MODEL BASED FAULT DIAGNOSIS USING STRUCTURED HYPOTHESIS TESTS

Mattias Nyberg

Department of Electrical Engineering, Linköping University,
Linköping, Sweden
Email: matny@isy.liu.se

Abstract: A new framework for fault diagnosis, called *structured hypothesis tests*, is proposed. The basic idea is to construct the diagnosis system by combining a set of 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. Arbitrary types of faults, including multiple faults, can be handled. That means that one single diagnosis system can diagnose faults of many different types. When using structured hypothesis tests, existing diagnosis methods such as residual generation, parameter estimation, and statistical methods, become parts of one common framework.

Keywords: Fault Diagnosis, Isolation, Hypothesis Testing, Decoupling

1. INTRODUCTION

In this paper, a new framework for fault diagnosis is proposed: *structured hypothesis tests*. The basic idea is to construct the diagnosis system by combining a set of hypothesis tests. The isolation strategy becomes simple and a great advantage is that the concept of using hypothesis tests allows the user naturally to model faults in arbitrary ways. This means that throughout this paper, there will be no restriction to any special types of faults and also, no restriction will be made regarding the multiplicity of faults. An additional advantage is that structured hypothesis tests has a strong mathematical foundation since it is theoretically grounded in hypothesis testing (seen from either a statistical or decision theoretic standpoint), and also in mathematical logic. Further, previous methods for fault diagnosis, e.g. structured residuals, statistical methods, parameter estimation, and different residual generation methods, becomes naturally parts or special cases of this common framework.

The motivation for this work comes from an industrial application, diagnosis of an automotive engine (Nyberg, 1999*b*). This application inspired much of the development of the framework of structured hypothesis tests, since previous frameworks could not provide an elegant solution to the requirement of diagnosing several different kinds of faults in the same diagnosis system.

This paper deals primarily with *design* of diagnosis systems. Analysis, or evaluation, of diagnosis system is also an important topic. Methods for this, using the same general framework as in this paper, are given in (Nyberg, 1999*a*).

The first part of the paper, i.e. Sections 2 and 3, presents a general and formalized description of the diagnosis problem. The basics of structured hypothesis tests is given in Sections 4 to 5. Some general principles for design of the hypothesis tests are then discussed in Section 6.

2. FAULT MODELING

The system to be diagnosed is modeled as a plant $G(\theta_G, \phi_G)$ and the vector valued signal $z(t, \theta_z, \phi_z)$. The parameters θ_G and θ_z describe faults and the parameters ϕ_G and ϕ_z describe other unknown variables, e.g. disturbances or unknown constant parameters. The plant $G(\theta_G, \phi_G)$ is modeled by differential equations. It has known inputs $u(t)$, e.g. control signals, and measurable outputs $y(t)$. In addition, the plant can be affected by other signals, which are collected in $z(t, \theta_z, \phi_z)$. These additional signals are assumed to be unknown or at least partially unknown. Some of the signals $z(t, \theta_z, \phi_z)$ may be modeled as stochastic processes.

2.1 *Fault State*

The parameter vector $\theta = [\theta_G \ \theta_z]$ is called the *fault state* and represents the true but unknown fault situation of the plant $G(\theta_G, \phi_G)$ and the signal $z(t, \theta_z, \phi_z)$. One or possibly several fault states 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 as we will see later. We will be quite liberal regarding the definition of the parameter vector θ, e.g. we will allow elements that are functions.

Corresponding to θ there is the constant parameter vector $\phi = [\phi_G \ \phi_z]$, which represents other unknown variables affecting the system. This paper will mostly *not* be focused on handling of unknown variables. Therefore, the parameter ϕ will often be neglected. However, it will be realized that the techniques to deal with unknown variables ϕ are the same as the ones needed for fault isolation (i.e. to decouple faults).

The whole system model (with ϕ_G and ϕ_z neglected) will be denoted $\mathcal{M}(\theta)$ and thus

$$\mathcal{M}(\theta) = \langle G(\theta_G), z(t, \theta_z) \rangle$$

The model $\mathcal{M}(\theta)$ with a fixed value of θ then exactly specifies the system when a specific fault state is present.

2.2 *Fault Modeling Principles*

Many different principles for fault modeling have been used in the 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. Fault modeling by signals is very general and can 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 $f(t)$. If discrete time and finite data is considered, then θ_i becomes a vector $\theta_i = [f(t_1) \ldots f(t_n)]$.

Another common fault modeling principle is to model faults by deviations in constant parameters. One further, also common, fault model is to consider abrupt changes of variables, e.g. see (Basseville and Nikiforov, 1993). For an example of how this can be described with the parameter θ, see Example 3. More discussions on how the here mentioned fault modeling principles, and also other, can be formulated using the parameter θ is found in (Nyberg, 1999*c*; Nyberg, 1999*b*).

No matter which fault modeling principle that is used, it is generally advantageous to include restrictions on θ into the fault models. For example, the size of a bias or a gain-error is usually limited by the system, and this restriction is easy to model by using inequalities including θ. Another example is that the bandwidth of a fault signal $f(t)$ is limited to some value. The reason to include restrictions is that the isolation task becomes easier the more restrictive fault models we have.

3. FAULT MODES

Different faults will be classified into different *fault modes*. For example, consider a system containing a water tank and leakages in the bottom of this tank. All such leakages, regardless of their area, belong to the same fault mode "water-tank bottom-leakage".

The classification of different faults into fault modes corresponds to a *partition* of the fault-state space Θ. This means that each fault mode γ is associated with a subset Θ_γ of Θ. One of the fault modes corresponds to the fault-free case and this fault mode will be denoted "no fault" or **NF**. Further, all sets Θ_γ are pairwise disjoint and $\Theta = \cup_{\gamma \in \Omega} \Theta_\gamma$ where Ω is used to denote the set of all fault modes.

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 means that only one fault mode can be present at the same time. For notational convenience we will to each fault mode associate an abbreviation, e.g. "no fault" was abbreviated **NF**.

Example 1. Consider a system described by the following equations:

$$\dot{x} = f(x, u) \tag{1a}$$
$$y_1 = h_1(x) + b_1 \tag{1b}$$
$$y_2 = h_2(x) + b_2 \tag{1c}$$

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" **NF**, "bias in sensor 1" **B1**, "bias in sensor 2" **B2**, "bias both sensor 1 and sensor 2" **B1&B2**. The sets Θ, $\Theta_{\mathbf{NF}}$, $\Theta_{\mathbf{B1}}$, $\Theta_{\mathbf{B2}}$, and $\Theta_{\mathbf{B1\&B2}}$ become

$$\Theta = \{[b_1\ b_2];\ b_1 \geq 0, b_2 \geq 0\} \tag{2a}$$
$$\Theta_{\mathbf{NF}} = \{[0\ 0]\} \tag{2b}$$
$$\Theta_{\mathbf{B1}} = \{[b_1\ 0];\ b_1 > 0\} \tag{2c}$$
$$\Theta_{\mathbf{B2}} = \{[0\ b_2];\ b_2 > 0\} \tag{2d}$$
$$\Theta_{\mathbf{B1\&B2}} = \{[b_1\ b_2];\ b_1 > 0, b_2 > 0\} \tag{2e}$$

Note that with the formalism described here, multiple fault-modes, e.g. **B1&B2**, comes in naturally and requires no special treatment.

4. STRUCTURED HYPOTHESIS TESTS

In this section, we will see how classical hypothesis testing can be utilized for model based diagnosis and especially fault isolation. The literature is quite sparse on this subject but some related contributions can be found in (Riggins and Rizzoni, 1990; Grainger *et al.*, 1995; Bøgh, 1995; Basseville, 1997). It can be noted that the idea of structured hypothesis tests has a close connection to the general hypothesis testing method *intersection-union test* (Casella and Berger, 1990).

Inputs to the diagnosis system are the signals $u(t)$ and $y(t)$. The task of the diagnosis system is to generate a *diagnosis S*, which contains information about which fault modes that can explain the behavior of the process. Note that it is assumed that the diagnosis system is *passive*, i.e. it can by no means affect the plant. That is, the diagnosis system is a function from u and y to S and we write $S = \delta(x) = \delta([u\ y])$. Here, x is used to denote the whole measured data-set, which can also contain several samples of u and y from different times.

Model based diagnosis is a complex task and it is therefore advantageous to divide the task in smaller tasks. Therefore the whole diagnosis system $\delta(x)$ is divided into smaller parts $\delta_k(x)$, which we will assume to be hypothesis tests. Each hypothesis test $\delta_k(x)$ generates a sub-diagnosis S_k, i.e. $S_k = \delta_k(x)$. The purpose of the *decision logic* is then to combine the information of the sub-diagnoses to form the diagnosis S. The procedure for this will be described in Section 4.1 below.

The diagnosis S and the individual sub-diagnoses 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. Another possibility, discussed in (Nyberg, 1999*c*), is to let the diagnoses be expressed by logic formulas. Actually, there is a quite close connection to diagnosis methods based on logic (Reiter, 1987).

A diagnosis S can in general contain more than one fault mode. This is in contrast to most fault diagnosis literature, in which the diagnosis can only contain *one* fault mode. For example $S = \{F1, F2\}$ means

that both fault modes $F1$ and $F2$ can explain the behavior of the system. This feature corresponds well to a desired functionality since in cases where it is difficult or even impossible to decide which fault mode that is present, it is very useful for a service technician to get to know that there are more than one fault mode that can explain the behavior of the process. If the diagnosis system was forced to pick out one fault mode in cases like this, it is highly probable that a mistake is made and wrong fault mode is picked out.

4.1 Basic Principle

The classical, statistical or decision theoretic, definition of *hypothesis test* is adopted, e.g. see (Berger, 1985; Lehmann, 1986; Casella and Berger, 1990), which is to be distinguished from "multiple hypothesis testing" that is often found in literature, e.g. (Basseville and Nikiforov, 1993). Note that when using hypothesis testing, we can have a probabilistic (statistical) or a deterministic view. Therefore, the method structured hypothesis tests is valid either we have probabilistic knowledge, in terms of probability density functions of e.g. the signal z (described in Section 2.1), or not.

Let F_p denote the present fault mode. Then for the k:th hypothesis test, the null hypothesis and the alternative hypothesis can be written

$H_k^0 : F_p \in M_k$ "some fault mode in M_k can explain meas. data"

$H_k^1 : F_p \in M_k^C$ "no fault mode in M_k can explain meas. data"

Note that an alternative representation of the hypothesis tests is $H_k^0 : \theta \in \Theta_k^0$ and $H_k^1 : \theta \notin \Theta_k^0$, where $\Theta_k^0 = \cup_{\gamma \in M_k} \Theta_\gamma$.

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, i.e. it must belong to M_k^C. Further, when H_k^0 is *not* rejected, we will for the present not assume anything. This latter fact will be slightly modified in Section 4.2, where we discuss how we also can assume something when H_k^0 is *not* rejected. How the hypothesis tests are used to diagnose and isolate faults is illustrated by the following example.

Example 2. Assume that the diagnosis system contains the following set of three hypothesis tests:

$H_1^0 : F_p \in M_1 = \{NF, F_1\}$ $H_1^1 : F_p \in M_1^C = \{F_2, F_3\}$
$H_2^0 : F_p \in M_2 = \{NF, F_2\}$ $H_2^1 : F_p \in M_2^C = \{F_1, F_3\}$
$H_3^0 : F_p \in M_3 = \{NF, F_3\}$ $H_3^1 : F_p \in M_3^C = \{F_1, F_2\}$

Then if only H_1^0 is rejected, we can draw the conclusion that $F_p \in M_1^C = \{F_2, F_3\}$, 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 can draw the conclusion that $F_p \in M_1^C \cap M_2^C = \{F_2, F_3\} \cap \{F_1, F_3\} = \{F_3\}$, i.e. the present fault mode is F_3.

From the example above, it is clear that the decision logic is a simple intersection operation. In the example, there is one hypothesis test for each of the possible fault modes (except for NF). This should be the natural choice in most situations.

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

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

where Ω denotes the set of all fault modes. The diagnosis S can now be expressed as $S = \bigcap_k S_k$. We will in Section 4.2 below, relax the definition of S_k^0 such that it may be a subset of Ω, i.e. $S_k^0 \subseteq \Omega$.

4.2 Hypothesis Tests

For each hypothesis test δ_k, we need to find a *test quantity* and a *rejection region*. The test quantity is a function $T_k(\mathbf{x})$ from the sample data $\mathbf{x}$, to a scalar value which is to be thresholded by a threshold J_k. 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.

The test quantity $T_k(\mathbf{x})$ is in many texts instead called a *test statistic*. However, the name *test statistic* indicates that $T_k(\mathbf{x})$ is a random variable which in general may not be a desired view. The test quantity $T_k(\mathbf{x})$ may for example be a *residual generator* or a sum of squared prediction errors of a parameter estimator. In many applications, a deterministic view is taken and $T_k(\mathbf{x})$ is seen just as a function of the data and not as a random variable.

The sample data $\mathbf{x}$ can be all measured data up to present time or a subset of this data. One choice is to use a fixed size time window.

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. Using traditional fault-diagnosis terminology, the fault modes in M_k are said to be *decoupled*. 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.

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 significance level is small. This implies that the threshold J_k must be set relatively high. This in turn means that the value of the *power function* $P(\text{reject } H_k^0 \mid \theta)$ does not necessarily become large for all values $\theta \notin \Theta_k^0$. For instance, if the present fault mode is F_i and $F_i \in M_k^C$, then for some $\theta \in \Theta_{F_i}$, the probability to reject H_k^0 may be very small. This is the reason why we up to now, have assumed that $S_k^0 = \Omega$, i.e. we can not assume anything when H_k^0 is not rejected.

Now 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 has not occurred when H_k^0 is not rejected. If this is the case, F_i should be excluded from S_k^0. The relation between the power function and the decisions S_k^0 and S_k^1 is further investigated in (Nyberg, 1999c).

It turns out that some fault modes are related to other system fault modes such that for some values of θ they are impossible to separate. This relation, called the *submode relation*, is formally defined in (Nyberg, 1999c). Submode relations between fault modes have implications on how the sets M_γ (i.e. the null hypotheses) can be chosen. For example for most fault modes, the limit when the fault size goes to zero is equal to the fault mode "no fault". This means that

when fault mode **NF**, i.e. no fault, is present, most null hypothesis can not be rejected. The implication is that almost all sets M_γ must include **NF**. In other words, the choice of null hypotheses is not a completely free choice. For more information on how the submode relation restricts the choice of null hypotheses, see (Nyberg, 1999c).

5. INCIDENCE AND DECISION STRUCTURE

Common in the fault diagnosis literature, e.g. (Gertler, 1998), 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 or formalization of structured residuals. When using structured residuals, the *residual structure* is usually 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 affect the test quantities, and a *decision structure*, describing how the diagnosis S is formed from the thresholded test quantities. We will 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.

5.1 Incidence Structure

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

$$
\begin{array}{c|cccc}
 & NF & F_1 & F_2 & F_3 \\
\hline
T_1(\mathbf{x}) & 0 & 0 & 1 & 0 \\
T_2(\mathbf{x}) & 0 & 0 & 1 & 1 \\
T_3(\mathbf{x}) & 0 & X & 0 & 1
\end{array}
\tag{3}
$$

A 0 in the k:th row and the j:th column means that if the system 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.

Although a distinction has not been made between incidence structures and decision structures in previous literature, the basic idea of using incidence structures (or residual structures) is not new. However, compared to previous works involving incidence structures, a 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 or semantics of 0:s, 1:s, and X:s can be formally written as

$$
\begin{aligned}
F_p = F_j \;\rightarrow\; T_k(\mathbf{x}) = 0 \qquad &\text{if } s_{kj} = 0 \qquad &(4a) \\
F_p = F_j \;\rightarrow\; T_k(\mathbf{x}) \neq 0 \qquad &\text{if } s_{kj} = 1 \qquad &(4b)
\end{aligned}
$$

where F_p, as before, denotes the present fault mode. Note that the implication, denoted by the arrow, is not symmetric. Note also that the interpretation of X is implicitly contained in these two formulas.

In the next section, we will also define interpretations of 1:s, 0:s, and X:s for the decision structure. To the author's knowledge, no such strict interpretation has been defined in previous literature. The motivation for these strict definitions, is that we can discuss relations to for example diagnosis methods based on logic and hypothesis testing. In addition, these "local" interpretations of 1:s, 0:s, and X:s, together with a decision structure, is enough to define the isolation functionality of the whole diagnosis system.

5.2 Decision Structure

The incidence structure corresponds to the case where ideal conditions holds. If this was the case, we could derive the diagnosis S by using the incidence structure, the formulas (4), and the values of the test quantities $T_k(\mathbf{x})$. In practice, the model is not perfect; unmodeled disturbances affects the process, and there is measurement noise. All this means that the formulas (4) are not valid and can therefore not be used to form the diagnosis S.

In practice, we have to relax the assumptions of ideal conditions and the formulas (4) can be replaced by a formulation based on the use of thresholds, i.e. hypothesis testing. Doing this, we obtain a *decision structure*. The new interpretation or of 0:s, 1:s, and X:s becomes

$$
\begin{aligned}
F_p = F_j \;\rightarrow\; T_k(\mathbf{x}) < J_k \qquad &\text{if } s_{kj} = 0 \qquad &(5a) \\
F_p = F_j \;\rightarrow\; T_k(\mathbf{x}) \geq J_k \qquad &\text{if } s_{kj} = 1 \qquad &(5b)
\end{aligned}
$$

The implications are not completely true, but we assume that they holds. This corresponds to the basic assumptions, discussed in Section 4.2, that when H_k^0 is rejected, we assume that H_k^1 holds. However, there is a conflict between the two rules (5a) and (5b). To make the assumption that (5a) holds reasonable, the significance level α_k of all tests must be low. This means that the thresholds must be chosen relatively high. Further, this violates the assumption that (5b) holds. To achieve reasonable assumptions, some or probably most 1:s from the incidence structure must therefore be replaced by X:s. It might seem that another choice is to replace 0:s by X:s, but the problem with this is that for all small faults, the assumption of (5b) still not becomes reasonable.

An example of a decision structure is obtained by considering the incidence structure (3) which can be transformed to (by replacing most 1:s with X:s), for instance the following decision structure:

$$
\begin{array}{c|cccc}
 & NF & F_1 & F_2 & F_3 \\
\hline
\delta_1(\mathbf{x}) & 0 & 0 & X & 0 \\
\delta_2(\mathbf{x}) & 0 & 0 & X & 1 \\
\delta_3(\mathbf{x}) & 0 & X & 0 & X
\end{array}
\tag{6}
$$

Because the decision structure is related to the whole hypothesis tests and not only the test quantities, we use δ_k to label the rows instead of T_k.

By using the formulas (5), it is possible to formally define the interpretation of a whole decision structure. We will exemplify this below, by giving the interpretation of the decision structure (6), but note first that $F_p \notin \{F_2\}$ is the same as $F_p \in \Omega - \{F_2\}$. The symbol $\Leftrightarrow$ will be used to denote tautological equivalence. Now, the interpretation of the decision structure (6) becomes

$$T_1 < J_1 \leftarrow F_p \in \{NF, F_1, F_3\} \quad \Leftrightarrow \quad T_1 \geq J_1 \rightarrow F_p = F_2$$
$$T_2 < J_2 \leftarrow F_p \in \{NF, F_1\} \quad \Leftrightarrow \quad T_2 \geq J_2 \rightarrow F_p \in \{F_2, F_3\}$$
$$T_2 \geq J_2 \leftarrow F_p = F_3 \quad \Leftrightarrow \quad T_2 < J_2 \rightarrow F_p \in \{NF, F_1, F_2\}$$
$$T_3 < J_3 \leftarrow F_p \in \{NF, F_2\} \quad \Leftrightarrow \quad T_3 \geq J_3 \rightarrow F_p \in \{F_1, F_3\}$$

This interpretation of the decision structure (6) can now be used to derive the diagnosis S. For example if $T_1 < J_1$, $T_2 \geq J_2$, and $T_3 \geq J_3$, 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 is clear that there must be a strong relationship between this procedure, i.e. forming the diagnosis S by using the decision structure, and how the diagnosis S is formed by using the individual diagnoses S_k. For example, the sets S_k^0 and S_k^1 for the decision structure (6), are

$$S_1^0 = \{NF, F_1, F_2, F_3\} \qquad S_1^1 = \{F_2\}$$
$$S_2^0 = \{NF, F_1, F_2\} \qquad S_2^1 = \{F_2, F_3\}$$
$$S_3^0 = \{NF, F_1, F_2, F_3\} \qquad S_3^1 = \{F_1, F_3\}$$

That is, the set S_k^0 contains all fault modes which have 0 or X in row δ_k of the decision structure. Also S_k^1 contains all fault modes which have 1 or X in the same row. In this way, the decision structure can be seen as an overview of a diagnosis system based on structured hypothesis tests.

6. DESIGN OF TEST QUANTITIES

In the previous sections, the diagnosis-system architecture *structured hypothesis tests* was proposed. To get a complete diagnosis system, also the individual hypothesis tests need to be constructed. 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). In addition, many methods in the diagnosis literature have been developed.

From Section 4.2 it is realized that the test quantities should be constructed according to the following general principle:

> The test quantity $T_k(\mathbf{x})$ should be small if the data $\mathbf{x}$ matches any of the models $\mathcal{M}(\theta)$, $\theta \in \Theta_k^0$, and large otherwise.

Thus the test quantity can be seen as a measure of the validity of some models $\mathcal{M}(\theta)$. Below, we will exemplify such a model validity measure based on comparison of signals. 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).

6.1 Test Quantities based on Comparison of Signals

Using this principle, the calculation of the test quantity is based on a model validity measure $V_k(\theta, \mathbf{x})$,

which in turn is based on a comparison between signals and/or predictions (or estimates) of signals. The function $V_k(\theta, \mathbf{x})$, where θ is fixed, is a measure of the validity of the model $\mathcal{M}(\theta)$, for a fixed θ, in respect to the measurement data $\mathbf{x}$. To construct $V_k(\theta, \mathbf{x})$, we can compare an output signal y with an estimate $\hat{y}$, but it is also possible to for example compare two different estimates of the same signal. An example is

$$V_k(\theta, \mathbf{x}) = \frac{1}{N} \sum_{t=1}^{N} \|y(t) - \hat{y}(t|\theta, \mathbf{x})\| \qquad (7)$$

where $y(t|\theta, \mathbf{x})$ is the prediction of the output $y(t)$, derived from an assumption of a specific θ and the measured data $\mathbf{x}$. The test quantity can then be calculated as

$$T_k(\mathbf{x}) = \min_{\theta \in \Theta_k^0} V_k(\theta, \mathbf{x}) \qquad (8)$$

To calculate (8), we need in principle to perform a parameter (or signal) estimation. The prime interest here is fault isolation but it is obvious that this parameter estimation means that fault identification implicitly becomes a part of fault isolation. Note that the term *decoupling* in principle corresponds to estimation. The faults (or fault modes) that are decoupled are the fault modes described by the parameters we estimate.

Note that although the model validity measure $V_k(\theta, \mathbf{x})$ in (8) 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).

Often, the effect of disturbances, and also unknown or uncertain parameters, must be handled. These can, in accordance with Section 2.1, be modeled by using the parameter ϕ. The solution is to expand the set of parameters that are estimated, to also include ϕ. In this way, decoupling of disturbances and adaptation to unknown parameters is achieved.

6.1.1. *The Minimization of $V_k(\theta, \mathbf{x})$* In many cases the minimization procedure required in (8) is quite straightforward and can be performed by standard procedures. However, in some cases, more specific solutions are required. Often it is advantageous to first find an estimate $\hat{\theta}$, by using some other algorithm. For example, we can use a state-observer or try to minimize another simpler function $\bar{V}(\theta, \mathbf{x})$. Then the test quantity can be calculated as $T(\mathbf{x}) = V(\hat{\theta}, \mathbf{x})$. Sometimes it is not necessary to explicitly calculate an estimate $\hat{\theta}$, i.e. to carry out the minimization. Instead a closed form expression is found for (8) directly. This is a typical solution in residual generation, which can be seen as a special case of (8).

Example 3. This example illustrates how the principle from Section 6.1 can be applied to a change detection problem. Consider a signal $y(t)$ which can be modeled as

$$y(t) = v(t) + a(t) \qquad v(t) \text{ is iid } N(0, \sigma(t))$$

The functions $a(t)$ and $\sigma(t)$ are equal to known constant values a_0 and σ_0 in the fault free case, but can contain an abrupt change to unknown values a_1 and σ_1 if a fault occurs. We assume $\sigma_1 \geq \sigma_0$.

Further assume that we want to consider three fault modes:

NF "no fault"

F_a "an abrupt change in $a(t)$ at the time t_{ch}"

F_σ "an abrupt change in $\sigma(t)$ at the time t_{ch}"

This means that the fault-state vector can be described as $\theta = [t_{ch}, a_1, \sigma_1]$.

Further we want to design a test quantity for the hypotheses $H_0 : F_p \in \{NF, F_a\}$ and $H_1 : F_p \in \{F_\sigma\}$. By using the general expression (8), the test quantity becomes

$$T(\mathbf{x}) = \min_{\theta \in \Theta^0} V(\theta, \mathbf{x}) = \min_{[t_{ch}, a_1]} \sum_{t=1}^{N} (y(t) - \hat{y}(t|t_{ch}, a_1))^2$$

where
$$\hat{y}(t|t_{ch}, a_1) = \begin{cases} a_0 & \text{if } t < t_{ch} \\ a_1 & \text{if } t \geq t_{ch} \end{cases}$$

This test quantity will be small under H^0 and larger under H^1. Thus all faults belonging to F_a are decoupled in $T(\mathbf{x})$.

7. CONCLUSIONS

This paper has introduced a general formalism for describing and analyzing diagnosis problems. This formalism can be used with any type of fault modeling technique, including for example faults modeled as signals, deviations in constants, and abrupt changes of variables. Also multiple faults are naturally integrated so that no special treatment is needed. Further, the formalism is equally useful for both statistical and deterministical models.

Using this formalism, the diagnosis-system architecture *structured hypothesis tests* has been presented. Such a diagnosis system consists of a set of hypothesis tests. The procedure for how the diagnosis, i.e. the output from a diagnosis system, is formed from the real-valued test quantities (or residuals), is formalized by using a standard interpretation of the functionality of each hypothesis test. 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. One feature becomes that the diagnosis can contain several possible faults that can explain the observed behavior of the system.

Structured hypothesis test is in most cases *not* a replacement to existing principles for fault diagnosis, e.g. parameter estimation, observer schemes, structured residuals, and statistical methods. Instead it is a framework in which these other principles for diagnosis becomes naturally parts, and can be fully exploited. Also, there is a close connection to diagnosis methods based on logic.

Since structured hypothesis tests is defined using a general formalism, we can within the same diagnosis system, and using a general systematic approach, diagnose several different types of faults, e.g. at the same time diagnose faults modeled as additive signals and faults modeled as parameter deviations. Many papers in the field of fault diagnosis discuss decoupling of faults modeled as additive arbitrary signals. The principle of decoupling has in this paper been generalized to include decoupling of faults modeled in arbitrary ways.

Structured hypothesis tests is in a sense a generalization and formalization of the well-known method structured residuals. In contrast to structured residuals, a distinction is introduced, between the incidence structure, describing how faults ideally affect the test quantities, and the decision structure, describing how the faults affect the formation of the diagnosis. By doing so, it has been possible to define meanings of the 0:s, 1:s, and X:s, present in the incidence/decision structure. An introduction of X:s (don't care) in the incidence/decision structure is necessary since only using 0:s and 1:s often places unrealistic requirements on the test quantities (or residuals).

Except for the small examples included in this paper, the theory developed has also been successfully applied to a real application: diagnosis of the air intake system of an automotive engine (Nyberg, 1999b; Nyberg, 2000). In the first work, it is seen how structured hypothesis tests is used to systematically design a diagnosis system capable of diagnosing a large variety of different kinds of faults. In the second, the same framework is used to evaluate different methods for leakage detection. These works have shown that the theory has practical relevance for both design and analysis of diagnosis systems.

8. REFERENCES

Basseville, M. (1997). Information criteria for residual generation and fault detection and isolation. *Automatica* **33**(5), 783–803.

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

Berger, James O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer.

Blanke, M. (1999). *Fault Tolerant Control Systems*. Chap. 6. in Frank (1999).

Bøgh, S. (1995). Multiple hypothesis-testing approach to fdi for the industrial actuator benchmark. *Control Engineering Practice* **3**(12), 1763–1768.

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

Ding, S.X., T. Jeinsch, E.L. Ding, D. Zhou and G. Wang (1999). Detection of observer based fdi schemes to the three tank system. Proc. of the ECC'99.

Frank, P.M., Ed.) (1999). *Advances of Control: Highlights of ECC'99*. Springer.

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

Grainger, R.W., J. Holst, A.J. Isaksson and B.M. Ninnes (1995). A parametric statistical approach to fdi for the industrial actuator benchmark. *Control Engineering Practice* **3**(12), 1757–1762.

Lehmann, E. L (1986). *Testing Statistical Hypotheses*. second ed.. Springer Verlag.

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. (2000). Evaluation of test quantities for leakage diagnosis in the air path of an automotive engine. SAFEPROCESS. IFAC. Budapest, Hungary.

Nyberg, Mattias (1999c). Model Based Fault Diagnosis: Methods, Theory, and Automotive Engine Applications. PhD thesis. Linköping University. URL: http://www.fs.isy.liu.se/Publications/.

Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence* **32**(1), 57–95.

Riggins, R.N. and G. Rizzoni (1990). The distinction between a special class of multiplicative events and additive events: Theory and application to automotive failure diagnosis. American Control Conf.. San Diego, California. pp. 2906–2911.