Quantitative Fault Diagnosability Performance of Linear Dynamic Descriptor Models

Daniel Eriksson, Mattias Krysander, and Erik Frisk
Department of Electrical Engineering, Linköping University, Linköping, SE-581 83, Sweden
{daner,matkr,frisk}@isy.liu.se

Abstract

A theory is developed for quantifying fault detectability and fault isolability properties of time discrete linear dynamic models. Based on the model, a stochastic characterization of system behavior in different fault modes is defined and a general measure, called distinguishability, based on the Kullback-Leibler information, is used to quantify the difference between the modes. An analysis of distinguishability as a function of the number of observations is discussed. This measure is also shown to be closely related to the fault to noise ratios in residual generators. Further, the distinguishability of the model is shown to give upper limits of the fault to noise ratios of residual generators.

1 INTRODUCTION

Diagnosis and supervision of industrial systems concerns detecting and isolating faults that occurs in the system. When developing a diagnosis system, knowledge of achievable diagnosability performance given the model of the system, such as detectability and isolability, is useful. Such information tells for example if a test with certain properties can be created or if more sensors are needed to get satisfactory diagnosability performance.

A main limiting factor of diagnosis performance is the model uncertainty. Large uncertainties makes it difficult to detect or isolate a small fault. Without sufficient information of possible diagnosability properties, time could be wasted on, for example, developing tests to detect a fault that in reality is impossible to detect or isolate.

There exist systematic methods for analyzing isolability performance in dynamic systems, e.g., (Pucel *et al.*, 2009), (Frisk *et al.*, 2010) and (Travè-Massuyés *et al.*, 2006), however these approaches are deterministic and only give qualitative statements whether a fault is isolable or not. The methods tell nothing of how difficult it is to detect or isolate the faults. 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.

There are several works describing methods from classical detection theory, for example the books (Basseville and Nikiforov, 1993) and (Kay, 1998), which can be used for quantified detectability analysis using a stochastic characterization of faults. A main contribution with respect to these works is that here, isolability performance is also considered.

A framework for quantified diagnosability analysis was introduced in (Eriksson *et al.*, 2011). The method uses a stochastic characterization of noise and can analyze both detectability and isolability. It was also shown that the analysis could be used to design residuals which maximized the fault to noise ratio, FNR. The paper was focusing on analyzing linear static models but the method was also exemplified on a nonlinear dynamic model of a diesel engine. This paper extends the static results in (Eriksson *et al.*, 2011) and presents a theory for quantified isolability analysis of time discrete linear *dynamic* models.

2 PROBLEM FORMULATION

The purpose here is to develop a method for quantified diagnosability analysis of time discrete linear dynamic models. The linear dynamic model is represented by a time-discrete descriptor model

$$\begin{aligned} Ex_{k+1} &= Ax_k + B_u u_k + B_f f_k + B_v v_k \\ y_k &= Cx_k + D_u u_k + D_f f_k + D_\varepsilon \varepsilon_k \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^{l_x}$ are state variables, $y_k \in \mathbb{R}^{l_y}$ are measured signals, $u_k \in \mathbb{R}^{l_u}$ are input signals, $f_k \in \mathbb{R}^{l_f}$ are modeled faults, $v_k \sim \mathcal{N}(0, \Lambda_v)$ and $\varepsilon_k \sim \mathcal{N}(0, \Lambda_\varepsilon)$ are white Gaussian distributed random vectors with zero mean and symmetric positive definite covariance matrices $\Lambda_v \in \mathbb{R}^{l_v \times l_v}$ and $\Lambda_\varepsilon \in \mathbb{R}^{l_\varepsilon \times l_\varepsilon}$. The model matrices are of appropriate dimensions. Note that matrix E can be singular. It is assumed that noise-free residuals can not be created. One sufficient criteria is that

$$D_\varepsilon \text{ has full row-rank and } \lambda E - A \text{ has full rank,} \tag{2}$$

i.e. all sensors have measurement noise and the model has a unique solution for a given initial state.

To describe the behavior of the system, the term fault mode is used. A fault mode represents whether 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. The mode when no fault is present, i.e. $f = 0$, is denoted NF.

A small example illustrates detectability and isolability performance in the deterministic case before presenting the problem formulation.

Example 1. Consider a simple linear dynamic model

$$\begin{aligned} x_{k+1} &= u_k + f_k^u \\ y_k^1 &= x_k + f_k^1 + \varepsilon_k^1 \\ y_k^2 &= x_k + f_k^2 + \varepsilon_k^2 \end{aligned} \tag{3}$$

where f_k^u is an actuator fault, f_k^1 and f_k^2 are sensor faults and $\varepsilon_k^1, \varepsilon_k^2 \sim \mathcal{N}(0, \sigma^2)$ are white Gaussian noise with variance σ^2. It can be shown that all faults are detectable and isolable from each other by creating the following three residuals

$$\begin{aligned} r_1 : y_{k+1}^1 - u_k &= f_k^u + f_k^1 + \varepsilon_k^1 \\ r_2 : y_{k+1}^2 - u_k &= f_k^u + f_k^2 + \varepsilon_k^2 \\ r_3 : y_k^1 - y_k^2 &= f_k^1 - f_k^2 + \varepsilon_k^1 - \varepsilon_k^2. \end{aligned} \tag{4}$$

If a fault is present then the corresponding residuals will have a non-zero mean. If f^u is present then $f^u \neq 0$ then r_1 and r_2 will have non-zero mean and r_3 has zero mean. This can be explained by fault mode f^u but not by fault modes f^1 or f^2 and thus f^u is isolable from the other fault modes. The same can be stated about the other faults f^1 and f^2. ◇

The example above illustrates detectability and isolability performance using a set of residuals. How well the residuals (4) will detect the faults depend on the ratio between fault size and the standard deviation of the noise which is called the fault to noise ratio, FNR. The FNR is defined as

$$\text{FNR} = \frac{\lambda(\theta)}{\sigma} \tag{5}$$

where $\lambda(\theta)$ is the fault amplification of the residual, θ is the fault amplitude and σ is the standard deviation of the noise, see (Eriksson *et al.*, 2011). If the noise has a large variance in relation to the fault size then it is harder to detect the fault compared to a small noise variance.

The purpose here is, given a model (1), to quantify how difficult it is to isolate a fault f_i described by a fault vector $\bar{f}_i = (\theta_{k-n+1}, \ldots, \theta_k)$ from another mode f_j with an unknown fault vector. Thus, the main objective is to analyze quantified detectability and isolability performance directly from the model (1) and not from a set of residuals (4).

3 BACKGROUND THEORY

Before addressing the main goal in this paper, some background theory is introduced. First some definitions on detectability and isolability are discussed. Then, to develop the method for quantified diagnosability analysis for time-discrete linear dynamic models, results used for static linear models in (Eriksson *et al.*, 2011) are presented.

3.1 Diagnosability Properties

Detectability and isolability performance was defined in Section 2 as whether a test with a certain property can be created or not. This is used for example in (Chen and Patton, 1994). The diagnosability properties can also be considered as model properties, see for example (Massoumnia *et al.*, 1989), (Nikoukhah, 1989) and (Frisk *et al.*, 2009). In (Frisk *et al.*, 2009) detectability and isolability are defined as follows: Consider a deterministic model and a set of observations z. A fault f_i is detectable if and only if

$$\mathcal{O}(f_i) \not\subseteq \mathcal{O}(NF)$$

where $\mathcal{O}(f_i)$ is the set of all observations z consistent with the fault mode f_i. Thus a detectable fault will result in observations z which can not be explained by a fault free system. In the same way, a fault f_i is isolable from another fault f_j if and only if

$$\mathcal{O}(f_i) \not\subseteq \mathcal{O}(f_j).$$

One limitation is that the definitions do not state how difficult it is for a developed diagnosis system to fulfill these properties because they do not take model uncertainties into consideration.

3.2 Distinguishability of Linear Static Models

In (Eriksson *et al.*, 2011), static linear models written in the form

$$Lz = Hx + Ff + Ne, \tag{6}$$

were considered where $z \in \mathbb{R}^{l_z}$ are known variables, $x \in \mathbb{R}^{l_x}$ are unknown variables, $f \in \mathbb{R}^{l_f}$ are additive faults and $e \sim \mathcal{N}(0, \Lambda)$ is a white Gaussian distributed random vector with zero mean and a symmetric positive covariance matrix $\Lambda \in \mathbb{R}^{l_e \times l_e}$. If the number of equations in (6) is b, the matrices have dimensions $L \in \mathbb{R}^{b \times l_z}$, $H \in \mathbb{R}^{b \times l_x}$, $F \in \mathbb{R}^{b \times l_f}$ and $N \in \mathbb{R}^{b \times l_e}$.

To guarantee that no noise-free residuals can be created, it is assumed that

$$(H \quad N) \text{ has full row-rank} \tag{7}$$

which corresponds to (2) in the dynamic case. To simplify the computations, it is assumed that the covariance matrix $\bar{\Sigma}$ of variable $N_H L \bar{e}$ is equal to the identity matrix, that is

$$\bar{\Sigma} = N_H N \Lambda N^T N_H^T = I \tag{8}$$

where the rows of N_H forms an orthonormal basis for the left null-space of H. Note that any model satisfying (7) can be transformed into fulfilling $\bar{\Sigma} = I$. The choice of an invertible transformation matrix T is non-unique and one possibility is

$$T = \begin{pmatrix} \Gamma^{-1} N_H \\ T_2 \end{pmatrix} \tag{9}$$

where Γ is non-singular and satisfying

$$N_H N \Lambda N^T N_H^T = \Gamma \Gamma^T \tag{10}$$

and T_2 is any matrix ensuring invertability of T.

It is convenient to eliminate the unknown variables x in (6) by multiplying with N_H from the left such that

$$N_H L z = N_H F f + N_H N e \tag{11}$$

For any solution z_0, f_0, e_0 to (11) there exists an x_0 such that it also is a solution to (6). Thus no information about the model behavior is lost when rewriting (6) as (11).

To be able to make a quantitative statement about detectability and isolability, model uncertainties must be considered. Let $r = N_H L z \in \mathbb{R}^d$, which is the left hand side in (11), be used to analyze diagnosability performance of the model. The vector r describes the behavior of the model and depends on faults and model uncertainties. Let $p(r;\mu)$ be the probability density function, pdf, describing r defined as

$$p(r,\mu) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp(-\frac{1}{2}(r-\mu)^T(r-\mu)) \tag{12}$$

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

$$\mathcal{Z}_{f_i} = \{p(r,\mu) | \exists f_i : \mu = N_H F_i f_i\}, \tag{13}$$

where F_i is the ith column of F. Each fault mode f_i result in a set $\mathcal{Z}_{f_i}$. A fixed fault $f_i = \theta$ corresponds to one pdf in $\mathcal{Z}_{f_i}$ denoted

$$p^i_\theta = p(r, N_H F_i \theta). \tag{14}$$

The difference between the pdf's, $p^i_{\theta_1}$ and $p^j_{\theta_2}$, 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^i_θ and a pdf $p^j \in \mathcal{Z}_{f_j}$. 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 (Kullback and Leibler, 1951), between two pdf's p_1 and p_2 is defined as

$$K(p_1 \| p_2) = \int_{-\infty}^{\infty} p_1(v) \log \frac{p_1(v)}{p_2(v)} dv = \mathrm{E}_{p_1}\left[\log \frac{p_1}{p_2}\right] \tag{15}$$

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_1. Since all fault modes are described as multivariate Gaussian pdf's, the Kullback-Leibler information (15) of two multivariate Gaussian pdf's with the same covariance, $p^1_{\mu_1} = p(r;\mu_1)$ and $p^2_{\mu_2} = p(r;\mu_2)$, can be written as

$$K(p^1_{\mu_1} \| p^2_{\mu_2}) = \frac{1}{2} \|\mu_1 - \mu_2\|^2_{\Sigma^{-1}}. \tag{16}$$

Note that (16) is invariant to linear transformations which allows that the model (6) is multiplied with (9) without affecting the result of (16).

By using the stochastic characterization of fault modes 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 were defined.

Definition 1 (Distinguishability). Given a static linear model (6) under assumption (7), distinguishability $\mathcal{D}_{i,j}(\theta)$ of a fault $f_i = \theta$ from a fault mode f_j is defined as

$$\mathcal{D}_{i,j}(\theta) = \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p^i_\theta \| p^j\right) \tag{17}$$

where the set $\mathcal{Z}_{f_j}$ is defined in (13) and p^i_θ in (14). ■

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.

To compute (17), an explicit expression of $\mathcal{D}_{i,j}(\theta)$ is needed which is provided by the following result.

Theorem 1. The distinguishability for a static linear model (6) under assumption (8) is given by

$$\mathcal{D}_{i,j}(\theta) = \frac{1}{2} \|N_{\bar{H}} F_i \theta\|^2 \tag{18}$$

where $\bar{H} = (H \quad F_j)$ and the rows of $N_{\bar{H}}$ is an orthonormal basis for the left null space of $\bar{H}$. □

The proof of Theorem 1 is omitted here but can be found in (Eriksson *et al.*, 2011).

4 DISTINGUISHABILITY OF TIME DISCRETE DYNAMIC MODELS

The basic idea to analyze time discrete models (1) is to reformulate the dynamic model using a sliding window, see (Gustafsson, 2002), similar to the parity space approach, see (Gertler, 1997). This results in an augmented *static* model on a time window, where the results for static models from Section 3.2 can be applied to compute distinguishability.

4.1 Sliding Window Model

Before analyzing the time-discrete descriptor model (1) it is written as a sliding window model, i.e., a sliding window of length n is applied to (1). Define the vectors

$$\begin{aligned}
\bar{z} &= (y^T_{k-n+1}, \dots, y^T_k, u^T_{k-n+1}, \dots, u^T_k)^T \\
\bar{x} &= (x^T_{k-n+1}, \dots, x^T_{k+1})^T, \bar{f} = (f^T_{k-n+1}, \dots, f^T_k)^T \\
\bar{e} &= (v^T_{k-n+1}, \dots, v^T_k, \varepsilon^T_{k-n+1}, \dots, \varepsilon^T_k)^T,
\end{aligned}$$

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

$$L_n \bar{z} = H_n \bar{x} + F_n \bar{f} + N_n \bar{e} \tag{19}$$

where

$$L_n = \begin{pmatrix}
0 & 0 & \dots & 0 & -B_u & 0 & \dots & 0 \\
I & 0 & & 0 & -D_u & 0 & & 0 \\
0 & 0 & & 0 & 0 & -B_u & & 0 \\
0 & I & & 0 & 0 & -D_u & & 0 \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\
0 & & 0 & 0 & 0 & & 0 & -B_u \\
0 & \dots & 0 & I & 0 & \dots & 0 & -D_u
\end{pmatrix},$$

$$H_n = \begin{pmatrix}
A & -E & 0 & \dots & 0 & 0 \\
C & 0 & 0 & & 0 & 0 \\
0 & A & -E & & 0 & 0 \\
0 & C & 0 & & 0 & 0 \\
\vdots & & \ddots & \ddots & & \vdots \\
0 & 0 & & 0 & A & -E \\
0 & 0 & \dots & 0 & C & 0
\end{pmatrix}, F_n = \begin{pmatrix}
B_f & 0 & \dots & 0 \\
D_f & 0 & & 0 \\
0 & B_f & & 0 \\
0 & D_f & & 0 \\
\vdots & & \ddots & \vdots \\
0 & & 0 & B_f \\
0 & & 0 & D_f
\end{pmatrix},$$

$$N_n = \begin{pmatrix}
B_v & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
0 & 0 & & 0 & D_\varepsilon & 0 & & 0 \\
0 & B_v & & 0 & 0 & 0 & & 0 \\
0 & 0 & & 0 & 0 & D_\varepsilon & & 0 \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\
0 & 0 & & B_v & 0 & 0 & & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & D_\varepsilon
\end{pmatrix}.$$

The sliding window model (19) is a static representation of the dynamic behavior on the window given the time indexes $(k-n+1, ..., k)$.

Like in the static case in Section 3.2, without loss of generality, it is assumed that the covariance matrix $\bar{\Sigma}_n$ of variable $N_{H_n} L_n \bar{e}$ is equal to the identity matrix, that is

$$N_{H_n} N_n \Lambda_n N_n^T N_{H_n}^T = I. \quad (20)$$

This assumption is imposed since it will simplify the computations in the following sections.

Note that any model (19) satisfying (2) can be transformed into fulfilling $\bar{\Sigma}_n = I$. The choice of a non-singular transformation matrix T corresponds to choosing (9) in the static case.

4.2 Augmented Definition of Fault Models

By observing a system during several time samples, not only constant faults but faults that varies over time can be analyzed. A sliding window model (19) describes the behavior of a system over a time window of length n and therefore an augmented definition of fault modes in Section 3.2 is needed.

The sliding window model (19) is in the same form as (11). Multiplying with N_{H_n} from the left eliminates the unknown variables $\bar{x}$ and gives

$$N_{H_n} L_n \bar{z} = N_{H_n} F_n \bar{f} + N_{H_n} N_n \bar{e}. \quad (21)$$

As for (11), no model information is lost when rewriting (19) as (21). Let $r = N_{H_n} L_n \bar{z}$, which is the left hand side in (21). If b' is the total number of equations in (1) then the dimension of r is typically $d' = n(b' - l_y - l_u)$. The vector $r \in \mathbb{R}^{d'}$ depends on the fault vector $\bar{f}$ and the noise $\bar{e}$.

In Section 3.2 when analyzing distinguishability, only one sample is considered. For dynamic models, the fault time profile is affecting detectability and isolability performance and therefore it is interesting to analyze a time window of samples. The fault vector $\bar{f}$ describes how a fault varies over time. The fault free case is described by $\bar{f} = \bar{0} = (0, \ldots, 0)^T$ which is the zero vector. A fault mode f_i represents all fault vectors $\bar{f}_i \neq \bar{0}$ and $\bar{f}_j = \bar{0}$ for all $j \neq i$. Fig. 1 shows examples of possible fault vectors $\bar{f}_i$ representing different time profiles given a fault mode f_i. The definition of faults and fault modes are similar to the static case but considers the behavior during the time period given by the indexes $(k - n + 1, ..., k)$.

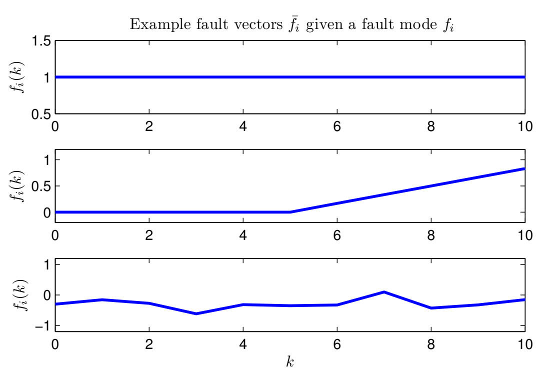

Figure 1: Examples of possible realizations of fault vectors $\bar{f}_i$ given a fault mode f_i.

Let $p(r; \bar{\mu})$ be the probability density function, pdf, describing the vector r defined as

$$p(r; \bar{\mu}) = \frac{1}{(2\pi)^{\frac{d'}{2}}} \exp\left(-\frac{1}{2}(r - \bar{\mu})^T (r - \bar{\mu})\right) \quad (22)$$

which is the multivariate normal distribution with unit covariance matrix. The set of pdf's of r representing the fault mode f_i, corresponding to all fault vectors $\bar{f}_i = (f_{k-n+1}^T, \ldots, f_k^T)^T$ is defined as

$$\mathcal{Z}_{f_i} = \{p(r; \bar{\mu}) | \exists \bar{f}_i : \bar{\mu} = N_{H_n} F_{n,i} \bar{f}_i\} \quad (23)$$

where $F_{n,i}$ are the columns of F_n corresponding to the elements $\bar{f}_i$ in $\bar{f}$. The fault free mode, NF, is a special case which is only described by a single pdf,

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

and corresponds to $\bar{f} = \bar{0}$. Each fault mode f_i result in a set $\mathcal{Z}_{f_i}$. The vector $\bar{\theta}$ will be used to denote a non-constant vector $(\theta_{k-n+1}^T, \ldots, \theta_k^T)^T$ and $\hat{\theta}^c$ will denote a constant vector $(\theta^T, \ldots, \theta^T)^T$

A fixed fault vector $\bar{f}_i = \bar{\theta}$ corresponds to one pdf in $\mathcal{Z}_{f_i}$, denoted

$$p_{\bar{\theta}}^i = p(r; N_{H_n} F_{n,i} \bar{\theta}) \in \mathcal{Z}_{f_i}. \quad (24)$$

The fault models in Section 3.2 can be seen as a special case of the augmented definition here when $n = 1$.

4.3 Distinguishability of sliding window models

By rewriting the dynamic model (1) as a sliding window model (19) of length n, distinguishability for a specific fault vector $\bar{f}_i$ from a fault mode f_j, where $\bar{f}_j$ can be any fault vector of length n, is defined as:

Definition 2 (Distinguishability). Given a sliding window model (19) of length n, under assumption (2), distinguishability $\mathcal{D}_{i,j}(\bar{\theta}; n)$ of a fault vector $\bar{f}_i = \bar{\theta} = (\theta_{k-n+1}, \ldots, \theta_k)$ from a fault mode f_j is defined as

$$\mathcal{D}_{i,j}(\bar{\theta}; n) = \min_{p^j \in \mathcal{Z}_{f_j}} K\left(p_{\bar{\theta}}^i \| p^j\right) \quad (25)$$

where the set $\mathcal{Z}_{f_j}$ is defined in (23) and $p_{\bar{\theta}}^i$ in (24). ■

Fig. 2 shows a graphical interpretation of distinguishability. The measure represents the smallest difference, given the Kullback-Leibler information, between a pdf $p_{\bar{\theta}}^i$, describing the influence of the fault vector $\bar{f}_i$, and all possible pdf's $p^j \in \mathcal{Z}_{f_j}$, that could be described by fault mode f_j.

Since the sliding window model (19) is static, the results of distinguishability in the static case hold.

An explicit computation of (25) follows from Theorem 1 and is stated in the following proposition.

Proposition 1. Distinguishability for a sliding window model (19) under assumption (20) is given by

$$\mathcal{D}_{i,j}(\bar{\theta}; n) = \frac{1}{2} \|N_{\bar{H}_n} F_{n,i} \bar{\theta}\|^2 \quad (26)$$

where $\bar{H}_n = (H_n \quad F_{n,j})$ and the rows of $N_{\bar{H}_n}$ is an orthonormal basis for the left null space of $\bar{H}_n$. □

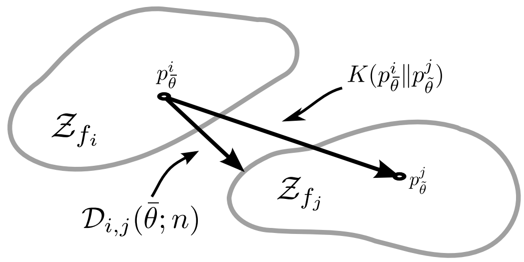

Figure 2: A graphical visualization where distinguishability represents the smallest difference between $p^i_{\bar\theta} \in \mathcal{Z}_{f_i}$ and a pdf $p^j \in \mathcal{Z}_{f_j}$.

Note that it follows from Proposition 1 that if the amplitude of the fault vector $\bar{f}_i = \bar\theta$ is varied by multiplying a constant $c \in \mathbb{R}$ then $\mathcal{D}_{i,j}(c\bar\theta; n) \propto c^2$.

A detectability and isolability analysis of the descriptor model (1) can be made using distinguishability if written as a sliding window model (1). The distinguishability depends on the window length n and the fault time profile defined by the fault vector $\bar{f}_i = \bar\theta$. The following example shows how distinguishability can be used to analyze detectability and isolability of constant faults, $\bar{f}_i = \bar\theta^{\text{c}}$, in a small time discrete dynamic model.

Example 2. An example model of a DC motor used in (Gustafsson, 2002) is considered. The system is sampled with sample interval $t_s = 0.4s$ and is described by a descriptor model (1) where

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0.33 \\ 0 & 0.67 \end{pmatrix}, B_u = \begin{pmatrix} 0.07 \\ 0.33 \end{pmatrix}, B_v = \begin{pmatrix} 0.08 \\ 0.16 \end{pmatrix} \\ B_f &= \begin{pmatrix} 0.07 & 0 \\ 0.33 & 0 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, D_u = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ E &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, D_f = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, D_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{27}$$

The modeled faults are input voltage disturbance, f_1, and a velocity sensor offset, f_2. In the analysis consider a window length of $n = 2$ and fault modes where only one fault is present. The window length $n = 2$ is chosen as the model order. Distinguishability is computed for constant faults $\bar{f}_i = \bar\theta^{\text{c}} = \bar{1}$ from f_j where $\bar{f}_j$ can be any fault vector. Distinguishability between the different fault modes is summarized as the isolability matrix in Table 1. A positive value represents if the fault in the row is isolable from the fault mode f_j in the column. The NF column corresponds to fault detectability, i.e., when $\bar{f}_j = \bar{0}$. A higher distinguishability corresponds to a fault that is easier to detect which gives that f_2 is easier to detect than f_1. Zeros represents that a fault can not be isolated from the other fault mode and thus f_1 can not be isolated from f_2.

The results of using distinguishability on the model when the window length is increased to $n = 5$ is shown in Table 2. Detectability and isolability performance are better than if $n = 2$, which can be interpreted as more measurements should give more information about the state of the system, i.e. isolating a fault should be easier. The fault f_1 is only isolable from f_2 for the longer window length. Thus, isolating f_1 from f_2 requires more measurements than to isolate f_2 from f_1 assuming that the fault is constant. ◇

Table 1: Results from analyzing a sliding window model of length $n = 2$ of the example model (27) using distinguishability. A fault $\bar{f}_1 = \bar{1}$ is not isolable from f_2.

$\mathcal{D}_{i,j}(1;2)$	NF	f_1	f_2
f_1	0.0369	0	0
f_2	0.0720	0.0384	0

Table 2: Results from analyzing a sliding window model of length $n = 5$ of the example model (27) using distinguishability. Distinguishability is higher, compared to Table 1, because the time window is longer. A fault $\bar{f}_1 = \bar{1}$ is now isolable from f_2.

$\mathcal{D}_{i,j}(1;5)$	NF	f_1	f_2
f_1	0.5931	0	0. 0403
f_2	0.9153	0.5794	0

The example illustrates how distinguishability can be used to get a quantified detectability and isolability analysis given the model (27). The analysis shows that some faults are easier to detect and isolate than others and that distinguishability increases with increasing window length n. Notice that the results from the analysis show that fault f_1 is not isolable if the window length is too short, for example when $n = 2$. The analysis shows some interesting properties of distinguishability that will be investigated further in the following sections.

5 ANALYZING DISTINGUISHABILITY OF DIFFERENT FAULT TIME PROFILES

Consider a sliding window model (19) of length n. Distinguishability of a fault f_i from another fault mode f_j depends on the fault time profile, i.e., how the fault varies over time, represented by the vector $\bar{f}_i = \bar\theta$. Analyzing distinguishability of the model (19) tells how difficult it is to detect and isolate a fault f_i depending on how it varies over time. In the following example, distinguishability will be used to analyze how different fault time profiles affect detectability and isolability performance.

Example 3. Consider the example model (27) and a sliding window model of length $n = 5$. Three different fault time profiles, representing how fast a fault enters the system, are analyzed, a step $\bar{f}_i = \bar\theta^{\text{step}} = (0, 0, 1, 1, 1)^T$, a ramp $\bar{f}_i = \bar\theta^{\text{ramp}} = (0, 0.6, 0.8, 1, 1)^T$ and a constant $\bar{f}_i = \bar\theta^{\text{const}} = 0.77 \cdot (1, 1, 1, 1, 1)^T$. The fault vectors are selected to have equal energy, i.e., $\bar{f}_i^T \bar{f}_i$ is equal for all faults.

Distinguishability is computed for each of the fault time profiles. Table 3 shows the computed distinguishability when a fault enters as a step, Table 4 as a ramp and Table 5 as a constant. The result from analyzing the different fault time profiles shows that f_2 is easier to detect if it occurs abruptly and f_1 if it is constant. It is easier to isolate any of the faults f_i if it is constant.

The conclusion for this model is that it is easier to detect a fault f_2 if it enters abruptly but it is easier to detect a fault f_1 or isolate any of the faults from the other if it is constant. ◇

The result of analyzing different fault behaviors using distinguishability, shows that the fault behavior can have a significant impact on detectability and isolability performance. Computing distinguishability is a straight forward approach to analyze detectability and isolability performance for different fault time profiles.

Table 3: Analyzing a sliding window model of length $n = 5$ of the example model (27) using distinguishability when the fault enters as a step.

$\mathcal{D}^{\text{step}}_{i,j}(\bar{1}; 5)$	NF	f_1	f_2
f_1	0.1943	0	0.0119
f_2	1.2313	0.2857	0

Table 4: Analyzing a sliding window model of length $n = 5$ of the example model (27) using distinguishability when the fault enters as a ramp.

$\mathcal{D}^{\text{ramp}}_{i,j}(\bar{1}; 5)$	NF	f_1	f_2
f_1	0.2494	0	0.0176
f_2	1.0861	0.3370	0

Table 5: Analyzing a sliding window model of length $n = 5$ of the example model (27) using distinguishability when the fault is assumed constant.

$\mathcal{D}^{\text{const}}_{i,j}(\bar{1}; 5)$	NF	f_1	f_2
f_1	0.3559	0	0.0242
f_2	0.5492	0.3477	0

6 STATIC VS DYNAMIC MODELS FOR FAULT DIAGNOSIS

An interesting aspect when developing a diagnosis system is the usage of a proper model. Here some non-trivial results are presented when choosing a model to use for fault diagnosis.

Assume for example that a system is modeled as a descriptor model (1). If the system seldom deviates from the equilibrium point, i.e. where $x_{k+1} \approx x_k$, the model (1) can be simplified by replacing x_{k+1} with x_k and thus creating a static model in the form

$$\begin{aligned} Ex_k &= Ax_k + B_u u_k + B_f f_k + B_v v_k \\ y_k &= Cx_k + D_u u_k + D_f f_k + D_\varepsilon \varepsilon_k. \end{aligned} \tag{28}$$

The question is which of the static model (28) and the dynamic model (1) that would give a higher distinguishability. If the static model sufficiently describes the system behavior then there is maybe no advantage in using the dynamic model when creating a diagnosis system. That a static model could be better than a dynamic model to detect faults is non-trivial and the following example illustrates such a case where a static model could be preferable. In the example, FNR is computed to simplify the computations but the same conclusions is made if analyzing distinguishability instead. The relation between distinguishability and FNR will be discussed further in Section 8.

Example 4. Consider a small time discrete dynamic model

$$\begin{aligned} x_{k+1} &= \frac{1}{2}x_k + u_k + f^u_k \\ y_k &= x_k + e_k \end{aligned} \tag{29}$$

where f^u is an actuator fault and $e_k \sim \mathcal{N}(0, \sigma^2)$ is measurement noise. Assume that the system seldom deviates from its equilibrium point, i.e. $x_{k+1} \approx x_k$. A residual shall be designed to detect the fault $f^u = \theta$, using as few time samples as possible. Given the model (29) the following residual can be created:

$$r_a : y_{k+1} - \frac{1}{2}y_k - u_k = f^u_k - \frac{1}{2}e_k + e_{k+1}. \tag{30}$$

The FNR of the residual (30) is computed as

$$\text{FNR}_a = \frac{\theta}{\sqrt{\frac{\sigma^2}{4} + \sigma^2}} = \frac{2}{\sqrt{5}}\frac{\theta}{\sigma}. \tag{31}$$

If also taking into considering that the system is close to the equilibrium point, i.e. $x_{k+1} \approx x_k$, the following residual can be created:

$$r_b : y_k - 2u_k = 2f^u_k + e_k. \tag{32}$$

which has a FNR as

$$\text{FNR}_b = \frac{2\theta}{\sigma} > \frac{2}{\sqrt{5}}\frac{\theta}{\sigma} = \text{FNR}_a$$

which is higher than (31). The residual (30) have a FNR which is higher than (32) and is therefore preferable when detecting f^u.

In this example creating a residual without taking into consideration that the system is close to its equilibrium point requires more measurements and will result in a residual with more noise. Here the residual based on the static model is more preferable to detect f^u than the dynamic model. ◇

The example illustrates that a simplified model can be better for diagnosis purposes, even if the more complex model is better at describing the system behavior, because the more complex model gives lower FNR and distinguishability. Computing distinguishability can be used in a wider sense to compare different models to evaluate which is better when detecting or isolating a certain fault type or behavior.

7 WINDOW LENGTH ANALYSIS

By using more measurements, measurement noise can be reduced and this will improve diagnosability performance. Increasing the window length n of a sliding window model (19) results in increased distinguishability. The relation between distinguishability and window length is discussed in this section.

The following example compares distinguishability of a time discrete dynamic model (1) to the corresponding static model (28), i.e., when $x_{k+1} = x_k$. The fault is assumed constant $\bar{f}_i = \bar{1}$.

Example 5. Consider again the example model (27). Fig. 3 shows the result of analyzing $\mathcal{D}_{i,j}(\bar{\theta}; n)$ for a constant fault $\bar{\theta}^c = \bar{1}$ as a function of window length n. The solid line represents the dynamic model and the dashed line represents the model (27) if $x_{k+1} = x_k$, i.e., it is rewritten as a static model in the form (28) which is observed for a time window of length n. Asymptotically, distinguishability of both models increases linearly. The slopes of the curves in Fig. 3 are shown in Fig. 4 where the distinguishability of the dynamic model converges to the distinguishability of the static model. ◇

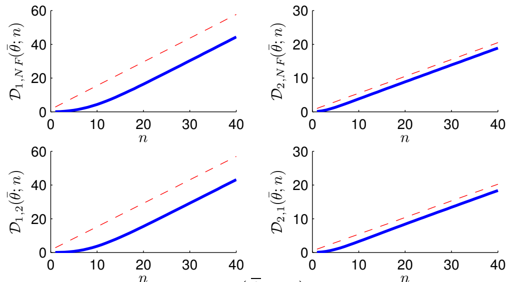

Figure 3: The value of $\mathcal{D}_{i,j}(\bar{\theta}^c;n)$ as a function of n for the example model (27) (solid line) and when (27) is assumed static, i.e. $x_{k+1} = x_k$, (dashed line).

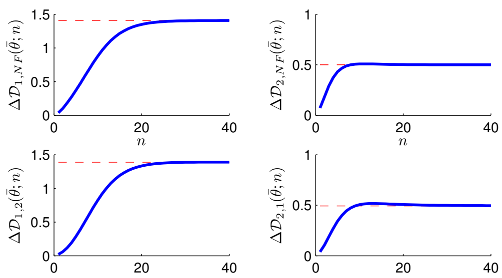

Figure 4: The slope of $\mathcal{D}_{i,j}(\bar{\theta}^c;n)$ as a function of n for the example model (27) (solid line) and when (27) is assumed static, i.e. $x_{k+1} = x_k$, (dashed line). The slope of the dynamic model converges to the slope of the static model.

Comparing the two models in Fig. 3 and Fig. 4, it seems that distinguishability of the faults in the dynamic model converge to the same slopes as for the static model. The asymptotic behavior of $\mathcal{D}_{i,j}(\bar{\theta};n)$ for the dynamic model, assuming a constant fault $\bar{f}_i = \bar{\theta}^c$, when the window length n increases can be written as

$$\lim_{n\to\infty} \mathcal{D}_{i,j}(\bar{\theta}^c;n+1) - \mathcal{D}_{i,j}(\bar{\theta}^c;n).$$

First consider the static model. It can be shown that the asymptotic behavior of distinguishability when the window length increases can be computed as

$$\lim_{n\to\infty} \mathcal{D}_{i,j}(\bar{\theta}^c;n+1) - \mathcal{D}_{i,j}(\bar{\theta}^c;n) = \lim_{n\to\infty} (n+1)\mathcal{D}_{i,j}(\theta) - n\mathcal{D}_{i,j}(\theta) = \mathcal{D}_{i,j}(\theta).$$

Distinguishability increases equally for each time step since the static model has no connection between two different samples due to the white noise assumption.

An analysis of dynamic models is illustrated by a small example. Consider a stable system where only measurement noise $\varepsilon_k \sim \mathcal{N}(0,\Sigma_\varepsilon)$ is considered

$$\begin{aligned} x_{k+1} &= Ax_k + B_u u_k + B_f f_k \\ y_k &= Cx_k + D_\varepsilon \varepsilon_k. \end{aligned} \tag{33}$$

Let q denote the shift operator, $qx_k = x_{k+1}$, then (33) can be written as

$$y_k - C(qI-A)^{-1}B_u u_k = C(qI-A)^{-1}B_f f_k + D_\varepsilon \varepsilon_k$$

where the left hand side corresponds to $r = N_H L z$. Then if $f_k = f$ is constant, the right hand side can be written as

$$r = C(I-A)^{-1}B_f f + D_\varepsilon \varepsilon_k + \text{transient} \tag{34}$$

where the transient goes to zero when $k \to \infty$. If $x_{k+1} = x_k$, i.e. a static model , then (33) is written as

$$y_k - C(I-A)^{-1}B_u u_k = C(I-A)^{-1}B_f f_k + D_\varepsilon \varepsilon_k$$

where the right hand side is equivalent to (34) when the transient goes to zero.

This special case shows that the slope of the distinguishability curve, when n increases, converges to the slope of the corresponding static model. In such a case it is enough to compute distinguishability of the static model, for $n = 1$, to analyze whether the faults are detectable and isolable or not. The result of analyzing the static model also tells how much diagnosability performance will increase asymptotically when the window length is increased.

8 RELATION TO RESIDUAL GENERATORS

The previous sections have discussed how to analyze different diagnosability properties using distinguishability. It will show that distinguishability is related to the maximum FNR of residual generators.

A residual generator of (19) is any function of the known variables $\bar{z}$ with zero mean in the fault free case. A residual generator that isolates f_i from f_j, detects f_i but is not sensitive to f_j. To design a residual generator isolating faults from fault mode f_j multiply (19) with $\gamma N_{(H_n\ F_{n,j})}$ from the left where γ is a row vector to obtain

$$\begin{aligned} &\gamma N_{(H_n\ F_{n,j})} L_n \bar{z} = \\ &\gamma N_{(H_n\ F_{n,j})} F_n \bar{f} + \gamma N_{(H_n\ F_{n,j})} N_n \bar{e}. \end{aligned} \tag{35}$$

where the rows of $N_{(H_n\ F_{n,j})}$ forms an orthonormal basis of the left null-space of $(H_n\ F_{n,j})$, i.e. it decouples f_j. Here, $\gamma N_{(H_n\ F_{n,j})} L_n \bar{z}$ is a residual generator that isolates *from* fault mode f_j. If only detectability, and not isolability, of f_i is considered, $N_{(H_n\ F_{n,j})}$ is replaced by N_{H_n}.

The residual generator (35) can be seen as a scalar model and thus distinguishability can be computed to analyze its performance. The relation between distinguishability and the FNR of the residual generator is described by the following theorem.

Theorem 2. A residual (35), for a model (19) under assumption (7), is $\mathcal{N}(\lambda(\bar{\theta}),\sigma^2)$ and

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

where $\bar{\theta} = (\theta_{k-n+1},\ldots,\theta_k)$ is the fault vector of fault f_i, and $\lambda(\bar{\theta})/\sigma$ is the fault to noise ratio with respect to fault f_i in (35). □

Because of the static model adaptation, a proof of Theorem 2 can be found in (Eriksson *et al.*, 2011). Distinguishability can be used to analyze diagnosability performance of both the model (19) and the residual generator (35). The relation between $\mathcal{D}_{i,j}(\bar{\theta};n)$ and $\mathcal{D}^{\gamma}_{i,j}(\bar{\theta};n)$ is described by the following theorem.

Theorem 3. For a window model (19) under assumption (8), an upper bound for $\mathcal{D}^{\gamma}_{i,j}(\bar{\theta};n)$ in (35) is given by

$$\mathcal{D}^{\gamma}_{i,j}(\bar{\theta};n) \le \mathcal{D}_{i,j}(\bar{\theta};n)$$

with equality if and only if γ and $N_{(H_n\ F_{n,j})} F_{n,i}\bar{\theta}$ are parallel. □

Because of the static model adaptation a proof to Theorem 3 can be found in (Eriksson *et al.*, 2011). Theorem 3 states that the maximum FNR a residual can have is upper bounded by the distinguishability of the model. It also tells how to choose γ to create a residual with maximum FNR. An example will show how the performance of residuals created to have maximum FNR, is related to distinguishability of the model.

Example 6. Data is simulated using the example model (27). A sliding window model of (27) with three different window lengths, $n = 10, 20, 30$, is used. Residuals with maximized FNR for detecting a constant fault $\bar{f}_1 = \bar{\theta}^c$ are created using Theorem 3 for each window length.

The three residuals are evaluated using simulated data. Fig. 5 shows the result of a simulation of the three residuals when the fault enters at 20 s. The residuals are normalized so that the variance is 1 to visualize the FNR. For a longer window length n, the FNR is higher which is shown as a larger deviation in the figure, i.e. f_i is easier to detect. The FNR of the residuals are computed using Theorem 2, and visualized by the horizontal dashed lines in the figure, and are consistent with the deviations of the residuals when the fault is present. $\diamond$

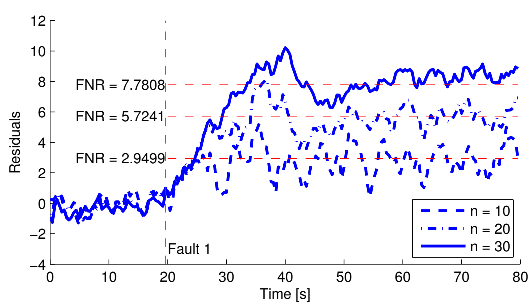

Figure 5: Residuals for different n. A constant fault $\bar{\theta}^c = \bar{1}$ i simulated at 20 s. The maximum FNR of the residual increases with increasing n. The vertical dashed lines represents the computed FNR using distinguishability which are consistent with the deviations of the residuals.

The residuals in the example are selected by choosing γ parallel to $N_{(H_n\ F_{n,j})} F_{n,i} \bar{\theta}^c$, where $\bar{\theta}^c = \bar{1}$, which maximizes the FNR of the residuals for a constant fault $\bar{f}_i$. The results of the example shows that the maximum FNR, computed using distinguishability, are consistent with the performance of the residuals. The computed distinguishability of a model gives information of the maximum FNR performance of a residual given the model and how to design those residuals.

9 CONCLUSIONS

The method, distinguishability, is able to capture useful detectability and isolability properties without the need of first implementing a diagnosis algorithm. The theory of diagnosability is exemplified in (Eriksson *et al.*, 2011) on an industrial sized model of a diesel engine for heavy trucks. The definition is general but is here adapted to time discrete linear dynamic models with white Gaussian noise.

The framework of analyzing static models using distinguishability is applicable to time discrete linear dynamic models. By writing the model as a sliding window model distinguishability can be analyzed as a function of window length and fault behavior.

Distinguishability can also be used to compare different models, for example a dynamic and a static model, to see which will give the best potential diagnosability performance for a given fault signature and window length.

The connection between residual generators and distinguishability can be used to get information of how to create an residual which maximizes the FNR for a specific fault signature.

REFERENCES

(Basseville and Nikiforov, 1993) M. Basseville and I. V. Nikiforov. *Detection of abrupt changes: Theory and application*. Prentice Hall, 1993.

(Chen and Patton, 1994) J. Chen and R. J. Patton. A reexamination of fault detectability and isolability in linear dynamic systems. In *Fault Detection, Supervision and Safety for Technical Processes*, 1994.

(Eriksson *et al.*, 2011) D. Eriksson, M. Krysander, and E. Frisk. Quantitative stochastic fault diagnosability analysis. Submitted to CDC'2011, 2011.

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

(Frisk *et al.*, 2010) E. Frisk, A. Bregon, J. Åslund, M. Krysander, B. Pulido, and G. Biswas. Diagnosability analysis considering causal interpretations for differential constraints. In *Proc. 21st International Workshop on the Principles of Diagnosis (DX-10)*, 2010.

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

(Gustafsson, 2002) F. Gustafsson. Stochastic fault diagnosability in parity spaces. In *Proc. 15th Triennal IFAC World Congress on Automatic Control*, 2002.

(Kay, 1998) S. M. Kay. *Fundamentals of statistical signal processing : detection theory*. Prentice Hall, 1998.

(Kullback and Leibler, 1951) S. Kullback and R. A. Leibler. On information and sufficiency. 1951.

(Massoumnia *et al.*, 1989) M. A. Massoumnia, G. C. Verghese, and A. S. Willsky. Failure detection and identification. In *IEEE Transactions on Automatic Control*, 1989.

(Nikoukhah, 1989) R. Nikoukhah. Innovations generation in the presence of unknown inputs: Application to robust failure detection. 1989.

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

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