



HAL
open science

On-Board Component Fault Detection and Isolation Using the Statistical Local Approach

Michèle Basseville

► **To cite this version:**

Michèle Basseville. On-Board Component Fault Detection and Isolation Using the Statistical Local Approach. [Research Report] RR-3252, INRIA. 1997. inria-00073437

HAL Id: inria-00073437

<https://inria.hal.science/inria-00073437v1>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***On-board component fault detection and isolation
using the statistical local approach***

Michèle Basseville

N° 3252

Septembre 1997

_____ THÈME 4 _____



***Rapport
de recherche***

On-board component fault detection and isolation using the statistical local approach

Michèle Basseville*

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Sigma2

Rapport de recherche n3252 — Septembre 1997 — 39 pages

Abstract: We describe both the key principles and real application examples of a unified theory which allows us to perform the on-board incipient fault detection and isolation tasks involved in monitoring for condition-based maintenance. We stress that, when designing detection algorithms, the main conceptual task is to select a convenient estimating function. ML, LS, IV and subspace identification methods are addressed in this perspective.

Key-words: Fault detection and isolation, component faults, statistical methods, local approach, linear and nonlinear dynamic systems.

(Résumé : *tsvp*)

This paper has been presented as a plenary talk at the *11th IFAC/IFORS Symposium on Identification and System Parameter Estimation - SYSID'97*, Kitakyushu, Japan, July 8-11, 1997, under the title '*Statistical approaches to industrial monitoring problems - Fault detection and isolation*'.

* IRISA/CNRS – basseville@irisa.fr.

Traitements embarqués pour la détection et le diagnostic de pannes de composants - Utilisation de l'approche statistique locale

Résumé : Cet article contient les développements de l'exposé présenté en séance plénière lors du *11ème Symposium IFAC/IFORS sur Identification and System Parameter Estimation - SYSID'97* qui s'est tenu à Kitakyushu, Japon, du 8 au 11 Juillet 1997.

On décrit les principes fondamentaux et des exemples d'application réels d'une théorie unifiée de conception de traitements embarqués pour la détection précoce et le diagnostic de pannes de composants. De tels traitements sont nécessaires pour la surveillance de systèmes en vue de leur maintenance conditionnelle. On met en évidence le fait que la principale tâche conceptuelle pour la conception de tels algorithmes consiste en la sélection d'une fonction d'estimation convenable. Les méthodes d'identification du maximum de vraisemblance, des moindres carrés, des variables instrumentales, et de type sous-espaces, sont discutées dans cette perspective.

Mots-clé : Détection et diagnostic de pannes, pannes de composants, méthodes statistiques, approche locale, systèmes dynamiques linéaires et non-linéaires.

1 Introduction

Before outlining the paper, we introduce model-based fault detection and isolation, distinguishing three types of faults in static or dynamic systems, discussing the design of FDI algorithms, and pointing up three different detection problems.

1.1 Condition monitoring and model-based FDI

In many applications, the problem of fault detection and isolation (FDI) is a crucial issue which has been theoretically and experimentally investigated with different types of approaches, as can be seen from the survey papers [63, 38, 29, 6, 24] and the books [55, 9] among other references. Moreover, an increasing interest in condition-based maintenance has appeared in a large number of industrial applications. The key idea there is to replace regular systematic inspections by condition-based inspections, i.e. inspections decided upon the continuous monitoring of the considered system (machine, structure, process or plant), based on the sensors data, in order to prevent from a possible malfunction or damage before it happens and optimize the maintenance cost. As shown in the examples described below, a possible solution consists in the early detection of slight deviations with respect to a parametric characterization of the system in its usual working conditions, without artificial excitation, speeding down, or stop. Indeed, if such an early detection can be performed while preserving robustness with respect to changes in normal operating conditions, one can reasonably expect being able to prevent from larger deviations resulting from malfunction, fatigue, faults or damage before they happen, and consequently to increase the availability of the system.

It turns out that there exists a mathematical statistics theory providing us with tools which perform this early detection task. This theory is known under the name of local approach. It aims at transforming a large class of detection problems concerning a parameterized stochastic *process* into the universal problem of monitoring the mean of a Gaussian *vector*.

It is the main purpose of this paper to describe both the key components of this theory and real industrial examples which have been successfully addressed by applying its principles. Of course, this theory assumes that a model of the monitored system is available. This is a reasonable assumption, since many industrial processes rely on physical principles, which write in terms of equations, providing us with models. But when physical models are either too complex or even not known at all, semi-physical or black-box models can be used instead, in the framework of the local approach, enlarging its initial scope. We assume the model structure to be given throughout, and we discuss further this important modelling issue at the end of the paper.

In the sequel, we equally use the words deviation, change, failure, fault, damage, malfunction, considering that all these events are reflected by a change in the parameter vector of a model of the system [6, 9].

1.2 Three types of faults

Several types of multidimensional parameterized stochastic processes are considered in this paper. They may be static or dynamic, linear or nonlinear, with a state-space description or an input-output representation given, in the latter case, in an explicit or an implicit form. In section 2, we describe several industrial examples of such processes. In all cases, as depicted on figure 1, the observed data Y are viewed as the output of a system parameterized by vector θ and having different kinds of inputs, which, neglecting the dynamics for the moment, we write :

$$Y = g(\theta, U + \Upsilon_i, W_s) + \Upsilon_o + W_o \quad (1)$$

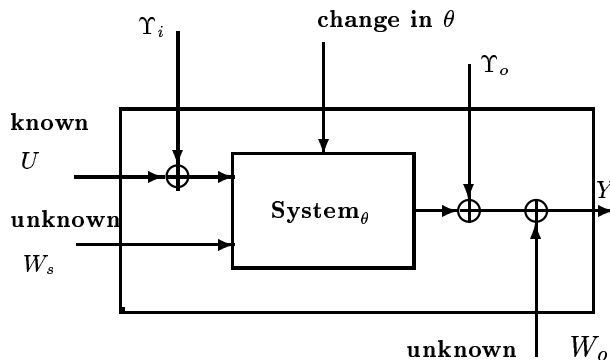


Figure 1: Three types of faults : input faults Υ_i and output faults Υ_o act additively, whereas system faults act as a change in the system parameter vector θ .

Input U is assumed to be known (measured). The unknown quantity W_s stands for non-measured inputs, unknown non-stationary excitation on or perturbation of the system, and input noise. The unknown quantity W_o stands for output noise.

We distinguish between three types of faults. The first type of faults Υ_o occur at the output of the system in an additive manner. It has been agreed that such faults appropriately model sensor faults. Speaking in terms of the distribution of the observed output data Y , the first type of faults only *affect its mean value*, which shifts to a different value after the occurrence of a fault. The second type of faults Υ_i occur at the input of the system in an additive manner. It has been agreed that such faults appropriately model actuator faults¹. If the system is linear, the second type of faults affect only the mean value of the output distribution. The third type of faults are modeled by *any change in the parameter θ* . They affect the generating mechanism itself and thus the variance, correlations, spectrum or higher order characteristics of the output distribution. This type of faults is often referred as component or system faults [55]. In the case of linear dynamic systems, it may be useful to think of these faults as multiplicative faults, because they affect the input-output transfer function in a multiplicative manner [9]. In the case of nonlinear static or dynamic systems, it should be clear that there are only two types of faults, namely the first and the third ones.

It is one of the purposes of this paper to outline how the fault type and, to a less extent, the process type, influence the design of FDI algorithms.

1.3 Design of model-based FDI algorithms

Residual generation and residual evaluation. It has been widely acknowledged - see the above mentioned references - that the FDI problem can be split into two steps : *generation of residuals*, which are ideally close to zero under no-fault conditions, minimally sensitive to noises and disturbances, and maximally sensitive to faults; and *residual evaluation*, namely design of decision rules based on these residuals.

When stated in terms of a (generally state-space) model of the monitored system, the FDI problem has been addressed using either deterministic or stochastic approaches. Deterministic methods for residual generation have been proposed both in the time domain (parity checks, resulting from analytical redundancy relations) and in the frequency domain (detection filters, fault observers), and both for linear and nonlinear systems [55]. The basic statistical approach to residual generation consists in deriving sufficient statistics [9], namely transformations of the measurements which capture the entire information about the fault contained in the original data.

¹Actuators faults, as well as sensors faults, can sometimes usefully be modeled as changes in the input or output variances, which we do not distinguish from the third type of change defined next.

Residual evaluation is typically answerable to statistical methods, which are basically aimed at deciding if a residual discrepancy from zero is significant. However deterministic approaches to residual evaluation have also been proposed recently [23].

The main advantage of the statistical approach is its ability in assessing the level of significance of discrepancies with respect to uncertainties. The accuracy of parameter estimates provides us with the relative size of the estimation error w.r.t. the noises on the data measured on the system. Similarly, the tests based on the local approach tell us if the relative size of the parameter discrepancy in the monitored system w.r.t. to the accuracy of the reference parameter value, is significant or not. However, it turns out that the estimation error or the prediction error are generally *not* the relevant functions of the model parameter and the measured data to be computed for stating this significance. It is one of the main purposes of this paper to show alternative functions of the model and the data which should be used as residuals, and how to evaluate these residuals. This alternative approach to residual generation and evaluation is called the local approach.

We find it useful, following [30] and although with a different meaning, to distinguish between a *primary residual*, and an *improved residual*, for making the decision.

Primary residuals and improved residuals. Neglecting the dynamics in the notation again, we summarize the two steps of the local approach to residual generation and evaluation as follows.

- *Primary residual* : a vector-valued function or process $K(\theta, U, Y)$, whose mean value switches from zero when the system is safe towards a non-zero value when a fault occurs. As explained in sections 3 and 4, such a function is tightly related to an estimating function used for the identification of parameter θ ;
- *Improved residual* : a vector-valued function $\zeta(\theta, U, Y)$, built on K , with known distribution, making possible its evaluation. Moreover, in all cases, this known distribution turns out to be the Gaussian distribution, with known covariance matrix.

In other words, any dynamic process is summarized via an (asymptotically) equivalent Gaussian vector (and *not* process). This performs a reduction of FDI problems to a universal one : any monitoring problem concerning a parameterized random process can be reduced to the problem of monitoring the mean of a Gaussian random vector.

1.4 Three types of detection problems

Three types of detection problems may occur in practice, according to the relative time constants of the processes to be monitored, of the sampling of the data and of the events to be detected. These three problems are depicted on figure 2.

- *Model validation* : Given, on one hand, a reference value θ_0 of the model parameter (for instance, identified with data from the safe system) and, on the other hand, a new data sample, the problem is to decide whether the new data are still well described by this parameter value or not. This problem may be stated either off-line (fixed sample size N) or on-line (varying sample size n);
- *Off-line change detection* : Given a data sample of size N , the problem is to decide whether, somewhere in this sample, a change in the parameter has occurred, from the value θ_0 to the value θ_1 , at an unknown time instant v ;
- *On-line change detection* : At every time instant n , the problem is to decide whether, before this instant, a change in the parameter has occurred, from the value θ_0 to the value θ_1 , at an unknown time instant v .

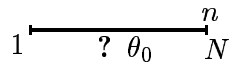
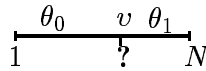
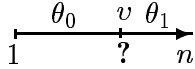
Model validation**Off-line change detection****On-line change detection**

Figure 2: Three detection problems.

It should be clear that the most difficult problem is the third one, because in this problem the amount of information in the data about the new parameter value θ_1 is the lowest. Also, the criteria for designing the detection algorithms and analyzing their performances depend on the detection problem. These are : mean time between false alarms, probability of wrong detection, mean delay to detection, probability of non detection, accuracy of the estimates of the fault onset time and of the magnitude of the change. For precise definition of detection and isolation criteria, we refer the reader to [9][chap.4], [61, 15, 50, 52].

However, we argue that, even though the decision functions for solving these three problems are not the same, they all reduce to different implementations of the same primary residual. In order not to blur the main purpose of the paper, we postpone further discussion of this implementation issue in section 9, and we put main emphasis on the model validation problem, which is the simplest. We do not discuss the issue of quickest detection of a change, and concentrate the discussion on how the fault type influences the choice of the estimating function K which is the basis of the primary residual to be handled.

In the framework of this paper, the fault isolation problem consists in determining which (subsets of) components of the parameter vector θ have changed. How this relates to diagnostics is discussed in section 10.

1.5 Outline of the paper

In section 2, we describe three real examples of static or dynamic industrial processes, on which different types of faults have to be detected and isolated. Then, in section 3, we explain how the three types of FDI problems can be reduced to the universal problem of detecting a change in the mean of a Gaussian vector, using either a finite sample or an asymptotic argument (local approach). Section 4 is devoted to the local approach to the design of detection algorithms, based on either the likelihood function or alternative estimating functions. The reduced Gaussian problem is investigated in section 5, where statistical tests for detection and isolation are described.

We then return to the three examples introduced in section 2. In section 6, we describe two types of detection algorithms for vibration monitoring, based on either instrumental variable or subspace identification methods. In section 7, we consider the case of faults in an error-in-variables nonlinear static model, and describe a least-squares (LS) based detection algorithm developed for monitoring the combustion chambers of a gas turbine. Section 8 is devoted to the case of nonlinear dynamic processes. We introduce two possible solutions based on a LS estimating function, and involving either a full state observer (as done for a catalytic converter), or an input-output equivalent representation. Finally, in section 9, we discuss several additional issues, among

which the use of models and the optimum sensor location problem. Some conclusions are drawn in section 10.

2 Industrial examples

In this section, we describe three real examples of industrial processes subject to faults or other undesirable deviations which have to be detected and isolated. These processes are linear dynamic, nonlinear static, and nonlinear dynamic respectively. They are all subject to *component faults*. We refer the reader to [9, 51] for interesting applications of statistical methods to *sensor faults* in linear systems, for inertial navigation systems and GPS integrity monitoring.

2.1 Linear dynamic system : vibration monitoring

Vibration monitoring² turns out to be an important example of the problem of detection and isolation of non-additive faults (changes in the eigenstructure) in linear dynamic systems, as depicted on figure 3.

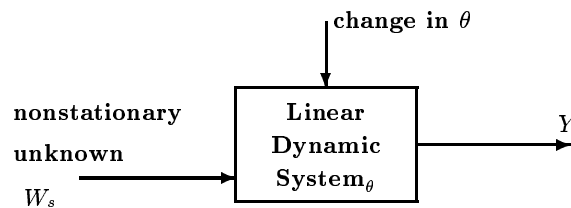


Figure 3: Vibration monitoring.

The monitoring problem. In-situ damage monitoring and predictive maintenance for complex mechanical structures and rotating machines is of key importance in many industrial areas : power engineering (rotating machines, core and pipes of nuclear power plants), civil engineering (large buildings subject to hurricanes or earthquakes, bridges, dams, offshore structures), aeronautics (wings and other structures subject to strength), automobile, rail transportation. These systems are subject to both fast and unmeasured variations in their environment and small slow variations in their vibrating characteristics. Of course, only the latter are to be detected, using the available data (accelerometers), and noting that the changes of interest (1% in eigenfrequencies) are not visible on spectra.

For example, offshore structures are subject to the turbulent action of the swell which cannot be considered as measurable. Moreover, the excitation is highly time-varying, according to wind and weather conditions. Therefore the challenge is as follows. The available measurements do not separate the effects of the external forces from the effect of the structure, the external forces vary more rapidly than the structure itself (fortunately!), damages or fatigues on the structure are of interest, while any change in the excitation is meaningless. Expert systems based on a human-like exploitation of recorded spectra can hardly work in such a case, the FDI method must rather rely on a model which will help in discriminating between the two mixed causes of the changes that are contained in the accelerometer measurements.

A different example concerns rotating machines such as huge alternators in electricity power plants. Most of the available vibration monitoring techniques require to stop the production,

²Here we refer to a work which was supported by several grants from IFREMER (Institut Français de Recherche pour la Mer), Brest, France and EDF (Électricité de France), Direction des Études et Recherches, Clamart, France. Further work is being accomplished in the context of Eureka project no 1562 SINOPSYS (Model based Structural monitoring using **in-operation system identification**) coordinated by LMS, Leuven, Belgium.

in order to vary the rotation speed and, by the way, get reliable spectra. When monitoring the machine without speeding it down, one is faced with the following challenge. The non-stationary and non-measured excitation is due to load unbalancing which creates forces with known frequency range (harmonics of the rotation speed) but unknown geometry, and to turbulence caused by steam flowing through the alternator and frictions at the bearings. The problem is again to discriminate between two mixed causes of changes in vibrating characteristics.

Modeling issues. We assume that the behavior of the mechanical system can be described by a stationary linear dynamic system, and that, in the frequency range of interest, the input forces can be modeled as a non-stationary white noise. Consequently the relevant model is the following matrix differential equation :

$$\begin{cases} M\ddot{\mathcal{X}}(t) + C\dot{\mathcal{X}}(t) + K\mathcal{X}(t) &= \nu(t) \\ Y(t) &= L\mathcal{X}(t) \end{cases} \quad (2)$$

where t denotes continuous time, M, C, K are the mass, damping and stiffness matrices respectively, the (high dimensional) vector \mathcal{X} collects the displacements of the degrees of freedom of the structure; the external (non measured) force ν is modeled as a non-stationary white noise with time-varying covariance matrix $Q_\nu(t)$, measurements are collected in the (low dimensional) vector Y , and matrix L states where the sensors are located. The mechanical characteristics (M, C, K) of the system cannot be recovered from the output measurements. Hence, identifiable modal characteristics of the system are introduced : the vibration modes or eigenfrequencies denoted generically by μ , and the modal shapes or eigenvectors denoted generically by ψ_μ . These quantities are solutions of the following equation :

$$(M\mu^2 + C\mu + K)\Psi_\mu = 0 \quad , \quad \psi_\mu = L\Psi_\mu \quad (3)$$

Sampling model (2) at rate $1/\tau$ yields the discrete time model in state space form :

$$\begin{cases} X_{k+1} &= FX_k + V_k \\ Y_k &= HX_k \end{cases} \quad (4)$$

where :

$$\begin{aligned} Y_k &= Y(k\tau) \quad , \quad X_k = \begin{bmatrix} \mathcal{X}(k\tau) \\ \dot{\mathcal{X}}(k\tau) \end{bmatrix} \quad , \quad H = \begin{bmatrix} L & 0 \end{bmatrix} \quad , \\ F &= \exp(\mathcal{A}\tau) \quad , \quad \mathcal{A} = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix} \quad , \\ \mathbf{E}(V_k) &= 0 \quad , \quad \mathbf{E}(V_k V_k^T) = \int_{k\tau}^{(k+1)\tau} e^{\mathcal{A}s} \tilde{Q}(s) e^{\mathcal{A}^T s} ds \\ \text{with : } \tilde{Q}(s) &= \begin{bmatrix} 0 & 0 \\ 0 & M^{-1}Q_\nu(s)M^{-T} \end{bmatrix} \end{aligned}$$

The modal characteristics defined in (3) are equivalently found in the eigenstructure (λ, Φ_λ) of F :

$$e^{\tau\mu} = \lambda, \quad \psi_\mu = \phi_\lambda \triangleq H \Phi_\lambda \quad (5)$$

Such a pair (λ, ϕ_λ) is called a mode.

A third useful model is the ARMA time series :

$$Y_k = \sum_{i=1}^p A_i Y_{k-i} + \sum_{j=0}^{p-1} B_j(k) W_{k-j} \quad (6)$$

where $(W_k)_k$ is a standard stationary white noise. In this model, it is easily verified that the AR coefficients A_i are related to the pair (H, F) of model (4) via :

$$HF^p = \sum_{i=1}^p A_i HF^{p-i} \quad (7)$$

On the other hand, since the covariance matrix of the excitation noise V is non-stationary, the MA coefficients $B_j(k)$ are *time-varying*.

We stress that sinusoidal or colored noise excitation can be encompassed as well [57].

From these modelling issues, it results that vibration monitoring can be equally stated as the problem of detecting changes in the eigenstructure of the state transition matrix of a linear dynamic system with non-stationary state noise, or the problem of detecting changes in the AR part of a multidimensional ARMA process with non-stationary MA part.

2.2 Nonlinear static system : gas turbine

Detecting faults in the combustion set of a gas turbine³ is a problem which can be addressed within the framework of system changes in a nonlinear static model with noisy input and output measurements, and also unknown inputs. Figure 4 means that the problem is again to infer the internal state of a complex physical process using only the available input and output measurements.

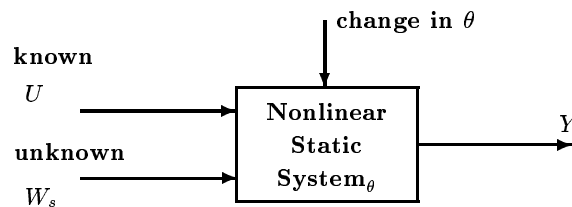


Figure 4: Monitoring combustion chambers.

The monitoring problem. Gas turbines are widely used as industrial motors, in electrical power generators and aircraft engines for example. As shown on figure 5, a gas turbine system is composed of a compressor, a set of combustion chambers, an expansion turbine and an exhauster. The problem here is to monitor the (ten) combustion chambers and the expansion turbine using the measurements given by the (eighteen) thermocouples at the output of the exhauster, to detect both sudden failures and gradual incipient faults, and to isolate which combustion chamber(s) failed.

This monitoring problem is complex for the following reasons. First, the knowledge about the physics of the involved phenomena (combustion, expansion, diffusion) is not complete and in some sense less mature than in the previous example of mechanical systems. There is thus a need for a semi-physical model, which provides us with a reasonable approximation of the measurements. Second, as explained below, this model has both noisy input and output measurements (errors-in-variable model). Third, one of the key difficulties here is that the considered gas turbines are used as secondary energy production devices in electricity power plants. Thus their functioning mode is changing quite often, and the effects of this non-stationarity on the system parameters are non negligible with respect to those of the faults.

³Here we refer to a work which was supported by several grants from Alcatel-Alsthom-Recherche, Marcoussis, France and European Gas Turbine S.A., Belfort, France.

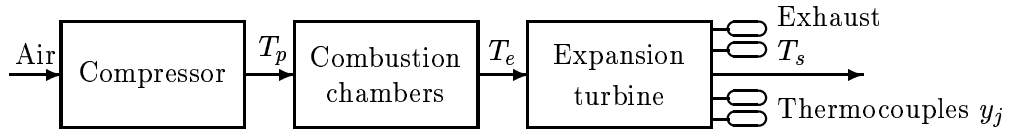


Figure 5: Gas turbine system.

Modeling issues. The two key ideas used for deriving a semi-physical model are to describe the output measurements (thermocouples at the output of the exhauster) as a linear combination of temperature profiles contributed by each combustion chamber, and to compress complex phenomena (rotation and turbulence in the turbine and the exhauster) into three phase shift coefficients.

The shape of the temperature profile of the i -th chamber contributed on the j -th thermocouple is modeled by $g(\phi_j^t - \phi_i^c - d)$, where $g(\cdot)$ is a zero mean nonlinear function, ϕ_i^c and ϕ_j^t are the angular coordinates of the chamber and the thermocouple respectively, and the phase shift d is the rotation angle of the gas during the expansion. Using prior knowledge, we can compute an approximation of the phase shift, using a known function $D(T_e, T_s, N)$, where N is the rotation speed of the expansion turbine, T_e is the mean temperature at the entrance of the turbine, T_s denotes the mean temperature at the exhaust of the turbine (average of the eighteen thermocouples). To compensate for this approximation in $D(T_e, T_s, N)$, we use an additional parameter d_0 , and a corrective term $d_1 \sin(\phi_j^t - \phi^s)$ for the asymmetry in the shape of the exhauster. Each chamber contribution is weighted with a coefficient a_i and the total contribution is scaled by the relative increase in temperature between the compressor output (T_p) and the turbine input (T_e), and then by the drop in temperature T_s/T_e during the expansion. An additional term $b \cos(\phi_j^t - \phi^s)$ takes into account the geometric asymmetry of the expansion turbine.

The j -th measurement is thus modeled as [47] :

$$y_j = \frac{T_s}{T_e} \left(T_e + \frac{T_e - T_p}{C} \sum_{i=1}^{10} a_i g(\phi_j^t - \phi_i^c - d) \right) + b \cos(\phi_j^t - \phi^s) + w_j \quad (8)$$

$$d = D(T_e, T_s, N) + d_0 + d_1 \sin(\phi_j^t - \phi^s)$$

where C is a normalizing constant, and w_j is a white noise capturing modeling and measurement errors.

To summarize, the model parameter vector is :

$$\theta = (a_1, \dots, a_{10}, b, d_0, d_1)^T$$

and the measured⁴ inputs and outputs are :

$$U_k = (T_e, T_p, T_s, N)^T(k) \quad (9)$$

$$Y_k = (y_1, \dots, y_{18})^T(k) \quad (10)$$

We note that this model is static (hence any data sample is considered as an i.i.d. sequence), and linear in the parameters, *except* for the phase shift coefficients d_0, d_1 . A key difficulty arises because this model has both noisy outputs *and* inputs (errors-in-variable model). Thus the parameter vector cannot be identified without bias. This identification bias could be thought as an obstacle to monitoring small deviations. However it is not the case, as we show in section 7.

2.3 Nonlinear dynamic system : catalyst

Detecting faults in the catalytic converter and the oxygen sensors of an automobile is a problem which can be addressed within the framework of changes in a parametric nonlinear dynamic

⁴Actually, temperature T_e is not measured, but estimated.

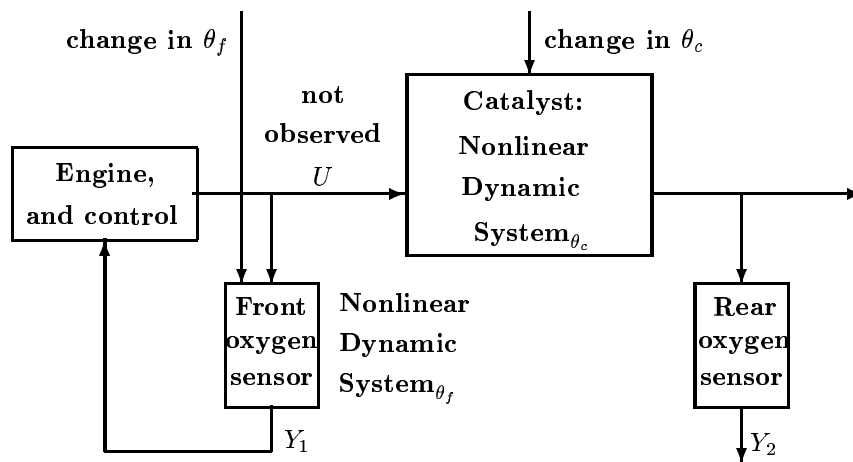


Figure 6: Monitoring the catalytic converter and the front oxygen sensor of an automobile.

model⁵. As obvious from figure 6, this system is still more tricky, because of a closed-loop which may blur some of the changes of interest, as do also varying operating conditions.

The monitoring problem. On-board fault diagnosis is a crucial issue for the automotive industry. New on-board diagnostic system requirements (OBD II) have been introduced in California [20, 31] which require monitoring vehicle emissions. Among the vehicle components to be monitored are the catalytic converter and the oxygen sensors.

A catalytic converter is aimed at processing the engine exhaust gases and reducing the quantity of pollutants. In particular, three way catalytic converters (TWC) oxidize the hydrocarbons HC and carbon monoxide CO and reduce nitrous oxides NO to produce water H_2O , carbon dioxide CO_2 and nitrogen N_2 . A catalytic converter is characterized by its efficiency in converting HC , CO and NO_x . Thanks to a front oxygen sensor (located upstream from the catalytic converter), a closed-loop controller maintains the air-fuel ratio in the engine close to the stoichiometric value. Because of the nonlinear characteristic of the oxygen sensor, a limit-cycle appears in the closed-loop system. Some possible causes of catalyst deterioration are a high temperature, due to the misfire for instance (an excess of fuel and air entering the catalyst increases the temperature), the malfunctions of the fuel system, a failure in the ignition system.

Two oxygen sensors – front and rear – measure the lack or excess of oxygen in the exhaust gases, in front of and after the catalytic converter respectively; see figure 6. We consider lambda sensors, which behave as follows. Their output voltage suddenly switches from zero to one when the gas mixture changes from lean to rich, and their response time is a function of temperature and air-fuel ratio, with a nearly bang-bang static characteristic. The response times for both lean-to-rich and rich-to-lean switches increase during the life cycle of a lambda sensor.

Due to the OBD recommendations, the problem is to monitor the catalytic converter and the front oxygen sensor.

Modeling issues. Models of all the elements of the de-pollution line (engine, catalytic converter and oxygen sensors) have been developed.

The engine is described by a nonlinear discrete-time dynamic model [1], whose inputs are the throttle position, the ignition advance and the fuel injection duration, and whose states and outputs are the engine speed, the intake manifold pressure, and the intake normalized air-fuel ratio. A PI regulator, based on the front oxygen sensor, makes the air-fuel ratio oscillate.

⁵Here we refer to a work which was supported by Renault, Direction de la Recherche, Rueil-Malmaison, France.

The catalytic converter is described by a nonlinear continuous-time state-space model [2], whose inputs are the front pollutants (hydrocarbons, carbon monoxide and nitrous oxides) and the oxygen, and whose states are the rear pollutants, the oxygen and the fraction of oxygen storage capacity filled with O_2 . This model is parameterized with parameter θ_c .

The front oxygen sensor is described by a sigmoid function fed with a first order transfer function, with the normalized air-fuel ratio as input and the sensor voltage as output. This model is parameterized with parameter θ_f .

Consequently, the two-dimensional (lambda sensors) measurements Y can be modeled with the aid of state-space model of the form :

$$\begin{cases} \dot{X} &= f(\theta, X, U) \\ Y &= h(\theta, X, U) \end{cases}$$

where functions f and h are strongly nonlinear, the parameter vector is : $\theta = \begin{pmatrix} \theta_c \\ \theta_f \end{pmatrix}$, and where, because of the closed-loop shown on figure 6, the input U depends on θ .

3 Residual generation and problem reduction

We have argued that the core of the statistical approach to FDI is problem reduction : a large class of FDI problems can be reduced to the universal static problem of monitoring the mean value of a Gaussian vector. Such a problem reduction is achieved with the aid of a convenient residual generation, which we explain further now. The key tool is the local approach. This approach has a wide scope of interest, since it can encompass all the types of FDI problems, for sensors, actuators and components faults. However, for additive faults, a non-asymptotic reasoning suffices, even though the local approach also applies in that case. 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.

Thus, in this section, we first briefly recall this elementary argument for additive faults. Then, we give a sketchy description of the local approach to problem reduction, for which further details can be found in section 4. Finally, we state the reduced FDI problems, for which solutions can be found in section 5.

In this and the following sections, we consider samples $(Y_k)_k$ of stochastic processes belonging to the class of controlled semi-Markov processes. These processes are generated with the aid of a finite dimensional state space, and can be thought of as the nonlinear counterpart of ARMA processes. We refer the reader to [18] for precise definitions and properties. For data samples with finite size ρ , we use the following notations :

$$\mathcal{Y}_{k,\rho}^- \triangleq \begin{pmatrix} Y_k \\ Y_{k-1} \\ \vdots \\ Y_{k-\rho+1} \end{pmatrix}, \quad \mathcal{Y}_{k,\rho}^+ \triangleq \begin{pmatrix} Y_k \\ Y_{k+1} \\ \vdots \\ Y_{k+\rho-1} \end{pmatrix} \quad (11)$$

A N -size data sample is denoted by : $\mathcal{Y}_N \triangleq \mathcal{Y}_{N,N}^-$. We stress that the dimension of vector \mathcal{Y}_k increases with time index k .

3.1 Additive faults : repeated equations

The argument for reducing the FDI problem for additive faults to the problem of monitoring the mean of a Gaussian vector is the following. The system is assumed to be perturbed by white

Gaussian noises. The argument, 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 .

To begin with, we consider the case of observable linear dynamic systems subject to the first and second types of faults. These faults act additively on the system, which we write :

$$\begin{cases} X_{k+1} = FX_k + GU_k + \Gamma \Upsilon_x + W_k \\ Y_k = HX_k + JU_k + \Xi \Upsilon_y + V_k \end{cases} \quad (12)$$

where the state X , the controlled input U , and the output Y have dimensions n , m , r , respectively, 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, where Υ_x and Υ_y are the assumed additive faults, of dimensions l_x and l_y , respectively, and the fault gains Γ and Ξ are matrices of dimensions $n \times l_x$ and $r \times l_y$, respectively.

It has been argued in [46] that, introducing the input-adjusted outputs :

$$\underline{y}_{k,\varrho} = \mathcal{Y}_{k,\varrho}^+ - \mathcal{M}_\varrho(G, J) U_{k,\varrho}^+$$

and the noise :

$$\underline{v}_{k,\varrho} = \mathcal{M}_\varrho(I_n, 0) \mathcal{W}_{k,\varrho}^+ + \mathcal{V}_{k,\varrho}^+$$

where $\mathcal{M}_\varrho(G, J)$ is the lower triangular block-Toeplitz matrix associated with the impulse response of the system, we can write the set of ϱ successive equations (12) as the *regression model* :

$$\zeta_{k,\varrho}(\Upsilon_x, \Upsilon_y) \triangleq \underline{y}_{k,\varrho} = \mathcal{O}_\varrho X_k + \underline{\mathcal{M}}_\varrho(\Gamma) \Upsilon_x + (\mathbf{1}_\varrho \otimes \Xi) \Upsilon_y + \underline{v}_{k,\varrho} \quad (13)$$

where \mathcal{O}_ϱ is observability matrix of order $(\varrho - 1)$ and $\underline{\mathcal{M}}_\varrho(\Gamma)$ is a matrix function which depends on Γ in a known fashion.

Under convenient assumptions, regression model (13) has three basic properties. First, the (unknown) state vector X_k is *independent* of noise $\underline{v}_{k,\varrho}$. Second, the failure gain matrices $\underline{\mathcal{M}}_\varrho(\Gamma)$ and $(\mathbf{1}_\varrho \otimes \Xi)$ are full column rank (f.c.r.). Third, noise $\underline{v}_{k,\varrho}$ is Gaussian, with a covariance matrix Σ_ϱ which is generally *not block-diagonal* [9, 7]. Therefore, vector ζ in (13) is Gaussian distributed, with covariance matrix Σ_ϱ and mean value $\mathcal{O}_\varrho X_k + M \Upsilon$, which we write as :

$$\zeta_{k,\varrho}(\Upsilon) \sim \mathcal{N}(\mathcal{O}_\varrho X_k + M \Upsilon, \Sigma_\varrho) \quad (14)$$

where :

$$M \triangleq \begin{pmatrix} \underline{\mathcal{M}}_\varrho(\Gamma) & (\mathbf{1}_\varrho \otimes \Xi) \end{pmatrix}, \quad \Upsilon \triangleq \begin{pmatrix} \Upsilon_x \\ \Upsilon_y \end{pmatrix}$$

The FDI problem for additive faults in linear Gaussian dynamical systems is thus reduced to the universal problem of monitoring the mean value of a Gaussian vector. As will become clear below, the problem of monitoring fault Υ in the residual vector $\zeta_{k,\varrho}$ of (14), where X_k is unknown, is identical to the isolation problem 3.2 in the basic Gaussian case, where $X_k = 0$ [7].

The same argument (off-line approach, repeated equations) and assumption (Gaussian noise) apply to the problem of monitoring the first type of faults in static nonlinear systems. But these arguments do not apply to the second type of faults in nonlinear systems.

3.2 Non-additive faults : local approach

The argument for reducing the FDI problem for any type of faults to the problem of monitoring the mean of a Gaussian vector is based on the local approach. This approach is of particular interest for non-additive faults, namely component faults in any system and actuator faults in

nonlinear systems. We introduce here a simplified sketch of its basic concepts and results. Details are given in section 4. Moreover, we refer the reader to [35] for a tutorial introduction to this approach.

As mentioned in section 1, from a statistical point of view, we find it useful to distinguish between primary residuals, which are functions of the model parameter and the observed data, and improved residuals for making the decision. To begin with, we define primary residuals and we discuss the strong connections with estimating functions used for parameter identification. Then we define improved residuals, of which the evaluation is much easier, since they have a Gaussian distribution in both safe and faulty cases.

Primary residuals and estimating functions. Because we consider the wide class of semi-Markov processes, and for reasons which become clear below, for defining primary residuals we need to introduce an auxiliary process Z_k . This process is a finite-dimensional vector-valued output of some causal dynamical system parameterized by vector θ and having as inputs the known or measured data U_k and Y_k , which we write :

$$\begin{cases} \xi_{k+1} &= \varphi(\theta, \xi_k, U_k, Y_k) \\ Z_k &= \vartheta(\theta, \xi_k, U_k, Y_k) \end{cases} \quad (15)$$

As explained below, the auxiliary vector Z_k may be simply composed of a finite size sample of observed outputs : $Z_k = \mathcal{Y}_{k,\varrho}^-$ and the innovation, as in the case of linear AR processes. The nonlinear dynamic system of example 2.3 is an instance of more complex cases where a more sophisticated auxiliary process Z_k is needed. Auxiliary processes relevant for semi-Markov processes are given in sections 4.3 and 8. We stress that the dimension of vector Z_k is constant and does not increase with time index k as the dimension of \mathcal{Y}_k does.

Primary residuals associated with parametric models for the purpose of FDI should satisfy two basic requirements.

Definition 3.1 (Primary residuals.) *A finite-dimensional vector-valued function $K(\theta, Z_k)$ ⁶ is a valid primary residual for monitoring parameter vector θ , if it is differentiable in θ , and if there exists a neighborhood $\omega(\theta_0)$ such that :*

$$\mathbf{E}_\theta K(\theta_0, Z_k) = 0 \text{ if } \theta = \theta_0 \quad (16)$$

$$\mathbf{E}_\theta K(\theta_0, Z_k) \neq 0 \text{ if } \theta \in \omega(\theta_0) \setminus \theta_0, \quad (17)$$

where \mathbf{E}_θ is the expectation when the actual system parameter value is θ . For isolating (subsets of) components of θ , we define the mean deviation :

$$M(\theta_0) \triangleq - \mathbf{E}_{\theta_0} \left. \frac{\partial}{\partial \theta} K(\theta, Z_k) \right|_{\theta=\theta_0} \quad (18)$$

which can be viewed as a Jacobian matrix⁷. An additional requirement for isolation is that its rank should equal the dimension of the parameter space Θ , that is :

$$\text{rank } M(\theta_0) = \dim \Theta = l \quad (19)$$

Some comments are in order about the requirements (16)-(17) and (19).

⁶It should be clear that there are several possible pairs (Z, K) which provide us with the same estimating function $K(\theta, Z)$. The Gaussian scalar AR process in section 4.1 is a simple instance of this statement.

⁷It can also be viewed as a Hessian matrix, when function $K(\theta, Z_k)$ is itself a gradient, w.r.t. θ . This is further discussed in section 4.

First, the empirical counterpart of (16), that is :

$$\frac{1}{N} \sum_{k=1}^N K(\theta, Z_k) = 0 \quad (20)$$

writes as an estimating equation : an estimate $\hat{\theta}$ is found as a root of this equation in θ . There is thus a tight connection between primary residuals and M -estimates [37] and estimating [32] or inference [48] functions. This issue is discussed further in section 4.

Second, in statements (16)-(17) and definition (18), it is assumed that expectation \mathbf{E}_θ exists and does not depend on time index k , at least for large k . In other words, the process $K(\theta, Z_k)$ is assumed to be asymptotically stationary. Moreover, requirement (16)-(17) is exactly a local identifiability assumption.

Third, for condition (19) to hold, vector-valued function K should have a dimension greater than or equal to the parameter dimension l .

As discussed further in section 4, a typical example of such a function K is the gradient, w.r.t. θ , of the squared LS-prediction error, and a typical *counter-example* of function K is the innovation.

Improved residuals : local approach. We now turn to residual evaluation. The statistical approach to residual evaluation requires the knowledge of the statistical properties of the FDI residuals. Unfortunately, even for simple linear models and classical estimating functions, the distribution of $K(\theta, Z_k)$ is often unknown. One manner to circumvent this difficulty is to assume close hypotheses :

$$(\textit{Safe system}) \mathbf{H}_0 : \theta = \theta_0 \quad \text{and} \quad (\textit{Faulty system}) \mathbf{H}_1 : \theta = \theta_0 + \frac{\Upsilon}{\sqrt{N}} \quad (21)$$

where vector Υ is unknown, but fixed. Note that for large N , hypothesis \mathbf{H}_1 in (21) corresponds to *small* deviations in θ .

Definition 3.2 (Improved residuals.) *Given a primary residual $K(\theta, Z)$ and a N -size sample of data, the improved residual is defined as :*

$$\zeta_N(\theta) \triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^N K(\theta, Z_k) \quad (22)$$

As formally stated in section 4, for primary residuals K satisfying the identifiability assumption (16)-(17) and regular enough, the improved residual ζ_N turns out to be asymptotically Gaussian distributed, with mean zero when the system is safe (hypothesis \mathbf{H}_0), with mean $M \Upsilon$ when the system failed (hypothesis \mathbf{H}_1), and with the same covariance matrix, under both \mathbf{H}_0 and \mathbf{H}_1 , making easy its evaluation. Actually, as explained in section 5, the resulting test is a quadratic form in the improved residual ζ_N .

It is one of the purposes of this paper to outline some examples of estimating functions which in practice lead to powerful detection algorithms when used as primary residuals and plugged in the local approach. Indeed, it should be clear that, thanks to the above asymptotic Gaussianity, which transforms any detection problem concerning a parameterized stochastic process into the problem of monitoring the mean of a Gaussian vector, the only conceptual task which remains to be done for solving a given detection problem is to select a relevant estimating function which, of course, we'd like to call a *monitoring function*.

3.3 Reduced FDI problem

Based on measured data \mathcal{Y} and a model parameterized by vector θ , we have sketched how to build a vector ζ , defined in (22) (resp. (13)). Its asymptotic distribution (resp. its distribution in (14)) is Gaussian, with the same covariance matrix Σ under both hypotheses, with mean zero under \mathbf{H}_0 and $M \Upsilon$ under \mathbf{H}_1 , which we summarize by :

$$\zeta \sim \mathcal{N}(M \Upsilon, \Sigma)$$

Consequently, the reduced detection and isolation problems can be stated as follows. We partition the fault vector as : $\Upsilon = \begin{pmatrix} \Upsilon_a \\ \Upsilon_b \end{pmatrix}$, with Υ_a, Υ_b of known dimensions⁸ l_a, l_b respectively ($l_a + l_b = l$), and the incidence matrix M accordingly : $M = \begin{pmatrix} M_a & M_b \end{pmatrix}$. Matrix M is assumed to be full column rank (f.c.r.).

Problem 3.1 (Detection.) *Decide between $\Upsilon = 0$ and $\Upsilon \neq 0$.*

Problem 3.2 (Isolation.) *Decide between $\Upsilon_a = 0$ and $\Upsilon_a \neq 0$, with Υ_b being an unknown vector with known dimension.*

4 Local approach to problem reduction

In this section, we go further into the details of the local approach to problem reduction and to residual generation and evaluation.

We have outlined above the tight connection between primary residuals and estimating functions. Moreover, we have stated in (21) that the local approach consists in testing between hypotheses which are close to each other. This is quite well in correspondence with the practical requirement of monitoring small deviations. On the other hand, it is well known that a useful tool for analyzing the effect of small perturbations on a given function is Taylor's expansion of this function. As a matter of fact, first and second order Taylor's expansions of estimating functions are the key tools for understanding the essential results of the local approach. Therefore, in this paper, we focus on these expansions and some of their consequences, referring the reader to more technical references for the precise derivation and proofs of the asymptotic convergence results. We discuss first the log-likelihood function, and then we consider alternative estimating functions.

4.1 Local approach : log-likelihood function

We consider stationary processes with finite and constant size ϱ of dependence range, and we note $p_\theta(Y_k | \mathcal{Y}_{k-1})$ the conditional probability distribution of an observation Y_k given past data.

Basic concepts. To begin with, we recall the three basic concepts of the likelihood theory : the log-likelihood function, the efficient score and Fisher information matrix.

The *log-likelihood function* of a N -size sample \mathcal{Y}_N is :

$$\ln p_\theta(\mathcal{Y}_N) \triangleq \sum_{k=1}^N \ln p_\theta(Y_k | \mathcal{Y}_{k-1})$$

⁸The important issue of inferring the (generally unknown) number of failed components in Υ is beyond the scope of this paper. The interested reader is referred to [11, 47] for a discussion of and a possible solution to this problem.

The *efficient score* is the normalized gradient, w.r.t. θ , of this functional :

$$\zeta_N(\theta_0) \triangleq \frac{1}{\sqrt{N}} \left. \frac{\partial \ln p_\theta(\mathcal{Y}_N)}{\partial \theta} \right|_{\theta=\theta_0} = \frac{1}{\sqrt{N}} \sum_{k=1}^N \left. \frac{\partial \ln p_\theta(Y_k | \mathcal{Y}_{k-1})}{\partial \theta} \right|_{\theta=\theta_0} \quad (23)$$

The maximum likelihood estimate (MLE) is based on the maximization of the log-likelihood function. The ML estimating function is the sum over k of the terms :

$$\frac{\partial}{\partial \theta} \ln p_\theta(Y_k | \mathcal{Y}_{k-1}) \quad (24)$$

How this writes under the form of an estimating function $K(\theta, Z_k)$ with an auxiliary vector Z_k as in (15), depends on the particular form of the conditional distribution $p_\theta(Y_k | \mathcal{Y}_{k-1})$ and is out of the scope of the present paper. However, the efficient score (23) can be viewed as the improved residual (22) associated with the ML estimating function (24). It satisfies the local identifiability assumption (16)-(17).

Fisher information matrix is the limit of the covariance matrix of the efficient score :

$$\mathbf{I}(\theta) \triangleq \lim_{N \rightarrow \infty} \mathbf{I}_N(\theta), \quad \mathbf{I}_N(\theta) \triangleq \mathbf{E}_\theta \left(\zeta_N(\theta) \zeta_N^T(\theta) \right) \quad (25)$$

and can easily be shown to write also as the expectation of the second order derivative :

$$\mathbf{I}_N(\theta) = - \frac{1}{N} \mathbf{E}_\theta \frac{\partial^2 \ln p_\theta(\mathcal{Y}_N)}{\partial \theta^2} \quad (26)$$

As explained below, a key property of the efficient score is that it contains (asymptotically) the same amount of information about the parameter vector θ as the original data. We stress that, in the case of non-additive faults and in particular for component faults, the efficient score is *different* from the innovation⁹. A simple illustration of this fundamental fact is provided by the case of a Gaussian scalar AR process : $Y_k = \sum_{i=1}^p a_i Y_{k-i} + E_k$, in which the AR parameter vector $\theta^T = (a_1 \dots a_p)$ is subject to changes, whereas the variance σ^2 of the input E is assumed known and constant. In that case, directly from the definition, using the innovation $\varepsilon(\theta)$ and the auxiliary vector :

$$Z_k \triangleq \begin{pmatrix} \mathcal{Y}_{k-1,p}^- \\ \varepsilon_k(\theta) \end{pmatrix}, \quad (27)$$

the efficient score is found [3] to write as in (22) with¹⁰ :

$$K(\theta, Z_k) = \mathcal{Y}_{k-1,p}^- \varepsilon_k(\theta) \quad (28)$$

Therefore, the efficient score ζ is a vector-valued function, whereas the innovation ε is a scalar function, which certainly contains much less information about the parameter vector θ .

Taylor expansion of the LLR and the efficient score. As stated in (21), we wish to detect small deviations w.r.t. the null hypothesis $\theta = \theta_0$ (reference parameter value), and we express small deviations as local alternative hypotheses, namely : $\theta = \theta_0 + \frac{\Upsilon}{\sqrt{N}}$. On the other hand, the basic detection concept is the log-likelihood ratio (LLR) [42]. Under mild conditions, the log-likelihood ratio of a N -size data sample admits the following second order Taylor expansion :

$$S_N(\theta_0, \theta) \triangleq \ln \frac{p_\theta(\mathcal{Y}_N)}{p_{\theta_0}(\mathcal{Y}_N)} \approx \Upsilon^T \zeta_N(\theta_0) - \frac{1}{2} \Upsilon^T \mathbf{I}(\theta_0) \Upsilon$$

⁹As a consequence, none of the three industrial monitoring problems in section 2 can be addressed with the aid of an innovation.

¹⁰The same estimating function can be written with the aid of $\mathcal{Y}_{k,p+1}$ as the auxiliary vector.

From this expansion, it is easy to deduce the first and second order moments of the LLR S_N under both the null and the local alternative hypotheses :

$$\begin{aligned} \mathbf{E}_{\theta_0} S_N &\approx -\frac{1}{2} \Upsilon^T \mathbf{I}(\theta_0) \Upsilon \\ \mathbf{E}_{\theta} S_N &\approx +\frac{1}{2} \Upsilon^T \mathbf{I}(\theta_0) \Upsilon \approx -\mathbf{E}_{\theta_0} S_N \\ \text{cov}_{\theta_0} S_N &\approx \Upsilon^T \mathbf{I}(\theta_0) \Upsilon \approx \text{cov}_{\theta} S_N \end{aligned}$$

where \mathbf{E}_{θ} and cov_{θ} are the expectation and the covariance, respectively, when the actual system parameter value is θ .

We have recalled the above expansion of the LLR mainly for historical reasons [41]. But the fact which generalizes to other estimating functions is the first order expansion of the efficient score (23), which writes :

$$\zeta_N(\theta) \approx \zeta_N(\theta_0) + \frac{1}{N} \left. \frac{\partial^2 \ln p_{\theta}(\mathcal{Y}_N)}{\partial \theta^2} \right|_{\theta=\theta_0} \Upsilon$$

Of course, there is nothing new here, since the score is the first order derivative of the likelihood function. Computing the expectation of both sides and using (26), we get :

$$\mathbf{E}_{\theta_0} \zeta_N(\theta) \approx -\mathbf{I}(\theta_0) \Upsilon$$

Note that this expression is coherent with the ML identifiability assumption.

Asymptotic Gaussianity of the LLR and the score. These elementary facts are now put into more formal statements. For a large class of parameterized conditional probability distributions¹¹ \mathbf{P}_{θ} , the above informal approximations to the first two moments of the LLR and the efficient score can be formally stated (and extended) as the following central limit theorems (CLT) [41, 58] :

$$\begin{aligned} S_N(\theta_0, \theta) &\rightarrow \begin{cases} \mathcal{N}\left(-\frac{1}{2} \Upsilon^T \mathbf{I}(\theta_0) \Upsilon, \Upsilon^T \mathbf{I}(\theta_0) \Upsilon\right) & \text{under } \mathbf{P}_{\theta_0} \\ \mathcal{N}\left(+\frac{1}{2} \Upsilon^T \mathbf{I}(\theta_0) \Upsilon, \Upsilon^T \mathbf{I}(\theta_0) \Upsilon\right) & \text{under } \mathbf{P}_{\theta_0 + \frac{\Upsilon}{\sqrt{N}}} \end{cases} \\ \zeta_N(\theta_0) &\rightarrow \begin{cases} \mathcal{N}\left(0, \mathbf{I}(\theta_0)\right) & \text{under } \mathbf{P}_{\theta_0} \\ \mathcal{N}\left(\mathbf{I}(\theta_0) \Upsilon, \mathbf{I}(\theta_0)\right) & \text{under } \mathbf{P}_{\theta_0 + \frac{\Upsilon}{\sqrt{N}}} \end{cases} \end{aligned}$$

These theorems mean that a small deviation in the parameter vector is reflected into a change in the sign of the mean value of the LLR vector S_N , and into a change in the mean value of the efficient score ζ_N which switches from zero to a non-zero value. Both vectors are asymptotically Gaussian distributed with equal covariance matrices under both null and local alternative hypotheses. An interesting byproduct of this result is that the efficient score ζ_N is asymptotically a sufficient statistics, that is it contains the same information about the parameter vector as the original data, which the innovation generally does not do.

Moreover, the change magnitude estimate $\hat{\Upsilon}$ is also asymptotically Gaussian distributed, and the optimized LLR : $2S_N\left(\theta_0, \theta_0 + \frac{\hat{\Upsilon}}{\sqrt{N}}\right)$, which turns out to be asymptotically a quadratic form in the score vector : $\zeta_N^T(\theta_0) \mathbf{I}_N^{-1}(\theta_0) \zeta_N(\theta_0)$, is asymptotically distributed as a χ^2 variable with $l(= \dim \Theta)$ degrees of freedom and, under $\mathbf{P}_{\theta_0 + \frac{\Upsilon}{\sqrt{N}}}$, non-centrality parameter : $\Upsilon^T \mathbf{I}(\theta) \Upsilon$.

¹¹This class is called the class of locally asymptotically normal (LAN) families. The main examples of LAN families are i.i.d. sequences, stationary Gaussian processes, and stationary Markov processes.

Design of optimum tests. The main use of the above CLT concerns the design of asymptotically optimum tests between composite (multi-valued) hypotheses. Actually, what the local approach basically states is, first, an equivalence between two hypotheses testing problems and, second, an equivalence between several possible solutions to these problems, among which one is computationally cheap.

Local approach states an asymptotic equivalence between the two following testing problems. The first one is the original detection problem, and concerns a N -size sample \mathcal{Y}_N of a process in one of the above mentioned LAN families, and the two hypotheses :

$$\mathbf{H}_i = \{\theta \in \Theta_i\}, \quad \Theta_i = \theta_0 + \frac{\Gamma_i}{\sqrt{N}} \quad (i = 0, 1)$$

The second one concerns a single observation of a random variable with Gaussian distribution $\mathcal{N}(\Upsilon, \mathbf{I}^{-1}(\theta_0))$, and the hypotheses $\mathbf{H}_i = \{\Upsilon \in \Gamma_i\}$ ($i = 0, 1$). We stress that the equivalence between these two problems is quite strong for two reasons. First, it means that whatever the probability distribution of the original data, local approach leads you to end up with a Gaussian distribution. Second, it also means that, whatever the level of dependence in the original data, local approach leads you to end up with an instantaneous random vector, and no longer with a time-dependent random process.

Local approach also states an asymptotic equivalence between several possible solutions to the original hypotheses testing problem. Assuming ellipsoidal shapes¹² for the hypotheses :

$$\Theta_0 = \left\{ \theta : (\theta - \theta_0)^T \mathbf{I}(\theta_0) (\theta - \theta_0) \leq \frac{a^2}{N} \right\}, \quad \Theta_1 = \left\{ \theta : (\theta - \theta_0)^T \mathbf{I}(\theta_0) (\theta - \theta_0) \geq \frac{b^2}{N} \right\},$$

the following four test statistics are asymptotically equivalent, and the tests are asymptotically optimum (uniformly most powerful) :

1. The GLR (generalized log-likelihood ratio) :

$$\ln \frac{\max_{\theta \in \Theta_1} p_{\theta}(\mathcal{Y}_N)}{\max_{\theta \in \Theta_0} p_{\theta}(\mathcal{Y}_N)}$$

2. The optimized LLR :

$$\ln \frac{p_{\hat{\theta}_N}(\mathcal{Y}_N)}{p_{\theta_0}(\mathcal{Y}_N)}$$

3. The quadratic form¹³ in the parameter estimation error :

$$N (\hat{\theta}_N - \theta_0)^T \mathbf{I}_N(\theta_0) (\hat{\theta}_N - \theta_0)$$

4. The quadratic form in the efficient score :

$$\zeta_N^T(\theta_0) \mathbf{I}_N^{-1}(\theta_0) \zeta_N(\theta_0) \quad (29)$$

We stress that, from a computational point of view, the last of these four tests is the most interesting. It requires neither the computation and optimization of the log-likelihood ratio as the first one does, nor the re-identification of the parameter vector as the second and third do. It only involves the computation of a quadratic form, in terms of the function of the reference parameter vector and the data which is the efficient score. Such a quadratic test may very well be computed on-board. This is one of the key interests of the local approach.

¹²Simple arguments favoring the definition of the metric in terms of Fisher information matrix can be found in [9].

¹³The reason why the GLR test for monitoring the mean of a Gaussian vector is a quadratic form in that vector is recalled in section 5.

4.2 Local approach - Other estimating/monitoring functions

It turns out that similar CLT results hold for large classes of dynamical processes, and for estimating functions $K(\theta, Z_k)$ different from the efficient score [4, 5, 17, 18]. As an example of such an alternative function, let $G(\theta, Z_k)$ be a contrast or quasi-likelihood [36] which minimization provides us with another estimate of θ :

$$\hat{\theta}_N = \arg \min_{\theta} \sum_{k=1}^N G(\theta, Z_k) \quad (30)$$

Then the estimating equation has the form (20) with the estimating function : $K \triangleq \partial G / \partial \theta$. We stress however that some estimates are obtained by solving estimating equations of the form (20) where K is *not* the gradient of any function G . Typical examples are the IV and subspace identification methods, which we discuss in section 6.

Let us thus consider an estimating function $K(\theta, Z_k)$, which we assume to be asymptotically stationary and to satisfy the local identifiability assumption (16)-(17). We define the improved residual or quasi-score ζ_N as in (22) :

$$\zeta_N(\theta_0) = \frac{1}{\sqrt{N}} \sum_{k=1}^N K(\theta_0, Z_k) \quad (31)$$

We introduce the mean deviation (18) :

$$M(\theta_0) \triangleq - \mathbf{E}_{\theta_0} \left. \frac{\partial}{\partial \theta} K(\theta, Z_k) \right|_{\theta=\theta_0} \quad (32)$$

and the covariance matrix :

$$\Sigma(\theta_0) \triangleq \lim_{N \rightarrow \infty} \Sigma_N(\theta_0), \quad \Sigma_N(\theta_0) \triangleq \mathbf{E}_{\theta_0} \left(\zeta_N(\theta_0) \zeta_N^T(\theta_0) \right) = \frac{1}{N} \sum_{k=1}^N \sum_{j=1}^N \mathbf{E}_{\theta_0} \left(K(\theta_0, Z_k) K^T(\theta_0, Z_j) \right) \quad (33)$$

It is of interest to note that, because of (16), Jacobian matrix (32) also writes :

$$\begin{aligned} M(\theta_0) &= - \left. \frac{\partial}{\partial \theta} \mathbf{E}_{\theta_0} K(\theta, Z_k) \right|_{\theta=\theta_0} \\ &= + \left. \frac{\partial}{\partial \theta} \mathbf{E}_{\theta} K(\theta, Z_k) \right|_{\theta=\theta_0} \end{aligned} \quad (34)$$

In some cases, the expression (34) may be more appropriate for deriving necessary and sufficient conditions for isolation requirement (19) to hold [10].

First order Taylor expansion of the quasi-scores. We now play with this improved residual the same game as before with the efficient score, namely first order Taylor expansion. We consider again local alternative hypothesis as in (21), and we write the first order expansion of the quasi-score ζ_N in (31) as :

$$\zeta_N(\theta) \approx \zeta_N(\theta_0) + \underbrace{\sqrt{N} \frac{1}{N} \left(\sum_{k=1}^N \left. \frac{\partial}{\partial \theta} K(\theta, Z_k) \right|_{\theta=\theta_0} \right)}_{\frac{\Upsilon}{\sqrt{N}}} \quad (35)$$

By the CLT under p_{θ_0} , the first term of the right hand-side is seen to be asymptotically Gaussian distributed with mean 0 and covariance matrix $\Sigma(\theta_0)$. By the law of large numbers under p_{θ_0} , the under-braced part of the second term of the right hand-side is seen to converge towards $\left(\mathbf{E}_{\theta_0} \left. \frac{\partial}{\partial \theta} K(\theta, Z_k) \right|_{\theta=\theta_0} \right)$ which is exactly what we defined as $(-M(\theta_0))$. From this can be shown the asymptotic Gaussianity of the quasi-scores under both null and local alternative hypotheses.

Asymptotic Gaussianity of the quasi-scores. As for the efficient score, from the above Taylor expansion and provided that $\Sigma(\theta_0)$ is positive definite, the following CLT can be shown to hold [48, 36, 18, 22] :

$$\zeta_N(\theta_0) \rightarrow \begin{cases} \mathcal{N}(0, \Sigma(\theta_0)) & \text{under } \mathbf{P}_{\theta_0} \\ \mathcal{N}(M(\theta_0) \Upsilon, \Sigma(\theta_0)) & \text{under } \mathbf{P}_{\theta_0 + \frac{\Upsilon}{\sqrt{N}}} \end{cases}$$

Again, this theorem means that a small deviation in the parameter vector is reflected into a change in the mean value of the quasi-score vector ζ_N , which switches from zero to a non-zero value. Also, this quasi-score is asymptotically Gaussian distributed with equal covariance matrices under both null and local alternative hypotheses.

Before investigating the design of the tests based on quasi-scores, we briefly describe another interesting byproduct of the above Taylor expansion.

Local approach and estimation efficiency. As extensively discussed in [36, 22], the local approach for a general estimating function $K(\theta, Z_k)$ is closely related with the CLT providing us with the accuracy of an identification method based on K . A sketchy argument is as follows. Noting that the solution of the estimating equation (20) based on K is characterized by $\zeta_N(\hat{\theta}_N) = 0$, from the Taylor expansion (35), we get :

$$\sqrt{N} (\hat{\theta}_N - \theta_0) \approx - \left(\mathbf{E}_{\theta_0} \left. \frac{\partial}{\partial \theta} K(\theta, Z_k) \right|_{\theta=\theta_0} \right)^{-1} \zeta_N(\theta_0) = M(\theta_0)^{-1} \zeta_N(\theta_0)$$

This relates the distribution of the estimation error to the Gaussian distribution of the score $\zeta_N(\theta)$ given above. In particular, it is found that the covariance of the parameter estimation error is :

$$\text{cov}(\hat{\theta}_N - \theta_0) = \left(M^T(\theta_0) \Sigma^{-1}(\theta_0) M(\theta_0) \right)^{-1}$$

Using empirical means of cross-products of residual ζ_N , this provides us with an estimate of the covariance of the parameter estimation error, which may happen to be easier to compute than those resulting from other theoretical formulas.

Design of optimum tests. As for the log-likelihood function, the main use of the above CLT concerns the design of asymptotically optimum tests between composite (multi-valued) hypotheses [4, 17]. When Jacobian matrix M in (32) is also assumed to be f.c.r., the quasi-score-based test between the hypotheses \mathbf{H}_0 and \mathbf{H}_1 defined in (21) is :

$$\zeta_N^T \Sigma_N^{-1} M \left(M^T \Sigma_N^{-1} M \right)^{-1} M^T \Sigma_N^{-1} \zeta_N \quad (36)$$

where the dependence on θ_0 has been removed for simplicity. This is the counterpart of the ML-based test (29) based on the efficient score (23). Its non-centrality parameter is : $\Upsilon^T M^T \Sigma^{-1} M \Upsilon$. As far as actual computation is concerned, we only mention that the estimation of $M(\theta)$ in (32) from a data sample is obtained, replacing the expectation by sample average. The estimation of Σ is more tricky. More details on that issue can be found in [66].

When M is square and invertible, test (36) boils down to : $\zeta_N^T \Sigma_N^{-1} \zeta_N$.

We stress that this local approach, based on alternative estimating functions, reduces to the classical likelihood-based one only in particular cases. More precisely, test (36) is identical to (29) when the estimating function K results from the minimization of the log-likelihood functional $(-\mathbf{E}_\theta \ln p_\theta)$ through a stochastic gradient (or stochastic Newton) algorithm for a Gaussian process. As shown in [17, 9], the ELS and IV identification methods provide us with detectors which are *not* equivalent to the ML-based one.

Invariance property of the local tests. This type of χ^2 -tests enjoy a simple but basic invariance property, which is of particular importance for subspace-based FDI algorithms as we discuss further in section 6.

Lemma 4.1 *The χ^2 -test in (36) is left unchanged when the estimating function K in (31) is pre-multiplied by an invertible matrix gain.*

4.3 Monitoring functions : which ones ?

Some comments are in order about the choice of a monitoring (estimating) function.

First, as already mentioned in the introduction, the basic statistical approach to residual generation consists in deriving sufficient statistics. In the case of additive faults on linear dynamic systems, the main example of such a lossless residual is Kalman filter innovation [33]. But it should be clear from the above results that the efficient score is (asymptotically) a sufficient statistics, and the innovation generally not. The innovation is a relevant monitoring function only in few cases, that are sensor faults in any system, and actuator faults in linear systems. For component faults in any system, and also for actuator faults in nonlinear systems, it is no longer true that the innovation is a lossless statistics. We remind the reader with the simple example of a scalar Gaussian AR process mentioned above, and the comments made on the requirements for valid primary residuals in section 3. The same statement applies to observer-based state estimation errors [9][chap.6]. Thus innovations and estimation errors are not convenient as primary residuals. The methods proposed in [49] are thus of interest only for the first type of faults in any system, and the second type of faults in linear systems. In this section, we have explained that detection problems (testing between two hypotheses) should be addressed using the likelihood ratio, or its first order expansion, and not a single likelihood. Remembering that Kalman filter provides us with the log-likelihood function as a quadratic form in the innovation, it is easy to check that the only case where the log-likelihood ratio boils down to a linear function of the innovation¹⁴ is the case of a change in the mean value of the observed output data. Moreover, in the case of a component fault, the log-likelihood ratio is a function of *two* innovations generated by different filters.

Second, when the likelihood function is too complex, a typical example of alternative monitoring function is the gradient of the squared LS-*prediction error* :

$$-\frac{1}{2} \frac{\partial}{\partial \theta} \left(\epsilon_k^T(\theta) \epsilon_k(\theta) \right)$$

where the prediction error is :

$$\epsilon_k(\theta) \triangleq Y_k - \hat{Y}_{k|k-1}(\theta), \quad (37)$$

and where $\hat{Y}_{k|k-1}(\theta)$ is a one-step ahead prediction of the output data, computed on the basis of the considered parameterized model. Note that the requirement that process $\epsilon_k(\theta)$ should be asymptotically stationary by no means implies that neither Y_k nor $\hat{Y}_{k|k-1}(\theta)$ should be so too. In the sequel, we refer to this function as the LS-score. Based on the auxiliary vector :

$$Z_k \triangleq \begin{pmatrix} \text{vec} \frac{\partial \hat{Y}_{k|k-1}(\theta)}{\partial \theta} \\ \epsilon_k(\theta) \end{pmatrix} \quad (38)$$

where vec is the column stacking operator, the LS-score also writes :

$$K(\theta, Z_k) = \left(\frac{\partial \hat{Y}_{k|k-1}(\theta)}{\partial \theta} \right)^T \epsilon_k(\theta) \quad (39)$$

¹⁴As in any matched filter, this linear function is nothing but the correlation between the innovation and the fault signature [64].

Note that, in the scalar Gaussian AR case and with the predictor given by the mean of the conditional distribution of the data, we recover exactly (27)-(28).

Third, assuming that the estimating function K is itself the gradient of a function, we stress that the assumptions on the Hessian matrix M in (32), which are necessary for computing the above χ^2 -test, are much weaker than those required for its stability (stated e.g. in [44, 18]). This means that a given estimating function may very well be of poor help for designing efficient recursive identification algorithms, but be of great interest, when plugged in the local approach, for designing efficient (possibly on-line) FDI algorithms. An illustrative example of this statement is the LS score in the gas turbine example in section 7.

Finally, we note that the identification algorithms based on linear algebraic decompositions (eigen analysis, SVD, QR) are usually not written under the form of an estimating equation (20) to be solved (or a contrast function (30) to be minimized), and do not seem to fit the local approach at a first glance. However, this apparent obstacle can be circumvented using their subspace interpretations written as (possibly constrained) estimating equations. An example of this is the subspace-based vibration monitoring algorithm described in section 6.

5 Residual evaluation : monitoring the mean of a Gaussian vector

This section contains several basic elementary results on likelihood ratios for testing between two (simple or composite) hypotheses concerning the mean value of a Gaussian vector with known covariance matrix. Because of the problem reduction outlined in sections 3 and 4, these tests are of universal interest for solving detection and isolation problems in any parameterized process.

5.1 Fault detection : likelihood ratio

Let notation $p_{\Upsilon}(\zeta)$ stand for probability density of a Gaussian vector $\zeta \sim \mathcal{N}(M\Upsilon, \Sigma)$ with mean $M\Upsilon$, M f.c.r., and covariance matrix $\Sigma = \bar{\Sigma} \bar{\Sigma}^T$. The log-likelihood function is (up to a constant) given by :

$$l_{\Upsilon}(\zeta) \triangleq -2 \ln p_{\Upsilon}(\zeta) = (\zeta - M\Upsilon)^T \Sigma^{-1} (\zeta - M\Upsilon)$$

Let $\mathbf{I} \triangleq M^T \Sigma^{-1} M$ be the corresponding Fisher information matrix.

Lemma 5.1 *For deciding between $\Upsilon = 0$ and $\Upsilon \neq 0$, the optimum test statistics t is the generalized log-likelihood ratio (GLR) :*

$$t \triangleq 2 \ln \frac{\max_{\Upsilon} p_{\Upsilon}(\zeta)}{p_0(\zeta)} = \zeta^T \Sigma^{-1} M \mathbf{I}^{-1} M^T \Sigma^{-1} \zeta$$

Under both hypotheses, test statistics t is distributed as a χ^2 -random variable with l ($= \dim \Upsilon$) degrees of freedom, and, under $\Upsilon \neq 0$, non-centrality parameter : $\gamma = \Upsilon^T \mathbf{I} \Upsilon$. The power (probability of correct decision) of test t is an increasing function of this quantity.

5.2 Fault isolation : projection or rejection

We now partition the parameter vector and the incidence matrix as done for stating problem 3.2, and Fisher information matrix accordingly, namely :

$$\mathbf{I} = \begin{pmatrix} \mathbf{I}_{aa} & \mathbf{I}_{ab} \\ \mathbf{I}_{ba} & \mathbf{I}_{bb} \end{pmatrix}$$

Let notation \mathbf{I}_a^{*-1} stand for the upper-left term of \mathbf{I}^{-1} , where $\mathbf{I}_a^* = \mathbf{I}_{aa} - \mathbf{I}_{ab} \mathbf{I}_{bb}^{-1} \mathbf{I}_{ba}$, and notation $p_{\Upsilon_a, \Upsilon_b}(\zeta)$ stands for the density of a Gaussian vector $\zeta \sim \mathcal{N}(M_a \Upsilon_a + M_b \Upsilon_b, \Sigma)$.

The problem is to isolate changes either in Υ_a , or in Υ_b , with the other component being unknown and considered as a nuisance parameter.

Statistical projection (sensitivity). A rather intuitive statistical solution to the isolation problem, which can be called sensitivity approach, consists in projecting the deviations in Υ onto the subspace generated by the components Υ_a to be isolated.

Lemma 5.2 *Sensitivity test \tilde{t}_a for monitoring Υ_a is GLR test between $\Upsilon = (0, 0)$ and $\Upsilon = (\Upsilon_a, 0)$, where $\Upsilon_a \neq 0$, namely :*

$$\tilde{t}_a \triangleq 2 \ln \frac{\max_{\Upsilon_a} p_{\Upsilon_a, 0}(\zeta)}{p_{0,0}(\zeta)} = \tilde{\zeta}_a^T \mathbf{I}_{aa}^{-1} \tilde{\zeta}_a$$

where : $\tilde{\zeta}_a \triangleq M_a^T \Sigma^{-1} \zeta$ is nothing but the efficient score w.r.t. Υ_a .

Under both hypotheses, test statistics \tilde{t}_a is distributed as a χ^2 -random variable with l_a degrees of freedom, and, under $\Upsilon_a \neq 0$, non-centrality parameter : $\Upsilon_a^T \mathbf{I}_{aa} \Upsilon_a$.

The isolation of the faulty component Υ_a or Υ_b is then performed according to the largest sensitivity test \tilde{t}_a or \tilde{t}_b .

Statistical rejection (min-max). Another statistical solution to the problem of isolating Υ_a consists in viewing parameter Υ_b as a nuisance, and using an existing method for inferring part of the parameters while ignoring and being robust to the complementary part. This method is called *min-max approach* and consists in replacing the nuisance parameter component Υ_b by its least favorable value, namely the value which minimizes the power of test t , or equivalently the non-centrality parameter γ . Actually, this is equivalent [9] to the GLR method :

$$t_a^* \triangleq 2 \ln \frac{\max_{\Upsilon_a, \Upsilon_b} p_{\Upsilon_a, \Upsilon_b}(\zeta)}{\max_{\Upsilon_b} p_{0, \Upsilon_b}(\zeta)} \quad (40)$$

which consists in replacing the nuisance parameter component Υ_b by its ML estimate.

Lemma 5.3 *Let Υ_b be an unknown nuisance parameter with known dimension. For deciding between $\Upsilon_a = 0$ and $\Upsilon_a \neq 0$, the optimum test is t_a^* in (40), which can be shown to write as :*

$$t_a^* = \zeta_a^{*T} \mathbf{I}_a^{*-1} \zeta_a^*$$

where : $\zeta_a^* \triangleq \tilde{\zeta}_a - \mathbf{I}_{ab} \mathbf{I}_{bb}^{-1} \tilde{\zeta}_b$ is nothing but the effective score ζ_a^* [35] [19][chap.2], which is the residual from regression of the score in Υ_a on the (nuisance) score in Υ_b .

Under both hypotheses, test statistics t_a^* is distributed as a χ^2 -random variable with l_a degrees of freedom and, under $\Upsilon_a \neq 0$, non-centrality parameter : $\Upsilon_a^T \mathbf{I}_a^* \Upsilon_a$.

Note that the above regression on the nuisance parameter is a very natural way of dealing with a nuisance, and actually can be traced back to Neyman [35].

The isolation of the faulty component Υ_a or Υ_b is then performed according to the largest min-max test t_a^* or t_b^* . It turns out that this decision rule is optimal w.r.t. two criteria [61, 47].

Moreover, sensitivity and min-max tests enjoy the following relationship [7].

Lemma 5.4 *Global test writes : $t = t_a^* + \tilde{t}_b$, where the decomposition is orthogonal in the sense that the two terms are uncorrelated, namely : $\text{cov}_0(t_a^*, \tilde{t}_b) = \text{cov}_1(t_a^*, \tilde{t}_b) = 0$ under both safe and faulty hypotheses. This additive decomposition holds for general distributions and more than two faults, whereas the orthogonal decomposition holds in the Gaussian case for only two faults.*

Further details can be found in [7], where the link between deterministic and stochastic approaches to FDI of additive faults in linear systems is shown to be a factorization of a stochastic rejector by a deterministic rejector, and where mixed deterministic/stochastic algorithms are suggested.

6 Back to example 2.1

Based on the results described in section 4, in this section we describe estimating functions which turn out to be useful for the vibration monitoring problem introduced in section 2. Before proceeding, we must state the definition of the parameter vector in that case. Referring to the definition of a mode in (5), the set of the m modes is considered as the (canonical) system parameter θ :

$$\theta \triangleq \begin{pmatrix} \Lambda \\ \text{vec } \Phi \end{pmatrix} \quad (41)$$

In (41), Λ is the vector whose elements are the eigenvalues λ and Φ is the matrix whose columns are the ϕ_λ 's defined in (5). We stress that this parameterization does not depend on the state-space basis [16].

6.1 Input-output monitoring functions

We have shown in section 2 that the vibration monitoring problem can be stated as monitoring the AR part of an ARMA model with non-stationary MA part. But it is well known that AR and MA parts are coupled in the likelihood, since Fisher information matrix is not block-diagonal [3]. Thus a local test based on the efficient score is of no help for solving this problem. A first solution to vibration monitoring [12] consists in building a local test based on the primary residual associated with this input-output representation and the instrumental variables (IV) identification method. The IV estimating equation can be written as the following system of delayed Yule-Walker relations :

$$\begin{pmatrix} A^T & -I_r \end{pmatrix} \hat{\mathcal{H}}_{p+1,m}^0(N) = 0 \quad (42)$$

where A^T contains the AR (matrix) parameters : $A^T \triangleq (A_p \ \dots \ A_1)$, and where :

$$\hat{\mathcal{H}}_{p+1,m}(N) = \begin{pmatrix} \hat{R}_0 & \hat{R}_1 & \dots & \hat{R}_{m-1} \\ \hat{R}_1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \hat{R}_p & \dots & \dots & \hat{R}_{p+m-1} \end{pmatrix}$$

is the empirical $(p+1) \times m$ block-Hankel matrix filled with covariance matrices of a N -size sample of data, and subscript ⁰ stands for data collected on the safe system. Let A_0^T be the AR parameters estimated by solving this equation (possibly in LS sense when $m > p$). This identification method is known to provide us with consistent estimate of the AR part even in case of non-stationary MA part [16]. Then the corresponding set of modes θ_0 can then be recovered by eigen analysis. We stress that $A(\theta_0) \triangleq A_0$ is characterized by :

$$\begin{pmatrix} A^T(\theta_0) & -I_r \end{pmatrix} \mathcal{O}_{p+1}(\theta_0) = 0 \quad (43)$$

where $\mathcal{O}_{p+1}(\theta_0)$ is the observability matrix written in modal basis [16].

Now, let $\widehat{\mathcal{H}}_{p+1,p}(N)$ the empirical $(p+1) \times p$ block-Hankel matrix filled with covariance matrices of new measurements collected on the (possibly failed) system. The improved residual corresponding to the IV estimating equation is :

$$\zeta_N(\theta_0) \triangleq \sqrt{N} \operatorname{vec} \left(\begin{pmatrix} A^T(\theta_0) & -I_r \end{pmatrix} \widehat{\mathcal{H}}_{p+1,p}(N) \right) \quad (44)$$

Thanks to the CLT of section 4, residual ζ_N is asymptotically Gaussian distributed, and its mean value reflects the possible change in the modal parameters θ_0 . The test statistic is thus a quadratic form in ζ_N .

As the empirical Hankel matrix writes :

$$\widehat{\mathcal{H}}_{p+1,m}(N) = \frac{1}{N} \sum_k \mathcal{Y}_{k,p+1}^+ \mathcal{Y}_{k,m}^{-T} \quad (45)$$

the residual ζ_N also writes :

$$\zeta_N(\theta_0) = \frac{1}{\sqrt{N}} \sum_k \operatorname{vec} \left(\begin{pmatrix} A^T(\theta_0) & -I_r \end{pmatrix} \mathcal{Y}_{k-p,p+1}^+ \mathcal{Y}_{k-p,m}^{-T} \right)$$

Using the auxiliary vector :

$$Z_k \triangleq \begin{pmatrix} \mathcal{Y}_{k-p,m}^- \\ V_k(\theta_0) \end{pmatrix} \quad (46)$$

where $V_k(\theta_0)$ is the MA part of the observed process Y_k :

$$V_k(\theta_0) \triangleq \begin{pmatrix} I_r & -A^T(\theta_0) \end{pmatrix} \mathcal{Y}_{k-p,p+1}^+ \quad (47)$$

the above improved residual ζ_N is found in the form (22), with estimating function K defined as :

$$K(\theta_0, Z_k) = \mathcal{Y}_{k-p,m}^- \otimes V_k(\theta_0) \quad (48)$$

6.2 Subspace-based monitoring functions

We have also shown in section 2 that the vibration monitoring problem can be equally stated as monitoring the eigenstructure of a state-space system with non-stationary state noise. A second solution to vibration monitoring thus consists in building a local test based on a primary residual associated with a subspace-based identification method. As explained in section 4, the key idea is to use the subspace interpretation of SVD in order to write the associated estimating equation.

Actually, the general form of output-only and covariance-driven subspace-based identification algorithms can be written as [62] :

$$\text{SVD of } W_1 \widehat{\mathcal{H}}_{p+1,m}^0 W_2^T$$

where $\widehat{\mathcal{H}}_{p+1,m}^0(N)$ is as in (42), and W_1, W_2 are user-defined matrices with given rank properties. Using the subspace interpretation of the SVD, more precisely all but the first n singular vectors, it is possible to show [10] that parameter vector θ can be characterized by the following estimating equation :

$$S^T W_1 \widehat{\mathcal{H}}_{p+1,m}^0(N) W_2^T = 0 \quad (49)$$

where the orthonormal matrix S is subject to the constraint :

$$S^T W_1 \mathcal{O}_{p+1}(\theta) = 0 \quad (50)$$

We stress that matrix S depends implicitly on parameter θ and is not unique. However, we treat it as a function of parameter θ , denoted by $S(\theta)$, which we justify below. Let θ_0 the

matrix resulting from this identification, and let $\widehat{\mathcal{H}}_{p+1,m}(N)$ the empirical Hankel matrix filled with covariance matrices of new measurements collected on the (possibly failed) system. The improved residual corresponding to the above SVD estimating equation is :

$$\zeta_N(\theta_0) \triangleq \sqrt{N} \operatorname{vec} \left(S^T(\theta_0) W_1 \widehat{\mathcal{H}}_{p+1,m}(N) W_2^T \right) \quad (51)$$

Thanks to the CLT of section 4, residual ζ_N is asymptotically Gaussian distributed, and its mean value reflects the possible change in the modal parameter θ . The test statistic is thus a quadratic form in ζ_N .

Using the auxiliary vector :

$$Z_k \triangleq \begin{pmatrix} W_2 \mathcal{Y}_{k-p,m}^- \\ V_k(\theta_0) \end{pmatrix} \quad (52)$$

where $V_k(\theta_0)$ is now defined as :

$$V_k(\theta_0) \triangleq S^T(\theta) W_1 \mathcal{Y}_{k-p,p+1}^+ \quad (53)$$

the above improved residual ζ_N is found in the form (22), with estimating function K defined as :

$$K(\theta_0, Z_k) = W_2 \mathcal{Y}_{k-p,m}^- \otimes V_k(\theta_0) \quad (54)$$

which is similar to (48).

Note that matrix S satisfying (50) is not unique, and thus cannot be considered as a function $S(\theta)$. However, thanks to lemma 4.1, the subspace-based χ^2 -tests can easily be shown [10] to be invariant w.r.t. the particular choice of S in (50). Hence the notation $S(\theta)$ in (51). Actually, the tests are also invariant w.r.t. W_1 and W_2 for the same reason. We refer the reader to [10] for more details.

7 Back to example 2.2

We now discuss estimating functions for error-in-variables models and explain how to design algorithms for the problem of monitoring the combustion set of a gas turbine described in section 2.

7.1 LS-score

The *static* model is given in (8). An identification procedure is described in [65] to identify parameter θ using a N -size sample of input and output data U_k, Y_k . It consists in minimizing the LS cost function or equivalently in using the LS-score K defined in (37)-(38)-(39). In this case, the one-step ahead predictor is nothing but the vector of measurements predicted by model (8), since this model is static. Hence, the LS-score K is based on :

$$\epsilon_k(\theta) = Y_k - Y_k(\theta), \quad Z_k \triangleq \begin{pmatrix} \operatorname{vec} \frac{\partial Y_k(\theta)}{\partial \theta} \\ \epsilon_k(\theta) \end{pmatrix}$$

and writes :

$$K(\theta, Z_k) = \left(\frac{\partial Y_k(\theta)}{\partial \theta} \right)^T \epsilon_k(\theta) \quad (55)$$

A Robbins-Monro stochastic gradient algorithm would be of the form :

$$\theta_k = \theta_{k-1} + \beta_k K(\theta_{k-1}, Z_k)$$

where β_k is a suitable gain factor. It turns out in practice that this algorithm behaves poorly, because of the nonlinearity in the model, and the presence of noise at both the output *and* the input of the considered system. Consequently, an off-line stable identification procedure [65] relies also on LS scores, but works instead in two steps, separating the identification of the nonlinear coefficients d_0, d_1 from the identification of the other (linear) parameters.

One problem remains however, since it is known that the parameters of a (static nonlinear) model with noisy outputs *and* inputs cannot be identified without a bias [60].

7.2 Bias-adjusted estimating function

We stress that this identification bias does not prevent from detecting slight deviations in the system parameters, although we must assume a small identification bias if isolation is also of interest [66].

The idea for dealing with such a bias is the following. Let θ_0 be the reference parameter obtained through the off-line two-steps identification algorithm mentioned above. Let K be the primary residual defined in (55)¹⁵. Furthermore, let :

$$\kappa(\theta_0) = \frac{1}{N} \sum_{k=1}^N K(\theta_0, Z_k)$$

be the empirical bias of this residual, computed on the data collected on the safe system and used for identification. This bias would be zero for an unbiased θ_0 .

Then, given a new N -size data sample, the bias-adjusted improved residual :

$$\zeta_N(\theta_0) \triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^N (K(\theta_0, Z_k) - \kappa(\theta_0))$$

can be shown [66] to obey the CLT result stated in section 4.2, and a χ^2 -test can be computed as in (36), which turns out to be powerful in practice. This bias-adjusted improved residual has been fed into two algorithms tuned and implemented differently, one on-line version tuned for detecting large sudden drops in temperatures, the other off-line version for detecting smaller and slower drops in temperatures. We refer to [11, 47] for more details.

8 Back to example 2.3

In this section, we address the design of monitoring algorithms for nonlinear dynamic state-space processes. We describe two possible solutions based on a least-squares estimating function, and involving either a full state observer (as proposed in [21] for the problem of monitoring the catalytic converter of an automobile described in section 2), or an equivalent input-output model.

We consider continuous time nonlinear dynamical systems in state-space form :

$$\begin{cases} \dot{X} &= f(\theta, X, U) \\ Y &= h(\theta, X, U) \end{cases} \quad (56)$$

where f and h are nonlinear functions having continuous derivatives; X is the non-measured state variable; U and Y are the measured input and output signals; \dot{X} is the time derivative of X ; θ collects the unknown system parameters. Note that, as in example 2.3, input U may also depend on θ , and that $\dot{\theta} = 0$ when no change occurs in θ . Because in practice one has to deal

¹⁵We stress that this monitoring function is global in all the parameters, whereas the identification is achieved in two steps.

with data sampled at rate $1/\tau$: $U_k = U(k\tau)$, $Y_k = Y(k\tau)$, we consider the following discrete time approximation :

$$\begin{cases} \delta X_k &= f(\theta, X_k, U_k) \\ Y_k &= h(\theta, X_k, U_k) \end{cases} \quad (57)$$

where δ is a discrete approximation of the differentiation operator. For taking into account model uncertainty and measurement noise, we introduce a stochastic assumption. For nonlinear systems, such an assumption should be based on physical knowledge. For the sake of simplicity, we assume a white Gaussian noise added on the observation equation :

$$\begin{cases} \delta X_k &= f(\theta, X_k, U_k) \\ Y_k &= h(\theta, X_k, U_k) + \varepsilon_k^{(Y)} \end{cases} \quad (58)$$

8.1 Full state observer and LS-score

We explain first how a full state observer is involved in the LS-score defined in (37)-(38)-(39), which requires a one-step ahead predictor $\hat{Y}_{k|k-1}(\theta)$. As stated above, the prediction error should be asymptotically stationary. This calls for the use of a full order observer in the form :

$$\begin{cases} \hat{X}_{k|k-1} &= \hat{f}(\theta, \hat{X}_{k-1|k-2}, U_{k-1}, Y_{k-1}) \\ \hat{Y}_{k|k-1}(\theta) &= h(\theta, \hat{X}_{k|k-1}, U_{k-1}, Y_{k-1}) \end{cases} \quad (59)$$

for a suitably chosen \hat{f} . Then matrix $\partial \hat{Y}_{k|k-1}(\theta)/\partial \theta$ is found by differentiating (59) w.r.t. θ , that is :

$$\begin{cases} \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} &= \hat{f}'_X \frac{\partial \hat{X}_{k-1|k-2}}{\partial \theta} + \hat{f}'_U \frac{\partial U_{k-1}}{\partial \theta} + \hat{f}'_\theta \\ \frac{\partial \hat{Y}_{k|k-1}(\theta)}{\partial \theta} &= \hat{h}'_X \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} + \hat{h}'_U \frac{\partial U_{k-1}}{\partial \theta} + \hat{h}'_\theta \end{cases}$$

Note that, because of the definition of the auxiliary vector for the LS-score in (38), these two dynamical systems correspond to what we have required in (15). Of course, this may be computationally complex. The requirement for a full order observer when implementing the LS-score thus calls for the use of *reduced* state-space models. This is what is done in [21].

An alternative solution, circumventing the need of a full order observer, is based on the use of an equivalent input-output representation of the *nonlinear* state-space model (56) for the design of monitoring algorithms, which we discuss next. It should be reminded that an input-output representation of a *linear* state-space model has been used in section 6.

8.2 Input-output representation and LS-score

The idea is to consider the nonlinear dynamical state-space model in (56) as a particular case of differential-algebraic equations (DAE) model, and to find an input-output representation of the form $F(\theta, U, Y, p) = 0$. Notation p stands for the differentiation operator, and function F is a vector of differential polynomials in θ, Y, U . How to obtain such an input-output relationship relies on Ritt's algorithm and is explained in [45, 67]. As above, we consider its discrete time counterpart :

$$F(\theta, U_k, Y_k, \delta) = 0 \quad (60)$$

based on the sampled data and the approximated differentiation operator δ .

Next, we modify the above noise modeling assumption, and we now assume that the stochastic model of the system is :

$$F(\theta, U_k, Y_k, \delta) = \varepsilon_k^{(F)} \quad (61)$$

where noise $\varepsilon_k^{(F)}$ is only assumed to be white. The LS-score for this model is :

$$K(\theta, U_k, Y_k, \delta) = \left(\frac{\partial F(\theta, U_k, Y_k, \delta)}{\partial \theta} \right)^T F(\theta, U_k, Y_k, \delta) \quad (62)$$

This can be written as $K(\theta, Z_k)$ using the auxiliary vector :

$$Z_k \triangleq \begin{pmatrix} \text{vec} \frac{\partial F(\theta, U_k, Y_k, \delta)}{\partial \theta} \\ F(\theta, U_k, Y_k, \delta) \end{pmatrix}$$

Note that the prediction $\hat{Y}_{k|k-1}(\theta)$ is not involved in this monitoring function¹⁶. If we additionally assume the noise in (61) to be Gaussian with covariance matrix proportional to identity matrix, quantity (62) is nothing but an efficient score for model (61).

The particular case where F is *linear* in the parameter vector θ , namely :

$$F(\theta, U_k, Y_k, \delta) = P(U_k, Y_k, \delta) \theta - Q(U_k, Y_k, \delta)$$

with polynomial matrices (or vectors) P and Q , is of special interest. Actually, the existence of such a DAE is the sufficient and necessary condition for θ to be globally identifiable [45]. In that case, we have that :

$$K(\theta, U_k, Y_k, \delta) = P^T(U_k, Y_k, \delta) P(U_k, Y_k, \delta) \theta - P^T(U_k, Y_k, \delta) Q(U_k, Y_k, \delta)$$

9 Discussion

In this section, we discuss further several key issues, concerning the role and use of (parametric) models, off-line and on-line FDI algorithms, and optimal sensor location.

9.1 On-board model-based FDI : which (use of) models ?

Throughout the paper, we have assumed that (the structure of) the model of the monitored system is given. Several points should be made, regarding the definition, the choice and the use of parametric models for on-board monitoring.

Models for simulation, control, monitoring. Whatever the goal of modeling, should it be simulation, control, or monitoring, building a model based on physical principles is a relevant approach [44]. It should be clear, however, that the choice of a particular model should be driven by the processing objective. It may very well happen that a model which is appropriate for designing a realistic simulator is *not* relevant for monitoring purposes – e.g. because it is too complex – exactly as it might not be convenient for control purposes, possibly for other reasons. It is of help keeping in mind that the only requirement for monitoring is that the chosen model, when fed into the FDI algorithm, should allow us to discriminate between the safe/failed states of the system which we wish to distinguish. Also, as far as on-board monitoring is concerned, it may be useful to distinguish between two (or more) modeling levels. This statement is made clearer in the discussion of on-board diagnostics below.

¹⁶Actually, as noted in [67], a gain in robustness w.r.t. modeling errors can be expected with (62), since the approximated differentiation operator δ in (60) requires filters which are *not* based on the model.

On-board detection : monitoring rather than repeated identification. Assuming the model (structure) is given and a reference parameter value is identified, the question arises of how to exploit this information for monitoring. Since we have argued that the core of statistical monitoring methods is a parametric model together with an estimating function, the question we are faced with is the following. Why monitoring cannot be simply stated as a problem of repeated identification of the parameter vector? In that case, a straightforward solution would be to compare two successive estimates with the aid of a convenient distance measure. This is of course a valid FDI approach, but with some limitations, for the following reasons.

First, parameter identification may be computationally expensive, especially when it cannot be completely achieved automatically and requires human operation. This is the case for complex plants, and in particular for vibration monitoring of complex mechanical structures and machines, as described in section 6. Also, since we are willing not to detect the non-stationarities due to changes in the environment of the monitored system, it may happen that the identification of the reference parameter should be performed on data mixing as many different functioning modes of the system as possible. This may very well perturb or even stop the production, which is another type of undesirable identification cost. This is the case for the gas turbine example introduced in section 2. There is thus an actual need for designing FDI algorithms which do not perform re-identification on each data set for which a decision has to be made.

Second, comparing directly two parameter estimates may provide us with a poor solution to the fault isolation problem. For example, it is well known that it is very difficult to match two sets of modal shapes when some of the modes are close to each other.

Third, it should be clear that the key issue is how to measure the discrepancy between the new estimated parameter vector and the reference one, and how to decide that this deviation is significant, especially in the presence of noises and disturbances on the system. This basically requires an estimate of the accuracy of the identification procedure, namely an asymptotic Gaussianity type of result. But, as explained in section 4, the local approach is based on such type of asymptotic Gaussianity results, and provides us with tests which are computationally much cheaper than those resulting from re-identification.

On-board diagnostics without solving inverse problems. In many applications, it may happen that a complex physical system, characterized by a generally non identifiable parameter vector Φ has to be monitored using a simple (black-box) model characterized by an identifiable parameter vector θ . A typical example is the vibration monitoring problem in section 2, for which complex finite elements models are often available but not identifiable, whereas the small number of existing sensors calls for identifying only simplified input-output (black-box) representations. In such a situation, two different diagnosis problems can arise, namely diagnosis in terms of the black-box parameter θ and diagnosis in terms of the parameter vector Φ of the underlying physical model.

As explained in section 5.2, several methods can be used for solving the first problem, based upon either a sensitivity technique with respect to a given fault¹⁷, or a (statistical) method for rejecting the alternative faults. This type of diagnosis brings informations which are interesting in themselves, but can be of too complex interpretation : for instance, in vibration mechanics, it is well known that two changes very different from a mechanical point of view can lead to similar deviations in the modal parameters.

The second diagnosis problem can be stated as follows. Given a reference black-box model plus a physical model and new measurements, how to diagnose the faults in terms of the physical model? Of course, this problem is more tricky, because the physical model parameter Φ is often not identifiable and has a dimension much higher than the black-box parameter θ has. We

¹⁷We remind the reader that, in this problem setting, a fault is a change in one component or in a subset of components of vector θ .

stress that on-board diagnosis is nevertheless possible in this situation. The solution proposed in [27, 28] circumvents the resolution of the corresponding ill-posed inverse problem. It relies upon the projection of the elementary changes in the physical space onto the space of the black-box model and the computation of convenient Jacobian matrices. In other words, residual vector ζ_N is asymptotically Gaussian with mean : $M \delta\theta = M \mathcal{J} \delta\Phi$. The projection and Jacobian computation steps are performed once for all when designing the monitoring algorithms, whereas the isolation (sensitivity) tests are of low computational cost and can be used on-board [14].

Robustness in FDI methods. It is often argued that a major drawback of model-based FDI methods lies in the requirement of a precise model of the system to be monitored [24, 55]. Many attempts towards increased robustness with respect to unmodelled disturbances and model parameter uncertainties have been developed these last ten years. Two main types of investigations have been reported in the literature [26]. The first type concerns the residual generation step, and involves either system theoretic approaches to unknown input observers and disturbance decoupling rejection methods [53, 54, 25], or the fuzzy logic concept of knowledge observer. The second type concerns the residual evaluation step, and involves either adaptive threshold selection [23, 25], or fuzzy set membership functions or fuzzy decision logic [26].

An alternative approach to the robustness problem is a statistical model-based one. The main idea [47, 11] consists in using confidence ellipsoids for plugging the uncertainty in the shape of the (no-fault and fault) hypotheses between which one is willing to test, instead of using a confidence interval around the decision function itself. Let these ellipsoids be defined by :

$$\begin{aligned} \Theta_0 & : & (\theta - \theta_0)^T \Sigma^{-1} (\theta - \theta_0) & \leq \rho_0 \\ \Theta_1 & : & (\theta - \theta_0)^T \Sigma^{-1} (\theta - \theta_0) & \geq \rho_1 \end{aligned}$$

where θ_0 is the reference parameter, the values ρ_0 and ρ_1 have to be chosen. The generalized likelihood ratio (GLR) approach consists in computing :

$$t = \min_{\theta_0 \in \Theta_0} l_\theta - \min_{\theta_1 \in \Theta_1} l_\theta$$

where : $l_\theta(\zeta) = (\zeta - \theta)^T \Sigma^{-1} (\zeta - \theta)$ under Gaussian assumption. Using the Lagrange multipliers method, it is easy to show [9] that :

$$t = \begin{cases} -(\chi_0 - \rho_1)^2 & \text{for } \chi_0 \leq \rho_0 \\ -(\chi_0 - \rho_1)^2 + (\chi_0 - \rho_0)^2 & \text{for } \rho_0 \leq \chi_0 \leq \rho_1 \\ +(\chi_0 - \rho_0)^2 & \text{for } \chi_0 \geq \rho_1 \end{cases}$$

where : $\chi_0 = \sqrt{l_{\theta_0}(\zeta)}$. Experiments on the above mentioned real applications have shown that this simple idea turns out to be a useful tool for obtaining robustness against both changes in the functioning modes and under-modeling of the system to be monitored, which are often reported as being a crucial issue for model-based on-board FDI algorithms [8].

Nonlinear black-box models. When dealing with a parametric model-based characterization of the monitored system in its usual working conditions, up to now we have basically referred to physical modeling, possibly with reduced dimension. It may happen that black-box modeling, which allows processing data without considering the physics, is a tool of interest, e.g. when the phenomena are complex and the knowledge on the physics too weak. When referring to nonlinear black-box models, and following [59, 39], we mean parameterized networks such as neural networks, wavelet networks, or even fuzzy rules. Of course, such a network can be viewed as a particular type of parametric model. For example, in a wavelet network, the parameters

are the weights and the shifts and scales coefficients. It should be clear that the above general (local) approach to the design of FDI algorithms does apply in the same manner [68]. However, the interest of such a parametric representation for fault isolation and diagnosis heavily depends upon the existence of a link between the parameters of the network and a (possibly partial) physical knowledge. We stress that, in a given application where the available expert physical knowledge is limited, the choice of the most relevant model for FDI is not obvious. It may very well happen [47, 59] that a parametric network is actually much better identified than a simplified physical model, in terms of prediction error and/or of simulation performances, but less efficient from the point of view of detection performances [68].

9.2 Off-line and on-line FDI algorithms

Now we make more precise a statement made in section 1.4, that the solution to the three types of detection problems are nothing but different implementations of the same estimating function (primary residual). We explain in section 4 that the model validation problem (testing between the local hypotheses defined in (21)) can be solved with the χ^2 -test (36) based on the improved residual ζ_N defined in (22) :

$$\zeta_N(\theta_0) = \frac{1}{\sqrt{N}} \sum_{k=1}^N K(\theta_0, Z_k)$$

For the sake of simplicity and without loss of generality, we assume that this test boils down to $\zeta_N^T \Sigma^{-1} \zeta_N \geq \lambda$.

The corresponding on-line change detection algorithm, which searches for a time instant v at which the parameter vector switches from θ_0 to a close value, is based on the stopping rule [9] :

$$\arg \min_{n \geq 1} \{ \chi_n^2 \geq \lambda \}, \quad \chi_n^2 = \max_{1 \leq j \leq n} (\zeta_j^n)^T \Sigma^{-1} \zeta_j^n$$

where :

$$\zeta_j^n = \frac{1}{\sqrt{n-j+1}} \sum_{k=j}^n K(\theta_0, Z_k)$$

It should be clear that this on-line algorithm is actually a different implementation of the estimating function K which is used in the model validation algorithm above.

9.3 Optimal sensor location

Another interesting byproduct of the statistical approach to FDI is the definition of a quantitative criterion, based upon the power of the tests [13, 27], which allows us to assess the quality of a given sensor location for monitoring and diagnosis purposes. This criterion can be used in two manners :

- For a given sensor pool location (it is not always possible to choose!), find the faults which are detectable and diagnosable;
- Alternatively, for one given (set of) fault(s), find the sensor pools and locations which allow us to detect and diagnose them.

10 Conclusion

We have described a general statistical approach to the design of FDI algorithms for dynamical systems, illustrated with real industrial examples. We have stressed why and how the main conceptual task of this design is the selection of an estimating function. We have discussed the

relevance of the efficient score and of the LS (prediction error) score for system faults, and outlined the limitation of the relevance of innovations and (state) estimation errors to sensor faults (and actuator faults in linear systems). For nonlinear state-space systems, we have exhibited two solutions, based either on a full state observer or on an input-output equivalent relationship. An open question of interest would be to design intermediate solutions involving reduced observers.

References

- [1] J. ABIDA and D. CLAUDE (1994). Spark ignition engine and pollution emissions : new approaching in modeling and control. *Int. Jal Vehicle Design*, vol.15, no 3-4-5, pp.484-508.
- [2] F. AIMARD, S. LI and M. SORINE (1995). Mathematical modeling of automotive three way catalytic converters with oxygen storage capacity. *IFAC Workshop Intelligent Components for Autonomous and Semi-autonomous Vehicles*, Toulouse, F.
- [3] T.W. ANDERSON (1971). *The Statistical Analysis of Time Series*. Wiley, New York.
- [4] I.V. BASAWA (1985). Neyman-Le Cam tests based on estimating functions. In *Proc. Berkeley Conf. in Honor of Neyman and Kiefer*, vol.2, Le Cam and Olshen (Eds), pp.811-825.
- [5] I.V. BASAWA (1991). Generalized score tests for composite hypotheses. In [32], pp.121-132.
- [6] M. BASSEVILLE (1988). Detecting changes in signals and systems - A survey. *Automatica*, vol.24, no 3, pp.309-326.
- [7] M. BASSEVILLE (1997). Information criteria for residual generation and fault detection and isolation. *Automatica*, vol.33, no 5, pp.783-803.
- [8] M. BASSEVILLE (1997). On robustness in statistical FDI methods. *IFAC/IMACS Symp. on Fault detection, supervision and safety for technical process - SAFEPROCESS'97*, Kingston upon Hull, U.K., Aug. 26-28, 1997.
- [9] M. BASSEVILLE and I. NIKIFOROV (1993). *Detection of Abrupt Changes - Theory and Applications*. Prentice Hall, N.J.
- [10] M. BASSEVILLE, M. ABDELGHANI and A. BENVENISTE (1997). Subspace-based fault detection and isolation methods - Application to vibration monitoring. *Research Report IRISA*, to appear.
- [11] M. BASSEVILLE, A. BENVENISTE, G. MATHIS and Q. ZHANG (1994). Monitoring the combustion set of a gas turbine. *IFAC/IMACS Symp. on Fault detection, supervision and safety for technical process - SAFEPROCESS'94*, Helsinki, Finland, June 13-16, 1994.
- [12] M. BASSEVILLE, A. BENVENISTE, G. MOUSTAKIDES and A. ROUGÉE (1987a). Detection and diagnosis of changes in the eigenstructure of non-stationary multivariable systems. *Automatica*, vol.23, pp.479-489.
- [13] M. BASSEVILLE, A. BENVENISTE, G. MOUSTAKIDES and A. ROUGÉE (1987b). Optimal sensor location for detecting changes in dynamical behavior. *IEEE Trans. on Automatic Control*, vol.AC-32, no 12, pp.1067-1075.
- [14] M. BASSEVILLE, A. BENVENISTE, B. GACH-DEVAUCHELLE, M. GOURSAT, D. BONNECASE, P. DOREY, M. PREVOSTO and M. OLAGNON (1993). Damage monitoring in vibration mechanics : issues in diagnostics and predictive maintenance. *Mech. Syst. Sign. Proc.*, vol.7, no 5, pp.401-423.
- [15] C.W. BAUM and V.V. VEERAVALLI (1994). A sequential procedure for multihypothesis testing. *IEEE Trans. Information Theory*, vol.IT-40, no 6, pp.1994-2007.
- [16] A. BENVENISTE and J-J. FUCHS (1985). Single sample modal identification of a non-stationary stochastic process. *IEEE Trans. Autom. Control*, vol.AC-30, pp.66-74.

- [17] A. BENVENISTE, M. BASSEVILLE and G. MOUSTAKIDES (1987). The asymptotic local approach to change detection and model validation. *IEEE Trans. Autom. Control*, vol.AC-32, no 7, pp.583-592.
- [18] A. BENVENISTE, M. MÉTIVIER and P. PRIOURET (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer, NY.
- [19] P.J. BICKEL, C.A.J. KLAASEN, Y. RITOV and J.A. WELLNER (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore.
- [20] M.H. COSTIN (1991). On-board diagnostics of vehicle emission system components : review of upcoming government regulation. *IFAC/IMACS Symp. on Fault detection, supervision and safety for technical process – SAFEPROCESS'91*, Baden-Baden, Sept. 10-13, 1991.
- [21] C. CUSSENOT, M. BASSEVILLE and F. AIMARD (1996). Monitoring the vehicle emission system components. *IFAC'96 World Conf.*, San Francisco, CA.
- [22] B. DELYON, A. JUDITSKY and A. BENVENISTE (1997). On the relationship between identification and local tests. *Research Report IRISA no 1104*.
- [23] A. EMAMI-NAEINI, M.M. AKHTER and S.M. ROCK (1988). Effect of model uncertainty on failure detection : the threshold selector. *IEEE Trans. Autom. Control*, vol.AC-33, no 12, pp.1106-1115.
- [24] P.M. FRANK (1990). Fault diagnosis in dynamic systems using analytical and knowledge based redundancy – A survey and new results. *Automatica*, vol.26, pp.459-474.
- [25] P.M. FRANK (1994a). Enhancement of robustness in observer-based fault detection. *Int. J. Control*, vol.59, no 4, pp.955-981.
- [26] P.M. FRANK (1994b). Application of fuzzy logic to process supervision and fault diagnosis. *IFAC/IMACS Symp. on Fault detection, supervision and safety for technical process – SAFEPROCESS'94*, Helsinki, Finland, June 13-16, 1994.
- [27] B. GACH-DEVAUCHELLE (1991a). *Diagnostic Mécanique des Fatigues sur les Structures Soumises à des Vibrations en Ambiance de Travail*. Thesis, Paris IX Dauphine University (in French).
- [28] B. GACH-DEVAUCHELLE (1991b). Diagnosing mechanical changes in vibrating systems. *IFAC/IMACS Symp. on Fault detection, supervision and safety for technical process – SAFEPROCESS'91*, Baden-Baden, FRG, September 1991.
- [29] J.J. GERTLER (1988). Survey of model-based failure detection and isolation in complex plants. *IEEE Control Syst. Mag.*, vol.8, no 6.
- [30] J.J. GERTLER and D. SINGER (1990). A new structural framework for parity equation based failure detection and isolation. *Automatica*, vol.26, pp.381-388.
- [31] J.J. GERTLER, M. COSTIN, X. FANG, Z. KOWALCZUK, M. KUNWER and R. MONAJEMY (1995). Model based diagnosis for automotive engines - Algorithm development and testing on a production vehicle. *IEEE Trans. Control Syst. Techn.*, vol.CST-3, no 1, pp.61-69.
- [32] V.P. GODAMBE, Ed. (1991). *Estimating Functions*. Clarendon Press, Oxford.
- [33] G.C. GOODWIN and K.S. SIN (1984). *Adaptive Filtering, Prediction and Control*. Prentice Hall.

- [34] P. HALL and C.C. HEYDE (1980). *Martingale Limit Theory and its Application*. Academic Press, NY.
- [35] W.J. HALL and D.J. MATHIASON (1990). On large sample estimation and testing in parametric models. *Int. Stat. Review*, vol.58, no 1, pp.77-97.
- [36] C.C. HEYDE (1989). Fisher lecture : quasi-likelihood and optimality for estimating functions : some current unifying themes. *Proc. Int. Stat. Inst., 47th Session*, Paris, pp.19-29.
- [37] P.J. HUBER (1981). *Robust Statistics*. Wiley.
- [38] R. ISERMANN (1984). Process fault detection based on modeling and estimation methods – A survey. *Automatica*, vol.20, pp.387-404.
- [39] A. JUDITSKY, H. HJALMÄRSSON, A. BENVENISTE, B. DELYON, L. LJUNG, J. SJÖBERG and Q. ZHANG (1995). Nonlinear black-box modeling in system identification : mathematical foundations. *Automatica*, vol.31, no 12, pp.1725-1750.
- [40] T.L. LAI (1995). Sequential change-point detection in quality control and dynamical systems. *Jal Royal Stat. Soc. B*, vol.57, no 4, pp.613-658.
- [41] L. LE CAM (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
- [42] E.L. LEHMANN (1986). *Testing Statistical Hypotheses*, 2d ed. Wiley, New York.
- [43] B.C. LÉVY, A. BENVENISTE and R.N. NIKOUKHAH (1996). High-level primitives for recursive maximum likelihood estimation. *IEEE Trans. Autom. Control*, vol.AC-41, pp.1125-1145.
- [44] L. LJUNG (1987). *System Identification - Theory for the User*. Prentice Hall, NJ.
- [45] L. LJUNG and T. GLAD (1994). On global identifiability for arbitrary model parameterizations. *Automatica*, vol.30, no 2, pp.265-276.
- [46] X.-C. LOU, A.S. WILLSKY and G.C. VERGHEESE (1986). Optimally robust redundancy relations for failure detection in uncertain systems. *Automatica*, vol.22, no 3, pp.333-344.
- [47] G. MATHIS (1994). *Outils de Détection de Ruptures et de Diagnostic – Application à la Surveillance de Turbines à Gaz*. Thesis, Rennes I University, (In French).
- [48] D.L. MCLEISH and C.G. SMALL (1988). *The Theory and Applications of Statistical Inference Functions*. Lect. Notes Stat. 44, Springer.
- [49] R.K. MEHRA and J. PESCHON (1971). An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica*, vol.7, pp.637-640.
- [50] I.V. NIKIFOROV (1995). A generalized change detection problem. *IEEE Trans. Information Theory*, vol.IT-41, no 1, pp.171-187.
- [51] I.V. NIKIFOROV (1996). A new optimal approach to GPS/DGPS integrity monitoring. *AIAA Jal Guidance, Control, and Dynamics*, vol.19, no 5, pp. 1023-1033.
- [52] I.V. NIKIFOROV (1997). Two strategies in the problem of change detection and isolation. *IEEE Trans. Information Theory*, vol.IT-43, no 3, pp.770-776.
- [53] R.N. NIKOUKHAH (1994). Innovations generation in the presence of unknown inputs : application to robust failure detection. *Automatica*, vol.30, pp.1851-1868.

- [54] R.J. PATTON (1994). Robust model-based fault diagnosis : the state of the art. *IFAC/IMACS Symp. on Fault detection, supervision and safety for technical process – SAFEPROCESS'94*, Helsinki, Finland, June 13-16, 1994.
- [55] R.J. PATTON, P. FRANK and R. CLARK, Eds. (1989). *Fault Diagnosis in Dynamic Systems – Theory and Application*. Prentice Hall.
- [56] D. PICARD (1985). Testing and estimating change-points in time series. *Adv. Applied Proba.*, vol.17, pp.841-867.
- [57] M. PREVOSTO, M. OLAGNON, A. BENVENISTE, M. BASSEVILLE AND G. LEVEY (1991). "State-space formulation, a solution to modal parameter estimation". *Jal Sound and Vibration*, vol.148, no 2, pp.329-342.
- [58] G.G. ROUSSAS (1972). *Contiguity of Probability Measures, Some Applications in Statistics*. Cambridge Univ. Press, New York.
- [59] J. SJÖBERG, Q. ZHANG, L. LJUNG, A. BENVENISTE, B. DELYON, P.-Y. GLORENNEC, H. HJALMARSSON and A. JUDITSKY (1995). Nonlinear black-box modeling in system identification : a unified overview. *Automatica*, vol.31, no 12, pp.1691-1724.
- [60] T. SÖDERSTRÖM (1981). Identification of stochastic linear systems in presence of input noise. *Automatica*, vol.17, pp.713-725.
- [61] E. SPJØTVOLL (1972). On the optimality of some multiple comparison procedures. *Annals Statistics*, vol.21, no 3, pp.1486-1521.
- [62] P. VAN OVERSCHEE and B. DE MOOR (1996). *Subspace Identification for Linear Systems : Theory - Implementation - Methods*. Kluwer.
- [63] A.S. WILLSKY (1976). A survey of design methods for failure detection in dynamic systems. *Automatica*, vol.12, pp.601-611.
- [64] A.S. WILLSKY and H.L. JONES (1976). A generalized likelihood ratio approach to detection and estimation of jumps in linear systems. *IEEE Trans. Autom. Control*, vol.AC-21, no 1, pp.108-112.
- [65] Q. ZHANG (1991). *Contribution à la Surveillance de Procédés Industriels*. Thesis, Rennes I University, (in French).
- [66] Q. ZHANG, M. BASSEVILLE and A. BENVENISTE (1994). Early warning of slight changes in systems and plants with application to condition based maintenance. *Automatica*, vol.30, no 1, pp.95-114.
- [67] Q. ZHANG, M. BASSEVILLE and A. BENVENISTE (1996). Fault detection and isolation in nonlinear dynamic systems : a combined input-output and local approach. *Research Report IRISA no 1074/ INRIA no 3120*.
- [68] Q. ZHANG (1997). Fault detection and isolation with nonlinear black-box models. *11th IFAC/IFORS Symposium on Identification and System Parameter Estimation - SYSID'97*, Kitakyushu, Japan, July 8-11, 1997.

Contents

1	Introduction	3
1.1	Condition monitoring and model-based FDI	3
1.2	Three types of faults	3
1.3	Design of model-based FDI algorithms	4
1.4	Three types of detection problems	5
1.5	Outline of the paper	6
2	Industrial examples	7
2.1	Linear dynamic system : vibration monitoring	7
2.2	Nonlinear static system : gas turbine	9
2.3	Nonlinear dynamic system : catalyst	10
3	Residual generation and problem reduction	12
3.1	Additive faults : repeated equations	12
3.2	Non-additive faults : local approach	13
3.3	Reduced FDI problem	16
4	Local approach to problem reduction	16
4.1	Local approach : log-likelihood function	16
4.2	Local approach - Other estimating/monitoring functions	20
4.3	Monitoring functions : which ones ?	22
5	Residual evaluation : monitoring the mean of a Gaussian vector	23
5.1	Fault detection : likelihood ratio	23
5.2	Fault isolation : projection or rejection	23
6	Back to example 2.1	25
6.1	Input-output monitoring functions	25
6.2	Subspace-based monitoring functions	26
7	Back to example 2.2	27
7.1	LS-score	27
7.2	Bias-adjusted estimating function	28
8	Back to example 2.3	28
8.1	Full state observer and LS-score	29
8.2	Input-output representation and LS-score	29
9	Discussion	30
9.1	On-board model-based FDI : which (use of) models ?	30
9.2	Off-line and on-line FDI algorithms	33
9.3	Optimal sensor location	33
10	Conclusion	33
	References	35



Unit e de recherche INRIA Lorraine, Technop le de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS L ES NANCY
Unit e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit e de recherche INRIA Rh ne-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

 diteur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399