A General Framework for Fault Diagnosis Based on Statistical Hypothesis Testing

Mattias Nyberg
Department of Electrical Engineering, Linköping University,
SE-58183 Linköping, Sweden
Email: matny@isy.liu.se

Abstract

A framework for fault diagnosis, called *structured hypothesis tests*, is presented. It has earlier been developed within the area of automatic control, but is in fact very much inspired by the ideas developed in the AI area. The motivation was originally to handle dynamic systems with noise. However, it is here shown that also the noise-free case can be perfectly handled. The system to be diagnosed, and also the different faults, are described by differential equations, algebraic equations, and probability distribution functions. By using the framework, it is in the isolation possible to utilize all such modeled knowledge about the faults. The diagnosis system is constructed by combining a set of different hypothesis tests. In this way, the task of diagnosis is transferred to the task of validating a set of different models with respect to the measured data.

1 Introduction

An ongoing effort in the fault diagnosis community is to investigate relations between the model-based diagnosis methods used by researchers from the AI and automatic-control areas respectively, e.g. see (Cordier *et al.* 2000). In this context it can be interesting to study the framework of *structured hypothesis tests* (SHT) (Nyberg 1999b; 1999c; 1999a). This framework was developed from the perspective of automatic control but also uses inspiration from the AI area. It is primarily based on statistical hypothesis testing (Lehmann 1986) and decision theory (Berger 1985). The basic idea is to combine a set of different (binary) hypothesis tests, and in this way solve complicated diagnosis problems. Hypothesis testing, but from a slightly different perspective, have also been used in AI-based approaches to model based diagnosis, e.g. see (McIlraith & Reiter 1992; Struss 1994).

Originally, the SHT framework was developed for diagnosis of noisy systems. However, in this paper it is shown how it also can be used for diagnosis of noise-free systems. It is proved that with the SHT framework, we can in a noise-free environment design a diagnosis system that always produce a *complete* and *logically sound* diagnosis statement, i.e. the diagnosis system will always tell exactly which faults that can explain the observed behavior.

The motivation to develop the SHT framework comes from work with real applications, namely diagnosis of automotive engines (Nyberg 1999b; Nyberg & Stutte 2001). An example of requirements in such an application is that we have 1 sensor, 5 different faults modeled in different ways, a system that is dynamic, non-linear and contains noise and model uncertainties. Since the goal in these applications is to make decisions in a noisy environment it is natural to utilize the framework and theory of statistical hypothesis testing, which was developed exactly for this purpose.

In statistical hypothesis testing, and therefore also in the SHT framework, the reasoning is about *models*, which can be dynamic or static, and deterministic or stochastic. This reasoning about models has the advantage that all type of faults can be handled. Further on, by using models, we can into the diagnosis system, include detailed knowledge about how the faults affect the system. This can be a significant advantage since the more knowledge about the faults that can be considered, the larger are the possibilities to isolate between different faults.

The following section will introduce the basic concepts in the SHT framework. Then Section 3 will discuss fault modeling. Section 4 goes into some details about hypothesis testing. Section 5 presents the *incidence structure*, which is closely related to *structured residuals*, a common automatic-control approach to fault diagnosis. In Section 6, it will be shown that the incidence structure ideally is a representation of a set of relations between the faults. Based on these results, a sufficient condition for obtaining a complete and logically sound diagnosis statement is proven in Section 7. While the first part of the paper assumes a noise-free environment, Section 8 finally extends the discussion to systems with noise.

2 Basic Idea of Structured Hypothesis Tests

When using the SHT framework, different faults are classified into different *fault modes*. This is similar to *behavioral modes* as defined in (de Kleer 1989). Here we briefly introduce the concept of fault modes but more formal definitions will follow later in the paper. For an illustrative example, consider a system consisting of a gas tank with potential leakages. The tank is also equipped with a pressure sensor. We decide that all leakages, regardless of their area, belong to the same fault mode ”leakage”. We also decide that all faults in the pressure sensor belong to the fault mode

"pressure sensor fault". Further, one fault mode is always the "no-fault" case. Then the complete list of fault modes is

NF	"no fault"
PSF	"pressure sensor fault"
L	"leakage"
$PSF\&L$	"pressure sensor fault" and "leakage"

As seen each fault mode is associated with one abbreviation. We distinguish between single fault modes: PSF and L, and multiple fault modes: $PSF\&L$. The set of all fault modes is denoted Ω, and in the example, $\Omega = \{NF, PSF, L, PSF\&L\}$. A convention used, is that only one fault mode can be present at the same time. As we will see later, this originates from the theory of hypothesis testing.

2.1 The Diagnosis System

The diagnosis problem, and also the objective of the diagnosis system can be expressed as follows:

> Given a set of observations, the task of the diagnosis system is to generate a *diagnosis statement* S, which contains information about which fault modes that can explain the observations.

Note that it is assumed that the diagnosis system is *passive*, i.e. it can by no means affect the plant. We also assume that the diagnosis system is *static*, i.e. the same observations will always give the same diagnosis statement. In terms of *decision theory* (e.g. see (Berger 1985)), the diagnosis system is then a *decision rule* $\delta(x)$, i.e. a function from the observations to the diagnosis statement:

$$\delta : \mathcal{X} \longrightarrow \mathcal{P}(\Omega)$$

where $\mathcal{X}$ is the set of all possible observations and $\mathcal{P}(\Omega)$ is the *power set* of Ω. Here, x is used to denote the whole measured data-set, which usually consists of all known and measured variables of the system up to present time or a subset of this data. One choice is to use a fixed size time window.

Model based diagnosis is a complex task and it is therefore advantageous to divide the task in smaller subtasks. Thus the whole diagnosis system $\delta(x)$ is divided into smaller parts $\delta_k(x)$, which we will assume to be hypothesis tests. The classical, statistical or decision theoretic, definition of *hypothesis test* is adopted, e.g. see (Berger 1985; Lehmann 1986; Casella & Berger 1990), which is to be distinguished from "multiple hypothesis testing" that can also be found in literature, e.g. (Basseville & Nikiforov 1993). This means that we use hypothesis tests that are "binary" in the sense that the outcome of a hypothesis test is one, out of two possible decisions.

Each hypothesis test $\delta_k(x)$ generates a sub-diagnosis statement S_k, i.e. $S_k = \delta_k(x)$. The diagnosis statement S is then formed by combining the information of the sub-diagnosis statements. The procedure for this will be described later.

The diagnosis statements S and S_k do all contain information about which fault-modes that can explain the behavior of the system. In this paper, the representation and reasoning about this information are based on sets of fault modes, i.e. $S_k \subseteq \Omega$. Another possibility, discussed in (Nyberg 1999c), is to let the diagnosis statements be expressed by logic formulas.

A diagnosis statement S can in general contain more than one fault mode. For example $S = \{F1, F2\}$ means that both fault modes $F1$ and $F2$ can explain the behavior of the system.

Let F_p denote the present fault mode. Then for the k:th hypothesis test, the *null hypothesis* and the *alternative hypothesis* can with the help of a set M_k be written

$$H_k^0 : F_p \in M_k \quad \text{"some fault mode in } M_k \text{ can explain meas. data"} \tag{1a}$$

$$H_k^1 : F_p \in M_k^C \quad \text{"no fault mode in } M_k \text{ can explain meas. data"} \tag{1b}$$

where M_k^C denotes the complement of M_k. For the two possible decisions of a hypothesis test δ_k, we use the notation S_k^0 and S_k^1. This means that

$$S_k = \begin{cases} S_k^1 = M_k^C & \text{if } H_k^0 \text{ is rejected } (H_k^1 \text{ accepted}) \\ S_k^0 \subseteq \Omega & \text{if } H_k^0 \text{ is } \textit{not} \text{ rejected} \end{cases}$$

The convention used here and also commonly used in hypothesis testing literature, is that when H_k^0 is rejected, we *assume* that H_k^1 is true. This implies that the present fault mode can not belong to M_k, and therefore $S_k^1 = M_k^C$. What we can assume when H_k^0 is *not* rejected depends on the actual hypothesis tests, and will be discussed in Sections 5 and 8. However, it always holds that $M_k \subseteq S_k^0 \subseteq \Omega$.

How the hypothesis tests are used to diagnose and isolate faults is illustrated by the following example.

Example 1 *Assume that $\Omega = \{NF, F_1, F_2, F_3\}$ and that the diagnosis system contains the following set of three hypothesis tests:*

$$\begin{aligned} H_1^0 &: F_p \in M_1 = \{NF, F_1\} & S_1^0 &= \Omega \\ H_1^1 &: F_p \in M_1^C = \{F_2, F_3\} & S_1^1 &= \{F_2, F_3\} \end{aligned}$$

$$\begin{aligned} H_2^0 &: F_p \in M_2 = \{NF, F_2\} & S_2^0 &= \Omega \\ H_2^1 &: F_p \in M_2^C = \{F_1, F_3\} & S_2^1 &= \{F_1, F_3\} \end{aligned}$$

$$\begin{aligned} H_3^0 &: F_p \in M_3 = \{NF, F_3\} & S_3^0 &= \Omega \\ H_3^1 &: F_p \in M_3^C = \{F_1, F_2\} & S_3^1 &= \{F_1, F_2\} \end{aligned}$$

Then if only H_1^0 is rejected, we draw the conclusions that $F_p \in S_1^1$, $F_p \in S_2^0$, $F_p \in S_3^0$. That is, $F_p \in S_1^1 \cap S_2^0 \cap S_3^0 = \{F_2, F_3\} \cap \Omega \cap \Omega = \{F_2, F_3\}$, i.e. the present fault mode is either F_2 or F_3. If both H_1^0 and H_2^0 are rejected, we draw the conclusion that $F_p \in \{F_2, F_3\} \cap \{F_1, F_3\} \cap \Omega = \{F_3\}$, i.e. the present fault mode is F_3.

From the example above, it is clear that the diagnosis statement S can in general be expressed as $S = \bigcap_k S_k$.

3 Fault Modeling and Fault Modes

The plant to be diagnosed is modeled with a model $\mathcal{M}(\theta)$. The parameter vector θ is called the *fault state* and represents the true but unknown fault situation of the plant. The

model $\mathcal{M}(\theta)$ consists of differential and algebraic equations. In this paper we assume that no disturbances affect the plant and that there are no unknown parameters. However, the general case, including disturbances, unknown parameters, and also stochastic models, is described in (Nyberg 1999c). The effect of noise is included in this paper but will be handled later in Section 8.

One or possibly several fault states θ always corresponds to the fault-free case. The *fault-state space*, i.e. the parameter space of θ, will be denoted Θ. Note that we have chosen the convention that θ is not dependent on time which corresponds to an assumption that the fault state of the system never changes. Even though this may seem to be a limitation, this is not the case since we will be quite liberal regarding the definition of the parameter vector θ, e.g. elements are allowed to be functions.

The model $\mathcal{M}(\theta)$ can now formally be defined as

$$\mathcal{M} : \Theta \longrightarrow \mathcal{P}(\mathcal{X}) \tag{2}$$

That is, for a fixed value of θ, the model specifies the set of observations that are possible to observe.

3.1 Fault Modeling Principles

Many different principles for fault modeling have been used in the automatic-control literature. One of the most common is to model faults by unrestricted arbitrary fault signals. When fault signals are used, a specific fault is usually modeled as a scalar fault signal. For example consider an adder described by the equation

$$y(t) = u_1(t) + u_2(t) + f(t) \tag{3}$$

The fault free case can be represented by $f(t) \equiv 0$ and then any fault can be modeled by an $f(t) \not\equiv 0$. Obviously, fault modeling by signals is very general and can in principle describe all types of faults, but as has been noted in e.g. (Blanke 1999; Ding *et al.* 1999), this can cause problems with the isolation. In the formalism described above, a fault signal $f(t)$ can be an element in the θ-vector, i.e. $\theta_i = f(t)$. Note that θ_i is still constant but its value is the whole signal $f(t)$.

Another common fault modeling principle is to model faults by deviations in constant parameters. For an example of how this can be described with the parameter θ, see Example 2. One further, also common, fault model is to consider abrupt changes of variables, e.g. see (Basseville & Nikiforov 1993). More discussions on how the here mentioned fault modeling principles, and also other, can be formulated using the parameter θ is found in (Nyberg 1999c; 1999b).

Note that although we in SHT have the possibility to utilize fault models, there is no *must* that all faults are enumerated and precisely modeled. It is always possible to use a fault mode "unknown fault", either alone together with the NF fault mode or together with other fault modes representing more detailed fault models. For an example, consider the adder in (3). There can for example be three fault modes: NF (no fault), $S0$ (stuck at zero), and AF (arbitrary fault). The model for $S0$ is obviously $y(t) \equiv 0$, and the model for AF can be written $y(t) = f_a(t)$, where $f_a(t)$ is an unknown arbitrary signal. To represent this with the θ-vector, we can for example assume the following model:

$$y(t) = g_1\big(u_1(t) + u_2(t)\big) + g_2 f_a(t)$$

where $\theta = [g_1, g_2]$. Then the model for NF is obtained with $\theta = [1, 0]$, the model for $S0$ with $\theta = [0, 0]$, and the model for AF with $\theta = [0, 1]$. Note that in contrast to the fault model in (3), the fault in fault mode AF is not assumed to always affect the adder. The assumption that a fault always affect a system is called *fault exoneration* (Cordier *et al.* 2000). As was seen in the adder example, we can with the help of choosing fault models, chose to assume *fault exoneration* or not.

3.2 Fault Modes

The classification of different faults into fault modes corresponds to a *partition* of the fault-state space Θ. Each fault mode γ is associated with a subset Θ_γ of Θ. Thus all sets Θ_γ are pairwise disjoint and $\Theta = \cup_{\gamma\in\Omega}\Theta_\gamma$. If fault mode γ is present in the system, then we know that $\theta \in \Theta_\gamma$. The fact that all sets Θ_γ are pairwise disjoint reflects the fact that only one fault mode can be present at the same time. Note however that, even though two different fault modes always have disjoint Θ_γ-sets, they can result in identical observations. With the model (2), each fault mode γ can now be seen as a model of the process, namely the model $\mathcal{M}(\theta)$, where $\theta \in \Theta_\gamma$.

Example 2 *Consider a system described by the following equations:*

$$\begin{aligned}\dot{x}(t) =& f(x(t), u(t))\\ y_1(t) =& h_1(x(t)) + b_1\\ y_2(t) =& h_2(x(t)) + b_2\end{aligned}$$

The constants b_1 and b_2 represents sensor bias faults and it is assumed that only positive biases can occur. It is natural to let $\theta_1 = b_1$ and $\theta_2 = b_2$, and thus $\theta = [\theta_1\ \theta_2] = [b_1\ b_2]$. Four fault modes are considered: "no fault" $\mathbf{NF}$*, "bias in sensor 1"* $\mathbf{B1}$*, "bias in sensor 2"* $\mathbf{B2}$*, "bias in both sensor 1 and sensor 2"* $\mathbf{B1}\&\mathbf{B2}$*. The sets Θ, $\Theta_{\mathbf{NF}}$, $\Theta_{\mathbf{B1}}$, $\Theta_{\mathbf{B2}}$, and $\Theta_{\mathbf{B1}\&\mathbf{B2}}$ become*

$$\begin{aligned}\Theta =& \{[b_1\ b_2];\ b_1 \geq 0, b_2 \geq 0\}\\ \Theta_{\mathbf{NF}} =& \{[0\ 0]\}\\ \Theta_{\mathbf{B1}} =& \{[b_1\ 0];\ b_1 > 0\}\\ \Theta_{\mathbf{B2}} =& \{[0\ b_2];\ b_2 > 0\}\\ \Theta_{\mathbf{B1}\&\mathbf{B2}} =& \{[b_1\ b_2];\ b_1 > 0, b_2 > 0\}\end{aligned}$$

3.3 Component Fault Modes

So far, we have only considered fault modes and models on a "system level". Sometimes it is desirable to have a more component-oriented view of the system. Assume that the system is separated into a number of *components*. For each of these components a number of faults can occur. Each of these faults can be classified into different *component fault modes*. To avoid confusion, the fault modes on the system level can then be denoted *system fault-modes*.

Let F^i_j be the j:th component fault-mode of the i:th component. Further, let NF^i denote the no-fault fault-mode

of the i:th component. A system fault-mode can then be composed by a vector of component fault-modes. Some examples of system fault-modes are

$$\begin{aligned}\mathbf{NF} &= [NF^1, NF^2, \ldots NF^p] \\ \mathbf{F_1^1} &= [F_1^1, NF^2, \ldots NF^p] \\ \mathbf{F_1^2} &= [NF^1, F_1^2, NF^3, \ldots NF^p] \\ \mathbf{F_2^1 \& F_1^2} &= [F_2^1, F_1^2, NF^3, \ldots NF^p]\end{aligned}$$

Note the strong relationship with how failure/behavioral modes are treated in (de Kleer, Mackworth, & Reiter 1992). Here we have shortly discussed a representation based on component fault modes instead of system fault-modes, and also the relation between the two. Actually also the complete logical reasoning can be done using only the component fault modes. This topic will not be discussed here but is further investigated in (Nyberg 1999c).

4 Construction of the Hypothesis Tests

To develop the actual hypothesis tests, we first need to decide the set of hypotheses to test. We will here assume that the set of hypothesis tests is already specified with the help of sets M_k.

By using the sets Θ_γ, an alternative representation of the hypothesis test (1) can be written as

$$\begin{aligned}H_k^0 &: \theta \in \bigcup_{\gamma \in M_k} \Theta_\gamma \\ H_k^1 &: \theta \notin \bigcup_{\gamma \in M_k} \Theta_\gamma\end{aligned}$$

This is the representation commonly used in statistical hypothesis testing literature. For each hypothesis test δ_k, we then need to find a *rejection region*, i.e. a subset of $\mathcal{X}$ where the null-hypothesis is rejected. This is usually done via a *test quantity* (often also called test statistic). The test quantity is a function $T_k(\mathbf{x})$ from the *sample data* $\mathbf{x}$ (i.e. the observations), to a scalar value which is to be compared with a threshold J_k. Typically if $T_k(\mathbf{x}) \geq J_k$, then H_k^0 is rejected and otherwise not rejected. The rejection region of each test is thereby implicitly defined.

4.1 Construction of the Test Quantities

According to what has been said above, we need to design a test quantity $T_k(\mathbf{x})$ such that it is low or at least below the threshold if the data $\mathbf{x}$ matches the hypothesis H_k^0, i.e. a fault mode in M_k can explain the data. Also if the data come from a fault mode not in M_k, $T_k(\mathbf{x})$ should be large or at least above the threshold.

Design of test quantities, primarily from a statistical point of view, has been extensively discussed in general hypothesis testing literature, e.g. see (Lehmann 1986). Many of these ideas are applicable to fault diagnosis. In addition it can be useful to view the test quantity as a *model validity measure*. From the text above it should be realized that the test quantity is a model validity measure for the the model

$$\mathcal{M}(\theta), \quad \theta \in \bigcup_{\gamma \in M_k} \Theta_\gamma$$

i.e. the model defined by the null hypothesis. Below, we will exemplify such a model validity measure based on using the *prediction error*. Another example of a commonly used model validity measure is the likelihood function. For further discussions about different model validity measures useful for fault diagnosis, see (Nyberg 1999c).

4.2 Test Quantities based on Prediction Errors

Here we will assume that the observations can be divided into inputs u and outputs y. The calculation of the test quantity is then based on a comparison between the measured and predicted output y, over a time window of length N:

$$V_k(\theta, \mathbf{x}) = \frac{1}{N} \sum_{t=1}^{N} \big(y(t) - \hat{y}(t|\theta, \mathbf{x})\big)^2 \tag{4}$$

where $\hat{y}(t|\theta, \mathbf{x})$ is the prediction of the output $y(t)$, derived from an assumption of a specific θ and the measured data $\mathbf{x}$. The function $V_k(\theta, \mathbf{x})$, where θ is fixed, is then a measure of the validity of the model $\mathcal{M}(\theta)$, for a fixed θ, in respect to the measurement data $\mathbf{x}$.

The test quantity can then be calculated as

$$T_k(\mathbf{x}) = \min_{\theta \in \Theta_k^0} V_k(\theta, \mathbf{x}) \tag{5}$$

Note that although the model validity measure $V_k(\theta, \mathbf{x})$ in (5) is indexed by k, meaning that it is specific for the hypothesis test δ_k, it is often possible (and also quite elegant) to use the same $V(\theta, \mathbf{x})$ for all hypothesis tests. In that case, the only thing that differs test quantities in different tests, is the set Θ_k^0 over which the minimization is performed. This approach is demonstrated in (Nyberg 1999b).

5 Representing the Diagnosis System Using an Incidence Structure

A standard approach in the fault diagnosis literature within the automatic control community, e.g. (Gertler 1998; Chen & Patton 1999), is to use the principle of *structured residuals* to achieve fault isolation. In this section, we will see that structured hypothesis tests can actually be seen as a generalization and formalization of structured residuals. When using structured residuals, the *residual structure* (also called e.g. *fault-signature matrix*, *incidence matrix*) is an important concept. A consequence of formalizing the diagnosis procedure, as is done in structured hypothesis tests, is that the concept of residual structure must be modified. The solution here is to introduce a distinction between an *incidence structure*, describing how the faults ideally affect the test quantities, and a *decision structure*, describing how the diagnosis S is formed from the thresholded test quantities. This section primarily describes the incidence structure but later in Section 8, also the decision structure will be discussed. However, we will already here see that representing a diagnosis system with a decision structure, is equivalent to a representation using the sets M_k, S_k^0, and S_k^1.

To get an overview of how faults in different fault modes *ideally* affect the test quantities, it is useful to set up an *incidence structure*. With *ideally*, we mean that the system behaves exactly in accordance with the model and all

stochastic parts have been neglected, e.g. no unmodeled disturbances exists and there is no measurement noise. The incidence structure is derived by studying the equations describing the process model and how the test quantities $T_k(\mathbf{x})$ are calculated.

An incidence structure is a table or matrix containing 0:s, 1:s, and X:s. The X:s will be called *don't care*. An example of an incidence structure is

	NF	F_1	F_2	F_3
$T_1(\mathbf{x})$	0	0	1	0
$T_2(\mathbf{x})$	0	0	X	1
$T_3(\mathbf{x})$	0	X	0	X

(6)

A 0 in the k:th row and the j:th column means that if the fault mode present in the system, is equal to the fault mode of the j:th column, then the test quantity $T_k(\mathbf{x})$ will not be affected, i.e. it will be exactly zero. A 1 in the k:th row and the j:th column means that for *all* faults belonging to the fault mode of the j:th column, $T_k(\mathbf{x})$ will always be affected, i.e. it will be non-zero. An X in the k:th row and the j:th column means that for *some* faults belonging to the fault mode of the j:th column, $T_k(\mathbf{x})$ will under *some* operating conditions be affected, i.e. it will be non-zero. The dependence on operating condition typically arise in non-linear systems. Another reason for X:s is multiple fault modes, where the individual faults may compensate out each other. Compared to previous works involving residual structures (or fault-signature matrices etc.), the major difference is that we have here added the use of *don't care*.

Let s_{kj} denote the entry in the k:th row and the j:th column of an incidence structure. Then the interpretation of 0:s, 1:s, and X:s can be formally written as

$$F_p = F_j \;\rightarrow\; T_k(\mathbf{x}) = 0 \qquad \text{if } s_{kj} = 0 \tag{7a}$$
$$F_p = F_j \;\rightarrow\; T_k(\mathbf{x}) \neq 0 \qquad \text{if } s_{kj} = 1 \tag{7b}$$

where F_p, as before, denotes the present fault mode. Note that the interpretation of X is implicitly contained in these formulas, since if $s_{kj} = X$ then none of the two formulas are valid.

These "local" interpretations of 1:s, 0:s, and X:s, together with an incidence structure, is enough to define the isolation functionality of the whole diagnosis system. For example the interpretation of the incidence structure (6) becomes

$$\begin{aligned}
&T_1 = 0 \leftrightarrow F_p \in \{NF, F_1, F_3\}\\
&T_2 = 0 \leftarrow F_p \in \{NF, F_1\}\\
&T_2 \neq 0 \leftarrow F_p = F_3\\
&T_3 = 0 \leftarrow F_p \in \{NF, F_2\}\}
\end{aligned}$$

or equivalently

$$\begin{aligned}
&T_1 \neq 0 \leftrightarrow F_p = F_2\\
&T_2 \neq 0 \rightarrow F_p \in \{F_2, F_3\}\\
&T_2 = 0 \rightarrow F_p \in \{NF, F_1, F_2\}\\
&T_3 \neq 0 \rightarrow F_p \in \{F_1, F_3\}
\end{aligned}$$

This interpretation of the incidence structure (6) can now be used to derive the diagnosis statement S. For example if $T_1 = 0$, $T_2 \neq 0$, and $T_3 \neq 0$, we know by using the rules, that $F_p \in \{F_2, F_3\}$ and $F_p \in \{F_1, F_3\}$. This means that F_3 must be the present fault mode.

It can be realized that there is a one-to-one relationship between this procedure, i.e. forming S by using the incidence structure, and how S is formed via the individual sub-diagnoses statements S_k. For example, the sets S_k^0 and S_k^1 for the incidence structure (6), are

$$\begin{aligned}
S_1^0 &= \{NF, F_1, F_3\} & S_1^1 &= \{F_2\}\\
S_2^0 &= \{NF, F_1, F_2\} & S_2^1 &= \{F_2, F_3\}\\
S_3^0 &= \{NF, F_1, F_2, F_3\} & S_3^1 &= \{F_1, F_3\}
\end{aligned}$$

That is, the set S_k^0 contains all fault modes which have 0 or X in the k:th row of the incidence structure. Also S_k^1 contains all fault modes which have 1 or X in the same row. When assuming ideal conditions, the incidence structure can in this way be seen as an overview of a diagnosis system based on structured hypothesis tests.

6 Relations Between Fault Modes

It turns out that some fault modes are related to other fault modes such that in some cases they are impossible to separate. Consider for example a system modeled as

$$y = abu \tag{8}$$

where one fault mode $\mathbf{F_a}$ corresponds to that $a \neq 1$ and fault mode $\mathbf{F_b}$ corresponds to that $b \neq 1$. It is obvious that both $\mathbf{F_a}$ and $\mathbf{F_b}$ can equally well describe the system, and that it is impossible to isolate between these two fault modes.

For both analysis and design of diagnosis systems, this kind of relations play a fundamental role. They tell us for example when isolation is possible, and also controls how it is possible to chose the null-hypothesis, i.e. the set M_k.

To investigate this relation between fault modes, let us first formally define fault modes:

Definition 1 (Fault Mode) *A fault mode F_i is a function $F_i : \Theta_{F_i} \longrightarrow \mathcal{P}(\mathcal{X})$.*

Further we need the notion of *observation set*:

Definition 2 (Observation Set) *The* observation set *of a fault mode F_i is denoted $\mathcal{O}_{F_i}$ and defined by*

$$\mathcal{O}_{F_i} = \bigcup_{\theta \in \Theta_{F_i}} F_i(\theta)$$

Then the relations of interest are:

$$\mathcal{O}_{F1} = \mathcal{O}_{F2} \tag{9a}$$
$$\mathcal{O}_{F1} \subseteq \mathcal{O}_{F2} \tag{9b}$$
$$\mathcal{O}_{F1} \cap \mathcal{O}_{F2} = \emptyset \tag{9c}$$

For example, $\mathcal{O}_{F1} \subseteq \mathcal{O}_{F2}$ means that any fault belonging to fault mode $F1$ can also be "explained" by a fault in $F2$. Further, $\mathcal{O}_{F1} \cap \mathcal{O}_{F2} = \emptyset$ means that no fault in $F1$ can be "explained" by a fault in $F2$, and vice verse.

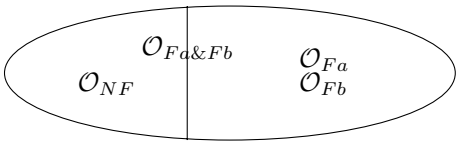
Example 3 *Consider the system (8) and the fault modes*

$$\begin{aligned}
NF:&\quad a = 1, b = 1\\
Fa:&\quad a \neq 1, b = 1\\
Fb:&\quad a = 1, b \neq 1\\
Fa\&Fb:&\quad a \neq 1, b \neq 1
\end{aligned}$$

Of the relations (9), the only *relations that hold in this example are the following:*

$$\mathcal{O}_{Fa} = \mathcal{O}_{Fb} \quad (10a)$$
$$\mathcal{O}_{NF} \subseteq \mathcal{O}_{Fa\&Fb} \quad (10b)$$
$$\mathcal{O}_{Fa} \subseteq \mathcal{O}_{Fa\&Fb} \quad (10c)$$
$$\mathcal{O}_{Fb} \subseteq \mathcal{O}_{Fa\&Fb} \quad (10d)$$
$$\mathcal{O}_{Fa} \cap \mathcal{O}_{NF} = \emptyset \quad (10e)$$
$$\mathcal{O}_{Fb} \cap \mathcal{O}_{NF} = \emptyset \quad (10f)$$

In a Venn-diagram, this can be illustrated as



6.1 Design of the Influence Structure

The influence structure for each hypothesis test and thereby also the sets M_k are more or less determined from the relations between the fault modes. To study this, we assume that we have *ideal test quantities*. An ideal test quantity is zero if the measured data can be explained by the null hypothesis and non-zero otherwise.

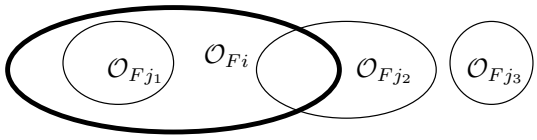
Assume we want to design a hypothesis test with the *desired* null-hypothesis $H_k^0 : F_p \in M_k = \{Fi_1, Fi_2, \ldots, Fi_n\}$. This may be possible, but depending on the relations between the fault modes, it is sometimes necessary to add some fault modes to the set M_k.

The following theorem tells us the relation between the fault mode relations and the incidence structure. It is here assumed that the desired null hypotheses only contain one fault modes. However the extension to more complex null hypotheses is trivial.

Theorem 1 *Given a hypothesis test with a desired null-hypothesis $H_k^0 : F_p = Fi$, and an ideal test quantity, the actual set M_k and the influence structure are uniquely determined by the knowledge of the relations $\mathcal{O}_{F1} \subseteq \mathcal{O}_{F2}$ and $\mathcal{O}_{F1} \cap \mathcal{O}_{F2} = \emptyset$ as follows:*
The entry in the column corresponding to fault mode Fj is

0 if $\mathcal{O}_{Fj} \subseteq \mathcal{O}_{Fi}$
1 if $\mathcal{O}_{Fi} \cap \mathcal{O}_{Fj} = \emptyset$
X if $\mathcal{O}_{Fi} \cap \mathcal{O}_{Fj} \neq \emptyset$ and $\mathcal{O}_{Fj} \not\subseteq \mathcal{O}_{Fi}$

PROOF. Let $M_k' = \{Fi\}$ represent the desired null hypothesis. Assume that Fj is the present fault mode. There are three principle ways the fault mode Fj can be related to Fi, as is illustrated below:



The ideal test quantity for M_k' will be zero for all observations inside $\mathcal{O}_{Fi}$ and non-zero for all observations outside $\mathcal{O}_{Fi}$. This means that the test quantity will be zero for all observations originating from fault mode Fj_1. Thus, the influence structure should contain a 0 in the position for Fj_1.

For observations originating from Fj_3, the test quantity will be always non-zero. That is, the influence structure should contain a 1 in the position for Fj_3. Finally for observations originating from Fj_2, the test quantity will be sometimes zero and sometimes non-zero. Therefore, the influence structure should contain an X in the position for Fj_2.

□

Example 4 *Consider again the system in Example 3. Assume that we can use ideal test quantities and want to construct four hypothesis tests with the desired sets $M_1' = \{NF\}$, $M_2' = \{Fa\}$, $M_3' = \{Fb\}$, and $M_4' = \{Fa\&Fb\}$. Using the relations (10) and Theorem 1, the influence structure becomes*

	NF	Fa	Fb	$Fa\&Fb$
T_1	0	1	1	X
T_2	1	0	0	X
T_3	1	0	0	X
T_4	0	0	0	0

By using the relations between the incidence structure and the sets S_k^1 and S_k^0, we realize that also the sets M_k, related as $M_k = S_k^{1^C}$, are determined from Theorem 1. For example, in the example, we have $M_2 = \{Fa, Fb\}$.

7 Completeness and Soundness of Structured Hypothesis Tests

It is desirable that a diagnosis system produces diagnosis statements that are *complete* and *logically sound*. That is, all fault modes that can explain the observations are contained in S (completeness), and all fault modes in S can explain the observations (logical soundness). The following theorem contains a sufficient condition for producing such diagnosis statements when using structured hypothesis tests.

Theorem 2 *Let a diagnosis system be constructed with one hypothesis test for each fault mode Fi, i.e. the desired null hypotheses are $H_k^0 : F_p = Fi$. Assume that ideal test quantities are used and let the incidence/decision structure be chosen according to Theorem 1. Then the diagnosis statement S will always be complete and logically sound.*

PROOF. We need to prove that $Fi \in S$ if and only if some fault state in Fi can explain the measured data. Completeness, i.e. the if-part of the proof, follows from the fact that ideal test quantities are used and that the incidence/decision structure is constructed with Theorem 1.

For the only-if part of the proof, assume that no fault state belonging to Fi can explain the measured data. Consider now the hypothesis test with desired null hypothesis $H_k^0 : F_p = Fi$. Because of Theorem 1, the actual M_k will look like $M_k = \{Fi, F_1, F_2 \ldots\}$, where for all $F_j \in M_k$ it holds that $\mathcal{O}_{F_j} \subseteq \mathcal{O}_{Fi}$. This means that no fault states belonging to any of the fault modes in M_k can explain the data. This further means that $T_k \neq 0$ and $S_k = S_k^1 = M_k^C$, and therefore $F_i \notin S_k$. This also implies that $F_i \notin S$ which concludes the proof. □

Remark A diagnosis system based on the framework of structured residuals uses a residual structure (i.e. an incidence/decision structure) with only 0:s and 1:s, and no X:s. This has the effect that in the general case, the diagnosis statement will not be complete.

8 Diagnosis in a Noisy Environment

So far, only noise-free systems have been considered. However, most real systems are in fact affected by noise and model uncertainties. Since hypothesis testing theory is primarily developed for making decisions in a noisy and uncertain environment, it is quite easy to extend the discussion to the noisy case. Basically, instead of checking if the test quantities are equal to zero, as was done in Section 6 to 7, we have to use thresholds.

Consider the following system with noise:

$$y = \theta + n$$

Here, y is measured and n is a stochastic term with some probability distribution function. There are two fault modes: $\theta = 0$ (NF) and $\theta > 0$ (F), i.e. $\Omega = \{NF, F\}$. For diagnosing this system, we use a hypothesis test with the following hypotheses:

$$H^0 : F_p = NF \qquad (\theta = 0)$$
$$H^1 : F_p = F \qquad (\theta > 0)$$

Assume that the test quantity is chosen as $T = y$. In a noise-free environment (i.e. $n \equiv 0$), we could easily draw the conclusion that $F_p = NF$ if $T = 0$ and $F_p = F$ if $T \neq 0$.

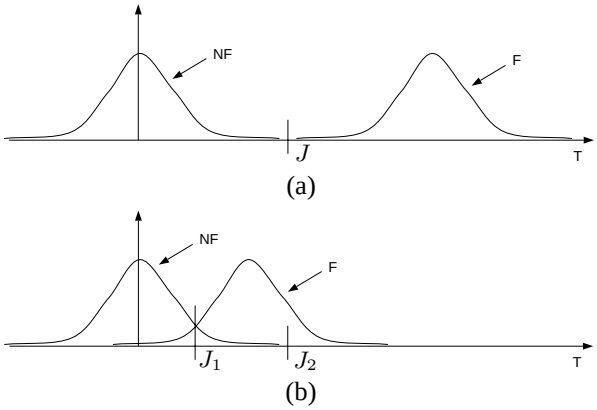


Figure 1: Probability distribution functions for a large fault and a small fault.

Now consider the noisy case but assume that when the fault mode F is present, θ is always large. Then the probability distribution functions for NF and F would look as is illustrated in Figure 1(a). Then by placing a threshold J in the middle between the two distributions, the following conclusions can be drawn:

$$T < J \Rightarrow S = \{NF\}$$
$$T > J \Rightarrow S = \{F\}$$

The tails of the distributions (on the wrong side of J) will cause errors in the decisions. However the probability of making errors will in the example be very small.

Assume now that faults with also small θ must be handled. The probability distribution functions for NF and F would look as is illustrated in Figure 1(b). If we use a threshold J_1 in the middle of the distributions, we would make errors with a high probability. The only possibility to avoid errors is to only consider the distribution of NF and therefore use the threshold J_2. The following conclusions can then be made:

$$T < J_2 \Rightarrow S = \{NF, F\} \tag{11a}$$
$$T > J_2 \Rightarrow S = \{F\} \tag{11b}$$

That is, when the $T < J_2$ we do not draw any conclusion since $S = \Omega$.

8.1 Hypothesis Tests in a Noisy Environment

The solution of using an "asymmetric" test, such as (11), is a standard solution in hypothesis testing. Therefore a hypothesis test in a noisy environment normally becomes:

$$\text{not reject } H_k^0 \;\Rightarrow\; S_k = S_k^0 = \Omega$$
$$\text{reject } H_k^0 \text{ (=accept } H_1) \Rightarrow S_k = S_k^1 = M_k^C$$

This means that a fault mode is either an element in only S_k^0 or an element in both S_k^0 and S_k^1. In terms of the incidence structure, this means that all 1:s must be replaced with X:s.

Remember that the incidence structure corresponds to the case where ideal conditions holds. In a more realistic case, the model is not perfect; unmodeled disturbances affects the process, and there is measurement noise. All this means that the formulas (7) are not valid and the incidence structure can therefore not be used to form the diagnosis S. That is, the structure used for deriving the diagnosis decision should not be the incidence structure, but instead the *decision structure*. The decision structure is in most cases the incidence structure but with all 1:s replaced with X:s.

In the noise-free case, the decisions S_k made by the hypothesis tests were always true. In the noisy case, even though we use a good threshold and S_k^0 chosen as $S_k^0 = \Omega$, the decisions S_k can not be guaranteed to be true. To be able to make the assumption that H_k^1 is true when H_k^0 is rejected, we need to design the hypothesis test such that the so called *significance level* $\alpha_k = P(\text{reject } H_k^0 \mid H_k^0 \text{ true})$ is small.

Sometimes, it is also in a noisy environment reasonable to make the assumption that H_k^0 is true when it is not rejected. This is controlled by the *power function* $\beta_k(\theta) = P(\text{reject } H_k^0 \mid \theta)$. For example, if it actually holds that $P(\text{reject } H_k^0 \mid \theta)$ is large for all $\theta \in \Theta_{F_i}$, then we do not take any large risk if we assume that F_i is not present when H_k^0 is not rejected. If this is the case, F_i should be excluded from S_k^0. In other words, given a hypothesis test, it is the power function that determines the choices of the sets S_k^0 and S_k^1 (i.e. the choices of 0, 1, and X in the decision structure). The relation between the power function and the decisions S_k^0 and S_k^1 is further investigated in (Nyberg 1999c).

9 Conclusions

In this paper, we have seen how statistical hypothesis testing and decision theory can be used to form a general framework for fault diagnosis. One advantage of using these existing theories is that all already developed theory for design and also evaluation of hypothesis tests and general decision functions can quite easily be applied to the diagnosis problem. This advantage can clearly be seen in (Nyberg 1999a; 2000a) which use methods from hypothesis testing and decision theory for evaluations and comparisons of diagnosis systems.

Two consequences of using hypothesis testing are that X:s must be used in the incidence/decision structure, and that the reasoning to produce the diagnosis statement is about *models*. The X:s (*don't care*) are necessary to get a *complete* diagnosis statement. Interesting is that the established framework *structured residuals* in the area of automatic control, does not use X:s and can therefore not produce complete diagnosis statements. The reasons to include X:s are nonlinearities, noise, and multiple fault compensation. Also in (Cordier *et al.* 2000), it is argued that X:s are needed. The reason there is to handle multiple fault compensation and to relax the *fault exoneration assumption*, i.e. the assumption that a fault always affects the system. In the SHT framework, relaxing of the fault exoneration assumption is not directly a reason to use X:s. Instead, we choose if we want to make the fault exoneration assumption or not, by using different fault modeling approaches.

To work with fault models is a powerful tool to handle in principle all types of faults. Thus, in the SHT framework we can for example diagnose faults that are modeled as deviations in constant parameters, arbitrary signals, abrupt changes, a change in signal variance, and also a mix between different types. Another reason to work with fault models is the increased possibility to isolate different faults. For example, by knowing that two different faults are acting in a different way, we can distinguish between the two even though they are acting on the same component.

In this paper, the SHT framework has only been exemplified on small toy examples. A more complete, and AI oriented example, can be found in (Nyberg 2000b) which investigates the well known polybox example from (de Kleer & Williams 1987). However the theory has also been successfully applied to real applications: diagnosis of the air intake system of different kinds of automotive engines (Nyberg 1999b; 1999a; 2000a; Nyberg & Stutte 2001). These works have shown that the theory has practical relevance for both design and analysis of diagnosis systems.

10 References

Basseville, M., and Nikiforov, I. 1993. *Detection of Abrupt Changes*. PTR Prentice-Hall, Inc.

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer.

Blanke, M. 1999. *Fault Tolerant Control Systems*, In Frank (1999). chapter 6.

Casella, G., and Berger, R. 1990. *Statistical Inference*. Duxbury Press.

Chen, J., and Patton, R. J. 1999. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers.

Cordier, M.; Dague, P.; Dumas, M.; Levy, F.; Montmain, J.; Staroswiecki, M.; and Trave-Massuyes, L. 2000. AI and automatic control approaches of model-based diagnosis: Links and underlying hypothesis. SAFEPROCESS, 274–279. Budapest, Hungary: IFAC.

de Kleer, J., and Williams, B. 1987. Diagnosing multiple faults. *Artificial Intelligence* 32(1):97–130.

de Kleer, J.; Mackworth, A.; and Reiter, R. 1992. Characterizing diagnoses and systems. *Artificial Intelligence* 56(2-3):197–222.

de Kleer, J. 1989. Diagnosis with behavioral modes. Proc. IJCAI-89, 104–109.

Ding, S.; Jeinsch, T.; Ding, E.; Zhou, D.; and Wang, G. 1999. Detection of observer based FDI schemes to the three tank system. Proc. of the ECC'99.

Frank, P., ed. 1999. *Advances of Control: Highlights of ECC'99*. Springer.

Gertler, J. 1998. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker.

Hamscher, W.; Console, L.; and de Kleer, J., eds. 1992. *Readings in Model Based Diagnosis*. Morgan Kaufmann Publishers.

Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Springer Verlag, second edition.

McIlraith, S., and Reiter, R. 1992. *On Tests for Hypothetical Reasoning*, In Hamscher et al. (1992). 89–96.

Nyberg, M., and Stutte, T. 2001. Model based diagnosis of the air path of an automotive diesel engine. IFAC Automotive Workshop. Karlsruhe, Germany: IFAC.

Nyberg, M. 1999a. Automatic design of diagnosis systems with application to an automotive engine. *Control Engineering Practice* 7(8):993–1005.

Nyberg, M. 1999b. Model based diagnosis of both sensor-faults and leakage in the air-intake system of an SI-engine. *SAE Paper 1999-01-0860*.

Nyberg, M. 1999c. *Model Based Fault Diagnosis: Methods, Theory, and Automotive Engine Applications*. Ph.D. Dissertation, Linköping University. URL: http://www.fs.isy.liu.se/Publications/.

Nyberg, M. 2000a. Evaluation of test quantities for leakage diagnosis in the air path of an automotive engine. SAFEPROCESS, 143–148. Budapest, Hungary: IFAC.

Nyberg, M. 2000b. The polybox example using the framework of structured hypothesis tests. Technical Report LiTH-ISY-R-2335, Linköping, Sweden.

Struss, P. 1994. Testing physical systems. 12:th National Conference on Artificial Intelligence. Seattle, Washington: AAAI.