Subspace-based Fault Detection and Isolation Methods - Application to Vibration Monitoring

Michèle Basseville, Maher Abdelghani, Albert Benveniste

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Subspace-based fault detection and isolation methods - Application to vibration monitoring

Michèle Basseville, Maher Abdelghani, Albert Benveniste

N ° 3299

Novembre 1997

THÈME 4

ISSN 0249-6399

INRIA
RENNES

Subspace-based fault detection and isolation methods - Application to vibration monitoring

Michèle Basseville*, Maher Abdelghani, Albert Benveniste†

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Sigma2

Rapport de recherche n3299 — Novembre 1997 — 29 pages

Abstract: We address the problem of detecting and isolating faults modeled as changes in the eigenstructure of a linear dynamical system. The purpose of the paper is to describe and analyze new fault detection and isolation algorithms, based on recent stochastic subspace-based identification methods and the statistical local approach to the design of detection algorithms. The application to vibration monitoring of mechanical structures and rotating machines is discussed. A conceptual comparison is made with another detection algorithm based on the instrumental variables identification method, and previously proposed by two of the authors.

Key-words: Fault detection and isolation, subspace-based stochastic identification methods, statistical local approach, structural vibration monitoring.

(Résumé : tsvp)

This work is supported by Eureka project no 1562 Sinopsys (Model based **S**tructural monitoring using **in-op**eration **sys**tem identification) coordinated by Lms, Leuven, Belgium.

* Irisa/Cnrs – basseville@irisa.fr.
† Irisa/Inria – Maher.Abdelghani,benveniste@irisa.fr.

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : 02 99 84 71 00 - International : +33 2 99 84 71 00
Télécopie : 02 99 84 71 71 - International : +33 2 99 84 71 71

Détection et diagnostic de pannes à l'aide de méthodes par sous-espaces - Application à la surveillance des vibrations.

Résumé : On considère le problème de la détection et du diagnostic de pannes modélisées par des changements de la structure propre d'un système dynamique linéaire. Dans cet article sont décrits et analysés de nouveaux algorithmes de détection et diagnostic de telles pannes, basés sur les méthodes récentes d'identification par sous-espaces et sur l'approche statistique locale. L'application à la surveillance vibratoire des structures mécaniques et machines tournantes est discutée. Ces méthodes sont comparées conceptuellement à un autre algorithme basé sur la méthode d'identification par variables instrumentales, et proposé précédemment par deux des auteurs.

Ce travail est effectué dans le cadre du projet Eureka no 1562 SINOPSYS (Model based **S**tructural monitoring using **in-op**eration **sys**tem identification), qui est coordonné par la société LMS, Leuven, Belgique.

Mots-clé : Détection et diagnostic de pannes, méthodes stochastiques d'identification par sous-espaces, approche statistique locale, surveillance des vibrations.

1 Introduction

Before outlining the paper, we describe our motivations and introduce the models which we use throughout.

1.1 Motivations

The problem of fault detection and isolation (FDI) is a crucial issue which has been investigated with different types of approaches, as can be seen from the survey papers [32, 18] and the books [24, 8] 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. It has been recognized that a relevant approach to condition-based maintenance consists in the early detection of *slight deviations* with respect to a (parametric) characterization of the system in usual working conditions, without artificial excitation, speeding down, or stop. Processing long samples of multi-sensor output measurements is often necessary for this purpose.

It turns out that, in many applications, the FDI problem of interest is to detect and diagnose changes in the eigenstructure of a linear dynamical system. An important example is structural vibration monitoring, which we address in section 6. The key issue there is to identify and monitor vibrating characteristics (modes and modal shapes) of mechanical structures subject to *unmeasured* and *non-stationary* natural excitation. Typical examples are offshore structures subject to swell, buildings subject to wind or earthquake, bridges, dams, wings subject to flutter in flight, and turbines subject to steam turbulence, friction in bearings, and imperfect balancing.

A relevant approach to the vibration monitoring problem has been shown to be based on the modeling of modes and modal shapes through state space representations [25], the use of *output-only* and *covariance-driven* identification methods (such as instrumental variables or balanced realization algorithms) [11], and the computation of specific χ^2-type tests based on the so-called instrumental statistics [9]. Moreover, the design of this χ^2-test is a particular case of a general approach to FDI which builds a detector from a convenient estimating function [12, 33, 6].

On the other hand, during the last decade there has been a growing interest in subspace-based linear system identification methods [28, 30, 29] and their relationship with instrumental variables [31]. Because of what has been argued above, the question arises to design FDI algorithms based on this class of identification methods, to investigate their theoretical properties and check their relevance for the vibration monitoring problem. It is the purpose of this paper to address these three issues.

1.2 System models and parameters

We consider linear multi-variable systems described by a discrete-time state space model :

$$\left\{ \begin{array}{rcl} X_{k+1} & = & FX_k + \varepsilon_{k+1} \\ Y_k & = & HX_k \end{array} \right. \tag{1}$$

where state X and observed output Y have dimensions m and r respectively. The state noise process $(\varepsilon_k)_k$ is an *unmeasured* Gaussian white noise sequence with zero mean and covariance matrix :

$$\mathrm{cov}(\varepsilon_k) = \mathbf{E}\left(\varepsilon_k\, \varepsilon_k^T\right) \stackrel{\mathrm{def}}{=} Q$$

where $\mathbf{E}(.)$ denotes the expectation operator. For the double sake of simplicity and validity of the arguments we use, we assume noise ε_k to be stationary, that is constant covariance matrix Q. We address the issue of robustness with respect to non-stationary excitation in section 5.

Let $G \stackrel{\text{def}}{=} \mathbf{E}\left(X_k Y_k^T\right)$ be the cross-correlation between state X_k and observation Y_k, and let :

$$\mathcal{O}_p = \begin{pmatrix} H \\ HF \\ \vdots \\ HF^{p-1} \end{pmatrix} \text{ and } \mathcal{C}_p = \begin{pmatrix} G & FG & \dots & F^{p-1}G \end{pmatrix} \tag{2}$$

be the pth-order observability matrix of system (1) and controllability matrix of pair (F, G), respectively. We assume that, for p large enough, both observability and controllability matrices are full rank m.

As mentioned above, and assuming the observation matrix H to be given, the problem is to identify and monitor the observed system eigenstructure. The observed system eigenstructure is the collection of m pairs (λ, ϕ_λ), where λ ranges over the set of eigenvalues of state transition matrix F, $\phi_\lambda = H\varphi_\lambda$ and φ_λ is the corresponding eigenvector. In all what follows, we assume that the system has no multiple eigenvalues. We stress that the collection of (λ, ϕ_λ) provides us with a canonical parameterization of the pole part of system (1). In particular, it does not depend on the state space basis [11]. In the sequel, referring to vibration monitoring, such a pair (λ, ϕ_λ) is called a mode. The set of the m modes is considered as the system parameter θ :

$$\theta \stackrel{\text{def}}{=} \begin{pmatrix} \Lambda \\ \text{vec}\,\Phi \end{pmatrix} \tag{3}$$

In (3), Λ is the vector whose elements are the eigenvalues λ, Φ is the matrix whose columns are the ϕ_λ's, and vec is the column stacking operator. Parameter θ has size $(r+1)m$, but since the eigenvectors φ_λ, and thus the ϕ_λ, are defined up to a constant complex number, the parameter space $\Theta = \text{Span}\,\theta$ has dimension $l = rm$. The FDI problem is to detect and isolate changes in parameter vector θ. By isolation, we mean deciding which mode(s) has(ve) changed.

This FDI problem can also be addressed using an input-output ARMA re-writing of state space model (1) :

$$Y_k = \sum_{i=1}^{p} A_i \; Y_{k-i} + \sum_{j=0}^{p-1} B_j \; E_{k-j} \tag{4}$$

where $(E_k)_k$ is a Gaussian white noise sequence with zero mean and identity covariance matrix, and where the AR coefficients A_i are related to the pair (H, F) of model (1) via :

$$HF^p = \sum_{i=1}^{p} A_i \; HF^{p-i} \tag{5}$$

The FDI problem is then to detect and isolate changes in the matrix AR parameters of a multivariable ARMA process with unknown MA part.

1.3 Outline of the paper

We first recall the two basic concepts which we use for designing the new subspace-based FDI algorithms proposed in this paper. Section 2 summarizes the basic principles and computational forms of the stochastic weighted subspace-based identification methods on the basis of which we propose to design FDI algorithms. In particular, we discuss useful system parameter characterizations, and the role of weighting in model reduction. Section 3 summarizes a general method

for designing detection algorithms, based on estimating functions and a statistical local approach (assumption of asymptotically close hypotheses). Then we turn to the proposed subspace-based FDI algorithms. Section 4 is devoted to their design and analysis. In particular, we exhibit an invariance property of the χ^2-tests. Comparison is made with the instrumental variables-based detection algorithm proposed in [9]. The case of non-stationary excitation is addressed in section 5. The structural vibration monitoring problem is addressed in section 6. Finally, some conclusions are drawn in section 7.

2 Subspace-based identification methods

We begin this section with a summary of the basic principles and computational forms of stochastic subspace-based identification methods, assuming known system order. Then we discuss the role of weighting for model reduction. We outline a system parameter characterization in terms of an estimating function, which we use for designing the test statistics.

2.1 Basic principles - Known system order

As explained in the introduction, because of the FDI problem we address, we are interested in output-only and covariance-driven algorithms. We are thus given a sequence of covariances :

$$R_i \stackrel{\text{def}}{=} \mathbf{E}\left(Y_k\, Y_{k-i}^T\right) \tag{6}$$

of output Y_k of a state space model (1). We begin with several possible characterizations of the system parameter θ (3). Then we summarize the estimation procedures, and useful interpretations.

For $q \geq p+1$, let $\mathcal{H}_{p+1,q}$ be the block-Hankel matrix :

$$\mathcal{H}_{p+1,q} = \begin{pmatrix} R_0 & R_1 & \dots & R_{q-1} \\ R_1 & R_2 & \dots & R_q \\ \vdots & \vdots & \ddots & \vdots \\ R_p & \dots & \dots & R_{p+q-1} \end{pmatrix} \tag{7}$$

Notation. At this point, a notational convention should be stressed, which holds throughout the paper. As obvious from (7) and (6), notation $\mathcal{H}$ is used for 'true' block-Hankel matrices, whereas notation $\widehat{\mathcal{H}}$ is used for empirical block-Hankel matrices filled with empirical covariances.

2.1.1 Several system parameter characterizations

Direct computations of the R_i's from the model equations lead to the following well known factorization property [27] :

$$\mathcal{H}_{p+1,q} = \mathcal{O}_{p+1}\, \mathcal{C}_q \tag{8}$$

Choosing the eigenvectors of F as a basis for the state space[1] yields the following particular representation of the observability matrix [9] :

$$\mathcal{O}_{p+1}(\theta) = \begin{pmatrix} \Phi \\ \Phi\Delta \\ \vdots \\ \Phi\Delta^p \end{pmatrix} \tag{9}$$

[1] This is called the modal basis in the vibration monitoring application of section 6.

where diagonal matrix Δ is defined as $\Delta = \text{diag}(\Lambda)$, and Λ and Φ are as in (3). For any other state basis, the observability matrix $\mathcal{O}_{p+1}$ can be written as :

$$\mathcal{O}_{p+1} = \mathcal{O}_{p+1}(\theta)\, T \tag{10}$$

for a suitable $m \times m$ invertible matrix T.

Now we are given, on the one hand, a system parameter θ, and on the other hand a block-Hankel matrix $\mathcal{H}_{p+1,q}$, and we wish to characterize whether $\mathcal{H}_{p+1,q}$ can correspond to parameter θ. From (8), (9) and (10), we get the following. Whether $\mathcal{H}_{p+1,q}$ corresponds to θ can be checked in the following four equivalent ways :

1. $\mathcal{O}_{p+1}$ is the observability matrix $\mathcal{O}_{p+1}(H, F)$ for some pair (H, F) with eigenstructure θ, and $\mathcal{H}_{p+1,q}$ and $\mathcal{O}_{p+1}$ have the same range space.

2. $\mathcal{O}_{p+1}$ is the observability matrix $\mathcal{O}_{p+1}(H, F)$ for some pair (H, F) with eigenstructure θ, and $\mathcal{H}_{p+1,q}$ and $\mathcal{O}_{p+1}$ have the same left kernel space[2].

3. $\mathcal{O}_{p+1}(\theta)$ is defined by (9), and $\mathcal{H}_{p+1,q}$ and $\mathcal{O}_{p+1}(\theta)$ have the same range space.

4. $\mathcal{O}_{p+1}(\theta)$ is defined by (9), and $\mathcal{H}_{p+1,q}$ and $\mathcal{O}_{p+1}(\theta)$ have the same left kernel space, spanned, say, by the columns of a matrix A.

Point of view 1. gives rise to the usual stochastic realization methods as we recall next. But, as discussed further below, point of view 4. provides us with a much more direct characterization of parameter θ, which is the basis for our approach to the design of statistical tests for eigenstructure monitoring.

Principle of the proposed approach. Before proceeding, we stress that our approach to vibration monitoring proceeds as follows [9, 10]. Given data collected from the safe system, build the empirical block-Hankel matrix $\widehat{\mathcal{H}}^{\text{SAFE}}$ and extract a modal parameter estimate θ_0 using one of the subspace-based identification methods described below. Then, compute the observability matrix $\mathcal{O}(\theta_0)$ in this modal basis, as defined in (9), and compute a matrix $A(\theta_0)$ whose columns span the left kernel space of $\mathcal{O}(\theta_0)$. Then, given newly collected data, build the empirical block-Hankel matrix $\widehat{\mathcal{H}}^{\text{NEW}}$ and check whether $A^T(\theta_0)\, \widehat{\mathcal{H}}^{\text{NEW}}$ is significantly different from zero.

2.1.2 Stochastic realization and balancing

Stochastic output-only and covariance-driven subspace-based identification methods can be described in the following manner[3] [29, 21].

Let W_1 and W_2 be two user-defined invertible weighting matrices of size $(p+1)r$ and qr, respectively. Regarding the weighted block-Hankel matrix, these conditions imply that :

$$\text{rank}\left(W_1\, \mathcal{H}_{p+1,q}\, W_2^T\right) = \text{rank}\left(\mathcal{H}_{p+1,q}\right) (= m) \tag{11}$$

Of course, for known system order, requiring a square invertible weighting matrix W_2 is not necessary for (11) to hold. The following weaker condition :

$$W_2 \in \mathbb{R}^{\iota \times qr}, \quad \text{rank}\left(\mathcal{H}_{p+1,q}\, W_2^T\right) = \text{rank}\left(\mathcal{H}_{p+1,q}\right) (= m)$$

[2]The left kernel space of matrix M is the kernel space of matrix M^T.

[3]The presentation here is somewhat unusual. It is motivated by our objective of designing FDI algorithms.

with $\iota \geq qr$, is sufficient for this purpose. However, since in practice the actual rank of the empirical Hankel matrix is unknown, the only way to achieve (11) is to impose an invertible weighting matrix W_2 also.

Different choices of the weighting matrices lead to different identification methods (e.g., BRand CVA) and, for our purpose, to different detection methods. The nature and influence of these choices on identification and FDI is discussed in this section and in section 4, respectively.

Let :

$$W_1 \, \mathcal{H}_{p+1,q} \, W_2^T = \begin{pmatrix} P & P^\perp \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V^T \tag{12}$$

be the singular value decomposition (SVD) of this matrix, where $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ contains the m non-zero singular values in decreasing order, the m columns of matrix P are the corresponding left singular vectors, and the $(qr - m)$ columns of matrix $P^\perp$ are the left singular vectors associated with singular value 0. Note that P is full column rank (f.c.r.). The factorization property of the Hankel matrix results in :

$$W_1 \, \mathcal{H}_{p+1,q} \, W_2^T = W_1 \, \mathcal{O}_{p+1} \, \mathcal{C}_q \, W_2^T \tag{13}$$

with, for example :

$$W_1 \, \mathcal{O}_{p+1} = P D^{1/2} T \tag{14}$$

$$\mathcal{C}_q \, W_2^T = T^{-1} D^{1/2} V^T \tag{15}$$

where T is a non-singular transformation. The observability matrix can then be recovered, up to a change of basis or equivalently for some pair (H, F), as :

$$\mathcal{O}_{p+1}(H, F) = W_1^{-1} P D^{1/2} T \tag{16}$$

The particular choice of factor $D^{1/2}$, together with T identity, corresponds to the balanced realization (BR) case when weighting matrices W_1 and W_2 are identity. Any other non-singular factor may be used.

The pair (H, F) is recovered from $\mathcal{O}_{p+1}(H, F)$ in the following way. From the definition and the shift-invariant structure of $\mathcal{O}_p$ in (2), we deduce that the observation matrix H can be recovered as the first block-row of $\mathcal{O}_{p+1}$, and the state transition matrix F can be recovered as the LS solution of :

$$\mathcal{O}_p \, F = \mathcal{O}_{p+1}^{\uparrow} \tag{17}$$

In equation (17), matrices $\mathcal{O}_{p+1}^{\uparrow}$ and $\mathcal{O}_p$ are obtained by deleting the first and last block-row in $\mathcal{O}_{p+1}$, respectively. The eigenstructure of the state transition matrix F is then recovered from :

$$F = \Omega \, \Delta \, \Omega^{-1} \tag{18}$$

where Ω is the matrix whose columns are the φ_λ's. This is obtained, e.g., by eigenvalue solving.

To summarize, thanks to the definition of the parameter vector θ in (3) and of the observability matrix $\mathcal{O}_{p+1}(\theta)$ in modal basis in (9), we may re-write (13) as :

$$W_1 \, \mathcal{H}_{p+1,q} \, W_2^T = W_1 \, \mathcal{O}_{p+1}(\theta) \, \mathcal{C}_q \, W_2^T \tag{19}$$

2.1.3 Direct parameter characterization

Now, point of view 4. provides us with a much more direct characterization of parameter θ in terms of subspaces, namely :

$$\mathcal{H}_{p+1,q} \text{ and } \mathcal{O}_{p+1}(\theta) \text{ have the same left kernel space with co-rank } m \tag{20}$$

This property is true only for a system order known equal to m. It can be used for characterizing parameter θ as follows.

- Form $\mathcal{O}_{p+1}(\theta)$, and pre-multiply it by some invertible weighting matrix W_1;
- Pick an orthonormal basis of the left kernel space of matrix $W_1\mathcal{O}_{p+1}(\theta)$, in terms of the columns of some matrix S of co-rank m such that :

$$S^T\,S = I_s \tag{21}$$
$$S^T\,W_1\,\mathcal{O}_{p+1}(\theta) = 0 \tag{22}$$

  Matrix S has dimensions $(p+1)r \times s$, where $s = (p+1)r - m$, and it is not unique; it can be obtained, for example, by selecting $P^\perp$ in the SVD-factorization (12) of $W_1\mathcal{O}_{p+1}(\theta)W_2$. Two such matrices are related through a pre-multiplication with an orthonormal matrix T. Also, the choice of such a matrix depends on the weighting matrix W_1. We comment further on this below. We stress that, because of (22), matrix S depends implicitly on parameter θ. We bring the reader's attention to the fact that, even though S is not unique, in the sequel we treat it as a function of parameter θ, denoted by $S(\theta)$, which we fully justify in section 4.

- The parameter θ which actually corresponds to the output covariance sequence $(R_i)_i$ of model (1) is characterized by :

$$S^T(\theta)\,W_1\,\mathcal{H}_{p+1,q}\,W_2^T = 0 \tag{23}$$

The choice of weighting matrix W_1 only influences the particular basis of the left kernel space of $\mathcal{O}_{p+1}(\theta)$, since it is given by the rows of S^TW_1 where S is orthonormal (21). However, characterization (20) does not depend, of course, on the particular weighting matrices W_1 and W_2.

2.2 Model reduction : the role of weighting

Consider now the situation in which we seek for a model of order m, while the true system has a larger, possibly infinite, order. In this situation, relation (20) no longer holds, and the above approach has to be adapted in the following way. Observability matrix $\mathcal{O}_{p+1}(\theta)$ is constrained to be rank m, and parameter θ is searched such that the left kernel space of $\mathcal{O}_{p+1}(\theta)$ is as much orthogonal to $\mathcal{H}_{p+1,q}$ as possible.

In other words, it is still possible to find an orthonormal matrix S with co-rank m such that equation (22) holds. But relation (23) no longer holds, whatever θ is. Parameter θ should now be characterized by an approximation :

$$S^T(\theta)\,W_1\,\mathcal{H}_{p+1,q}\,W_2^T \approx 0 \tag{24}$$

reflecting that the left hand-side should be as close to zero as possible, namely that $S(\theta)$ should be as much orthogonal to $W_1\,\mathcal{H}_{p+1,q}\,W_2^T$ as possible. Of course, we need a precise definition for this approximation, which we introduce now.

Following [2], we consider the vectors containing the future and past data, respectively :

$$\mathcal{Y}^+_{k,p+1} \stackrel{\text{def}}{=} \begin{pmatrix} Y_k \\ Y_{k+1} \\ \vdots \\ Y_{k+p} \end{pmatrix}, \quad \mathcal{Y}^-_{k,q} \stackrel{\text{def}}{=} \begin{pmatrix} Y_k \\ Y_{k-1} \\ \vdots \\ Y_{k-q+1} \end{pmatrix} \tag{25}$$

Thanks to the stationarity assumption, Hankel matrix writes :

$$(\forall k) \quad \mathcal{H}_{p+1,q} = \mathbf{E}\left(\mathcal{Y}^+_{k,p+1}\,{\mathcal{Y}^-_{k,q}}^T\right) \tag{26}$$

It is of interest to introduce the normalized random vectors $\mathcal{E}^+_{k,p+1}$ and $\mathcal{E}^-_{k,q}$:

$$\mathcal{E}^+_{k,p+1} \stackrel{\text{def}}{=} \mathcal{T}^{-1/2}_{p+1,+}\, \mathcal{Y}^+_{k,p+1}, \quad \mathcal{E}^-_{k,q} \stackrel{\text{def}}{=} \mathcal{T}^{-1/2}_{q,-}\, \mathcal{Y}^-_{k,q}, \quad \text{where} \quad \mathcal{T}_{p+1,+} = \operatorname{cov} \mathcal{Y}^+_{k,p+1}, \quad \mathcal{T}_{q,-} = \operatorname{cov} \mathcal{Y}^-_{k,q},$$

which span the same subspaces as $\mathcal{Y}^+_{k,p+1}$ and $\mathcal{Y}^-_{k,q}$. A natural choice for the weighting matrices is thus :

$$W_1 = \mathcal{T}^{-1/2}_{p+1,+}, \quad W_2 = \mathcal{T}^{-1/2}_{q,-} \tag{27}$$

for which we have :

$$W_1\, \mathcal{H}_{p+1,q}\, W_2^T = \mathcal{T}^{-1/2}_{p+1,+}\, \mathbf{E}\left(\mathcal{Y}^+_{k,p+1}\, \mathcal{Y}^{-\,T}_{k,q}\right)\, \mathcal{T}^{-T/2}_{q,-} = \mathbf{E}\left(\mathcal{E}^+_{k,p+1}\, \mathcal{E}^{-\,T}_{k,q}\right) \tag{28}$$

Using point of view 4., we now characterize the m-th order parameter θ as follows. Let matrix $S(\theta)$ defined as in (22). Because of (28), checking that $S(\theta)$ is as much orthogonal to $W_1\, \mathcal{H}_{p+1,q}\, W_2^T$ as possible boils down to check that vector $\mathcal{F}^+_{k,p+1}$:

$$\mathcal{F}^+_{k,p+1} \stackrel{\text{def}}{=} S^T(\theta)\, \mathcal{E}^+_{k,p+1}$$

is as much orthogonal to $\mathcal{E}^-_{k,q}$ as possible. But, since the column vectors of $S(\theta)$ are orthonormal, vector $\mathcal{F}^+_{k,p+1}$ has identity covariance matrix. Hence, how orthogonal to $W_1\, \mathcal{H}_{p+1,q}\, W_2^T$ is $S(\theta)$ is evaluated by inspecting the principal angles between the subspaces generated by random vectors $\mathcal{F}^+_{k,p+1}$ and $\mathcal{E}^-_{k,q}$. Note that pre-multiplying $\mathcal{F}^+_{k,p+1}$ by an invertible matrix does not change these angles, which are thus not affected by the particular choice of basis in matrix S. These angles are provided [29] by the singular values σ_i of :

$$\mathbf{E}\left(\mathcal{F}^+_{k,p+1}\, \mathcal{E}^{-\,T}_{k,q}\right) = S^T(\theta)\, W_1\, \mathcal{H}_{p+1,q}\, W_2^T$$

where W_1 and W_2 are given by (27).

In the framework of stochastic realization and identification, this approach is known as the canonical variate analysis (CVA) [2, 3, 29]. Note that it proceeds exactly as in (12)-(17) for a specific choice of weighting matrices W_1 and W_2 [21, 22]. It is thus natural to use these particular weights for defining the approximation in (24), as done above.

To summarize, the minimization :

$$\min_{\theta,S} \quad g_\sigma\left(S^T\, W_1\, \mathcal{H}_{p+1,q}\, W_2^T\right) \tag{29}$$

$$\text{subject to} \qquad S^T\, S = I_s, \;\; S^T\, W_1\, \mathcal{O}_{p+1}(\theta) = 0 \tag{30}$$

yields the true system parameter $\theta_\star$ and the set of S such that the two constraints in (30) hold. In (29), function g_σ is defined as :

$$g_\sigma(A) = \sigma_{\max}(A) \quad \text{or} \quad g_\sigma(A) = \sum_i \sigma_i^2(A) \tag{31}$$

and notation $\sigma_{\max}(A)$ stands for the largest singular value of matrix A.

It is important to stress that, for the case of model reduction, characterization (29) depends on the particular choice of weighting matrices W_1 and W_2 defined in (27).

2.3 Subspace-based identification algorithms

We are now given a n-size data sample $Y_1, \ldots, Y_n$ and the associated empirical covariance matrices $\widehat{R}_i$. From (26), we deduce that the corresponding empirical block-Hankel matrix also writes :

$$\widehat{\mathcal{H}}_{p+1,q} = \frac{1}{n-p-q+1} \sum_{k=q}^{n-p} \mathcal{Y}^+_{k,p+1}\, \mathcal{Y}^{-\,T}_{k,q} \tag{32}$$

2.3.1 Characterization of parameter estimates

Since the empirical Hankel matrix of a m-th order system generically has rank greater than m, we characterize subspace-based parameter estimates as follows. One cannot expect the empirical counterpart of relationship (20) to hold exactly, even if θ is the true system parameter and no model reduction is enforced. Parameter θ should again be characterized by an approximation :

$$S^T(\theta)\, W_1\, \widehat{\mathcal{H}}_{p+1,q}\, W_2^T \approx 0 \tag{33}$$

Proceeding as in (29)-(30), we characterize the subspace-based estimate $\widehat{\theta}$ by :

$$(\widehat{\theta}, \widehat{S}) \stackrel{\text{def}}{=} \arg\min_{\theta,S} \; g_\sigma \left(S^T\, W_1\, \widehat{\mathcal{H}}_{p+1,q}\, W_2^T \right) \tag{34}$$

where matrix S is subject to constraints (30) :

$$S^T\, S = I_s \tag{35}$$

$$S^T\, W_1\, \mathcal{O}_{p+1}(\theta) = 0 \tag{36}$$

and g_σ is as in (31). This kind of estimates is a particular instance of the class of subspace fitting estimates investigated in [14].

Since (34) amounts to solving a nonlinear optimization problem, the practical implementation is slightly different, as explained below. Nevertheless, for the design and theoretical analysis of the proposed test statistics in section 4, we use characterization (34) and the properties of such subspace-based parameter estimates as established in [14].

2.3.2 Estimating functions

As will be made clear below, for designing our statistical FDI methods, it is of interest to re-write the left hand-side of system parameter characterization (33) as a vector-valued function of the parameter and the observed data, which can be viewed as an estimating function. Substituting (32) into (33), we obtain such a function in the form :

$$\frac{1}{n-\varrho+1} \sum_{k=\varrho}^{n} K(\theta, \mathcal{Y}_{k,\varrho}^-) \approx 0$$

with :

$$K(\theta, \mathcal{Y}_{k,\varrho}^-) \stackrel{\text{def}}{=} \text{vec} \left(S^T(\theta)\, W_1\, N(\mathcal{Y}_{k,\varrho}^-)\, W_2^T \right) \tag{37}$$

where $\varrho = p + q$ is the size of the window involved in data function N, and :

$$N(\mathcal{Y}_{k,\varrho}^-) \stackrel{\text{def}}{=} \mathcal{Y}_{k-p,p+1}^+\, {\mathcal{Y}_{k-p,q}^-}^T \tag{38}$$

Vector K has dimension qrs, where $s = (p+1)r - m$.

2.3.3 Practical implementation

As mentioned above, the practical implementation of subspace-based estimation algorithms does not rely on solving the nonlinear optimization problem in (34). It proceeds as follows [2, 3, 29]. The SVD[4] of the weighted empirical Hankel matrix splits up in two terms :

$$W_1\, \widehat{\mathcal{H}}_{p+1,q}\, W_2^T = UDV^T + \sum_{k=m+1}^{qr} \sigma_k u_k v_k^T \tag{39}$$

[4] Or alternatively the QR-decomposition.

where the last $(qr - m)$ singular values σ_k go to zero when the sample size n goes to infinity. The following approximate factorization holds : $\widehat{\mathcal{H}}_{p+1,q} = \widehat{\mathcal{O}}_{p+1}\, \widehat{\mathcal{C}}_q + \epsilon(n)$, where $\epsilon(n)$ converges to zero at infinity [2, 3, 21]. Then, the remaining steps of the algorithm are similar to those described in equations (14), (16), (17) and (18), where theoretical quantities are replaced with empirical ones.

3 Design of statistical FDI algorithms

In the problem statement given in the introduction, we have implicitly assumed that all the faults of interest are reflected by a change in the parameters of a model of the system [8]. The statistical local approach to the design of FDI algorithms provides us with tools which perform the early detection of slight changes in the system parameters. In this section, we briefly recall its basic principles.

From now on, we are given a reference value θ_0 of the model parameter (for instance, identified with data from the safe system), and a new data sample of size n. The detection problem is to decide whether or not the new data are still well described by this parameter value. The isolation problem is to decide which (subsets of) components of the parameter vector are affected by the change.

It has been widely acknowledged [32, 18, 24] that the FDI problem can be split in 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. In this perspective, taking a statistical point of view, we find it useful to distinguish between a *primary residual*, which is a function of the model parameter and the observed data, and an *improved residual* for making the decision.

We first define these two residuals. Then we describe the χ^2-type test statistics which solves the detection problem and a sensitivity approach to the isolation problem.

3.1 FDI residuals and estimating functions

To begin with, we define primary residuals and we discuss the strong connections with estimating functions used for parameter identification [12, 19, 6]. Then we define improved residuals, of which the evaluation is much easier.

Primary residuals associated with parametric models satisfy two basic requirements.

Definition 3.1 (Primary residuals.) *A vector-valued function $K(\theta, \mathcal{Y}_{k,\varrho}^-)$ is a valid* primary residual *for monitoring parameter vector θ, if it is differentiable in θ, and if there exists a neighborhood $v(\theta_0)$ such that :*

$$\mathbf{E}_\theta\, K(\theta_0, \mathcal{Y}_{k,\varrho}^-) = 0 \;\; \textit{if} \;\; \theta = \theta_0 \tag{40}$$

$$\mathbf{E}_\theta\, K(\theta_0, \mathcal{Y}_{k,\varrho}^-) \neq 0 \;\; \textit{if} \;\; \theta \in v(\theta_0) \setminus \theta_0\ , \tag{41}$$

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

$$M(\theta_0) \stackrel{\text{def}}{=} -\,\mathbf{E}_{\theta_0}\, \frac{\partial}{\partial \theta}\, K(\theta, \mathcal{Y}_{k,\varrho}^-)\Big|_{\theta=\theta_0} \tag{42}$$

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

$$\operatorname{rank} M(\theta_0) = \dim \Theta = l \tag{43}$$

Some comments are in order on the requirements (40), (41) and (43). First, the empirical counterpart of (40), that is :

$$\frac{1}{n-\varrho+1}\sum_{k=\varrho}^{n} K(\theta, \mathcal{Y}_{k,\varrho}^{-}) = 0 \tag{44}$$

writes as an estimating equation : an estimate $\widehat{\theta}$ is found as a root, in θ, of this equation. Actually, requirements (40)-(41) characterize an *estimating function* [19, 6]. Moreover, referring to the discussion in subsection 2.2, requirement (40) means that the exact model order is known and there is no model reduction.

Second, in statements (40)-(41) and definition (42), it is assumed that expectation $\mathbf{E}_\theta$ exists and does not depend on time index k, at least for large k. In other words, the process $K(\theta, \mathcal{Y}_{k,\varrho}^{-})$ is assumed to be asymptotically stationary. Moreover, requirement (40)-(41) is exactly a local identifiability assumption.

Third, for condition (43) to hold, vector-valued function K should have dimension greater than or equal to the parameter space dimension l. Finally, because of (40), Jacobian matrix (42) also writes :

$$\begin{aligned} M(\theta_0) &= -\frac{\partial}{\partial\theta}\,\mathbf{E}_{\theta_0}\, K(\theta, \mathcal{Y}_{k,\varrho}^{-})\Big|_{\theta=\theta_0} \\ &= +\frac{\partial}{\partial\theta}\,\mathbf{E}_{\theta}\, K(\theta_0, \mathcal{Y}_{k,\varrho}^{-})\Big|_{\theta=\theta_0} \end{aligned} \tag{45}$$

As discussed further in section 4.1, in the case of estimating functions K of the form (37), the last expression is more appropriate for deriving necessary and sufficient conditions for isolation requirement (43) to hold.

The statistical approach to residual evaluation requires the knowledge of the statistical properties of the residuals. Unfortunately, even for simple linear models and classical estimating functions, the distribution of $K(\theta, \mathcal{Y}_{k,\varrho}^{-})$ is unknown. One manner to circumvent this difficulty is to accumulate primary residuals and to use a local approach, that is to assume close hypotheses :

$$(\textit{Safe system})\ \mathbf{H}_0 :\ \theta = \theta_0 \quad \text{and} \quad (\textit{Faulty system})\ \mathbf{H}_1 :\ \theta = \theta_0 + \frac{\Upsilon}{\sqrt{n}} \tag{46}$$

where vector Υ is unknown, but fixed. Note that for large n, hypothesis $\mathbf{H}_1$ in (46) corresponds to small deviations in θ.

Definition 3.2 (Improved residuals.) *Given a primary residual K and a n-size sample of data $\mathcal{Y}_1^n$, the* improved residual *is defined as :*

$$\zeta_n(\theta) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n-\varrho+1}}\sum_{k=\varrho}^{n} K(\theta, \mathcal{Y}_{k,\varrho}^{-}) \tag{47}$$

As formally stated below, for primary residuals (estimating functions) K satisfying (40)-(41) and regular enough, this improved residual turns out to be asymptotically Gaussian distributed, under both $\mathbf{H}_0$ and $\mathbf{H}_1$, making easy its evaluation.

3.2 Fault detection

The following central limit theorem (CLT) holds for large classes of dynamical processes and estimating functions K, as proven in [4, 12, 13, 33, 15]. Let $\Sigma(\theta_0)$ be defined by :

$$\Sigma(\theta_0) \stackrel{\text{def}}{=} \lim_{n\to\infty} \mathbf{E}_{\theta_0}\left(\zeta_n\zeta_n^T\right) = \lim_{n\to\infty} \frac{1}{n-\varrho+1}\sum_{k=\varrho}^{n}\sum_{j=\varrho}^{n} \mathbf{E}_{\theta_0}\left(K(\theta_0, \mathcal{Y}_{k,\varrho}^{-})\, K^T(\theta_0, \mathcal{Y}_{j,\varrho}^{-})\right) \tag{48}$$

Theorem 3.1 (CLT.) *Provided that $\Sigma(\theta_0)$ is positive definite, the improved residual ζ_n defined in (47) is asymptotically Gaussian distributed under both hypotheses defined in (46); that is :*

$$\zeta_n(\theta_0) \rightsquigarrow \begin{cases} \mathcal{N}(\ 0,\ \Sigma(\theta_0)\ ) & \textit{under} \quad \mathbf{H}_0 \\ \\ \mathcal{N}(\ M(\theta_0)\ \Upsilon,\ \Sigma(\theta_0)\ ) & \textit{under} \quad \mathbf{H}_1 \end{cases} \tag{49}$$

where $M(\theta_0)$ is defined in (42).

This theorem means that a deviation in the system parameter θ is reflected into a change in the mean value of residual ζ_n, which is asymptotically Gaussian distributed with constant covariance matrix. The sensitivity of residual ζ_n w.r.t. parameter θ is captured by the Jacobian matrix $M(\theta)$. The uncertainty in residual ζ_n is captured by the covariance matrix $\Sigma(\theta)$. Note that $M(\theta_0)$ and $\Sigma(\theta_0)$ do not depend on the particular fault vector Υ in hypothesis $\mathbf{H}_1$. Thus we do not need to re-estimate them when testing the hypotheses. They can be estimated prior to testing, using data on the safe system (exactly as the reference parameter θ_0). Matrix M can be estimated from a data sample using a simple sample average, whereas the estimation of matrix Σ is more tricky [33]. Let $\widehat{M}, \widehat{\Sigma}$ consistent estimates[5] of M, Σ. Deciding that residual ζ_n is *significantly* different from zero can be achieved as follows.

Theorem 3.2 (χ^2**-test.**) *Assume additionally that Jacobian matrix M is f.c.r., that is its rank is equal to parameter space dimension l. Then the test between the hypotheses $\mathbf{H}_0$ and $\mathbf{H}_1$ defined in (46) is achieved through :*

$$\chi^2_n \stackrel{\text{def}}{=} \zeta_n^T\ \widehat{\Sigma}^{-1}\ \widehat{M}\left(\widehat{M}^T\ \widehat{\Sigma}^{-1}\ \widehat{M}\right)^{-1}\widehat{M}^T\ \widehat{\Sigma}^{-1}\ \zeta_n \tag{50}$$

which should be compared to a threshold. In (50), the dependence on θ_0 has been removed for simplicity. Test statistics χ^2_n is asymptotically distributed as a χ^2-variable, with l degrees of freedom[6]*, and with non-centrality parameter under $\mathbf{H}_1$:*

$$\Upsilon^T\ M^T\ \Sigma^{-1}\ M\ \Upsilon \tag{51}$$

The following lemma states a basic invariance property enjoyed by this type of χ^2-tests, which is of particular importance for subspace-based FDI algorithms as we discuss further below.

Lemma 3.1 *The χ^2-test in (50) is left unchanged when the estimating function K in (47) is pre-multiplied by an invertible matrix gain T.*

Actually, ζ_n and $\widehat{M}$ are then also pre-multiplied by T, matrix $\widehat{\Sigma}$ becomes $T\ \widehat{\Sigma}\ T^T$, and invertible matrix T factors out when substituting these transformed matrices in test (50).

3.3 Fault isolation

The improved residual ζ is distributed as a Gaussian vector $\mathcal{N}(M\Upsilon, \Sigma)$. We now partition the fault vector as :

$$\Upsilon = \begin{pmatrix} \Upsilon_a \\ \Upsilon_b \end{pmatrix}$$

with Υ_a, Υ_b of dimensions l_a, l_b respectively ($l_a + l_b = l$), and the fault incidence and Fisher information matrices accordingly, namely :

$$M = \begin{pmatrix} M_a & M_b \end{pmatrix}$$

[5] Explicit formulas are given in section 4.

[6] If Jacobian matrix M is not f.c.r., it should be reduced to a f.c.r. matrix. The number of degrees of freedom of the χ^2-test is then the column rank of the reduced matrix.

and :

$$\mathbf{I} \stackrel{\text{def}}{=} M^T \, \Sigma^{-1} \, M = \begin{pmatrix} \mathbf{I}_{aa} & \mathbf{I}_{ab} \\ \mathbf{I}_{ba} & \mathbf{I}_{bb} \end{pmatrix}$$

The problem now is to detect changes in Υ_a, with Υ_b being unknown and considered as a nuisance parameter.

A rather intuitive statistical solution to the isolation problem consists in projecting the deviations in Υ onto the subspace generated by the components to be isolated. This can be viewed as a sensitivity test [9, 5].

Lemma 3.2 *Sensitivity test $\widetilde{t}_a$ for monitoring Υ_a is* GLR *test between $\Upsilon = (0,0)$ and $\Upsilon = (\Upsilon_a, 0)$, where $\Upsilon_a \neq 0$, namely :*

$$\widetilde{t}_a = 2 \, \ln \frac{\max_{\Upsilon_a} p_{\Upsilon_a,0}(\zeta)}{p_{0,0}(\zeta)}$$

Straightforward computations [5] show that this test writes :

$$\widetilde{t}_a = \widetilde{\zeta}_a^T \, \mathbf{I}_{aa}^{-1} \, \widetilde{\zeta}_a$$

where :

$$\widetilde{\zeta}_a = \widetilde{\mathcal{G}}_a \, \zeta \stackrel{\text{def}}{=} M_a^T \, \Sigma^{-1} \, \zeta$$

and :

$$\widetilde{\mathcal{G}}_a \, M_a = \mathbf{I}_{aa} = \widetilde{\mathcal{G}}_a \, \Sigma \, \widetilde{\mathcal{G}}_a^T$$

Under $\Upsilon_a \neq 0$, it is distributed as a χ^2-random variable with l_a degrees of freedom and non-centrality parameter $\Upsilon_a^T \, \mathbf{I}_{aa} \, \Upsilon_a$.

Another possible solution to the isolation problem consists in a statistical rejection of the nuisance parameter Υ_b, which is called min-max test, and has connections with sensitivity test. The set of min-max tests monitoring the components of Υ is known to enjoy optimality properties [5]. However, since in the structural vibration monitoring application, we are mainly interested in the diagnostics of the faults in terms of the (non identifiable) finite elements model [17], the sensitivity tests are more appropriate than the min-max ones, which we do not introduce here despite their optimality.

4 Subspace-based FDI algorithms

In this section, we describe the application of the design of FDI algorithms in section 3 to the subspace identification algorithms presented in section 2.

First, we define our test statistics, based on the general form of subspace fitting estimating functions introduced in (37). We investigate several invariance properties of the corresponding χ^2-tests, and we explain how to compute estimates of Jacobian matrix $M(\theta)$ and covariance matrix $\Sigma(\theta)$. Conditions under which the isolation requirement (43) holds are outlined.

Second, we turn back to the case of particular interest in this paper, where Hankel matrix plays the role of data function N. We make use of the system parameter characterization described in section 2, assuming known system order.

Third, we exhibit three important instances of these tests, based on IV, BR and CVA identification methods. Finally, we address the robustness of the proposed subspace-based FDI algorithms w.r.t. model reduction, and we discuss how the subspace-based FDI algorithms proposed in this paper relate to another subspace-based detection algorithm recently proposed [20].

An important comment holds throughout the section. As explained in the introduction, we are interested in the pole part of the system and consider the zero part as a nuisance parameter.

Since the distribution of the considered processes depend on the pole *and* the zero parts, we should index all expectations by both parts. For simplicity, and by an abuse of notation, we note only $\mathbf{E}_\theta$. Any statement involving $\mathbf{E}_\theta$ should thus be understood as holding for *every* given zero part.

4.1 Subspace-based test statistics

Estimating functions (40)-(41) of special interest in the present paper is the class of subspace fitting algorithms, when the estimating function K is, as in (37), based on the product of a matrix-valued function of the parameter and a matrix-valued function of the data [14], that is :

$$K(\theta, \mathcal{Y}^-_{k,\varrho}) = \mathrm{vec}\left(S^T(\theta)\ N(\mathcal{Y}^-_{k,\varrho})\right) \tag{52}$$

where matrices S and N have s and qr columns, respectively, and matrices S and $\mathbf{E}_\theta N$ are assumed f.c.r. Estimating functions of the form (52) generalize those of the form (37), and several of our results apply to this generalized form as well.

For reasons which already appear in section 2 and are further discussed in subsections 4.2 and 4.3, it is of interest to consider weighted versions $W_1\ N(\mathcal{Y}^-_{k,\varrho})\ W_2^T$ of the function of the data, and to investigate the case where S is an implicit function of parameter vector θ, that is when S is defined by :

$$S^T\ W_1\, U(\theta) = 0 \tag{53}$$

$$S^T\ S = I_s \tag{54}$$

where $U(\theta)$ is a given f.c.r. matrix with dimensions $(p+1)r \times u$. Of course, such a matrix S is non unique : two such matrices are related through pre-multiplication with an orthonormal matrix T. We stress that, because of (53), matrix S depends implicitly on parameter θ. As already mentioned, even though S is not unique, in the sequel we treat it as a function of parameter θ, denoted by $S(\theta)$, which we fully justify with the aid of the invariance property stated below.

We thus consider the class of estimating functions of the form :

$$K_{W_1,W_2,T}(\theta, \mathcal{Y}^-_{k,\varrho}) \stackrel{\mathrm{def}}{=} \mathrm{vec}\left(T\ S^T(\theta)\ W_1\ N(\mathcal{Y}^-_{k,\varrho})\ W_2^T\right) \tag{55}$$

where W_1, W_2, T are three invertible matrices with compatible dimensions. These matrices are design choices.

Using lemma 3.1, we first state invariance properties of χ^2-tests (50)-(47) built on (55) and (53)-(54).

4.1.1 Invariance properties

We show how weighting N, the function of the data, affects the estimating function in(55), and consequently the resulting χ^2-test. We altogether discuss the effect of the indetermination in S, the function of the parameter satisfying constraints (53)-(54).

Lemma 4.1 *Let $S_I^T(\theta)$ be an orthonormal matrix satisfying constraint (53) with identity weight $W_1 = I$, and let :*

$$K_I(\theta, \mathcal{Y}^-_{k,\varrho}) \stackrel{\mathrm{def}}{=} K_{I,I,I}(\theta, \mathcal{Y}^-_{k,\varrho}) = \mathrm{vec}\left(S_I^T(\theta)\ N(\mathcal{Y}^-_{k,\varrho})\right)$$

be the corresponding estimating function. Then :

$$K_{W_1,W_2,T}(\theta, \mathcal{Y}^-_{k,\varrho}) = (W_2 \otimes T\, T_I)\ K_I(\theta, \mathcal{Y}^-_{k,\varrho})$$

where T_I is the unique invertible matrix such that : $S^T(\theta)\ W_1 = T_I\ S_I^T(\theta)$.

Since matrix $(W_2 \otimes T\ T_I)$ is invertible, from lemmas 4.1 and 3.1, we deduce the following invariance properties of subspace-based χ^2-tests.

Theorem 4.1 (Invariance of subspace-based χ^2-tests.) *χ^2-tests in (50)-(47) based on estimating functions K defined in (55) under the constraints (53)-(54) are, for given functions U and N, invariant with respect to particular choices of invertible matrices T, W_1 and W_2 of compatible dimensions.*

This theorem means that, for any subspace-based estimating function (55) satisfying (40)-(41), the corresponding χ^2-tests introduced in section 3 are invariant with respect to the weighting matrices in the data function N and the choice of the particular orthonormal matrix S satisfying constraint (53).

Actually, this invariance property of these subspace-based χ^2-tests statistics – or equivalently of their non-centrality parameter – can also be seen as a consequence of two fairly general results. The first one [15] links the asymptotic covariance of parameter estimates based on an estimating function K to the non-centrality parameter of local χ^2-tests based on the same function[7]. Based also on estimating functions (called pseudo-scores), the second result [14] states that the asymptotic covariance matrix of subspace fitting estimates depend only on the subspaces spanned by the matrix-valued functions of the parameter and of the data respectively, and not on the particular choices of these functions.

We now explain how to compute estimates of Jacobian matrix $M(\theta)$ and covariance matrix $\Sigma(\theta)$, still in the general framework of estimating functions (55). Other formulas are given below, in the case of subspace-based algorithms, where Hankel matrix plays the role of data function N.

4.1.2 Jacobian matrix

We note that function K in (55) also writes :

$$K(\theta, \mathcal{Y}^-_{k,\varrho}) = \Big(W_2\ N^T(\mathcal{Y}^-_{k,\varrho})\ W_1^T \otimes T\Big)\ \mathcal{S}(\theta) \tag{56}$$

$$= \Big(W_2 \otimes T\ S^T(\theta)\ W_1\Big)\ \mathcal{N}(\mathcal{Y}^-_{k,\varrho}) \tag{57}$$

where $\mathcal{S} \stackrel{\text{def}}{=} \text{vec}(S^T)$ and $\mathcal{N} \stackrel{\text{def}}{=} \text{vec}(N)$, and that constraint (53) also writes :

$$\Big(U^T(\theta)\ W_1^T \otimes I_s\Big)\ \mathcal{S}(\theta) = \Big(I_u \otimes S^T(\theta)\ W_1\Big)\ \text{vec}\, U(\theta) = 0 \tag{58}$$

A first computation of the Jacobian matrix is based on the mean deviation formula (42). We explain below that, when Hankel matrix plays the role of data function N, a second computation of Jacobian matrix can usefully be based on formula (45).

From (56), we find that Jacobian matrix (42) writes :

$$M(\theta) = -\ \Big(W_2\ \mathbf{E}_\theta N^T(\mathcal{Y}^-_{k,\varrho})\ W_1^T \otimes T\Big)\ \frac{\partial \mathcal{S}(\theta)}{\partial \theta} \tag{59}$$

When S satisfies constraint (53), or equivalently (58), Jacobian matrix M is computed by substituting in (59) a solution of the linear equation :

$$\Big(U^T(\theta)\ W_1^T \otimes I_s\Big)\ \frac{\partial \mathcal{S}(\theta)}{\partial \theta} = -\ \Big(I_u \otimes S^T(\theta)\ W_1\Big)\ \mathcal{U}'(\theta) \tag{60}$$

[7] The result in [15] has an interesting consequence on modal identification, as outlined in the conclusion of this paper.

where :

$$\mathcal{U}'(\theta) \stackrel{\text{def}}{=} \frac{\partial \operatorname{vec} U(\theta)}{\partial \theta} \tag{61}$$

We stress that, even though the solution $\partial \mathcal{S}(\theta)/\partial\theta$ of (60) is non unique, the right hand-side of (59) can easily be checked to be unique, provided that :

$$\operatorname{Ker}\left(U^T(\theta)\, W_1^T \otimes I_s\right) \subset \operatorname{Ker}\left(W_2\, \mathbf{E}_\theta N^T(\mathcal{Y}^-_{k,\varrho})\, W_1^T \otimes T\right) \tag{62}$$

Necessary conditions for the isolation requirement. From (59) and (60), we deduce conditions on f.c.r. matrices $\mathbf{E}_\theta N, S$ and U in (55)-(53)-(54), which are *necessary* for isolation requirement (43) to hold. Remembering that $\operatorname{rank} S = s$, we must have :

$$\operatorname{rank} \mathbf{E}_\theta N \times \operatorname{rank} S \geq \dim \Theta \tag{63}$$

$$\mathcal{U}'(\theta) \quad \text{f.c.r.} \tag{64}$$

$$\operatorname{rank} \frac{\partial \mathcal{S}(\theta)}{\partial \theta} \geq \dim \Theta \tag{65}$$

4.1.3 Covariance matrix

From (57), we find that covariance matrix (48) writes :

$$\Sigma(\theta) = \frac{1}{n-\varrho+1}\, \left(W_2 \otimes T\, S^T(\theta)\, W_1\right)\, \left(\sum_{k=\varrho}^{n} \sum_{j=\varrho}^{n} \mathbf{E}_\theta \left(\mathcal{N}(\mathcal{Y}^-_{k,\varrho})\, \mathcal{N}^T(\mathcal{Y}^-_{j,\varrho})\right)\right)\, \left(W_2^T \otimes W_1^T\, S(\theta)\, T^T\right)$$

A consistent[8] estimate is :

$$\widehat{\Sigma}(\theta) = \frac{1}{n-\varrho+1}\, \left(W_2 \otimes T\, S^T(\theta)\, W_1\right)\, \left(\sum_{k=\varrho}^{n} \sum_{j=\varrho}^{n} \mathcal{N}(\mathcal{Y}^-_{k,\varrho})\, \mathcal{N}^T(\mathcal{Y}^-_{j,\varrho})\right)\, \left(W_2^T \otimes W_1^T\, S(\theta)\, T^T\right)$$

4.2 Hankel matrix as data function - Known system order

We now turn back to the case of particular interest in this paper, where Hankel matrix plays the role of data function N, namely :

$$N(\mathcal{Y}^-_{k,\varrho}) \stackrel{\text{def}}{=} \mathcal{Y}^+_{k-p,p+1}\, {\mathcal{Y}^-_{k-p,q}}^T \tag{66}$$

Of course, we make use of the system parameter characterization (35)-(36), assuming known system order. This means that the parameter constraint function is now :

$$U(\theta) \stackrel{\text{def}}{=} \mathcal{O}_{p+1}(\theta)$$

with column rank $u = m$.

We stress that the invariance theorem 4.1 means that, theoretically, when the exact system order is known and no model reduction is enforced, the subspace-based χ^2-tests introduced in this paper are invariant with respect to the weighting matrices in the Hankel matrix $\mathcal{H}$ and the choice of the particular orthonormal matrix S satisfying constraint (22). We note however the above theorem states no invariance property in the case of model reduction, generically encountered in practice.

Now we first discuss the above necessary conditions for isolation requirement to hold. Then we describe a second computation of Jacobian matrix M, from which we derive necessary *and* sufficient conditions. Finally, we suggest another estimate of covariance matrix Σ.

[8] Sufficient conditions for this to be true are mentioned below.

4.2.1 Jacobian matrix - First computation

From (59) and (26), we deduce that Jacobian matrix writes :

$$M(\theta) = - \left(W_2 \, \mathcal{H}_{p+1,q}^T \, W_1^T \otimes T \right) \, \frac{\partial \mathcal{S}(\theta)}{\partial \theta}$$

and a consistent estimate, based on a data sample, is found using (32) :

$$\widehat{M}(\theta) = - \left(W_2 \, \widehat{\mathcal{H}}_{p+1,q}^T \, W_1^T \otimes T \right) \, \frac{\partial \mathcal{S}(\theta)}{\partial \theta} \tag{67}$$

where $\partial \mathcal{S}(\theta)/\partial\theta$ is a solution of (60).

In this case, the above necessary conditions for isolation requirement (43) to hold write as follows. First, for N in (66), we have : $\operatorname{rank} N = m$, and thus condition (63) writes[9] : $sm \geq l = rm$, and thus $s \geq r$. Remembering that $s = (p+1)r - m$, from $s \geq r$, we get : $pr \geq m$. This last condition is stronger than condition : $(p+1)r \geq m$, which results from (11). This stronger condition is intuitively reasonable, since it means that we should use more equations for isolation than for identification.

Assuming that the exact model order is known, and using factorization property (19) of Hankel matrix, it is straightforward to show that (62) holds, and thus Jacobian matrix $M(\theta)$ does not depend on the particular solution $\partial \mathcal{S}(\theta)/\partial\theta$.

4.2.2 Jacobian matrix - Second computation

As mentioned above, Jacobian matrix can also be computed from formula (45) which, from (55), writes :

$$M(\theta_0) = \frac{\partial}{\partial \theta} \, \operatorname{vec} \left(T \, S^T(\theta_0) \, W_1 \, \mathbf{E}_\theta N(\mathcal{Y}_{k,\varrho}^-) \, W_2^T \right) \Big|_{\theta=\theta_0} \tag{68}$$

and, in the present case :

$$M(\theta_0) = \frac{\partial}{\partial \theta} \, \operatorname{vec} \left(T \, S^T(\theta_0) \, W_1 \, \mathcal{H}_{p+1,q} \, W_2^T \right) \Big|_{\theta=\theta_0}$$

Thanks to the factorization (19) of the Hankel matrix, this also writes :

$$M(\theta_0) = \frac{\partial}{\partial \theta} \, \operatorname{vec} \left(T \, S^T(\theta_0) \, W_1 \, \mathcal{O}_{p+1}(\theta) \, \mathcal{C}_q \, W_2^T \right) \Big|_{\theta=\theta_0} \tag{69}$$

From this formula, and using :

$$\mathcal{O}'_{p+1}(\theta) \stackrel{\text{def}}{=} \frac{\partial \operatorname{vec} \mathcal{O}_{p+1}(\theta)}{\partial \theta} \, , \tag{70}$$

we easily get :

$$\begin{aligned} M(\theta) &= \left(W_2 \, \mathcal{C}_q^T \otimes T \, S^T(\theta) \, W_1 \right) \, \mathcal{O}'_{p+1}(\theta) \\ &= \left(W_2 \otimes T \, S^T(\theta) \, W_1 \right) \, \left(\mathcal{C}_q^T \otimes I_{(p+1)r} \right) \, \mathcal{O}'_{p+1}(\theta) \\ &= \left(W_2 \otimes T \, S^T(\theta) \, W_1 \right) \, \left(\mathcal{H}_{p+1,q}^T \, \mathcal{O}_{p+1}^{\dagger \, T}(\theta) \otimes I_{(p+1)r} \right) \, \mathcal{O}'_{p+1}(\theta) \end{aligned} \tag{71}$$

The last equality, where $\mathcal{O}_{p+1}^\dagger(\theta)$ is the pseudo-inverse of $\mathcal{O}_{p+1}(\theta)$, is obtained using the factorization property (19) again.

[9] We here recover the usual assumption on the number of parity checks, which should be greater than or equal to the number of faults.

A consistent estimate, based on a data sample, is found using (32) :

$$\widehat{M}(\theta) = \left(W_2 \otimes T\, S^T(\theta)\, W_1\right)\, \left(\widehat{\mathcal{H}}^T_{p+1,q}\, \mathcal{O}^{\dagger\, T}_{p+1}(\theta) \otimes I_{(p+1)r}\right)\, \mathcal{O}'_{p+1}(\theta) \tag{72}$$

Some comments are in order on formula (71), which also apply to (72).

First, it is easily checked that $M(\theta)$ does *not* depend on the particular choice of the eigenvectors ϕ_λ contained in the parameter vector θ defined in (3). Multiplying the ϕ_λ's by constant complex numbers amounts to post-multiply observability matrix $\mathcal{O}_{p+1}(\theta)$ in (9) by an invertible diagonal matrix D, to post-multiply matrix $\mathcal{O}^{\dagger\, T}_{p+1}(\theta)$ by D^T, and to pre-multiply matrix $\mathcal{O}'_{p+1}(\theta)$ by $(D^T \otimes I_{(p+1)r})$. And all the terms in D cancel out in (71).

Second, matrix $\mathcal{O}'_{p+1}(\theta)$ is full rank $(r+1)m$, as can be checked directly from the explicit formula (90). Conditions regarding the rank of $M(\theta)$ are thus as follows.

4.2.3 Conditions for the isolation requirement

We now investigate *necessary and sufficient* conditions for $M(\theta)$ to be full rank mr, or equivalently for isolation requirement (43) to hold.

Formula (69) also writes :

$$M(\theta_0) = \left(W_2\, \mathcal{C}_q^T \otimes T\right)\, \left(\frac{\partial}{\partial \theta}\, \mathrm{vec}\left(S^T(\theta_0)\, W_1\, \mathcal{O}_{p+1}(\theta)\right)\Big|_{\theta=\theta_0}\right)$$

For matrix $(W_2\, \mathcal{C}_q^T \otimes T)$ to be f.c.r. *and* dimension of $\mathrm{vec}\left(S^T(\theta_0)\, W_1\, \mathcal{O}_{p+1}(\theta)\right)$ to be large enough, we need the single condition : $sm \geq l = rm = \dim \Theta$. We thus deduce the necessary conditions :

$$sm \;\geq\; \dim \Theta \tag{73}$$

$$\mathrm{rank}\left(\frac{\partial}{\partial \theta}\, \mathrm{vec}\left(S^T(\theta_0)\, W_1\, \mathcal{O}_{p+1}(\theta)\right)\Big|_{\theta=\theta_0}\right) \;\geq\; \dim \Theta \tag{74}$$

These conditions are sufficient for isolation requirement (43) to hold, since $(W_2\, \mathcal{C}_q^T \otimes T)$ is f.c.r.

4.2.4 Covariance matrix.

When N is as in (66), function K in (55) also writes :

$$K(\theta, \mathcal{Y}^-_{k,\varrho}) = Z_k \otimes V_k \tag{75}$$

where :

$$Z_k \;\stackrel{\mathrm{def}}{=}\; W_2\, \mathcal{Y}^-_{k-p,q} \tag{76}$$

$$V_k \;\stackrel{\mathrm{def}}{=}\; T\, S^T(\theta)\, W_1\, \mathcal{Y}^+_{k-p,p+1} \tag{77}$$

From (75), we find that covariance matrix writes :

$$\Sigma(\theta) = \frac{1}{n-\varrho+1} \sum_{k=\varrho}^{n} \sum_{j=\varrho}^{n} \mathbf{E}_\theta \left(Z_k\, Z_j^T \otimes V_k\, V_j^T\right) \tag{78}$$

Remembering the design of matrix $S(\theta)$ in (22) and the factorization property in (19), it is easy to show that, for every k, vector V_k is orthogonal to $\mathcal{Y}^-_{k-p-1,\infty}$. Also, for every $i \geq 0$, vector Z_{k-i} belongs to $\mathrm{Span}\, \mathcal{Y}^-_{k-p,\infty}$, and vector V_{k-i} belongs to $\mathrm{Span}\, \mathcal{Y}^+_{k-p-i,p+1}$ and thus to $\mathrm{Span}\, \mathcal{Y}^-_{k-i,\infty}$.

Consequently, for every $i \geq p$, vector V_{k-i} belongs to Span $\mathcal{Y}^-_{k-p,\infty}$. Thus, V_k is, by construction, independent of Z_k and of (Z_j, V_j) for $|k-j| \geq p$, and the above expression boils down to :

$$\Sigma(\theta) = \frac{1}{n-\varrho+1} \sum_{k=\varrho}^{n} \sum_{i=-p+1}^{p-1} \mathbf{E}_\theta \left(Z_k \; Z_{k-i}^T \otimes V_k \; V_{k-i}^T \right)$$

An estimate is :

$$\widehat{\Sigma}(\theta) = \frac{1}{n-\varrho+1} \sum_{k=\varrho}^{n} \sum_{i=-p+1}^{p-1} \left(Z_k \; Z_{k-i}^T \otimes V_k \; V_{k-i}^T \right) \tag{79}$$

Some important comments are in order about matrices $\Sigma(\theta)$ and $\widehat{\Sigma}(\theta)$. For a particular but difficult case[10], sufficient conditions are given in [23] under which Σ in (78) is asymptotically uniformly positive definite and bounded, and the estimate $\widehat{\Sigma}(\theta)$ in (79) is consistent. We do not know, however, how far from necessary are these conditions, and how they translate on the data function $N(\mathcal{Y}^-_{k,\varrho})$. From a numerical point of view, estimates as in (79) turn out to be often ill-conditioned. A much better solution consists in estimating $\widehat{\Sigma}$ using empirical means of cross products of residual ζ, which can be computed using only covariance data [17].

4.3 IV, BR, and CVA-based FDI algorithms

As an illustrative example, we now turn to the FDI problem we address in this paper, that is monitoring the eigenstructure of a linear dynamical system. First, we discuss which estimating functions should be used as primary residuals in this case. Then we recall a FDI algorithm previously proposed by some of the authors [9], which is based on the instrumental variable approach to eigenstructure identification. Finally, we give two other particular examples of the family of subspace-based FDI algorithms introduced above, namely the BR and CVA-algorithms.

4.3.1 Residuals for eigenstructure monitoring

Classical detection theory, based on the likelihood ratio, suggests that the gradient of the likelihood function, called efficient score, is a valid primary residual. This is generically true, even though this may lead to computationally complex detectors [8]. However, for the type of faults in models (1) or equivalently (4) which we consider in this paper, the likelihood function is no longer a relevant primary residual. As a matter of fact, the present monitoring problem is restricted to the AR parameters of the multi-dimensional ARMA model (4), the MA part playing the role of nuisance parameters. The prominent feature of this case is the intrinsic coupling, in the likelihood function, between the AR and MA parameters. This coupling shows up in the non-zero block-off-diagonal terms of Fisher information matrix. From this rises the need for alternative functions of the model parameters and the data which should be used as primary residuals.

4.3.2 IV-based FDI algorithm

It is well known that the identification of the AR parameters of a multi-dimensional ARMA model :

$$Y_k = \sum_{i=1}^{p} A_i \; Y_{k-i} + \sum_{j=0}^{p-1} B_j \; E_{k-j}$$

can be achieved by solving, in least squares sense if $q > p$, the following system of delayed Yule-Walker equations :

$$\left(\begin{array}{cc} A^T & -I_r \end{array} \right) \widehat{\mathcal{H}}_{p+1,q} = 0 \tag{80}$$

[10] The observed process is scalar, and its MA part is non-stationary.

where :

$$A^T \stackrel{\text{def}}{=} \left(\begin{array}{ccc} A_p & \dots & A_1 \end{array} \right)$$

From the estimated autoregressive matrices, a modal parameter θ (3) is deduced by eigen-analysis of (5). We stress that, for a given (reference) value θ, matrix A is an implicit function of parameter θ, defined by the equation :

$$\left(\begin{array}{cc} A^T(\theta) & -I_r \end{array} \right) \mathcal{O}_{p+1}(\theta) = 0 \tag{81}$$

which can be shown to write under the form (53) with a matrix S satisfying (54) and with $U = \mathcal{O}_{p+1}$. To this end, we consider the QR-decomposition :

$$\begin{array}{rcl} \left(\begin{array}{c} A(\theta) \\ -I_r \end{array} \right) & = & \left(\begin{array}{cc} S & \widetilde{S} \end{array} \right) \left(\begin{array}{c} B \\ 0 \end{array} \right) \\ & = & S\, B \end{array} \tag{82}$$

where matrices $\left(\begin{array}{cc} S & \widetilde{S} \end{array} \right)$ and B are orthonormal and upper triangular, respectively. Of course, matrices $S, \widetilde{S}, B$ depends on θ. From the second equation, we deduce that matrices S and B are orthonormal (namely $S^T S = I_s$) and invertible, respectively, and that the left hand-side of (81) writes :

$$\left(\begin{array}{cc} A^T(\theta) & -I_r \end{array} \right) \mathcal{O}_{p+1}(\theta) = B^T\, S^T\, \mathcal{O}_{p+1}(\theta)$$

Turning back to (80), we get the IV estimating function in the form (55), namely :

$$K(\theta, \mathcal{Y}^-_{k,\varrho}) = \mathrm{vec}\left(B^T(\theta)\, S^T(\theta)\, N(\mathcal{Y}^-_{k,\varrho}) \right),$$

with $B(\theta)$ and $S(\theta)$ defined in (82) and $N(\mathcal{Y}^-_{k,\varrho})$ as in (66). The dimension of K is now $qrs = qr^2$. Note that, thanks to the invariance property stated in theorem 4.1, the particular choice of invertible matrix B plays no role in the resulting χ^2-test.

An estimate of Jacobian matrix M is computed by substituting in (67) a solution of (60)-(61), with $U = \mathcal{O}_{p+1}$ as defined in (81). The comment made above of the invariance of M w.r.t. the particular solution of (60) holds in this case.

For computing the covariance matrix, we write the estimating function K under the alternative form in (75), with :

$$Z_k \stackrel{\text{def}}{=} \mathcal{Y}^-_{k-p,q} \tag{83}$$

$$V_k \stackrel{\text{def}}{=} -\, B^T(\theta)\, S^T(\theta)\, \mathcal{Y}^+_{k-p,p+1} \tag{84}$$

and with B and S defined in (82). Note that vector V_k in (84) is nothing but the MA part of the observed process $(Y_k)_k$. Using the same reasoning as above, we find that an estimate of the covariance matrix is given in (79), with Z_k and V_k defined in (83)-(84). Also a covariance-driven formula is given in [17].

4.3.3 BR and CVA-based FDI algorithms

We now turn to weighted estimating functions of the form (55)-(53)-(54) with constraint function : $U = \mathcal{O}_{p+1}$, and we discuss two particular choices of weighting matrices W_1 and W_2.

The first choice corresponds to the BR-identification method, for which weights W_1 and W_2 are identity matrices. The second choice corresponds to the CVA-identification method, for which weights W_1 and W_2 are given in (27).

The main difference with the IV-based FDI algorithm is in the dimension of orthonormal matrix S, and thus in the number of degrees of freedom of the χ^2-test. Additional differences might result from numerical issues on one hand, and from the effect of model reduction on the other one.

4.4 Model reduction : bias-adjusted estimating functions

We now turn to the (actual) situation in which, for the purpose of monitoring, we use a model of order m, while the true system has a larger, possibly infinite, order. As explained in section 2.2, in this situation, relation (23) no longer holds. This means that condition (40) is not fulfilled. Moreover, Hankel matrix no longer factorizes exactly. Thus condition (62) is not fulfilled, and Jacobian matrix $M(\theta)$ is not unique.

Consequently, in the case of model reduction, the above approach to FDI has to be adapted. We proceed in the following way, based on a straightforward intuitive idea. In order to recover requirement (40), when estimating the reference parameter value θ_0, we compute the empirical mean value $\kappa(\theta_0)$ of the estimating function K. Then, when we are given a new data sample, we define the bias-adjusted improved residual as :

$$\zeta_n(\theta_0) = \frac{1}{\sqrt{n-\varrho+1}} \sum_{k=\varrho}^{n} \Big(K(\theta_0, \mathcal{Y}_{k,\varrho}^{-}) - \kappa(\theta_0) \Big)$$

Obviously, this residual satisfies (40). We have shown in [33] that proceeding that way is a valid solution to the fault detection problem. More precisely, in [33] this improved residual is shown to obey the CLT result in section 3, and the performance loss, due to model reduction, of the corresponding χ^2-test is analyzed. Of course, the above bias adjustment may have consequences on fault isolation.

4.5 Discussion

Now we discuss how the subspace-based FDI algorithms proposed in this paper relate to another subspace-based detection algorithm recently proposed [20].

The problem addressed in [20] is to detect changes in the pole *and* and the zero parts of a transfer function from the inputs and the innovation towards the output of a multivariable system, whereas ours is to detect changes in the pole parts of the transfer function from a possibly nonstationary excitation towards the output of a multivariable system.

The design of the detection algorithm is based on subspace identification methods, nicely re-written as a prediction error method. As a consequence, since the subspace-based formulation is not directly used, no analysis of the possible effect of different weighting matrices is performed, whereas in this paper we provide an invariance result in theorem 4.1.

The test statistics in [20] are based on the correlation between this prediction error and the past outputs, whereas ours are based (75) on the correlation between past outputs and the zero part.

One interesting feature of the approach in [20] is that it does not use any canonical parameterization, whereas the modal parameterization does play a key role in our intended application to vibration monitoring. As a consequence, since no parameter is introduced, no analysis of the effect of the change on the test statistics is performed, and the isolation problem is not addressed, whereas in this paper we provide an explicit formula for the mean value of the test statistics under the alternative through the CLTresult in theorem 3.1 and the formulas for the Jacobian matrix M.

5 Robustness with respect to non-stationary excitation

So far, models (1)-(5) have been introduced with time-invariant covariance matrix Q for the excitation ε_k. However, as discussed before, for application to structural vibration monitoring, it is of interest to consider situations in which the excitation, while still being modeled as white

noise, is assumed *non-stationary*, meaning that the covariance matrix :

$$\mathrm{cov}(\varepsilon_k) = Q(k)$$

depends on time index k. This has the following consequences. First, since state and observation processes X_k, Y_k are no longer stationary, matrix :

$$G(k) = \mathbf{E}\left(X_k Y_k^T\right)$$

is also time varying, and the same is true of controllability matrix $\mathcal{C}_p(k)$ in formula (2). Referring to the ARMA form (4), this has as the consequence that the MA coefficients $B_j(k)$ are now time varying also.

Consistency of IV and BR methods (two particular instances of subspace methods) in such a case of non-stationary excitation has been proved in [11], using an appropriate form of uniform controllability assumption for controllability matrix $\mathcal{C}_p(k)$.

Also, IV-based eigenstructure change detection for SISO ARMA processes with non-stationary MA part has been studied in [23], and, with a shorter proof in [15], also encompassing convergence rate for the IV identification procedure. Finally, in [9] it has been argued why likelihood based techniques won't work in such a non-stationary framework, due to coupling between pole and zero parts in the Fisher information matrix of an ARMA process.

We believe that the results of [11] easily extend to the more general subspace methods we study here. The same conjecture seems reasonable for [23, 15], although we think it to be much more difficult.

6 Application to structural vibration monitoring

As mentioned in the introduction, in-situ damage monitoring and predictive maintenance for complex mechanical structures and rotating machines is of key importance in many industrial areas. 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 eigen-frequencies) 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, 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.

The main modeling issues can be summarized in the following manner.

We assume that the behavior of the mechanical system can be described by a stationary linear dynamical 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 :

$$\left\{ \begin{array}{rcl} M\ddot{\mathcal{Z}}(t) + C\dot{\mathcal{Z}}(t) + K\mathcal{Z}(t) & = & \nu(t) \\ Y(t) & = & L\mathcal{Z}(t) \end{array} \right. \tag{85}$$

where t denotes continuous time, M, C, K are the mass, damping and stiffness matrices respectively, (high dimensional) vector $\mathcal{Z}$ 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 eigen-frequencies denoted generically by μ, and the modal shapes or eigenvectors denoted generically by ψ_μ. These quantities are solutions of the following equation :

$$(M\mu^2 + C\mu + K)\ \Psi_\mu = 0 \ \ , \ \psi_\mu = L\ \Psi_\mu \tag{86}$$

Sampling model (85) at rate $1/\tau$ yields the discrete time model in state space form :

$$\left\{ \begin{array}{rcl} X_{k+1} & = & FX_k + \varepsilon_{k+1} \\ Y_k & = & HX_k \end{array} \right. \tag{87}$$

where the state and the output are :

$$X_k = \left[\begin{array}{c} \mathcal{Z}(k\tau) \\ \dot{\mathcal{Z}}(k\tau) \end{array} \right] \ , \ Y_k = Y(k\tau),$$

the state transition and observation matrices are :

$$F = e^{\mathcal{L}\tau} \ , \ \mathcal{L} = \left[\begin{array}{cc} 0 & I \\ -M^{-1}K & -M^{-1}C \end{array} \right] \ , \ H = \left[\begin{array}{cc} L & 0 \end{array} \right],$$

and where state noise ε_{k+1} is zero-mean, white, with covariance matrix :

$$Q_{k+1} \stackrel{\text{def}}{=} \mathbf{E}\left(\varepsilon_{k+1}\ \varepsilon_{k+1}^T\right) = \int_{k\tau}^{(k+1)\tau} e^{\mathcal{L}s}\ \widetilde{Q}(s)\ e^{\mathcal{L}^T s} ds$$

with :

$$\widetilde{Q}(s) = \left[\begin{array}{cc} 0 & 0 \\ 0 & M^{-1}Q_\nu(s)M^{-T} \end{array} \right]$$

The modal characteristics defined in (86) are equivalently found in the eigenstructure $(\lambda, \varphi_\lambda)$ of F :

$$e^{\tau\mu} = \lambda, \quad \psi_\mu = \phi_\lambda \stackrel{\text{def}}{=} H\varphi_\lambda$$

Because of the structure of the state in (87), modes are pairwise complex conjugate. Eigenvectors are real if proportional damping is assumed, that is $C = \alpha M + \beta K$. The corresponding ARMA model writes :

$$Y_k = \sum_{i=1}^{p} A_i\ Y_{k-i} + \sum_{j=0}^{p-1} B_j(k)\ E_{k-j} \tag{88}$$

where $(E_k)_k$ is a standard stationary white noise. Since the covariance matrix of the excitation noise ε 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 [25].

Therefore, 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.

State X and observed output Y have dimensions $2m$ and r respectively, with r (much) smaller than $2m$ in practice[11]. From now on, parameter vector θ is defined by filling the right hand-side of (3) with the half of the $(2m)$ complex conjugate modes. Observability matrix in modal basis (9) now writes [9] :

$$\mathcal{O}_{p+1} = \begin{pmatrix} \Phi & \overline{\Phi} \\ \Phi\Delta & \overline{\Phi\Delta} \\ \vdots & \vdots \\ \Phi\Delta^p & \overline{\Phi\Delta^p} \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{O}_{p+1}(\theta) & \overline{\mathcal{O}_{p+1}(\theta)} \end{pmatrix} \tag{89}$$

It results from (67)-(60), or alternatively from (72) that, whatever the subspace-based FDI algorithm, the only thing we need for computing an estimate of Jacobian matrix $M(\theta)$ is quantity $\mathcal{U}'$ in (61) when parameter constraint matrix U is equal to matrix $\mathcal{O}_{p+1}$ defined in (89). In this case :

$$\mathcal{U}'(\theta) = \left(\begin{array}{c} \mathcal{J}(\theta) \\ \hline \overline{\mathcal{J}(\theta)} \end{array} \right)$$

where :

$$\mathcal{J}(\theta) \stackrel{\text{def}}{=} \frac{\partial \operatorname{vec} \mathcal{O}_{p+1}(\theta)}{\partial \theta} = \left(\begin{array}{ccc|ccc} \Lambda_1'^{(p)} \otimes \phi_1 & & 0 & \Lambda_1^{(p)} \otimes I_r & & 0 \\ & \ddots & & & \ddots & \\ 0 & & \Lambda_m'^{(p)} \otimes \phi_m & 0 & & \Lambda_m^{(p)} \otimes I_r \end{array} \right) \tag{90}$$

with :

$$\Lambda_i^{(p)} \stackrel{\text{def}}{=} \begin{pmatrix} 1 \\ \lambda_i \\ \lambda_i^2 \\ \vdots \\ \lambda_i^p \end{pmatrix}, \qquad \Lambda_i'^{(p)} \stackrel{\text{def}}{=} \begin{pmatrix} 0 \\ 1 \\ 2\,\lambda_i \\ \vdots \\ p\,\lambda_i^{p-1} \end{pmatrix} \qquad (1 \leq i \leq m)$$

From this expression, matrix $\mathcal{U}'(\theta)$ is easily shown to be f.c.r.

7 Conclusion

In this paper, we have systematically investigated how subspace-based methods for eigenstructure identification can be turned into fault detection and isolation methods. This work is based on the general approach advocated by two of the authors for designing algorithms to detect and diagnose slight deviations from nominal behavior in dynamical systems.

From the analysis in this paper and collecting results from [11, 15, 22, 23, 26], we summarize our understanding of subspace-based methods for eigenstructure identification and change detection in the following manner.

- *Relation between identification and change detection.* As our general approach makes a strong interconnection between identification techniques, based on estimating functions,

[11] As explained in [10], when few sensors are available, we identify and monitor a number of modes which is much smaller than the number of solutions of (86), but much larger than the number of sensors.

and associated change detection algorithms, it is not a surprise that the performances of both are intrinsically related. Based on [15], we can assert that, when true model order is used, the covariance matrix of subspace-based eigenstructure identification algorithms is identical to the matrix in the (quadratic) non-centrality parameter (51) of the associated change detection procedure. In particular, we obtain the accuracy of the subspace-based estimates of the parameter vector (3) considered in this paper, namely the modal parameter vector. Up to our knowledge, this seems to be a new result.

- *The role of weighting when true model order is used.* From the present study, we can assert that, when true model order is used for eigenstructure (or pole part) identification, the choice of the weighting matrices in subspace-based methods does not influence the performance, for given sizes of the Hankel matrix $\mathcal{H}$.

- *The role of the number of columns in the Hankel matrix,* again when the true model order is used. From [26], we can assert that the performance (of both the test and identification procedures) increases with the number of columns of the Hankel matrix $\mathcal{H}$, and that Cramer-Rao bound is achieved when this number tends to infinity[12]. This can be actually achieved by filtering the instruments using some appropriate filter, whose choice requires, however, prior knowledge of the zero part of the system.

- *The role of weighting when model reduction is enforced.* From results in stochastic realization theory [22], we know that weighting matrices in subspace methods do play a role for model approximation, and thus in the bias of the estimating function. Finally, from the asymptotic decomposition of estimators into *bias* + *variance* [15], we can assert that weighting matrices contribute to reduce the bias, but not the variance.

- *Robustness.* The following claim is more like a (sensible) conjecture, based on the results of [11, 23][13]. Subspace-based methods for eigenstructure identification or testing are robust with respect to time variations of the zero (or MA) part of the considered system.

- *Range of application.* Our experience [1, 9, 17, 10] suggests that subspace-based methods for eigenstructure identification and testing are useful for in-operation analysis of mechanical structures subject to vibration.

[12] In [26], these results are proved for IV methods, but we now know that it extends to any subspace-based method, thanks to the results of subsection 4.1.1 of this paper.

[13] Considering the technical nature of the referred papers, the actual generalization from IV to subspace-based methods is not expected to be that easy.

References

[1] M. Abdelghani, M. Basseville and A. Benveniste (1997). In-operation damage monitoring and diagnostics of vibrating structures, with applications to offshore structures and rotating machinery. *IMAC-XV, International Modal Analysis Conference*, Orlando, FL.

[2] H. Akaïke (1973). Markovian representation of stochastic processes by canonical variables. *SIAM Jal Control*, vol.13, no 1, pp.162-173.

[3] K.S. Arun and S.Y. Kung (1986). Generalized principal components analysis and its application in approximate stochastic realization. In *Modeling and Application of Stochastic Processes* (U.B. Desai, Ed.), Kluwer Academic Publishers, pp.75-103.

[4] I.V. Basawa (1985). Neyman-Le Cam tests based on estimating functions. In *Proc. Berkeley Conf. honor of Neyman and Kiefer*, vol.2, Le Cam and Olshen (Eds), pp.811-825.

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

[6] M. Basseville (1997). Statistical approaches to industrial monitoring problems - Fault detection and isolation. *Plenary talk, 11th IFAC/IFORS Symposium on Identification and System Parameter Estimation - SYSID'97*, Kitakyushu, Japan, July 8-11, 1997.

[7] M. Basseville (1997). On-board component fault detection and isolation using the statistical local approach. *Research Report* Irisa no 1122/ Inria no 3252.
ftp://ftp.irisa.fr/techreports/1997/PI-1122.ps.gz.

[8] M. Basseville and I.V. Nikiforov (1993). *Detection of Abrupt Changes – Theory and Applications*. Prentice Hall, Englewood Cliffs, N.J.

[9] M. Basseville, A. Benveniste, G. Moustakides and A. Rougée (1987). Detection and diagnosis of changes in the eigenstructure of non-stationary multi-variable systems. *Automatica*, vol.23, no 3, pp.479-489.

[10] 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. *Mechanical Systems and Signal Processing*, vol.7, no 5, pp.401-423.

[11] A. Benveniste and J.-J. Fuchs (1985). Single sample modal identification of a non-stationary stochastic process. *IEEE Trans. on Automatic Control*, vol.AC-30, pp.66-74.

[12] A. Benveniste, M. Basseville and G. Moustakides (1987). The asymptotic local approach to change detection and model validation. *IEEE Trans. Automatic Control*, vol.AC-32, no 7, pp.583-592.

[13] A. Benveniste, M. Métivier and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer, NY.

[14] J.-F. Cardoso and É. Moulines (1996). Invariance of subspace based estimators. Submitted to *IEEE Trans. on Signal Processing*.

[15] B. Delyon, A. Juditsky and A. Benveniste (1997). On the relationship between identification and local tests. *Research Report* Irisa no 1104.
ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz.

[16] U.B. Desai, P. Debajyoti and R.D. Kirkpatrick (1985). A realization approach to stochastic model reduction. *Int. Jal Control*, vol.42, no 4, pp.821-838.

[17] B. Devauchelle-Gach (1991). *Diagnostic Mécanique des Fatigues sur les Structures Soumises à des Vibrations en Ambiance de Travail*. Thesis, Paris IX Dauphine University (in French).

[18] 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.

[19] V.P. Godambe, Ed. (1991). *Estimating Functions*. Clarendon Press, Oxford.

[20] T. Gustafsson (1996). A 4SID-based change detection algorithm. *IFAC'96 World Conference*, San Francisco, CA.

[21] A. Lindquist and G. Picci (1996a). Geometric methods for state space identification. In *Identification, Adaptation, Learning - The Science of Learning Models from Data*, NATO ASI Series F, vol. 153, S. Bittanti and G. Picci (Eds.).

[22] A. Lindquist and G. Picci (1996b). Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica*, Vol. 32, no 5, pp.709-733.

[23] G. Moustakides and A. Benveniste (1986). Detecting changes in the AR part of a non-stationary ARMA process. *Stochastics*, vol.16, pp.137-155.

[24] R.J. Patton, P. Frank and R. Clark, Eds. (1989). *Fault Diagnosis in Dynamic Systems – Theory and Application*. Prentice Hall.

[25] 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, pp.329-342.

[26] A. Rougée, M. Basseville, A. Benveniste and G. Moustakides (1987). Optimum robust detection of changes in the AR part of a multivariable ARMA process. *IEEE Trans. on Automatic Control*, vol. AC-32, no 12, pp.1116-1120.

[27] P. Stoïca and R.L. Moses (1997). *Introduction to Spectral Analysis*. Prentice Hall.

[28] P. Van Overschee and B. De Moor (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, vol.30, no 1, pp. 75-93.

[29] P. Van Overschee and B. De Moor (1996). *Subspace Identification for Linear Systems : Theory - Implementation - Applications*. Kluwer Academic Publishers.

[30] M. Viberg (1995). Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, vol.31, no 12, pp.1835-1853.

[31] M. Viberg, B. Wahlberg and B. Ottersten (1997). Analysis of state space system identification methods based on instrumental variables and subspace fitting. *Automatica*, vol.33, no 9.

[32] A.S. Willsky (1976). A survey of design methods for failure detection in dynamic systems. *Automatica*, vol.12, pp.601-611.

[33] 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.

Contents

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-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

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