

Information Criteria for Residual Generation and Fault Detection and Isolation

Michèle Basseville

► **To cite this version:**

Michèle Basseville. Information Criteria for Residual Generation and Fault Detection and Isolation. [Research Report] RR-2890, INRIA. 1996. <inria-00073800>

HAL Id: inria-00073800

<https://hal.inria.fr/inria-00073800>

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.

***Information criteria for residual generation
and fault detection and isolation***

Michèle Basseville

N° 2890

Mai 1996

————— THÈME 4 —————



***Rapport
de recherche***

Information criteria for residual generation and fault detection and isolation

Michèle Basseville*

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet AS

Rapport de recherche n° 2890 — Mai 1996 — 39 pages

Abstract: Using an information point of view, we discuss deterministic versus stochastic tools for residual generation and evaluation for fault detection and isolation (FDI) in linear time invariant (LTI) state-space systems. In both types of approaches to off-line FDI, residual generation can be viewed as the design of a linear transformation of a Gaussian vector (the finite-window input-adjusted observations). Several statistical isolation methods are revisited, using both a linear transform formulation and the information content of the corresponding residuals. We formally state several multiple fault cases, with or without causality assumptions, and discuss an optimality criterion for the most general one. New information criteria are proposed for investigating the residual optimization problem.

Key-words: Fault detection and isolation, residual generation, residual evaluation, residual optimization, information criteria.

(Résumé : *tsvp*)

Part of this paper is accepted for presentation at IFAC'96 World Conference, under the title *Information criteria for FDI*.

* IRISA/CNRS – basseville@irisa.fr.

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

Critères d'information pour la génération de résidus et la détection et le diagnostic de pannes

Résumé : On prend un point de vue informationnel pour discuter des outils déterministes et stochastiques servant à la génération et l'évaluation de résidus en vue de la détection et du diagnostic de pannes dans les systèmes dynamiques représentés par des modèles d'état linéaires invariant dans le temps. Dans les deux types d'approches au problème hors-ligne, la génération de résidus peut être vue comme la conception d'une transformation linéaire d'un vecteur Gaussien. Plusieurs méthodes de diagnostic statistiques sont ré-examinées, à l'aide de leur formulation en terme de cette transformation et aussi du contenu informationnel des résidus correspondants. On formule précisément plusieurs cas de pannes multiples, avec ou sans hypothèse de causalité entre elles, et on discute un critère d'optimalité pour le cas le plus général. On propose de nouveaux critères d'information pour l'étude du problème de l'optimisation des résidus.

Mots-clé : Détection et diagnostic de pannes, génération de résidus, évaluation de résidus, optimisation de résidus, critères d'information.

1 Introduction

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 [61, 34, 28, 5, 24] and the books [58, 51, 6] among other references. It has been widely acknowledged that the FDI problem can be splitted 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 restricted to the case of additive faults occurring on Gaussian dynamical controlled processes modelled by linear time invariant (LTI) state-space systems, 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). The isolation problem is addressed through directional residuals designed with deterministic rejection (decoupling) methods. The basic statistical approach to residual generation consists in deriving sufficient statistics, namely transformations of the measurements which capture the entire information about the fault contained in the original data. The main example of such a lossless residual is Kalman filter innovation¹. Statistical rejection (decoupling) methods are based on standard techniques for dealing with nuisance parameters. 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.

Even though a clear understanding of how each approach links to and complements the other one is of interest, very few attempts have been made in this direction [60, 6, 49]. It is one of the purposes of this paper to outline some connections between those deterministic and stochastic methods for additive faults. These connections can be more conveniently investigated when restricting the problem to *off-line FDI*, that is assuming that measured observations Y and known controlled inputs U are given over a finite time-window. As we show below, it results from this assumption that the dynamical problem reduces to a static one, and that, in *both types of approaches*, residual generation can be viewed as the design of a *linear transformation* $\zeta = \mathcal{A} \mathcal{Y}$ of the finite-window input-adjusted observations \mathcal{Y} with some desired detection and isolation properties. Writing \mathcal{Y} as

$$\mathcal{Y} = \mathcal{O} X + \mathcal{M} \Upsilon + \mathcal{E}$$

where Υ contains the additive fault vectors and $\text{cov}(\mathcal{E}) = \Sigma$, any relevant residual ζ is normally distributed

$$\zeta \sim \mathcal{N}(L \Upsilon, R), \quad L = \mathcal{A} \mathcal{M}, \quad R = \mathcal{A} \Sigma \mathcal{A}^T \quad (1)$$

Here $L \Upsilon$ is the signature of the fault on the residual through incidence matrix² L , and R is the covariance matrix of the residual. The two matrices \mathcal{M} and Σ depend on the dynamics of the system and on the fault gains; the exact formulas are given below.

Based on an information theoretic point of view, the main purpose of this paper is to discuss several possible *criteria for optimizing the choice of transformation* \mathcal{A} , among which several are new, at least in this context. They are based on the quantity

$$\mathbf{F}_\zeta = L^T R^{-1} L = \mathcal{M}^T \mathcal{A}^T (\mathcal{A} \Sigma \mathcal{A}^T)^{-1} \mathcal{A} \mathcal{M} \quad (2)$$

¹This is *not* true in the case of multiplicative faults, such as changes in the data spectrum or changes in the input-output transfer function. For these faults, the innovation is not a sufficient statistics, the transformation from observations to innovations loses some information about the fault. A detailed discussion of this issue can be found in [6].

²Note that this incidence matrix is real valued, as opposed to the incidence matrix used in [29, 35], which results from thresholding and is thus binary valued.

which is Fisher information about fault Υ contained in residual ζ , whereas

$$\mathbf{F}_y = \mathcal{M}^T \Sigma^{-1} \mathcal{M} \quad (3)$$

is the information contained in the initial data. The key reason for using this quantity is that sufficient statistics, which are the basis of statistical detection methods, preserves Fisher information, which also turns out to play a crucial role in performance criteria for detection and isolation methods.

The paper is organized in the following manner. In section 2, we recall how to reduce off-line additive fault detection in a dynamic state-space system to additive fault detection in a static regression model, and recall a standard deterministic residual generation method. In section 3, we revisit stochastic tools for FDI in the basic case of a Gaussian vector, emphasizing transformation forms and information properties. In section 4, we address the case of Gaussian regression models, and we discuss how to combine the results of sections 2 and 3 for investigating the links between deterministic and stochastic approaches, and deriving possible mixed methods. In particular, both stochastic and deterministic approaches basically rely upon projection and rejection operators, and we discuss connections between them. In section 5, we formally state several multiple fault cases, with or without causality assumptions, and an optimality criterion for the most general one. The methods of section 3 are placed in this perspective, and the properties of the corresponding residuals summarized. From these properties, we deduce in section 6 new information criteria for optimizing the residuals, and we discuss the resulting optimization problems. Section 7 is devoted to conclusions and questions open for future research.

2 FDI in linear dynamic systems

In this section, we show that off-line residual generation for a *dynamic* LTI system observed over a finite time-window reduces to residual generation for a *static* system, and thus to appropriate linear transformations of a Gaussian random *vector* (and not a random process). We introduce problem statements for detection, isolation, and residual generation and evaluation which are used throughout, and we recall a standard deterministic residual generation method.

2.1 Reducing off-line FDI to a static problem

We assume that the model of the safe dynamic system is

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

where the state X , the controlled input U , and the observation Y have dimensions n , m , r , respectively, and where $(W_k)_k$ and $(V_k)_k$ are two independent white noise sequences, with covariance matrices Q_x and Q_y , respectively. The observation noise covariance matrix Q_y is assumed to be positive definite. The necessity of the observability assumption is discussed below.

Moreover, we assume that the model of the faulty system is

$$\begin{cases} X_{k+1} &= FX_k + GU_k + \Gamma \Upsilon_x + W_k \\ Y_k &= HX_k + JU_k + \Xi \Upsilon_y + V_k \end{cases} \quad (5)$$

where Υ_x and Υ_y are the assumed additive faults, of dimensions q_x and q_y , respectively, and the fault gains Γ and Ξ are matrices of dimensions $n \times q_x$ and $r \times q_y$, respectively. The total number of faults is denoted by $q = q_x + q_y$. It is well known that such a model is appropriate for sensors and actuators faults³. Note that we assume the fault vectors Υ_x and Υ_y to be *constant*; thus we do not need the usual fault isolation assumption (total number of faults lower than the number of sensors: $q = q_x + q_y \leq r$). Assuming time-varying fault vectors $\Upsilon_x(k)$ and $\Upsilon_y(k)$ would only change the expressions of the fault gains in the stacked model, but the analysis of the subsequent sections would still apply and the proposed criteria still be used, provided that the corresponding rank assumptions are enforced. We only assume that matrices Γ and Ξ are full column rank (f.c.r.), and thus $q_x \leq n$, $q_y \leq r$.

We address the *finite horizon off-line FDI problem*. In other words, we are given measured observations Y and known controlled inputs U over a finite time-window with size ℓ^4 . Using obvious notations, e.g. $\mathcal{Y}_k^{(\ell)}$ is the column vector made of $Y_{k-\ell+1}, \dots, Y_k$, and standard computations, we rewrite the set of ℓ successive equations (5) as

$$\mathcal{Y}_k^{(\ell)} = \mathcal{O}_\ell X_{k-\ell+1} + \mathcal{M}_\ell(G, J) \mathcal{U}_k^{(\ell)} + \mathcal{M}_\ell(\Gamma, 0) (\mathbf{1}_\ell \otimes \Upsilon_x) + \mathcal{M}_\ell(0, \Xi) (\mathbf{1}_\ell \otimes \Upsilon_y) + \mathcal{M}_\ell(I_n, 0) \mathcal{W}_k^{(\ell)} + \mathcal{V}_k^{(\ell)} \quad (6)$$

where \mathcal{O}_ℓ is observability matrix of order $(\ell - 1)$

$$\mathcal{O}_\ell(H, F) = \begin{pmatrix} H \\ HF \\ HF^2 \\ \vdots \\ HF^{\ell-1} \end{pmatrix} \quad (7)$$

³Typical examples are to assume that the gain matrices Γ and Ξ are proportional to the identity matrix I up to an unknown fault magnitude ν , and that fault vector Υ has zero components except for the j -th one which equals one.

⁴Considering off-line FDI means that we do *not* address the quickest detection of the fault and the estimation of its onset time, even though these problems might be of great importance in some applications. The interested reader is referred to [6] for a detailed description of on-line detection algorithms.

and $\mathcal{M}_\ell(G, J)$ is the lower triangular block-Toeplitz matrix associated with the impulse response of the system

$$\mathcal{M}_\ell(G, J) = \begin{pmatrix} J & \dots & \dots & \dots & \dots & \dots \\ HG & J & \dots & 0 & \dots & \dots \\ HFG & HG & J & \dots & \dots & \dots \\ HF^2G & HFG & HG & J & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ HF^{\ell-2}G & \dots & \dots & HFG & HG & J \end{pmatrix} \quad (8)$$

Introducing the input-adjusted observations

$$\underline{\mathcal{Y}}_k^{(\ell)} = \mathcal{Y}_k^{(\ell)} - \mathcal{M}_\ell(G, J) \mathcal{U}_k^{(\ell)} \quad (9)$$

and the noise

$$\underline{\mathcal{Y}}_k^{(\ell)} = \mathcal{M}_\ell(I_n, 0) \mathcal{W}_k^{(\ell)} + \mathcal{V}_k^{(\ell)}, \quad (10)$$

we rewrite (6) as the *regression model* [40]

$$\mathcal{Y} = \underline{\mathcal{Y}}_k^{(\ell)} = \mathcal{O}_\ell X_{k-\ell+1} + \underline{\mathcal{M}}_\ell(\Gamma) \Upsilon_x + (\mathbb{1}_\ell \otimes \Xi) \Upsilon_y + \underline{\mathcal{Y}}_k^{(\ell)} \quad (11)$$

where

$$\underline{\mathcal{M}}_\ell(\Gamma) = \begin{pmatrix} 0 \\ H\Gamma \\ HFG + H\Gamma \\ \vdots \\ HF^{\ell-2}\Gamma + \dots + HFG + H\Gamma \end{pmatrix} \quad (12)$$

In (11), noise $\underline{\mathcal{Y}}_k^{(\ell)}$ has covariance matrix

$$\Sigma_\ell = \mathcal{M}_\ell(I_n, 0) (I_\ell \otimes Q_x) \mathcal{M}_\ell^T(I_n, 0) + I_\ell \otimes Q_y$$

It results from the recursive formula

$$\Sigma_\ell = \begin{pmatrix} & & & 0 \\ & & & HQ_x F^{(\ell-2)T} H^T \\ & & & \vdots \\ & & & \vdots \\ & & & HQ_x F^T H^T \\ 0 & HF^{\ell-2} Q_x H^T & \dots & \dots & HF Q_x H^T & HQ_x H^T + Q_y \end{pmatrix} \quad (13)$$

for $\ell \geq 2$, and the matrix

$$\Sigma_2 = \begin{pmatrix} Q_y & 0 \\ 0 & HQ_x H^T + Q_y \end{pmatrix}$$

that matrix Σ_ℓ is *positive definite* for every ℓ , and is *not block-diagonal* whenever $\ell \geq 3$, even if the state and observation noise covariance matrices Q_x and Q_y are diagonal – which is a common assumption in practice when few *a priori* information is available. This point has important consequences on residual generation as explained below.

Regression model (11) has two basic properties, under convenient assumptions. First, state $X_{k-\ell+1}$ is *independent* of noise $\underline{\mathcal{Y}}_k^{(\ell)}$, basically because the first block-line of $\mathcal{M}_\ell(I_n, 0)$ in (10) is zero, and because of the assumptions regarding noises \mathcal{W} and \mathcal{V} . Second, let us discuss the ranks of the failure

gain matrices $\underline{\mathcal{M}}_\ell(\Gamma)$ and $\mathbf{1}_\ell \otimes \Xi$ in (11). The latter is clearly f.c.r. with Ξ . As for the former, which is defined in (12), we note that

$$\text{rank}(\underline{\mathcal{M}}_\ell(\Gamma)) = \text{rank} \begin{pmatrix} 0 \\ \mathcal{O}_{\ell-2} \Gamma \end{pmatrix}$$

A necessary and sufficient condition for $\underline{\mathcal{M}}_\ell(\Gamma)$ to be full column rank (f.c.r.) is thus

$$\begin{cases} \Gamma \text{ f.c.r.} \\ \text{Range}(\Gamma) \cap \text{Ker}(\mathcal{O}_{\ell-2}) = \{0\} \end{cases}$$

A simpler, and only sufficient, condition is

$$\begin{cases} \Gamma \text{ f.c.r.} \\ \text{rank}(\mathcal{O}_{\ell-2}) = n \end{cases}$$

If system (4) is assumed observable, then, for window size ℓ large enough (and possibly smaller than state dimension n), this sufficient condition is fulfilled.

2.2 Static FDI problem statement

We have thus re-written a finite number of dynamical state-space equations (5) as a *regression* model

$$\mathcal{Y} = \mathcal{O}_\ell X + \mathcal{M}_\ell \Upsilon + \mathcal{E}_\ell \quad (14)$$

where matrix \mathcal{O}_ℓ is observability matrix of order $(\ell-1)$ of the original dynamic system (4); regressor X is unknown; the fault gain matrix and fault vector are

$$\mathcal{M}_\ell = \begin{pmatrix} \underline{\mathcal{M}}_\ell(\Gamma) & \mathbf{1}_\ell \otimes \Xi \end{pmatrix}, \quad \Upsilon = \begin{pmatrix} \Upsilon_x \\ \Upsilon_y \end{pmatrix} \quad (15)$$

with $\underline{\mathcal{M}}_\ell(\Gamma)$ defined in (12); Gaussian white noise \mathcal{E}_ℓ is independent of regressor X , and has positive definite covariance matrix Σ_ℓ . Regression model (14) shows up particular structures in its matrix $(\mathcal{O}_\ell, \mathcal{M}_\ell, \Sigma_\ell)$ and vector (\mathcal{E}_ℓ) components. These structures are referred to with calligraphic letters, and are discussed later. From now on, we investigate as such the detection and isolation problems in a general regression model

$$Z = H X + M \eta + \epsilon, \quad \text{cov}(\epsilon) = \Sigma = \bar{\Sigma} \bar{\Sigma}^T, \quad \bar{\Sigma} > 0 \quad (16)$$

where white noise ϵ is again independent of regressor X , and has positive definite covariance matrix Σ . Vector X is not assumed to be known, and is processed while solving the detection and isolation problems. As we explain below, its (least-squares) estimation is performed by stochastic approaches, whereas a rejection technique is used in deterministic approaches.

Formal problem statements can be found in [6]. In section 5, we discuss single and multiple faults with or without causality assumptions, we state the isolation problem as a simultaneous multiple hypotheses testing problem, and we discuss relevant isolation optimality criteria [54].

According to the above discussions, we consider the following problem statements.

Problem 2.1 (Detection) *Decide between $\eta = 0$ and $\eta \neq 0$ in $Z = H X + M \eta + \epsilon$ where M is a f.c.r. matrix and $\dim \eta = q$ and $\dim Z = \ell r$.*

Problem 2.2 (Isolation) *Decide between $\eta_a = 0$ and $\eta_a \neq 0$ in $Z = H X + M_a \eta_a + M_b \eta_b + \epsilon$ where $\dim Z = \ell r$, $[M_a, M_b]$ is a f.c.r. matrix, η_a and η_b have known dimensions q_a and q_b , $q_a + q_b = q$, and η_b is an unknown vector⁵.*

Again, no assumption on failure dimensions and ranks of matrix gains M_i is made at this point. We introduce these assumptions when they are necessary in the corresponding methods.

Residual generation is the process of transforming data Z in quantities ζ ideally close to zero under no-fault conditions, minimally sensitive to noises and disturbances, and maximally sensitive to faults η , in such a way to exhibit relevant detection and/or isolation properties. Residual evaluation refers to the quantization of these different sensitivities. Residual generation and evaluation can be stated as follows.

Problem 2.3 (Residual generation and evaluation) *Find a linear transform \mathcal{A} producing a residual $\zeta = \mathcal{A} Z$ such that $\|\zeta\|$ is minimal when $\eta = 0$ and maximal when $\eta \neq 0$.*

It should be clear that this statement refers to a choice of a norm for assessing for the residual sizes $\|\zeta\|$ under no-fault and fault conditions respectively. Several choices are discussed below : Frobenius norm of the fault incidence matrix $\mathcal{A} M$ arises in deterministic approaches, whereas scalar functions of Fisher information matrix result from a statistical point of view. Actually this issue of (vector or matrix) norms is central in our discussions on both residual optimization criteria and comparison between deterministic and stochastic methods.

We now recall some deterministic residual generation and evaluation methods.

2.3 Deterministic approaches

There are two main classes of deterministic approaches, based either on parity checks (analytical redundancy relations) or observers (failure detection filters) [61, 28, 24]. These two types of methods are recognized to be structurally equivalent, whereas exhibiting possibly different numerical behavior [41, 52]. As explained in the introduction, one of our main goals in this paper is to outline relationships between deterministic and stochastic approaches to FDI. Having this goal in mind, we already restricted the problem statement to finite horizon off-line FDI, and we now restrict the discussion of deterministic residual generation methods to the design of parity checks. These methods are basically finite horizon methods, and can be viewed as performing a deterministic rejection of unknown quantities.

Even though the basic principle originated for regression models (16), we describe it directly for the stacked dynamical model (14). It proceeds as follows.

Parity checks or deterministic rejection. *Parity space of order $(\ell - 1)$ is the space of ℓr -dimensional vectors v such that [40]*

$$v^T \mathcal{O}_\ell = 0$$

It is thus the left null space of observability matrix of order $(\ell - 1)$ defined in (7). From a geometrical point of view, it is the orthogonal complement of the range of this observability matrix. Its dimension is n_ℓ , where

$$n_\ell = \ell r - \text{rank}(\mathcal{O}_\ell)$$

We know that $n_\ell \geq 0$, that n_ℓ is non-decreasing, and that $n_\ell = \ell r - n$ if the system is assumed observable and ℓ is large enough, as discussed above.

⁵The important issue of inferring the (generally unknown) number of failed components in η is beyond the scope of this paper. The interested reader is referred to [8] for a discussion of and a possible solution to this problem.

A (*scalar*) *parity check* is defined to be $v^T \mathcal{Y}$, where \mathcal{Y} is an abbreviation for the input-adjusted observations $\tilde{\mathbf{y}}_k^{(\ell)}$ in (11). Therefore, a parity check, which is defined at any time k , performs a linear combination of the observations contained in a finite time window of size ℓ .

Let s be an integer such that $s \leq n_\ell$, and let $\mathcal{P}_{\ell,s}$ a matrix whose s columns form an orthonormal sub-basis in the parity space. Note that matrix $\mathcal{P}_{\ell,n_\ell}$ defines the (deterministic) orthogonal projection onto the parity space. We call (*vector*) *parity check* the s -dimensional vector defined by

$$\zeta = \mathcal{P}_{\ell,s}^T \mathcal{Y}$$

Defining $\mathcal{A}_\ell = \mathcal{P}_{\ell,s}^T$, parity check $\zeta = \mathcal{A}_\ell \mathcal{Y}$ thus results from the following design constraints

$$\mathcal{A}_\ell \mathcal{O}_\ell = 0 \quad (17)$$

$$\mathcal{A}_\ell \mathcal{A}_\ell^T = I_s \quad (18)$$

In these equations, the window size ℓ and parity vector dimension s are part of the design.

For fixed ℓ, s , two such matrices are related through pre-multiplication with an orthogonal matrix. Consequently, quantity (2)

$$\mathbf{F}_\zeta = L^T R^{-1} L = \mathcal{M}^T \mathcal{A}^T (\mathcal{A} \Sigma \mathcal{A}^T)^{-1} \mathcal{A} \mathcal{M}$$

is left unchanged, and thus *all such parity checks contain the same Fisher information about the fault*. Nevertheless, nothing can be said about the possible information discrepancy between fault information \mathbf{F}_ζ in the residual and fault information \mathbf{F}_y (3) in the original data. This point will be further addressed in section 7.

Because of this forthcoming discussion, we now need to recall that such parity checks can be computed with any selection of s left singular vectors of \mathcal{O}_ℓ corresponding to singular value 0 [40]. The only point here is that singular value decomposition (SVD) is known to be an efficient method for computing the orthogonal complement of the range of the matrix of interest, among other subspaces [36].

Lemma 2.1 (Parity checks through SVD) *Let*

$$\mathcal{O}_\ell = \begin{pmatrix} S_1 & S_2 \end{pmatrix} \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} T_1^T \\ T_2^T \end{pmatrix}$$

be the SVD of observability matrix of order $(\ell - 1)$, where the columns of the orthonormal matrices (S_1, S_2) and (T_1, T_2) contain the singular vectors, and $\Delta = \text{diag}(\delta_1, \dots, \delta_{\ell-r-n_\ell})$ is the diagonal matrix containing the (non-zero) singular values in decreasing order.

Then parity space of order $(\ell - 1)$ is the range of matrix S_2 .

Moreover, let \underline{S}_2 be a matrix made of any selection of s columns of S_2 , with $s \leq n_\ell$, and define $\mathcal{A}_\ell = \underline{S}_2^T$. Then $\zeta = \mathcal{A}_\ell \mathcal{Y}$ is a s -dimensional parity check.

In practice, when it turns out that the last singular values of \mathcal{O}_ℓ are not zero but small, a useful byproduct of this computational design is then to search for a transform \mathcal{A}_ℓ which minimizes $\|\mathcal{A}_\ell \mathcal{O}_\ell\|_F$ under constraint (18), where Frobenius norm of a matrix M is defined as :

$$\|M\|_F^2 = \text{tr} (MM^T) = \sum_i \sum_j |m_{ij}|^2 \quad (19)$$

and where notation $\text{tr}(A)$ stands for trace of matrix A . Such a transform also minimizes $|\mathcal{A}_\ell \mathcal{O}_\ell \mathcal{O}_\ell^T \mathcal{A}_\ell^T|$, where notation $|A|$ stands for determinant of matrix A . This Frobenius norm minimization property motivates the choice of Frobenius norm for solving the residual generation and evaluation problem stated above.

Actually, this minimizing property is used for defining parity checks which are as much robust as possible with respect to uncertainties in the system matrices, and thus in the observability matrix [40]. Referring to the statement of problem 2.3, and defining the residual norms as

$$\begin{aligned} \text{under } \Upsilon = 0, \quad & \|\zeta_0\| = \|\mathcal{A}_l \mathcal{O}_l\|_F^2 = \text{tr} \left(\mathcal{A}_l \mathcal{O}_l \mathcal{O}_l^T \mathcal{A}_l^T \right) \\ \text{under } \Upsilon \neq 0, \quad & \|\zeta_\Upsilon\| = \|\mathcal{A}_l \mathcal{M}_l\|_F^2 = \text{tr} \left(\mathcal{A}_l \mathcal{M}_l \mathcal{M}_l^T \mathcal{A}_l^T \right) \end{aligned}$$

the (multi-criteria) optimization problem is then to simultaneously minimize $\|\zeta_0\|$ and maximize $\|\zeta_\Upsilon\|$. Frank [25, 26] suggests to minimize the criterion

$$J_r = \frac{\|\zeta_0\|}{\|\zeta_\Upsilon\|} \quad (20)$$

following [23] where a transfer function statement is rather used. Patton [18] additionally suggests to minimize the alternative criterion

$$J_\alpha = \alpha \|\zeta_0\| + (1 - \alpha) \|\zeta_\Upsilon\| \quad (21)$$

where the minimizations should be understood under constraint (18). Both problems can be solved using singular value decompositions of matrices \mathcal{O}_ℓ and \mathcal{M}_ℓ [36, 40, 52].

Related discussions in the framework of detection filters (and not parity checks) can be found in [22] where \mathcal{H}_2 or \mathcal{H}_∞ norms of the noise-to-residual transfer function are minimized. Another reference is [47]. \mathcal{H}_∞ norm is also used in [23]. It is interesting to note that this norm has strong connections with information theoretic concepts [56].

3 Basic statistical tools

In this section, we describe key classical statistic tools for detection and isolation in the basic case of a Gaussian vector⁶ ($X = 0$). The case of Gaussian regression vectors is addressed in section 4. We give several basic results which are used throughout, concerning likelihood ratio test for Gaussian distributions (detection) and two statistical isolation methods (fault sensitivity and nuisance rejection). Before proceeding, we emphasize that the general stochastic approach to FDI *jointly* considers the residual generation *and* evaluation problems.

We use notation $Z \sim \mathcal{N}(\mu, \Sigma)$ for a (Gaussian) random vector which is normally distributed with mean μ and covariance matrix Σ . For computing all noncentrality parameters and residual covariances, we make repeated use of the following results [32].

Lemma 3.1 (Expected values of quadratic forms) *Let $Z \sim \mathcal{N}(\mu, \Sigma)$ and A, B be constant symmetric matrices. Then*

$$\begin{aligned} \mathbf{E}(Z^T A Z) &= \text{tr}(A \Sigma) + \mu^T A \mu \\ \text{var}(Z^T A Z) &= 2 \text{tr}(A \Sigma)^2 + 4 \mu^T A \Sigma A \mu \\ \text{cov}(Z^T A Z, Z^T B Z) &= 2 \text{tr}(A \Sigma B \Sigma) + 4 \mu^T A \Sigma B \mu \end{aligned}$$

Let notation $p_\eta(Z)$ stand for probability density of a Gaussian vector $Z \sim \mathcal{N}(M \eta, \Sigma)$. Similarly, $p_{\eta_a, \eta_b}(Z)$ is the density of a Gaussian vector $Z \sim \mathcal{N}(M_a \eta_a + M_b \eta_b, \Sigma)$. The log-likelihood function is given by

$$l_\eta(Z) \triangleq -2 \ln p_\eta(z) = (Z - M\eta)^T \Sigma^{-1} (Z - M\eta) \quad (22)$$

up to a constant which we do not need in the sequel.

Lemma 3.2 (Stochastic projection and rejection) *Let $\hat{\eta}$ be the (maximum likelihood or) least-squares estimate of η , namely*

$$\hat{\eta} = \arg \max_{\eta} p_\eta(Z) = \arg \min_{\eta} l_\eta(Z)$$

Factorize the (generally non-diagonal) covariance matrix as $\Sigma = \bar{\Sigma} \bar{\Sigma}^T$. Let P_M be the associated (stochastic) orthogonal projection

$$P_M = I - \bar{\Sigma}^{-1} M (M^T \Sigma^{-1} M)^{-1} M^T \bar{\Sigma}^{-T} \quad (23)$$

which is idempotent and symmetric. The corresponding innovation writes

$$Z - M\hat{\eta} = \tilde{P}_M Z$$

where the idempotent matrix

$$\tilde{P}_M \triangleq \bar{\Sigma} P_M \bar{\Sigma}^{-1} \quad (24)$$

being nonsymmetric, is not a projection, except if covariance Σ is diagonal. However it is a ‘stochastic rejector’, namely such that

$$\tilde{P}_M M = 0 \quad (25)$$

The optimized log-likelihood function is

$$\min_{\eta} l_\eta(Z) = Z^T \tilde{P}_M^T \Sigma^{-1} \tilde{P}_M Z = Z^T \bar{\Sigma}^{-T} P_M \bar{\Sigma}^{-1} Z$$

⁶Because of the local approaches to detection discussed in [12, 6, 4], this basic case is of much wider importance than for the present case of additive faults in linear dynamical systems. The detection and isolation methods discussed here are thus also convenient for FDI in nonlinear dynamical systems as well [8, 44, 7].

We define

$$\mathbf{F} = M^T \Sigma^{-1} M \quad (26)$$

which is *Fisher information* matrix, namely the matrix quantifying the information about η contained in random variable Z .

We now revisit existing detection (likelihood ratio) and isolation (sensitivity, rejection) methods. The detection result is given under both a projection form and a transformation form. The projection form is useful for outlining the link with deterministic methods. The transformation form is useful for discussing information properties and residual optimization. The projection form results straightforwardly from the likelihood methodology as we explain now.

3.1 Statistical detection

The key statistical detection tool is likelihood ratio, that computes, from the available observations, which hypothesis is most likely to hold.

Lemma 3.3 (Basic Gaussian detection) *For deciding between*

$$\mathcal{H}_0 : \eta = 0 \quad \text{and} \quad \mathcal{H}_1 : \eta \neq 0$$

in $Z \sim \mathcal{N}(M\eta, \Sigma)$ where $q = \dim \eta \leq \dim Z$, M is a f.c.r. matrix, and $\Sigma = \bar{\Sigma}\bar{\Sigma}^T$, the optimum test statistics t_1 , referred as the global test, is based on two times the generalized log-likelihood ratio (GLR)

$$t_1 = 2 \ln \frac{\max_{\eta} p_{\eta}(Z)}{p_0(Z)} = l_0(Z) - \min_{\eta} l_{\eta}(Z) \quad (27)$$

and can be written in two manners.

- The projection form is

$$t_1 = Z^T \bar{\Sigma}^{-T} (I - P_M) \bar{\Sigma}^{-1} Z \quad (28)$$

where P_M is given in lemma 3.2.

- The transformation form is based on the transformed variable

$$\zeta = \mathcal{G} Z \triangleq M^T \Sigma^{-1} Z \quad (29)$$

where transform \mathcal{G} is such that

$$\mathcal{G} M = \mathbf{F} = \mathcal{G} \Sigma \mathcal{G}^T$$

This vector ζ is distributed, under $\eta \neq 0$, as

$$\zeta \sim \mathcal{N}(\mathbf{F} \eta, \mathbf{F})$$

Test statistics t_1 writes

$$t_1 = \zeta^T \mathbf{F}^{-1} \zeta = Z^T \Sigma^{-1} M \mathbf{F}^{-1} M^T \Sigma^{-1} Z \quad (30)$$

Test statistics t_1 is distributed as a noncentral χ^2 -random variable $\chi'^2(q, \tau)$, with $q (= \dim \eta)$ degrees of freedom and noncentrality parameter

$$\tau = \mathbf{K}(\eta) = \eta^T \mathbf{F} \eta \quad (31)$$

The power (probability of correct decision) of test (30) is an increasing function of quantity (31). Noncentrality parameter (31) is the expectation of quadratic form (30) under hypothesis $\eta \neq 0$, and its value results from lemma 3.1.

3.2 Statistical isolation

We now address problem 2.2 in the case $X = 0$, namely for

$$Z = M_a \eta_a + M_b \eta_b + \epsilon \sim \mathcal{N}(M \eta, \Sigma)$$

In other words, we partition the matrix gain and parameter vector as

$$M = \begin{pmatrix} M_a & M_b \end{pmatrix}, \quad \eta = \begin{pmatrix} \eta_a \\ \eta_b \end{pmatrix}$$

where the sizes q_a and q_b of the components are assumed to be fixed and known.

We already noticed that matrix

$$\mathbf{F} = M^T \Sigma^{-1} M = \begin{pmatrix} M_a^T \Sigma^{-1} M_a & M_a^T \Sigma^{-1} M_b \\ M_b^T \Sigma^{-1} M_a & M_b^T \Sigma^{-1} M_b \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{F}_{aa} & \mathbf{F}_{ab} \\ \mathbf{F}_{ba} & \mathbf{F}_{bb} \end{pmatrix}$$

plays a key role in statistical detection. Its importance as an information measure is a central issue in the derivation of the new criteria proposed in the paper.

We make extensive use of the following results [32].

Lemma 3.4 (Partitioned matrix inversion and determinant formulas)

$$\mathbf{F}^{-1} = \begin{pmatrix} \mathbf{F}_a^{*-1} & -\mathbf{F}_{aa}^{-1} \mathbf{F}_{ab} \mathbf{F}_b^{*-1} \\ -\mathbf{F}_{bb}^{-1} \mathbf{F}_{ba} \mathbf{F}_a^{*-1} & \mathbf{F}_b^{*-1} \end{pmatrix}$$

and

$$|\mathbf{F}| = |\mathbf{F}_{aa}| |\mathbf{F}_b^*| = |\mathbf{F}_{bb}| |\mathbf{F}_a^*|$$

where

$$\mathbf{F}_i^* = \mathbf{F}_{ii} - \mathbf{F}_{ij} \mathbf{F}_{jj}^{-1} \mathbf{F}_{ji} \quad (32)$$

is such that

$$\mathbf{F}_i^{*-1} \mathbf{F}_{ij} \mathbf{F}_{jj}^{-1} = \mathbf{F}_{ii}^{-1} \mathbf{F}_{ij} \mathbf{F}_j^{*-1}$$

for $i, j = a, b; i \neq j$.

We now re-formulate statistical isolation methods using the transformation form. We thus consider several residuals

$$\zeta = \mathcal{A} Z \sim \mathcal{N}(L \Upsilon, R), \quad L = \mathcal{A} M, \quad R = \mathcal{A} \Sigma \mathcal{A}^T$$

Viewing ζ as the output of a regression model with regressor η , a classical statistical argument would first lead to diagonalize its covariance matrix $\mathcal{A} \Sigma \mathcal{A}^T$. On the other hand, a common isolation requirement [29, 35] is to constrain fault incidence matrix $\mathcal{A} M$ to be diagonal, in order to have an easy separation of the regressor components. For each transformation \mathcal{A} , we thus discuss the properties of these two matrices. We also compute Fisher information about the fault η contained in the residual ζ :

$$\mathbf{F}_\zeta = M^T \mathcal{A}^T (\mathcal{A} \Sigma \mathcal{A}^T)^{-1} \mathcal{A} M$$

which allows to ensure whether the residual is a sufficient statistics or not, and the mutual information⁷ [20] among the residual components ζ_a and ζ_b

$$\mathcal{I}_\zeta = -\frac{1}{2} \ln \frac{|R|}{|R_a| |R_b|}$$

⁷Mutual information between Gaussian vectors, exactly as their entropy, does not depend on their means. Therefore the formula given here holds under no-fault and any of the fault hypotheses considered in the paper. Mutual information can also be viewed as Kullback distance between the joint distribution of ζ_a, ζ_b and the product of their individual distributions. Kullback distance is the base of the residual optimization criteria we propose.

which is another measure of the dependence between the residual components [53].

3.2.1 Karhunen-Loeve decomposition

A principal component analysis (or Karhunen-Loeve decomposition), based on the eigendecomposition

$$\Sigma = \check{\mathcal{G}} \Delta \check{\mathcal{G}}^T$$

leads to define a transformed vector

$$\check{\zeta} = \check{\mathcal{G}}^T Z \sim \mathcal{N}(\check{\mathcal{G}}^T M \eta, \Delta) \quad (33)$$

having a diagonal covariance matrix Δ , but a generally non-diagonal incidence matrix $\check{\mathcal{G}}^T M$. This transformed vector keeps Fisher information about η :

$$\check{\mathbf{F}} = M^T \check{\mathcal{G}} \Delta^{-1} \check{\mathcal{G}}^T M = \mathbf{F}$$

However, because of the non-diagonal aspect of the incidence matrix, this transform is of limited interest for isolation purposes.

In what follows, we show how a straightforward application of likelihood ratio methodology (lemma 3.3) to the hypotheses testing problem statements given above, leads to different transformed vectors, among which some show up a triangular or even diagonal fault incidence matrix.

3.2.2 Sensitivity tests

A rather intuitive statistical solution to the isolation problem consists in projecting the deviations in η onto the subspace generated by the components to be isolated. It has proven to give satisfactory results for vibration monitoring [9]. It proceeds as follows.

Lemma 3.5 (Basic Gaussian sensitivity) *Sensitivity test \tilde{t}_a for monitoring η_a is GLR test between $\eta = (0, 0)$ and $\eta = (\eta_a, 0)$, where $\eta_a \neq 0$, namely*

$$\tilde{t}_a = 2 \ln \frac{\max_{\eta_a} p_{\eta_a,0}(Z)}{p_{0,0}(Z)} = l_{0,0}(Z) - \min_{\eta_a} l_{\eta_a,0}(Z) \quad (34)$$

It can be written under two forms :

- Projection form :

$$\tilde{t}_a = Z^T \bar{\Sigma}^{-T} (I - P_{M_a}) \bar{\Sigma}^{-1} Z \quad (35)$$

where P_{M_a} is defined as in lemma 3.3;

- Transformation form : Let $\tilde{\zeta}_a$ be the transformed variable

$$\tilde{\zeta}_a = \tilde{\mathcal{G}}_a Z \triangleq M_a^T \Sigma^{-1} Z \quad (36)$$

where transform $\tilde{\mathcal{G}}_a$ is such that

$$\tilde{\mathcal{G}}_a M_a = M_a^T \Sigma^{-1} M_a = \mathbf{F}_{aa} = \tilde{\mathcal{G}}_a \Sigma \tilde{\mathcal{G}}_a^T$$

Residual $\tilde{\zeta}_a$ is distributed, under $\eta_a \neq 0$, as

$$\tilde{\zeta}_a \sim \mathcal{N}(\mathbf{F}_{aa} \eta_a, \mathbf{F}_{aa})$$

Test statistics \tilde{t}_a writes

$$\tilde{t}_a = \tilde{\zeta}_a^T \mathbf{F}_{aa}^{-1} \tilde{\zeta}_a = Z^T \tilde{\mathcal{G}}_a^T \mathbf{F}_{aa}^{-1} \tilde{\mathcal{G}}_a Z \quad (37)$$

$$= Z^T \Sigma^{-1} M_a \left(M_a^T \Sigma^{-1} M_a \right)^{-1} M_a^T \Sigma^{-1} Z \quad (38)$$

Under $\eta_a \neq 0$, it is distributed as a χ^2 -random variable with q_a degrees of freedom and noncentrality parameter

$$\tilde{\mathbf{K}}(\eta_a) = \eta_a^T \mathbf{F}_{aa} \eta_a$$

given by lemma 3.1.

Now comments on fault isolation versus residual decorrelation are in order. Let

$$\tilde{\zeta}_b = \tilde{\mathcal{G}}_b Z \triangleq M_b^T \Sigma^{-1} Z \quad (39)$$

It should be noted that neither the two random variables $\tilde{\zeta}_a$ and $\tilde{\zeta}_b$ nor the two test statistics \tilde{t}_a and \tilde{t}_b are decorrelated under $\eta = 0$. In the following lemma, we summarize the joint properties of the transformed variables $\tilde{\zeta}_a$ and $\tilde{\zeta}_b$, from the double point of view of decoupling and information content.

Lemma 3.6 (Decoupling and information in sensitivity-based transformed variables) *Let $\tilde{\zeta}$ be the stacked residual made of the two transformed variables $\tilde{\zeta}_i$ corresponding to sensitivity w.r.t. the two subsets of parameter components, namely*

$$\tilde{\zeta} \triangleq \begin{pmatrix} \tilde{\zeta}_a \\ \tilde{\zeta}_b \end{pmatrix} = \tilde{\mathcal{G}} Z \quad (40)$$

where

$$\tilde{\mathcal{G}} \triangleq \begin{pmatrix} \tilde{\mathcal{G}}_a \\ \tilde{\mathcal{G}}_b \end{pmatrix} = M^T \Sigma^{-1} Z \quad (41)$$

The distribution of $\tilde{\zeta}$ under $\eta \neq 0$ ⁸ is

$$\tilde{\zeta} \sim \mathcal{N}(\tilde{L} \eta, \tilde{\Sigma})$$

where

$$\tilde{L} = \tilde{\mathcal{G}} M = \mathbf{F}$$

and

$$\tilde{\Sigma} = \tilde{\mathcal{G}} \Sigma \tilde{\mathcal{G}}^T = \mathbf{F}$$

Furthermore residual $\tilde{\zeta}$ contains the same Fisher information about η as the original vector Z , namely

$$\tilde{\mathbf{F}} = \tilde{L}^T \tilde{\Sigma}^{-1} \tilde{L} = \mathbf{F}$$

Residual $\tilde{\zeta}$ is thus a sufficient statistics for η . Also mutual information between $\tilde{\zeta}_a$ and $\tilde{\zeta}_b$ is

$$\tilde{\mathcal{I}} = \mathcal{I}(\tilde{\zeta}_a, \tilde{\zeta}_b) = -\frac{1}{2} \ln \frac{|\tilde{\Sigma}|}{|\mathbf{F}_{aa}| |\mathbf{F}_{bb}|} = -\frac{1}{2} \ln |\mathbf{F}_{aa}^{-1} \mathbf{F}_a^*| = -\frac{1}{2} \ln |\mathbf{F}_{bb}^{-1} \mathbf{F}_b^*| \quad (42)$$

⁸ And not under $\eta_a \neq 0, \eta_b = 0$ as above.

Consider now the correlation issue in the quadratic forms \tilde{t}_i themselves. Lemma 3.1, definitions (36), and repeated use of (32) result in that, under $\eta = 0$,

$$\begin{aligned} \text{cov}_0(\tilde{t}_a, \tilde{t}_b) &= 2 \text{tr} \left(\tilde{\mathcal{G}}_a^T \mathbf{F}_{aa}^{-1} \tilde{\mathcal{G}}_a \Sigma \tilde{\mathcal{G}}_b^T \mathbf{F}_{bb}^{-1} \tilde{\mathcal{G}}_b \Sigma \right) \\ &= 2 \text{tr} \left(\mathbf{F}_{aa}^{-1} \mathbf{F}_{ab} \mathbf{F}_{bb}^{-1} \mathbf{F}_{ba} \right) \\ &= 2 \text{tr} \left(I_{q_a} - \mathbf{F}_{aa}^{-1} \mathbf{F}_a^* \right) \end{aligned} \quad (43)$$

$$= 2 \text{tr} \left(I_{q_b} - \mathbf{F}_{bb}^{-1} \mathbf{F}_b^* \right) \quad (44)$$

The last formula of lemma 3.1 provides us with the correlation under $\eta \neq 0$

$$\text{cov}_1(\tilde{t}_a, \tilde{t}_b) = \text{cov}_0(\tilde{t}_a, \tilde{t}_b) + 4 \eta^T \begin{pmatrix} \mathbf{F}_{aa} - \mathbf{F}_a^* & \mathbf{F}_{ab} \\ \mathbf{F}_{ba} - \mathbf{F}_b^* \mathbf{F}_{bb}^{-1} \mathbf{F}_{ba} & \mathbf{F}_{bb} - \mathbf{F}_b^* \end{pmatrix} \eta$$

Note that the correlation between the \tilde{t}_i is tightly connected to the mutual information (42) between the $\tilde{\zeta}_i$.

3.2.3 Statistical rejection

Another statistical solution to the isolation problem consists in viewing parameter η_b as 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 *minmax approach* and consists in replacing the unknown parameter components by their least favorable value, namely the value which minimizes the power of test (30), or equivalently the quantity in (31). Actually it is equivalent [6] to a GLR method, which consists in replacing the unknown parameter components η_b by their maximum likelihood estimates.

Lemma 3.7 (Basic Gaussian rejection) *For deciding between $\eta_a = 0$ and $\eta_a \neq 0$ in*

$$Z \sim \mathcal{N}(M_a \eta_a + M_b \eta_b, \Sigma)$$

where $M = \begin{pmatrix} M_a & M_b \end{pmatrix}$ is a f.c.r. matrix, η_b is an unknown nuisance parameter with known dimension⁹, and $\Sigma = \bar{\Sigma} \bar{\Sigma}^T$, the optimum test is as follows [13, 60, 6].

- Projection form :

$$t_a^* = Z^T \bar{\Sigma}^{-T} (P_{M_b} - P_M) \bar{\Sigma}^{-1} Z \quad (45)$$

which also writes

$$t_a^* = Z^T \left(\tilde{P}_{M_b}^T \Sigma^{-T} M_a (M_a^T \Sigma^{-1} \tilde{P}_{M_b} M_a)^{-1} M_a^T \Sigma^{-1} \tilde{P}_{M_b} \right) Z$$

where \tilde{P}_{M_b} , given in (24), is the ‘stochastic rejector’ achieving (25).

- Transformation form : Let ζ_a^* be the transformed variable

$$\zeta_a^* = \mathcal{G}_a^* Z \triangleq M_a^T \Sigma^{-1} \tilde{P}_{M_b} Z \quad (46)$$

where transform \mathcal{G}_a^* also writes

$$\mathcal{G}_a^* = \begin{pmatrix} I_{q_a} & -\mathbf{F}_{ab} \mathbf{F}_{bb}^{-1} \end{pmatrix} M^T \Sigma^{-1} \quad (47)$$

and is such that

$$\mathcal{G}_a^* M_b = 0, \quad \mathcal{G}_a^* M_a = \mathbf{F}_a^* = \mathcal{G}_a^* \Sigma \mathcal{G}_a^{*T} \quad (48)$$

⁹See statement of problem 2.2.

Residual ζ_a^* is distributed, under $\eta_a \neq 0$, as

$$\zeta_a^* \sim \mathcal{N}(\mathbf{F}_a^* \eta_a, \mathbf{F}_a^*)$$

The optimum test statistics is

$$t_a^* = \zeta_a^{*T} \mathbf{F}_a^{*-1} \zeta_a^* = Z^T \mathcal{G}_a^{*T} \mathbf{F}_a^{*-1} \mathcal{G}_a^* Z \quad (49)$$

which is distributed as a χ^2 -random variable with q_a degrees of freedom and noncentrality parameter

$$\mathbf{K}^*(\eta_a) = \eta_a^T \mathbf{F}_a^* \eta_a \quad (50)$$

Formulas (45) and (49) results from elementary computations

- either on two times the GLR

$$t_a^* = 2 \ln \frac{\max_{\eta_a, \eta_b} p_{\eta_a, \eta_b}(Z)}{\max_{\eta_b} p_{0, \eta_b}(Z)} = \min_{\eta_b} l_{0, \eta_b}(Z) - \min_{\eta_a, \eta_b} l_{\eta_a, \eta_b}(Z) \quad (51)$$

- or on the ‘least favorable’ GLR

$$t_a^* = 2 \ln \frac{\max_{\eta_a} p_{\eta_a, \eta_b^*}(Z)}{p_{0, \eta_b^*}(Z)} = l_{0, \eta_b^*}(Z) - \min_{\eta_a} l_{\eta_a, \eta_b^*}(Z)$$

resulting from a minimization of noncentrality parameter (31), namely

$$\eta_b^* = \arg \min_{\eta_b} \mathbf{K}((\eta_a, \eta_b)) = -\mathbf{F}_{bb}^{-1} \mathbf{F}_{ba} \eta_a$$

with minimum value

$$\min_{\eta_b} \mathbf{K}((\eta_a, \eta_b)) = \eta_a^T \mathbf{F}_a^* \eta_a = \eta_a^T \mathcal{G}_a^* \Sigma \mathcal{G}_a^{*T} \eta_a$$

Test statistic t_a^* is often said to achieve *statistical rejection* or *decoupling* of η_b . Noncentrality parameter (50) is the expectation of quadratic form (49) under hypothesis $\eta_a \neq 0$, and its value results either from lemma 3.1 or from

$$\mathbf{K}^*(\eta_a) = \mathbf{K}((\eta_a, \eta_b^*)) \quad (52)$$

Now comments on fault rejection versus residual decorrelation are in order. Let

$$\begin{aligned} \zeta_b^* &= \mathcal{G}_b^* Z \stackrel{\Delta}{=} M_b^T \Sigma^{-1} \tilde{P}_{M_a} Z \\ &= \begin{pmatrix} -\mathbf{F}_{ba} \mathbf{F}_{aa}^{-1} & I_{q_b} \end{pmatrix} M^T \Sigma^{-1} Z \end{aligned} \quad (53)$$

Even though each of t_a^* and t_b^* achieves the best statistical decoupling between η_a and η_b , neither the two random variables ζ_a^* and ζ_b^* nor the two test statistics t_a^* and t_b^* are decorrelated under $\eta = 0$. In the following lemma, we summarize the joint properties of the transformed variables ζ_a^* and ζ_b^* , from the double point of view of decoupling and information content.

Lemma 3.8 (Decoupling and information in rejection-based transformed variables) *Let ζ^* be the stacked residual made of the two transformed variables ζ_i^* corresponding to statistical rejection of the two subsets of parameter components, namely*

$$\zeta^* \stackrel{\Delta}{=} \begin{pmatrix} \zeta_a^* \\ \zeta_b^* \end{pmatrix} = \mathcal{G}^* Z \quad (54)$$

where

$$\mathcal{G}^* \triangleq \begin{pmatrix} \mathcal{G}_a^* \\ \mathcal{G}_b^* \end{pmatrix} \quad (55)$$

The transformation matrix writes :

$$\mathcal{G}^* = \begin{pmatrix} I_{q_a} & -\mathbf{F}_{ab} \mathbf{F}_{bb}^{-1} \\ -\mathbf{F}_{ba} \mathbf{F}_{aa}^{-1} & I_{q_b} \end{pmatrix} M^T \Sigma^{-1} = \begin{pmatrix} \mathbf{F}_a^* & 0 \\ 0 & \mathbf{F}_b^* \end{pmatrix} \mathbf{F}^{-1} M^T \Sigma^{-1}$$

The distribution of ζ^* under $\eta \neq 0$ is

$$\zeta^* \sim \mathcal{N}(L^* \eta, \Sigma^*)$$

where

$$L^* = \mathcal{G}^* M = \begin{pmatrix} \mathbf{F}_a^* & 0 \\ 0 & \mathbf{F}_b^* \end{pmatrix}$$

and

$$\Sigma^* = \mathcal{G}^* \Sigma \mathcal{G}^{*T} = \begin{pmatrix} \mathbf{F}_a^* & -\mathbf{F}_a^* \mathbf{F}_{aa}^{-1} \mathbf{F}_{ab} \\ -\mathbf{F}_b^* \mathbf{F}_{bb}^{-1} \mathbf{F}_{ba} & \mathbf{F}_b^* \end{pmatrix} = \begin{pmatrix} I_{q_a} & -\mathbf{F}_{ab} \mathbf{F}_{bb}^{-1} \\ -\mathbf{F}_{ba} \mathbf{F}_{aa}^{-1} & I_{q_b} \end{pmatrix} L^* \quad (56)$$

In other words, residual ζ^* has a block-diagonal incidence matrix L^* , but a non-block-diagonal covariance matrix.

Furthermore the information content of ζ^* can be summarized as follows. Residual ζ^* contains exactly the same Fisher information about η as the original vector Z , namely

$$\mathbf{F}^* = L^{*T} \Sigma^{*-1} L^* = M^T \Sigma^{-1} M = \mathbf{F}$$

as straightforward computations show. Residual ζ^* is thus a sufficient statistics for η . Also mutual information between ζ_a^* and ζ_b^* is

$$\mathcal{I}(\zeta_a^*, \zeta_b^*) = -\frac{1}{2} \ln \frac{|\Sigma^*|}{|\mathbf{F}_a^*| |\mathbf{F}_b^*|} = -\frac{1}{2} \ln |\mathbf{F}_{aa}^{-1} \mathbf{F}_a^*| = -\frac{1}{2} \ln |\mathbf{F}_{bb}^{-1} \mathbf{F}_b^*| \quad (57)$$

It is interesting to note that the mutual information between the two types of transformed variables is the same:

$$\mathcal{I}(\zeta_a^*, \zeta_b^*) = \mathcal{I}(\tilde{\zeta}_a, \tilde{\zeta}_b)$$

Consider now the correlation issue in the quadratic forms t_i^* themselves. Lemma 3.1, definitions (46) and (53), and repeated use of (32) result in that, under $\eta = 0$,

$$\begin{aligned} \text{cov}_0(t_a^*, t_b^*) &= 2 \text{tr} \left(\mathcal{G}_a^{*T} \mathbf{F}_a^{*-1} \mathcal{G}_a^* \Sigma \mathcal{G}_b^{*T} \mathbf{F}_b^{*-1} \mathcal{G}_b^* \Sigma \right) \\ &= 2 \text{tr} \left(\mathbf{F}_{aa}^{-1} \mathbf{F}_{ab} \mathbf{F}_{bb}^{-1} \mathbf{F}_{ba} \right) \\ &= 2 \text{tr} \left(I_{q_a} - \mathbf{F}_{aa}^{-1} \mathbf{F}_a^* \right) \end{aligned} \quad (58)$$

$$= 2 \text{tr} \left(I_{q_b} - \mathbf{F}_{bb}^{-1} \mathbf{F}_b^* \right) \quad (59)$$

Note that, under no-fault hypothesis $\eta = 0$, the correlation between the two types of quadratic forms is the same:

$$\text{cov}_0(t_a^*, t_b^*) = \text{cov}_0(\tilde{t}_a, \tilde{t}_b)$$

The last formula of lemma 3.1 and expression (49) provide us with the correlation under $\eta \neq 0$

$$\text{cov}_1(t_a^*, t_b^*) = \text{cov}_0(t_a^*, t_b^*) - 4 \eta_a^T \mathbf{F}_a^* \mathbf{F}_{aa}^{-1} \mathbf{F}_{ab} \eta_b$$

Finally, the first formula of lemma 3.1 leads to the expectation of t_a^* under $\eta \neq 0$ ¹⁰

$$\mathbf{E}_1(t_a^*) = q_a + \eta_a^T \mathbf{F}_a^* \eta_a$$

At this point, it is interesting to note that these formulas and the corresponding ones for the sensitivity tests above, allow us to compare the two approaches.

Lemma 3.9 (Comparison between sensitivity and rejection methods) *Under $\eta \neq 0$,*

$$\text{var}_1(\tilde{t}_a) - \text{var}_1(t_a^*) = 4 \left(\mathbf{E}_1(\tilde{t}_a) - \mathbf{E}_1(t_a^*) \right) \quad (60)$$

$$= \eta^T \begin{pmatrix} \mathbf{F}_{aa} - \mathbf{F}_a^* & \mathbf{F}_{ab} \\ \mathbf{F}_{ba} & \mathbf{F}_{bb} - \mathbf{F}_b^* \end{pmatrix} \eta \quad (61)$$

$$\text{cov}_1(\tilde{t}_a, \tilde{t}_b) - \text{cov}_1(t_a^*, t_b^*) = 4 \eta^T \begin{pmatrix} \mathbf{F}_{aa} - \mathbf{F}_a^* & \mathbf{F}_{ab} + \mathbf{F}_a^* \mathbf{F}_{aa}^{-1} \mathbf{F}_{ab} \\ \mathbf{F}_{ba} - \mathbf{F}_b^* \mathbf{F}_{bb}^{-1} \mathbf{F}_{ba} & \mathbf{F}_{bb} - \mathbf{F}_b^* \end{pmatrix} \eta \quad (62)$$

Now, for $(\eta_a \neq 0, \eta_b = 0)$, we easily get that

$$\mathbf{E}_1(\tilde{t}_a) - \mathbf{E}_1(t_a^*) \geq 0 \quad (63)$$

$$\text{var}_1(\tilde{t}_a) - \text{var}_1(t_a^*) \geq 0 \quad (64)$$

$$\text{cov}_1(\tilde{t}_a, \tilde{t}_b) - \text{cov}_1(t_a^*, t_b^*) \geq 0 \quad (65)$$

In other words, we recover here well known experimental properties of sensitivity and minmax tests [38, 44] : under a single fault hypothesis, the mean value of a sensitivity test is higher than the mean value of the corresponding minmax test, but its variance is also higher; moreover, the correlation between different sensitivity tests is higher than the correlation between the corresponding minmax tests.

We now introduce a last transformation which shows up an interesting tradeoff between the issues of diagonal fault incidence matrix and diagonal residual covariance matrix.

3.2.4 Mixed rejection and sensitivity approach

Let $\check{\zeta}$ be the residual made of the two transformed variables ζ_a^* and $\tilde{\zeta}_b$, corresponding to statistical rejection of and sensitivity to the second subset of parameter components, namely

$$\check{\zeta} \triangleq \begin{pmatrix} \zeta_a^* \\ \tilde{\zeta}_b \end{pmatrix} = \check{\mathcal{G}} Z \quad (66)$$

where

$$\check{\mathcal{G}} \triangleq \begin{pmatrix} \mathcal{G}_a^* \\ \tilde{\mathcal{G}}_b \end{pmatrix} \quad (67)$$

The transformation matrix writes :

$$\check{\mathcal{G}} = \begin{pmatrix} \left(I_{q_a} & -\mathbf{F}_{ab} \mathbf{F}_{bb}^{-1} \right) M^T \Sigma^{-1} \\ M_b^T \Sigma^{-1} \end{pmatrix}$$

The distribution of $\check{\zeta}$ under $\eta \neq 0$ is

$$\check{\zeta} \sim \mathcal{N}(\check{L} \eta, \check{\Sigma})$$

¹⁰And *not* under $\eta_a \neq 0$ as in lemma 3.7.

where incidence matrix

$$\check{L} = \check{G} M = \begin{pmatrix} \mathbf{F}_a^* & 0 \\ \mathbf{F}_{ba} & \mathbf{F}_{bb} \end{pmatrix}$$

is *block-triangular*, and covariance matrix

$$\check{\Sigma} = \check{G} \Sigma \check{G}^T = \begin{pmatrix} \mathbf{F}_a^* & 0 \\ 0 & \mathbf{F}_{bb} \end{pmatrix}$$

is *block-diagonal*. Furthermore residual $\check{\zeta}$ contains exactly the *same Fisher information* about η as the original vector Z , namely

$$\check{\mathbf{F}} = \check{L}^T \check{\Sigma}^{-1} \check{L} = \mathbf{F}$$

Mutual information between ζ_a^* and $\tilde{\zeta}_b$ is of course $\mathcal{I}(\zeta_a^*, \tilde{\zeta}_b) = 0$.

Let us comment further the residual decorrelation issue. In the present Gaussian framework, the structure of $\check{\Sigma}$ means that ζ_a^* and $\tilde{\zeta}_b$ are not only uncorrelated, but also *independent*, which is a stronger property. This implies that t_a^* and \tilde{t}_b are also independent, as regular functions of independent variables. Consequently, we get the following result.

Lemma 3.10 (Orthogonal decomposition of global test) *Global test (30) writes*

$$t_1 = t_a^* + \tilde{t}_b \tag{68}$$

where t_a^* is given by lemma 3.7 and \tilde{t}_b is given by lemma 3.5. This decomposition is orthogonal in the sense that

$$\text{cov}_1(t_a^*, \tilde{t}_b) = \text{cov}_0(t_a^*, \tilde{t}_b) = 0 \tag{69}$$

under both *unfailed and failed hypotheses*.

Actually, the sum of (51) and (34) is obviously (27), namely

$$t_1 = t_a^* + \tilde{t}_b$$

In other words, directly from the definitions, we get that the additive decomposition of global test is valid in much more general cases than the Gaussian one, even though the independence issue (orthogonality of the sum) is more tricky in the general case. We discuss further this point in section 5.

3.2.5 Summarizing residuals properties

We now summarize the properties of the residuals of this section in table 1. The last line of this table is discussed in section 5.

Table 1: Properties enjoyed by four types of residuals ζ . The properties of $\mathbf{F}, L, \Sigma, \mathcal{I}$ in the first three columns concern the case of two fault vectors η_a, η_b . The case of three or more fault vectors is discussed in section 5.

	Sensitivity $\tilde{\zeta}$ (40)	Rejection ζ^* (54)	Mixed $\check{\zeta}$ (66)	Karhunen-L. $\check{\zeta}$ (33)
Fisher information matrix \mathbf{F}	Unchanged	Unchanged	Unchanged	Unchanged
Fault incidence matrix L	Equals Fisher	Diagonal	Triangular	
Residual covariance matrix Σ	Equals Fisher		Diagonal	Diagonal
Mutual information \mathcal{I} in ζ	$\tilde{\mathcal{I}}$	$\mathcal{I}^* = \tilde{\mathcal{I}}$	0	0
Optimality for isolation (section 5)	Single faults isolation	Multiple faults isolation		

4 Extensions and discussion

In this section, we address the case of Gaussian regression models, and we discuss how to combine the results of sections 2 and 3 for investigating the links between deterministic and stochastic approaches, and deriving possible mixed methods.

4.1 Gaussian regression models

First, it should be obvious that the solution to detection problem 2.1 in the regression case is a direct consequence of the solution to isolation problem 2.2 in the basic Gaussian case given in lemma 3.7, where $M_b \eta_b$ stands for HX and $M_a \eta_a$ for $M\eta$. We summarize it in the following lemma.

Lemma 4.1 (Gaussian detection in regression models) *For deciding between $\eta = 0$ and $\eta \neq 0$ in*

$$Z \sim \mathcal{N}(HX + M\eta, \Sigma)$$

where $\dim X \leq \dim Z$, $\dim \eta \leq \dim Z$, H and M are f.c.r. matrices and $\Sigma = \bar{\Sigma} \bar{\Sigma}^T$ with $\bar{\Sigma}$ positive definite, the projection form of the test is

$$t_{R1} = Z^T \Sigma^{-1} \left(\tilde{P}_H M (M^T \Sigma^{-1} \tilde{P}_H M)^{-1} M^T \tilde{P}_H^T \right) \Sigma^{-T} Z \quad (70)$$

where \tilde{P}_H is ‘stochastic rejector’ (24)-(25) associated with LS estimation of X .

The transformation form of the test is based on transform \mathcal{G}_M^*

$$\mathcal{G}_M^* = M^T \Sigma^{-1} \tilde{P}_H \quad (71)$$

such that

$$\mathcal{G}_M^* H = 0, \quad \mathcal{G}_M^* M = \mathcal{G}_M^* \Sigma \mathcal{G}_M^{*T}$$

4.2 Links between deterministic and stochastic approaches

It turns out that ‘stochastic rejector’ \tilde{P}_H (24) can be factorized in a manner which is useful for outlining the interaction between the deterministic and stochastic approaches. We summarize this result in the following lemma, which is only a particular case of a much more general result concerning descriptor systems with possibly singular noise covariance [37].

Lemma 4.2 (Factorization of ‘stochastic rejector’) *Let ε be a zero-mean Gaussian vector with unit covariance matrix. For*

$$Z = HX + \bar{\Sigma} \varepsilon \sim \mathcal{N}(HX, \Sigma)$$

there exist an invertible matrix A and an orthonormal matrix T , such that transformed observation AZ can be written as

$$AZ = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} Z = \begin{pmatrix} H_1 \\ 0 \end{pmatrix} X + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \quad (72)$$

where

$$\begin{aligned} A_2 &= D^{-1} S^T P, \\ A_1 &= B^{-1} (N - C A_2), \\ H_1 &= B^{-1} N H, \\ \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} &= T^T \varepsilon \end{aligned} \quad (73)$$

In these equations, matrices N, P result from QR-factorization of H

$$\begin{aligned} P H &= 0 \\ N H &\text{ full row rank} \\ N^T N + P^T P &= I, \end{aligned} \quad (74)$$

matrices D, S, T result from SVD decomposition of matrix $P \bar{\Sigma}$

$$P \bar{\Sigma} = S \begin{pmatrix} 0 & D \end{pmatrix} T^T$$

matrices B, C are defined by

$$\begin{pmatrix} B & C \end{pmatrix} \triangleq N \bar{\Sigma} T$$

and matrix B can be shown invertible.

Two consequences of this result are of particular interest for our purpose. First, comparing lemmas 4.2 and 3.3 leads to

$$\max_X l_X(Z) = Z^T A_2^T A_2 Z = Z^T \tilde{P}_H^T \Sigma^{-1} \tilde{P}_H Z$$

which results in

$$\tilde{P}_H = \bar{\Sigma} A_2$$

Second, matrix A_2 factorizes as in (73), where the right-hand term P is such that (74) holds. In other words, ‘deterministic rejector’ P factorizes ‘stochastic rejector’ \tilde{P}_H . Furthermore, because of (71), P also factorizes stochastic transform \mathcal{G}_M^* . It is interesting to note that the ‘deterministic rejector’ P considered in lemma 4.2 is obtained through QR-factorization of H , whereas the analytical redundancy approach usually relies on SVD of H , as recalled in lemma 2.1.

Considering the isolation problem 2.2, it can be shown similarly that a deterministic rejector P_b of M_b factorizes statistical rejection (minmax) transform \mathcal{G}_a^* when written as in (48).

This extends the connection between deterministic and stochastic approaches to both detection and isolation problems discussed in [60, 6].

4.3 Mixed deterministic/stochastic approaches

Pursuing further this line of results on modular estimation developed in [37], we now suggest that several combinations of the above methods can be used for computing GLR solution t_{Ra}^* to isolation problem 2.2 which extends lemma 3.7, namely

$$t_{Ra}^* = \min_{X, \eta_b} l_{X, 0, \eta_b}(Z) - \min_{X, \eta_a, \eta_b} l_{X, \eta_a, \eta_b}(Z)$$

The idea is to gradually increase the deterministic part of the processing, replacing (LS) statistical estimation by deterministic rejection as in (74), which results in

- Estimate X, η_a, η_b , as above;
- Reject η_b , estimate X, η_a ;
- Reject X, η_b , estimate η_a .

5 Multiple faults : definition and optimality criteria

In this section, we formally state several multiple fault cases and we state the isolation problem as a simultaneous multiple hypotheses testing problem. We then discuss optimality criteria, which are the relevant counterpart of the tradeoff between false alarms and detection probabilities in the ordinary hypotheses testing (detection) problem. When the fault size is small, we show that the asymptotic expansion of one of them involves only the diagonal elements of Fisher information matrix, motivating further the criteria proposed in section 6.

5.1 Multiple faults with or without causality assumptions

We now describe different types of faults, and discuss how the above sensitivity and rejection approaches can handle them. We discuss further the additive decomposition of global test reported in lemma 3.10, and show that, even for only three faults, this decomposition is no longer orthogonal.

As before, vector η is the q -dimensional fault vector. For *each component* η_j , we define the two following hypotheses

$$\begin{aligned} \mathcal{H}_j &: \eta_j = 0 \cdot e_j \\ \mathcal{K}_j &: \eta_j = \alpha_j \cdot e_j, \quad \alpha_j \neq 0 \text{ and unknown,} \end{aligned} \quad (75)$$

where e_j is the j -th coordinate vector. The null (no fault) and global alternative hypotheses are of course

$$\begin{aligned} \mathcal{H}_0 &: \cap_j \mathcal{H}_j \\ \mathcal{H}_0^c &: \cup_j \mathcal{H}_j^c = \cap_j \mathcal{K}_j \end{aligned} \quad (76)$$

Global test t_1 (28), (30) tests between \mathcal{H}_0 and \mathcal{H}_0^c . The j -th rejection test t_j^* (46), (48) tests between \mathcal{H}_j and \mathcal{K}_j . Let β_j^* be its power (probability of correct decision).

5.1.1 Single faults and sensitivity tests

A single fault on the j -th component of η corresponds to hypothesis

$$\tilde{\mathcal{H}}_j = \mathcal{K}_j \cap \left(\cap_{i \neq j} \mathcal{H}_i \right) \quad (77)$$

Sensitivity test \tilde{t}_j (35), (38) tests between \mathcal{H}_0 and $\tilde{\mathcal{H}}_j$. Let $\tilde{\beta}_j$ be its power.

The set of all sensitivity tests $(\tilde{t}_1, \dots, \tilde{t}_q)$ simultaneously solves the set of testing problems $\cup_j (\mathcal{H}_0 / \tilde{\mathcal{H}}_j)$. The point here is that the global alternative hypothesis \mathcal{H}_0^c cannot be expressed in terms of the ‘individual’ alternatives $\tilde{\mathcal{H}}_j$. However, as these ‘individual’ alternatives do not intersect, Neyman-Pearson lemma [13] tells us that

$$(\tilde{t}_1, \dots, \tilde{t}_q) \text{ maximizes } \min_j \tilde{\beta}_j$$

Moreover, achieving the fault isolation by selecting the maximizing fault number $\arg \max_j \tilde{t}_j$ corresponds to a maximum *a posteriori* decision rule, assuming a flat prior for all the single faults.

5.1.2 Multiple faults and rejection tests

Two basic types of multiple faults can be distinguished : embedded multiple faults with causality constraints, and independent multiple faults.

Embedded multiple faults Actually, we already considered the embedded case for two faults in paragraph 3.2.4 when mixing rejection of the first one and sensitivity to the second. In lemma 3.10, we have shown an *orthogonal* decomposition of global test

$$\begin{aligned} t_1 &= -2 \ln \frac{\max_{\eta_a, \eta_b} p_{\eta_a, \eta_b}}{p_{0,0}} \\ &= -2 \ln \frac{\max_{\eta_a, \eta_b} p_{\eta_a, \eta_b}}{\max_{\eta_b} p_{0, \eta_b}} - 2 \ln \frac{\max_{\eta_b} p_{0, \eta_b}}{p_{0,0}} \\ &= t_a^* + \tilde{t}_b \end{aligned}$$

where t_a^* is decorrelated from \tilde{t}_b . It turns out that, for three or more faults, the additive decomposition of global test is no longer orthogonal, as we discuss in the case of three faults for the sake of simplicity. The additive decomposition again results from

$$\begin{aligned} t_1 &= -2 \ln \frac{\max_{\eta_a, \eta_b, \eta_c} p_{\eta_a, \eta_b, \eta_c}}{p_{0,0,0}} \\ &= -2 \ln \frac{\max_{\eta_a, \eta_b, \eta_c} p_{\eta_a, \eta_b, \eta_c}}{\max_{\eta_b, \eta_c} p_{0, \eta_b, \eta_c}} - 2 \ln \frac{\max_{\eta_b, \eta_c} p_{0, \eta_b, \eta_c}}{\max_{\eta_c} p_{0,0, \eta_c}} - 2 \ln \frac{\max_{\eta_c} p_{0,0, \eta_c}}{p_{0,0,0}} \\ &= t_a^* + \underline{t}_b + \tilde{t}_c \end{aligned}$$

Here test \underline{t}_b is a rejection test, correlated with t_a^* . Nevertheless, test \tilde{t}_c is decorrelated from tests t_a^* and \underline{t}_b , and the counterpart of paragraph 3.2.4 can be stated as follows. Let \mathcal{G} the joint transformation corresponding to $\mathcal{G}_a^*, \underline{\mathcal{G}}_b, \tilde{\mathcal{G}}_c$. The incidence and covariance matrices of associated residual $\mathcal{G}Z$ are structured as

$$L = \begin{pmatrix} x & 0 & 0 \\ x & x & 0 \\ x & x & x \end{pmatrix}, \quad \Sigma = \begin{pmatrix} x & x & 0 \\ x & x & 0 \\ 0 & 0 & x \end{pmatrix}$$

where x denotes non-zero blocks. The incidence matrix L is still triangular as in the two-faults case summarized in table 1, but the covariance matrix Σ is no longer diagonal.

Independent multiple faults Here we assume no causality between the ‘individual’ faults η_j , which can occur simultaneously in any manner. The isolation problem is then to solve simultaneously the set of testing problems $\bigcup_j (\mathcal{H}_j / \mathcal{K}_j)$. Actually the set of all rejection tests (t_1^*, \dots, t_q^*) solves this multiple hypotheses testing problem [54].

We summarize this subsection in table 2. The last line of this table is the purpose of the next subsection.

5.2 Optimality criteria

It should be clear that defining an optimality criterion for such a multiple hypotheses testing problem is not straightforward. When dealing with tests powers optimized with constraints on false alarms probabilities, the main issue with multiple hypotheses tests is to maintain high powers for the individual tests. The interested reader is referred to the discussions and references in [54, 48], where several optimality criteria are investigated.

We use the following definitions [13] and notations. The density of a noncentral χ^2 -test $\chi'^2(q, \tau)$ with q degrees of freedom and noncentrality parameter τ is $\gamma_y(\frac{q}{2}, \frac{\tau}{2}, \frac{1}{2})$, where $\gamma_y(\kappa, \tau, \omega)$ is the density of the Gamma distribution

$$\begin{aligned} \gamma_y(\kappa, \tau, \omega) &\triangleq \sum_{i=0}^{\infty} \frac{1}{i!} e^{-\tau} \tau^i \gamma_y(\kappa + i, \omega) \\ \gamma_y(\kappa, \omega) &\triangleq \frac{1}{\Gamma(\kappa)} \omega^\kappa e^{-\omega y} y^{\kappa-1} \\ \Gamma(\kappa) &\triangleq \int_0^{\infty} e^{-x} x^{\kappa-1} dx \end{aligned} \tag{78}$$

Table 2: Fault hypotheses, isolation tests and their properties.

Test	Between hypotheses	With ‘power’
t_1 (30)	$\mathcal{H}_0/\mathcal{H}_0^c$ (76)	β
t_j^* (48)	$\mathcal{H}_j/\mathcal{K}_j$ (75)	β_j^*
\tilde{t}_j (37)	$\mathcal{H}_0/\tilde{\mathcal{H}}_j$ (77)	$\tilde{\beta}_j$
$(\tilde{t}_1, \dots, \tilde{t}_q)$	$\bigcup_j (\mathcal{H}_0/\tilde{\mathcal{H}}_j)$	$\min_{1 \leq j \leq q} \tilde{\beta}_j$
(t_1^*, \dots, t_q^*)	$\bigcup_j (\mathcal{H}_j/\mathcal{K}_j)$	$\min_{1 \leq j \leq q} \beta_j^*, \quad \sum_{j=1}^q \beta_j^*$

Assuming a threshold λ , the power of such a test is defined by

$$\beta = f\left(\kappa, \frac{\tau}{2}, \omega, \lambda\right) \triangleq \int_{\lambda}^{\infty} \gamma_y\left(\kappa, \frac{\tau}{2}, \omega\right) dy \quad (79)$$

with $\kappa = \frac{q}{2}, \omega = \frac{1}{2}$. It is an increasing function of noncentrality parameter τ .

5.2.1 Optimality of rejection tests

It turns out [54] that the set of all rejection tests (t_1^*, \dots, t_q^*) satisfies two optimality criteria. Roughly speaking,

$$(t_1^*, \dots, t_q^*) \text{ maximizes both } \min_{1 \leq j \leq q} \beta_j^* \text{ and } \sum_{j=1}^q \beta_j^*,$$

where β_j^* is the power of test t_j^* . We now show that the second criterion involves only the diagonal elements of Fisher information matrix, and then we give an explicit formula which is valid asymptotically (for small faults).

Before showing this, we should note that these optimality results are technically more involved : each alternative hypothesis \mathcal{K}_j in (75) needs to be moved to

$$\mathcal{K}_j : \eta_j = \alpha_j \cdot e_j, \quad |\alpha_j| = \|\eta_j\| \geq \rho_j \quad (80)$$

Under this assumption, and denoting by λ_j^* the threshold for test t_j^* , its power β_j^* writes

$$\beta_j^* = f\left(\frac{1}{2}, \tau_j^*, \frac{1}{2}, \lambda_j^*\right) \quad (81)$$

$$\begin{aligned} \text{where } \tau_j^* &= \frac{1}{2} \min_{\alpha_j \geq \rho_j} \left(\alpha_j^2 e_j^T \mathbf{F} e_j \right) \\ &= \frac{1}{2} \rho_j^2 \mathbf{F}_{jj} \end{aligned} \quad (82)$$

as results from (52).

We finally mention that, in (81), all the thresholds λ_j^* are chosen equal to a constant λ , for ensuring equal individual false alarm probabilities for the tests t_j^* .

5.2.2 Isolation optimality and Fisher information

Now, asymptotically (for small faults), we can derive an explicit formula for $\bar{\beta}$. We consider the optimality criterion

$$\begin{aligned}\bar{\beta} &\triangleq \sum_{j=1}^q f\left(\frac{1}{2}, \tau_j^*, \frac{1}{2}, \lambda\right) \\ &= \int_{\lambda}^{\infty} \sum_{j=1}^q \gamma_y\left(\frac{1}{2}, \tau_j^*, \frac{1}{2}\right) dy\end{aligned}$$

with τ_j^* defined in (82).

We now consider small faults (and thus small noncentrality parameters τ_j^*), and we investigate the asymptotic form of this criterion.

First, following [54], we select the sizes ρ_j of the alternatives \mathcal{K}_j in (80), so as to have equal individual powers

$$\int_{\lambda}^{\infty} \gamma_y\left(\frac{1}{2}, \tau_j^*, \frac{1}{2}\right) dy = \text{cst in } j$$

Second, we let the resulting fault sizes ρ_j^* go to zero in a normalized fashion, assuming that

$$\tau_j^* = \frac{1}{2} \rho_j^{*2} \mathbf{F}_{jj} = \alpha d_j, \quad \alpha \rightarrow 0$$

Under these assumptions, and because of (78), the criterion writes

$$\begin{aligned}\bar{\beta} &= q - \int_0^{\lambda} \sum_{j=1}^q \sum_{i=0}^{\infty} \frac{1}{i!} e^{-\tau_j^*} \tau_j^{*i} \gamma_y\left(\frac{1}{2} + i, \frac{1}{2}\right) dy \\ &= q - \int_0^{\lambda} \sum_{i=0}^{\infty} g_i(\alpha) \gamma_y\left(\frac{1}{2} + i, \frac{1}{2}\right) dy\end{aligned}$$

where function g_i is defined by

$$g_i(\alpha) \triangleq \frac{1}{i!} \sum_{j=1}^q e^{-\tau_j^*} \tau_j^{*i} = \frac{\alpha^i}{i!} \sum_{j=1}^q e^{-\alpha d_j} d_j^i$$

and is such that $g'_i(\alpha)|_{\alpha=0} = 0$ whenever $i \geq 2$.

Now, the following first order Taylor expansion of $\bar{\beta}$ holds :

$$\begin{aligned}\bar{\beta} &\approx \left(1 - f\left(\frac{3}{2}, 0, \frac{1}{2}, \lambda\right)\right) \left(\sum_{j=1}^q d_j\right) \alpha \\ &\approx \left(1 - f\left(\frac{3}{2}, 0, \frac{1}{2}, \lambda\right)\right) \left(\frac{1}{2} \sum_{j=1}^q \rho_j^{*2} \mathbf{F}_{jj}\right)\end{aligned}\tag{83}$$

We thus end up, for small faults, with an explicit formula for a criterion which always rules out the off-diagonal terms of Fisher information matrix. Note that we have considered here the case of q one-dimensional fault vectors, and that a similar result could be obtained for a set of vector (and not scalar) faults with *equal sizes* and total size q .

This is a further motivation, in addition to the results summarized in table 1, for the residual optimization criteria we propose in the next section.

6 Information criteria for residual optimization

In the previous sections, we have explained how to design linear transformations of the initial data, and discussed several ways to establish links between deterministic and stochastic designs. We now take an information point of view in order to introduce several residual optimization criteria, which merge the generation and evaluation steps in a unique framework. Before proceeding, we recall that other residual optimization criteria have been proposed in [29, 30, 25, 26, 18, 19, 55, 35, 33, 50]. We have discussed some of them in the previous sections.

6.1 Background and motivation

It results from the discussions of the previous sections that, for a q -dimensional fault vector Υ , any s -dimensional residual

$$\zeta = \mathcal{A}_{s,\ell} \mathcal{Y}$$

based on a ℓ -size stacked observation vector \mathcal{Y} , is normally distributed

$$\zeta \sim \mathcal{N}(L_\ell \Upsilon, \mathcal{R}_\ell)$$

with fault incidence and residual covariance matrices

$$L_\ell = \mathcal{A}_{s,\ell} \mathcal{M}_\ell, \quad \mathcal{R}_\ell = \mathcal{A}_{s,\ell} \Sigma_\ell \mathcal{A}_{s,\ell}^T$$

We now discuss the residual optimization problem, namely the problem of selecting the window-size ℓ , the residual dimension s and the linear transform $\mathcal{A}_{s,\ell}$ which are optimal with respect to a given criterion.

Before introducing criteria for this purpose, let us comment further on the above problem statement. First, one could argue that in practice covariance matrix Σ_ℓ computed as in (13) is not known, especially because the state and observation noise covariance matrices Q_x and Q_y in (4) themselves are unknown. We rather think that it is practically simpler and more efficient to tune such unknown but meaningful parameters Q_x and Q_y , than to tune other values such as multiple thresholds for test functions whose distributions are unknown. Second, in order to outline the importance of the residual correlation issue, we insist again on that, even when the state and observation noise covariance matrices are assumed to be diagonal, covariance matrix Σ_ℓ is not block-diagonal, whenever the window size ℓ is greater than or equal to 3. Furthermore, residual covariance matrix \mathcal{R}_ℓ is not diagonal neither, whatever the transformation matrix $\mathcal{A}_{s,\ell}$ is : it is well known that the only observer which provides us with residuals having a diagonal covariance matrix is Kalman filter! Finally, we also recall that, for both deterministic and stochastic approaches to residual generation, transformation $\mathcal{A}_{s,\ell}$ is subject to the linear constraint (17) and that the additional constraint (18) holds in the deterministic case.

We now introduce several criteria for residual optimization. As explained in the introduction, we base our criteria on Fisher information

$$\begin{aligned} \mathbf{F}_\zeta &= L_\ell^T \mathcal{R}_\ell^{-1} L_\ell \\ &= \mathcal{M}_\ell^T \mathcal{A}_{s,\ell}^T \left(\mathcal{A}_{s,\ell} \Sigma_\ell \mathcal{A}_{s,\ell}^T \right)^{-1} \mathcal{A}_{s,\ell} \mathcal{M}_\ell \end{aligned} \quad (84)$$

This quantity is unchanged when transformation matrix \mathcal{A} is pre-multiplied by any invertible matrix. Fisher information matrix is known to play a key role in statistical inference, because of Cramer-Rao lower bound in parameter estimation. Several scalar functions of this matrix have been widely used for asymptotic performance analysis of identification [39] and direction finding [17] algorithms, optimal design of experiments [3] and optimal sensor location [45] including for fault detection [10]. Most of these criteria have the form [3, 46]

$$c_z(\mathbf{F}) = \left(\frac{\text{tr } \mathbf{F}^z}{q} \right)^{1/z} \quad (85)$$

for *negative* z . Particular cases are determinant, trace and extremal eigenvalue norms

$$\begin{aligned}\lim_{z \rightarrow 0} c_z(\mathbf{F}) &= |\mathbf{F}|^{1/q} \\ c_{-1}(\mathbf{F}) &= \frac{q}{\text{tr}(\mathbf{F}^{-1})} \\ \lim_{z \rightarrow -\infty} c_z(\mathbf{F}) &= \lambda_{\min}(\mathbf{F})\end{aligned}$$

In this section, we show that these and other criteria are also useful residual optimization criteria.

Two comments are in order concerning FDI. First, as noticed in section 2, existing parity checks generally do not keep Fisher information about the fault contained in the original data; again compare (2) with (3). There is thus a need for controlling this information in some way. Second, the power of test (30) is an increasing function of noncentrality parameter (31), which writes

$$\mathbf{K}(\Upsilon) = \Upsilon^T L_\ell^T \mathcal{R}_\ell^{-1} L_\ell \Upsilon = \Upsilon^T \mathbf{F} \Upsilon \quad (86)$$

in terms of the possible fault vector Υ^{11} . Because of (86), a diagonal matrix \mathbf{F} would indicate that the total ‘inertia’ due to the fault Υ is explained by the sum of the contributions of the individual components Υ_j , without any cross-term. Actually, one of the optimal isolation criteria in [54] is of this form, asymptotically for small faults Υ_j , as shown in (83). Moreover, real data processing [8] show that isolation is difficult for non-zero off-diagonal terms in \mathbf{F} . We thus suggest that minimizing a distance between Fisher information matrix \mathbf{F} and a diagonal matrix is thus an appropriate optimization criterion, in order to balance the sensitivities of the residual components ζ_j with respect to the individual fault components Υ_j [29, 35].

We now propose three such criteria minimizing a distance between Fisher information matrix \mathbf{F} and a diagonal matrix. They are based on a particular distance between matrices, and show up sound tradeoffs between the simplest scalar measures of multivariable scatter, namely trace and determinant criteria.

6.2 A distance between matrices

We define a distance between two q -dimensional symmetric positive definite matrices Σ_1 and Σ_2 as Kullback distance between two Gaussian vectors, having the same mean μ and those matrices as covariances¹². This distance

$$\mathbf{K}(\Sigma_1, \Sigma_2) = \frac{1}{2} \left(\text{tr} \left(\Sigma_1 \Sigma_2^{-1} - I_q \right) - \ln \left| \Sigma_1 \Sigma_2^{-1} \right| \right) = \mathbf{K}(\Sigma_2^{-1} \Sigma_1, I_q) \quad (87)$$

does not depend on the mean value μ , as straightforward computations show.

Distance (87) is tightly connected to maximum likelihood estimation [15] : the log-likelihood of a sample of independent zero-mean Gaussian vectors with unknown covariance matrix Σ is equal to Kullback distance between the empirical covariance matrix $\hat{\Sigma}$ and Σ , up to a constant in Σ . It has been used for array calibration [27].

Kullback distance is invariant under a change of basis, namely

$$\mathbf{K}(A \Sigma_1 A^T, A \Sigma_2 A^T) = \mathbf{K}(\Sigma_1, \Sigma_2)$$

whatever invertible matrix A is. In particular, this distance is invariant under orthogonal transforms, and it is scale-invariant

$$\mathbf{K}(\delta \Sigma_1, \delta \Sigma_2) = \mathbf{K}(\Sigma_1, \Sigma_2)$$

¹¹This quantity is nothing but Kullback distance between two Gaussian vectors with different means (0 and $L\Upsilon$) and the same covariance matrix Σ , and thus Kullback distance between the unfailed and failed models [6].

¹²Kullback distance may be defined for non invertible positive matrices also.

for any positive scalar δ .

When the second matrix is diagonal, Kullback distance can be written as

$$\mathbf{K}(\Sigma, \Delta) = \frac{1}{2} \left(\sum_{i=1}^{\dim \Sigma} \frac{\sigma_{ii}}{\delta_i} + \sum_{i=1}^{\dim \Sigma} \ln \delta_i - \ln |\Sigma| - \dim \Sigma \right) \quad (88)$$

where $\Delta = \text{diag}(\delta_i)$, and σ_{ii} are the diagonal elements of Σ .

When matrix Σ is close to identity, the following approximations hold up to first order

$$\mathbf{K}(\Sigma, I) \approx \|\Sigma - I\|_F^2 \approx \|\Sigma^{-1} - I\|_F^2$$

in terms of Frobenius norm (19).

Finally, note that a more geometric distance to the identity could be defined by

$$\frac{\|\Sigma\|_F^2}{\dim \Sigma} - \left(\frac{\text{tr} \Sigma}{\dim \Sigma} \right)^2$$

as in [43]. However, this matrix complexity measure is invariant under orthogonal transforms, but not scale-invariant.

6.3 Three isolation criteria

We now show that Kullback distance can be used in several ways for deriving residual design criteria which reflect the asymptotic expression (83) of Spjøtvoll's isolation optimality criterion. We introduce three criteria based on Kullback distance. Up to our knowledge, the use of this distance is new in the context of residual optimization for fault detection and isolation.

The first criterion is based on Kullback distance between Fisher information matrix \mathbf{F} and the identity matrix

$$C_1(\mathbf{F}) = \mathbf{K}(\mathbf{F}, I_q) = \frac{1}{2} (\text{tr} \mathbf{F} - \ln |\mathbf{F}| - q) \quad (89)$$

Kullback distance to identity has been used for model approximation [2] and source separation [16]. But it corresponds to a very strong isolation constraint, because the faults generally have different units¹³.

The second isolation criterion relaxes the normalization constraint. It only requires balanced sensitivities [29] by minimizing Kullback distance with respect to an unknown matrix proportional to the identity

$$C_2(\mathbf{F}) = \min_{\delta > 0} \mathbf{K}(\mathbf{F}, \delta I_q)$$

where

$$\mathbf{K}(\mathbf{F}, \delta I_q) = \frac{1}{2} \left(\frac{1}{\delta} \text{tr} \mathbf{F} - \ln |\mathbf{F}| - q + q \ln \delta \right)$$

The minimizing value is

$$\delta^* = \frac{\text{tr} \mathbf{F}}{q}$$

and results in

$$C_2(\mathbf{F}) = \frac{1}{2} \left(q \ln \frac{\text{tr} \mathbf{F}}{q} - \ln |\mathbf{F}| \right) = \frac{1}{2} \ln \frac{\left(\frac{\text{tr} \mathbf{F}}{q} \right)^q}{|\mathbf{F}|} \quad (90)$$

This criterion has been considered in [43, 14] as a measure of maximal complexity. Actually it turns out that

$$C_2(\mathbf{F}) = \max_{AA^T=I} C_0(A \mathbf{F} A^T)$$

¹³One typical example is in vibration mechanics, where the order of magnitudes of the faults are definitely different for volumic masses and Young moduli [11, 21].

where

$$C_0(\mathbf{F}) = -\frac{1}{2} \ln \frac{|\mathbf{F}|}{\prod_{i=1}^q \mathbf{F}_{ii}}$$

is mutual information between the components of a Gaussian vector with covariance matrix \mathbf{F}^{14} , and is not invariant under orthogonal transforms [59, 14]

$$C_0\left(A \mathbf{F} A^T\right) \neq C_0(\mathbf{F})$$

for $AA^T = I$.

For the third criterion, we relax the assumption of a diagonal matrix proportional to identity. We rather compute the smallest possible Kullback distance with respect to an unknown positive q -dimensional diagonal matrix¹⁵

$$C_3(\mathbf{F}) = \min_{\Delta_q > 0} \mathbf{K}(\mathbf{F}, \Delta_q)$$

Computing the componentwise derivatives of (88), we obtain the optimum values

$$\delta_i^* = \mathbf{F}_{ii} \quad (1 \leq i \leq q)$$

and

$$C_3(\mathbf{F}) = \mathbf{K}(\mathbf{F}, \Delta_q^*) = -\frac{1}{2} \ln \frac{|\mathbf{F}|}{\prod_{i=1}^q \mathbf{F}_{ii}} = C_0(\mathbf{F}) \quad (91)$$

The difference between the three criteria is made clearer in

$$\begin{aligned} C_1(\mathbf{F}) &= \frac{1}{2} (\sum_{i=1}^q \mathbf{F}_{ii}) - \ln |\mathbf{F}| - q \\ C_2(\mathbf{F}) &= \frac{1}{2} (q \ln (\sum_{i=1}^q \mathbf{F}_{ii}) - \ln |\mathbf{F}| - q \ln q) \\ C_3(\mathbf{F}) &= \frac{1}{2} (\sum_{i=1}^q \ln \mathbf{F}_{ii}) - \ln |\mathbf{F}| \end{aligned} \quad (92)$$

Only the first two criteria are invariant under orthogonal transforms.

6.4 Residual optimization

The problem is now to optimize each of these criteria C_i with respect to window size ℓ , residual dimension s and data transformation $\mathcal{A}_{s,\ell}$ matrix, taking into account the structure of Fisher matrix in (84) and the linear constraints (17) and (18), that is :

$$\left\{ \begin{array}{l} \min_{\mathcal{A}_{s,\ell}, s, \ell} C_i(\mathbf{F}), \\ \mathbf{F} = \mathcal{M}_\ell^T \mathcal{A}_{s,\ell}^T \left(\mathcal{A}_{s,\ell} \Sigma_\ell \mathcal{A}_{s,\ell}^T \right)^{-1} \mathcal{A}_{s,\ell} \mathcal{M}_\ell \\ \mathcal{A}_{s,\ell} \mathcal{O}_\ell = 0 \\ \mathcal{A}_{s,\ell} \mathcal{A}_{s,\ell}^T = I_s \end{array} \right. \quad (93)$$

¹⁴Note that sensitivity residual $\tilde{\zeta}$ in (40) is one such vector!

¹⁵Because actual isolation algorithms sometimes take advantage of a rejection of faults having very small magnitudes [21], a fourth criterion of the form

$$C_4(\mathbf{F}) = \min_{\inf(\Delta_q) > \rho} \mathbf{K}(\mathbf{F}, \Delta_q)$$

could be of interest, but will not be discussed further here.

It should be reminded that the ‘known’ matrices \mathcal{M}_ℓ , Σ_ℓ , \mathcal{O}_ℓ have particular structures, and dimensions increasing with ℓ ($\ell r \times q$, $\ell r \times \ell r$, $\ell r \times n$, respectively). Matrix $\mathcal{A}_{s,\ell}$ has dimensions $s \times \ell r$, where $s \leq n_\ell$ and $(\ell r - n_\ell)$ is the rank of \mathcal{O}_ℓ .

The interested reader is referred to [42] for an extensive treatment of matrix differential calculus. Before proceeding, we mention that another optimization of criterion (89) under constraints similar to but simpler than (84) has been used for source separation [16], and that a stronger (Toeplitz) structural constraint has been considered for array calibration [27].

6.4.1 Computing the gradients of the criteria

We first compute the gradients $\frac{\partial C_i}{\partial \mathbf{F}}$. Let J_i be the diagonal matrix whose only nonzero element is 1 in i -th position. Remembering that \mathbf{F} is symmetric, and plugging the elementary formulas

$$\begin{aligned}\frac{\partial \mathbf{F}_{ii}}{\partial \mathbf{F}} &= J_i \\ \frac{\partial \operatorname{tr} \mathbf{F}}{\partial \mathbf{F}} &= I_q \\ \frac{\partial \ln |\mathbf{F}|}{\partial \mathbf{F}} &= \mathbf{F}^{-1}\end{aligned}$$

in the derivatives of criteria (89), (90) and (91), we get

$$\frac{\partial C_1}{\partial \mathbf{F}} = \frac{1}{2} (I_q - \mathbf{F}^{-1}) \quad (94)$$

$$\frac{\partial C_2}{\partial \mathbf{F}} = \frac{1}{2} \left(\frac{q}{\operatorname{tr} \mathbf{F}} I_q - \mathbf{F}^{-1} \right) \quad (95)$$

$$\frac{\partial C_3}{\partial \mathbf{F}} = \frac{1}{2} \left(\sum_i \frac{J_i}{\mathbf{F}_{ii}} - \mathbf{F}^{-1} \right) \quad (96)$$

Then, we assume integers ℓ and s to be fixed, and we compute the gradients $\frac{\partial C_i}{\partial \mathcal{A}_{s,\ell}}$. To this end, it is sufficient to use the definition

$$C_i(\mathcal{A}_{s,\ell} + \delta \mathcal{A}_{s,\ell}) = C_i(\mathcal{A}_{s,\ell}) + \operatorname{tr} \left[\left(\frac{\partial C_i}{\partial \mathcal{A}_{s,\ell}} \right)^T (\delta \mathcal{A}_{s,\ell}) \right] + o(\delta \mathcal{A}_{s,\ell}) \quad (97)$$

and to insert the expression of the variation $\delta \mathbf{F}$ in terms of the variation $\delta \mathcal{A}_{s,\ell}$ in the second term of the right hand-side of

$$C_i(\mathbf{F} + \delta \mathbf{F}) = C_i(\mathbf{F}) + \operatorname{tr} \left[\left(\frac{\partial C_i}{\partial \mathbf{F}} \right) (\delta \mathbf{F}) \right] + o(\delta \mathbf{F}) \quad (98)$$

simply because the left hand-sides of equations (97) and (98) are equal. Writing symmetric matrix $\delta \mathbf{F}$ as $\delta \mathbf{F} = \delta \mathbf{G} + \delta \mathbf{G}^T$, simple but tedious computations lead to

$$\delta \mathbf{G} = \mathcal{M}_\ell^T \mathcal{A}_{s,\ell}^T \left(\mathcal{A}_{s,\ell} \Sigma_\ell \mathcal{A}_{s,\ell}^T \right)^{-1} (\delta \mathcal{A}_{s,\ell}) \left(I_{\ell r} - \Sigma_\ell \mathcal{A}_{s,\ell}^T \left(\mathcal{A}_{s,\ell} \Sigma_\ell \mathcal{A}_{s,\ell}^T \right)^{-1} \mathcal{A}_{s,\ell} \right) \mathcal{M}_\ell$$

Elementary manipulations then give the gradients

$$\frac{\partial C_i}{\partial \mathcal{A}_{s,\ell}} = 2 \left(\mathcal{A}_{s,\ell} \Sigma_\ell \mathcal{A}_{s,\ell}^T \right)^{-1} \mathcal{A}_{s,\ell} \mathcal{M}_\ell \frac{\partial C_i}{\partial \mathbf{F}} \mathcal{M}_\ell^T \left(I_{\ell r} - \mathcal{A}_{s,\ell}^T \left(\mathcal{A}_{s,\ell} \Sigma_\ell \mathcal{A}_{s,\ell}^T \right)^{-1} \mathcal{A}_{s,\ell} \Sigma_\ell \right) \quad (99)$$

Note that the relative gradients are equal to zero, namely

$$\nabla_{\mathcal{A}_{s,\ell}} C_i \triangleq \frac{\partial C_i}{\partial \mathcal{A}_{s,\ell}} \mathcal{A}_{s,\ell}^T = 0 \quad (100)$$

6.4.2 Taking the constraints into account

We now consider the Lagrangian

$$\mathcal{L}_i(\mathcal{A}_{s,\ell}) = C_i(\mathcal{A}_{s,\ell}) - \text{tr} \left(\Lambda^T \mathcal{A}_{s,\ell} \mathcal{O}_\ell \right) - \text{tr} \left(\Omega \left(\mathcal{A}_{s,\ell} \mathcal{A}_{s,\ell}^T - I_s \right) \right)$$

where the Lagrange multipliers are the elements of the $s \times n$ matrix Λ and the $s \times s$ symmetric matrix Ω . The gradient of this Lagrangian is

$$\frac{\partial \mathcal{L}_i}{\partial \mathcal{A}_{s,\ell}} = \frac{\partial C_i}{\partial \mathcal{A}_{s,\ell}} - \Lambda \mathcal{O}_\ell^T - 2 \Omega \mathcal{A}_{s,\ell} \quad (101)$$

An optimal transformation $\mathcal{A}_{s,\ell}^*$ is a solution of equation

$$\frac{\partial \mathcal{L}_i}{\partial \mathcal{A}_{s,\ell}} = 0 \quad (102)$$

and is a function of the Lagrange multipliers, which are such that

$$\mathcal{A}_{s,\ell}^* \mathcal{O}_\ell = 0, \quad \text{and} \quad \mathcal{A}_{s,\ell}^* \mathcal{A}_{s,\ell}^{*T} = I_s$$

Because of (100), we also have

$$\Lambda \mathcal{O}_\ell^T \mathcal{A}_{s,\ell}^{*T} + 2 \Omega \mathcal{A}_{s,\ell}^* \mathcal{A}_{s,\ell}^{*T} = 0$$

from which we conclude that the Lagrange multipliers are

$$\begin{cases} \Omega = 0 \\ \Lambda \text{ is any full rank } s \times n \text{ matrix} \end{cases} \quad (103)$$

Inserting (99) and (103) in (101), and solving (102), we thus get a parametrization of the optimal transformations $\mathcal{A}_{s,\ell}^*$ which are defined by

$$\left(\mathcal{A}_{s,\ell}^* \Sigma_\ell \mathcal{A}_{s,\ell}^{*T} \right)^{-1} \mathcal{A}_{s,\ell}^* \mathcal{M}_\ell \frac{\partial C_i}{\partial \mathbf{F}} \mathcal{M}_\ell^T \left(I_{lr} - \mathcal{A}_{s,\ell}^{*T} \left(\mathcal{A}_{s,\ell}^* \Sigma_\ell \mathcal{A}_{s,\ell}^{*T} \right)^{-1} \mathcal{A}_{s,\ell}^* \Sigma_\ell \right) - \Lambda \mathcal{O}_\ell^T = 0 \quad (104)$$

where $\frac{\partial C_i}{\partial \mathbf{F}}$ should be replaced by any of the expressions (94), (95) or (96).

Now, some comments are in order, for which we distinguish the case where the residual dimension is maximum, namely $s = n_\ell$, and the case $s \leq n_\ell$. In the former, two any solutions are related by a pre-multiplication with an invertible matrix. Thus the only remaining optimization is in terms of the window size ℓ .

6.4.3 Optimizing the dimensions

We now consider how to optimize with respect to the residual dimension s and window size ℓ . It is important to notice that, as our criteria are based on the comparison of noncentrality parameters of χ^2 -tests (and thus on Fisher information matrix of the corresponding residuals), they implicitly assume these tests to have the same number of degrees of freedom. Therefore, for comparing residual of different sizes, and thus tests with different numbers of degrees of freedom, it is necessary to add to criteria $C_i(\mathbf{F})$ a term which compensates for this effect, as done for example in [1, 10].

7 Conclusion

We have taken an information point of view for discussing links and possible interactions between deterministic and stochastic approaches to FDI, and for proposing new criteria for residual optimization. This paper should be viewed only as a first step towards a broader picture. Questions open for future research are of different kinds.

First, the optimization of $C_i(\mathbf{F})$ should be given numerical solutions, based on the general forms (104) and exploiting the structures of matrices $\mathcal{O}_\ell, \mathcal{M}_\ell, \Sigma_\ell$ defined in section 2.

Second, the actual benefit of using these criteria for practical design should of course be investigated further.

Third, alternative criteria should be investigated, based on both controlling the information discrepancy through $\mathbf{K}(\mathbf{F}_\zeta, \mathbf{F}_y)$ and improving the isolation through $C_i(\mathbf{F}_\zeta)$.

Finally, the new proposed criteria should be investigated as possible alternatives to previously proposed criteria for optimum sensor location as in [10] and references therein on one hand, and for fault detectability as in [57, 6] on the other one.

References

- [1] H. AKAIKE (1971). Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Intern. Symposium on Information Theory*, Tsahkadsor, Arménie, pp.267-281.
- [2] B.D.O. ANDERSON, J.B. MOORE and R.M. HAWKES (1978). Model approximations via prediction error identification. *Automatica*, vol.14, pp.615-622.
- [3] A.C. ATKINSON (1989). Recent developments in the method of optimum and related experimental designs. *Int. Stat. Review*, vol.56, no 2, pp.99-115.
- [4] M. BASAWA (1985). Neyman-Le Cam tests based on estimating functions. *Proc. Berkeley Conf. in honor of Neyman and Kiefer*, vol.2, L.M. Le Cam and R.A. Olshen (Eds.), Wadsworth Inc., pp.811-825.
- [5] M. BASSEVILLE (1988). Detecting changes in signals and systems – A survey. *Automatica*, vol.24, no 3, pp.309-326.
- [6] M. BASSEVILLE and I. NIKIFOROV (1993). *Detection of Abrupt Changes – Theory and Applications*. Prentice Hall Information and System Sciences Serie (T.Kailath, ed.), Englewood Cliffs, N.J.
- [7] M. BASSEVILLE and Q. ZHANG (1996). Monitoring nonlinear dynamical systems: a combined input-output and local approach. Submitted to *IEEE Decision and Control Conference, CDC 96*, Kobe, Japan, Dec.1996.
- [8] M. BASSEVILLE, A. BENVENISTE, G. MATHIS and Q. ZHANG (1994). Monitoring the combustion set of a gas turbine. *Proc. Safeprocess'94*, Helsinki, Finland.
- [9] M. BASSEVILLE, A. BENVENISTE, G. MOUSTAKIDES and A. ROUGÉE (1987a). Detection and diagnosis of changes in the eigenstructure of nonstationary multivariable systems. *Automatica*, vol.23, no 3, pp.479-489.
- [10] 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.
- [11] 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.
- [12] A. BENVENISTE, M. BASSEVILLE and G. MOUSTAKIDES (1987). The asymptotic local approach to change detection and model validation. *IEEE Trans. on Automatic Control*, vol.AC-32, no 7, pp.583-592.
- [13] A.A. BOROVKOV (1984). *Theory of Mathematical Statistics - Estimation and Hypotheses Testing*, Naouka, Moscow (in Russian). Translated in French under the title *Statistique Mathématique – Estimation et Tests d'Hypothèses*, Mir, Paris, 1987.
- [14] H. BOZDOGAN (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics, Part A: Theory and Methods*, vol.19, no 1, pp.221-278.
- [15] J.P. BURG, D.G. LUENBERGER and D.L. WENGER (1982). Estimation of structured covariance matrices. *Proc. IEEE*, vol.70, no 9, pp.963-974.

- [16] J.F. CARDOSO and B. LAHELD (1994). Equivariant adaptive source separation. Submitted.
- [17] J.F. CARDOSO and É. MOULINES (1994). How much more DOA information in higher-order statistics? *Proc. IEEE Workshop on Statistical Signal and Array Processing*, Québec, July 1994.
- [18] J. CHEN, R.J. PATTON and H.Y. ZHANG (1993). A multi-criteria optimization approach to the design of robust fault detection algorithms. *Proc. Tooldiag'93*, Int. Conf. on Fault Diagnosis, Toulouse, F., April 1993.
- [19] J. CHEN, R.J. PATTON and H.Y. ZHANG (1996). Design of unknown input observers and robust fault detection filters. *Int. J. Control*, vol.63, no 1, pp.85-105.
- [20] T.M. COVER and J.A. THOMAS (1991). *Elements of Information Theory*. Wiley Series in Telecommunications.
- [21] 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).
- [22] R.K. DOUGLAS and J.L. SPEYER (1991). Process and sensor noise robustness in detection filter design. *Proc. American Control Conf.*, Boston, MA, pp.2274-2279.
- [23] A. EMAMI-NAEINI, M.M. AKHTER and S.M. ROCK (1988). Effect of model uncertainty on failure detection : the threshold selector. *IEEE Trans. on Automatic 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 (1991). Enhancement of robustness in observer-based fault detection. *Proc. Safe-process'91*, Baden-Baden, FRG., pp.275-287.
- [26] P.M. FRANK and X. DING (1994). Frequency domain approach to optimally robust residual generation and evaluation for model-based fault diagnosis. *Automatica*, vol.30, no 5, pp.789-804.
- [27] J.J. FUCHS (1993). Shape calibration for a nominally linear equispaced array. *Proc. Icassp'93*, vol.4, pp.300-303.
- [28] J.J. GERTLER (1988). Survey of model-based failure detection and isolation in complex plants. *IEEE Control Systems Magazine*, vol.8, no 6, pp.3-11.
- [29] J.J. GERTLER and Q. LUO (1989). Robust isolable models for failure diagnosis. *AIChE*, vol.35, pp.1856-1868.
- [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 and J. MONAJEMY (1995). Generating directional residuals with dynamic parity relations. *Automatica*, vol.31, no 4, pp.627-635.
- [32] F.A. GRAYBILL (1983). *Matrices with Applications in Statistics*. Wadsworth Statistics/Probability Series, 2nd ed.
- [33] M. HOU and P.C. MÜLLER (1994). Fault detection and isolation observers. *Int. J. Control*, vol.60, no 5, pp.827-846.
- [34] R. ISERMANN (1984). Process fault detection based on modeling and estimation methods – A survey. *Automatica*, vol.20, pp.387-404.

- [35] M. KINNAERT (1996). Design of redundancy relations for failure detection and isolation by constrained optimization. *Int. J. Control*, vol.63, no 3, pp.609-622.
- [36] V.C. KLEMA and A.J. LAUB (1980). The singular value decomposition : its computation and some applications. *IEEE Trans. on Automatic Control*, vol.AC-25, no 2, pp.164-176.
- [37] B.C. LÉVY, A. BENVENISTE and R.N. NIKOUKHAH (1996). High-level primitives for recursive maximum likelihood estimation. *IEEE Trans. on Automatic Control*, vol.AC-41, no 8, to appear.
- [38] G. LE VEY (1988). *Analyse Modale et Surveillance Vibratoire des Machines Tournantes*. Thesis, Rennes I University. (In French).
- [39] L. LJUNG (1987). *System Identification – Theory for the User*. Prentice Hall Information and System Sciences Serie (T.Kailath, ed.), Englewood Cliffs, N.J.
- [40] X.-C. LOU, A.S. WILLSKY and G.C. VERGHESE (1986). Optimally robust redundancy relations for failure detection in uncertain systems. *Automatica*, vol.22, no 3, pp.333-344.
- [41] J.F. MAGNI and P. MOUYON (1994). On residual generation by observer and parity space approaches. *IEEE Trans. on Automatic Control*, vol.AC-39, no 2, pp.441-447.
- [42] J.R. MAGNUS and H. NEUDEKER (1994). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Proba. Math. Stat.
- [43] M.S. MAKLAD and T. NICHOLS (1980). A new approach to model structure discrimination. *IEEE Trans. on Systems, Man and Cybernetics*, vol.SMC-10, no 2, pp.78-84.
- [44] 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).
- [45] R.K. MEHRA (1976). Optimization of measurement schedules and sensor design for linear dynamic systems. *IEEE Trans. on Automatic Control*, vol.AC-21, no 1, pp.55-64.
- [46] P.C. MÜLLER and H.I. WEBER (1972). Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica*, vol.8, pp.237-246.
- [47] C.N. NETT, C.A. JACOBSON and A.T. MILLER (1988). An integrated approach to controls and diagnostics : the 4-parameter controller. *Proc. American Control Conf.*, Atlanta, Georgia, pp.824-835.
- [48] I.V. NIKIFOROV (1995). A generalized change detection problem. *IEEE Trans. on Information Theory*, vol.IT-41, no 1, pp.171-187.
- [49] I.V. NIKIFOROV, M. STAROSWIECKI and B. VOZEL (1996). Duality of analytical redundancy and statistical approach in fault diagnosis. *IFAC'96 World Congress*, San Francisco, CA, June 30-July 4, 1996.
- [50] R.N. NIKOUKHAH (1994). Innovations generation in the presence of unknown inputs : application to robust failure detection. *Automatica*, vol.30, no 12, pp.1851-1868.
- [51] R.J. PATTON, P. FRANK and R. CLARK, eds. (1989). *Fault Diagnosis in Dynamic Systems – Theory and Application*, International Series in Systems and Control Engineering, Prentice Hall International, London, UK.
- [52] R.J. PATTON (1994). Robust model-based fault diagnosis : the state of the art. *Proc. Safeprocess'94*, Helsinki, Finland.

- [53] A. RÉNYI (1959). On measures of dependence. *Acta Mathematica*, vol.10, pp.441-451.
- [54] E. SPJØTVOLL (1972). On the optimality of some multiple comparison procedures. *Annals Statistics*, vol.21, no 3, pp.1486-1521.
- [55] M. STAROSWIECKI, J.P. CASSAR and V. COCQUEMPOT (1993). A general approach for multicriteria optimization of structured residuals. *Proc. Tooldiag'93*, Int. Conf. on Fault Diagnosis, Toulouse, F., April 1993, pp.800-807.
- [56] A.A. STOOORVOGEL and J.H. VAN SCHUPPEN (1995). System identification with information theoretic criteria. *CWI Report BS-R9513*.
- [57] S. TANAKA (1989). Diagnosability of systems and optimal sensor location. In [51].
- [58] S.G. TZAFESTAS, M.G. SINGH and G. SCHMIDT, eds. (1987). *System Fault Diagnosis, Reliability and Related Knowledge-based Approaches* – vol.1. *Fault Diagnostics and Reliability* – vol.2. *Knowledge-based and Fault-tolerant Techniques*. Reidel, Dordrecht, Holland.
- [59] M.H. VAN EMDEN (1971). An analysis of complexity. *Amsterdam Mathematical Center Tracts*, vol.35.
- [60] E. WAHNON, M. BASSEVILLE and A. BENVENISTE (1991). On failure detection and identification : an optimum robust min-max approach. *Proc. Safeprocess'91*, Baden-Baden, FRG., pp.319-324.
- [61] A.S. WILLSKY (1976). A survey of design methods for failure detection in dynamic systems. *Automatica*, vol.12, pp.601-611.

Contents

1	Introduction	3
2	FDI in linear dynamic systems	5
2.1	Reducing off-line FDI to a static problem	5
2.2	Static FDI problem statement	7
2.3	Deterministic approaches	8
3	Basic statistical tools	11
3.1	Statistical detection	12
3.2	Statistical isolation	13
3.2.1	Karhunen-Loeve decomposition	14
3.2.2	Sensitivity tests	14
3.2.3	Statistical rejection	16
3.2.4	Mixed rejection and sensitivity approach	19
3.2.5	Summarizing residuals properties	20
4	Extensions and discussion	21
4.1	Gaussian regression models	21
4.2	Links between deterministic and stochastic approaches	22
4.3	Mixed deterministic/stochastic approaches	23
5	Multiple faults : definition and optimality criteria	24
5.1	Multiple faults with or without causality assumptions	24
5.1.1	Single faults and sensitivity tests	24
5.1.2	Multiple faults and rejection tests	24
5.2	Optimality criteria	25
5.2.1	Optimality of rejection tests	26
5.2.2	Isolation optimality and Fisher information	27
6	Information criteria for residual optimization	28
6.1	Background and motivation	28
6.2	A distance between matrices	29
6.3	Three isolation criteria	30
6.4	Residual optimization	31
6.4.1	Computing the gradients of the criteria	32
6.4.2	Taking the constraints into account	33
6.4.3	Optimizing the dimensions	33
7	Conclusion	34
	References	35



Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité 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)
ISSN 0249-6399