

FAULT ISOLATION FOR DIAGNOSIS: NUISANCE REJECTION AND MULTIPLE HYPOTHESES TESTING

Michèle Basseville * Igor Nikiforov **

* IRISA/CNRS, 35042 Rennes Cedex, F. basseville@irisa.fr

** LM2S/UTT, BP 2060, 10010 Troyes Cedex, F. nikiforov@utt.fr

Abstract Fault detection, fault isolation and fault diagnosis are addressed within a statistical framework. The corresponding inference problems are stated. Several statistical tools for solving these inference problems are described. Particular emphasis is put on dealing with nuisance parameters and deciding between multiple hypotheses. How to use these tools for solving FDI problems is discussed. An example illustrates some of the proposed methods.

Keywords: Hypotheses testing, nuisance parameters, multiple hypotheses, fault isolation and diagnosis.

1. INTRODUCTION

Monitoring complex structures and processes is necessary for fatigue prevention, aided control and condition-based maintenance. Many industrial processes rely on physical principles, which write in terms of (differential) equations and thus (dynamical) models. Moreover, the use of (physical) parameters is mandatory for fault isolation and diagnosis. Faults can thus often be modeled as deviations, w.r.t. a nominal reference value, in the parameter vector of a stochastic system. A crucial issue is to state the significance of the observed changes w.r.t. noises, uncertainties, and changes in the environment of the monitored process.

On the other hand, there exist mathematical statistics theories and tools for solving hypotheses testing problems. Key features of these methods are their ability to handle noises and uncertainties, to reject nuisance parameters, to select one among several hypotheses. The purpose of this paper is to describe the key components of these theories and to explain how to use these tools for fault detection, isolation and diagnosis (FDI).

The paper is organized as follows. Several inference problems are introduced in section 2. Section 3 is devoted to major statistical tools for solving these hypotheses testing problems. Composite hypotheses, nuisance parameters and multiple hypotheses are dis-

cussed, minimax, invariant, and most stringent tests are introduced, and some asymptotic approaches are described. The use of these hypotheses testing tools for solving fault detection, isolation and diagnosis problems is addressed in section 4. An example illustrating the relevance of some of the proposed tools is described in section 5. Some discussions and conclusions are drawn in section 6.

2. SEVERAL INFERENCE PROBLEMS

We introduce the detection, isolation and diagnosis problems, stated in a statistical framework, distinguishing the test of hypotheses for a whole data sample and the test for the presence of a change within the sample. It is assumed throughout that the signature of the faults on the model of the monitored system is a *change* in its *parameter* vector. The models considered in this paper are described, and two change types are defined. Then several assumptions on the parameter, corresponding to different types of monitoring problems, are introduced.

2.1 Parameterized model

The measured data Y are viewed as the output of a continuous-time system, which, ignoring the dynam-

ics for a while, we write

$$Y(t) = h(X(t), U(t), \theta) + \xi(t) \quad (1)$$

where X is an unknown variable (typically a state), U the input vector, θ the parameter vector, Y the measured output, h a known function, and ξ a white noise. Sampling the data at period δ results in the discrete-time model

$$Y_k = H(X_k, U_k, \theta) + \xi_k \quad (2)$$

with $Y_k = Y(k\delta)$, $U_k = U(k\delta)$. At time n , the data sample is $(Y_0, \dots, Y_n, U_0, \dots, U_n)$ if the input U is measured, or (Y_0, \dots, Y_n) if U is unknown.

Emphasis is put throughout on the simpler model

$$Y = HX + M\theta + \xi, \quad \xi \sim \mathcal{N}(0, \Sigma) \quad (3)$$

where $Y \in \mathbb{R}^r$, $X \in \mathbb{R}^p$, $\theta \in \mathbb{R}^m$, with $p + m \leq r$, matrices H and M are full column rank (f.c.r.), and ξ is a zero mean Gaussian noise. In this static model, it is assumed that the processed data Y are Gaussian vectors, and that the faults affect their mean value. As explained in section 4, the relevance of this assumption is of much wider scope than it seems to be: a large class of FDI problems in *dynamic* systems can be reduced to the universal *static* problem of monitoring the mean value of a Gaussian vector. This is achieved with the aid of a convenient residual generation.

2.2 Two change situations

The two following situations are distinguished.

Hypotheses testing. In the first situation, parameter vector θ is assumed to be *constant* within the entire data sample. The *null hypothesis* \mathcal{H}_0 corresponds to the fault-free case, and thus to the nominal value θ_0 of the parameter, possibly within a set Θ_0 :

$$\mathcal{H}_0 : \theta \in \Theta_0 \subset \mathbb{R}^m \quad (4)$$

The *alternative hypotheses* \mathcal{H}_i correspond to different fault modes:

$$\mathcal{H}_i : \theta \in \Theta_i \subset \mathbb{R}^m, \quad (i = 1 : K) \quad (5)$$

where $\Theta_i \cap \Theta_j = \emptyset$ for $i \neq j$. In case of a single fault mode, the only problem to solve is the detection one. When $K > 1$, the isolation and diagnosis problems have to be solved also.

Occurrence of a change. In the second situation, parameter θ can *change* within the data sample at an unknown instant v ($1 \leq v \leq n$), and thus is a function of time: for one fault mode i ($i = 1 : K$),

$$\theta(t) \in \begin{cases} \Theta_0 & \text{if } t < v \\ \Theta_i & \text{if } t \geq v \end{cases} \quad (6)$$

These two situations lead to different formal statements of the hypotheses and of the criteria to be used

for the design and performance evaluation of the decision algorithms. In this paper, we concentrate on the first situation. The interested reader is referred to (Nikiforov, 2002) for the second one.

2.3 Hypotheses testing problems for FDI

The following testing problems are central to FDI.

Detection. Detection refers to deciding whether the monitored system is in its nominal (safe) state or not. This amounts to testing the null hypothesis \mathcal{H}_0 in (4) against a *global alternative* hypothesis \mathcal{H}_1 . In case of no information about the fault modes, the alternative hypothesis writes

$$\mathcal{H}_1 : \theta \in \Theta_0^c \stackrel{\text{def}}{=} \mathbb{R}^m \setminus \Theta_0 \quad (7)$$

When some information about the fault modes is available under the form (5), the alternative writes

$$\mathcal{H}_1 : \theta \in \left(\bigcup_{i=1}^K \Theta_i \right) \quad (8)$$

Several statistical testing tools are described in 3.2 and 3.3. This detection problem refers to the case of two hypotheses. Any detection procedure should perform a tradeoff between two incorrect decisions: false alarm (false rejection of the null hypothesis) and non detection (missed acceptance, or equivalently false rejection, of the alternative hypothesis). If *a priori* probabilities of the safe and fault modes are available, the probabilities of these two errors can be weighted and combined into a single performance index for evaluating the detection scheme. If not, both errors are handled, as made precise below.

Isolation. In case of two fault modes or more, isolation refers to deciding which fault mode occurred. Two basic approaches can be undertaken for this purpose. The first one consists in deciding in favor of one fault mode while considering the other fault modes as *nuisance information*. Isolating fault i then amounts to testing the null hypothesis \mathcal{H}_0 in (4) against the alternative \mathcal{H}_i in (5). For handling nuisances, minimax and *invariant* tests are shown in 3.4 to provide us with algorithms which detect (isolate) a given fault i while being insensitive to the other faults. As explained in section 4, running simultaneously several isolation tests can make up a diagnosis procedure for simultaneous faults.

Multiple hypotheses testing is the second approach, when a single fault at a time is assumed. Here, the null hypothesis \mathcal{H}_0 is tested against K alternatives $\mathcal{H}_1, \dots, \mathcal{H}_K$. Efficient solutions for this are described in section 3.5.

Any isolation procedure should balance all the possible combinations of errors in making the decision. Should *a priori* probabilities of the fault modes be

available or not, handling simultaneously all the errors is a hard issue. To mention but one point, is it preferable to put a constraint on the probability of at least one false rejection, or on the expected number of false rejections? Several performance indexes are investigated below.

3. HYPOTHESES TESTING APPROACHES

Hypotheses testing methods are now described. Problem statements and criteria are introduced in 3.1. In 3.2, several approaches to the design of optimum tests for composite hypotheses are discussed. Asymptotic approaches for dealing with large data samples are described in 3.3, where a reparameterization result of Wald turns out to be powerful and relevant to FDI. Dealing with nuisance parameters is the topic of 3.4, based on some of the approaches presented before. The case of multiple hypotheses is discussed in 3.5. This forms the algorithmic basis of statistical FDI.

3.1 Problem statement and criteria

A (fixed) n -size sample of independent observations Y_1, \dots, Y_n is available and supposed to be generated by one among $(K + 1)$ probability distributions P_0, \dots, P_K . Possible inputs U_1, \dots, U_n are assumed non random; the corresponding data sample is denoted by $Y_1, \dots, Y_n, U_1, \dots, U_n$ though. The hypotheses testing problem consists in deciding which distribution P_i is the true one. To each distribution P_i corresponds an hypothesis \mathcal{H}_i . A *statistical test* for testing between the \mathcal{H}_i 's is any measurable mapping $\delta : (\mathcal{Y}, \mathcal{U}) \rightarrow \{\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_K\}$ from the observation space onto the set of hypotheses. We concentrate on parameterized distributions P_θ with density f_θ .

Two types of hypotheses have to be distinguished. A *simple hypothesis* \mathcal{H}_i is defined by a unique value of the parameter vector: $\mathcal{H}_i : \theta = \theta_i$. A *composite hypothesis* refers to a set of parameters

$$\mathcal{H}_i : \theta \in \Theta_i \quad (9)$$

with $\Theta_i \subset \Theta \subseteq \mathbb{R}^m$. We assume that $\Theta_i \cap \Theta_j = \emptyset$ for $i \neq j$. Composite hypotheses are more relevant than simple ones in practice, because of limited available amount of information, especially for the alternatives (fault modes).

The quality of a statistical test is defined with a set of *error probabilities*:

$$\alpha_i = \mathbb{P}_i(\delta \neq \mathcal{H}_i), \quad i = 0 : K,$$

where \mathbb{P}_i stands for observations Y_1, \dots, Y_n being generated by distribution P_i . The *power* is defined with a set of probabilities of correct decisions:

$$\beta_i = \mathbb{P}_i(\delta = \mathcal{H}_i)$$

For a non-randomized test, the *critical function* is $\pi(Y) = i$ when $\delta(Y) = \mathcal{H}_i$.

Until 3.5, we assume $K=1$ alternative hypothesis. The pair $(\alpha \stackrel{\text{def}}{=} \alpha_0, \beta \stackrel{\text{def}}{=} \beta_1)$ is then a sufficient performance index, and the *decision function*

$$\delta(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda(Y) < h(\alpha) \\ \mathcal{H}_1 & \text{if } \Lambda(Y) \geq h(\alpha) \end{cases}, \quad (10)$$

is defined by the *statistics* Λ and the threshold h .

3.2 Composite hypotheses testing

For composite hypotheses, tests are searched within the class of tests with upper-bounded maximum false alarm probability

$$\mathcal{K}_\alpha = \left\{ \delta : \sup_{\theta \in \Theta_0} \mathbb{P}_0(\delta \neq \mathcal{H}_0) \leq \alpha \right\}, \quad (11)$$

and a test δ is evaluated with the *power function*:

$$\forall \theta \in \Theta_1, \beta_\delta(\theta) = \mathbb{P}_\theta(\delta = \mathcal{H}_1) = 1 - \alpha_\delta(\theta).$$

A test δ^* is said *uniformly most powerful* (UMP) in the class \mathcal{K}_α if:

$$\forall \delta \in \mathcal{K}_\alpha, \forall \theta \in \Theta_1 : \beta_{\delta^*}(\theta) \geq \beta_\delta(\theta) \quad (12)$$

Unfortunately, UMP tests scarcely exist, except when parameter θ is scalar, the family of distributions $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ has a monotone likelihood ratio, and the test is one-sided, namely: $\mathcal{H}_0 : \{\theta \leq \theta_0\}$ and $\mathcal{H}_1 : \{\theta > \theta_1\}$ with $\theta_1 \geq \theta_0$ (Lehmann, 1986; Borovkov, 1987).

In case of a vector parameter θ , the crucial issue is to find an optimal solution over a set of alternatives which is rich enough. Actually, a UMP test often exists only for a subset $\bar{\Theta}_1 \subset \Theta_1$ but not for the whole set Θ_1 . For example, a UMP test exists if this subset is defined by a straight line in the parametric space and the alternative is one-sided.

To overcome this difficulty, and find an optimal test for multidimensional composite hypotheses, several approaches exist.

- *Bayesian approach*: Assume that θ is a random vector and introduce some *a priori* information on the distribution of θ .
- *Minimax approach*: Consider only the worst case situation, which often amounts to consider the closest alternatives.
- *Invariant tests*: Take advantage of an invariance of the distributions under some transformations, and impose the corresponding restrictions on the class of statistical tests in order to simplify the initial problem and find an optimum (invariant) test.
- *Constant power approach*: Impose some additional constraints on the class \mathcal{K}_α , in order to avoid tests UMP over a subset $\bar{\Theta}_1$ of Θ_1 and very inefficient over $\Theta_1 \setminus \bar{\Theta}_1$. For instance, require the best *constant power* over a family of surfaces defined on Θ_1 .

- *Most stringent tests:* Find a test δ in \mathcal{K}_α which minimizes a convenient difference between $\sup_{\delta \in \mathcal{K}_\alpha} \beta_\delta(\theta)$ and $\beta_\delta(\theta)$.

Bayesian approach. The Bayesian approach exploits some *a priori* information on the distribution of θ , e.g. $\mathbb{P}(\mathcal{H}_j) = q_j, q_0 + q_1 = 1$. It leads to relatively simple and well investigated theoretical schemes, which implementation involves only tuning parameters which are functions of the *a priori* information (Borovkov, 1987). The drawbacks for FDI lay in that the *a priori* information on the faults may be not reliable; and that, for safety-critical applications or when intentional faults should be considered, the risk function

$$R(Q, \pi) = 1 - q_0 \mathbb{E}_0(1 - \pi(Y)) - q_1 \mathbb{E}_1(\pi(Y))$$

is not a convenient criterion.

Nevertheless, the Bayesian approach is useful at a preliminary stage in the design of FDI algorithms, in order to obtain optimal schemes, using *invariant* properties or minimax approaches. An example is given in 3.5.

Minimax approach. It consists in optimizing the worst case situation (Lehmann, 1986; Borovkov, 1987). A test $\bar{\delta}$ is *minimax* in \mathcal{K}_α if it maximizes the minimum power in this class:

$$\forall \delta \in \mathcal{K}_\alpha : \inf_{\theta \in \Theta_1} \beta_{\bar{\delta}}(\theta) \geq \inf_{\theta \in \Theta_1} \beta_\delta(\theta) \quad (13)$$

This often amounts to consider the closest alternatives (Borovkov, 1987). Actually, for model:

$$Y = M\theta + \xi, \quad \xi \sim \mathcal{N}(0, \Sigma) \quad (14)$$

and the two sets $\Theta_0 = \{\theta : \theta^T M^T \Sigma^{-1} M \theta \leq a^2\}$ and $\Theta_1 = \{\theta : \theta^T M^T \Sigma^{-1} M \theta \geq b^2\}$ ($a^2 < b^2$), it is easy to show that the worst parameters lay on the boundaries of Θ_0 and Θ_1 , and the minimax test $\bar{\delta}$ writes as in (10) with:

$$\bar{\Lambda} \stackrel{\text{def}}{=} Y^T \Sigma^{-1} M (M^T \Sigma^{-1} M)^{-1} M^T \Sigma^{-1} Y \quad (15)$$

and thus is based on a χ^2 statistics.

In more general cases, the design of minimax tests (13) is often more complex. One reason is the lack of invariance property for exhibiting the worst case parameters (Wald, 1943; Borovkov, 1987).

Invariant tests. Let \mathcal{G} be a group of transformations of \mathbb{R}^r . A parametric family of distributions $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ is *invariant under* \mathcal{G} if (Lehmann, 1986; Borovkov, 1987):

$$\forall g \in \mathcal{G}, \forall \theta \in \Theta, \exists \theta_g \stackrel{\text{def}}{=} \bar{g}(\theta) \in \Theta : \mathbb{P}_\theta(g(Y) \in A) = \mathbb{P}_{\bar{g}(\theta)}(Y \in A)$$

The transformations \bar{g} form a group $\bar{\mathcal{G}}$ on Θ . An hypotheses testing problem, between \mathcal{H}_0 and \mathcal{H}_1 as in (9), is said to be *invariant* if :

- the family \mathcal{P} remains invariant under \mathcal{G} ;
- the subsets Θ_i remain invariant under $\bar{\mathcal{G}}$.

A test δ is said to be invariant if its critical function $\pi(Y)$ remains invariant under \mathcal{G} :

$$\forall Y \in \mathbb{R}^r, \forall g \in \mathcal{G} : \pi(Y) = \pi(g(Y))$$

The design of invariant tests is based on *maximal invariant statistics*. A statistics Λ is maximal invariant under \mathcal{G} if it is invariant and if

$$\Lambda(X) = \Lambda(Y) \Rightarrow \exists g \in \mathcal{G} : Y = g(X)$$

The important feature of this approach is the possibility to reduce a more general (and mathematically more complex) statistical problem to another one less general and often less complex (Bernshtein, 1980; Lehmann, 1986; Borovkov, 1987; Scharf, 1991; Porat and Friedlander, 1993; Scharf and Friedlander, 1994; Burgess and Veen, 1996; King, 1996). For instance, all the results presented in subsection 3.2 for the case of *known* covariance matrix Σ can be directly obtained from the case of unit covariance matrix by using the invariance of the Gaussian distribution $\mathcal{N}(\theta, I)$ under the group of linear transformations $g(Y) = \bar{\Sigma}Y$, where $\bar{\Sigma}\bar{\Sigma}^T = \Sigma$.

Using the invariance of a parametric family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ under a suitable group of transformations is a powerful and efficient method to solve an hypotheses testing problem when a UMP test does not exist. This is especially relevant in case of nuisance parameters, as discussed below.

Constant power approach. The constraint of constant power over a family of surfaces can be introduced as follows. Assume a simple null hypothesis : $\mathcal{H}_0 : \{\theta = \theta_0\}$, and the alternative $\mathcal{H}_1 : \{\theta \neq \theta_0\}$, and define a parametric family of surfaces $\mathcal{S} = \{S_c\}$ around the point θ_0 . A test $\delta^* \in \mathcal{K}_\alpha$ is said to be *UMP with constant power* over a family $\mathcal{S} = \{S_c\}$ of surfaces, if:

$$\forall \theta_1, \theta_2 \in S_c : \beta_{\delta^*}(\theta_1) = \beta_{\delta^*}(\theta_2) \quad (16)$$

$$\forall \delta \in \mathcal{K}_\alpha \text{ s.t. } \forall \theta_1, \theta_2 \in S_c : \beta_\delta(\theta_1) = \beta_\delta(\theta_2),$$

$$\text{then : } \beta_{\delta^*}(\theta) \geq \beta_\delta(\theta).$$

The rationale is depicted on Fig. 1 for model (14) where $\theta_0 = 0$. Surfaces S_c are ellipsoids:

$$S_c : \theta^T M^T \Sigma^{-1} M \theta = c^2 \quad (c > 0) \quad (17)$$

The upper part of the figure is devoted to the scalar case $r=1$ with $\sigma^2=1$. The dotted line is the power function of a UMP one-sided test, whereas the solid line is the best constant power function $\beta_{\delta^*}(\theta)$ over a family of surfaces $\mathcal{S} = \{S_c : |\theta| = c |c > 0\}$. It can be seen that the one-sided UMP test outperforms the UMP test with constant power, only for the alternatives ($\theta > 0$), but is very inefficient for other alternatives ($\theta < 0$). Therefore, constant power

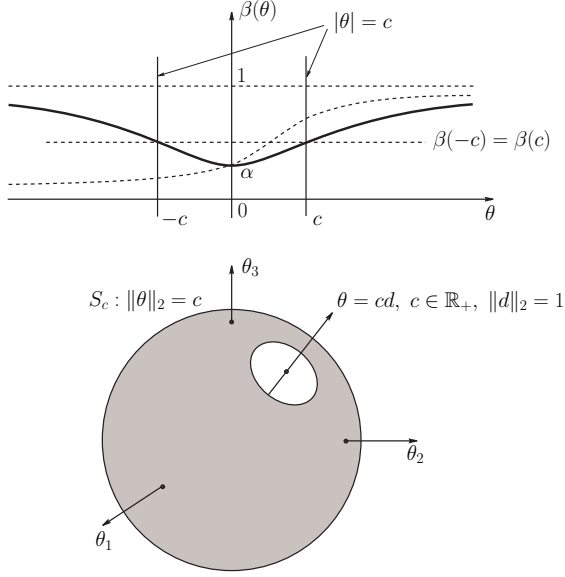


Figure 1. Best constant power function over a family of surfaces.

constraint (16) eliminates tests like UMP one-sided ones, for achieving a uniform performance over the surfaces S_c in the alternative Θ_1 . The lower part of Fig. 1 is devoted to a higher dimensional case with $M = \Sigma = I$, and a surface (17) is depicted. For any direction defined by a unit vector d , a one-sided UMP test can be designed for testing $\mathcal{H}_0 : \{\theta = 0\}$ against a limited alternative, say $\tilde{\mathcal{H}}_1 : \{\theta = cd | c > 0\}$. As before, this directional UMP test outperforms the UMP test with constant power if $\theta = cd, c > 0$, but is inefficient for other directions. In the general case of model (14) ($\Sigma \neq I$), the test with the best constant power over the family of surfaces (17) is given by (15).

Most stringent tests. Let

$$\beta(\theta, \alpha) = \sup_{\delta \in \mathcal{K}_\alpha} \beta_\delta(\theta)$$

be the maximum power among the tests in the class \mathcal{K}_α . A test δ^* is said to be *most stringent* in \mathcal{K}_α if it minimizes a convenient difference between this maximum power and its own one

$$\sup_{\theta \in \Theta_1} \{\beta(\theta, \alpha) - \beta_{\delta^*}(\theta)\} \leq \sup_{\theta \in \Theta_1} \{\beta(\theta, \alpha) - \beta_\delta(\theta)\}$$

This is illustrated in Fig. 2. In the Gaussian case (14), a remarkable fact is that the most stringent test is given again by (15). The (involved) proof can be found in (Wald, 1943).

3.3 Asymptotic approaches

When the number of observations is large, several asymptotic approaches can be used to design tests between composite hypotheses.

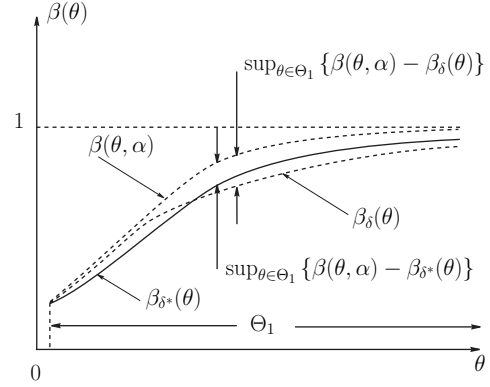


Figure 2. Most stringent test.

Likelihood ratio tests. Assume a simple $\mathcal{H}_0 : \{\theta = \theta_0\}$, and the composite $\mathcal{H}_1 : \{\theta \neq \theta_0\}$. Let

$$\mathcal{F}(\theta) = \mathbb{E}_\theta(\partial \log f_\theta(Y) / \partial \theta) (\partial \log f_\theta(Y) / \partial \theta)^T \quad (18)$$

be Fisher information matrix, and let $\mathcal{S} = \{S_c\}$,

$$S_c : (\theta - \theta_0)^T \mathcal{F}(\theta_0) (\theta - \theta_0) = c^2 \quad (c > 0),$$

be the family of surfaces, counterpart of (17). The definitions of asymptotic optimality are somewhat complex (Wald, 1943). All asymptotic results are associated with a sequence of tests $\{\delta_n\}$, and should be interpreted as a limit when $n \rightarrow \infty$. In the independent case, it can be shown (Wald, 1943; Rao, 1973; Zacks, 1971; Hall and Mathiason, 1990) that both the test with asymptotically best constant power over the family of surfaces S_c and the asymptotically most stringent test $\bar{\delta}$ are given by rule (10) with

$$\bar{\Lambda}(Y_1, \dots, Y_n) = n(\hat{\theta}_n - \theta_0)^T \mathcal{F}(\theta_0) (\hat{\theta}_n - \theta_0) \quad (19)$$

where $\hat{\theta}_n = \arg \sup_\theta f_\theta(Y_1, \dots, Y_n)$ is the maximum likelihood estimate (MLE) of θ . Note that Wald's test (19) is the generalization of the minimax test (15). It also writes as the generalized likelihood ratio (GLR) test $\tilde{\delta}$ defined by (10) with

$$\tilde{\Lambda}(Y_1, \dots, Y_n) = \log \frac{\sup_\theta f_\theta(Y_1, \dots, Y_n)}{f_{\theta_0}(Y_1, \dots, Y_n)} \quad (20)$$

The tests $\bar{\delta}$ and $\tilde{\delta}$ are asymptotically equivalent (Wald, 1943; Rao, 1973).

Reparameterization. Another approach uses a reparameterization for imposing some additional constraints and simplifying the testing problem (Wald, 1943). Let $\mathcal{P} = \{P_\theta, \theta \in \Theta \subseteq \mathbb{R}^m\}$ be a family of distributions, and the *composite null hypothesis*

$$\mathcal{H}_0 : \{\theta \in \Theta : \eta(\theta) = 0\} \quad (21)$$

where η is a l -dimensional vector-valued function, $l < m$. The existence of $(m - l)$ additional functions over \mathbb{R}^m is enforced, for the extended vector-valued function $\tilde{\eta} = (\eta_1, \dots, \eta_l, \eta_{l+1}, \dots, \eta_m)$ to play the role of a reparameterization. This is depicted in Fig. 3, where hypothesis \mathcal{H}_0 is the shading surface. As explained below, the degrees of freedom therein can be

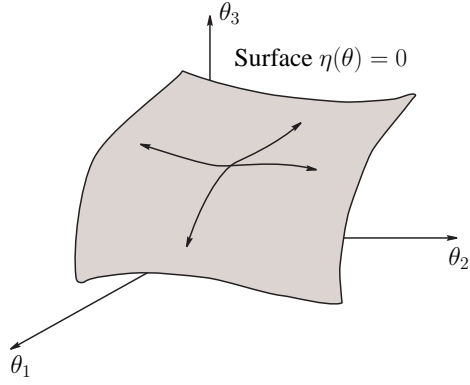


Figure 3. The general null composite hypothesis.

exploited to deal with nuisance parameters. The alternative hypothesis \mathcal{H}_1 is $\{\theta \in \Theta : \eta(\theta) \neq 0\}$. Under some regularity conditions on function $\tilde{\eta} : \mathbb{R}^m \rightarrow \mathbb{R}^m$, both the test with the asymptotically best constant power and the asymptotically most stringent test can be shown (Wald, 1943) to be given by (10), with:

$$\bar{\Lambda}(Y_1, \dots, Y_n) = n \eta(\hat{\theta}_n)^T \tilde{\mathcal{F}}_l^{-1}(\hat{\theta}_n) \eta(\hat{\theta}_n) \quad (22)$$

where $\hat{\theta}_n = \arg \sup_{\theta} f_{\theta}(Y_1, \dots, Y_n)$ is the MLE of θ , matrix $\tilde{\mathcal{F}}_l$ is the $l \times l$ -dimensional upper-left block of matrix $J\mathcal{F}^{-1}J^T$ and J is the $m \times m$ -dimensional Jacobian matrix $\partial\tilde{\eta}/\partial\theta$. The sensitivity method discussed in (Basseville, 1997) is a particular case of this approach.

Local approach. In case of *dependent* observations, another asymptotic approach can be used for designing optimal tests between composite hypotheses (Roussas, 1972; Le Cam, 1986; Le Cam and Yang, 1990). This approach assumes that the sets $\Theta_i(n)$ are getting closer as $n \rightarrow \infty$, namely : $\Theta_i(n) = \theta^* + \vartheta_i/\sqrt{n}$, where $\vartheta_i \subset \mathbb{R}^m$. For testing $\mathcal{H}_0(n) : \{\theta = \theta^*\}$ against $\mathcal{H}_1(n) : \{\theta = \theta_n \stackrel{\text{def}}{=} \theta^* + \Upsilon/\sqrt{n}\}$, the local approach replaces the log-likelihood ratio

$$S(\Upsilon) = \log(f_{\theta_n}(Y_1, \dots, Y_n)/f_{\theta^*}(Y_1, \dots, Y_n))$$

by the expansion (Roussas, 1972; Le Cam, 1986; Le Cam and Yang, 1990)

$$S(\Upsilon) \simeq \zeta_n^T(\theta^*)\Upsilon - 1/2 \Upsilon^T \mathcal{F}(\theta^*)\Upsilon, \text{ where}$$

$$\zeta_n(\theta) = 1/\sqrt{n} \sum_{k=1}^n \partial \log f_{\theta}(Y_k)/\partial\theta \quad (23)$$

is the *efficient score* and Fisher information $\mathcal{F}(\theta)$ is positive definite ($\forall\theta \in \Theta$). Vector ζ_n is asymptotically Gaussian distributed under both hypotheses

$$\zeta_n(\theta^*) \rightarrow \begin{cases} \mathcal{N}(0, \mathcal{F}(\theta^*)) & \text{if } P_{\theta^*} \\ \mathcal{N}(\mathcal{F}(\theta^*)\Upsilon, \mathcal{F}(\theta^*)) & \text{if } P_{\theta^* + \Upsilon/\sqrt{n}} \end{cases} \quad (24)$$

The GLR test between $\mathcal{H}_0(n)$ and $\mathcal{H}_1(n)$ is easily found to be based on $\Lambda(\zeta) = \zeta_n^T \mathcal{F}^{-1}(\theta^*)\zeta_n$. Moreover, testing between K local hypotheses \mathcal{H}_i can be reduced to testing between K Gaussian hypotheses

defined by the mean vector of the efficient score (23). This approach has been extended to other parameter estimating functions than the efficient score (Basawa, 1985; Benveniste *et al.*, 1987). The application of these approaches to fault detection and isolation is discussed in (Basseville and Nikiforov, 1993; Basseville, 1998).

3.4 Dealing with nuisance parameters

The parameter vector is now partitioned as:

$$\theta^T = (\phi^T, \psi^T), \phi \in \mathbb{R}^{m_{\phi}}, \psi \in \mathbb{R}^{m_{\psi}}, m_{\phi} + m_{\psi} = m \quad (25)$$

where ϕ (resp. ψ) is the informative (resp. nuisance) parameter. Assume the hypotheses:

$$\mathcal{H}_0 : \{\phi = \phi_0, \psi \in \mathbb{R}^{m_{\psi}}\}, \mathcal{H}_1 : \{\phi \neq \phi_0, \psi \in \mathbb{R}^{m_{\psi}}\} \quad (26)$$

about ϕ , while considering ψ as an *unknown* vector. Since fault isolation can be viewed as distinguishing subsets of components of θ , testing hypotheses with nuisance parameters is an important FDI issue. There are several approaches to dealing with nuisance parameters.

Minimax approach. This consists in maximizing the minimum power β over the unknown vector ψ . Since the power is an increasing function of Kullback-Leibler “distance” between the two densities f_{ϕ_0, ψ_0} and f_{ϕ_1, ψ_1} (Lehmann, 1986):

$$\rho(\phi_1, \phi_0, \psi_1, \psi_0) = \int \log \frac{f_{\phi_1, \psi_1}(Y)}{f_{\phi_0, \psi_0}(Y)} f_{\phi_1, \psi_1}(Y) dY,$$

this amounts to search for two least favorable values ψ_0^* and ψ_1^* which minimize ρ , and to compute the optimal GLR test (20) for these values. For linear Gaussian models (14), this boils down to a quadratic minimization problem. In this case, partition matrix M and Fisher information

$$M = (M_{\phi} \ M_{\psi}), \mathcal{F} \stackrel{\text{def}}{=} M^T \Sigma^{-1} M = \begin{pmatrix} \mathcal{F}_{\phi\phi} & \mathcal{F}_{\phi\psi} \\ \mathcal{F}_{\psi\phi} & \mathcal{F}_{\psi\psi} \end{pmatrix} \quad (27)$$

Assume a null hypothesis (26) and the simple alternative $\mathcal{H}_1 : \{\phi = \phi_1, \psi \in \mathbb{R}^{m_{\psi}}\}$. The expectation φ_j of the observation Y under \mathcal{H}_j , and assuming a constant nuisance ψ_j , is

$$\varphi_j = \mathbb{E}(Y) = M_{\phi}\phi_j + M_{\psi}\psi_j, \quad j = 0, 1.$$

Then Kullback-Leibler information writes:

$$\rho(\phi_1, \phi_0, \psi_1, \psi_0) = (\varphi_1 - \varphi_0)^T \Sigma^{-1} (\varphi_1 - \varphi_0) / 2.$$

Since ρ is a function of the difference $x = \psi_1 - \psi_0$, its minimum should be searched w.r.t. x . It is reached at $x^* = -\mathcal{F}_{\psi\psi}^{-1} \mathcal{F}_{\psi\phi} (\phi_1 - \phi_0)$ and is equal to $\rho^* = (\phi_1 - \phi_0)^T \mathcal{F}_{\phi\phi}^* (\phi_1 - \phi_0) / 2$, where $\mathcal{F}_{\phi\phi}^*$ is the upper left term of \mathcal{F}^{-1} . Last, the LR for the least favorable value x^* of the nuisance is

$$\begin{aligned} \bar{\Lambda}(Y) &= \log \frac{f_{\phi_1, \psi_1^*}(Y)}{f_{\phi_0, \psi_0^*}(Y)} \\ &= (\phi_1 - \phi_0)^T M_{\phi}^T \bar{\Sigma}^{-T} P_{\bar{\Sigma}}^{-1} (Y - M_{\phi}\phi_0) \\ &\quad - (\phi_1 - \phi_0)^T \mathcal{F}_{\phi\phi}^* (\phi_1 - \phi_0) / 2 \end{aligned}$$

where $P_{\tilde{M}} = I - \bar{\Sigma}^{-1} M_\psi (M_\psi^T \bar{\Sigma}^{-1} M_\psi)^{-1} M_\psi^T \bar{\Sigma}^{-T}$ and $\bar{\Sigma} \bar{\Sigma}^T = \Sigma$. This minimax test is independent of the unknown values ψ_1 and ψ_0 , and coincides with the LR test where the likelihoods are maximized w.r.t. the nuisance parameter ψ :

$$\bar{\Lambda}(Y) = \tilde{\Lambda}(Y) = \log \frac{\sup_{\psi} f_{\phi_1, \psi}(Y)}{\sup_{\psi} f_{\phi_0, \psi}(Y)}$$

For a composite alternative $\mathcal{H}_1 : \{\phi \neq \phi_0\}$, the minimax test coincides with the GLR test (20) based also on the maximized likelihoods:

$$\bar{\Lambda}(Y) = \tilde{\Lambda}(Y) = \log \frac{\sup_{\phi, \psi} f_{\phi, \psi}(Y)}{\sup_{\psi} f_{\phi_0, \psi}(Y)}$$

Invariance. Since the nuisance parameter $\psi \in \mathbb{R}^{m_\psi}$ is completely unknown, the decision function $\Lambda(Y)$ should be independent of its value. To obtain a statistics which is independent of the nuisance parameter, the theory of invariance can be used. For instance, if the distribution of the observation Y depends on $g(\psi)$, where g is a vector-valued function, then it is natural to state the hypotheses testing problem as invariant under the group of transformations $G = \{g : \psi_g = g(\psi)\}$. To show how this principle works, we consider the simplest partitioning in (14):

$$Y = M_\phi \phi + \mathbf{1}_r \psi + \xi, \quad \xi \sim \mathcal{N}(0, \Sigma) \quad (28)$$

where $\psi \in \mathbb{R}$. The statistics which remains invariant under the group of translations $G = \{Y \mapsto g(Y) = Y + \mathbf{1}_r \psi\}$ is (Lehmann, 1986):

$$Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{r-1} \end{pmatrix} \stackrel{\text{def}}{=} VY = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{pmatrix}$$

and it is maximal invariant. Matrix V plays the role of a rejector of the nuisance parameter ψ , and such an invariant vector Z corresponds to a *parity vector* in the FDI literature (see 4.2). The family $\mathcal{P} = \{\mathcal{N}(M_\phi \phi + \mathbf{1}_r \psi, \Sigma), \phi \in \mathbb{R}^{m_\phi}\}$ remains invariant under G , and the group \bar{G} is defined by: $\bar{g}(\phi) = M_\phi \phi + \mathbf{1}_r c$ ($c \in \mathbb{R}$). The testing problem between $\mathcal{H}_0 : \{VM_\phi \phi = 0\}$ and $\mathcal{H}_1 : \{VM_\phi \phi \neq 0\}$ remains invariant under G . From (28), we obtain:

$$Z \stackrel{\text{def}}{=} VY = VM_\phi \phi + V\xi, \quad V\xi \sim \mathcal{N}(0, V\Sigma V^T).$$

It results from formula (19) applied to model (14), that the invariant statistics is:

$$\bar{\Lambda}(Y) = Y^T V^T (V\Sigma V^T)^{-1} VM_\phi \cdot (M_\phi^T V^T (V\Sigma V^T)^{-1} VM_\phi)^{-1} \cdot M_\phi^T V^T (V\Sigma V^T)^{-1} VY$$

and the invariant test $\bar{\delta}$ with best constant power over the family of surfaces

$$S_c : \phi^T M_\phi^T V^T (V^T \Sigma V)^{-1} VM_\phi \phi = c^2$$

is given by (10) with this statistics $\bar{\Lambda}$. In this case the model is linear-Gaussian and Wald's theory can be applied non asymptotically.

Reparameterization. The approach described for the null hypothesis (21) can be used for testing the hypotheses with nuisance parameters in (26). For the model: $Y = M\theta + \xi = M_\phi \phi + M_\psi \psi + \xi$, where θ is partitioned as in (25), constraint (21) writes, for $l = m_\phi$: $\eta(\theta) = 0 \iff \theta_i - \phi_{0,i} = 0$ ($i = 1 : m_\phi$). The application to fault isolation and diagnosis is investigated in section 4.

3.5 Multiple hypotheses testing

Testing between $K \geq 2$ hypotheses is a difficult problem, for which few theoretic results exist.

Bayesian approach. Assume $(K + 1)$ simple hypotheses \mathcal{H}_i ($i = 0 : K$), defined by known densities $f_i(Y)$. Let the *a priori* distribution Q over the hypotheses be defined by $\mathbb{P}(\mathcal{H}_j) = q_j$, with $\sum_{j=0}^K q_j = 1$. Let $\pi(Y)$ be the vector of critical functions (Ferguson, 1967; Borovkov, 1987)

$$\pi(Y) = (\pi(j | Y))_{(j=0:K)}, \quad \sum_{j=0}^K \pi(j | Y) = 1, \quad \text{with}$$

$$\pi(j | Y) = \begin{cases} 0 & \text{if } q_j f_j(Y) \leq \max_{i \neq j} q_i f_i(Y) \\ 1 & \text{if } q_j f_j(Y) > \max_{i \neq j} q_i f_i(Y) \end{cases} \quad (29)$$

The Bayesian decision minimizes the risk function

$$R(Q, \pi) = 1 - \sum_{j=0}^K q_j \mathbb{E}_j(\pi(j | Y)).$$

When distribution Q is difficult to obtain, this is not a convenient criterion. Sometimes the invariance approach allows to circumvent this difficulty and to design a test optimal in some sense.

Slippage problem - Invariant tests. Assume that the observation is $Y = (y_1, \dots, y_r)$, where the scalar y_i 's are independent. Assume also that the simple null hypothesis is $\mathcal{H}_0 : \{y_1, \dots, y_r \sim P_0\}$ and r simple alternatives state that only one distribution has switched: $\mathcal{H}_j : \{y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_r \sim P_0, y_j \sim P_1\}$. This problem is invariant under the group of permutations of the y_i 's, or equivalently the \mathcal{H}_i 's ($i = 1 : r$). The *a priori* distribution invariant under that group is $q_0 = 1 - rq$ and $q_j = q$ for $j = 1 : r$. It can be shown (Ferguson, 1967) that the corresponding invariant Bayesian rule, optimal over the class \mathcal{K}_α (11), is:

$$\hat{\delta}(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \max_{i=1:r} \frac{f_1(y_i)}{f_0(y_i)} < h \\ \mathcal{H}_j & \text{if } j = \arg \max_{i=1:r} \frac{f_1(y_i)}{f_0(y_i)} \geq h \end{cases} \quad (30)$$

where $h = h(\alpha)$ and $\alpha = 1 - \mathbb{E}_0(\pi(0|Y))$. It maximizes the common power $\beta = \mathbb{P}_i(\delta = \mathcal{H}_i) = \mathbb{P}_j(\delta = \mathcal{H}_j), \quad \forall i, j \neq 0$.

Turning back to model (28), for testing $\mathcal{H}_0 : \{\theta_i = 0 \ (i = 1 : r)\}$ against r alternatives $\mathcal{H}_j : \{\theta_j = a, \theta_i = 0 \ (i = 1 : r, i \neq j)\}$ where $a > 0$, the Bayesian test (30) invariant under the group of translations $G = \{Y \mapsto g(Y) = Y + \mathbf{1}_r x\}$ writes

$$\widehat{\delta}(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \max_{i=1:r} (y_i - \bar{y}) < h \\ \mathcal{H}_j & \text{if } j = \arg \max_{i=1:r} (y_i - \bar{y}) \geq h \end{cases},$$

where $\bar{y} = \sum_{i=1}^r y_i / r$. It is worth noting that this test is independent of the value of a .

4. FROM HYPOTHESES TESTING TO FDI

How the hypotheses testing methods introduced above can be used for FDI is addressed now. First we explain to reduce FDI problems in dynamic systems to the static problem of monitoring the mean of a Gaussian vector. Then the application of the methods of section 3 to fault detection, fault isolation and fault diagnosis problems is investigated. Finally, the computation of the performance indices (error probabilities) is illustrated on a specific test instance.

4.1 Reducing FDI to a static problem

Reducing FDI problems in dynamic systems to the universal static problem of monitoring the mean value of a Gaussian vector can be achieved with the aid of a convenient residual generation. For additive faults (affecting the mean value of the measured outputs), a non-asymptotic reasoning suffices. For component faults (affecting the dynamics of the system), the asymptotic point of view introduced above under the name of local approach can be used to this end.

Additive faults. The argument for problem reduction, which applies to both linear and nonlinear systems, is based on an off-line point of view, assuming that measured observations Y and known inputs U are given over a finite time-window with size ρ . Assuming that the system is perturbed by Gaussian noises, the argument is simply to use repeated equations, as done for the design of a dead-beat observer.

Let an observable linear dynamic system be subject to faults acting additively as modeled in

$$\begin{cases} X_{k+1} = FX_k + GU_k + \Gamma \Upsilon_x + W_k \\ Y_k = HX_k + JU_k + \Xi \Upsilon_y + V_k \end{cases} \quad (31)$$

where $(W_k)_k$ and $(V_k)_k$ are two independent white noise sequences, with positive definite covariance matrices Q_x and Q_y respectively, Υ_x and Υ_y are the assumed additive faults, and the fault gains Γ and Ξ are full column rank matrices. Such a model is appropriate for sensors and actuators faults.

Let $\mathcal{Y}_{k,\rho}$ contain the stacked measured outputs $Y_k, \dots, Y_{k+\rho-1}$. Introducing the input-adjusted outputs $\underline{\mathcal{Y}}_{k,\rho} = \mathcal{Y}_{k,\rho} - \mathcal{M}_\rho(G, J) \mathcal{U}_{k,\rho}$ and the noise

$\underline{\mathcal{V}}_{k,\rho} = \mathcal{M}_\rho(I_n, 0) \mathcal{W}_{k,\rho} + \mathcal{V}_{k,\rho}$, where $\mathcal{M}_\rho(G, J)$ is the lower triangular block-Toeplitz matrix associated with the system impulse response, we can write ρ successive equations (31) as the *regression model* (Lou et al., 1986)

$$\zeta_{k,\rho}(\Upsilon) \stackrel{\text{def}}{=} \underline{\mathcal{Y}}_{k,\rho} = \mathcal{O}_\rho X_k + M\Upsilon + \underline{\mathcal{V}}_{k,\rho} \quad (32)$$

where \mathcal{O}_ρ is the observability matrix of order $(\rho - 1)$, $M \stackrel{\text{def}}{=} (\underline{\mathcal{M}}_\rho(\Gamma) (\mathbf{1}_\rho \otimes \Xi))$, $\Upsilon \stackrel{\text{def}}{=} \begin{pmatrix} \Upsilon_x \\ \Upsilon_y \end{pmatrix}$, and $\underline{\mathcal{M}}_\rho(\Gamma)$ is a matrix function which depends on Γ and on the system dynamics in a known fashion. Under convenient assumptions, regression model (32) has three basic properties. First, the (unknown) state vector X_k is *independent* of noise $\underline{\mathcal{V}}_{k,\rho}$. Second, the failure gain matrices $\underline{\mathcal{M}}_\rho(\Gamma)$ and $(\mathbf{1}_\rho \otimes \Xi)$ are f.c.r. Third, noise $\underline{\mathcal{V}}_{k,\rho}$ is Gaussian, with a covariance matrix Σ_ρ which is generally *not block-diagonal* (Basseville and Nikiforov, 1993; Basseville, 1997).

The FDI problem for additive faults (31) is thus reduced to a static problem for the Gaussian vector in (32): $\zeta_{k,\rho}(\Upsilon) \sim \mathcal{N}(\mathcal{O}_\rho X_k + M\Upsilon, \Sigma_\rho)$.

Component faults. The argument for reducing the FDI problem for component faults in *dynamic* systems to the problem of monitoring the mean of a Gaussian vector (14) is based on the local approach introduced at the end of 3.3. The idea is to build a primary residual from a parameter estimating function (e.g. the efficient score), and then to define an improved residual, of which the evaluation is much easier, since it is Gaussian distributed with the same covariance matrix in both safe and faulty cases as in (24). The details of this approach are outside the scope of the present paper. The interested reader is referred to (Basseville and Nikiforov, 1993; Basseville, 1998).

From now on we concentrate on the static Gaussian model (3), possibly reduced to (14), and on the parameter partition (25). The processed data are either the measurements themselves or some residuals. Note that, in all cases, the term $M\theta$ captures the signature of the fault on the processed data, and matrix M depends on the system (and possibly the residual generator) dynamics.

4.2 Fault detection

Model (3) can be re-written as:

$$Y = \tilde{H}\zeta + \xi, \quad \tilde{H} = (M H), \quad \zeta = \begin{pmatrix} \theta \\ X \end{pmatrix} \quad (33)$$

where the $r \times (m + p)$ -dimensional matrix \tilde{H} is assumed to be f.c.r. To simplify the formulas, from now on the covariance matrix of ξ is assumed to be: $\Sigma = \sigma^2 I$. The hypotheses are

$$\mathcal{H}_0 : \{\theta = 0\}, \quad \mathcal{H}_1 : \{\theta \neq 0\}$$

It should be obvious that the solution to this detection problem is a direct consequence of the solution to problem (26) for model (14) investigated in 3.4, where $M_\psi\psi$ stands for HX and $M_\phi\phi$ stands for $M\theta$.

Using the reparameterization approach to dealing with the nuisance X , we define $\eta_i(\zeta) = \zeta_i (i = 1 : m + p)$. The hypotheses can be re-written as:

$$\mathcal{H}_0 : \{\eta_i(\zeta) = \zeta_i = 0; i = 1 : m\}$$

$$\text{and } \mathcal{H}_1 : \{\eta_i(\zeta) = \zeta_i \neq 0; i = 1 : m\}.$$

The Jacobian matrix writes $J = I_{m+p}$ and matrix $J\mathcal{F}^{-1}J$ is equal to $\mathcal{F}^{-1} = \sigma^2(\tilde{H}^T\tilde{H})^{-1}$. Finally, the MLE of ζ is given by $\hat{\zeta} = (\tilde{H}^T\tilde{H})^{-1}\tilde{H}^TY$. Let \mathcal{F}_m be the $m \times m$ -dimensional upper-left block of matrix \mathcal{F}^{-1} and $\hat{\zeta}_m$ contain the first m components of $\hat{\zeta}$. It results from (22) that Wald's statistics writes:

$$\bar{\Lambda}(Y) = \eta(\hat{\zeta})^T \tilde{\mathcal{F}}_m^{-1}(\hat{\zeta}) \eta(\hat{\zeta}) = \hat{\zeta}_m^T \tilde{\mathcal{F}}_m^{-1} \hat{\zeta}_m. \quad (34)$$

The test based on this statistics has optimality properties analogous to the test based on (22). It is equivalent to the GLR test based on $\log \frac{\sup_{\theta, X} f_{HX+M\theta}(Y)}{\sup_X f_{HX}(Y)}$ and to the minimax test. Sometimes the equations $\eta(\theta) = 0$ result from some structural (geometric) properties of the system, energy or mass balance equations for instance. If the balance equations involve the only observations Y , some additional work should be done to carry the relation from the observations to the parameters θ and X . For linear systems, such equations are known under the name of analytical redundancy relations in the FDI literature (Potter and Suman, 1977; Chow and Willsky, 1984; Lou *et al.*, 1986; Patton *et al.*, 1989; Frank, 1990; Patton and Chen, 1991).

4.3 Fault isolation

As mentioned in 2.3, in case of two fault modes or more, isolation refers to deciding which fault mode occurred. Two basic approaches can be undertaken for this purpose. The first one consists in partitioning θ as in (25), with ϕ scalar, and deciding in favor of the fault mode ϕ , while considering the other fault modes collected in ψ as nuisance information. Model (3) writes as in (33) with $\tilde{H} \stackrel{\text{def}}{=} (M_\phi \ M_\psi \ H)$. The $r \times (m_\phi + m_\psi + p)$ -dimensional matrix \tilde{H} is assumed to be f.c.r. Fault isolation can thus be seen as fault detection in the presence of a nuisance parameter. The test to be used is (34) with $m = m_\phi$ and this \tilde{H} . The problem of detecting and isolating faults in the presence of disturbances modeled exactly as the faults can be handled in a similar manner.

The second approach is multiple hypotheses testing. In model (3), the $r \times 1$ -dimensional matrix M is now the fault direction and the scalar value θ the fault magnitude. Hence, the problem is to test $\mathcal{H}_0 : \{Y \sim \mathcal{N}(HX, \sigma^2 I_r); X \in \mathbb{R}^p\}$ against

$$\mathcal{H}_j : \{Y \sim \mathcal{N}(HX + M_j\theta_j, \sigma^2 I_r); X \in \mathbb{R}^p, \theta_j \neq 0\}$$

($j = 1 : K$). Unfortunately, this hypotheses testing problem is not invariant under the group of permutations of the \mathcal{H}_j 's. For this reason, in the absence of *a priori* information on the alternatives, we assume equal prior probabilities: $\mathbb{P}(\mathcal{H}_j) \stackrel{\text{def}}{=} q_j = (1 - q_0)/K$, where q_0 is the prior on the null hypothesis. The corresponding Bayesian test is not optimal over the class \mathcal{K}_α (11). It results from (29) that this test is:

$$\hat{\delta}(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \max_{i=1:K} \frac{f_{HX+M_i\theta_i}(Y)}{f_{HX}(Y)} < h \\ \mathcal{H}_j & \text{if } j = \arg \max_{i=1:K} \frac{f_{HX+M_i\theta_i}(Y)}{f_{HX}(Y)} \geq h \end{cases} \quad (35)$$

where $f_\theta(Y)$ is the density of the Gaussian distribution $\mathcal{N}(\theta, \sigma^2 I_r)$ and $h = h(q_0)$. There are two unknown parameters in (35): the nuisance X and the fault magnitude θ . To reject the effect of the nuisance X , the test statistics should remain invariant under the group of translations $G = \{Y \mapsto g(Y) = Y + HX\}$. Let W the $(r - p) \times r$ matrix which rows span the left null space of matrix H , namely such that $WH = 0, WW^T = I_{r-p}, W^T W = I_m - H(H^T H)^{-1}H^T$. Vector $\varepsilon \stackrel{\text{def}}{=} WY$ is known as the *parity vector* in the analytical redundancy approach (Potter and Suman, 1977). From (3), we get the model:

$$\varepsilon \stackrel{\text{def}}{=} WY = WM\theta + W\xi$$

that remains invariant under the group of translations G . Of course, it is assumed that $r - p > 1$ and that the cross products $(WM_j) \times (WM_l)$ are nonzero for every $1 \leq j \neq l \leq K$. Next, to cope with the unknown fault magnitude θ_j , the MLE can be plugged into (35), which results in:

$$\hat{\delta}(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \max_{i=1:K} \frac{f_{WM_i\hat{\theta}_i}(\varepsilon)}{f_0(\varepsilon)} < h \\ \mathcal{H}_j & \text{if } j = \arg \max_{i=1:K} \frac{f_{WM_i\hat{\theta}_i}(\varepsilon)}{f_0(\varepsilon)} \geq h \end{cases} \quad (36)$$

where $f_\theta(\varepsilon)$ is the density of $\mathcal{N}(\theta, \sigma^2 I_{r-p})$ and

$$\hat{\theta}_j = \arg \min \|\varepsilon - WM_j\theta_j\|^2 = \frac{M_j^T W^T \varepsilon}{M_j^T W^T WM_j}.$$

Decision rule (36) has a simple geometric interpretation. Assume that the unit vectors $V_j = WM_j / \sqrt{M_j^T W^T WM_j}$ define K different directions in \mathbb{R}^{r-p} . Then, the inner product $V_j \cdot \varepsilon$ represents the projection of the parity vector ε along V_j . Finally, for testing the alternatives $\mathcal{H}_1, \dots, \mathcal{H}_K$ against the null \mathcal{H}_0 , we have to compute the differences $\varepsilon - (V_j \cdot \varepsilon) V_j$ and to choose the index j that minimizes the norm $\|\varepsilon - (V_j \cdot \varepsilon) V_j\|$. This is equivalent to minimize the angle between ε and V_j in \mathbb{R}^{r-p} . If the ratio $f_{WM_j\hat{\theta}_j}(\varepsilon)/f_0(\varepsilon)$ is greater than h , then the hypothesis \mathcal{H}_j is chosen.

4.4 Fault diagnosis

At this point, some comments are in order on different types of faults.

A *single fault* affects a single component of the fault vector θ or several components of this vector in a specific direction or subspace. When a single fault at a time is assumed, for isolating - or diagnosing - the faulty component, the multiple hypotheses testing approach described above is to be preferred: precise criteria (performance indexes) can be defined and algorithms exist.

Two basic types of *multiple faults* can be distinguished : embedded multiple faults with causality constraints, which we do not address here, and independent multiple faults. The latter type assumes no causality between the faults on the individual components, which can occur simultaneously in any manner. In this case, and assuming that the number of failed components is known, running simultaneously several isolation tests can make up a diagnosis procedure for simultaneous faults. Actually, viewing all but one faults as nuisance parameters and exploiting possible invariance properties is a relevant approach, for which some optimality property has been established (Spjotvoll, 1972; Basseville, 1997). Roughly speaking, the set of all ‘individual’ tests maximizes both the minimum and the sum of the individual powers. Running these tests in parallel has proven useful in practice (Basseville *et al.*, 1994).

Inferring the (generally unknown) number of failed components in θ is beyond the scope of this paper. The interested reader is referred to (Basseville *et al.*, 1994) for a discussion of and a possible solution to this problem.

4.5 Statistical properties

For computing the performance indices, we have to define the distributions of the test statistics under the hypotheses \mathcal{H}_j , $j = 0, \dots, K$. Wald’s statistics (19), which should be used for fault detection in the absence of nuisance parameter, is χ^2 -distributed with m degrees of freedom. This χ^2 distribution is central under \mathcal{H}_0 and noncentral under \mathcal{H}_1 . Hence, we get

$$\alpha_0 = \mathbb{P}_0(\delta \neq \mathcal{H}_0) = \mathbb{P}_0(\bar{\Lambda}(Y) \geq h) = \int_h^\infty p_0(x) dx \quad (37)$$

where $p_0(x) = \frac{x^{m/2-1} e^{-x/2}}{2^{m/2} \Gamma(m/2)}$ is the density of the central $\chi^2(m)$ and $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ is the gamma function, and

$$\alpha_1 = \mathbb{P}_1(\delta \neq \mathcal{H}_1) = \mathbb{P}_\theta(\bar{\Lambda}(Y) < h) = \int_0^h p_\lambda(x) dx \quad (38)$$

where $p_\lambda(y) = p_0(y) e^{-\frac{\lambda}{2}} G\left(\frac{n}{2}, \frac{\lambda y}{4}\right)$, and G is the hypergeometric function $G(a, y) = \sum_{i=0}^\infty \frac{\Gamma(a)y^i}{\Gamma(a+i)i!}$.

The error probabilities in (37) and (38) are also valid for fault detection in the presence of a nuisance parameter and for the first approach to fault isolation. In both cases, Wald’s statistics (22) is χ^2 -distributed with m (respectively m_ϕ) degrees of freedom. It is a difficult problem to compute the error probabilities in case of multiple hypotheses (36). Some asymptotic results and particular cases can be found in (Nikiforov, 1997; Lai, 2000).

5. EXAMPLE

How to apply some of the above methods to an integrity monitoring problem is now discussed. Integrity monitoring, a major issue for the Global Positioning System (GPS) in many safety-critical applications, requires that a navigation system *detects*, *isolates* faulty measurement sources, and *removes* them from the navigation solution before they significantly contaminate the output.

The GPS is based on accurate measuring of the distance (*range*) from several satellites with known locations to a user (vehicle). Assume r satellites located at known positions $X_j = (x_j, y_j, z_j)^T$ ($j = 1 : r$) and a user at $X_u = (x_u, y_u, z_u)^T$. The distance from the j -th satellite to the user is defined as

$$d_j = \|X_j - X_u\|,$$

and the *pseudorange* r_j from the j -th satellite to the user writes:

$$r_j = d_j + cs + \xi_j,$$

where s is a user clock bias, c is the light speed and ξ_j is an additive pseudorange error at the user position. The pseudorange r_j is thus a function of X_u and s . Let $X = (X_u^T, s)^T$ be the state, and $R = (r_1, \dots, r_n)^T$ be the measurement. Linearizing the function $R(X)$ around the working point $X_0 = (X_{u_0}^T, s_0)^T$, we get the measurement equation

$$Y \stackrel{\text{def}}{=} R - R_0 \simeq H_0(X - X_0) + \xi,$$

where $R_0 = (r_{1_0}, \dots, r_{n_0})^T$, with

$$r_{j_0} = \|X_j - X_{u_0}\| + cs_0,$$

$\xi = (\xi_1, \dots, \xi_n)^T$, and where

$$H_0 = \left. \frac{\partial R}{\partial X} \right|_{X=X_0}$$

is the $(n \times 4)$ -dimensional Jacobian matrix. Hence the GPS model with a fault is given by:

$$Y = H_0(X - X_0) + M_j \theta_j + \xi, \quad \xi \sim \mathcal{N}(0, \sigma^2 I_r)$$

where θ_j is an additional bias (fault) on the pseudorange r_j and where the fault direction is $M_j = (0, \dots, 0, 1, 0, \dots, 0)^T$ ($j = 1 : r$). Assume that only one measurement vector Y is available to the observer, situation often called snapshot. In this case, the problem of fault detection/isolation is solved by using the decision rule (36) with $K = r$. If the integrity

monitoring problem is to be solved in real time, the performances of the decision algorithm can be crucially improved using statistical sequential methods. The interested reader is referred to (Nikiforov, 2002).

6. DISCUSSION

We now discuss some methodological features of the methods presented above.

A priori information. The first step in designing a FDI algorithm is the derivation of the model and the integration of the available *prior information*. The model design includes the definition of the model's structure and the partition of the parameter vector θ into the informative parameter ϕ and the nuisance ψ . The integration of *a priori* information on the hypotheses \mathcal{H}_i includes,

- For Bayesian approaches: the *a priori* probabilities $\mathbb{P}(\mathcal{H}_i)$ and the probability distributions of the parameters θ_i . The existence of the $\mathbb{P}(\mathcal{H}_i)$ is especially important in case of multiple faults, as shown in section 3;
- For non Bayesian approaches: the parametric domains Θ_i , typically given by points, curves or ellipsoids in \mathbb{R}^m .

The prior information on the nuisance parameter vector ψ is usually its dimension m_ψ .

Optimality criteria. The second design step is the choice of an optimality criterion. This should achieve a tradeoff between the practical needs and the existing theoretical results. Here the model structure plays a crucial role: the geometric properties of the monitored system (energy or mass balance equations and other deterministic - static or dynamic - properties) not only affect the term HX in (3), but also the natural parameter (the mean) of the processed data distribution through the term $M\theta$, signature of the fault.

The possible criteria usually result from the prior information on the hypotheses. For instance, in case of binary decision scheme, if the alternative hypothesis \mathcal{H}_1 is defined as the outside of an ellipsoid, then the power function is defined over a family of closed ellipsoidal surfaces.

Theoretical tools. Some comments are in order on the design approaches presented in section 3. The minimax approach can often be used, because it is not very demanding: there is no constraint on the shape of the sets Θ_i , and this method can thus be applied to deal with nuisance parameters. But, finding the least favorable values ψ_0^* and ψ_1^* might not be that easy. In general (non linear) cases, it might not be sufficient for the resulting test be optimal in the class \mathcal{K}_α either.

The constant power and the most stringent test approaches are much more demanding. First, the alternative hypothesis \mathcal{H}_1 should be given as the outside of an ellipsoid defined by Fisher information matrix. In the Gaussian or asymptotic cases, the three methods (minimax, constant power and most stringent test) lead to the same (Wald's) statistics. Second, the invariance theory and reparameterization approaches to deal with nuisance parameters in this context have the following properties. The invariance theory does not warrant any optimal property of the resulting (invariant) test; but it can be used to reduce the initial problem to another one, usually simpler or having a known solution. Instead, the reparameterization approach warrants an optimality in the sense given in 3.2. But this approach should be supplied with the definition of a composite null hypothesis (21), namely with a vector-valued function $\theta \mapsto \eta(\theta)$. The role of the invariance theory is to investigate the invariant properties of the hypotheses testing problem under a group of transformations and to prepare some variants of the above function $\eta(\theta)$. This investigation should be finalized by using the reparameterization approach to design an optimal decision rule.

It is worth to note that the two approaches, (statistical) invariance theory and (engineering) analytical redundancy, are actually very close to each other. Both approaches exploit the model structure at a preliminary stage in order to define the function $\eta(\theta)$, where θ contains e.g. the additive faults and the nuisance state.

Conclusion. The FDI problem has been addressed from a statistical point of view, with faults modeled as deviations in the parameter vector of a stochastic system. Fault detection, isolation and diagnosis have been stated as hypotheses testing problems. Several major statistical tools for solving these testing problems have been introduced. Particular emphasis has been put on nuisance rejection and deciding between multiple hypotheses. The application to GPS integrity monitoring has been described. The advantages and drawbacks of the different methods have been discussed.

7. REFERENCES

- Basawa, I.V. (1985). Neyman-Le Cam tests based on estimating functions. In: *Proc. Berkeley Conf. in Honor of Neyman and Kiefer*. Vol. 2. pp. 811–825.
- Basseville, M. (1997). Information criteria for residual generation and fault detection and isolation. *Automatica* **33**(5), 783–803.
- Basseville, M. (1998). On-board component fault detection and isolation using the statistical local approach. *Autom* **34**(11), 1391–1416.
- Basseville, M., A. Benveniste, G. Mathis and Q. Zhang (1994). Monitoring the combustion set of a gas turbine. In: *Proc. Safeprocess'94, Helsinki, FI*.

- Basseville, M. and I.V. Nikiforov (1993). *Detection of Abrupt Changes: Theory and Application*. Prent.Hall, www.irisa.fr/sigma2/kniga.
- Benveniste, A., M. Basseville and G. Moustakides (1987). The asymptotic local approach to change detection and model validation. *IEEE Trans. on Automatic Control* **32**, 583–592.
- Bernshtein, A.V. (1980). On verifying composite hypotheses with nuisance parameters in the multivariate case. *Theory of Probability and Its Applications* **25**(2), 287–298.
- Borovkov, A. (1987). *Statistique Mathématique*. Mir Editions, Moscow. (In French).
- Burgess, K.A. and B.D. Van Veen (1996). Subspace-based adaptive generalized likelihood ratio detection. *IEEE Trans. Sign. Proc.* **44**(4), 912–927.
- Chow, E.Y. and A.S. Willsky (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Trans. Aut. Control* **29**(7), 603–614.
- Ferguson, T.S. (1967). *Mathematical Statistics. A Decision Theoretic Approach*. Academ. Press.
- Frank, P.M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge based redundancy - a survey and some new results. *Automatica* **26**, 459–474.
- Hall, W.J. and D.J. Mathiason (1990). On large sample estimation and testing in parametric models. *Int. Stat. Review* **58**(1), 77–97.
- King, M.L. (1996). Hypothesis testing in the presence of nuisance parameters. *Jal Stat. Planning and Inference* **50**(1), 103–120.
- Lai, T.L. (2000). Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Trans. Info. Theory* **46**(2), 595–608.
- Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Series in Statistics, Springer, New York.
- Le Cam, L. and G. L. Yang (1990). *Asymptotics in Statistics. Some Basic Concepts*. Series in Statistics, Springer, New York.
- Lehmann, E.L. (1986). *Testing Statistical Hypotheses*. Chapman and Hall.
- Lou, X.-C., A.S. Willsky and G.C. Verghese (1986). Optimally robust redundancy relations for failure detection in uncertain systems. *Automatica* **22**(3), 333–344.
- Nikiforov, I.V. (1997). Two strategies in the problem of change detection and isolation. *IEEE Trans. Info. Theory* **43**(2), 770–776.
- Nikiforov, I.V. (2002). Optimal sequential change detection and isolation. In: *These Proc.*
- Patton, R.J. and J. Chen (1991). A review of parity space approaches to fault diagnosis. In: *Proc.Safeprocess'91, Baden-Baden, FRG*.
- Patton, R.J., Frank, P.M. and Clark, R.N., Eds.) (1989). *Fault Diagnosis in Dynamic Systems, Theory and Application*. Prentice Hall.
- Porat, B. and B. Friedlander (1993). On the generalized likelihood ratio test for a class of nonlinear detection problems. *IEEE Trans. Sign. Proc.* **41**(11), 3186–3190.
- Potter, J. E. and M. C. Suman (1977). Thresholdless redundancy management with arrays of skewed instruments. In: *Integrity in Electronic Flight Control Systems, Agardograph-224*. pp. 15.11–15.25. NATO.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. John Wiley & Sons.
- Roussas, G. G. (1972). *Contiguity of Probability Measures, Some Applications in Statistics*. Cambridge University Press, Mass.
- Scharf, L.L. (1991). *Statistical Signal Processing*. Addison-Wesley. Reading, MA.
- Scharf, L.L. and B. Friedlander (1994). Matched subspace detectors. *IEEE Trans. Sign. Proc.* **42**(8), 2146–2157.
- Spjotvoll, E. (1972). On the optimality of some multiple comparison procedures. *Annals Statistics* **21**(3), 1486–1521.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54**, 426–482.
- Zacks, S. (1971). *The Theory of Statistical Inference*. Wiley.