

# Dynamic System Fault Diagnosis under Sparseness Assumption

Qinghua Zhang

► To cite this version:

Qinghua Zhang. Dynamic System Fault Diagnosis under Sparseness Assumption. IEEE Transactions on Signal Processing, Institute of Electrical and Electronics Engineers, 2021, pp.10. 10.1109/TSP.2021.3072004 . hal-03195954

HAL Id: hal-03195954

<https://hal.inria.fr/hal-03195954>

Submitted on 12 Apr 2021

**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.

# Dynamic System Fault Diagnosis under Sparseness Assumption

Qinghua Zhang\*

\*Univ. Gustave Eiffel, Inria, 35042 Rennes, France. Qinghua.Zhang@inria.fr

This paper has supplementary material available at <https://www.mathworks.com/matlabcentral/fileexchange/89847-sparse-fault-diagnosis>, provided by the author. The material includes **Matlab files** illustrating the sparse fault diagnosis approach presented in this paper. This material is 162KB in size.

**Abstract**—Dynamic system fault diagnosis is often faced with a large number of possible faults. The purpose of this paper is to propose an efficient method for such situations. To avoid intractable combinatorial problems, sparse estimation techniques appear to be a powerful tool for isolating faults, under the assumption that only a small number of possible faults can be simultaneously active. However, sparse estimation is often studied in the framework of *linear algebraic equations*, whereas model-based fault diagnosis is usually investigated for *dynamic systems modeled with state equations involving internal states*. The main contribution of this paper is a link between these two formalisms through efficient and reliable algorithms, mainly relying on advanced analyses of residuals generated with the Kalman and Kitanidis filters. Based on these results, it becomes straightforward to solve fault diagnosis problems by applying well known sparse estimation techniques, in the framework of general time varying state-space systems involving unknown inputs.

**Index Terms**— fault detection and isolation, fault diagnosis, sparse estimation, Kalman filter, unknown inputs, residual analysis.

## I. INTRODUCTION

Due to increasing demands for reliable engineering systems and in response to the trend towards more strict environmental norms, researches on fault diagnosis theory have been continuously developing since several decades, as demonstrated by the vast literature on this topic, for instance, [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. Model-based fault diagnosis typically consists of a residual generation step and a residual evaluation step. Residuals are usually designed to stay close to zero when the monitored system is healthy, and to deviate significantly from zero in faulty situations. Depending on the assumptions about modeling and measurement uncertainties, residual evaluation can be based on statistics [1] and on set membership reasoning theories [11], [12].

Fault detection is a non trivial task due to modeling and measurement uncertainties, and fault diagnosis is even more tricky. In the statistical approach, hypothesis testing is applied to fault diagnosis. However, because different faults are *not independent* in general, hypotheses should be considered for subsets of possible faults instead of individual faults. Quite often these subsets correspond to a large number of combinations, which are difficultly tractable in practice. For example, consider the case of 30 possible faults, among which up to 3 faults may be simultaneously active, then  $C_{30}^1 + C_{30}^2 + C_{30}^3 = 4525$  possible combinations would have

to be tested. The situation is similar when designing residuals focusing on subsets of possible faults for fault diagnosis. A large number of possible faults is typically encountered with distributed parameter systems. For example, in mechanical structure health monitoring [13], in electrical cable monitoring [14] and in petrochemical infrastructure monitoring [15]. It is important to develop fault diagnosis methods capable of addressing large scale problems.

Because of this complexity issue, existing fault diagnosis methods are effective only when a small number of possible faults are assumed. To address a large number of possible faults while avoiding intractable combinatorial problems, these methods would have to resort to the following remedies, with serious limitations and drawbacks:

- Assume that only one of the possible faults can be active (known as *single fault assumption*), see e.g., [16], [17], so that each possible fault can be detected individually. Though it is unlikely that many components of a system fail simultaneously, yet the single fault assumption is too restrictive.
- In a parametric setting, estimate simultaneously all the fault parameters and apply a threshold to them in order to isolate actually active faults [6]. The estimation of a large number of parameters is often an ill-posed problem (typically involving the inversion of some large singular matrix). When this estimation is numerically feasible, fault isolation by thresholding the estimated parameters is not reliable due to high uncertainties in the large set of estimated parameters.
- Design statistical tests focusing on individual faults following the minmax approach. When there is a large number of possible faults, such tests require the inversion of large matrices, typically singular, like in the previous case relying on large parametric estimation. If luckily no singular matrix is involved, such minmax tests lack sensitivity, because each of them rejects a large number of faults with the least favorable hypothesis in terms of test power [1, Section 4.2.8].

*The purpose of this paper* is to propose a method to efficiently address a large number of possible faults, relaxing the limitations of the existing methods and overcoming their drawbacks summarized above. Sparse estimation [15], [16], [17] appears an efficient tool for the diagnosis of a large number of possible faults under the sparseness assumption

(only a small number of faults can be active simultaneously). Such techniques can solve simultaneously for a large number of unknowns by assuming that most of them are equal to zero. By appropriately modeling the severity of each possible fault with a parameter value, in particular with the zero value representing an inactive fault, sparse estimation will be able to isolate and to estimate non-zero parameters representing active faults. A related topic is signal sparse representation by combining a small number of elements selected from a large dictionary. While the same algorithms may be used for sparse representation and for sparse estimation, the former aims to obtain a compact representation of a given signal [18], [19], [20], whereas the latter is for the purpose of estimating model parameters under the assumption that most of them are zeros [21], [22], [23].

However, efficient sparse estimation algorithms are usually designed in the framework of *linear algebraic equations* involving a large number of unknowns, whereas fault diagnosis problems are typically formulated for *dynamic* systems modeled with *state equations* involving *internal states*. This formalism incompatibility represents a major difficulty for the application of sparse estimation techniques to fault diagnosis. *The main contribution of this paper* is to establish a link between these two formalisms for efficient fault diagnosis, in the framework of general time varying state-space systems involving unknown inputs.

Applications of sparse representation techniques have been reported in the fault detection and diagnosis literature. A Bayesian compressive sensing approach to fault diagnosis was proposed in [24]. Feature extractions based on sparse signal representation for fault classification have been investigated in [25], [26], [27], [28]. Impulse response monitoring with convolution sparse representations was studied in [29]. Sparse principal component analysis was applied in [30] for data dimension reduction. Sparse based reconstruction was applied to compensate data loss due to sensor failures in [31]. These reported works all avoid state-space representation of the monitored system, and their results are either strongly relying on massively available data or with limited physical interpretation. On the other hand, state-space systems are considered in [32], [33]. In these works, *hidden state variables* are eliminated through numerical differentiation of sensor signals. This approach cannot address unknown inputs.

In this paper, sparse estimation techniques will be efficiently applied to state-space system fault diagnosis *without* resorting to numerical differentiation. The main results will be based on advanced analyses of residuals generated with the Kalman filter, and more generally, with the Kitaniadis filter for systems involving unknown inputs, *connecting residuals to fault parameters through simple algebraic equations*. Based on the main results presented in Propositions 1 and 2, it will become straightforward to solve fault diagnosis problems in general time varying state-space systems involving unknown inputs, by applying well known sparse estimation techniques.

The remaining part of this paper is organized as follows. Section II states the considered fault diagnosis problem. Section III contains the main results for analyzing the considered

fault diagnosis residuals, establishing the link with sparse estimation. Section IV is about fault diagnosis with standard sparse estimation tools. Section V presents numerical examples. Finally Section VI concludes the paper.

## II. PROBLEM STATEMENT

Physical laws governing the behavior of engineering systems are typically described by differential equations [34]. Formulated in state-space form (vectorial first order differential equations), such equations are often discretized in time in order to be easily manipulated in digital computers together with sampled sensor data, resulting in difference equations. This paper will consider linear time varying (LTV) stochastic systems in the form of discrete time state-space equations:

$$x_{k+1} = A_k x_k + B_k u_k + E_k d_k + w_k + \Phi_k \theta \quad (1a)$$

$$y_k = C_k x_k + v_k, \quad (1b)$$

where  $x_k \in \mathbb{R}^n$  is the state,  $y_k \in \mathbb{R}^m$  the output,  $u_k \in \mathbb{R}^l$  the known input,  $d_k \in \mathbb{R}^s$  some unknown input,  $w_k \in \mathbb{R}^n$  the state noise of covariance  $Q_k$ ,  $v_k \in \mathbb{R}^m$  the output noise of covariance  $R_k$ ,  $\theta \in \mathbb{R}^p$  the parameter vector characterizing possible faults affecting the system, and  $A_k, B_k, C_k, E_k, \Phi_k$  are known matrix sequences of appropriate sizes for  $k = 0, 1, 2, \dots$ . The internal state vector  $x_k$  is not directly accessible, but partially observed through the sensor signal vector  $y_k$ .

The *unknown input*  $d_k$  is a totally arbitrary and unknown vector sequence. Typically, such unknown inputs represent disturbances or some non parametric faults to be rejected. See, e.g., [3].

The considered fault diagnosis problem consists in isolating non-zero components of  $\theta$  based on input-output data analysis.

Fault diagnosis problems as formulated above with the state-space model (1) have been widely investigated in the literature [1], [2], [3], [5], [7]. Sometimes restricted to time invariant systems (with constant matrices instead of the matrix sequences  $A_k, B_k$ , etc.), this problem formulation is a well-known general framework for researches on fault diagnosis. Effectively tractable solutions are only available when a small number of possible faults are assumed, *i.e.*, when  $p$  is a small number.

*The particularity* of the present paper is to consider the case of many possible faults ( $\theta$  is a large dimensional vector) under the assumption that only a small number of faults can be active simultaneously ( $\theta$  has few non-zero components).

*Assumptions:*

- (i)  $A_k, B_k, C_k, E_k$  and  $\Phi_k$  are bounded matrix sequences for all  $k \geq 0$ .
- (ii) The initial state  $x_0 \in \mathbb{R}^n$  is a random vector following the Gaussian distribution  $\mathcal{N}(\bar{x}_0, \bar{P}_0)$ , with a mean vector  $\bar{x}_0$  and a positive definite covariance matrix  $\bar{P}_0$ .
- (iii)  $w_k$  and  $v_k$  are zero mean white Gaussian noises independent of each other and of  $x_0$ , with bounded covariance matrices  $Q_k$  and  $R_k$  for all  $k \geq 0$ . The inverse matrix  $R_k^{-1}$  is also bounded for all  $k \geq 0$ .

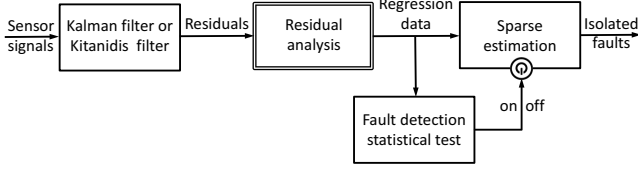


Fig. 1. Fault diagnosis flow chart. This paper is focused on residual analysis.

- (iv) For all  $k \geq 0$ , the matrix product  $C_{k+1}E_k$  has a full column rank and a bounded Moore-Penrose inverse.

These assumptions are typically made for classical Kalman filter implementations [35], [36], except part of Assumption (i) about  $E_k$  and Assumption (iv), both for the implementation of the Kitanidis filter [37], which will be recalled later in this paper.

The flow char of the proposed fault diagnosis approach is illustrated in Figure 1. Sensor data are processed by the Kalman filter or by the Kitanidis filter generating residuals, as recalled in Sections III-A and III-B. Then the residuals are analyzed in order to form the data of a standard linear regression problem. Fault detection can be made by simply thresholding residuals. For more reliable detections, in the stochastic framework formulated in this paper, the well known *generalized likelihood ratio test* [1, Section 4.2.7] is applied to the analyzed residuals. When a fault is detected by the fault detection module, the sparse estimation module processes the regression data in order to isolate active faults among a large set of possible faults.

This paper is focused on residual analysis, yielding regression data enabling the application of standard sparse estimation algorithms.

### III. RESIDUAL GENERATION AND ANALYSIS

As most readers are more familiar with the Kalman filter than with the Kitanidis filter, in order to convey the essential idea behind the results of this paper, let us first consider a simpler case to which the Kalman filter can be applied, before extending the result to the general case requiring the Kitanidis filter.

#### A. LTV system without unknown input

In this reduced case, consider the following system model resulting from a simplification of system (1) by omitting the unknown input term:

$$x_{k+1} = A_k x_k + B_k u_k + w_k + \Phi_k \theta \quad (2a)$$

$$y_k = C_k x_k + v_k. \quad (2b)$$

The considered fault diagnosis problem is as previously formulated for system (1), but omitting the unknown input term.

The purpose here is to transform the fault diagnosis problem to a standard sparse estimation problem, by building, *from the available I-O data and the system model* (2), a matrix  $X$  and

a vector  $y$  of appropriate sizes, such that the *standard sparse regression*

$$y = X\theta + e \quad (3)$$

can be solved under the assumption that the unknown vector  $\theta$  has few non-zero components, by minimizing in some sense the error term  $e$  accounting for measurement noise and modeling uncertainties.

Due to the linearity of the considered system model (2), the output  $y_k$  results from the sum of linear contributions from the input  $u_k$ , the (possibly present) fault term  $\Phi_k \theta$ , the noises  $w_k, v_k$ , and the initial state. If the contribution from the fault term  $\Phi_k \theta$  can be isolated from the others, then intuitively it is possible to build a matrix  $X$  and a vector  $y$  satisfying (3) with the error term  $e$  depending on  $w_k$  and  $v_k$ . However, the dependence of  $y_k$  on  $\theta$  through the *recursive* state equation (2a) makes this task non trivial. The computation of  $X$  and  $y$  from available information should be made *reliably and efficiently*, avoiding unstable recursions. These requirements may appear too demanding, but they can indeed be satisfied, thanks to the Kalman filter.

#### (A-1) The Kalman filter and its residuals

The Kalman filter cannot be applied to system (2) in the usual way for state estimation, because of the unknown fault parameter vector  $\theta$ . A Kalman filter will be designed under the fault-free assumption, *i.e.*, assuming  $\theta = 0$ . This Kalman filter will *always* be applied to system (2), no matter if  $\theta = 0$  or  $\theta \neq 0$  in reality. In the latter case, the Kalman filter will produce a biased residual because of the wrong fault-free assumption. The analysis of this residual bias will lead to a matrix  $X$  and a vector  $y$  satisfying (3), with an error term  $e$  resulting from linearly filtered noises  $w_k$  and  $v_k$ .

The well known Kalman filter, designed for system (2) under the fault-free assumption ( $\theta = 0$ ) consists of the following recursive computations for  $k = 0, 1, 2, \dots$ , after the initialization  $P_{0|0} = \bar{P}_0$  and  $\hat{x}_{0|0} = \bar{x}_0$ :

$$P_{k+1|k} = A_k P_{k|k} A_k^T + Q_k \quad (4a)$$

$$\Sigma_{k+1} = C_{k+1} P_{k+1|k} C_{k+1}^T + R_{k+1} \quad (4b)$$

$$K_{k+1} = P_{k+1|k} C_{k+1}^T \Sigma_{k+1}^{-1} \quad (4c)$$

$$P_{k+1|k+1} = (I_n - K_{k+1} C_{k+1}) P_{k+1|k} \quad (4d)$$

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + B_k u_k \quad (4e)$$

$$\tilde{y}_{k+1|k} = y_{k+1} - C_{k+1} \hat{x}_{k+1|k} \quad (4f)$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} \tilde{y}_{k+1|k}. \quad (4g)$$

In this section, the residual analyzed for fault diagnosis will be the updated output error

$$\tilde{y}_{k|k} \triangleq y_k - C_k \hat{x}_{k|k}, \quad k = 0, 1, 2, \dots \quad (5)$$

Another natural choice would be the innovation sequence  $\tilde{y}_{k+1|k}$ , also known as prediction error. However, the innovation  $\tilde{y}_{k+1|k}$  is not available for the general problem formulated in (1), because it is impossible to make prediction for the system involving an unknown input term  $E_k d_k$ . For this reason, the updated output error  $\tilde{y}_{k|k}$  is considered below in order to preserve the possibility of extending the results to the

general problem formulated in (1). Nevertheless, it is worth mentioning the innovation residual analysis made in [38] in the context of statistical tests for fault diagnosis. Of course, it cannot be extended to systems involving unknown inputs.

(A-2) *Updated output error residual*

The following result, directly relating the residual  $\tilde{y}_{k|k}$  to the fault parameter vector  $\theta$ , will be the basis for building a matrix  $X$  and a vector  $y$  satisfying (3).

*Proposition 1:* The residual  $\tilde{y}_{k|k}$  defined in (5) and generated by the Kalman filter (4), which is designed under the fault-free assumption ( $\theta = 0$ ), but applied to the possibly faulty system (2), is related to the fault parameter vector  $\theta$  through

$$\tilde{y}_{k|k} = C_k \Upsilon_k \theta + \tilde{y}_{k|k}^0, \quad (6)$$

where  $\Upsilon_k \in \mathbb{R}^{n \times p}$  is a matrix sequence recursively generated by

$$\Upsilon_0 = 0 \quad (7a)$$

$$\begin{aligned} \Upsilon_{k+1} &= (I_n - K_{k+1} C_{k+1}) A_k \Upsilon_k \\ &\quad + (I_n - K_{k+1} C_{k+1}) \Phi_k, \end{aligned} \quad (7b)$$

with  $K_k$  as in the Kalman filter (4),  $A_k, C_k, \Phi_k$  as in system (2), and  $\tilde{y}_{k|k}^0$  corresponds to the residual that the Kalman filter *would* generate if it *was* applied to the *fault-free* system characterized by  $\theta = 0$ .  $\square$

*Proof of Proposition 1.:* The main steps of the proof are as follows. First the state estimation error dynamics equation (10) will be derived by combining the Kalman filter (4) and the system model (2). This is a classical step in residual analysis. Then the key step will be the linear transformation (12) to decouple the state estimation error  $\tilde{x}_{k|k}$  and the parameter vector  $\theta$ , yielding the dynamics of the decoupled state estimation error  $\eta_k$  in (17). The decoupling transformation will be used again to obtain the direct relationship between the residual  $\tilde{y}_{k|k}$  and  $\theta$  in (20). The last step will show that the last term in (20) is equal to the residual that the Kalman filter would generate if it was applied to the fault-free system characterized by  $\theta = 0$ .

Combining (4e), (4f) and (4g) yields the recursive equation of  $\hat{x}_{k|k}$ :

$$\begin{aligned} \hat{x}_{k+1|k+1} &= A_k \hat{x}_{k|k} + B_k u_k \\ &\quad + K_{k+1} (y_{k+1} - C_{k+1} A_k \hat{x}_{k|k} - C_{k+1} B_k u_k). \end{aligned} \quad (8)$$

Define the updated state estimation error

$$\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k} \quad (9)$$

then it follows from (2) and (8) that

$$\begin{aligned} \tilde{x}_{k+1|k+1} &= (I_n - K_{k+1} C_{k+1}) A_k \tilde{x}_{k|k} + e_k \\ &\quad + (I_n - K_{k+1} C_{k+1}) \Phi_k \theta, \end{aligned} \quad (10)$$

where

$$e_k \triangleq (I_n - K_{k+1} C_{k+1}) w_k - K_{k+1} v_{k+1}. \quad (11)$$

Somehow inspired by the adaptive observer design approach in [39], define the decoupling transformation

$$\eta_k \triangleq \tilde{x}_{k|k} - \Upsilon_k \theta, \quad (12)$$

with  $\Upsilon_k$  as defined in (7). This  $\Upsilon_k$  behaves like a Jacobian of  $\tilde{x}_{k|k}$  with respect to  $\theta$ , hence  $\eta_k = \tilde{x}_{k|k} - \Upsilon_k \theta$  is the part of  $\tilde{x}_{k|k}$  unaffected by  $\theta$ . Indeed,

$$\begin{aligned} \eta_{k+1} &= (I_n - K_{k+1} C_{k+1}) A_k \tilde{x}_{k|k} + e_k \\ &\quad + (I_n - K_{k+1} C_{k+1}) \Phi_k \theta - \Upsilon_{k+1} \theta \end{aligned} \quad (13)$$

$$\begin{aligned} &= (I_n - K_{k+1} C_{k+1}) A_k (\eta_k + \Upsilon_k \theta) + e_k \\ &\quad + (I_n - K_{k+1} C_{k+1}) \Phi_k \theta - \Upsilon_{k+1} \theta \end{aligned} \quad (14)$$

$$\begin{aligned} &= (I_n - K_{k+1} C_{k+1}) A_k \eta_k + e_k \\ &\quad + [(I_n - K_{k+1} C_{k+1}) A_k \Upsilon_k \\ &\quad \quad + (I_n - K_{k+1} C_{k+1}) \Phi_k - \Upsilon_{k+1}] \theta \end{aligned} \quad (15)$$

$$= (I_n - K_{k+1} C_{k+1}) A_k \eta_k + e_k, \quad (16)$$

where the last equality is due to (7b) annihilating the content of the brackets  $[\cdot \cdot \cdot]$ .

The above successive equations are then summarized as

$$\eta_{k+1} = (I_n - K_{k+1} C_{k+1}) A_k \eta_k + e_k. \quad (17)$$

Combine (2b), (5) and (9) to obtain

$$\tilde{y}_{k|k} = C_k \tilde{x}_{k|k} + v_k. \quad (18)$$

A simple substitution with (12) then leads to

$$\tilde{y}_{k|k} = C_k (\eta_k + \Upsilon_k \theta) + v_k \quad (19)$$

$$= C_k \Upsilon_k \theta + \tilde{y}_{k|k}^0 \quad (20)$$

with

$$\tilde{y}_{k|k}^0 \triangleq C_k \eta_k + v_k. \quad (21)$$

The claimed equality (6) is then obtained in (20). It remains to confirm that the above defined  $\tilde{y}_{k|k}^0$  is equal to the residual that the Kalman filter *would* generate if it *was* applied to the fault-free system, as stated in Proposition 1.

This  $\tilde{y}_{k|k}^0$  as defined in (21) depends on  $C_k, \eta_k$  and  $v_k$ , with  $\eta_k$  satisfying (17). The last term of (17) is defined in (11) from  $K_{k+1}, C_{k+1}, w_k$  and  $v_{k+1}$ . With (17),  $\eta_k$  is recursively determined by the sequences  $K_k, C_k, A_k, e_k$ , and by

$$\eta_0 = \tilde{x}_{0|0} - \Upsilon_0 \theta = \tilde{x}_{0|0} - 0 \theta = \tilde{x}_{0|0} = x_0 - \hat{x}_{0|0} \quad (22)$$

following (12) and (7a). Hence  $\eta_k$  is independent of  $\theta$ , so is  $\tilde{y}_{k|k}^0$  defined in (21). In other words, if the system was “replayed” under the same conditions except the value of  $\theta$ , then  $\tilde{y}_{k|k}^0$  would remain unchanged.

When  $\theta = 0$ , (20) becomes  $\tilde{y}_{k|k} = \tilde{y}_{k|k}^0$ . In this particular case ( $\theta = 0$ ),  $\tilde{y}_{k|k}$  is the residual of the Kalman filter applied to the fault-free system, so is  $\tilde{y}_{k|k}^0$ , which is independent of  $\theta$ . Therefore, regardless of the value of  $\theta$ ,  $\tilde{y}_{k|k}^0$  is equal to the residual that the Kalman filter *would* generate if it *was* applied to the fault-free system.

The proof of Proposition 1 is then completed.  $\square$

### B. Residual analysis rejecting unknown inputs

Now let us go back to the system formulated in (1), which is copied below for ease of presentation and reading:

$$x_{k+1} = A_k x_k + B_k u_k + E_k d_k + w_k + \Phi_k \theta \quad (23a)$$

$$y_k = C_k x_k + v_k. \quad (23b)$$

The Kalman filter cannot be applied to this system, because of the involved unknown input term  $E_k d_k$ . Instead, the Kitanidis filter [37] will be used in what follows. In the fault-free case ( $\theta = 0$ ), the Kitanidis filter minimizes the state estimation error covariance while rejecting the unknown input. It is thus a natural extension of the Kalman filter to systems involving unknown inputs.

Like the Kalman filter in Section III-A, here the Kitanidis filter is designed under the fault-free assumption ( $\theta = 0$ ), and will be applied to the possibly faulty system (23), no matter if  $\theta = 0$  or  $\theta \neq 0$  in reality. The residual generated by this Kitanidis filter will be analyzed in order to establish a link with sparse regression.

The Kitanidis filter [37] assuming  $\theta = 0$  is as follows:

$$\bar{P}_{k+1|k} = A_k \bar{P}_{k|k} A_k^T + Q_k \quad (24a)$$

$$\bar{\Sigma}_{k+1} = C_{k+1} \bar{P}_{k+1|k} C_{k+1}^T + R_{k+1} \quad (24b)$$

$$\Gamma_{k+1} = E_k - \bar{P}_{k+1|k} C_{k+1}^T \bar{\Sigma}_{k+1}^{-1} C_{k+1} E_k \quad (24c)$$

$$\Xi_{k+1} = E_k^T C_{k+1}^T \bar{\Sigma}_{k+1}^{-1} C_{k+1} E_k \quad (24d)$$

$$\Lambda_{k+1} = \Gamma_{k+1} \Xi_{k+1}^{-1} \quad (24e)$$

$$\begin{aligned} \bar{P}_{k+1|k+1} &= \bar{P}_{k+1|k} - \bar{P}_{k+1|k} C_{k+1}^T \bar{\Sigma}_{k+1}^{-1} C_{k+1} \bar{P}_{k+1|k} \\ &\quad + \Lambda_{k+1} \Xi_{k+1} \Lambda_{k+1}^T \end{aligned} \quad (24f)$$

$$\begin{aligned} \bar{K}_{k+1} &= \bar{P}_{k+1|k} C_{k+1}^T \bar{\Sigma}_{k+1}^{-1} \\ &\quad + \Gamma_{k+1} \Xi_{k+1}^{-1} E_k^T C_{k+1}^T \bar{\Sigma}_{k+1}^{-1} \end{aligned} \quad (24g)$$

$$\begin{aligned} \hat{x}_{k+1} &= A_k \hat{x}_k + B_k u_k \\ &\quad + \bar{K}_{k+1} (y_{k+1} - C_{k+1} A_k \hat{x}_k - C_{k+1} B_k u_k). \end{aligned} \quad (24h)$$

In the Kalman filter (4), state estimation is decomposed into a prediction step and an update step, yielding respectively the predicted state estimate  $\hat{x}_{k+1|k}$  and the updated state estimate  $\hat{x}_{k+1|k+1}$ . In the case of the Kitanidis filter, however, state prediction is impossible, because of the involved unknown input. A single state estimate  $\hat{x}_{k+1}$  is computed with the recursive equation (24h), corresponding to the Kalman filter recursion (8) combining the prediction step (4e) and the update step (4g).

While the double index notations  $\hat{x}_{k+1|k}$  and  $\hat{x}_{k+1|k+1}$  are reserved for the Kalman filter, the single index notation  $\hat{x}_k$  is used for the Kitanidis filter.

*Proposition 2:* Let

$$\tilde{y}_k \triangleq y_k - C_k \hat{x}_k, \quad k = 0, 1, 2, \dots, \quad (25)$$

be the residual generated with the Kitanidis filter (24), designed under the fault-free assumption ( $\theta = 0$ ), but applied to the possibly faulty system (23). Then  $\tilde{y}_k$  is related to the fault parameter vector  $\theta$  through

$$\tilde{y}_k = C_k \tilde{\Upsilon}_k \theta + \tilde{y}_k^0, \quad (26)$$

where  $\tilde{\Upsilon}_k \in \mathbb{R}^{n \times p}$  is a matrix sequence generated by

$$\tilde{\Upsilon}_0 = 0 \quad (27a)$$

$$\begin{aligned} \tilde{\Upsilon}_{k+1} &= (I_n - \bar{K}_{k+1} C_{k+1}) A_k \tilde{\Upsilon}_k \\ &\quad + (I_n - \bar{K}_{k+1} C_{k+1}) \Phi_k \end{aligned} \quad (27b)$$

with  $\bar{K}_k$  as in the Kitanidis filter (24),  $A_k, C_k, \Phi_k$  as in system (23), and  $\tilde{y}_k^0$  corresponds to the residual that the Kitanidis filter (24) would generate if it was applied to the fault-free system characterized by  $\theta = 0$ .  $\square$

*Proof of Proposition 2.:* Define

$$\tilde{x}_k \triangleq x_k - \hat{x}_k. \quad (28)$$

Then, combining (23) and (24h) yields

$$\begin{aligned} \tilde{x}_{k+1} &= (I_n - \bar{K}_{k+1} C_{k+1}) A_k \tilde{x}_k \\ &\quad + (I_n - \bar{K}_{k+1} C_{k+1}) E_k d_k + (I_n - \bar{K}_{k+1} C_{k+1}) \Phi_k \theta \\ &\quad + (I_n - \bar{K}_{k+1} C_{k+1}) w_k - \bar{K}_{k+1} v_{k+1}. \end{aligned} \quad (29)$$

It turns out that the Kitanidis filter gain  $\bar{K}_{k+1}$  satisfies

$$(I_n - \bar{K}_{k+1} C_{k+1}) E_k = 0. \quad (30)$$

This equality can be checked as follows. It follows from (24d) and (24g) that

$$\begin{aligned} \bar{K}_{k+1} C_{k+1} E_k &= \bar{P}_{k+1|k} C_{k+1}^T \bar{\Sigma}_{k+1}^{-1} C_{k+1} E_k \\ &\quad + \Gamma_{k+1} \Xi_{k+1}^{-1} \Xi_{k+1} \end{aligned} \quad (31)$$

$$= \bar{P}_{k+1|k} C_{k+1}^T \bar{\Sigma}_{k+1}^{-1} C_{k+1} E_k + \Gamma_{k+1} \quad (32)$$

$$= E_k, \quad (33)$$

where the last equality is due to (24c). Then, (30) is confirmed by a simple rearrangement.

The important equality (30) cancels the term involving the unknown input  $d_k$  out of the error dynamics equation (29), which then becomes

$$\begin{aligned} \tilde{x}_{k+1} &= (I_n - \bar{K}_{k+1} C_{k+1}) A_k \tilde{x}_k + \bar{e}_k \\ &\quad + (I_n - \bar{K}_{k+1} C_{k+1}) \Phi_k \theta \end{aligned} \quad (34)$$

with

$$\bar{e}_k \triangleq (I_n - \bar{K}_{k+1} C_{k+1}) w_k - \bar{K}_{k+1} v_{k+1}. \quad (35)$$

The recursive equation (34) satisfied by  $\tilde{x}_k$  is the same as (10) satisfied by  $\tilde{x}_{k|k}$ , except that here  $K_k$  is replaced by  $\bar{K}_k$ . The same similarity is also observed between (11) and (35). By copying the steps of the proof of Proposition 1 after equation (11) with obvious notation adaptations, Proposition 2 is then proved.  $\square$

### C. Residual generator boundedness

The residuals generated by the Kalman filter and by the Kitanidis filter are recursively computed. In practice, it is important to ensure that the numerical computations are bounded. The boundedness results of the Kalman filter and of the Kitanidis filter have been established in the literature essentially based on observability and controllability assumptions, as shortly recalled below.

For the state-space system (2), the Kalman filter is usually analyzed under the conditions that the matrix sequence pair  $[A_k, C_k]$  is uniformly completely observable, and the matrix sequence pair  $[A_k, Q_k^{\frac{1}{2}}]$  is uniformly completely controllable [40], [36], where  $Q_k^{\frac{1}{2}}$  is a symmetric matrix square root of the state noise covariance  $Q_k$ . Note that the controllability condition *does not* concern the input  $u_k$ , but the state noise  $w_k$ . See [40], [36] for the definitions of uniform complete observability and controllability. A state-space model is both observable and controllable if and only if it has the minimum number of states [41]. In other words, if such a model is not observable or not controllable, it can be reduced to an equivalent model which is observable and controllable.

Under the observability and controllability conditions, the Kalman filter is stable, implying that its error covariance  $P_{k|k}$  and its gain matrix  $K_k$  are bounded [40], [36], so is the covariance of the fault-free residual  $\tilde{y}_{k|k}^0$ . As expressed in (6), the residual  $\tilde{y}_{k|k}$  has an extra term  $C_k \Upsilon_k \theta$  compared to the fault-free residual  $\tilde{y}_{k|k}^0$ . The matrix sequence  $\Upsilon_k$  is generated by filtering  $\Phi_k$  through the linear filter (7), which has the same transition matrix  $(I_n - K_{k+1} C_{k+1}) A_k$  as the Kalman filter error dynamics equation (10). According to [36], the Kalman filter error dynamics (10) is stable, so is the linear filter (7). Then the matrix sequence  $\Upsilon_k$  generated by the stable linear filter (7) is bounded, given the boundedness of  $\Phi_k$  assumed in Assumption (i).

For the LTV system (1) subject to unknown inputs, similar results about the Kitanidis filter have been established [42], under a more involved observability condition, ensuring the boundedness of the residual computation with the Kitanidis filter.

#### IV. FAULT DIAGNOSIS BY SPARSE ESTIMATION

By modeling faults as deviations of the parameter vector  $\theta$  from its nominal value, fault detection amounts to detecting changes in the mean value of the residual  $\tilde{y}_{k|k}$  as expressed in (6) or  $\tilde{y}_k$  in (26), respectively based on Proposition 1 or 2. In this work the generalized likelihood ratio test [1, Section 4.2.7] is applied for this purpose.

After the detection of any fault, fault diagnosis is performed through sparse estimation based on the results developed in the previous sections, under the assumption that  $\theta$  contains few non-zero components. The number of non-zero components of  $\theta$  is unknown. Let  $q$  denote this unknown number in order to ease the presentation.

Given an LTV system (1), a Kitanidis filter (24) is designed under the fault-free assumption ( $\theta = 0$ ). The output error  $\tilde{y}_k$  of this Kitanidis filter is then characterized by Proposition 2.

Let  $N$  denote the data sample length. With data collected for  $k = 1, 2, \dots, N$ , a matrix  $\mathbf{X}$  and a vector  $\mathbf{y}$  are built as

$$\mathbf{X} = \begin{bmatrix} C_1 \tilde{\Upsilon}_1 \\ C_2 \tilde{\Upsilon}_2 \\ \vdots \\ C_N \tilde{\Upsilon}_N \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_N \end{bmatrix} \quad (36)$$

where  $C_k \tilde{\Upsilon}_k$  and  $\tilde{y}_k$  are as in Proposition 2. Then, according to (26),

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{e} \quad (37)$$

with the error vector

$$\mathbf{e} = \begin{bmatrix} \tilde{y}_1^0 \\ \tilde{y}_2^0 \\ \vdots \\ \tilde{y}_N^0 \end{bmatrix}. \quad (38)$$

As already discussed in the introduction, existing sparse estimation algorithms can be applied to solve for  $\theta$  under the assumption that  $\theta$  has few non-zero components. The numerical examples in the next section will be based on Lasso [43]. It is also possible to apply other algorithms, such as Stability selection [44], Dantzig selector [45], Orthogonal Matching Pursuit [46] and CoSaMP [47]. The Lasso solution is given by

$$\hat{\theta} = \arg \min_{\theta} \left( \frac{1}{2N} \|\mathbf{X}\theta - \mathbf{y}\|_2^2 + \lambda \|\theta\|_1 \right) \quad (39)$$

where  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote respectively the  $l_1$  norm and  $l_2$  norm, and  $\lambda > 0$  is a weighting coefficient.

The term  $\lambda \|\theta\|_1$  controls the number of non-zero components in the Lasso estimate  $\hat{\theta}$ . A tricky point is how to choose the value of  $\lambda$  in practice. It is possible to choose  $\lambda$  by cross-validation [23] if the available data set is large enough. In this paper the following choice is adopted. The number of non-zero components in the Lasso solution  $\hat{\theta}$  decreases when  $\lambda$  increases, and there is a wide range of  $\lambda$  in which  $\hat{\theta}$  keeps the same number of non-zero components. The value of  $\lambda$  is chosen at the middle of this range, by numerically exploring different values of  $\lambda$ . Numerical examples will be given in the next section.

Like any fault detection and diagnosis problem in a stochastic framework, decision making is subject to random errors. When the data sample size  $N$  tends to infinity, Lasso is able to asymptotically isolate the non-zero components of  $\theta$  in the sparse regression problem (37) under the *irrepresentable condition* [48], [23]. Basically this condition means that the columns of  $\mathbf{X}$  in (37) corresponding to non-zero components of  $\theta$  are not too much correlated with the other columns of  $\mathbf{X}$ . In practice it is difficult to check this condition. In this paper,  $\mathbf{X}$  is filled in (36) with  $C_k \tilde{\Upsilon}_k$ . Each column of the matrix sequence  $\tilde{\Upsilon}_k$  is generated from the corresponding column of  $\Phi_k$  through (27), without interaction between different columns, because the recursive *matrix* equation (27b) is indeed a collection of *vector* equations in parallel, one for each corresponding column of  $\tilde{\Upsilon}_{k+1}$ ,  $\tilde{\Upsilon}_k$  and  $\Phi_k$ . If the columns of  $\Phi_k$  are weakly correlated, so are the columns of  $\tilde{\Upsilon}_k$ , and therefore of  $\mathbf{X}$ .

When the state-space system (2) is considered, the Kitanidis filter is replaced by the Kalman filter, and accordingly, Proposition 2 is replaced by Proposition 1, yielding a similar matrix  $\mathbf{X}$  and a similar vector  $\mathbf{y}$ , in which  $\tilde{\Upsilon}_k$  is replaced by  $\Upsilon_k$  and  $\tilde{y}_k$  by  $\tilde{y}_{k|k}$ .

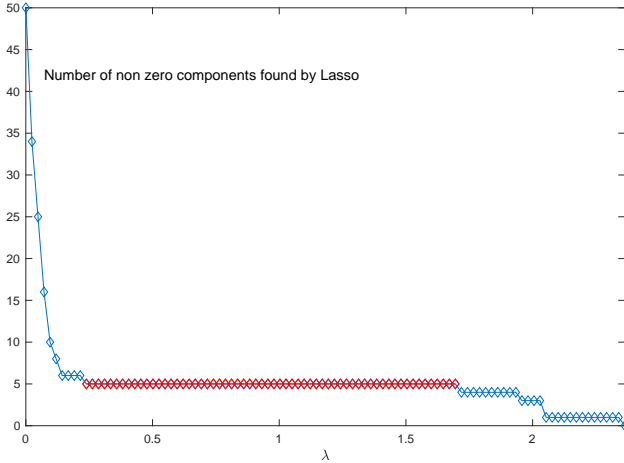


Fig. 2. Number of non-zero components of the Lasso estimate  $\hat{\theta}$  per  $\lambda$ . This number remains 5 in the range represented by red diamonds.

## V. NUMERICAL EXAMPLES

### Example 1 – simulations with asymptotic behavior evaluation

This first example is an LTV system in the form of (1), which, as stated in Section II, is a well-known general framework for researches on fault diagnosis. The considered example is

$$x_{k+1} = A_k x_k + E d_k + w_k + \Phi_k \theta \quad (40a)$$

$$y_k = C x_k + v_k, \quad (40b)$$

with

$$A_k = \begin{bmatrix} -0.8 & 0.7 & 0.1 \exp(-k) \\ -0.6 & 0 & 0.1 \cos(k) \\ 0.1 \sin(k) & -0.5 & -0.4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad (41)$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}. \quad (42)$$

The input term  $B_k u_k = 0$  for all  $k \geq 0$  has been omitted from (40). The *unknown* input

$$d_k = 10 \sin(0.1k) + \epsilon_k + 100,$$

with a zero mean white Gaussian noise  $\epsilon_k$  of unitary variance. The constant 100 is added so that this simulated unknown input is far from a zero mean noise. The matrix  $\Phi_k \in \mathbb{R}^{3 \times 50}$  is randomly generated, with each entry following the uniform distribution within  $[-17.3, 17.3]$ . The two white Gaussian noise vectors  $w_k$  and  $v_k$  have their covariance matrices equal to  $I_3$  and  $0.5I_2$ , respectively.

The fault parameter vector  $\theta$  has 50 components. In each simulation trial, 5 out of the 50 components of  $\theta$  are randomly drawn to be set to 1, while the other components are set to 0.

In every simulation trial, data are simulated for  $k = 1, 2, \dots, 100$ , *i.e.*, the data sample size  $N = 100$ . The Kitanidis filter (24) designed under the fault-free assumption ( $\theta = 0$ ) then processes the simulated data.

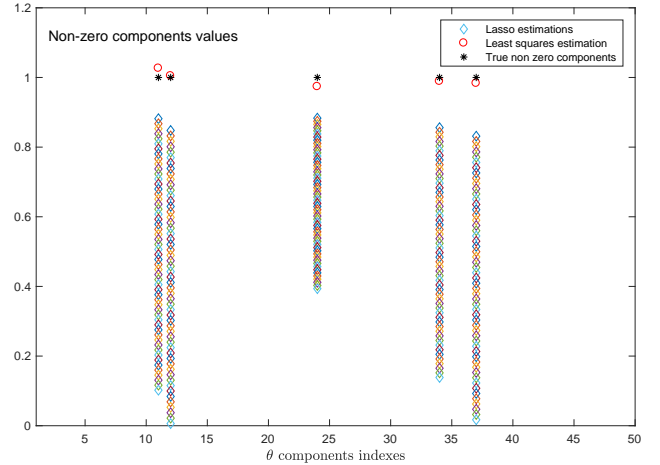


Fig. 3. Lasso estimation and least squares estimation. Diamonds: non-zero components of the Lasso estimate  $\hat{\theta}$ , with each color corresponding to one particular value of  $\lambda$  within the flat range at the middle of Figure 2. Circles: least squares estimate of the non-zero components selected by Lasso. Asterisks: simulated true non-zero components.

The Lasso estimate (39) is applied to (37), based on Proposition 2, with different values of  $\lambda$ . For one of the simulation trials, the number of non-zero components in the Lasso estimate  $\hat{\theta}$  is depicted against the value of  $\lambda$  in Figure 2. These results show a large range of  $\lambda$  (represented by red diamonds) within which the number of non-zero components remains 5. For each value of  $\lambda$  within this range, do the 5 non-zero components of the Lasso estimate  $\hat{\theta}$  correspond to the true non-zero components of  $\theta$ ? The results shown in Figure 3 (the diamonds) confirm that, for all these values of  $\lambda$ , the Lasso estimate yields the same non-zero components, which turn out to be in agreement with the simulated true non-zero components of  $\theta$  (the asterisks) in this example.

The Lasso estimate is biased due to the penalty term  $\lambda \|\theta\|_1$  in (39). After having isolated the non-zero components of  $\theta$  with the value of  $\lambda$  in the middle of the flat range in Figure 2, these non-zero components of  $\theta$  are then estimated by the classical least squares method. The result is illustrated in Figure 3 by circles.

The numerical results presented above correspond to a *single realization* of random noises  $w_k, v_k$ , randomly drawn non-zero components of  $\theta$  and randomly drawn signals in  $\Phi_k$ . To statistically evaluate the behavior of the proposed approach, the simulation is repeated a large number of times. Moreover, these experiments are made for different numbers of non-zero components in  $\theta$ , denoted by  $q$ . The simulation is repeated 1000 times for each value of  $q$ , in order to evaluate the successful fault isolation rate. In each simulation, fault isolation is considered successful if all the non-zero components of  $\theta$  estimated by Lasso correspond to the true simulated non-zero components. The results are reported in Figure 4 and Table I. It is in particular shown that the successful isolation rate is 99.5% for  $q = 5$ , and 81.7% for  $q = 9$ . The sparseness assumption becomes less appropriate when  $q$  increases, therefore the performance of sparse estimation



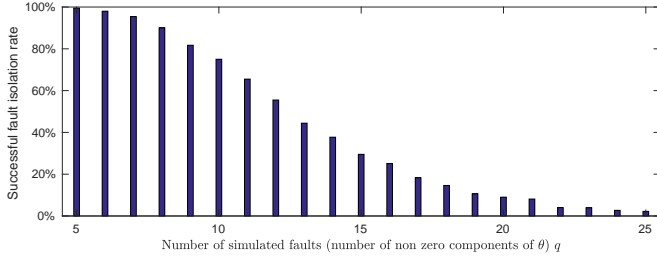


Fig. 4. Example 1. Successful fault isolation rates and the corresponding number of simulated faults (number of non-zero components of  $\theta$ )  $q$ . For each number  $q$  from 5 to 25, the isolation rate is evaluated over 1000 random realizations. For each random realization, fault isolation is successful if every non-zero component of  $\theta$  is correctly isolated (no misdetection, nor over-detection). Data sample length  $N = 100$ . The covariance matrix of the output noise  $v_k$  is  $\text{diag}([0.5, 0.5])$ . The covariance matrix of the state noise  $w_k$  is the  $3 \times 3$  identity matrix. See Table I for numerical values.

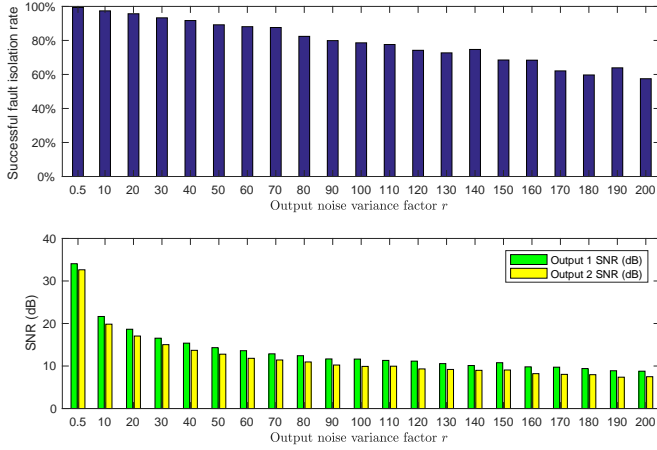


Fig. 5. Example 1. Successful fault isolation rates and the corresponding average signal-to-noise ratio (SNR) levels of the two output sensors. At each SNR level, the isolation rate is evaluated over 1000 random realizations. For each random realization, fault isolation is successful if every non-zero component of  $\theta$  is correctly isolated (no misdetection, nor over-detection). Data sample length  $N = 100$ . Number of simulated faults  $q = 5$ . The covariance matrix of the output noise  $v_k$  is  $rI_2$  ( $I_2$  is the  $2 \times 2$  identity matrix). The covariance matrix of the state noise  $w_k$  is the  $3 \times 3$  identity matrix. See Table II for numerical values.

decreases.

More experiments are then made at different signal-to-noise ratio (SNR) levels. More specifically, in these simulations, the covariance matrix of the sensor noise vector  $v_k$  is set to  $rI_2$ , with  $I_2$  being the  $2 \times 2$  identity matrix and with a scalar factor  $r$  varying from 0.5 to 200. For each value of  $r$ , the simulation is randomly simulated 1000 times. The successful fault isolation rate over the 1000 realizations for each value of  $r$  is illustrated in Figure 5. In each of the 1000 simulations, the SNR for each of the two output sensors is computed, and the average SNR for each sensor over the 1000 simulations for each value of  $r$  is also shown in Figure 5. The numerical values corresponding to Figure 5 are presented in Table II. These results show that, the successful isolation rate is 97.4% when the SNR is about 21 dB, and 82.4% when the SNR is about 12 dB, with the data sample size  $N = 100$ .

### A. Example 2 – leakage diagnosis in a gas transportation pipeline

This example of natural gas transportation pipeline is borrowed from [15]. Leakage detection and localization are a crucial issue for safe operation of pipelines, especially in high pressure natural gas transportation. A major difficulty for leakage monitoring is due to the distributed nature of possible leakages along long pipelines. In this example, sparse estimation will be applied to localize leakages. In [15], a linear parameter varying model of the gas dynamics in a gas transportation pipeline is experimentally established as follows:

$$x_{k+1} = (A_0 + A_p \rho_k)x_k + (B_0 + B_p \rho_k)u_k + w_k \quad (43a)$$

$$y_k = (C_0 + C_p \rho_k)x_k + (D_0 + D_p \rho_k)u_k + v_k, \quad (43b)$$

where  $u_k \in \mathbb{R}$  and  $y_k \in \mathbb{R}$  are respectively the input mass flow and the output mass flow, the state vector  $x_k \in \mathbb{R}^2$  is composed of the internal mass flow and the pressure drop in the pipeline. The scheduling variable  $\rho_k \in \mathbb{R}$  is the average of the pressure measurements. The sampling period is 2 minutes. The details about  $\rho_k$  and parameter values contained in  $A_0, A_p, B_0, B_p, C_0, C_p, D_0, D_p$  are available in [15].

Let

$$A_k = A_0 + A_p \rho_k \quad (44)$$

$$B_k = B_0 + B_p \rho_k \quad (45)$$

$$C_k = C_0 + C_p \rho_k \quad (46)$$

$$D_k = D_0 + D_p \rho_k \quad (47)$$

$$\Phi_k = - \begin{bmatrix} u_{k-1} & u_{k-2} & \cdots & u_{k-p} \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (48)$$

then the state-space model (43) is rewritten as

$$x_{k+1} = A_k x_k + B_k u_k + w_k + \Phi_k \theta \quad (49a)$$

$$y_k = C_k x_k + D_k u_k + v_k \quad (49b)$$

in which the additional term  $\Phi_k \theta$  represents possible leakages proportional to delayed inputs  $u_k$ , with delays related to the distance between the leakage positions and the sensor position. The non-zero components of  $\theta \in \mathbb{R}^p$  correspond to leakages.

The term  $D_k u_k$  was not considered in the previous sections. By moving it to the left hand side of (49b) and by replacing  $y_k$  by  $y'_k \triangleq y_k - D_k u_k$ , then (49b) becomes  $y'_k = C_k x_k + v_k$ .

Two leakage faults are simulated, corresponding to the 10-th and the 15-th components of  $\theta$ , both set to  $10^{-4}$ , while the other components of  $\theta \in \mathbb{R}^{30}$  are zeros. The input signal  $u_k$  and the scheduling variable  $\rho_k$  are as in [15], and the output signal  $y_k$  is simulated with the state noise  $w_k$  of covariance  $[3.31 \ -2.18; -2.18 \ 1.44]$  and the output noise  $v_k$  of variance  $10^4$ . The data sample length  $N = 500$ .

The Lasso estimate (39) is then applied to the innovation sequence of the Kalman filter, based on Proposition 1, with different values of  $\lambda$ . The number of non-zero components in the Lasso estimate  $\hat{\theta}$  is depicted against the value of  $\lambda$  in Figure 6. These results show a large range of  $\lambda$  (represented by red diamonds) within which the number of non-zero components remains 2. For each value of  $\lambda$  within this range, the 2

Number of faults $q$	5	6	7	8	9	10	11	12	13	14	15
Successful fault isolation rate	99.5%	98.0%	95.4%	90.0%	81.7%	75.0%	65.5%	55.5%	44.4%	37.7%	29.5%

Number of faults $q$	16	17	18	19	20	21	22	23	24	25
Successful fault isolation rate	25.1%	18.3%	14.6%	10.6%	9.0%	8.1%	4.0%	3.9%	2.7%	2.2%

TABLE I

Example 1. Successful fault isolation rates and the corresponding number of simulated faults (number of non-zero components of  $\theta$ )  $q$ . For each number  $q$  from 5 to 25, the isolation rate is evaluated over 1000 random realizations. For each random realization, fault isolation is successful if every non-zero component of  $\theta$  is correctly isolated (no misdetection, nor over-detection). Data sample length  $N = 100$ . The covariance matrix of the output noise  $v_k$  is  $\text{diag}([0.5, 0.5])$ . The covariance matrix of the state noise  $w_k$  is the  $3 \times 3$  identity matrix. See Figure 4 for a graphical illustration.

Noise factor $r$	0.5	10	20	30	40	50	60	70	80	90	100
Output 1 SNR (dB)	34.06	21.65	18.64	16.54	15.37	14.32	13.60	12.86	12.42	11.66	11.63
Output 2 SNR (dB)	32.63	19.84	17.05	15.02	13.69	12.78	11.81	11.42	10.95	10.23	9.91
Successful fault isolation rate	99.4%	97.4%	95.7%	93.3%	91.7%	89.2%	88.1%	87.6%	82.4%	79.9%	78.6%

Noise factor $r$	110	120	130	140	150	160	170	180	190	200
Output 1 SNR (dB)	11.32	11.15	10.56	10.11	10.78	9.808	9.727	9.403	8.892	8.784
Output 2 SNR (dB)	9.959	9.320	9.193	8.988	9.065	8.206	8.037	7.965	7.374	7.489
Successful fault isolation rate	77.6%	74.2%	72.7%	74.7%	68.5%	68.4%	62.1%	59.7%	63.9%	57.5%

TABLE II

Example 1. Successful fault isolation rates and the corresponding average signal-to-noise ratio (SNR) levels of the two output sensors. At each SNR level, the isolation rate is evaluated over 1000 random realizations. For each random realization, fault isolation is successful if every non-zero component of  $\theta$  is correctly isolated (no misdetection, nor over-detection). Data sample length  $N = 100$ . Number of simulated faults  $q = 5$ . The covariance matrix of the output noise  $v_k$  is  $rI_2$  ( $I_2$  is the  $2 \times 2$  identity matrix). The covariance matrix of the state noise  $w_k$  is the  $3 \times 3$  identity matrix. See Figure 5 for a graphical illustration.

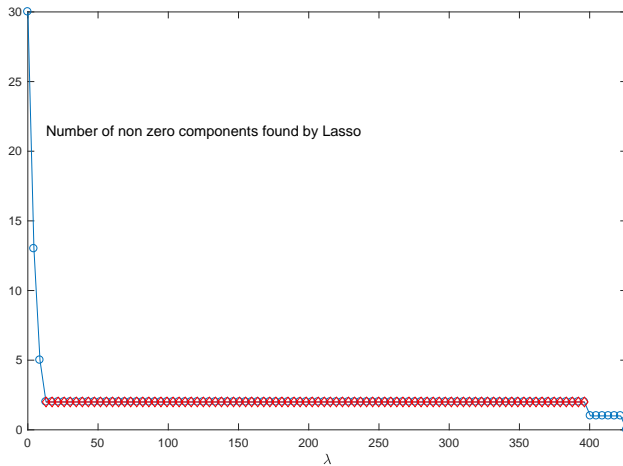


Fig. 6. Example 2 – number of non-zero components of the Lasso estimate  $\hat{\theta}$  per  $\lambda$ . This number remains 2 in the range represented by red diamonds.

non-zero components of the Lasso estimate  $\hat{\theta}$  are represented in Figure 7 by diamonds. For all these values of  $\hat{\theta}$ , the Lasso estimate yields the same non-zero components in agreement with the simulated true non-zero components of  $\theta$  (represented by asterisks).

The Lasso estimate is biased due to the penalty term  $\lambda\|\theta\|_1$  in (39). After having isolated the non-zero components of  $\theta$ , their values are estimated by the classical least squares method, as illustrated in Figure 7 by circles. This result shows that the two simulated leakages are correctly isolated and estimated with a good accuracy.

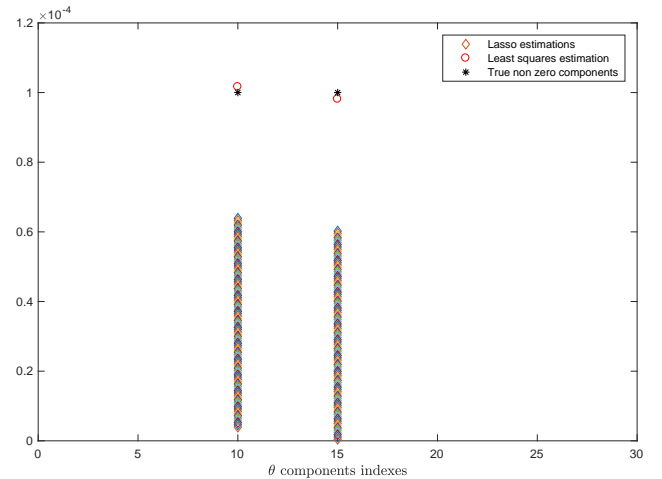


Fig. 7. Example 2 – Diamonds: non-zero components of the Lasso estimate  $\hat{\theta}$ , with each color corresponding to one particular value of  $\lambda$  within the flat range at the middle of Figure 6. Circles: least squares estimate of the non-zero components selected by Lasso. Asterisks: simulated true non-zero components.

## VI. CONCLUSION

To overcome the incompatibility between efficient sparse estimation techniques available in the framework of *linear algebraic equations* and fault diagnosis problems formulated for *dynamic systems modeled with state equations*, a connection between them has been established in this paper based on advanced residual analyses of the Kalman filter and of the Kitaniadis filter. The numerical efficiency and the stability of

the proposed approach rely on the related properties of the involved filters. These results provide an effective solution to fault diagnosis by avoiding large combinatorial problems, in the framework of general time varying state-space systems involving unknown inputs. For future works, applications to mechanical structure vibration analysis will be investigated. Some previous studies on fault diagnosis will also be revisited in order to relax the limitation about fault parameter vectors, such as in [49].

## REFERENCES

- [1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes - Theory and Application*. Englewood Cliffs: Prentice Hall, 1993.
- [2] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. New York: Marcel Dekker, 1998.
- [3] J. Chen and R. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Boston: Kluwer, 1999.
- [4] R. Patton, P. Frank, and R. Clarke, eds., *Issues of Fault Diagnosis for Dynamic Systems*. London: Springer, 2000.
- [5] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control (2nd ed.)*. Berlin: Springer, 2006.
- [6] R. Isermann, *Fault Diagnosis Systems: An Introduction From Fault Detection To Fault Tolerance*. Berlin: Springer, 2006.
- [7] S. Ding, *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Berlin: Springer, 2008.
- [8] C. Volosencu, ed., *Fault Detection and Diagnosis*. IntechOpen, 2018.
- [9] T. Escobet, A. Bregon, B. Pulido, and V. Puig, eds., *Fault Diagnosis of Dynamic Systems*. Springer, 2019.
- [10] S. Huang, K. K. Tan, P. V. Er, and T. H. Lee, *Intelligent Fault Diagnosis and Accommodation Control*. Oxon: CRC Press, 2020.
- [11] W. Tang, Z. Wang, and Y. Shen, "Fault detection and isolation for discrete-time descriptor systems based on  $h_-/h_\infty$  observer and zonotopic residual evaluation," *International Journal of Control*, 2018.
- [12] V. Puig and M. Poursasgar, "Fault diagnosis using set-membership approaches," in *Fault Diagnosis of Dynamic Systems*, Springer, 2019.
- [13] S. Allahdadian, M. Döhler, C. Ventura, and L. Mevel, "Towards robust statistical damage localization via model-based sensitivity clustering," *Mechanical Systems and Signal Processing*, vol. 134, p. 106341, 2019.
- [14] F. Loete, Q. Zhang, and M. Sorine, "Experimental validation of the inverse scattering method for distributed characteristic impedance estimation," *IEEE Trans. on Antennas and Propagation*, vol. 63, no. 6, pp. 2532–2538, 2015.
- [15] P. Lopes dos Santos, T.-P. Azevedo-Perdicóulis, J. Ramos, J. Martins de Carvalho, G. Jank, and J. Milhinhos, "An LPV modeling and identification approach to leakage detection in high pressure natural gas transportation networks," *IEEE Trans. Control Systems Technology*, vol. 19, no. 1, pp. 77–92, 2011.
- [16] S. Mir, M. Lubaszewski, V. Kolarik, and B. Courtois, "Fault-based testing and diagnosis of balanced filters," *Analog Integrated Circuits and Signal Processing*, vol. 11, no. 1, pp. 5–19, 1996.
- [17] D. Dubois, M. Grabisch, O. De Mouzon, and H. Prade, "A possibilistic framework for single-fault causal diagnosis under uncertainty," *International Journal of General Systems*, vol. 30, no. 2, pp. 167–192, 2001.
- [18] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.
- [19] E. C. Marques, N. Maciel, L. Naviner, H. Cai, and J. Yang, "A review of sparse recovery algorithms," *IEEE Access*, vol. 7, pp. 1300–1322, 2019.
- [20] S. Foucart, R. Gribonval, L. Jacques, and H. Rauhut, "Jointly low-rank and bisparsity recovery: Questions and partial answers," *Analysis and Applications*, vol. 18, no. 1, pp. 25–48, 2020.
- [21] K. Knight and W. Fu, "Asymptotics for Lasso-type estimators," *The Annals of Statistics*, vol. 28, no. 28, pp. 1356–1378, 2000.
- [22] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.
- [23] H. Liu and B. Yu, "Asymptotic properties of Lasso+mLS and Lasso+Ridge in sparse high-dimensional linear regression," *Electronic Journal of Statistics*, vol. 7, pp. 3124–3169, 2013.
- [24] G. Oliveri, P. Rocca, and A. Massa, "Reliable diagnosis of large linear arrays – a Bayesian compressive sensing approach," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 10, 2012.
- [25] Y. Wang, J. Xiang, Q. Mo, and S. He, "Compressed sparse time-frequency feature representation via compressive sensing and its applications in fault diagnosis," *Measurement*, vol. 68, pp. 70–81, 2015.
- [26] L. Ren, W. Lv, S. Jiang, and Y. Xiao, "Fault diagnosis using a joint model based on sparse representation and SVM," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 10, pp. 2313 – 2320, 2016.
- [27] Y. Qin, "A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2716–2726, 2018.
- [28] D. Zhong, C. Yi, H. Xiao, H. Zhang, and A. Wu, "A novel fault diagnosis method for rolling bearing based on improved sparse regularization via convex optimization," *Complexity*, vol. 2018, 2018.
- [29] J. Ding, Z. Zhang, and Y. Yin, "Wheeler-bearing fault detection using adaptive convolution sparse representation," *Shock and Vibration*, vol. 2019, 2019.
- [30] S. Gajjara, M. Kulahcib, and A. Palazoglu, "Real-time fault detection and diagnosis using sparse principal component analysis," *Journal of Process Control*, vol. 67, pp. 112–128, 2018.
- [31] X. Li, H. Shen, L. Zhang, and H. Li, "Sparse-based reconstruction of missing information in remote sensing images from spectral/temporal complementary information," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 106, pp. 1–15, 2015.
- [32] S. Nateghi, Y. Shtessel, J.-P. Barbot, G. Zheng, and L. Yu, "Cyber-attack reconstruction via sliding mode differentiation and sparse recovery algorithm: Electrical power networks application," in *15th International Workshop on Variable Structure Systems (VSS)*, 2018.
- [33] S. Derbel, N. Feki, F. Nicolau, J.-P. Barbot, M. Abbes, and M. Haddar, "Application of homogeneous observers with variable exponent to a mechatronic system," *Journal of Mechanical Engineering Science*, vol. 233, no. 18, pp. 6491–6502, 2019.
- [34] E. Beltrami, *Mathematics for Dynamic Modeling*. Boston: Academic Press, 2nd edition ed., 1998.
- [35] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, vol. 64 of *Mathematics in Science and Engineering*. New York: Academic Press, 1970.
- [36] J. B. Moore and B. D. O. Anderson, "Coping with singular transition matrices in estimation and control stability theory," *International Journal of Control*, vol. 31, no. 3, pp. 571–586, 1980.
- [37] P. K. Kitaniadis, "Unbiased minimum variance linear state estimation," *Automatica*, vol. 23, no. 6, pp. 775–778, 1987.
- [38] Q. Zhang and M. Basseville, "Statistical detection and isolation of additive faults in linear time-varying systems," *Automatica*, vol. 50, no. 10, pp. 2527–2538, 2014.
- [39] Q. Zhang, "Adaptive observer for multiple-input-multiple-output (MIMO) linear time varying systems," *IEEE Trans. on Automatic Control*, vol. 47, no. 3, pp. 525–529, 2002.
- [40] R. E. Kalman, "New methods in Wiener filtering theory," in *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability* (J. L. Bogdanoff and F. Kozin, eds.), (New York), John Wiley & Sons, 1963.
- [41] T. Kailath, *Linear Systems*. Prentice-Hall, 1980.
- [42] Q. Zhang, "Boundedness of the Kitaniadis filter for optimal robust state estimation," in *21st IFAC World Congress*, 2020.
- [43] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [44] N. Meinshausen and P. Bühlmann, "Stability selection," *Journal of the Royal Statistical Society: Series B*, vol. 72, no. 4, pp. 417–473, 2010.
- [45] E. Candès and T. Tao, "Rejoinder: The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ ," *Annals of Statistics*, vol. 35, no. 6, pp. 2392–2404, 2007.
- [46] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [47] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [48] P. Zhao and B. Yu, "On model selection consistency of Lasso," *Journal of Machine Learning Research*, vol. 7, pp. 2541–2563, 2006.
- [49] Q. Zhang and H. Tang, "Diagnosis of inhomogeneous insulation degradation in electric cables by distributed shunt conductance estimation," *Control Engineering Practice*, vol. 21, no. 9, pp. 1195–1203, 2013.