2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Quantitative Stochastic Fault Diagnosability Analysis

Daniel Eriksson, Mattias Krysander and Erik Frisk

***Abstract*— A theory is developed for quantifying fault detectability and fault isolability properties of static linear stochastic models. Based on the model, a stochastic characterization of system behavior in different fault modes is defined and a general measure, based on the Kullback-Leibler information, is proposed to quantify the difference between the modes. This measure, called distinguishability, of the model is shown to give sharp upper limits of the fault to noise ratios of residual generators. Finally, a case-study of a diesel engine model shows how the general framework can be applied to a dynamic and nonlinear model.**

I. INTRODUCTION

Diagnosis and supervision of industrial systems concerns detecting and isolating faults that occur in the system. When designing a diagnosis system, information of detectability and isolability properties of the system, before actually designing any diagnosis system, is useful. Such information states whether a test with certain properties can be created or if more sensors are needed to get satisfactory diagnosability performance. Because of system noise, time could be wasted on developing tests to detect a fault that in reality is impossible to detect or isolate.

There are several works describing methods from classical detection theory, for example the books [1] and [2], which can be used for quantified detectability analysis using a stochastic characterization of faults. A main contribution with respect to these works is that here, *isolability* performance is also considered.

There exist systematic methods for analyzing isolability performance in dynamic systems, e.g., [3], [4] and [5], however these approaches are deterministic and only give qualitative statements whether a fault is isolable or not. This gives an optimistic result of isolability performance because an isolable fault can be hard to detect in practice due to low fault to noise ratio.

In [7], a quantitative analysis of isolability performance in dynamic linear models is made using parity spaces. Main differences compared to this paper are that the analysis in [7] assumes a known fault size, which is not assumed in this work, and also focuses on the performance of a set of tests rather than properties of the model.

This paper presents a theory for quantified isolability analysis of linear static models. The theory has close connections to isolability performance of residual generators and the fault to noise ratio of these. An example in Section II introduces problems faced when analyzing isolability performance. The problem formulation is specified in Section III. The method and theory are presented in Section IV. A relation between the theory and residual generators and how to choose optimal residuals are shown in Section V. The method is then used to analyze isolability performance of a diesel engine model in Section VI.

Dept. of Electrical Engineering, Linköping University, Sweden Email: {daner,matkr,frisk}@isy.liu.se

II. AN INTRODUCTORY EXAMPLE

An academic example will be used to illustrate important isolability properties encountered and to discuss limitations in using only a deterministic analysis method. The discussion will be used as a basis for the problem formulation in Section III.

Consider a static model described by

$$\begin{aligned} x_1 &= u + f_1 + f_2 + \varepsilon_1 \\ x_2 &= 2(x_1 - f_1) + f_3 + \varepsilon_2 \\ y_1 &= x_2 + f_4 + \varepsilon_3 \\ y_2 &= x_2 + \varepsilon_4 \end{aligned} \tag{1}$$

where x_i are unknown variables, y_i are measured variables, u is a known actuator variable, f_i are modeled faults, and ε_i are independent stochastic variables modeling uncertainty. A fault mode represents if a fault f_i is present, i.e., $f_i \neq 0$. With a little abuse of notation, f_i will also be used to denote the fault mode when f_i is the present fault. For simplicity in this example, let $f_i = 1$ if f_i is present and 0 otherwise. The stochastic variables are assumed to be normally distributed with means and variances given by $\varepsilon_1, \varepsilon_2 \sim \mathcal{N}(0, 0.1)$ and $\varepsilon_3, \varepsilon_4 \sim \mathcal{N}(0, 1)$.

Given how the faults enter the model, information of isolability performance is useful from a diagnostic perspective. Table I shows the results of a deterministic isolability analysis of (1) obtained by neglecting the stochastic variables ε_i and using for example the method described in [6]. An X in the table marks if the fault in the row is not isolable from the fault in the column. The NF (No Fault) column represents if the fault mode is detectable. The analysis states that f_1 is not detectable, since no test sensitive to f_1 can be created. Therefore f_1 is not isolable from the other faults. Fault modes f_2 and f_3 are not isolable from each other because getting information about f_3 requires information about x_1 which is affected by f_2 and vice verse. Fault mode f_4 is isolable from all other fault modes since comparing y_1 and y_2 gives a test that is sensitive only to f_4.

The deterministic analysis gives a rather coarse description of the diagnosability properties of the set of residuals. Next,

TABLE I
A DETERMINISTIC DETECTABILITY AND ISOLABILITY ANALYSIS OF (1).

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

978-1-61284-799-3/11/$26.00 ©2011 IEEE

the analysis will be extended to also take the uncertainties, modeled by the stochastic variables ε_i, into account. First, consider a set of residuals based on (1),

$$\begin{aligned} r_1 &= y_1 - 2u = 2f_2 + f_3 + f_4 + 2\varepsilon_1 + \varepsilon_2 + \varepsilon_3 \\ r_2 &= y_2 - 2u = 2f_2 + f_3 + 2\varepsilon_1 + \varepsilon_2 + \varepsilon_4 \\ r_3 &= y_1 - y_2 = f_4 + \varepsilon_3 - \varepsilon_4. \end{aligned} \tag{2}$$

The residuals in (2) are consistent with the results in Table I which states that no test can detect f_1, no test can detect f_2 but not f_3 or vice verse, and it is possible to isolate f_4.

By taking noise into consideration when analyzing the residuals, isolability and detectability performance can be quantified using fault to noise ratio (FNR). FNR is, in the normally distributed case, the ratio between fault influence of the system, λ, and the standard deviation, σ, of the noise, see e.g. [7]. For simplicity in this example, for each fault pair the highest FNR of the residuals which isolate the fault pair is used to quantify isolability. For example, both residuals r_1 and r_3 can be used to detect fault f_4, but residual r_1 has a higher FNR than r_3. The FNR of r_1 with respect to f_4 is

$$\text{FNR}_{r_1} = \frac{\lambda}{\sigma} = \frac{f_4}{\sqrt{\text{var}(2\varepsilon_1 + \varepsilon_2 + \varepsilon_3)}} = \frac{1}{\sqrt{1.5}} \approx 0.82.$$

The residual r_2, which isolates f_2 and f_3 from f_4, has a FNR approximately 1.63 and 0.82 respectively for the two faults. Thus a quantified isolability analysis using FNR of the residuals (2) is summarized in Table II. A fault not isolable from another fault is quantified as 0 and a high FNR value represents an easily detectable fault. Table II gives more detailed information of isolability performance compared to Table I, for example that f_2 and f_3 are easier to isolate from f_4 than the opposite using the residuals (2).

III. PROBLEM FORMULATION

The example in Section II showed the advantages of taking noise into consideration to give more detailed information of detectability and isolability performance. The example illustrated how to analyze a set of given residuals. Now, the approach will be generalized to analyze the model equations directly. Here, a static model in the form

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

is considered where $z \in \mathbb{R}^k$ are known variables, $x \in \mathbb{R}^l$ are unknown variables, $f \in \mathbb{R}^n$ are additive faults, and $e \sim \mathcal{N}(0, \Lambda)$ is a normally distributed random vector with zero mean and a symmetric positive definite covariance matrix $\Lambda \in \mathbb{R}^{m \times m}$. If q is the number of equations, the model matrices have dimensions $H \in \mathbb{R}^{q \times l}$, $L \in \mathbb{R}^{q \times k}$, $F \in \mathbb{R}^{q \times n}$ and $N \in \mathbb{R}^{q \times m}$.

TABLE II

A QUANTIFIED ISOLABILITY ANALYSIS OF $f_i = 1$, FOR $i = 1, 2, 3, 4$, BY COMPUTING FNR OF THE RESIDUALS (2).

	NF	f_1	f_2	f_3	f_4
f_1	0	0	0	0	0
f_2	1.63	1.63	0	0	1.63
f_3	0.82	0.82	0	0	0.82
f_4	0.82	0.82	0.71	0.71	0

The example also indicates that the FNR of different residuals are related to isolability performance. To avoid that all stochastic variables can be eliminated in a residual giving the possibility for infinite FNR:s, it is assumed that

$$\begin{pmatrix} H & N \end{pmatrix} \text{ has full row-rank.} \tag{4}$$

This assumption is for example fulfilled if all sensors have measurement noise. Without loss of generality, it is assumed that the covariance matrix $\bar{\Sigma}$ of variable $N_H N e$ equal the identity matrix, that is

$$N_H N \Lambda N^T N_H^T = I \tag{5}$$

where the rows of matrix N_H forms an orthonormal basis for the left null-space of matrix H. The notation, N_A, will be used henceforth in this paper to denote a matrix where the rows form an orthonormal basis for the left null-space of any matrix A. Assumption (5) is imposed since it will simplify the presentation of the results in the following sections. To see that no generality is lost, note that any model satisfying assumption (4) can be transformed into fulfilling $\bar{\Sigma} = I$ by multiplying (3) from the left with a suitable, invertible, transformation matrix T. The choice of T is not unique and one possibility is

$$T = \begin{pmatrix} \Gamma^{-1} N_H \\ T_2 \end{pmatrix} \tag{6}$$

where Γ is a non-singular factorization matrix satisfying

$$N_H N \Lambda N^T N_H^T = \Gamma \Gamma^T \tag{7}$$

and T_2 is any matrix ensuring invertibility of T. The factorization of (7) is always possible since the left hand side, according to assumption (4), is positive definite and symmetric. Thus, Γ can be found for example from a singular value decomposition.

Deterministic detectability and isolability is a relation over pairs of behavioral modes of the system as can be seen in Table I and this motivates a formal discussion of how modes and their behaviors are described. The behavior of the fault free mode is given by (3) when $f = 0$. If only fault j is present, the behavior is given by (3) when the fault size $f_j \neq 0$ is any non-zero value and $f_i = 0$ for all $i \neq j$.

In a theory of quantified isolability also the size of faults must be considered since faults are compared to noise levels. For example, a large fault is easier to detect than a small one and this should be reflected in the quantified detectability and isolability analysis. Thus the goal of this paper is, given a system description in the form (3), to quantify how easy it is to isolate a fault f_i of size $f_i = \theta$ from another mode f_j with an unknown fault size.

IV. DISTINGUISHABILITY

To solve this problem, a method will be presented for analyzing detectability and isolability performance of a model where noise is taken into account. The method is exemplified on (1) and the results are compared to the example in Section II.

A. Stochastic Characterization of Fault Modes

It proves useful to write (3) in a different form. To motivate the rewrite, a small example is considered.

Example 1: Consider a small model

$$x = u + f$$
$$y = x + \varepsilon$$

with an unknown variable x, a measured variable y, an actuator variable u, a modeled fault f, and a stochastic variable $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Both y and u are known and thus by rewriting the model as,

$$y - u = f + \varepsilon, \tag{8}$$

where x has been eliminated, the left hand side can be used to analyze the fault detectability properties of the model, for example to compute the FNR. $\diamond$

The example shows that by eliminating the unknown variables the relations between the known variables can be used to get information about the modeled FNR. Elimination of x in (3) is achieved by multiplying with N_H from the left, where the rows of N_H are an orthonormal basis that spans the left null-space of H.

Now, the model rewrite can in the general case be written as

$$N_H L z = N_H F f + N_H N e. \tag{9}$$

For any solution z_0, f_0, e_0 to (9) there exists an x_0 such that it also is a solution to (3). Thus no information about the model behavior is lost when rewriting (3) as (9).

Let $r = N_H L z \in \mathbb{R}^d$, which corresponds to the left hand side of (8) in the example, be used to analyze diagnosability performance of the model. The dimension d of the signal r corresponds to the degree of redundancy in the model and is typically $d = q - k$. This representation can be related to the definition of observation sets for the deterministic case, see [11]. Note that the rows in $N_H L$ span all possible residual generators.

The vector r depends on faults and noise, and describes the *behavior* of the model, see [8]. The normal distribution of r has a positive definite covariance matrix $\bar{\Sigma} = N_H \Sigma N_H^T$, if (4) is fulfilled. According to the assumption in Section III, it holds that $\bar{\Sigma} = I$. In (9), all modeled faults f only affect the mean of r. Let $p(r; \mu)$ be the probability density function, pdf, describing r defined as

$$p(r; \mu) = \frac{1}{|2\pi|^{\frac{d}{2}}} \exp\left(-\frac{1}{2}(r-\mu)^T(r-\mu)\right) \tag{10}$$

which is the multivariate normal distribution with unit covariance matrix. The set of pdf's of r, representing the different fault sizes of f_i that can be explained by fault mode f_i, is defined as

$$\mathcal{Z}_{f_i} = \{p(r; \mu) | \exists f_i : \mu = N_H F_i f_i\}, \tag{11}$$

where F_i is the i:th column of F. The fault free mode, NF, is a special case which is only described by a single pdf,

$$\mathcal{Z}_{NF} = \{p_{\text{NF}}\} = \{p(r; 0)\}$$

and corresponds to $f = 0$. Each fault mode f_i results in a set $\mathcal{Z}_{f_i}$. A fixed fault $f_i = \theta$ corresponds to one pdf in $\mathcal{Z}_{f_i}$, denoted

$$p_\theta^i = p(r; N_H F_i \theta). \tag{12}$$

B. Quantified Detectability and Isolability

The difference between the pdf's, $p_{\theta_1}^1$ and $p_{\theta_2}^2$, of r for two faults $f_1 = \theta_1$ and $f_2 = \theta_2$ respectively, can be seen as a measure of isolability. Thus, the isolability of $f_i = \theta$ from a fault mode f_j with unknown fault size can be quantified by the smallest difference between p_θ^i and a pdf $p^j \in \mathcal{Z}_{f_i}$. The Kullback-Leibler information is a measure of the difference between two pdf's, and this measure will be used here.

The Kullback-Leibler information, see [9], between two pdf's p_1 and p_2 is defined as

$$K(p_1 \| p_2) = \int_{-\infty}^{\infty} p_1(v) \log \frac{p_1(v)}{p_2(v)} dv = \mathrm{E}_{p_1}\left[\log \frac{p_1}{p_2}\right] \tag{13}$$

where $\mathrm{E}_{p_1}\left[\log \frac{p_1}{p_2}\right]$ is the expected value of $\log \frac{p_1}{p_2}$ given p_i. Equation (13) is non-symmetric, i.e., $K(p_1 \| p_2) \neq K(p_2 \| p_1)$ in the general case, and has the following properties

$$K(p_1 \| p_2) > 0 \text{ if } p_1 \neq p_2,\ K(p_1 \| p_2) = 0 \text{ if } p_1 = p_2.$$

These properties are consistent with Table II where isolable fault modes have FNR > 0 and the FNR for isolability of f_3 from f_4 is not the same as isolability of f_4 from f_3. The Kullback-Leibler information thus have necessary properties needed for a quantified isolability analysis.

All fault modes are described as multivariate normal pdf's. Thus the Kullback-Leibler information of two multivariate normal pdf's with the same covariance, $p_1 \sim \mathcal{N}(\mu_1, \bar{\Sigma})$ and $p_2 \sim \mathcal{N}(\mu_j, \bar{\Sigma})$, are considered. Then (13) can be written as

$$K(p_1 \| p_2) = \frac{1}{2}\|\mu_1 - \mu_2\|_{\bar{\Sigma}^{-1}}^2. \tag{14}$$

Note that (14) is invariant to linear transformations. Let $p_1^T \sim \mathcal{N}(T\mu_1, T\bar{\Sigma}T^T)$ and $p_2^T \sim \mathcal{N}(T\mu_2, T\bar{\Sigma}T^T)$ where T is a non-singular transformation matrix, then

$$\begin{aligned} K(p_1^T \| p_2^T) &= \frac{1}{2}\|T(\mu_1 - \mu_2)\|_{(T\bar{\Sigma}T^T)^{-1}}^2 = \\ &= \frac{1}{2}\|\mu_1 - \mu_2\|_{\bar{\Sigma}^{-1}}^2 = K(p_1 \| p_2). \end{aligned} \tag{15}$$

By using the stochastic characterization of fault modes in Section IV-A together with the Kullback-Leibler information to measure the distance between a fault $f_i = \theta$ and a fault mode f_j with an unknown fault size, a measure for isolability properties can be defined.

Definition 1 (Distinguishability): Given a static linear model (3) under assumption (4), distinguishability $\mathcal{D}_{i,j}(\theta)$ of a fault $f_i = \theta$ from a fault mode f_j is defined as

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

where the set $\mathcal{Z}_{f_j}$ is defined in (11) and p_θ^i in (12).

Figure 1 shows a graphical interpretation of distinguishability. 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.

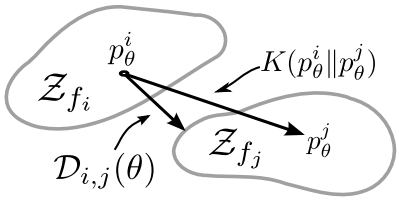

Fig. 1. A graphical visualization where distinguishability represents the smallest difference between $p^i_\theta \in \mathcal{Z}_{f_i}$ and a pdf $p^j \in \mathcal{Z}_{f_j}$.

In [10], it is shown that (13) can be written as

$$K(p^i_\theta \| p^j) = L(p^i_\theta, p^i_\theta) - L(p^i_\theta, p^j) \tag{17}$$

where $L(p^i_\theta, p^i_\theta) = E_{p^i_\theta}[\log p^i_\theta]$ and $L(p^i_\theta, p^j) = E_{p^i_\theta}[\log p^j]$ are log-likelihood functions. Thus minimizing the Kullback-Leibler information is the same as maximizing the maximum likelihood estimate of $p^j \in \mathcal{Z}_{f_j}$ to p^i_θ. Equation (17) gives that if $\exists p^j \in \mathcal{Z}_{f_j}$ such that $K\left(p^i_\theta \| p^j\right) = 0$ then $p^i_\theta \in \mathcal{Z}_{f_j}$, i.e., fault mode f_i can be explained by fault mode f_j and thus f_i is not isolable from f_j.

Example 2: Consider the example in Section II. The model can be written in the form (3) as

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 2 & -1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \tag{18}$$

where $\Lambda = \text{diag}\left(\begin{bmatrix} 0.1 & 0.1 & 1 & 1 \end{bmatrix}\right)$.

To fulfill that the covariance matrix of $N_H N e$ is equal to the identity matrix, (18) is multiplied with a matrix T, defined in (6), from the left. An example is

$$T = \begin{bmatrix} 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 2 & 1 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The analysis, computing $\mathcal{D}_{i,j}(\theta)$ with $\theta = 1$, can be seen in Table III. Comparing Table I and Table III, $\mathcal{D}_{i,j}(\theta) = 0$ corresponds to an X in the deterministic analysis and represents a non-detectable or non-isolable fault pair. A detectable or isolable fault gives $\mathcal{D}_{i,j}(\theta) > 0$ if $|\theta| > 0$, where higher distinguishability represents a fault that is easier to detect or isolate.

TABLE III
DISTINGUISHABILITY OF THE FAULTS FROM THE OTHER FAULTS AND THE FAULT FREE CASE WHEN $f_i = 1$, FOR $i = 1, 2, 3, 4$.

$\mathcal{D}_{i,j}(1)$	NF	f_1	f_2	f_3	f_4
f_1	0	0	0	0	0
f_2	2.000	2.000	0	0	1.333
f_3	0.500	0.500	0	0	0.333
f_4	0.375	0.375	0.250	0.250	0

The results in Table III are consistent with Table II since the calculated FNR:s in Table II give

$$\begin{aligned} \mathcal{D}_{2,4}(1) &= \frac{1}{2}(1.63)^2 = 1.33 \\ \mathcal{D}_{3,4}(1) &= \frac{1}{2}(0.82)^2 = 0.33 \\ \mathcal{D}_{4,2}(1) &= \mathcal{D}_{4,3}(1) = \frac{1}{2}(0.71)^2 = 0.25 \end{aligned} \tag{19}$$

which are the same values as the corresponding elements in Table III. However, distinguishability of f_2, f_3 and f_4, from the fault free case and f_1 is higher in the model analysis compared to the calculated FNR for the set of residuals (2),

$$\begin{aligned} \mathcal{D}_{2,\text{NF}}(1) &= \mathcal{D}_{2,1}(1) = 2 \geq \frac{1}{2}(1.63)^2 = 1.33 \\ \mathcal{D}_{3,\text{NF}}(1) &= \mathcal{D}_{3,1}(1) = 0.5 \geq \frac{1}{2}(0.82)^2 = 0.33 \\ \mathcal{D}_{4,\text{NF}}(1) &= \mathcal{D}_{4,1}(1) = 0.375 \geq \frac{1}{2}(0.82)^2 = 0.33, \end{aligned} \tag{20}$$

The inequalities means that the residuals in Section II is not an optimal set of residuals. These relations are discussed in detail in Section V. The analysis shows that fault f_2 is easiest to detect and fault f_4 is hardest to detect. Isolability of f_2, f_3, and f_4 from f_1 is equal to their detectability because $\mathcal{Z}_{f_1} = \mathcal{Z}_{NF}$. ◇

The example shows that distinguishability gives a quantified isolability analysis by analyzing the model instead of the set of residuals used in Section II.

C. Computation of Distinguishability

To compute (16), an explicit expression of $\mathcal{D}_{i,j}(\theta)$ is provided by the following result.

Theorem 1: The distinguishability for a static linear model (3) under assumption (5) is given by

$$\mathcal{D}_{i,j}(\theta) = \frac{1}{2}\|N_{\bar{H}} F_i \theta\|^2 \tag{21}$$

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

Before proving Theorem 1, note that the distinguishability for a general model in the form (3) under assumption (4) can be computed by:

1) applying the distinguishability invariant transformation (6),
2) redefining the matrices L, H, F, and N given the transformed model fulfilling assumption (5), and
3) computing the distinguishability using (21).

Note also that distinguishability is proportional to the square of the fault size, i.e., $\mathcal{D}_{i,j}(\theta) \propto \theta^2$. By varying θ, $\mathcal{D}_{i,j}(\theta)$ can be used to analyze faults of different sizes which is more realistic than just assuming a fixed fault size. For example, by choosing θ as the minimum fault size required to be detected, the minimum distinguishability is computed.

To prove Theorem 1, the following lemma 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

$$\min_x \|Ax - b\|^2 = \|N_A b\|^2. \tag{22}$$

where the rows of N_A is an orthonormal basis for the left null space of A.

Proof: Minimizing the left hand side of (22) is equivalent to projecting b into the orthogonal complement of A, Ker(A), with the projection matrix $P = N_A^T N_A$. This gives that

$$\min_x \|Ax - b\|^2 = \|Pb\|^2 = b^T Pb = b^T N_A^T N_A b = \|N_A b\|^2.$$

■

Theorem 1 can now be proved using Lemma 1.

Proof: The set $\mathcal{Z}_{f_j}$ is parametrized by f_j, thus minimizing (16) with the respect to $p^j \in \mathcal{Z}_{f_j}$ is equal to

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

Assumption (5) gives that $\bar{\Sigma} = I$. Then, with $\bar{H} = \begin{pmatrix} H & F_j \end{pmatrix}$,

$$\begin{aligned}&\min_{f_j} \frac{1}{2}\|N_H F_i \theta - N_H F_j f_j\|^2_{\bar{\Sigma}^{-1}} = \\ &= \min_{f_j} \frac{1}{2}\|N_H (F_i \theta - F_j f_j)\|^2 = \\ &= \min_{f_j, x} \frac{1}{2}\|Hx - F_i\theta + F_j f_j\|^2 = \\ &= \min_{f_j, x} \frac{1}{2}\|\bar{H}\begin{pmatrix} x \\ f_j \end{pmatrix} - F_i\theta)\|^2 = \\ &= \frac{1}{2}\|N_{\bar{H}} F_i \theta\|^2\end{aligned}$$

where Lemma 1 is used in the second and fourth equalities. ■

V. RELATION TO RESIDUAL GENERATORS

A residual generator of a static model is any function of the known variables z with zero mean in the fault free case. A residual generator that isolates fault f_i from f_j, is a residual that detects f_i but is not sensitive to fault f_j. To design a residual generator isolating faults from fault mode f_j, multiply (3) from the left with $\gamma N_{(H\,F_j)}$ where γ is a row-vector to obtain

$$\gamma N_{(H\,F_j)} L z = \gamma N_{(H\,F_j)} F f + \gamma N_{(H\,F_j)} N e \qquad (23)$$

Here, $\gamma N_{(H\,F_j)} L z$ is a residual generator that isolates *from* fault mode f_j. If only detectability, and not isolability, of f_i is considered, $N_{(H\,F_j)}$ is replaced by 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. Note that (23) is in the same form as (3) and can be seen as a scalar model. Therefore distinguishability can directly be used to analyze isolability performance of a residual. A superscript γ is used, $\mathcal{D}^\gamma_{i,j}(\theta)$, to emphasize that it is distinguishability of a residual with a given γ.

Theorem 2: A residual (23), for a model (3) under assumption (4), is normally distributed $\mathcal{N}(\lambda(\theta), \sigma^2)$ and

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

where θ is the size of fault f_i, and $\lambda(\theta)/\sigma$ is the fault to noise ratio with respect to fault f_i in (23).

Proof: Assumption (4) on the model (3) directly implies that (4) is fulfilled also for the residual generator (23). However, there is no guarantee that (23) fulfills (5) and the 3 step procedure after Theorem 1 must be used. After the transformation, the model is

$$\underbrace{\frac{\gamma N_{(H\,F_j)} L}{\sigma}}_{=:L} z = \underbrace{\frac{\gamma N_{(H\,F_j)} F}{\sigma}}_{=:F} f + \underbrace{\frac{\gamma N_{(H\,F_j)} N}{\sigma}}_{=:N} e \qquad (24)$$

where σ is the standard deviation of the residual in (23). Note that the matrices L, F, and N are redefined in (24) and the new H is the empty matrix. Model (24) fulfills (5) and Theorem 1 gives that

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

where it has been used that $N_{\bar{H}} = 1$ in the first equality. ■

Theorem 2 shows a direct relation between FNR in a residual isolating fault f_i from fault f_j and the distinguishability $\mathcal{D}^\gamma_{i,j}(\theta)$ for the residual. This also confirms the connection between Table II and Table III.

Distinguishability can be used to analyze isolability performance of both the model (3) and the residual generator made of the model (23). The relation between $\mathcal{D}^\gamma_{i,j}(\theta)$ and $\mathcal{D}_{i,j}(\theta)$ is described by the following theorem.

Theorem 3: For a model (3) under assumption (5), an upper bound for $\mathcal{D}^\gamma_{i,j}(\theta)$ in (23) is given by

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

with equality if and only if γ and $N_{(H\,F_j)} F_i$ are parallel.

Proof: Since both N_H and $N_{(H\,F_j)}$ define orthonormal bases and the row vectors of $N_{(H\,F_j)}$ are in the span of the row vectors of N_H, there exists an α such that $N_{(H\,F_j)} = \alpha N_H$ and

$$I = N_{(H\,F_j)} N^T_{(H\,F_j)} = \alpha N_H N_H^T \alpha^T = \alpha\alpha^T$$

Using this result and assumption (5), the variance σ^2 in Theorem 2 can be written as

$$\begin{aligned}\sigma^2 &= \gamma N_{(H\,F_j)} N \Lambda N^T N^T_{(H\,F_j)} \gamma^T = \\ &= \gamma\alpha N_H N \Lambda N^T N_H^T \alpha^T \gamma^T = \gamma\gamma^T\end{aligned}$$

Finally, Cauchy-Schwarz inequality gives

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

with equality if and only if γ and $N_{(H\,F_j)} F_i$ are parallel. ■

Theorem 3 shows that an optimal residual for isolating a fault mode f_i from a fault mode f_j is obtained if $\gamma = kN_{(H\,F_j)} F_i$ for any non-zero scalar k. Such residual has the highest FNR of fault f_i that any residual decoupling f_j can have. To implement a diagnosis algorithm with optimal isolability to isolate n single faults from each other thus requires at most n^2 tests. It is also shown that distinguishability of a residual can never exceed the distinguishability of the model.

Example 3: The residual r_2 in (2) is the only test which can isolate a fault f_2 from a fault f_4. Therefore, r_2 is an optimal test which was also confirmed in (19). There are more possibilities to create a residual which detects f_2, for example both r_1 and r_2 in (2). The inequalities in (20) states that none of the residuals in (2) has maximum distinguishability compared to Table III. An optimal residual to detect f_2 is

$$r_4 = y_1 + y_2 - 4u = 4f_2 + 2f_3 + f_4 + 4\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3 + \varepsilon_4.$$

The fault to noise ratio with respect to f_2 is here

$$\text{FNR}_{r_4} = \frac{4}{\sqrt{4^2 \cdot 0.1 + 2^2 \cdot 0.1 + 1 + 1}} = 2$$

which is equivalent to the corresponding model property in Table III. Note that this residual also is optimal to detect f_3. Similarly, an optimal residual to detect f_4 is

$$r_5 = 3y_1 - y_2 - 4u = 4f_2 + 2f_3 + 3f_4 + 4\varepsilon_1 + 2\varepsilon_2 + 3\varepsilon_3 - \varepsilon_4.$$

since it has maximum distinguishability. ◇

VI. DIESEL ENGINE MODEL ANALYSIS

Distinguishability, as a measure of quantified isolability, has been defined for a static linear model (3). Many industrial systems contain dynamic behavior and non-linearities. The purpose here is to show how the theory can be used also to analyze dynamic nonlinear models.

A. Model Description

The considered model is an industrial model of a heavy duty diesel engine. The model is documented in [12] and an overview is shown in Fig. 2. The model has 11 internal states; four actuators: fuel injection u_δ, valve control 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 in [12] has been extended with 13 possible faults. The faults are briefly described in Table IV and can be divided into four groups: $f_1, \ldots, f_4$ are actuator faults, $f_5, \ldots, f_8$ are sensor faults, $f_9, \ldots, f_{12}$ are leakages, and f_{13} is degraded efficiency of the compressor. Actuator faults and sensor faults are modeled as additive faults. Leakage flow is modeled as proportional to the pressure difference over the leaking hole. Degraded efficiency of the compressor is modeled as as an additive negative fault to the compressor efficiency map. The fault sizes $f_i = \theta_i$ have been selected in the order of 10% of a nominal value of the corresponding variable.

Uncertainties must be introduced in the model, and it is important how it is made because it can greatly affect the result of the analysis. In this case, process noise, actuator noise, and measurement noise have been modeled as independent additive normally distributed noise. The standard deviations of the process noise are selected proportional to the uncertainties in the model according to [12]. The standard deviation of actuator noise is 5% of maximum value and sensor noise is 5% of a nominal value.

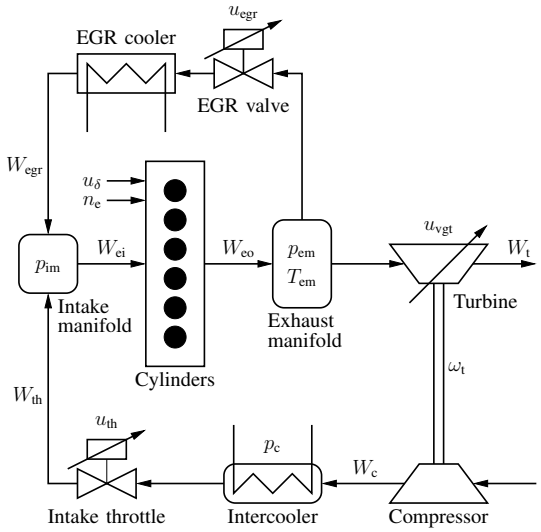

Fig. 2. Overview of diesel engine model.

TABLE IV
IMPLEMENTED FAULTS IN THE DIESEL ENGINE MODEL

Fault	Enters model	Description
f_1	$u_\delta = u_\delta^{\text{nom}} + f_1$	Actuator fault
f_2	$u_{\text{egr}} = u_{\text{egr}}^{\text{nom}} + f_2$	Actuator fault
f_3	$u_{\text{vgt}} = u_{\text{vgt}}^{\text{nom}} + f_3$	Actuator fault
f_4	$u_{\text{th}} = u_{\text{th}}^{\text{nom}} + f_4$	Actuator fault
f_5	$y_{\omega_t} = y_{\omega_t}^{\text{nom}} + f_5$	Sensor fault
f_6	$y_{p_{\text{em}}} = y_{p_{\text{em}}}^{\text{nom}} + f_6$	Sensor fault
f_7	$y_{p_{\text{im}}} = y_{p_{\text{im}}}^{\text{nom}} + f_7$	Sensor fault
f_8	$y_{W_c} = y_{W_c}^{\text{nom}} + f_8$	Sensor fault
f_9	$W_{c,\text{leak}} = f_9(p_c - p_{\text{atm}})$	Leakage
f_{10}	$W_{\text{egr,leak}} = f_{10}(p_{\text{em}} - p_{\text{atm}})$	Leakage
f_{11}	$W_{\text{th,leak}} = f_{11}(p_c - p_{\text{atm}})$	Leakage
f_{12}	$W_{t,\text{leak}} = f_{12}(p_{\text{em}} - p_{\text{atm}})$	Leakage
f_{13}	$\eta_c = \eta_c^{\text{nom}} - f_{13}$	Degraded efficiency

B. Diagnosability Analysis of the Model

The dynamic nonlinear diesel engine model will be analyzed to see how distinguishability of the faults varies with the operating point of the engine. The analysis is made by computing distinguishability when the model is linearized at different linearization points. The points are selected from the World Harmonized Transient Cycle (WHTC), which covers the whole operating range of the engine, see [13]. WHTC is used world-wide in the certification of heavy duty diesel engines.

Distinguishability depends on all actuators and states of the model. To simplify visualization of how distinguishability varies, it is compared to a single closely connected state variable or actuator signal. As an example, Fig. 3 shows the distinguishability of a leakage after the compressor, f_9, from the fault free case against the pressure after the compressor. To clarify how the model behavior affects distinguishability, the y-axis shows the square root of distinguishability. The linear relation between $\sqrt{\mathcal{D}_{9,\text{NF}}(\theta_9)}$ and p_c is reasonable since the leakage flow is proportional to $p_c - p_{\text{atm}}$, see Table IV, where p_{atm} is assumed constant.

Fig. 4 shows an example where distinguishability of an

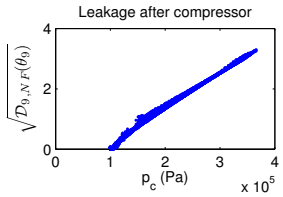

Fig. 3. The square root of distinguishability of a leakage after the compressor is proportional to the pressure in the compressor.

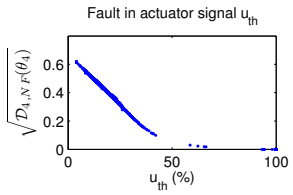

Fig. 4. The square root of distinguishability of an additive fault in the actuator signal u_{th} is better with a lower value of u_{th}.

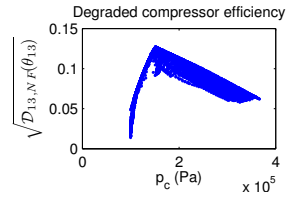

Fig. 5. The square root of distinguishability of degradation of compressor efficiency to the pressure after the compressor.

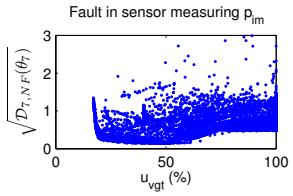

Fig. 6. The square root of distinguishability of an additive fault in the sensor measuring p_{im} compared to u_{vgt}.

additive fault in the actuator signal u_{th} from the fault free case is compared to the amplitude of u_{th}. Distinguishability of the fault is high when the actuator signal is low. Note that comparison of Fig. 3 and Fig. 4 shows that maximum distinguishability of the leakage is higher than the maximum distinguishability of the actuator fault. This implies that it should be easier to detect a leakage after the compressor than a fault in the throttle control signal.

There are also cases where the relation between distinguishability of a fault and a state variable or actuator signal is not as simple as in Fig. 3 and Fig. 4. Fig. 5 shows distinguishability of compressor efficiency degradation from the fault free case where the distinguishability has a maximum at $p_c \approx 1.5 \cdot 10^5$ Pa. Distinguishability goes to zero when $p_c \approx 101$ kPa which is reasonable because $p_c \approx p_{\text{atm}}$ implies no flow in the compressor and then a degradation cannot be detected.

Fig. 6 shows distinguishability of an additive sensor fault measuring p_{im} from the fault free case compared to the actuator signal u_{vgt}. Distinguishability of the sensor fault does not have a clear relation to u_{vgt}, but compared to Fig. 5 it is always higher than the distinguishability of a degradation of the compressor and should thus be easier to detect.

All examples above concern detectability of faults from the fault free case. An example of analyzing distinguishability of a fault from another fault can be seen in Table V. The table shows distinguishability of a fault in the fuel injector, f_1, from each of the other faults during idling. The highest distinguishability value of f_1 is from f_2 and lowest value from f_{12}. An fault f_1 is thus easiest to isolate from f_2 and hardest to isolate from f_{12} during idling.

TABLE V

DISTINGUISHABILITY OF A FAULT IN THE FUEL INJECTION, f_1, FROM EACH OF THE OTHER FAULTS. THE VALUES ARE $\times 10^{-3}$.

10^{-3}	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}
f_1	0.0	8.1	6.8	6.7	7.5	4.0	8.0	7.9	8.0	7.7	6.7	1.4	6.8

VII. CONCLUSIONS

The basic question discussed in this paper is how to quantify diagnosability properties of a given model. Here, static linear models with normally distributed uncertainties are considered but the developed framework is general and is applied to a non-linear dynamic model.

A key contribution is the definition of *distinguishability* which is based on a distance measure of probability distributions of observations for different faults. This is formalized using a stochastic characterization of the system behavior under different fault modes. Distinguishability is a *model* property which means that diagnosis performance can be evaluated without designing any diagnosis algorithms. Based on the definition, a simple computational algorithm is developed to efficiently compute distinguishability. Another key contribution is an analysis of the relation between distinguishability and residual generators. It is proved that distinguishability has a direct relation to the fault to noise ratio in a residual generator and that the model property forms a tight upper bound on fault to noise ratio for any residual generator.

An industrial sized model of a diesel engine for heavy trucks is used to evaluate and exemplify applications of the developed theory and algorithms. Here, as an example, non-trivial results are derived on how detectability and isolability performance varies with the operating point of the diesel engine.

REFERENCES

[1] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: Theory and application*, Prentice Hall, Eaglewood Cliffs, 1993.

[2] S. M. Kay, *Fundamentals of statistical signal processing : detection theory*, Prentice Hall, 1998.

[3] X. Pucel, W. Mayer and M. Stumptner, "Diagnosability analysis without fault models", *Proc. 20st International Workshop on the Principles of Diagnosis (DX-09)*, 2009.

[4] E. Frisk, A. Bregon, J. Åslund, M. Krysander, B. Pulido and G. Biswas, "Diagnosability Analysis Considering Causal Interpretations for Differential Constraints", *Proc. 21st International Workshop on the Principles of Diagnosis (DX-10)*, 2010.

[5] L. Travè-Massuyés, T. Escobet and X. Olive, "Diagnosability analysis based on component supported analytical redundancy relations", *IEEE Trans. Syst. Man Cy. A.*, 36(6), 2006.

[6] E. Frisk, M. Krysander and J. Åslund, "Sensor Placement for Fault Isolation in Linear Differential-Algebraic Systems", 2008.

[7] F. Gustafsson, "Stochastic fault diagnosability in parity spaces", *Proc. 15th Triennial IFAC World Congress on Automatic Control*, 2002.

[8] J. W. Polderman and J. C. Willems, *Introduction to Mathematical Systems Theory - A Behavioral Approach*, Springer-Verlag, New York, 1998.

[9] S. Kullback and R. A. Leibler, "On Information And Sufficiency", 1951.

[10] S. Eguchi and J. Copas, "Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma", *Journal of Multivariate Analysis*, 97(9), 2034-2040, 2006.

[11] M. Nyberg and E. Frisk, "Residual Generation for Fault Diagnosis of Systems Described by Linear Differential-Algebraic Equations", *IEEE Transactions on Automatic Control*, 51(12), 1995-2000, 2006.

[12] J. Wahlström and L. Eriksson, "Modeling of a Diesel Engine with Intake Throttle, VGT, and EGR", 2010.

[13] Economic Commission for Europe – Inland Transport Committee, "Regulation No 49 of the Economic Commission for Europe of the United Nations (UN/ECE)", *Official Journal of the European Union*, August 2010.