Residual Generation Methods for Fault Diagnosis with Automotive Applications

Carl Svärd

Department of Electrical Engineering
Linköpings universitet, SE-581 83 Linköping, Sweden
Linköping 2009
Residual Generation Methods
for Fault Diagnosis with Automotive Applications

© 2009 Carl Svärd

carl@isy.liu.se
http://www.vehicular.isy.liu.se
Department of Electrical Engineering,
Linköpings universitet,
SE–581 83 Linköping,
Sweden.

ISBN 978-91-7393-608-8
ISSN 0280-7971
LIU-TEK-LIC-2009:14

Printed by LiU-Tryck, Linköping, Sweden 2009
Abstract

The problem of fault diagnosis consists of detecting and isolating faults present in a system. As technical systems become more and more complex and the demands for safety, reliability and environmental friendliness are rising, fault diagnosis is becoming increasingly important. One example is automotive systems, where fault diagnosis is a necessity for low emissions, high safety, high vehicle uptime, and efficient repair and maintenance.

One approach to fault diagnosis, providing potentially good performance and in which the need for additional hardware is minimal, is model-based fault diagnosis with residuals. A residual is a signal that is zero when the system under diagnosis is fault-free, and non-zero when particular faults are present in the system. Residuals are typically generated by using a mathematical model of the system and measurements from sensors and actuators. This process is referred to as residual generation.

The main contributions in this thesis are two novel methods for residual generation. In both methods, systems described by Differential-Algebraic Equation (DAE) models are considered. Such models appear in a large class of technical systems, for example automotive systems. The first method considers observer-based residual generation for linear DAE-models. This method places no restrictions on the model, such as e.g. observability or regularity, in comparison with other previous methods. If the faults of interest can be detected in the system, the output from the design method is a residual generator, in state-space form, that is sensitive to the faults of interest. The method is iterative and relies on constant matrix operations, such as e.g. null-space calculations and equivalence transformations.

In the second method, non-linear DAE-models are considered. The proposed method belongs to a class of methods, in this thesis referred to as sequential residual generation, which has shown to be successful for real applications. This method enables simultaneous use of integral and derivative causality, and is able to handle equation sets corresponding to algebraic and differential loops in a systematic manner. It relies on a formal framework for computing unknown variables in the model according to a computation sequence, in which the analytical properties of the equations in the model as well as the available tools for equation solving are taken into account. The method is successfully applied to complex models of an automotive diesel engine and a hydraulic braking system.
Acknowledgments

This work has been performed as a part of a collaborative industrial research project between Scania CV AB in Södertälje and the division of Vehicular Systems, Department of Electrical Engineering, Linköping University.

First of all, I would like to express my gratitude to Dr. Mattias Nyberg, my supervisor, for great guidance into the world of research, his never-ending enthusiasm, and for always taking his time for discussions. I would also like to thank Dr. Erik Frisk, my assistant supervisor, for giving and interesting discussions, proof-reading, and help with numerous \LaTeX{} issues. Professor Lars Nielsen is acknowledged for letting me join his research group Vehicular Systems.

Thanks also goes to all colleagues at Vehicular Systems and NESD for inspiring working atmospheres and nice coffee-breaks. Furthermore, I would like to thank Anna Pernestål for proof-reading parts of this manuscript, and Anders Eriksson, Peter Madsen, and Peter Vansölin, my managers at Scania, for letting me be a part of this project and do research.

This work is jointly financed by Scania CV AB and VINNOVA, Swedish Governmental Agency for Innovation Systems, who are also acknowledged.

Finally, I would like to thank my parents, Åsa and Kjell, my sister Anna and my friends for their support and encouragement.

Carl Svärd
Linköping, May 2009
# Contents

## I Introduction to Model-Based Fault Diagnosis 1

### 1 Introduction 3

1.1 Overview and Contributions 4  
1.1.1 Paper 1 - Linear Observer-Based Residual Generation 4  
1.1.2 Paper 2 - Non-Linear Sequential Residual Generation 5

### 2 Model-Based Fault Diagnosis 7

2.1 Models 7  
2.2 Diagnostic Tests 7  
2.3 Diagnostic Tests Based on Residuals 9  
2.3.1 Test Quantity 9  
2.4 Fault Isolation 10  
2.5 Fault Decoupling 11  
2.6 Residual Generation 11  
2.6.1 Observer-Based Residual Generation 11  
2.6.2 Sequential Residual Generation 12

### 3 Model-Based Fault Diagnosis in Automotive Systems 13

3.1 Why Fault Diagnosis is Important 13  
3.1.1 Emissions 14  
3.1.2 Vehicle Uptime 15  
3.1.3 Efficient Repair and Maintenance 15  
3.2 Faults to Diagnose 16
6 Application Studies ......................................................... 72
  6.1 Implementation and Configuration of the Method ........... 72
  6.2 Performance Measures ........................................... 73
  6.3 Automotive Diesel Engine ....................................... 74
  6.4 Hydraulic Braking System ........................................ 75
  6.5 Realization of a Residual Generator for the Diesel Engine 76
7 Conclusions ................................................................. 80
References ................................................................. 82
A Proofs of Theorems and Lemmas .................................... 86
Part I

Introduction to Model-Based Fault Diagnosis
Fault diagnosis is the act of detecting and isolating faults present in a system. With the rising demand for safety and reliability of technical systems, driven by economical and environmental incentives, fault diagnosis has become increasingly important. One example is automotive systems, and in particular engines, that are by regulations required to have on-board diagnosis of all faults that may lead to increased emissions, see e.g. [United Nations, 2008]. In addition, fault diagnosis in automotive systems is essential to maintain high vehicle uptime, low fuel consumption, high safety, and efficient service and maintenance.

One approach to fault diagnosis that provides potentially good performance and in which the need for additional hardware is avoided, is model-based fault diagnosis with residuals. A residual is a signal that is zero when the system under diagnosis is fault-free, and non-zero when particular faults are present in the system. Residuals are often generated by utilizing a mathematical model of the system under diagnosis and measurements from sensors and actuators, a process referred to as residual generation. To enable fault isolation, a diagnosis system typically contains a set of residuals designed to respond to different subsets of faults. Meaning that some faults in a residual must be decoupled. Decoupling of faults in residuals is thus a fundamental problem in residual generation for fault isolation.

One important class of residual generation methods is observer-based residual generation. In these methods, the approach is to base residual generators on state-observers. A state-observer utilizes a model of the system and measurements to obtain an estimate of the states in the system. A residual can then
be formed as the difference between estimated and measured states. Several methods exists for design of observers for both linear and non-linear state-space models, i.e. ordinary differential equations with additional equations relating the states and measurements, as well as for linear and non-linear Differential-Algebraic Equation (DAE) models. DAE-models contains both differential and algebraic equations, and are of interest since these general models appear in a large class of technical systems, e.g. electrical-, mechanical-, and chemical systems. DAE-models are also the result when using object-oriented modeling tools, e.g. Modelica. In most of the observer-based residual generation methods, for both state-space and DAE-models, decoupling of faults is obtained by transforming the original model into a sub-model where only the faults of interest are present.

Another class of residual generation methods, that has shown to be successful for real applications, is in this thesis referred to as sequential residual generation. In sequential residual generation, the unknown variables in a model, or sub-model, are computed by solving equations one at a time in a sequence and a residual is then obtained by evaluating a redundant equation. In this class of methods, the original model is often divided into sub-models with specific properties and then residual generators are designed for each sub-model. Since a residual generator is only sensitive to those faults affecting its corresponding sub-model, all other faults are decoupled. Sequential residual generation methods has the potential to be automated to an high extent, making them especially important for the automotive applications studied in this thesis.

1.1 Overview and Contributions

Chapter 2 gives a brief introduction of theoretical concepts in model-based fault diagnosis that are central in this thesis. The aim of Chapter 2 is to provide a theoretical background to the rest of the thesis and to place its contributions in a context. Chapter 3 focuses on model-based fault diagnosis in automotive systems and intends to give an application oriented background and motivation to this work. Two papers are enclosed in Part II. These constitute the main contributions and are summarized below.

1.1.1 Paper 1 - Linear Observer-Based Residual Generation

In Paper 1, residual generation for linear DAE-models is considered. The main contribution is a new systematic design method for observer-based residual generation for systems described by linear DAE-models. By constant matrix operations, the original DAE-model is transformed into a sub-model in state-space form, of lower dimension than the DAE-model, where only faults that should be detected are present. Thus, faults not present in the transformed sub-model are decoupled. The transformation is iterative and straightforward
1.1. Overview and Contributions

to implement. In contrast to other methods no restrictions, such as e.g. observability or regularity, are placed on the model of the system to be diagnosed. An illustrative numerical example is included, where the design method is applied to a non-observable model of a robot manipulator.

Paper 1 has been submitted to European Journal of Control. The paper is based on [Svärd and Nyberg, 2008c]:


The work in the above conference paper has also been presented at *Reglermöte 2008*, Luleå, Sweden, [Svärd and Nyberg, 2008d].

1.1.2 Paper 2 - Non-Linear Sequential Residual Generation

The main contribution of Paper 2 is a novel method for sequential residual generation for non-linear DAE-models. The method relies on a formal framework for computing unknown variables according to a computation sequence, in which the analytical properties of the equations in the model and the available tools for algebraic equation solving are taken into account. An initial step in the method is to divide the original model into sub-models with specific properties, and residual generators for each sub-model are then designed. In this way, all faults not affecting the sub-model are decoupled in the corresponding residual generator. The proposed method is successfully applied to two models of automotive systems, a Scania diesel engine and a hydraulic braking system.

Paper 2 has been submitted to IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans. The paper is an extended and revised version of the work presented in [Svärd and Nyberg, 2008a]:


An extended version of the above conference paper can be found in [Svärd and Nyberg, 2008b].
Chapter 1. Introduction
The aim of this chapter is to introduce some theoretical concepts in model-based fault diagnosis that are central in this thesis, and to place the contributions presented in Part II in a context.

2.1 Models

To perform model-based diagnosis, a model of the system under diagnosis is needed. In this thesis, a model is a set of equations relating sets of unknown and known variables. The equations may be linear or non-linear, static or dynamic. That is, linear and non-linear Differential-Algebraic Equation (DAE) models are considered. Typically, faults that may affect the system are also included in the model. Faults are often classified into behavioral modes. For example, behavioral modes for a simple system containing one sensor and one actuator may be “sensor fault”, “actuator fault”, and “no fault”. Behavioral modes are usually assigned to components, here we instead use them for systems.

2.2 Diagnostic Tests

A typical diagnosis system consists of a set of diagnostic tests and a fault isolation scheme, see Figure 2.1. A diagnostic test utilizes observations, i.e. measurements, from the system under diagnosis to determine if a specific behavioral mode is present in the system or not. A diagnostic test $\delta_i$, can be viewed
as a hypothesis test [Berger, 1985] with the hypothesis

\[ H_0^i : F_p \in B_i \]
\[ H_1^i : F_p \in B_C^i \]

where \( F_p \) denotes the present behavioral mode in the system, \( B_i \) a set of behavioral modes corresponding to faults not monitored by \( \delta_i \), and \( B_C^i \) the complement of \( B_i \), see e.g. [Nyberg, 1999]. The common convention used is that when the hypothesis \( H_0^i \) is rejected, it is assumed that \( H_1^i \) is true. When \( H_0^i \) is not rejected, nothing is assumed which means that the present behavioral mode can be any of the behavioral modes for the system under diagnosis. The outcome of the diagnostic test \( \delta_i \) is thus a decision

\[ S_j = \begin{cases} 
S_1^j = B_C^i & \text{if } H_0^i \text{ is rejected} \\
S_0^j = \Omega & \text{if } H_0^i \text{ is not rejected}
\end{cases} \]  

where \( \Omega \) denotes all behavioral modes for the system.

Common traditional approaches for construction of diagnostic tests are for example limit checking, i.e. to check if a sensor is within its normal operating range, or to employ hardware redundancy. For instance, if two sensors are used to measure the same physical quantity, it is possible to test if one of the sensors is faulty by comparing the values of the sensors. Another approach, providing potentially increased diagnosis performance and in which the need of additional, redundant, hardware is avoided, is to use diagnostic tests based on residuals.

![Diagram](image.png)

**Figure 2.1:** A typical diagnosis system consists of a set of diagnostic test and a fault isolation scheme.
2.3 Diagnostic Tests Based on Residuals

A residual is a signal ideally zero in the non-faulty case and non-zero else. A residual generator takes measurements from the system under diagnosis as input, and produces a residual as output, see Figure 2.2. Residual generators are typically constructed by using a mathematical model of the system. For instance, a residual can be obtained as the comparison between a value estimated by a model and the corresponding measured quantity. The residual generator consists in this case of the model used for the estimation and the equation describing the comparison, referred to as the residual equation. Two methods for residual generation are presented in this thesis. The method presented in Paper 1 handles linear DAE-models, and the method in Paper 2 non-linear DAE-models.

![Residual Generator Diagram](image)

Figure 2.2: A residual can be generated by utilizing a mathematical model of the system under diagnosis and measurements.

2.3.1 Test Quantity

A common way to construct a diagnostic test based on a residual is to form a test quantity from the residual, and then threshold the test quantity, see Figure 2.3. A test quantity is a constant value, in comparison with a residual which is a trajectory, i.e. a function of time. A test quantity can for example be formed as the mean-effect or mean-value of the residual in some time-window, or just as a sample of the residual at a specific time. Simply, given a residual \( r \), generated by using a model and measurements \( z \), a diagnostic test \( \delta_i \) constructed...
Chapter 2. Model-Based Fault Diagnosis

via a test quantity \( T \), based on \( r \), is defined as

\[
S_i = \delta(z) = \begin{cases} 
S_i^1 & \text{if } T(r(z)) \geq J \\
S_i^0 & \text{if } T(r(z)) < J 
\end{cases}
\]

where \( J \) is a threshold. Typically, residuals are not perfectly zero in the non-faulty case due to for example noisy measurements and modeling errors. Thus, the approach used to form the test quantities and the thresholds are important design parameters in a diagnosis system.

![Diagnostic Test Diagram](image)

Figure 2.3: A diagnostic test based on a residual via a test quantity.

2.4 Fault Isolation

There are several approaches for fault isolation, most originating from the field of Artificial Intelligence (AI), see e.g. [de Kleer and Williams, 1987]. Another approach is Bayesian fault isolation, see e.g. [Pernestål, 2007]. Here, in order to briefly illustrate the concept of fault isolation, we will use a straight-forward method referred to as structured residuals, [Gertler, 1991], or structured hypothesis tests, [Nyberg, 1999].

To enable isolation of faults, the diagnostic tests used in a diagnosis system are designed to test different behavioral modes. Consider a diagnosis system containing the diagnostic tests \( \{\delta_1, \delta_2, \ldots, \delta_n\} \). The outcome of the diagnostic test \( \delta_i \) is a decision \( S_i \), according to (2.1). Under a single fault assumption, we simply obtain the total diagnosis statement \( S \) as

\[
S = \bigcap_{i=1}^{n} S_i
\]

for multiple faults please refer to e.g. [de Kleer and Williams, 1987]; [Reiter, 1987]; [Greiner et al., 1989].

For an example, consider a set of tests, \( \{\delta_1, \delta_2, \delta_3\} \), constructed to detect and isolate three faults, \( \{f_1, f_2, f_3\} \). The following fault signature matrix,

\[
\begin{array}{ccc}
\delta_1 & f_1 & f_2 & f_3 \\
\delta_2 & 1 & 1 & \\
\delta_3 & 1 & 1 & 
\end{array}
\]  
\text{(2.2)}
shows which tests that are sensitive to which faults, i.e. test $\delta_1$ is sensitive to faults $f_2$ and $f_3$, and so on. Now assume a situation where tests $\delta_1$ and $\delta_2$, but not $\delta_3$ has reacted. We then obtain the decisions $S_1 = \{f_2, f_3\}$, $S_2 = \{f_1, f_3\}$, and $S_3 = \{f_1, f_2, f_3, NF\}$, where $NF$ is used to denote the behavioral mode corresponding to that no faults are present. The diagnosis statement thus becomes

$$S = S_1 \cap S_2 \cap S_3 = \{f_2, f_3\} \cap \{f_1, f_3\} \cap \{f_1, f_2, f_3, NF\} = f_3$$

and we can conclude that fault $f_3$ is present.

### 2.5 Fault Decoupling

To achieve a specific fault signature matrix, for example one similar to (2.2), decoupling of faults in diagnostic tests is needed. The faults that are decoupled in a test are often referred to as non-monitored faults, whereas the faults not decoupled are called monitored faults. In the example above, fault $f_1$ is decoupled in test $\delta_1$, which means that for $\delta_1$, fault $f_1$ is a non-monitored fault and $f_2$ and $f_3$ are monitored faults. Decoupling of faults in a set of tests based on residuals, means that the residuals must respond to, or similarly be sensitive to, different subsets of faults. Thus, fault decoupling is a fundamental problem in residual generation for fault isolation.

### 2.6 Residual Generation

In this thesis, two classes of residual generation methods are considered, observer-based residual generation and sequential residual generation. These both classes have the potential to handle DAE-models, and to handle fault decoupling in a systematic manner. DAE-models are of interest since such models appear in a large class of technical systems, e.g. automotive systems, and also are the result when using object-oriented modeling tools such as e.g. Modelica, [Fritzon, 2004].

#### 2.6.1 Observer-Based Residual Generation

A common approach is, as said in Section 1, to base residual generators on state-observers. A residual is in this case formed as the difference between estimated and measured states. Several methods exists for design of observers for state-space models, see e.g. [Kailath et al., 2000] for linear models, and [Hendey, 2008]; [Misawa and Hedrick, 1989]; [Walcott et al., 1987]; [Slotine et al., 1987]; [Khalil, 1999] for non-linear models. For linear DAE-models see e.g. [Hou and Müller, 1999]; [Hou and Müller, 1995]; [Darouach and Boutayeb, 1995];
Chapter 2. Model-Based Fault Diagnosis

[Müller and Hou, 1993]; [Shields, 1992]; [Dai, 1989]. For non-linear DAE-models the list of works is not that extensive, but includes for example [Åslund and Frisk, 2006]; [Becerra et al., 2001]; [Zimmer and Meier, 1997]. For the specific application of using the observer for diagnosis, see for example [Massoumnia, 1986]; [Massoumnia et al., 1989]; [Hammouri et al., 2001] for linear state-space models, and [Hammouri et al., 1999]; [De Persis and Isidori, 2001]; [Martínez-Guerra et al., 2005]; [Kaboré et al., 2000] for non-linear state-space models. Several methods also exists for observer-based residual generation in linear DAE-models, for example [Hou, 2000]; [Patton and Hou, 1998]; [Shields, 1994]; [Marx et al., 2003] and some for non-linear DAE-models e.g. [Gao and Ding, 2007]; [Vemuri et al., 2001]; [Shields, 1997]. In most of the works above, for both state-space models and DAE-models, decoupling of faults is obtained by transforming the original model into a sub-model where only the faults of interest are present. Observer-based residual generation for linear DAE-models is considered in Paper 1.

2.6.2 Sequential Residual Generation

Sequential residual generation, [Staroswiecki and Declerck, 1989], is of interest since it has shown to be successful for real applications, [Dustegor et al., 2004]; [Izadi-Zamanabadi, 2002]; [Cocquempot et al., 1998]; [Hansen and Molin, 2006]; [Kingstedt and Johansson, 2008]; [Dagsson and Nissilä-Källström, 2009], and in addition also has the potential to be automated to a high extent, [Frisk et al., 2006]; [Einarsson and Arhennius, 2005]; [Krigsman and Nilsson, 2005]; [Eriksson, 2005]; [Svård and Wassén, 2006]. In sequential residual generation, the unknown variables in a model, or sub-model, are computed by solving equations one at a time in a sequence and a residual is then obtained by evaluating a redundant equation. Similar approaches as in [Staroswiecki and Declerck, 1989], have been described and exploited in e.g. [Staroswiecki, 2002]; [Blanke et al., 2003]; [Pulido and Alonso-González, 2004]. In this class of methods, the original model is often divided into sub-models with specific properties and residual generators are then designed for each sub-model. Since a residual generator is only sensitive to those faults affecting its corresponding sub-model, all other faults are decoupled. Sequential residual generation is considered in Paper 2.
Modern automotive systems are complex. One example is automotive diesel engines, see Figure 3.1, that in order to have low fuel consumption, produce low emissions, and offer good driveability, are equipped with for example Exhaust Gas Recirculation (EGR) and a Variable Geometry Turbocharger (VGT). To purify exhausts, diesel engines interact with and are dependent on one or several advanced after-treatment systems such as a Diesel Particulate Filter (DPF), and a Selective Catalytic Reduction (SCR) system, see Figure 3.2(b). In addition, to provide optimum fuel economy, good safety, and further increase driveability, they interact with other complex systems in the powertrain like an automatic gearbox and an auxiliary hydraulic braking system, see Figure 3.3. Even small faults in the engine or in any of the systems mentioned above may have undesirable effects, such as increased emissions or reduced safety. The objectives of this chapter are to provide a background and motivation to this thesis and to place its contributions into an application oriented context.

3.1 Why Fault Diagnosis is Important

Faults affecting the engine or any of the systems mentioned above may lead to

- increased emissions,
- decreased safety,
- increased fuel consumption,
• decreased driveability, or
• vehicle off-road.

These consequences may be prevented, or at least reduced, if faults can be detected and isolated in time. In addition, beside these more or less obvious gains, good diagnosis is a requirement for high vehicle uptime and efficient maintenance, regarding both cost and time. These aspects are further discussed below.

3.1.1 Emissions

Automotive engines are by regulations required to have high-precision On-Board Diagnosis (OBD) of faults that are harmful for the environment, see e.g. [United Nations, 2008]. The legislations states that all manufactured vehicles must be equipped with an OBD-system capable of detecting faults in all components that, if broken, leads to emissions over pre-defined OBD thresholds during a specific driving cycle. For heavy-duty trucks, emissions of especially nitrogen oxides (NOx) and particulate matter (PM) are crucial. Coming legislations in the European Union, EUVI, require substantially lowered emission and OBD thresholds, see Figures 3.4 and 3.5, and in addition that faults leading to increased emissions can be isolated.
3.1. Why Fault Diagnosis is Important

Figure 3.2: Usage of EGR and/or SCR in diesel engine reduces the generation of NOx. Illustrations are due to Semcon Informatic Graphic Solutions.

3.1.2 Vehicle Uptime

To reduce, or preferably eliminate, the impact of faults by taking appropriate actions on the road, referred to as fault tolerant control, see e.g. [Blanke et al., 2003], on-board diagnosis is essential. For example, if a fault occurs but the fault can be detected and isolated on-board so that the effects of the fault can be eliminated, the vehicle can continue on its driving mission and stop by the workshop later. On-board diagnosis therefore increases the vehicle uptime. Vehicle uptime is important for vehicle owners, since even a stationary vehicle costs money, and can not be used to earn money.

3.1.3 Efficient Repair and Maintenance

On-board diagnosis of faults is also important to provide efficient service when the vehicle visits the workshop. If faults have been correctly detected and isolated, additional troubleshooting at the workshop is unnecessary. However, as automotive systems become more and more complex it is utopian that all necessary fault detection and isolation can be performed on-board the vehicle. Therefore off-board fault detection and isolation of faults, i.e. at the workshop, is becoming more and more important. Due to hardware limitations on-board the vehicle and the ability to actively excite systems when the vehicle is at the workshop, off-board fault detection and isolation of faults may also give better and more precise results. Nevertheless, fault detection and isolation, on-or/and off-board, decreases the repair and maintenance costs for the vehicle, since the time at the workshop is minimized and no unnecessary parts are changed.
3.2 Faults to Diagnose

To investigate which faults that need to be considered, Failure Mode Effect Analysis (FMEA) [Stamatis, 1995] and Fault Tree Analysis (FTA) [Haasl et al., 1981] may be successful approaches. Furthermore, as said above, legislations require that all faults in the engine, or in its surroundings, that results in increased emissions must be detected and in some cases also isolated. Much effort is therefore also spent on testing engines in test-cells where faults can be injected and emissions measured, with the objective to see which faults that may lead to increased emissions.

3.2.1 Fault Types

Faults that must be diagnosed in, or around, the engine are for instance faults affecting the fuel injection system, the cooling system, and the gas-flow system, faults in all sensors and actuators, and faults affecting after-treatment systems like the SCR-system and the DPF. Common fault types are electrical faults,
3.3. Residual Generation for Automotive Systems

Due to economical reasons and space limitations, it is not a desired option to mount additional hardware in order to diagnose faults. As said in Chapter 2, one approach to fault diagnosis, providing potentially good performance and in which no additional hardware is needed, is model-based fault diagnosis with residuals.
Chapter 3. Model-Based Fault Diagnosis in Automotive Systems

Figure 3.5: Legislations require lowered OBD thresholds for heavy-duty trucks in the European Union. The line with circle markers shows NOx emission thresholds. The line with dotted markers shows thresholds for PM emissions scaled with a factor 10.

3.3.1 Models

For model-based fault diagnosis, a model of the system under diagnosis is needed. Above we concluded that modern automotive diesel engines as well as their surrounding systems are complex. To describe the dynamic courses in these systems, physical modeling is an often utilized approach, in which models are based on first principles of physics. One popular and successful approach is to use physical object-oriented modeling tools, e.g. Modelica [Fritzon, 2004]. For a large class of technical systems, such as mechanical-, electrical, and chemical systems, these approaches generally result in complex non-linear equation systems containing both algebraic and differential equations, i.e. non-linear DAE-systems. When considering complex automotive systems, it is thus important that the residual generation method is able to handle such models.

3.3.2 Design Process

The view taken in this thesis, as in e.g. [Nyberg and Krysander, 2008], [Nyberg, 1999], is that design of a diagnosis system is a two-step approach, see
3.4. Industrial Relevance

Figure 3.6. In a first step, a large number of candidate residual generators are found, and in a second step the set of residual generators most suitable to be included in the final diagnosis system is picked out. The set of residual generators to be used in the final diagnosis system must in the second step be chosen so that desired fault detection and isolation performance is achieved. To do this, it is necessary to evaluate many candidate residual generators with real measurement data in order to investigate sensitivity to faults in the presence of disturbances, modeling errors, measurement noise, etc. Therefore it is for the second step important that there is a large selection of different candidate residual generators to choose between. Thus, the initial set of candidate residual generators should be as large as possible.

![Diagram](image)

Figure 3.6: Design of a diagnosis system is a two-step approach. In the first step, a large number of candidate residual generators are found, and in the second step the set of residual generators to be used in the final diagnosis system is picked out. This is done by evaluating candidate residual generators with measurement data.

3.3.3 Methods

As argued above, it is desirable that a method for residual generation intended to be used for automotive systems is able to handle DAE-models. In addition, as said in Section 2.5, decoupling of faults is a fundamental problem in residual generation. When aiming at finding as many candidate residual generators as possible, it is also highly desirable that the method used for residual generation is automated. One class of residual generation methods having the potential to handle all of these issues, is sequential residual generation. Sequential residual generation with application to automotive systems is considered in Paper 2.

3.4 Industrial Relevance

As said earlier, model-based diagnosis with residuals is one of many approaches for design of diagnosis systems. The main argument for not using model-
based approaches is the lack of adequate accurate models. It is true that modeling may be time-consuming and also that once a model is created, it must be validated and tuned which may require additional effort and access to measurement data. In addition, models also need to be kept updated to be useful. Therefore, model-based approaches require a well defined engineering process that supports this way of working and takes mentioned aspects into account. However, there are efficient object-oriented modeling tools such as Simulink or Modelica that can be used to facilitate this process, and several established engineering tools supporting model-based development, e.g. Real-Time Workshop. Furthermore, as systems become more and more complex, models are needed for other purposes than diagnosis system development, for example simulation and development of control systems. These models can likewise, perhaps with small modifications, be used for development of diagnosis systems as long as they describe the system under diagnosis. This is for example the case for the models used in the application study in Paper 2, which are developed for simulation purposes.

Another argument is that model-based diagnostic tests based on residuals tend to be hard to run in real-time in computers on-board for example trucks. This is due to severe hardware limitations, in terms of CPU power and memory. The matter of the fact is however that technical systems become more and more advanced and complex. It is therefore reasonable that the hardware on-board these systems evolves in the same pace. The method presented in Paper 2 gives residual generators where variables are computed sequentially. Computing variables in this way is suitable for real-time execution. Further, it is not necessary to run all tests in a diagnosis system at the same time or even in real-time. For example, one set of tests can be run for fault detection and once a fault is detected, another set of tests can be run for fault isolation. If there is memory available, data can be saved and the isolation tests can as well be run later, i.e. not in real-time. In addition, for a given detection and isolation performance, it is not that certain that a diagnosis system developed with a systematic model-based approach requires more CPU power and memory, in comparison with a diagnosis system developed through some “ad-hoc” approach. It is likely that the set of tests contained in the model-based system can be more tailor made, through for example fault decoupling. Moreover, for a given detection and isolation performance, a model-based diagnosis system would probably require fewer sensors than a non model-based diagnosis system, since models instead of hardware are used to provide necessary redundancy. This means an over-all cost reduction. Another aspect is that it is not necessary that all diagnosis is performed on-board, see Section 3.1.3. When diagnosis instead is done off-board, diagnostic tests can be run in ordinary computers stationed in the workshop, and thus limitations in hardware are not an issue.

It is the author’s strong belief that model-based fault diagnosis, or model-based development in general, is a necessity for being able to meet future de-
3.4. Industrial Relevance

mands on safety, reliability, environmental friendliness, and performance of automotive systems. It is believed that usage of model-based approaches for design of diagnosis systems increases productivity and simplifies the overall design process for diagnosis systems, since many steps in the process can be automated. For instance, a large set of diagnostic tests, or residual generators, can be automatically generated given a model with the method presented in Paper 2. If conditions are changed, the model can be updated and a new set of tests can easily be generated, meaning reconfigurability.
Chapter 3. Model-Based Fault Diagnosis in Automotive Systems
Bibliography


Bibliography


Bibliography


use in vehicles, and the emission of gaseous pollutants from positive-ignition engines fuelled with natural gas or liquefied petroleum gas for use in vehicles. ECE-R49.

