

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

Daniel Eriksson, Erik Frisk and Mattias Krysander

Linköping University Post Print



N.B.: When citing this work, cite the original article.

Original Publication:

Daniel Eriksson, Erik Frisk and Mattias Krysander, A method for quantitative fault diagnosability analysis of stochastic linear descriptor models, 2013, Automatica, (49), 6, 1591-1600.

<http://dx.doi.org/10.1016/j.automatica.2013.02.045>

Copyright: Elsevier

<http://www.elsevier.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-89941>

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

Daniel Eriksson ^{*}, Erik Frisk, Mattias Krysander

*Department of Electrical Engineering, Linköping University
SE-581 83 Linköping, Sweden. {daner,frisk,matkr}@isy.liu.se.*

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.

Key words: Fault diagnosability analysis, Fault detection and isolation, Model-based diagnosis.

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 [16], [19], and [25].

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 [4] and [28]. In [6], 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 literature, see for example [3], [5], [18], and [31]. 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 [1] and [20],

[★] The work is partially supported by the Swedish Research Council within the Linnaeus Center CADICS.

^{*} Corresponding author. Tel: +46 13 - 28 5743

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 [11], [27], and [29]. 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 [8] and [9] 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

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

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 [12]. A simple discrete-time dynamic

model of a spring-mass system is considered,

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

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. \diamond

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.

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

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 .

With a little abuse of notation, define the vectors

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

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

$$Lz = Hx + Ff + Ne \quad (4)$$

where

$$\begin{aligned} L &= \begin{pmatrix} 0 & 0 & \dots & 0 & -B_u & 0 & \dots & 0 \\ 0 & 0 & & 0 & 0 & -B_u & & 0 \\ \vdots & \ddots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & -B_u \\ I & 0 & \dots & 0 & -D_u & 0 & \dots & 0 \\ 0 & I & & 0 & 0 & -D_u & & 0 \\ \vdots & \ddots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 & 0 & \dots & -D_u \end{pmatrix}, \\ H &= \begin{pmatrix} A & -E & 0 & \dots & 0 \\ 0 & A & -E & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & A & -E \\ C & 0 & 0 & \dots & 0 \\ 0 & C & 0 & & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & C & 0 \end{pmatrix}, \quad F = \begin{pmatrix} B_f & 0 & \dots & 0 \\ 0 & B_f & & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & B_f \\ D_f & 0 & \dots & 0 \\ 0 & D_f & & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & D_f \end{pmatrix}, \\ N &= \begin{pmatrix} B_v & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & B_v & & 0 & 0 & 0 & & 0 \\ \vdots & \ddots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & B_v & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & D_\varepsilon & 0 & \dots & 0 \\ 0 & 0 & & 0 & 0 & D_\varepsilon & & 0 \\ \vdots & \ddots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & D_\varepsilon \end{pmatrix}, \end{aligned}$$

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 $(t-n+1, t-n+2, \dots, t)$.

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.,

$$f_i = (f[i], f[l_f + i], f[2l_f + i], \dots, f[(n-1)l_f + i])^T. \quad (5)$$

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

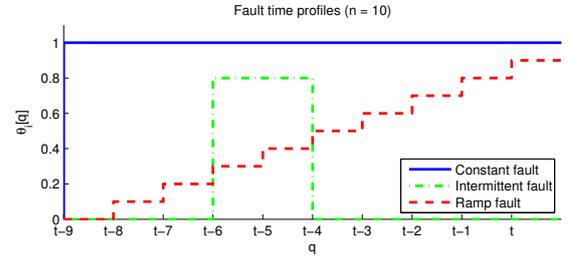


Fig. 1. Fault time profiles representing a constant fault, an intermittent fault, and a fault entering the system like a ramp.

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

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

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

$$D_\varepsilon \text{ is full row-rank and } \exists \lambda \in \mathbb{C}: \lambda E - A \text{ is full rank,} \quad (7)$$

i.e., all sensors have measurement noise and the model has a unique solution for a given initial state, see [22]. 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 [15] and [32]. The input-output model can, in the general case, then be written as

$$\mathcal{N}_H Lz = \mathcal{N}_H Ff + \mathcal{N}_H Ne. \quad (8)$$

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 [26].

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 $\tau = \mathcal{N}_H Lz$, which is the left hand side of (8). The vector $\tau \in \mathbb{R}^{n_l y - l_x}$ depends linearly on the fault vector f and the noise vector e and represents the *behavior* of the model, see [26]. 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 Fig. 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.

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

The definition of \mathcal{Z}_{f_i} is a stochastic counterpart to observation sets in the deterministic case, see [24]. Each fault mode f_i , including NF, can be described by a set \mathcal{Z}_{f_i} . The set \mathcal{Z}_{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

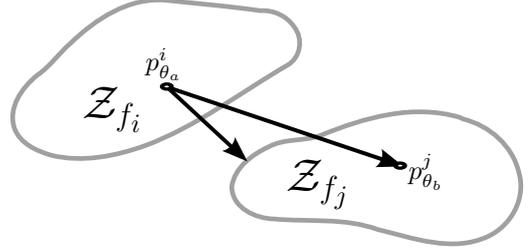


Fig. 2. A graphical visualization of the sets \mathcal{Z}_{f_i} and \mathcal{Z}_{f_j} and the smallest difference between $p_{\theta_a}^i \in \mathcal{Z}_{f_i}$ and a pdf $p^j \in \mathcal{Z}_{f_j}$.

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

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

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

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

Similarly, if $p_{\theta}^i \notin \mathcal{Z}_{\text{NF}}$, 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, $p_{\theta_a}^i \in \mathcal{Z}_{f_i}$ and $p_{\theta_b}^j \in \mathcal{Z}_{f_j}$, 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 Fig. 2 where the difference of $p_{\theta_a}^i$ and $p_{\theta_b}^j$ 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

$$\begin{aligned} H_0 : p &= p_{\theta_b}^j \\ H_1 : p &= p_{\theta_a}^i. \end{aligned} \quad (11)$$

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

$$\lambda(\tau) = \log \frac{L_{p_{\theta_a}^i}(\tau)}{L_{p_{\theta_b}^j}(\tau)}$$

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 $p_{\theta_b}^j$ 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 $p_{\theta_a}^i$ and $E[\lambda(\tau)] \geq 0$, see [2]. Thus, λ changes sign, in the mean, with the two hypotheses. Therefore, the mean of $\lambda(\tau)$, under H_1

$$E_{p_{\theta_a}^i}[\lambda(\tau)] = E_{p_{\theta_a}^i} \left[\log \frac{L_{p_{\theta_a}^i}(\tau)}{L_{p_{\theta_b}^j}(\tau)} \right] \quad (12)$$

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 [21] and will be denoted $K(p_{\theta_a}^i \| p_{\theta_b}^j)$.

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

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

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

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

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 $K(p_{\theta}^i \| p^j)$, for any $p^j \in \mathcal{Z}_{f_j}$. Also, minimizing $K(p_{\theta}^i \| p^j)$ 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_{θ}^i , see [7]. 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*

$$\mathcal{D}_{i,j}(\theta) = \min_{p^j \in \mathcal{Z}_{f_j}} K(p_{\theta}^i \| p^j) \quad (15)$$

where the set \mathcal{Z}_{f_j} is defined in Definition 1 and p_{θ}^i 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 $\mathcal{D}_{i,j}(\theta)$ 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*

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

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

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

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

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

$$\lim_{p^j \rightarrow p_{NF}} K(p_{\theta}^i \| p^j) = K(p_{\theta}^i \| p_{NF}).$$

Then

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

which proves Proposition 2. \square

From now on, it is assumed that $\Theta_i = \mathbb{R}^n \setminus \{0\}$ for all $i = 1, 2, \dots, l_f$ 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 disturbances 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 [17]. 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

$$\Sigma = \mathcal{N}_H N \Lambda_e N^T \mathcal{N}_H^T = I \quad (18)$$

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

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

where Γ is non-singular and

$$\mathcal{N}_H N \Lambda_e N^T \mathcal{N}_H^T = \Gamma \Gamma^T \quad (20)$$

is satisfied, and T_2 is any matrix ensuring invertibility 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

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

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

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

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

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

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 , $\text{Ker}(A)$, 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

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

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

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

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

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

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

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 $n = 3$.

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

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.

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

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. \diamond

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 [23].

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 [13,14,33], as

Definition 4 (Linear residual generator) *A linear function $r = Rz$, 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 $\gamma\mathcal{N}_{(HF_j)}$ where γ is a row-vector to obtain

$$r = \gamma\mathcal{N}_{(HF_j)}Lz = \gamma\mathcal{N}_{(HF_j)}Ff + \gamma\mathcal{N}_{(HF_j)}Ne \quad (25)$$

Here, $\gamma\mathcal{N}_{(HF_j)}Lz$ is a residual generator that isolates from fault mode f_j . If only detectability, and not isolability, of f_i is considered, $\mathcal{N}_{(HF_j)}$ 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.

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

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, $\lambda(\theta) = \gamma\mathcal{N}_{(HF_j)}F\theta$ and the standard deviation of the noise σ as*

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

where

$$\sigma^2 = \gamma\mathcal{N}_{(HF_j)}N\Lambda_eN^T\mathcal{N}_{(HF_j)}^T\gamma^T.$$

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

$$\underbrace{\gamma\mathcal{N}_{(HF_j)}L}_{=:L} z = \underbrace{\gamma\mathcal{N}_{(HF_j)}F}_{=:F} f + \underbrace{\gamma\mathcal{N}_{(HF_j)}N}_{=:N} e \quad (27)$$

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

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

□

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 $(\mathcal{N}_{(H F_j)} F_i \theta)^T$ 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 $(\mathcal{N}_{(H F_j)} F_i \theta)^T$ are parallel. \square

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

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

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

$$r = \gamma \mathcal{N}_{(H F_j)} L z = k (\mathcal{N}_{(H F_j)} F_i \theta)^T \mathcal{N}_{(H F_j)} L z. \quad (28)$$

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

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

which has a FNR, with respect to f_3 ,

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

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. \diamond

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 [20]. 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 Fig. 3. 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 mod-

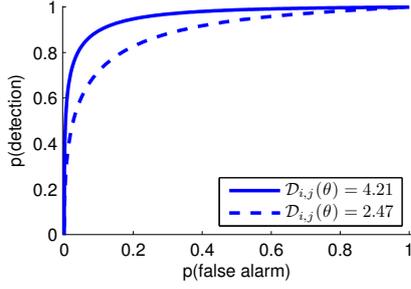


Fig. 3. A ROC-curve comparing probability of detection and false alarm for two residuals with different computed values of distinguishability. A higher distinguishability corresponds to a higher ratio between the probability of detection and false alarm.

els 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 [30] and an overview is shown in Fig. 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 Fig. 4. The faults are briefly described in Table 6 and can be divided into four groups: f_1, \dots, f_4 are actuator faults, f_5, \dots, f_8 are sensor faults, f_9, \dots, 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 [30], 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 pro-

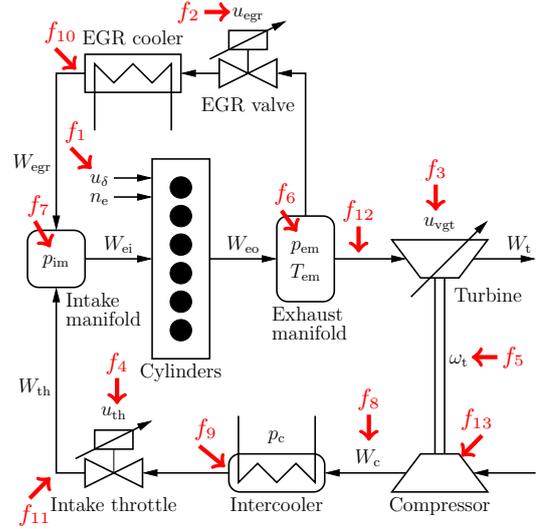


Fig. 4. Overview of diesel engine model. The arrows indicate the locations in the model of the modeled faults in Table 6.

Table 6

Implemented faults in the diesel engine model where nom = "nominal value" and $\Delta p_x = p_x - p_{atm}$.

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

portional to the uncertainties in the model according to [30]. 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 [30]. 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.

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

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

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.,

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

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

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

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 [10]. 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.

Fig. 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 Fig. 5 correspond to the detectability performance which increases with pressure after the compressor and seems proportional to $\sqrt{p_c - p_{atm}}$. 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 Fig. 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 Fig. 5 and Fig. 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 Fig. 5 and Fig. 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.

7 Conclusions

The topic addressed in this paper is how to quantify diagnosability properties of a given model, without design-

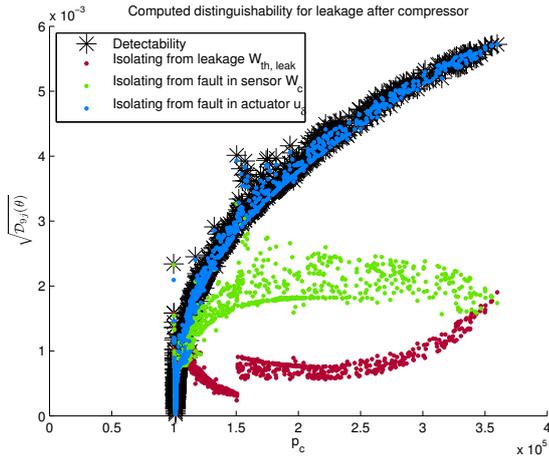


Fig. 5. Computed distinguishability for a leakage after the compressor. Distinguishability for a leakage after the compressor from the no fault case, i.e., detectability performance, increases by increasing compressor pressure. Isolating the leakage from a leakage after the throttle, or a fault in the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

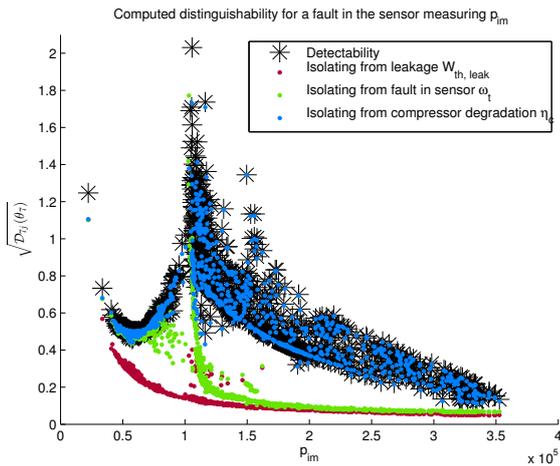


Fig. 6. Computed distinguishability for an additive fault in sensor measuring p_{im} . Distinguishability is not changed when isolating from a degradation in the compressor compared to detectability. There is a peak for distinguishability around $p_{im} \approx p_{atm}$ except when isolating from a leakage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

ing 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 *distinguishabil-*

ity, 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 derived¹.

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.

References

- [1] Michèle Basseville and Igor V. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
- [2] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Resource Center, Pacific Grove, CA, 2001.
- [3] Robert H. Chen, D.Lewis Mingori, and Jason L. Speyer. Optimal stochastic fault detection filter. *Automatica*, 39(3):377 – 390, 2003.
- [4] 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.
- [5] 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.
- [6] Dilek Düşteğör, 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.
- [7] Shinto Eguchi and John Copas. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *J. Multivar. Anal.*, 97:2034–2040, October 2006.
- [8] 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, 2011.
- [9] Daniel Eriksson, Mattias Krysander, and Erik Frisk. Quantitative stochastic fault diagnosability analysis. In *50th IEEE Conference on Decision and Control*, Orlando, Florida, USA, 2011.

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

- [10] Economic Commission for Europe Inland Transport Committee. Regulation No 49 of the Economic Commission for Europe of the United Nations (UN/ECE), August 2010.
- [11] 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.
- [12] Erik Frisk, Mattias Krysander, and Jan Åslund. Sensor Placement for Fault Isolation in Linear Differential-Algebraic Systems. *Automatica*, 45(2):364–371, 2009.
- [13] Erik Frisk and Lars Nielsen. Robust residual generation for diagnosis including a reference model for residual behavior. *Automatica*, 42(3):437 – 445, 2006.
- [14] 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.
- [15] Janos Gertler. "Fault detection and isolation using parity relations". *Control Engineering Practice*, 5(5):653 – 661, 1997.
- [16] Janos Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker Inc., Upper Saddle River, NJ, USA, 1998.
- [17] Torkel Glad and Lennart Ljung. *Control Theory : Multivariable & Nonlinear Methods*. CRC, March 2000.
- [18] Frédéric Hamelin and Dominique Sauter. Robust fault detection in uncertain dynamic systems. *Automatica*, 36(11):1747 – 1754, 2000.
- [19] Rolf Isermann. "Model-based fault-detection and diagnosis - status and applications". *Annual Reviews in Control*, 29(1):71 – 85, 2005.
- [20] Steven M. Kay. *Fundamentals of statistical signal processing: Detection theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998.
- [21] Kullback, S. and Leibler, R. A. On Information and Sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.
- [22] Peter Kunkel and Volker Mehrmann. *Differential-Algebraic Equations: Analysis and Numerical Solution*. European Mathematical Society, Zürich, 2006.
- [23] Mattias Nyberg. Criteria for detectability and strong detectability of faults in linear systems. *International Journal of Control*, 75(7):490–501, May 2002.
- [24] 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.
- [25] Ron J. Patton, Paul M. Frank, and Robert N. Clark. *Issues of Fault Diagnosis for Dynamic Systems*. Springer Publishing Company, Incorporated, 1st edition, 2010.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] Timothy J. Wheeler. *Probabilistic Performance Analysis of Fault Diagnosis Schemes*. PhD thesis, University of California, Berkeley, 2011.
- [32] Ping Zhang and Steven X. Ding. "Disturbance decoupling in fault detection of linear periodic systems". *Automatica*, 43(8):1410 – 1417, 2007.
- [33] 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.