Robust Fault Detection With Statistical Uncertainty in Identified Parameters

Jianfei Dong, Michel Verhaegen and Fredrik Gustafsson

Linköping University Post Print

N.B.: When citing this work, cite the original article.

©2012 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Jianfei Dong, Michel Verhaegen and Fredrik Gustafsson, Robust Fault Detection With Statistical Uncertainty in Identified Parameters, 2012, IEEE Transactions on Signal Processing, (60), 10, 5064-5076. http://dx.doi.org/10.1109/TSP.2012.2208638

Postprint available at: Linköping University Electronic Press http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-84886

Robust Fault Detection with Statistical Uncertainty in Identified Parameters

Jianfei Dong, Michel Verhaegen, and Fredrik Gustafsson

Abstract

Detection of faults that appear as additive unknown input signals to an unknown LTI discrete-time MIMO system is considered. State of the art methods consist of the following steps. First, either the state space model or certain projection matrices are identified from data. Then, a residual generator is formed based on these identified matrices, and this residual generator is used for online fault detection. Existing techniques do not allow for compensating for the identification uncertainty in the fault detection. This contribution explores a recent data-driven approach to fault detection. We show first that the identified parametric matrices in this method depend linearly on the noise contained in the identification data, and then that the on-line computed residual also depends linearly on the noise. This allows an analytic design of a robust fault detection scheme, that takes both the noise in the online measurements as well as the identification uncertainty into account. We illustrate the benefits of the new method on a model of aircraft dynamics extensively studied in literature.

Index Terms

Fault detection; Parameter uncertainty; Statistical analysis; Additive faults; Closed-form solution.

I. INTRODUCTION

Model-based fault detection and isolation (FDI) is a well-established technique in literature, see [1]–[4]. The main task for the practitioner is to derive the model, that is generally assumed to be given in linear state space form represented by the matrices (A, B, C, D). Then the algorithms are easily implemented; and given the correctness of the model, the tests satisfy various optimality criteria. However, these models

Corresponding author J. Dong, email: jfeidong@hotmail.com. J. Dong was with Delft University of Technology when this work was carried out, and was with Philips Research, 5656 AE, Eindhoven, The Netherlands. M. Verhaegen is with Delft University of Technology, 2628CD, Delft, the Netherlands. F. Gustafsson is with Department of Electrical Engineering, Linköpings Universitet, SE-581 83 Linköping, Sweden.

might be quite complex [5], [6]; and there might be uncertain parameters in the model which have to be estimated. But the stochastic uncertainty in the estimated parameters is not taken care of in the approaches described in the references above.

Model-free, or data-driven, FDI methods have been investigated in various tracks. One track, e.g. [7], is via testing the changes in the statistics of signals, which are assumed to be stationary in fault-free case. Another track, e.g. [8], [9], aims at overcoming the modeling stage by using system identification techniques, e.g. prediction error methods (PEM) [10] and subspace identification (SID) methods [11]. Taking SID as an example, two steps need to be taken. First, either the range space of the extended observability matrix [11] or the unknown state sequence [12], [13] is identified. Then, the state space matrices, e.g. (A,B,C,D), need to be estimated using standard realization theory. In fact, the recent literature shows that the state space matrices, (A,B,C,D), are not really needed in designing an FDI, e.g. [8], [9]. Based on SID [11], [13], these approaches require computing the left null of the extended observability matrix, directly identified from measurement data, without realizing (A,B,C,D). The projection of a residual vector onto the left null space aims at annihilating the influence of unknown initial states on the residual, which is known as parity space analysis (PSA) [14]–[17].

One of the drawbacks of computing the parity relation from data is that the residuals hence generated are sensitive to the approximation errors attributed to the model reduction step in identifying the range space of the extended observability matrix [18]. We have in [19], [20] recently developed a data-driven *FdI* scheme Connected to Subspace Identification (FICSI), whose parameters can be directly identified from data, with neither realizing (A, B, C, D), nor computing the range space of the extended observability matrix together with its left null space. A key step therein is to replace in the parity relation the product between the extended observability matrix and the initial states with a product of a matrix that depends in an affine manner on the least squares estimated model parameters and a matrix constructed from past input and output (I/O) signals, subject to a bias error vanishing exponentially with past horizon. In this way the annihilation of this product does not require a projection as in PSA methods, that very much complicates the statistical analysis of the residual.

On the other hand, despite the large amount of existing work on FDI methods, the robustness of a detection scheme against model identification errors has not yet been investigated. The major challenge lies in quantifying the uncertainties in the residual generator from the parameter identification errors. As proposed in [21], an empirical estimate of the PSA residual covariance may be computed by the identification data from a fault-free system. But an analysis of the residual distribution subject to model errors is still an open problem in the literature. As pointed out in [18], the difficulties in analyzing the

parameter error effect on a data-driven PSA residual include the nonlinear dependence of the identified projection matrix on the parameter errors and the multiplication of the erroneous residual generator with this matrix. As to be shown later in this paper, the main advantage of the FICSI formulation [19] is that the uncertainties in the residual generator linearly depend on the parameter errors. It is hence possible to develop a robust data-driven fault detection scheme coping with the identification errors in a closed-form solution. It shall be mentioned here that approaches coping with stochastic parameter errors are also seen in filtering methods, e.g. [22], [23], which are called *cautious* filtering, due to the penalties imposed to the covariance of the parameter errors. Although [22] gave a closed-form solution to cautious Wiener filters, this solution was based on the assumption that parameter errors only exist in the numerator polynomials. Instead of providing closed-form solutions, the approach in [23] relies on particle filters (see e.g. [24]) for linear and nonlinear systems.

The rest of the paper is organized as follows. We start in Sec. II with describing the Vector ARX model for linear dynamic systems and linking this model to residual generation for fault detection. The error effects on the residual generator using the identified parameters is then analyzed in Sec. III. The statistical distribution of this residual vector is also quantified in this section, which then leads to a statistically optimal fault detection test. Sec. IV verifies all the main arguments via simulation studies.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Notations

The notations used in this paper are standard. $\mathbb{R}^{h \times q}$ denotes the set of all real *h*-by-*q* matrices. *I*,0 will respectively denote an identity and a zero matrix with proper dimensions; while with subscripts, $I_h, 0_{h \times q}$ shall denote respectively an $h \times h$ identity matrix and an $h \times q$ zero matrix. $\mathcal{N}(\mu, \Sigma)$ represents a normal distribution with mean μ and covariance Σ . χ_h^2 stands for a χ^2 distribution with *h* degrees of freedom. \mathbb{E} and Cov denote respectively expectation and covariance operator. vec(M) is the column vector concatenating the columns of a matrix M. M^{\dagger} stands for the pseudo-inverse of M. The operator " \otimes " shall denote Kronecker product. For two matrices M_1 and M_2 , $M_1 \succeq M_2$ means that $M_1 - M_2$ is positive semi-definite.

B. Additive faults in linear Gaussian systems

Faults that can occur in a dynamic system are categorized into two classes [4], [25], [26]. The first class of faults is called additive, which are signals additive on the model of a dynamic system, and hence change the mean value of the distribution of the observed signals [25, Chapter 7]. For instance, drifts



Fig. 1. Fault detection scheme, where r represents residual.

and bias in sensor measurements and leakage in pipelines are typical additive faults [9]. Another class is called multiplicative, which change the transfer function of a dynamic system [25, Chapters 8,9]. A typical example is the changes in the vibrating characteristics of a mechanical structure [25, Chapter 11]. This paper especially considers additive faults, and assumes that the system transfer function is unchanged, which can hence be identified offline using I/O signals from the nominal system.

We shall describe the faulty system dynamics with the following discrete-time state-space model [25, Chapter 7]:

$$x(k+1) = Ax(k) + Bu(k) + Ef(k) + Fw(k),$$
(1)

$$y(k) = Cx(k) + Du(k) + Gf(k) + v(k).$$
 (2)

We consider MIMO systems; i.e. $x(k) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^\ell$, and $u(k) \in \mathbb{R}^m$. $f(k) \in \mathbb{R}^{n_f}$ represents fault signals. *A*,*B*,*C*,*D*,*E*,

F,*G* are real, with bounded norms and appropriate dimensions. The disturbances are represented by the process noise $w(k) \in \mathbb{R}^{n_w}$ and the measurement noise $v(k) \in \mathbb{R}^{\ell}$. w(k) and v(k) are assumed to be white zero-mean Gaussian [25, Chapter 7]. The following assumptions are standard in Kalman filtering [10], [27], [28] and subspace identification [12], [13], [29].

Assumption 1: The pair (C,A) is detectable; and there are no uncontrollable modes of $(A, FQ_w^{1/2})$ on the unit circle, where $Q_w^{1/2} \cdot (Q_w^{1/2})^T$ is the covariance matrix of w(k).

Fault detection can be regarded as a residual generation and evaluation problem. A residual is the deviation between measurements and model-equation based computations [30]. A fault detection scheme can be illustrated as in Fig. 1. The essential goals of this paper are to develop the output estimator based on least-squares (LS) estimates of system parameters, and moreover to robustify the residual evaluator against the stochastic errors in the LS estimates of the parameters. To this end, we will first review the parameter identification problem and the residual generation problem in this section.

C. Vector ARX description of linear dynamic systems and parameter identification

Since both parameter identification and residual generation are based on a nominal system model, we will first set f(k) = 0 in (1,2), and consider the following model:

$$x(k+1) = Ax(k) + Bu(k) + Fw(k),$$
 (3)

$$y(k) = Cx(k) + Du(k) + v(k).$$
 (4)

In system identification literature [10]–[13], the I/O relationship of the model (3,4) is usually reformulated by the following innovation form [27], [28], with the innovation signal defined as $e(k) = y(k) - C\hat{x}(k) - Du(k)$.

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + Ke(k),$$
 (5)

$$y(k) = C\hat{x}(k) + Du(k) + e(k).$$
 (6)

Here, *K* is the Kalman gain. The innovation e(k) is white Gaussian with a covariance matrix denoted by Σ_e , as determined by w(k) and v(k), [10], [11], [13].

A closed-loop observer thus results from (5,6); i.e.

$$\hat{x}(k+1) = (A - KC)\hat{x}(k) + (B - KD)u(k) + Ky(k),$$
(7)

$$y(k) = C\hat{x}(k) + Du(k) + e(k).$$
 (8)

For brevity, denote $\Phi \triangleq A - KC$, $\tilde{B} \triangleq B - KD$ in what follows. Under Assumption 1, Φ is stable.

For the well-posedness of the problem, we assume that the plant (3,4) is internally stable, with or without closed-loop stabilizing control. Under this assumption, x(k), u(k), y(k) are bounded for any k. Due to the stability of Φ , $\mathbb{E}\hat{x}(k)$ is bounded for any k (see e.g. [28, Theorem 9.2.2]).

Starting from the time instant k - p and solving by recursion for p sampling instants till the time instant k, it is easy to derive:

$$\hat{x}(k) = \Phi^{p}\hat{x}(k-p) + \sum_{\tau=0}^{p-1} \Phi^{\tau} \cdot \begin{bmatrix} \tilde{B} & K \end{bmatrix} \cdot \begin{bmatrix} u(k-\tau-1) \\ y(k-\tau-1) \end{bmatrix}.$$
(9)

The output equation can hence be written in the following form, by substituting (9) into (8):

$$y(k) = C\Phi^{p}\hat{x}(k-p) + \sum_{\tau=0}^{p-1} C\Phi^{\tau} \cdot \begin{bmatrix} \tilde{B} & K \end{bmatrix} \cdot \begin{bmatrix} u(k-\tau-1) \\ y(k-\tau-1) \end{bmatrix} + \begin{bmatrix} D & 0 \end{bmatrix} \cdot \begin{bmatrix} u(k) \\ y(k) \end{bmatrix} + e(k).$$
(10)

This equation is known as a vector autoregressive model with exogenous inputs (Vector ARX or VARX), which is generally used as a first step in subspace identification [12], [31], [32].

The output equation (10) establishes a linear transformation from past I/Os to outputs. We have shown in our previous work that this equation can be used in developing predictive controllers [33] and identifying estimation filters for additive actuator and sensor faults [20]. There are two main benefits in using this VARX description. First, the parameters therein can be consistently identified from data measured in closed-loop plants by solving a single least-squares problem, and hence do not contain model reduction errors in realizing the state-space matrices (A, B, C, D). Second, and more importantly, this output equation has a linear dependence on the parameter identification errors, and hence enables analyzing and developing FDI solutions robust to the stochastic parameter identification errors. It is hence the objective of this current paper to link this output equation to fault detection, where the parameters are corrupted with stochastic identification errors.

We emphasize that in the output equation (10), the parameters that need to be identified are $\{D, C\Phi^{j}\tilde{B}, C\Phi^{j}K, j = 0, \dots, p-1\}$, instead of (A, B, C, D) or $C\Phi^{p}$.

Replace the time index k in (10) respectively by a sequence of time indices $t, t + 1, \dots, t + N - 1$. Collect $y(t), y(t+1), \dots, y(t+N-1)$ into a block row vector, and denote it by Y_{id} ; i.e.

$$\mathbf{Y}_{id} = \left[\begin{array}{ccc} y(t) & y(t+1) & \cdots & y(t+N-1) \end{array} \right].$$

Here, the subscripts "*id*" indicate that Y_{id} contains the output signals collected from an identification experiment, and will be used to identify model parameters. The following data equation thus results:

$$\begin{bmatrix} y(t) & y(t+1) & \cdots & y(t+N-1) \end{bmatrix} = C\Phi^{p} \cdot \underbrace{\left[\hat{x}(t-p) & \hat{x}(t-p+1) & \cdots & \hat{x}(t+N-p-1) \right]}_{X_{id}} + \underbrace{\left[C\Phi^{p-1}\tilde{B} & C\Phi^{p-1}K & \cdots & C\tilde{B} & CK + D \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} u(t-p) & u(t-p+1) & \cdots & u(t+N-p-1) \\ y(t-p) & y(t-p+1) & \cdots & y(t+N-p-1) \\ \vdots & \vdots & & \vdots \\ u(t-1) & u(t) & \cdots & u(t+N-2) \\ y(t-1) & y(t) & \cdots & y(t+N-2) \\ u(t) & u(t+1) & \cdots & u(t+N-1) \\ \end{array} \right]}_{Z_{id}} + \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & \cdots & e(t+N-1) \\ E_{id} & e(t+1) & \cdots & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}{cccc} e(t) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}[c] e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}[c] e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+1) & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}[c] e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}[c] e(t) & e(t+1) & e(t+N-1) \\ E_{id} & e(t+N-1) \\ E_{id} & e(t+N-1) \end{array} \right]}_{E_{id}} \cdot \underbrace{\left[\begin{array}[c] e(t) & e(t+N-1) \\ E_{id} & e($$

With the notations of the data matrices as defined in the equation above, (11) can be further written in

a compact form:

$$\mathbf{Y}_{id} = C\Phi^{p} \cdot \mathbf{X}_{id} + \begin{bmatrix} C\Phi^{p-1}\tilde{B} & C\Phi^{p-1}K & \cdots & C\tilde{B} & CK & D \end{bmatrix} \cdot \mathbf{Z}_{id} + \mathbf{E}_{id}.$$
 (12)

For brevity, we shall denote the sequence of Markov parameters in (12) as

$$\Xi = \begin{bmatrix} C\Phi^{p-1}\tilde{B} & C\Phi^{p-1}K & \cdots & C\tilde{B} & CK & D \end{bmatrix}.$$

From (12), the estimation of Ξ can be formulated in a least-squares sense as:

$$\hat{\boldsymbol{\Xi}} \triangleq \arg\min_{\boldsymbol{\Xi}} \|\boldsymbol{Y}_{id} - \boldsymbol{\Xi} \cdot \boldsymbol{Z}_{id}\|_2^2.$$
(13)

It has been proven in subspace identification literature that the state-space model (5,6) can be consistently identified from data measured in closed-loop plants via first solving the LS problem (13), e.g. [12], [34].

As a standard assumption of persistent excitation in system identification (see e.g. [10]), the data matrix Z_{id} is bounded and has full row rank; i.e. $\exists \underline{\rho}_{z,id} > 0$ such that

$$\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^{T} \succeq \underline{\boldsymbol{\rho}}_{z,id}^{2} \boldsymbol{I}.$$
⁽¹⁴⁾

Besides, we will use the following explicit bound on the covariance matrix of the unknown state sequence $vec(\mathbf{X}_{id})$ in the analysis; i.e. $\exists \bar{\sigma}_{xx} > 0$ such that

$$\|\operatorname{Cov}(\operatorname{vec}(\boldsymbol{X}_{id}))\|_2 \le \bar{\sigma}_{xx}.$$
(15)

If the data matrix Z_{id} has full row rank, then the LS problem (13) has a solution with the following structure,

$$oldsymbol{Y}_{id}\cdotoldsymbol{Z}_{id}^{\dagger}=C\Phi^{p}\cdotoldsymbol{X}_{id}\cdotoldsymbol{Z}_{id}^{\dagger}+\Xi+oldsymbol{E}_{id}\cdotoldsymbol{Z}_{id}^{\dagger},$$

which contains the parameter estimate,

$$\hat{\Xi} = Y_{id} \cdot Z_{id}^{\dagger}, \tag{16}$$

and the errors,

$$\Delta \hat{\Xi} \triangleq \Xi - \hat{\Xi} = C \Phi^p \boldsymbol{X}_{id} \cdot \boldsymbol{Z}_{id}^{\dagger} + \boldsymbol{E}_{id} \cdot \boldsymbol{Z}_{id}^{\dagger}.$$
⁽¹⁷⁾

The minus signs on the right hand side of (17) are absorbed into the unknown random variables X_{id} , E_{id} for simplicity.

Treating Σ_e as known, or using the estimate,

$$\hat{\Sigma}_{e} = \operatorname{Cov}\left(\boldsymbol{Y}_{id} - \hat{\Xi} \cdot \boldsymbol{Z}_{id}\right), \tag{18}$$

in its place, is a standard practice in the statistical signal detection literature [3], [35]. We shall hence not distinguish between $\hat{\Sigma}_e$ and Σ_e in the rest of the paper.

D. Output estimator based on the VARX model

We can hence link the VARX model (10) to fault detection, by employing (10) to compute output signals. In fault detection literature, residuals are often generated in a sliding window of more than one sampling instants; i.e. $[k - L + 1, \dots, k]$ up to the current time instant k. We shall hence refer to $[k - L + 1, \dots, k]$ as the detection window, and call L the detection horizon.

Similar to deriving (12), replace the time index k in (10) respectively by the sequence of L time indices, $k-L+1, \dots, k$. Collect $y(k-L+1), \dots, y(k)$ into a column vector, and denote it by $y_{k,L}$; i.e.

$$\boldsymbol{y}_{k,L} = \left[\begin{array}{ccc} \boldsymbol{y}^T(k-L+1) & \boldsymbol{y}^T(k-L+2) & \cdots & \boldsymbol{y}^T(k) \end{array} \right]^T.$$

Similarly, denote the lumped input vector and lumped innovation vector along the detection window respectively by $u_{k,L}$ and $e_{k,L}$. Then, the following lumped output equation follows:

$$\begin{bmatrix} y(k-L+1) \\ y(k-L+2) \\ \vdots \\ y(k) \end{bmatrix} = \begin{bmatrix} C\Phi^{p} \cdot \hat{x}(k-L-p+1) \\ C\Phi^{p} \cdot \hat{x}(k-L-p+2) \\ \vdots \\ C\Phi^{p} \cdot \hat{x}(k-p) \end{bmatrix} + \begin{bmatrix} C\Phi^{p-1}\tilde{B} & C\Phi^{p-1}K & \cdots & C\Phi\tilde{B} & C\Phi K & C\tilde{B} & CK \\ 0 & 0 & C\Phi^{p-1}\tilde{B} & C\Phi^{p-1}K & \cdots & C\Phi\tilde{B} & C\Phi K \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & C\Phi^{p-1}\tilde{B} & \cdots & \cdots & C\Phi^{L-1}K \end{bmatrix}$$

$$\frac{D}{k_{k,L}} = \begin{bmatrix} u(k-L-p+1) \\ y(k-L-p+1) \\ \vdots \\ u(k-L) \\ y(k-L-1) \\ \vdots \\ u(k-L+1) \\ y(k-L+1) \\ \vdots \\ u(k) \\ y(k) \end{bmatrix} + \begin{bmatrix} e(k-L+1) \\ e(k-L+2) \\ \vdots \\ e(k) \end{bmatrix}.$$
(19)

Note that the big matrix containing the Markov parameters has a structure, with the block to the left of the vertical line representing a block Hankel matrix, and the block to its right showing a block lower triangular Toeplitz structure. Besides, the Hankel matrix corresponds to a time window with p sampling instants, i.e. [k - L - p + 1, k - L]; while the Toeplitz matrix coincides with the detection window [k - L + 1, k]. We will therefore refer to [k - L - p + 1, k - L] as the past window, and call p the past horizon.

9

For clarity, we shall explicitly denote the block Hankel matrix as

$$H_{z}^{L,p} = \begin{bmatrix} C\Phi^{p-1}\tilde{B} & C\Phi^{p-1}K & \cdots & \cdots & C\tilde{B} & CK \\ 0 & 0 & C\Phi^{p-1}\tilde{B} & C\Phi^{p-1}K & \cdots & C\Phi\tilde{B} & C\Phi K \\ \vdots & \vdots & & \vdots & \\ 0 & \cdots & 0 & C\Phi^{p-1}\tilde{B} & \cdots & \cdots & C\Phi^{L-1}K \end{bmatrix},$$
(20)

where the superscripts "L, p" remind respectively the detection horizon L (number of block rows) and the past horizon p (number of block columns). Similarly, define the following two Toeplitz matrices, respectively corresponding to the inputs and outputs in the detection window:

$$\boldsymbol{T}_{u}^{L} = \begin{bmatrix} D & & & \\ C\tilde{B} & D & & \\ \vdots & \vdots & \ddots & \\ C\Phi^{L-2}\tilde{B} & C\Phi^{L-3}\tilde{B} & \cdots & D \end{bmatrix}, \boldsymbol{T}_{y}^{L} = \begin{bmatrix} 0 & & & \\ CK & 0 & & \\ \vdots & \vdots & \ddots & \\ C\Phi^{L-2}K & C\Phi^{L-3}K & \cdots & 0 \end{bmatrix},$$
(21)

where the superscript "L" reminds that there are L block columns and rows.

It is convenient to introduce a new notation to denote the lumped I/Os in the past window, i.e. $z(k) = [u^T(k), y^T(k)]^T$, and collect the lumped I/Os in the past window into $z_{k-L,p}$, i.e.

$$z_{k-L,p} = \begin{bmatrix} z^T(k-L-p+1) & z^T(k-L-p+2) & \cdots & z^T(k-L) \end{bmatrix}^T$$

Now, the lumped L-step output equation (19) can be expressed in a compact form as

$$\boldsymbol{y}_{k,L} = \boldsymbol{b}_{k,L} + \boldsymbol{H}_{z}^{L,p} \boldsymbol{z}_{k-L,p} + \boldsymbol{T}_{u}^{L} \boldsymbol{u}_{k,L} + \boldsymbol{T}_{y}^{L} \boldsymbol{y}_{k,L} + \boldsymbol{e}_{k,L}.$$
(22)

Remark 1: It is tedious but straightforward to show that (22) is equivalent to the output equation in classical parity space approaches [14]–[17], if it is also derived from the closed-loop observer form (7,8); i.e.

$$\boldsymbol{y}_{k,L} = \boldsymbol{O}^{L} \cdot \hat{\boldsymbol{x}}(k-L+1) + \boldsymbol{T}_{u}^{L} \boldsymbol{u}_{k,L} + \boldsymbol{T}_{y}^{L} \boldsymbol{y}_{k,L} + \boldsymbol{e}_{k,L},$$
(23)

where O^L denotes the extended observability matrix:

$$\left[\begin{array}{ccc} C^T & (C\Phi)^T & \cdots & (C\Phi^{L-1})^T \end{array}\right]^T.$$

The benefit of using Eq. (22) is that all the parameters in the matrices $H_z^{L,p}, T_u^L, T_y^L$ only depend on the sequence of Markov parameters Ξ , which can be consistently identified from data by solving the single LS problem (16). Besides, instead of computing the left null space of O^L to annihilate the unknown product $O^L \cdot \hat{x}(k-L+1)$, as in the classical PSA methods [14]–[17], this product is replaced by $b_{k,L} + H_z^{L,p} z_{k-L,p}$ in (22), where the bias term $b_{k,L}$ is negligible if p is large enough.

E. Residual generator and its distribution

A residual generator for fault detection is a direct consequence of comparing the measured and computed outputs in the sliding window [k - L + 1, k]:

$$\boldsymbol{r}_{k,L} = (\boldsymbol{I} - \boldsymbol{T}_{y}^{L})\boldsymbol{y}_{k,L} - \boldsymbol{H}_{z}^{L,p}\boldsymbol{z}_{k-L,p} - \boldsymbol{T}_{u}^{L}\boldsymbol{u}_{k,L}.$$
(24)

In the fault free case, the residual has two components, i.e.

$$\boldsymbol{r}_{k,L} = \boldsymbol{b}_{k,L} + \boldsymbol{e}_{k,L}. \tag{25}$$

Since the innovation signals *e* are white, the covariance matrix of $e_{k,L}$ can be written as $\Sigma_e^L = I_L \otimes \Sigma_e$.

In the presence of additive faults, the residual is still computed by (24), but the unknown fault signals will contribute an additive term to (25). To see this, we shall turn to Eqs. (1,2). Similarly, define the innovation signal as $e(k) = y(k) - C\hat{x}(k) - Du(k) - Gf(k)$. The following closed-loop observer form results.

$$\hat{x}(k+1) = \Phi \hat{x}(k) + Bu(k) + (E - KG)f(k) + Ky(k),$$
(26)

$$y(k) = C\hat{x}(k) + Du(k) + Gf(k) + e(k).$$
 (27)

Remark 2: The additive fault signals, f(k), can be considered as extra external inputs. We shall assume f(k) to be deterministic but unknown, which only change the mean of the residual (31), instead of its covariance. Therefore, following standard Kalman filtering theory, e.g. [27, Chapter 7] and [28, Chapter 9], the innovation signals, e(k), are white. Besides, $\mathbb{E}(e(k)|\text{past I/Os, faults}) = 0$; and Cov(e(k)|past I/Os, faults) is only determined by w(k) and v(k), [27], [28].

For brevity, denote $\tilde{E} = E - KG$. Now, similar to the derivation of (19), (26,27) lead to the following lumped output equation with faults:

$$\boldsymbol{y}_{k,L} = \boldsymbol{b}_{k,L} + \boldsymbol{H}_{z}^{L,p} \boldsymbol{z}_{k-L,p} + \boldsymbol{T}_{u}^{L} \boldsymbol{u}_{k,L} + \boldsymbol{T}_{y}^{L} \boldsymbol{y}_{k,L} + \boldsymbol{\varphi}_{f} + \boldsymbol{e}_{k,L},$$
(28)

where $\varphi_f = \left[\boldsymbol{H}_f^{L,p} \boldsymbol{T}_f^L \right] \cdot \boldsymbol{f}_{k,p+L}$, with $\boldsymbol{f}_{k,p+L} = \left[f^T (k - L - p + 1), \cdots, f^T (k) \right]^T$. The matrix, $\left[\boldsymbol{H}_f^{L,p} \mid \boldsymbol{T}_f^L \right]$, explicitly reads as

$$\begin{bmatrix} C\Phi^{p-1}\tilde{E} & \cdots & C\Phi\tilde{E} & C\tilde{E} & G \\ 0 & C\Phi^{p-1}\tilde{E} & \cdots & C\Phi\tilde{E} & C\tilde{E} & G \\ \vdots & & \ddots & & \vdots & \ddots & \ddots \\ 0 & \cdots & C\Phi^{p-1}\tilde{E} & \cdots & C\Phi^{L-2}\tilde{E} & C\Phi^{L-3}\tilde{E} & \cdots & G \end{bmatrix}$$

_

Thus, in the presence of additive faults, the residual contains the following three components:

$$\boldsymbol{r}_{k,L} = \boldsymbol{\varphi}_f + \boldsymbol{b}_{k,L} + \boldsymbol{e}_{k,L}. \tag{29}$$

Here, φ_f changes the mean of $r_{k,L}$, as long as $f_{k,p+L} \notin \operatorname{Ker}\left(\left[H_f^{L,p} T_f^L\right]\right)$. Moveover, as will be further detailed in the companion paper [36], $\left[H_f^{L,p} T_f^L\right]$ is needed in computing the projection directions to isolate individual fault input channels.

Under Assumption 1 and noting that the innovation signals e(k) are white Gaussian based on standard Kalman filter theories [10], [27], [28], the distribution of $r_{k,L}$ belongs to the following parametric family [3]:

$$\boldsymbol{r}_{k,L} \sim \begin{cases} \mathcal{N}(\mathbb{E}\boldsymbol{b}_{k,L},\boldsymbol{\Sigma}_{e}^{L}), & \text{fault free}, \\ \mathcal{N}\left(\mathbb{E}\boldsymbol{b}_{k,L} + \boldsymbol{\varphi}_{f}, \boldsymbol{\Sigma}_{e}^{L}\right), & \text{faulty.} \end{cases}$$
(30)

More details of the bias effect on the residual distribution will be analyzed in Sec. III.

If $L \leq p$, then Σ_e and the identified parameters $\hat{\Xi}$ from (16) can fully parameterize (24). We shall denote the Hankel and Toeplitz parametric matrices comprising $\hat{\Xi}$ by $\bar{H}_z^{L,p}, \bar{T}_u^L, \bar{T}_y^L$, and rewrite the residual generator built by the identified parameters as

$$\boldsymbol{r}_{k,L} = (\boldsymbol{I} - \bar{\boldsymbol{T}}_{y}^{L})\boldsymbol{y}_{k,L} - \bar{\boldsymbol{H}}_{z}^{L,p}\boldsymbol{z}_{k-L,p} - \bar{\boldsymbol{T}}_{u}^{L}\boldsymbol{u}_{k,L}.$$
(31)

Since the residual is a linear function of the estimated matrices, the residual again becomes Gaussian with an additional covariance matrix denoted by $\Sigma_{\Lambda \hat{=}}^{L}$; i.e.

$$\boldsymbol{r}_{k,L} \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}_{k,L}, \boldsymbol{\Sigma}_{e}^{L} + \boldsymbol{\Sigma}_{\Delta \hat{\Xi}}^{L}), & \text{fault free,} \\ \mathcal{N}\left(\boldsymbol{\mu}_{k,L} + \boldsymbol{\varphi}_{\boldsymbol{f}}, \boldsymbol{\Sigma}_{e}^{L} + \boldsymbol{\Sigma}_{\Delta \hat{\Xi}}^{L}\right), & \text{faulty.} \end{cases}$$
(32)

Here, $\mu_{k,L} = \mathbb{E} r_{k,L}$, when no additive faults are present, and it will be analyzed later.

Remark 3: In comparison, the data-driven PSA methods proposed in [8], [9] require first identifying the range of O^L from identification data, and then computing its left null space by SVD. Here, the first step is prone to stochastic parameter identification errors; while the second one contains model reduction errors. Quantifying the statistical distribution of this computed left null space would be a difficult task. In fact, even if such a distribution can be obtained, it is still needed to project the computed $y_{k,L}$ by (23) with erroneous parameters onto this left null space. This procedure further complicates the analysis of the statistical distribution of the projected output vector. The benefit of the residual defined in (31) is that it avoids an error analysis of the SVD and the products of statistical variables.

We are now ready to formulate the main problem to be solved in this paper.

Problem 1: What is the additional covariance $\Sigma_{\Delta \hat{\Xi}}^{L}$ in (32) by using the uncertain data model in (31), and is this distribution significantly different from the one in (30)? What is then the optimal test for the residual in (31)?

III. ROBUST DATA-DRIVEN FAULT DETECTION

With the assumption that the lumped I/O z(k) is a quasi-stationary process (see e.g. [10, Chapter 2]), the asymptotic analyses in system identification methods [10], [12] have established the following facts:

- $\lim_{p\to\infty} \mathbb{E}\hat{\Xi} = \Xi;$
- lim_{N,p→∞} Cov(vec(ΔΞ̂)|Z_{id}) = lim_{N→∞} (¹/_N · R⁻¹_{zz} ⊗ Σ_e) = 0, where R_{zz} denotes the correlation matrix of z(k), and has bounded inverse due to the quasi-stationarity of z(k).

In other words, with $N, p \to \infty$, $\hat{\Xi} = \Xi$ holds exactly. Therefore, $\lim_{N,p\to\infty} \text{Cov}(r_{k,L}) = I_L \otimes \Sigma_e$; i.e. the covariance of the data-driven residual vector $r_{k,L}$ is only determined by the innovation vector $e_{k,L}$.

However, in practice, the duration of an identification experiment cannot be infinite; and the past horizon p cannot be too long for the computation of the residual vector to be practical. But when N and p are finite, $\hat{\Xi}$ is both biased and stochastic, as can be seen in (17). The error effects on the residual distribution and fault detection are hence the objectives of this section.

A. Dependence of the residual vector on parameter identification errors

Since in (29) the fault-dependent term φ_f only changes the residual mean, instead of its covariance, we shall take $\varphi_f = 0$ in the analysis of the residual covariance, without loss of generality. The preliminary step before deriving the additional covariance $\Sigma_{\Delta \hat{\Xi}}^L$ in (32) is to explicitly write the uncertainties in $r_{k,L}$ in terms of the stochastic errors, $\Delta \hat{\Xi}$.

Lemma 1: If the Markov parameters contained in \bar{T}_u^L , \bar{T}_y^L , and $\bar{H}_z^{L,p}$ are identified by (16) from finitely many I/O samples, then the residual vector (31) computed at time instant k has the following structure:

$$\boldsymbol{r}_{k,L} = \boldsymbol{\zeta}_{k,L} + \boldsymbol{e}_{k,L} + \boldsymbol{\beta}_{k,L} + \boldsymbol{b}_{k,L}. \tag{33}$$

Here, $\zeta_{k,L}$ and $\beta_{k,L}$ are defined as:

$$\boldsymbol{\zeta}_{k,L} = \left(\boldsymbol{Z}_{ol}^T \otimes \boldsymbol{I}_{\ell} \right) \cdot \left(\boldsymbol{Z}_{id}^{\dagger,T} \otimes \boldsymbol{I}_{\ell} \right) \cdot \operatorname{vec}(\boldsymbol{E}_{id}), \tag{34}$$

$$\boldsymbol{\beta}_{k,L} = \left(\boldsymbol{Z}_{ol}^T \otimes I_{\ell} \right) \cdot \left[\boldsymbol{Z}_{id}^{\dagger,T} \otimes (C \Phi^p) \right] \cdot \operatorname{vec}(\boldsymbol{X}_{id}), \tag{35}$$

with the data matrix Z_{ol} given by

$$Z_{ol} = \begin{bmatrix} z_{k-L,p} & z_{k-L+1,p} & \cdots & z_{k-1,p} \\ \hline u(k-L+1) & u(k-L+2) & \cdots & u(k) \end{bmatrix}$$

$$\in \mathbb{R}^{(p(m+\ell)+m)\times L}.$$

Proof: See Appendix B.

Here, the subscripts, "*ol*", remind that the elements of Z_{ol} are the I/Os measured online. Besides, for the convenience of analyzing $\text{Cov}(r_{k,L})$, $\zeta_{k,L}$ and $\beta_{k,L}$ are expressed in terms of respectively $\text{vec}(E_{id})$ and $\text{vec}(X_{id})$, instead of the matrices $E_{id} \in \mathbb{R}^{\ell \times N}$ and $X_{id} \in \mathbb{R}^{n \times N}$.

Note that the bias terms $\beta_{k,L}$ and $b_{k,L}$ are unknown, and stochastic by the definition of Kalman filter states [27], [28]. For the boundedness of $\mathbb{E}b_{k,L}$ during the implementation of the fault detection scheme, we assume that both the nominal system and the system under the influence of additive faults are internally stable; i.e. $\exists \infty > \bar{p}_{z,ol} > 0$ such that

$$\|\boldsymbol{Z}_{ol}\|_2 \le \bar{\boldsymbol{\rho}}_{z,ol}, \forall k. \tag{36}$$

Under this assumption, the state sequence,

$$\hat{\boldsymbol{x}}_{k-p,L} = \left[\hat{\boldsymbol{x}}^T (k-p-L+1) \cdots \hat{\boldsymbol{x}}^T (k-p) \right]^T,$$

of the observer (7,8) has a bounded covariance, according to the standard Kalman filter theory [27], [28]. For the analysis to follow, we shall assume that

$$\|\operatorname{Cov}(\hat{\boldsymbol{x}}_{k-p,L})\|_2 \leq \bar{\sigma}_{xx}, \forall k, \tag{37}$$

where for simplicity but without loss of generality, the same bound $\bar{\sigma}_{xx}$ as in (15) is used.

B. Analysis of the residual covariance

We shall now analyze and solve Problem 1 based on the residual structure (33). The key ideas are to quantify the composite covariance matrix $\Sigma_e^L + \Sigma_{\Delta \hat{\Xi}}^L$ in (32) with regard to the stochastic term $\zeta_{k,L} + e_{k,L}$, and to derive an error bound in approximating $\text{Cov}(\mathbf{r}_{k,L})$ by its computable components, $\Sigma_e^L + \Sigma_{\Delta \hat{\Xi}}^L$. We shall also show that this error bound exponentially decays as the past horizon p increases.

Since the covariance of both $vec(E_{id})$ and $e_{k,L}$ are known, $Cov(\zeta_{k,L} + e_{k,L})$ can be explicitly quantified as follows. First, note that $vec(E_{id})$ is due to the noise in the identification data, prior to the implementation of the detection algorithm. The two white noise sequences, $e_{k,L}$ and $vec(E_{id})$, are independent. Besides, the elements in Z_{id} , Z_{ol} are all measured quantities, respectively during identification experiment and online implementation. Hence, $vec(E_{id})$ is independent of Z_{ol} ; and $e_{k,L}$ is independent of Z_{id} . Therefore,

$$\operatorname{Cov}(\boldsymbol{\zeta}_{k,L} + \boldsymbol{e}_{k,L}) = \operatorname{Cov}(\boldsymbol{\zeta}_{k,L}) + \operatorname{Cov}(\boldsymbol{e}_{k,L}) \triangleq \boldsymbol{\Sigma}_{\Delta \hat{\Xi}}^{L} + \boldsymbol{\Sigma}_{e}^{L} \triangleq \boldsymbol{\Sigma}_{\Delta \hat{\Xi},e}^{L}$$

Based on (34) in Lemma 1 and Property (47) in Lemma 2 in the appendix, $\Sigma_{\Delta \hat{\Xi}}^{L}$ is derived as follows.

$$\begin{split} \Sigma_{\Delta \widehat{\Xi}}^{L} &= \mathbb{E}\left\{ \left(\boldsymbol{Z}_{ol}^{T} \otimes \boldsymbol{I}_{\ell} \right) \cdot \left(\left(\boldsymbol{Z}_{id}^{\dagger,T} \otimes \boldsymbol{I}_{\ell} \right) \cdot \operatorname{vec}(\boldsymbol{E}_{id}) \right) \cdot \left(\operatorname{vec}^{T}(\boldsymbol{E}_{id}) \cdot \left(\boldsymbol{Z}_{id}^{\dagger} \otimes \boldsymbol{I}_{\ell} \right) \right) \cdot \left(\boldsymbol{Z}_{ol} \otimes \boldsymbol{I}_{\ell} \right) \right\} \\ &= \left(\boldsymbol{Z}_{ol}^{T} \otimes \boldsymbol{I}_{\ell} \right) \cdot \mathbb{E}\left[\left(\boldsymbol{Z}_{id}^{\dagger,T} \otimes \boldsymbol{I}_{\ell} \right) \cdot \operatorname{vec}(\boldsymbol{E}_{id}) \cdot \operatorname{vec}^{T}(\boldsymbol{E}_{id}) \cdot \left(\boldsymbol{Z}_{id}^{\dagger} \otimes \boldsymbol{I}_{\ell} \right) \right] \cdot \left(\boldsymbol{Z}_{ol} \otimes \boldsymbol{I}_{\ell} \right). \end{split}$$

Here, the second equality is due to the statistical independence between $vec(E_{id})$ and Z_{ol} . Since $vec(E_{id})$ contains a sequence of white and zero-mean random variables, by standard derivations, e.g. in [10, Sec. 9.3] and [12] and references therein, we can write

$$\mathbb{E}\left[(\boldsymbol{Z}_{id}^{\dagger,T} \otimes \boldsymbol{I}_{\ell}) \cdot \operatorname{vec}(\boldsymbol{E}_{id}) \cdot \operatorname{vec}^{T}(\boldsymbol{E}_{id}) \cdot (\boldsymbol{Z}_{id}^{\dagger} \otimes \boldsymbol{I}_{\ell}) \right] \\ = \mathbb{E}\left[\left(\boldsymbol{Z}_{id}^{\dagger,T} \otimes \boldsymbol{I}_{\ell} \right) \cdot \mathbb{E}\left[\operatorname{vec}(\boldsymbol{E}_{id}) \cdot \operatorname{vec}^{T}(\boldsymbol{E}_{id}) \right] \cdot \left(\boldsymbol{Z}_{id}^{\dagger} \otimes \boldsymbol{I}_{\ell} \right) \right] \\ = \mathbb{E}\left[\left(\boldsymbol{Z}_{id}^{\dagger,T} \otimes \boldsymbol{I}_{\ell} \right) \cdot (\boldsymbol{I}_{N} \otimes \boldsymbol{\Sigma}_{e}) \cdot \left(\boldsymbol{Z}_{id}^{\dagger} \otimes \boldsymbol{I}_{\ell} \right) \right] \\ = \mathbb{E}\left(\boldsymbol{Z}_{id}^{\dagger,T} \cdot \boldsymbol{Z}_{id}^{\dagger} \right) \otimes \boldsymbol{\Sigma}_{e} \\ = \mathbb{E}\left(\boldsymbol{Z}_{id}^{T} \boldsymbol{Z}_{id}^{T} \right)^{-1} \otimes \boldsymbol{\Sigma}_{e}.$$

In the third equality, Property (45) in Lemma 2 is used. It is also a standard practice in least squares and system identification to use sample estimates to replace the expectation $\mathbb{E}(Z_{id}Z_{id}^T)$. See e.g. [10, Sec. 9.6]. Besides, due to the Hankel structure of the data matrix Z_{id} , the matrix multiplication in $Z_{id}Z_{id}^T$ naturally leads to sample estimates of the expectations, i.e. $Z_{id}Z_{id}^T \approx \mathbb{E}(Z_{id}Z_{id}^T)$ and $(Z_{id}Z_{id}^T)^{-1} \approx \mathbb{E}(Z_{id}Z_{id}^T)^{-1}$; therefore,

$$\mathbb{E}\left(\boldsymbol{Z}_{id}\boldsymbol{Z}_{id}^{T}\right)^{-1} \approx \mathbb{E}\left[\mathbb{E}\left(\boldsymbol{Z}_{id}\boldsymbol{Z}_{id}^{T}\right)\right]^{-1} = \left[\mathbb{E}\left(\boldsymbol{Z}_{id}\boldsymbol{Z}_{id}^{T}\right)\right]^{-1} \approx (\boldsymbol{Z}_{id}\boldsymbol{Z}_{id}^{T})^{-1}.$$

As $N \to \infty$, the approximation in the above equation becomes strict equality. For simplicity, we shall ignore the approximation error in using the sample estimates as in standard practice. We therefore have $\Sigma_{\Delta \hat{\Xi}}^{L} = \left[Z_{ol}^{T} \left(Z_{id} Z_{id}^{T} \right)^{-1} Z_{ol} \right] \otimes \Sigma_{e}$, and

$$\Sigma_{\Delta \hat{\Xi}, e}^{L} = \left[\boldsymbol{Z}_{ol}^{T} \left(\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^{T} \right)^{-1} \boldsymbol{Z}_{ol} \right] \otimes \Sigma_{e} + \boldsymbol{I}_{L} \otimes \Sigma_{e}.$$
(38)

Since $(Z_{id}Z_{id}^T)^{-1}$ is determined by Z_{id} , which and Z_{ol} and Σ_e are all known, $\Sigma_{\Delta \hat{\Xi}, e}^L$ can be computed. However, $\Sigma_{\Delta \hat{\Xi}, e}^L$ does not fully characterize $\text{Cov}(r_{k,L})$, which also depends on the covariance of $\beta_{k,L} + b_{k,L}$. The latter cannot be estimated, but indeed decays with the past horizon p. The following theorem quantifies the error bound in approximating $\text{Cov}(r_{k,L})$ by $\Sigma_{\Delta \hat{\Xi}, e}^L$. *Theorem 1:* Let the signals in the identification experiment and fault detection respectively satisfy (14,15) and (36,37). Then the following matrix

$$\Sigma_{\Delta \hat{\Xi}, e}^{L} = \left(\boldsymbol{Z}_{ol}^{T} \left(\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^{T} \right)^{-1} \boldsymbol{Z}_{ol} + I_{L} \right) \otimes \Sigma_{e}$$

approximates the covariance matrix of the residual $r_{k,L}$ computed by (31) in the following sense:

$$\left\|\operatorname{Cov}(\boldsymbol{r}_{k,L}) - \boldsymbol{\Sigma}_{\Delta \hat{\boldsymbol{\Xi}}, e}^{L}\right\|_{2} \leq \left(1 + \frac{\bar{\rho}_{z,ol}^{2}}{\underline{\rho}_{z,id}^{2}}\right) \cdot \|C\Phi^{p}\|_{2}^{2} \cdot \bar{\sigma}_{xx}^{2}$$

Proof: See Appendix C.

Due to the boundedness of $\bar{\rho}_{z,ol}, \bar{\sigma}_{xx}$ and $\underline{\rho}_{z,id}^2 > 0$, $\left(1 + \frac{\bar{\rho}_{z,ol}^2}{\underline{\rho}_{z,id}^2}\right) \cdot \bar{\sigma}_{xx}^2$ is bounded. Hence, the approximation error decays exponentially with the past horizon p.

Remark 4: Note from (38) that $\Sigma_{\Delta \hat{\Xi}}^{L}$ relies not only on the identification data, but also on the online I/O signals measured during the implementation of the residual generator. This means that the covariance matrix of the data-driven residual vector cannot be entirely determined by the identification data. This is fundamentally different from the robust fault detection scheme in [37], where it is the model changes (or multiplicative faults), instead of additive faults, that are detected. The fact making this difference is that a residual generator (a filter with uncertain parameters) is involved in the data-driven method proposed in this paper, but not required by the scheme in [37].

C. Statistically optimal fault detection test against identification errors

Now, due to (25), when $\varphi_f = 0$, $\mu_{k,L} = \mathbb{E} r_{k,L} = \mathbb{E} (\beta_{k,L} + b_{k,L})$. And according to Theorem 1, $\text{Cov}(r_{k,L})$ can be approximated by $\Sigma_{\Delta \hat{\Xi}, e}^L$ with an arbitrarily small error if p is chosen sufficiently large. The distribution of $r_{k,L}$ can hence be expressed as:

$$oldsymbol{r}_{k,L} \sim \left\{ egin{array}{ll} \mathcal{N}\left(\mathbb{E}oldsymbol{eta}_{k,L} + \mathbb{E}oldsymbol{b}_{k,L}, \Sigma^L_{\Delta \hat{\Xi},e}
ight), & ext{fault free}, \ \mathcal{N}\left(\mathbb{E}oldsymbol{eta}_{k,L} + \mathbb{E}oldsymbol{b}_{k,L} + oldsymbol{arphi}_{f}, \Sigma^L_{\Delta \hat{\Xi},e}
ight), & ext{faulty}. \end{array}
ight.$$

Whiten $r_{k,L}$ by $\left(\Sigma_{\Delta \hat{\Xi}, e}^{L}\right)^{-\frac{1}{2}}$, i.e. $\tilde{r}_{k,L} \triangleq \left(\Sigma_{\Delta \hat{\Xi}, e}^{L}\right)^{-\frac{1}{2}} r_{k,L}$. The test statistic for the changing mean in $\tilde{r}_{k,L}$ can be written as [3], [35]

$$\tau(k) \triangleq \tilde{\boldsymbol{r}}_{k,L}^T \cdot \tilde{\boldsymbol{r}}_{k,L} = \left\| \left(\boldsymbol{\Sigma}_{\Delta \hat{\Xi}, e}^L \right)^{-\frac{1}{2}} \boldsymbol{r}_{k,L} \right\|_2^2.$$
(39)

However, $\tau(k)$ defines a noncentral χ^2 distribution even in the fault free case, with a non-centrality parameter

$$\lambda_{\boldsymbol{r}_{k,L}} = \left\| \left(\Sigma_{\Delta \hat{\Xi}, e}^{L} \right)^{-\frac{1}{2}} \cdot \left(\mathbb{E} \boldsymbol{\beta}_{k,L} + \mathbb{E} \boldsymbol{b}_{k,L} \right) \right\|_{2}^{2}.$$

Denote this distribution by $\tau(k) \sim \chi^2_{L\ell,\lambda_{r_{k,L}}}$, with $L\ell$ degrees of freedom (DoF) [3], [35]. The parameter $\lambda_{r_{k,L}}$ is again unknown. But if the past horizon p also satisfies

$$\|C\Phi^{p}\|_{2}^{2} < \frac{\zeta \cdot L\ell}{\|\Sigma_{e}^{-1/2}\|_{2}^{2} \left(\|\mathbb{E}\left[\operatorname{vec}(\boldsymbol{X}_{id})\right]\|_{2} + \|\mathbb{E}\hat{\boldsymbol{x}}_{k-p,L}\|_{2}\right)^{2}},\tag{40}$$

for an arbitrarily small $\zeta > 0$; then the cumulative distribution function (cdf) of $\tau(k)$ under the fault free case can be approximated by that of a central χ^2 distribution. Detailed analysis is given in Appendix D.

The fault detection test can hence be derived based on the central χ^2 distribution, i.e. $\chi^2_{L\ell}$, as follows.

$$au(k) \hspace{0.2cm} \underset{ ext{no fault}}{\gtrless} \hspace{0.2cm} \gamma_{lpha},$$

where γ_{α} denotes the detection threshold, determined by a chosen false alarm rate (FAR) α .

Remark 5: In practice, the inequality (40) cannot be explicitly checked, due to the unknown parameters therein, e.g. the mean of the Kalman filter states. A practical way is to tune p and L such that in nominal case, the test statistic $\tau(k)$ defined in (39) leads to an FAR as close as possible to the chosen FAR, when compared to the theoretical threshold required by the $\chi^2_{L\ell}$ test under the chosen FAR. This tuning procedure can be carried out using the identification data.

Remark 6: In the proposed method, the past horizon p is required to be large. It is known in system identification literature that estimating a higher order VARX model (12), i.e. with bigger p, leads to larger parameter covariance, and thus more uncertainties in the residual generator. But in this paper, this increased covariance has been explicitly taken into account in the covariance of the residual vector.

By (38) and the matrix inversion lemma, the test statistic $\tau(k)$ defined in (39) can be decomposed into two terms, i.e.

$$\boldsymbol{\tau}(k) = \boldsymbol{r}_{k,L}^{T} (\boldsymbol{\Sigma}_{e}^{L})^{-1} \boldsymbol{r}_{k,L} - \boldsymbol{r}_{k,L}^{T} \left[\left(\boldsymbol{Z}_{ol}^{T} \left(\boldsymbol{Z}_{ol} \boldsymbol{Z}_{ol}^{T} + \boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^{T} \right)^{-1} \boldsymbol{Z}_{ol} \right) \otimes \boldsymbol{\Sigma}_{e}^{-1} \right] \boldsymbol{r}_{k,L},$$

where the first term is normalized only by the innovation covariance, and represents the nominal design; while the second term is a correction against the parameter errors. The robust design can hence be schematically shown in Fig. 2.

Algorithm 1 summarizes the robust data driven fault detection approach.



Fig. 2. Robust fault detection against stochastic parameter errors. To the right of the vertical dashed line: above the dotted line is the nominal detection scheme, below the dotted line is the correction of the statistic for robustness.

Algorithm 1 (Robust Data-Driven Fault Detection):

Design Phase:

- 1) choose the horizons $L \leq p$ and the false alarm rate α , and determine the threshold γ_{α} ;
- 2) measure the I/O signals from a plant, and form the data matrix Z_{id} ;
- 3) compute $\hat{\Xi}$ by (16) and $\hat{\Sigma}_e$ by (18);
- 4) form the Hankel and Toeplitz matrices $\bar{H}_{z}^{L,p}, \bar{T}_{u}^{L}, \bar{T}_{v}^{L}$.

Implementation Phase at time instant k:

- 1) measure the past p + L I/Os and u(k) from the plant, and form the data matrix Z_{ol} ;
- 2) compute the residual $r_{k,L}$ by (31) and the covariance $\sum_{\Delta \hat{\Xi}, e}^{L}$ by (38);
- 3) compute the test statistic $\tau(k)$ by (39) and compare it to the threshold γ_{α} .

IV. SIMULATION STUDIES

A. Illustration of the main idea with a simple example

Consider a SISO zeroth order system; i.e.

$$y(k) = d \cdot u(k) + e(k),$$

where e(k) is zero mean white Gaussian with variance σ_e^2 . The scalar *d* is a static gain. Now, the purpose is to design a residual generator of the following form from data; i.e.

$$r(k) = y(k) - \hat{d} \cdot u(k). \tag{41}$$

Here, the detection horizon is L = 1; and \hat{d} is from the LS estimate (16):

$$\hat{d} = Y_{id} \cdot U_{id}^{\dagger}$$

DRAFT

where U_{id} and Y_{id} collect respectively N I/O data samples measured from this plant from a previous time instant t. Due to the static property of the system, the past horizon is p = 0. Thus, $Z_{id} = U_{id}$ in (16) and $Y_{id}, U_{id} \in \mathbb{R}^{1 \times N}$. It is easy to derive the error statistics of the estimate; i.e.

$$\Delta \hat{d} = d - \hat{d} = \boldsymbol{E}_{id} \cdot \boldsymbol{U}_{id}^{\dagger},$$
$$\operatorname{var}(\Delta \hat{d}) = \frac{\sigma_e^2}{U_{id}U_{id}^T}.$$

The variance, $var(\Delta \hat{d})$, is determined by the SNR, defined as $10 \cdot \log_{10} \frac{1/N \cdot U_{id} U_{id}^T}{\sigma_e^2}$.

Now, the residual in (41) has the following structure, when no fault signal is added:

$$r(k) = du(k) + e(k) - \hat{d}u(k) = \Delta \hat{d} \cdot u(k) + e(k)$$

Obviously, both $\Delta \hat{d} \cdot u(k)$ and e(k) contribute to var(r(k)). Due to the independence of $\Delta \hat{d}$ and e(k) and the fact that u(k) is a known quantity,

$$\operatorname{var}(r(k)) = \operatorname{var}(\Delta \hat{d} \cdot u(k)) + \operatorname{var}(e(k))$$
$$= u^2(k) \cdot \operatorname{var}(\Delta \hat{d}) + \sigma_e^2$$
$$= \frac{u^2(k) \cdot \sigma_e^2}{U_{id}U_{id}^T} + \sigma_e^2.$$

The first term in the last equality corresponds to the additional covariance information targeted in Problem 1, which consists of two factors; i.e. $var(\Delta \hat{d})$ of the identified parameter and u(k) measured during the implementation of the residual generator. Moreover, $r^2(k)/var(r(k))$ is χ^2 distributed with one DoF, denoted as χ_1^2 .

Let us now consider the numerical values, d = 2, $\sigma_e = 1$, N = 2. The input was chosen as $u \equiv 0.001$ in the identification experiment, leading to an SNR of -60dB, and $var(\Delta \hat{d}) = 5 \times 10^5 \gg \sigma_e^2$. We did 10^4 Monte Carlo (MC) simulations with independent innovation sequences in the identification experiment, and collected 10^4 different estimates of \hat{d} . The distribution of $\Delta \hat{d}$ is plotted in Fig. 3(a). We then used the 10^4 estimates, \hat{d} , in computing the residual via (41), where u(k) was set to 2. Hence, $u^2(k) \cdot var(\Delta \hat{d}) =$ $2 \times 10^6 \gg \sigma_e^2$. The distribution of $r^2(k)/var(r(k))$ is plotted in Fig. 3(b), and found to perfectly follow the theoretical χ_1^2 distribution. In contrast, the distribution of $r^2(k)/\sigma_e^2$ shown in Fig. 3(c) significantly deviates from χ_1^2 , due to the dominant majority of its population valued much higher than 10.8, or the 99.9% probability bound of the χ_1^2 distribution. The few nontrivial values in this plot are due to the very rare events where $r^2(k)/\sigma_e^2$ evaluates lower than 10.8.

B. Fault detection in a MIMO dynamic system

1) Model and simulation parameters: Consider a linearized VTOL (vertical take-off and landing) model, originally appeared in [38] and also studied in [3], [21], in the form of (3,4); with $D = 0, F = I_4$,



Fig. 3. Distribution of respectively $\Delta \hat{d}$ in (a), $r^2(k)/\text{var}(r(k)$ in (b), and $r^2(k)/\sigma_e^2$ in (c). Solid: theoretical distributions. Dots: empirical probabilities evaluated in continuous bins, as the ratio between the number of points in each bin and the total number of MC simulations 10⁴ (283 bins within the 6 σ bound for (a), and 217 bins within the 99.9% cdf bound for (b) and (c)).

and A, B, C as discretized at a sampling rate of 0.5 seconds, from the following continuous time model (distinguished from discrete-time model by the subscript "c"),

$$\begin{split} \dot{x}_{c}(t) &= A_{c}x_{c}(t) + B_{c}u_{c}(t) \\ y_{c}(t) &= C_{c}x_{c}(t) \\ A_{c} &= \begin{bmatrix} -0.0366 & 0.0271 & 0.0188 & -0.4555 \\ 0.0482 & -1.01 & 0.0024 & -4.0208 \\ 0.1002 & 0.3681 & -0.707 & 1.42 \\ 0 & 0 & 1 & 0 \end{bmatrix}, C_{c} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \\ B_{c} &= \begin{bmatrix} 0.4422 & 3.5446 & -5.52 & 0 \\ 0.1761 & -7.5922 & 4.49 & 0 \end{bmatrix}^{T}. \end{split}$$

The process and measurement noise, w(k), v(k), are assumed to be zero mean white noise, respectively with a covariance of $Q_w = 0.25 \cdot I_4$ and $Q_v = 2 \cdot I_2$.

In the identification experiment, an empirical stabilizing output feedback controller was used; i.e.

$$u(k) = -\begin{bmatrix} 0 & 0 & -0.5 & 0 \\ 0 & 0 & -0.1 & -0.1 \end{bmatrix} \cdot y(k) + \psi(k)$$

where $\psi(k)$ is a zero-mean white noise with a covariance of diag(1,1), which ensures that the system is persistently excited to any order. 2000 data samples were collected from the simulation.

The past horizon was chosen as p = 15. The covariance of the identified parameters (see e.g. [18]), i.e.

$$\operatorname{Cov}(\operatorname{vec}(\hat{\Xi})) = (\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^T)^{-1} \otimes \boldsymbol{\Sigma}_e, \tag{42}$$

November 7, 2012

DRAFT

has a maximum singular value of 0.044, corresponding to an SNR of $10\log_{10}(1/N \cdot 1/\|\text{Cov}(\text{vec}(\hat{\Xi}))\|_2) = -19.4dB$. Based on the steady-state Kalman filter (see e.g. [27], [28]) of the discrete-time VTOL model, $\|C\Phi^p\|_2^2$ was found to be 2.1×10^{-5} . The other parameters of Algorithm 1 were chosen as $L = 10, \alpha = 0.005$. The design phase of this algorithm produced the parameters, $\hat{\Xi}, \hat{\Sigma}_e, (Z_{id}Z_{id}^T)^{-1}, \bar{H}_z^{L,p}, \bar{T}_u^L, \bar{T}_y^L, \bar{H}_f^{L,p}, \bar{T}_f^L$.

The online detection experiment is designed as follows. The initial states of the system were set as $[10\ 10\ 1\ 1]^T$. An LQG tracking controller was used, to maintain a vertical velocity (the second output) of 40. The purpose of this closed-loop experiment is to show the advantage of the robust fault detection method, when the I/O signals in the system are large enough for $\Sigma_{\Delta \hat{\Xi}, e}^L$ to be significantly different from Σ_e^L .

2) *Fault detection results:* Consider the following actuator faults. The first actuator was stuck at -3 in the interval of $301 \le k \le 600$; and the second actuator had a bias of -5 in the interval of $901 \le k \le 1200$. The simulation was run for 1200 sampling instants.

We checked (40) based on the steady-state Kalman filter:

$$\frac{L\ell}{\|\boldsymbol{\Sigma}_{e}^{-1/2}\|_{2}^{2} \left(\|\operatorname{vec}(\boldsymbol{X}_{id})\|_{2} + \|\hat{\boldsymbol{x}}_{k-p,L}\|_{2}\right)^{2}} = 5.4 \times 10^{-5} > \|C\Phi^{p}\|_{2}^{2}.$$

Here, we used $\hat{x}_{k-p,L}$, instead of $\mathbb{E}\hat{x}_{k-p,L}$, since the latter cannot be evaluated. In practice, (40) cannot be computed. Practical method for tuning p,L is suggested in Remark 5.

We tested and compared four data-driven fault detection solutions all with p = 15, L = 10. These are P_{ss} , the classical model-based PSA, with (A, B, C, D, K) identified by the PBSID-OPT method of [12]; P_{mp} , the data-driven PSA, proposed in Sec. 3 of [18]; F_{no} , the nominal data-driven method proposed in Sec. II of this paper, without considering the parameter identification errors; F_{rb} , the robust data-driven method, as proposed in Sec. III of this paper. The test thresholds were all computed with the false alarm rate of 0.5%, as in the χ^2 test. The results are illustrated in Fig. 4 and compared in the following table, where FA and MA means the number of respectively false alarms and missed alarms; and SCR is the overall successful classification rate.

	FA	MA	Delay	SCR
P _{ss}	584	0	0	51.33%
P_{mp}	254	3	0	78.58%
Fno	556	0	0	53.67%
F_{rb}	16	1	0	98.58%

As indicated by their large magnitudes above the thresholds under the fault-free case in Fig. 4, none

of the nominal data-driven algorithms can correctly quantify the distribution of their test statistics, since only the innovation signals are considered therein to be stochastic. Besides, the statistics of P_{ss} under the nominal cases were even further away from the threshold than those of P_{mp} , because the identified state space matrices not only inherit the errors from the Markov parameters identified first, but also contain model reduction errors and the errors from solving a second least-squares [12]. Consequently, P_{ss} become highly nonlinear in the noise originally contained in the identification data.

The robust data-driven method correctly accounts the composite covariance information contained both in the innovation signals and in the parameter identification errors, as can be seen from the correct theoretical threshold separating the fault-free case from the faulty one. The FAR is significantly reduced, but is slightly higher than the design parameter, $\alpha = 0.5\%$, mainly due to the FAs from sampling instant 601 to 610. Since L = 10, it took 10 sampling instants for the fault signals to entirely retreat from the detection window. However, the robust data-driven method has a lowered sensitivity to faults, compared with its nominal counterpart, as can be seen from the reduced gap between the test statistics of respectively the fault-free case and the faulty one. This is the conservativeness to pay for the robustness.

V. CONCLUSIONS

In this paper, we have explicitly analyzed the effect of parameter identification errors on the data-driven fault detection design. We have also derived in closed-form the residual vector covariance in terms of both the innovation signals and the parameter identification errors. Besides, we have established the error bound in neglecting the contribution of the bias terms due to initial states to the residual; and showed that this error decays with the past horizon. All the analytical results are tested in the simulation studies, which have validated that the data-driven fault detection method developed in this paper has clearly improved performance compared to the nominal data-driven solutions without taking into account the identification uncertainty, especially when the SNR of the identification data is low. A robust fault isolation method against parameter errors is developed in the companion paper [36]. Possible future directions are to extend the robust fault detection method to deal with multiplicative faults and to linear parameter varying systems.

REFERENCES

[1] J. Chen and R. Patton, Robust Model-Based Fault Diagnosis for Dynamic Systems. Kluwer Academic Publishers, 1999.

- [2] S. Ding, Model-based Fault Diagnosis Techniques Design Schemes, Algorithms, and Tools. Springer, 2008.
- [3] F. Gustafsson, Adaptive filtering and change detection. John Wiley & Sons, 2001.
- [4] R. Isermann, Fault-Diagnosis Systems: An introduction from Fault Detection to Fault Tolerance. Springer Verlag, 2006.

- [5] H. Hjalmarsson, "System identification of complex and structured systems," *European Journal of Control*, vol. 15, pp. 275–310, 2009.
- [6] S. Qin, "Data-driven fault detection and diagnosis for complex industrial processes," in the Proceedings of the 7th IFAC SafeProcess, Barcelona, Spain, 2009, pp. 1115–1125.
- [7] H. Bassily, R. Lund, and J. Wagner, "Fault detection in multivariate signals with applications to gas turbines," *IEEE Transactions on Signal Processing*, vol. 57, pp. 835–842, 2009.
- [8] S. Ding, P. Zhang, A. Naik, E. Ding, and B. Huang, "Subspace method aided data-driven design of fault detection and isolation systems," *Journal of Process Control*, vol. 19, pp. 1496–1510, 2009.
- [9] S. Qin and W. Li, "Detection and identification of faulty sensors in dynamic processes," *AIChE Journal*, vol. 47, pp. 1581–1593, 2001.
- [10] L. Ljung, System Identification Theory for the User. Prentice-Hall, 1987.
- [11] M. Verhaegen and V. Verdult, Filtering and System Identification: A Least Squares Approach. Cambridge University Press, 2007.
- [12] A. Chiuso, "On the relation between CCA and predictor-based subspace identification," *IEEE Transactions on Automatic Control*, vol. 52, pp. 1795–1812, 2007.
- [13] P. van Overschee and B. de Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 36, pp. 75–93, 1994.
- [14] E. Chow and A. Willsky, "Analytical redundancy and the design of robust failure detection systems," *IEEE Transactions on Automatic Control*, vol. 29, pp. 603–614, 1984.
- [15] P. Frank, "Enhancement of robustness in observer-based fault detection," *International Journal of control*, vol. 59, pp. 955–981, 1994.
- [16] J. Gertler and D. Singer, "A new structural framework for parity equation based failure detection and isolation," *Automatica*, vol. 26, pp. 381–388, 1990.
- [17] R. Patton, P. Frank, and R. Clarke, Fault diagnosis in dynamic systems: theory and application. Prentice-Hall, 1989.
- [18] J. Dong, M. Verhaegen, and F. Gustafsson, "Data driven fault detection with robustness to uncertain parameters identified in closed loop," in *Proceedings of the 49th IEEE Conference on Decision and Control*, Atlanta, USA, 2010, pp. 750–755.
- [19] J. Dong and M. Verhaegen, "Subspace based fault identification for LTI systems," in the Proceedings of the 7th IFAC SafeProcess, Barcelona, Spain, 2009, pp. 330–335.
- [20] —, "Identification of fault estimation filter from I/O data for systems with stable inversion," *IEEE Transactions on Automatic Control*, vol. In Press, 2011.
- [21] F. Gustafsson, "Statistical signal processing approaches to fault detection," Annual Reviews in Control, vol. 31, pp. 41–54, 2007.
- [22] K. Öhrn, A. Ahlén, and M. Sternad, "A probabilistic approach to multivariable robust filtering and open-loop control," *IEEE Transactions on Automatic Control*, vol. 40, pp. 405–418, 1995.
- [23] U. Orguner and F. Gustafsson, "Risk-sensitive particle filters for mitigating sample impoverishment," *IEEE Transactions on Signal Processing*, vol. 56, pp. 5001 –5012, 2008.
- [24] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 425–437, 2002.
- [25] M. Basseville and I. Nikiforov, Detection of Abrupt Changes Theory and Application. Prentice-Hall, 1993.

- [26] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. Heidelberg: Springer Verlag, 2003.
- [27] G. Goodwin and K. Sin, Adaptive Filtering Prediction and Control. Prentice-Hall, 1984.
- [28] T. Kailath, A. Sayed, and B. Hassibi, Linear Estimation. Prentice-Hall, 2000.
- [29] M. Verhaegen, "Application of a subspace model identification technique to identify LTI systems operating in closed-loop," *Automatica*, vol. 29, pp. 1027–1040, 1993.
- [30] R. Isermann and R. Ballé, "Trends in the application of model-based fault detection and diagnosis of technical processes," *Control Engineering Practice*, vol. 5, pp. 709–719, 1997.
- [31] A. Dahlen and W. Scherrer, "The relation of the CCA subspace method to a balanced reduction of an autoregressive model," *Journal of Econometrics*, vol. 118, pp. 293–312, 2004.
- [32] S. Qin and L. Ljung, "On the role of future horizon in closed-loop subspace identification," in *the Proceedings of the 14th. IFAC Symposium on System Identification*, New Castle, Australia, 2009, pp. 1080–1084.
- [33] J. Dong and M. Verhaegen, "Cautious *H*₂ optimal control using uncertain markov parameters identified in closed loop," *Systems & Control Letters*, vol. 58, pp. 378–388, 2009.
- [34] A. Chiuso, "The role of vector autoregressive modeling in predictor based subspace identification," *Automatica*, vol. 43, pp. 1034–1048, 2007.
- [35] S. Kay, Fundamentals of Statistical Signal Processing: Detection Theory. Prentice-Hall, 1998.
- [36] J. Dong, M. Verhaegen, and F. Gustafsson, "Robust fault isolation with statistical uncertainty in identified parameters," *IEEE Transactions on Signal Processing*, vol. 60, pp. 5556–5561, 2012.
- [37] O. Kwon, G. Goodwin, and W. Kwon, "Robust fault detection method accounting for modelling errors in uncertain systems," *Control Engineering Practice*, vol. 2, pp. 763–771, 1994.
- [38] K. Narendra and S. Tripathi, "Identification and optimization of aircraft dynamics," *Journal of Aircraft*, vol. 10, pp. 193–199, 1973.
- [39] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Automatic Control*, vol. 25, pp. 772–781, 1978.
- [40] M. Sankaran, "Approximations to the non-central chi-square distribution," Biometrika, vol. 50, pp. 199–204, 1963.

APPENDIX A

PROPERTIES OF KRONECKER PRODUCTS

In the paper, the following properties of Kronecker products [39] are frequently used.

Lemma 2: For matrices M_1, M_2, M_3, M_4 with proper dimensions, the following properties hold:

$$\operatorname{vec}(M_1 \cdot M_2 \cdot M_3) = (M_3^T \otimes M_1) \cdot \operatorname{vec}(M_2), \tag{43}$$

$$(M_1 \otimes M_2) \otimes M_3 = M_1 \otimes (M_2 \otimes M_3), \tag{44}$$

$$(M_1 \otimes M_2) \cdot (M_3 \otimes M_4) = (M_1 \cdot M_3) \otimes (M_2 \cdot M_4), \tag{45}$$

$$(M_1+M_2)\otimes M_3 = M_1\otimes M_3 + M_2\otimes M_3, \tag{46}$$

$$(M_1 \otimes M_2)^T = M_1^T \otimes M_2^T.$$
⁽⁴⁷⁾

DRAFT

APPENDIX B

PROOF OF LEMMA 1

We only consider the fault-free case. By the output equation (10), for $i = 1, \dots, L$, we can write

$$y(k-L+i) = C\Phi^{p}\hat{x}(k-p-L+i) + e(k-L+i) + (\hat{\Xi} + \Delta \hat{\Xi}) \cdot \begin{bmatrix} z_{k-L+i-1,p} \\ u(k-L+i) \end{bmatrix}.$$

Here, Ξ is replaced by $\hat{\Xi} + \Delta \hat{\Xi}$. Since $\Delta \hat{\Xi} \cdot \begin{bmatrix} z_{k-L+i-1,p} \\ u(k-L+i) \end{bmatrix}$ is a column vector, it equals vec $\left(\Delta \hat{\Xi} \cdot \begin{bmatrix} z_{k-L+i-1,p} \\ u(k-L+i) \end{bmatrix}\right)$
which by Property (43), can be written as

$$\left(\begin{bmatrix} \boldsymbol{z}_{k-L+i-1,p}^T & \boldsymbol{u}^T(k-L+i)\end{bmatrix} \otimes \boldsymbol{I}_\ell\right) \cdot \operatorname{vec}(\Delta \hat{\boldsymbol{\Xi}})$$

On the other hand, by Property (43), Eq. (17) can be rewritten in a vectorized form; i.e.

$$\operatorname{vec}(\Delta \hat{\Xi}) = (\mathbf{Z}_{id}^{\dagger,T} \otimes I_{\ell}) \cdot \operatorname{vec}(\mathbf{E}_{id}) + [\mathbf{Z}_{id}^{\dagger,T} \otimes (C\Phi^{p})] \cdot \operatorname{vec}(\mathbf{X}_{id}).$$

We hence have

$$\Delta \hat{\Xi} \cdot \begin{bmatrix} \boldsymbol{z}_{k-L+i-1,p} \\ u(k-L+i) \end{bmatrix} = \left(\begin{bmatrix} \boldsymbol{z}_{k-L+i-1,p}^T & \boldsymbol{u}^T(k-L+i) \end{bmatrix} \otimes I_\ell \right) \cdot \left\{ (\boldsymbol{Z}_{id}^{\dagger,T} \otimes I_\ell) \operatorname{vec}(\boldsymbol{E}_{id}) + [\boldsymbol{Z}_{id}^{\dagger,T} \otimes (C\Phi^p)] \operatorname{vec}(\boldsymbol{X}_{id}) \right\}$$

Now, assemble y(k-L+i), $i = 1, \dots, L$ into $y_{k,L}$, and arrange the terms, the residual structure (33) follows.

APPENDIX C

PROOF OF THEOREM 1

Again, $e_{k,L}$ and $vec(E_{id})$ are independent, white, and have zero mean. With $\varphi_f = 0$, the mean of the residual equals

$$\mathbb{E}\boldsymbol{r}_{k,L} = \mathbb{E}\boldsymbol{\beta}_{k,L} + \mathbb{E}\boldsymbol{b}_{k,L} = [\boldsymbol{I}_L \otimes (C\Phi^p)] \cdot \mathbb{E}\hat{\boldsymbol{x}}_{k-p,L} + \left[(\boldsymbol{Z}_{ol}^T \boldsymbol{Z}_{id}^{\dagger,T}) \otimes (C\Phi^p) \right] \cdot \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right].$$

For simplicity, we will also use the sample estimates $Z_{id}^{\dagger} Z_{ol}$ to replace $\mathbb{E}(Z_{id}^{\dagger} Z_{ol})$ in this proof.

With the cross correlations of the independent terms left out, the covariance of $r_{k,L}$ takes the following form,

$$\operatorname{Cov}(\boldsymbol{r}_{k,L}) = \operatorname{Cov}(\boldsymbol{\beta}_{k,L}) + \operatorname{Cov}(\boldsymbol{b}_{k,L}) + \operatorname{Cov}(\boldsymbol{\zeta}_{k,L}) + \operatorname{Cov}(\boldsymbol{e}_{k,L}).$$

To see $\mathbb{E}\left(\left(\beta_{k,L} - \mathbb{E}\beta_{k,L}\right) \cdot \zeta_{k,L}^{T}\right) = 0$, note that the closed-loop observer states in X_{id} are independent of the innovation signals contained in E_{id} , given the measured past I/O signals. This can be seen from (7), and is similar to the standard discussions in Kalman filtering; e.g. [27, Chapter 7]. However, vec (X_{id}) is stochastic with bounded covariance matrix as given in (15), but distributed independently of vec (E_{id}) .

In $\text{Cov}(r_{k,L})$, $\text{Cov}(\zeta_{k,L}) + \text{Cov}(e_{k,L})$ is derived in (38). The other two terms are derived as follows.

$$\operatorname{Cov}(\boldsymbol{b}_{k,L}) = [I_L \otimes (C\Phi^p)] \cdot \operatorname{Cov}(\hat{\boldsymbol{x}}_{k-p,L}) \cdot [I_L \otimes (C\Phi^p)^T]$$

Here, $C\Phi^p$ is the true constant parameter of the plant, and can hence be moved outside the "Cov" operator. Besides, since $x(k) - \hat{x}(k)$ has zero mean, and is uncorrelated with the past I/Os [27, Chapter 7], we have

$$\operatorname{Cov}(\boldsymbol{\beta}_{k,L}) = \left[(\boldsymbol{Z}_{ol}^T \boldsymbol{Z}_{id}^{\dagger,T}) \otimes (C\Phi^p) \right] \cdot \operatorname{Cov}(\operatorname{vec}(\boldsymbol{X}_{id})) \cdot \left[(\boldsymbol{Z}_{id}^\dagger \boldsymbol{Z}_{ol}) \otimes (C\Phi^p)^T \right].$$

Then due to (15,37), and Properties (44,45),

$$\begin{aligned} \operatorname{Cov}(\boldsymbol{b}_{k,L}) + \operatorname{Cov}(\boldsymbol{\beta}_{k,L}) \\ & \leq [I_L \otimes (C\Phi^p)] \cdot \left[I_L \otimes (\bar{\sigma}_{xx}^2 I_n)\right] \cdot \left[I_L \otimes (C\Phi^p)^T\right] + \left[(\boldsymbol{Z}_{ol}^T \boldsymbol{Z}_{id}^{\dagger,T}) \otimes (C\Phi^p)\right] \cdot \left[I_N \otimes (\bar{\sigma}_{xx}^2 I_n)\right] \cdot \left[(\boldsymbol{Z}_{id}^{\dagger} \boldsymbol{Z}_{ol}) \otimes (C\Phi^p)^T\right] \\ & = \left(I_L + \boldsymbol{Z}_{ol}^T \boldsymbol{Z}_{id}^{\dagger,T} \boldsymbol{Z}_{id}^{\dagger} \boldsymbol{Z}_{ol}\right) \otimes \left[(C\Phi^p)(\bar{\sigma}_{xx}^2 I_n)(C\Phi^p)^T\right] \\ & = \left(I_L + \boldsymbol{Z}_{ol}^T (\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^T)^{-1} \boldsymbol{Z}_{ol}\right) \otimes \left[(C\Phi^p)(\bar{\sigma}_{xx}^2 I_n)(C\Phi^p)^T\right] .\end{aligned}$$

Now, due to the inequalities (14,36),

$$\boldsymbol{Z}_{ol}^{T} \left(\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^{T} \right)^{-1} \boldsymbol{Z}_{ol} \preceq \frac{\bar{\rho}_{z,ol}^{2}}{\underline{\rho}_{z,id}^{2}} \boldsymbol{I}.$$

The error bound can finally be quantified as

$$\begin{split} \|\operatorname{Cov}(\boldsymbol{r}_{k,L}) - \boldsymbol{\Sigma}_{\Delta \hat{\Xi}, e}^{L} \|_{2} \\ &= \|\operatorname{Cov}(\boldsymbol{b}_{k,L}) + \operatorname{Cov}(\boldsymbol{\beta}_{k,L})\|_{2} \\ &\leq \left\|I_{L} + \boldsymbol{Z}_{ol}^{T} \left(\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^{T}\right)^{-1} \boldsymbol{Z}_{ol}\right\|_{2} \cdot \left\| (C\Phi^{p})(\bar{\sigma}_{xx}^{2}I_{n})(C\Phi^{p})^{T} \right\|_{2} \\ &\leq \left(1 + \frac{\bar{\rho}_{z,ol}^{2}}{\underline{\rho}_{z,id}^{2}}\right) \cdot \|C\Phi^{p}\|_{2}^{2} \cdot \bar{\sigma}_{xx}^{2}. \end{split}$$

This completes the proof.

APPENDIX D

Analysis of the condition to ignore the noncentrality in the distribution of au(k)

First, recall that the cdf of the noncentral χ^2 distribution, denoted as $P(\tau(k); L\ell, \lambda_{r_{k,L}})$, has no closedform expression in $L\ell$ and $\lambda_{r_{k,L}}$, but can be approximated by the cdf of a normal distribution (denoted by P_N) [40]; i.e.

$$P_{N}\left(\frac{1-ab(1-a+0.5(2-a)cb)-\left(\frac{\tau(k)}{L\ell+\lambda_{r_{k,L}}}\right)^{a}}{a\sqrt{2b(1+cb)}}\right).$$

Here, the scalars a, b, c are defined as

$$a = 1 - \frac{2}{3} \frac{(L\ell + \lambda_{r_{k,L}})(L\ell + 3\lambda_{r_{k,L}})}{(L\ell + 2\lambda_{r_{k,L}})^2}, b = \frac{L\ell + 2\lambda_{r_{k,L}}}{(L\ell + \lambda_{r_{k,L}})^2}, c = (a-1)(1-3a).$$

Clearly, P_N is parameterized by the linear combinations of the DoF, $L\ell$, and the non-centrality parameter, $\lambda_{r_{k,L}}$; i.e. $L\ell + i \cdot \lambda_{r_{k,L}}$, i = 1, 2, 3. Hence, if $\lambda_{r_{k,L}} \ll L\ell$, then $L\ell + i \cdot \lambda_{r_{k,L}} \approx L\ell$, i = 1, 2, 3; and thus $\chi^2_{L\ell, \lambda_{r_{k,L}}}$ reduces to $\chi^2_{L\ell}$. This is what remains to show.

We shall now derive an upper bound of $\|\mathbb{E}\tilde{r}_{k,L}\|_2$.

$$\|\mathbb{E} ilde{m{r}}_{k,L}\|_2 \leq \left\| \left(\Sigma_{\Delta \hat{\Xi},e}^L
ight)^{-rac{1}{2}} \mathbb{E}m{eta}_{k,L} \right\|_2 + \left\| \left(\Sigma_{\Delta \hat{\Xi},e}^L
ight)^{-rac{1}{2}} \mathbb{E}m{b}_{k,L} \right\|_2.$$

Similar to the derivations in proving Theorem 1, we have

$$\begin{split} & \left\| \left(\Sigma_{\Delta \hat{\Xi}, e}^{L} \right)^{-\frac{1}{2}} \mathbb{E} \boldsymbol{\beta}_{k, L} \right\|_{2}^{2} \\ &= \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right]^{T} \cdot \left\{ \left[\boldsymbol{Z}_{id}^{\dagger} \boldsymbol{Z}_{ol} \cdot \left(\boldsymbol{Z}_{ol}^{T} (\boldsymbol{Z}_{id} \boldsymbol{Z}_{id}^{T})^{-1} \boldsymbol{Z}_{ol} + I_{\ell} \right)^{-1} \cdot \boldsymbol{Z}_{ol}^{T} \boldsymbol{Z}_{id}^{\dagger, T} \right] \otimes \left[(C \Phi^{p})^{T} \Sigma_{e}^{-1} (C \Phi^{p}) \right] \right\} \cdot \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right] \\ &\leq \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right]^{T} \cdot \left\{ I_{N} \otimes \left[(C \Phi^{p})^{T} \Sigma_{e}^{-1} (C \Phi^{p}) \right] \right\} \cdot \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right] \\ &\leq \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right]^{T} \cdot \left(\left\| \Sigma_{e}^{-1/2} \right\|_{2}^{2} \cdot \left\| C \Phi^{p} \right\|_{2}^{2} \cdot I_{Nn} \right) \cdot \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right] \\ &= \left\| \Sigma_{e}^{-1/2} \right\|_{2}^{2} \cdot \left\| C \Phi^{p} \right\|_{2}^{2} \cdot \left\| \mathbb{E} \left[\operatorname{vec}(\boldsymbol{X}_{id}) \right] \right\|_{2}^{2}. \end{split}$$

In the first inequality, we use the fact that for an arbitrary matrix M, $M(I + M^T M)^{-1}M^T \prec I$.¹ Similarly,

$$\left\| \left(\boldsymbol{\Sigma}_{\Delta \hat{\Xi}, e}^{L} \right)^{-\frac{1}{2}} \mathbb{E} \boldsymbol{b}_{k, L} \right\|_{2}^{2} \leq \| \boldsymbol{\Sigma}_{e}^{-1/2} \|_{2}^{2} \cdot \| \boldsymbol{C} \boldsymbol{\Phi}^{p} \|_{2}^{2} \cdot \| \mathbb{E} \hat{\boldsymbol{x}}_{k-p, L} \|_{2}^{2}$$

Collecting these inequalities together and using the condition (40), we have

$$\lambda_{\boldsymbol{r}_{k,L}} = \|\mathbb{E}\tilde{\boldsymbol{r}}_{k,L}\|_{2}^{2} < \|\boldsymbol{\Sigma}_{e}^{-1/2}\|_{2}^{2} \cdot \|C\Phi^{p}\|_{2}^{2} \cdot \left(\|\mathbb{E}[\operatorname{vec}(\boldsymbol{X}_{id})]\|_{2} + \|\mathbb{E}\hat{\boldsymbol{x}}_{k-p,L}\|_{2}\right)^{2} < \zeta \cdot L\ell,$$

for $0 < \zeta \ll 1$.

¹By Schur complement,
$$M(I+M^TM)^{-1}M^T \prec I \Leftrightarrow \begin{bmatrix} I & M \\ M^T & I+M^TM \end{bmatrix} \succ 0 \Leftrightarrow I+M^TM-M^TM = I \succ 0.$$



Fig. 4. Test statistics of the four different data-driven fault detection scheme. Upper: F_{rb} (solid) and F_{no} (dash-dotted). Middle: P_{ss} (solid). Lower: P_{mp} (solid). Horizonal dashed lines: the statistical detection threshold determined by the FAR $\alpha = 0.5\%$.