on the NO_x emissions (Heywood, 1988), that is the control system controls the engine to run with less EGR than needed to fulfill the legislated NO_x levels.

An analogous analysis can be made for λ close to $\lambda_{\text{smoke lim}}$ which further supports the statement that the an accurate estimate of the air mass-flow is important.

The large spread among the calibration curves for the V8 engine plot indicates that an ad hoc approach for compensating the sensor signal using only a calibration curve might not be enough. The quality also has to be improved in a way that the spread is reduced as well. These observations together with the importance of the estimates of λ and EGR-fractions necessitate an accurate estimate of the air mass-flow. This is one of the motives for the discussion in Sections 2.2 to 2.4.

As Figure 2.2 shows there are no obvious trends in the data over time. However, due to the relatively short time span, over which the data is collected, it is hard to draw any conclusions regarding long term aging of the air mass-flow sensors.

2.2 Estimators

To evaluate different methods for improving the air mass-flow sensor signal, four different estimators are designed. The aim with this is to address issues regarding what quality measure to use when designing and evaluating estimators, not

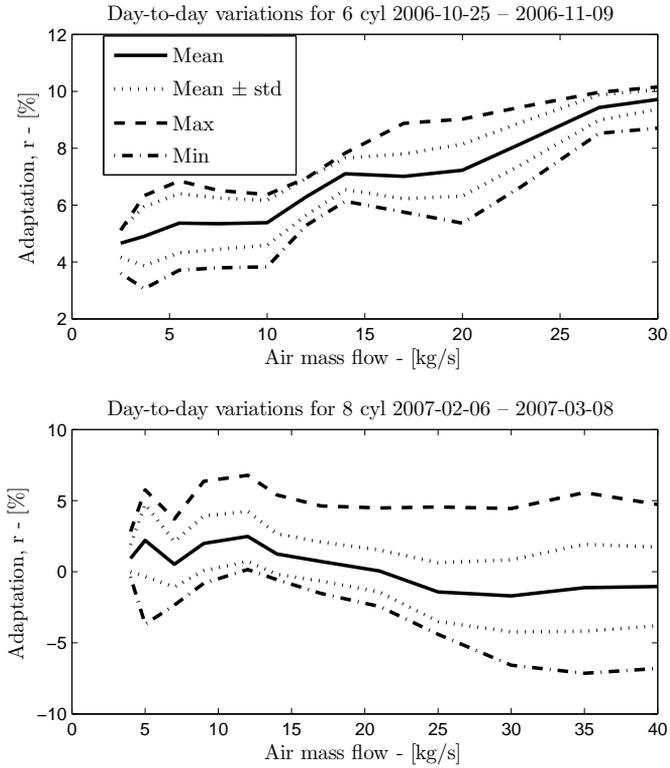


Figure 2.1: Min, max, mean, and the standard deviation over all collected calibration curves are presented for a 6 cylinder engine (upper plot) and a V8 engine (lower plot). It can be seen that the variation is quite large for both engine configurations, especially for the V8 engine.

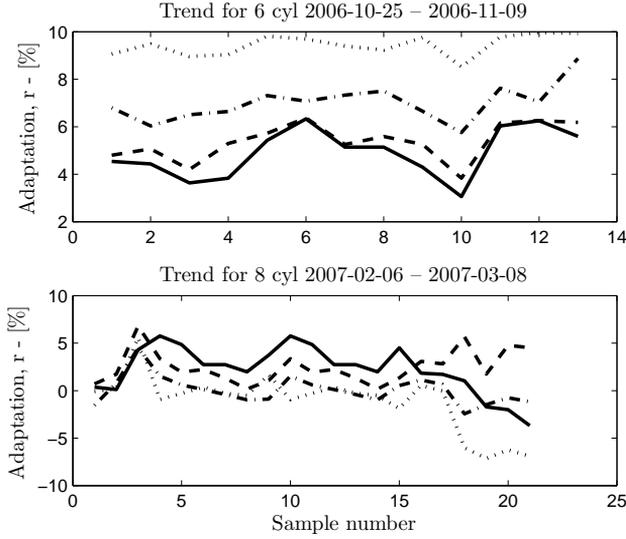


Figure 2.2: In this figure the trend of 4 grid points are presented for a 6 cylinder engine (upper plot) and a V8 engine (lower plot). It indicates that there is no particular trend in neither of the engine configurations.

to design the best estimator for any particular air mass-flow application, for example λ or EGR-fraction estimation.

2.2.1 Model Output

The first estimator is the model output of a forward Euler discretization of the model presented in Appendix A,

$$\hat{x}_{t+1} = x_t + T_s f(\hat{x}_t, u_t)$$

$$\hat{W}_{\text{model}, t} = (0 \ 0 \ 0 \ 1) h(\hat{x}_t),$$

where T_s is the sample time. It has the main purpose of enabling an analysis of the model quality. This estimator is driven by the control and parameterization inputs only, giving what would be called an open loop simulation of the system.

2.2.2 Air Mass-flow Sensor

Two estimators are based directly upon the measured air mass-flow. The first is the raw air mass-flow measurement, \hat{W}_{raw} . The second uses the calibration curve, r , to correct a filtered sensor signal, \hat{W}_{filt} , where \hat{W}_{filt} is \hat{W}_{raw} filtered with a low pass filter with a cut off frequency of 6.8 Hz. The cut of frequency is chosen to 6.8 Hz since this signal was available in the engine control unit. An

example of a calibration curve is presented in Figure 2.1 and the adapted air mass-flow is computed according to

$$\hat{W}_{\text{adapt}, t} = \hat{W}_{\text{filt}, t}(1 + r(\hat{W}_{\text{filt}, t})).$$

The idea of this estimator is to use the calibration curve from Section 2.1 to correct the signal from the air mass-flow sensor.

2.2.3 Extended Kalman Filter - EKF

A natural choice of estimator design method for a nonlinear system is the extended Kalman filter (EKF) (Kailath et al., 2000). The designed EKF is based on the model described in Appendix A and utilizes measurements from all model states, that is the intake and exhaust manifold pressures and the turbine speed for feedback. Further, the covariance matrices for the system and measurement noise, Q and R , are used as tuning parameters, and since the main objective is not to design an optimal estimator they are chosen in an ad hoc manner. The model used in the EKF design is the continuous time model,

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

presented in Appendix A. This model is then discretized with forward Euler and a sampling time of T_s seconds,

$$x_{t+1} = x_t + T_s f(x_t, u_t) \tag{2.2a}$$

$$y_t = h(x_t). \tag{2.2b}$$

Now a discrete time EKF is designed on Equation (2.2). The EKF equations for a time discrete model looks as follows, starting with the internal variables

$$S_t = H_t P_{t|t-1} H_t^T + R_t$$

$$K_t = P_{t|t-1} H_t^T S_t^{-1}$$

$$e_t = y_t - h(\hat{x}_{t|t-1}),$$

that is the innovation covariance, the Kalman gain, and the estimation error respectively. Continuing with the update equations

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t e_t$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} H_t^T S_t^{-1} H_t P_{t|t-1},$$

and the prediction equations

$$\hat{x}_{t+1|t} = \hat{x}_{t|t} + T_s f(\hat{x}_{t|t}, u_t)$$

$$P_{t+1|t} = (I - T_s A_t) P_{t|t} (I - T_s A_t)^T + Q_t$$

where

$$A_t = \left. \frac{\partial f}{\partial x} \right|_{x=\hat{x}_t|t}, \quad H_t = \left. \frac{\partial h}{\partial x} \right|_{x=\hat{x}_t|t-1}.$$

Finally the estimator can be written

$$\hat{W}_{\text{EKF},t} = h_W(\hat{x}_t),$$

where h_W is a nonlinear function of the states describing the air mass-flow through the compressor, see Appendix A.2.

2.3 Quality Measures

This section describes the different measures used for evaluating the quality of the estimated signal. Different applications of estimators require different properties. Therefore several different measures are used in the evaluation of the estimators. For example, if the estimator is used to estimate the amount of EGR fed through the engine it is suitable to use an absolute measure, while if it is used to estimate a concentration it might be better to choose a relative measure. In some applications, like for example if the estimated signal is differentiated and used as input to another estimator, estimation bias is not crucial. In such cases it may be better to use a signal noise based measure, for example by estimating the variance of the estimated signal.

The measures used are grouped into two groups, *reference signal based measures* and a *signal noise based measure*. In all equations below, $\hat{W}(t_i)$ is the estimated signal, $W_{\text{ref}}(t_i)$ is the measured reference signal and N is the number of samples.

2.3.1 Reference Signal Based Measures

Different measures have different properties, for example, a maximum norm can be used to capture robustness. Below the quality measures used in this chapter are presented.

Absolute Measures

An absolute measure is often preferred in diagnosis applications where the estimator is used to create residuals. The residual, together with a threshold, is then used to detect faults in the system. Here the following absolute measures are used.

- Mean absolute error

$$\varepsilon = \frac{1}{N} \sum_{i=1}^N \left| \hat{W}(t_i) - W_{\text{ref}}(t_i) \right|$$

- Maximum absolute error

$$\varepsilon = \max_{1 \leq i \leq N} \left| \hat{W}(t_i) - W_{\text{ref}}(t_i) \right|$$

- Root Mean Square Error (RMSE)

$$\varepsilon = \sqrt{\frac{1}{N} \sum_i \left(\hat{W}(t_i) - W_{\text{ref}}(t_i) \right)^2}$$

Relative Measures

Relative measures are preferred in applications where the absolute error varies with the level of the underlying signal. An example of an application like this is the air mass-flow where it is used to compute the air to fuel ratio, λ . The discussion in Section 2.1 highlighted the importance of a correct λ estimation, especially in transients where the torque demand is high. In these cases $W_{\text{fuel}} \propto \hat{W}_{\text{air}}$ and the error in λ , Δ_λ , becomes

$$\begin{aligned} \lambda &= \frac{W_{\text{air}}}{W_{\text{fuel}}} = \left[W_{\text{fuel}} \propto \hat{W}_{\text{air}} \right] \Rightarrow \\ \lambda &\propto \frac{\hat{W}_{\text{air}} + \Delta_W}{\hat{W}_{\text{air}}} \Rightarrow \\ \Delta_\lambda &\propto \frac{\Delta_W}{\hat{W}_{\text{air}}}, \end{aligned}$$

where Δ_W is the air mass-flow estimation error. In these situations the emissions are critical and it is important that the error in λ is small, hence a low relative measure is preferable. The following relative measures are used.

- Mean relative error

$$\varepsilon = \frac{1}{N} \sum_{i=1}^N \frac{\left| \hat{W}(t_i) - W_{\text{ref}}(t_i) \right|}{\left| W_{\text{ref}}(t_i) \right|}$$

- Maximum relative error

$$\varepsilon = \max_{1 \leq i \leq N} \frac{\left| \hat{W}(t_i) - W_{\text{ref}}(t_i) \right|}{\left| W_{\text{ref}}(t_i) \right|}$$

2.3.2 Signal Noise Based Measure

Some control applications require smooth input signals. In such cases, quality can be measured through signal noise.

As a measure of signal noise, the variance on a dynamic sequence is used. It is determined by estimating the variance of the difference between the original sequence and a non-causal low pass filtered version of the same, that is

$$\varepsilon = \text{var} \left(\hat{W}(t_i) - \bar{W}(t_i) \right)$$

$$\bar{W}(t_i) = H_{\text{LP}}(z)\hat{W}(t_i).$$

To remove the underlying trend, a cut-off frequency of 2 Hz is chosen for the filter since all system dynamics are slower than this. Because of this high pass filtering this measure does not capture any low frequency error.

2.4 Experimental Evaluation

The results from the comparison between the estimators are presented in this section using Figure 2.3 that compare the different estimators and Figure 2.4 that presents the results from the different measures. As previously stated, the main objective in this chapter is to examine the use of different quality measures when evaluating estimators and the influence of model errors in estimator design.

The experimental set-up and data used in the evaluation are presented in Appendix B. The measured outputs are the model outputs complemented with an extra air mass-flow sensor, W_{ref} , used as a reference.

2.4.1 Model Errors

All model based estimators are highly dependent on the accuracy of the model used, which is especially true if the assumptions made in the design method do not hold. If for example an EKF is used, the measurement and model error have to be described by zero mean white noise processes, i.e., biased measurements is not handled. The upper plot in Figure 2.3 presents the air mass-flow estimates from the estimators designed in Section 2.2 together with the air mass-flow measured with the reference sensor and the lower plot presents the corresponding estimation errors. One observation in Figure 2.3 is that \hat{W}_{model} does not follow W_{ref} well, \hat{W}_{model} has an obvious offset both for low and high air mass-flows but manages to capture the system dynamics. From this observation it is obvious that the model does not fully describe the engine and these model errors violate the assumptions made when utilizing the model to design an EKF, i.e., zero mean Gaussian system error. This will explain some of the results for the model based estimator, discussed in Section 2.4.2.

The model itself is not a well performing estimator which is clear also in Figure 2.4, where all reference signal based measures are higher than for the other estimators. That the reference signal based measures are higher for \hat{W}_{model} is of course a consequence of the stationary errors in the model. However, the variance measure is lower as a result of the absence of noisy feedback.

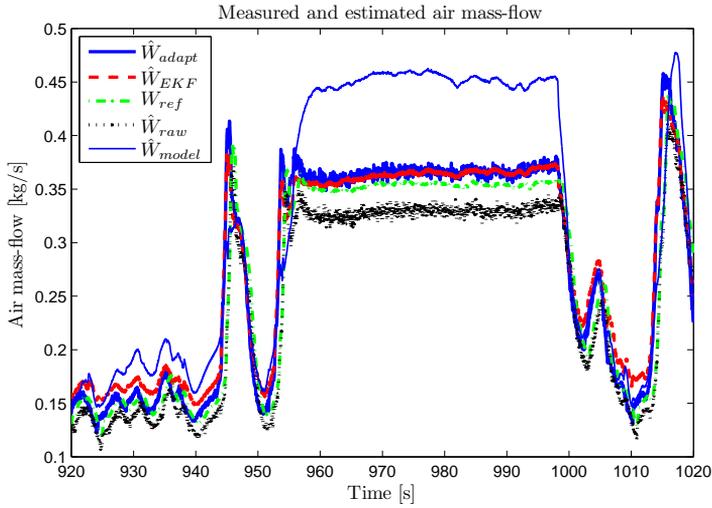
2.4.2 Comparison of Estimators

Again looking at Figure 2.3, it can be seen that \hat{W}_{raw} is poor, it has a negative bias which is larger for high air mass-flows than it is for low air mass-flows but still manages to capture the dynamics well. As far as it goes for \hat{W}_{adapt} , it can be noticed that by applying an correction based on the calibration curves determined in Section 2.1.1, it manages to remove most of the measurement bias in \hat{W}_{raw} . However the estimate is a bit high at the stationary part in the middle of the segment but besides that follows W_{ref} well. Further, \hat{W}_{EKF} is less noisy than \hat{W}_{raw} and \hat{W}_{adapt} , and performs well with respect to offset for high air mass-flows but not for low air mass-flows. The last observation in Figure 2.3, i.e., that the offset error differs for low and high air mass-flows, can be explained by a combination of model inaccuracies and feedback from the engine outputs. In this particular case a closer look at the estimation errors shows that the model and the raw sensor signal have different signs of the error for high flows and when they are combined in the EKF the result comes close to the reference. For low flows the raw signal is closer to the reference while the model still has a positive offset giving a positive error for the EKF. That is, an operating point dependent model error in the feedback loop, i.e., the EKF measurement update step, would in this case give a larger correction for high air mass-flows than it does for low air mass-flows.

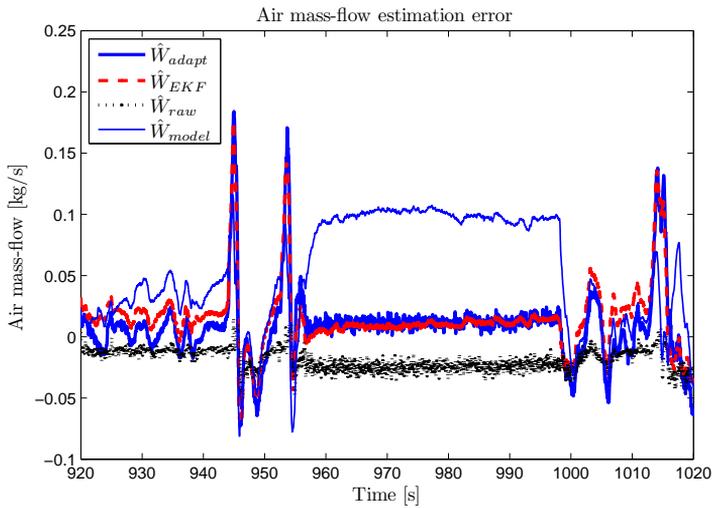
From Figure 2.3 it is also obvious that the estimators \hat{W}_{adapt} and \hat{W}_{raw} , based directly on the measured air mass-flow, have a significantly higher variance than \hat{W}_{model} and \hat{W}_{EKF} , which is even more evident in Figure 2.4.

The quality measures described in Section 2.3 are presented in Figure 2.4 for the different estimators, where the different measures have been normalized with respect to the \hat{W}_{adapt} measures. The figure not only verifies the observations made in Figure 2.3, but also further illuminates the differences between the different estimators with respect to the quality measures. Especially the maximum relative and absolute measures are presented that are hard to examine in a plot like Figure 2.3. These errors occur in transients, which is not surprising. However, the fact that these measures point out different estimators as the best is interesting. The reason for this is that the estimator responses differ in transients, i.e., \hat{W}_{adapt} utilizes a filtered version of \hat{W}_{raw} which introduces some time delay.

To conclude, the \hat{W}_{adapt} is the best estimator according to all reference signal based measures, except those using maximum norm, where \hat{W}_{raw} is the best with respect to the maximum relative error and \hat{W}_{EKF} is the best according to the maximum absolute error. With respect to noise \hat{W}_{model} is the best closely followed by \hat{W}_{EKF} .



(a) Air mass-flow estimates



(b) Air mass-flow estimation error

Figure 2.3: In the upper plot \hat{W}_{model} , \hat{W}_{EKF} , \hat{W}_{adapt} , \hat{W}_{raw} and W_{ref} are plotted for a 100 s segment of an ETC, and in the lower plot the corresponding estimation errors are plotted. It is obvious that the model has stationary errors and that \hat{W}_{adapt} follows W_{ref} well.

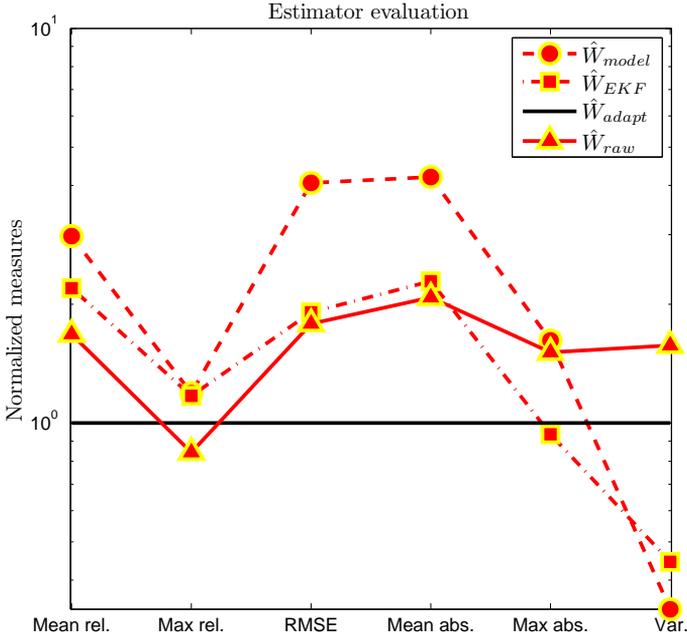


Figure 2.4: Presents the quality measures presented in Section 2.3 for the different estimators. The measures are normalized by division with \hat{W}_{adapt} .

2.5 Conclusions

An analysis of experimental air mass-flow sensor data has been performed. It is concluded that accuracy demands on the air mass-flow measurement necessitate continuous monitoring and adaptation of the state of the mass-flow characteristics. In particular an adaptive scheme should be used to properly account and compensate for the time variations in the mass-flow characteristics both due to operating point dependent sensor bias, and effects of aging and environmental conditions on the system.

Further, the model quality is of great importance when designing model based estimators. A model with stationary errors have the same effect on the estimates as biased measurements and a way to handle, these common model and measurement deficiencies, is desirable. When evaluating estimators designed for a particular application the choice of quality measure is central. The choice has to reflect the important properties in that particular application.

The objective of this thesis is to develop systematic approaches for handling stationary measurement and model deficiencies.

MODEL AUGMENTATION FOR BIAS COMPENSATION

The objective of this chapter is to develop a systematic method for reducing estimation bias in estimators without involving further modeling efforts. Chapter 2 considered an output estimation problem in a diesel engine while this chapter focuses on state reconstruction, however the method developed here applies also to output estimation.

The method utilizes an observable model and measurement data from the true system. The given model, referred to as the default model, and the measured inputs and outputs from the true system are used to estimate a suitable model augmentation. Then, the augmented model is used to design an observer that is shown to give estimates with reduced bias compared to an observer based on the default model. Three approaches for estimating a bias compensating augmentation are developed and evaluated with respect to measurement noise and model errors. Key results are a theoretical characterization of all possible augmentations from observability perspectives and a parametrization of the estimated augmentations. Finally the method is evaluated on a non-linear diesel engine model with experimental data from an engine test cell.

3.1 Problem Formulation

Chapter 2 revealed that designing an observer based on a model that captures dynamic behavior reasonably well but suffers from stationary errors results in biased estimates. How to reduce the bias in a systematic manner without involving further modeling efforts is the topic of this chapter.

The starting point is an existing model, referred to as the default model,

that is provided in state-space form

$$\dot{x} = f(x, u) \quad (3.1a)$$

$$y = h(x), \quad (3.1b)$$

where x is the state-vector, u the known control inputs, y the measurement vector, and f and h are non-linear functions.

The objective is to find a systematic way to design an observer that gives an unbiased estimate of either the complete state x or a function of the state $z = g(x)$. This should be done even though the default model is subject to significant bias errors. A direct approach to compensate for constant, or slowly varying, biases is to augment the default model with bias variables q as

$$\dot{x} = \tilde{f}(x, u, q) \quad (3.2a)$$

$$\dot{q} = 0 \quad (3.2b)$$

$$y = \tilde{h}(x, q), \quad (3.2c)$$

and design the observer using this augmented model. If the augmentation captures the true modeling errors and the augmented system is observable, the estimates will be unbiased. An obvious question is then how to introduce the bias variable q in the model equations. One way could be through process knowledge, which have been successfully applied in (Andersson and Eriksson, 2004; Tseng and Cheng, 1999). To automatize this an estimation procedure based on available measurement data is proposed.

Besides the natural restriction, that the augmented model (3.2) is observable, it is also desirable not to introduce more bias states than necessary. It is therefore desirable to find a bias vector q with as low dimension as possible that manages to reduce the bias. Another reason for finding a low-dimensional bias is that, since the model often is a first-principles physical model, bias in multiple states may be explained by one underlying bias affecting all these states. For example, bias in two modeled pressures can originate from a bias in the modeled mass flow between the two volumes or an incorrect modeling of energy conservation can give rise to bias in several states connected to the energy. Note that, the bias is necessarily not the same in the entire operating region of the system and may vary between operating points. This is part of the reason for introducing the bias as new states, rather than just as a fixed parameter, which allows the observer to have a tracking ability.

In model (3.1) there are two natural ways to introduce biases, in the dynamic equation (3.1a) or in the measurement equation (3.1b). In the truck engine application the sensors, intake and exhaust manifold pressures and turbine speed, are considered more reliable than the model presented in Appendix A and the bias augmentation is therefore introduced in the dynamic equations according

to

$$\dot{x} = f(x - A_q q, u) \quad (3.3a)$$

$$\dot{q} = 0 \quad (3.3b)$$

$$y = h(x), \quad (3.3c)$$

where a stationary point of the system is moved by $A_q q$. The matrix A_q is thus a description of how the underlying bias variable q influences the stationary value of the state variable x . The model (3.3) will be referred to as the augmented model. It is worth mentioning that although the focus is on biases in the dynamic equation, it is straightforward to modify the approach to also cover sensor biases.

3.1.1 Problem and Chapter Outline

Based on the discussion above, the problem studied in the sections to follow can now be stated as: Given an observable default model (3.1) and available measurement data, find a low order bias augmented model (3.3) and design an observer that estimates x with reduced bias compared to using the default model.

To solve the problems, some issues need to be addressed. First, which matrices A_q are possible at all? All are not possible since it is required that the augmented system is observable and a characterization of possible augmentations is derived in Section 3.3. Among these possible bias augmentations, which should be used? Section 3.4 describes three approaches for how to estimate a, for bias compensation, suitable low order A_q based on measurement data.

Section 3.5 presents two examples of the proposed estimator design methodology applied to a Scania diesel engine using simulated and real measurement data, respectively.

3.2 Discretization and Linearization

As a first step, the non-linear augmented model (3.3) is transformed into a linearized time discrete model. A reason for the discretization is the demand on the implementation, where sampled data with fixed sampling time is used. Here, a simple forward Euler discretization with step size T_s seconds is used. Note that observability does not depend on the choice of discretization method. For example, as long as T_s is chosen small enough the observability results are valid also for zero-order-hold (Kalman et al., 1963).

One objective of the chapter is to find a suitable A_q such that (3.3) is locally observable and to be able to use simple observability conditions. The observability analysis is here performed on a linearization of the non-linear model (3.3). Of course, non-linear observability is not guaranteed from observability of the linearization. Nevertheless, observability of a linearization in a stationary

point is a sufficient condition for local observability of the non-linear system, see Theorem 6.4 in (Lee and Markus, 1968). Even though observability is not strictly guaranteed, for example in transient mode when moving between operating points, the referred result gives theoretical support for using the linearized system in the observability analysis. Thus, when analyzing (3.3) the following model will be used

$$\begin{pmatrix} x_{t+1} \\ q_{t+1} \end{pmatrix} = \begin{pmatrix} I + T_s A & -T_s A A_q \\ 0 & I \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix} + \begin{pmatrix} T_s B \\ 0 \end{pmatrix} u_t \quad (3.4a)$$

$$y_t = \begin{pmatrix} C & 0 \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix}, \quad (3.4b)$$

where

$$A = \left. \frac{\partial f}{\partial x} \right|_{\substack{x=x_0 \\ u=u_0}}, \quad B = \left. \frac{\partial f}{\partial u} \right|_{\substack{x=x_0 \\ u=u_0}}, \quad \text{and} \quad C = \left. \frac{\partial h}{\partial x} \right|_{\substack{x=x_0 \\ u=u_0}}.$$

In the following, $I + T_s A$ is substituted for F to increase readability, and (3.4) now becomes

$$\begin{pmatrix} x_{t+1} \\ q_{t+1} \end{pmatrix} = \begin{pmatrix} F & -(F - I)A_q \\ 0 & I \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix} + \begin{pmatrix} T_s B \\ 0 \end{pmatrix} u_t \quad (3.5a)$$

$$y_t = \begin{pmatrix} C & 0 \end{pmatrix} \begin{pmatrix} x_t \\ q_t \end{pmatrix}. \quad (3.5b)$$

3.3 Possible Augmentations

Augmenting a model with more states may affect the observability of the model. Since the purpose of the augmented model is to use it for estimation, observability has to be maintained also after the augmentation. An observability investigation of the augmented model is performed to find which augmentations that are possible. The aim is to derive necessary and sufficient conditions on A_q such that the augmented model is observable. Similar results can be found in (Bembenek et al., 1998), which also includes a discussion regarding the observability results, similar to the short discussion in the end of this section. The observability criterion used in the analysis is known as the Popov-Belevitch-Hautus (PBH)-test (Kailath, 1980).

Theorem 3.1. *A pair (C, F) is observable if and only if*

$$\begin{pmatrix} C \\ \lambda I - F \end{pmatrix}$$

has full column rank $\forall \lambda \in \mathbb{C}$.

Now, using Theorem 3.1 and the assumption that the default model is observable the main result of this section can be formulated as

Theorem 3.2. *Assume that (C, F) in (3.5) is an observable pair then the augmented system (3.5) is observable if and only if*

$$\text{Ker}((F - I) \begin{pmatrix} N_C & A_q \end{pmatrix}) = \{0\},$$

where the columns of N_C span $\text{Ker} C$.

Proof. From Theorem 3.1 it follows that the augmented model (3.5) is observable if and only if $x = 0, q = 0$ is the only solution to

$$Cx = 0 \tag{3.6a}$$

$$(\lambda I - F)x + (F - I)A_q q = 0 \tag{3.6b}$$

$$(\lambda I - I)q = 0 \tag{3.6c}$$

for all $\lambda \in \mathbb{C}$. For $\lambda \neq 1$ it is immediate from (3.6c) that $q = 0$. Then the assumption that (C, F) is an observable pair together with (3.6a), (3.6b), and Theorem 3.1 gives that $x = 0$. Thus, only $\lambda = 1$ needs to be investigated further.

For $\lambda = 1$ in (3.6) the augmented model is observable if and only if $x = 0, q = 0$ is the only solution to

$$\begin{aligned} (F - I)(x - A_q q) &= 0 \\ Cx &= 0. \end{aligned}$$

Let the columns of N_C be a basis for $\text{Ker} C$, then $x = N_C \xi$ for some ξ and observability is equivalent to $q = 0, \xi = 0$ being the only solution to the equation

$$(F - I)(N_C \xi - A_q q) = 0.$$

This is equivalent to that the matrix

$$(F - I) \begin{pmatrix} N_C & A_q \end{pmatrix}$$

has full column rank which ends the proof. \square

This means that the space spanned by the columns in A_q can lie neither in $\text{Ker} C$ nor in $\text{Ker}(F - I)$ for the augmented model to be observable. These interpretations of the rank condition can be understood by analyzing the two requirements separately. First, the requirement that A_q can not lie in $\text{Ker} C$ is easily seen by studying the following linear example.

Example 3.1

Starting with a linear model with a stationary bias

$$x_{t+1} = Fx_t - (F - I)A_q q_t$$

$$q_{t+1} = q_t$$

$$y_t = Cx_t$$

and performing a change of variables, $z_t = x_t - A_q q_t$, gives

$$\begin{aligned} z_{t+1} &= Fx_t - (F - I)A_q q_t - A_q q_t = Fz_t \\ q_{t+1} &= q_t \\ y_t &= Cz_t + CA_q q_t, \end{aligned}$$

which shows that columns of A_q in $\text{Ker } C$ are not observable.

Second, a non-empty $\text{Ker}(F - I)$ implies that the system contains pure integrators, and a bias in $\text{Ker}(F - I)$ is not distinguishable from an unknown initialization of the integrator and is therefore not observable.

A closer look at the requirement that $(F - I)(A_q N_C)$ has to have full column rank conveys some other interesting results. One is, assuming full column rank of $(F - I)$, that the number of augmented states n_q can never exceed the number of linearly independent measurement signals n_y since

$$\begin{aligned} \text{rank}(F - I)(A_q N_C) &= \text{rank}(A_q N_C) \\ &\leq \text{rank } A_q + \text{rank } N_C = n_q + n_x - n_y \leq n_x, \end{aligned} \quad (3.8)$$

i.e., $n_q \leq n_y$. Another, assume once again that $(F - I)$ has full rank which means that the model does not have any pure integrators, then the full column rank condition on $(F - I)(A_q N_C)$ reduces to requiring full column rank of $(A_q N_C)$ or, equivalently, full column rank of the product CA_q . Now if C has one or several zero columns, then CA_q will not contain any information from those rows in A_q corresponding to zero columns in C . That is, those rows in A_q that correspond to zero columns in C will not contribute to the observability. This is illustrated in the following example.

Example 3.2

Illustration of possible augmentations for a default model without pure integrators and

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Let $*$ denote a non-zero element, then some possible augmentations are

$$A_q^1 = \begin{pmatrix} * & 0 \\ 0 & * \\ 0 & 0 \end{pmatrix}, \text{ and } A_q^2 = \begin{pmatrix} 0 & * \\ * & 0 \\ 0 & 0 \end{pmatrix}$$

since

$$CA_q^1 = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}, \text{ and } CA_q^2 = \begin{pmatrix} 0 & * \\ * & 0 \end{pmatrix},$$

which have full column rank. While an augmentation

$$A_q^3 = \begin{pmatrix} * & 0 \\ 0 & 0 \\ 0 & * \end{pmatrix} \text{ is not possible since } CA_q^3 = \begin{pmatrix} * & 0 \\ 0 & 0 \end{pmatrix}$$

does not have full column rank.

3.4 Augmentation Estimation

Now that all possible model augmentations have been characterized by Theorem 3.2, the next question is how to find a suitable augmentation, using measured data from the real system, that fulfills the requirements derived in Section 3.3. The proposed augmentation estimation procedure is divided into two steps, i) from measured data estimate samples of the bias and ii) compute a basis for the bias samples. Three approaches for how to conduct the first step are developed. In the second step a low order augmentation is computed by performing a Singular Value Decomposition (SVD) on selected samples of the bias found in step one.

3.4.1 Bias Estimation

The first step in the estimation of a low order model augmentation deals with estimating the bias, i.e., collect samples of the bias $\beta_t = A_q q_t$. The first approach is quite simple and its main purpose is to illustrate the basic ideas for the estimation of bias samples, whereas the second and third approach are applicable to more general systems. Since the method aims at reducing bias in stationary operating points only stationary behavior and data are studied.

Approach 1

The first approach utilizes the discretized linearization directly and the assumptions that all states are measured, i.e., C_t has full column rank, and that the system does not have any pure integrators, i.e., $(I - F_t)$ has full column rank. The linearized and time discretized augmented model is

$$x_{t+1} = F_t x_t + (I - F_t) A_q q_t + T_s B_t u_t \quad (3.9a)$$

$$y_t = C_t x_t. \quad (3.9b)$$

Due to the full column rank assumptions on C_t and $(I - F_t)$ it is possible to invert the measurement equation and insert the resulting x_t in the dynamic equation. This gives that

$$\beta_t = A_q q_t = (I - F_t)^{-1} (C_{t+1}^\dagger y_{t+1} - F_t C_t^\dagger y_t - T_s B_t u_t),$$

where \dagger denotes the Moore-Penrose inverse (Lancaster and Tismenetsky, 1984).

This approach shows, in a direct way, that the estimation approaches are about finding ways to solve $A_q q_t$ from (3.9). However, it requires a full column rank C_t and $(I - F_t)$ and, since no filtering of the measurements is involved, it is sensitive to low signal to noise ratio (SNR).

Therefore two other approaches are proposed for estimating bias samples. Common for both these approaches are that they utilize the residuals from an observer and the assumption that the true bias enters the model according to Equation (3.3). The fact that they are based on observers makes them less

sensitive to low SNR and imply that they do not require full column rank C_t to work. The first employ an observer based on the default model and the bias samples are computed by inverting the observer system. The second employ a fully augmented model fulfilling the observability requirements developed in Section 3.3.

Approach 2

The second approach uses the residuals originating from an observer based on the default model. Here, the observer is an extended Kalman filter (EKF) (Kailath et al., 2000), where the noise covariance matrices Q and R are design parameters tuned by the user. Of course, other observer designs are equally possible, for example an Unscented Kalman Filter (UKF) (Julier and Uhlmann, 1997, 2004). Let K_t be the EKF feedback gain, then the estimation error becomes,

$$\begin{aligned}
 e_{t+1} &= x_{t+1} - \hat{x}_{t+1|t+1} \\
 &= F_t x_t + (I - F_t) A_q q + T_s B_t u_t - \\
 &\quad (F_t \hat{x}_{t|t} + T_s B_t u_t + K_t (y_{t+1} - C_t F_t \hat{x}_{t|t} - C_t T_s B_t u_t)) \\
 &= \{y_{t+1} = C_t F_t x_t + C_t (I - F_t) A_q q + C_t T_s B_t u_t\} \\
 &= (F_t - K_t C_t F_t) e_t + (I - K_t C_t) (I - F_t) A_q q. \tag{3.10}
 \end{aligned}$$

Equation (3.10) can not be used directly since the state estimation error is not known. Therefore, the output error

$$r_t = y_t - \hat{y}_{t|t} = C_t (x_t - \hat{x}_{t|t}) = C_t e_t, \tag{3.11}$$

is used for estimating the bias.

As previously stated, solely stationary parts of the residuals are involved in the bias estimation. It would be possible to use also dynamic parts of the residuals and a dynamic inverse. The reason for not utilizing these here is to prevent dynamic estimation errors from affecting the estimation of the constant or slowly varying bias.

Now, utilizing that only stationary data is considered, (3.10) and (3.11) can be combined resulting in

$$\begin{aligned}
 r_{\text{stat}} &= C_{\text{stat}} e_{\text{stat}} \\
 &= C_{\text{stat}} (I - F_{\text{stat}} + K_{\text{stat}} C_{\text{stat}} F_{\text{stat}})^{-1} \times \\
 &\quad (I - K_{\text{stat}} C_{\text{stat}}) (I - F_{\text{stat}}) A_q q_{\text{stat}}
 \end{aligned}$$

and the bias can be estimated as

$$\begin{aligned}
 \beta_t = A_q q_t &= (C_{\text{stat}} (I - F_{\text{stat}} + K_{\text{stat}} C_{\text{stat}} F_{\text{stat}})^{-1} \times \\
 &\quad (I - K_{\text{stat}} C_{\text{stat}}) (I - F_{\text{stat}}))^\dagger r_t. \tag{3.12}
 \end{aligned}$$

Approach 3

An alternative to Approach 2 for finding β_t is to augment the default model with as many extra states as possible. According to Theorem 3.2, the requirement on A_q is that $(F-I)(A_q N_C)$ has to have full column rank. This means that A_q can have a maximum of n_y columns. These columns have to be linearly independent of the columns of N_C and can not lie in $\text{Ker}(F-I)$. One way to construct such an augmentation is to use C^\dagger and leave out those columns that become zero when multiplied by $(F-I)$ from the left. Then run the observer based on the augmented model, estimating both \hat{x} and \hat{q} , and assemble $\beta_t = C^\dagger \hat{q}_t$.

An advantage with this approach is that no inversions as those in (3.12) are needed. A disadvantage though is that since a fully augmented model is used the order of the observer might be unnecessarily high.

3.4.2 Augmentation Computation

As stated in the problem formulation in Section 3.1, the bias is necessarily not the same in the entire operating region of the system. This makes it important to collect samples of the bias from stationary operating points selected such that the entire operating region is covered. From the first step of the proposed procedure, bias samples are collected according to this. Based on the discussion of only a few underlying biases affecting several states in Section 3.1, the task of step two is to find a low order basis spanning the space in which these bias samples are located.

To start with bias samples from N stationary operating points are assembled

$$\bar{\beta}^{n_x \times N} = (\beta^1 \quad \dots \quad \beta^N),$$

Then the SVD of $\bar{\beta}$ is computed,

$$\bar{\beta} = U\Sigma V^*,$$

where U contains orthogonal vectors spanning the space in which the bias moves and Σ are the corresponding singular values. The singular values in Σ are ordered in non-increasing order which means that the far left columns of U , corresponding to large singular values, represent the most dominating directions along which the bias moves. Therefore the dimension of q can be found by comparing the singular values in Σ , and picking the most significant ones. Then the corresponding columns of U are used to assemble \hat{A}_q .

This way of computing an augmentation from bias samples is optimal with respect to the Frobenius norm, i.e.,

$$\hat{A}_q = \arg \min \|\bar{\beta} - \hat{A}_q\|_F.$$

3.4.3 Properties of the Estimated Augmentation

According to the discussion in the end of Section 3.3, the properties of C place restrictions on which A_q :s that are possible to find. The conclusion of that

discussion is that rows in A_q corresponding to zero columns in C become zero in the estimation step. However, a more thorough analysis shows that more can be said.

Theorem 3.3. *Assume that the observer in Approach 2 is strictly stable and does not have any poles in the origin, then, in absence of noise, the bias samples for all three approaches are spanned by the rows of C , that is*

$$\beta_t = C^T \Gamma.$$

Proof. Since Approach 1 is applicable only if C has full column rank and due to the augmentation, C^\dagger , used in Approach 3 the theorem automatically holds for these cases. It is therefore sufficient to prove the result for Approach 2.

Now, starting with the output error and rewriting it

$$\begin{aligned} r_t &= C \overbrace{(I - F + KCF)}^W{}^{-1} (I - KC)(I - F)\beta_t \\ &= CW^{-1}(W - KC)A_q q = (I - CW^{-1}K)C\beta_t, \end{aligned} \quad (3.13)$$

where the assumption that K is chosen such that the observer system, $(I - F + KCF)$, is strictly stable and does not have any eigenvalues equal to zero which assures that W^{-1} exists, is used. Then, using the Moore-Penrose inverse, (3.13) can be written as

$$C\beta_t = (I - CW^{-1}K)^\dagger r_t = \bar{r}_t. \quad (3.14)$$

A unique solution to (3.14) is received by computing the minimum square solution with least Euclidean norm. Writing

$$\beta_t = \beta_t^o + \beta_t^\perp, \quad (3.15)$$

where

$$\beta_t^o \in (\text{Ker } C)^\perp = \text{span}\{C^T\} \quad (3.16)$$

and

$$\beta_t^\perp \in \text{Ker } C, \quad (3.17)$$

the solution with least Euclidean norm is the solution with $\beta_t^\perp = 0$, i.e.,

$$\beta_t = \beta_t^o = C^T \Gamma \quad (3.18)$$

which concludes the proof. \square

Note that Theorem 3.3 holds for the Moore-Penrose inverse and is not generally true for an arbitrary left inverse.

As a consequence of Theorem 3.3, the observer based on an estimated augmentation may not be able to reduce the bias in the estimates to acceptable levels. This problem can be circumvented in, for example one of the two following ways. The first is for an engineer to design an A_q not possible to find through estimation, for example through knowledge of the underlying physics and knowledge about the modeling errors. The second is to, during the design phase, add extra sensors to the true system to acquire a full column rank C which enables estimation of all rows in A_q . When utilizing this possibility one must be cautious and check the observability of the augmented system that in the end will not rely on the additional sensors used for estimating \hat{A}_q . That is, check the column rank of $(F - I)(\hat{A}_q N_C)$, and in case of column rank deficiency remove those columns in \hat{A}_q causing rank deficiency. Since SVD is used, the columns in \hat{A}_q are arranged in non-increasing significance order which makes it appropriate to remove the columns in \hat{A}_q starting from the right to get an augmentation that is observable.

The example below illustrates the remarks regarding the effects that properties of C have on the augmentation estimation.

Example 3.3

Consider a true system with

$$F = \begin{pmatrix} 1 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix}, \text{ and } C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \end{pmatrix}$$

and a true bias,

$$A_q q = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Then the estimate of A_q , according to Theorem 3.3, will be

$$\hat{A}_q = C^T \Gamma = \frac{1}{\sqrt{1^2 + (14/5)^2 + (7/5)^2}} \begin{pmatrix} 1 \\ 2 \times 7/5 \\ 1 \times 7/5 \end{pmatrix},$$

found by minimizing

$$\left\| A_q q - C^T \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} \hat{q} \right\|_2 = (1 - \Gamma_1)^2 + (2 - 2\Gamma_2)^2 + (3 - \Gamma_2)^2$$

with respect to Γ_1 and Γ_2 .

That is, because of the structure of C the estimated bias direction will be incorrect resulting in a stationary estimation error

$$x - \hat{x} = A_q q - \hat{A}_q \hat{q},$$

which with

$$CA_q q = C\hat{A}_q \hat{q} \Rightarrow \hat{q} = \sqrt{1^2 + (14/5)^2 + (7/5)^2}$$

becomes

$$x - \hat{x} = A_q q - \hat{A}_q \sqrt{1^2 + (14/5)^2 + (7/5)^2} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 14/5 \\ 7/5 \end{pmatrix} = \begin{pmatrix} 0 \\ -4/5 \\ 8/5 \end{pmatrix}.$$

This example illustrates the result in Theorem 3.3, and shows that an incorrect estimation of the bias direction result in stationary estimation error even though the system output is estimated correctly. In this case the correct estimation of system output result in correct estimation of x_1 while x_2 and x_3 are not estimated correctly, due to the structure of C .

3.4.4 Evaluation of Augmentation Estimation Approaches

Two main approaches, approaches 2 and 3, for estimating the bias have been proposed. It is important to understand how these approaches perform under varying operating conditions and model uncertainty. Therefore, the approaches are evaluated with respect to robustness against model errors and robustness to changes in noise levels. This is done by introducing noise and modeling errors in a non-linear simulation model of a Scania diesel engine with exhaust gas recirculation (EGR) and variable geometry turbine (VGT), and performing Monte Carlo simulations. In the simulations, a one-dimensional q is also introduced, i.e., A_q is a vector with three elements.

Modeling errors can be introduced in many ways and it is difficult to obtain a comprehensive evaluation of robustness properties of a non-linear method. Therefore, a more pragmatic approach is adopted. First, model errors are introduced by manipulating physical constants in the simulation model and thus making the simulation model, that generates the observations, different from the default model used for designing the observer. Another way model errors are introduced is by pre-multiplying the vector field f in (3.1a) by a slowly varying sinusoid, i.e., the simulation is done with $\bar{f}(x, u)$ defined as $\bar{f}(x, u) = (1 + \gamma \sin(\Lambda t))f(x, u)$, where Λ is the model error frequency, and γ is a small number varied between 0.1 and 0.5. Doing Monte Carlo simulations with such model errors reveal that both approaches react similarly to the model errors with respect to degraded performance in bias estimation and variance in the estimation. No certain conclusion can be drawn concerning which approach is more robust against modeling errors and the overall picture is that both approaches have similar graceful performance degradation with increased modeling errors.

Examining the effect of measurement noise is done by introducing white Gaussian noise with different noise levels in the simulation model and estimating

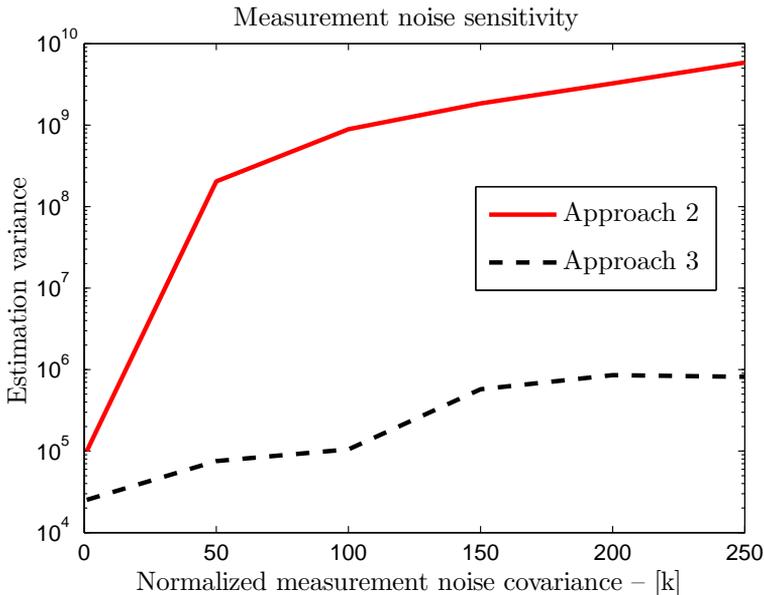


Figure 3.1: Estimation variance for bias estimation approaches 2 and 3 with different measurement noise – $k \cdot \mathcal{N}(0, R)$.

the effect on the augmentation estimation by computing the variance in the β^i :s. In Figure 3.1 the effect of increased measurement-noise level on the variance in the estimated β^i :s is shown. It is seen that Approach 3 is significantly less sensitive to measurement noise and therefore preferable when estimating an augmentation.

3.4.5 Method Summary

The procedure can be summarized in three steps.

Step 1 - Linearize and discretize the model if necessary. Normally, the default model is a non-linear time continuous model such as (3.1) and has to be linearized and discretized.

Step 2 - Find an appropriate augmentation, A_q , and compile an augmented model (3.4). Here the designer has three options, either to estimate an augmentation from measured data, introduce an augmentation found in some other way, or to combine an estimated augmentation with one found through system knowledge.

The estimation procedure contains two steps, i) estimation of bias samples utilizing one of the three approaches presented in Section 3.4.1, ii) compute a basis for the bias samples using SVD.

With good knowledge of the system, the designer might have some idea of what is causing the bias in the estimates and can choose an appropriate A_q .

To combine an augmentation found through process knowledge with one found through estimation can be desirable if some model deficiencies are known but does not manage to achieve satisfactory bias reduction. In this case the estimation approach can be applied to the, by the engineer, partly augmented model to find an additional augmentation that captures the remaining dominating bias.

Step 3 - Design an observer based on the augmented model (3.3) and the A_q found in Step 2. In this chapter, an EKF is used but any non-linear observer design methodology is possible.

3.5 Experimental Evaluation

To evaluate the method experiments are performed using a non-linear model of a heavy-duty truck engine. The experiments consist of both a simulation study of the non-linear model, and an evaluation of the method on measurement data from an engine test cell.

The non-linear model of the diesel engine has three states: p_{im} , p_{em} , and n_{trb} , that represent intake and exhaust manifold pressures, and turbine speed, respectively. See Appendix A for more information about the engine and engine model. In the second experiment, real data from the engine is used together with the engine model to illustrate the properties of the proposed approach in a real application. In both experiments the stationary parts of the data, used in the augmentation estimation, are separated out through visual inspection and estimation Approach 3 is chosen to estimate the bias.

3.5.1 Evaluation Using Simulated Data

The objective of the first experiment is to illustrate how the approach, which is based on linearization procedures, performs when fed with data from a non-linear simulation model. Thus, synthetic data is created where known biases are introduced in the simulation. The method is then applied to show how biases in non-linear systems can be estimated.

The introduced bias is represented by a matrix

$$A_q = \begin{pmatrix} 1 & -2 \\ 2 & 1 \\ 0 & 0.2 \end{pmatrix},$$

and two slowly varying biases q_1 and q_2 . This A_q means that there are two independent biases affecting the model states which varies between approximately 0 and 10% of the state values. The default system has linear measurement

equations where $y_1 = p_{\text{im}}$ and $y_2 = n_{\text{trb}}$. However, according to the discussion in Section 3.4.3, an augmentation as the one introduced in this example cannot be estimated without a direct connection between p_{em} and y . Therefore the measurement equation is extended with an extra sensor for p_{em} for the augmentation estimation. Note that this extra sensor is not used for feedback neither in the observer based on the default model nor in the observer based on the augmented model. This reflects the situation that a lab environment or development system may be equipped with extra sensors to achieve a better augmentation estimation.

The observer based on the default model is referred to as the default observer while the observer based on the augmented model is referred to as the augmented observer. Both observers only use the p_{im} and p_{em} measurements. To make the simulation more realistic, white system and measurement noise are added in the creation of the synthetic data.

Using the simulated data and the default model, the augmentation estimation results in

$$\Sigma \approx \begin{pmatrix} 5.0259 & 0 & 0 \\ 0 & 4.8669 & 0 \\ 0 & 0 & 0.0024 \end{pmatrix} 10^5,$$

and

$$U \approx \begin{pmatrix} -0.8295 & -0.5527 & 0.0800 \\ 0.5515 & -0.8233 & -0.0388 \\ 0.0881 & 0.0123 & 0.9960 \end{pmatrix},$$

where Σ indicates that there are two slowly varying biases present. Hence, \hat{A}_q is estimated using the first two columns of U .

At a first look \hat{A}_q does not appear similar to A_q . However, the crucial fact is that the columns of \hat{A}_q and A_q span, approximately, the same space. A closer look reveals that the elements in the bottom row is significantly smaller than the other elements, and that the factor between row one and two are approximately 2. That is, the only thing that differs between A_q and \hat{A}_q is a scaling. The objective was not only to estimate the bias, but rather to obtain an observer that compensated for the model bias.

In the third step, an observer is created using EKF methodology for a model augmented according to the estimated \hat{A}_q . The performance is compared to the default observer. The state estimates are presented in Figure 3.2 together with the true states. Note that the estimated states from the augmented observer coincide with the true states. It is easily seen that the augmented observer estimates p_{im} and n_{trb} better than the default observer. To obtain a better view of the observer performance, the estimation errors are plotted in Figure 3.3. Here it is clear that all three state estimates become better with the augmented observer than with the default observer.

The conclusion of this small simulation example is that the approach managed to get a good enough estimate of a bias in a non-linear model to improve the state estimates.

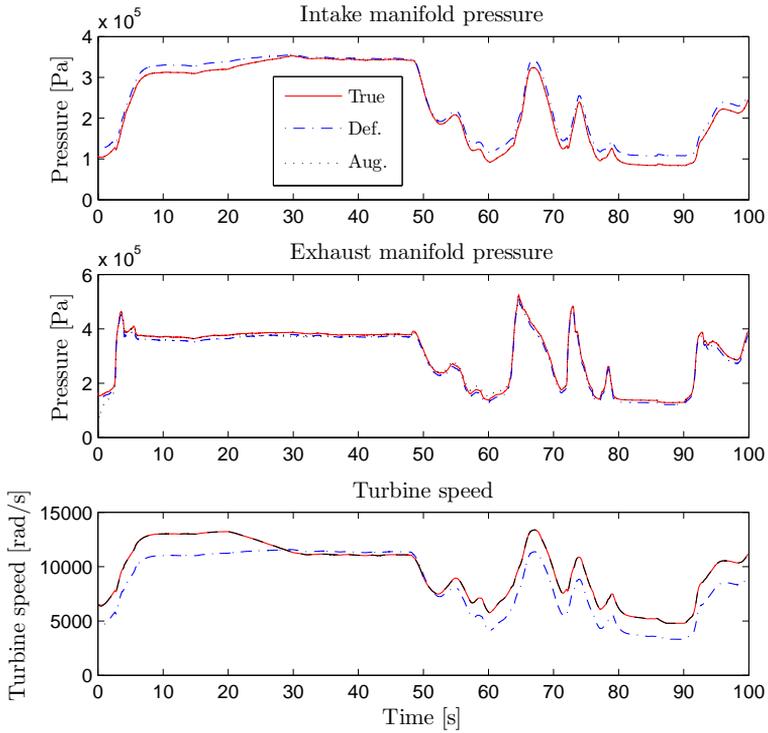


Figure 3.2: True states and estimated states using default and augmented observer in the simulation study. Note that the estimated states from the augmented observer coincides with the true states.

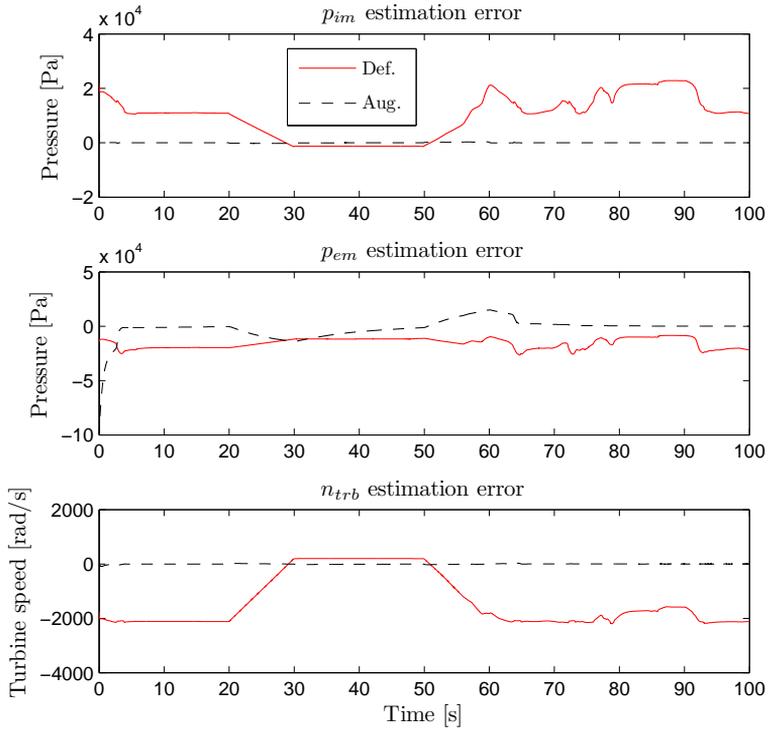


Figure 3.3: Estimation errors using default and augmented observer in the simulation study.

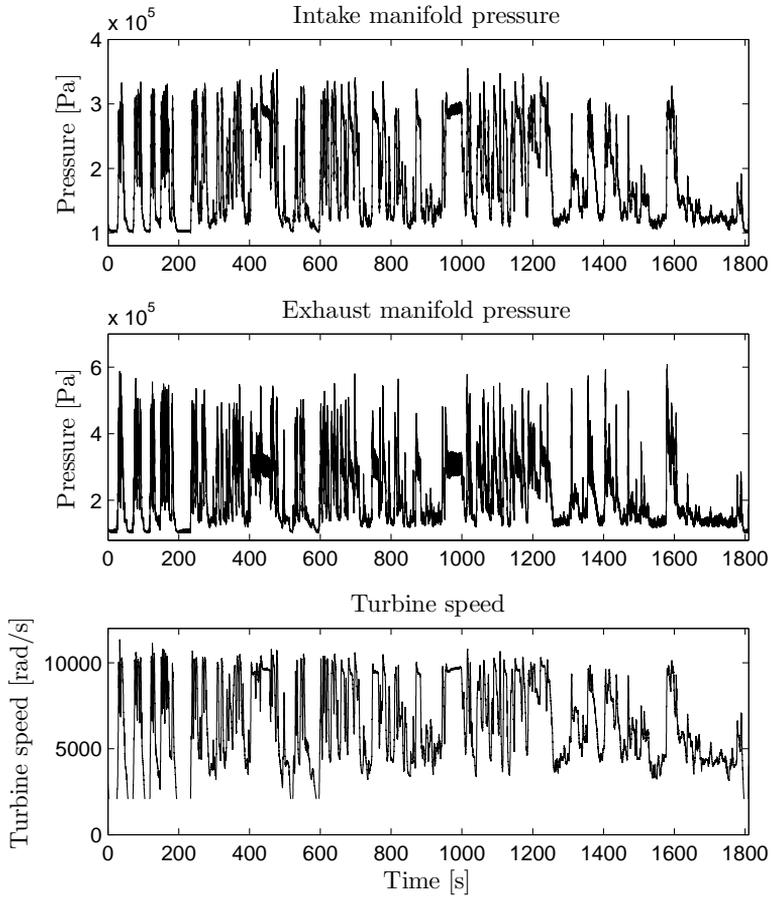


Figure 3.4: Measurements of p_{im} , p_{em} , and n_{trb} from the ETC used in the experimental evaluation. Note that turbine speeds below approximately 2100 rad/s are missing. This is due to the limited measurement range of the turbine speed sensor.

3.5.2 Two Experimental Evaluations

The experimental data described in Appendix A is used to evaluate the augmentation estimation and observer performance. The true states are approximated by non-causal, zero-phase, low-pass filtered measurements, where the filter has a cut off frequency of 2 Hz, see Figure 3.4. In the measurements parts of the turbine speed data is missing, which is due to the fact that the measuring range of the turbine speed sensor is limited. Speeds below 20 000 rpm, or approximately 2 100 rad/s cannot be measured. Since the observers use feedback from this sensor this issue has to be solved. Here it is solved by setting the elements in the observer gain from the turbine speed equal to zero when the turbine speed measurement is equal to zero.

Based on the measurement data, an augmentation is estimated using data from two stationary operating points in the European transient cycle (ETC) of about 1000 samples each. All states are measured and the augmentation estimation results in

$$\Sigma \approx 10^5 \begin{pmatrix} 5.3230 & 0 & 0 \\ 0 & 0.3739 & 0 \\ 0 & 0 & 0.0044 \end{pmatrix},$$

and

$$U \approx \begin{pmatrix} -0.2610 & 0.9650 & -0.0249 \\ -0.9648 & -0.2671 & -0.0274 \\ -0.0329 & 0.0169 & 0.9993 \end{pmatrix}, \quad (3.19)$$

where Σ indicates that there is one dominant slowly varying bias present. Hence, \hat{A}_q is selected to be only the first column of U .

Reduced Augmentation Order

In this system it is possible to augment the system with three extra states and still have an observable system if all states are measured. One interesting question is if the proposed method that estimates a lower dimension augmentation can still capture most of the bias. Therefore, three observers are designed: the default observer, a fully augmented observer, and a one-dimensional augmentation observer.

The aim of this comparison is thus to conclude whether the proposed method works, and is performed by analyzing the estimation errors from the three observers. The resulting probability density functions (PDF) of the estimation errors are shown in Figure 3.5 and the mean and maximum absolute errors for the entire ETC are presented in Table 3.1. From the data it is clear that the default observer has a bias and that the augmented observers reduce the bias. Now comparing the two augmented observers it is seen that the observer with only a one-dimensional augmentation delivers close to the same reduction in bias as the fully augmented observer. This is a clear illustration that the method succeeds in finding the dominant bias in the model.

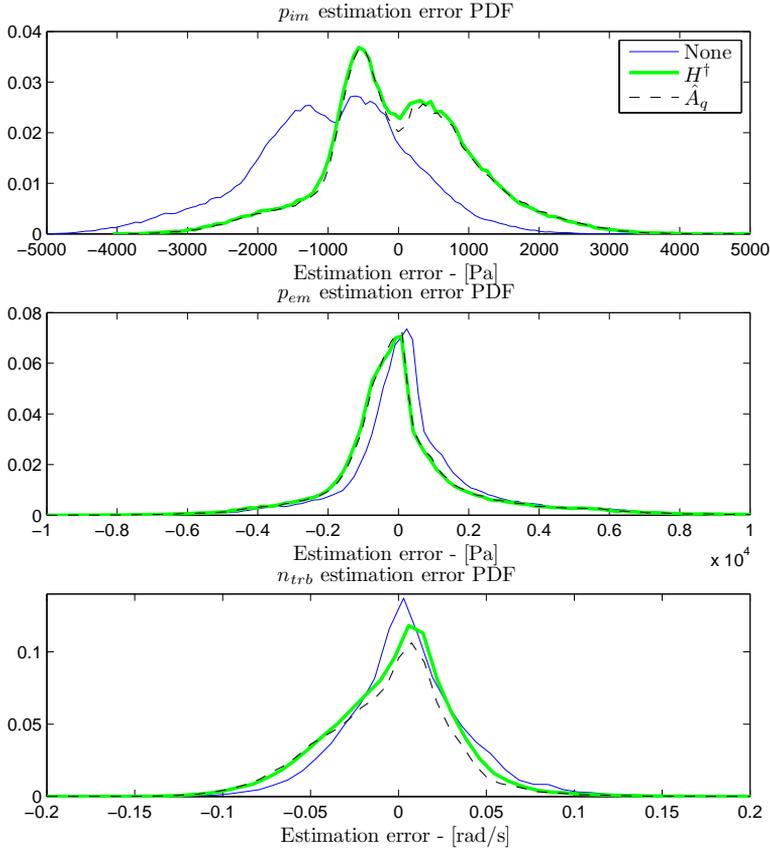


Figure 3.5: Probability density functions for three observers: None – default observer, H^\dagger – observer augmented with three states, and \hat{A}_q – observer augmented with one state and the estimated \hat{A}_q .

Table 3.1: Data from observers None – default observer, H^\dagger – fully augmented observer, and \hat{A}_q – observer using reduced dimension augmentation found using augmentation estimation approach 3. All observers use feedback from all states.

States	Max abs. error			Mean error		
	None	H^\dagger	\hat{A}_q	None	H^\dagger	\hat{A}_q
p_{im} [Pa]	5459	6840	6599	-985	11	37
p_{em} [Pa]	14411	14277	14278	443	86	132
n_{trb} [rad/s]	0.8	0.7	0.6	0.005	-0.003	-0.007

Table 3.2: Static data from observers None – default observer, \bar{A}_q – observer using augmentation estimated using only measurements of p_{im} , and n_{trb} , and \hat{A}_q – observer using augmentation estimated using measurements of all states. All observers use feedback from p_{im} , and p_{em} only.

States	Max abs. error			Mean error		
	None	\bar{A}_q	\hat{A}_q	None	\bar{A}_q	\hat{A}_q
$p_{\text{im}}[\text{Pa}]$	4191	3650	3641	-622	-80	-176
$p_{\text{em}}[\text{Pa}]$	58758	58197	51322	6810	6328	-678
$n_{\text{trb}}[\text{rad/s}]$	0.1	0.1	0.1	0.02	0.004	0.006

Table 3.3: Dynamic data from observers None – default observer, \bar{A}_q – observer using augmentation estimated using only measurements of p_{im} , and n_{trb} , and \hat{A}_q – observer using augmentation estimated using measurements of all states. All observers use feedback from p_{im} , and p_{em} only.

States	Max abs. error			Mean error		
	None	\bar{A}_q	\hat{A}_q	None	\bar{A}_q	\hat{A}_q
$p_{\text{im}}[\text{Pa}]$	5748	6828	6316	-533	2	-34
$p_{\text{em}}[\text{Pa}]$	180279	177982	174486	16604	16479	8922
$n_{\text{trb}}[\text{rad/s}]$	0.9	0.6	0.5	0.02	0.0007	-0.001

Benefits of Additional Sensor During Design

Another interesting question is what can be achieved by allowing extra sensors, compared to what is used in the final observer, while estimating an augmentation. The application chosen is to estimate the exhaust manifold pressure with reduced bias compared to a default observer without having a sensor measuring it. That is, design an observer for p_{em} using feedback from p_{im} and n_{trb} . The analysis is performed by comparing the estimates from two observers; one based on the augmentation

$$\hat{A}_q = (-0.2610 \quad -0.9648 \quad -0.0329)^T$$

estimated using measurements of p_{im} , p_{em} , and n_{trb} , i.e., column one in (3.19), and another based on an augmentation

$$\bar{A}_q = (-0.9864 \quad 0 \quad -0.1644)^T$$

estimated using measurements of p_{im} , and n_{trb} only.

The two augmented observers are compared to the default observer and the results are shown in Figure 3.6 and Table 3.2 and 3.3. Figure 3.6 shows the probability density function for the estimation errors for the default observer, the observer based on the model augmented with \hat{A}_q , and the observer based on

the model augmented with \bar{A}_q . It is seen that both augmented observers reduce the mean of the bias for p_{im} and n_{trb} compared to the default observer and that the observer based on the model augmented with \hat{A}_q significantly reduces also the bias in p_{em} . Table 3.2 and 3.3 show the mean and maximum absolute estimation errors for selected stationary parts of an ETC and for the entire ETC respectively. In both tables it is obvious that the observer based on a model augmented with \hat{A}_q significantly reduces the estimation bias. The mean error is reduced by approximately 50 % during an entire ETC and by approximately 90 % for selected stationary parts, while the maximum absolute errors are almost unaffected. These, quite large, differences in the different measures are all explained by the fact that the suggested method reduces stationary bias and, as can be seen in Figure 3.4 the ETC is a rather dynamic sequence and the maximum absolute errors occur in transients.

3.6 Conclusions

A method for bias compensation in observers is developed. The idea is to, based on measurement data, compute a low dimension augmentation of the model that describes the most significant model biases. This augmented model is used to design an augmented observer that results in a state estimate with reduced bias. Three main results are a characterization of possible augmentations from observability perspectives, a parameterization of the estimated augmentations, and a robustness analysis of the proposed augmentation estimation method.

The method is successfully applied to a diesel engine with VGT and EGR, using a non-linear default model and measurement data from an engine in a test cell. It is shown that an augmentation according to the suggested augmentation procedure reduces the mean estimation error, that is the bias, by approximately 50 % in an ETC.

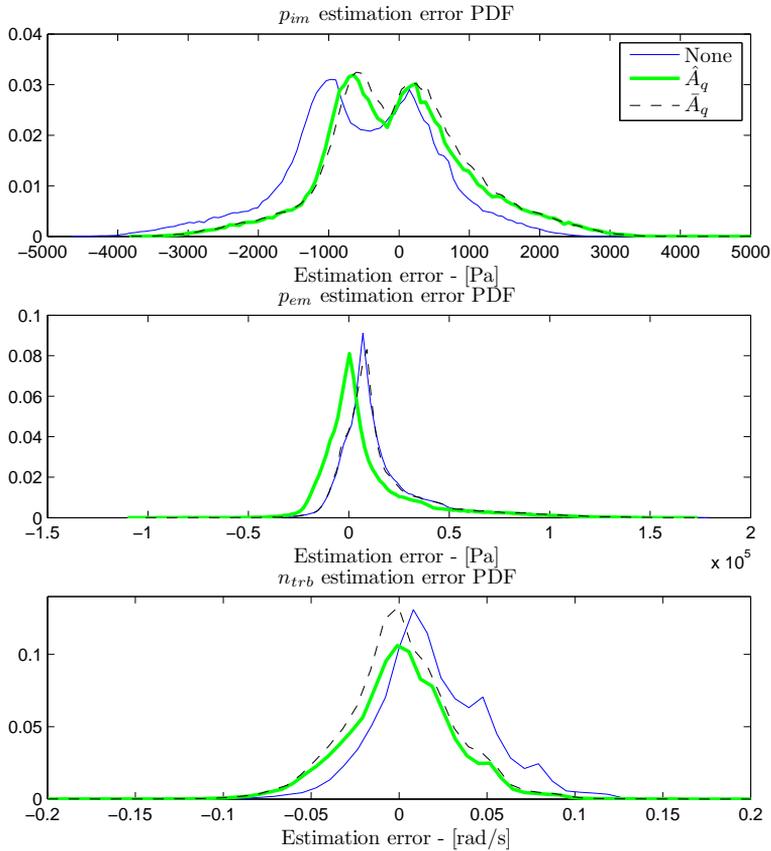


Figure 3.6: Probability density functions for default and augmented observers applied to real measurement data using feedback from p_{im} and n_{trb} . The two augmented observers are Red. – augmentation estimated measuring p_{im} and n_{trb} and Full – augmentation estimated measuring p_{im} , p_{em} , and n_{trb} respectively.

PARAMETERIZING THE BIAS

The observer designed in Chapter 3 handles bias by modeling it as a random walk. In cases where the bias is operating point dependent, an observer designed according to this is connected with some issues. One issue is that as soon as the system changes operating point, all information about the bias in the previous operating point is discarded. It would be desirable to have an observer that remembers the system bias in each operating point. An observer with a bias memory would have several advantages, for example the robustness can be increased and it can be used in engine map adaptation algorithms, compared to an observer that models the bias as a random walk, and how to design an observer with a memory is the topic of this chapter.

4.1 Method Outline

Based on the method developed in Chapter 3, an information preserving observer, like the one discussed above, can be obtained by exchanging the assumption of a bias modeled as a slowly varying bias state

$$q_{t+1} = q_t,$$

driven by white noise, for a parameterized function, or map, describing the bias

$$\begin{aligned} q_t &= q_{\text{fcn}}(x_t, y_t, u_t, \theta_t) \\ \theta_{t+1} &= \theta_t. \end{aligned}$$

Here, $q_{\text{fcn}}(x_t, y_t, u_t, \theta_t)$ is a parameterized function, or map, with unknown parameters θ that describe the bias dependence on the system states x , outputs

y , and inputs u . Further, the parameter states are modeled as slowly varying states in the same way as the bias state in Chapter 3.

If this is done with the augmented model developed in Chapter 3, the following system description is obtained,

$$x_{t+1} = x_t + T_s f(x_t - A_q q_t, u_t) \quad (4.1a)$$

$$\theta_{t+1} = \theta_t \quad (4.1b)$$

$$q_t = q_{\text{fcn}}(x_t, u_t, \theta_t) \quad (4.1c)$$

$$y_t = h(x_t). \quad (4.1d)$$

Note that the only difference with this formulation compared to the one used in Chapter 3 is that the bias states have been exchanged for a parameterized function and the function parameters are introduced as new states. Further this is, if (4.1c) is inserted in (4.1a), a standard state space form which means that any suitable observer design can be applied. For example, one way of estimating states while at the same time handle unknown parameters is to apply a joint parameter and state estimating extended Kalman filter (Kopp and Orford, 1963), or unscented Kalman filter (Wan et al., 2000). There the parameters are introduced as new states with constant time derivatives and augmented to the original states just as in a system described by Equation (4.1).

This formulation can be used in several different applications, for example to reduce the lag in estimation during transients or as a new way of updating engine maps (Wu, 2006; Peyton Jones and Muske, 2007). However, to develop a model like (4.1) entails that some new questions have to be addressed:

- i) –What inputs should be used in the parameterized function and how should the function be parameterized with respect to these inputs?
- ii) –How can the function parameters be updated in a controlled manner?

This chapter addresses the second of these problems and leaves the first for future work. A short discussion regarding the parameterization and how to find it is presented in Section 5.2. Further, the possibilities as well as the affinity of these ideas to other research areas, such as for example simultaneous localization and mapping (SLAM) (Durrant-Whyte, 1988; Smith et al., 1990), and switched linear systems (Vidal et al., 2002), are discussed in Section 4.2.

Even though any suitable observer design can be applied to this system the choice here is to use a stochastic filter, i.e., an EKF. An advantage of stochastic filters, like for example EKF and UKF, compared to deterministic observers is that, not only the state estimate, but also an estimate of the estimation error distribution is computed. Further, the estimation error statistics is used in the computation of the filter feedback gain, which gives the stochastic filters natural tuning parameters that allow filter tailoring to handle unknown initialization of states, time dependent model and measurement quality etc.

4.2 Observability

In all estimation problems observability or, at least, detectability of the system at hand is central. Since the system considered in this chapter is a further extension of the augmented system addressed in Chapter 3 it is natural to assume that the system

$$x_{t+1} = x_t + T_s f(x_t - A_q q_t, u_t) \quad (4.2a)$$

$$q_{t+1} = q_t \quad (4.2b)$$

$$y_t = h(x_t) \quad (4.2c)$$

addressed in Chapter 3 is observable and analyze how the replacement of (4.2b) for a parameterized function, (4.1b – 4.1c), affects the observability. The observability is therefore directly dependent on the properties of the parameterized function (4.1c) that describes the dependence of the bias q_t on operating point.

If for example (4.1c) is an engine map, implemented as a look-up table with the grid points as parameters, and with an interpolation function that computes the output. Then the system will be neither locally, nor weakly observable, i.e., parameters that are not used in the interpolation are not locally observable. In Figure 4.1 the local observability of the parameter states in linear interpolation is illustrated using the air mass-flow correction map discussed in Chapter 2. If the operating point, in this case the air mass-flow measurement, is in the shaded region of the plot then only the two grid points, constituting the border of the shaded region, are locally observable. This way of reasoning can be applied also to other interpolation methods like for example cubic spline with the only difference that more than the two closest parameter states, denoted θ_i and θ_{i+1} with $i = 8$ in Figure 4.1, are locally observable in each interpolation.

Definitions of the terms weak and local observability can be found in for example (Hermann and Krener, 1977). Weak observability refers to being able to distinguish one state from its immediate neighbors at all times while local observability means that it is possible to distinguish every state from each other instantaneously.

In switched, or jump, linear systems (Vidal et al., 2002) observability is studied for trajectories, and for a system to be observable there has to exist a trajectory such that the switching sequence is uniquely defined. Based on this, the observability of the systems studied here can be assessed by studying a set of trajectories, where for each of these trajectories the default states x and some of the parameter states θ_i are locally observable. Consider for example a trajectory in the shaded region of Figure 4.1. For such a trajectory x , θ_8 , and θ_9 are locally observable. Then, a system is observable if for all parameters there exist an interval on the trajectory where the parameter is locally observable.

4.2.1 Relations to SLAM

To facilitate the discussion of affinity between SLAM and the estimation problem dealt with in this chapter, further discussed in this section and the section

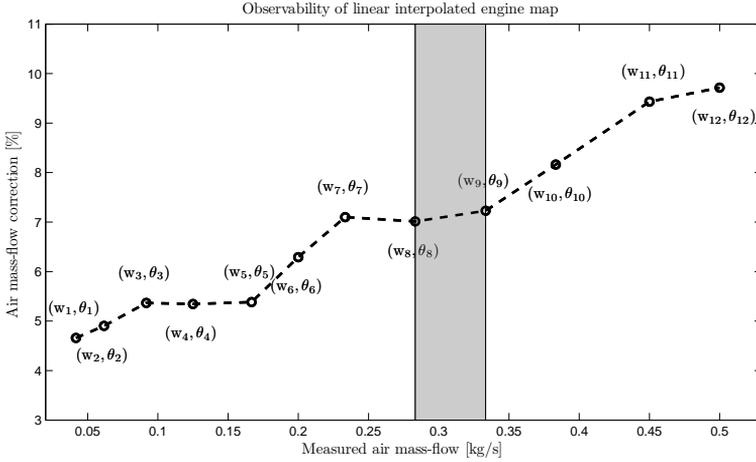


Figure 4.1: Air mass-flow correction map with the grid points marked with a circle and denoted with the pair (w_i, θ_i) corresponding to a correction factor of θ_i at a mass-flow of w_i . If the measurement is in the shaded area of the plot only the two parameters on the border line are locally observable while the other remain unobservable.

to come, some relations to SLAM are presented. The problem of introducing a parameterized function describing the bias have several resemblances, and some differences, with SLAM. Before discussing the relations to SLAM a short presentation of SLAM is conducted. SLAM refers to the problem of estimating a position of for example a robot, and at the same time building a map over the environment in which the robot moves. The map is constructed from so called features, or landmarks, that are sensed by the robot.

Figure 4.2 illustrates SLAM by presenting the solution to a SLAM problem. Figure 4.2(a) presents the true map and trajectory of the robot, while figures 4.2(b) – 4.2(e) shows the initialization and build up of the map.

In SLAM the term local observability used throughout this chapter would refer to the fact that not all features are measured at every time sample, i.e., if the robot leaves a room the features in that room are not detected until the robot returns to that room, which corresponds to only having the nearest grid points active when interpolating in an engine map.

However, when it comes to observability in SLAM the only aspect analyzed in the literature is that the position is only estimated relatively to the map which gives a corresponding unobservable mode. In SLAM this is solved by for example fixating one feature in the map. Further, the SLAM problem is marginally stable, meaning that for the parameter estimation error to converge to zero, the robot trajectory has to be a closed loop, illustrated in Figure 4.2. This loop closing is not necessary for a stable system like the ones studied in this chapter.

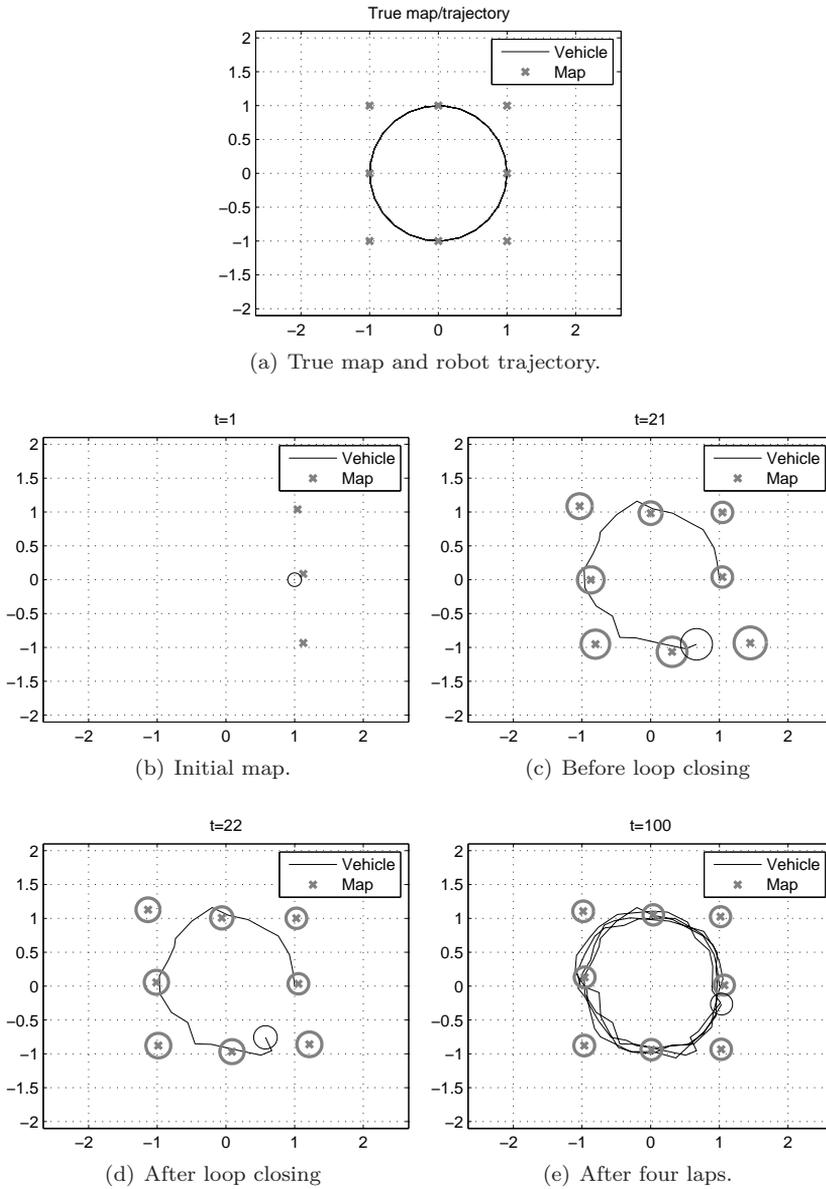


Figure 4.2: Example of a solution to a SLAM problem. The robot trajectory is plotted together with the mapped features, indicated by crosses, and their associated covariance, indicated by ellipses. *Illustration: (Törnqvist, 2008).*

However, the problem of locally unobservable features that is of interest in this chapter is not discussed. Instead the problem of local unobservable features is discussed as a convergence issue and handled by having zero system noise on the parameters describing the map features. In an engine map application this is not an option since the whole idea is to be able to capture and track engine aging and changed operating conditions. This force an introduction of system noise also on the parameter states describing the parameterized function.

Another difference between SLAM and the estimation problem addressed in this chapter is that in SLAM the number of map features, corresponding to the number of unknown function parameters in Equation (4.1b), increase as the robot moves. Which means that the number of features can become very large, while in Equation (4.1) the number of parameters, θ , is constant and typically not large, at least for the one- and two-dimensional engine maps used in modern engine control systems today.

4.2.2 Growing Estimation Error Covariance

A problem that comes with a locally unobservable system is the propagation of the estimation error covariance matrix in an EKF/UKF framework. In regions where the system seldom operates the estimation error covariance matrix coefficients, corresponding to locally unobservable parameter states, will grow linearly without bound. This linear growth in covariance matrix elements can convey numerical problems like overflow when considering the time span of an engine.

Example 4.1

A model that contains a parameterized function describing a linearly interpolated engine map is used to illustrate the linear growth in the covariance estimate P in an EKF. Analyzing the equations describing the feedback gain and estimation error covariance propagation give insight into the covariance growth.

$$K_t = P_{p,t-1} C_t^T (C_t P_{p,t-1} C_t^T + R)^{-1} \quad (4.3a)$$

$$P_{u,t} = P_{p,t-1} - P_{p,t-1} C_t^T (C_t P_{p,t-1} C_t^T + R)^{-1} C_t P_{p,t-1} \quad (4.3b)$$

$$= P_{p,t-1} - K_t C_t P_{p,t-1} \quad (4.3c)$$

$$P_{p,t} = A_t P_{u,t} A_t^T + Q \quad (4.3d)$$

Here K_t is the Kalman gain, and $P_{u,t}$ and $P_{p,t}$ are the estimation error covariance estimates in the measurement update and prediction step of the EKF respectively. The potential problem of locally unobservable parameter states is demonstrated using diagonal P_0 , Q , R and a linear map interpolation in the measurement equation. Here it can be concluded that the positions in the covariance matrix corresponding to areas in the system operating region where the system seldom operates will increase linearly with time.

Consider a state vector

$$\bar{x} = (x^T \quad \theta_o^T \quad \theta_u^T)^T,$$

where x is the state vector in the original model, θ_o are parameter states that have been observed, and θ_u are the parameter states not yet observed. Start in time step k with

$$P_{p,k-1} = \begin{pmatrix} *^{n_x \times n_x} & 0 & 0 \\ 0 & *^{n_{\theta_o} \times n_{\theta_o}} & 0 \\ 0 & 0 & D_k \end{pmatrix}, \quad (4.4a)$$

$$C_k = \begin{pmatrix} *^{n_y \times n_x} & * & 0 \end{pmatrix} \quad (4.4b)$$

$$A_k = \begin{pmatrix} *^{n_x \times n_x} & 0 \\ 0 & I^{n_{\theta} \times n_{\theta}} \end{pmatrix}, \quad (4.4c)$$

where D on the diagonal of $P_{p,k}$ is a diagonal matrix corresponding to parameter states not yet observed and the zero column of C_k corresponds to parameters not used in the interpolation in time step k , i.e., with

$$y = h(x) + q_{\text{fcn}}(x, \theta),$$

$$C_k = \left(\frac{\partial h}{\partial x} + \frac{\partial q_{\text{fcn}}}{\partial x} \quad \frac{\partial q_{\text{fcn}}}{\partial \theta_o} \quad \frac{\partial q_{\text{fcn}}}{\partial \theta_u} \right) \Big|_{\bar{x}=\bar{x}_k} = \begin{pmatrix} * & * & 0 \end{pmatrix}$$

which can be compared to Figure 4.1. Hereafter, the dimensions of the matrix sub-blocks are left out to increase readability.

Now (4.3a) becomes

$$K_k = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} \begin{pmatrix} * \\ * \\ 0 \end{pmatrix} (*) = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} \begin{pmatrix} * \\ * \\ 0 \end{pmatrix} = \begin{pmatrix} * \\ * \\ 0 \end{pmatrix}, \quad (4.5)$$

that is, due to the zero on row three in K_k , there is no update of the unobservable states θ_u . Further, (4.3b) becomes

$$\begin{aligned} P_{u,k} &= \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} - \begin{pmatrix} * \\ * \\ 0 \end{pmatrix} \begin{pmatrix} * & * & 0 \end{pmatrix} \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} \\ &= \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} - \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} \\ &= \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} - \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix}, \quad (4.6) \end{aligned}$$

i.e., no reduction of estimation error covariance for the unobserved parameter states. Nevertheless, due to (4.3d) a linear increase in estimation error covari-

ance is obtained

$$P_{p,k} = \begin{pmatrix} * & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} \begin{pmatrix} * & 0 \\ 0 & I \end{pmatrix}^T + Q = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & D_k \end{pmatrix} + Q. \quad (4.7)$$

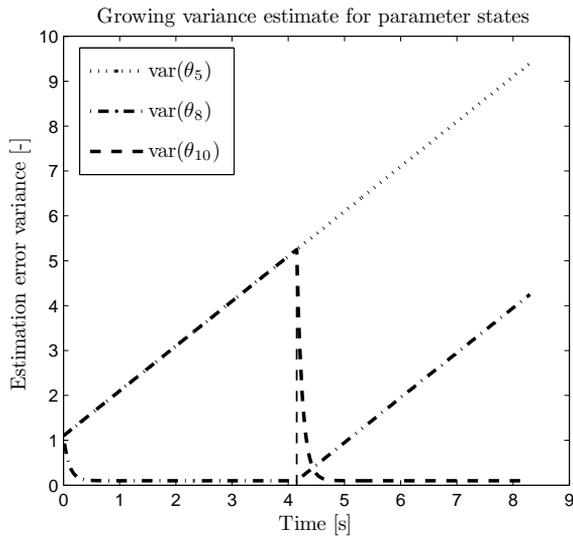
This effect is also illustrated in Figure 4.3, where the variance of three parameter states are plotted, θ_5 , θ_8 , and θ_{10} . They illustrate how the variance of each parameter state develops with time for three cases. θ_5 corresponds to a parameter that is not observed at all for the particular trajectory, presented in Figure 4.3(b). While the trajectory is such that the parameter corresponding to θ_8 is observable during the first half of the trajectory and is then becoming unobservable where instead the parameter corresponding to θ_{10} becomes observable for the second half of the trajectory, which is seen in Figure 4.3(a). Note that Figure 4.3(a) is constructed, and the purpose is to illustrate what happens when the sets of locally observable and unobservable parameter states are changed. This is similar to what will be shown later in Figure 4.9 where the variance propagation is shown for an augmented observer that is run on experimental data.

In SLAM this is, as stated above, not a problem since the map features are considered to be fixed, which is achieved by setting the system noise of the map features equal to zero. In this way the estimation error covariance for these parameters will be limited on a stationary level decided by the measurement noise and ordinary system noise only. In an engine application however it is desirable to be able to track for example system aging, which is not possible if there is no noise driving the engine map parameters.

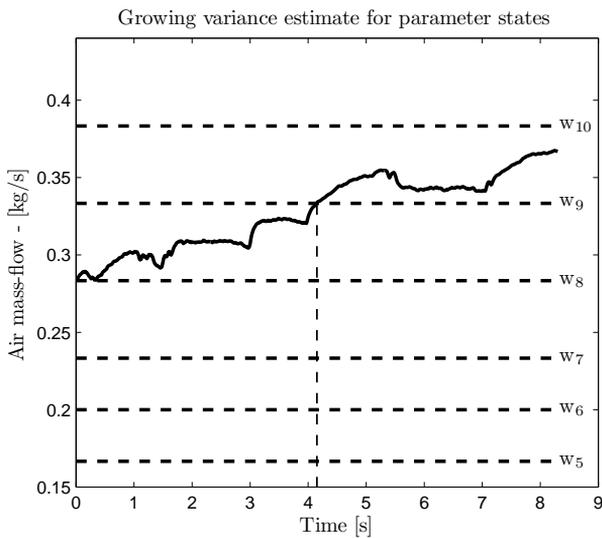
One idea of how to handle the linear growth of estimation covariance of locally unobservable parameter states is to introduce an upper limit for the corresponding estimation error covariance matrix elements. A possible upper limit is the initializing error covariance matrix, P_0 . Since it is desirable to limit the estimation error covariance of only those parameters that are locally unobservable it is appropriate to do the limitation element wise, i.e., compare $P_{i,i}$ to $P_{0,i,i}$, and limiting $P_{i,i}$ by setting $Q_{i,i} = 0$ if $P_{i,i} > P_{0,i,i}$. By choosing P_0 as an upper limit, the introduction of yet another filter tuning parameter is avoided.

One may think that by just analyzing the diagonal elements, $P_{i,i}$, the information of the parameter corresponding to the i :th diagonal element of P is incorrect due to possible cross correlations in P . However, it is straightforward to show that the off-diagonal elements in P do not affect the estimation error covariance for a single parameter, which is done in Appendix C.

Further, this property of the EKF and locally unobservable states can solve some of the engine map adaptation problems that control engineers struggle with in modern engine control systems. For example, experiences from adaptive maps in engine applications, not using the EKF/UKF and joint parameter and state estimation, are for example how to handle the fact that an engine during normal



(a) Parameter variance.



(b) Air mass-flow trajectory.

Figure 4.3: The figure shows the development of the estimation error variance for three parameter states. One that is unobservable during the entire trajectory – θ_5 , one that is observable for the first half of the trajectory – θ_8 , and one that is observable for the second half of the trajectory – θ_{10} .

operation does not cover the entire parameter space used to span the map and only occasionally enters some areas. Many of today's adaptation schemes apply the same adaptation algorithm in each update step and do not adjust the update procedure with respect to when the parameter was last updated. This may result in undesired system behavior caused by old parameters, not handled by the adaptation scheme. In cases like this, a linearly growing uncertainty for seldom updated parameters could be helpful, enabling a fast parameter update without risking large errors in the state estimates.

4.3 Noise Sensitivity

Even though the method developed in this chapter is quite similar to the one developed in Chapter 3, for example new states are introduced to compensate for stationary biases, there are some differences. One difference between the observer developed and analyzed in this chapter compared to that treated in Chapter 3 is the time scale in which the bias and parameter states have to be updated.

For the observer designed in Chapter 3 it is necessary for the bias state to change approximately as fast as the system dynamics, otherwise it will not be able to track a change in operating point. However, a rapidly changing bias state captures also high frequency disturbances which makes the bias state in this approach inappropriate for usage as engine adaptation map estimation.

In an observer utilizing a parameterized function to describe the bias, the parameter states have to work in a time scale determined by system aging, which is substantially longer than for a bias state, and can therefore be made slow. This makes the observer based on a model containing a parameterized function or map less sensitive to temporary disturbances, compared to the observer designed in Chapter 3 since a temporary disturbance does not affect the parameter states significantly. Figure 4.7 presents the true parameterized map, a map estimated with a joint state and parameter estimating EKF, and a bias state augmented EKF from Chapter 3. It is seen that a filter utilizing a parameterized map describing the bias manages to estimate the true map well and that a filter utilizing a bias state in mean captures the adaptation map but with significant variance.

Another issue that is straightforwardly handled by stochastic filters is initialization of unknown bias or function parameters. By proper initialization of the corresponding elements in the estimation error covariance matrix, P_0 , a temporary faster update of unknown bias or function parameter states is achieved. That is, due to an initially faster update rate of unknown parameters, in the same way as old parameters are allowed a faster update rate, a rapid convergence of the parameter states is achieved.

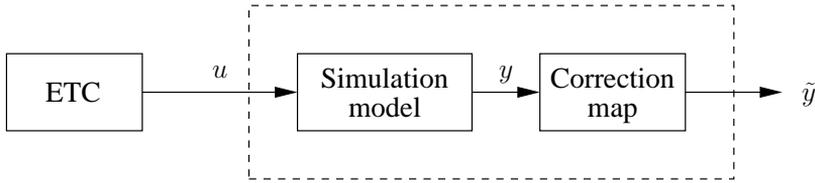


Figure 4.4: Simulation set-up.

4.4 Experimental Evaluation

To evaluate the bias parameterization, a simulation study is performed on a non-linear model of the heavy duty truck engine presented in Appendix A. Inputs to the simulation model are measured signals from an engine in an engine test cell, collected during an ETC.

Three observers are designed and evaluated. The first observer is based on the default model directly, the second observer is based on the idea in Chapter 3, and the third observer utilizes the ideas of a parameterized map describing the bias.

Figure 4.4 shows the simulation set-up used for creating the data used in the observer evaluation. The data is created by simulating the model with input data from an ETC segment, presented in Figure 4.5. The segment is chosen to contain a wide range of air mass-flows such that a trajectory that makes the system observable is created. Apart from the model, a correction map is applied on the simulation output to distort the measured air mass-flow with the aim to mimic the sensor error illustrated by the calibration curve studied in Chapter 2. Further the choice of introducing the bias in the measurement equation shows that the developed methods work, not only for bias in the dynamic equation, but also for bias in the measurement equation. The calibration curve from Chapter 2 corrects for a too low measured air mass-flow W_{meas} compared to the true W_{true} ,

$$W_{\text{true}} = (1 + r(W_{\text{meas}}))W_{\text{meas}}, \quad (4.8)$$

where the typical appearance of r is presented in Figure 2.1. To simulate this incorrect air mass-flow measurement the correction map used in the simulation adjusts the true air mass-flow according to

$$W_{\text{meas}} = (1 + f(W_{\text{true}}))W_{\text{true}}, \quad (4.9)$$

where $f(W_{\text{true}})$ is presented in Figure 4.1. This distorted air mass-flow is then used for feedback in the observers.

All these observers use feedback from the air mass-flow sensor only whilst the estimation performance evaluation is with respect to all states and outputs, i.e., intake and exhaust manifold pressures, turbine speed, and air mass-flow through the compressor.

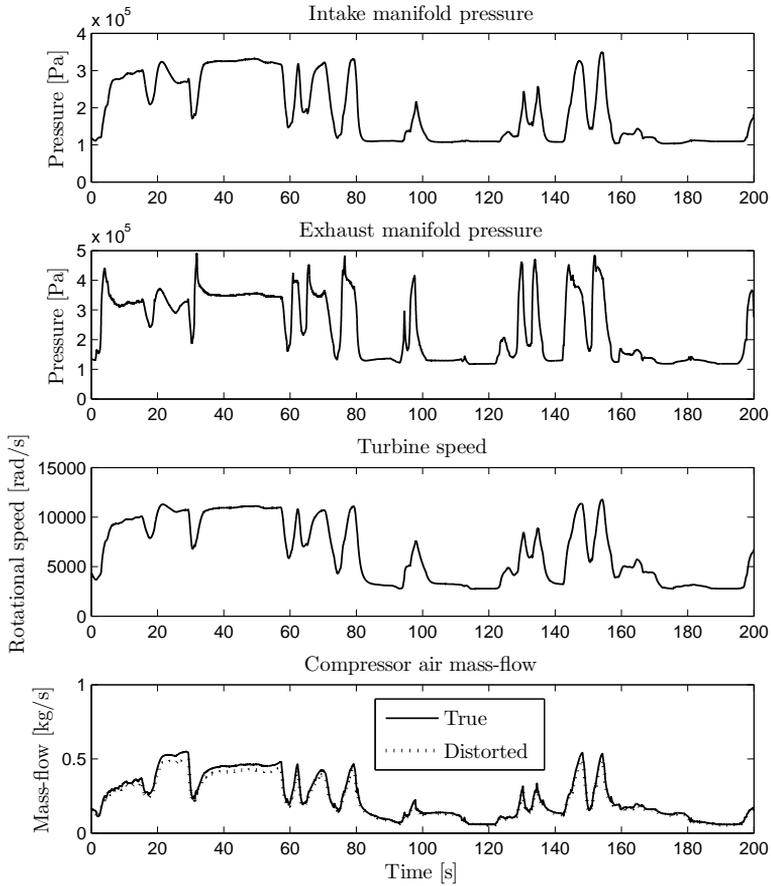


Figure 4.5: Simulated states and outputs from the ETC segment used in the experimental evaluation. The true states and outputs are plotted with solid lines and the distorted output used for feedback is plotted with a dotted line.

4.4.1 Observers

The observer designs are: an EKF observer based on the default model described in Appendix A directly

$$\begin{aligned}x_{t+1} &= x_t + T_s f(x_t, u_t) \\ y_t &= h(x_t),\end{aligned}$$

referred to as Def., an EKF observer with an extra bias state introduced in the measurement equation to reduce the estimation error from the approach developed in Chapter 3

$$\begin{aligned}x_{t+1} &= x_t + T_s f(x_t, u_t) \\ q_{t+1} &= q_t \\ y_t &= h(x_t) + q_t,\end{aligned}$$

referred to as Aug., and a joint state and parameter estimating EKF based on the default model and a parameterized bias,

$$\begin{aligned}x_{t+1} &= x_t + T_s f(x_t, u_t) \\ \theta_{t+1} &= \theta_t \\ y_t &= (1 + q_{\text{fcn}}(h(x_t), u_t, \theta_t))h(x_t),\end{aligned}$$

referred to as Map.

The EKF:s are tuned according to the discussion in Section 4.3 with respect to the driving noise for the bias and parameter states. That is, all three observers have the same measurement noise R , and system noise corresponding to the default states x , i.e., the $n_x \times n_x$ upper left block of Q . While the difference in the bias modeling between the Map. and Aug. observers gives that the $n_q \times n_q$ lower right block in Q corresponding to the Aug. observer is thousand times greater than the $n_\theta \times n_\theta$ lower right block of Q corresponding to the Map. observer. This due to the different time scales in which the bias and parameter states operate. By this tuning the Aug. observer becomes sufficiently fast and is able to track the bias.

4.4.2 Convergence Speed

One issue with the estimation bias correction method developed in Chapter 3 is that since no information about the bias in each operating point is saved the observer convergence is directly dependent on the speed of the bias states. While the extension presented in this chapter use a parameterized function and the convergence will therefore not be dependent on the speed of the parameter states.

To analyze and compare the convergence speed of filters utilizing a parameterized bias to filters utilizing bias states is a difficult task. A reason for this is

that the convergence is highly dependent on the filter tuning. Since the maximum errors occur in transients, see for example the transient at the time 5 s in Figure 4.6, they give an indication on how well the estimate is able to track transient behavior. That is, a coarse way of analyzing convergence speed is to analyze the maximum absolute errors. Figure 4.6 presents the estimation errors from an ill tuned Aug. observer, where the bias state deliberately has been tuned to be too slow and does not manage to track the change in bias, together with the estimation errors from the Map. and Def. observers. From Section 4.3 where the difference in variance of the driving noise of the augmented states between the methods developed in chapters 3 and 4 is discussed, the convergence speed is just a matter of filter tuning and it is hard to make a fair comparison. Since a bias state has to be approximately as fast as the system dynamics and the parameters as fast as the system aging, filters tuned with these aspects in mind will have about the same performance, which is confirmed in Table 4.1. Nevertheless, since the bias state is allowed to change much faster than the parameter states, a filter utilizing that approach will be more sensitive to disturbances. A filter with a parameterized function with slow parameters does not allow disturbances to affect the estimation of neither model nor parameter states, i.e., have a stronger smoothing effect.

Figure 4.7 shows the true and estimated map, and the correction made by the slowly varying bias from Chapter 3 computed according to

$$f = \frac{\hat{q}_t}{\hat{y}_t - \hat{q}_t}.$$

From this figure it is seen that the Map. observer manages to estimate a correction map out of a cycle without any post processing. Also the correction made by the Aug. observer captures the true map but some post processing, like for example mean value computations, is needed to get a map that can be used for interpolation etc.

4.4.3 Adaptation Map Development

In an application where the method is used for engine map adaptation it is of great importance that the method converges toward the true parameterized map. Figure 4.8 shows the development of the adaptation map over time and shows how it converges as air mass-flows from the entire operating region spans the adaptation space.

In Table 4.1 it is seen that all measures, both maximum and mean, are approximately the same for the Aug. and Map. observers for all system states and the system output, while the Def. observer has significantly larger errors. From this it can be concluded that the estimation performance with respect to the default states and outputs are similar for the two observers Aug. and Map. Though, the Map. observer also estimates a parameterized map which is quite common in modern engine control systems.

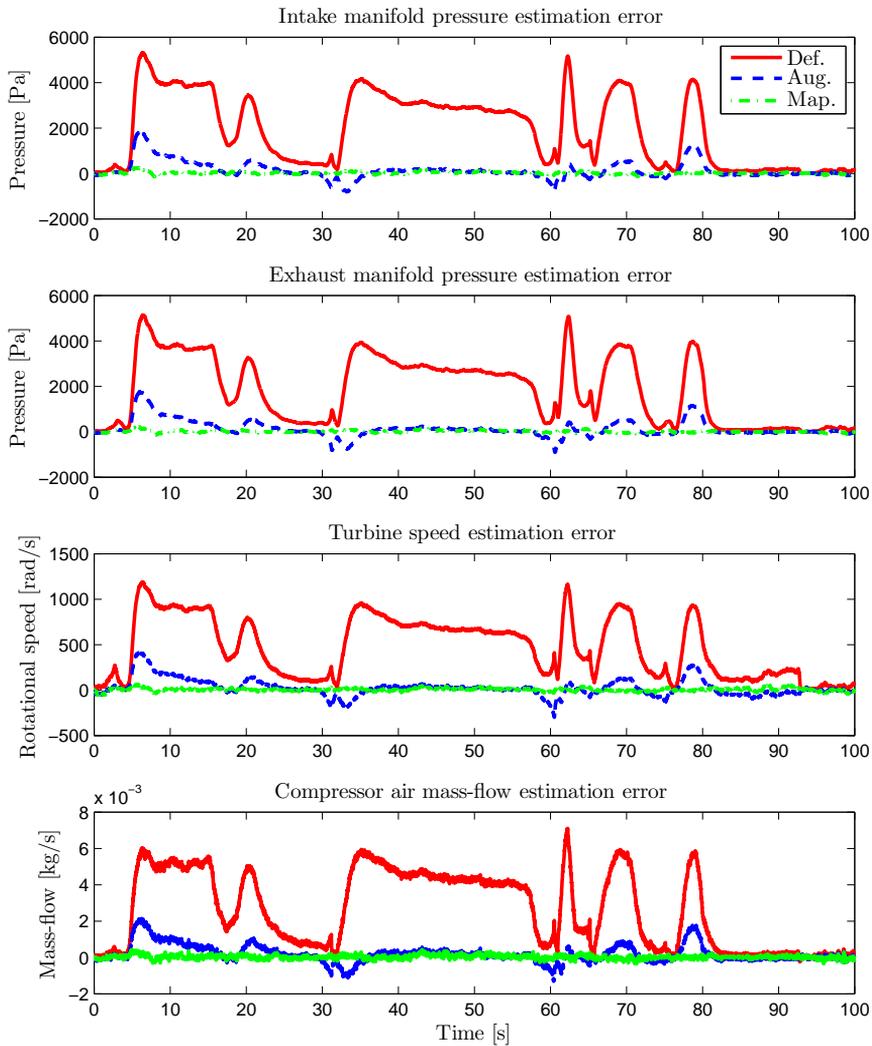


Figure 4.6: Estimation error of an ill tuned Aug. observer, where the bias deliberately is modeled to be too slow which gives large estimation errors in transients, together with estimation errors of the Map. and Def. observers.

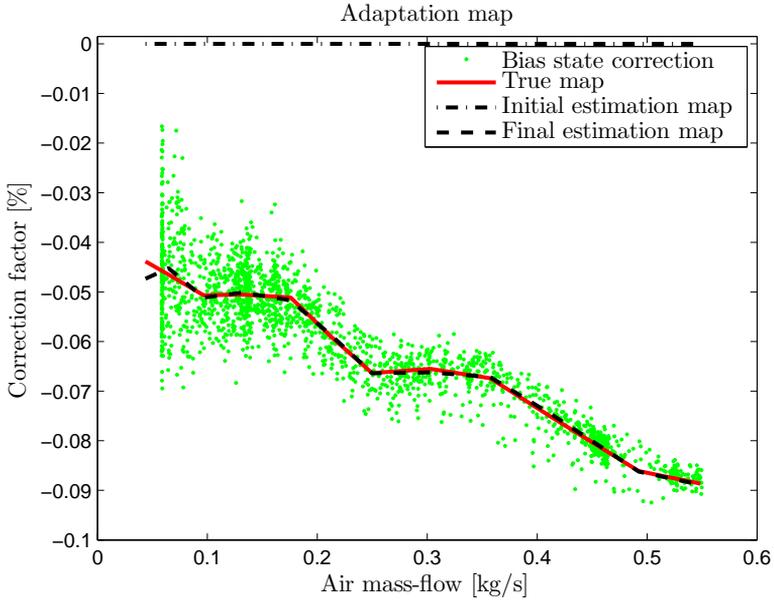


Figure 4.7: True and estimated air mass-flow correction maps together with the correction made by the slowly varying bias from Chapter 3.

Table 4.1: Estimation error for default, augmented and mapped observer respectively, where the filters are tuned according to Section 4.3.

Measurements	Max abs. error			Mean error		
	Def.	Aug.	Map.	Def.	Aug.	Map.
p_{im} [Pa]	5286	264	230	1145	5	4
p_{em} [Pa]	5169	241	205	1086	4	4
n_{trb} [rpm]	1183	99	99	304	-0.5	-0.3
W_{air} [kg/s]	0.007	0.0005	0.0006	0.002	0.00001	0.00001

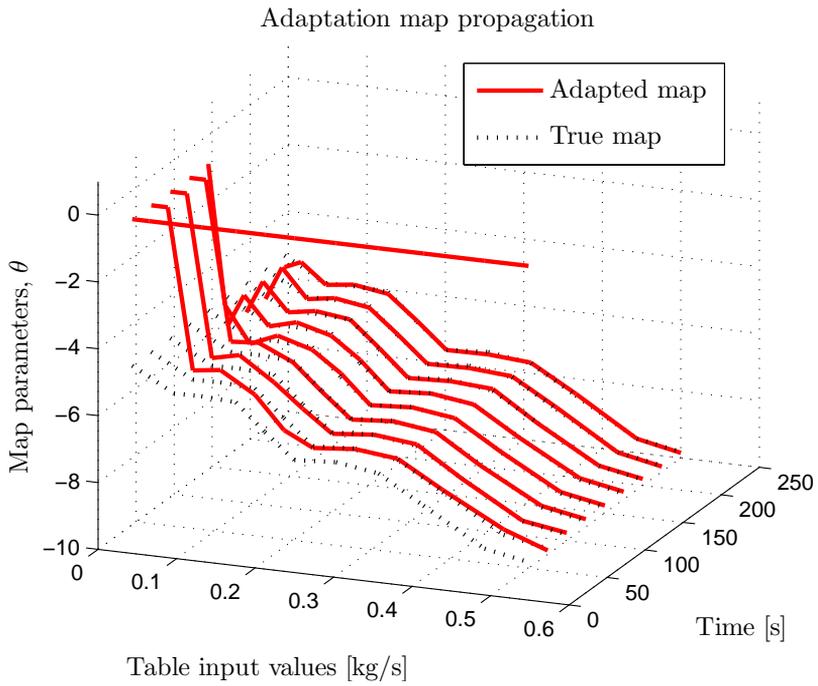


Figure 4.8: Adaptation map propagation

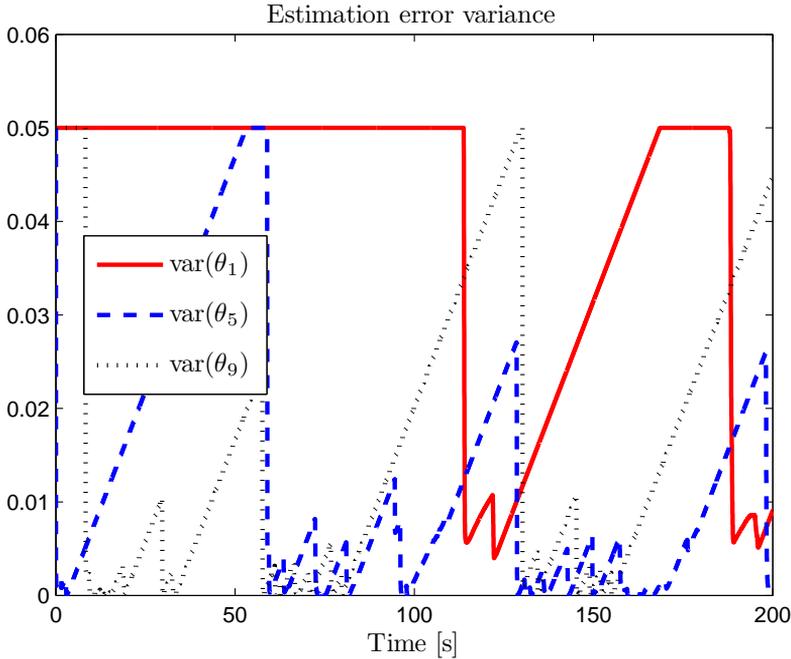


Figure 4.9: Propagation of the variance for three parameter states. Here the variance of the driving noise is significantly increased compared to the filter which estimation error is presented in Table 4.1. This increase in system noise is just for illustrating the linear increase in uncertainty for the locally unobservable parameters and an exponential decrease in uncertainty when the parameters become locally observable.

Figure 4.9 shows how the estimation error covariance is updated. Notice that the estimation error covariance matrix element corresponding to a locally unobservable parameter grows approximately linearly until it reaches the upper bound defined by P_0 , according to Section 4.2.2. However, for short time periods, like the ones simulated here, this linear growth of estimation error covariance is hardly a problem whilst it might become a problem when considering the entire life-time of an engine. In the figure the parameter noise covariance has been increased by a factor of 10^5 to clearly illustrate the growth in variance of the locally unobservable parameters.

4.5 Conclusions

The method developed in Chapter 3 does not keep the information collected about the bias in each operating point, i.e., as soon as the operating point

is changed the information about the bias in that point is discarded. In this chapter a method for storing bias information from different operating points is developed. With this method it is possible to achieve simultaneous estimation of original model states and parameters, like for example adaptation of engine maps.

If a stochastic filter is used, like for example an EKF, then the estimation error stochastics, which in an EKF is the estimation error covariance, can be used to achieve variable update rate of the map parameters. For instance by proper tuning of the initialization covariance, a fast initial update of the parameters is achieved. Further, the use of stochastic filters together with a parameterized bias offers a straightforward way to handle the problem of seldom updated parameters in ordinary engine map adaptation algorithms. The original algorithms do not consider when the parameters were last updated, which may result in unwanted system behavior due to old parameters. With a combination of the parameterized bias, a stochastic filter and the fact that a seldom updated parameter has been locally unobservable for some time, the update rate of these old parameters is increased. The linear growth of estimation error covariance, that comes as a result of local unobservability of the parameters, form a potential problem for the filter and a way to limit this growth is provided. This to avoid numerical problems like for example over flow.

Besides the method itself, its relations to simultaneous localization and mapping (SLAM) are discussed, mainly with respect to local observability of the parameters states. In SLAM there is for example no problem of growing estimation error covariance since the map parameters are considered fixed and the system noise of the parameters are set to zero. This is not an option in an engine map adaptation application since its purpose is to track system aging and changes in environmental conditions.

The method shows promising results in a simulation study using an EKF, where it manages to estimate the engine states while at the same time estimating a parameterized air mass-flow adaptation map. Further it is shown that while maintaining the same estimation quality with respect to mean and maximum absolute error, as the method developed in Chapter 3, an engine adaptation map can be estimated as well.

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

The task of estimating fundamental variables needed for control, or diagnosis in heavy duty truck engines is crucial to fulfill customer demands on low fuel consumption and high torque output. This have to be done without significant time delay and with high accuracy to keep the emissions below legislated levels.

It is empirically established that continuous adaptation of the air mass-flow signal in heavy duty diesel engines is needed and it is further concluded that the air mass-flow adaptation is operating point dependent. The sensor correction needed increases for higher air mass-flows and varies with outer conditions such as summer, winter, high altitudes etc.

In model based observer designs the model quality has a crucial effect on the resulting observer. In the engine application studied in Chapter 2 the model deficiency is that the model captures dynamics well but is biased. Further, the stationary errors often varies with operating point.

An idea of how to handle stationary model errors is to augment the model with extra states that describe the bias. Methods for estimating a bias reducing model augmentation are developed which handles operating point dependent stationary errors. The methods utilize a default model that captures the dynamics of the modeled system well but suffers from stationary errors, and measurements from the system. Based on this a low dimension model augmentation is estimated. Any suitable observer design can then utilize the augmented model to get an observer that has reduced stationary errors compared to using the default model directly. An analysis of robustness against model errors, and

noise is also conducted. This analysis indicates that one of the methods is less sensitive to noise than the others.

The main results regarding the augmentation estimation are a theorem stating which bias augmentations are possible while maintaining observability and another theorem stating what augmentations are possible to find through estimation. The latter theorem states that it is the sensor set-up of the system that limits the space in which bias is observable.

When using a pure bias state augmentation to compensate for an operating point dependent bias, like in Chapter 3, the information about the bias in each operating point is discarded as soon as the operating point is changed. A way to better take care of this bias information is to replace the bias state with a parameterized function. In this way the bias information is stored in the function parameters that are introduced as new states instead of the bias state. This approach falls under what is usually called joint parameter and state estimation where some common methods are extended and unscented Kalman filters (EKF/UKF). By applying these stochastic filters the approach can be used as a new way of updating engine adaptation maps. The stochastic part of the filters give a frame-work suitable for controlling the parameter update rate, that is the filters handle system aging and locally unobservable parameter states by increasing the update rate of these seldom updated parameters once they become locally observable.

The first method developed utilize introduction of new bias states to reduce the stationary estimation errors and is shown to give an approximate mean estimation error reduction of 50 % in an engine application using real measurements. In this application the exhaust manifold pressure is estimated using an EKF and feedback from the intake manifold pressure and turbine speed.

In the second method the bias states are replaced by parameterized functions with unknown parameters, and these parameters are then introduced as new states. This method shows promising results in a simulation study where both the original engine states as well as an air mass-flow correction map, like the one discussed in Chapter 2, is estimated. The promising results refers to that the method finds the correction map fast, and that the ideas fits well in established frame works without any need for ad hoc choices.

5.2 Future work

During the development process of this thesis several interesting questions and thoughts have arisen that have not been given their deserved attention, and this section serves as a compilation of the most important of these ideas.

Starting with the question disregarded in Chapter 4, namely how to find a suitable parameterized function $q_{fcn}(x, y, u, \theta)$ and corresponding inputs x , y , and/or u . A simple idea of how to find these is to analyze the correlation between the biases found in Chapter 3 and the states, outputs, and inputs. This however only answers parts of the questions namely which inputs should the

parameterized function have. Nevertheless, in the case of using the approach for continuous engine map adaptation, suitable parameterizations are often already found from physical relations or other engineering experiences.

Further, in many applications the model and/or measurements are of different quality depending on system operating point. In these cases it can be desirable to be able to control the update of the parameters, for example in operating points where the measurements are known to be inaccurate they should be ignored and when the model is inaccurate some other filter modification might be appropriate to achieve a globally better performing observer. This can be achieved by introducing time varying system and measurement noise descriptions, Q_t and R_t , and thereby in some sense extend the procedure of limiting the magnitude of the estimation error covariance. Another related term on the same topic that could be useful is Q -boosting, which means a temporal increase in system noise when the model has become invalid.

It would also be interesting to apply the developed methods to other estimation problems both in the vehicle power train, for example turbine speed, intake manifold temperature, as well as in other applications such as for example selective catalytic reduction (SCR) catalysts and diesel particulate filters (DPF).

Throughout this thesis the EKF have been used and some other related filters with different properties are for example the UKF and the particle filter (PF) (Gordon et al., 1993) and it would be interesting to also analyze what possible gains there would be in using one of these instead of the EKF.

REFERENCES

- Andersson, P. and Eriksson, L. (2004). Cylinder air charge estimator in turbocharged SI-engines. In *Electronic Engine Controls*, number 2004-01-1366 in SAE Technical paper series SP-1822.
- Bembenek, C., Chmielewski, T. A. J., and Kalata, P. R. (1998). Observability conditions for biased linear time invariant systems. In *Proceedings of the American Control Conference*, pages 1180–1184, Philadelphia, Pennsylvania.
- Council of European Parliament (2005). Directive 2005/55/EC of the european parliament and of the council of 28 september 2005.
- Durrant-Whyte, H. F. (1988). Uncertain geometry in robotics. *IEEE Journal of Robotics and Automation Magazine*, 4(1):23–31.
- García-Nieto, S., Martínez, M., Blasco, X., and Sanchis, J. (2008). Nonlinear predictive control based on local model networks for air management in diesel engines. *Control Engineering Practice*, 16(12):1399–1413.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings-F Radar and Signal Processing*, volume 140, pages 107–113.
- Hermann, R. and Krener, A. J. (1977). Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, AC-22(5):728–740.
- Heywood, J. B. (1988). *Internal Combustion Engine Fundamentals*. McGraw-Hill, Inc.

- Höckerdal, E., Eriksson, L., and Frisk, E. (2008a). Air mass-flow measurement and estimation in diesel engines equipped with EGR and VGT. In *Electronic Engine Controls*, number 2008-01-0992 in SAE Technical paper series SP-2159, SAE World Congress, Detroit, USA.
- Höckerdal, E., Frisk, E., and Eriksson, L. (2008b). Observer design and model augmentation for bias compensation applied to an engine. IFAC World Congress, Seoul, Korea.
- Höckerdal, E., Frisk, E., and Eriksson, L. (2008c). Observer design and model augmentation for bias compensation with a truck engine application. *Control Engineering Practice*.
- Julier, S. J. and Uhlmann, J. K. (1997). A new extension of the kalman filter to nonlinear systems. In *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, pages 182–193.
- Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3).
- Kailath, T. (1980). *Linear Systems*. Prentice-Hall, Inc, Englewood Cliffs, New Jersey 07632.
- Kailath, T., Sayed, A. H., and Hassibi, B. (2000). *Linear Estimation*. Prentice-Hall, Inc, Upper Saddle River, New Jersey 07458, 2 edition.
- Kalman, R., Ho, B. L., and Narendra, K. (1963). Controllability of linear dynamical systems. *Contributions to Differential Equations*, 1.
- Kopp, R. E. and Orford, R. J. (1963). Linear regression applied to system identification for adaptive control systems. *AIAA*, 1(10):2300–2306.
- Lancaster, P. and Tismenetsky, M. (1984). *The Theory of Matrices*. Academic Press, 2 edition.
- Lee, E. B. and Markus, L. (1968). *Foundations of Optimal Control Theory*. John Wiley & Sons, Inc, New York.
- Lino, P., Maione, B., and Amorese, C. (2008). Modelling and predictive control of a new injection system for compressed natural gas engines. *Control Engineering Practice*, 16(10):1216–1230.
- Peyton Jones, J. C. and Muske, K. R. (2007). Automatic calibration of 1 and 2-d look-up tables using recursive least-squares identification techniques. In *Electronic Engine Controls*, number 2007-01-1343 in SAE Technical paper series SP-2087, SAE World Congress, Detroit, USA.
- Smith, R., Self, M., and Cheeseman, P. (1990). Estimating uncertain spatial relationships in robotics. In Cox, I. J. and Wilfong, G. T., editors, *Autonomous Robot Vehicles*, pages 167–193, New York. Springer-Verlag, Inc.

- Törnqvist, D. (2008). *Estimation and Detection with Applications to Navigation*. Dissertation no. 1216, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.
- Tseng, T.-C. and Cheng, W. K. (1999). An adaptive air/fuel ratio controller for si engine throttle transients. In *Electronic Engine Controls*, number 1999-01-0552 in SAE Technical paper series SP-1419, SAE International Congress and Exposition, Detroit, USA.
- Vidal, R., Chiuso, A., and Soatto, S. (2002). Observability and identifiability of jump linear systems. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 3614–3619.
- Wahlström, J. (2006). Control of EGR and VGT for emission control and pumping work minimization in diesel engines. Technical report. LiU-TEK-LIC-2006:52, Thesis No. 1271.
- Wahlström, J. and Eriksson, L. (2006). Modeling of a diesel engine with VGT and EGR including oxygen mass fraction. Technical Report LiTH-R-2747, Department of Electrical Engineering, Linköpings Universitet, SE-581 83 Linköping, Sweden.
- Wan, E. A., Merwe, R. V. D., and Nelson, A. T. (2000). Dual estimation and the unscented transformation. In *Neural Information Processing Systems*, pages 666–672. MIT Press.
- Wang, J. (2008). Air fraction estimation for multiple combustion mode diesel engines with dual-loop egr systems. *Control Engineering Practice*, 16(12):1479–1486.
- Wu, G. (2006). A table update method for adaptive knock control. In *Electronic Engine Controls*, number 2006-01-0607 in SAE Technical paper series SP-2003, SAE World Congress, Detroit, USA.

PLANT MODEL

The engine model, used throughout this thesis, is a third order non-linear state-space model of an inline six cylinder Scania diesel engine with variable geometry turbo (VGT) and exhaust gas recirculation (EGR). The model states are intake manifold pressure, p_{im} , exhaust manifold pressure, p_{em} , and turbine speed, n_{trb} . The inputs are injected amount of fuel, VGT and EGR positions. Besides these control inputs the model is parameterized by the ambient pressure and temperature, p_{amb} and T_{amb} , and the engine speed, n_e . It is based on a model developed in (Wahlström and Eriksson, 2006) and slightly simplified. The simplifications are that the states for the EGR mass fraction and actuator dynamics are removed. The states for EGR mass fraction are removed since they are unobservable and the actuator dynamics are removed to get a simpler model. A plant overview is presented in Figure A.1.

A.1 Basic Structure

The basic structure of the model is

$$\dot{x} = \begin{pmatrix} \dot{p}_{\text{im}} \\ \dot{p}_{\text{em}} \\ \dot{n}_{\text{trb}} \end{pmatrix} = \begin{pmatrix} f_{p_{\text{im}}}(p_{\text{im}}, p_{\text{em}}, n_{\text{trb}}, u_{\text{egr}}, n_e) \\ f_{p_{\text{em}}}(p_{\text{im}}, p_{\text{em}}, n_{\text{trb}}, u_{\delta}, u_{\text{vgt}}, u_{\text{egr}}, n_e) \\ f_{n_{\text{trb}}}(p_{\text{im}}, p_{\text{em}}, n_{\text{trb}}, u_{\text{vgt}}) \end{pmatrix}$$

$$y = (p_{\text{im}} \quad p_{\text{em}} \quad n_{\text{trb}} \quad h_{W_c}(p_{\text{im}}, n_{\text{trb}}))^T,$$

where $f_{p_{\text{im}}}$, $f_{p_{\text{em}}}$, $f_{n_{\text{trb}}}$, and h_{W_c} are described in Section A.2. Further, the model inputs, outputs, and states used are collected in Table A.1.

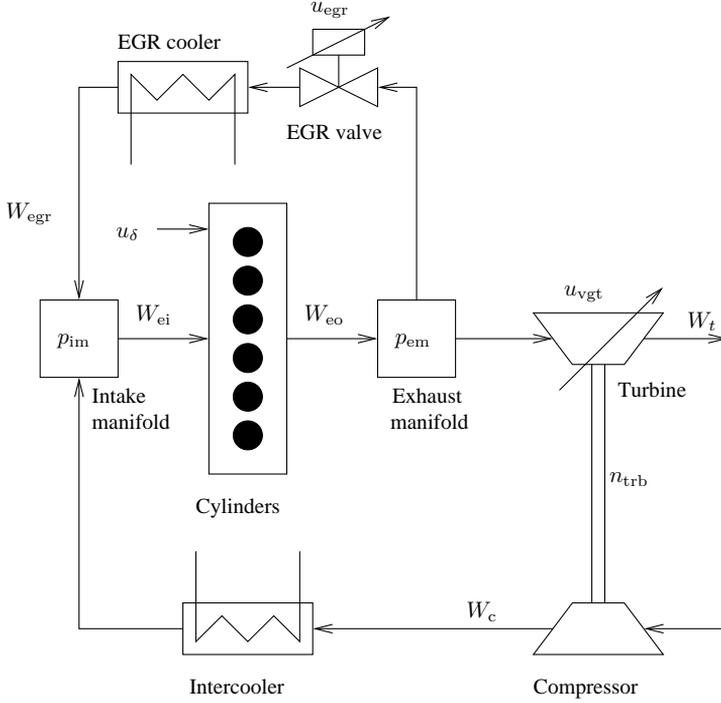


Figure A.1: The model structure of the diesel engine. It has three control inputs, and three states related to the engine. In addition, it has one external parameterization input, the engine speed n_e .

Table A.1: Model inputs and outputs

Inputs	Outputs	States
u_δ	p_{im}	p_{im}
n_e	p_{em}	p_{em}
u_{vgt}	n_{trb}	n_{trb}
u_{egr}	W_c	

A.2 Model Details

Below follow a summary of the model equations using the symbols and indices presented in Table A.2 and Table A.3. More details about the model is found in (Wahlström and Eriksson, 2006).

Manifolds

$$\begin{aligned}\frac{d}{dt} p_{\text{im}} &= \frac{R_a T_{\text{im}}}{V_{\text{im}}} (W_c + W_{\text{egr}} - W_{\text{ei}}) \\ \frac{d}{dt} p_{\text{em}} &= \frac{R_e T_{\text{em}}}{V_{\text{em}}} (W_{\text{eo}} - W_t - W_{\text{egr}})\end{aligned}$$

Cylinder

Cylinder flow

$$\begin{aligned}W_{\text{ei}} &= \frac{\eta_{\text{vol}} p_{\text{im}} n_e V_d}{120 R_a T_{\text{im}}} \\ \eta_{\text{vol}} &= c_{\text{vol1}} \sqrt{p_{\text{im}}} + c_{\text{vol2}} \sqrt{n_e} + c_{\text{vol3}} \\ W_{\text{f}} &= \frac{10^{-6}}{120} u_{\delta} n_e n_{\text{cyl}} \\ W_{\text{eo}} &= W_{\text{f}} + W_{\text{ei}}\end{aligned}$$

Cylinder out temperature

Since the equations used to model the cylinder out temperature are non-linear and dependent on each other, the temperature is computed numerically using a fix point iteration using the equations

$$\begin{aligned}q_{\text{in},k+1} &= \frac{W_{\text{f}} q_{\text{HV}}}{W_{\text{ei}} + W_{\text{f}}} (1 - x_{\text{r},k}) \\ x_{\text{p},k+1} &= 1 + \frac{q_{\text{in},k+1} x_{\text{cv}}}{c_{\text{va}} T_{1,k} r_c^{\gamma_a - 1}} \\ x_{\text{v},k+1} &= 1 + \frac{q_{\text{in},k+1} (1 - x_{\text{cv}})}{c_{\text{pa}} \left(\frac{q_{\text{in},k+1} x_{\text{cv}}}{c_{\text{va}}} + T_{1,k} r_c^{\gamma_a - 1} \right)} \\ x_{\text{r},k+1} &= \frac{\Pi_e^{1/\gamma_a} x_{\text{p},k+1}^{-1/\gamma_a}}{r_c x_{\text{v},k+1}} \\ T_{\text{e},k+1} &= \eta_{\text{sc}} \Pi_e^{1-1/\gamma_a} r_c^{1-\gamma_a} x_{\text{p},k+1}^{1/\gamma_a - 1} \left(q_{\text{in},k+1} \left(\frac{1 - x_{\text{cv}}}{c_{\text{pa}}} + \frac{x_{\text{cv}}}{c_{\text{va}}} \right) + T_{1,k} r_c^{\gamma_a - 1} \right) \\ T_{1,k+1} &= x_{\text{r},k+1} T_{\text{e},k+1} + (1 - x_{\text{r},k+1}) T_{\text{im}}.\end{aligned}$$

For each time sample the iteration is initialized by setting the initial values $x_{r,0}$ and $T_{1,0}$ to the solutions of x_r and T_1 from the previous time sample.

$$T_{em} = T_{amb} + (T_e - T_{amb}) e^{-\frac{h_{tot} \pi d_{pipe} l_{pipe} n_{pipe}}{W_{eo} c_{pe}}}$$

EGR-valve

$$W_{egr} = \frac{A_{egr} p_{em} \Psi_{egr}}{\sqrt{T_{em}} R_e}$$

$$\Psi_{egr} = 1 - \left(\frac{1 - \Pi_{egr}}{1 - \Pi_{egropt}} - 1 \right)^2$$

$$\Pi_{egr} = \begin{cases} \Pi_{egropt} & \text{if } \frac{p_{im}}{p_{em}} < \Pi_{egropt} \\ \frac{p_{im}}{p_{em}} & \text{if } \Pi_{egropt} \leq \frac{p_{im}}{p_{em}} \leq 1 \\ 1 & \text{if } 1 < \frac{p_{im}}{p_{em}} \end{cases}$$

$$A_{egr} = A_{egrmax} f_{egr}(u_{egr})$$

$$f_{egr}(u_{egr}) = \begin{cases} c_{egr1} u_{egr}^2 + c_{egr2} u_{egr} + c_{egr3} & \text{if } u_{egr} \leq -\frac{c_{egr2}}{2 c_{egr1}} \\ c_{egr3} - \frac{c_{egr2}^2}{4 c_{egr1}} & \text{if } u_{egr} > -\frac{c_{egr2}}{2 c_{egr1}} \end{cases}$$

Turbo

Turbo inertia

$$\frac{d}{dt} n_{trb} = \frac{P_t \eta_m - P_c}{J_t n_{trb}}$$

Turbine efficiency

$$P_t \eta_m = \eta_{tm} W_t c_{pe} T_{em} \left(1 - \Pi_t^{1-1/\gamma_e} \right)$$

$$\Pi_t = \frac{p_{amb}}{p_{em}}$$

$$\eta_{tm} = \eta_{tm,max} - c_m (BSR - BSR_{opt})^2$$

$$BSR = \frac{R_t n_{trb}}{\sqrt{2 c_{pe} T_{em} \left(1 - \Pi_t^{1-1/\gamma_e} \right)}}$$

$$c_m = c_{m1} (n_{trb} - c_{m2})^{c_{m3}}$$

Turbine mass flow

$$W_t = \frac{A_{\text{vgtmax}} p_{\text{em}} f_{\Pi_t}(\Pi_t) f_{\text{vgt}}(u_{\text{vgt}})}{\sqrt{T_{\text{em}}}}$$

$$f_{\Pi_t}(\Pi_t) = \sqrt{1 - \Pi_t^{K_t}}$$

$$f_{\text{vgt}}(u_{\text{vgt}}) = c_{f2} + c_{f1} \sqrt{1 - \left(\frac{u_{\text{vgt}} - c_{\text{vgt}2}}{c_{\text{vgt}1}} \right)^2}$$

Compressor efficiency

$$P_c = \frac{W_c c_{\text{pa}} T_{\text{amb}}}{\eta_c} \left(\Pi_c^{1-1/\gamma_a} - 1 \right)$$

$$\Pi_c = \frac{p_{\text{im}}}{p_{\text{amb}}}$$

$$\eta_c = \eta_{\text{cmax}} - \chi^T Q_c \chi$$

$$\chi = \begin{bmatrix} W_c - W_{\text{copt}} \\ \pi_c - \pi_{\text{copt}} \end{bmatrix}$$

$$\pi_c = (\Pi_c - 1)^{\text{pow}_\pi}$$

$$Q_c = \begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix}$$

Compressor mass flow

$$W_c = \frac{p_{\text{amb}} \pi R_c^3 n_{\text{trb}}}{R_a T_{\text{amb}}} \Phi_c$$

$$\Phi_c = \sqrt{\frac{1 - c_{\Psi 1} (\Psi_c - c_{\Psi 2})^2}{c_{\Phi 1}}} + c_{\Phi 2}$$

$$\Psi_c = \frac{2 c_{\text{pa}} T_{\text{amb}} \left(\Pi_c^{1-1/\gamma_a} - 1 \right)}{R_c^2 n_{\text{trb}}^2}$$

$$c_{\Psi 1} = c_{\omega \Psi 1} n_{\text{trb}}^2 + c_{\omega \Psi 2} n_{\text{trb}} + c_{\omega \Psi 3}$$

$$c_{\Phi 1} = c_{\omega \Phi 1} n_{\text{trb}}^2 + c_{\omega \Phi 2} n_{\text{trb}} + c_{\omega \Phi 3}$$

Table A.2: Symbols used in the plant model

Symbol	Description	Unit
A	Area	m^2
BSR	Blade speed ratio	–
c_p	Spec. heat capacity, constant pressure	$\text{J}/(\text{kg} \cdot \text{K})$
c_v	Spec. heat capacity, constant volume	$\text{J}/(\text{kg} \cdot \text{K})$
J	Inertia	$\text{kg} \cdot \text{m}^2$
n_{cyl}	Number of cylinders	–
n_e	Rotational engine speed	rpm
n_{trb}	Rotational speed	rad/s
p	Pressure	Pa
P	Power	W
q_{HV}	Heating value of fuel	J/kg
r_c	Compression ratio	–
R_a, R_y	Gas constant	$\text{J}/(\text{kg} \cdot \text{K})$
R_c, R_t	Radius	m
T	Temperature	K
x_{cv}	Fuel mass fraction burned during constant volume	–
u_{egr}	EGR control signal. 100:open 0:closed	%
u_{vgt}	VGT control signal. 100:open 0:closed	%
u_δ	Injected amount of fuel	mg/cycle
V	Volume	m^3
W	Mass flow	kg/s
γ	Specific heat capacity ratio	–
η	Efficiency	–
Π	Pressure quotient	–
ρ	Density	kg/m^3
Φ_c	Volumetric flow coefficient	–
Ψ_c	Energy transfer coefficient	–

Table A.3: Indices used in the plant model.

Index	Description
a	air
amb	ambient
c	compressor
d	displaced
e	exhaust
egr	EGR
ei	engine cylinder in
em	exhaust manifold
eo	engine cylinder out
f	fuel
im	intake manifold
m	mechanical
t	turbine
vgt	VGT
vol	volumetric
δ	fuel injection

EXPERIMENTAL SET-UP AND DATA

The data are collected in an engine test cell at Scania CV AB in Södertälje, Sweden and are from an inline six cylinder Scania diesel engine with VGT and EGR. Data was collected during a European transient cycle (ETC) defined in (Council of European Parliament, 2005). The sensor signals used in all experimental evaluations are; intake and exhaust manifold pressures, turbine speed, and engine speed. Actuator signals used are; VGT and EGR positions, and injected amount of fuel. All these signals are available on a standard engine, i.e. no extra lab sensors were used, and collected with a sampling rate of 100 Hz.

For the experimental evaluation in Chapter 2 an extra air mass-flow sensor, W_{ref} , is used as a reference. This signal is logged using a different measurement system with a sampling frequency of 10 Hz. For the evaluation in Section 2.4 the measurements are synchronized. The synchronization is made by comparing measurements of the same signal with the two systems and performing a time shift.

Sensor Dynamics

To justify that it is the system dynamics that is captured by in the measurements, i.e., the sensors is fast enough to be able to track the system dynamics, a brief presentation of the sensor data is presented. The sensor specifications are provided by Scania.

The pressure sensors are capacitive pressure sensors and have a first order step response with a time constant of approximately 15 ms for the intake manifold

pressure, and 20 ms for the exhaust manifold pressure.

The mass-flow sensor measuring the air mass-flow through the compressor is a hot wire sensor also with a first order response and a time constant of 20 ms.

The rotational speed sensors are inductive and measures the time between two interrupts. For the engine speed the interrupts occur every sixth crankshaft degree and the signal used throughout this thesis is the mean value of 20 consecutive interrupts. This gives a time constant of approximately $20 \cdot (1 - e^{-1}) \approx 13$ samples. For the turbine speed the interrupts occur ones for every revolution and the signal used throughout this thesis is the median of three consecutive interrupts. That is, the maximum lag is roughly 13 times six crankshaft degrees and 2 times 360 turbine shaft degrees respectively. For engine speeds over 500 rpm, which is the engine idle speed, this gives a maximal time constant of approximately $13 \cdot (500/60 \cdot (360/6))^{-1} = 26$ ms, and for turbine speeds over 20 000 rpm, which is the minimum revolution speed for which the sensor works, this gives a maximum time delay of approximately $2 \cdot (20\,000/60)^{-1} = 6$ ms. Since these sensor responses are significantly faster than the dynamics seen in measurements they are neglected throughout this thesis.

Reference Signal – W_{ref}

The measured reference output W_{ref} is a cell sensor measuring the air mass flow into the engine. It is a hot wire sensor, a Sensyflow P-Tube with type no. 14241-7962638 and a measuring range of 0.055-1.111 kg/s. The uncertainty is less than 1 % of reading and the sensor has a response time of 12 ms. This sensor is placed approximately 4 meters in front of the engine air mass-flow sensor on a straight pipe with a diameter of 0.28 meters. The fact that the reference sensor is mounted in an environment well suited for air mass-flow measurements, i.e., no restrictions on packing etc. have to be fulfilled, makes this sensors reading closer to the actual air mass-flow. The volume and distance between the two sensors give rise to unwanted dynamics. Calculations show that the filling and emptying dynamics from this volume has a time constant of approximately 10 ms and the effect from wave propagation has approximately the same traveling time, which is small in comparison to the time constants of the system. With these facts and the fact that the reference signal has to be manually synchronized with the other signals, these effects are not considered.

STATE MARGINALIZATION

In Section 4.2.2 the problem of growing estimation error covariance for locally unobservable states are discussed. An idea on how to handle this problem is to introduce an upper limit on the estimation error covariance. Since it is desirable to limit only the error covariance of the locally unobservable parameters the limitation is done element wise, that is by comparing $P_{i,i}$ to $P_{0,i,i}$, and limiting $P_{i,i}$ by setting $Q_{i,i} = 0$ if $P_{i,i} > P_{0,i,i}$, where P_0 is the upper limit.

To deduce that it is sufficient to analyze the diagonal elements in P when limiting the estimation error covariance in the EKF it is convenient to introduce

$$P = \begin{pmatrix} P_x & P_{x\theta} \\ P_{x\theta}^T & P_\theta \end{pmatrix}, \quad \bar{x} = \begin{pmatrix} x \\ \theta \end{pmatrix}, \quad \bar{\mu} = \begin{pmatrix} \mu_x \\ \mu_\theta \end{pmatrix},$$

and the variable transformation $\bar{x}_{\theta_i} = T_{\theta_i} \bar{x}$, where

$$T_{\theta_i} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ I_{n_x} & 0 & 0 & 0 \\ 0 & I_{\theta_{j < i}} & 0 & 0 \\ 0 & 0 & 0 & I_{\theta_{j > i}} \end{pmatrix},$$

to move θ_i to the top of the state vector.

Then the probability density function for θ_i can be computed according to

$$\begin{aligned}
f_{\Theta_i}(\theta_i) &= \int_{x, \theta_{j \neq i}} \frac{e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T P^{-1}(\bar{x}-\bar{\mu})}}{(2\pi)^{(n_x+n_\theta)/2} \sqrt{\det(P)}} dx d\theta_{j \neq i} \\
&= \int_{x, \theta_{j \neq i}} \frac{e^{-\frac{1}{2} \overbrace{(T_{\theta_i}^{-1} T_{\theta_i}(\bar{x}-\bar{\mu}))^T}_{=I} P^{-1} T_{\theta_i}^{-1} T_{\theta_i}(\bar{x}-\bar{\mu})}}{(2\pi)^{(n_x+n_\theta)/2} \sqrt{\det(P)}} dx d\theta_{j \neq i} \\
&= \int_{x, \theta_{j \neq i}} \frac{e^{-\frac{1}{2}(\bar{x}_{\theta_i}-\bar{\mu}_{\theta_i})^T \overbrace{(T_{\theta_i} P T_{\theta_i}^T)^{-1}}^{\Sigma_i} (\bar{x}_{\theta_i}-\bar{\mu}_{\theta_i})}}{(2\pi)^{(n_x+n_\theta)/2} \sqrt{\det(P)}} dx d\theta_{j \neq i} \\
&= \int_{x, \theta_{j \neq i}} \frac{e^{-\frac{1}{2}(\bar{x}_{\theta_i}-\bar{\mu}_{\theta_i})^T \Sigma_i^{-1} (\bar{x}_{\theta_i}-\bar{\mu}_{\theta_i})}}{(2\pi)^{(n_x+n_\theta)/2} \sqrt{\det(\Sigma_i)}} dx d\theta_{j \neq i}, \tag{C.2}
\end{aligned}$$

where the change of variables $\bar{x}_{\theta_i} = T_{\theta_i} \bar{x}$ have been used.

Now setting

$$\Sigma_i^{-1} = \begin{pmatrix} \Sigma_{\theta_i} & \Sigma_{\theta_i \zeta} \\ \Sigma_{\theta_i \zeta}^T & \Sigma_{\zeta} \end{pmatrix}^{-1} = M_i = \begin{pmatrix} M_{\theta_i} & M_{\theta_i \zeta} \\ M_{\theta_i \zeta}^T & M_{\zeta} \end{pmatrix},$$

and

$$\bar{x}_{\theta_i} - \bar{\mu}_{\theta_i} = q_i = \begin{pmatrix} \theta_i - \mu_{\theta_i} \\ \zeta \end{pmatrix},$$

$$\begin{aligned}
(\bar{x}_{\theta_i} - \bar{\mu}_{\theta_i})^T \Sigma_i^{-1} (\bar{x}_{\theta_i} - \bar{\mu}_{\theta_i}) &= q_i^T M_i q_i = \\
&= (\theta_i - \mu_{\theta_i})^T M_{\theta_i} (\theta_i - \mu_{\theta_i}) + (\theta_i - \mu_{\theta_i})^T M_{\theta_i \zeta} \zeta + \zeta^T M_{\theta_i \zeta}^T (\theta_i - \mu_{\theta_i}) + \zeta^T M_{\zeta} \zeta
\end{aligned}$$

(C.2) becomes

$$\begin{aligned}
f_{\Theta_i}(\theta_i) &= e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T M_{\theta_i} (\theta_i - \mu_{\theta_i})} \times \\
&\int_{x, \theta_{j \neq i}} \frac{e^{-\frac{1}{2}((\theta_i - \mu_{\theta_i})^T M_{\theta_i \zeta} \zeta + \zeta^T M_{\theta_i \zeta}^T (\theta_i - \mu_{\theta_i}) + \zeta^T M_{\zeta} \zeta)}}{(2\pi)^{(n_x+n_\theta)/2} \sqrt{\det(\Sigma_i)}} dx d\theta_{j \neq i} \tag{C.3}
\end{aligned}$$

which by completing the squares,

$$\begin{aligned}
(\theta_i - \mu_{\theta_i})^T M_{\theta_i \zeta} \zeta + \zeta^T M_{\theta_i \zeta}^T (\theta_i - \mu_{\theta_i}) + \zeta^T M_{\zeta} \zeta &= (\zeta - \hat{\zeta})^T M_{\zeta} (\zeta - \hat{\zeta}) + C \\
&= \zeta^T M_{\zeta} \zeta - \zeta^T M_{\zeta} \hat{\zeta} - \hat{\zeta}^T M_{\zeta} \zeta + \hat{\zeta}^T M_{\zeta} \hat{\zeta} + C
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \\
M_{\theta\zeta}^T(\theta_i - \mu_{\theta_i}) &= -M_{\zeta}\hat{\zeta} \\
(\theta_i - \mu_{\theta_i})^T M_{\theta\zeta} &= -\hat{\zeta}^T M_{\zeta} \\
\hat{\zeta}^T M_{\zeta}\hat{\zeta} + C &= 0 \\
&\Rightarrow \\
C &= -(\theta_i - \mu_{\theta_i})^T M_{\theta\zeta} M_{\zeta}^{-1} M_{\theta\zeta}^T (\theta_i - \mu_{\theta_i}),
\end{aligned}$$

can be written

$$f_{\Theta_i}(\theta_i) = e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T (M_{\theta} - M_{\theta\zeta} M_{\zeta}^{-1} M_{\theta\zeta}^T) (\theta_i - \mu_{\theta_i})} \times \quad (\text{C.4})$$

$$\begin{aligned}
&\int_{x, \theta_{j \neq i}} \frac{e^{-\frac{1}{2}(\zeta - \hat{\zeta})^T M_{\zeta} (\zeta - \hat{\zeta})}}{(2\pi)^{(n_x + n_{\theta})/2} \sqrt{\det(\Sigma_i)}} dx d\theta_{j \neq i} \\
&= \frac{e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T (M_{\theta} - M_{\theta\zeta} M_{\zeta}^{-1} M_{\theta\zeta}^T) (\theta_i - \mu_{\theta_i})}}{(2\pi)^{(n_x + n_{\theta})/2} \sqrt{\det(\Sigma_i)}} \frac{(2\pi)^{(n_x + n_{\theta} - 1)/2}}{\sqrt{\det(M_{\zeta})}} \\
&= \frac{e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T (M_{\theta} - M_{\theta\zeta} M_{\zeta}^{-1} M_{\theta\zeta}^T) (\theta_i - \mu_{\theta_i})}}{\sqrt{2\pi} \sqrt{\det(\Sigma_i)} \sqrt{\det(M_{\zeta})}}. \quad (\text{C.5})
\end{aligned}$$

Finally, using the equalities

$$\begin{aligned}
\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} &= \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}, \\
\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} &= \det(A - BD^{-1}C) \det(D),
\end{aligned}$$

and

$$\det(A^{-1}) = \det(A)^{-1}, \quad \text{if } \det(A) \neq 0,$$

(C.4) becomes

$$\begin{aligned}
f_{\Theta_i}(\theta_i) &= \frac{e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T (M_\theta - M_{\theta_\zeta} M_\zeta^{-1} M_{\theta_\zeta}^T) (\theta_i - \mu_{\theta_i})}}{\sqrt{2\pi}} \times \\
&\quad \frac{\sqrt{\det(M_\theta - M_{\theta_\zeta} M_\zeta^{-1} M_{\theta_\zeta}^T) \det(M_\zeta)}}{\sqrt{\det(M_\zeta)}} \\
&= e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T (M_\theta - M_{\theta_\zeta} M_\zeta^{-1} M_{\theta_\zeta}^T) (\theta_i - \mu_{\theta_i})} \sqrt{\frac{\det(M_\theta - M_{\theta_\zeta} M_\zeta^{-1} M_{\theta_\zeta}^T)}{\sqrt{2\pi}}} \\
&= \frac{e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T (M_\theta - M_{\theta_\zeta} M_\zeta^{-1} M_{\theta_\zeta}^T) (\theta_i - \mu_{\theta_i})}}{\sqrt{2\pi \det\left(\left(M_\theta - M_{\theta_\zeta} M_\zeta^{-1} M_{\theta_\zeta}^T\right)^{-1}\right)}} \\
&= \frac{e^{-\frac{1}{2}(\theta_i - \mu_{\theta_i})^T \Sigma_{\theta_i}^{-1} (\theta_i - \mu_{\theta_i})}}{\sqrt{2\pi \det(\Sigma_{\theta_i})}}. \tag{C.6}
\end{aligned}$$

Which shows that $\theta_i \sim \mathcal{N}(\mu_{\theta_i}, \Sigma_{\theta_i}) = \mathcal{N}(\mu_{\theta_i}, P_{\theta_i})$.

