Evaluation of Observers for Fault Diagnosis on an Automotive Engine

Niten Olofsson

Reg nr: LiTH-ISY-EX-3118
31st October 2000

Evaluation of Observers for Fault Diagnosis on an Automotive Engine

Examensarbete utfört i Fordonssystem
vid Tekniska Högskolan i Linköping
av

Niten Olofsson

Reg nr: LiTH-ISY-EX-3118

Handledare: **Mattias Nyberg**

Examinator: **Mattias Nyberg**

Linköping, 31st October 2000

Avdelning, Institution
Division, Department

Vehicular Systems
Dept. of Electrical Engineering

Datum:
Date:

31st October 2000

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English
☐ ______

Rapporttyp
Report category

☐ Licentiatavhandling
☒ Examensarbete
☐ C-uppsats
☐ D-uppsats
☐ Övrig rapport
☐ ______

ISBN

ISRN

Serietitel och serienummer
Title of series, numbering

ISSN

LiTH-ISY-EX-3118

URL för elektronisk version

`http://www.fs.isy.liu.se/`

Titel:
Title: Evaluation of Observers for Fault Diagnosis on an Automotive Engine

Författare: Niten Olofsson
Author:

Sammanfattning
Abstract

Fault diagnosis is an important area in automotive applications. A diagnosis system is taken forth using the method of structured hypothesis tests. The test quantities used in the hypothesis tests are calculated from prediction error from fault models. A parameter in each model is estimated using a state space observer, where the parameter is viewed upon as an extra state with zero time derivative. Two observer design methods are compared: linearisation of the system in combination with pole placement and an observer with feedback gain only on the fault parameter state. Simulations based on real data from an automotive diesel engine indicates that the method of linearisation in combination with pole placement gives the better diagnosis system of the two tested approaches. However, the method of feedback gain only on the fault parameter is less complex and easier to implement.

Nyckelord
Keywords observer,fault diagnosis,diesel engine

Acknowledgements

I wish to take the opportunity to thank my supervisor Mattias Nyberg for making it a pleasure to work with this project, and for pleasant trips around Germany. Also, I thank the people at DaimlerChrysler, Thomas Stutte, Michael Back, Frank Kirscbaum, Dirk Mehlfeldt, Michael Frölich, Rainer Müller, Jürgen Hoffmann, Markus Butschek, Bernd Schneider, Frank Heister, Martin Burgbacher and Michel Willems (the order is according to my e-mail address book) for biergarten visits, a very pleasant working atmosphere and showing me the good side of Germany. A special thanks goes to Nicole Miller who helped me with the DaimlerChrysler bureaucracy. A thank goes also of course to Maria Andrée. Thanks also to the people at Vehicular System, Linköping University, and especially Erik Frisk, with whom I had many fruitful discussions.

(Thanks also to the German health care system, which made it possible for me to take out my wisdom tooth for free...)

Thank you Lotta, for encouraging me to take the opportunity to go to Germany.

Contents

1 Introduction **1**
1.1 About the thesis 2

2 Introduction to model based diagnosis **3**
2.1 Diagnosis system 4
2.2 Test quantities 5
2.3 Decision structure 5
2.4 Power function 7

3 System model **9**
3.1 Brief description of a modern diesel engine 9
3.2 Model and fault models 10
3.3 Summary of the models 12

4 Test quantities **13**
4.1 Observability 14
4.2 Linearisation in combination with pole placement 14
4.3 Feedback on the fault state only 15
4.4 Calculation of the test quantities 17

5 Experiments **19**
5.1 Experimental setup 19
5.2 Evaluation of the two methods 21
5.3 Fault state feedback observer test quantity 22

5.4 Pole placement observer based test quantity 26
5.5 Comparison 29

6 Conclusion **33**

Chapter 1
Introduction

Diagnosis is an important area in automotive applications for at least two reasons: Legislation demands that faults in components in an engine can be detected. This is due to the fact that the major contribution to the pollution from modern cars is from cars with engines with faulty components. It is also desirable to detect severe faults in order to protect the car from damage. Early detection of faults can lead to avoidance of serious damage of the car. If a fault affects some parameters of the engine, the control system can use this information when determining the control signal, and thus achieving better performance.

In the thesis, methods of model based diagnosis as described in [1] is applied on the intake and exhaust system of a diesel engine. Two observer methods (Chapter 4) will be applied to estimate fault parameters, later to be used in the diagnosis system. The purpose of this study is to compare these two observer design methods applied to the diagnosis problem of the air intake-exhaust systems on a diesel engine.

The method of using observers to estimate a constant parameter is based on the observation that the parameter can be introduced as an extra state, with zero derivative. The two observer design methods are:

Linearisation in combination with pole placement
The system is linearised, and then the observer gain is chosen such that the poles/eigenvalues of the linear observer are some values corresponding to a tradeoff between time response and disturbance rejection.

Feedback gain only on the parameter state
Only the fourth state, the parameter, is observed using an observer gain. The other states are purely simulated.

1.1 About the thesis

In Chapter 2, a short introduction to model based diagnosis is given. The structure of the model of the system, extended with fault models, is then given in Chapter 3. The two observer methods are discussed in Chapter 4, and in Chapter 5, the actual evaluation between the two observer methods is carried out. In the final chapter, Chapter 6, the results of the thesis is presented.

Chapter 2
Introduction to model based diagnosis

In some applications, it is crucial that a process runs under certain conditions for both safety and performance reasons. When those conditions are violated, notification hereof is mandatory so that proper actions can be carried out to prevent damage to the process and its environment. Therefore there is a need for diagnosis. Diagnosis can be divided into three parts:

- fault detection
- fault isolation
- fault size estimation

First of all, it is important to determine if a fault has occurred or not (detection). When a fault is present, the location of the fault is needed (isolation). It is also desired to know, where applicable, the size of the fault (identification). In the literature, the term *Fault detection and isolation* (*FDI*) is often used for the two first parts.

The isolation of a fault can be very crucial for determining what actions are to be taken.

Example 2.1 *Assume that a pump for the cooling water in a nuclear plant breaks down. The flow of water halts, and the temperature is increased. The diagnosis system warns for a broken pump, for stop in the cooling water flow, for increased temperature in the plant, and for 30 less important faults resulting from the stop of water flow. If there is an auxiliary pump, the proper action would be to start it. However, the system triggers on the heating warning and decreases the effect of the* *plant, eventually leading to a full stop. On the other hand, if the diagnosis system could deduce that the behaviour was caused by the faulty pump, the auxiliary pump would be started.*

2.1 Diagnosis system

In Figure 2.1, a general structure for a diagnosis system is shown. A plant, affected by disturbances and maybe also faults, is observed during its operation. The input u and the output y are collected, and they are fed to a diagnosis system. The diagnosis system then makes a diagnosis statement. The diagnosis statement shows

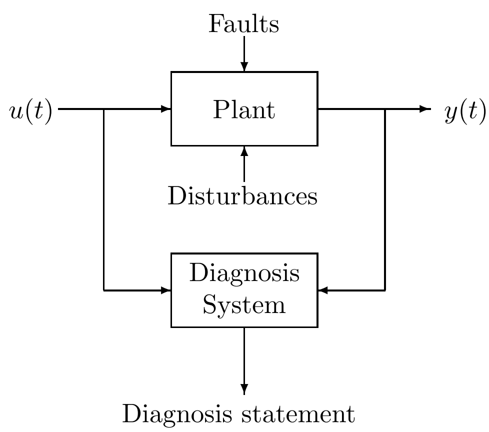

Figure 2.1 *General structure of a diagnosis application.*

in which *fault mode* [1] the system can be operating. The fault mode says which fault is present. For instance, if there are no faults, the fault mode is “No fault mode”. If there is a sensor fault, the fault mode is “Sensor fault mode”. Note that “No fault” is considered to be a fault mode.

Summarising the discussion so far, it is desired to find an explanation that agrees with the collected data of the system. The diagnosis statement is the chosen explanation of the observed behaviour. In order to find such an explanation, a model of the system is needed. For every fault mode the diagnosis system is to detect, a fault model is derived.

Every fault model is fed with the observed data, and the question “Can the observed data be explained by the model?” is posed. This is actually a hypothesis test, known from statistics [2]. Denote the set of all known fault modes by

$$\Omega = \{NF, F_1, F_2, \dots, F_n\}$$

and denote the present fault mode with F_p. Assume the null hypothesis is [1]

$$H_0 : F_p \in \{NF, F_i\}. \tag{2.1}$$

If H_0 is rejected, then the alternative hypothesis

$$H_1 : F_p \in \bigcup_{j \neq i} \{F_j\} \tag{2.2}$$

is accepted. However, if H_0 is *not* rejected, nothing can, without further investigations, be said about which fault mode explains the observed data. For example, if a fault affects a derivative in the system, and the system is working in a stationary operating point, then the fault will not affect the system. Therefore, when H_0 is not rejected, Ω is said to explain the observed data. In this thesis, the hypotheses are always on the form

$$\begin{aligned} H_0^i : F_p &\in \{NF, F_i\} \\ H_1^i : F_p &\in \textstyle\bigcup_{j \neq i} \{F_j\} \end{aligned}$$

If there are model uncertainties present, then the fault models model usually only NF and one other fault F_i. Otherwise, there is a risk that the fault models become over-parameterised. For example, if a model models $\{NF, F_1, F_2, F_3, \}$, but the fault mode is F_4, then the model might have enough parameters to variate to compensate for the faulty behaviour of the system, and the null hypothesis is not rejected.

2.2 Test quantities

In order to perform the hypothesis tests, a test quantity, $T(x)$, based on the observed data $x = (u\ y)^T$ is required. If the value of the test quantity is outside some predetermined range, then the null hypothesis is rejected. From a modelling point of view, a test quantity is a model validity measure. There are several methods to design test quantities and some of them are described below.

Prediction error: The model is fed with the observed data and predicts the output y with $\hat{y}$. Then some measure is used on the prediction error $(y - \hat{y})$. For example, the measure can be the mean effect, the mean energy or the average of $(y - \hat{y})$.

Estimation: Assume that in a model there is a parameter, with nominal value θ_0. The parameter is estimated, for example with maximum likelihood or least squares methods. If $|\theta - \theta_0| > \epsilon$, for some ϵ, then the null hypothesis is rejected.

2.3 Decision structure

The next question that arises is how the faults affect the test quantities. The answer is seen in the alternative hypothesis (the complement of the null hypothesis). If a null hypothesis is rejected, then one or some of the fault models listed in the alternative hypothesis is said to explain the observed data. Let S_i denote the result of the i:th hypothesis test, that is

$$S_i = \begin{cases} S_i^0 & = \Omega \text{ when } H_0^i \text{ is } \textit{not} \text{ rejected} \\ S_i^1 & = \bigcup_{j \neq i} \{F_j\} \text{ when } H_0^i \text{ is rejected} \end{cases} \tag{2.3}$$

Since the S_i:s holds the result of all the hypothesis tests, the diagnosis statement is

$$S = \bigcap_i S_i. \tag{2.4}$$

Another interpretation of the faults listed in the alternative hypothesis, is that those faults may affect the test quantity so that that it is outside its predetermined range, but not necessarily always. For bookkeeping reasons, or to obtain a graphical view of the hypothesis tests structure, the null and alternative hypothesis can be displayed in a decision structure[1] shown in Figure 2.2. In the top of the decision

	NF	F_1	F_2	F_3
$T_1(x)$	0	0	X	0
$T_2(x)$	0	0	X	X
$T_3(x)$	0	X	0	1

Figure 2.2 *Example of a decision structure*

structure, NF and all the faults are listed. To the left, the test quantities are listed. In the intersection between the column of fault mode F_i and the row of test quantity T_j, an X is put if the fault mode F_i is both in the null and the alternative hypothesis. If there is no doubt that the fault F_i affects the test quantity when present, i.e.

$$F_i \notin S_j^0 \tag{2.5}$$

when the null hypothesis is *not* rejected, then the X is replaced by the number 1. This means that the fault mode always affects the test quantity when present.

Example 2.2 *If the the test quantities T_2 and T_3 in the diagnosis system described by Figure 2.2 indicates that H_0^2 and H_0^3 should be rejected, then the diagnosis statement is*

$$S = \Omega \bigcap \{F_2, F_3\} \bigcap \{F_1, F_3\} = F_3. \tag{2.6}$$

On the other hand, if only H_0^2 is rejected, then the diagnosis statement becomes

$$S = \Omega \bigcap \{F_2, F_3\} \bigcap (\Omega \backslash F_3) = F_2. \tag{2.7}$$

2.4 Power function

To determine how well a test quantity responds to a fault, a so called *power function* [3] can be used. The power function is defined as

$$\beta(\theta) = P(\text{reject } H_0 \text{ for the given } \theta) \tag{2.8}$$

$$= P(T_i(x) \in R|\theta) \tag{2.9}$$

where R is the rejection region of the test quantity and θ is the fault value. In Chapter 5, two different test quantities for every fault case will be compared using an estimation of the power function.

Chapter 3

System model

In order to make a diagnosis system using the methodology adopted in Chapter 2, a model of the system and fault models are needed. Starting with a model of a diesel engine with an EGR system, the system, its models and fault models will be presented. These models will be used in Chapter 4 to design the test quantities used by the diagnosis system.

3.1 Brief description of a modern diesel engine

In Figure 3.1, a schematic overview of the air intake and exhaust system of a diesel engine with an EGR system is shown. On the left, air from a compressor (not shown) enters the inlet manifold, passing an air flow sensor (HFM) and a temperature sensor. In the inlet manifold, a sensor measures the inlet manifold pressure. If the engine is equipped with an EGR (exhaust gas recirculation) system, exhaust fumes mix with the air from the compressor. The gas mixture flows into the cylinder, where it is compressed. Fuel (diesel) is injected into the cylinder. Due to the high pressure, the diesel ignition temperature is lower than the temperature in the cylinder, and therefore the diesel ignites. This is called combustion. The burnt gases are then let out of the cylinder into the exhaust manifold. Some of the exhaust are recirculated via the EGR valve to the inlet manifold, and the rest exits at the right to the turbo turbine (not shown) and then to the exhaust system.

To control the combustion, it is required to know the amount of air flowing into the cylinder (deduced from air flow and pressure sensors), the engine speed (measured by a cog wheel) and the amount of fuel injected. In our particular diesel engine, it is also needed to know the amount of recirculated exhaust gases (EGR) for

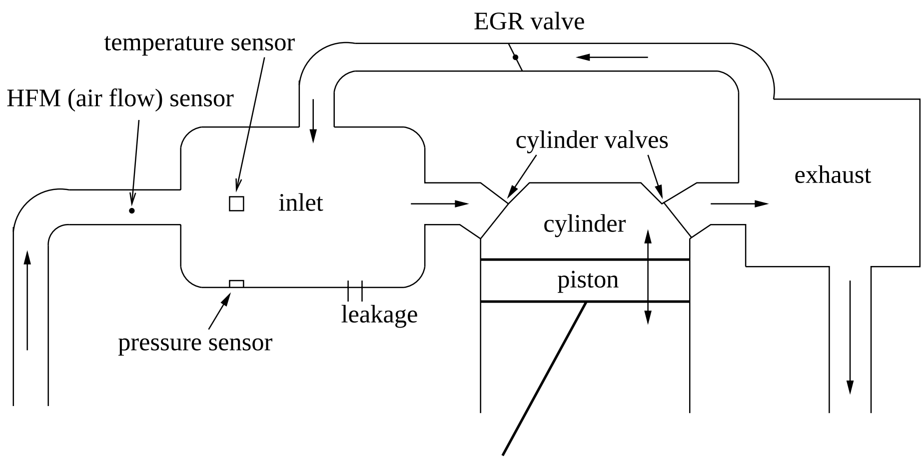

Figure 3.1 *Schematic view of the air intake and exhaust system of a modern diesel engine*

the following reason: in the combustion process, nitrogen oxides, NO_x, are formed. External requirements (law, environmental opinion) put constraints on how much NO_x is allowed to be produced. By recirculating some of the exhaust gases and letting them mix with the intake air, the combustion process will take place at a lower temperature. This lowers the amount of NO_x produced. But if too little or too much exhaust gases are recirculated, less decrement in NO_x production will take place or the engine will not run smoothly.

What to diagnose

The aim of the diagnosis system is to detect and isolate

- leakages in the inlet manifold
- HFM (air flow) sensor fault
- IPS (pressure) sensor fault,

all indicated in Figure 3.1.

3.2 Model and fault models

Faults are treated as if they are stationary, that is, $\dot{\theta}_{fault} = 0$, where θ_{fault} is the fault parameter. The sensors measure input or output signals. Sensor faults are modelled as a gain acting on the "true value", $y_{sensor} = \theta_{sensor-fault} y_{true-value}$, where $\theta_{sensor-fault} \neq 1$ implies a fault. Leakages are considered to affect the pressure and mass build-up in the intake manifold, according to basic equations in fluid dynamics. The fault parameter is viewed upon as an effective leakage area, where $\theta_{leakage} > 0$ denotes a fault.

Basic system model

The original model is described by a system of nonlinear differential equations, given by the general structure

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

$$y = x_1. \tag{3.1b}$$

The states are

$$x = \begin{pmatrix} p_{inlet} \\ m_{inlet} \\ p_{exhaust} \end{pmatrix} = \begin{pmatrix} \text{inlet manifold pressure } [Pa] \\ \text{inlet manifold mass } [kg] \\ \text{exhaust manifold pressure } [Pa] \end{pmatrix} \tag{3.2}$$

The output of the system is the inlet manifold pressure, the first component of the state vector x. The input signals are

$$u = \begin{pmatrix} w_{HFM} \\ N_{engine} \\ \chi_{EGR} \\ \chi_{VGT} \\ \chi_{fuel} \\ T_{inter} \end{pmatrix} = \begin{pmatrix} \text{air flow } [kg/s] \\ \text{engine speed } [revolutions/min] \\ \text{EGR valve opening } (0-1) \\ \text{vane position (on the turbo turbine) } (0-1) \\ \text{injected fuel}[] \\ \text{temperature before inlet manifold} \end{pmatrix} \tag{3.3}$$

Leakage model

Manifold leakage is modelled as an extra constant parameter $\theta_{leakage}$ added to the model. Its effect depends only on the value of $\theta_{leakage}$ and of the state x, and it is described by

$$\dot{x} = f(x, u) + g(x, \theta_{leakage}) \tag{3.4a}$$

$$\dot{\theta}_{leakage} = 0 \tag{3.4b}$$

$$y = x_1 \tag{3.4c}$$

$$\theta_{leakage} \in [0, \inf[, \ \theta_{leakage} > 0 \text{ implies a fault.} \tag{3.4d}$$

If several leakages occur, they are lumped together and treated as one effective leakage.

HFM sensor fault model

Sensor faults are described by (constant) gains. Note that the air flow is considered an input signal, therefore the HFM sensor fault acts on the input signal of the system! The physically true input signal is denoted by u and the measured input signal by u_{sensor}. From the beginning of this section, we know that

$$u_{i,sensor} = \theta_{sensor-fault} u_i. \tag{3.5}$$

However, the only information about the system comes from the sensors, but the real system is affected by the true signal (the real air flow). The problem is thus to compute the number with which to multiply the sensor signal (value) to get the true signal (value). Thus, from (3.5), the basic model equations (3.1) are changed to

$$\dot{x} = f(x, \begin{pmatrix} \theta_{HFM}^{-1} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} u_{sensor}) \tag{3.6a}$$

$$\dot{\theta}_{HFM} = 0 \tag{3.6b}$$

$$y = x_1 \tag{3.6c}$$

$$\theta_{HFM}^{0-1} \in [0, \inf[, \ \theta_{HFM} \neq 1 \text{ implies a fault.} \tag{3.6d}$$

where θ is the fault parameter.

IPS sensor fault model

Regarding the pressure sensor fault, it is seen from the basic model equations (3.1) that only the output is affected by the fault. The pressure sensor fault model thus becomes

$$\dot{x} = f(x, u) \tag{3.7a}$$

$$\dot{\theta}_{IPS} = 0 \tag{3.7b}$$

$$y = \theta_{IPS} x_1 \tag{3.7c}$$

$$\theta_{IPS} \in [0, \inf[, \ \theta_{IPS} \neq 1 \text{ implies a fault.} \tag{3.7d}$$

3.3 Summary of the models

The fault models above have the general model structure

$$z_i = \begin{pmatrix} x \\ \theta_i \end{pmatrix} \tag{3.8a}$$

$$\dot{z}_i = g'(z_i, u) = \begin{pmatrix} g(z_i, u) \\ 0 \end{pmatrix} \tag{3.8b}$$

$$y_i = h(z_i), \tag{3.8c}$$

where $i \in \{HFM, IPS, leakage\}$. This model structure will be used in the development of the observers, discussed in Chapter 4. For the NF model, $z_{NF} = x$ and $g(z_{NF}, u) = f(x, u)$.

Chapter 4
Test quantities

The diagnosis system which will be used on the provided diesel engine, will pose the question "Can the model structure $M(\theta)$ explain the observed data?" rather than "Can the model $M(\theta^*)$, θ^* fixed, explain the observed data?" This will be accomplished by a two stage method. First, the fault parameter θ_i is estimated, and then the prediction error is used to calculate the test quantities, where fault model i uses the estimated θ_i. The test quantity is

$$T(y,u) = \frac{1}{\tau}\int_{t_0}^{t_0+\tau}(y-\hat{y})^2 dt, \tag{4.1}$$

that is, the mean power of the residual $(y-\hat{y})$.

θ_i is estimated using an observer, where the parameter θ_i is viewed upon as an extra state, as in (3.8). A simple and straightforward approach is used for the observer,

$$\dot{\hat{z}} = g(\hat{z},u) + K(y-\hat{y}) \tag{4.2a}$$

$$\hat{y} = h(\hat{z}) \tag{4.2b}$$

$$K = \begin{pmatrix} K_x \\ K_\theta \end{pmatrix}, \tag{4.2c}$$

that is, a vector valued observer gain K is applied on the difference between the estimated and the measured values. The observer gain K is chosen either as the result of a pole placement based on the linearised system or as a feedback term only on the fault state θ_i; in the latter observer, $K_x = 0$. These two methods will be discussed in detail later in this chapter.

4.1 Observability

The system function $f(x, u)$ is partially known analytically, and partially known by measured maps. To compute a K in a systematic manner, numerical methods must therefore be used. Since the system is nonlinear, it must be ensured that representative data are used during the validation of the chosen observer gain.

A concern that should be dealt with first is observability. Is the system observable? What is meant by observability for a nonlinear system? The representation of the model has made it difficult to answer the first question using the analytical tools, Lie derivatives [4, 5] for instance, available in the literature [5], so a pragmatic approach has been taken. The system is tested against representative data, and if the system does what is required of it, then it is said to work. The second question, what is meant by observability, is not dealt with due to the pragmatic approach. For the interested, the meaning of observability for nonlinear system is dealt with in [5] for instance.

4.2 Linearisation in combination with pole placement

Given a linear system (A, B, C, D), the observer is

$$\dot{\hat{\zeta}} = A\hat{\zeta} + Bv + K(\xi - \hat{\xi}) = (A - KC)\hat{\zeta} + Bv + K\xi \tag{4.3a}$$

$$\hat{\xi} = C\hat{\zeta}. \tag{4.3b}$$

The observer gain K can be obtained with the method of *pole placement* [6], that is, the poles of the closed system $(A - KC)$ is chosen. If a linear system has non-observable modes, the poles of those modes cannot be changed. Remember that, if the matrix $(A - KC)$ has all its poles, or eigenvalues, in the left half plane, then the estimation error $(\zeta - \hat{\zeta})$ will asymptotically tend to zero! Assume that the poles of the observable modes are chosen so that they all have negative real parts. The conclusion is then that if a system has non-observable modes, the estimation error $(\zeta - \hat{\zeta})$ will tend to zero if and only if the poles of the non-observable modes all have negative real parts.

It is known that some functions can be expanded in a Taylor series in a neighbourhood of an operating point

$$f(x, u) = f(x_0, u_0) + \frac{\partial f(x_0, u_0)}{\partial x}(x - x_0) + \frac{\partial f(x_0, u_0)}{\partial u}(u - u_0) + \ldots \tag{4.4}$$

If (x_0, u_0) is chosen so that (x_0, u_0) is a stationary point for the system

$$\dot{x} = f(x, u), \tag{4.5}$$

then the system can be approximated, or *linearised*, by

$$\dot{\zeta} = A\zeta + B\upsilon \tag{4.6a}$$
$$\xi = C\zeta + D\upsilon \tag{4.6b}$$
$$\zeta = x - x_0, \ \upsilon = u - u_0 \tag{4.6c}$$
$$\xi = y - y_0 = h(x,u) - h(x_0,u_0) \tag{4.6d}$$
$$A = \frac{\partial f(x_0,u_0)}{\partial x}, \ B = \frac{\partial f(x_0,u_0)}{\partial u} \tag{4.6e}$$
$$C = \frac{\partial h(x_0,u_0)}{\partial x}, \ D = \frac{\partial h(x_0,u_0)}{\partial u} \tag{4.6f}$$

4.3 Feedback on the fault state only

It is valid for any asymptotically stable system [7], that the observer

$$\dot{\hat{x}} = f(\hat{x}, u) \tag{4.7}$$

where u constant (that is, simulation of the system) will tend to the correct value of x, as long as the initial conditions are in the stability region of the system. The diesel engine is assumed to be an asymptotically stable system, the fault modes included. If the system

$$\dot{x} = g((x \ \theta)^T, u) \tag{4.8a}$$
$$\dot{\theta} = 0 \tag{4.8b}$$

is asymptotically stable, it is not farfetched, in view of the above statement, to ask if the system

$$\dot{\hat{x}} = g(\begin{pmatrix} \hat{x} \\ \hat{\theta} \end{pmatrix}, u) \tag{4.9a}$$
$$\dot{\hat{\theta}} = K_\theta(y - \hat{y}) \tag{4.9b}$$
$$\hat{y} = h(\hat{x}, \hat{\theta}) \tag{4.9c}$$

tends asymptotically to the correct values of x and θ.

For the system, and in all the fault models, it holds that $y > 0$ and $\theta \geq 0$. We must now look into the four different fault models to determine their properties.

NF (no fault)

The system is asymptotically stable, and no fault state has been added. The observer is equivalent to simply simulate the system, with wrong initial conditions. However, the system is asymptotically stable when u constant, so the estimation error will probably tend to a small value.

IPS sensor fault

Assume that the $\hat{x}$-part of the system state vector $\hat{z}$ is correct! Then the observer equation becomes

$$\begin{aligned}\dot{\hat{\theta}}_{IPS} &= K_\theta(y_{IPS} - \hat{y}_{IPS}) && (4.10)\\ &= K_\theta(x_1\theta_{IPS} - \hat{x}_1\hat{\theta}_{IPS}) && (4.11)\\ &= K_\theta x_1(\theta_{IPS} - \hat{\theta}_{IPS}) && (4.12)\\ &= K_\theta y(\theta_{IPS} - \hat{\theta}_{IPS}) && (4.13)\end{aligned}$$

Denoting the estimation error with $\theta_e = \theta_{IPS} - \hat{\theta}_{IPS}$, the dynamics of the estimation error is described by

$$\dot{\theta}_e = -Ky\theta_e \quad (4.14)$$

If $\theta_e > 0$, then $\dot{\theta}_e$ must be negative, and vice versa. Since it always holds that $y > 0$, then a necessary condition for the observer to work is that the gain K is positive.

HFM sensor fault

Air and exhausts are flowing into the inlet manifold from the turbo charger and the exhaust via the EGR system. This contributes to a mass build, which leads to a pressure buildup. From the inlet manifold, the gas mixture flows into the cylinder. This gives a mass loss. The net mass buildup, $\dot{m}_{inlet} = w_{HFM} + w_{EGR} - w_{cylinder}$ is the main contributor to the pressure buildup $\dot{y}_{IPS} = \dot{x}_1 = \dot{p}_{inlet}$. When a sensor fault is present, discrepancies between the model and the measured pressure will be seen. However, the true pressure buildup in the inlet manifold is unaffected.

The aim of the observer is thus to compensate for the effect of the sensor fault. From (3.6) it is seen that this is described by

$$\dot{\hat{x}} = f(\hat{x}, \begin{pmatrix} \hat{\theta}_{HFM}^{-1} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} u_{sensor}) \quad (4.15a)$$

$$\dot{\hat{\theta}} = K(y - \hat{y}) \quad (4.15b)$$

$$\hat{y} = \hat{x}_1. \quad (4.15c)$$

When $\theta > 1$, the simulated pressure buildup is greater than the true (measured) pressure buildup. Similarly, the simulated pressure fall is slower than the real (measured) one so when $y < \hat{y}$, $\hat{\theta}$ should increase, that is $\dot{\hat{\theta}} > 0$. A necessary condition for the observer to work is thus that K is negative.

Leakage

Leakages in the inlet manifold affects the true pressure buildup. If the pressure $p = x_1$ is greater than the ambient air pressure ($\approx 100\ kPa$), air and exhaust gases will flow out from the inlet manifold and the pressure will consequently drop. (It happens rarely that the inlet manifold pressure in a diesel engine is lower than the ambient pressure, since diesel engines usually do not have throttles, and have therefore no way to reduce the pressure by letting less air into the inlet manifold.)

The observer equations are derived from (3.4), and are obtained simply by replacing x with $\hat{x}$:

$$\dot{\hat{x}} = f(\hat{x}, u) + g(\hat{x}, \hat{\theta}_{leakage}) \tag{4.16a}$$

$$\dot{\hat{\theta}}_{leakage} = K(y - \hat{y}) \tag{4.16b}$$

$$\hat{y} = \hat{x}_1. \tag{4.16c}$$

Here, $\theta_{leakage} > 0$ denotes a fault, and $\theta_{leakage} = 0$ no fault. When a leakage $\theta_{leakage} > 0$ is present, but the model assumes no leakage, $\hat{\theta}_{leakage} = 0$, the simulated pressure buildup will be higher than the real (measured) one. Then $(y - \hat{y})$ will be negative, but $\dot{\hat{\theta}}_{leakage}$ should be positive, so K must be negative.

4.4 Calculation of the test quantities

As described earlier in this chapter, a fault parameter is estimated using an observer. With this estimated fault parameter fixed, the model is then rerun on the same data set from which the fault parameter was estimated. On this rerun, the test quantity (4.1)

$$T(y, u) = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} (y - \hat{y}(u, \hat{\theta}_i))^2 dt,$$

is calculated.

Chapter 5
Experiments

In the previous chapters, two methods have been described how to perform diagnosis using test quantities and how to design these test quantities. In this chapter, the two methods are tried on real data, and a comparison between the two observer approaches in Chapter 4 are made in perspective of which one gives the overall best performance of the resulting diagnosis statement. The performance measure is the power function defined in (2.8) and an estimate of the power function is used.

5.1 Experimental setup

During the work with this thesis, an engine model of a diesel engine of the type described in Chapter 2 was developed. At the time when the methods of this thesis were investigated, the model was not properly tuned so instead of working directly on data from measurements on a real car, a modified approach had to be taken.

The non-tuned model contained about 10 states. It modelled the whole engine and not only the behaviour of the inlet and exhaust manifolds. The inlet and exhaust manifold parts of the non-tuned model became the model described in Section 3.2, hereafter denoted $Navelludd$. Measurements on a car with a measurement system installed were then made. The input signal to the $Navelludd$ model is given in (3.3). From the measurement system, N_{engine}, χ_{EGR} and χ_{VGT} are obtained. The data were fed to the non-tuned model, and the signals N_{engine}, χ_{EGR}, χ_{VGT} and χ_{fuel} in combination with the output of the non-tuned model, w_{HFM}, T_{inter} and $y = P_{inlet}$ became the input and reference signals to the $Navelludd$ model. See Figure 5.1.

The measurements were done in city traffic and on a German highway. The

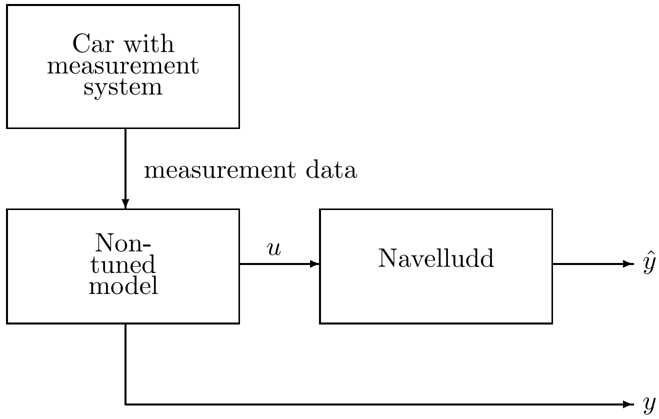

Figure 5.1 *Experimental setup*

data was then fed, as described above, to the non-tuned model for every fault value listed below. The obtained data set was then divided into 29 one-minute runs. The two approaches were then tried for each fault value on these 29 runs. The faults tested for had the values

$$\theta_{HFM} \in \{0.8, 0.9, 1.0, 1.1, 1.2\} \quad (5.1a)$$

$$\theta_{IPS} \in \{0.8, 0.9, 1.0, 1.1, 1.2\} \quad (5.1b)$$

$$\theta_{LEAK} \in \{0, 0.5, 1.0, 1.5, 2.0\} \times 10^{-5}[m^2]. \quad (5.1c)$$

The actual calculation of the test quantity is done in two steps. First the fault parameter θ_i is estimated using one of the described observer methods over the one-minute run in question. $\hat{\theta}_i$ is taken as the average value over the last 20 seconds of the fault state in the observer. The system is then *simulated* with the estimated fault parameter $\hat{\theta}_i$ fixed, and the estimated output is denoted $\hat{y}$. The test quantity is then determined using (4.1),

$$T_i(y, u) = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} (y - \hat{y}(u, \hat{\theta}_i))^2 dt,$$

with $\tau = 60s$.

To be able to perform diagnosis, it has to be checked whether the test quantities are in the rejection region or not. In this particular case, the null hypothesis is rejected if the test quantity is greater than a certain value, that is a threshold is used:

$$\text{Reject } H_0^i \text{ if } T_i(y_j, u_j) > J_i, j = j\text{:th run}$$

The index i on the test quantity denotes which fault model is used to calculate $\hat{y}$ in (4.1). For each fault model i, $i \in \{NF, HFM, IPS, LEAK\}$, there is a corresponding threshold J_i. The threshold for each fault model is chosen according to the following procedure:

1. The test quantity T_i is fed with data from the 29 runs where the fault parameter θ_i is varied over the values given in (5.1), and the other fault parameters are kept to their no fault values

2. The threshold J_i is set to the maximum of the set of test quantities obtained in the first step

Depending on how crucial it is that a fault is detected or how many false alarms that can be tolerated, different choices of the threshold J can be made. When using a max function to determine the threshold, no false alarms are wanted, but some missed detections are accepted.

This choice of threshold is made because the underlying probability distribution is not known, and it is chosen to keep the number of false alarms low. However, it can be dangerous to use the maximum of the test quantity, $\max_j T(y_j, u_j)$, as an threshold since it can be determined by a value which is much larger than what typically is the case, a so called *outlier*. The probability of such an outlier may be low enough to be acceptable when wanting a low false alarm risk. If more measurements were made, the distribution could be estimated with a histogram for instance. The threshold could then, for example, be chosen as

$$J = k\hat{\sigma}, \tag{5.2}$$

where σ is the standard deviation and k is a real constant chosen so that the probability of false alarm or missed detection, whichever is important for the application, is chosen to an appropriate value.

5.2 Evaluation of the two methods

If the number of times a test quantity is larger than its threshold is counted, and then divided by the number of trials, an estimation of the power function is obtained. In Figure 5.2, the estimated power function for the hypothesis test using the test quantity for the NF fault model is plotted against three faults. From the left are HFM sensor fault, IPS sensor fault and leakage fault, with the corresponding θ_i values on the x-axis.

The corresponding values of the test quantities, from which the estimated power function $\hat{\beta}(\theta_i)$ is determined, is shown in Figure 5.3. The test quantities are plotted against the three different fault parameters, in the same manner as for the power functions plotted in Figure 5.2. The threshold J_{NF} used to obtain $\hat{\beta}_{NF}(\theta)$ in Figure 5.2 is 6500. This threshold, J_{NF}, can be derived from Figure 5.3 by studying the maximum of the values of the test quantities corresponding to the values of the fault parameters when they have the values $\theta_{HFM} = 1$, $\theta_{IPS} = 1$ and $\theta_{LEAK} = 0$.

On the following pages, the estimated power function for the two methods are presented. Note that the power function in Figure 5.2 is the power function for the NF model for both the feedback on the fault state and the pole placement estimation observer methods, since the NF fault model has no fault parameter. Plots for the underlying test quantities are also shown.

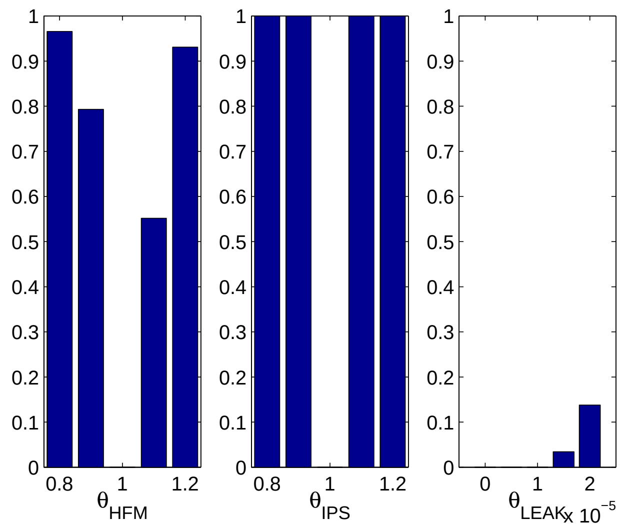

Figure 5.2 *Power function $\beta_{NF}(\theta_i)$ when the null hypothesis is $H_0 : F_p = NF$ and the NF fault model is simulated to obtain $\hat{y}$ used in (4.1).*

5.3 Fault state feedback observer test quantity

In this section, results from the fault state feedback observer method is shown. The parameter is first estimated, and then the model is simulated with the estimated fault parameter. The thresholds are

$$\begin{aligned} J_{NF} &= 6500 \\ J_{HFM} &= 4600 \\ J_{IPS} &= 3500 \\ J_{LEAK} &= 6400. \end{aligned}$$

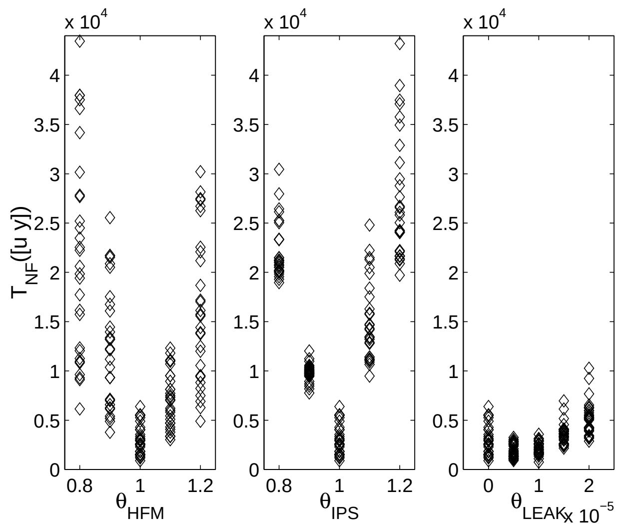

Figure 5.3 *Test quantity calculated according to* (4.1) *where* $\hat{y}$ *is determined by simulation of the* NF *fault model.*

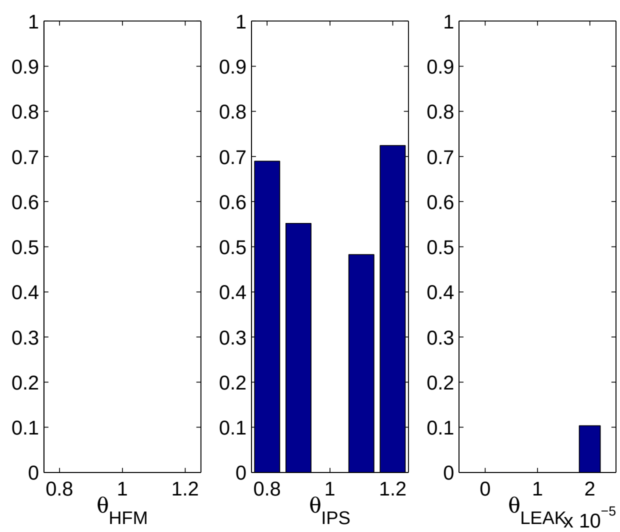

Figure 5.4 *Power function for the hypothesis test* $H_0 : F_p \in \{NF, HFM\}$ *when* θ_{HFM} *is estimated with feedback on the fault state only observer and* $\hat{y}$ *used in* (4.1) *is obtained by simulating the* HFM *fault model.*

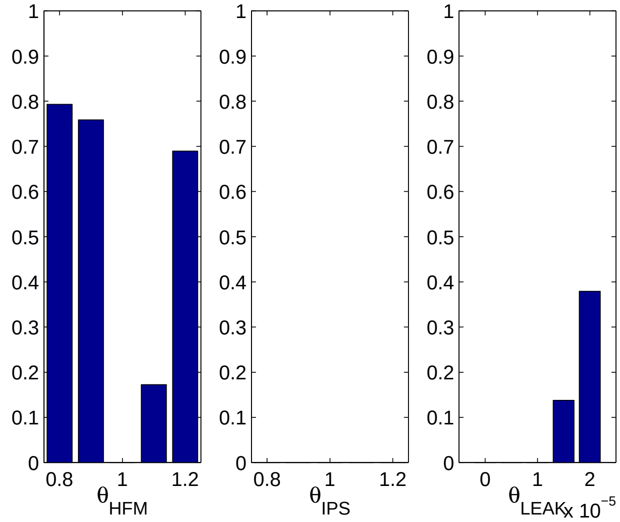

Figure 5.5 *Power function for the hypothesis test* $H_0 : F_p \in \{NF, IPS\}$ *when* θ_{IPS} *is estimated with feedback on the fault state only observer and* $\hat{y}$ *used in* (4.1) *is obtained by simulating the* IPS *fault model.*

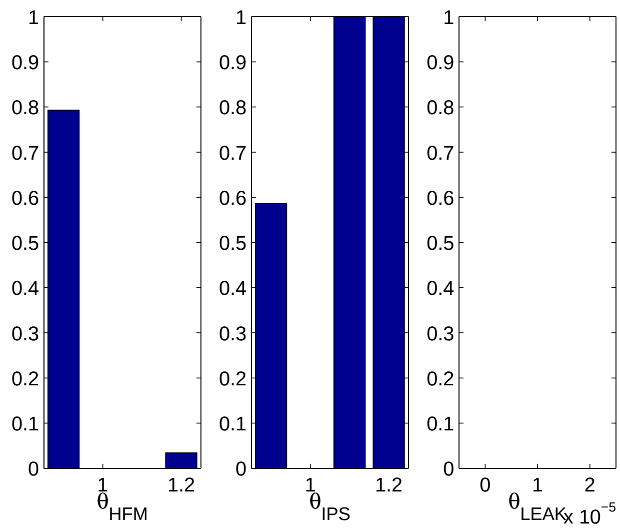

Figure 5.6 *Power function for the hypothesis test* $H_0 : F_p \in \{NF, LEAK\}$ *when* θ_{LEAK} *is estimated with feedback on the fault state only observer and* $\hat{y}$ *used in* (4.1) *is obtained by simulating the* $LEAK$ *fault model.*

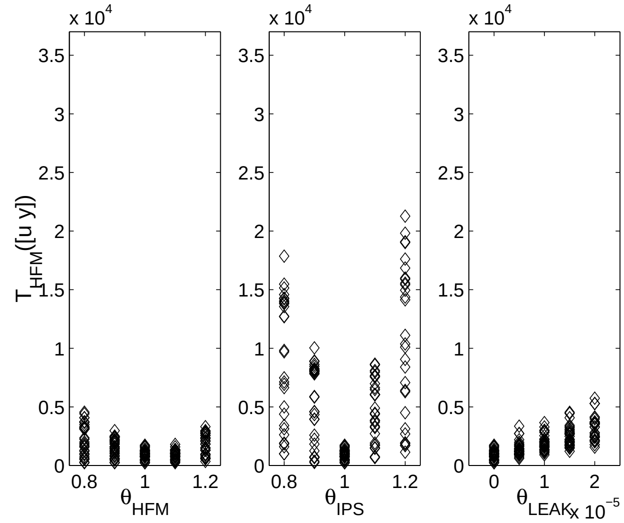

Figure 5.7 *Test quantities calculated according to* (4.1) *where* $\hat{y}$ *is determined by simulation of the* HFM *fault model and* θ_{HFM} *is estimated by feedback on the fault state only observer.*

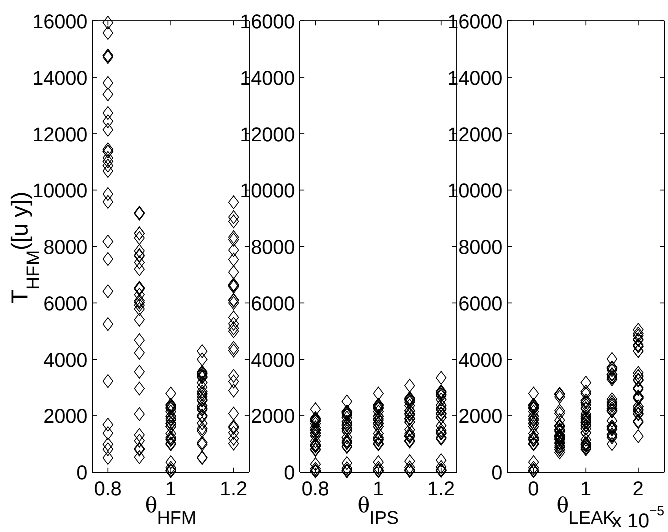

Figure 5.8 *Test quantities calculated according to* (4.1) *where* $\hat{y}$ *is determined by simulation of the* IPS *fault model and* θ_{IPS} *is estimated by feedback on the fault state only observer.*

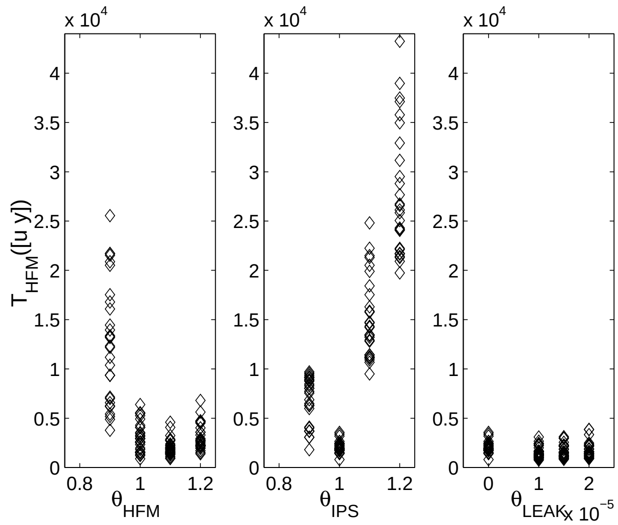

Figure 5.9 *Test quantities calculated according to* (4.1) *where* $\hat{y}$ *is determined by simulation of the* $LEAK$ *fault model and* θ_{LEAK} *is estimated by feedback on the fault state only observer.*

5.4 Pole placement observer based test quantity

In this section, the results from the pole placement estimation observer is shown. First the fault parameter θ_i is estimated with an observer with an observer gain determined by the method of pole placement applied to a linearised version of the system. In the second run, the test quantity is calculated when the system is simulated without an observer gain. The thresholds obtained are

$$\begin{aligned} J_{NF} &= 6500 \\ J_{HFM} &= 2200 \\ J_{IPS} &= 3000 \\ J_{LEAK} &= 6400. \end{aligned}$$

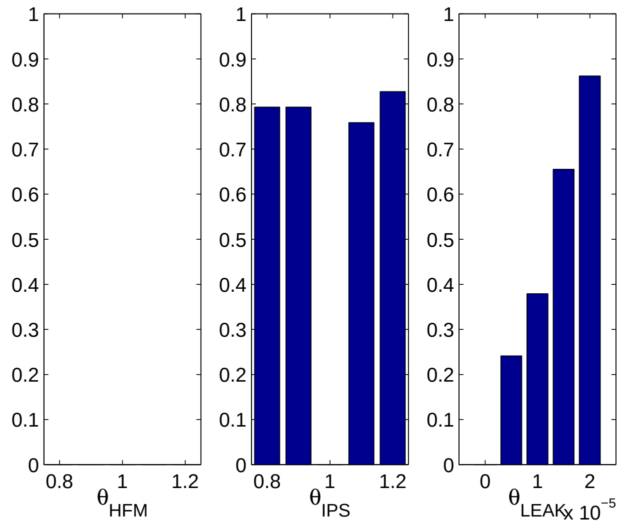

Figure 5.10 *Power function for the hypothesis test* $H_0 : F_p \in \{NF, HFM\}$ *when* θ_{HFM} *is estimated with pole placement observer and* $\hat{y}$ *used in* (4.1) *is obtained by simulating the* HFM *fault model.*

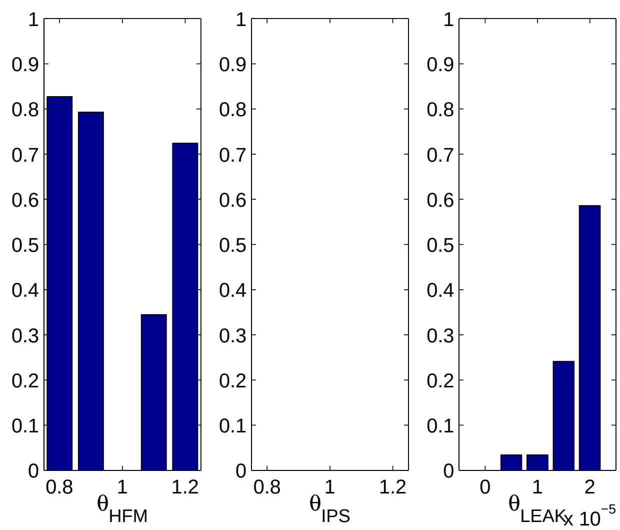

Figure 5.11 *Power function for the hypothesis test* $H_0 : F_p \in \{NF, IPS\}$ *when* θ_{IPS} *is estimated with pole placement observer and* $\hat{y}$ *used in* (4.1) *is obtained by simulating the* IPS *fault model.*

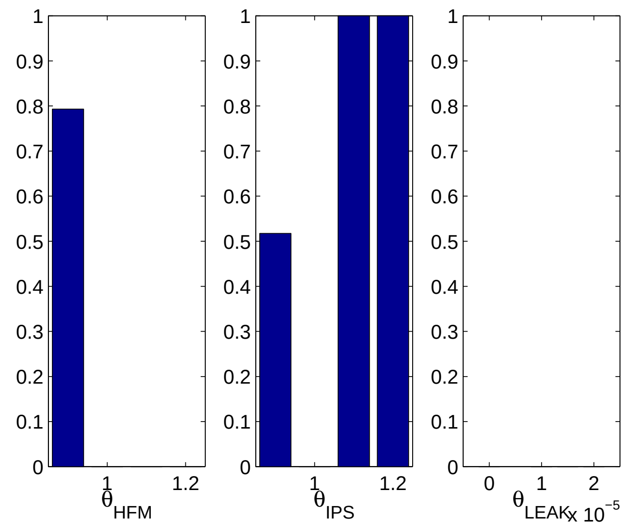

Figure 5.12 *Power function for the hypothesis test $H_0 : F_p \in \{NF, LEAK\}$ when θ_{LEAK} is estimated with pole placement observer and $\hat{y}$ used in (4.1) is obtained by simulating the $LEAK$ fault model.*

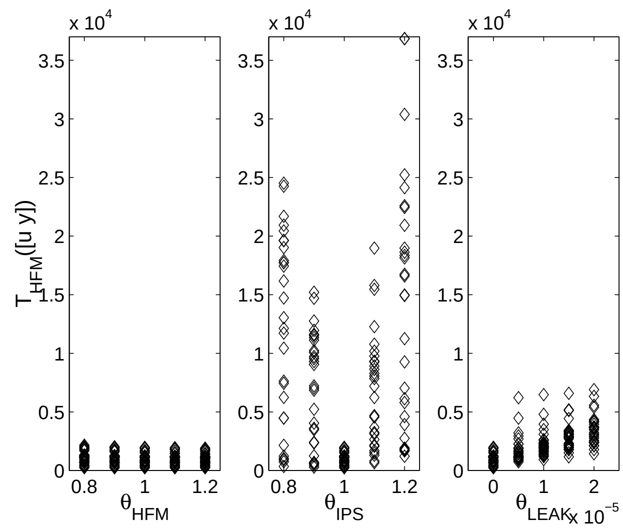

Figure 5.13 *Test quantities calculated according to (4.1) where $\hat{y}$ is determined by simulation of the HFM fault model and θ_{HFM} is estimated by the pole placement observer.*

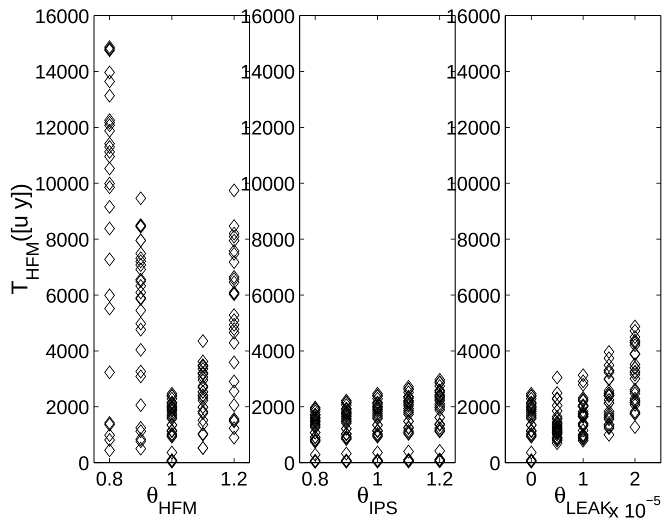

Figure 5.14 *Test quantities calculated according to* (4.1) *where* $\hat{y}$ *is determined by simulation* … *pole placement observer.*

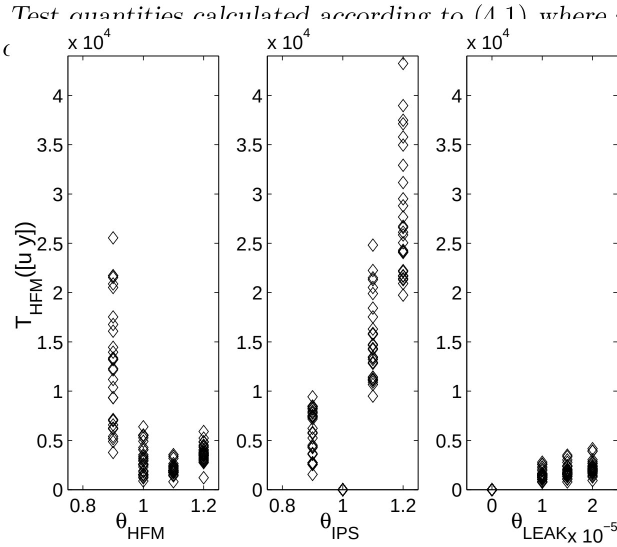

Figure 5.15 *Test quantities calculated according to* (4.1) *where* $\hat{y}$ *is determined by simulation of the* $LEAK$ *fault model and* θ_{LEAK} *is estimated by the pole placement observer.*

5.5 Comparison

A brief summary of the result of the experiments are presented in Table 5.1. The results of the two methods will now be compared for each fault mode.

	Fault state feedback	Pole placement	Comment
NF	n/a	n/a	no fault parameter estimated
HFM		better	overall better performance
IPS		better	better for smaller faults
$LEAK$	slightly better		better for IPS faults and HFM faults $\theta_{HFM} = 1.2$

Table 5.1 *Summarised results for the comparison between the two discussed methods. See the text in Section 5.5 for further details.*

NF

The result for the two methods are identical since the two methods are identical for the NF fault model. No parameter is estimated, and the test quantity (4.1) is obtained by simply simulating the system. It is interesting to note that the NF model can describe the effect of small leakages, thus obtaining a small probability of rejecting the null hypothesis $H_0 : F_p \in \{NF\}$ when a leakage of the small sizes tested are present. This is an undesired property, since it means that leakages sizes under 1.5 10^{-5} m^2 cannot be detected with the present model.

HFM

Comparing the two power functions for the two methods, Figure 5.4 and Figure 5.10, the power function based on the pole placement method is around 0.8 for all IPS faults, while the corresponding power function based on the fault state feedback varies between the values 0.45 and 0.65. For the $LEAK$ fault, the feedback state observer method gives virtually a power function of 0, while the pole placement method gives a power function with rising probabilities with increasing leakage sizes. The differences are significant for the hypothesis test $H_0 : F_p \in \{NF, HFM\}$, in favour of the pole placement method.

IPS

The pole placement observer method gives a slightly better power function regarding HFM faults than the fault state feedback observer method, as seen when comparing the plots in Figure 5.5 and in Figure 5.11. However, comparing the power functions for the leakage faults, the pole placement methods outperforms the fault state feedback observer method. The pole placement method thus gives the best performance for the hypothesis test $H_0 : F_p \in \{NF, IPS\}$.

$LEAK$

In Figures 5.6 and 5.12, the two power function resulting from the two methods applied to the leakage model are presented. The fault state feedback observer method gives a slightly better performance, but the differences are not significant.

Comment on the $LEAK$ case

In both Figures 5.6 and 5.12, it is seen that the power function for

$$\hat{\beta}_{LEAK}(\theta_{HFM}),\ \theta_{HFM} > 1$$

are almost always zero. According to the *Navelludd* model, a higher intake air flow, which $\theta_{HFM} > 1$ suggests, gives a higher pressure. No such pressure buildup is observed, but according to the leakage fault model, this can be explained by air leaking out to the surrounding environment. This could explain the obtained power functions.

There was also a problem with the leakage fault model when testing against sensor fault values of $\theta_{HFM}, \theta_{IPS} = 0.8$, see Figures 5.9, 5.15 . When the test quantity were to be calculated, the simulation of $\hat{y}$ could not be done because the simulation stalled. This may be due to the fact that a higher pressure buildup than expected has to be explained by a negative leakage area, which is not allowed in the model.

A serious concern is that the NF fault model can describe leakages better than the other fault models, which have one adjustable parameter to minimise discrepancies between the reference and predicted values. This can be seen by comparing Figure 5.2 with the other power functions, Figures 5.4,5.5, 5.10 and 5.11. Another choice of $\hat{\theta}_i$ should probably be made.

Interpretation of the comparison

For the two given observers, the pole placement method works better. However, both the observers were tuned in an ad hoc manner, and the tuning was only iterated ≈ 5 times. In order to reach conclusive results, more effort has to be put in systematically tuning of the two observers. Although input signals from real driving with a car was used, no analysis has been made how typical these kind of input signals are in everyday driving. The aim of the diagnosis system is, after all, that it is to be used in production cars.

Because of the non-completed tuning of the observers, it can only be speculated about the cause of the differences seen in the performances of the two observer methods. The pole placement observer uses all known information of the behaviour of the model. This might explain the better performance of the model.

On the other hand, the answer may lie in the two stage method applied to calculate the test quantities. The thresholds for the pole placement observer is lower than the thresholds for the fault state feedback gain observer, for the HFM case even much lower. Compare for instance Figures 5.7 and 5.13. This explains why the power functions differs som much. The remaining, and unanswered, question is thus why the test quantities, and therefore alos the thresholds, differ so much.

Remark

It was also tried to estimate $\hat{y}$ in (4.1) with the pole placement observer, but this resulted in the same power functions shown in Section 5.4.

Chapter 6

Conclusion

In the thesis, a model based diagnosis has been performed using the method of structured hypothesis tests [1]. The test quantities were determined by using the prediction error from a model structure, with an unknown parameter value taken forth with an observer. Two observer design methods were tried, linearisation in combination with pole placement [6], and an observer gain feedback only on the fault state.

With the obtained observers, the method of linearisation in combination with pole placement performed best. However, although the pole placement approach works better when regarding the performance against faults, the feedback gain only on the fault state method also gives a working diagnosis system. In fact, both methods gives the decision structure in Figure 6.1 for the resulting diagnosis system. Since the fault state feedback gain method is less complex than the pole placement method, the result, in view of the unfinished tuning, is that no conclusive result can be given at this point which of the two described methods is to be favoured.

	NF	HFM	IPS	$LEAK$
T_{NF}	0	X	X	X
T_{HFM}	0	0	X	X
T_{IPS}	0	X	0	X
T_{LEAK}	0	X	X	0

Figure 6.1 *Decision structure for the diagnosis system obtained by using the methods described in the thesis.*

Bibliography

[1] Mattias Nyberg. *Model Based Fault Diagnosis: Methods, Theory, and Automotive Engine Applications*. Phd thesis liu-tek-phd-1999:591, Linköping University, 1999.

[2] Gunnar Blom. *Sannolikhetsteori och statistikteori med tillämpningar*. Studentlitteratur, 1989.

[3] George Casella and Roger L. Berger. *Statistical inference*. Brooks/Cole Pub Co, 1990.

[4] Torkel Glad and Lennart Ljung. *Control Theory : Multivariable & Nonlinear Methods*. Taylor & Francis, 2000.

[5] Torkel Glad. Nonlinear control theory. Technical Report LiTH-ISY-R-1686, Dept of EE. Linköping University, S-581 83 Linköping, Sweden, Feb 1994.

[6] Nichols Kautsky and Van Dooren. Robust pole assignment in linear state feedback,. *Intl. J. Control, 41(1985)5, pp 1129-1155*, 41(5), May 1985.

[7] Karl Gustav Andersson and Lars-Christer Böiers. *Ordinära differentialekvationer*. Studentlitteratur, 1992.