Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Handling nuisance parameters in systems monitoring

Michèle Basseville and Igor Nikiforov

Abstract— **Dealing with nuisance parameters is an important issue in monitoring safety-critical complex systems and detecting events that affect their functioning. Several tools for solving statistical inference problems in the presence of nuisance parameters are described. The application of these tools to (off-line) hypotheses testing and (on-line) change detection is discussed. The usefulness of some of the proposed methods is illustrated on a couple of monitoring problems.**

I. Introduction

Monitoring complex structures and processes is necessary for fatigue prevention, aided control and condition-based maintenance. We have argued [1], [2] that *i)* faults can often be modeled as deviations, w.r.t. a nominal reference value, in the parameter vector of a stochastic system; *ii)* mathematical statistics theories and tools for solving hypotheses testing and change detection problems are relevant for addressing monitoring problems; *iii)* key features of these methods are their ability to handle noises and uncertainties, to select one among several hypotheses, to reject nuisance parameters.

Handling the presence of nuisance parameters is indeed an important issue in this framework. Distinguishing two subsets of components of the parameter vector, the parameters of interest and the nuisance parameters, may be necessary for at least two reasons. First, some parameters of no interest for monitoring, if not of no physical meaning, may appear in the model for e.g. model flexibility or specification, or for data interpretation reasons. Second, the fault isolation problem (deciding which fault mode occurred) can be approached as deciding in favor of one fault mode while considering the other fault modes as nuisance information.

The purpose of this paper is to investigate the nuisance parameters issue in monitoring and fault detection, isolation and diagnosis (FDI) problems. A particular emphasis is put on the invariant and generalized likelihood ratio (GLR) approaches to detection in the presence of nuisance parameters. Moreover, whereas the discussion in [2] has concentrated on (off-line) hypotheses testing, in the present paper the (on-line) change detection problem is addressed.

The paper is organized as follows. A basic model and the two off-line and on-line inference problems are introduced in section II, together with key criteria for evaluating decision algorithms. Section III is devoted to the handling of nuisance parameters : classical tools are briefly reviewed and three statistical approaches – invariant tests, GLR tests and minimax tests – are discussed. How to use these three approaches for change detection and hypotheses testing is explained in section IV. Hypotheses testing is addressed in IV-B. Change detection is the subject of IV-C, where on-line detection and isolation of changes occurring within a data sample is investigated. Some examples of FDI problems with nuisance parameters are described in section V. Some conclusions are drawn in section VI.

M. Basseville is with IRISA, UMR CNRS 6074, Campus de Beaulieu, 35042 Rennes Cedex, F. basseville@irisa.fr

I. Nikiforov is with ISTIT, FRE CNRS 2732, 12 rue Marie Curie, BP 2060, 10010 Troyes Cedex, F. nikiforov@utt.fr

II. Statistical tests for monitoring

As argued in [2], two situations are relevant to monitoring : *i)* hypotheses testing, namely deciding between two (or more) hypotheses, for detecting (or isolating) faults; *ii)* detection and isolation of changes soon after their onset time. These two situations are now described.

A. Hypotheses testing

1) Problem statement: The parameterized distribution of the observations is noted as $(Y_1, \ldots, Y_N) = \mathcal{Y} \sim \mathcal{P}_\theta$. The parameter vector θ is partitioned as:

$$\theta^T = (\phi^T,\ \psi^T),\ \phi \in \mathbb{R}^m,\ \psi \in \mathbb{R}^q \qquad (1)$$

where ϕ (resp. ψ) is the informative (resp. nuisance) parameter. Here, the informative parameter vector ϕ is assumed to be *constant* within the entire data sample $\mathcal{Y} = \{Y_1, \ldots, Y_N\}$.

The hypotheses testing problem consists in deciding which family of distributions $\mathcal{P}_i = \{\mathcal{P}_\theta,\ \phi \in \Phi_i,\ \psi \in \mathbb{R}^q\}$ is the true one. The *null hypothesis* $\mathcal{H}_0$ corresponds to the fault free case, e.g. a nominal parameter ϕ within a set Φ_0:

$$\mathcal{H}_0 \ :\quad \phi \in \Phi_0 \subset \mathbb{R}^m, \psi \in \mathbb{R}^q. \qquad (2)$$

The *alternative hypotheses* stand for different fault modes:

$$\mathcal{H}_i \ :\quad \phi \in \Phi_i \subset \mathbb{R}^m, \psi \in \mathbb{R}^q \ (i = 1, \ldots, K), \qquad (3)$$

where $\Phi_i \bigcap \Phi_j = \emptyset$ 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.

A *statistical test* for testing between the $\mathcal{H}_i$'s is any measurable mapping $\delta : (\mathcal{Y}) \to \{\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_K\}$ from the observation space onto the set of hypotheses. The quality of a statistical test is defined with a set of *error probabilities*: $\alpha_i(\phi, \psi) = \mathbb{P}_i\,(\delta \neq \mathcal{H}_i)\,,\ \ i = 0, \ldots, K$, where $\mathbb{P}_i$ stands for observations $Y_1, \ldots, Y_n$ being generated by distribution $\mathcal{P}_i$. The *test power* is defined with a set of probabilities of correct decisions: $\beta_i(\phi, \psi) = \mathbb{P}_i\,(\delta = \mathcal{H}_i)\,,\ i = 1, \ldots, K$ in the class of tests with upper-bounded maximum false alarm probability $\mathcal{K}_\alpha = \big\{\delta : \sup_{\phi \in \Phi_0,\ \psi} \mathbb{P}_0(\delta \neq \mathcal{H}_0) \leq \alpha\big\}$.

2) Roles of informative and nuisance parameters: As made explicit in the two equations above, the performance indexes of statistical tests are functions of both the informative parameters ϕ and nuisance parameters ψ. The desirable relations between the error probabilities or the power of a test and the informative vector ϕ usually result from the application. Sometimes, the statistician must define some additional constraints (possibly artificial w.r.t. the application) resulting from the statistical nature of the problem, in order to achieve optimal properties of the test. For example, a family of surfaces parameterized by Kullback-Leibler (KL) distance between the densities corresponding to $\mathcal{H}_0$ and $\mathcal{H}_i$ is assumed and a constant power of the test over such a family is imposed. This results in a test power which is an increasing function of KL distance.

The main difference between ϕ and ψ is the following. In contrast to the informative parameter, the nuisance parameter ψ has no desirable impact on the performance indexes. When designing a test for deciding between hypotheses in the presence of a nuisance parameter, the goal is to achieve performance indexes independent from the actual value of ψ.

3) UBCP tests: For introducing the uniformly best constant power (UBCP) test, the nuisance-free case is considered, namely θ is entirely composed of informative parameters: $\theta = \phi$. The first solution to the composite hypotheses testing problem for a vector parameter traces back to [3], and summarizes as follows for a Gaussian mean. The observation Y is generated by a Gaussian distribution $\mathcal{N}(\theta, \Sigma)$ with mean θ and positive definite covariance matrix Σ, and the problem consists in deciding between $\mathcal{H}_0 : \{\theta = 0\}$ and $\mathcal{H}_1 : \{\theta \neq 0\}$. The UBCP test defined on the family of surfaces (ellipsoids):

$$\mathcal{S} = \{S_c : \theta^T \Sigma^{-1} \theta = c^2,\ c > 0\} \tag{4}$$

is given by [3]:

$$\delta^*(Y) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \Lambda(Y) = Y^T \Sigma^{-1} Y < h(\alpha_0) \\ \mathcal{H}_1 & \text{if} \quad \Lambda(Y) = Y^T \Sigma^{-1} Y \geq h(\alpha_0) \end{cases}, \tag{5}$$

where $h(\alpha_0)$ is tuned from the false alarm probability α_0.

When the number of observations is large ($N \to \infty$), an asymptotic approach can be used for deciding between $\mathcal{H}_0 : \{\theta = \theta_0\}$ and $\mathcal{H}_1 : \{\theta \neq \theta_0\}$. In the general case $Y \sim \mathcal{P}_\theta$, the relevant family of surfaces then writes:

$$\mathcal{S} = \{S_c :\ (\theta - \theta_0)^T\ \mathcal{F}(\theta_0)\ (\theta - \theta_0) = c^2,\ c > 0\} \tag{6}$$

$$\text{where } \mathcal{F}(\theta) = \mathbb{E}_\theta \left(\partial \log f_\theta(Y)/\partial\theta\right) \cdot \left(\partial \log f_\theta(Y)/\partial\theta\right)^T \tag{7}$$

is Fisher information matrix. Defining asymptotic optimality is somewhat complex [3], and always involves a sequence of tests $\{\delta_N\}$. It can be shown [3–6] that the test based on:

$$\overline{\Lambda}(Y_1, \ldots, Y_N) = N\ (\widehat{\theta}_N - \theta_0)^T\ \mathcal{F}(\theta_0)\ (\widehat{\theta}_N - \theta_0), \tag{8}$$

where $\widehat{\theta}_N$ is the maximum likelihood estimate (MLE) of θ, is asymptotically UBCP over the family of surfaces $\mathcal{S}$.

The statistics $\Lambda(Y)$ in (5) obeys a χ^2 distribution with m degrees of freedom, central under $\mathcal{H}_0$, with noncentrality parameter c^2 under $\mathcal{H}_1$, and the power function:

$$\beta_{\delta^*}(c^2) = \mathbb{P}_{c^2}\left(\Lambda(Y) \geq h(\alpha)\right) \tag{9}$$

is constant over the surface S_c [7, Ch.2.7]. Under some regularity conditions, $\overline{\Lambda}(Y_1, \ldots, Y_N)$ in (8) is also asymptotically χ^2-distributed with m degrees of freedom, central under $\mathcal{H}_0$ and noncentral under $\mathcal{H}_1$. Finally, since the definition of the family (4) (resp. (6)) involves the KL distance $\rho(\theta, \theta_0) \triangleq 1/2\ (\theta - \theta_0)^T\ \mathcal{F}(\theta_0)\ (\theta - \theta_0)$ (resp. its second order approximation), the (asymptotic) power β_{δ^*} of the (asymptotically) UBCP test is a function of the KL distance.

B. Change detection/isolation

For introducing the second situation, the nuisance-free case $\theta_0 = \phi_0$ is considered as above. The sequence of observations $\mathcal{Y} = \{Y_1, \ldots, Y_N\}$ is generated by the distribution $\mathcal{P}_{\theta_0}$. Until time $k_0 - 1$, the parameter is $\theta(k) = \theta_0$ and, from k_0 onwards, it becomes $\theta(k) = \theta_l$ for some l, $1 \leq l \leq K$. The fault onset time k_0 and fault index l are assumed unknown and non random[1]. The problem is to *detect* and *isolate* the change in θ (in other words, to determine the fault type index l) as soon as possible.

The change detection/isolation algorithm should compute a *pair* (N, ν) based on the observations $(Y_k)_{k \geq 1}$, where ν, $1 \leq \nu \leq K$, is the *final decision* and N is the *alarm time* at which a ν-type change is detected. There are several different criteria to evaluate a change detection/isolation algorithm.

1) Worst case conditional detection/isolation delay: For $k_0 = 1, 2, \cdots$, let $\mathcal{P}^l_{k_0}$ be the distribution of the observations $Y_1, Y_2, \cdots, Y_{k_0}, Y_{k_0+1}, \cdots$ when Y_{k_0} is the first observation with distribution $\mathcal{P}_l$, and let $\mathbb{E}^l_{k_0}$ (resp. $\mathbb{E}_0$) be the expectation w.r.t. the distribution $\mathcal{P}^l_{k_0}$ (resp. $\mathcal{P}_0 = \mathcal{P}^0_\infty$). Following [10], the worst case conditional detection/isolation delay[2] :

$$\overline{\mathbb{E}}^*(N) \triangleq \sup_{k_0 \geq 1, 1 \leq l \leq K} \operatorname{esssup} \mathbb{E}^l_{k_0} \left((N - k_0 + 1)^+ | \mathcal{Y}_1^{k_0 - 1} \right) \tag{10}$$

where $x^+ = \max(0, x)$, is required [11] to be *as small as possible* for a given minimum γ of the mean times before false alarm or false isolation:

$$\mathbb{E}_0 \left(\inf_{r \geq 1} \{N_r : \nu_r = j\} \right) \geq \gamma, \mathbb{E}^l_1 \left(\inf_{r \geq 1} \{N_r : \nu_r = j\} \right) \geq \gamma \tag{11}$$

for $1 \leq l, j \neq l \leq K$. For a more tractable performance index, the isolation constraint in (11) has been replaced [12] by the probability of false isolation:

$$\mathbb{P}^l_1(\nu = j \neq l) \leq \beta \sim \gamma^{-1} \text{ as } \gamma \to \infty. \tag{12}$$

An asymptotic lower bound for the worst case delay (10)-(12), which extends the result in [10], is [9], [11]:

$$\overline{\mathbb{E}}^*(N; \gamma)\ \gtrsim\ \log \frac{\gamma}{\rho^*}\ \text{ as }\ \gamma \to \infty, \tag{13}$$

[1]The results of a Bayesian approach can be found in [8], [9].

[2]We say that $y = \operatorname{esssup} x$ if : *i)* $\mathbf{P}(x \leq y) = 1$; *ii)* if $\mathbf{P}(x \leq y') = 1$ then $\mathbf{P}(y \leq y') = 1$, where y, y', x are three random variables and $\mathbf{P}(A)$ is the probability of the event A.

$$\text{where}\quad \rho^* \triangleq \min_{1\le l\le K} \min_{0\le j\ne l\le K} \rho(\theta_l,\theta_j)$$

$$\text{and}\quad 0<\rho(\theta_l,\theta_j) \triangleq \mathbb{E}_1^l\left(\log\frac{f_{\theta_l}(Y_i)}{f_{\theta_j}(Y_i)}\right)<\infty$$

is the KL information, which definition, in the general case of dependent observations, is more complicated [9].

2) Uniformly constrained conditional probability of false isolation: The drawback of criterion (10)-(12) lies in that the probability of false isolation is constrained only if the change time is $k_0 = 1$. A more tractable criterion [13], [14] consists in minimizing the maximum mean delay for detection/isolation:

$$\overline{\mathbb{E}}(N) \triangleq \sup_{k_0\ge 1, 1\le l\le K} \mathbb{E}_{k_0}^l(N-k_0+1 \mid N\ge k_0) \tag{14}$$

subject to the constraints:

$$\mathbb{E}_0(N)\ge\gamma,\ \sup_{k_0\ge 1} \mathbb{P}_{k_0}^l(\nu=j\ne l|N\ge k_0)\le\beta, \tag{15}$$

for $1\le l, j\ne l\le K$. An asymptotic lower bound $n(\gamma,\beta)$ for the maximum mean delay (14)-(15) is given by [14]:

$$\overline{\mathbb{E}}(N;\gamma,\beta) \gtrsim \max\left\{\log\frac{\gamma}{\rho_{\rm d}^*},\ \log\frac{\beta^{-1}}{\rho_{\rm i}^*}\right\} \tag{16}$$

as $\min\{\gamma,\beta^{-1}\}\to\infty$, where $\rho_{\rm d}^* = \min_{1\le j\le K}\rho(\theta_j,\theta_0)$ and $\rho_{\rm i}^* = \min_{1\le l\le K}\min_{1\le j\ne l\le K}\rho(\theta_l,\theta_j)$.

3) Uniformly constrained probabilities of false alarm and false isolation within a time window: For some safety-critical applications, it is necessary to warrant that the false alarm and false isolation probabilities within a time window with size m_α are lower than a prescribed upper bound. Since the constraint $\mathbb{E}_0(N)\ge\gamma$ does not necessarily imply that the probability of having a false alarm before some specified time instant is small, it is proposed in [9] to minimize the mean delay for detection/isolation for every $1\le l\le K$:

$$\mathbb{E}_{k_0}^l(N-k_0+1)^+ \tag{17}$$

subject to the following constraints :

$$\sup_{k\ge 1} \mathbb{P}_0(k\le N<k+m_\alpha)\le\alpha\ m_\alpha \tag{18}$$

$$\sup_{k_0\ge 1} \mathbb{P}_{k_0}^l(k_0\le N<k_0+m_\alpha\cap\nu\ne l)\le\alpha\ m_\alpha \tag{19}$$

on the false alarm and false isolation probabilities. For every $1\le l\le K$, an asymptotic lower bound for (17) under (18)-(19) which holds uniformly in k_0 when $\alpha\to 0$ is [9] :

$$\mathbb{E}_{k_0}^l\ (N-k_0+1)^+\ \ge\ \frac{\mathbb{P}_0(N\ge k_0)\ |\log\alpha|}{\rho_l+o(1)} \tag{20}$$

where $\rho_l \triangleq \min_{j\ne l}\ \rho(\theta_l,\theta_j)$.

Note that, because of (13), (16) and (20), the performance indexes of the change detection/isolation algorithms discussed in this section are functions of KL distances, as it is the case for the hypotheses testing algorithms in II-A.

III. Dealing with nuisance parameters

After a brief review of available methods for handling nuisance parameters, we discuss two statistical approaches to monitoring in the presence of nuisance parameters : invariant tests, GLR and minimax tests. The emphasis here is put on the GLR test, both for its connection with the invariant one, and for its usefulness in the bounded and non-linear cases. Other approaches to testing are described in [2].

A. Nuisance parameter elimination

Eliminating nuisance parameters is a long-standing and major issue in statistical inference [15]. So many methods have been proposed so far that it is impossible to describe all of them in this section. Instead, we overview some key ideas for reducing or eliminating the effect of nuisance parameters.

From a Bayesian point of view, computing the marginal posterior distribution of the parameter of interest should help eliminating a nuisance parameter [16], but in practice the situation might be much less simple [17].

In the likelihood approach, the problem is to find a likelihood function for the parameter of interest only [18]. In special cases, it may happen that the marginal distribution of some components of the observation do not depend on the nuisance parameter ψ. Another idea is to base the inference for ϕ on a conditional distribution of the observations given a sufficient statistics for ψ for fixed ϕ [19]. Such distributions, called *marginal likelihood* and *conditional likelihood* respectively, can then be used as pseudo-likelihoods for ϕ. Such approaches should be used with care [17] (a marginal and a conditional distribution may provide different results [15]). The use of conditional score functions is supported by optimality results, whereas marginal score functions are not [19]. A detailed discussion of modern developments can be found in [20][Chap.8]. How the nuisance information is parameterized may be of key importance for that.

A property of particular interest is *parameter orthogonality* : the parameter of interest ϕ is said to be orthogonal (w.r.t. Fisher information) to the nuisance parameter ψ if their corresponding score functions are uncorrelated, or equivalently when the block off-diagonal terms $\mathcal{F}_{\phi\psi}$ of Fisher matrix is zero [21], [22]. Local (resp. global) orthogonality stands for these conditions holding for one (resp. all) parameter values. In case of orthogonality, the MLE's of both parameters ϕ and ψ are asymptotically independent. When ϕ is scalar, a transformation to orthogonal parameterization can always be found [22]. Parameter orthogonality is a special case of estimating functions orthogonality discussed in [19]. Taking advantage of an invariance property of the probability distribution under some transformations is known under the name of *invariant approach* [23] and further investigated in III-B.

The most general likelihood-based approach, further developed in III-C for testing, consists in maximizing the likelihood over the nuisance parameter. The *profile likelihood*, also called *concentrated* or *peak likelihood*, is $\mathcal{L}_p(\phi)\triangleq\mathcal{L}(\phi,\hat{\psi}_\phi)$ where $\hat{\psi}_\phi$ is the MLE of ψ for fixed ϕ [24]. The score function corresponding to the profile likelihood is no longer zero-mean. When the number of nuisance parameters is small w.r.t. the sample size, this bias is often negligible. This no longer holds in case of many nuisance parameters. Corrections to the profile likelihood have been proposed [22].

B. Invariant tests

When testing, as in (2) - (3), in the presence of a nuisance parameter $\psi \in \mathbb{R}^q$ completely unknown and non-random, a test statistics δ is wanted independent of the value of ψ. The theory of invariance can be used for this purpose. 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 apply the invariant test theory, it is necessary to check that the family of distributions $\mathcal{P}_\theta$ remains invariant under a group of transformations G (see [23] for details and definitions), which induces in the parameter space the group $\overline{G} = \{\overline{g}\}$ that leaves both Θ_0 and Θ_1 unchanged. The optimal invariant tests are based on the maximal invariants (invariance principle [23]). Let $T = T(Y)$ be a maximal invariant. A statistics S is invariant if it depends on the observation Y via the maximal invariant $T : S = \varphi(T(Y))$.

C. GLR tests

The invariant approach to testing may be of poor if not no help when a group of transformations G under which the problem remains invariant is hard to find, or when the maximal invariant does not exist. For example, as shown in section IV, for a linear model with bounded nuisance parameters $\psi \in \Psi$, it is difficult to show that the hypotheses testing problem is invariant, whereas the unbounded nuisance case is usually to be solved with invariant tests. The same kind of difficulties takes place in case of stochastic model with non-linear nuisance parameters.

The likelihood ratio (LR) is a standard decision statistics in a wide class of hypotheses testing problem. But in the case of nuisance parameter, the LR cannot be directly computed because of the unknown vector ψ. Hence, an adaptive testing method (i.e. a method based on the estimation of ψ from input data) as the GLR test $\widehat{\delta}$ can be used :

$$\widehat{\delta}(Y) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \widehat{\Lambda}(Y) < h(\alpha) \\ \mathcal{H}_1 & \text{if} \quad \widehat{\Lambda}(Y) \geq h(\alpha) \end{cases} \tag{21}$$

$$\text{where} \quad \widehat{\Lambda}(Y) \triangleq 2 \log \frac{\sup_{\phi \in \Phi_1, \psi \in \Psi} f_{\phi,\psi}(Y)}{\sup_{\phi \in \Phi_0, \psi \in \Psi} f_{\phi,\psi}(Y)}.$$

D. Minimax tests

The minimax method considers the worst case situation, e.g. the closest alternatives. It consists in optimizing the worst case situation [23], [25]: a test $\overline{\delta}$ is *minimax* in $\mathcal{K}_\alpha$ if it maximizes the minimum power in this class:

$$\forall \delta \in \mathcal{K}_\alpha \ : \quad \inf_{\phi \in \Phi_1,\ \psi} \beta_{\overline{\delta}}(\phi, \psi) \geq \inf_{\phi \in \Phi_1,\ \psi} \beta_{\delta}(\phi, \psi) \tag{22}$$

In the case of nuisance parameters, this requires, first, to define a probabilistic distance (KL, for instance) and find a set of closest alternatives; to design a test, optimal in some sense over that set corresponding to the worst case.

IV. Application to hypotheses testing and change detection/isolation

We now address the application of techniques described in III to the two testing situations introduced in II, and first define a generic model with nuisance parameters.

A. Generic observation model with nuisance parameters

Consider the discrete time stochastic system:

$$Y_k = \mathcal{F}(X_k, \phi(k), \xi_k, k), \tag{23}$$

where $Y \in \mathbb{R}^r$ is the measured output, $\phi \in \mathbb{R}^m$ the parameter of interest, $X_k \in \mathbb{R}^q$ an unknown vector (typically a state or a nuisance fault parameter), ξ a zero-mean white noise. This system is observed sequentially: at time n, $Y_1, \dots, Y_n$ are available. Several cases of model (23), which turn out to be relevant in practice, are discussed in section V.

B. Hypotheses testing

Testing hypotheses in the presence of nuisance parameters is first addressed in a particular case.

1) Linear model - Invariant test: A linear instance of (23) of wide interest [26], [27] is:

$$Y_k = H \ X_k \ + \ M \ \phi \ + \ \xi_k, \tag{24}$$

where $r > \max\{m, q\}$, H is a $r \times q$ full column rank (f.c.r.) matrix, M is a $r \times m$ f.c.r. matrix, and the white noise $\xi_k \sim \mathcal{N}(0, \sigma^2 I_r)$ is Gaussian distributed with known $\sigma^2 > 0$. The application of the invariance principle to the hypotheses testing problem (2) - (3) for model (24) involves the projection of Y onto the orthogonal complement $R(H)^\perp$ of the column space of matrix H. It results from [27] that the invariant UBCP test is given by:

$$\delta^*(Y) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \Lambda(Y) < h(\alpha) \\ \mathcal{H}_1 & \text{if} \quad \Lambda(Y) \geq h(\alpha) \end{cases} \tag{25}$$

$$\text{where } \Lambda(Y) = Y^T P_H M \left(M^T P_H M\right)^{-1} M^T P_H Y / \sigma^2 \tag{26}$$

and $P_H = I_r - H(H^T H)^{-1} H^T$. Test (25) is UBCP over the family of surfaces:

$$\mathcal{S}_{WM} = \left\{ S_c : 1/\sigma^2 \ \ \|WM\phi\|_2^2 = c^2, \ c > 0 \right\}. \tag{27}$$

2) Linear model - GLR test: For the linear model (24), it can be shown that the GLR test writes:

$$\widehat{\Lambda}(Y) = 2 \log \frac{\sup_{\phi, X} f_{\phi, X}(Y)}{\sup_X f_X(Y)} = \frac{Y^T (P_H - P_{\widetilde{H}}) Y}{\sigma^2} = \Lambda(Y)$$

where $\widetilde{H} \triangleq (H \ M)$, $P_{\widetilde{H}} = I_r - \widetilde{H}(\widetilde{H}^T \widetilde{H})^{-1} \widetilde{H}^T$ is assumed to be f.c.r., and $\Lambda(Y)$ is in (26). Hence, the GLR test for model (24) is an optimal invariant test.

3) Linear model - Minimax test: For the linear model (24) with $M = I_r$, it is easy to see that the KL distance is decomposed in the following manner:

$$\rho(\theta, 0) \triangleq \rho_1(\theta, 0) + \rho_2(\theta, 0) = \frac{1}{2\sigma^2} \theta^T P_H \theta + \frac{1}{2\sigma^2} \theta^T P_H^\perp \theta,$$

where $P_H^\perp = H(H^T H)^{-1} H^T$. The cylindric surface of equal power can be interpreted as a surface of equal component $\rho_1(\theta, 0)$ of KL distance which defines the set of closest to $R(H)$ alternatives.

4) Non-linear model and/or bounded parameters: As mentioned above, for the non-linear model:

$$Y_k = H(X_k) \; + \; M(\phi) \; + \; \xi_k, \tag{28}$$

or for the linear model (24) with bounded nuisance parameters $X_k \in \mathcal{X} \subset \mathbb{R}^q$, the application of the invariant tests theory is usually very difficult. The solution is the GLR test, and sometimes the minimax test. If the non-linearity of functions $H(.)$ and $M(.)$ is moderate, then an ε-optimal solution with a guarranted loss of optimality w.r.t. a linear model can be obtained by using the linearization of $H(.)$ and $M(.)$ and applying the invariant solution as an approximation to an optimal one for limited sets of values of X and ϕ [28].

5) Local asymptotic model: The linear model (24) is also of interest when monitoring non-linear systems and/or non-Gaussian noises, thanks to the asymptotic local approach [29]. This approach, which assumes *small changes*, is aimed at circumventing the (difficult) issue of unknown distribution of most decision functions, including the likelihood ratio. Other estimation methods than ML can be considered in building residuals for monitoring component faults affecting the dynamics of a system. Performing the early detection and isolation of slight deviations of a process, w.r.t. a reference behavior, is of crucial importance for condition-based maintenance. The local approach provides tools which perform the early warning task [29].

Typically, a residual writes under the form:

$$\zeta_N(\theta) \triangleq \sum_{k=1}^{N} K(\theta, Z_k)/\sqrt{N} \tag{29}$$

where K is an estimating function for θ and Z_k is an auxiliary process based on (Y_k) [29]. Under some conditions, the residual ζ_N is asymptotically Gaussian distributed, and reflects a small fault by a change in its *mean* vector.

Based on a first order Taylor expansion of (29):

$$\zeta_N(\theta) \approx \zeta_N(\theta_0) + 1/\sqrt{N} \sum_{k=1}^{N} \partial/\partial\theta K(\theta_0, Z_k)\Upsilon/\sqrt{N}$$

the following CLT can be shown to hold [6]:

$$\zeta_N(\theta_0) \rightarrow \left\{ \begin{array}{lll} \mathcal{N}(0, & \Sigma(\theta_0)) & \text{under} \quad \mathbf{P}_{\theta_0} \\ \mathcal{N}(M(\theta_0)\Upsilon, & \Sigma(\theta_0)) & \text{under} \quad \mathbf{P}_{\theta_0+\Upsilon/\sqrt{N}} \end{array} \right.$$

provided $\Sigma(\theta_0)$ is positive definite,

$$\text{with } M(\theta_0) \triangleq - \mathbb{E}_{\theta_0} \, \partial/\partial\theta \; K(\theta, Z_k)|_{\theta=\theta_0} \tag{30}$$

and: $\Sigma(\theta_0) \triangleq \lim_{N\rightarrow\infty} \Sigma_N(\theta_0)$, where:

$$\begin{array}{l} \Sigma_N(\theta_0) \triangleq \mathbb{E}_{\theta_0} \left(\zeta_N(\theta_0) \; \zeta_N^T(\theta_0) \right) \\ = 1/N \sum_{k=1}^{N} \sum_{j=1}^{N} \mathbb{E}_{\theta_0} \left(K(\theta_0, Z_k) K^T(\theta_0, Z_j) \right) \end{array} \tag{31}$$

This theorem means that a small deviation in θ is reflected into a change in the mean of ζ_N (29), which is asymptotically Gaussian distributed with the same covariance matrix under both null and local alternative hypotheses.

The main use of this asymptotic Gaussianity result is the design of asymptotically optimum tests between composite hypotheses [6]. For deciding between $\Upsilon{=}0$ and $\Upsilon{\neq}0$, the optimum test statistics is the GLR which, in this case, writes:

$$\chi_N^2 = \zeta_N^T \; \Sigma_N^{-1} M \left(M^T \Sigma_N^{-1} M \right)^{-1} M^T \Sigma_N^{-1} \; \zeta_N \tag{32}$$

where the dependence on θ_0 has been removed for simplicity. Under both hypotheses, test (32) is distributed as a χ^2-random variable with $l(= \dim \Upsilon)$ degrees of freedom, and non-centrality parameter $\gamma = \Upsilon^T M^T \Sigma^{-1} M \Upsilon$ under $\mathbf{H}_1$. The estimation of $M(\theta)$ in (30) is obtained using sample averaging. Estimating $\Sigma(\theta)$ in (31) is more tricky [31].

C. Change detection/isolation

Pursuing the discussion of the linear model (24), we are now interested in detecting a change from 0 to $\phi_l \neq 0$, while considering X as an *unknown* nuisance parameter.

1) Invariant change detection/isolation: Motivated by the navigation system integrity monitoring discussed in section V, consider again the case of model (24) with $M = I_r$. This now writes:

$$Y_k = HX_k + \phi_l(k, k_0) + \xi_k$$

where $\phi_l(k, k_0)$ is the l-type change occurring at time k_0:

$$\phi_l(k, k_0) = \left\{ \begin{array}{lll} 0 & \text{if} & k < k_0 \\ \phi_l & \text{if} & k \geq k_0 \end{array} \right. , \; 1 \leq l \leq K. \tag{33}$$

The recursive invariant algorithm asymptotically attaining the lower bound for the maximum detection/isolation delay (14) is given by the alarm time and final decision (N_r, ν_r) [14]:

$$N_r \triangleq \min_{1\leq l\leq K-1} \{N_r(l)\}, \quad \nu_r \triangleq \arg \min_{1\leq l\leq K-1} \{N_r(l)\}, \tag{34}$$

where

$$N_r(l) = \inf \left\{ t \geq 1 : \min_{0\leq j\neq l\leq K-1} [g_t(l,0) - g_t(j,0) - h_{l,j}] \geq 0 \right\}$$

and the recursive decision function $g_t(l, 0)$ is defined by:

$$g_t(l, 0) = (g_{t-1}(l, 0) + Z_t(l, 0))^+ \tag{35}$$

with the invariant log-LR

$$S(Y_k; l, 0) = \frac{1}{\sigma^2} \phi_l^T P_H Y_k - \frac{1}{2\sigma^2} \phi_l^T P_H \phi_l, \tag{36}$$

$g_0(l, 0) = 0$ for every $1 \leq l \leq K - 1$ and $g_t(0, 0) \equiv 0$. The thresholds $h_{l,j}$ are chosen by the following formula :

$$h_{l,j} = \left\{ \begin{array}{lllll} h_\text{d} & \text{if} & 1 \leq l \leq K-1 & \text{and} & j = 0 \\ h_\text{i} & \text{if} & 1 \leq j, l \leq K-1 & \text{and} & j \neq l \end{array} \right. , \tag{37}$$

where h_d and h_i are the detection and isolation thresholds.

2) Minimax change detection/isolation: Let:

$$S_k(l, j) = \log f_{\phi_l}(Y_k; X^l)/f_{\phi_j}(Y_k; X^j)$$

be the likelihood ratio between $\mathcal{H}_l$ and $\mathcal{H}_j$, where $\mathcal{H}_l = \{Y \sim \mathcal{N}(HX^l + \phi_l, \sigma^2 I_r)$ and $0 \leq l, j \neq l \leq K$. It results from sections II-B and III-D that the lower bound for the detection/isolation delay is a monotone decreasing function of the KL information. Therefore, designing the *minimax algorithm* (minimizing the detection/isolation delay) consists in finding a pair of *least favorable values* X^l and X^j for which the KL information $\rho_{l,j} = \rho(X^l, X^j)$ is minimum, and in computing the LR $S_k(l,j)$ corresponding to these values. Since the KL information is $\rho_{l,j}(x) = \|Hx + \phi_{l,j}\|^2 / 2\sigma^2$, where $x = X^l - X^j$ and $\phi_{l,j} = \phi_l - \phi_j$, this boils down to minimize $\rho_{l,j}(x)$ w.r.t. x [11]. The log-LR between $\mathcal{H}_l$ and $\mathcal{H}_j$ under the least favorable value x^* writes:

$$S(x^*; l, j) = \frac{1}{\sigma^2}\phi_{l,j}^T P_H Y_k - \frac{1}{2\sigma^2}\phi_{l,j}^T P_H \phi_{l,j} \quad (38)$$

It is of interest to note that $S(x^*; l, j)$ coincides with the log-LR (36) of the invariant change detection/isolation algorithm.

V. Navigation system integrity monitoring

We now describe an example outlining the importance of dealing with nuisance parameters for system monitoring. Integrity monitoring requires that a navigation system detects, isolates faulty measurement sources, and removes them from the navigation solution before they sufficiently contaminate the output. A conventional multi-sensor integrated navigation system includes different sensors (or subsystems): INS, GPS, Loran-C, air-data subsystem, Doppler radar and others. A wide group of navigation (sub)systems is described by measurement equations which can be reduced to the static (slightly non) linear regression model:

$$Y_k = H(X_k) + \phi_l(k, k_0) + \xi_k, \quad (39)$$

where $H(.)$ is a known vector-valued function, the state vector X contains the unknown position of the vehicle (or its velocities, accelerations,...) and $\phi_l(k, k_0)$ is the l-type change occurring at time k_0 as in (33). Physically, the vector X is completely unknown, non-random and it ranges over a large domain of possible values. Therefore, it is considered as a nuisance parameter. Integrity monitoring is then reduced to the detection of faults ϕ_l that leads to a positioning failure which is usually defined in terms of protection zone around the aircraft (vertical and horizontal alarm limits). By linearization of (39), the FDI integrity monitoring algorithm is given by equations (34)-(36) in subsection IV-C.1.

VI. 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 and isolation have been discussed, in both frameworks of (off-line) hypotheses testing and (on-line) change detection and onset time estimation. Several major statistical tools for solving these problems have been introduced. Particular emphasis has been put on handling nuisance parameters. The application to GPS integrity monitoring and vibration-based structural health monitoring has been addressed.

References

[1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, www.irisa.fr/sisthem/kniga, 1993.

[2] ——, "Fault isolation for diagnosis : Nuisance rejection and multiple hypothesis testing," *Annual Reviews in Control*, vol. 26, no. 2, pp. 189–202, 2002.

[3] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, vol. 54, pp. 426–482, 1943.

[4] C. R. Rao, *Linear Statistical Inference and its Applications*. John Wiley & Sons, 1973.

[5] S. Zacks, *The Theory of Statistical Inference*. Wiley, 1971.

[6] W. J. Hall and D. J. Mathiason, "On large sample estimation and testing in parametric models," *Int. Stat. Review*, vol. 58, no. 1, pp. 77–97, 1990.

[7] K.-R. Koch, *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer-Verlag, Berlin, 1999.

[8] D. P. Malladi and J. L. Speyer, "A generalized Shiryayev sequential probability ratio test for change detection and isolation," *IEEE Trans. Automat. Contr.*, vol. 44, no. 8, pp. 1522–1534, 1999.

[9] T. L. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 595–608, 2000.

[10] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Stat.*, vol. 42, pp. 1897–1908, 1971.

[11] I. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 171–187, Jan. 1995.

[12] ——, "On two new criteria of optimality for the problem of sequential change diagnosis," in *Proc. Am. Contr. Conf., Seattle, USA*, 1995, pp. 97–101.

[13] ——, "A simple recursive algorithm for diagnosis of abrupt changes in random signals," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2740–2746, Nov. 2000.

[14] ——, "A lower bound for the detection/isolation delay in a class of sequential tests," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 3037–3046, Nov. 2003.

[15] D. Basu, "On the elimination of nuisance parameter," *Jal Am. Stat. Assoc.*, vol. 72, no. 358, pp. 355–366, 1977.

[16] J. O. Berger, B. Liseo, and R. L. Wolpert, "Integrated likelihood methods for eliminating nuisance parameters (with discussion)," *Statistical Science*, vol. 14, no. 1, pp. 1–28, 1999.

[17] B. Liseo, "The elimination of nuisance parameters," in *Handbook of Statistics*, D. Dey and C. Rao, Eds. Elsevier, 2005, vol. 25, pp. 811–825.

[18] T. A. Severini, "Likelihood functions for inference in the presence of a nuisance parameter," *Biometrika*, vol. 85, pp. 507–522, 1998.

[19] K.-Y. Liang and S. L. Zeger, "Inference based on estimating functions in the presence of nuisance parameters," *Statistical Science*, vol. 10, no. 2, pp. 158–172, 1995.

[20] T. A. Severini, *Likelihood Methods in Statistics*. Oxford Univ. Press, 2000.

[21] V. S. Huzurbazar, "Probability distributions and orthogonal parameters," *Proc. Camb. Phil. Soc.*, vol. 46, pp. 281–284, 1950.

[22] D. R. Cox and N. Reid, "Parameter orthogonality and approximate conditional inference (with discussion)," *Jal Royal Stat. Soc.*, vol. 49, no. 1, pp. 1–39, 1987.

[23] E. L. Lehmann, *Testing Statistical Hypotheses*. Wiley, 1986.

[24] S. A. Murphy and A. W. van der Vaart, "On profile likelihood (with discussion)," *Jal Am. Stat. Assoc.*, vol. 95, no. 450, pp. 449–485, 2000.

[25] A. Borovkov, *Statistique Mathématique*. Mir Editions, Moscow, 1987, (In French).

[26] M. Basseville, "Information criteria for residual generation and fault detection and isolation," *Automatica*, vol. 33, no. 5, pp. 783–803, 1997.

[27] M. Fouladirad and I. Nikiforov, "Optimal statistical fault detection with nuisance parameters," *Automatica*, vol. 41, no. 7, pp. 1157–1171, July 2005.

[28] H. Lacresse, A. Grall, and I. Nikiforov, "Statistical fault detection in a certain class of non-linear systems with nuisance parameters," in Proc. Amer. Control Conf., June 2003.

[29] M. Basseville, "On-board component fault detection and isolation using the statistical local approach," *Automatica*, vol. 34, no. 11, pp. 1391–1416, 1998.

[30] M. Basseville, M. Abdelghani, and A. Benveniste, "Subspace–based fault detection algorithms for vibration monitoring," *Automatica*, vol. 36, no. 1, pp. 101–109, Jan. 2000.

[31] Q. Zhang and M. Basseville, "Advanced numerical computation of χ^2–tests for fault detection and isolation," in *5th Symp. Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Washington*, June 2003, pp. 211–216.