Linköping studies in science and technology
Licentiate Thesis. No. 1584

Diagnosability analysis and FDI system design for uncertain systems

Daniel Eriksson

Linköping University
INSTITUTE OF TECHNOLOGY

Department of Electrical Engineering
Linköping University, SE-581 33 Linköping, Sweden

Linköping 2013

Linköping studies in science and technology
Licentiate Thesis. No. 1584

This is a Swedish Licentiate's Thesis.
Swedish postgraduate education leads to a Doctor's degree and/or a Licentiate's degree.
A Doctor's degree comprises 240 ECTS credits (4 years of full-time studies).
A Licentiate's degree comprises 120 ECTS credits,
of which at least 60 ECTS credits constitute a Licentiate's thesis.

Daniel Eriksson
daniel.eriksson@liu.se
www.vehicular.isy.liu.se
Division of Vehicular Systems
Department of Electrical Engineering
Linköping University
SE-581 33 Linköping, Sweden

Eriksson, Daniel
Diagnosability analysis and FDI system design for uncertain systems
ISBN 978-91-7519-652-7
ISSN 0280-7971
LIU-TEK-LIC-2013:18

Typeset with LaTeX 2ε
Printed by LiU-Tryck, Linköping, Sweden 2013

To my family

Abstract

Our society depends on advanced and complex technical systems and machines, for example, cars for transportation, industrial robots in production lines, satellites for communication, and power plants for energy production. Consequences of a fault in such a system can be severe and result in human casualties, environmentally harmful emissions, high repair costs, or economical losses caused by unexpected stops in production lines. Thus, a diagnosis system is important, and in some applications also required by legislations, to monitor the system health in order to take appropriate preventive actions when a fault occurs. Important properties of diagnosis systems are their capability of detecting and identifying faults, i.e., their fault detectability and isolability performance.

This thesis deals with quantitative analysis of fault detectability and isolability performance when taking model uncertainties and measurement noise into consideration. The goal is to analyze diagnosability performance given a mathematical model of the system to be monitored before a diagnosis system is developed. A measure of fault diagnosability performance, called distinguishability, is proposed based on the Kullback-Leibler divergence. For linear descriptor models with Gaussian noise, distinguishability gives an upper limit for the fault to noise ratio of any linear residual generator. Distinguishability is used to analyze fault detectability and isolability performance of a non-linear mean value engine model of gas flows in a heavy duty diesel engine by linearizing the model around different operating points.

It is also shown how distinguishability is used for determine sensor placement, i.e, where sensors should be placed in a system to achieve a required fault diagnosability performance. The sensor placement problem is formulated as an optimization problem, where minimum required diagnosability performance is used as a constraint. Results show that the required diagnosability performance greatly affects which sensors to use, which is not captured if not model uncertainties and measurement noise are taken into consideration.

Another problem considered here is the on-line sequential test selection problem. Distinguishability is used to quantify the performance of the different test quantities. The set of test quantities is changed on-line, depending on the output of the diagnosis system. Instead of using all test quantities the whole time, changing the set of active test quantities can be used to maintain a required diagnosability performance while reducing the computational cost of the diagnosis system. Results show that the number of used test quantities can be greatly reduced while maintaining a good fault isolability performance.

A quantitative diagnosability analysis has been used during the design of an engine misfire detection algorithm based on the estimated torque at the flywheel. Decisions during the development of the misfire detection algorithm are motivated using quantitative analysis of the misfire detectability performance. Related to the misfire detection problem, a flywheel angular velocity model for misfire simulation is presented. An evaluation of the misfire detection algorithm show results of good detection performance as well as low false alarm rate.

Populärvetenskaplig Sammanfattning

Vårt samhälle är idag beroende av avancerade tekniska system, till exempel bilar för transport, industrirobotar vid produktionslinor, satelliter för kommunikation och kraftverk för energiproduktion. Ett fel i något av dessa system kan leda till allvarliga konsekvenser och resultera i att människor skadas, miljöskadliga utsläpp, dyra reparationskostnader eller ekonomiska förluster på grund av oväntade produktionsstopp. Diagnossystem för att övervaka sådana tekniska system är därför viktiga för att kunna identifiera när ett fel inträffar så att lämpliga åtgärder kan vidtas. I vissa fall, till exempel inom fordonsindustrin, finns det även lagkrav på att specifika funktioner hos fordonet måste övervakas av ett diagnossystem.

Ett diagnossystem använder mätsignaler från systemet som ska övervakas för att detektera om fel har uppstått och beräknar sedan möjliga förklaringar på vilka fel som kan finnas i systemet. Finns det en matematisk modell som beskriver systemet går det att jämföra mätsignaler med förväntat beteende givet modellen för att detektera och isolera fel. När ett diagnossystem använder metoder baserade på modeller för att övervaka ett system kallas det för modellbaserad diagnos. Osäkerheter i modellen, mätbrus och var sensorer är placerade i systemet begränsar hur bra diagnosprestanda som kan uppnås av ett diagnossystem. Med hjälp av kunskap om osäkerheter, och var sensorer kan placeras, kan ett diagnossystem konstrueras så att den negativa påverkan av osäkerheterna begränsas.

I denna avhandling analyseras hur ett diagnossystem ska konstrueras med hjälp av information om vad för feldetektions- och felisoleringsprestanda som kan uppnås givet en matematisk modell av systemet som ska övervakas. Diagnosprestanda analyseras kvantitativt genom att ta hänsyn till osäkerheterna i modellen, mätbrus och hur varje fel ser ut. Ett mått för att analysera kvantifierad diagnosprestanda för en given modell av ett system presenteras, som kallas *distinguishability*, och exempel visar hur detta mått kan användas för att analysera diagnosegenskaper givet en modell där modellosäkerheter och mätbrus är kända. Distinguishability har tillämpats bland annat för att hitta den billigaste uppsättning sensorer som uppfyller en önskad diagnosprestanda.

Att kunna analysera och kvantifiera diagnosprestanda under utvecklingen av ett diagnossystem ger möjlighet att välja en design som ger bäst diagnosprestanda. En applikation som analyserats i detta arbete är detektion av missständningar i bensinmotorer. Missständning sker i en cylinder exempelvis på grund av ett trasigt tändstift och orsakar skador på katalysatorn samt förhöjda avgasutsläpp. Detektion av missständningar försvåras bland annat av störningar i drivlinan, variationer i last och hastighet i motorn och fel på mätutrustningen. Ett diagnossystem har utvecklats för att detektera när missständning inträffar utifrån varvtalsmätningar på motorns svänghjul med hjälp av kvantitativ analys av diagnosprestanda för att maximera detektionsprestandan. Dessutom har en modell av drivlinan utvecklats för att kunna simulera mätsignaler från svänghjulet när missständning inträffar.

Acknowledgments

This work has been carried out at the Division of Vehicular Systems at the Department of Electrical Engineering, Linköping University.

First of all I would like to express my gratitude to my supervisor Dr. Erik Frisk and co-supervisor Dr. Mattias Krysander for all guidance and inspiration that you have given me. I appreciate our interesting discussions and your insightful inputs. I also want to acknowledge Dr. Lars Eriksson for all your support during the work with the misfire project.

I want to thank Prof. Lars Nielsen for letting me join his research group. I also want to thank all of my colleagues for the nice atmosphere, all fun and interesting discussions, and the joy when being at work. A special thanks to Lic. Christofer Sundström, Lic. Emil Larsson, and Ylva Jung for help with proofreading parts of the manuscript. I also want to thank Maria Hamnér and Maria Hoffstedt for all help with administrational issues.

I also want to acknowledge Dr. Sasa Trajkovic and his co-workers at Volvo Cars in Torslanda for all help during the work with the misfire detection project. I appreciate all useful inputs and help with data collection during my visits. I also enjoyed the nice time and all good lunch restaurants you brought me to.

I will forever be in dept to my family, especially my parents Bonita and Mats and my brother Mikael, for all your support. If it was not for you I would not have been here today. I also want to thank all of my friends that bring my life much joy and happiness.

Last but not least, I want to express my deep gratitude and love to Ylva Jung for all her support, encouragement, patience, and for being by my side when I need you the most.

Linköping, February 2013
Daniel Eriksson

Contents

1 Introduction **1**
1.1 Fault diagnosis . . . 1
1.1.1 Model based diagnosis . . . 4
1.2 Fault diagnosability analysis . . . 6
1.2.1 Utilizing diagnosability analysis for design of diagnosis systems . . . 7
1.2.2 The Kullback-Leibler divergence . . . 8
1.2.3 Engine misfire detection . . . 9
1.3 Scope . . . 10
1.4 Contributions . . . 12
1.5 Publications . . . 13
References . . . 15

Publications **21**

A A method for quantitative fault diagnosability analysis of stochastic linear descriptor models **23**
1 Introduction . . . 26
2 Problem formulation . . . 27
3 Distinguishability . . . 28
3.1 Reformulating the model . . . 28
3.2 Stochastic characterization of fault modes . . . 31
3.3 Quantitative detectability and isolability . . . 32
4 Computation of distinguishability . . . 34
5 Relation to residual generators . . . 39
6 Diesel engine model analysis . . . 43
6.1 Model description . . . 43
6.2 Diagnosability analysis of the model . . . 45
7 Conclusions . . . 48
References . . . 50

B Using quantitative diagnosability analysis for optimal sensor placement 53
1 Introduction . . . 56
2 Introductory example . . . 56
2.1 Sensor placement using deterministic method . . . 57
2.2 Analysis of minimal sensor sets using distinguishability . . . 58
3 Problem formulation . . . 59
4 Background theory . . . 60
4.1 Model . . . 60
4.2 Quantified diagnosability performance . . . 62
5 The small example revisited . . . 63
6 A greedy search approach . . . 65
7 Sensor placement using greedy search . . . 66
7.1 Model . . . 66
7.2 Analysis of the underdetermined model . . . 67
7.3 Analysis of the exactly determined model . . . 69
8 Conclusion . . . 69
References . . . 71

C A sequential test selection algorithm for fault isolation 73
1 Introduction . . . 76
2 Problem formulation . . . 78
3 Background theory . . . 79
3.1 Distinguishability . . . 79
3.2 Relation of residual generators . . . 81
4 Generalization of distinguishability . . . 81
5 Sequential test selection . . . 83
5.1 Principles . . . 83
5.2 Algorithm . . . 84
6 Case study: DC circuit . . . 87
6.1 System . . . 87
6.2 Diagnosis algorithm . . . 88
6.3 Evaluation . . . 88
7 Tuning the test selection algorithm . . . 93
7.1 Off-line . . . 93
7.2 On-line . . . 93
7.3 Other measures of diagnosability performance . . . 94
8 Conclusion . . . 94
9 Acknowledgment . . . 95
References . . . 96

D Flywheel angular velocity model for misfire simulation 99
1 Introduction . . . 102
2 Model requirements . . . 103
3 Model . . . 104
3.1 Model outline . . . 104
3.2 Engine . . . 105
3.3 Driveline . . . 108
3.4 Modeling disturbances . . . 109
4 Model validation . . . 110
4.1 Experimental data . . . 110
4.2 Validation . . . 110
5 Conclusions . . . 113
References . . . 117

E Analysis and optimization with the Kullback-Leibler divergence for misfire detection using estimated torque 119
1 Introduction . . . 122
2 Vehicle control system signals . . . 123
3 Analysis of the flywheel angular velocity signal . . . 125
4 The Kullback-Leibler divergence . . . 130
5 Torque estimation based on the angular velocity signal . . . 131
5.1 Analyzing misfire detectability performance of estimated torque signal . . . 135
6 An algorithm for misfire detection . . . 144
6.1 Algorithm outline . . . 144
6.2 Design of test quantity . . . 144
6.3 Thresholding . . . 149
7 Evaluation of the misfire detection algorithm . . . 151
8 Conclusions . . . 159
9 Future works . . . 159
10 Acknowledgment . . . 160
References . . . 161

Chapter 1

Introduction

Many parts in our society depend on advanced and complex technical systems and machines, for example, cars for transportation, industrial robots in production lines, satellites for communication, and power plants for energy production. Consequences of a *fault* in such a system can be severe and result in human casualties, environmentally harmful emissions, high repair costs, or economical losses caused by unexpected stops in production lines. Thus, a *diagnosis system* is important, and in some applications also required by legislations, to monitor the system health and detect faults in order to take appropriate preventive measures.

This thesis addresses the issue of analyzing fault diagnosability performance by taking model uncertainties and measurement noise into consideration. The information from a diagnosability analysis is utilized in the development process of a diagnosis system, before the actual diagnosis system has been developed, to improve the final performance. As an application, an engine misfire detection algorithm is developed using methods for quantitative analysis of fault diagnosability performance. Related to the misfire detection problem, a model for simulating the flywheel angular velocity signal when affected by misfires is presented.

1.1 Fault diagnosis

A diagnosis system uses information from sensors and actuators of the monitored system to detect abnormal behaviors caused by a fault in the system. Detecting if there is a fault in the system without locating the root cause is referred to as *fault detection*. In some applications, a diagnosis system that is only able to detect faults is not sufficient. For example, different faults might require different types of actions, thus requiring that the diagnosis system is able to make a

correct identification of which faults that are present in the system. Identifying the faults in the system is referred to as *fault isolation*. In this thesis, when considering both fault detection and isolation is referred to as *fault diagnosis*.

The following example is used to introduce the principles of fault detection and isolation. The example will also be referred back to later in this chapter.

Example 1 (Fault detection and isolation)**.** *Consider two thermometers, y_1 and y_2, measuring the outdoor temperature T,*

$$\begin{aligned} y_1 &= T \\ y_2 &= T. \end{aligned}$$

If the two thermometers are working properly they should show the same temperature, i.e.

$$y_1 - y_2 = 0.$$

Assume that the thermometer y_1 is faulty showing the wrong temperature

$$y_1 = T + f$$

where $f \neq 0$ represents the sensor deviation from the true temperature T. Since the two thermometers show different temperatures,

$$y_1 - y_2 \neq 0. \tag{1.1}$$

By comparing the two sensors, a fault in any of them can be detected when the difference is non-zero. However, it is not possible to isolate which of the sensors that is faulty because (1.1) *will be non-zero in either case.*

Assume that there is a third thermometer $y_3 = T$. By comparing all thermometers pairwise as in (1.1) *gives that*

$$\begin{aligned} y_1 - y_2 &\neq 0 \\ y_1 - y_3 &\neq 0 \\ y_2 - y_3 &= 0, \end{aligned} \tag{1.2}$$

i.e., it is only when comparing y_1 to any other thermometer that the difference is non-zero. By comparing the outputs of the three thermometers, the sensor y_1 can be isolated as the faulty thermometer. However, note that the case where the two thermometers y_2 and y_3 are faulty and measuring the same faulty temperature is also consistent to the observations in (1.2)*. Thus, the observations* (1.2) *can be explained by either a fault in y_1 or two faults in y_2 and y_3, where the first case is the true scenario in this example. This means that there can be more than one diagnosis in a given situation.*

Monitoring the system health requires general knowledge of how the system works and behaves, in order to detect when something is not working properly. Knowledge about a system can be incorporated into a mathematical model

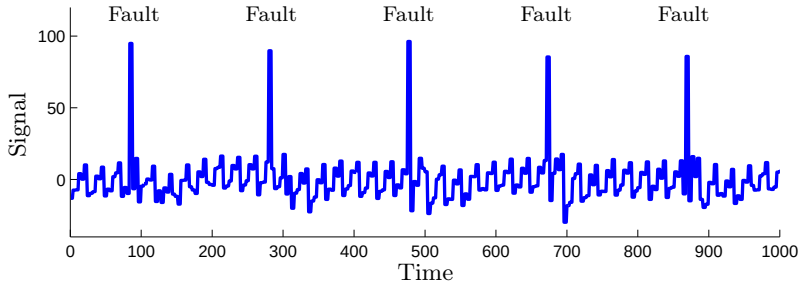

Figure 1.1: A signal with little noise compared to the amplitudes of the intermittent faults.

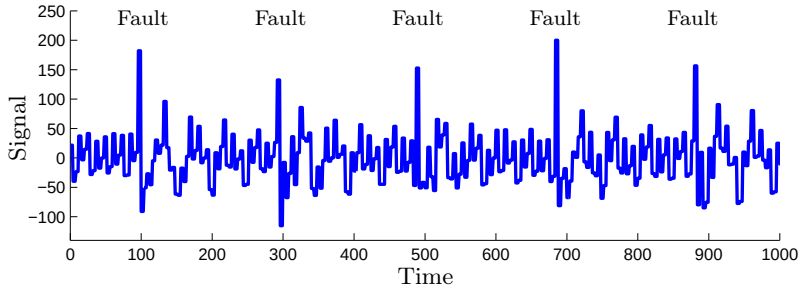

Figure 1.2: A signal with much noise compared to the amplitudes of the intermittent faults.

of the system based on, for example, physical principles, measurements, and experience. If a mathematical model of the system is available, sensor data can be compared to predicted behavior given by the model, actuators to the system, and other sensors in order to detect and isolate faults. Fault diagnosis when utilizing models of the system is referred to as *model based diagnosis*.

Even though a model is able to describe the general behavior of the system, it is seldom perfect. Also, the sensor outputs are not necessarily the same as the values of the measured states of the system. *Model uncertainties* and *measurement noise* are common issues when working with applications utilizing models of systems. For fault diagnosis purposes, a model should be able to distinguish between model uncertainties and an actual fault of a required minimum magnitude. A more accurate model with less uncertainties will increase the possibility of detecting even smaller faults. Two examples, where model uncertainties and noise affects the fault diagnosability performance, are shown in Figure 1.1 and Figure 1.2. The two figures show two noisy signals with intermittent faults. In both signals, it is possible to detect the faults, which are visible as high peaks in the signals. However, it is more difficult to distinguish the intermittent

faults from the noise in Figure 1.2 since the ratio of the noise amplitude and the amplitudes of the intermittent faults in Figure 1.1 is relatively small compared to the signal in Figure 1.2. Thus, considering model uncertainties and measurement noise is important when developing a diagnosis system because it will affect the fault detectability and isolability performance.

There are many factors to consider when developing a diagnosis system. The purpose of the diagnosis system determines the requirements of the fault diagnosability performance. To be able to predict how difficult it will be to detect or isolate a certain fault at an early stage of the development of the diagnosis system, can save lots of development time and money. It might be necessary to add new sensors or hardware to meet the diagnosability requirements. At the same time, identifying unnecessary sensors and using a smart design of the diagnosis algorithm, which reduces the required computational power, can reduced hardware costs. Hardware changes are possible to deal with early in the development process of a product but more complicated and expensive to deal with later or once the product is manufactured. Thus, methods for evaluating diagnosability performance early in the development process are important to efficiently obtain a diagnosis system with satisfactory performance.

1.1.1 Model based diagnosis

Fault diagnosis of technical systems covers many different approaches from different fields. This thesis will focus on model based diagnosis, i.e., there exists a mathematical model describing the system. A trivial example of a mathematical model is shown in Example 1, where it is assumed that all thermometers y_1, y_2, and y_3 measure the same temperature.

The term fault detection and isolation, FDI, often relates to model based diagnosis methods founded in control theory and focuses on the application of *residual generators* for fault detection, see for example Gertler (1991), Isermann (1997), Gertler (1998), Chen and Patton (1999), Isermann (2005), Blanke et al. (2006), Gustafsson (2000), and Patton et al. (2010). A residual is a function of known signals and is zero in the fault-free case. The three pairwise comparisons of measured temperatures in (1.2) are simple examples of residuals.

Within the field of artificial intelligence, model based diagnosis, DX, focuses more on fault isolation and the use of logics to identify faulty behavior, see for example Reiter (1987), de Kleer and Williams (1987), Feldman and van Gemund (2006), and de Kleer (2011). In this thesis, diagnosis systems are considered where fault detection is mainly performed using methods from the FDI community and fault isolation is performed using methods from the DX community, see for example Cordier et al. (2004).

Other approaches for fault diagnosis not considered here are, for example, data-driven methods such as PCA and neural networks, see Qin (2012) and Venkatasubramanian et al. (2003), and probabilistic approaches such as Bayesian networks, see Pernestål (2009) and Lerner et al. (2000).

A schematic overview of a typical diagnosis system considered in this thesis

is shown in Figure 1.3. The diagnosis algorithm takes observations from the monitored system as input and computes possible statements about the system health that are consistent with the observations, called *diagnoses*, see Gertler (1991) and de Kleer and Williams (1987).

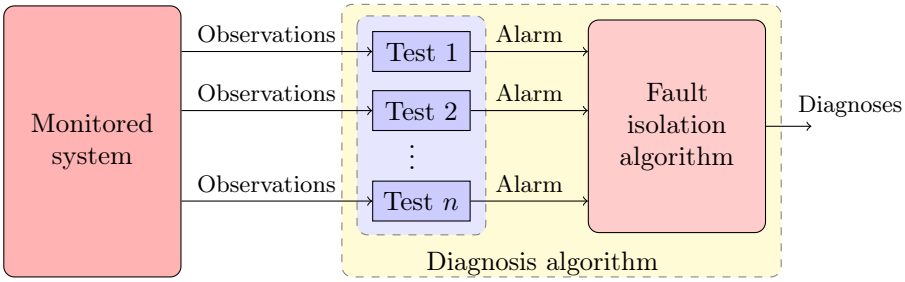

Figure 1.3: An example of a diagnosis algorithm where observations from the system are used to compute possible statements about the system health consistent with the observations, referred to as diagnoses.

In Figure 1.3, the diagnosis system uses a set of tests to detect when there is a fault present in the system. A test typically consists of a test quantity, e.g., a residual, and a decision logic for triggering an alarm. The test quantity T can be described as a function of known signals z, such as sensor outputs and known actuators, and indicate if there is a fault present or not. If a mathematical model of the system is available, a test quantity can be designed, for example, based on a residual generator which uses observations from the system to compare model predictions with measurements, see Blanke et al. (2006), Chen and Patton (1999), Patton et al. (2010), Svärd (2012), and Frisk (2001).

To detect if there is a fault in the system, the test quantity T is evaluated and compared to, for example, a threshold J. The test will generate an alarm if the value of the test quantity exceeds the threshold, i.e, if

$$T(z) > J$$

where z is a vector of known signals. If there is no fault in the system, the test quantity T should be below the threshold J, and above the threshold when there is a fault. If the test alarms even though there is no fault in the system, it is referred to as a *false alarm*, and if the test do not alarm when there is a fault is referred to as a *missed detection*. Each test can be seen as a hypothesis test where the null hypothesis is the fault-free case or a nominal case, see Nyberg (1999) and Nyberg (2001).

Depending on how each test quantity is designed, different test quantities can be sensitive to different sets of faults, i.e., each test quantity is designed to not detect all faults. Then, fault isolation can be performed by using the knowledge that different test quantities are sensitive to different sets of faults. A fault

isolation algorithm, see Figure 1.3, combines the information of all triggered tests to compute statements about the system health which are consistent with the triggered tests, see Gertler (1991), de Kleer and Williams (1987), Nyberg (2006), and de Kleer (2011). In Example 1, the three residuals (1.2) are sensitive to different sets of faults given by which sensors that are used in each residual. By combining the information from the residuals that are non-zero, two possible diagnoses are a single fault in y_1 or multiple faults in y_2 and y_3.

1.2 Fault diagnosability analysis

Analyzing fault diagnosability performance is an important part for the design and evaluation of diagnosis systems. Before the development of the diagnosis system, an analysis can give an understanding of how good performance that can be achieved given the model of the system. This knowledge can be used, for example, when designing the diagnosis system or specific test quantities, or to decide if more sensors are needed. Here, a description of existing measures and methods for analyzing diagnosability performance is presented.

Two important properties in fault diagnosis when evaluating the diagnosability performance are fault detectability and fault isolability. Fault detectability and isolability can be evaluated, both for a given model and for a given diagnosis system. Evaluating fault detectability and isolability performance for a given diagnosis system are described in, e.g., Chen and Patton (1999).

Measures used in classical detection theory for analyzing detectability performance of specific tests are probabilities of detection, false alarm, and missed detection, see Basseville and Nikiforov (1993) and Kay (1998). Other methods for analyzing the performance of tests are, for example, Receiver operating characteristics, or ROC curves, see Kay (1998), and power functions, see Casella and Berger (2001). Computing the probabilities requires that the distributions of the test quantity for the fault-free case and faulty case are known, or that realistic approximations are available. These methods take the uncertainties into consideration and give a quantitative measure of the fault detectability performance. However, they do not consider fault isolability performance.

Methods for analyzing fault detectability and isolability performance for a given diagnosis system, using probabilistic measures, are found in several works, e.g., Wheeler (2011), Krysander and Nyberg (2008), Chen and Patton (1996), Willsky and Jones (1976), and Emami-Naeini et al. (1988). The methods in these works take model uncertainties and measurement noise into consideration. However, the methods analyze diagnosability performance for a given diagnosis system and not for a given model.

Methods for fault diagnosability analysis for a given model mainly considers fault detectability and isolability. Fault detectability and isolability analysis of linear systems can be found in, for example, Nyberg (2002). In Krysander (2006), Trave-Massuyes et al. (2006), Ding (2008), Dustegör et al. (2006), and Frisk et al. (2012), fault detectability and isolability analysis is performed by

analyzing a structural representation of the model equations which enables the analysis of non-linear systems. However, fault detectability and isolability are in these works analyzed as deterministic properties of the models and do not take the behavior of the fault, model uncertainties, or measurement noise into consideration. The two signals in Figure 1.1 and Figure 1.2 show an example where diagnosability performance is affected by noise. In both figures, the intermittent faults are detectable, even though it is not as easy to detect them in each case. To answer questions like: "How difficult is it to isolate a fault f_i from another fault f_j?", when analyzing a model of the system, a method for analyzing fault diagnosability performance is required where model uncertainties and measurement noise are taken into consideration.

1.2.1 Utilizing diagnosability analysis for design of diagnosis systems

In this section, some examples are presented where fault diagnosability analysis can be utilized during the design process of a diagnosis system. This thesis addresses the optimal sensor placement problem and on-line sequential test selection problem.

The possibility of designing test quantities to detect and isolate a certain fault, depends on which sensors that are available. Depending on where a fault occurs in the system, it is not be possible to design a test quantity that is able to detect the fault if the effects of the fault are not measured by a sensor. Thus, having a suitable set of sensors which is able to monitor the whole system is an important aspect when designing a diagnosis system. One example is given in Example 1, where a third thermometer is required to isolate a fault in one of the thermometers. Finding a set of sensors that fulfills a required fault detectability and isolability performance is often referred to as the sensor placement problem, see e.g. Raghuraj et al. (1999), Trave-Massuyes et al. (2006) Krysander and Frisk (2008), Frisk et al. (2009), and Rosich (2012).

If there is a large number of test quantities in the diagnosis algorithm maybe not all test quantities are necessary to be active all the time. Using all test quantities can be computationally expensive and test quantities with poor detectability performance that are mainly designed for isolating faults are not useful unless a fault is detected. By sequentially updating which test quantities to be used by the diagnosis algorithm, the computational power can be reduced while maintaining sufficient fault isolability performance. This is referred to as the on-line sequential test selection problem in Krysander et al. (2010).

There are several examples where diagnosability analysis methods are applied during the design of the diagnosis system. For example, the sensor placement problem is often formulated as to find a minimum set of sensors which achieves a minimum required fault detectability and isolability performance, see Rosich (2012), Raghuraj et al. (1999), Commault et al. (2008), Trave-Massuyes et al. (2006), and Frisk et al. (2009). Another example is the design and selection of test quantities such that the diagnosis system fulfills a required fault detectability

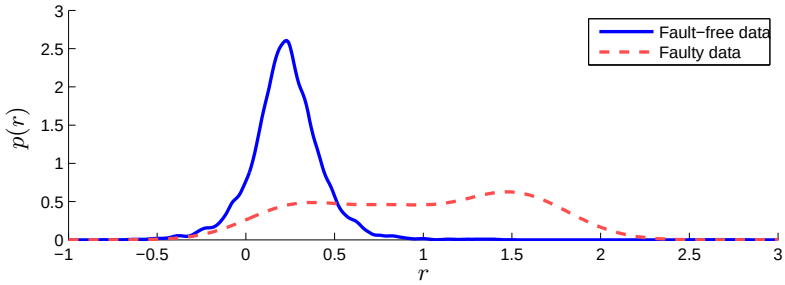

Figure 1.4: Probability density functions of the output y of a residual generator r for the fault-free case and the faulty case.

and isolability performance, see Staroswiecki and Comtet-Varga (2001), Svärd (2012), Krysander (2006), and Rosich et al. (2012).

1.2.2 The Kullback-Leibler divergence

One important part in this thesis is the use of the Kullback-Leibler divergence, see Kullback, S. and Leibler, R. A. (1951), to quantify diagnosability performance given a model or diagnosis system. This section, motivates why the Kullback-Leibler divergence is suitable for quantitative analysis of fault diagnosability performance.

Consider the task of determine if a signal, e.g., a residual, r is useful for detecting a specific fault f. The performance of a diagnosis system, or a single test, depend on many different design parameters, such as which tests to use and how to design the thresholds. To make such choices early in the development process is not desirable if the design of the whole diagnosis algorithm, or a single test, is not yet determined. Thus, the analysis of r should separate the performance of r and the performance of the test based on r. Figure 1.4 shows the probability density functions of the output of r when no fault is present, $p_{\text{no fault}}(r)$, and when a fault f is present, $p_{\text{fault}}(r)$. The value of r lies around 0.25 with little variance in the fault-free case but when a fault is present the output lies around one with higher variance.

One way to measure the performance of a residual r is to evaluate how likely it is that a value of r comes from the fault-free distribution when a fault is present. The most powerful test, for any given probability of false alarm determined by J, is the likelihood ratio test by using the Neyman-Pearson lemma, see e.g., Casella and Berger (2001). If r has the distribution $p_{\text{no fault}}$ in the fault-free case and p_{fault} when a fault is present, then the likelihood ratio test, which rejects $H_0 : p_{\text{no fault}}(r)$ in favor of the alternative hypothesis $H_0 : p_{\text{no fault}}(r)$, can be written as

$$\Lambda(r) = \frac{p_{\text{fault}}(r)}{p_{\text{no fault}}(r)} \geq J. \tag{1.3}$$

The likelihood ratio tells how much more likely the value of r comes from the faulty case than from the fault-free case. A high ratio corresponds to that a fault is more likely. Thus, a high likelihood is good if a fault is present. Here, the log-likelihood ratio,

$$\log \frac{p_{\text{fault}}(r)}{p_{\text{no fault}}(r)}, \tag{1.4}$$

is considered instead of the likelihood ratio. For different values of r, the log-likelihood ratio is positive if a fault is more likely than the fault-free case and negative if the fault-free case is more likely. In the example in Figure 1.4, the log-likelihood ratio is positive for $r > 0.5$ and negative for $r < 0.5$.

The performance of r for detecting a fault f can be quantified by computing the expected value of the log-likelihood ratio (1.4) when a fault is present. If r in general has a large log-likelihood ratio in the faulty case, it should be easy to select a threshold with high probability of detecting the fault while having a low false alarm rate compared to if the log-likelihood ratio is small. The expected value of (1.4) when a fault is present can be written as

$$E_{p_{\text{fault}}}\left[\log\left(\frac{p_{\text{fault}}(r)}{p_{\text{no fault}}(r)}\right)\right] = \int_{-\infty}^{\infty} p_{\text{fault}}(x) \log \frac{p_{\text{fault}}(x)}{p_{\text{no fault}}(x)} dx \tag{1.5}$$

where $E_p[q(x)]$ is the expected value of the function $q(x)$ when the distribution of x is given by p. Equation (1.5) is known as the Kullback-Leibler divergence from p_{fault} to $p_{\text{no fault}}$ denoted $K(p_{\text{fault}} \| p_{\text{no fault}})$, see Kullback, S. and Leibler, R. A. (1951) and Eguchi and Copas (2006).

The Kullback-Leibler divergence can be related to the expected value of how many times more likely that the output of r comes from the faulty case than the fault-free case when a fault is present. If r is not sensitive to the fault f, the probability density functions p_{fault} and $p_{\text{no fault}}$ are equal which gives that (1.5) is zero. The benefit of using the Kullback-Leibler divergence is that the distributions of r, i.e., model uncertainties and measurement noise, are taken into consideration when analyzing how difficult it is to detect a fault.

Some examples of recent works where the Kullback-Leibler divergence has been applied in fault diagnosis applications are Carvalho Bittencourt (2012) and Svärd (2012). In Carvalho Bittencourt (2012), the Kullback-Leibler divergence is used for condition monitoring of industrial robots, and in Svärd (2012) for detecting faults in a diesel engine when the fault-free distribution varies for different operating points of the engine. An application related to fault diagnosis is change detection where the Kullback-Leibler divergence is used to measure if a change has occurred, see for example Takeuchi and Yamanishi (2006) and Afgani et al. (2008).

1.2.3 Engine misfire detection

In practice, the freedom when designing a diagnosis system is often limited by computational power and which sensors that are available. Feedback from fault

diagnosability analysis can be helpful in order to design a diagnosis system using all available information as good as possible to detect and isolate faults.

In the automotive industry, the on-board diagnosis (OBDII) legislations require that many systems of a vehicle are monitored on-line in order to detect if a fault occurs. An overview of automotive diagnosis research is found in Mohammadpour et al. (2012). One example of an automotive diagnosis application is engine misfire detection.

Misfire refers to an incomplete combustion inside a cylinder and can be caused by many different factors, for example a fault in the ignition system, see Heywood (1988). Misfire detection is an important part of the OBDII legislations in order to reduce exhaust emissions and avoid damage to the catalytic converters. The legislations require that the on-board diagnosis system is able to both detect misfires and identify in which cylinder the misfire occurred, see Heywood (1988) and Walter et al. (2007).

The OBDII legislations define strict requirements in terms of the amount of allowed missed misfire detections. Also, to avoid unnecessary visits to the garage and annoyed customers, requires that the number of false alarms is minimized. To fulfill both conditions is difficult and impose tough requirements on the development and tuning of the misfire detection algorithm.

There are several approaches to detect misfires using different types of sensors, e.g., ion current sensors, see Lundström and Schagerberg (2001), or crankshaft angular velocity measured at the flywheel, see Osburn et al. (2006), Naik (2004), and Tinaut et al. (2007). Misfire detection based on torque estimation using the flywheel angular velocity signal has been studied in, e.g, Connolly and Rizzoni (1994), Kiencke (1999), and Walter et al. (2007). A picture of a flywheel is shown in Figure 1.5, where a Hall effect sensor is mounted close to the flywheel and triggers when the punched holes, or teeth, passes the sensor. The flywheel angular velocity signal measures the time it takes between two holes on the flywheel to pass the Hall sensor. An example of the flywheel angular velocity signal and the effect of a misfire is shown in Figure 1.6. A quantitative analysis of diagnosability performance can be used when designing a test quantity based on the flywheel angular velocity signal to improve misfire detectability performance.

Detecting misfires is a non-trivial problem which is complicated by, for example, changes in load, speed, and flywheel manufacturing errors, see Naik (2004) and Kiencke (1999). Flywheel errors results in that the measured time periods between the holes, or teeth, are measured over nonuniform angular intervals. The flywheel errors varies for different vehicles and this must be compensated for by the misfire detection algorithm, see e.g. Kiencke (1999).

1.3 Scope

This thesis deals with quantitative analysis of fault detectability and isolability performance when taking model uncertainties and measurement noise into consideration. The fault diagnosability performance is analyzed for a given

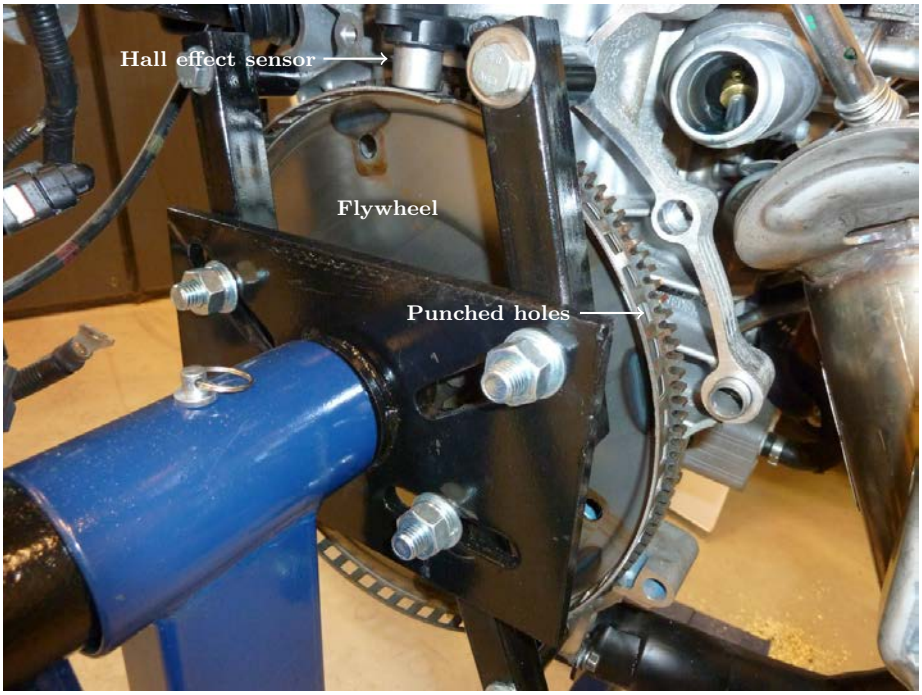

Figure 1.5: A picture of a flywheel and the Hall effect sensor.

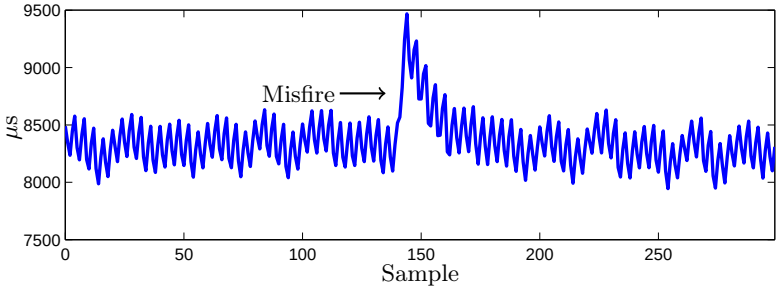

Figure 1.6: Flywheel angular velocity signal when a misfire occurs.

model of the system. The purpose of the quantitative analysis to utilize the information of achievable diagnosability performance during the design of the diagnosis system to improve the diagnosability performance. In this thesis, a quantitative measure of fault detectability and isolability performance *given a model* is proposed. The quantitative diagnosability measure is applied to the *sensor placement problem*, i.e, where should sensors be placed in a system to achieve a required fault diagnosability performance, and *on-line sequential test selection*, i.e., updating the diagnosis system on-line to maintain a sufficiently good fault isolability performance while reducing the computational cost.

Quantitative diagnosability analysis has also been applied during the design of an *engine misfire detection algorithm*. Decisions during the development of the misfire detection algorithm are motivated using quantitative analysis of the misfire detectability performance. Related to the engine misfire detection application, a flywheel angular velocity model for misfire simulation is presented.

1.4 Contributions

The main contributions of Paper A - E are summarized below.

Paper A

Paper A is an extended version of Eriksson et al. (2011a) and Eriksson et al. (2011b). The main contribution is the definition of *distinguishability*, based on the Kullback-Leibler divergence, which is used for quantitative fault detectability and isolability analysis of stochastic models. The second main contribution is the connection between distinguishability and linear residual generators for linear descriptor models with Gaussian noise.

Paper B

The main contribution of Paper B is the use of distinguishability for optimal sensor placement in time-discrete linear descriptor systems with Gaussian noise. The sensor placement problem is formulated as an optimization problem, where required fault detectability and isolability performance is taken into consideration by using minimum required diagnosability performance as a constraint.

Paper C

Paper C proposes an on-line test selection algorithm where the active set of residuals is updated depending on the present output of the diagnosis system. The main contribution is that the performance of each residual is evaluated using distinguishability and included in the test selection algorithm. The test selection problem is formulated as a minimal hitting set problem where the best residuals are selected to detect or isolate any present faults. The second contribution is a

generalization of distinguishability to quantify fault isolability of one fault from multiple faults.

Paper D

Paper D is an initial work considering the engine misfire detection problem where a model of the crankshaft and driveline for misfire analysis is developed. The main contribution in Paper D is a flywheel angular velocity model for misfire simulation. The model is a multi-mass model of the crankshaft and driveline where the cylinder pressure is computed using an analytical model in order to model cylinder variations and misfire.

Paper E

Paper E presents an engine misfire detection algorithm based on torque estimation using the flywheel angular velocity signal. The contribution is a misfire detection algorithm based on the estimated torque at the flywheel. A second contribution is the use of the Kullback-Leibler divergence for analysis and optimization of misfire detection performance.

1.5 Publications

The following papers have been published.

Journal papers

- Daniel Eriksson, Erik Frisk, and Mattias Krysander. A method for quantitative fault diagnosability analysis of stochastic linear descriptor models. *Automatica* (Accepted for publication). **(Paper A)**

Conference papers

- Daniel Eriksson, Mattias Krysander, and Erik Frisk. Using quantitative diagnosability analysis for optimal sensor placement. In *Proceedings of the 8th IFAC Safe Process*. Mexico city, Mexico, 2012. **(Paper B)**

- Daniel Eriksson, Erik Frisk, and Mattias Krysander. A sequential test selection algorithm for fault isolation. In *Proceedings to the 10th European Workshop on Advanced Control and Diagnosis*. Copenhagen, Denmark, 2012. **(Paper C)**

- Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative Fault Diagnosability Performance of Linear Dynamic Descriptor Models. In *Proceedings of the 22nd International Workshop on Principles of Diagnosis* (DX-11). Murnau, Germany, 2011.

- Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative Stochastic Fault Diagnosability Analysis. In *Proceedings of the 50th IEEE Conference on Decision and Control*. Orlando, Florida, USA, 2011.

- Erik Almqvist, Daniel Eriksson, Andreas Lundberg, Emil Nilsson, Niklas Wahlström, Erik Frisk, and Mattias Krysander. Solving the ADAPT Benchmark Problem - A Student Project Study. In *Proceedings of the 21st International Workshop on Principles of Diagnosis* (DX-10). Portland, Oregon, USA, 2010.

Technical reports

- Daniel Eriksson, Lars Eriksson, Erik Frisk, and Mattias Krysander. Analysis and optimization with the Kullback-Leibler divergence for misfire detection using estimated torque. *Technical Report LiTH-ISY-R-3057*. Department of Electrical Engineering, Linköpings Universitet, SE-581 83 Linköping, Sweden, 2013. **(Paper E)**

Submitted

- Daniel Eriksson, Lars Eriksson, Erik Frisk, and Mattias Krysander. Flywheel angular velocity model for misfire simulation. Submitted to *7th IFAC Symposium on Advances in Automotive Control*. Tokyo, Japan, 2013. **(Paper D)**

References

Mostafa Afgani, Sinan Sinanovic, and Harald Haas. Anomaly detection using the Kullback-Leibler divergence metric. In *Applied Sciences on Biomedical and Communication Technologies, 2008. ISABEL'08. First International Symposium on*, pages 1–5. IEEE, 2008.

Michèle Basseville and Igor V. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

Mogens Blanke, Michel Kinnaert, Jan Lunze, Marcel Staroswiecki, and J. Schröder. *Diagnosis and Fault-Tolerant Control*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

André Carvalho Bittencourt. *On Modeling and Diagnosis of Friction and Wear in Industrial Robots*. Licentiate thesis, Linköping University, Automatic Control, The Institute of Technology, 2012.

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Resource Center, Pacific Grove, CA, 2001.

C.J. Chen and R.J. Patton. *Robust Model-Based Fault Diagnosis For Dynamic Systems*. Kluwer International Series on Asian Studies in Computer and Information Science, 3. Kluwer, 1999.

J. Chen and R.J. Patton. Optimal filtering and robust fault diagnosis of stochastic systems with unknown disturbances. *Control Theory and Applications, IEE Proceedings* -, 143(1):31 –36, jan 1996.

Christian Commault, Jean-Michel Dion, and Sameh Yacoub Agha. Structural analysis for the sensor location problem in fault detection and isolation. *Automatica*, 44(8):2074 – 2080, 2008.

Francis T. Connolly and Giorgio Rizzoni. Real time estimation of engine torque for the detection of engine misfires. *Journal of Dynamic Systems, Measurement, and Control*, 116(4):675–686, 1994.

M.-O. Cordier, P. Dague, F. Levy, J. Montmain, M. Staroswiecki, and L. Trave-Massuyes. Conflicts versus analytical redundancy relations: a comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 34(5):2163 –2177, oct 2004.

Johan de Kleer. Hitting set algorithms for model-based diagnosis. 22nd International Workshop on Principles of Diagnosis (DX-11), Murnau, Germany, 2011.

Johan de Kleer and Brian C. Williams. Diagnosing Multiple Faults. *Artif. Intell.*, 32(1):97–130, 1987.

S.X. Ding. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Springer, 2008.

Dilek Dustegör, Erik Frisk, Vincent Coquempot, Mattias Krysander, and Marcel Staroswiecki. Structural analysis of fault isolability in the DAMADICS benchmark. *Control Engineering Practice*, 14(6):597–608, 2006.

Shinto Eguchi and John Copas. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *J. Multivar. Anal.*, 97:2034–2040, October 2006.

A. Emami-Naeini, M.M. Akhter, and S.M. Rock. Effect of model uncertainty on failure detection: the threshold selector. *Automatic Control, IEEE Transactions on*, 33(12):1106 –1115, dec. 1988.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative stochastic fault diagnosability analysis. In *50th IEEE Conference on Decision and Control*, Orlando, Florida, USA, 2011a.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative fault diagnosability performance of linear dynamic descriptor models. 22nd International Workshop on Principles of Diagnosis (DX-11), Murnau, Germany, 2011b.

Alexander Feldman and Arjan van Gemund. A two-step hierarchical algorithm for model-based diagnosis. In *Proceedings of the 21st national conference on Artificial intelligence - Volume 1*, AAAI'06, pages 827–833. AAAI Press, 2006.

Erik Frisk. *Residual Generation for Fault Diagnosis*. PhD thesis, Linköpings Universitet, November 2001.

Erik Frisk, Mattias Krysander, and Jan Åslund. Sensor placement for fault isolation in linear differential-algebraic systems. *Automatica*, 45(2):364–371, 2009.

Erik Frisk, Anibal Bregon, Jan Åslund, Mattias Krysander, Belarmino Pulido, and Gautam Biswas. Diagnosability analysis considering causal interpretations for differential constraints. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 42(5):1216–1229, September 2012.

Janos Gertler. Analytical redundancy methods in fault detection and isolation. In *IFAC Fault Detection, Supervision and Safety for Technical Processes*, pages 9–21, Baden-Baden, Germany, 1991.

Janos Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker Inc., Upper Saddle River, NJ, USA, 1998.

F. Gustafsson. *Adaptive filtering and change detection*. Wiley, 2000.

J.B. Heywood. *Internal combustion engine fundamentals*. McGraw-Hill series in mechanical engineering. McGraw-Hill, 1988.

R. Isermann. Supervision, fault-detection and fault-diagnosis methods - An introduction. *Control Engineering Practice*, 5(5):639 – 652, 1997.

R. Isermann. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer, 2005.

Steven M. Kay. *Fundamentals of statistical signal processing: Detection theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998.

Uwe Kiencke. Engine misfire detection. *Control Engineering Practice*, 7(2):203 – 208, 1999.

Mattias Krysander. *Design and Analysis of Diagnosis Systems Using Structural Methods*. PhD thesis, Linköpings universitet, June 2006.

Mattias Krysander and Erik Frisk. Sensor placement for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 38(6):1398–1410, 2008.

Mattias Krysander and Mattias Nyberg. Statistical properties and design criterions for fault isolation in noisy systems. 19th International Workshop on Principles of Diagnosis (DX-08), Sydney, Australia, 2008.

Mattias Krysander, Fredrik Heintz, Jacob Roll, and Erik Frisk. FlexDx: A reconfigurable diagnosis framework. *Engineering Applications of Artificial Intelligence*, 23(8):1303–1313, October 2010.

Kullback, S. and Leibler, R. A. On Information and Sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

Uri Lerner, Ronald Parr, Daphne Koller, and Gautam Biswas. Bayesian Fault Detection and Diagnosis in Dynamic Systems. In Henry A. Kautz and Bruce W. Porter, editors, *AAAI/IAAI*, pages 531–537. AAAI Press / The MIT Press, 2000.

D. Lundström and S. Schagerberg. Misfire Detection for Prechamber SI Engines Using Ion-Sensing and Rotational Speed Measurements. *SAE Technical Paper 2001-01-0993*, 2001.

J Mohammadpour, M Franchek, and K Grigoriadis. A survey on diagnostic methods for automotive engines. *International Journal of Engine Research*, 13 (1):41–64, 2012.

Sanjeev Naik. Advanced misfire detection using adaptive signal processing. *International Journal of Adaptive Control and Signal Processing*, 18(2):181–198, 2004.

Mattias Nyberg. *Model Based Fault Diagnosis: Methods, Theory, and Automotive Engine Applications*. PhD thesis, Linköpings Universitet, June 1999.

Mattias Nyberg. A general framework for model based diagnosis based on statistical hypothesis testing (revised version). 12:th International Workshop on Principles of Diagnosis, pages 135–142, Sansicario, Via Lattea, Italy, 2001.

Mattias Nyberg. Criterions for detectability and strong detectability of faults in linear systems. *International Journal of Control*, 75(7):490–501, May 2002.

Mattias Nyberg. A fault isolation algorithm for the case of multiple faults and multiple fault types. In *Proceedings of IFAC Safeprocess'06*, Beijing, China, 2006.

Andrew W. Osburn, Theodore M. Kostek, and Matthew A. Franchek. Residual generation and statistical pattern recognition for engine misfire diagnostics. *Mechanical Systems and Signal Processing*, 20(8):2232 – 2258, 2006.

Ron J. Patton, Paul M. Frank, and Robert N. Clark. *Issues of Fault Diagnosis for Dynamic Systems*. Springer Publishing Company, Incorporated, 1st edition, 2010.

Anna Pernestål. *Probabalistic Fault Diagnosis with Automotive Applications*. PhD thesis, Linköping University, 2009.

S. Joe Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2):220 – 234, 2012.

Rao Raghuraj, Mani Bhushan, and Raghunathan Rengaswamy. Locating sensors in complex chemical plants based on fault diagnostic observability criteria. *AIChE Journal*, 45(2):310 – 322, 1999.

R Reiter. A theory of diagnosis from first principles. *Artif. Intell.*, 32(1):57–95, April 1987.

A. Rosich, E. Frisk, J. Aslund, R. Sarrate, and F. Nejjari. Fault diagnosis based on causal computations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 42(2):371 –381, march 2012.

Albert Rosich. Sensor placement for fault detection and isolation based on structural models. 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Process, Safeprocess'12, Mexico City, Mexico, 2012.

M. Staroswiecki and G. Comtet-Varga. Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica*, 37(5): 687 – 699, 2001.

Carl Svärd. *Methods for Automated Design of Fault Detection and Isolation Systems with Automotive Applications*. PhD thesis, Linköping University, 2012.

J. Takeuchi and K. Yamanishi. A unifying framework for detecting outliers and change points from time series. *Knowledge and Data Engineering, IEEE Transactions on*, 18(4):482 – 492, april 2006.

Francisco V. Tinaut, Andrés Melgar, Hannes Laget, and José I. Domínguez. Misfire and compression fault detection through the energy model. *Mechanical Systems and Signal Processing*, 21(3):1521 – 1535, 2007.

L. Trave-Massuyes, T. Escobet, and X. Olive. Diagnosability Analysis Based on Component-Supported Analytical Redundancy Relations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 36(6): 1146 –1160, nov 2006.

Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, and Kewen Yin. A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers and Chemical Engineering*, 27(3): 327 – 346, 2003.

Andreas Walter, Uwe Kiencke, Stephen Jones, and Thomas Winkler. Misfire Detection for Vehicles with Dual Mass Flywheel (DMF) Based on Reconstructed Engine Torque. *SAE Technical Paper 2007-01-3544*, 2007.

Timothy J. Wheeler. *Probabilistic Performance Analysis of Fault Diagnosis Schemes*. PhD thesis, University of California, Berkeley, 2011.

A. Willsky and H. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *Automatic Control, IEEE Transactions on*, 21(1):108 – 112, feb 1976.

Publications

Paper A

A method for quantitative fault diagnosability analysis of stochastic linear descriptor models*

A

*Accepted for publication in *Automatica*.

A method for quantitative fault diagnosability analysis of stochastic linear descriptor models

Daniel Eriksson, Erik Frisk, and Mattias Krysander

Vehicular Systems, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.

Abstract

Analyzing fault diagnosability performance for a given model, before developing a diagnosis algorithm, can be used to answer questions like "How difficult is it to detect a fault f_i?" or "How difficult is it to isolate a fault f_i from a fault f_j?". The main contributions are the derivation of a measure, *distinguishability*, and a method for analyzing fault diagnosability performance of discrete-time descriptor models. The method, based on the Kullback-Leibler divergence, utilizes a stochastic characterization of the different fault modes to quantify diagnosability performance. Another contribution is the relation between distinguishability and the *fault to noise ratio* of residual generators. It is also shown how to design residual generators with maximum fault to noise ratio if the noise is assumed to be i.i.d. Gaussian signals. Finally, the method is applied to a heavy duty diesel engine model to exemplify how to analyze diagnosability performance of non-linear dynamic models.

1 Introduction

Diagnosis and supervision of industrial systems concern detecting and isolating faults that occur in the system. As technical systems have grown in complexity, the demand for functional safety and reliability has drawn significant research in model-based fault detection and isolation. The maturity of the research field is verified by the amount of existing reference literature, for example Gertler (1998), Isermann (2005), and Patton et al. (2010).

When developing a diagnosis algorithm, knowledge of achievable diagnosability performance given the model of the system, such as detectability and isolability, is useful. Such information indicates if a test with certain diagnosability properties can be created or if more sensors are needed to get satisfactory diagnosability performance, see Commault et al. (2008) and Raghuraj et al. (1999). In Düştegör et al. (2006), a structural diagnosability analysis is used during the modeling process to derive a sufficiently good model which achieves a required diagnosability performance. In these previous works, information of diagnosability performance is required before a diagnosis algorithm is developed.

The main limiting factor of fault diagnosability performance of a model-based diagnosis algorithm is the model uncertainty. Model uncertainties exist because of, for example, non-modeled system behavior, process noise, or measurement noise. Models with large uncertainties make it difficult to detect and isolate small faults. Without sufficient information of possible diagnosability properties, engineering time could be wasted on, e.g., developing tests to detect a fault that in reality is impossible to detect.

The main contribution of this work is a method to quantify detectability and isolability properties of a model when taking model uncertainties and fault time profiles into consideration. It can also be used to compare achievable diagnosability performance between different models to evaluate how much performance is gained by using an improved model.

Different types of measures to evaluate the detectability performance of diagnosis algorithms exists in the litterature, see for example Chen et al. (2003), dos Santos and Yoneyama (2011), Hamelin and Sauter (2000), and Wheeler (2011). A contribution of this work with respect to these previously published papers, is to quantify diagnosability performance given the model without designing a diagnosis algorithm.

There are several works describing methods from classical detection theory, for example, the books Basseville and Nikiforov (1993) and Kay (1998), which can be used for quantified detectability analysis using a stochastic characterization of faults. In contrast to these works, *isolability* performance is also considered here which is important when identifying the faults present in the system.

There exist systematic methods for analyzing fault isolability performance in dynamic systems, see Frisk et al. (2010), Pucel et al. (2009), and Travé-Massuyès et al. (2006). However these approaches are deterministic and only give qualitative statements whether a fault is isolable or not. These methods give an optimistic result of isolability performance and they tell nothing about how

difficult it is to detect or isolate the faults in practice, due to model uncertainties.

The results in this paper are based on the early work done in Eriksson et al. (2011b) and Eriksson et al. (2011a) where a measure is derived, named *distinguishability*, for quantitative fault detectability and isolability analysis. First, the problem is formulated in Section 2. The measure is derived in Section 3 for linear discrete-time descriptor models. How to compute the special case when the noise is i.i.d. Gaussian is discussed in Section 4. In Section 5 the relation between distinguishability and the performance of linear residual generators is derived. Finally, it is shown in Section 6, via a linearization scheme, how the developed methodology can be used to analyze a non-linear dynamic model of a heavy duty diesel engine.

2 Problem formulation

The objective here is to develop a method for quantitative diagnosability analysis of discrete-time descriptor models in the form

$$\begin{aligned} Ex[t+1] &= Ax[t] + B_u u[t] + B_f f[t] + B_v v[t] \\ y[t] &= Cx[t] + D_u u[t] + D_f f[t] + D_\varepsilon \varepsilon[t] \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^{l_x}$ are state variables, $y \in \mathbb{R}^{l_y}$ are measured signals, $u \in \mathbb{R}^{l_u}$ are input signals, $f \in \mathbb{R}^{l_f}$ are modeled faults, $v \sim \mathcal{N}(0, \Lambda_v)$ and $\varepsilon \sim \mathcal{N}(0, \Lambda_\varepsilon)$ are i.i.d. Gaussian random vectors with zero mean and symmetric positive definite covariance matrices $\Lambda_v \in \mathbb{R}^{l_v \times l_v}$ and $\Lambda_\varepsilon \in \mathbb{R}^{l_\varepsilon \times l_\varepsilon}$. Model uncertainties and noise are represented in (1) by the random vectors v and ε. The notation l_α denotes the number of elements in the vector α. To motivate the problem studied in this paper, fault isolability performance is analyzed for a small example using a deterministic analysis method. Then a shortcoming of using this type of method is highlighted, based on the example.

Example 1. *The example will be used to discuss the result when analyzing fault detectability and isolability performance of a model by using a deterministic analysis method from Frisk et al. (2009). A simple discrete-time dynamic model of a spring-mass system is considered,*

$$\begin{aligned} x_1[t+1] &= x_1[t] + x_2[t] \\ x_2[t+1] &= x_2[t] - x_1[t] + u[t] + f_1[t] + f_2[t] + \varepsilon_1[t] \\ y_1[t] &= x_1[t] + f_3[t] + \varepsilon_2[t] \\ y_2[t] &= x_1[t] + f_4[t] + \varepsilon_3[t], \end{aligned} \tag{2}$$

where x_1 is the position and x_2 the velocity of the mass, y_1 and y_2 are sensors measuring the mass position, u is a control signal, f_i are possible faults, and ε_i are model uncertainties modeled as i.i.d. Gaussian noise where $\varepsilon_1 \sim \mathcal{N}(0, 0.1)$, $\varepsilon_2 \sim \mathcal{N}(0, 1)$, and $\varepsilon_3 \sim \mathcal{N}(0, 0.5)$. For simplicity, the mass, the spring constant, and sampling time are set to one. The faults are assumed additive and represent

faults in the control signal, f_1, a change in rolling resistance, f_2, and sensor biases, f_3 and f_4.

Analyzing fault isolability performance for the model (2) *using a deterministic method gives that all faults are detectable, f_3 and f_4 are each isolable from all other faults, and f_1 and f_2 are isolable from the other faults but not from each other. The result from the isolability analysis is summarized in Table 1. An X in position (i, j) represents that the fault mode f_i is isolable from fault mode f_j and a 0 represents that fault mode f_i is not isolable from fault mode f_j. The NF column indicates whether the corresponding fault mode is detectable or not.*

A shortcoming with an analysis like the one in Table 1 is that it does not take model uncertainties into consideration, i.e., the analysis does not state how difficult it is to detect and isolate the different faults depending on model uncertainties and how the fault changes over time.

The example highlights a limitation when using a deterministic diagnosability analysis method to analyze a mathematical model. Model uncertainties, process noise, and measurement noise are affecting diagnosability performance negatively and therefore it would be advantageous to take these uncertainties into consideration when analyzing diagnosability performance.

3 Distinguishability

This section defines a stochastic characterization for the fault modes and introduces a quantitative diagnosability measure based on the Kullback-Leibler divergence.

3.1 Reformulating the model

First, the discrete-time dynamic descriptor model (1) is written as a sliding window model of length n.

Table 1: A deterministic detectability and isolability analysis of (2) where an X in position (i, j) represents that a fault f_i is isolable from a fault f_j and 0 otherwise.

	NF	f_1	f_2	f_3	f_4
f_1	X	0	0	X	X
f_2	X	0	0	X	X
f_3	X	X	X	0	X
f_4	X	X	X	X	0

With a little abuse of notation, define the vectors

$$\begin{aligned} z &= \left(y[t-n+1]^T, \ldots, y[t]^T, u[t-n+1]^T, \ldots, u[t]^T\right)^T \\ x &= \left(x[t-n+1]^T, \ldots, x[t]^T, x[t+1]^T\right)^T, \\ f &= \left(f[t-n+1]^T, \ldots, f[t]^T\right)^T \\ e &= \left(v[t-n+1]^T, \ldots, v[t]^T, \varepsilon[t-n+1]^T, \ldots, \varepsilon[t]^T\right)^T, \end{aligned} \tag{3}$$

where $z \in \mathbb{R}^{n(l_y+l_u)}$, $x \in \mathbb{R}^{(n+1)l_x}$, $f \in \mathbb{R}^{nl_f}$ and e is a stochastic vector of a known distribution with zero mean. Note that in this section the additive noise will not be limited to be i.i.d. Gaussian as assumed in (1). Then a sliding window model of length n can be written as

$$Lz = Hx + Ff + Ne \tag{4}$$

where

$$L = \begin{pmatrix} 0 & 0 & \ldots & 0 & -B_u & 0 & \ldots & 0 \\ 0 & 0 & & 0 & 0 & -B_u & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & -B_u \\ I & 0 & \ldots & 0 & -D_u & 0 & \ldots & 0 \\ 0 & I & & 0 & 0 & -D_u & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & I & 0 & 0 & \ldots & -D_u \end{pmatrix}, H = \begin{pmatrix} A & -E & 0 & \ldots & 0 \\ 0 & A & -E & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & A & -E \\ C & 0 & 0 & \ldots & 0 \\ 0 & C & 0 & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & C & 0 \end{pmatrix},$$

$$F = \begin{pmatrix} B_f & 0 & \ldots & 0 \\ 0 & B_f & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & B_f \\ D_f & 0 & \ldots & 0 \\ 0 & D_f & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & D_f \end{pmatrix}, N = \begin{pmatrix} B_v & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & B_v & & 0 & 0 & 0 & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & B_v & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & D_\varepsilon & 0 & \ldots & 0 \\ 0 & 0 & & 0 & 0 & D_\varepsilon & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & D_\varepsilon \end{pmatrix},$$

and I is the identity matrix. Note that the sliding window model (4) is a static representation of the dynamic behavior on the window given the time indexes $(t-n+1, t-n+2, ..., t)$.

The sliding window model (4) represents the system (1) over a time window of length n. By observing a system during a time interval, not only constant faults, but faults that vary over time can be analyzed. Let $f_i \in \mathbb{R}^n$ be a vector containing only the elements corresponding to a specific fault i in the vector $f \in \mathbb{R}^{nl_f}$, i.e,

$$f_i = (f[i], f[l_f + i], f[2l_f + i], \ldots, f[(n-1)l_f + i])^T. \tag{5}$$

A vector $\theta = (\theta[t-n+1], \theta[t-n+2], \ldots, \theta[t])^T \in \mathbb{R}^n$ is used to represent how a fault, $f_i = \theta$, changes over time and is called a *fault time profile*. Figure 1 shows some examples of different fault time profiles where $n = 10$.

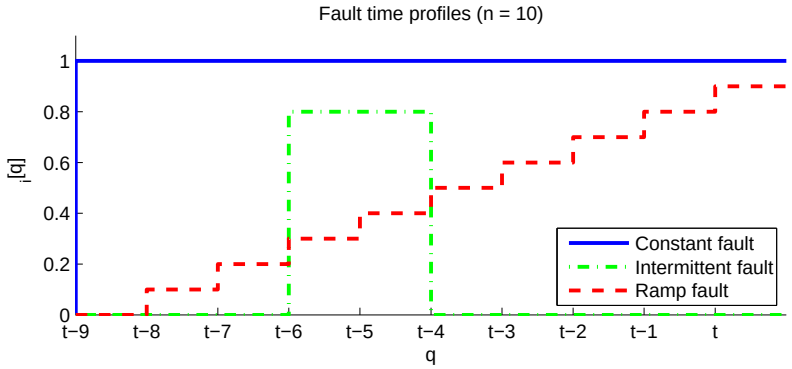

Figure 1: Fault time profiles representing a constant fault, an intermittent fault, and a fault entering the system like a ramp.

It is assumed that model (4) fulfills the condition that

$$\begin{pmatrix} H & N \end{pmatrix} \text{ is full row-rank.} \tag{6}$$

One sufficient criteria for (1) to satisfy (6) is that

$$D_\varepsilon \text{ is full row-rank and } \exists \lambda \in \mathbb{C}\colon \lambda E - A \text{ is full rank,} \tag{7}$$

i.e., all sensors have measurement noise and the model has a unique solution for a given initial state, see Kunkel and Mehrmann (2006). Assumption (7) assures that model redundancy can only be achieved when sensors y are included. The technical condition (6) is non-restrictive since it only excludes models where it is possible to design ideal residual generators, i.e., residuals that are not affected by noise.

It proves useful to write (4) in an input-output form where the unknowns, x, are eliminated. If the model (4) fulfills assumption (6), the covariance matrix of e for the model in input-output form will be non-singular. Elimination of x in (4) is achieved by multiplying with $\mathcal{N}_H$ from the left, where the rows of $\mathcal{N}_H$ is an orthonormal basis for the left null-space of H, i.e.,

$$\mathcal{N}_H H = 0,$$

This operation is also used, for example, in classical parity space approaches, see Gertler (1997) and Zhang and Ding (2007). The input-output model can, in the general case, then be written as

$$\mathcal{N}_H L z = \mathcal{N}_H F f + \mathcal{N}_H N e. \tag{8}$$

It is important to note that for any solution z_0, f_0, e_0 to (8) there exists an x_0 such that it also is a solution to (4), and also if there exists a solution z_0, f_0, e_0, x_0 to (4) then z_0, f_0, e_0 is a solution to (8). Thus no information about the model behavior is lost when rewriting (4) as (8), see Polderman and Willems (1998).

3.2 Stochastic characterization of fault modes

To describe the behavior of system (4), the term *fault mode* is used. A fault mode represents whether a fault f_i is present, i.e., $f_i \neq \bar{0}$, where $\bar{0}$ denotes a vector with only zeros. With a little abuse of notation, f_i will also be used to denote the fault mode when f_i is the present fault. The mode when no fault is present, i.e., $f = \bar{0}$, is denoted NF.

Let $\tau = \mathcal{N}_H L z$, which is the left hand side of (8). The vector $\tau \in \mathbb{R}^{nl_y - l_x}$ depends linearly on the fault vector f and the noise vector e and represents the *behavior* of the model, see Polderman and Willems (1998). A non-zero fault vector f only affect the mean of the probability distribution of the vector τ.

Let $p(\tau; \mu)$, denote a multivariate probability density function, pdf, with mean μ describing τ, where μ depends on f. The mean $\mu = \mathcal{N}_H F_i \theta$, where the matrix $F_i \in \mathbb{R}^{n(l_x+l_y)\times n}$ contains the columns of F corresponding to the elements of f_i in (5), is a function of the fault time profile $f_i = \theta$. Let Θ_i denote the set of all fault time profiles θ corresponding to a fault mode f_i which for example could look like the fault time profiles in Figure 1. For each fault time profile $f_i = \theta \in \Theta_i$ which could be explained by a fault mode f_i, there is a corresponding pdf $p(\tau; \mathcal{N}_H F_i \theta)$. According to this, each fault mode f_i can be described by a set of pdf's $p(\tau; \mu)$, giving the following definition.

Definition 1. *Let $\mathcal{Z}_{f_i}$ denote the set of all pdf's $p(\tau; \mu(\theta))$, for all fault time profiles $\theta \in \Theta_i$, describing τ which could be explained by the fault mode f_i, i.e.*

$$\mathcal{Z}_{f_i} = \{p(\tau; \mathcal{N}_H F_i \theta) | \forall \theta \in \Theta_i\}. \tag{9}$$

The definition of $\mathcal{Z}_{f_i}$ is a stochastic counterpart to observation sets in the deterministic case, see Nyberg and Frisk (2006). Each fault mode f_i, including NF, can be described by a set $\mathcal{Z}_{f_i}$. The set $\mathcal{Z}_{\text{NF}}$ describing the fault-free mode typically only includes one pdf, $p_{\text{NF}} = p(\tau; \bar{0})$. Note that the different sets, $\mathcal{Z}_{f_i}$, does not have to be mutually exclusive since different fault modes could affect the system in the same way, resulting in the same pdf. A specific fault time profile $f_i = \theta$ corresponds to one pdf in $\mathcal{Z}_{f_i}$ and is denoted

$$p_\theta^i = p(\tau; \mathcal{N}_H F_i \theta). \tag{10}$$

Using Definition 1 and (10), isolability (and detectability) of a window model (4) can be defined as follows.

Definition 2 (Isolability (detectability))**.** *Consider a window model* (4)*. A fault f_i with a specific fault time profile $\theta \in \Theta_i$ is isolable from fault mode f_j if*

$$p_\theta^i \notin \mathcal{Z}_{f_j}.$$

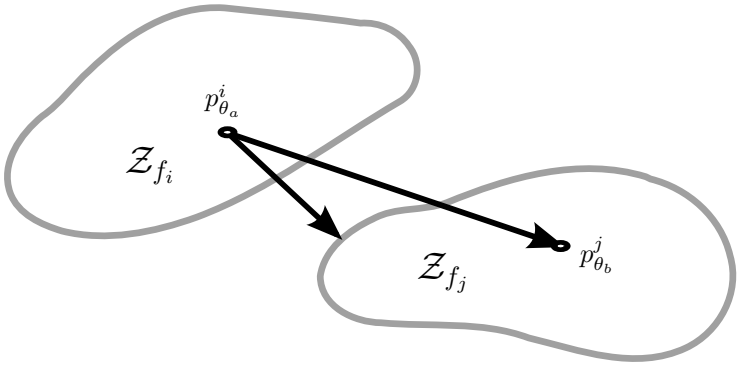

Figure 2: A graphical visualization of the sets $\mathcal{Z}_{f_i}$ and $\mathcal{Z}_{f_i}$ and the smallest difference between $p^i_{\theta_a} \in \mathcal{Z}_{f_i}$ and a pdf $p^j \in \mathcal{Z}_{f_j}$.

Similarly, if $p^i_\theta \notin \mathcal{Z}_{NF}$, the fault is detectable, i.e., the fault is isolable from the fault-free mode.

3.3 Quantitative detectability and isolability

Consider two pdf's, $p^i_{\theta_a} \in \mathcal{Z}_{f_i}$ and $p^j_{\theta_b} \in \mathcal{Z}_{f_j}$, describing τ for two different faults with given fault time profiles $f_i = \theta_a$ and $f_j = \theta_b$ respectively. The more the distribution of the observations differ between two fault scenarios, the easier it is to isolate the faults. Therefore, a measure of difference between probability distributions can be used to quantify isolability between faults. The basic idea is illustrated in Figure 2 where the difference of $p^i_{\theta_a}$ and $p^j_{\theta_b}$ can graphically be interpreted as the distance and the closer the pdf's are, the more similar are their distributions.

To motivate the distance measure that will be used, consider the task of isolating a given fault time profile $f_i = \theta_a$ from fault time profile $f_j = \theta_b$. Therefore, consider the hypothesis test

$$\begin{aligned} H_0 &: \ p = p^j_{\theta_b} \\ H_1 &: \ p = p^i_{\theta_a}. \end{aligned} \tag{11}$$

and, to solve it, consider as test statistic the log-likelihood ratio

$$\lambda(\tau) = \log \frac{L_{p^i_{\theta_a}}(\tau)}{L_{p^j_{\theta_b}}(\tau)}$$

where $L_p(\tau)$ is the likelihood of τ given the pdf p. In case that hypothesis H_0 is true, i.e., fault $f_j = \theta_b$ is the true fault, then observations τ are drawn from a distribution $p^j_{\theta_b}$ and $E[\lambda(\tau)] \leq 0$. In case hypothesis H_1 is true, i.e., fault $f_i = \theta_a$ is the true fault, observations are drawn from $p^i_{\theta_a}$ and $E[\lambda(\tau)] \geq 0$, see

Casella and Berger (2001). Thus, λ changes sign, in the mean, with the two hypotheses. Therefore, the mean of $\lambda(\tau)$, under H_1

$$E_{p^i_{\theta_a}}\left[\lambda(\tau)\right] = E_{p^i_{\theta_a}}\left[\log \frac{L_{p^i_{\theta_a}}(\tau)}{L_{p^j_{\theta_b}}(\tau)}\right] \tag{12}$$

is an indicator on how difficult it is to isolate fault $f_i = \theta_a$ from fault $f_j = \theta_b$. The right hand side of (12) can be identified as the Kullback-Leibler divergence Kullback, S. and Leibler, R. A. (1951) and will be denoted $K(p^i_{\theta_a} \| p^j_{\theta_b})$.

Generally, the Kullback-Leibler divergence between two pdf's p^i and p^j is defined as

$$K(p^i \| p^j) = \int_{-\infty}^{\infty} p^i(v) \log \frac{p^i(v)}{p^j(v)} dv = \mathrm{E}_{p^i}\left[\log \frac{p^i}{p^j}\right] \tag{13}$$

where $\mathrm{E}_{p^i}\left[\log \frac{p^i}{p^j}\right]$ is the expected value of $\log \frac{p^i}{p^j}$ given p^i. The Kullback-Leibler divergence has the following properties

$$\begin{aligned} &K(p^i \| p^j) \geq 0, \\ &K(p^i \| p^j) = 0 \text{ iff } p^i = p^j. \end{aligned} \tag{14}$$

In the hypotheses in (11), fault sizes are completely specified but in the general case the fault sizes are not known and we want to isolate a particular fault $f_i = \theta$ from a *fault mode* f_j. It is then natural to quantify the isolability performance as the minimal expected value of $\lambda(r)$, i.e., the Kullback-Leibler divergence $K(p^i_\theta \| p^j)$, for any $p^j \in \mathcal{Z}_{f_j}$. Also, minimizing $K(p^i_\theta \| p^j)$ with respect to $p^j \in \mathcal{Z}_{f_j}$ is the same as maximizing the maximum likelihood estimate of $p^j \in \mathcal{Z}_{f_j}$ to p^i_θ, see Eguchi and Copas (2006). Based on this discussion, the measure of isolability performance is then defined as follows.

Definition 3 (Distinguishability)**.** *Given a sliding window model* (4)*, distinguishability $\mathcal{D}_{i,j}(\theta)$ of a fault f_i with a given fault time profile θ from a fault mode f_j is defined as*

$$\mathcal{D}_{i,j}(\theta) = \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p^i_\theta \| p^j\right) \tag{15}$$

where the set $\mathcal{Z}_{f_j}$ is defined in Definition 1 and p^i_θ in (10)*.*

Note that in Definition 3, the additive noise in (1) is not required to be i.i.d. Gaussian. Distinguishability can be used to analyze either isolability or detectability performance depending on whether $\mathcal{Z}_{f_j}$ describes a fault mode or the fault-free case.

The measure defined in Definition 3 does not directly relate to, for example, probability of detection or isolation of a specific fault. However, it turns out that distinguishability is, in the Gaussian case, closely related to the performance of optimal residual generators. This fact is explained in detail and discussed further in Section 5.

Two properties of the index $D_{i,j}(\theta)$ are given in the following propositions. Proposition 1 presents a necessary and sufficient condition for isolability, while Proposition 2 shows that it is easier to detect faults than isolate faults.

Proposition 1. *Given a window model* (4)*, a fault* $f_i = \theta \in \Theta_i$ *is isolable from a fault mode* f_j *if and only if*

$$\mathcal{D}_{i,j}(\theta) > 0 \tag{16}$$

Proof. The Kullback-Leibler divergence $K(p^i_\theta \| p^j)$ is zero if and only if $p^i_\theta = p^j$. Given Definition 2, $f_i = \theta$ is isolable from f_j if $p^i_\theta \neq p^j$ for all $p^j \in \mathcal{Z}_{f_j}$ and (16) holds. If (16) holds then $p^i_\theta \neq p^j$ for all $p^j \in \mathcal{Z}_{f_j}$ and $f_i = \theta$ is isolable from f_j which proves Proposition 1. □

Proposition 2. *If* $\bar{0}$ *is a boundary point of* Θ_j *for a fault mode* f_j *then*

$$\mathcal{D}_{i,j}(\theta) \leq \mathcal{D}_{i,NF}(\theta). \tag{17}$$

Proof. If $\bar{0}$ is a boundary point of Θ_j, then p_{NF} is a boundary point of $\mathcal{Z}_{f_j}$ and there exists a limit $p^j \in \mathcal{Z}_{f_j}$ such that

$$\lim_{p^j \to p_{\text{NF}}} K\left(p^i_\theta \| p^j\right) = K\left(p^i_\theta \| p_{\text{NF}}\right).$$

Then

$$\mathcal{D}_{i,j}(\theta) = \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p^i_\theta \| p^j\right) \leq K\left(p^i_\theta \| p_{\text{NF}}\right) = \mathcal{D}_{i,\text{NF}}(\theta)$$

which proves Proposition 2. □

From now on, it is assumed that $\Theta_i = \mathbb{R}^n \setminus \{\bar{0}\}$ for all $i = 1, 2, \ldots, l_f$ and then Proposition 2 holds.

4 Computation of distinguishability

The definition of distinguishability in Section 3.3, given a model in the form (4), is general for any type of multivariate pdf (10) describing the vector τ for a given $\theta \in \Theta_i$. Computing (15) requires solving a minimization problem which can in general be difficult. In this section, the vector e in (4) is assumed to be Gaussian distributed with covariance $\Lambda_e \in \mathbb{R}^{n(l_v+l_\varepsilon)\times n(l_v+l_\varepsilon)}$. Model uncertainties and disturbancies only containing a limited band of frequencies, e.g., low-frequency disturbances, can be included in (1) by adding noise dynamics to the model, see Glad and Ljung (2000). If e is Gaussian and (6) is fulfilled, then $\mathcal{N}_H N e$ in (8) is Gaussian distributed with positive definite covariance matrix

$$\Sigma = \mathcal{N}_H N \Lambda_e N^T \mathcal{N}_H^T,$$

and (15) can be computed explicitly.

To simplify the computations of (15), it is assumed without loss of generality that Σ is equal to the identity matrix, that is

$$\Sigma = \mathcal{N}_H N \Lambda_e N^T \mathcal{N}_H^T = I \tag{18}$$

Note that any model in the form (4), satisfying (6), can be transformed into fulfilling $\Sigma = I$ by multiplying (4) with an invertible transformation matrix T from the left. The choice of matrix T is non-unique and one possibility is

$$T = \begin{pmatrix} \Gamma^{-1}\mathcal{N}_H \\ T_2 \end{pmatrix} \tag{19}$$

where Γ is non-singular and

$$\mathcal{N}_H N \Lambda_e N^T \mathcal{N}_H^T = \Gamma\Gamma^T \tag{20}$$

is satisfied, and T_2 is any matrix ensuring invertability of T. Matrix Γ can, for example, be computed by a Cholesky factorization of the left hand side of (20).

Given the assumption in (18), it holds that $\Sigma = I$. In (8), all modeled faults f are additive and only affect the mean of τ. Then the pdf, $p(\tau;\mu)$, describing τ in the Gaussian case is defined as

$$p(\tau;\mu) = \frac{1}{|2\pi|^{\frac{d}{2}}} \exp\left(-\frac{1}{2}(\tau-\mu)^T(\tau-\mu)\right)$$

which is the multivariate Gaussian pdf with unit covariance matrix.

The vector τ is described, for any fault time profile, by a multivariate Gaussian pdf. Thus the Kullback-Leibler divergence is computed for two multivariate Gaussian pdf's with equal covariance $\Sigma = I$, $p^i \sim \mathcal{N}(\mu_i, I)$ and $p^j \sim \mathcal{N}(\mu_j, I)$. Then (13) can be written as

$$K(p^i \| p^j) = \frac{1}{2}\|\mu_i - \mu_j\|_{I^{-1}}^2 = \frac{1}{2}\|\mu_i - \mu_j\|^2. \tag{21}$$

Note that (21) is invariant to linear transformations, i.e., multiplying (4) from the left by an invertible matrix T will not affect the computed distinguishability. The invariance is easily verified by using $\tilde{p}^i \sim \mathcal{N}(T\mu_i, TT^T)$ and $\tilde{p}^j \sim \mathcal{N}(T\mu_j, TT^T)$ where T is a non-singular transformation matrix, then

$$\begin{aligned} K(\tilde{p}^i \| \tilde{p}^j) &= \frac{1}{2}\|T(\mu_i - \mu_j)\|_{(TT^T)^{-1}}^2 = \\ &= \frac{1}{2}\|\mu_i - \mu_j\|^2 = K(p^i \| p^j). \end{aligned} \tag{22}$$

To derive an explicit expression of (15), the following standard result will be used.

Lemma 1. *For a matrix $A \in \mathbb{R}^{n\times m}$ and a vector $b \in \mathbb{R}^n$, with $n > m$, it holds that*

$$\min_x \|Ax - b\|^2 = \|\mathcal{N}_A b\|^2. \tag{23}$$

where the rows of $\mathcal{N}_A$ is an orthonormal basis for the left null space of A.

Proof. Minimizing the left hand side of (23) is equivalent to projecting b onto the orthogonal complement of A, $\mathrm{Ker}(A)$, with the projection matrix $P = \mathcal{N}_A^T \mathcal{N}_A$. This gives that

$$\min_x \|Ax - b\|^2 = \|Pb\|^2 = b^T P b = b^T \mathcal{N}_A^T \mathcal{N}_A b = \|\mathcal{N}_A b\|^2.$$

□

Theorem 1. *Distinguishability for a sliding window model* (4) *with Gaussian distributed stochastic vector* e*, under assumption* (18)*, is given by*

$$\mathcal{D}_{i,j}(\theta) = \frac{1}{2}\|\mathcal{N}_{(H\ F_j)} F_i \theta\|^2 \tag{24}$$

where the rows of $\mathcal{N}_{(H\ F_j)}$ *is an orthonormal basis for the left null space of* $(H\ F_j)$*.*

Before proving Theorem 1, note that distinguishability for a general model in the form (4) under assumption (6) can be computed by:

1. applying the transformation (19),
2. redefining the matrices L, H, F, and N given the transformed model fulfilling assumption (18), and
3. computing distinguishability using (24).

Proof. The set $\mathcal{Z}_{f_j}$ is parametrized by $f_j = \theta_j$, thus minimizing (15) with the respect to $p^j \in \mathcal{Z}_{f_j}$ is equal to

$$\begin{aligned}\mathcal{D}_{i,j}(\theta) &= \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p_\theta^i \| p^j\right) = \\ &= \min_{\theta_j} \frac{1}{2}\|\mathcal{N}_H F_i \theta - \mathcal{N}_H F_j \theta_j\|^2_{\Sigma^{-1}}\end{aligned}$$

Assumption (18) gives that $\Sigma = I$. Then,

$$\begin{aligned}\mathcal{D}_{i,j}(\theta) &= \min_{\theta_j} \frac{1}{2}\|\mathcal{N}_H (F_i \theta - F_j \theta_j)\|^2 = \\ &= \min_{\theta_j, x} \frac{1}{2}\|Hx - F_i\theta + F_j\theta_j\|^2 = \\ &= \min_{\theta_j, x} \frac{1}{2}\| \begin{pmatrix} H & F_j \end{pmatrix} \begin{pmatrix} x \\ \theta_j \end{pmatrix} - F_i\theta\|^2 = \\ &= \frac{1}{2}\|\mathcal{N}_{(H\ F_j)} F_i \theta\|^2\end{aligned}$$

where Lemma 1 is used in the second and fourth equality. □

Note that, if a fault time profile θ is multiplied with a scalar $\alpha \in \mathbb{R}$, Theorem 1 gives that distinguishability is proportional to the square of the parameter α, i.e., $\mathcal{D}_{i,j}(\alpha\theta) = \alpha^2\mathcal{D}_{i,j}(\theta)$.

Example 2. *In this example, distinguishability is computed to analyze diagnosability performance and the result is compared to the deterministic analysis of the spring-mass model* (2) *made in Example 1. The model is rewritten as a window model of length three in the form* (4)*. Then, distinguishability is computed for each fault pair where the fault time profile is assumed to be constant of amplitude one, i.e.,* $\theta = (1, 1, 1)^T$*. The computed distinguishability is summarized in Table 2.*

Table 2: Computed distinguishability of (2) when rewritten on the form (4) where $n = 3$ and $\theta = (1, 1, 1)^T$.

$\mathcal{D}_{i,j}(\theta)$	NF	f_1	f_2	f_3	f_4
f_1	0.16	0	0	0.11	0.05
f_2	0.16	0	0	0.11	0.05
f_3	1.02	1.00	1.00	0	0.05
f_4	1.07	1.00	1.00	0.11	0

Table 2 shows that it is easier to detect the sensor faults f_3 *and* f_4 *than the actuator fault* f_1 *and a change in rolling resistance* f_2 *for the constant fault time profile, since* $1.02 > 0.16$ *and* $1.07 > 0.16$*. A comparison of Table 2 and Table 1 shows that all positions marked with X in Table 1 correspond to nonzero distinguishability in Table 2. Table 2 also shows that distinguishability of isolating each of the faults from the other faults never exceeds distinguishability of detecting the faults, which follows from Proposition 2.*

If instead a window model of length $n = 6$ *is analyzed, i.e., the window length is doubled, then the computed distinguishability is shown in Table 3. A comparison of Table 3 and Table 2 shows as expected that the increased window length results in higher distinguishability for the different fault pairs. Note also that, for example, distinguishability is higher for detecting* f_1 *and* f_2*,* 4.21*, than* f_3 *and* f_4*,* 2.47 *and* 3.87 *respectively, which were the opposite situation for* $n = 3$*.*

If instead the window length is decreased from $n = 3$*, detectability and isolability performance is lost. In Table 4 distinguishability is computed where* $n = 2$*. The analysis shows that, for the spring-mass model, if the window length is lower than three then a constant fault* f_1 *or* f_2 *can not be detected or isolated. When analyzing the model* (2) *it turns out that to have enough redundancy in the data to detect* f_1 *and* f_2*, a window length of at least* $n = 3$ *is needed due to the model dynamics. To detect and isolate* f_3 *and* f_4 *from* f_1 *and* f_2 *requires only* $n = 1$ *because it is sufficient to take the difference of the two sensors* y_1 *and* y_2 *for obtaining redundancy.*

To illustrate how a different fault time profile effects the distinguishability consider as a comparison to the constant fault time profile used in the computation of Table 3 a step where $\theta = (0, 0, 0, \sqrt{2}, \sqrt{2}, \sqrt{2})^T$. *The amplitude of the step is chosen such that the energy of the fault time profile here is equal to a constant fault of amplitude one, i.e.,* $\theta^T\theta$ *is equal for the two fault time profiles in the comparison. The distinguishability for this step fault is shown in Table 5. The distinguishability in Table 3 is lower except for isolating* f_3 *and* f_4 *from* f_1 *and* f_2 *then in Table 5. Thus it is more difficult to detect or isolate a fault behaving like a step even though the amplitude is higher compared to a constant fault. Finally, in Table 5, distinguishability for detecting* f_1 *and* f_2*, is lower than detecting* f_3 *and* f_4 *which is the opposite to Table 3. This is due to the fact that only three time instances in the time window are effected by the fault and hence is similar to the case in Table 2.*

Table 4 contains more zeros than Table 1 which states that $n = 2$ is not enough to detect and isolate some faults. If only Table 4 is used to state which faults that are theoretically isolable in (1), then it could be wrongly concluded that f_1 and f_2 are not isolable at all. Therefore, distinguishability should be computed for $n \geq l_x+1$, or a deterministic isolability analysis could be performed, see Nyberg (2002).

Table 3: Computed distinguishability of (2) when rewritten on the form (4) where $n = 6$ and $\theta = (1, 1, 1, 1, 1, 1)^T$.

$\mathcal{D}_{i,j}(\theta)$	NF	f_1	f_2	f_3	f_4
f_1	4.21	0	0	3.04	1.65
f_2	4.21	0	0	3.04	1.65
f_3	2.47	2.00	2.00	0	1.65
f_4	3.87	2.00	2.00	3.04	0

Table 4: Computed distinguishability of (2) when rewritten on the form (4) where $n = 2$ and $\theta = (1, 1)^T$.

$\mathcal{D}_{i,j}(\theta)$	NF	f_1	f_2	f_3	f_4
f_1	0	0	0	0	0
f_2	0	0	0	0	0
f_3	0.67	0.67	0.67	0	0
f_4	0.67	0.67	0.67	0	0

5 Relation to residual generators

An important property of the computed distinguishability for a model (4) with Gaussian distributed stochastic vector e is the connection to the performance of residual generators. The connection between distinguishability and residual generators shows the relation between computed distinguishability and achievable diagnosability performance. In this section these relations are derived.

A linear residual generator is here defined, in a direct stochastic extension to the definitions in Frisk and Nielsen (2006); Frisk and Nyberg (2001); Zhong et al. (2003), as

Definition 4 (Linear residual generator)**.** *A linear function $r = R\,z$, with the scalar r as output and z as defined in* (3)*, is a residual generator for* (4) *if r is zero mean in the fault free case. A residual is sensitive to a fault if the transfer function from fault to residual is non-zero.*

A residual generator that isolates a fault f_i from f_j, is a residual that is sensitive to f_i but not to f_j. To design a residual generator isolating faults from fault mode f_j, multiply (4) from the left with $\gamma\mathcal{N}_{(H\,F_j)}$ where γ is a row-vector to obtain

$$r = \gamma\mathcal{N}_{(H\,F_j)}Lz = \gamma\mathcal{N}_{(H\,F_j)}Ff + \gamma\mathcal{N}_{(H\,F_j)}Ne \tag{25}$$

Here, $\gamma\mathcal{N}_{(H\,F_j)}Lz$ is a residual generator that isolates *from* fault mode f_j. If only detectability, and not isolability, of f_i is considered, $\mathcal{N}_{(H\,F_j)}$ is replaced by $\mathcal{N}_H$. The vector γ parametrizes the space of all linear residual generators decoupling f_j, and is a design parameter selected to achieve fault sensitivity.

To quantify the performance of a residual generator (25), the following definition is used.

Definition 5 (Fault to noise ratio)**.** *For a residual generator* (25) *where e is a stochastic vector with covariance Λ_e. The fault to noise ratio, FNR, for a given fault $f_i = \theta$, is defined as the ratio between the amplified fault time profile, $\lambda(\theta) = \gamma\mathcal{N}_{(H\,F_j)}F\theta$ and the standard deviation of the noise σ as*

$$FNR = \frac{\lambda(\theta)}{\sigma} \tag{26}$$

Table 5: Computed distinguishability of (2) when rewritten on the form (4) where $n = 6$ and $\theta = (0, 0, 0, \sqrt{2}, \sqrt{2}, \sqrt{2})^T$.

$\mathcal{D}_{i,j}(\theta)$	NF	f_1	f_2	f_3	f_4
f_1	0.63	0	0	0.44	0.23
f_2	0.63	0	0	0.44	0.23
f_3	2.08	2.00	2.00	0	0.26
f_4	2.32	2.00	2.00	0.50	0

where

$$\sigma^2 = \gamma \mathcal{N}_{(H\,F_j)} N \Lambda_e N^T \mathcal{N}_{(H\,F_j)}^T \gamma^T.$$

Note that (25) is in the same form as (4) and can be seen as a scalar model. Therefore distinguishability, and Theorem 1 can directly be used to analyze isolability performance of a residual generator. A superscript γ is used, $\mathcal{D}_{i,j}^{\gamma}(\theta)$, to emphasize that it is computed distinguishability of a specific residual generator with a given γ. The connection between distinguishability and the FNR is given by the following result, which also gives an alternative way of computing distinguishability for a scalar model.

Theorem 2. *A residual generator* (25)*, for a model* (4) *where* e *is Gaussian distributed under assumption* (6)*, is also Gaussian distributed* $\mathcal{N}(\lambda(\theta), \sigma^2)$ *and*

$$\mathcal{D}_{i,j}^{\gamma}(\theta) = \frac{1}{2}\left(\frac{\lambda(\theta)}{\sigma}\right)^2$$

where θ *is the fault time profile of a fault* f_i*, and* $\lambda(\theta)/\sigma$ *is the fault to noise ratio with respect to fault* f_i *in* (25)*.*

Proof. Assumption (6) on the model (4) directly implies that (6) is fulfilled also for the residual generator (25). However, there is no guarantee that (25) fulfills (18) and the 3-step procedure after Theorem 1 must be used. After the transformation, the model is

$$\underbrace{\frac{\gamma \mathcal{N}_{(H\,F_j)} L}{\sigma}}_{=:L} z = \underbrace{\frac{\gamma \mathcal{N}_{(H\,F_j)} F}{\sigma}}_{=:F} f + \underbrace{\frac{\gamma \mathcal{N}_{(H\,F_j)} N}{\sigma}}_{=:N} e \tag{27}$$

where σ is the standard deviation of the residual in (25). Note that the matrices L, F, and N are redefined in (27) and the new corresponding H is the empty matrix. Model (27) fulfills (18) and Theorem 1 gives that

$$\mathcal{D}_{i,j}^{\gamma}(\theta) = \frac{1}{2}\left\|\frac{\gamma \mathcal{N}_{(H\,F_j)} F_i \theta}{\sigma}\right\|^2 = \frac{1}{2}\left(\frac{\lambda(\theta)}{\sigma}\right)^2.$$

□

Theorem 2 shows a direct relation between FNR in a residual isolating fault f_i from fault f_j and the computed distinguishability $\mathcal{D}_{i,j}^{\gamma}(\theta)$ for the residual.

An important connection between $\mathcal{D}_{i,j}^{\gamma}(\theta)$ and $\mathcal{D}_{i,j}(\theta)$ is given by the inequality described by the following theorem.

Theorem 3. *For a model* (4) *under assumption* (18)*, an upper bound for* $\mathcal{D}_{i,j}^{\gamma}(\theta)$ *in* (25) *is given by*

$$\mathcal{D}_{i,j}^{\gamma}(\theta) \leq \mathcal{D}_{i,j}(\theta)$$

with equality if and only if γ *and* $\left(\mathcal{N}_{(H\,F_j)} F_i \theta\right)^T$ *are parallel.*

Proof. Since both $\mathcal{N}_H$ and $\mathcal{N}_{(H\,F_j)}$ define orthonormal bases and the row vectors of $\mathcal{N}_{(H\,F_j)}$ are in the span of the row vectors of $\mathcal{N}_H$, there exists an α such that $\mathcal{N}_{(H\,F_j)} = \alpha\mathcal{N}_H$ and

$$I = \mathcal{N}_{(H\,F_j)}\mathcal{N}_{(H\,F_j)}^T = \alpha\mathcal{N}_H\mathcal{N}_H^T\alpha^T = \alpha\alpha^T$$

Using this result and assumption (18), the variance σ^2 in Theorem 2 can be written as

$$\begin{aligned}\sigma^2 &= \gamma\mathcal{N}_{(H\,F_j)}N\Lambda N^T\mathcal{N}_{(H\,F_j)}^T\gamma^T = \\ &= \gamma\alpha\mathcal{N}_H N\Lambda N^T\mathcal{N}_H^T\alpha^T\gamma^T = \gamma\gamma^T\end{aligned}$$

Finally, Cauchy-Schwarz inequality gives

$$\begin{aligned}\mathcal{D}_{i,j}^{\gamma}(\theta) &= \frac{1}{2}\frac{(\gamma\mathcal{N}_{(H\,F_j)}F_i\theta)^2}{\gamma\gamma^T} = \frac{1}{2}\frac{\langle\gamma^T, \mathcal{N}_{(H\,F_j)}F_i\theta\rangle^2}{\|\gamma\|^2} \leq \\ &\leq \frac{1}{2}\|\mathcal{N}_{(H\,F_j)}F_i\theta\|^2 = \mathcal{D}_{i,j}(\theta)\end{aligned}$$

with equality if and only if γ and $\left(\mathcal{N}_{(H\,F_j)}F_i\theta\right)^T$ are parallel. □

Theorem 3 shows that distinguishability of a residual never can exceed the distinguishability of the corresponding model. The result of Theorem 3 shows that an optimal residual for isolating a fault mode f_i from a fault mode f_j is obtained if $\gamma = k\left(\mathcal{N}_{(H\,F_j)}F_i\theta\right)^T$ for any non-zero scalar k. Such a residual has the highest FNR of fault f_i that any residual decoupling f_j can have. A key observation here is that by computing distinguishability for a model (4), maximum achievable FNR of a residual generator (25) is known. To implement a diagnosis algorithm with optimal single fault distinguishability to detect and isolate l_f single faults from each other thus requires at most

$$\underbrace{l_f}_{\text{detect}} + \underbrace{(l_f - 1)l_f}_{\text{isolate}} = l_f^2 \text{ tests.}$$

Example 3. *Now, Theorem 3 is applied to the spring-mass model* (2)*, with* $n = 3$*, to generate residual generators which achieves maximum FNR. The fault time profile is chosen as* $\theta = (1, 1, 1)^T$*, i.e., a constant fault with amplitude one. Maximum distinguishability is given in Table 2 and shows the upper limit of FNR which can be achieved.*

The vector γ *is computed as*

$$\gamma = k\left(\mathcal{N}_{(H\,F_j)}F_i\theta\right)^T$$

where $k \in \mathbb{R}$ *is non-zero, and an optimal residual generator, isolating a constant fault* f_i *from any fault* f_j*, is computed using* (25) *as*

$$r = \gamma\mathcal{N}_{(H\,F_j)}Lz = k\left(\mathcal{N}_{(H\,F_j)}F_i\theta\right)^T\mathcal{N}_{(H\,F_j)}Lz. \tag{28}$$

Using (28) *and a suitable* k*, a residual generator isolating* f_3 *from* f_1 *with maximum FNR is*

$$r = \sum_{m=t-2}^{t} (y_1[m] - y_2[m]) \tag{29}$$

which has a FNR, with respect to f_3*,*

$$FNR = \frac{1+1+1}{\sqrt{1+1+1+0.5+0.5+0.5}} = 1.41.$$

Distinguishability of r *in* (29) *for isolating* f_3*, with a constant fault time profile of amplitude one, from* f_1 *is computed using Theorem 2 as* 1.00 *which is equal to the corresponding position in Table 2. This means that* (29) *is also optimal in isolating* f_3 *from* f_2 *and also isolating* f_4 *from* f_1 *and* f_2 *respectively.*

The performance of a residual generator can also be visualized using a ROC-curve which shows the ratio between the probability of detection and false alarm, see Kay (1998). Consider the computed distinguishability for the spring mass model in Table 3. Distinguishability for detecting f_2 is higher than detecting f_3, $4.21 > 2.47$, and the ROC-curve for the corresponding optimal residuals using Theorem 3 is shown in Figure 3. A higher distinguishability corresponds to a higher ratio between the probability of detection and false alarm.

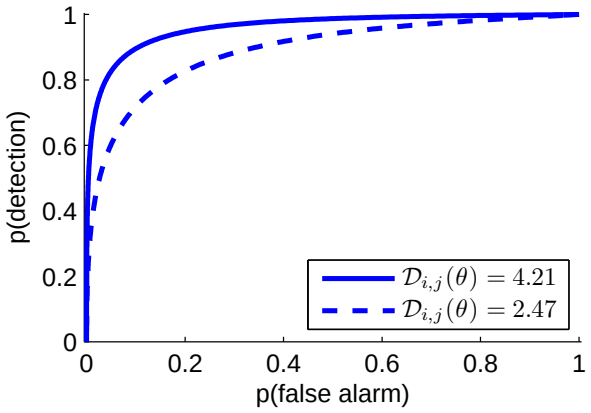

Figure 3: A ROC-curve comparing probability of detection and false alarm for two residuals with different computed values of distinguishability. A higher distinguishability corresponds to a higher ratio between the probability of detection and false alarm.

6 Diesel engine model analysis

Distinguishability, as a measure of quantified isolability, has been defined for discrete-time linear descriptor models written in the form (4). Typically, many industrial systems exhibits non-linear behavior. The purpose here is to demonstrate how the developed theory can be used also to analyze dynamic non-linear models of industrial complexity. Here, a model of a heavy duty diesel engine is analyzed by linearizing at different operating points of the engine and then computing distinguishability for each linearization point.

6.1 Model description

The considered model is a mean value engine model of gas flows in a heavy duty diesel engine. The model is documented in Wahlström and Eriksson (2011) and an overview is shown in Figure 4. The model considered here has 11 internal states; four actuators: fuel injection u_δ, valve controls u_{egr} and u_{vgt}, and throttle control u_{th}; and four measured signals: turbine speed ω_t, pressures p_{em} and p_{im}, and air mass-flow past the compressor W_c. The model has been extended with 13 possible faults indicated by arrows in Figure 4. The faults are briefly described in Table 6 and can be divided into four groups: $f_1, \ldots, f_4$ are actuator faults, $f_5, \ldots, f_8$ are sensor faults, $f_9, \ldots, f_{12}$ are leakages, and f_{13} is degraded compressor efficiency. Actuator faults and sensor faults are modeled as a bias of the nominal value. Leakage flow is modeled as proportional to the square root of the pressure difference over the leaking hole. Degraded efficiency of the compressor is modeled as a proportional negative fault to the compressor efficiency map. The faults f_i are shown in Table 6 and the fault sizes $f_i = \theta_i$ have been selected in the order of 10% of a nominal value of the corresponding model variable.

Uncertainties must be introduced in the model, and it is important how it is made because it significantly affects the result of the analysis. In this case, model uncertainties, actuator noise and measurement noise have been modeled as i.i.d. Gaussian noise. In Wahlström and Eriksson (2011), the model uncertainties for each sub model were analyzed by comparing simulation with measurement data. The model uncertainties are modeled as process noise where the standard deviations of the process noise are selected proportional to the uncertainties in the model according to Wahlström and Eriksson (2011). The model uncertainties are assumed proportional to the amplitude of the submodel outputs, e.g., the flow out of the throttle. More detailed information of the sub models are described in Wahlström and Eriksson (2011). Also, actuator noise and sensor noise were added, where the standard deviation of the actuator noise is chosen as 5% of maximum value and sensor noise as 5% of a nominal value.

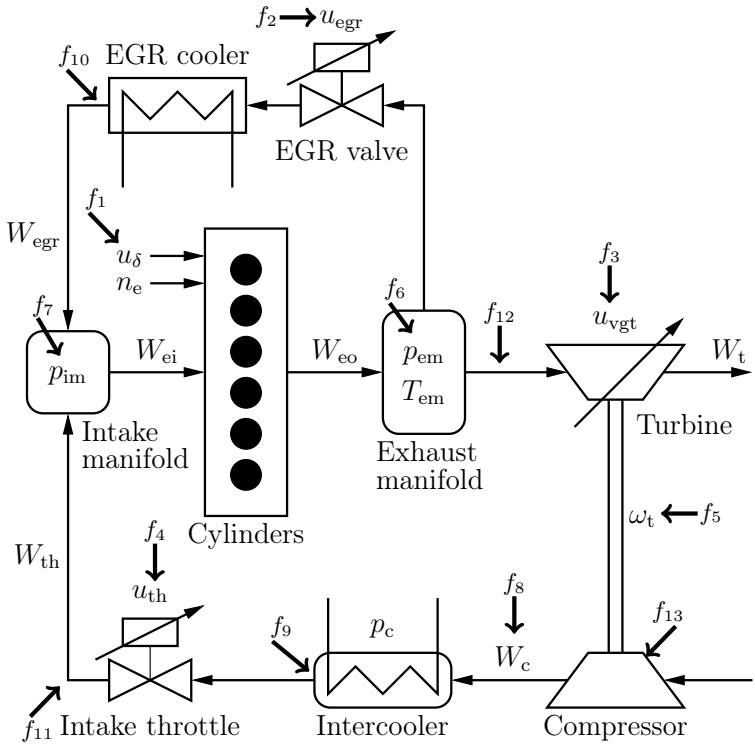

Figure 4: Overview of diesel engine model. The arrows indicate the locations in the model of the modeled faults in Table 6.

Table 6: Implemented faults in the diesel engine model where nom = "nominal value" and $\Delta p_{\mathrm{x}} = p_{\mathrm{x}} - p_{\mathrm{atm}}$.

Fault	Modeling	Descr.
f_1	$u_\delta = u_\delta^{\mathrm{nom}} + f_1$	Act. fault
f_2	$u_{\mathrm{egr}} = u_{\mathrm{egr}}^{\mathrm{nom}} + f_2$	Act. fault
f_3	$u_{\mathrm{vgt}} = u_{\mathrm{vgt}}^{\mathrm{nom}} + f_3$	Act. fault
f_4	$u_{\mathrm{th}} = u_{\mathrm{th}}^{\mathrm{nom}} + f_4$	Act. fault
f_5	$y_{\omega_t} = y_{\omega_t}^{\mathrm{nom}} + f_5$	Sensor fault
f_6	$y_{p_{\mathrm{em}}} = y_{p_{\mathrm{em}}}^{\mathrm{nom}} + f_6$	Sensor fault
f_7	$y_{p_{\mathrm{im}}} = y_{p_{\mathrm{im}}}^{\mathrm{nom}} + f_7$	Sensor fault
f_8	$y_{W_c} = y_{W_c}^{\mathrm{nom}} + f_8$	Sensor fault
f_9	$W_{c,\mathrm{leak}} = \mathrm{sgn}(\Delta p_c) f_9 \sqrt{\Delta p_c}$	Leakage
f_{10}	$W_{\mathrm{egr,leak}} = \mathrm{sgn}(\Delta p_{\mathrm{em}}) f_{10} \sqrt{\Delta p_{\mathrm{em}}}$	Leakage
f_{11}	$W_{\mathrm{th,leak}} = \mathrm{sgn}(\Delta p_c) f_{11} \sqrt{\Delta p_c}$	Leakage
f_{12}	$W_{t,\mathrm{leak}} = \mathrm{sgn}(\Delta p_{\mathrm{em}}) f_{12} \sqrt{\Delta p_{\mathrm{em}}}$	Leakage
f_{13}	$\eta_c = \eta_c^{\mathrm{nom}}(1 - f_{13})$	Degr. eff.

6.2 Diagnosability analysis of the model

The dynamic non-linear diesel engine model is analyzed to see how the distinguishability for the different faults varies with the operating point of the engine.

The non-linear model is time-continuous and in the form

$$\begin{aligned} \dot{x}(t) &= g\left(x(t), u(t), f(t), v(t)\right) \\ y(t) &= h\left(x(t), f(t), \varepsilon(t)\right). \end{aligned} \tag{30}$$

To linearize (30), the system is simulated when a constant actuator signal $u[t] = u_s$ is applied to the fault-free and noise-free system until steady-state is reached, i.e.,

$$\begin{aligned} 0 &= g(x_s, u_s, 0, 0) \\ y_s &= h(x_s, u_s, 0, 0). \end{aligned} \tag{31}$$

Then (31) is static and linearized around $x(t) = x_s$, $u(t) = u_s$, $f(t) = 0$, $v(t) = 0$, and $\varepsilon(t) = 0$ and written as a static version of (1) where $E = 0$ and

$$\begin{aligned} A &= \frac{\partial g(x, u_s, 0, 0)}{\partial x}\Big|_{x=x_s}, \; B_u = \frac{\partial g(x_s, u, 0, 0)}{\partial u}\Big|_{u=u_s}, \\ B_f &= \frac{\partial g(x_s, u_s, f, 0)}{\partial f}\Big|_{f=0}, \; B_v = \frac{\partial g(x_s, u_s, 0, v)}{\partial v}\Big|_{v=0}, \\ C &= \frac{\partial h(x, u_s, 0, 0)}{\partial x}\Big|_{x=x_s}, \; D_u = \frac{\partial h(x_s, u, 0, 0)}{\partial u}\Big|_{u=u_s}, \\ D_f &= \frac{\partial h(x_s, u_s, f, 0)}{\partial f}\Big|_{f=0}, \; D_\varepsilon = \frac{\partial h(x_s, u_s, 0, \varepsilon)}{\partial v}\Big|_{\varepsilon=0}. \end{aligned}$$

In this analysis, the model is static and no fault dynamics are considered. Therefore, the window length is chosen as $n = 1$.

Then, distinguishability can be applied for each linearization point. The operating points are selected from the World Harmonized Transient Cycle (WHTC), see for Europe Inland Transport
Committee (2010). WHTC is used world-wide in the certification of heavy duty diesel engines and should therefore be suitable to get linearization points which covers most of the operating points of the engine.

Computing distinguishability for each linearization point results in a huge amount of data that is difficult to visualize. For the engine case, 13 faults result in a 13×14 sized table of data for each linearization point, if only single faults are considered. Here, to illustrate some of the analysis results, distinguishability for each fault is plotted against a relevant system state variable to see how it varies depending on the operating point. For easier physical interpretation of the result, the square root of the computed distinguishability is plotted. We expect that distinguishability, when trying to isolate a fault $f_i = \theta_i$ from another fault mode f_j, never exceeds the detectability performance for the fault according to Proposition 2. What we do not know is how much distinguishability will decrease when isolating from the different faults.

Figure 5 shows the computed distinguishability of a leakage after the compressor f_9 from four different fault modes: no fault, a leakage after the throttle f_{11}, a fault f_8 in the sensor measuring W_c and a fault f_1 in control signal u_δ. The stars in Figure 5 correspond to the detectability performance which increases with pressure after the compressor and seems proportional to $\sqrt{p_c - p_{\text{atm}}}$. This is expected since the detection is easier with increasing flow. Also, the shape corresponds to how the fault was modeled, see Table 6. The computed distinguishability for the leakage from the fault in u_δ does not differ noticeably, from the no fault case, which is expected since the locations of the faults are not close to each other. Instead, the computed distinguishability for the leakage from the sensor fault or the leakage after the throttle are much lower since they are physically closer to each other in the model. This means that isolating a leakage after the compressor from a leakage after the throttle, or from a fault in the sensor measuring W_c, should be more difficult than only trying to detect the leakage. However, there is little difference between detectability performance and isolating from an actuator fault in u_δ.

In Figure 6, computed distinguishability for a fault f_7 in the sensor measuring the pressure p_{im} is shown. Detectability of the sensor fault, represented by the stars, has a peak around $p_{im} \approx p_{atm}$ where the fault is relatively large compared to p_{im}. When isolating the sensor fault from a compressor degradation f_{13} distinguishability is not changed which could be explained by that the faults are located far from each other. Note that distinguishability is clearly lower when isolating the sensor fault from a fault f_5 in the sensor measuring ω_t, comparing to the degradation, except an increase around $p_{im} \approx p_{atm}$. Even though the faults are close to each other the isolability performance is different. When

isolating the sensor fault from a leakage after the throttle f_{11}, distinguishability is low but increases when p_{im} decreases. This behavior could be explained by that the sensor fault becomes relatively larger when the measured pressure decreases. The increase of distinguishability around $p_{im} \approx p_{atm}$ seems to depend on the feedback from the compressor which is decoupled when isolating from the leakage.

The computed distinguishability in Figure 5 and Figure 6 shows how diagnosability performance of a non-linear model is analyzed. Non trivial results are presented but also physical interpretations of the analyzes which shows that distinguishability also could be used in the non-linear case. Both Figure 5 and Figure 6 show how distinguishability varies for different operating points. This information is useful when designing a diagnosis algorithm because it tells when it should be easiest to detect the different faults and isolate them from the other faults.

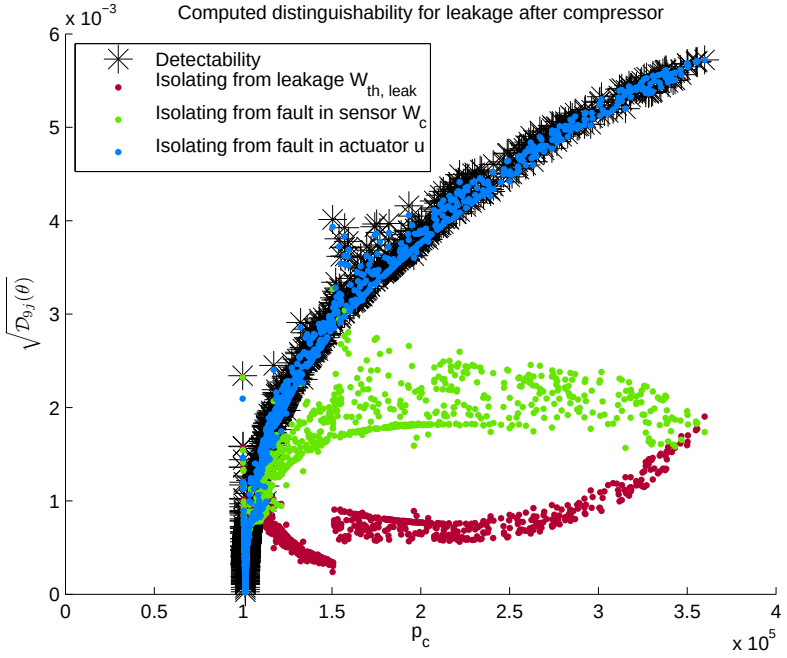

Figure 5: Computed distinguishability for a leakage after the compressor. Distinguishability for a leakage after the compressor from the no fault case, i.e., detectability performance, increases by increasing compressor pressure. Isolating the leakage from a leakage after the throttle, or a fault in the sensor measuring the mass flow through the compressor, affects the performance negatively while isolating from a fault in control signal u_δ does not affect the performance noticeably.

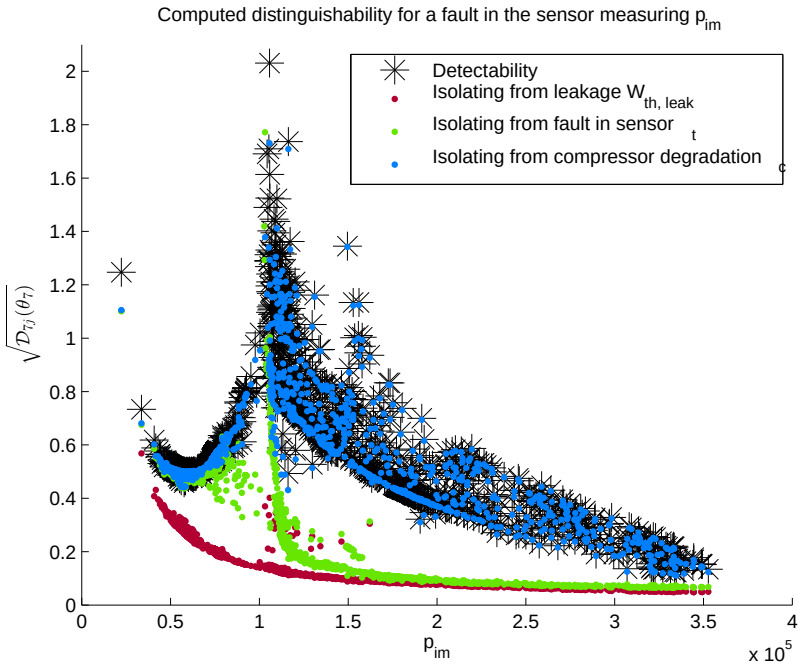

Figure 6: Computed distinguishability for an additive fault in sensor measuring p_{im}. Distinguishability is not changed when isolating from a degradation in the compressor compared to detectability. There is a peak for distinguishability around $p_{im} \approx p_{atm}$ except when isolating from a leakage after the compressor. Distinguishability increases when p_{im} goes to zero because the sensor fault becomes relatively large compared to the measured pressure.

7 Conclusions

The topic addressed in this paper is how to quantify diagnosability properties of a given model, without designing any diagnosis system. Here, discrete-time dynamic descriptor models are considered where uncertainties are described by stochastic processes with known characteristics. The descriptor model is written as a window model by considering the model dynamics for a time window of certain length.

A key contribution is the definition of *distinguishability*, a detectability and isolability performance measure, which is based on the Kullback-Leibler divergence to measure the difference between probability distributions of observations under different fault modes. It is important that distinguishability is a *model* property. Also, a method to analyze quantitative diagnosability performance

using distinguishability is derived[1].

A second key contribution is the analysis of the connection between distinguishability and residual generators. If the model uncertainties are Gaussian distributed then it is proved that distinguishability of the model gives an upper bound to the fault to noise ratio (FNR) for any residual generator. It is also shown how to design a residual generator with maximum FNR.

Finally, the developed theory and algorithms are applied to a non-linear industrial sized model of a diesel engine. The analysis is used to evaluate and exemplify an application of the methodology derived in this paper. Non-trivial results are derived on how detectability and isolability performance varies with the operating point of the diesel engine.

[1] A Matlab implementation can be found at http://www.vehicular.isy.liu.se/Software/.

References

Michèle Basseville and Igor V. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Resource Center, Pacific Grove, CA, 2001.

Robert H. Chen, D.Lewis Mingori, and Jason L. Speyer. Optimal stochastic fault detection filter. *Automatica*, 39(3):377 – 390, 2003.

Christian Commault, Jean-Michel Dion, and Sameh Yacoub Agha. Structural analysis for the sensor location problem in fault detection and isolation. *Automatica*, 44(8):2074 – 2080, 2008.

Davi Antônio dos Santos and Takashi Yoneyama. A Bayesian solution to the multiple composite hypothesis testing for fault diagnosis in dynamic systems. *Automatica*, 47(1):158 – 163, 2011.

Dilek Düştegör, Erik Frisk, Vincent Cocquempot, Mattias Krysander, and Marcel Staroswiecki. Structural analysis of fault isolability in the DAMADICS benchmark. *Control Engineering Practice*, 14(6):597 – 608, 2006.

Shinto Eguchi and John Copas. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *J. Multivar. Anal.*, 97:2034–2040, October 2006.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative stochastic fault diagnosability analysis. In *50th IEEE Conference on Decision and Control*, Orlando, Florida, USA, 2011a.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative fault diagnosability performance of linear dynamic descriptor models. 22nd International Workshop on Principles of Diagnosis (DX-11), Murnau, Germany, 2011b.

Economic Commission for Europe Inland Transport
Committee. Regulation No 49 of the Economic Commission for Europe of the United Nations (UN/ECE), August 2010.

Erik Frisk and Lars Nielsen. Robust residual generation for diagnosis including a reference model for residual behavior. *Automatica*, 42(3):437 – 445, 2006.

Erik Frisk and Mattias Nyberg. A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems. *Automatica*, 37(9):1417 – 1424, 2001.

Erik Frisk, Mattias Krysander, and Jan Åslund. Sensor Placement for Fault Isolation in Linear Differential-Algebraic Systems. *Automatica*, 45(2):364–371, 2009.

Erik Frisk, Anibal Bregon, Jan Åslund, Mattias Krysander, Belarmino Pulido, and Gautam Biswas. Diagnosability Analysis Considering Causal Interpretations for Differential Constraints. 21st International Workshop on Principles of Diagnosis (DX-10), Portland, Oregon, USA, 2010.

Janos Gertler. Fault detection and isolation using parity relations. *Control Engineering Practice*, 5(5):653 – 661, 1997.

Janos Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker Inc., Upper Saddle River, NJ, USA, 1998.

Torkel Glad and Lennart Ljung. *Control Theory : Multivariable & Nonlinear Methods*. CRC, March 2000.

Frédéric Hamelin and Dominique Sauter. Robust fault detection in uncertain dynamic systems. *Automatica*, 36(11):1747 – 1754, 2000.

Rolf Isermann. Model-based fault-detection and diagnosis - status and applications. *Annual Reviews in Control*, 29(1):71 – 85, 2005.

Steven M. Kay. *Fundamentals of statistical signal processing: Detection theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998.

Kullback, S. and Leibler, R. A. On Information and Sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

Peter Kunkel and Volker Mehrmann. *Differential-Algebraic Equations: Analysis and Numerical Solution*. European Mathematical Society, Zürich, 2006.

Mattias Nyberg. Criterions for detectability and strong detectability of faults in linear systems. *International Journal of Control*, 75(7):490–501, May 2002.

Mattias Nyberg and Erik Frisk. Residual Generation for Fault Diagnosis of Systems Described by Linear Differential-Algebraic Equations. *IEEE Transactions on Automatic Control*, 51(12):1995–2000, 2006.

Ron J. Patton, Paul M. Frank, and Robert N. Clark. *Issues of Fault Diagnosis for Dynamic Systems*. Springer Publishing Company, Incorporated, 1st edition, 2010.

J. W. Polderman and J. C. Willems. *Introduction to Mathematical Systems Theory: A Behavioral Approach*, volume 26 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1998.

Xavier Pucel, Wolfgang Mayer, and Markus Stumptner. Diagnosability analysis without fault models. In *Proceedings of the 20th International Workshop on Principles of Diagnosis, DX'09*, pages 67–74, Stockholm, Sweden, June 2009.

Rao Raghuraj, Mani Bhushan, and Raghunathan
Rengaswamy. Locating sensors in complex chemical plants based on fault diagnostic observability criteria. *AIChE Journal*, 45(2):310–322, 1999.

Louise Travé-Massuyès, Teresa Escobet, and Xavier Olive. Diagnosability Analysis Based on Component-Supported Analytical Redundancy Relations. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 36(6):1146–1160, 2006.

Johan Wahlström and Lars Eriksson. Modelling diesel engines with a variable-geometry turbocharger and exhaust gas recirculation by optimization of model parameters for capturing non-linear system dynamics. *Proceedings of the Institution of Mechanical Engineers, Part D, Journal of Automobile Engineering*, 225(7):960–986, 2011.

Timothy J. Wheeler. *Probabilistic Performance Analysis of Fault Diagnosis Schemes*. PhD thesis, University of California, Berkeley, 2011.

Ping Zhang and Steven X. Ding. Disturbance decoupling in fault detection of linear periodic systems. *Automatica*, 43(8):1410 – 1417, 2007.

Maiying Zhong, Steven X. Ding, James Lam, and Haibo Wang. An LMI approach to design robust fault detection filter for uncertain LTI systems. *Automatica*, 39(3):543 – 550, 2003.

Paper B

Using quantitative diagnosability analysis for optimal sensor placement*

B

*Published in *Proceedings of the 8th IFAC Safe Process*, Mexico city, Mexico, 2012.

Using quantitative diagnosability analysis for optimal sensor placement

Daniel Eriksson, Mattias Krysander, and Erik Frisk

Vehicular Systems, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.

Abstract

A good placement of sensors is crucial to get good performance in detecting and isolating faults. Here, the sensor placement problem is cast as a minimal cost optimization problem. Previous works have considered this problem with *qualitative* detectability and isolability specifications. A key contribution here is that *quantified* detectability and isolability performance is considered in the optimization formulation. The search space for the posed optimization problem is exponential in size, and to handle complexity a greedy optimization algorithm that compute optimal sensor positions is proposed. Two examples illustrate how the optimal solution depends on the required quantified diagnosability performance and the results are compared to the solutions using a deterministic method.

1 Introduction

In model-based diagnosis, diagnosis is performed by comparing on-line system information and a system model. The on-line information is mostly obtained by installed sensors and therefore the placement of the sensors is important for the diagnosability performance.

Sensor placement for fault diagnosis has been treated in several papers. Example of previous works are Yassine et al. (2008), Commault and Dion (2007), Krysander and Frisk (2008), Raghuraj et al. (1999), and Trave-Massuyes et al. (2006) which all use structural descriptions of the model to find a set of sensors which achieves a required deterministic isolability performance. Deterministic isolability states whether a fault is isolable or not, given the selected set of sensors. In Rosich et al. (2010) and Debouk et al. (2002) the optimal minimum cost sensor set is sought given a required deterministic isolability performance. In Frisk et al. (2009) an analytical approach is used to find all sets of sensors fulfilling the required deterministic isolability performance.

A limitation of deterministic isolability analyses is that they only provide a yes or no answer to questions like: is a fault detectable? Sensor placement based on deterministic isolability can provide sensor sets that in practice are not good for diagnosis due to noise and model uncertainties.

A method for analyzing quantified diagnosability performance, *distinguishability*, was introduced in Eriksson et al. (2011b) for linear static models, and extended to time-discrete dynamic linear descriptor models in Eriksson et al. (2011a). Distinguishability is used in this paper to optimize sensor placement for fault diagnosis to find a cheapest sensor set which achieves a required quantified diagnosability performance. The proposed method is applied to two example models where the solutions are analyzed and compared to the results using a deterministic method.

2 Introductory example

Before presenting the problem formulation in this paper, the result of using a deterministic algorithm on a linear model for finding optimal sensor sets will be discussed. Then a discussion will follow on how the performance of a diagnosis algorithm, based the computed set of sensors, is affected by model uncertainties and why this should be considered when finding optimal sensor sets.

A discretized version of a small continuous linear dynamic example model, discussed in Krysander and Frisk (2008),

$$\begin{aligned}
x_1[t+1] &= x_2[t] + x_5[t] \\
x_2[t+1] &= -x_2[t] + x_3[t] + x_4[t] \\
x_3[t+1] &= -2x_3[t] + x_5[t] + f_1[t] + f_2[t] \\
x_4[t+1] &= -3x_4[t] + x_5[t] + f_3[t] \\
x_5[t+1] &= -4x_5[t] + f_4[t]
\end{aligned} \tag{1}$$

is considered where x_i are state variables and f_i are modeled faults.

A deterministic method finds sets of sensors that achieves maximum deterministic fault isolability, i.e., a set of sensors which makes it possible to isolate all faults that are isolable from each other. A set of sensors which fulfills maximum deterministic fault isolability, where no subset of sensors fulfills it, is called a *minimal sensor set*, see Krysander and Frisk (2008).

2.1 Sensor placement using deterministic method

If x_i in (1) are possible sensor locations and model uncertainties and measurement noise are ignored, a deterministic analysis of maximum fault isolation can be performed, e.g., using the method in Frisk et al. (2009). Maximum deterministic fault isolability can be computed by including all possible sensors, and the result is summarized in Table 1. An X in position i, j represents that fault mode f_i is isolable from f_j and a 0 if not. The NF column represents if the fault is detectable, i.e., if f_i is isolable from the no fault case then it is detectable. The analysis shows that all faults are detectable, f_1 and f_2 are isolable from the faults f_3 and f_4 but not from each other, and that f_3 and f_4 are fully isolable from the other faults.

Table 1: Achievable maximum fault isolability of the example model (1).

	NF	f_1	f_2	f_3	f_4
f_1	X	0	0	X	X
f_2	X	0	0	X	X
f_3	X	X	X	0	X
f_4	X	X	X	X	0

It is not necessary to measure all states x_i in (1) to achieve the isolability in Table 1. Applying the deterministic sensor placement method in Krysander and Frisk (2008), gives all minimal sensor sets,

$$\{x_1, x_3\}, \{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_4\}, \text{ and } \{x_3, x_4\}, \tag{2}$$

that achieve the deterministic fault isolability in Table 1. Each set in (2), and all supersets, represents all sensor sets that achieves maximum deterministic fault isolability.

The minimal sensor sets in (2) are found without taking model uncertainties and measurement noise into consideration. If model uncertainties and measurement noise are considered, the choice of minimal sensor set will affect the achieved diagnosability performance. The deterministic analysis does not state which sensor set in (2) that will give the best performance of a diagnosis system. It neither gives any information if the number of sensors is enough to get sufficient diagnosability performance.

2.2 Analysis of minimal sensor sets using distinguishability

In Eriksson et al. (2011b) and Eriksson et al. (2011a) a measure, distinguishability, for quantifying diagnosability performance of time-discrete linear dynamic systems was introduced. Distinguishability gives the upper limit for the fault to noise ratio, FNR, of an residual by considering model uncertainties and fault time profiles. A *fault time profile* is a vector $\theta_i = (\theta[t-n+1], \ldots, \theta[t])^T$ describing how the fault, $f_i = \theta_i$, varies during a time period of length n. A higher distinguishability value corresponds to a higher diagnosability performance.

Before computing distinguishability for the different minimal sensor sets, (2), some assumptions are made. First, all possible sensors in this example are assumed to have additive measurement noise which is i.i.d. Gaussian with variance one, i.e., $y_i = x_i + e_i$ where $e_i \sim \mathcal{N}(0,1)$. For simplicity, it is assumed that the added sensors can not become faulty, i.e., no new faults are introduced in the model. It is also assumed that the system is observed for a time window length of five samples, and that the faults to be isolated are constant faults with amplitude one, i.e. $\theta_i = \bar{1} = (1, 1, \ldots, 1)^T$ for each fault mode f_i.

Consider first the minimal sensor set $\{x_2, x_3\}$ in (2). The computed distinguishability is presented in Table 2. A non-zero value in position i, j corresponds to a constant fault f_i is isolable from the fault mode f_j. A higher distinguishability value means that the fault is easier to detect or isolate. The same information about deterministic isolability performance as in Table 1 can be stated in Table 2 since all non-isolable fault pairs have distinguishability value zero. Table 2 also shows that, for example, it is easier to detect f_1 than f_3, since 0.308 is greater than 0.033, and it is easier to isolate f_1 from f_3 than vice versa since 0.230 is greater than 0.020.

Table 2: Distinguishability for each fault pair $\{f_i, f_j\}$, if $f_i = \bar{1}$, given the sensor set $\{x_2, x_3\}$.

$\{x_2, x_3\}$	NF	f_1	f_2	f_3	f_4
f_1	0.308	0	0	0.230	0.017
f_2	0.308	0	0	0.230	0.017
f_3	0.033	0.020	0.020	0	0.017
f_4	0.018	0.001	0.001	0.010	0

If instead the minimal sensor set $\{x_2, x_4\}$ in (2) is used, the computed distinguishability is presented in Table 3. A comparison of Table 2 and Table 3 gives that the sensor set $\{x_2, x_4\}$ makes it easier to detect and isolate f_3 from f_1 than the sensor set $\{x_2, x_3\}$, since 0.123 is greater than 0.020, but more difficult to detect and isolate f_1 and f_3, since 0.037 is less than 0.230.

The analysis shows that no minimal sensor set in (2) gives the best diagnosability performance for all pairs of fault modes. It could also be that none of the

minimal sensor sets are sufficient to get satisfactory diagnosability performance in practice. If model uncertainties are considered, when finding an optimal sensor set, then the solution could be different from the solution of the deterministic analysis.

The example shows that if model uncertainties and measurement noise are not considered, when selecting a minimal sensor set, then sufficient diagnosability performance might not be achievable if the faults are too small. If process noise and measurement noise were considered then an optimal sensor set could be found which gives a required performance, for example FNR, of the diagnosis system.

3 Problem formulation

The objective here is to utilize *distinguishability* for quantified diagnosability performance to optimize sensor placement for fault diagnosis purposes. The type of models that will be considered are time-discrete linear dynamic descriptor models written as

$$\begin{aligned} Ex[t+1] &= Ax[t] + B_u u[t] + B_f f[t] + B_v v[t] \\ y[t] &= Cx[t] + D_u u[t] + D_f f[t] + D_\varepsilon \varepsilon[t] \end{aligned} \tag{3}$$

where $x \in \mathbb{R}^{l_x}$ are state variables, $y \in \mathbb{R}^{l_y}$ are measured signals, $u \in \mathbb{R}^{l_u}$ are input signals, $f \in \mathbb{R}^{l_f}$ are modeled faults, $v \sim \mathcal{N}(0, \Lambda_v)$ and $\varepsilon \sim \mathcal{N}(0, \Lambda_\varepsilon)$ are i.i.d. Gaussian vectors with zero mean and symmetric positive definite covariance matrices $\Lambda_v \in \mathbb{R}^{l_v \times l_v}$ and $\Lambda_\varepsilon \in \mathbb{R}^{l_\varepsilon \times l_\varepsilon}$. The model matrices are of appropriate dimensions. Note that the matrix E can be singular.

Assume that a model (3), denoted with $\mathcal{M}$, and a set of possible sensors $\mathcal{O}$ are given. Each sensor $s \in \mathcal{O}$ has a sensor position and a known noise variance. Let $\mathcal{D}^{\mathcal{S}}_{i,j}(\theta_i; n)$ define distinguishability for a fault f_i with a given fault time profile θ_i and a window length n from a fault mode f_j for a given sensor set $\mathcal{S}$. A formal definition of distinguishability will be presented in Section 4. The objective is to find a minimum cost sensor set which fulfills a *minimum required distinguishability*, $\mathcal{D}^{\text{req}}_{i,j}(\theta_i; n)$, for each fault pair $\{f_i, f_j\}$.

Table 3: Distinguishability for each fault pair $\{f_i, f_j\}$, if $f_i = \bar{1}$, given the sensor set $\{x_2, x_4\}$.

$\{x_2, x_4\}$	NF	f_1	f_2	f_3	f_4
f_1	0.062	0	0	0.037	0.023
f_2	0.062	0	0	0.037	0.023
f_3	0.171	0.123	0.123	0	0.023
f_4	0.014	0.005	0.005	0.002	0

The sensor placement problem is now formulated as an optimization problem,

$$\begin{aligned} &\min_{\mathcal{S}\subseteq\mathcal{O}} \; h(\mathcal{S}) \\ &\text{s.t. } \mathcal{D}^{\mathcal{S}}_{i,j}(\theta_i;n) \geq \mathcal{D}^{\text{req}}_{i,j}(\theta_i;n),\, \forall i,j, \end{aligned} \tag{4}$$

where $\mathcal{S} \subseteq \mathcal{O}$ is a set of selected sensors, $h(\mathcal{S})$ is a cost function, and $\mathcal{D}^{\mathcal{S}}_{i,j}(\theta_i;n)$ is the achieved distinguishability for each fault pair $\{f_i, f_j\}$ given the sensors $\mathcal{S}$. The cost function $h(s)$ could, for example, be the total sensor cost

$$h(s) = \sum_{s_l \in S} \text{cost}(s_l)$$

or the total number of sensors if $\text{cost}(s_l) = 1$ for all $s_l \in \mathcal{O}$.

The objective in this paper is, given a model $\mathcal{M}$ in the form (3) and an available set of sensors $\mathcal{O}$, to find a solution to (4). That is, finding a minimum cost sensor set which fulfills the required diagnosability performance defined by $\mathcal{D}^{\text{req}}_{i,j}(\theta_i;n)$.

4 Background theory

The theory presented here is needed to define distinguishability. A more thorough description can be found in Eriksson et al. (2011b) and Eriksson et al. (2011a).

4.1 Model

Before analyzing the time-discrete descriptor model (3) it is written as a sliding window model, i.e., a sliding window of length n is applied to (3), see, e.g., Eriksson et al. (2011a). Define the vectors

$$\begin{aligned} z &= (y[t-n+1]^T, \ldots, y[t]^T, u[t-n+1]^T, \ldots, u[t]^T)^T \\ x &= (x[t-n+1]^T, \ldots, x[t+1]^T)^T, \\ f &= (f[t-n+1]^T, \ldots, f[t]^T)^T \\ e &= (v[t-n+1]^T, \ldots, v[t]^T, \varepsilon[t-n+1]^T, \ldots, \varepsilon[t]^T)^T, \end{aligned}$$

where $z \in \mathbb{R}^{n(l_y+l_u)}$, $x \in \mathbb{R}^{(n+1)l_x}$, $f \in \mathbb{R}^{nl_f}$ and $e \in \mathcal{N}(0,\Lambda_e)$ is an i.i.d. Gaussian vector with zero mean and $\Lambda_e \in \mathbb{R}^{n(l_e+l_v)\times n(l_e+l_v)}$ is a positive definite symmetric covariance matrix. Then a sliding window model of length n can be written as

$$Lz = Hx + Ff + Ne \tag{5}$$

where

$$L = \begin{pmatrix} 0 & 0 & \dots & 0 & -B_u & 0 & \dots & 0 \\ I & 0 & & 0 & -D_u & 0 & & 0 \\ 0 & 0 & & 0 & 0 & -B_u & & 0 \\ 0 & I & & 0 & 0 & -D_u & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & & 0 & 0 & 0 & & 0 & -B_u \\ 0 & \dots & 0 & I & 0 & \dots & 0 & -D_u \end{pmatrix}, H = \begin{pmatrix} A & -E & 0 & \dots & 0 & 0 \\ C & 0 & 0 & & 0 & 0 \\ 0 & A & -E & & 0 & 0 \\ 0 & C & 0 & & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & & 0 & A & -E \\ 0 & 0 & \dots & 0 & C & 0 \end{pmatrix},$$

$$F_n = \begin{pmatrix} B_f & 0 & \dots & 0 \\ D_f & 0 & & 0 \\ 0 & B_f & & 0 \\ 0 & D_f & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & 0 & B_f \\ 0 & & 0 & D_f \end{pmatrix}, N = \begin{pmatrix} B_v & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & & 0 & D_\varepsilon & 0 & & 0 \\ 0 & B_v & & 0 & 0 & 0 & & 0 \\ 0 & 0 & & 0 & 0 & D_\varepsilon & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & B_v & 0 & 0 & & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & D_\varepsilon \end{pmatrix}.$$

The sliding window model (5) is a *static* representation of the dynamic behavior on the window given the time indexes $(k - n + 1, ..., k)$.

To guarantee, given the sliding window model (5), that no noise-free residuals can be created, it is assumed that

$$\begin{pmatrix} H & N \end{pmatrix} \text{ has full row-rank.} \tag{6}$$

Assumption (6) is fulfilled, e.g., if all sensors have measurement noise. To simplify the computations, it is assumed that the covariance matrix $\bar{\Sigma}$ of variable $N_H Le$ is equal to the identity matrix, that is

$$\bar{\Sigma} = N_H N \Lambda N^T N_H^T = I \tag{7}$$

where the rows of N_H forms an orthonormal basis for the left null-space of matrix H. Note that any model satisfying (6) can be transformed into fulfilling $\bar{\Sigma} = I$. The choice of an invertible transformation matrix T is non-unique and one possibility is

$$T = \begin{pmatrix} \Gamma^{-1} N_H \\ T_2 \end{pmatrix} \tag{8}$$

where Γ is non-singular and satisfying

$$N_H N \Lambda N^T N_H^T = \Gamma \Gamma^T \tag{9}$$

and T_2 is any matrix ensuring invertibility of T.

It is convenient to eliminate the unknown variables x in (5) by multiplying with N_H from the left such that

$$N_H Lz = N_H F f + N_H N e. \tag{10}$$

The model (10) is in an input-output form. For any solution z_0, f_0, e_0 to (10) there exists an x_0 such that it also is a solution to (5), and also if there exists a solution z_0, f_0, e_0, x_0 to (5) then z_0, f_0, e_0 is a solution to (10). Thus no information about the model behavior is lost when rewriting (5) as (10).

To quantify diagnosability performance, define the vector $r = N_H L z$. The vector $r \in \mathbb{R}^{n(l_y - l_x)}$ depends on the fault vector f and the noise vector e and represents the behavior of the model (5).

4.2 Quantified diagnosability performance

Let $p^i_{\theta_i}$ be the probability density function, pdf, describing the vector r when there is a fault f_i present in the system represented by the fault time profile θ_i.

The set of pdf's of r representing the fault mode f_i, corresponding to all possible fault time profiles θ_i is defined as

$$\mathcal{Z}_{f_i} = \left\{ p^i_{\theta_i} | p^i_{\theta_i} \text{ consistent with fault mode } f_i \right\}. \tag{11}$$

Each fault mode f_i is thus described by a set $\mathcal{Z}_{f_i}$ of all pdf's consistent with the fault mode. Consider two different sets, $\mathcal{Z}_{f_i}$ and $\mathcal{Z}_{f_j}$, for two fault modes f_i and f_j in Figure 1. Assume that there is a measure to quantify the distance from a specific pdf $p^i_{\theta_i} \in \mathcal{Z}_{f_i}$ given a fault time profile θ_i to any $p^j_{\theta_j} \in \mathcal{Z}_{f_j}$. Then, the shortest distance from $p^i_{\theta_i}$ to any pdf in $\mathcal{Z}_{f_j}$ is a quantified isolability performance of a fault $f_i = \theta_i$ from the fault mode f_j.

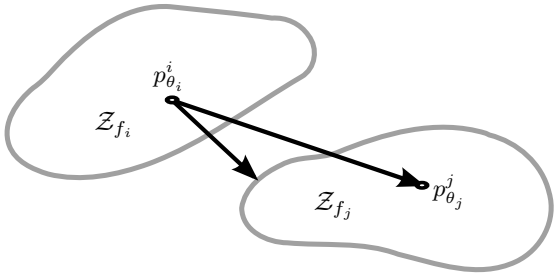

Figure 1: A graphical visualization where the smallest difference between $p^i_{\theta_i} \in \mathcal{Z}_{f_i}$ and a pdf $p^j \in \mathcal{Z}_{f_j}$ is the quantified diagnosability measure.

To quantify the difference between the pdf's, $p^i_{\theta_i}$ and $p^j_{\theta_j}$, of r for two faults $f_i = \theta_i$ and $f_j = \theta_j$ the Kullback-Leibler divergence

$$K(p^i_{\theta_i} \| p^j_{\theta_j}) = \int_{-\infty}^{\infty} p^i_{\theta_i}(v) \log \frac{p^i_{\theta_i}(v)}{p^j_{\theta_j}(v)} dv = \mathrm{E}_{p^i_{\theta_i}} \left[\log \frac{p^i_{\theta_i}}{p^j_{\theta_j}} \right] \tag{12}$$

is used, see Kullback and Leibler (1951).

Then, to quantify isolability of a fault mode f_i with fault time profile θ_i from a fault mode f_j with an unknown fault time profile, a measure for isolability performance is defined as follows.

Definition 1 (Distinguishability)**.** *Given a sliding window model* (5) *of length* n*, under assumption* (6)*, distinguishability* $\mathcal{D}_{i,j}(\theta_i; n)$ *of a fault* f_i *with a given fault time profile* θ_i *from a fault mode* f_j *is defined as*

$$\mathcal{D}_{i,j}(\theta_i; n) = \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p^i_{\theta_i} \| p^j\right). \tag{13}$$

Distinguishability can be used to analyze either isolability or detectability performance depending on whether $\mathcal{Z}_{f_j}$ describes a fault mode or the fault free case. Note that distinguishability is asymmetric in general, i.e., $\mathcal{D}_{i,j}(\theta_i; n) \neq \mathcal{D}_{j,i}(\theta_i; n)$, which is a natural property.

By using Theorem 2 and Theorem 3 in Eriksson et al. (2011a), distinguishability gives the maximum achievable FNR for any residual given the window model (5) of length n.

Theorem 1. *For a window model* (5) *of length* n *under assumption* (7)*, a tight upper bound for the fault to noise ratio of any residual based on* (5) *is given by*

$$\mathcal{D}_{i,j}(\theta_i; n) \geq \frac{1}{2}\left(\frac{\lambda}{\sigma}\right)^2$$

where $\lambda(\theta_i)/\sigma$ *is the fault to noise ratio for a residual with respect to fault* f_i *and a fault time profile* θ_i*.*

For a sliding window model (5) an explicit computation of (13) is stated in the following theorem.

Theorem 2. *Distinguishability for a sliding window model* (5) *under assumption* (6) *is given by*

$$\mathcal{D}_{i,j}(\theta_i; n) = \frac{1}{2}\|N_{\bar{H}} F_i \theta_i\|^2 \tag{14}$$

where $\bar{H} = \begin{pmatrix} H & F_j \end{pmatrix}$ *and the rows of* $N_{\bar{H}}$ *is an orthonormal basis for the left null space of* $\bar{H}$*.*

Proofs of Theorem 1 and Theorem 2 can be found in Eriksson et al. (2011b).

A detectability and isolability analysis of the descriptor model (3) can be made using distinguishability if the model is written as a sliding window model (5). The distinguishability measure depends on the window length n and the fault time profile θ_i.

5 The small example revisited

Consider again the example model (1). This time a minimal sensor set is sought which is a solution to the optimization problem (4). It is assumed that the faults to be detected are constants over time with amplitude one and a window model of length $n = 5$ is used when computing distinguishability.

Selecting the appropriate constraints, $\mathcal{D}^{\text{req}}_{i,j}(\bar{1};5)$ for each fault pair, $\{f_i, f_j\}$, can be difficult if $\mathcal{D}^{\text{req}}_{i,j}(\bar{1};5)$ contains many elements. A more convenient approach is to select $\mathcal{D}^{\text{req}}_{i,j}(\bar{1};5)$ as a fraction p of *maximum achievable distinguishability*, $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$ for each fault pair $\{f_i, f_j\}$, where $p \in [0,1]$ is a scalar. In this way only one parameter is required for all elements. Note that there is still complete freedom in selecting $\mathcal{D}^{\text{req}}_{i,j}(\bar{1};5)$ for each fault pair individually. As when maximum fault isolability in Section 2.1 was determined, the maximum achievable distinguishability can be computed by including all sensors in $\mathcal{O}$. The computed $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$ is shown in Table 4. If $\mathcal{D}^{\text{req}}_{i,j}(\bar{1};5)$ is selected higher than $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$ for any fault pair, then the optimization problem can not be solved. By comparing Table 2 and Table 3 to Table 4 show that none of the two minimum sensors sets reaches $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$.

Table 4: Maximum achievable distinguishability for each fault pair, if the maximum number of sensors is used.

$\mathcal{D}^{\max}_{i,j}(\bar{1};5)$	NF	f_1	f_2	f_3	f_4
f_1	0.385	0	0	0.341	0.275
f_2	0.385	0	0	0.341	0.275
f_3	0.213	0.187	0.187	0	0.161
f_4	0.251	0.177	0.177	0.187	0

Assume that a minimal sensor set is to be found which achieves at least 50% of $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$ for each fault pair. That is, the sensor placement problem can be written as

$$\begin{aligned} &\min_{\mathcal{S}\subseteq\mathcal{O}} |\mathcal{S}| \\ &\text{s.t. } \mathcal{D}^{\mathcal{S}}_{i,j} \geq 0.5 \times \mathcal{D}^{\max}_{i,j}(\bar{1};5),\ i,j = 1,2,3,4 \end{aligned} \tag{15}$$

where $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$, and $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$ can be found in Table 4. A global search gives that a solution to (15) is the unique optimal sensor set, of cardinality four, which measures the states: x_1, x_3, x_4, and x_5.

The analysis is expanded to see how the cardinality of the minimal sensor set, required to achieve a fraction p of $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$, depends on p. The result is presented in Figure 2. Note that the minimum number of required sensors coincides with the cardinality of the minimal sensor sets (2) given by the deterministic analysis when $p \to 0+$.

Since there is only one minimal sensor set with four sensors achieving at least 50% of maximum distinguishability, the analysis in Figure 2 gives that the minimal sensor set measuring the states: x_1, x_3, x_4, and x_5, achieves almost 80% of $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$. The number of sensors in the minimal sensor sets given by the structural analysis is two, which is not able to achieve more than 3% of $\mathcal{D}^{\max}_{i,j}(\bar{1};5)$.

The result of the analysis in this section shows that the minimum cost sensor sets (2) given by the deterministic analysis results in a solution where the

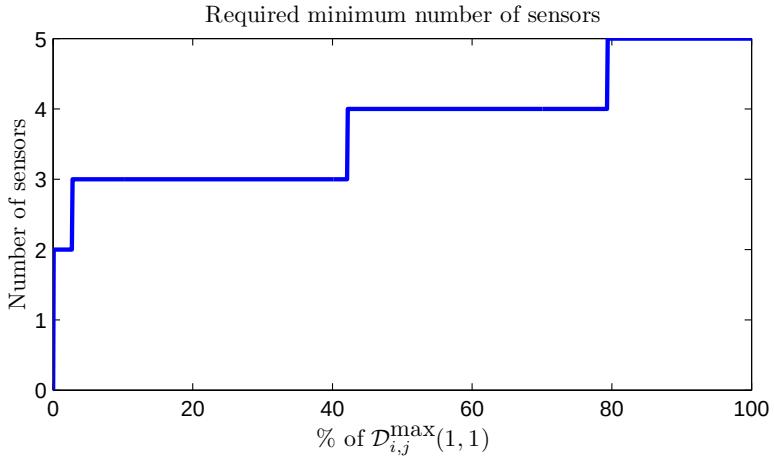

Figure 2: The least number of sensors required to exceed a certain percentage of maximum distinguishability given the example model (1). Note that the number of sensors when required performance goes to zero is equal to the cardinality of the minimal sensor sets (2) from the structural analysis.

achieved diagnosability performance is relatively low given $\mathcal{D}_{i,j}^{\max}(\bar{1};5)$ ($< 3\%$ of $\mathcal{D}_{i,j}^{\max}(\bar{1};5)$). By using minimum required distinguishability as the constraints of the optimization problem, a solution is found which better fits the requirements when designing a diagnosis system.

For the small system (1), a global search could be performed to find the solution. For larger systems, this is not realistic because of high computational complexity. The number of sensor combinations, 2^k where k is the number of possible sensors, grows exponentially with the total number of sensors. A more efficient algorithm to reduce complexity is needed to find the optimal solution. An algorithm which iteratively adds new sensors to the solution, would be more appealing since it reduces the complexity. A heuristic is needed to implement such an iterative approach.

6 A GREEDY SEARCH APPROACH

A heuristic greedy search algorithm starts with an empty set and iteratively adds the sensor with the largest utility to the solution. The iteration continues until the solution fulfills the constraints. Thus, a utility function must be defined to use the greedy search heuristic.

In the iterative search, the heuristic adds the sensor s which best improves the previously selected set of sensors $\mathcal{S}$ to fulfill the constraints, i.e., the algorithm adds the sensor $s \in \mathcal{O} \setminus \mathcal{S}$ that maximizes the utility function $\mu(s)$. The utility

function is here the sum, over all fault pairs, of the distinguishability improvements when adding a sensor s. There is no utility in improving distinguishability more than what is required by $\mathcal{D}_{i,j}^{\text{req}}$. Thus, the utility function can be written as

$$\mu(s) = \sum_{i,j} \max\left(\min\left(\mathcal{D}_{i,j}^{\text{req}}, \mathcal{D}_{i,j}^{\mathcal{S}\bigcup\{s\}}\right) - \mathcal{D}_{i,j}^{\mathcal{S}}, 0\right).$$

The algorithm SelMinSensSetGreedy for greedy selection of minimal sensor set is given below. The inputs to the algorithm are the model $\mathcal{M}$ in the form (3), a set of sensors $\mathcal{O}$ where each sensor measures one model variable and has a known noise variance, and a minimum required distinguishability $\mathcal{D}_{i,j}^{\text{req}}$. The output from the algorithm is a set of sensors $\mathcal{S}$. If the achieved distinguishability, $\mathcal{D}_{i,j}^{\mathcal{S}}$, given the set of sensors $\mathcal{S}$ fulfills the constraints $\mathcal{D}_{i,j}^{\text{req}}$ then the solution $\mathcal{S}$ is returned. If $\mathcal{D}_{i,j}^{\text{req}}$ is lower than the maximum achievable distinguishability, given $\mathcal{M}$ and $\mathcal{O}$ then SelMinSensSetGreedy will always return a set of sensors $\mathcal{S}$ fulfilling the constraints in (4).

```
1: function SelMinSensSetGreedy(M, O, D^req_{i,j}(θ_i; n))
2:     S := ∅
3:     while O ≠ ∅ do
4:         s* := arg max_{s∈O} μ(s)
5:         S := S ⋃ {s*}
6:         O := O \ s*
7:         if D^S_{i,j}(θ_i; n) ≥ D^req_{i,j}(θ_i; n), ∀i, j then
8:             return S
9:         end if
10:    end while
11:    return S
12: end function
```

The complexity of the algorithm SelMinSensSetGreedy is linear in the number of sensors in $\mathcal{O}$. This approach is faster than a global search, however, the approach can of course not guarantee that the found solution is optimal.

7 Sensor placement using greedy search

In this section, the greedy algorithm presented in Section 6 is applied to a slightly larger example, but still small enough to compare the solution to the global optimal solution.

Consider the sensor placement problem given a time-discrete linear model where *one* sensor can be selected to measure each unknown variable x_i. A minimum cost sensor set is to be found which achieves a minimum required distinguishability for each fault pair.

7.1 Model

The example is a static linear model describing a flow through a set of 15 branches visualized in Figure 3. The input flow u is known and x_i represents

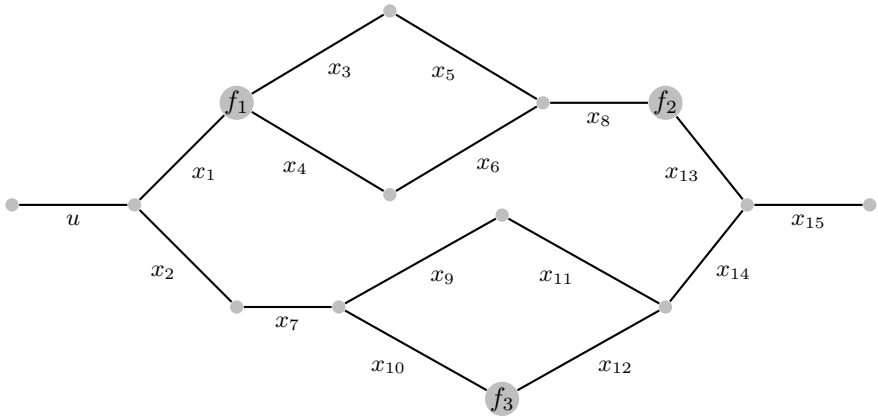

Figure 3: A schematic overview of a model describing a static flow through a number of nodes. The input u is known and the flows through the branches x_i, $i = 1, 2, \ldots, 15$, are unknown, and f_1, f_2, f_3 are three additive faults.

the flow through each branch. For each node the sum of all branches is zero, e.g. $x_8 = x_5 + x_6$. The equations describing the flow through each node, has a model uncertainty which is assumed additive i.i.d. Gaussian $\mathcal{N}(0, 0.1)$, defines the model $\mathcal{M}$. The set of available sensors $\mathcal{O}$, one for each flow x_i, has measurement noise $\mathcal{N}(0, 1)$. There are three possible leaks added to the model, f_i, $i = 1, 2, 3$, and they are assumed to be additive in the equation describing the flow through the specific node, e.g., $x_3 + x_4 = x_1 + f_1$. To compute distinguishability, a window model of length one is assumed since the model is static and all fault time profiles are assumed amplitude one.

Note that, in the model described above, there are no equations in the model describing how the flow splits when the flow branches. This underdetermined model is analyzed first, and then in a second step the model is extended with equations describing how flow splits in the branches. This exactly determined model is analyzed to illustrate how diagnosability performance changes with modelling effort.

7.2 Analysis of the underdetermined model

Using the global search algorithm, the minimum number of sensors will depend on the minimum required distinguishability $\mathcal{D}^{\text{req}}_{i,j}(1;1)$. A higher $\mathcal{D}^{\text{req}}_{i,j}(1;1)$ requires more sensors added to the system to be fulfilled. The global search is used to analyze how the cardinality of $\mathcal{S}$ depends on $\mathcal{D}^{\text{req}}_{i,j}(1;1) = p\mathcal{D}^{\max}_{i,j}(1;1)$. The maximum achievable distinguishability for each fault pair is shown in Table 5. The solid line in Figure 4 shows that the minimal number of sensors which achieves full deterministic isolability performance, i.e., four sensors, only is able

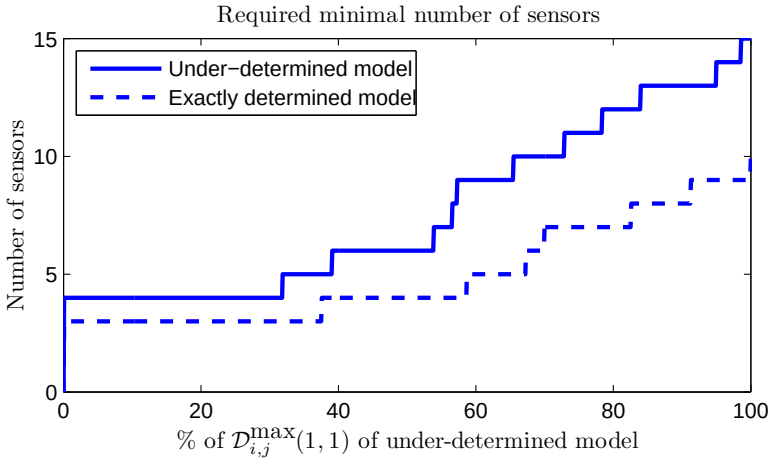

Figure 4: The least number of sensors required to exceed a certain percentage of maximum distinguishability given the underdetermined model described in Figure 3, for the underdetermined case and the exactly determined case, given a global search.

to achieve approximately 30% of maximum distinguishability, for each fault pair. To achieve 70% of $\mathcal{D}_{i,j}^{\max}(1;1)$ for each fault pair a solution requires at least that 10 of 15 unknown variables, x_i, are measured by a sensor.

Table 5: Maximum achievable distinguishability for the underdetermined model in Figure 3.

$\mathcal{D}_{i,j}^{\max}(1;1)$	NF	f_1	f_2	f_3
f_1	1.228	0	0.693	0.875
f_2	0.831	0.470	0	0.621
f_3	1.086	0.774	0.812	0

Assume that a minimal sensor set is to be found using the greedy algorithm which fulfills 50% of maximum distinguishability for each fault pair. The solution requires six sensors, selected by the greedy algorithm in the following order: x_{15}, x_{14}, x_8, x_7, x_{13}, and x_2. In this case the solution of the greedy algorithm is optimal since it has the same cardinality as the optimal solution given by the global search in Figure 4.

The greedy algorithm always tries to find the sensor which best improves the distinguishability to fulfill the constraints and therefore some solutions could be missed. Consider, for example, in Figure 3 that a set of sensors are to be found to isolate f_1 from f_2 and f_3. Assume that the optimal solution would be

to measure x_1, x_3 and x_4, and that the algorithm has already selected x_1. Since neither only measuring x_3 or x_4 is enough to isolate f_1, a sensor measuring, for example, x_8 is selected. This sensor selection will improve the solution of the greedy search locally but will miss the global optimal solution. A solution to this could be to use the solution of a deterministic analysis and then improving the result by adding more sensors. Alternatively, more advanced search methods could be used, see e.g. Russell and Norvig (2003).

7.3 Analysis of the exactly determined model

Assume now that the flows in Figure 3 where a branch is split are approximately equal, i.e., $x_3 \approx x_4$. This information is included in the underdetermined model, for example, by adding the equations $x_1 = x_2 + v_1$, $x_3 = x_4 + v_2$, and $x_9 = x_{10} + v_3$, where $v_i \sim \mathcal{N}(0, 1)$, which makes the model exactly determined. The solution of the greedy algorithm, given the exactly determined model, which fulfills 50% of the maximum achievable distinguishability of the underdetermined model, in Table 5, is: x_{13}, x_{14}, x_8, and x_{15}. Note that the exactly determined model has a higher maximum achievable distinguishability than the underdetermined model, see Table 6. This is expected as additional process knowledge is incorporated in the model.

Table 6: Maximum achievable distinguishability for the exactly determined model in Figure 3.

$\mathcal{D}_{i,j}^{\max}(1;1)$	NF	f_1	f_2	f_3
f_1	1.399	0	0.786	1.162
f_2	0.849	0.477	0	0.680
f_3	1.192	0.990	0.954	0

An extended analysis of the exactly determined case can be seen as a dashed line in Figure 4. The number of sensors required to achieve a certain percentage of maximum distinguishability is lower compared to the underdetermined case. For the exactly determined case, only ten sensors are needed to achieve $\mathcal{D}_{i,j}^{\max}(1;1)$ of the underdetermined model, in Table 5, where 15 sensors are required for the underdetermined case. Comparing the results analyzing the underdetermined and the exactly determined model shows that better diagnosability performance can be achieved using fewer sensors at the price of more modeling work.

8 Conclusion

A key contribution in this paper, is the use of quantitative diagnosability analysis, distinguishability, to find optimal sensor sets for diagnosis. The sensor placement problem is formulated as a minimum cost optimization problem and a main

observation is that the optimal solutions here differ significantly from solutions given by previously published deterministic methods.

The search space for the optimization problem is exponential in size and a heuristic greedy search algorithm is proposed as a solution to this for large problems. The algorithm iteratively adds the sensor which best improves diagnosability to fulfill the requirements.

Two examples are analysed to illustrate properties of the optimal solutions when using quantified diagnosability performance in the sensor placement optimization, e.g., how the number of sensors in the solution depends on the required diagnosability performance, and that better diagnosability performance can be achieved using fewer sensors by improving the model.

References

Christian Commault and Jean-Michel Dion. Sensor location for diagnosis in linear systems: A structural analysis. *IEEE Trans. on Automatic Control*, 52 (2):155 –169, feb. 2007.

Rami Debouk, Stéphane Lafortune, and Demosthenis Teneketzis. On an optimization problem in sensor selection. *Discrete Event Dynamic Ssystems: Theory and Applications*, 14(4):417–445, 2002.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative fault diagnosability performance of linear dynamic descriptor models. 22nd International Workshop on Principles of Diagnosis (DX-11), Murnau, Germany, 2011a.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative stochastic fault diagnosability analysis. In *50th IEEE Conference on Decision and Control*, USA, 2011b.

Erik Frisk, Mattias Krysander, and Jan Åslund. Sensor placement for fault isolation in linear differential-algebraic systems. *Automatica*, 45(2):364–371, 2009.

Mattias Krysander and Erik Frisk. Sensor placement for fault diagnosis. *IEEE Trans. Syst. Man, Cybern. A, Syst., Humans*, 38(6):1398–1410, 2008.

S. Kullback and R. A. Leibler. On information and sufficiency. 1951.

Rao Raghuraj, Mani Bhushan, and Raghunathan Rengaswamy. Locating sensors in complex chemical plants based on fault diagnostic observability criteria. *AIChE Journal*, 45(2):310 – 322, 1999.

Albert Rosich, Abed Alrahim Yassine, and Stéphane Ploix. Efficient optimal sensor placement for structural model based diagnosis. 21th International Workshop on Principles of Diagnosis (DX-10), USA, 2010.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition edition, 2003.

L. Trave-Massuyes, T. Escobet, and X. Olive. Diagnosability analysis based on component-supported analytical redundancy relations. *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, 36(6):1146 –1160, nov. 2006.

Abed Yassine, Stéphane Ploix, and Jean-Marie Flaus. A method for sensor placement taking into account diagnosability criteria. *Int. J. Appl. Math. Comput. Sci.*, 18:497–512, December 2008.

Paper C

A sequential test selection algorithm for fault isolation*

C

*Published in *Proceedings to the 10th European Workshop on Advanced Control and Diagnosis*, Copenhagen, Denmark, 2012.

A sequential test selection algorithm for fault isolation

Daniel Eriksson, Erik Frisk, and Mattias Krysander

Vehicular Systems, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.

Abstract

A sequential test selection algorithm is proposed which updates the set of active test quantities depending on the present minimal candidates. By sequentially updating the set of active test quantities, computational time and memory usage can be reduced. If test quantities are generated on-line, a sequential test selection algorithm gives information about which test quantities that should be created. The test selection problem is defined as an optimization problem where a set of active test quantities is chosen such that the cost is minimized while the set fulfills a required minimum detectability and isolability performance. A quantitative diagnosability measure, distinguishability, is used to quantify diagnosability performance of test quantities. The proposed test selection algorithm is applied to a DC-circuit where the diagnosis algorithm generates residuals on-line. Experiments show that the sequential test selection algorithm can significantly reduce the number of active test quantities during a scenario and still be able to identify the true faults.

1 Introduction

A diagnosis algorithm uses test quantities to detect and isolate faults present in the system. An overview of a diagnosis algorithm is shown in Figure 1. If different test quantities are sensitive to different sets of faults a fault isolation algorithm can be used to isolate and identify which faults that are present given the alarmed test quantities, see de Kleer and Williams (1987). Often, different sets of faults, *candidates*, are consistent with the set of alarmed tests. The goal is to have a set of tests which can identify the present faults.

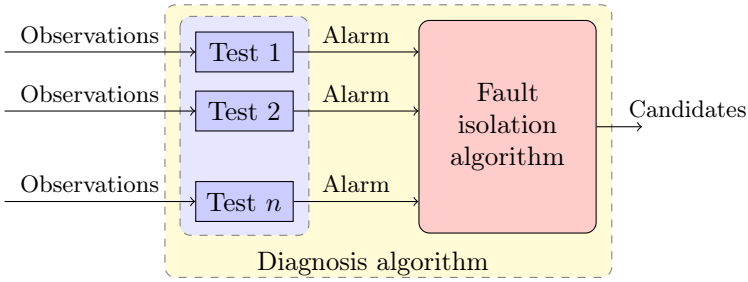

Figure 1: A diagnosis algorithm consists of a set of test quantities and a fault isolation algorithm to compute (minimal) candidates.

Since different test quantities are good at detecting and isolating different sets of faults, not all available test quantities are necessary to be active during a scenario in the diagnosis algorithm. By choosing different sets of test quantities depending on the present candidates, the total number of active test quantities can be reduced while still having sufficient diagnosability performance.

A sequential test selection algorithm becomes interesting, for example, if there exists a large number of available test quantities or if new test quantities should be automatically selected and generated on-line. To use all available test quantities during a scenario could be too computationally expensive. Therefore, a smart test selection algorithm can be used to reduce the number of active test quantities while maintaining a satisfactory diagnosability performance. As an example, a system considered here is a re-configurable DC-circuit where a number of relays are used to distribute power from three batteries to two load banks. A number of sensors are used to measure voltages and currents in the circuit. The DC-circuit is a part of the diagnosis test bed ADAPT which is described in Poll et al. (2007), see Figure 2. The number of test quantities that can be generated based on a model of the DC-circuit is large because of high redundancy in the system.

A quantitative diagnosability measure called distinguishability which is introduced in Eriksson et al. (2011a) and Eriksson et al. (2011b) is used when implementing the diagnosis algorithm to automatically generate residual generators with maximum fault to noise ratio. Distinguishability will be used to

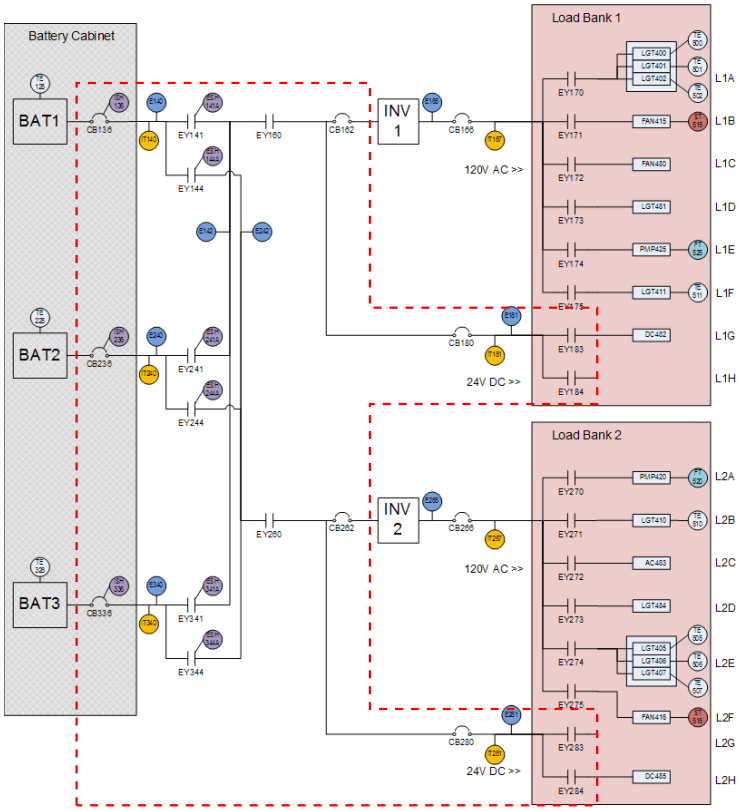

Figure 2: Schematics of the diagnosis test bed ADAPT. The dark area to the left represents the three batteries and the two dark areas to the right are two load banks. Sensors in the circuit are marked as circles. The DC-circuit considered in this work is inside the dotted line.

quantify detectability and isolability performance of the residual generators in order to choose the residual generators which fulfills a required quantitative diagnosability performance.

There are other works where sequential diagnosis are used to improve fault isolation, for example de Kleer and Williams (1987). In Krysander et al. (2010), a sequential test selection algorithm, FlexDX, updates the active set of test quantities during run-time to isolate the faults. In Svärd (2012), a greedy test selection algorithm is used off-line to find a set of test quantities which fulfills a required single fault isolability performance. In these previous works, only deterministic fault detectability and isolability performance are considered. In this work, a measure for quantitative diagnosability performance is used to quantify the performance of a test quantity when choosing the active set of test quantities.

First, the problem formulation is presented in Section 2. Then some previous results regarding distinguishability are presented in Section 3 and a generalization of distinguishability for multiple faults is presented in Section 4. A sequential test selection algorithm is presented in Section 5. Some experimental results when applying the test selection algorithm in a diagnosis algorithm for a DC-circuit are shown in Section 6. Then some tuning aspects of the test selection algorithm are discussed and compared to other works in Section 7 and finally some conclusions are presented in Section 8

2 Problem formulation

The purpose here is to develop a sequential test selection algorithm to automatically compute which available test quantities that should be active in the diagnosis algorithm during run-time depending on the present candidates. A candidate is a hypothesis about the system state given the alarmed test quantities, de Kleer and Williams (1987).

A candidate represents a set of faults $d = \{f_1, f_2, \ldots, f_l\}$, that is consistent with the alarmed test quantities. If d is a candidate where no subset $d' \subset d$ is a candidate then d is called a minimal candidate.

Let $\mathcal{T} = \{T_1, T_2, \ldots, T_n\}$ be the set of all available test quantities T. Each test quantity T is sensitive to a set of faults f_i. The cost to use a test quantity T_i is defined as $c(T_i)$. The cost could, for example, be the computational cost of generating and using T_i. Let D be a set of minimal candidates d_i, i.e., $D = \{d_1, d_2, \ldots, d_k\}$. We want to find a cheapest set of tests $\mathcal{T}^* \subseteq \mathcal{T}$ where each minimal candidate $d \in D$ can be rejected by at least one test quantity $T \in \mathcal{T}^*$. The test selection problem can be formulated as a binary integer programming (BIP) problem,

$$\begin{aligned} &\min \sum_{T \in \mathcal{T}} c(T) x_T \\ &s.t. \sum_{T \in \mathcal{E}_{d_i,d_j}} x_T \geq 1, \forall \mathcal{E}_{d_i,d_j} : d_j \in D \\ &x_T = \{0, 1\}, \forall T \end{aligned} \tag{1}$$

where $\mathcal{E}_{d_i,d_j} \subseteq \mathcal{T}$ contains a set of tests which can reject d_j for a new minimal candidate $d_i = \{f_a\} \cup d_j$ where $f_a \notin d_j$ and x_T is a binary variable determining whether a test quantity T should be active or not. The solution $\mathcal{T}^*$ contains all tests $T \in \mathcal{T}$ where $x_T = 1$.

To assure that each test $T \in \mathcal{E}_{d_i,d_j}$ have satisfactory performance, distinguishability is used to quantify the diagnosability performance for each test T. Each set $\mathcal{E}_{d_i,d_j}$ can be chosen such that, for example, only test quantities which have a sufficiently high distinguishability when rejecting d_j for d_i are included.

3 Background theory

Here, some useful results from Eriksson et al. (2011a) and Eriksson et al. (2011b) are recalled. A definition of distinguishability and some useful properties are presented here which will be used for quantifying detectability and isolability performance. There are also some results of how to generate residual generators with maximum fault to noise ratio, FNR. For more details and proofs, the reader is referred to previous mentioned papers.

3.1 Distinguishability

The models considered here are discrete-time linear models written in the form

$$Lz = Hx + Ff + Ne \tag{2}$$

where z are known signals, x are unknown signals, f are fault signals and e is Gaussian distributed with known co-variance matrix Λ. Discrete-time descriptor models when observed for a given time interval n can be written as (2), see Eriksson et al. (2011b). Thus, model dynamics and fault time profiles $\theta = (\theta(t-n+1), \theta(t-n+2), \ldots, \theta(t))^T$, describing how a fault changes over the time interval can be considered. It is assumed that the model (2) fulfills

$$\begin{pmatrix} H & N \end{pmatrix} \text{ is full row-rank.} \tag{3}$$

The model (2) is written in an input-output form, see Polderman and Willems (1998), by multiplying with $\mathcal{N}_H$ from the left, where the rows of $\mathcal{N}_H$ is an orthonormal basis for the left null-space of H, which gives

$$\mathcal{N}_H Lz = \mathcal{N}_H Ff + \mathcal{N}_H Ne. \tag{4}$$

If the linear model (2) fulfills assumption (3) the co-variance matrix of the vector $\mathcal{N}_H Ne$ will be non-singular.

To be able to quantify diagnosability performance, a stochastic representation of different fault modes is required. Let $\tau = \mathcal{N}_H Lz$, then let $p(\tau, \mu)$ denote a multivariate probability density function where $\mu(\theta) = \mathcal{N}_H F_i \theta$ is the mean of τ where F_i is the ith row of F corresponding to fault mode f_i.

Let Θ_i denote the set of all fault time profiles θ of fault $f_i = \theta$ which could be explained by fault mode f_i. Each fault mode can thus be described by a set of pdf's $p(\tau, \mu)$ given the following definition.

Definition 1. *Let $\mathcal{Z}_{f_i}$ denote the set of all pdf's, $p(\tau, \mu)$, for all fault time profiles $\theta \in \Theta_i$, describing τ which could be explained by the fault mode f_i, i.e.,*

$$\mathcal{Z}_{f_i} = \{p(\tau, \mathcal{N}_H F_i \theta) | \forall \theta \in \Theta_i\}. \tag{5}$$

■

Definition 1 is a stochastic counterpart to observation sets in the deterministic case, see Frisk et al. (2009). A specific pdf given that $f_i = \theta$ is denoted $p_\theta^i = p(\tau, \mathcal{N}_H F_i \theta)$ and $p_{\text{NF}} = p(\tau, \bar{0})$ corresponds to the no fault case.

Now, the measure for quantitative diagnosability performance can be defined by using the Kullback-Leibler divergence, $K(p\|q) = \int_{-\infty}^{\infty} p(\nu) \log \frac{p(\nu)}{q(\nu)} d\nu$, see Kullback and Leibler (1951), as follows.

Definition 2 (Distinguishability)**.** *Given a sliding window model* (2)*, distinguishability $\mathcal{D}_{i,j}(\theta)$ of a fault f_i with a given fault time profile θ from a fault mode f_j is defined as*

$$\mathcal{D}_{i,j}(\theta) = \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p_\theta^i \| p^j\right) \tag{6}$$

where the set $\mathcal{Z}_{f_j}$ is defined in Definition 1. ■

The definition of distinguishability fulfills the following two propositions.

Proposition 1. *Given a window model* (2)*, a fault $f_i = \theta \in \Theta_i$ is isolable from a fault mode f_j if and only if*

$$\mathcal{D}_{i,j}(\theta) > 0 \tag{7}$$

■

Proposition 2. *If $\bar{0}$ is a boundary point of Θ_j, where $\Theta_j = \mathbb{R}^n \setminus \{\bar{0}\}$, for a fault mode f_j then*

$$\mathcal{D}_{i,j}(\theta) \leq \mathcal{D}_{i,NF}(\theta). \tag{8}$$

■

It is assumed, without loss of generality, that

$$\text{cov}\left(\mathcal{N}_H N e\right) = I \tag{9}$$

where I is the identity matrix, see Eriksson et al. (2011a). Since the noise in (2) is Gaussian distributed, distinguishability can be computed explicitly given the following theorem.

Theorem 1. *Distinguishability for a sliding window model* (2) *with Gaussian distributed stochastic vector e, under assumption* (9)*, is given by*

$$\mathcal{D}_{i,j}(\theta) = \frac{1}{2} \|\mathcal{N}_{(H\, F_j)} F_i \theta\|^2 \tag{10}$$

where the rows of $\mathcal{N}_{(H\, F_j)}$ is an orthonormal basis for the left null space of (H, F_j). ■

3.2 Relation of residual generators

To design a residual generator r isolating faults from fault mode f_j, multiply (2) from the left with $\gamma\mathcal{N}_{(H\,F_j)}$ where γ is a row-vector to obtain

$$r = \gamma\mathcal{N}_{(H\,F_j)}Lz = \gamma\mathcal{N}_{(H\,F_j)}Ff + \gamma\mathcal{N}_{(H\,F_j)}Ne. \tag{11}$$

The residual generator (11) is a scalar model in the same form as (2). Therefore, distinguishability can be computed for the residual generator, which is denoted $\mathcal{D}^{\gamma}_{i,j}(\theta)$, where the superscript γ is used to distinguish from when computing distinguishability for the model. The connection between distinguishability and FNR of r is given by the following theorem, which also gives an alternative way of computing distinguishability for a scalar model.

Theorem 2. *A residual generator* (11)*, for a model* (2) *where* e *is Gaussian distributed under assumption* (3)*, is also Gaussian distributed* $\mathcal{N}(\lambda(\theta), \sigma^2)$ *and*

$$\mathcal{D}^{\gamma}_{i,j}(\theta) = \frac{1}{2}\left(\frac{\lambda(\theta)}{\sigma}\right)^2$$

where θ *is the fault time profile of a fault* f_i*, and* $\lambda(\theta)/\sigma$ *is the fault to noise ratio with respect to fault* f_i *in* (11)*.*

An important connection between $\mathcal{D}^{\gamma}_{i,j}(\theta)$ and $\mathcal{D}_{i,j}(\theta)$ is given by the inequality described in the following theorem.

Theorem 3. *For a model* (2) *under assumption* (9)*, an upper bound for* $\mathcal{D}^{\gamma}_{i,j}(\theta)$ *in* (11) *is given by*

$$\mathcal{D}^{\gamma}_{i,j}(\theta) \leq \mathcal{D}_{i,j}(\theta)$$

with equality if and only if γ *and* $\mathcal{N}_{(H\,F_j)}F_i\theta$ *are parallel.* ■

Theorem 3 shows that $\mathcal{D}_{i,j}(\theta)$ gives an upper limit of the FNR which can be achieved by any residual generator (11). Note that if a fault $f_i = \theta$ is isolable from a fault mode f_j, then the theorem shows how to design a residual generator with maximum FNR by choosing $\gamma = (\mathcal{N}_{(H\,F_j)}F_i\theta)^T$. This will be used when automatically generating new residual generators by the sequential test selection algorithm described in the following section.

4 Generalization of distinguishability

The definition of distinguishability in Section 3 only considers fault modes containing single faults. Here, it is also interesting to quantify how easy it is to isolate a fault f_a from several faults d_j. This corresponds to quantifying how easy it is to reject a minimal candidate d_j for a new minimal candidate $d_i = \{f_a\} \cup d_j$. In order to handle fault modes with multiple faults a generalization of the previous definition of distinguishability is presented here. Most results presented here can be derived in a similar way as for the single fault case in Section 3.

First, a generalization of the sets $\mathcal{Z}_{f_j}$ in Definition 1 is defined by considering combinations of different fault time profiles for each fault in the fault mode. Consider all faults present given a fault mode d_j and let $\bar{\theta} \in \Theta_{d_j}$ denote a matrix $\bar{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ where each column corresponds to a fault time profile $f = \theta$ for each fault $f \in d_j$. Then, a generalization of Definition 1 can be made as follows.

Definition 3. *Let $\mathcal{Z}_{d_j}$ denote the set of all pdf's, $p(\tau, \mu)$, for all combination of fault time profiles $\bar{\theta} \in \Theta_{d_j}$, describing τ which could be explained by the fault mode d_j,*

$$\mathcal{Z}_{d_i} = \left\{ p(\tau, \mathcal{N}_H F_{d_j} \bar{\theta}) | \forall \bar{\theta} \in \Theta_{d_j} \right\} \tag{12}$$

where F_{d_j} contains the rows of F corresponding to the faults $f \in d_j$. ■

Recall that $p_\theta^a = p(\tau, \mathcal{N}_H F_a \theta)$. Since isolating a fault f_a from all faults d_j corresponds to rejecting d_j for $d_i = \{f_a\} \cup d_j$, a generalization of Definition 2 can be formulated as follows.

Definition 4 (Distinguishability of multiple faults)**.** *Given a sliding window model* (2)*, distinguishability $\mathcal{D}_{d_i,d_j}(\theta)$ of a fault $f_a \notin d_j$ where $d_i = \{f_a\} \cup d_j$ with a fault time profile θ from a fault mode d_j with multiple faults is defined as*

$$\mathcal{D}_{d_i,d_j}(\theta) = \min_{p^j \in \mathcal{Z}_{d_j}} K\left(p_\theta^a \| p^j\right) \tag{13}$$

where the set $\mathcal{Z}_{d_j}$ is defined in Definition 3. ■

In the Gaussian case it follows from Theorem 1 that (13) can be computed explicitly using

$$\mathcal{D}_{d_i,d_j}(\theta) = \frac{1}{2} \| \mathcal{N}_{(H\, F_{d_j})} F_a \theta \|^2 \tag{14}$$

where F_{d_j} are the rows of F corresponding to the faults in d_j.

Proposition 1 holds given the new definition of distinguishability and a generalization of Proposition 2 can be formulated as follows.

Proposition 3. *Let $d_k \subset d_j$ and $f_a \notin d_k, d_j$. Then let $d_l = \{f_a\} \cup d_k$ and $d_i = \{f_a\} \cup d_j$, then the following inequality holds.*

$$\mathcal{D}_{d_i,d_j}(\theta) \leq \mathcal{D}_{d_l,d_k}(\theta). \tag{15}$$

■

The proposition states that it performance always decreases when isolating a specific fault from an increasing set of faults. If we want to generate a residual generator which is sensitive to few faults then a residual generator with maximum FNR can never have higher FNR than if we would create an residual with maximum FNR which is sensitive to a super-set of faults.

To design a residual generator which is sensitive to a fault f_a but no fault $f \in d_j$ multiply (2) from the left with $\gamma\mathcal{N}_{(H\,F_{d_j})}$ to obtain

$$\gamma\mathcal{N}_{(H\,F_{d_j})}Lz = \gamma\mathcal{N}_{(H\,F_{d_j})}Ff + \gamma\mathcal{N}_{(H\,F_{d_j})}Ne \tag{16}$$

Theorem 2 and Theorem 3 can be generalized to the case of multiple faults. As follows from Theorem 3, a residual generator with maximum FNR isolating a fault $f_a = \theta$ from the faults d_j is designed by choosing $\gamma = (\mathcal{N}_{(H\,F_{d_j})}F_a\theta)^T$.

5 Sequential test selection

Here, an algorithm for sequential test selection is presented. The algorithm updates the set of active test quantities in the diagnosis algorithm given the present minimal candidates. The algorithm applies the results in Section 4 to generate residual generators and to quantify their diagnosability performance. Finally, the sequential test selection algorithm is evaluated using an academic example.

5.1 Principles

A diagnosis algorithm computes minimal candidates consistent with the alarmed test quantities, see Figure 1. New test quantities are selected which are able to reject any present minimal candidates and fulfills a required minimum distinguishabiliy. Each time one or more test quantities alarms, D is updated and a new set of test quantities is selected.

The fault isolation algorithm used here is described in de Kleer and Williams (1987) and updates the set of minimal candidates each time new test quantities alarms. A rejected minimal candidate d_j is replaced by new minimal candidates d_i such that $d_i = \{f_a\} \cup d_j$ where $f_a \notin d_j$. If any of the new minimal candidates is a super-set of any other minimal candidate then the new candidate is removed. Note that it is assumed here, that the available test quantities are designed such that the risk of false alarms is minimized.

To be able to reject a minimal candidate d_j for another candidate $d_i = \{f_a\} \cup d_j$ requires a test quantity which is sensitive to f_a but not any $f \in d_j$, see de Kleer and Williams (1987). The following example shows how a lattice set representation of all possible candidates can be used to visualize what kind of isolablity performance that is required by the active set of test quantities.

Example 1. *Consider a system with three possible faults f_1,f_2, and f_3. All possible candidates can be described using a lattice set, see Figure 3. Each node is a candidate where all nodes at each level have the same cardinality. Each node have arrows going to all nodes at the next level which are super-sets of the first node. For example from node $\{f_1\}$ goes arrows to $\{f_1, f_2\}$ and $\{f_1, f_3\}$. The bottom node represents the fault-free case.*

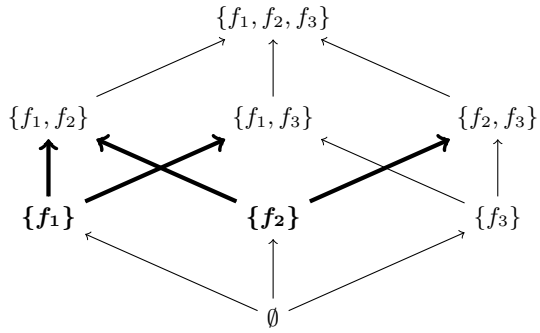

Figure 3: A lattice set representation of all possible candidates for a system with three possible faults where $\emptyset$ corresponds to the fault-free case. If $D = \{\{f_1\}, \{f_2\}\}$ then tests, able to reject $\{f_1\}$ for $\{f_1, f_2\}$ and $\{f_1, f_3\}$ and able to reject $\{f_2\}$ for $\{f_1, f_2\}$ and $\{f_2, f_3\}$, are needed to improve D.

If the present minimal candidates are $\{f_1\}$ and $\{f_2\}$, then tests need to be activated with the capability of rejecting $\{f_1\}$ for each of the candidates $\{f_1, f_2\}$ and $\{f_1, f_3\}$, and rejecting $\{f_2\}$ for each of the candidates $\{f_1, f_2\}$ and $\{f_2, f_3\}$ to be able to refine the candidates. This is visualized in Figure 3 by thicker arrows going from the present minimal candidates to new possible minimal candidates if a new test quantity alarms. ◇

Sequentially updating the minimal candidates in the fault isolation algorithm requires an active set of test quantities able to reject each $d_j \in D$ for each $d_i = \{f_a\} \cup d_j$ where $f_a \notin d_j$. By using the results in Section 3, the set of active test quantities can be selected and generated for each minimal candidate $d_j \in D$ if there exists a test quantity in the set which have maximum distinguishability, or fulfills a minimum required distinguishability, when rejecting that minimal candidate d_j.

5.2 Algorithm

In the previous subsection a discussion was made regarding which types of test quantities to use depending on the present set of minimal candidates D. How well the present faults can be identified, depends on which test quantities $T \in \mathcal{T}$ that are active in the diagnosis algorithm.

Let $\mathcal{E}_{i,j}$ contain each test quantity $T \in \mathcal{T}$ which can reject d_j for d_i, where $d_i = \{f_a\} \cup d_j$ and $f_a \notin d_j$, and fulfills a minimum required distinguishability. If there exists at least one test which is able to reject a minimal candidate d_j for a new minimal candidate d_i with at least minimum required distinguishability then $\mathcal{E}_{i,j} \neq \emptyset$. Otherwise if $\mathcal{E}_{i,j} = \emptyset$, then d_j can not be rejected for d_i.

The sequential test selection algorithm is summarized in the following steps.

I Initialize the set of minimal candidates, $D = \emptyset$.

II Solve (1) given D to compute a new set of test quantities.

III Replace the previous set of active test quantities by the new set of test quantities.

IV Run the new set of test quantities until a test alarms.

V Update D given the alarmed test quantities.

VI Return to step II.

The optimization problem (1) is a minimum hitting set problem which is NP-hard, see Moret and Shapiro (1985), but there exists many heuristic search methods to find a solution, see for example de Kleer (2011) and references. The following example considers a small system and is used to describe how the sequential test selection algorithm updates the set of active test quantities as the minimal candidates are updated.

Example 2. *Consider the system with three possible faults. Here, a diagnosis algorithm applying the sequential test selection algorithm is designed to detect and isolate faults while minimizing the number of active tests, i.e.,* $c(T) = 1$ *in* (1)*,* $\forall T \in \mathcal{T}$*. Assume that there are seven available test quantities,* $\mathcal{T} = \{T_1, T_2, \ldots, T_7\}$ *where the detectability performance of each test quantity and fault is quantified using distinguishability, see Table 1. First in the considered fault scenario, a fault* f_1 *enters the system and then later another fault* f_3 *also enters the system.*

When choosing each set $\mathcal{E}_{d_i,d_j}$*, the performance of each test quantity is chosen to fulfill a required minimum distinguishability. In this example, we choose that if the cardinality of* d_i *is 1 then distinguishability should be at least* 2.0*, if the cardinality of* d_i *is 2 then distinguishability should at least be* 1.0*, and if the*

Table 1: The performance of each test is quantified using distinguishability. Distinguishability of a test T_i detecting a fault $f_j = \theta$ is written in position (i, j).

	f_1	f_2	f_3
T_1	2.5	2.4	2.0
T_2	2.2	1.5	2.3
T_3	2.0	0.8	0
T_4	0	1.3	1.5
T_5	0	0	1.2
T_6	0.5	0	0
T_7	0	0.6	0

cardinality of d_i is 3 then distinguishability should at least be 0.5. *Based on the distinguishability of the test quantities given in Table 1, the sets will be*

$$\begin{aligned}
&\mathcal{E}_{\{f_1\},\emptyset} = \{T_1, T_2, T_3\}, && \mathcal{E}_{\{f_2\},\emptyset} = \{T_1\},\\
&\mathcal{E}_{\{f_3\},\emptyset} = \{T_1, T_2\}, && \mathcal{E}_{\{f_1,f_2\},\{f_1\}} = \{T_4\},\\
&\mathcal{E}_{\{f_1,f_3\},\{f_1\}} = \{T_4\}, && \mathcal{E}_{\{f_1,f_2\},\{f_2\}} = \emptyset,\\
&\mathcal{E}_{\{f_2,f_3\},\{f_2\}} = \{T_5\}, && \mathcal{E}_{\{f_1,f_3\},\{f_3\}} = \{T_3\},\\
&\mathcal{E}_{\{f_2,f_3\},\{f_3\}} = \emptyset, && \mathcal{E}_{\{f_1,f_2,f_3\},\{f_1,f_2\}} = \{T_5\},\\
&\mathcal{E}_{\{f_1,f_2,f_3\},\{f_1,f_3\}} = \{T_7\}, && \mathcal{E}_{\{f_1,f_2,f_3\},\{f_2,f_3\}} = \{T_6\}.
\end{aligned} \tag{17}$$

Note that some sets are empty, for example $\mathcal{E}_{\{f_1,f_2\},\{f_2\}} = \emptyset$ because there are no test quantities which are sensitive to f_1 but not to f_2 with sufficiently high *distinguishability.*

When the diagnosis algorithm is initialized the only minimal candidate is the empty set, i.e., $D = \{d_0\}$ where $d_0 = \emptyset$. The first set of active test quantities is the solution to the problem

$$\begin{aligned}
&\min \sum_{i=1}^{7} x_i\\
&s.t.\ x_1 + x_2 + x_3 \geq 1, && (\mathcal{E}_{\{f_1\},\emptyset})\\
&\quad x_1 \geq 1, && (\mathcal{E}_{\{f_2\},\emptyset})\\
&\quad x_1 + x_2 \geq 1, && (\mathcal{E}_{\{f_3\},\emptyset})\\
&\quad x_i = \{0, 1\}, && i = 1, 2, \ldots, 7.
\end{aligned}$$

The optimal solution is $x_1 = 1$, and $x_i = 0, \forall i \neq 1$, which means that only T_1 is activated. Then a fault f_1 enters the system and T_1 alarms. The new minimal candidates are $D = \{d_1, d_2, d_3\}$ where $d_1 = \{f_1\}$, $d_2 = \{f_2\}$, and $d_3 = \{f_3\}$. To update the set of test quantities, the following problem is solved

$$\begin{aligned}
&\min \sum_{i=1}^{7} x_i\\
&s.t.\ x_4 \geq 1, && (\mathcal{E}_{\{f_1,f_2\},\{f_1\}})\\
&\quad x_4 \geq 1, && (\mathcal{E}_{\{f_1,f_3\},\{f_1\}})\\
&\quad x_3 \geq 1, && (\mathcal{E}_{\{f_1,f_3\},\{f_3\}})\\
&\quad x_5 \geq 1, && (\mathcal{E}_{\{f_2,f_3\},\{f_2\}})\\
&\quad x_i = \{0, 1\}, && i = 1, 2, \ldots, 7.
\end{aligned} \tag{18}$$

Note that $\mathcal{E}_{\{f_1,f_2\},\{f_2\}}$ and $\mathcal{E}_{\{f_2,f_3\},\{f_3\}}$ are empty sets and therefore not included in (18). *The new set of active test quantities are T_3, T_4, and T_5. Later T_3 alarms and the remaining minimal candidates are $D = \{d_1, d_2\}$. The corresponding new set of active test quantities is T_4 and T_5. Since only f_1 is present in the system and there is no test quantity which have sufficient performance to reject*

the other candidate $\{f_2\}$ *for* $\{f_1, f_2\}$, $\{f_1\}$ *can not be isolated from* $\{f_2\}$. *Note that by using another selection criteria when defining each set* $\mathcal{E}_{d_i,d_j}$ *could let* $\mathcal{E}_{\{f_1,f_2\},\{f_2\}}$ *contain* T_6 *which is sensitive to* f_1 *but not* f_2.

Later a fault f_3 *also enters the system which first results in that* T_4 *alarms. The new minimal candidates becomes* $D = \{d_2, d_4\}$ *where* $d_4 = \{f_1, f_3\}$. *The corresponding new activated test quantities are* T_5 *and* T_7. *Later* T_5 *alarms which replaces* d_2 *with* $d_5 = \{f_2, f_3\}$ *and activates the test quantities* T_6 *and* T_7. *Finally* T_6 *alarms which removes* d_5 *since the new candidate* $\{f_1, f_2, f_3\}$ *is a super-set of* d_4 *which is the true and remaining candidate.* ◇

The properties of each set in (17) can be visualized using the lattice in Figure 3. If $\mathcal{E}_{d_i,d_j} = \emptyset$ then there exists no test quantity which have sufficiently high distinguishability to reject d_j for d_i, see Table 1. Each set $\mathcal{E}_{d_i,d_j}$ is related to an arrow from d_j to d_i, see Figure 4. All non-empty sets $\mathcal{E}_{d_i,d_j}$ are represented by thicker arrows.

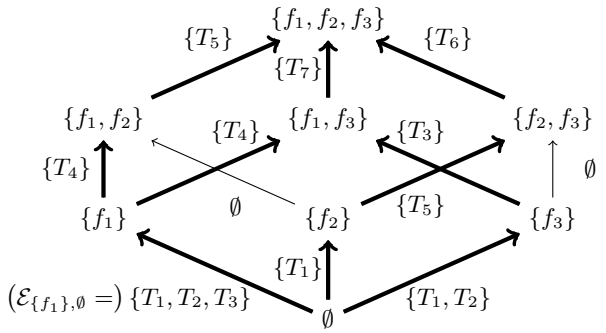

Figure 4: A lattice set representation of all possible candidates for a system with three possible faults. The set at each arrow represents which test quantities that are able to reject d_j for d_i and fulfills a specified distinguishability criteria.

6 Case study: DC circuit

The sequential test selection algorithm presented in Section 5 is applied to a DC-circuit. The DC-circuit is a part of the diagnosis test bed ADAPT, see Poll et al. (2007). The analyzed system is similar to the circuit analyzed in Gorinevsky et al. (2009) except that some additional faults are included here while outputs and faults in batteries and loads are unknown and not considered.

6.1 System

The system contains 19 sensors measuring voltages and currents in the circuit and 12 sensors measuring the positions of the relays. The number of faults

included in the model is 38, both relay faults and sensor faults. Not all faults are detectable because there are not enough sensors and some parts of the circuit are assumed unknown, e.g., the power consumption of the loads and the output from the batteries. Some faults are not fully isolable from each other because several faults affect the same part of the system.

A model of the DC-circuit has been developed by formulating the Sparse Tableau Analysis (STA) equations, see Gorinevsky et al. (2009). Model uncertainties have been included as i.i.d. Gaussian measurement noise and the faults are modeled as additive signals. The model of the DC-circuit is linear and static and is written in the form (2) where $n = 1$.

6.2 Diagnosis algorithm

The implemented diagnosis algorithm applies the result from Theorem 3, i.e., how to generate residuals with maximum FNR, to automatically design and generate residuals during run-time. One problem when sequentially updating the set of active test quantities is how to find residuals which have sufficiently high distinguishability given the present minimal candidates. When choosing which residual to generate the inequality (15) is used. Based on the inequality, residuals should be generated where as few faults are decoupled as possible while still being able to reject d_j for d_i to maximize distinguishability. Residuals with maximum distinguishability can be generated, using the result in Section 4, by considering each of the present minimal candidates. As the minimal candidates are updated, new tests with maximum distinguishability are automatically generated.

The distinguishability criteria when selecting the sets $\mathcal{E}_{d_i,d_j}$ are such that each set $\mathcal{E}_{d_i,d_j}$ contains the residual with highest distinguishability when rejecting d_j for d_i. Also, if there are more residuals able to reject d_j for d_i, they are included in $\mathcal{E}_{d_i,d_j}$ if they fulfill a minimum required distinguishability.

In this implementation, when a new set of active residuals is selected the residuals are activated when the next sample of data is received from the system. This implementation has been chosen to visualize how the number of active residuals changes each time one or more residuals alarms. Note that this way of implementation results in a delay in the fault isolation process because a fault could be isolated faster if new residuals are used directly on previous data samples, see Krysander et al. (2010). Such a change in the implementation can easily be made and do not affect the sequential test selection algorithm.

Here, the optimization problem (1) is solved by using a simple greedy test selection algorithm to find a solution. The greedy search algorithm iteratively picks the residual with maximum distinguishability in each set $\mathcal{E}_{d_i,d_j}$ if no other residual in the set is already selected.

6.3 Evaluation

The sequential test selection algorithm described in section 5.2 is evaluated using a number of different test scenarios. In the first scenario, a relay breaks, f_{10},

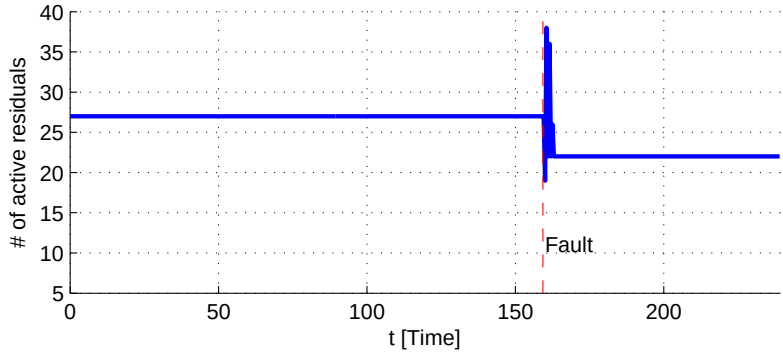

Figure 5: The number of active residuals during a scenario where a relay gets stuck open, f_{10}, occurs at 159 s. The total number of active residuals during the scenario is 64.

at 159 s. The number of active residuals at any time during the scenario is shown in Figure 5. In the first evaluation, residuals are included in $\mathcal{E}_{d_i,d_j}$ if their corresponding distinguishability is higher than 0.5, which corresponds to $FNR = 1$ if the fault amplitude is 1. Note that if only the residuals which have maximum distinguishability in each set $\mathcal{E}_{d_i,d_j}$ is active from the start the number of used residuals would exceed 370.

The total number of used residuals is 64 but the maximum number of active residuals is mostly below 30, except right after the fault is detected where the number of active residuals rises to almost 40.

If the required distinguishability is chosen lower, for example 0.045 ($FNR = 0.3$), the number of active residuals used during the scenario is reduced which is shown in Figure 6. The number of active residuals are less than 15 most of the time and less than 25 when the fault is detected. The total number of generated residuals during the whole scenario is 41.

The final minimal candidates for the two choices of minimum distinguishability are in this case equal and are shown in Figure 7. Each row represent a minimal candidate and the only candidate of minimum cardinality, $\{f_{10}\}$, is also the correct isolated fault.

In the second scenario two sensor faults, one voltage sensor f_{35} and one current sensor f_{24}, occurs simultaneously. The diagnosis algorithm have been run twice, first with minimum required distinguishability chosen to be 0.5 and the second time to be 0.045. The number of active residuals during the first run is shown in Figure 8. The total number of generated residuals is 54 and the number of active residuals is below 30 except right after the faults are detected.

In the second run the number of active residuals is reduced to 12 during the whole scenario except right after the faults are detected. The total number of generated residuals is 40.

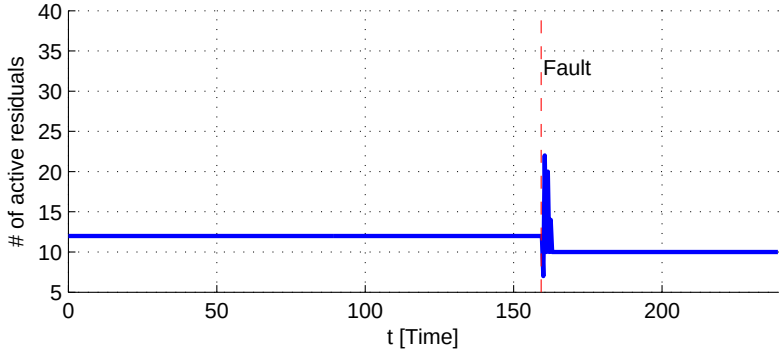

Figure 6: The number of active residuals during a scenario where a relay gets stuck open, f_{10}, occurs at 159 s. The required minimum distinguishability for each residual is lower compared to the result in Figure 5 resulting in fewer active residuals during the scenario. The total number of active residuals during the scenario is 41.

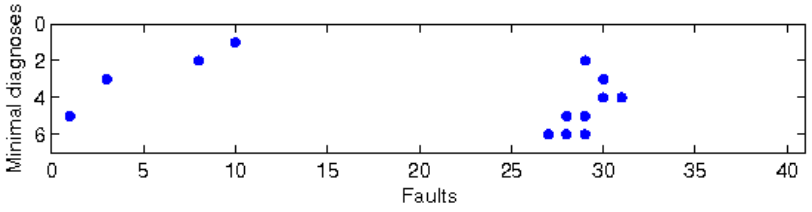

Figure 7: The final minimal candidates for the first scenario given by the diagnosis algorithm, sorted by cardinality. The final candidates are in this case the same for both choices of minimum required distinguishability.

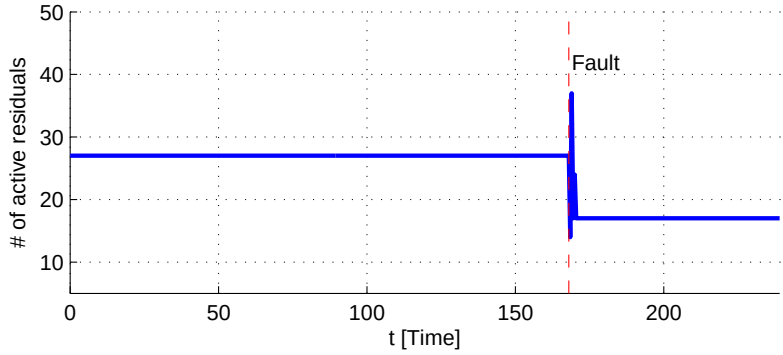

Figure 8: The number of active residuals during a scenario where two sensors fail, f_{24} and f_{35}, occurs at 168 s. The total number of active residuals during the scenario is 54.

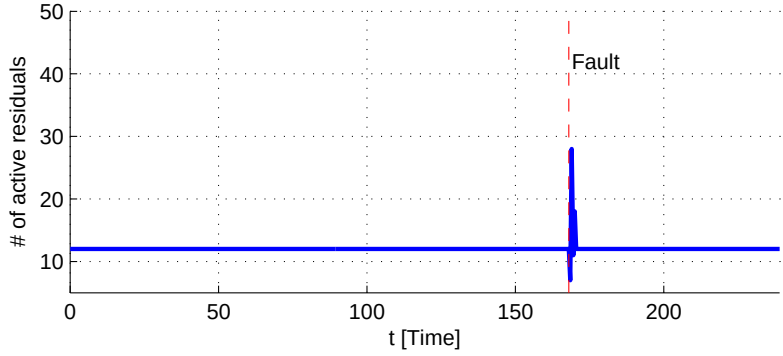

Figure 9: The number of active residuals during a scenario where two sensors fail, f_{24} and f_{35}, occurs at 168 s. The required minimum distinguishability for each residual is lower compared to the result in Figure 8 resulting in fewer active residuals during the scenario. The total number of active residuals during the scenario is 40.

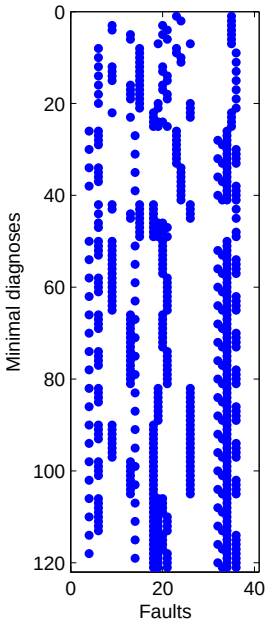

Figure 10: The final minimal candidates given by the diagnosis algorithm, sorted by cardinality. In this case, both choices of required minimum distinguishability resulted in the same final candidates.

The total number of minimal candidates is 121, see Figure 10. The number of candidates with minimum cardinality equal to two are $\{f_{24}, f_{35}\}$ and $\{f_{23}, f_{35}\}$ where $\{f_{24}, f_{35}\}$ is the true candidate. The other candidate, $\{f_{23}, f_{35}\}$, can not be rejected because f_{23} and f_{24} are not isolable from each other.

Choosing a lower required distinguishability when creating the sets $\mathcal{E}_{d_i,d_j}$ can result in solutions where the set of active test quantities has lower cardinality. A lower number of active test quantities is a trade off which will probably result in reduced diagnosability performance since the diagnosis algorithm will accept active test quantities with lower distinguishability when rejecting some minimal candidates. In this analysis, the actual faults in the measured data are relatively large compared to the fault amplitudes used when analyzing distinguishability. This could explain why the minimal candidates are equal in the two scenarios even though the required distinguishability is changed.

7 Tuning the test selection algorithm

Here, some extra tuning steps which could be added to the test selection algorithm are discussed. The proposed algorithm is also related to test selection algorithms presented in other works. It is shown how the problem formulation in this work can be used to describe other test selection algorithms.

7.1 Off-line

If the available set of test quantities $\mathcal{T}$ is large and many of the sets $\mathcal{E}_{d_i,d_j}$ have high cardinality then it will be computationally expensive to find an optimal solution to (1). There might be several test quantities that will never be used because there are other test quantities that have better performance. This might be the case if tests quantities are generated automatically off-line or sequentially during run-time or if a developer have not been able to remove not so good redundant test quantities.

If the set of available test quantities $\mathcal{T}$ is generated off-line, the computational time for solving (1) on-line, and the memory needed for storing all test quantities, could be reduced by first solving (1) for all sets of possible minimal candidates off-line, i.e., $D = \{\text{all possible minimal candidates}\}$. By using only the set of test quantities in the solution as available test quantities $\mathcal{T}$ will remove unnecessary test quantities when the diagnosis algorithm is on-line.

Another approach to reduce the computational time for solving (1) could be to neglect minimal candidates of higher cardinality. If multiple faults of cardinality higher than $q > 0$ are assumed to be unlikely then only test quantities able to reject minimal candidates of lower cardinality need to be considered. In Svärd (2012), an algorithm for automatically generating residuals off-line for a non-linear system uses a greedy search algorithm to find a set of test quantities able to detect and isolate single faults. This corresponds to solving (1) for all minimal candidates d_j of maximum cardinality one. In Svärd (2012), only deterministic performance of test quantities are considered which is the same as letting $\mathcal{E}_{d_i,d_j}$ contain all test quantities able to reject d_j for d_i independent of quantitative performance.

7.2 On-line

In Krysander et al. (2010), the proposed sequential test selection algorithm solves the optimization problem (1) every time the minimal candidates are updated, but do not consider any quantitative diagnosability performance which is done in this work. Although, in Krysander et al. (2010), an extra constraint to the solution of (1) requires that any combination of faults should always be detectable if possible. For example, there could be situations where two faults could cancel out each other which would in such a case not be detectable. An example is a residual $r = f_1 + f_2$ which would be zero if $f_1 = 1, f_2 = -1$. Such

cases could be assumed highly unlikely, and therefore neglected, but this extra constraint can also easily be included in the algorithm proposed in this paper.

One approach to save computational time is to let each set $\mathcal{E}_{d_i,d_j}$ only contain the best test quantity, i.e., the test quantity with highest distinguishability. Then, all sets $\mathcal{E}_{d_i,d_j}$ would either contain one element or no elements depending on whether there exists a test quantity able to reject d_j for d_i or not. Then the optimal solution to (1) is found by simply selecting the test quantity in each set $\mathcal{E}_{d_i,d_j}$ where $d_j \in D$. This assures that the diagnosis algorithm always uses the best set of tests given D at the cost of requiring more active tests used at the same time. Note that in such a case where each set $\mathcal{E}_{d_i,d_j}$ maximally contains one element, it is trivial to find the optimal solution.

7.3 Other measures of diagnosability performance

Here, distinguishability has been used to quantify diagnosability performance when selecting a set of test quantities with sufficient detectability and isolability performance. The proposed test selection algorithm is not limited to the use of distinguishability but could use any type of measure to quantified diagnosability performance to compare the performance of different sets of test quantities, see for example Wheeler (2011).

8 Conclusion

A sequential test selection algorithm to update an active set of test quantities given the present minimal candidates is proposed. A contribution here is the use of a measure for quantitative diagnosabiltiy performance, distinguishability, to choose which test quantities to be used to reject the present minimal candidates. A generalization of distinguishability is introduced in order to handle isolability from multiple faults.

The test selection algorithm proves useful when, for example, implementing a diagnosis algorithm where the set of tests are automatically selected and generated on-line. As an example, a diagnosis algorithm for a DC-circuit is evaluated which automatically generates new residuals with maximum distinguishability given the present minimal candidates. The example shows that the number of test quantities can be greatly reduced by using the test selection algorithm while fulfilling a required quantitative detectability and isolability performance.

The proposed sequential test selection algorithm is also compared and related to other works, including both off-line and on-line test selection algorithms. Different tuning parameters can be used to reduce computational time or change the properties of the solution. The proposed algorithm can be used for both on-line and off-line test selection algorithms.

9 Acknowledgment

This work was partially supported by the Swedish Research Council within the Linnaeus Center CADICS.

References

Johan de Kleer. Hitting set algorithms for model-based diagnosis. 22nd International Workshop on Principles of Diagnosis (DX-11), Murnau, Germany, 2011.

Johan de Kleer and Brian C. Williams. Diagnosing Multiple Faults. *Artif. Intell.*, 32(1):97–130, 1987.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative Stochastic Fault Diagnosability Analysis. In *50th IEEE Conference on Decision and Control*, USA, 2011a.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative Fault Diagnosability Performance of Linear Dynamic Descriptor Models. 22nd International Workshop on Principles of Diagnosis (DX-11), Murnau, Germany, 2011b.

Erik Frisk, Mattias Krysander, and Jan Åslund. Sensor Placement for Fault Isolation in Linear Differential-Algebraic Systems. *Automatica*, 45(2):364–371, 2009.

Dimitry Gorinevsky, Stephen Boyd, and Scott Poll. Estimation of Faults in DC Electrical Power System. In *Proceedings IEEE Conference on Decision and Control*, December 2009.

Mattias Krysander, Fredrik Heintz, Jacob Roll, and Erik Frisk. FlexDx: A reconfigurable diagnosis framework. *Engineering Applications of Artificial Intelligence*, 23(8):1303–1313, October 2010.

S. Kullback and R. A. Leibler. On Information and Sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

B. Moret and H. Shapiro. On Minimizing a Set of Tests. *SIAM Journal on Scientific and Statistical Computing*, 6(4):983–1003, 1985.

J. W. Polderman and J. C. Willems. *Introduction to Mathematical Systems Theory: A Behavioral Approach*, volume 26 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1998.

Scott Poll, Ann Patterson-hine, Joe Camisa, David Garcia, David Hall, Charles Lee, Ole J. Mengshoel, David Nishikawa, John Ossenfort, Adam Sweet, Serge Yentus, Indranil Roychoudhury, Matthew Daigle, Gautam Biswas, and Xenofon Koutsoukos. Advanced diagnostics and prognostics testbed. In *in Proceedings of the 18th International Workshop on Principles of Diagnosis*, pages 178–185, 2007.

Carl Svärd. *Methods for Automated Design of Fault Detection and Isolation Systems with Automotive Applications*. PhD thesis, Linköping University, 2012.

Timothy J. Wheeler. *Probabilistic Performance Analysis of Fault Diagnosis Schemes*. PhD thesis, University of California, Berkeley, 2011.

Paper D

Flywheel angular velocity model for misfire simulation*

D

*Submitted to *7th IFAC Symposium on Advances in Automotive Control.*

Flywheel angular velocity model for misfire simulation

Daniel Eriksson, Lars Eriksson, Erik Frisk, and Mattias Krysander

Vehicular Systems, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.

Abstract

A flywheel angular velocity model for misfire and disturbance simulation is presented. Applications of the model are, for example, initial parameter calibration or robustness analysis of misfire detection algorithms. An analytical model of cylinder pressure is used to model cylinder torque and a multi-body model is used to model crankshaft and driveline oscillations. Different types of disturbances, such as cylinder variations, changes in auxiliary load, and flywheel manufacturing errors can be injected in the model. A qualitative validation of the model shows that simulated angular velocity captures the amplitude and oscillatory behavior of real measurements and the effects of different types of disturbances, e.g. misfire and flywheel manufacturing errors.

1 Introduction

Engine misfire detection is an important part of the OBDII legislations that reduce exhaust emissions and avoid damage to the catalytic converters. Misfire detection based on angular velocity measured at the flywheel has been studied in several papers, e.g, Connolly and Rizzoni (1994), Kiencke (1999), and Tinaut et al. (2007). An overview of misfire detection research is found in Mohammadpour et al. (2012). Detecting misfire is a non-trivial problem which is complicated by, for example, changes in load and speed and flywheel manufacturing errors, see Naik (2004).

Development and validation of a misfire detection algorithm can require lots of resources using test rigs and real cars which is expensive and time consuming. A misfire simulation model is beneficial for reducing development costs by, for example, automating the initial calibration of the parameters of the misfire detection algorithm. One example is to investigate which teeth of the flywheel to measure the time difference between to best capture a misfire event while reducing the number of measurements during a revolution.

Another application is to make a quantitative analysis of how different sizes of disturbances affect the observations. This can be used for example for robustness analysis of a misfire detection algorithm to analyze how large disturbances it can handle.

In Minelli et al. (2004), a model for simulating misfire is proposed which considers the effects of misfire and the subsequent oscillations in the angular velocity signal. A contribution in this work with respect to the previous mentioned paper is that other types of disturbances, such as flywheel manufacturing errors and changes to auxiliary loads, can be simulated and the cylinder pressure is computed using an analytical model to make it possible to model cylinder variations.

In Schagerberg (2003) a model to estimate cylinder pressure using torque sensors is developed. A multi-body model of the crankshaft is used to model torsional vibrations in the crankshaft. In contrast to Schagerberg (2003) the focus in this work is the use of angular velocity measurements instead of measuring torque.

Another similar application of driveline modeling is torsional vibration analysis, e.g., Rabeih (1997), Nickmehr et al. (2012), and Crowther and Zhang (2005). In contrast to these works, a contribution here is the use of the cylinder pressure model in Eriksson and Andersson (2002). A further contribution is the addition of capabilities for simulating cycle to cycle variations in the cylinder pressure. This is used, e.g., to simulate the effects of misfire to the angular velocity measurements at the flywheel.

Here, a multi-body model similar to the model in Schagerberg (2003) together with the cylinder pressure model in Eriksson and Andersson (2002) is used to model crankshaft oscillations. Also, a driveline model, similar to the model in Nickmehr et al. (2012), is used to model the torsional vibration modes of the driveline. Experiments have been carried on parameter tuning of misfire detection

algorithms but the focus here is on the modeling work. The contribution in this work is a model to simulate angular velocity measurements at the flywheel when different types of disturbances are injected in the model such as misfire and change in auxiliary loads. The model is designed using a modular structure to easily extend the model depending on the vehicle configuration, e.g., the number of cylinders.

2 Model requirements

A common approach to detect misfire is to use a test quantity based on the crankshaft angular velocity measurements at the flywheel. To distinguish changes in the measurements caused by misfire from disturbances in the engine and driveline is a non-trivial problem, mainly due to complicating factors such as changes in load and speed, cold starts, engines with a large number of cylinders, and the resolution of the angular velocity measurements.

The purpose of the model, developed in this work, is to simulate flywheel angular velocity when misfire and disturbances are injected in the model. A list of implemented disturbances that can be injected in the model is

- Combustion variations:
  - Misfire
  - Cold starts
  - Cylinder variations
- Crankshaft torsional vibrations.
- Auxiliary load variations, e.g. turning on and off air conditioning.
- Disturbances in road load torques, e.g. crossing a railroad.
- Flywheel resolution and measurement errors.

The model can be used to analyze how different types of disturbances complicates misfire detection and to evaluate and optimize misfire detection algorithms by using data from different simulated scenarios.

For model analysis and validation, high resolution data from a vehicle with a four cylinder engine is used. Low resolution angular velocity measurements from a five cylinder vehicle is also used for validation. To handle different types of vehicle configurations, the model is designed using an extensible block structure describing different parts of the system to easily modify the model depending on the vehicle configuration.

3 Model

First in this section, an outline of the model is presented. Then each part of the model is described and finally a description is given of how the disturbances listed in the previous section are implemented in the model.

3.1 Model outline

The developed model is divided into two subsystems: engine and driveline, see Figure 1. The engine model consists of a crankshaft including a damping wheel, n_{cyl} cylinders, and the flywheel. The crankshaft is modeled as rotating masses connected with springs and dampers. Each rotating mass is represented by two circles connected by a vertical line. Each mass connected to the cylinders are affected by a cylinder torque $T_{\text{cyl},i}$.

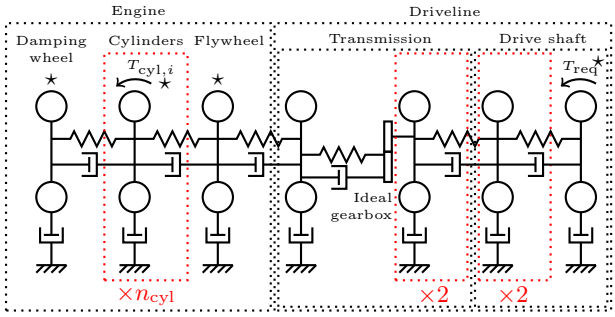

Figure 1: An overview of the model without disturbances. The engine driveline model is composed by a number of connected rotating masses. The connections between the masses are modeled as springs and dampers. The friction at each rotating mass is modeled as a damper connected to ground. The model has n_{cyl} cylinders where the torque from each cylinder affects a rotating mass of the crankshaft. The gearbox is assumed ideal. Modeled disturbances are marked as stars in the figure.

As input to the model, the mean angular velocity of the crankshaft and driveline is set by a required torque T_{req} at the drive shaft. The torque from each cylinder is modeled using an analytical pressure model, see Eriksson and Andersson (2002), describing the cylinder pressure during the combustion and a model of the moving piston mass, see Rizzoni and Zhang (1994).

Angular velocity measurements are simulated from the flywheel by computing time periods for angular intervals corresponding to the teeth angles of the flywheel.

The locations of disturbances in the model, listed in Section 2, are marked with stars in Figure 1. A more detailed description of how each disturbance is modeled can be found in Section 3.4.

3.2 Engine

Here, models of the different parts of the engine subsystem are described in detail.

Crankshaft

The crankshaft consists of $2+n_{\text{cyl}}$ rotating masses, corresponding to the damping wheel, n_{cyl} cylinders, and the flywheel. The connection between two masses is modeled as a spring and a damper. The friction at each mass is modeled as a damper connected to ground.

Each rotating mass connected to a cylinder is affected by a torque $T_{\text{cyl},i}$ related to the moving piston. A model of each rotating mass at position i, where $i = 1, \ldots, n_{\text{cyl}}$, is described as

$$\begin{aligned} J_i\,\dot{\omega}_i &= T_{\text{cyl},i} + c_{i+1,i}(\omega_{i+1} - \omega_i) - c_{i,i-1}(\omega_i - \omega_{i-1}) - \\ &\quad - c_i\,\omega_i + k_{i+1,i}(\theta_{i+1} - \theta_i) - k_{i,i-1}(\theta_i - \theta_{i-1}) \\ \dot{\theta}_i &= \omega_i \end{aligned} \tag{1}$$

where θ_i and ω_i are angular position and angular velocity respectively of the rotating mass i, J_i is the inertia, $c_{i-1,i}$ and $k_{i-1,i}$ are damping constant and spring constant respectively between the masses at position $i-1$ and i, c_i is the damping constant modeling friction, and $T_{\text{cyl},i}$ is the cylinder torque.

The damping wheel is positioned at the end of the crankshaft and is connected to auxiliary loads. A change in auxiliary load, for example if the AC is turned on, is modeled as a negative torque on the damping wheel T_{aux} which affects the damping wheel, represented by position index 0, as

$$\begin{aligned} J_0\,\dot{\omega}_0 &= -T_{\text{aux}} + c_{1,0}(\omega_1 - \omega_0) - c_0\,\omega_0 + k_{1,0}(\theta_1 - \theta_0) \\ \dot{\theta}_0 &= \omega_0 \end{aligned} \tag{2}$$

Cylinder

Each cylinder is modeled as a moving piston mass connected to the crankshaft by a rod. The resulting torque $T_{\text{cyl},i}$ on the rotating mass at position i is the sum of two force components: compression pressure force $F_{c,i}$ and piston mass force $F_{p,i}$, see Figure 2.

Each cylinder angle is modeled using a local angle $\tilde{\theta}_i$ around TDC. The angle θ_i of the corresponding rotating mass in the crankshaft model is translated to $\tilde{\theta}_i$ by adding a constant $\delta\theta_i$ to the angle θ_i. As an example, for a four cylinder engine the cylinder angles $\tilde{\theta}_i = \theta_i + \delta\theta_i$ where, depending on the firing order, $\delta\theta_i \in \{0, 180°, 360°, 540°\}$. The resulting torque $T_{\text{cyl},i}$ as a function of $\tilde{\theta}_i$, $F_{p,i}$,

and $F_{c,i}$, is given by

$$T_{\mathrm{cyl},i}(\tilde{\theta}_i) = \left(r\sin(\tilde{\theta}_i) + \frac{r^2\sin(2\tilde{\theta}_i)}{2\sqrt{l^2 - r^2\sin^2(\tilde{\theta}_i)}} \right) \times \times \left(F_{p,i}(\tilde{\theta}_i) + F_{c,i}(\tilde{\theta}_i) \right). \quad (3)$$

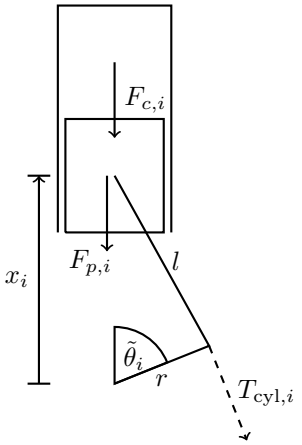

Figure 2: The cylinder model describes cylinder torque $T_{\mathrm{cyl},i}$ as a function of compression pressure force $F_{c,i}$ and the piston mass force $F_{p,i}$.

Compression pressure force The cylinder pressure force is modeled as the pressure difference between cylinder pressure and crankcase pressure multiplied with the cylinder area as

$$F_{c,i}(\tilde{\theta}_i) = A\left(p_{\mathrm{cyl,i}}(\tilde{\theta}_i) - p_{\mathrm{crank}} \right), \quad (4)$$

see Rizzoni and Zhang (1994). The cylinder pressure $p_{\mathrm{cyl,i}}$ is computed using an analytic model given in Eriksson and Andersson (2002) as

$$p_{\mathrm{cyl},i} = f\left(\tilde{\theta}_i, \theta_{\mathrm{ign},i}, \theta_{\mathrm{d},i}, \theta_{\mathrm{b},i}, p_{\mathrm{im},i}, p_{\mathrm{em},i}, \lambda_i, \omega_i, \chi_{\mathrm{mf}} \right) \quad (5)$$

describing the cylinder pressure for each cylinder at position i as a function of ignition angles: $\theta_{\mathrm{ign},i}$, $\theta_{\mathrm{d},i}$, $\theta_{\mathrm{b},i}$, representing ignition time, 10% fuel burned and 90% fuel burned, intake manifold pressure $p_{\mathrm{im},i}$, exhaust manifold pressure $p_{\mathrm{em},i}$, air to fuel ratio λ_i, crankshaft angle velocity ω_i, and a fuel conversion efficiency factor χ_{mf} to simulate misfire. The factor $\chi_{\mathrm{mf}} \in \{0, 1\}$ is added to the expression for computing η_f in (6) in Eriksson and Andersson (2002) as

$$\eta_f(\lambda, \chi_{\mathrm{mf}}) = \chi_{\mathrm{mf}}\left(0.95\ \min(1, 1.2\lambda - 0.2)\right), \quad (6)$$

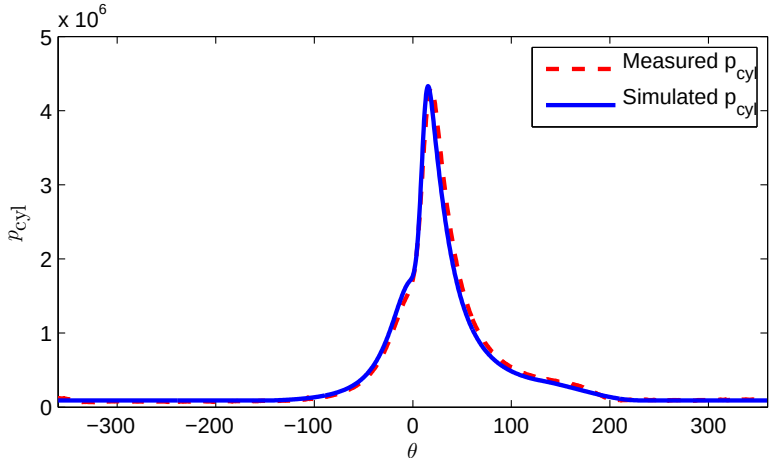

Figure 3: Cylinder pressure model compared to measurement data.

where $\chi_{\text{mf}} = 0$ when simulating a misfire.

In Figure 3, the output of the cylinder pressure model (5) is compared to a measured pressure trace. The model is able to simulate cylinder pressure and by varying the input parameters, different pressure traces can be modeled. For a more detailed description of the pressure model, the reader is referred to Eriksson and Andersson (2002).

Piston mass force The moving piston mass results in an additional force component $F_{p,i}$, see Figure 2. The significance of this force increases at higher speeds. The mass of the rod connecting the piston to the crankshaft is modeled such that it is translated to the piston mass and rotating mass of the crankshaft. The piston mass velocity and acceleration as functions of the angle $\tilde{\theta}_i$ are given by

$$\frac{dx_i(\tilde{\theta}_i)}{d\tilde{\theta}_i} = -r\sin(\tilde{\theta}_i) - \frac{r^2\sin(\tilde{\theta}_i)\cos(\tilde{\theta}_i)}{\sqrt{l^2 - r^2\sin^2(\tilde{\theta}_i)}} \tag{7}$$

$$\frac{d^2x_i(\tilde{\theta}_i)}{d\tilde{\theta}_i^2} = -r\cos(\tilde{\theta}_i) - \frac{r^2\left(\cos^2(\tilde{\theta}_i) - \sin^2(\tilde{\theta}_i)\right)}{\sqrt{l^2 - r^2\sin^2(\tilde{\theta}_i)}} - \\ - \frac{r^4\sin^2(\tilde{\theta}_i)\cos^2(\tilde{\theta}_i)}{\left(\sqrt{l^2 - r^2\sin^2(\tilde{\theta}_i)}\right)^3}, \tag{8}$$

see Rizzoni and Zhang (1994). Since the model is simulated in the time domain, the piston mass acceleration as a function of time is given by

$$\frac{d^2x_i(\tilde{\theta}_i)}{dt^2} = \frac{d^2x_i}{d\tilde{\theta}_i^2}\omega_i^2 + \frac{dx_i}{d\tilde{\theta}_i}\dot{\omega}_i, \tag{9}$$

and the resultant force is given by Newton's second law as

$$F(\tilde{\theta}_i) = m\frac{d^2x_i(\tilde{\theta}_i)}{dt^2}. \tag{10}$$

The term $\frac{dx_i}{d\tilde{\theta}_i}\dot{\omega}_i$ in (9) is modeled by a variable inertia in (1) as

$$J_i(\tilde{\theta}_i) = J_{i,c} - m\left(r\sin(\tilde{\theta}_i) + \frac{r^2\sin(2\tilde{\theta}_i)}{2\sqrt{l^2 - r^2\sin^2(\tilde{\theta}_i)}}\right)\frac{dx(\tilde{\theta}_i)}{d\tilde{\theta}_i} \tag{11}$$

where $J_{i,c}$ is the inertia of the rotating mass of the crankshaft and the piston mass force $F_{p,i}(\tilde{\theta}_i)$ is given by

$$F_{p,i}(\tilde{\theta}_i) = m\frac{d^2x_i}{d\tilde{\theta}_i^2}\omega_i^2. \tag{12}$$

Flywheel

The model describing the flywheel is the same as (1) excluding $T_{\mathrm{cyl},i}$. The timing and interrupts when the sensor passes the flywheel teeth are simulated by computing the time period between two specified angles of the flywheel, see Figure 4. The simulated angular velocity measurement can be generated off-line using simulated data.

To keep track of the angle of the flywheel, two teeth are removed which helps to identify the start of each revolution. To simulate low resolution measurements, time periods are computed over several teeth.

Manufacturing errors resulting in unequal distances between teeth angles of the flywheel are important to consider because they will affect the accuracy of the measurements between different vehicles, see Kiencke (1999).

3.3 Driveline

The driveline model is based on the model described in Nickmehr et al. (2012) and represents the system from the transmission to the drive shaft. The model consists of repeated blocks of connected rotating masses which means that the model can be easily adapted by adding or removing blocks for different system configurations, see Figure 1.

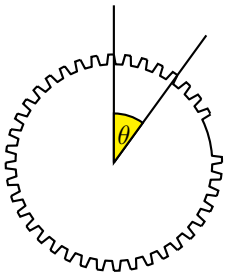

Figure 4: Angular velocity measurements are generated by computing the time difference between two teeth when the flywheel rotates. Two teeth are removed to keep track of the angular position of the flywheel.

The transmission is modeled using three rotating masses as in Figure 1, modeling clutch and gearbox. The transmission is modeled as ideal,

$$\begin{aligned} T_1 &= \gamma T_2 \\ \omega_2 &= \gamma \omega_1 \end{aligned} \tag{13}$$

where γ is the selected gear ratio.

The driveline after the transmission out to the wheels are modeled as additional rotating masses. In this implementation three rotating masses are included after the transmission to model the drive shaft as indicated in Figure 1. The required torque T_{req} at the wheel is modeled at the last rotating mass on the driveline.

3.4 Modeling disturbances

As discussed in Section 2, a main purpose of the developed model is to simulate flywheel angular velocity measurements and the effects of injected disturbances. The considered disturbances are related to combustion variations, flywheel errors, auxiliary load variations, and disturbances in road load torques. Here a description of how the disturbances are modeled is presented.

By varying the angles describing the ignition, $\theta_{\text{ign},i}$, $\theta_{\text{d},i}$, and $\theta_{\text{b},i}$, in the cylinder pressure model (5), cycle to cycle variations can be modeled. Different types of fuel quality affecting combustion can be modeled by generating the angles $\theta_{\text{d},i}$ and $\theta_{\text{b},i}$ as random variables changing from cycle to cycle. Late ignitions are modeled by using later ignition angles $\theta_{\text{ign},i}$. A misfire is modeled by setting χ_{mf} in (6) to zero which corresponds to a fuel conversion efficiency equal to zero. The occurance of these types of disturbances affecting combustion can be specified from cycle to cycle. The cycle to cycle variation is implemented by using a vector for each parameter in (5) where a counter, which is updated each cycle, specifies which element in the vector to use.

In Figure 5, a real pressure trace for one cylinder is compared with simulated pressure. In the simulation the ignition angle is varied from cycle to cycle and

during the last combustion the ignition is chosen to occur relatively late which is shown by the lower pressure and a small extra peak. Simulating with late ignition angles can be used to simulate for example cold starts and gear shifts.

Auxiliary load variations, for example turning on and off the AC, is modeled as an additional torque T_{aux} in (2). A disturbance at the driveline, for example when crossing a railroad, is modeled as an extra torque component added to the required road load torque T_{req}.

Errors in the angular velocity measurement or different resolutions can be simulated by making small changes to the angles where measurements are made. Manufacturing errors can result in inaccurate teeth angles which are cyclic for each vehicle but varying between different vehicles.

4 Model validation

First a short description of experimental data is presented. Then the results from a qualitative evaluation of the model is discussed.

4.1 Experimental data

Two types of validation data have been used. High resolution measurements from a four cylinders vehicle with angular resolution of 0.5 degrees is used for model validation. Data is used where the time period between two teeth is measured with angular resolution of 36 degrees. A typical data sequence is shown in the upper figure in Figure 6. The data is used to validate simulated flywheel measurement data from the model. Depending on what type of validation data that is used the number of cylinders in the model is adapted.

4.2 Validation

One problem with experimental data is that the disturbances are not measured making a quantitative validation impossible. This is also one intended application of the model, to investigate how different diagnosis methods can decouple these types of disturbances. A qualitative analysis of the model is performed by simulating the different types of disturbances and comparing the result with real data. Measurement data is not available for all types of disturbances.

In Figure 6, angular velocity data are simulated for a five cylinder engine shown in the lower figure which is compared to real data in the upper figure. The angles $\theta_{\text{d},i}$ and $\theta_{\text{b},i}$ are generated as Gaussian distributed random variables to simulate cycle to cycle variations, see Eriksson (2000). The simulated data resembles the real data capturing the oscillations from the cylinder combustions but also the amplitude of the oscillations.

In Figure 7, manufacturing errors on the flywheel are modeled by adding random teeth angle errors visible as the cyclic variations in the measurement

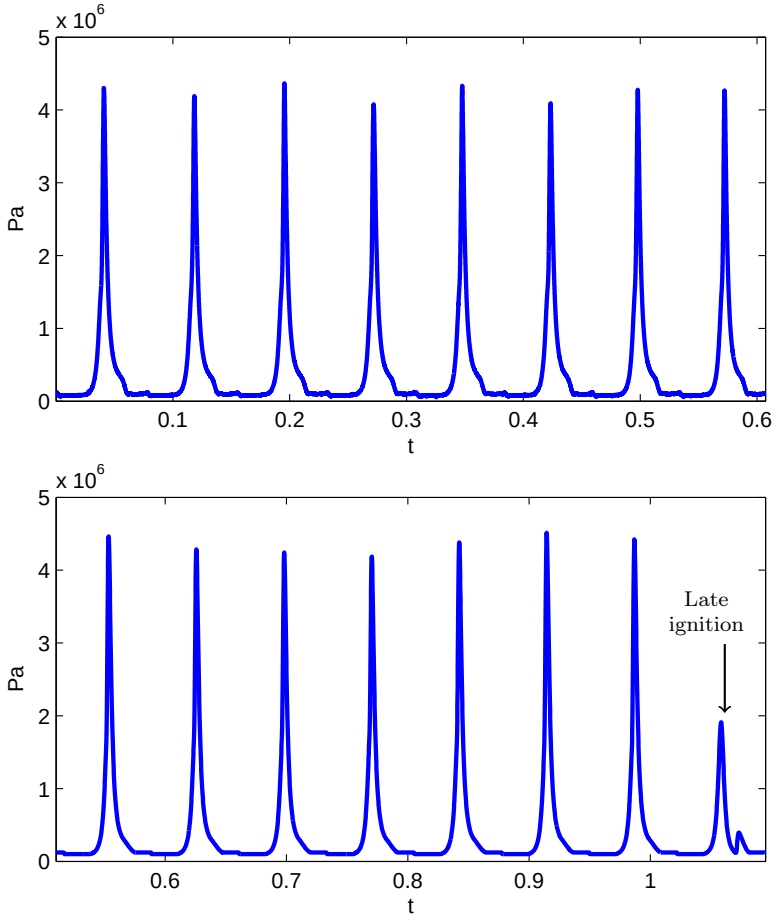

Figure 5: Measured cylinder pressure in one cylinder in the upper figure compared to simulated cylinder pressure with varying ignition angle in the lower figure. The last peak is a simulated late ignition

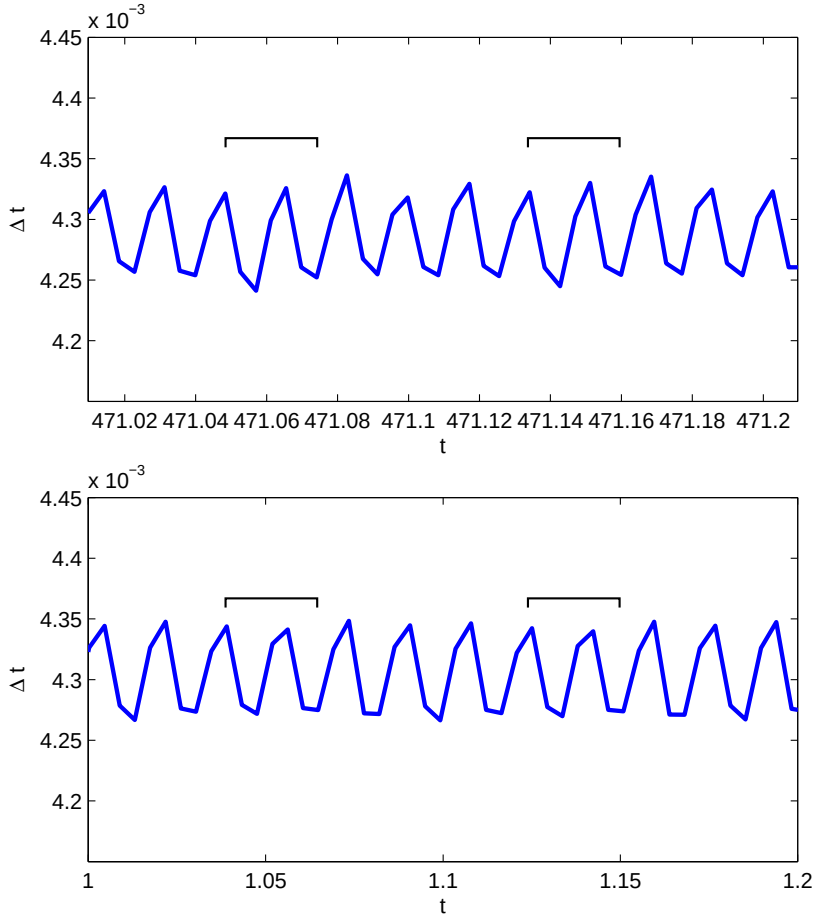

Figure 6: Simulated angular velocity measurements at the flywheel for a five cylinder engine. A qualitative comparison with the real measurement data in Figure 6 shows that the model is able to mimic the behavior of the real data. The cyclic measurement error caused by manufacturing error of the flywheel is marked in the upper figure and a simulation of a similar error in the lower figure.

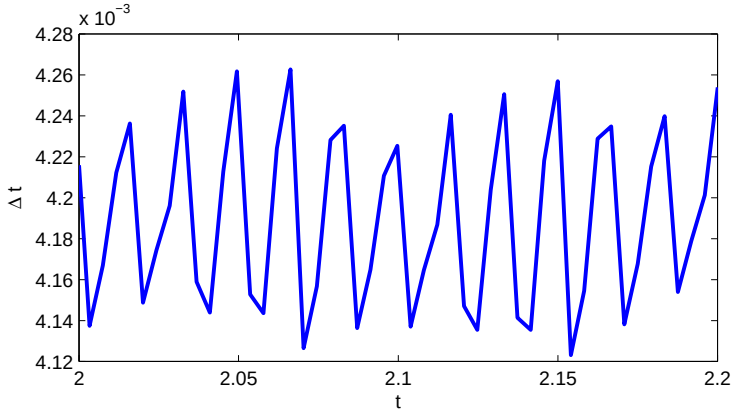

Figure 7: Errors have been added to the teeth angles where measurements on the flywheel are made.

data which repeats every fifth oscillation. An example of the same type of repetitive behavior is marked in the real measurement data in Fig 6.

A simulated misfire is visible at $t = 1.05$ in Figure 8 as a swift increase in time passed per angular interval. The amplitude of the signal during the misfire is almost equal for the real data and the simulation and the subsequent oscillation follows of the winding of the crankshaft which is captured by the model.

There is no data available to compare simulations of change in auxiliary loads and disturbances to the required torque to real data. Anyhow, data from simulations are provided to visualize how a disturbance of the required torque and a change in auxiliary load connected to the damping wheel affects the measurements, see Figure 9 and Figure 10 respectively. A driveline disturbance is simulated as an impulse to a constant required torque. The result in Figure 9 shows that the disturbance is shown but there is no large change during one combustion like for a misfire see Figure 8. Simulation of a negative step change in torque at the damping wheel is shown in Figure 10 which resembles more the oscillations caused by a misfire. The result indicates that a large sudden change in auxiliary load could be mistaken for a misfire.

5 Conclusions

A flywheel angular velocity model for misfire and disturbance simulation is developed. Different types of disturbances can be injected in the model to analyze their effects on the flywheel measurements. The model is modular to enable easy adaptation of the model structure to different vehicle configurations.

The disturbances that can be injected are misfire, cylinder variations, flywheel

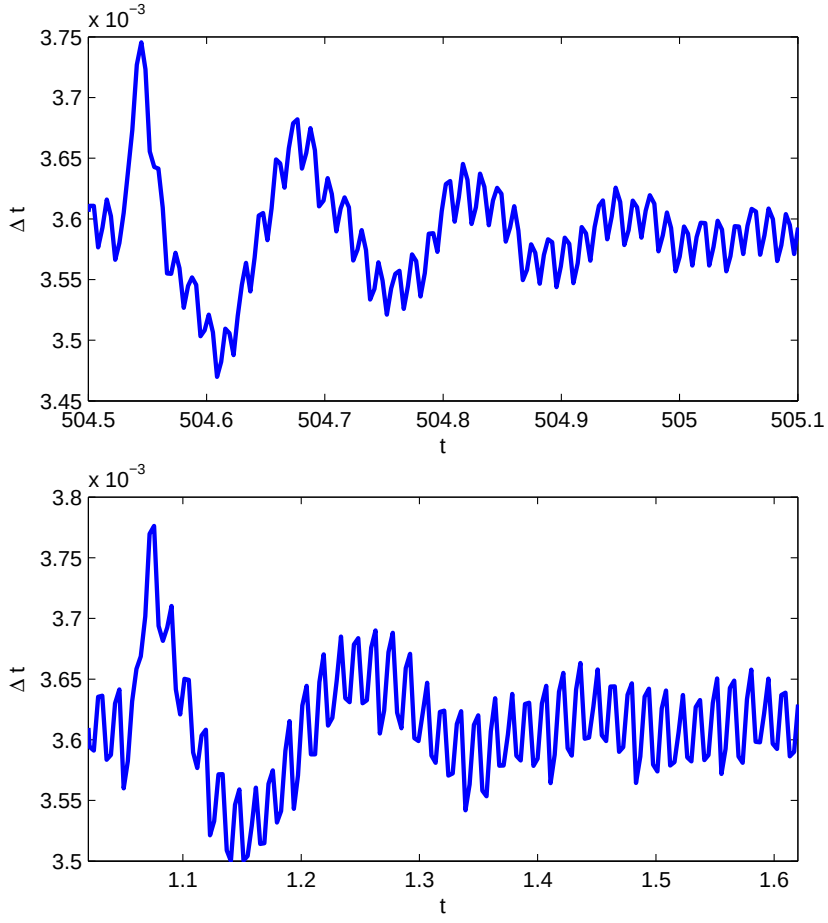

Figure 8: In the lower figure simulated angular velocity measurements at the flywheel are shown during a misfire and in the upper figure real measured angular velocity during a misfire.

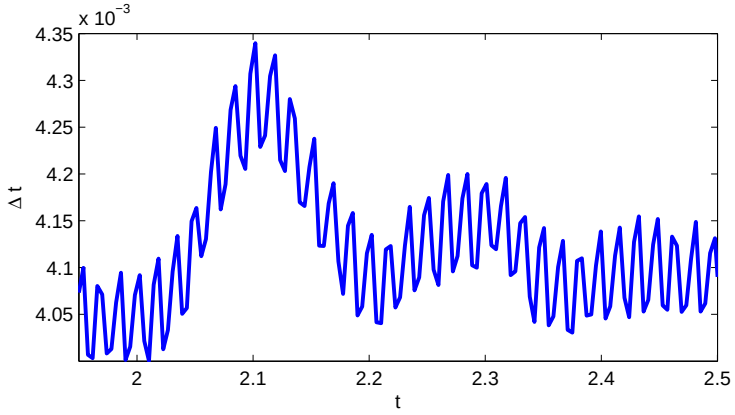

Figure 9: Simulated driveline disturbance at time 2 measured at the flywheel. Disturbance modeled as an impulse added to the constant required torque.

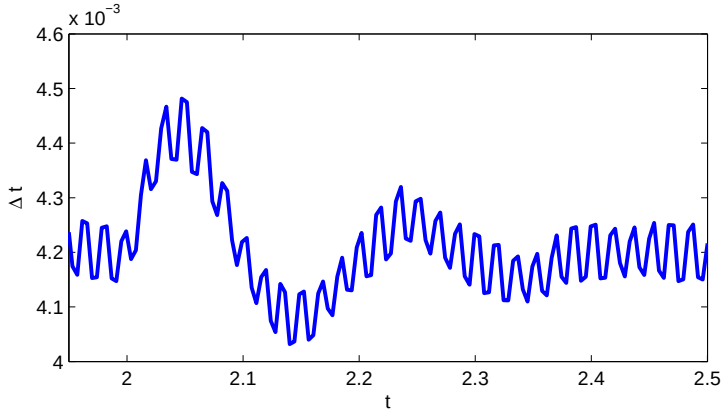

Figure 10: Simulated auxiliary load disturbance at time 2 measured at the flywheel. The auxiliary load is modeled as a step of negative torque at the damping wheel.

manufacturing errors, change in auxiliary load, and disturbances to road load torques.

A qualitative validation of the model is performed where simulations are compared to real data. Results show that the model captures the important behavior of the angular velocity measurements at the flywheel including misfire.

During the development of the misfire model, parameter tuning has been performed manually to capture the qualitative behavior shown in measured data which was validated in the previous section. A more systematic tuning of model parameters is useful to generate misfire data behaving like real measurement data from a specific car configuration.

Acknowledgements

This work has been supported by Volvo Cars. The authors would like to thank Sasa Trajkovic at Volvo Cars for helping us with the data collection.

References

Francis T. Connolly and Giorgio Rizzoni. Real time estimation of engine torque for the detection of engine misfires. *Journal of Dynamic Systems, Measurement, and Control*, 116(4):675–686, 1994.

A.R. Crowther and N. Zhang. Torsional finite elements and nonlinear numerical modelling in vehicle powertrain dynamics. *Journal of Sound and Vibration*, 284:825 – 849, 2005.

Lars Eriksson. Spark advance for optimal efficiency. *SAE Transactions, Journal of Engines, 1999-01-0548*, Volume 108:789–798, 2000.

Lars Eriksson and Ingemar Andersson. An analytic model for cylinder pressure in a four stroke SI engine. In *Modeling of SI engines*, number SAE Technical Paper 2002-01-0371 in SP-1702, 2002.

U. Kiencke. Engine misfire detection. *Control Engineering Practice*, 7(2):203 – 208, 1999.

Giorgio Minelli, Davide Moro, Nicolò Cavina, Enrico Corti, and Fabrizio Ponti. An Engine-Vehicle Simulator for the Calibration of Misfire Detection Algorithms. 59^o *Congresso annuale ATI, Genova*, pages 861 – 871, 2004.

J Mohammadpour, M Franchek, and K Grigoriadis. A survey on diagnostic methods for automotive engines. *International Journal of Engine Research*, 13 (1):41–64, 2012.

Sanjeev Naik. Advanced misfire detection using adaptive signal processing. *International Journal of Adaptive Control and Signal Processing*, 18(2):181–198, 2004.

N. Nickmehr, J. Åslund, L. Nielsen, and K. Lundahl. On experimental-analytical evaluation of passenger car ride quality subject to engine and road disturbances. 19th International Congress on Sound and Vibration, Vilnius, Lithuania, 2012.

El-Adl Mohammed Aly Rabeih. *Torsional vibration analysis of automotive drivelines*. PhD thesis, The University of Leeds, 1997.

Giorgio Rizzoni and Yuxuan Zhang. Identification of a non-linear internal combustion engine model for on-line indicated torque estimation. *Mechanical Systems and Signal Processing*, 8(3):275 – 287, 1994.

Stefan Schagerberg. *Torque sensors for engine applications*. Chalmers tekniska högsk., Göteborg, 2003. Lic.-avh. Göteborg : Chalmers tekn. högsk., 2003.

Francisco V. Tinaut, Andrés Melgar, Hannes Laget, and José I. Domínguez. Misfire and compression fault detection through the energy model. *Mechanical Systems and Signal Processing*, 21(3):1521 – 1535, 2007.

Paper E

Analysis and optimization with the Kullback-Leibler divergence for misfire detection using estimated torque*

E

*Published as *Technical Report LiTH-ISY-R-3057*, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 2013.

Analysis and optimization with the Kullback-Leibler divergence for misfire detection using estimated torque

Daniel Eriksson, Lars Eriksson, Erik Frisk, and Mattias Krysander

Vehicular Systems, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.

Abstract

Engine misfire detection is an important part of the On-Board Diagnostics (OBDII) legislations to reduce exhaust emissions and avoid damage to the catalytic converters. The flywheel angular velocity signal is analyzed, investigating how to use the signal in order to best detect misfires. An algorithm for engine misfire detection is proposed based on the flywheel angular velocity signal. The flywheel signal is used to estimate the torque at the flywheel and a test quantity is designed by weighting and thresholding the samples of estimated torque related to one combustion. During the development process, the Kullback-Leibler divergence is used to analyze the ability to detect a misfire given a test quantity and how the misfire detectability performance varies depending on, e.g., load and speed. The Kullback-Leibler divergence is also used for parameter optimization to maximize the difference between misfire data and fault-free data. Evaluation shows that the proposed misfire detection algorithm is able to have a low probability of false alarms while having a low probability of missed detections.

1 Introduction

Engine misfire detection is an important part of the On-Board Diagnostics (OBDII) legislations to reduce exhaust emissions and avoid damage to the catalytic converters, see Mohammadpour et al. (2012). Misfire refers to an incomplete combustion inside a cylinder and can be caused by many different factors, for example, a failure in the ignition system or a clogged fuel injector, see Heywood (1988). The On-Board Diagnostics system is required to both detect misfires and identify in which cylinder the misfire occurred, see Heywood (1988) and Walter et al. (2007).

An overview of engine misfire detection research is found in Mohammadpour et al. (2012). There are several approaches to detect misfires using different types of sensors, e.g., ion current sensors, see Lundström and Schagerberg (2001), or crankshaft angular velocity measured at the flywheel, see Osburn et al. (2006), Naik (2004), and Tinaut et al. (2007). Misfire detection based on torque estimation using the flywheel angular velocity signal has been studied in, e.g, Connolly and Rizzoni (1994), Kiencke (1999), and Walter et al. (2007). Detecting misfires is a non-trivial problem which is complicated by, for example, changes in load, speed, and flywheel manufacturing errors, see Naik (2004) and Kiencke (1999).

As a tool during the design process of the misfire detection algorithm, the Kullback-Leibler divergence is proposed to evaluate misfire detectability performance Eguchi and Copas (2006). The Kullback-Leibler divergence can be used to quantify the separation of two distributions, for example, to evaluate the misfire detectability performance before a test quantity is designed or when optimizing algorithm parameters to maximize the separation. Other examples where the Kullback-Leibler divergence is used for quantifying diagnosability performance are Eriksson et al. (Accepted for publication) and Eriksson et al. (2012).

A misfire detection algorithm is proposed based on torque estimation at the flywheel using flywheel angular velocity measurements. The test quantity is designed by weighting the estimated torque related to one combustion such that the distributions of fault-free data and misfire data are as separated as possible. To handle variations of the estimated torque due to cold starts and varying load and speed, the parameters of the misfire detection algorithm depend on the operating point. Parameterization of the misfire detection algorithm is performed automatically given training data and evaluations show that misfires are detected with low probability of missed detections while having few false alarms.

A short presentation of the vehicle control system signals is given in Section 2 and an analysis of the flywheel angular velocity signal and the effects of misfire is presented in Section 3. In Section 4, the Kullback-Leibler divergence is described which is used to analyze misfire detectability performance. The flywheel angular velocity signal is used to estimate the torque at the flywheel in Section 5 and the estimated torque is analyzed using the Kullback-Leibler divergence. Then, the

misfire detection algorithm is presented in Section 6 and evaluated in Section 7. Finally, the conclusions are presented in Section 8 and future works in Section 9.

2 Vehicle control system signals

A list of the vehicle control system signals and variables used by the misfire detection algorithm is presented in Table 1. All data used in this work are measured from real cars from Volvo. The flywheel angular velocity signal is used to detect misfires and the crank counter, load, speed, and the catalytic converter warming flag are used to identify the operating point of the engine, in which cylinder the combustion occurs, and if it is during a cold start or not. All signals are sampled crankshaft angle synchronous, instead of time synchronous, at the same rate which depends on the number of cylinders.

During the data collection, misfires have been injected by interrupting the combustion in different cylinders. The misfires are occurring with fixed intervals between cycles while switching which cylinder that is misfiring. Note that it is not known if the data contains more misfires than the injected misfire. Therefore, all samples not related to an injected misfire is assumed fault-free. Thus, only intermittent misfires are analyzed is this work and no occurrence of a constant misfiring cylinder. However, intermittent misfires are considered more difficult to detect since the fault appears and disappears randomly.

Table 1: A list of vehicle control system signals used by the misfire detection algorithm.

Signal	Unit
Flywheel angular velocity signal	$\mu s/\Delta\theta$
Crank counter	-
Speed	rpm
Load	g/rev
Catalytic converter warming flag	-

The flywheel angular velocity signal is the measured time period for different angular intervals of the flywheel, see Ken-jen Lang, Lela Liu, Alec L. Lang, and Louis Yizhang Liu (2005). The angular intervals are determined and limited by the angles between the teeth on the flywheel, which are usually 6°. Two teeth are removed to keep track of the angle of the flywheel as shown in Figure 1. The available flywheel angular velocity signals have the lower resolution of 30° for five cylinder engines and 36° for six cylinder engines, see Table 2. This results in 20 samples per cycle for five cylinder engines and 24 samples per cycle for six cylinder engines. In both cases, this results in four samples per cylinder combustion.

To keep track of the angle of the flywheel in the vehicle control system, a crank counter variable is used. The crank counter is increased by one for each

Table 2: A summary of the flywheel angular velocity signal resolution for different number of cylinders.

Num. cylinders	Resolution	samples/rev.	samples/cycle	samples/cyl.
5	36°	10	20	4
6	30°	12	24	4

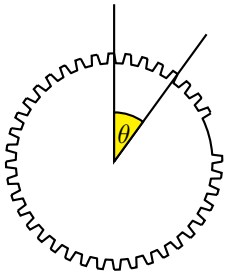

Figure 1: The flywheel angular velocity signal is generated by computing the time difference between two following teeth when the flywheel rotates. Two teeth are removed to keep track of the angular position of the flywheel.

new sample of the flywheel angular velocity signal. As shown in Table 2, there are four samples of the flywheel angular velocity signal between two combustions. The firing order of the five cylinder engine is 1-2-4-5-3 and for the six cylinder engine 1-5-3-6-2-4. The crank counts associated to each cylinder for five and six cylinder engines are presented in Table 3 and follows the firing order of each engine. The crank counts have been selected by analyzing when the signal is affected by misfire which is discussed more in Section 5.1.

Table 3: The crank counter values in the vehicle control system that are associated to each cylinder for five and six cylinder engines.

	Crank counts	
Cylinder	Five cylinder engine	Six cylinder engine
1	0, 1, 2, 3	1, 2, 3, 4
2	4, 5, 6, 7	17, 18, 19, 20
3	16, 17, 18, 19	9, 10, 11, 12
4	8, 9, 10, 11	21, 22, 23, 0
5	12, 13, 14, 15	5, 6, 7, 8
6	-	13, 14, 15, 16

The speed signal is an estimation of the crankshaft angular velocity measured in revolutions per minute (rpm). The load signal measures the air mass in the cylinders per revolution, measured in grams per revolution (r/rev). To handle dependencies of the flywheel angular velocity signal to different speeds and loads,

the samples related to one cylinder, see Table 3, are associated to one of the operating points in Table 4. The operating points have been chosen the same as in the misfire detection algorithm used by Volvo Cars. Samples related to one combustion are classified to the closest, in the Euclidean sense, operating point given the speed and load signal at the last sample for the cylinder. The last sample is given at the last crank count for each cylinder, see Table 3. As an example for a six cylinder engine, samples of the flywheel angular velocity signal from cylinder 1 is classified to the closest operating point corresponding to the value of the speed and load at crank count 4.

Table 4: Data is categorized in 42 different operating points depending on the speed and load.

Load	Speed	[rpm]				
[g/rev]	500	1000	1500	2000	2500	3000
0.3	$*$	$*$	$\cdots$		$\cdots$	$*$
0.4	$*$	$*$				$\vdots$
0.6	$\vdots$		$\ddots$			
0.8						
1.2						
1.6	$\vdots$				$\ddots$	$\vdots$
2.0	$*$	$\cdots$			$\cdots$	$*$

A catalytic converter warming flag in the vehicle control system is used to indicate when the ignition angle is chosen late in order to heat the catalytic converter during cold starts. Cold starts are known to be complicated when trying to detect misfires and the flag is used to identify when this occurs. Cold starts will be considered separately in the misfire detection algorithm using the catalytic converter warming flag.

3 Analysis of the flywheel angular velocity signal

Here, the flywheel angular velocity signal is analyzed and examples of how misfire and cold starts affect the flywheel are shown. The flywheel signal from a five cylinder engine is shown in Figure 2 where the samples related to one combustion of cylinder 3 are marked in the figure.

In Figure 3, two examples of the flywheel angular velocity signal from a five cylinder engine with misfire are shown. The effect of misfires are visible as a sudden increase in the measured time period. The increased time period is caused by the deceleration of the crankshaft and driveline due to the missed torque from the combustion. Flywheel angular velocity signals with misfires from

a six cylinder engine are shown in Figure 4. The same characteristic behavior is visible in the signals from a misfire as for the five cylinder engine. Comparing the two examples in Figure 3 and Figure 4, respectively, shows that the effects of a misfire to the flywheel signal varies depending on load and speed, which complicates misfire detection.

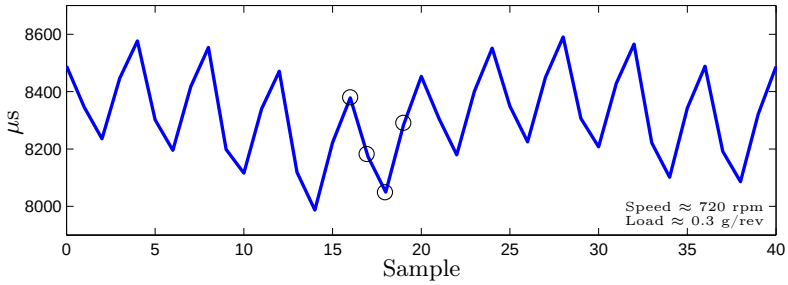

Figure 2: Example of angular velocity measurements at the flywheel from a five cylinder engine. Data related to one combustion from cylinder 3 is marked with circles.

As mentioned in the previous section, it is more difficult during cold starts to detect a misfire in the flywheel angular velocity signal, see Figure 5. The signal has a more stochastic nominal behavior and it is more difficult to distinguish an injected misfire from the other data because there are several jumps in the signal that are not related to any known misfire. Note that it is not known if these sudden increases in the signal are misfires or not.

The flywheel angular velocity signal depends on the speed and load which varies during a driving scenario. Figure 6 shows the flywheel angular velocity signal at one operating point, speed 1500 rpm and load 1.2 g/rev, for four of six cylinders. The blue curves represents the signal during the fault-free case and the red curves during the misfire case. The red curves are all within the variation of the blue curves and it is thus difficult to separate a misfire from fault-free data by just looking at the plots. Generally, in order to detect a misfire, the goal is to manipulate the data such that the fault-free data and misfire data are as separated as possible. The more the data from the two cases are separated, the easier it is to detect a misfire, using for example thresholding.

Comparing different methods to systematically improve misfire detectability requires a measure to quantify the separation between fault-free data and misfire data. To quantify the separation, the Kullback-Leibler divergence is used which is presented in the next section.

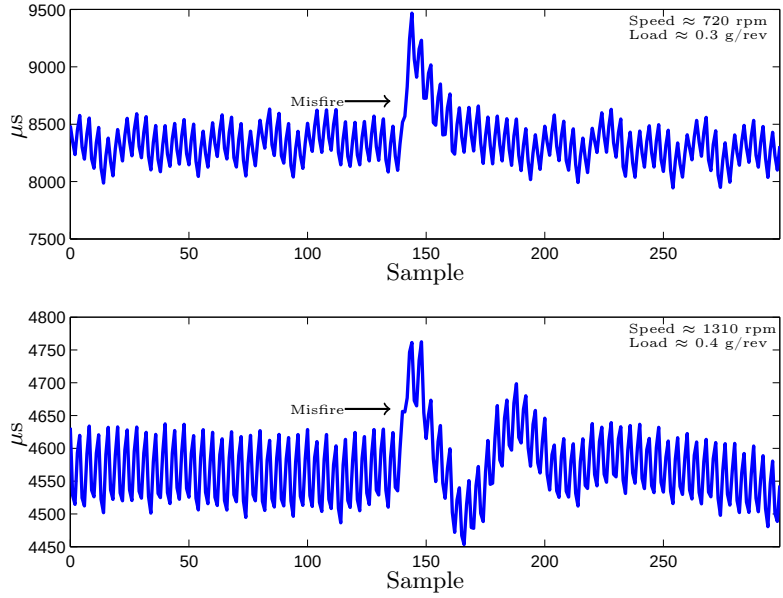

Figure 3: Examples of the flywheel angular velocity signal from a five cylinder engine. The signals contain misfires which are visible as increased measured time.

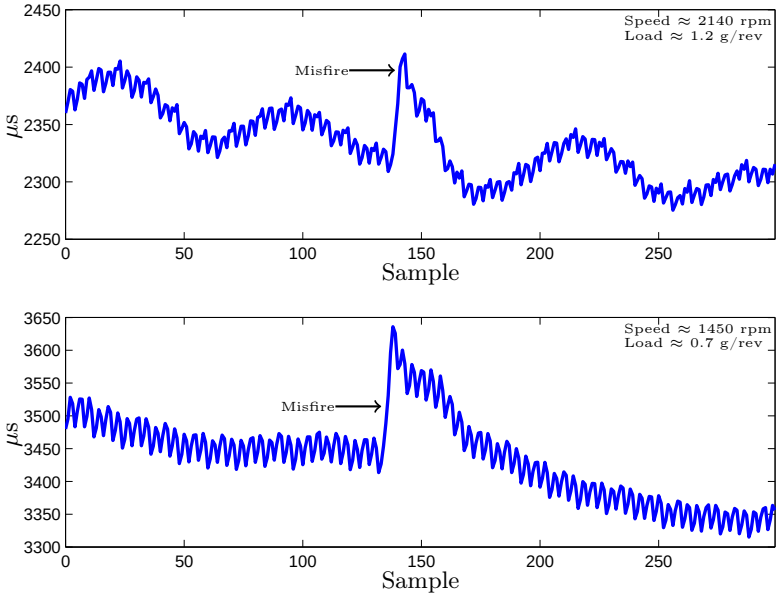

Figure 4: Examples of the flywheel angular velocity signal from a six cylinder engine. The signals contain misfires which are visible as increased measured time.

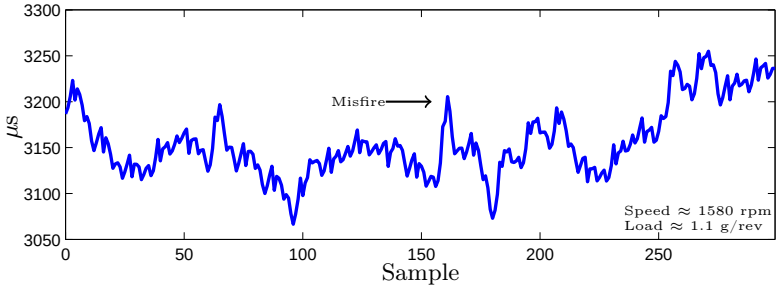

Figure 5: An example of the flywheel angular velocity signal from a six cylinder engine during cold start. Compared to signals in Figure 4, a misfire is more difficult to detect.

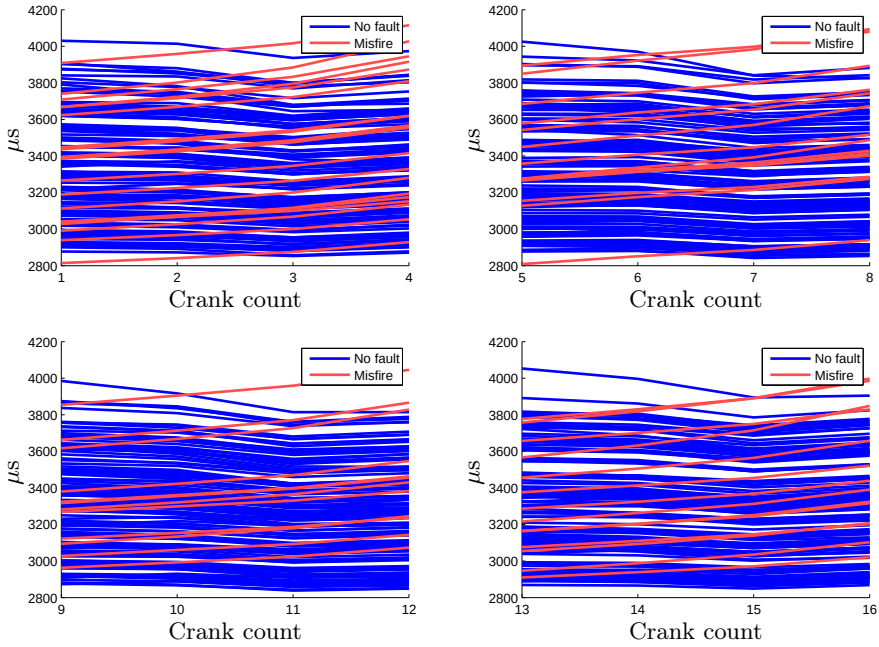

Figure 6: Flywheel angular velocity signal related to cylinder 1, 5, 3 and 6, see Table 3, during normal driving at operating point speed 1500 rpm and load 1.2 g/rev.

4 The Kullback-Leibler divergence

the Kullback-Leibler divergence is used here to quantify the misfire detectability performance of a signal or test quantity. The Kullback-Leibler divergence is suitable since it can be used to quantify the separation of the distributions of a given signal during the fault-free case and misfire case before constructing the misfire detection algorithm.

The Kullback-Leibler divergence Kullback, S. and Leibler, R. A. (1951) is a non-symmetric measure of the difference between two probability density functions, p_{f} and p_{nf} defined as

$$K(p_{\mathrm{f}}\|p_{\mathrm{nf}}) = \int_{-\infty}^{\infty} p_{\mathrm{f}}(x) \log \frac{p_{\mathrm{f}}(x)}{p_{\mathrm{nf}}(x)} dx \tag{1}$$

and is defined if:

- $p_{\mathrm{nf}}(x) > 0$ for all x where $p_{\mathrm{f}}(x) > 0$.
- If $p_{\mathrm{f}}(x) = 0$ then $p_{\mathrm{f}}(x) \log \frac{p_{\mathrm{f}}(x)}{p_{\mathrm{nf}}(x)} = 0$.

The Kullback-Leibler divergence is non-negative, i.e., $K(p_{\mathrm{f}}\|p_{\mathrm{nf}}) \geq 0$ with equality if and only if $p_{\mathrm{f}} = p_{\mathrm{nf}}$, and asymmetric, i.e, in general $K(p_{\mathrm{f}}\|p_{\mathrm{nf}}) \neq K(p_{\mathrm{nf}}\|p_{\mathrm{f}})$. A large value of (1) corresponds to that the distributions p_{f} and p_{nf} are well separated and it is easy to detect a misfire.

If p_{nf} is the distribution of the null hypothesis (no fault) and p_{f} is the distribution for the alternative hypothesis (misfire), then $K(p_{\mathrm{f}}\|p_{\mathrm{nf}})$ can be interpreted as the expected log-likelihood ratio if p_{f} is the true hypothesis, see Eguchi and Copas (2006). In Eriksson et al. (Accepted for publication), an alternative interpretation of the Kullback-Leibler divergence is: "How easy is it to distinguish if a sample is drawn from p_{f} (misfire) and not from p_{nf} (no fault)?". Thus, maximizing $K(p_{\mathrm{f}}\|p_{\mathrm{nf}})$ would make it easier to detect a misfire. A value close to zero corresponds to the probability density functions, p_{f} and p_{nf}, being similar, i.e. it is hard to distinguish misfire data from fault-free data.

Computing (1), requires that the probability distributions of p_{f} and p_{nf} are known. Often, an analytical solution of (1) is not available. However, a numerical approximation can be made using training data. There are several approaches to compute approximations of the distributions p_{f} and p_{nf} if data are available. The distribution of the data can be approximated either by assuming a known parametric distribution, for example, the Gaussian distribution Casella and Berger (2001), or by using a non-parametric method such as kernel density estimators, see Bishop (2006). A kernel density estimator approximates a distribution by summing up a number of kernels, e.g., Gaussian distributions, centered around each sample of the data. A kernel density estimator for a one dimensional distribution, where each sample x_n of the signal is represented by a Gaussian probability density function with a mean value equal to the value of

the sample, is

$$\hat{p}(x) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\left(2\pi\sigma^2\right)^{1/2}}e^{-\frac{(x-x_n)^2}{2\sigma^2}} \tag{2}$$

where $\hat{p}(x)$ is the estimated probability density function, N is the length of the signal, and σ is the standard deviation of the Gaussian components, see Bishop (2006). If data from the misfire case and the fault-free case are clearly separated then the computation of (1) will depend only of the tails of the distribution p_{nf} since p_{f} is close to zero for fault-free data. The tails of the distribution are difficult to estimate since there are few data measured from the tails. When the distributions are well separated, parametric approximated distribution of the data can be used to model the tails.

If p_{f} and p_{nf} are Gaussian distributions with known means, μ_{f} and μ_{nf}, and co-variance matrices, Σ_{f} and Σ_{nf}, then (1) can be computed exactly, see Casella and Berger (2001). For a one-dimensional Gaussian distributed variable where $p_{\mathrm{f}} \sim \mathcal{N}(\mu_{\mathrm{f}}, \sigma_{\mathrm{f}}^2)$ and $p_{\mathrm{nf}} \sim \mathcal{N}(\mu_{\mathrm{nf}}, \sigma_{\mathrm{nf}}^2)$, the Kullback-Leibler divergence $K(p_{\mathrm{f}}\|p_{\mathrm{nf}})$ is given by

$$K(p_{\mathrm{f}}\|p_{\mathrm{nf}}) = \frac{1}{2}\left(\frac{\sigma_{\mathrm{f}}^2}{\sigma_{\mathrm{nf}}^2} + \frac{(\mu_{\mathrm{nf}}-\mu_{\mathrm{f}})^2}{\sigma_{\mathrm{nf}}^2} - \log\frac{\sigma_{\mathrm{f}}^2}{\sigma_{\mathrm{nf}}^2} - 1\right). \tag{3}$$

Thus, by approximating the fault-free and misfire distributions of a signal as Gaussian distributions, the Kullback-Leibler divergence can be computed explicitly.

5 Torque estimation based on the angular velocity signal

As observed in Section 3, it is possible in most cases to manually identify a misfire in the flywheel angular velocity signal. However, when comparing several measurements at the same operating point, see Figure 6, it is difficult to select a fixed threshold to detect misfires while having a low probability of false alarms. Therefore, some additional signal processing is needed and here, the flywheel angular velocity signal is used to estimate the torque at the flywheel. Methods where cylinder torque is estimated in real-time for misfire detection can be found in, for example, Connolly and Rizzoni (1994) and Kiencke (1999). In Kiencke (1999), the torque affecting the flywheel is described as

$$J\dot{\omega} = \sum_{i=1}^{n_{\mathrm{cyl}}}\left(T_{c_i} + T_{p,i}\right) - T_{\mathrm{load}} - T_{\mathrm{fr}} \tag{4}$$

where J is the inertia of the flywheel, ω is the angular velocity of the flywheel, T_{c_i} is the torque from cylinder pressure at cylinder i, $T_{p,i}$ is the torque from piston mass at cylinder i, T_{load} is load torque and T_{fr} is friction. Each combustion results in a torque contribution T_{c_i} which is is not present during a misfire.

The flywheel angular velocity signal $x[\theta]$ is sampled at a fixed angular interval $\Delta\theta$. The flywheel angular velocity signal $x[\theta]$ measures the time period Δt for a fixed angular interval $\Delta\theta$. An approximation of the flywheel angular velocity during the angular interval $\Delta\theta$ is computed as $\omega = \frac{d\theta}{dt} \approx \frac{\Delta\theta}{\Delta t} = \frac{\Delta\theta}{x[\theta]}$. Since the signal is angle synchronous, the relation (4) is written as

$$J\frac{d\omega}{d\theta}\underbrace{\frac{d\theta}{dt}}_{=\omega} = T \tag{5}$$

where $T = \sum_{i=1}^{n_{\text{cyl}}} (T_{c_i} + T_{p,i}) - T_{\text{load}} - T_{\text{fr}}$. The factor $\frac{d\omega}{d\theta}$ is approximated using Euler forward as

$$\frac{d\omega}{d\theta} \approx \frac{\frac{\Delta\theta}{x[\theta+\Delta\theta]} - \frac{\Delta\theta}{x[\theta]}}{\Delta\theta}. \tag{6}$$

A simple estimate of T is then given by

$$J\frac{\Delta\theta}{x[\theta]}\left(\frac{\frac{\Delta\theta}{x[\theta+\Delta\theta]} - \frac{\Delta\theta}{x[\theta]}}{\Delta\theta}\right) = J\Delta\theta z[\theta]\left(z[\theta+\Delta\theta] - z[\theta]\right) = T \tag{7}$$

where $z[\theta] = x^{-1}[\theta]$ and $J\Delta\theta$ is assumed constant and can be seen as a scaling factor.

Two examples of estimated torque using the flywheel angular velocity signal are shown in Figure 7. The flywheel angular velocity signals from a five cylinder engine that have been used are shown in Figure 3. A combustion in a cylinder results in an increase in torque but a misfire results in a smaller peak. Compared to the flywheel angular velocity signal, the estimated torque oscillates around zero, although the mean value is slightly positive. This is convenient since there is no need of compensation in the signal for the variation in speed which will be shown in Section 5.1.

In Figure 8, two examples of estimated torque are shown using the flywheel angular velocity signal from the six cylinder engine shown in Figure 4. The estimated torque oscillates around zero but it shows a bigger torque drop from a misfire compared to Figure 7. The difference in shape between misfires in Figure 7 and Figure 8 are similar for different speeds and loads. The relative torque loss during misfire is bigger in general for the six cylinder engine case than for the five cylinder engine case.

Data from cold starts is only available from a six cylinder engine. Figure 9 shows the estimated torque from flywheel measurements during cold start where the corresponding flywheel angular velocity signal is shown in Figure 5. It is difficult to distinguish the misfire from other torque drops. However, there are some torque drops that are almost as low as the known misfire but it is not known if these drops are also caused by misfires.

To summarize this section, by estimating the torque at the flywheel, a misfire can be identified as a smaller torque gain from the misfiring cylinder. Also, since the estimated torque oscillates around zero, all data associated to the same

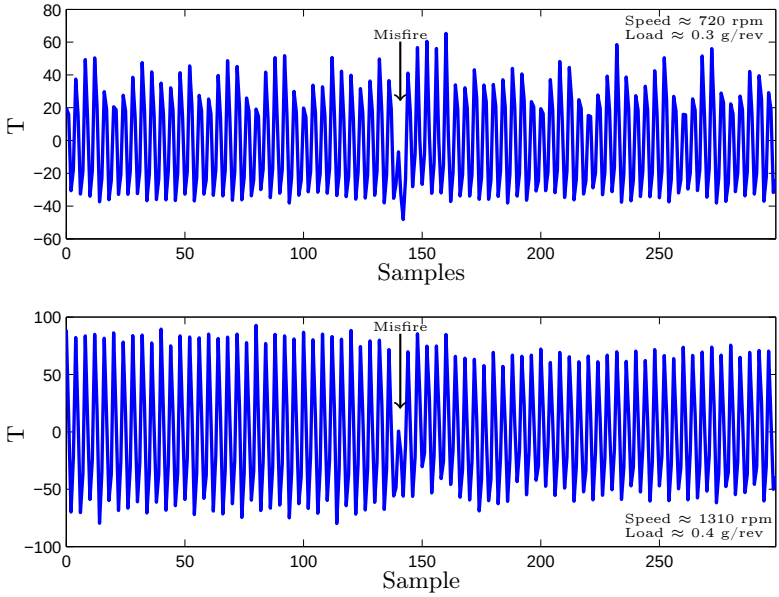

Figure 7: Examples of estimated torque from a five cylinder vehicle. The data contains misfires which are visible as smaller torque gain.

operating point will have the same mean value. Thus, it is not necessary to compensate for varying speed when comparing data from each operating point to identify a misfire. The misfire detectability performance of using the estimated torque is further analyzed in Section 5.1

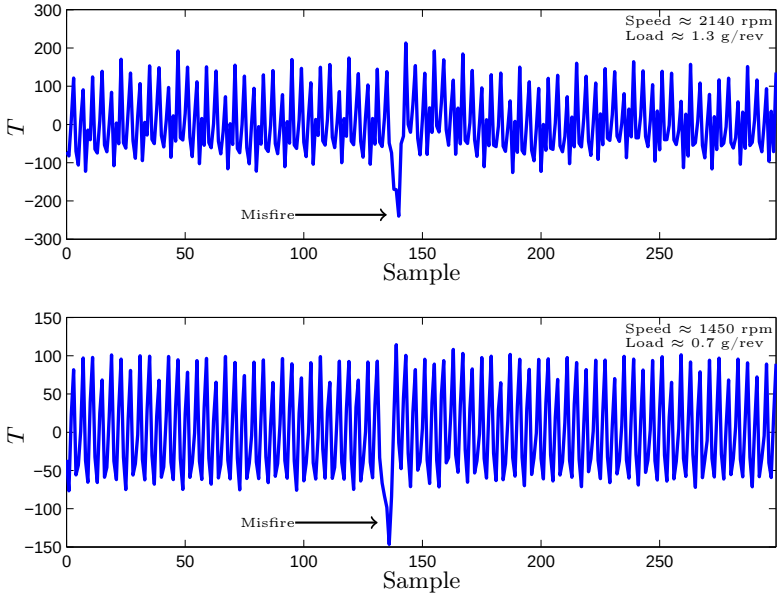

Figure 8: Examples of estimated torque from a six cylinder vehicle. The data contains misfires which are visible as smaller torque gain.

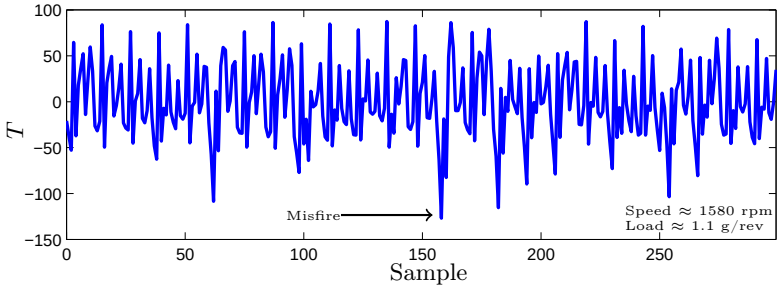

Figure 9: An example of estimated torque during cold start from a six cylinder vehicle. Compared to data in Figure 7, a misfire is more difficult to detect during cold starts.

5.1 Analyzing misfire detectability performance of estimated torque signal

In the previous section, the flywheel angular velocity signal is used to estimate the torque at the flywheel. An interesting question is how much performance is gained by using the estimated torque instead of the original flywheel signal. It is also interesting to analyze how the performance varies, depending on operating point and if it is during cold start or not. Here, the estimated torque based on fault-free data and misfire data is analyzed using the Kullback-Leibler divergence described in Section 4.

Signal distribution at one operating point

First, a comparison is made between torque estimation of fault-free data and misfire data for each cylinder separately at one operating point. An example of estimated torque for fault-free combustions and misfires for different cylinders at operating point speed 1500 rpm and load 1.2 g/rev are shown in Figure 10. Each plot shows the data from one cylinder where the upper left plot is cylinder 1, upper right plot is cylinder 5, lower left plot is cylinder 3 and lower right plot is cylinder 6. The plots show that the estimated torque clearly differs between normal behavior and misfire compared to the flywheel angular velocity signal shown in Figure 6. The difference between the distributions for each of the four cylinders in Figure 10 is largest at the second sample in each plot and smallest at the last sample.

Consider the upper left plot in Figure 10. The distribution of fault-free data and misfire data for each sample of cylinder 1 is plotted in Figure 11. The upper left plot is the distribution of the first sample of the combustion, the upper right is the second sample and so on. By comparing the four plots, it is clear that the distributions are separated the most at the second sample (upper right plot) and the least at the last sample (lower right plot) which is also shown in Figure 10.

Figure 12 shows the estimated torque during cold starts, at operating point speed 1500 rpm and load 1.2 g/rev. During cold starts, when the catalytic converter is heated, the difference between fault-free data and misfire data is smaller compared to the normal case which indicates that misfire detection is more difficult during cold starts compared to normal driving since the data is not as separated as in Figure 10. Note that some of the torque traces are similar to the torque during normal driving in Figure 10, which have maximum torque in the second sample. This might be caused by a latency in the catalytic converter warming flag.

A comparison of the sample distributions of cylinder 1 during cold start in the upper left plot in Figure 12 is shown in Figure 13. The distributions of fault-free data and misfire data are most separated in the last sample and the least in the first sample. A possible explanation for this is that during cold start the ignition angle is intentionally set later to heat the catalytic converter. Therefore, since the combustion occurs later, the effect of a combustion is visible

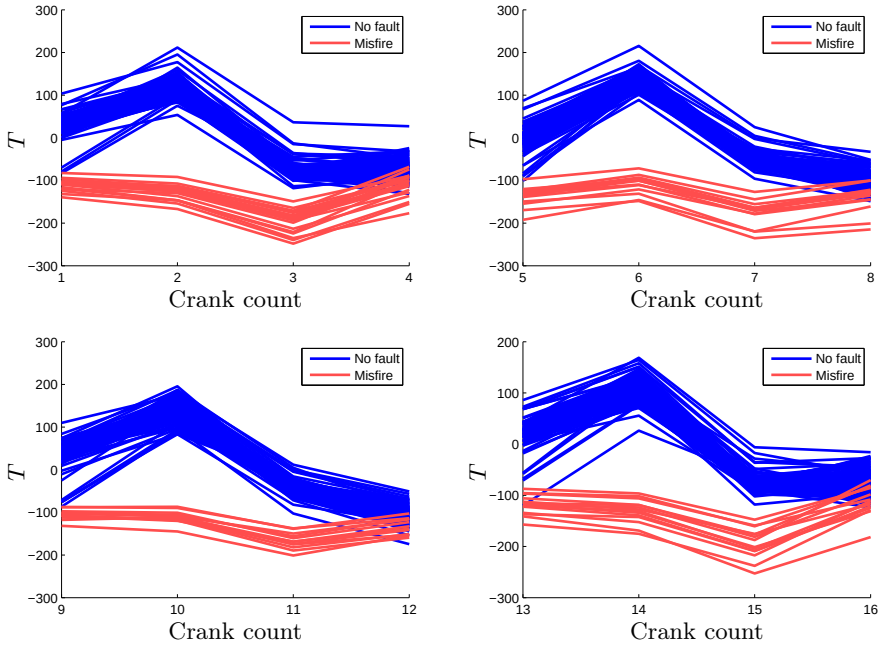

Figure 10: Estimated torque of cylinder 1, 5, 3, and 6 during normal driving at operating point speed 1500 rpm and load 1.2 g/rev.

later in the estimated torque compared to the results in Figure 11.

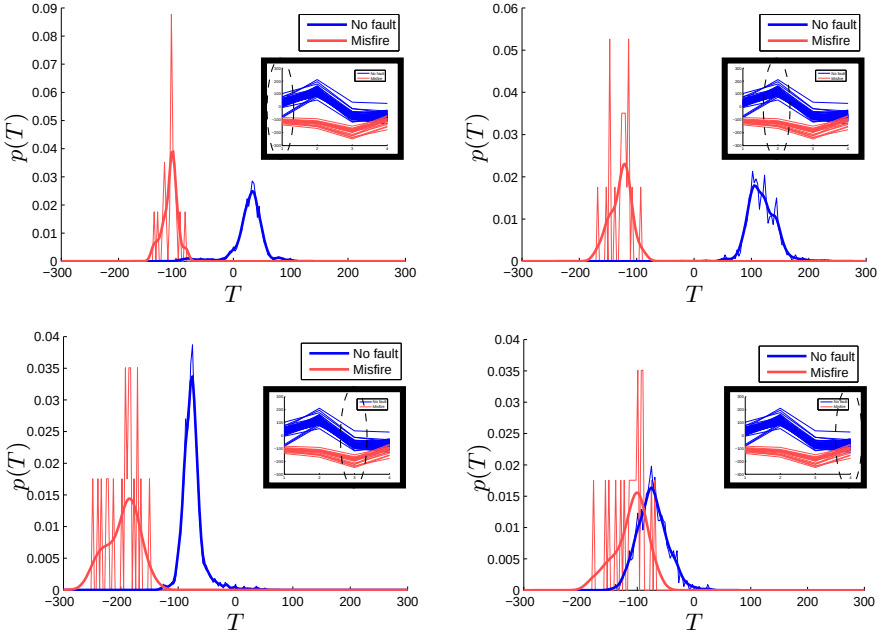

Figure 11: The distribution of the torque of cylinder 1, i.e., crank counts 1, 2, 3, and 4, during normal driving at operating point speed 1500 rpm and load 1.2 g/rev. A normalized histogram of the data and an approximation of each distribution using kernels is shown in each plot. The separation between torque from a fault-free combustion and a misfire is largest at the second sample and smallest at the last sample.

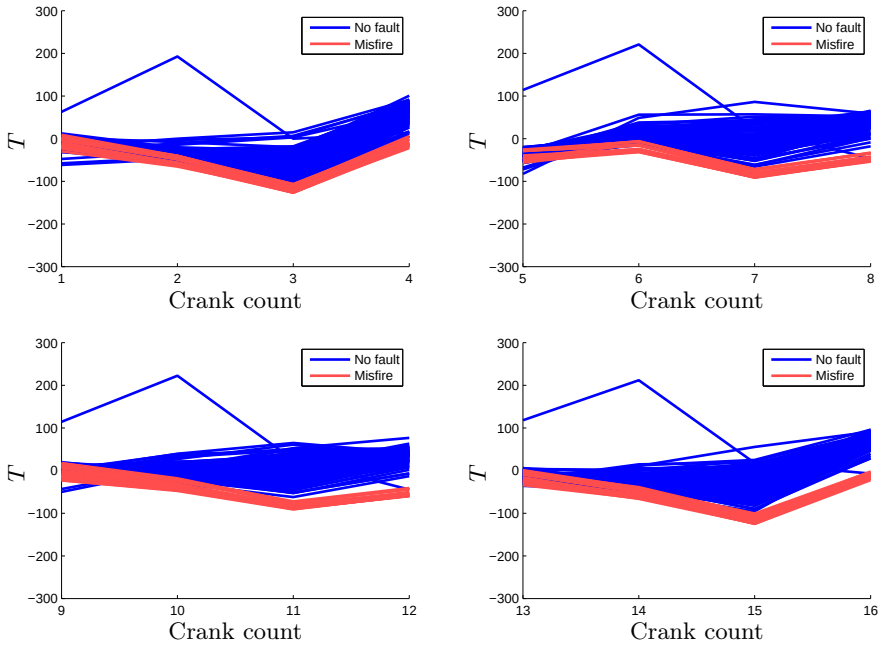

Figure 12: Estimated torque of cylinder 1, 5, 3, and 6 during cold start at operating point speed 1500 rpm and load 1.2 g/rev.

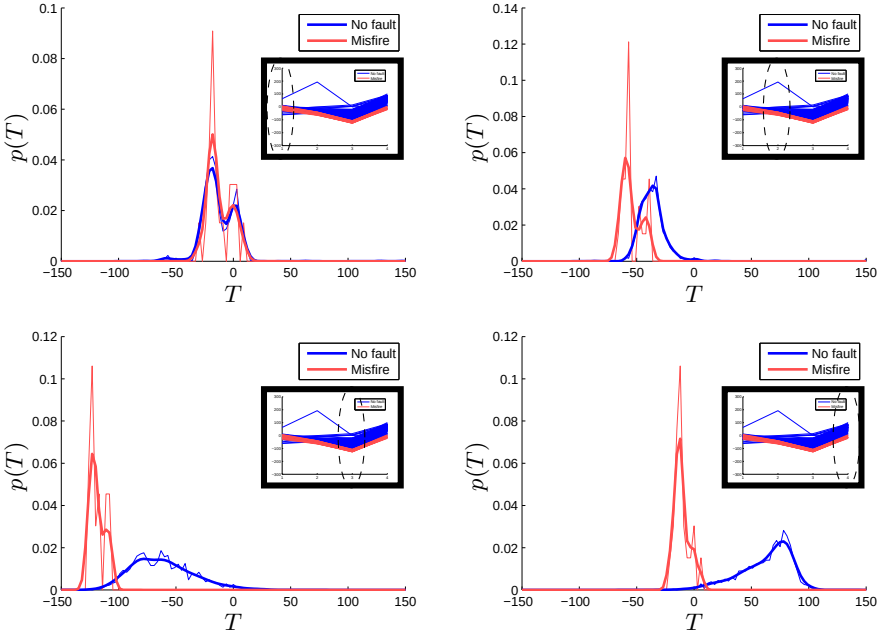

Figure 13: The distribution of the torque of cylinder 1, i.e., crank counts 1, 2, 3, and 4, during cold start at operating point speed 1500 rpm and load 1.2 g/rev. A normalized histogram of the data and an approximation of each distribution using kernels is shown in each plot. The separation between torque from fault-free combustion and a misfire is largest at the last sample and smallest at the first sample.

Analyzing signal distribution using the Kullback-Leibler divergence

Here, the Kullback-Leibler divergence is computed by approximating the fault-free data and misfire data as Gaussian distributed and using (3) to quantify the separation. The results are shown in Figure 14 for normal driving and in Figure 15 for cold starts. Comparing the normalized histograms of the data and the approximated Gaussian distributions shows that the approximation is not always representative of the data. However, a value of Kullback-Leibler divergence around 50 seems to correspond to well separated distributions but already around 20 the distributions are fairly separated such that few or no data from the fault-free case and misfire case are overlapping.

An interesting aspect is to analyze the gain by using the estimated torque instead of the original flywheel angular velocity signal. The computed Kullback-Leibler divergence in Figure 14 is compared to the Kullback-Leibler divergence of the corresponding flywheel angular velocity signal shown in the upper left plot in Figure 6. The distributions and computed Kullback-Leibler divergence are presented in Figure 16. The Kullback-Leibler divergence are all below 0.5 for all samples which is expected since the approximated distributions are similar. The differences in Kullback-Leibler divergence indicate that the estimated torque is more suitable to detect misfires compared to the original flywheel angular velocity signal.

Computing the Kullback-Leibler divergence, for example, for different cylinders or operating points, can be used to get an overview of how misfire detectability varies for different operating points. As an example, data from a city driving scenario in Gothenburg, Sweden, is used to compute the Kullback-Leibler divergence in Table 5. The Kullback-Leibler divergence is computed for fault-free data and misfire data for cylinder 1 and crank count 2 for different operating points. If data is not available from both the fault-free case and misfire case, the Kullback-Leibler divergence is not computed which is represented by an empty field. The computed Kullback-Leibler divergence is above 20 for all cases and increases for high speeds and high loads. The results indicate that it should be more difficult, in general, to detect a misfire at low loads and speeds around 1000 rpm compared to high speeds and loads.

Another example, using the same driving scenario, but estimated torque for cylinder 6 and crank count 14 is shown in Table 6. The Kullback-Leibler divergence is lower at low speeds and loads compared to Table 5 but increases for higher loads and speeds. The result from the two tables indicate that it should be in general more difficult to detect a misfire at low speed and load but easier at high speed and load.

In this section, the Kullback-Leibler divergence has been used to analyze the estimated torque to be used for misfire detection. The estimated torque is compared to the original flywheel angular velocity signal and the results show that the estimated torque is more suitable to use because it gives a higher value of Kullback-Leibler divergence. The Kullback-Leibler divergence is a useful

tool to get an overview of how misfire detectability varies during a combustion and for different operating points. The analyzes have shown that, for example, depending on if it is during normal driving or a cold start, different samples during the combustion should be used to detect a misfire.

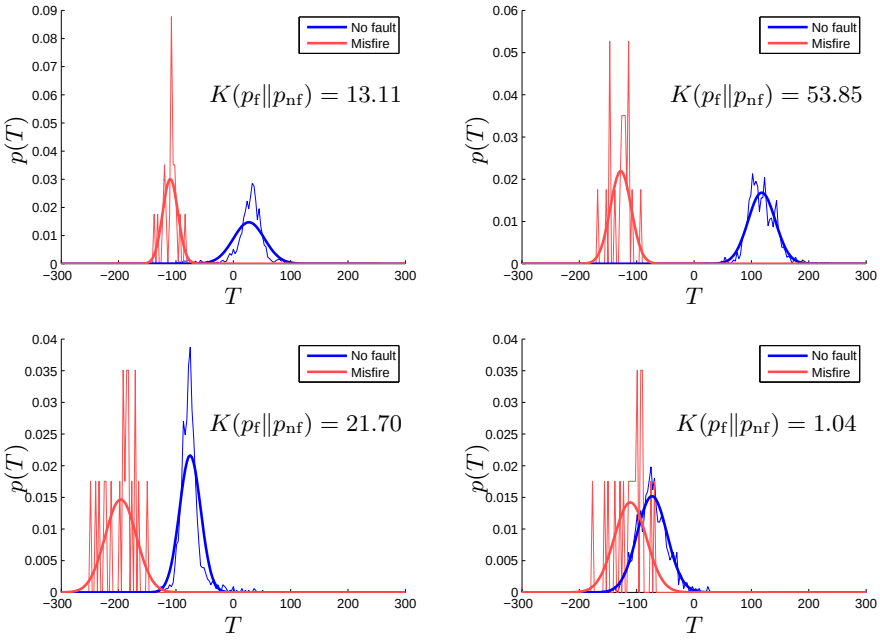

Figure 14: The distribution of the estimated torque of cylinder 1, i.e., crank counts 1, 2, 3, and 4, during normal driving at operating point speed 1500 rpm and load 1.2 g/rev. A normalized histogram of the data and an approximation of each distribution as a Gaussian is shown in each plot.

Table 5: Computed Kullback-Leibler divergence of Gaussian approximations of data from cylinder 1 (crank count 2) for different operating points.

Load	Speed	[rpm]				
[g/rev]	500	1000	1500	2000	2500	3000
0.3		20.94	39.60			
0.4	51.95	20.30	28.21			
0.6	31.05	24.48	32.64	37.81		
0.8		27.54	44.22	38.11		
1.2			53.85	72.69		
1.6			94.76	96.45		
2.0				108.43		

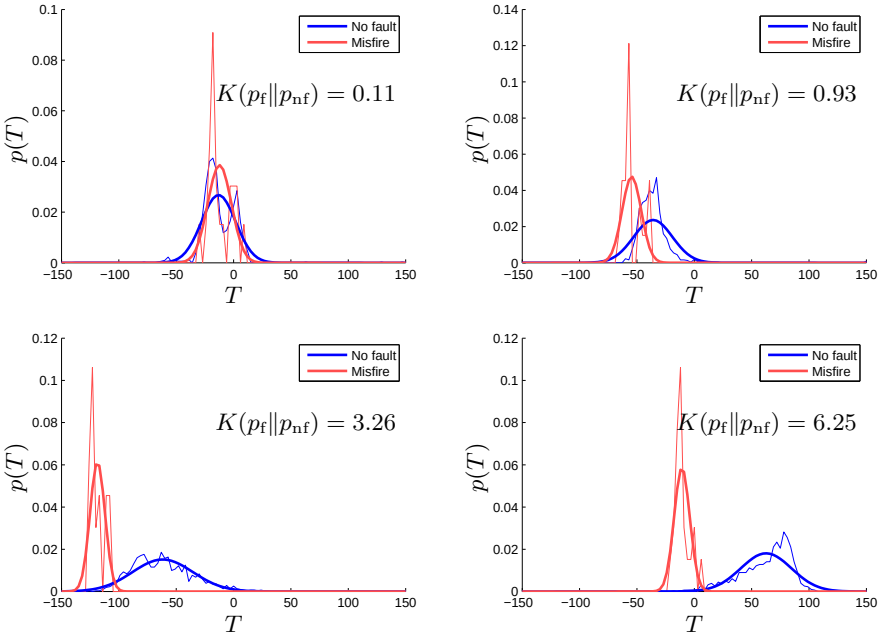

Figure 15: The distribution of the estimated torque of cylinder 1 during cold start at operating point speed 1500 rpm and load 1.2 g/rev. A normalized histogram of the data and an approximation of each distribution as a Gaussian is shown in each plot.

Table 6: Computed Kullback-Leibler divergence of Gaussian approximations of data from cylinder 6 (crank count 14) for different operating points.

Load	Speed	[rpm]				
[g/rev]	500	1000	1500	2000	2500	3000
0.3		9.57	32.22			
0.4	33.17	12.73	23.20			
0.6	15.45	25.50	30.91	31.70		
0.8		31.07	46.54	32.34		
1.2			54.62	80.84		
1.6		27.44	69.84	124.20		
2.0			162.28	117.68	151.52	

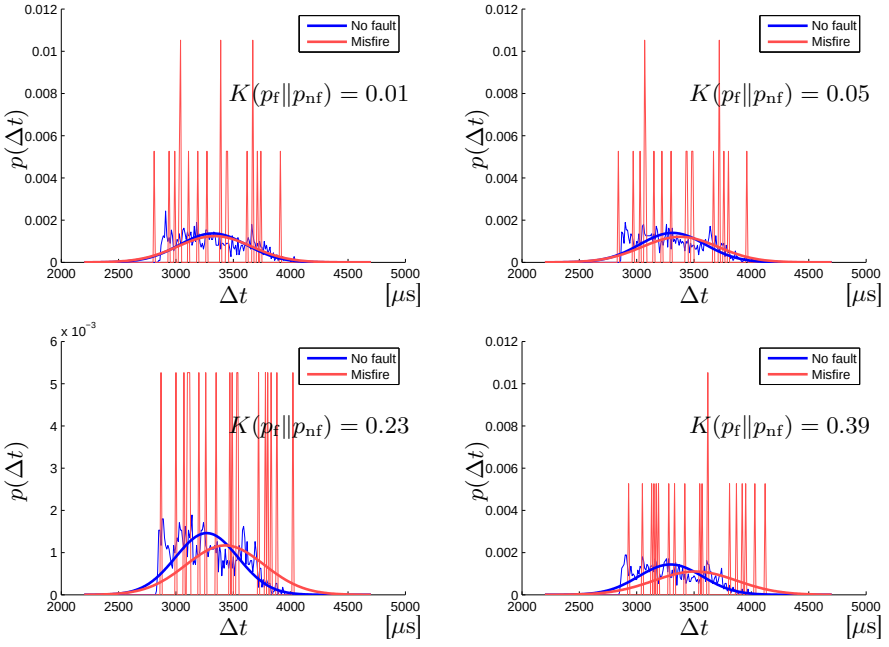

Figure 16: The distribution of the flywheel signal of cylinder 1 during normal driving at operating point speed 1500 rpm and load 1.2 g/rev. A normalized histogram of the data and an approximation of each distribution as a Gaussian is shown in each plot.

6 An algorithm for misfire detection

The flywheel angular velocity signal is used to estimate the torque at the flywheel. The Kullback-Leibler divergence has been used to show how misfire detectability performance depends on speed and load, and that fault-free data and misfire data are more separated at different samples depending on if it is during a cold start or not.

Here, a proposal for a misfire detection algorithm is presented based on the estimated torque. When designing the misfire detection algorithm, one goal has been to keep the solution simple in order to simplify parameter tuning. Another purpose is to let different steps in the design procedure be as independent from each other as possible. In that way, each step can be improved or replaced without the need of changing or removing other steps when tuning or running the algorithm.

6.1 Algorithm outline

Here, an outline of the misfire detection algorithm is presented. The algorithm can be summarized in the following steps.

1. Estimate the torque from the flywheel angular velocity signal.
2. Compute a test quantity using weights for data at different operating points.
3. Threshold the test quantity to detect a misfire.

The algorithm is designed such that all parameterization are automated, i.e., parameters and thresholds, are set automatically given training data. Step 1 of the algorithm has been described previously in Section 5. Step 2 is presented in Section 6.2 and step 3 is presented in Section 6.3.

6.2 Design of test quantity

Since the distribution of the estimated torque varies for different speeds and loads, the estimated torque from each operating point is considered separately. Also, estimated torque from normal driving and from cold starts are considered separately.

Step 1 of the algorithm described in Section 6.1, is to estimate the torque at the flywheel from the flywheel angular velocity signal. The torque at the flywheel is estimated using (7). The estimated torque from one combustion is classified to the closest operating point, see Table 4, given the value of the speed and load variables from the vehicle control system at the last sample of the combustion.

In this report, an adaptive compensation of flywheel manufacturing errors has not been considered. The difference in distribution of estimated torque for different cylinders at the same operating point is probably partly caused

by flywheel errors. If available, a flywheel error compensation can be included in the first step of the algorithm. Here, to handle the errors which affect the torque estimation, each cylinder is considered separately. For the given number of operating points based on speed and load, the number of cylinders, and the catalytic converter heating flag will result, in total of $42 * 5 * 2 = 420$, modes to be parameterized for a five cylinder engine. However, if an adaptation of the estimated torque is included, such that the distribution is compensated for all cylinders, the number of modes would reduce by a factor five to 84.

In Section 5.1, the separation between the distributions of fault-free data and misfire data have been compared for each sample separately. One simple approach to design a test quantity is to take the one data point where the separation between the distributions are maximally separated. However as shown in Section 5.1, a misfire is more or less visible in all samples related to the cylinder depending on how much the misfire and fault-free distributions are separated. Instead of looking at each sample independently, the different samples can be weighted together such that a good misfire detectability performance is achieved. In Figure 17, the three last samples of the estimated torque, denoted T_2, T_3, and T_4, shown in Figure 11 and Figure 13, are plotted in a three dimensional space. The three dimensional distributions of estimated torque during cold starts and normal driving are marked in the figure. The two cases are clearly distinct from each other, thus requiring different weights to detect misfires for each case.

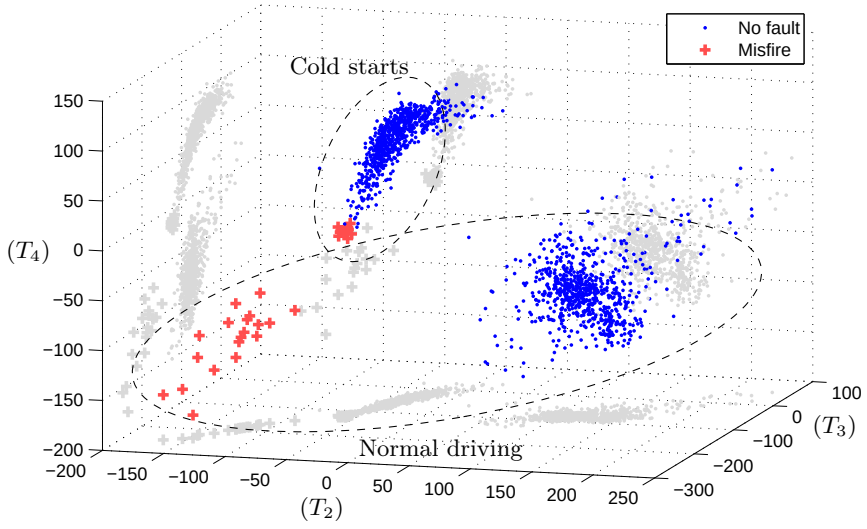

Figure 17: Estimated torque for three of the four samples shown in Figure 11 and Figure 13. The estimated torque at crank count 2 is T_2, the estimated torque at crank count 3 is T_3, and the estimated torque at crank count 4 is T_4.

There are many ways to utilize all samples from one combustion to find a test quantity where the distribution of fault-free data and misfire data are as separated as possible, e.g., data-driven methods, see Bishop (2006). An implementation of a misfire detection algorithm where computational power is limited, requires that the complexity of the misfire detection performed on-line should be as low as possible. To reduce complexity of the misfire detection algorithm, a linear weighting function is applied when designing the test quantity. During each combustion, the torque is estimated four times, $\{T_1, T_2, T_3, T_4\}$. A test quantity y for detecting misfires is designed using linear weights as

$$y = \sum_{i=1}^{4} \alpha_i T_i \tag{8}$$

where α_i is the weight for the estimated torque T_i and $\|\alpha\|^2 = 1$. The weights $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ are chosen such that the value of y during the fault-free case and during misfire are as separated as possible.

Finding the weights $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ can be formulated as an optimization problem

$$\begin{aligned} &\max_{\alpha} M(p_{\mathrm{f}}, p_{\mathrm{nf}}; \alpha) \\ &\text{s.t. } \|\alpha\| = 1 \end{aligned} \tag{9}$$

where $M(p_{\mathrm{f}}, p_{\mathrm{nf}}; \alpha)$ is a measure of the separation between the distributions of y for fault-free data p_{nf} and misfire data p_{f} for a given α. Since p_{nf} and p_{f} are unknown, the probability density functions are approximated using training data.

As an example to visualize the weights of the vector α, the direction of one vector α is plotted together with the training data. The training data in Figure 17 is separated in two sets shown in Figure 18, the upper plot contains data from normal driving and the lower plot contains data from cold starts since these two cases have clearly separated distributions. The outlier in the cold start data is the same as the outlier in Figure 12. As an example, the solid lines represents the vector of the weights α for each data set which are parallel to the direction in which the distributions are separated.

In the upper plot in Figure 18 a threshold, exemplified by a dashed line, can easily be chosen such that a classification can be made with low risk of missed detection or false alarms. In the lower plot, the distributions are more overlapped and it is not possible by linear weights α to find a threshold which completely separates the data from the two cases. The weights α are different for normal driving and cold starts. Note that the weights α are largest for samples where the distributions where separated the most in Figure 14 and Figure 15, for example α_2 and α_3 for normal driving data and α_3 and α_4 for cold start data.

Here follows two proposed measures $M(p_{\mathrm{f}}, p_{\mathrm{nf}}; \alpha)$ that can be used to find a suitable α.

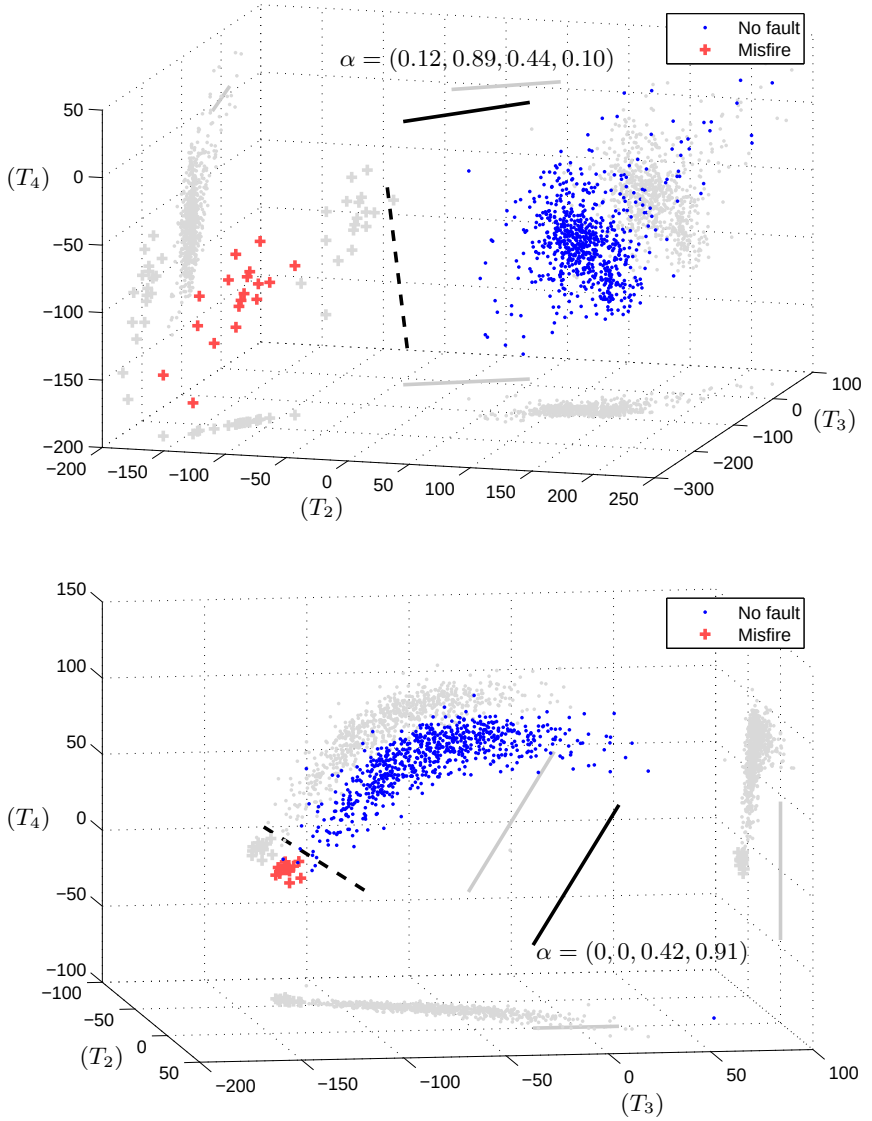

Figure 18: The data in Figure 17 is separated in normal driving, the upper plot, and cold starts, the lower plot. The solid lines represent the weights α such that the distributions of fault-free data and misfire data are well separated. Examples of thresholds to separate fault-free data and misfire data are shown as dashed lines.

Maximizing the Kullback-Leibler divergence

One choice of $M(p_{\mathrm{f}}, p_{\mathrm{nf}}; \alpha)$ could be to find weights α which maximizes the Kullback-Leibler divergence, i.e,

$$M(p_{\mathrm{f}}, p_{\mathrm{nf}}; \alpha) = K(p_{\mathrm{f}}(\alpha) \| p_{\mathrm{nf}}(\alpha)) \tag{10}$$

where $p_{\mathrm{f}}(\alpha)$ and $p_{\mathrm{f}}(\alpha)$ represents the distributions of y for the misfire case and fault-free case respectively for the given weights α. By maximizing (10), the distributions are optimally separated given the definition of Kullback-Leibler divergence.

As an example, optimized weights α which maximizes the Kullback-Leibler divergence for cylinder 1 at different operating points are presented in Figure 19. The weights are optimized given data from a driving scenario in Gothenburg. Each plot represents one sample for the different operating points. The Kullback-Leibler divergence is computed by estimating the distributions $p_{\mathrm{f}}(\alpha)$ and $p_{\mathrm{nf}}(\alpha)$ using the kernel density estimator in (2).

Comparing the weights for each sample in Figure 19 shows that the weights for each sample varies for different operating points. The second sample (crank count 2) has a high weight for most of the operating points while the other three samples (crank counts 1, 3, and 4) varies around zero. The misfire detection algorithm is evaluated when optimizing the weights using the Kullback-Leibler divergence in the next section.

Maximizing the gap between distributions

Let y_{nf} denote the vector of all fault-free data of the test quantity and y_{f} all misfire data.

$$M(p_{\mathrm{f}}, p_{\mathrm{nf}}; \alpha) = \begin{cases} \min(y_{\mathrm{nf}}) - \max(y_{\mathrm{f}}) & \text{if } \min(y_{\mathrm{nf}}) > \max(y_{\mathrm{f}}) \\ \min(y_{\mathrm{f}}) - \max(y_{\mathrm{nf}}) & \text{if } \max(y_{\mathrm{nf}}) < \min(y_{\mathrm{f}}) \end{cases} \tag{11}$$

Maximizing the gap is also applied in support vector machines to find planes which separate different classes given training data, see Bishop (2006). Note that maximizing the gap is sensitive to outliers since it only considers the closest data points of the two distributions.

In Figure 20, optimized weights are shown which maximizes the gap between distributions of the fault-free data and misfire data for cylinder 1 at different operating points. The data used here is the same data as in Figure 19. Comparing the optimized weights in Figure 20 and Figure 19, shows that the weights appear to vary less when maximizing the gap compared to maximizing the Kullback-Leibler divergence. The misfire detection algorithm is evaluated when optimizing the weights such that the gap is maximized in the next section.

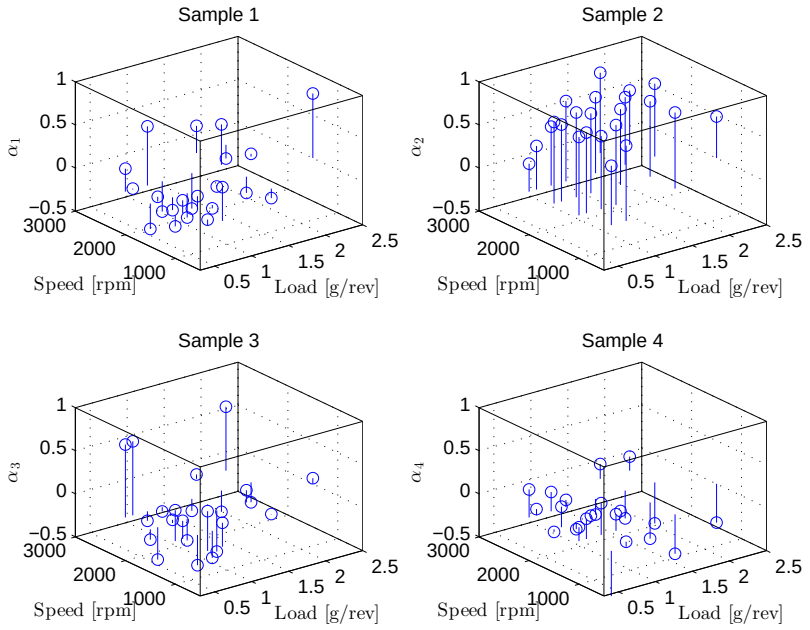

Figure 19: Optimized weights α for samples from cylinder 1 which maximizes the Kullback-Leibler divergence from the misfire distribution to the fault-free distribution.

6.3 Thresholding

The simplest and most common approach to detect a change in a test quantity y is the use of a fixed threshold J such that

$$\begin{aligned} y > J \quad & \text{concludes no misfire} \\ y \leq J \quad & \text{concludes misfire.} \end{aligned}$$

Depending on the fault-free and misfire distributions of the test quantity, the threshold are chosen to balance the risk of missed detections and false alarms.

By using training data to approximate the distributions of fault-free data and misfire data, thresholds can be automatically selected to balance the required performances.

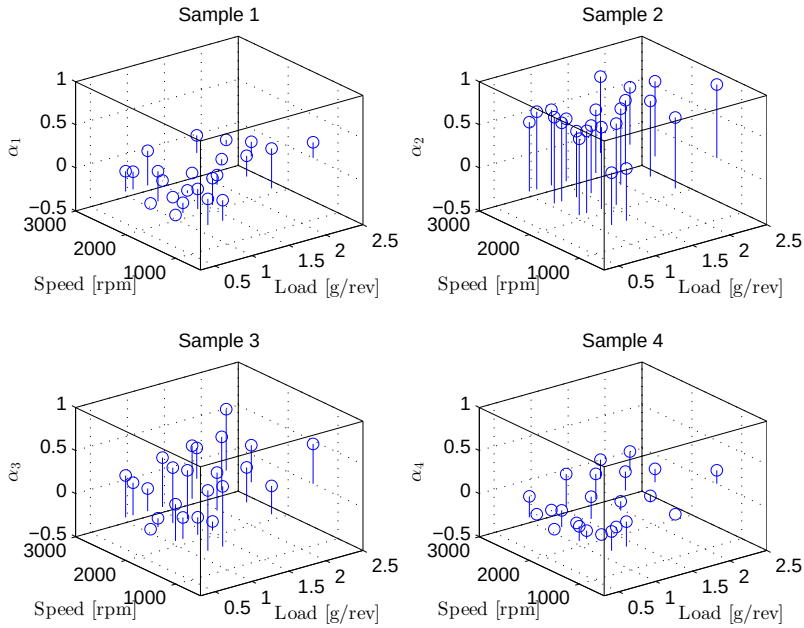

Figure 20: Optimized weights α for samples from cylinder 1 which maximizes the gap between the fault-free and misfire distributions.

7 Evaluation of the misfire detection algorithm

The purpose of this evaluation is to see how well separated misfire data and fault-free data are for a designed test quantity. One goal is to see if it is possible to select thresholds such that all misfires can be detected with few false alarms.

A test quantity is designed by estimating the torque from the flywheel angular velocity signal using (7). By using the speed and load variables, the crank counter, and the catalyst warming flag, from the vehicle control system, the estimated torque is associated to a specific operating point, cylinder, and if it is during a cold start or not. For each case, a vector α in (8) which solves (9), has been optimized using Matlab and an interior-point search algorithm, see Byrd et al. (2000). The distributions of the test quantity are estimated using a set of training data. Two versions of the misfire detection algorithm are evaluated using two different sets of vectors α. The first version has weights α such that the Kullback-Leibler divergence (10) is maximized and the second version has weights α such that the gap between fault-free data and misfire data (11) is maximized. Note that a misfire is detected by analyzing samples associated to one specific cylinder. Thus, identification of the misfiring cylinder is obtained automatically.

Constant thresholds, which decides when a misfire is detected, are selected for each operating point, cylinder, and state of the catalyst warming flag. Parameterization of the algorithm is made given a set of training data in the following order.

1. Estimate torque using (7).

2. Compute vectors α for the test quantity y in (8) for each operating point using (9).

3. Compute thresholds which fulfills required probabilities for false alarm and missed detection.

All parameterization is performed automatically using training data and the only necessary manual tuning is to choose probabilities for false alarm and missed detection.

To be sure that there is a sufficiently large set of training data for all operating points in the validation data, the validation data is included in the training data. Thresholds are chosen just above the highest value of the test quantity for any misfire in the training data, i.e., all misfires in the training data are detected. This choice of the thresholds is used to analyze how separated the distributions of misfire data and fault-free data are, by looking at the number of false alarms. If the number of false alarms is high when all misfires are detected then it is not possible to reduce the number of false alarms without increasing the number of missed detections.

For a six cylinder engine, the misfire detection algorithms are evaluated using one normal driving scenario and three cold start scenarios. The two versions of the misfire detection algorithm are compared to Volvo's misfire detection algorithm which is used as a reference where the number of detected true misfires, false alarms and missed detections are compared. Note that Volvo's misfire detection algorithm is the implemented algorithm for the vehicle and is not tuned using the training data as for the other algorithms. The results for each scenario are shown in Tables 7-10. Table 7 summarizes the result from the normal driving scenario. The two versions of the algorithm have no false alarms and an example of the test quantities and thresholds are shown in Figure 21 and Figure 22 respectively. In the cold starts scenarios, there are some false alarms for both versions of the algorithm which is inevitable since the fault-free data and misfire data are overlapping, as shown in Figure 18, and can not be completely separated. The results in Tables 8-10 show that both maximizing the Kullback-Leibler divergence and maximizing the gap give a better performance than the reference algorithm in almost all cases. However, maximizing the Kullback-Leibler divergence seems to give a slightly better performance compared to maximizing the gap.

Examples of the test quantities and thresholds when maximizing the Kullback-Leibler divergence and maximizing the gap are shown in Figure 21 and Figure 22 respectively. The figures show how the threshold varies depending on speed, load, and cylinder, where a large jump corresponds to a change in operating point and the small variations are different cylinders. The two dips of the test quantity are the result of misfires. Given the data in both Figure 21 and Figure 22, the thresholds can be chosen higher without significantly increasing the risk of false alarms. This is also visible when looking at the separation between the distributions of the test quantities shown in Figure 23 and Figure 24 which gives examples of the distributions for two operating points and two cylinders. The thin lines are normalized histograms of the fault-free data and misfire data and the thick lines are approximated distributions using (2). The test quantities during cold start when maximizing the Kullback-Leibler divergence and maximizing the gap are shown in Figure 25 and Figure 26 respectively. The distributions of the test quantities are shown in Figure 27 and Figure 28. Misfires during cold starts are most difficult to detect at the operating point load 1.2 g/rev and speed 1500 rpm but the distributions are better separated compared to looking at each sample individually, see Figure 12.

The evaluation of the two versions of the misfire detection algorithm for a five cylinder engine is shown in Table 11 and Table 12. The algorithms are evaluated using two driving scenarios. No cold start scenarios where available for the five cylinder engine and the performance in this case has not been evaluated. The number of false alarms for the two versions of the misfire detection algorithm are fewer compared to the reference algorithm.

The evaluations show some promising results that the misfire detection algorithms proposed in this work are able to detect all misfires while having

few false alarms. Two approaches to choose the weights α are compared and they seem to give similar performance considering number of false alarms when detecting all misfires.

Table 7: Results for a six cylinder engine from a normal driving scenario. The number of combustions during the scenario is 60356.

	Max. K-L	Max. gap	Ref.
Detected true misfires	1021	1021	871
Missed detections	0	0	150
False alarms	0	0	58

Table 8: Results for a six cylinder engine from a cold start scenario at $7°C$. The number of combustions during the scenario is 6137.

	Max. K-L	Max. gap	Ref.
Detected true misfires	127	127	61
Missed detections	0	0	66
False alarms	9	13	19

Table 9: Results for a six cylinder engine from a cold start scenario at $21°C$. The number of combustions during the scenario is 3962.

	Max. K-L	Max. gap	Ref.
Detected true misfires	82	82	14
Missed detections	0	0	68
False alarms	13	19	16

Table 10: Results for a six cylinder engine from a cold start scenario at $25°C$. The number of combustions during the scenario is 3251.

	Max. K-L	Max. gap	Ref.
Detected true misfires	66	66	10
Missed detections	0	0	56
False alarms	7	6	11

Table 11: Results for a five cylinder engine from a normal driving scenario. The number of combustions during the scenario is 122283.

	Max. K-L	Max. gap	Ref.
Detected true misfires	953	953	945
Missed detections	0	0	8
False alarms	9	24	78

Table 12: Results for a five cylinder engine from a normal driving scenario. The number of combustions during the scenario is 10339.

	Max. K-L	Max. gap	Ref.
Detected true misfires	2	2	2
Missed detections	0	0	0
False alarms	21	18	23

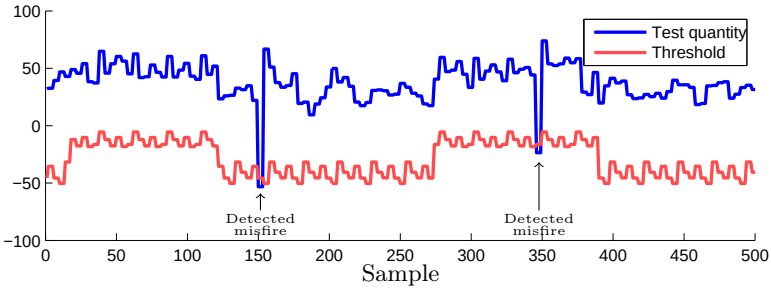

Figure 21: An example of the test quantity and threshold, when maximizing the Kullback-Leibler divergence, for a six cylinder engine during normal driving. The threshold variations depends on operating point and cylinder.

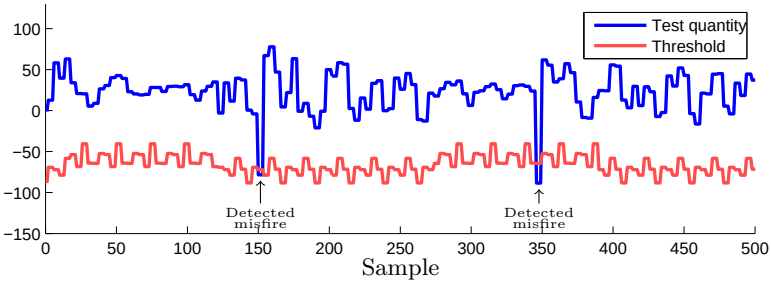

Figure 22: An example of the test quantity and threshold, when maximizing the gap, for a six cylinder engine during normal driving. The threshold variations depends on operating point and cylinder.

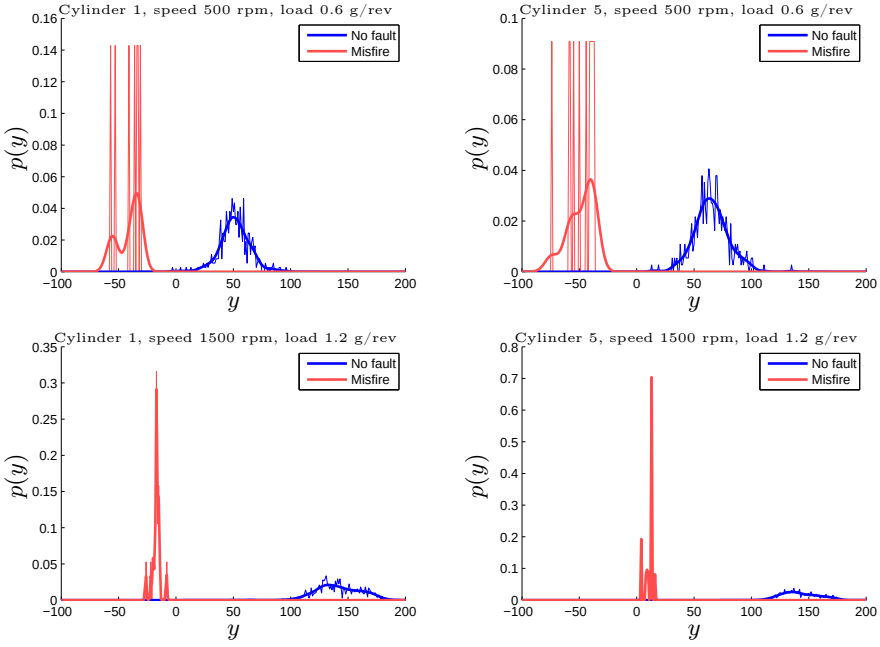

Figure 23: The distribution of the test quantity for different loads and speeds when maximizing the Kullback-Leibler divergence.

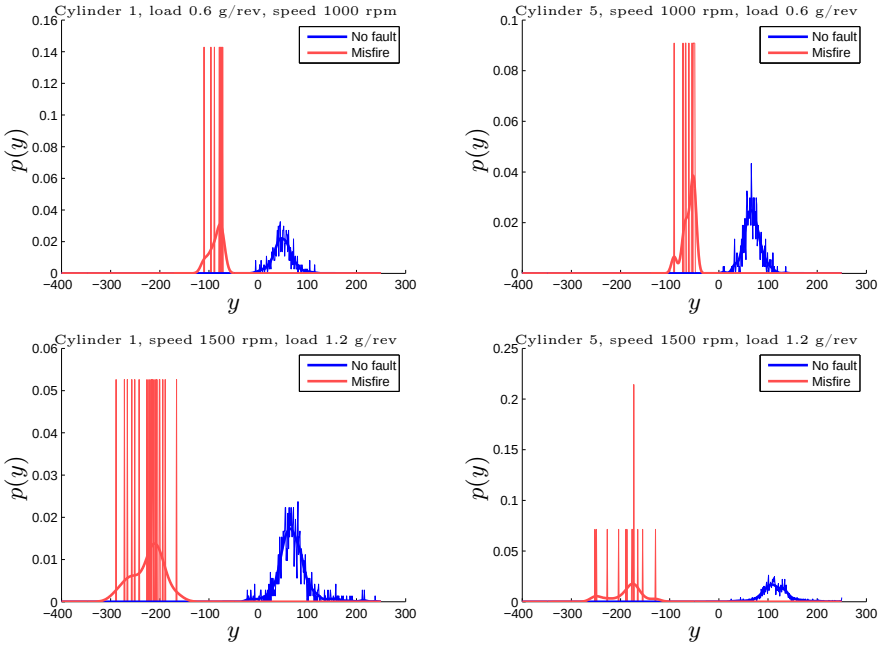

Figure 24: The distribution of the test quantity for different loads, speeds, and cylinders, when maximizing the gap.

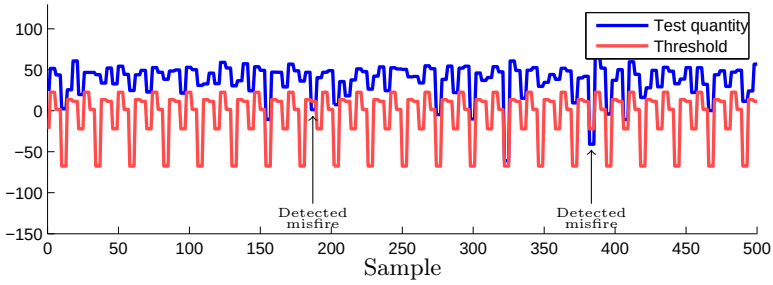

Figure 25: An example of the test quantity and threshold, when maximizing the Kullback-Leibler divergence, for a six cylinder engine during cold start. The threshold variations in the figure depend on the different cylinders.

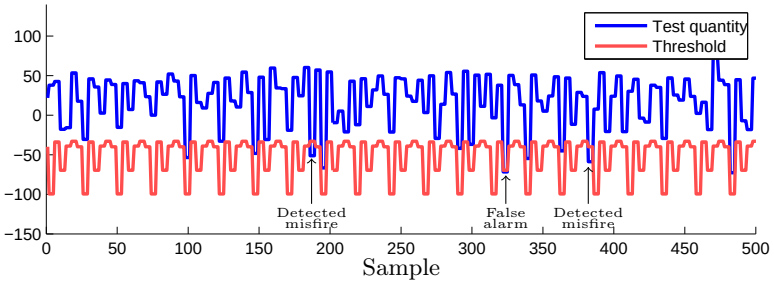

Figure 26: An example of the test quantity and threshold, when maximizing the gap, for a six cylinder engine during cold start. The threshold variations in the figure depend on the different cylinders.

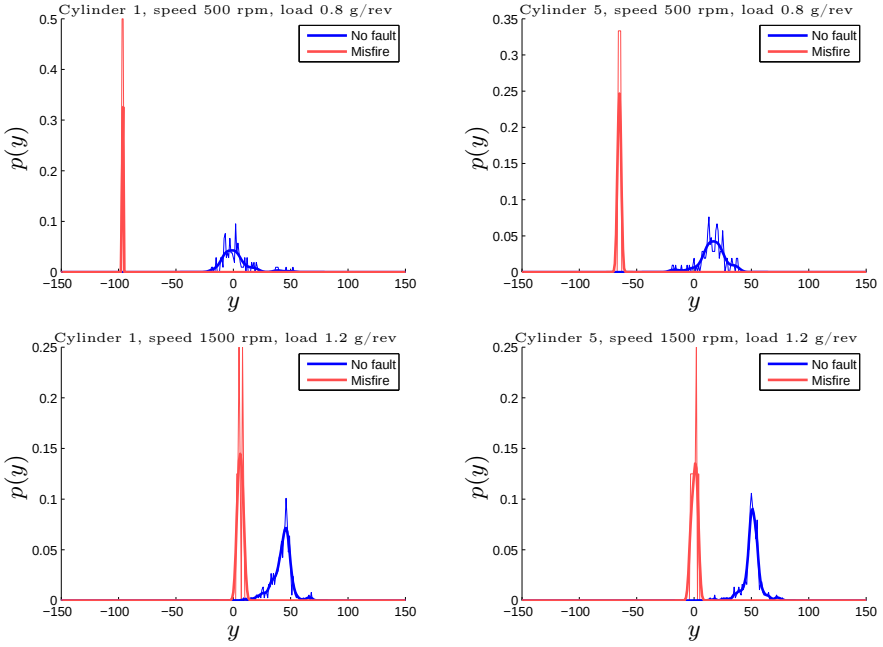

Figure 27: The distribution of the test quantity during cold start for different loads and speeds when maximizing the Kullback-Leibler divergence.

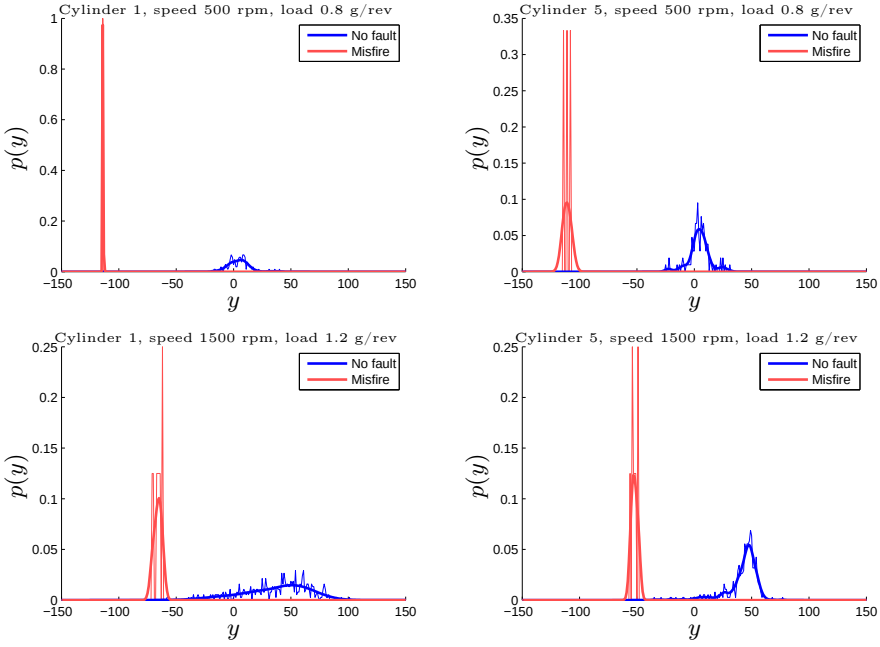

Figure 28: The distribution of the test quantity during cold start for different loads, speeds, and cylinders, when maximizing the gap.

8 Conclusions

A misfire detection algorithm is proposed where the torque at the flywheel is estimated based on the flywheel angular velocity signal. The estimated torque is associated to different operating points depending on speed, load, cylinder, and the catalytic converter warming flag in the vehicle control system. A test quantity is designed by weighting data from each combustion such that the distributions of the test quantity for fault-free data and misfire data are separated. Then, thresholds for detecting misfires are selected for each operating point which balances the probabilities of false alarms and missed detections. All tuning steps are performed automatically given a set of training data which simplifies the tuning of the algorithm.

As a tool for designing and evaluating the misfire detection algorithm, the Kullback-Leibler divergence is proposed. The Kullback-Leibler divergence quantifies the separation between two distributions, in this case describing fault-free data and misfire data, and gives information of when it is easy to detect a misfire. Using the Kullback-Leibler divergence has the advantage that the separation between the distributions is summarized into a scalar value which makes it possible to easily get an overview of the performance of several operating points at the same time. The Kullback-Leibler divergence has proven useful during the whole design process, from analyzing the misfire detectability performance for different samples and operating points to optimizing algorithm parameters.

Two versions of the proposed misfire detection algorithm are evaluated using real data from both a five cylinder engine and a six cylinder vehicle. The difference is how the weights are chosen. In the first version, the weights are chosen such that the Kullback-Leibler divergence is maximized from the misfire data to the fault-free data. In the second version, the weights are chosen such that the gap is maximized between the distributions of the test quantity for the fault-free data and misfire data. The performance of the algorithms are compared to a reference algorithm and evaluations show that the proposed algorithm is able to detect all misfires while having a low probability of false alarms.

9 Future works

In this work, flywheel manufacturing errors that affect the flywheel angular velocity signal are not considered. The errors are different for different vehicles which will affect the results of the misfire detection algorithm. A compensation for these errors should be included in the misfire detection algorithm before it can be used in a production vehicle.

Even though all parameter tuning is performed automatically there are many parameters in the misfire detection algorithm. Both weights for the estimated torque and thresholds at different operating points but also for different cylinders. The number of parameters can possibly be reduced by compensating for the

flywheel manufacturing errors. It should be analyzed if the number of parameters can be reduced without significantly degrading the performance of the misfire detection algorithm.

10 Acknowledgment

This work has been supported by Volvo Cars. The authors would like to thank Sasa Trajkovic at Volvo Cars for helping us with the data collection.

References

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

Richard H. Byrd, Jean Charles Gilbert, and Jorge Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89:149–185, 2000.

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Resource Center, Pacific Grove, CA, 2001.

Francis T. Connolly and Giorgio Rizzoni. Real time estimation of engine torque for the detection of engine misfires. *Journal of Dynamic Systems, Measurement, and Control*, 116(4):675–686, 1994.

Shinto Eguchi and John Copas. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *J. Multivar. Anal.*, 97:2034–2040, October 2006.

Daniel Eriksson, Mattias Krysander, and Erik Frisk. Using quantitative diagnosability analysis for optimal sensor placement. Proceedings of IFAC Safe Process 2012, Mexico city, Mexico, 2012.

Daniel Eriksson, Erik Frisk, and Mattias Krysander. A method for quantitative fault diagnosability analysis of stochastic linear descriptor models. *Automatica*, Accepted for publication.

J.B. Heywood. *Internal combustion engine fundamentals*. McGraw-Hill series in mechanical engineering. McGraw-Hill, 1988.

Ken-jen Lang, Lela Liu, Alec L. Lang, and Louis Yizhang Liu. Multi-Purpose Flywheel (MPF) and Misfire Detection. *SAE Technical Paper 2005-01-1141*, 2005.

U. Kiencke. Engine misfire detection. *Control Engineering Practice*, 7(2):203 – 208, 1999.

Kullback, S. and Leibler, R. A. On Information and Sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

D. Lundström and S. Schagerberg. Misfire Detection for Prechamber SI Engines Using Ion-Sensing and Rotational Speed Measurements. *SAE Technical Paper 2001-01-0993*, 2001.

J Mohammadpour, M Franchek, and K Grigoriadis. A survey on diagnostic methods for automotive engines. *International Journal of Engine Research*, 13 (1):41–64, 2012.

Sanjeev Naik. Advanced misfire detection using adaptive signal processing. *International Journal of Adaptive Control and Signal Processing*, 18(2):181–198, 2004.

Andrew W. Osburn, Theodore M. Kostek, and Matthew A. Franchek. Residual generation and statistical pattern recognition for engine misfire diagnostics. *Mechanical Systems and Signal Processing*, 20(8):2232 – 2258, 2006.

Francisco V. Tinaut, Andrés Melgar, Hannes Laget, and José I. Domínguez. Misfire and compression fault detection through the energy model. *Mechanical Systems and Signal Processing*, 21(3):1521 – 1535, 2007.

Andreas Walter, Uwe Kiencke, Stephen Jones, and Thomas Winkler. Misfire Detection for Vehicles with Dual Mass Flywheel (DMF) Based on Reconstructed Engine Torque. *SAE Technical Paper 2007-01-3544*, 2007.