Preprints, 9th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes
September 2-4, 2015. Arts et Métiers ParisTech, Paris, France

# Quantitative isolability analysis of different fault modes

**Daniel Jung, Erik Frisk, and Mattias Krysander**

*Linköping University, Linköping, Sweden (e-mail: daner, frisk, matkr@isy.liu.se).*

**Abstract:** To be able to evaluate quantitative fault diagnosability performance in model-based diagnosis is useful during the design of a diagnosis system. Different fault realizations are more or less likely to occur and the fault diagnosis problem is complicated by model uncertainties and noise. Thus, it is not obvious how to evaluate performance when all of this information is taken into consideration. Four candidates for quantifying fault diagnosability performance between fault modes are discussed. The proposed measure is called *expected distinguishability* and is based of the previous *distinguishability* measure and two methods to compute expected distinguishability are presented.

*Keywords:* Fault detection and isolation, quantitative diagnosability analysis, Kullback-Leibler divergence.

## 1. INTRODUCTION

In model-based diagnosis, there are several measures for analyzing fault diagnosability performance given a mathematical model of the system or a set of residuals. Two of the most common model properties related to fault diagnosis are fault *detectability* and *isolability* (Patton et al., 2000; Blanke and Schröder, 2006), which tell whether a fault can be detected or isolated from another fault or not. Methods for analyzing different types of mathematical models of systems are described in for example Frisk et al. (2009) for linear systems or in Krysander (2006); Travé-Massuyes et al. (2006); Bregon et al. (2014) for nonlinear systems using structural analysis. In these previous works, fault detectability and isolability are considered as deterministic properties where model uncertainties, measurement noise, and fault realizations are not taken into consideration. Thus, the deterministic diagnosability analysis only give yes/no answers to the question whether a fault is detectable or isolable or not.

Analysis of fault detectability and isolability performance based on a model is useful during the diagnosis system design, for example the sensor placement problem (Frisk et al., 2009; Sarrate et al., 2012) or to find a suitable set of diagnosis tests (Travé-Massuyes et al., 2006; Bregon et al., 2014). However, if the design methodologies are based on deterministic performance measures, such as fault detectability and isolability, it is not certain that a solution will fulfill the given quantitative fault detection and isolation performance requirements in practice due to model uncertainties and measurement noise. In this case, the achieved performance of the diagnosis system, such as fault-to-noise ratio or probability of detection, is evaluated first when the diagnosis system has been developed. Since model uncertainties and noise will have a negative impact on performance, this should be taken into consideration early in the design process to avoid unnecessary delays during the development.

In Eriksson et al. (2013); Harrou et al. (2014), a quantitative measure of detectability and isolability performance is proposed, called *distinguishability*, which takes fault time profiles and model uncertainties into consideration. The distinguishability measure has been applied to, for example, the sensor placement problem (Eriksson et al., 2012b; Huber et al., 2014) and the test selection problem (Eriksson et al., 2012a). In Huber et al. (2014), distinguishability was used to find a set of sensors for an extended Kalman filter to estimate faults in an IC engine. Other works proposing quantitative diagnosability measures taking uncertainties into consideration are, for example (Bhushan et al., 2008; Wheeler, 2011; Cui et al., 2014).

Distinguishability quantifies how difficult it is to isolate a fault mode $f_i$, with a specific fault time profile, or fault realization, $\theta$ from another fault mode $f_j$ of any fault time profile. Note that different fault time profiles will give different distinguishability values and that a particular fault mode typically covers many different fault profiles. Examples of fault time profiles are constant or ramp faults with different sizes and increase rate. Thus, a single fault mode consists of many, possibly infinitely many, different cases and analyzing each case separately is not feasible. The main objective of this work is to not only consider a specific fault time profile in the analysis, but to compute distinguishability performance for isolating a fault mode $f_i$, taking all possible fault realizations into consideration. Here, probability distributions over the set of fault time profiles are used to describe the set of possible fault profiles. The proposed measure should take the information that different fault time profiles have different probabilities to occur into consideration. Therefore, the goal is to quantify isolability performance

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

Copyright © 2015 IFAC

between fault modes instead of having to select specific fault time profiles.

A question when defining such a performance measure is how to represent the different fault modes and how to quantify the separation between the fault modes. In this work, four different candidates based on the distinguishability measure in Eriksson et al. (2013) are analyzed based on how well they can be used to quantify the separation.

The outline is as follows. First, the problem formulation is discussed using an example in Section 2. Some background theory is summarized in Section 3. Then, a probabilistic modeling of fault modes is presented in Section 4 and the different measure candidates are described in Section 5. The measure candidates are discussed in Section 6, and some additional results regarding computation of one of the measures, expected distinguishability, are presented in Section 7. Finally, a case study is analyzed in Section 8 and some conclusions are given in Section 9.

## 2. PROBLEM FORMULATION

The problem in this work is introduced by first considering the following example.

*Example 1.* A simple, normalized, discrete-time dynamic model of a spring-mass system used in Eriksson et al. (2013) is considered,

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

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 additive faults, and $\varepsilon_i$ 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 $\varepsilon_3 \sim \mathcal{N}(0, 0.5)$. $\mathcal{N}(\mu, \sigma)$ denotes a Gaussian random variable with mean $\mu$ and standard deviation $\sigma$. The faults are modeled as unknown additive signals in the model and represent faults in the control signal, $f_1$, a change in rolling resistance, $f_2$, and sensor biases, $f_3$ and $f_4$.

To analyze the performance using distinguishability (Eriksson et al., 2013), a time window model of length $N = 10$ samples is here selected. It is assumed that for each fault mode $f_i$, the fault can have any non-zero fault magnitude. A fault time profile or fault realization given a fault mode $f_i$ is here referring to a vector $\theta = (\theta[t-N+1], \theta[t-N+2], \ldots, \theta[t])$ describing the fault signal $f_i$ in (1). Distinguishability $\mathcal{D}_{i,j}(\theta)$ is computed for a specific fault time profile, in this case a constant fault with magnitude one, and the result is presented in Table 1. The notation $\bar{\mathbf{1}}_n$ denotes a column vector of length $n$ with ones, for example $\bar{\mathbf{1}}_4 = (1, 1, 1, 1)^{\mathrm{T}}$. A positive value at position $(i, j)$ in Table 1 corresponds to that the fault $f_i = \theta$ is isolable from another fault $f_j$ of any fault time profile. A higher value corresponds to that a fault is easier to isolate, and zero otherwise. Table 1 shows, for example, that it is easier isolate a constant fault $f_1$ from $f_3$ than vice versa, since $6.59 > 3.33$, and the faults $f_1$ and $f_2$ are not isolable from each other.

Table 1. Computed distinguishability of (1) for $\theta = \bar{\mathbf{1}}_{10}$.

| $\mathcal{D}_{i,j}(\theta)$ | NF | $f_1$ | $f_2$ | $f_3$ | $f_4$ |
|---|---|---|---|---|---|
| $f_1$ | 9.07 | 0 | 0 | 6.59 | 3.62 |
| $f_2$ | 9.07 | 0 | 0 | 6.59 | 3.62 |
| $f_3$ | 4.34 | 3.33 | 3.33 | 0 | 3.62 |
| $f_4$ | 7.36 | 3.33 | 3.33 | 6.59 | 0 |

Table 2. Distribution of fault time profiles for different fault modes.

| Fault | $p(\theta)$ |
|---|---|
| $f_1$ | $\theta \sim \mathcal{N}(0, I)$ |
| $f_2$ | $\theta = \bar{\mathbf{1}}_{10}\alpha$ where $\alpha \sim \mathcal{N}(0, 0.5)$ |
| $f_3$ | $\theta = \bar{\mathbf{1}}_{10}\alpha$ where $\alpha \sim \mathcal{N}(1, 0.5)$ |
| $f_4$ | $\theta = \bar{\mathbf{1}}_{10}\alpha$ where $\alpha \sim 0.2\mathcal{N}(0.5, 0.5) + 0.8\mathcal{N}(-1, 0.3)$ |

However, Table 1 only shows distinguishability given a specific fault time profile, here constant faults. Each fault mode $f_i$ is represented by non-zero values of $f_i$ in (1) and the fault-free case is usually when all fault signals are zero. Assume that, for example, during fault mode $f_2$ the fault time profile is almost always constant with a fault magnitude $\alpha$ that is $\mathcal{N}(0, 0.5)$-distributed. Then a probabilistic description of the fault profile $\theta$ is

$$\theta = \bar{\mathbf{1}}_{10}\alpha, \quad \alpha \sim \mathcal{N}(0, 0.5)$$

Table 2 shows an example how the fault modes in Example 1 can be represented using probability distributions. Different fault time profiles of each fault can occur and some are more likely than others. This is taken into consideration by modeling a probability to each fault time profile $\theta$ describing the conditional probability of observing that fault time profile $\theta$ given the system being in the given fault mode. The faults $f_2$, $f_3$, and $f_4$ are assumed constant where the magnitude is random and the magnitude of $f_4$ is modeled as a Gaussian mixture (Hastie et al., 2009). The fault $f_1$ is modeled as additive noise.

The distinguishability measure used in this example does not take the information about the distributions of the fault time profiles in Table 2 into consideration. The result in Table 1 is not representative of the isolability performance of the whole fault mode since it is computed only for a constant fault time profile. Thus, different fault time profiles would result in different distinguishability values compared to Table 2. □

The example shows that it is difficult to get an overview of overall diagnosability performance between fault modes by only using distinguishability for a given fault profile. The probabilities of different fault profiles should be taken into considerations to get a measure which represents the expected performance when taking all fault time profiles, and their probabilities, into considerations.

The purpose of this work is to derive a quantitative fault diagnosability measure candidate to evaluate quantitative detectability and isolability performance between fault modes. The measure should take different fault time profiles, and the conditional probability that a given fault profile will occur when the system is in a given fault mode, into consideration. To compute the measure, information such as Table 2 will be taken into consideration and the result should be presented similar as Table 1.

## 3. BACKGROUND

Before discussing how to quantify diagnosability performance between fault modes, a summary of the results in Eriksson et al. (2013), including the definition of distinguishability, is presented in this section.

### *3.1 Model*

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

$$\begin{aligned} x[t+1] &= Ax[t] + B_u u[t] + B_f f[t] + B_v v[t] \\ y[t] &= Cx[t] + D_u u[t] + D_f f[t] + D_\varepsilon \varepsilon[t] \end{aligned} \tag{2}$$

where $v \sim \mathcal{N}(0, \Sigma_v)$ and $\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$.

If observing the system modeled as (2) during a time interval of $n$ samples, the system can be represented using a sliding window model, or batch model, in the form

$$Lz = Hx + Ff + Ne \tag{3}$$

where

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

and $e \sim \mathcal{N}(0, \Sigma_e)$.

### *3.2 Modeling fault modes as sets of probability density functions*

Without losing any information about the system, (3) is multiplied from the left by $\mathcal{N}_H$, where $\mathcal{N}_M$ spans the left null space of a matrix $M$. For the analysis, define $\tau = \mathcal{N}_H Lz = \mathcal{N}_H Ff + \mathcal{N}_H Ne$ which is a function of the faults and model uncertainties. Then, the conditional distribution of $\tau$ given a fault mode $f_i$ and fault time profile $\theta$ is described by the pdf $p(\tau|\theta) = p(\tau; \mathcal{N}_H F_i \theta)$ which is Gaussian distributed with mean $\mathcal{N}_H F_i \theta$, where $F_i$ represents the columns in $F$ corresponding to fault $f_i$, and covariance matrix $\mathcal{N}_H N \Sigma_e N^T \mathcal{N}_H^T$. Note that since the covariance of $\tau$ is not affected by the fault, it is omitted for convenience when denoting the pdf $p(\tau; \mathcal{N}_H F_i \theta)$. Then, each fault mode $f_i$ is represented by a set of pdf's as follows.

*Definition 2.* Let $\mathcal{Z}_{f_i}$ denote the set of all pdf's $p(\tau; \mathcal{N}_H F_i \theta)$, for all fault time profiles $\theta \in \Theta_i$, describing $\tau$ which could be explained by the fault mode $f_i$, i.e.

$$\mathcal{Z}_{f_i} = \{p(\tau|\theta) | \forall \theta \in \Theta_i\}. \tag{4}$$

□

The definition can be interpreted as if the pdf of $\tau$ is $p^i$ and $p^i \in \mathcal{Z}_{f_i}$ then the observations $\tau$ can be explained by the system being in fault mode $f_i$.

A specific fault time profile $f_i = \theta$ corresponds to one pdf in $\mathcal{Z}_{f_i}$ and is denoted

$$p^i_\theta = p(\tau|\theta) = p(\tau; \mathcal{N}_H F_i \theta). \tag{5}$$

### *3.3 The Kullback-Leibler divergence*

The Kullback-Leibler divergence between two pdf's $p^i$ and $p^j$ is defined as

$$K(p^i \| p^j) = \int_{-\infty}^{\infty} p^i(v) \log \frac{p^i(v)}{p^j(v)} dv = \mathrm{E}_{p^i}\left[\log \frac{p^i}{p^j}\right] \tag{6}$$

and is zero if and only if $p^i = p^j$. The Kullback-Leibler divergence can be interpreted as the expected log-likelihood ratio when the pdf of $\tau$ is $p^i$, see Eguchi and Copas (2006). Thus, if $p^i$ represent the pdf of $\tau$ given a fault $f_i$ and $p^j$ the pdf given another fault mode $f_j$, a larger Kullback-Leibler divergence can be interpreted as it is easier to isolate $f_i$ from $f_j$.

### *3.4 Distinguishability*

Distinguishability is defined in Eriksson et al. (2013) based on the Kullback-Leibler divergence as follows.

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

$$\mathcal{D}_{i,j}(\theta) = \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p^i_\theta \| p^j\right) \tag{7}$$

where $\mathcal{Z}_{f_j}$ is defined in Definition 2 and $p^i_\theta$ in (5). □

Theorem 1 in Eriksson et al. (2013) states that for models in the form (3), distinguishability can be computed explicitly as

$$\mathcal{D}_{i,j}(\theta) = \frac{1}{2} \|\mathcal{N}_{(H\, F_j)} F_i \theta\|^2 \tag{8}$$

given the assumption that, without loss of generality, $\Sigma$ is equal to the identity matrix, that is

$$\Sigma = \mathcal{N}_H N \Sigma_e N^T \mathcal{N}_H^T = I. \tag{9}$$

Distinguishability quantifies how difficult it is to isolate a fault $f_i$ with a specific fault time profile $\theta$ from another fault mode $f_j$, and is related to the maximum fault to noise ratio of any linear residual generator (Eriksson et al., 2013). Since the Kullback-Leibler divergence is minimized with respect to $p^j \in \mathcal{Z}_{f_j}$, distinguishability can be viewed as the minimum Kullback-Leibler divergence from the pdf $p^i_\theta$ to any pdf that can be explained by the system being in fault mode $f_j$ as graphically represented in Fig. 1. The Kullback-Leibler divergence is computed between two specific pdf's as shown in the figure, while distinguishability gives the smallest Kullback-Leibler divergence from a given pdf in $\mathcal{Z}_{f_i}$ to any pdf in $\mathcal{Z}_{f_j}$. Based on this, it is later shown that the distinguishability measure is related to the performance of the generalized log-likelihood ratio test.

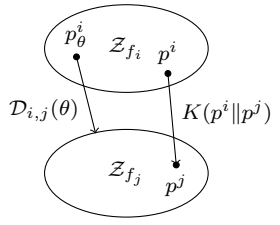


Fig. 1. A graphical representation of distinguishability which is the minimum Kullback-Leibler divergence from $p^i_\theta$ to any pdf $p^j \in \mathcal{Z}_{f_j}$.

## 4. REPRESENTING FAULT MODES USING PROBABILITIES OF FAULT TIME PROFILES

In its basic form (Eriksson et al., 2013), distinguishability is computed without taking any distribution of $\theta$ into consideration. Since distinguishability is computed for a given fault time profile $\theta$, there will be different distinguishability values for different fault time profiles.

Consider the situation where the occurrence of different fault realizations is modeled using probabilities as in Table 2. In order to describe a fault mode, the information that different fault realizations of a specific fault are more or less likely to occur can be taken into consideration using probabilities. For example, some faults are more likely to be constant with small magnitudes and less likely with higher magnitudes. Another case could be that a fault mostly occurs as one of a few specific fault realizations where some realizations are more likely than others. This is modeled by considering the conditional probability of each fault time profile given that the fault $f_i$ has occurred. The probability of $\tau$ distributed as the pdf $p^i_\theta \in \mathcal{Z}_{f_i}$ given that the system being in fault mode $f_i$ is denoted $q(\theta|f_i)$ where $\int_{\theta\in\Theta_i} q(\theta|f_i)d\theta = 1$. Thus, the probability that a fault of different magnitudes or fault time profiles can occur is represented by the conditional pdf $q(\theta|f_i)$ where a high probability represents a more probable fault realization $\theta$. An example of how different $q(\theta|f_i)$ can be described is shown in Table 2.

By taking the distribution of $\theta$ for a given fault mode $f_i$ into consideration, the distribution of $\tau$ given the fault mode $f_i$ can be written as

$$p(\tau|f_i) = \int_{\theta\in\Theta_i} p(\tau,\theta|f_i)d\theta = \int_{\theta\in\Theta_i} p(\tau|\theta)q(\theta|f_i)d\theta. \tag{10}$$

This means that the distribution of $\tau$ given that the system is in fault mode $f_i$ can be described by one pdf $p(\tau|f_i)$. For convenience $p(\tau|f_i)$ is sometimes denoted $p_{f_i}(\tau)$. Note that here, instead of representing a fault mode with a set of pdf's for the observation $\tau$, a single pdf is used to represent the complete fault mode.

In many cases, fault realizations with small magnitudes close to zero are significantly more probable compared to fault realizations with larger fault magnitudes. That is, small faults are more probable than bigger faults. This will result in a pdf of $\tau$, $p(\tau|f_i)$, where most observations $\tau$ will be close to zero even in the faulty case. That is, when a fault occur it will almost always be small. This gives that the pdf $p(\tau|f_i)$ in (10) will be similar to the pdf in the fault-free case since large faults are relatively rare, i.e. $p(\tau|f_i) \approx 0$.

In many applications, the pdf $q(\theta|f_i)$ is not known. Some knowledge might be from previous experiences of occurred faults but $q(\theta|f_i)$ could also be used to represent the type of faults that a diagnosis system is expected to detect. That is, instead of representing the probabilities of different fault time profiles, the pdf $q(\theta|f_i)$ can be used as a design parameter when representing different fault modes in the fault diagnosability analysis. For example, fault realizations and magnitudes that are required to be detected can have high probabilities and small otherwise. Thus, the diagnosability analysis can be designed such that the focus is on fault realizations that are required to be detected by a diagnosis system.

## 5. CANDIDATE MEASURES OF DISTINGUISHABILITY BETWEEN FAULT MODES

Two different ways of representing fault modes has been presented. Either as a set of pdf's (4) or as a single pdf (10). Therefore, there are more than one way of defining distinguishability between fault modes. Here, four candidates are discussed and a graphical representation of the candidates are shown in Fig. 2. The first two measures, a) and b) use a single pdf of $\tau$ to represent a fault mode. A fault mode that is represented by a single pdf is shown as a dot in the figure and otherwise a circle if the fault mode is represented by a set of pdf's. The third measure c), called minimum distinguishability, computes the worst-case distinguishability between two fault modes, and the last, called expected distinguishability, uses the conditional probabilities of $\tau$ given $\theta$.This is represented by an arrow between the closest elements in each set in Fig. 2. The last measure d) is called expected distinguishability and is the expected value of the distinguishability measure in (7) weighed with the probability of having that fault time profile.

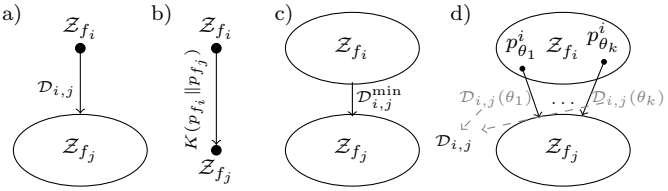


Fig. 2. Graphical representation of the quantitative diagnosability measures described in a) (11), b) (15), c) (16), and d) (17).

### 5.1 Modeling fault modes using mixture models

If the pdf $p_{f_i}(\tau)$ in (10) represents fault mode $f_i$ then a candidate to compute distinguishability of a fault mode $f_i$ from another fault mode $f_j$ can be formulated as given by the following definition.

*Definition 4.* Given a sliding window model (3), distinguishability of a fault mode $f_i$ from another fault mode $f_j$ is defined as

$$\mathcal{D}_{i,j} = \min_{p^j\in\mathcal{Z}_{f_j}} K\left(p_{f_i}\|p^j\right). \tag{11}$$

□

Note that the pdf $p^i_\theta$ for a given $\theta$ in (7) is here replaced by the pdf of the fault mode $p_{f_i}$ in (11). Then, (11) could be viewed as the minimum Kullback-Leibler divergence from the pdf $p_{f_i}$ to any $p^j \in \mathcal{Z}_{f_j}$. However, a problem with this formulation is for example when $\mathcal{Z}_{f_i} = \mathcal{Z}_{f_j}$ and $q(\theta|f_i) = q(\theta|f_j)$, i.e., the two fault modes are identical as described in the following example.

*Example 5.* Consider the small example

$$\tau = f_1 + f_2 + e \tag{12}$$

with $e \sim \mathcal{N}(0,\sigma^2)$ and

$$f_1, f_2 = \begin{cases} 1 & \text{with probability } q \\ 2 & \text{with probability } 1-q, \end{cases} \tag{13}$$

where the faults modes are represented by the sets $\mathcal{Z}_{f_1}, \mathcal{Z}_{f_2} = \{\mathcal{N}(1,\sigma^2), \mathcal{N}(2,\sigma^2)\}$. To compute (11), $p_{f_i} \sim q\mathcal{N}(1,\sigma^2) + (1-q)\mathcal{N}(2,\sigma^2)$. Since $\mathcal{Z}_{f_1} = \mathcal{Z}_{f_2}$, and also $p_{f_i} = p_{f_j}$, the distribution of $\tau$ given the two fault modes are identical, it is not possible to isolate the fault modes from each other. However, (11) will be non-zero since $p^j \neq p_{f_i}$ for all $p^j \in \mathcal{Z}_{f_j}$ since $p^j$ is Gaussian distributed while $p_{f_i}$ is given by (10). □

In Example 5, the fault modes $f_i$ and $f_j$ should be non-distinguishable since the two fault modes result in identical pdf's of $\tau$. However, $\mathcal{D}_{i,j} \neq 0$ since (11) does not take the whole fault mode $\mathcal{Z}_{f_j}$ into consideration but only the element $p^j$ which minimizes $K\left(p_{f_i} \| p^j\right)$. Then, distinguishability computed using (11) is a measure which gives non-intuitive results and is therefore not a suitable candidate for quantitative diagnosability performance between fault modes.

A solution to the problem with (11) is to also represent $\mathcal{Z}_{f_j}$ as one distribution (10), i.e.

$$p_{f_j}(\tau) = \int_{\theta\in\Theta_j} p(\tau,\theta|f_j)d\theta = \int_{\theta\in\Theta_j} p(\tau|\theta)q(\theta|f_j)d\theta, \tag{14}$$

which gives the following definition.

*Definition 6.* Given a sliding window model (3), distinguishability of a fault mode $f_i$ from another fault mode $f_j$ is defined as

$$\mathcal{D}_{i,j} = K\left(p_{f_i} \| p_{f_j}\right). \tag{15}$$

□

Then given Definition 6, $\mathcal{D}_{i,j} = 0$ in Example 5 since $p_{f_i}(\tau) = p_{f_j}(\tau)$. Also, since there is no minimization, (7) can be simplified to the Kullback-Leibler divergence. Note that the same notation is used for distinguishability between fault modes in (11) and (15) since the difference is how the set $\mathcal{Z}_{f_j}$ is defined. The set $\mathcal{Z}_{f_j}$ describing fault mode $f_j$ only contains one element in (15). Therefore to distinguish between the measures in (11) and (15), the notation $K\left(p_{f_i} \| p_{f_j}\right)$ is used for (15).

When representing each fault mode using only one pdf as in Definition 6, the Kullback-Leibler divergence between the fault modes in (15) mainly depends on the conditional pdf's $p^i_\theta \in \mathcal{Z}_{f_i}$ and $p^j_\theta \in \mathcal{Z}_{f_j}$ that have high probabilities $q(\theta|f_i)$ and $q(\theta|f_j)$ respectively. If a specific pdf $p^i_\theta$ has a low probability $q(\theta|f_i)$ it will have almost no impact on $p_{f_i}$. Thus, the measure in Definition 6 will mainly depend on the fault time profiles which have a higher probability.

However, it can also be observed that the pdf of $\tau$ for a specific fault realization $\theta$ is given by the conditional pdf $p^i_\theta$ and not $p_{f_i}$. Assume the case where the pdf $p^i_\theta$ can be explained by both fault modes $f_i$ and $f_j$, i.e. $p^i_\theta \in \mathcal{Z}_{f_i}$ and $p^i_\theta \in \mathcal{Z}_{f_j}$. This means that if $\tau$ has the pdf $p^i_\theta$ then none of the two fault could be rejected as the true fault mode. Then the measure in Definition 6 could give a positive value even if it is considered that the two fault modes can not be isolated from each other. However, if $\tau$ is assumed to have the pdf of either $p_{f_i}$ or $p_{f_j}$, defined in (10) and (14), instead of the conditional probability $p^i_\theta$, one fault mode could be more likely compared to the other one. Then, a test can be designed to reject the less likely fault mode. This corresponds to the distinguishability measure in Definition 6 which gives the expected value of the log-likelihood ratio test $\log \frac{p_{f_i}(\tau)}{p_{f_j}(\tau)}$ when $\tau$ has the pdf $p_{f_i}$. However, for the candidate measure (15) to be accurate, requires that the conditional pdf's $q(\theta|f_i)$ and $q(\theta|f_j)$ resemble reality. If $q(\theta|f_i)$ and $q(\theta|f_j)$ are used as design parameters for the analysis, then it is difficult to interpret the result of (15) since $p_{f_i}$ and $p_{f_j}$ do not resemble the true pdf's.

### *5.2 Minimum distinguishability*

A third alternative is to consider the worst case performance for any fault time profile given fault mode $f_i$, i.e., the element $p^i_\theta \in \mathcal{Z}_{f_i}$ that minimizes (7). Then, minimum distinguishability would be the minimal Kullback-Leibler divergence with respect to both fault modes $\mathcal{Z}_{f_i}$ and $\mathcal{Z}_{f_j}$ which gives the following definition.

*Definition 7.* (Minimum distinguishability). Given a sliding window model (3), minimum distinguishability of a fault mode $f_i$ from a fault mode $f_j$ is defined as

$$\mathcal{D}^{\min}_{i,j} = \min_{p^i\in\mathcal{Z}_{f_i}, p^j\in\mathcal{Z}_{f_j}} K\left(p^i \| p^j\right). \tag{16}$$

□

If fault magnitudes can be close to zero, or if there exists a $p^i \in \mathcal{Z}_{f_i}$ such that $p^i \in \mathcal{Z}_{f_j}$, (16) will be zero. Thus, a subset $\tilde{\mathcal{Z}}_{f_i} \subseteq \mathcal{Z}_{f_i}$ should be considered such that $\tilde{\mathcal{Z}}_{f_i} \cap \mathcal{Z}_{f_j} = \emptyset$. That is, the two fault modes can not result in the same pdf $p^i$ of $\tau$. For example, considering minimum required magnitudes of fault $f_i$ to be detected by a diagnosis system. In cases where the probabilities of different fault realizations are not known, minimum distinguishability can be a candidate for quantifying performance between fault modes. However since the probability of different fault realizations are used here to describe the fault modes, (16) is not a suitable measure of the separation between the fault modes in this case.

### *5.3 Expected distinguishability*

In Section 5.1, the fault modes are represented by one pdf (10). The candidate measures proposed in Section 5.1 can give different results compared to the distinguishability measure for a given fault time profile (7). For example, distinguishability for each specific fault time profile can be zero while distinguishability between fault modes, when representing each fault mode using one pdf, could be non-zero. This means that a fault mode can be isolable from another fault mode but not when considering a specific fault time profile.

Here, a candidate measure is sought which is consistent with the results of using the distinguishability measure for a given fault time profile (10). That is, if no fault time profile is isolable then the fault mode can not be isolable and if there exist an isolable fault time profile then the fault mode is isolable.

Instead of modeling each fault mode as one pdf as in Section 5.1, one candidate measure is to compute expected

distinguishability where the fault time profile $\theta$ is distributed as $q_\theta^i = q(\theta|f_i)$ and is defined as follows.

*Definition 8.* (Expected distinguishability). Given a sliding window model (3), expected distinguishability $\mathcal{D}_{i,j}$ of a fault mode $f_i$ from a fault mode $f_j$, where the distribution of fault time profile $\theta$ given $f_i$ has the pdf $q_\theta^i$, is defined as

$$\mathcal{D}_{i,j} = \mathrm{E}_{q_\theta^i}\left[\mathcal{D}_{i,j}(\theta)\right]. \tag{17}$$

□

To motivate expected distinguishability as a candidate measure, distinguishability (7) is interpreted using the generalized log-likelihood ratio test (GLLRT) based on $N$ i.i.d. samples of $\tau$ where $\tau \sim p_\theta^i \in \mathcal{Z}_{f_i}$ and $N \to \infty$.

$$\begin{aligned}\upsilon_N(\tau) &= \max_{p^i \in \mathcal{Z}_{f_i}} \min_{p^j \in \mathcal{Z}_{f_j}} \log \frac{\prod_{n=1}^N p^i(\tau_n)}{\prod_{n=1}^N p^j(\tau_n)} = \\ &= \max_{p^i \in \mathcal{Z}_{f_i}} \min_{p^j \in \mathcal{Z}_{f_j}} \left( \sum_{n=1}^N \log p^i(\tau_n) - \sum_{n=1}^N \log p^j(\tau_n) \right).\end{aligned} \tag{18}$$

Consider the GLLRT when $\upsilon_N(\tau)$ is normalized by the number of samples $N$ and $N$ goes to infinity, i.e.

$$\max_{p^i \in \mathcal{Z}_{f_i}} \min_{p^j \in \mathcal{Z}_{f_j}} \lim_{N\to\infty} \frac{1}{N} \left( \sum_{n=1}^N \log p^i(\tau_n) - \sum_{n=1}^N \log p^j(\tau_n) \right). \tag{19}$$

Note that the factor $\frac{1}{N}$ will not affect which elements in the sets that will solve the maximization and minimization in (19). Since the pdf that maximizes the first sum in (19) is $p_\theta^i \in \mathcal{Z}_{f_i}$,

$$\max_{p^i \in \mathcal{Z}_{f_i}} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \log p^i(\tau_n) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \log p_\theta^i(\tau_n) \tag{20}$$

which gives that

$$\begin{aligned}&\max_{p^i \in \mathcal{Z}_{f_i}} \min_{p^j \in \mathcal{Z}_{f_j}} \lim_{N\to\infty} \frac{1}{N} \left( \sum_{n=1}^N \log p^i(\tau_n) - \sum_{n=1}^N \log p^j(\tau_n) \right) = \\ &= \min_{p^j \in \mathcal{Z}_{f_j}} \lim_{N\to\infty} \frac{1}{N} \left( \sum_{n=1}^N \log p_\theta^i(\tau_n) - \sum_{n=1}^N \log p^j(\tau_n) \right) = \\ &= \min_{p^j \in \mathcal{Z}_{f_j}} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \log \frac{p_\theta^i(\tau_n)}{p^j(\tau_n)} = \\ &= \min_{p^j \in \mathcal{Z}_{f_j}} \mathrm{E}_{p_\theta^i} \left[ \log \frac{p_\theta^i(\tau_n)}{p^j(\tau_n)} \right] = \mathcal{D}_{i,j}(\theta)\end{aligned} \tag{21}$$

where the forth equality is based on the asymptotic relation (Bishop, 2006)

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \log \frac{p_\theta^i(\tau_n)}{p^j(\tau_n)} = \mathrm{E}_{p_\theta^i} \left[ \log \frac{p_\theta^i(\tau_n)}{p^j(\tau_n)} \right] \tag{22}$$

since $\tau \sim p_\theta^i \in \mathcal{Z}_{f_i}$. Thus, distinguishability (7) measures the mean contribution of each sample $\tau_n$ to the GLLRT if the true pdf of $\tau$ is $p_\theta^i \in \mathcal{Z}_{f_i}$ when $N \to \infty$.

Then, expected distinguishability could be interpreted as the expected asymptotic gain of each new sample in the generalized log-likelihood ratio test given that the system being in fault mode $f_i$. In Definition 8, only the pdf of $\theta$ given $f_i$ is taken into consideration but not the pdf $\theta$ given $f_j$. The argument for this is the same as for distinguishability (7), i.e. to be sure to isolate $f_i$ from $f_j$ we want to quantify how difficult it is to isolate $f_i$ from $f_j$ independent of realization of $f_j$. If the pdf of $\tau$ is $p(\tau|\theta)$ which is an element in both sets $\mathcal{Z}_{f_i}$ and $\mathcal{Z}_{f_j}$, then $\mathcal{D}_{i,j}(\theta) = 0$. This is the case, independent of $q(\theta|f_i)$, since the pdf of $\tau$ can be explained by both fault modes $f_i$ and $f_j$.

## 6. DISCUSSION

Based on the previous discussions in Sections 5.1 and 5.2, the candidate measures (11) and (16), graphically represented by a) and c) respectively in Fig. 2, are not considered here as measures to quantify diagnosability performance between fault modes. The other two candidate measures, (15) and (17), represented by b) and d) respectively in Fig. 2, are here further analyzed.

A conceptual difference between (15) and (17) is that (15) quantify diagnosability performance based on the pdf's of $\tau$ while (17) uses the conditional pdf's of $\tau$ given $\theta$.

*Example 9.* Consider the same model as in Example 5 but with the small difference that

$$\begin{aligned}f_1 &= \begin{cases} 1 & \text{with probability } q_1 \\ 2 & \text{with probability } 1-q_1, \end{cases} \\ f_2 &= \begin{cases} 1 & \text{with probability } q_2 \\ 2 & \text{with probability } 1-q_2, \end{cases}\end{aligned} \tag{23}$$

where $q_1 \neq q_2$. As before, the faults modes are represented by the sets $\mathcal{Z}_{f_1}, \mathcal{Z}_{f_2} = \{\mathcal{N}(1,\sigma^2), \mathcal{N}(2,\sigma^2)\}$ which contains the same two elements but the pdf's $p_{f_1}$ and $p_{f_2}$ are different. □

The candidate measure (15) will give that the two fault modes are isolable from each other. The measure (15) can be seen as the expected value of the log-likelihood ratio test $\log \frac{p_{f_1}(\tau)}{p_{f_2}(\tau)}$ when $p_{f_1}$ is the true pdf. Since different values of $\tau$ are more or less likely given the two fault modes it is possible to identify the most probable fault mode.

Given Example 9, expected distinguishability (17) will be zero since each pdf in $\mathcal{Z}_{f_1}$ also exist in $\mathcal{Z}_{f_2}$. This can be interpreted as there are no observations given any fault time profile $\theta$ in fault mode $f_i$ that cannot also be explained by the system being in fault mode $f_j$. This would correspond to the generalized log-likelihood ratio test in (18) equal to 0 for all pdf's in $p_\theta^1 \in \mathcal{Z}_{f_1}$. Thus, expected distinguishability (17) will only be non-zero if there exists a pdf $p_\theta^1 \in \mathcal{Z}_{f_1}$ that does not exist in $\mathcal{Z}_{f_2}$ and the probability for that pdf is $q(\theta|f_1) > 0$.

Both candidate measures, (15) and (17) can be used to quantify isolability performance between fault modes. However, the measure (15) requires that the true pdf's $q(\theta|f_i)$ and $q(\theta|f_j)$ are known which is usually not the case. For expected distinguishability, $q(\theta|f_i)$ does not have to represent the true pdf but can, for example, be chosen to represent the importance of different fault realizations in the analysis. Also, since the result of expected distinguishability is consistent with the results of the previous distinguishability measure for a given fault time profile it is here selected as the quantitative measure of diagnosability performance between fault modes.

## 7. COMPUTATION OF EXPECTED DISTINGUISHABILITY

Here, some methods to compute expected distinguishability $\mathcal{D}_{i,j}$ in (17), both analytically and numerically are presented. For a linear descriptor model in the form (3) and a fault time profile $\theta = \bar{\theta}\alpha$ where $\bar{\theta}$ is a fixed fault time profile and $\alpha$ is a stochastic fault magnitude with a defined mean and variance and described by a pdf $q^i_\alpha$, (17) can be computed explicitly given the following proposition.

*Proposition 10.* Given a sliding window model (3) with Gaussian distributed random vector $e$, under assumption (9), and $\theta = \bar{\theta}\alpha$ where $\alpha$ is a random variable with pdf $q^i_\alpha$. The expected distinguishability defined in Definition 8 is given by

$$\mathcal{D}_{i,j} = \frac{1}{2}\left\|\mathcal{N}_{(H\,F_j)}F_i\bar{\theta}\right\|^2\left[\mu_\alpha^2 + \sigma_\alpha^2\right] \tag{24}$$

where $\mu_\alpha = \mathrm{E}_{q^i_\alpha}[\alpha]$ and $\sigma^2_\alpha = \mathrm{E}_{q^i_\alpha}\left[(\alpha - \mu_\alpha)^2\right]$. □

**Proof.** Equation (8) gives that expected distinguishability (17) can be written as

$$\begin{aligned}\mathcal{D}_{i,j} &= \mathrm{E}_{q^i_\theta}\left[\mathcal{D}_{i,j}(\theta)\right] = \mathrm{E}_{q^i_\theta}\left[\frac{1}{2}\|\mathcal{N}_{(H\,F_j)}F_i\theta\|^2\right] \\ &= \frac{1}{2}\|\mathcal{N}_{(H\,F_j)}F_i\bar{\theta}\|^2\mathrm{E}_{q^i_\alpha}\left[\alpha^2\right] \\ &= \frac{1}{2}\left\|\mathcal{N}_{(H\,F_j)}F_i\bar{\theta}\right\|^2\left[\mu_\alpha^2 + \sigma_\alpha^2\right]\end{aligned}$$

where the last equality follows from

$$\mathrm{E}_{q^i_\alpha}\left[\alpha^2\right] = \mu_\alpha^2 + \sigma_\alpha^2 \tag{25}$$

which finishes the proof. □

Proposition 10 gives that expected distinguishability is a function of the mean and variance of $q^i_\alpha$ if the fault time profile is fixed with random magnitude.

Expected distinguishability can also be computed using Monte Carlo sampling (see for example Bishop (2006)) as

$$\mathcal{D}_{i,j}(q^i_\theta) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\mathcal{D}_{i,j}(\theta_k) \text{ where } \theta_k \sim q^i_\theta. \tag{26}$$

where $\mathcal{D}_{i,j}(\theta_k)$ is computed using (8). Equation (26) gives an approximation of expected distinguishability which, compared to (24), can be used for any pdf $q^i_\theta$ and not only for fixed fault time profiles with random magnitudes which is required for the analytical expression (24). Both (24) and (26) are used to compute expected distinguishability in the case study which is presented in the following section.

## 8. CASE STUDY

Here, the example described in Section 2 is analyzed where a window model of length $N = 10$ is considered. Expected distinguishability in Definition 8 is used to quantify diagnosability between fault modes. The pdf's of different fault time profiles are given in Table 2. The fault time profiles given $f_2$, $f_3$, and $f_4$, are here assumed constant with random magnitudes. For these cases, expected distinguishability can be computed analytically using (24).

The fault time profiles given fault mode $f_1$ are here assumed to be random with zero mean and covariance matrix $I$ as described in Table 2. Therefore, expected distinguishability is computed using Monte Carlo sampling (26). Expected distinguishability is computed using Monte Carlo sampling for the other faults too, i.e. $f_2$, $f_3$, and $f_4$, to analyze the convergence of (26). For the Monte Carlo sampling in this case study, the number of samples is selected to $N = 10000$.

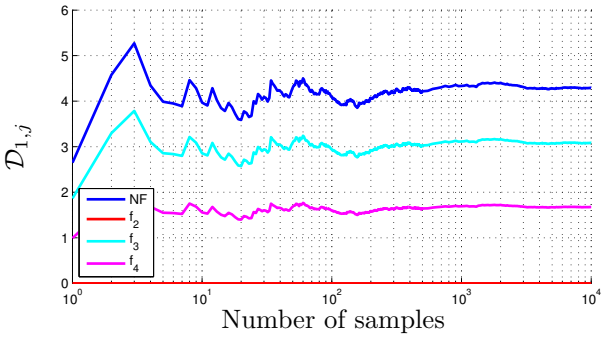


Fig. 3. Computed expected distinguishability of fault mode $f_1$ from the other fault modes using Monte Carlo sampling.

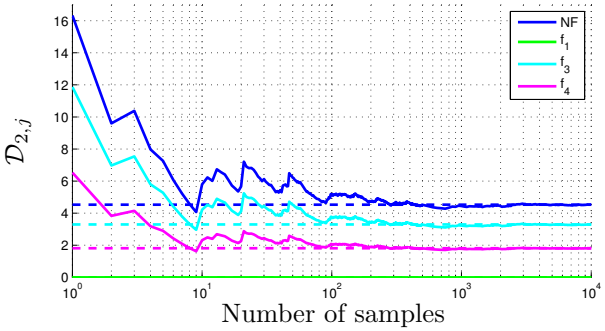


Fig. 4. Computed expected distinguishability of fault mode $f_2$ from the other fault modes using Monte Carlo sampling (solid) and analytical solution (dashed).

The results when computing distinguishability for each fault mode from all other fault modes are shown in Fig. 3-6 respectively. The solid curves represent the computed expected distinguishability using Monte Carlo sampling as a function of the number of samples. The analytically computed expected distinguishability values are presented by the dashed curves and when using Monte Carlo sampling are given by the solid lines. The figures show that the Monte Carlo sampling seems to converge within 5000 samples in all cases and the results are consistent with the analytical results.

A summary of the results from the analysis are shown in Table 3. Note that the first row is based in the approximations of the Monte Carlo sampling in Fig. 3 while the rest are computed using (17). The results show for example that it is easier to isolate fault mode $f_3$ from $f_1$ than vice versa and it is easier to detect $f_3$ and $f_4$ compared to $f_1$ and $f_2$ which is reasonable since the expected fault magnitudes are higher.

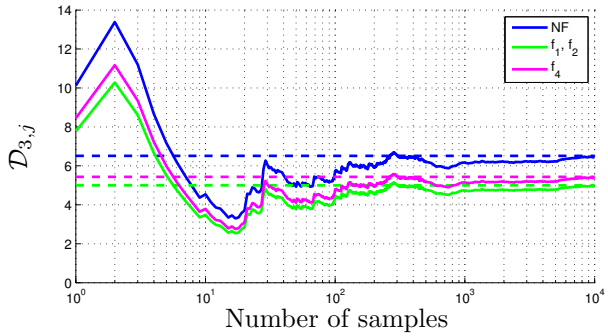


Fig. 5. Computed expected distinguishability of fault mode $f_3$ from the other fault modes using Monte Carlo sampling (solid) and analytical solution (dashed).

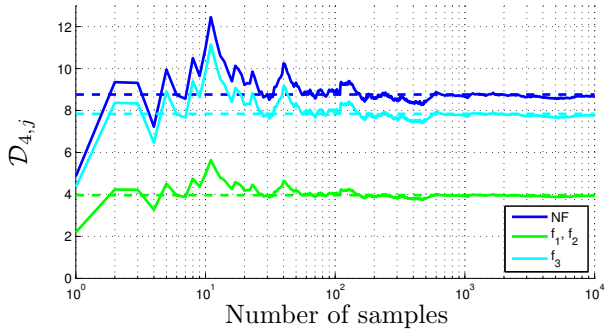


Fig. 6. Computed expected distinguishability of fault mode $f_4$ from the other fault modes using Monte Carlo sampling (solid) and analytical solution (dashed).

## 9. CONCLUSIONS

Four candidate measures are discussed to quantify fault diagnosability performance between fault modes where probabilities of different fault realizations are taken into consideration. Expected distinguishability is proposed as the best candidate and the measure is based on the previous definition of distinguishability and can be related to the performance of the generalized log-likelihood ratio test.

The expected distinguishability measure takes knowledge about different fault realizations, model uncertainties and noise into consideration to evaluate how difficult it is to isolate one fault mode $f_i$ from another fault mode $f_j$. This is important when analyzing fault diagnosability performance during the design of the diagnosis system. Two methods are presented to compute expected distinguishability, one analytical and one numerical, and both are used on a case study.

Table 3. Computed expected distinguishability of (1) given the fault distributions in Table 2.

| $\mathcal{D}_{i,j}(\theta)$ | NF | $f_1$ | $f_2$ | $f_3$ | $f_4$ |
|---|---|---|---|---|---|
| $f_1$ | 4.29 | 0 | 0 | 3.08 | 1.68 |
| $f_2$ | 4.53 | 0 | 0 | 3.29 | 1.81 |
| $f_3$ | 6.51 | 5.00 | 5.00 | 0 | 5.43 |
| $f_4$ | 8.76 | 3.97 | 3.97 | 7.84 | 0 |

## REFERENCES

Bhushan, M., Narasimhan, S., and Rengaswamy, R. (2008). Robust sensor network design for fault diagnosis. *Computers & Chemical Engineering*, 32(4), 1067–1084.

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Blanke, M. and Schröder, J. (2006). *Diagnosis and fault-tolerant control*, volume 2. Springer.

Bregon, A., Daigle, M., Roychoudhury, I., Biswas, G., Koutsoukos, X., and Pulido, B. (2014). An event-based distributed diagnosis framework using structural model decomposition. *Artificial Intelligence*, 210, 1–35.

Cui, Y., Shi, J., and Wang, Z. (2014). System-level operational diagnosability analysis in quasi real-time fault diagnosis: The probabilistic approach. *Journal of Process Control*, 24(9), 1444 – 1453.

Eguchi, S. and Copas, J. (2006). Interpreting kullback–leibler divergence with the neyman–pearson lemma. *Journal of Multivariate Analysis*, 97(9), 2034–2040.

Eriksson, D., Frisk, E., and Krysander, M. (2012a). A sequential test selection algorithm for fault isolation. In *Proc. of the 10th European Workshop on Advanced Control and Diagnosis, ACD 2012, Copenhagen, Denmark*.

Eriksson, D., Frisk, E., and Krysander, M. (2013). A method for quantitative fault diagnosability analysis of stochastic linear descriptor models. *Automatica*, 49(6), 1591–1600.

Eriksson, D., Krysander, M., and Frisk, E. (2012b). Using quantitative diagnosability analysis for optimal sensor placement. *Proceedings of IFAC Safe Process*.

Frisk, E., Krysander, M., and Åslund, J. (2009). Sensor placement for fault isolation in linear differential-algebraic systems. *Automatica*, 45(2), 364–371.

Harrou, F., Fillatre, L., and Nikiforov, I. (2014). Anomaly detection/detectability for a linear model with a bounded nuisance parameter. *Annual Reviews in Control*, 38(1), 32–44.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*, volume 2. Springer.

Huber, J., Kopecek, H., and Hofbaur, M. (2014). Sensor selection for fault parameter identification applied to an internal combustion engine. In *Proc. of the IEEE Multi-conference on Systems and Control, Antibes, France*.

Krysander, M. (2006). *Design and Analysis of Diagnosis Systems Using Structural Methods*. Ph.D. thesis, Linköpings universitet.

Patton, R.J., Clark, R.N., and Frank, P.M. (2000). *Issues of fault diagnosis for dynamic systems*. Springer.

Sarrate, R., Nejjari, F., and Rosich, A. (2012). Sensor placement for fault diagnosis performance maximization in distribution networks. In *Control & Automation (MED), 2012 20th Mediterranean Conf. on*, 110–115. IEEE.

Travé-Massuyes, L., Escobet, T., and Olive, X. (2006). 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.

Wheeler, T.J. (2011). *Probabilistic Performance Analysis of Fault Diagnosis Schemes*. Ph.D. thesis, University of California, Berkeley.