

Fault isolation schemes for a class of continuous-time stochastic dynamical systems

Ángela Castillo, Pedro J. Zufria

A B S T R A C T

In this paper a new method for fault isolation in a class of continuous-time stochastic dynamical systems is proposed. The method is framed in the context of model-based analytical redundancy, consisting in the generation of a residual signal by means of a diagnostic observer, for its posterior analysis. Once a fault has been detected, and assuming some basic a priori knowledge about the set of possible failures in the plant, the isolation task is then formulated as a type of on-line statistical classification problem. The proposed isolation scheme employs in parallel different hypotheses tests on a statistic of the residual signal, one test for each possible fault. This isolation method is characterized by deriving for the unidimensional case, a sufficient isolability condition as well as an upperbound of the probability of missed isolation. Simulation examples illustrate the applicability of the proposed scheme.

1. Introduction

In the last three decades, system control theory has experienced an important evolution thanks to the advances on computer control of complex processes (Aström et al., 2001; Nise, 2011). Since the design of control systems is getting more systematic, the automatic task of responding to abnormal events in a process is becoming a new important challenge. This task gets more involved due to the confluence of an increasing complexity of modern plants and the need of quick diagnosis procedures (Blanke et al., 2003; Chiang et al., 2001; Korbicz et al., 2004). Hence, fault diagnosis schemes must be improved in order to reliably support human operators in the management of malfunctions. The problem of fault diagnosis involves the timely detection of an abnormal event (*fault detection*), diagnosing its causal origins (*fault isolation*) and then taking appropriate supervisory control decisions and actions to bring the process back to a normal, safe, operating state (*fault accommodation*) (Ding, 2008; Iserman, 2006; Venkatasubramanian et al., 2003).

Fault Detection and Isolation (FDI) have deserved much attention from different perspectives (Chen and Patton, 1999; Palade et al., 2006; Simani et al., 2003). In general, the analytical tools employed so far for FDI can be classified into two main categories. On the one hand, stochastic discrete-time models inherited from the signal estimation and linear control fields have successfully com-

bined statistical schemes (mainly hypothesis testing) with geometrical tools in the design and characterization of FDI algorithms for linear systems (Basseville and Nikiforov, 1993; Gertler, 1998; Iserman, 2006). On the other hand, deterministic continuous-time models coming from the adaptive and robust control community have proved to be suitable for nonlinear system modelling, where detection and isolation algorithms rely on the use of diagnostic observers to generate residuals whose profiles are evaluated (Alcorta-García and Frank, 1997; Frank, 1996; De Persis and Isidori, 2001; Polycarpou and Trunov, 2000; Zhang et al., 2005; Zhang et al., 2002). Recent research is also being focused on the design of diagnosis schemes for nonlinear discrete-time stochastic systems either using computationally demanding *particle filters* (Tafazoli and Sun, 2006; Zhang et al., 2005) or adaptive estimators (Xu and Zhang, 2004).

The use of continuous-time stochastic models in system fault diagnosis provides a novel framework for taking into account system and sensor noises and disturbances, in order to construct new detection and isolation algorithms (Castillo et al., 2003; Castillo, 2006; Castillo and Zufria, 2009; Münz and Zufria, 2005; Münz and Zufria, 2009). The seminal work in Castillo et al. (2003) developed an initial study on both the detection and the isolation problems. An extended characterization of the detection problem was later derived in Castillo and Zufria (2009). The present paper accomplishes a parallel characterization task concerning the isolation problem.

Here, the *fault isolation* (FI) problem is addressed for a class of continuous-time stochastic models, where the stochastic framework will prove to be specially suitable for formulating and solving

the isolation problem. Although there are approaches which integrate detection and isolation in a unified process (see Venkatasubramanian et al., 2003 and references therein), typically, the fault detection and isolation tasks are accomplished sequentially: fault isolation goes into effect after a fault is detected, with the objective of determining the location/type of the fault (Zhang et al., 2005; Zhang et al., 2008). The fault isolation task is aimed to determine the particular type of fault among a set of known (or partially known) possible fault types, and to determine its location (the particular faulty subcomponents among the set of all subcomponents under consideration) (Zhang et al., 2005; Zhang et al., 2008). This paper addresses only the discrimination of the actual and previously detected fault among a list of possible faults in the system. been detected. This objective within such stochastic framework will naturally allow for the formulation of isolation as a classification or statistical decision task, or more generally, as a pattern recognition problem (Bishop, 2006; Duda et al., 2001).

Precisely, the nominal part will be in general a nonlinear time-variant system, perturbed by a random process which characterizes the model uncertainty. Additive faults will be considered (both abrupt and incipient) and also modelled as random processes. In such context, the presented FI schemes exploit the model-based analytical redundancy by generating a *residual* via a diagnostic observer, along the line proposed in Zhang et al. (2002) for deterministic nonlinear uncertain systems. Such residual defines the fundamental feature space where the fault classification task is to be performed.

It is worth mentioning that the whole residual-generation/detection/isolation process can be interpreted as the typical sequence of steps followed in a pattern recognition scheme (Bishop, 2006; Duda et al., 2001; Fukunaga, 1990): feature extraction, feature selection and statistical classification. Hence the isolation process can be seen as a special type of pattern recognition problem.

Following standard statistical classification theory (Bishop, 2006; Duda et al., 2001; Fukunaga, 1990), two main FI schemes could be developed in such context. On the one hand, well known Bayes rule based methods could be applied, but they would require the assumption that a prior distribution of possible faults to be available. On the other hand, a hypothesis testing based set-up can be constructed; this second perspective is the one employed in this paper. Precisely, we develop a hypothesis testing set-up on continuous-time statistics of the residual; this method does not require any a priori assumption and, despite that its main objective is to discriminate a fault from a previously given list, it can also alternatively determine that the fault type is a new unknown one. The development of this approach and its posterior analysis has been founded upon the results in Castillo et al. (2003) where fault isolation was first presented, and (Castillo and Zufria, 2009) where fault detection was rigorously characterized. Hence, the results presented in this work, applied to fault isolation, extend and complete those in Castillo and Zufria (2009) and Castillo et al. (2003).

The paper is organized as follows. In Section 2 the problem formulation is presented where the system model and possible faults are mathematically characterized. Then, the proposed FI scheme is put into a general context in Section 4.1, for its posterior analysis in Section 4.3. The practical validity of the proposed approach is illustrated in Section 5 with two simulation examples addressing abrupt and incipient faults. Concluding remarks are finally exposed in Section 6.

2. Problem statement

In this section, the mathematical formulation is presented for both the class of systems under consideration as well as the types of possible faults which may affect the system.

Let us consider a general nonlinear time-variant dynamical system described by:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), t) + \eta(t) + \mathcal{B}(t - T_0)\phi(t), \\ y(t) &= h(x(t), u(t), t), \\ x(0) &= x_0,\end{aligned}\tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the system state, which has known initial value $x_0 \in \mathbb{R}^n$; $u(t) \in \mathbb{R}^m$ is the control input; the known function $f: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$, $f \in C^1$, represents the dynamics of the nominal model; $\mathcal{B}(t - T_0)$ is a diagonal matrix representing the time profile of the fault, and it is made up of the following functions as diagonal elements:

$$\beta_i(t - T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 - e^{-\zeta_i(t - T_0)} & \text{if } t \geq T_0 \end{cases} \quad i = 1, \dots, n,$$

T_0 being the unknown instant when the fault occurs and ζ_i a positive value. Observe that for high values of ζ_i the function $\beta_i(t - T_0)$ will be similar to a step function. More specifically, in the case of an abrupt (sudden) fault, those functions will take the form of a step function and in the case of an incipient (slowly developing) fault they will be ramp-type functions; the fault process $\phi: \mathbb{R}^+ \rightarrow \mathbb{R}^n$ represents the changes in the system dynamics due to a fault, it is assumed to be an n -dimensional stochastic process whose components are Gaussian generalized processes (Larson and Shubert, 1979); in particular the fault process components are defined as $\phi_i(t) \stackrel{\text{def}}{=} X_i(t) + b_i(t)W_i(t)$, $i = 1, \dots, n$, where $X_i(t)$ is a continuous in mean squared (abbreviated MS-continuous) Gaussian random process, $b_i(t)$ is a positive and square integrable deterministic function and $W_i(t)$ is White Gaussian Noise (WGN). Detailed properties and implications of the Gaussian models can be found in Larson and Shubert (1979) and Duda et al. (2001). The uncertainty random vector $\eta: \mathbb{R}^+ \rightarrow \mathbb{R}^n$, which gathers external disturbances and modelling errors, corresponds also to an n -dimensional stochastic process whose components are of the same type. Statistical independence between the uncertainty and the fault processes is also assumed. The derivatives of stochastic system (1) are interpreted as mean square (MS) derivatives.

Gaussian generalized processes (see Larson and Shubert, 1979) are considered in the model with the aim to account for the widest variety of cases, preserving the ubiquitous Gaussianity condition. The hypotheses of MS-continuity and WGN (see Larson and Shubert, 1979) are needed to assure the existence of the stochastic integrals involved in the isolation approach presented in this paper.

Finally $y(t) \in \mathbb{R}^q$ is the measurable output, and the nonlinear mapping $h: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \rightarrow \mathbb{R}^q$ can represent different output availability situations (where output disturbances are not considered). If x can be computed from y , u and t via an implicit function theorem reasoning, this is equivalent to a full state availability; this is a standard assumption in most nonlinear deterministic system diagnosis schemes (Darkhovski and Staroswiecki, 2003; Li and Zhou, 2004; Mattone and De Luca, 2006; Polycarpou and Vemuri, 1995; Polycarpou and Helmicki, 1995; Polycarpou and Trunov, 2000; Trunov and Polycarpou, 2000; Zhang et al., 2004; Zhang et al., 2002), and it is also assumed in this paper. Nevertheless, the presented results can be generalized to the cases where the pair f, h allows for the construction of a robust observer which provides an accurate estimate of x with much faster dynamics than the proposed detection scheme. See (Reble et al., 2007) and references therein where the design of fast observers is addressed, and reference (Münz and Zufria, 2009) where the observation error is formulated and its implications are analytically illustrated.

3. Residual generation

Once a fault is detected in the system, for instance applying the Fault Detection (FD) schemes proposed in Castillo and Zufiria (2009), the next step is to find out as much information as possible concerning such a fault, in order to fix it or at least to avoid its consequences.

The isolation scheme proposed here is based on the analysis of the residual signal obtained from the process plant and a diagnostic observer, since such residual gathers basic information about the fault affecting the system. The choice of the diagnostic observer will depend on the characteristics of the system, and it is aimed to get a residual as easy to analyze as possible. In Jiang et al. (2002), Li and Zhou (2004), Venkatasubramanian et al. (2003), Witczak et al. (2002), and Zhang et al. (2002) and references therein, the existing work on the design of appropriate diagnostic observers for fault detection and isolation is illustrated.

Taking into account the structure of system (1), a convenient diagnostic observer can be constructed following the basis of the Luenberger observer, that is, the diagnostic observer is determined by the nominal part of the model plus a stabilizing term (see Zhang et al., 2002; Castillo and Zufiria, 2009):

$$\dot{\hat{x}}(t) = A(\hat{x}(t) - x(t)) + f(x(t), u(t), t), \quad \hat{x}(0) = x_0, \quad (2)$$

where $\hat{x}(t) \in \mathbb{R}^n$ is the state estimation and the matrix A is chosen to be a diagonal matrix (for the sake of simplicity) with negative diagonal elements in order to assure the diagnostic observer convergence: $A = \text{diag}(\lambda_1, \dots, \lambda_n)$, with $\lambda_i < 0$, $i = 1, \dots, n$. In the deterministic context, the selection of appropriate values for A has been addressed in Willsky (1976) and references therein; also, diagnostic observers (2) have been improved to get faster FDI schemes (Li and Zhou, 2004). Here, as a first step, the basic structure is employed looking for a simple compromise between fast response and numerical stability.

Note that since the state availability is assumed (see previous section), in (2) the function $f(x(\cdot), \cdot, \cdot)$ is used instead of $f(\hat{x}(\cdot), \cdot, \cdot)$ as it would correspond to a proper Luenberger observer. Such substitution does facilitate the residual generation.

Subtracting the diagnostic observer (2) from the system (1), and since the MS derivative is a linear operator, the differential equations system explaining the evolution of the residual process $\epsilon(t) = x(t) - \hat{x}(t)$ is obtained, namely

$$\dot{\epsilon}(t) = A\epsilon(t) + \eta(t) + B(t - T_0)\phi(t), \quad \epsilon(0) = \mathbf{0}. \quad (3)$$

This residual process contains fundamental information concerning the fault, that is the existence or not of a fault in the plant and, in affirmative case, about the fault features. As it will be seen later, while there is no fault in the system (and $\eta(t)$ is a WGN) $\epsilon(t)$ is a multidimensional Ornstein–Uhlenbeck stochastic process (Larson and Shubert, 1979), whose analysis has proved to be crucial in the detection procedure presented in Castillo and Zufiria (2009), and also constitutes the key of the isolation schemes presented here.

4. Isolation scheme

In this section we develop the proposed isolation scheme. We start considering some preliminary issues.

4.1. Preliminaries

The isolation scheme proposed in Section 4.3 is grounded on the following detection assumptions:

Assumption 4.1. The fault has been previously detected in the plant modelled by system (1) with any detection method – see (Castillo and Zufiria, 2009) for an example – at the time instant T_d .

Assumption 4.2. The detection time T_d is close enough to the fault occurrence time T_0 .

Concerning the isolation problem, the initial assumptions are:

Assumption 4.3. A list of possible faults is available:

$$\{\Phi_1, \Phi_2, \dots, \Phi_l\}, \quad (4)$$

with known (except for T_0) time profile functions $\beta_k(\cdot)$, $k = 1, \dots, l$, and totally determined fault processes $\phi_k(t)$, $k = 1, \dots, l$.

The probability distribution for each fault process $\phi_k(t)$ will be denoted by $p(\phi_k(t); \Phi_k)$ where Φ_k is the label which determines each fault. Hence, the corresponding moments or derived distributions will follow the same notation (e.g., $p(\epsilon(t); \Phi_k)$, $E[\epsilon(t); \Phi_k]$, etc.). In general, k -th fault will be referred either by process $\phi_k(t)$ or by its label Φ_k .

Assumption 4.4. The faults in the list (4) can occur just one at a time.

In practice, this assumption can be minimized if the isolation scheme is able to address unknown faults (e.g., the sum of known ones).

Assumption 4.5. The corresponding mean functions of the faults in the list (4) are significantly different.

Finally, if a fault not belonging to such list is occurring, we will refer to that fault as “unknown fault”.

The purpose of the isolation phase is to identify which one of the faults in the list is the one really acting in the system (or to conclude that a possible non-registered fault has occurred).

The residual process (3), will have different characteristics depending on the occurrence time T_0 and on which fault is acting in the system. Assuming that the k th fault occurs, the residual will evolve according to the following equations

$$\dot{\epsilon}_i(t) = \lambda_i \epsilon_i(t) + \eta_i(t) + \beta_{ki}(t - T_0)\phi_{ki}(t), \quad \epsilon_i(0) = 0, \quad t \geq 0, \\ i \in \{1, \dots, n\};$$

Note that the uncertainty on the value of T_0 (which lies in the essence of the *quickest detection* problem (Kailath and Poor, 1998; Poor and Hadjiliadis, 2009)) is not so relevant in this isolation formulation, since a previous detection step is being assumed. Hence, the value of T_0 can be estimated (for instance $\hat{T}_0 = T_d$) so that the residuals are to be considered since the detection of the fault, that is for $t \geq T_d$. Solving those equations, in case fault Φ_k has occurred in the system at time T_0 , the residual components result in

$$\epsilon_i(t) = \int_0^t e^{\lambda_i(t-\tau)} \eta_i(\tau) d\tau + \int_{T_0}^t e^{\lambda_i(t-\tau)} \beta_{ki}(\tau - T_0) \phi_{ki}(\tau) d\tau, \\ t \geq T_0, \quad i \in \{1, \dots, n\};$$

or equivalently

$$\epsilon_i(t) = c_i(T_d) e^{\lambda_i(t-T_d)} + \int_{T_d}^t e^{\lambda_i(t-\tau)} \eta_i(\tau) d\tau + \int_{T_d}^t e^{\lambda_i(t-\tau)} \beta_{ki}(\tau - T_0) \phi_{ki}(\tau) d\tau, \\ t \geq T_d, \quad i \in \{1, \dots, n\};$$

The analysis about the existence, continuity and Gaussianity of these type of integrals can be checked in Castillo (2006) and Larson and Shubert (1979). Since linearity preserves Gaussianity, $\epsilon(t)$ is a multivariate Gaussian process (see Larson and Shubert, 1979), whose mean vector components are determined by

$$E[\epsilon_i(t); \Phi_k] = \int_0^t e^{\lambda_i(t-\tau)} E[\eta_i(\tau)] d\tau + \int_{T_0}^t e^{\lambda_i(t-\tau)} \beta_{ki}(\tau - T_0) E[\phi_{ki}(\tau)] d\tau,$$

$$i \in \{1, \dots, n\},$$

and the covariance between two arbitrary components of the residual vector (i_1 and i_2) under the hypothesis that fault Φ_k is acting in the system is

$$\begin{aligned} \text{Cov}(\epsilon_{i_1}(t_1), \epsilon_{i_2}(t_2); \Phi_k) &= \int_0^{t_1} \int_0^{t_2} e^{\lambda_{i_1}(t_1-\tau_1)} \text{Cov}(\eta_{i_1}(\tau_1), \eta_{i_2}(\tau_2)) e^{\lambda_{i_2}(t_2-\tau_2)} d\tau_2 d\tau_1 \\ &+ \int_{T_0}^{t_1} \int_{T_0}^{t_2} e^{\lambda_{i_1}(t_1-\tau_1)} \beta_{ki_1}(\tau_1 - T_0) \text{Cov}(\phi_{ki_1}(\tau_1), \\ &\phi_{ki_2}(\tau_2)) \beta_{ki_2}(\tau_2 - T_0) e^{\lambda_{i_2}(t_2-\tau_2)} d\tau_2 d\tau_1 \\ &+ \int_0^{t_1} \int_{T_0}^{t_2} e^{\lambda_{i_1}(t_1-\tau_1)} \text{Cov}(\eta_{i_1}(\tau_1), \phi_{ki_2}(\tau_2)) \\ &\times \beta_{ki_2}(\tau_2 - T_0) e^{\lambda_{i_2}(t_2-\tau_2)} d\tau_2 d\tau_1 \\ &+ \int_{T_0}^{t_1} \int_0^{t_2} e^{\lambda_{i_1}(t_1-\tau_1)} \beta_{ki_1}(\tau_1 - T_0) \\ &\text{Cov}(\phi_{ki_1}(\tau_1), \eta_{i_2}(\tau_2)) e^{\lambda_{i_2}(t_2-\tau_2)} d\tau_2 d\tau_1. \end{aligned}$$

Assume now that the covariance matrices associated with the n -dimensional processes $\eta(t)$ and $\phi_k(t)$, respectively $\Sigma_{\eta}(t_1, t_2)$ and $\Sigma_{\phi_k}(t_1, t_2)$, are known. In addition, since we are assuming independence between faults and the system uncertainty (see Section 2), the last two terms in the preceding covariance expression must be zero.

Therefore, the probability distribution followed by the residual stochastic process under the hypothesis that the fault Φ_k is affecting the system is given by:

$$\epsilon(t) \triangleq \mathcal{N}(m_k(t), \Sigma_k(t)),$$

where

$$m_k(t) = \begin{pmatrix} E[\epsilon_1(t); \Phi_k] \\ \vdots \\ E[\epsilon_n(t); \Phi_k] \end{pmatrix},$$

and

$$\Sigma_k(t) = \begin{pmatrix} \text{Var}(\epsilon_1(t); \Phi_k) & \text{Cov}(\epsilon_1(t), \epsilon_2(t); \Phi_k) & \dots & \text{Cov}(\epsilon_1(t), \epsilon_n(t); \Phi_k) \\ \text{Cov}(\epsilon_1(t), \epsilon_2(t); \Phi_k) & \text{Var}(\epsilon_2(t); \Phi_k) & \dots & \text{Cov}(\epsilon_2(t), \epsilon_n(t); \Phi_k) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \text{Var}(\epsilon_n(t); \Phi_k) \end{pmatrix},$$

which characterize the distribution of $\epsilon(t)$ on each instant of time, depending on the fault process $\phi_k(t)$ and the profile function $\beta_k(t - T_0)$.

4.2. Statistical framework for detection and isolation

Under the assumptions in Section 2, the residual $\epsilon(t)$ is, in the most simple case, a multidimensional Ornstein-Uhlenbeck (OU) stochastic process. Hence, although it can be statistically characterized by the distribution provided above, the analysis of optimal detection and isolation schemes in such context is not straightforward. In Zufria (2012) time varying additive changes in such process are considered and the associated Likelihood Ratio (LR) is calculated, whose analytical expression is very cumbersome. In general, such complicated form of the LR makes optimal detection schemes to be computationally too expensive. In addition, such LR form does not allow a straightforward implementation of on-line schemes, as opposed, for instance, to the CUSUM schemes in the Brownian motion framework.

It is worth mentioning that the Generalized Likelihood Ratio (GLR) schemes proposed in Willsky and Jones (1976) for several types of additive faults in linear discrete-time Gaussian systems

serve as a good reference for the design of some detection/isolation schemes; nevertheless, they cannot be efficiently applied for continuous-time multivariable OU processes in a straightforward manner.

So far, practical FD schemes grounded on heuristic approaches have been successfully implemented (Castillo and Zufria, 2009; Münz and Zufria, 2009; Zufria, 2009). Under some additional assumptions such as the availability of an estimate of T_0 , or a partial knowledge of fault functions (e.g. their profile) it is possible to derive the approximations leading to the efficient isolation schemes proposed in this paper (see Zufria (2012) for more details).

The scheme proposed here represents a natural extension of the fault detection method presented in Castillo and Zufria (2009), by means of applying on the residual several hypothesis testing schemes in parallel, one for each possible fault; this method is specially suitable for noticing faults which are not registered in the list (4).

Next, this isolation approach is analyzed.

4.3. Proposed isolation approach

The approach proposed here employs some ideas from previous works in discrete-time systems (Basseville and Nikiforov, 1993; Willsky, 1976; Willsky and Jones, 1976) since it is based on the application of several hypotheses tests. One can interpret the parametrized distributions $p(\epsilon(t); \Phi_k) = p_{\Phi_k}(\epsilon(t))$ as likelihood functions. If these functions were fully available, one could employ the corresponding likelihood ratios in order to optimally construct such tests. Unfortunately, these overall likelihood functions are very difficult to obtain in general. Their computation has been addressed under some specific assumptions in Zufria (2012), where the problems of *classical detection* and *quickest detection* (Kailath and Poor, 1998; Poor and Hadjiliadis, 2009) have been studied using also the residual $\epsilon(t)$ of Eq. (3). In fact, the likelihood ratio between the null (no fault) and the fault hypotheses is calculated there (generalizing some results in Arató (1982)) for the case of deterministic faults and fixed interval of time. Also in Zufria (2012), this likelihood ratio is employed as the basis of some on-line suboptimal hypotheses tests for building fault detection schemes. Nevertheless, these results in Zufria (2012) may not be easily extended to the general isolation problem, specially for the case of stochastic faults. In such case, the analysis strongly depends on the statistical characterization of the faults: for instance, when they are non-stationary or non-Gaussian processes, it seems that as far as we know only suboptimal isolation schemes can be efficiently designed. (Some specific cases, such as the detection and isolation of parametric faults have been addressed in Münz and Zufria (2009), and a comparison of some existing schemes has also been performed in Zufria (2009).)

In this work, the analysis performed in Section 4.1, provides the probability density functions $p(\epsilon(t); \Phi_k)$ at each instant of time t . Hence, based on the corresponding family of likelihood functions (one for each instant of time and each fault in the list (4)), different isolation schemes could be defined. For instance, isolation could be accomplished when $p(\epsilon(t); \Phi_j) > p(\epsilon(t); \Phi_k), \forall k \neq j$ during a prescribed interval of time. Nevertheless, these schemes seem solely appropriate when only faults from the list (4), with a full statistical characterization, can happen.

Here, in order to consider that new faults not belonging to the list may show up, the proposed approach applies a bank of multi-dimensional hypothesis testing schemes, one for each fault in the list (4). Such tests are applied on the residual mean vector, in order to check which fault the actual residual mean vector corresponds to. The residual mean estimators given in Castillo and Zufria (2009) have been shown in Zufria (2012) to be valid overall likelihood ratio approximators, and consequently a valid residual fea-

ture; so they have been also successfully employed in Castillo and Zufria (2009). Under Assumption 4.5 the set of l simple tests

$$H_0^1 : E[\epsilon(t)] = E[\epsilon(t); \Phi_1] \cdots H_0^l : E[\epsilon(t)] = E[\epsilon(t); \Phi_l]$$

$$H_1^1 : E[\epsilon(t)] \neq E[\epsilon(t); \Phi_1] \cdots H_1^l : E[\epsilon(t)] \neq E[\epsilon(t); \Phi_l],$$

can be applied to check if the actual residual mean corresponds to the expected mean when each one of the faults in the given list was affecting the system. Substituting $\phi(t)$ by $\phi_k(t)$, $k = 1, \dots, l$, in (3), the distribution of the residual given the occurrence of the k th fault can be determined, so that the mentioned hypotheses tests for the means $\{E[\epsilon(t); \Phi_1], \dots, E[\epsilon(t); \Phi_l]\}$ can be appropriately constructed (Fukunaga, 1990; Gertler, 1998). Given the mean vectors and covariance matrices one can determine the tests acceptance regions following, for instance, the two strategies to construct a multidimensional hypothesis test presented in Castillo and Zufria (2009) (*component by component* and *squared Mahalanobis distance (SM-distance)* based strategies).

The test requires the construction of a mean estimator, $\mu(t) : \mathbb{R}^+ \rightarrow \mathbb{R}^n$, whose probability distribution must be determined, based on the distribution of the residual process assuming different faults. The acceptance region for each test, $B_k(t)$, $k = 1, \dots, l$, is obtained as a ball, which will depend on the selected test construction strategy, basically defined by a distance $d(\cdot, \cdot)$, see (Castillo and Zufria, 2009). Each ball $B_k(t)$ will be centered in the mean of the estimator under the hypothesis that the k th fault is occurring, $E[\mu(t); \Phi_k]$,

$$B_k(t) = \{\xi \in \mathbb{R}^n / d(\xi, E[\mu(t); \Phi_k]) \leq k_\gamma\}, \quad (5)$$

being k_γ such that $B_k(t)$ fulfills the condition

$$P(\mu(t) \in B_k(t); \Phi_k) = 1 - \gamma, \quad (6)$$

where the predefined small value γ is the *test size*.

For instance, in the unidimensional case, if choosing a Gaussian estimator $\mu(t)$, for example one of the following

$$\mu_a(t) = \epsilon(t); \quad \mu_b(t) = \frac{1}{t} \int_0^t \epsilon(\tau) d\tau, \quad t > 0;$$

$$\mu_c(t) = \frac{1}{T} \int_{t-T}^t \epsilon(\tau) d\tau, \quad t > T; \quad \mu_d(t) = \int_0^t e^{\rho(t-\tau)} \epsilon(\tau) d\tau, \quad \rho < 0 \quad (7)$$

(see Castillo and Zufria (2009) for details), the acceptance regions corresponding to each fault are

$$(l_k(t), u_k(t)) = \left(E[\mu(t); \Phi_k] - h_{\frac{\gamma}{2}} \sqrt{\text{Var}(\mu(t); \Phi_k)}, \right. \\ \left. \times E[\mu(t); \Phi_k] + h_{\frac{\gamma}{2}} \sqrt{\text{Var}(\mu(t); \Phi_k)} \right], \quad (8)$$

with $k = 1, \dots, l$. ($h_{\frac{\gamma}{2}}$ is the normal distribution value such that $P(z \leq h_{\frac{\gamma}{2}}) = 1 - \frac{\gamma}{2}$.)

Hence, provided a fault, $\Phi = \Phi_j$, from the list (4) happens, the realization of the residual mean estimator will remain between the limits of the test corresponding to such actual fault Φ_j with probability $1 - \gamma$, and hopefully outside the acceptance regions associated with the rest of the tests.

On the other hand, provided a fault Φ , which is not in the list (4) happens, the realization of the estimator will presumably end up outside the acceptance regions associated with all the faults in list (4). In this case the proposed isolation method will be useful to advertise that an unknown fault has appeared in the system.

Therefore, based on the specified bank of tests, the proposed isolation decision scheme is grounded on the following definition:

Definition 4.1. T_{iso}^j is the first (in the infimum sense) instant of time for which either

- (1) $\mu(t) \in B_j(t)$ and $\mu(t) \notin B_k(t)$, $\forall k \in \{1, \dots, l\} - \{j\}$, $\forall t \in [T_{iso}^j - \Delta t, T_{iso}^j]$. If there exists such T_{iso}^j , then it is considered that the j -th fault in (4) is affecting the system (1), and T_{iso}^j is the isolation time instant for that particular estimator realization.
- (2) $\mu(t) \notin B_k(t)$, $\forall k \in \{1, \dots, l\}$, $\forall t \in [T_{iso}^j - \delta t, T_{iso}^j]$. If there exists such T_{iso}^j , then it is considered that an unknown fault is affecting the system (1), and T_{iso}^j is the isolation time instant for that particular estimator realization.

(Note: the increments of time Δt and δt are parameters whose values must be fixed previously).

In the following we will consider the case a fault in the list (4) is happening, that is T_{iso}^j will correspond to the first item of Definition 4.1. Then, T_{iso}^j is determined as follows:

Proposition 4.1. Assuming a fault of list (4) is affecting the system, let it be

$$B_0(t) = \left(\bigcup_{k=1}^l B_k(t) \right)^c,$$

and

$$\mathbb{T}^j = \left\{ \bar{T} \geq T_d / \forall t \in [\bar{T} - \Delta t, \bar{T}], \mu(t) \in B_j(t) \cap \left(\bigcup_{k \neq j}^l B_k(t) \right)^c \right\}.$$

Hence

$$T_{iso}^j = \inf \mathbb{T}^j.$$

T_{iso}^j will depend on the size of the intersection region among the different acceptance regions, which could be quite big in case there is not much difference among the faults processes means: the most the faults differ the quickest is the isolation. Note that if there were not much difference among the faults means, the realization of the residual process might remain between the limits of the tests corresponding to more than a single fault Φ_k from the list (4), with a significant probability. In such case, one would need to resort to the comparison between the corresponding likelihoods $p_{\Phi_k}(\epsilon(t))$ in order to isolate the fault. Time T_{iso}^j will also depend on the time in which the residual realization gets its steady state; hence, high values of $|\lambda_i|$ (λ_i being the observer stabilizing parameter) are desirable to speed up the convergence to the steady state of the residual sample path. On the other hand, $|\lambda_i|$ must be small enough to guarantee convergence of the computational numerical methods employed for solving the differential equations; this leads to a trade-off.

We proceed now to characterize and analyze the presented isolation scheme for unidimensional systems as well as for multidimensional systems where only one component of the residual is considered: some analytical sufficient conditions which guarantee, with significant probability, the isolability of a fault from a predefined set, are given in Section 4.4. Besides that, the probability of missed isolation, for unidimensional systems under certain conditions, is studied in Section 4.5.

4.4. Isolability conditions for unidimensional systems

The results presented here for unidimensional systems are also valid for multidimensional systems where only one component of the residual is considered.

Let us define now an upperbound (in probabilistic sense) for the random variable T_{iso}^j , the instant of time when fault Φ_j is isolated among the faults of set $\{\Phi_1, \dots, \Phi_l\}$.

Definition 4.2. $\hat{T}_{iso}^{j,k} = \Delta t + \inf\{\tau \geq T_d / \forall t \geq \tau B_j(t) \cap B_k(t) = \emptyset\}$.

Definition 4.3. $\hat{T}_{iso}^j = \Delta t + \inf\{\tau \geq T_d / \forall t \geq \tau B_j(t) \cap (\bigcup_{n \in \{1, \dots, l\} - \{j\}} B_n(t)) = \emptyset\}$.

Proposition 4.2. Observe that

$$\hat{T}_{iso}^j = \max \left\{ \hat{T}_{iso}^{j,k}, \quad k = 1, \dots, l, k \neq j \right\}$$

Definition 4.4. If $\hat{T}_{iso}^j < +\infty$ then fault Φ_j is said to be \hat{T}_{iso}^j -**isolable**.

Definition 4.5. If $\hat{T}_{iso}^j = +\infty$ then fault Φ_j is said to be **hardly isolable** from the set of faults $\{\Phi_1, \dots, \Phi_l\}$.

Lemma 4.1. If a fault Φ_j is \hat{T}_{iso}^j -isolable then \hat{T}_{iso}^j is an upperbound in probabilistic sense, namely with certain probability, for T_{iso}^j , the isolation time instant of fault Φ_j . That is

$$P(\hat{T}_{iso}^j \in \mathbb{T}^j; \Phi_j) = 1 - P(\mu(t) \notin B_j(t) \text{ for some } t \in [\hat{T}_{iso}^j - \Delta t, \hat{T}_{iso}^j]; \Phi_j).$$

Observe that

- \hat{T}_{iso}^j depends on each particular mean estimator realization, that is \hat{T}_{iso}^j is a random variable, whereas T_{iso}^j is a deterministic value.
- \mathbb{T}^j is the set defined in Proposition 4.1.
- $P(\mu(t) \notin B_j(t) \text{ for some } t \in [\hat{T}_{iso}^j - \Delta t, \hat{T}_{iso}^j]; \Phi_j)$ can be upper-bounded (for systems fulfilling Assumption 4.6) following the reasoning in SubSection 4.5.

Proof. By construction of the j th acceptance region $B_j(t)$ and the definition of \hat{T}_{iso}^j ,

$$P(\hat{T}_{iso}^j \in \mathbb{T}^j; \Phi_j) = \frac{|\mathbb{T}^j|}{|\mathbb{T}^j|}$$

$$P(\forall t \in [\hat{T}_{iso}^j - \Delta t, \hat{T}_{iso}^j], \mu(t) \in B_0^c(t) \cap B_1^c(t) \dots \cap B_j(t) \cap \dots \cap B_l^c(t); \Phi_j) = \frac{|\mathbb{T}^j|}{|\mathbb{T}^j|}$$

$$P(\forall t \in [\hat{T}_{iso}^j - \Delta t, \hat{T}_{iso}^j], \mu(t) \in B_j(t); \Phi_j) =$$

$$1 - P(\mu(t) \notin B_j(t) \text{ for some } t \in [\hat{T}_{iso}^j - \Delta t, \hat{T}_{iso}^j]; \Phi_j) \quad \square$$

Next, based on the *isolability* concept, we develop a sufficient condition, aimed to guarantee, with certain probability, a fault isolation by the proposed scheme.

Theorem 4.1 (*Isolability sufficient condition*). Let us consider

(a) The time profile functions of two faults Φ_j and Φ_k , that is

$$\beta_j(t - T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 - e^{-\varsigma_j(t-T_0)} & \text{if } t \geq T_0 \end{cases}$$

and

$$\beta_k(t - T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 - e^{-\varsigma_k(t-T_0)} & \text{if } t \geq T_0 \end{cases}$$

(where $\varsigma_j, \varsigma_k > 0$);

(b) The corresponding random processes $\phi_j(t)$ and $\phi_k(t)$ such that $\exists T_0^* \geq T_0$ /

$$\begin{aligned} \beta_j(t - T_0)E[\phi_j(t)] - \beta_k(t - T_0)E[\phi_k(t)] &\geq 0 \quad \forall t \in [T_0, T_0^*] \text{ and} \\ \beta_j(t - T_0)E[\phi_j(t)] - \beta_k(t - T_0)E[\phi_k(t)] &\geq \varepsilon \quad \forall t > T_0^* \end{aligned} \quad (9)$$

or

$$\begin{aligned} \beta_j(t - T_0)E[\phi_j(t)] - \beta_k(t - T_0)E[\phi_k(t)] &\leq 0 \quad \forall t \in [T_0, T_0^*] \text{ and} \\ \beta_j(t - T_0)E[\phi_j(t)] - \beta_k(t - T_0)E[\phi_k(t)] &\leq -\varepsilon \quad \forall t > T_0^*; \end{aligned} \quad (10)$$

Then, a sufficient condition to assure the existence of a finite value $\hat{T}_{iso}^{j,k} \geq T_0^*$ such that the fault Φ_j given by $\beta_j(t - T_0)\phi_j(t)$ is $\hat{T}_{iso}^{j,k}$ -isolable by the FI scheme associated with estimator $\mu_\zeta(t)$, $\zeta \in \{a, b, c, d\}$ (see Definition 4.1 and estimator proposals in Eq. (7)), is that

$$\varepsilon > \omega_\zeta(\lambda) \limsup_{t \rightarrow +\infty} \left(h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu_\zeta(t); \Phi_j)} + h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu_\zeta(t); \Phi_k)} \right) \quad (11)$$

where

$$\omega_a(\lambda) = -\lambda, \quad \omega_b(\lambda) = -\lambda, \quad \omega_c(\lambda) = -\lambda, \quad \omega_d(\lambda) = \lambda\rho.$$

Observe that $\omega_\zeta(\lambda) > 0$, $\forall \zeta \in \{a, b, c, d\}$.

(Note: λ is the observer gain in the unidimensional case, and ρ the parameter in the expression of estimator μ_d .)

Proof. For the sake of simplicity let us choose for example estimator $\mu_d(t) = \mu(t) = \epsilon(t)$ and condition (9) to delineate the main steps of the proof.

- First, since

$$\begin{aligned} \mu(t) &= \epsilon(t) = \int_0^t e^{\lambda(t-\tau)} \eta(\tau) d\tau + \int_{T_0}^t e^{\lambda(t-\tau)} \beta(\tau - T_0) \phi(\tau) d\tau, \quad t \\ &> T_0 \end{aligned}$$

and taking into account the linearity of the mean operator,

$$\begin{aligned} E[\mu(t); \Phi_j] - E[\mu(t); \Phi_k] &= \int_{T_0}^t e^{\lambda(t-\tau)} E[\beta_j(\tau - T_0) \phi_j(\tau) \\ &\quad - \beta_k(\tau - T_0) \phi_k(\tau)] d\tau, \quad t > T_0, \end{aligned}$$

meaning that the estimator expectations assuming fault Φ_j and fault Φ_k differ in an integral term depending on the difference between the means of the corresponding fault processes $\beta_j(t - T_0)\phi_j(t)$ and $\beta_k(t - T_0)\phi_k(t)$. (This is also the case for estimators μ_b , μ_c and μ_d .)

- Assumption (9) implies

$$\begin{aligned} E[\mu(t); \Phi_j] - E[\mu(t); \Phi_k] &= \int_{T_0}^{T_0^*} e^{\lambda(t-\tau)} E[\beta_j(\tau - T_0) \phi_j(\tau) - \beta_k(\tau - T_0) \phi_k(\tau)] d\tau + \\ &\int_{T_0^*}^t e^{\lambda(t-\tau)} E[\beta_j(\tau - T_0) \phi_j(\tau) - \beta_k(\tau - T_0) \phi_k(\tau)] d\tau \geq \\ &0 + \frac{\varepsilon}{-\lambda} (1 - e^{\lambda(t-T_0)}) \nearrow_{t \rightarrow +\infty} \frac{\varepsilon}{-\lambda} = \frac{\varepsilon}{\omega(\lambda)}. \end{aligned}$$

(The diagonal arrow \nearrow indicates that the function $\frac{\varepsilon}{-\lambda} (1 - e^{\lambda(t-T_0)})$ grows up monotonically towards the value $\frac{\varepsilon}{-\lambda}$.)

- This result, along with condition (11), implies that $\exists t^* > T_0^*$ fulfilling that $\forall t \geq t^*$

$$\frac{\varepsilon}{-\lambda} (1 - e^{\lambda(t-T_0)}) > \left(h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)} + h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_k)} \right),$$

and so $\exists t^* > T_0^*$ such that $\forall t \geq t^*$

$$E[\mu(t); \Phi_j] - E[\mu(t); \Phi_k] > \left(h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)} + h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_k)} \right). \quad (12)$$

Therefore, remembering $(l_k(t), u_k(t))$ as defined in (8):

$$\begin{aligned} \inf\{\tau \geq T_0 / \forall t \geq \tau \ B_j(t) \cap B_k(t) = \emptyset\} &= \inf\{\tau \geq T_0 / \forall t \geq \tau \ l_j(t) \geq u_k(t)\} = \\ \inf\{\tau \geq T_0 / \forall t \geq \tau \ E[\mu(t); \Phi_j] - h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)} &\geq E[\mu(t); \Phi_k] + h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_k)}\} = \\ \inf\{\tau \geq T_0 / \forall t \geq \tau \ E[\mu(t); \Phi_j] - E[\mu(t); \Phi_k] &\geq h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)} + h_{\frac{\zeta}{2}} \sqrt{\text{Var}(\mu(t); \Phi_k)}\} \leq t^* \end{aligned}$$

so

$$\exists \hat{T}_{iso}^{j,k} = \Delta t + \inf\{\tau \geq T_0 / \forall t \geq \tau \ B_j(t) \cap B_k(t) = \emptyset\} \leq \Delta t + t^* < +\infty.$$

We conclude that the fault Φ_j is $\hat{T}_{iso}^{j,k}$ -**isolable** from the set $\{\Phi_j, \Phi_k\}$ by the isolation scheme described in this section (see Definition 4.1) corresponding to estimator $\mu(t)$.

Note that we can analytically determine the value of $\hat{T}_{iso}^{j,k}$ solving in \hat{T} the equation

$$l_j(\hat{T}) = u_k(\hat{T}).$$

That is possible thanks to the continuity of functions $u_k(t)$ and $l_j(t)$.

That equation can be rewritten as

$$E[\mu(\hat{T}); \Phi_j] - E[\mu(\hat{T}); \Phi_k] = h_{\frac{1}{2}} \sqrt{\text{Var}(\mu(\hat{T}); \Phi_j)} + h_{\frac{1}{2}} \sqrt{\text{Var}(\mu(\hat{T}); \Phi_k)},$$

when solving this equality it could be $\hat{T} \leq T_0$, in that case $\hat{T}_{iso}^{j,k} = T_0 + \Delta t$; or $\hat{T} > T_0$, and then $\hat{T}_{iso}^{j,k} = \hat{T} + \Delta t$. Obviously $\hat{T}_{iso}^{j,k} - T_0$ will be an upperbound of the time to be taken in the isolation of fault Φ_j versus Φ_k .

For the cases fulfilling assumption (10), we follow the same reasoning using

$$\inf\{\tau \geq T_0 / \forall t \geq \tau \ u_j(t) \leq l_k(t)\}$$

instead.

The same guidelines allow to prove the theorem for the rest of estimators $\mu_c(t)$, $\zeta \in \{b, c, d\}$. \square

Corollary 4.1. *If the isolability sufficient condition given in Theorem 4.1 fulfills for $k \in \{1, \dots, l\} - \{j\}$ then $\exists \hat{T}_{iso}^j$ and it is*

$$\hat{T}_{iso}^j = \max\{\hat{T}_{iso}^{j,k}, k = 1, \dots, l, k \neq j\} < \infty,$$

so fault Φ_j will be \hat{T}_{iso}^j -isolable.

An example of isolation in a unidimensional system, using the method just explained, is given in Section 5.2.

4.5. Missed isolation probability

According to Definition 4.1, the probability of missed isolation of the actual fault Φ_j in a time interval is

$$P(\mu(t) \notin B_j(t) \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\ = P(d(\mu(t), E[\mu(t); \Phi_j]) > k_y(t) \text{ for some } t \in [T - \Delta t, T]; \Phi_j).$$

For unidimensional systems, with a Gaussian mean estimator $\mu(t)$,

$$d(\mu(t), E[\mu(t); \Phi_j]) = d_M(\mu(t), E[\mu(t); \Phi_j]) \\ = \frac{|\mu(t) - E[\mu(t); \Phi_j]|}{\sqrt{\text{Var}(\mu(t); \Phi_j)}}, \quad (13)$$

and $k_y = h_{\frac{1}{2}}$.

A bound for the probability of missed isolation in an interval is calculated below, but under the assumption that the uncertainty process $\eta(t)$ and the fault process $\phi(t)$ fulfill the following (memoryless in mean) condition.

Assumption 4.6.

$$E[\eta(t)/\eta(\xi), 0 \leq \xi < t] = E[\eta(t)] \quad \forall t \in \mathbb{R}^+, \quad (14)$$

$$E[\phi(t)/\phi(\xi), 0 \leq \xi < t] = E[\phi(t)] \quad \forall t \in \mathbb{R}^+, \quad (15)$$

Assumption 4.7. The uncertainty process $\eta(\cdot)$ and the fault process $\phi(\cdot)$ are statistically independent (as established previously in Section 2).

Theorem 4.2 (*Missed isolation probability upperbound*). *As a consequence of Assumptions 4.6 and 4.7 the following results are derived for unidimensional systems.*

(1) For estimator $\mu_a(t) = \epsilon(t)$

$$P(\mu_a(t) \notin B_j(t) \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\ = P(|\mu_a(t) - E(\mu_a(t); \Phi_j)| \geq h_{\frac{1}{2}} \sqrt{\text{Var}(\mu_a(t); \Phi_j)}, \\ \text{for some } t \in [T - \Delta t, T]; \Phi_j) \leq \frac{e^{-2\lambda T} \text{Var}(\mu_a(T); \Phi_j)}{m_{[T-\Delta t, T]}^2}.$$

(2) For estimator $\mu_b(t) = \frac{1}{t} \int_0^t \epsilon(\tau) d\tau$

$$P(\mu_b(t) \notin B_j(t) \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\ = P(|\mu_b(t) - E(\mu_b(t); \Phi_j)| \geq h_{\frac{1}{2}} \sqrt{\text{Var}(\mu_b(t); \Phi_j)}, \\ \text{for some } t \in [T - \Delta t, T]; \Phi_j) \\ \leq \frac{T^2 \text{Var}(\mu_b(T); \Phi_j)}{m_{[T-\Delta t, T]}^2}$$

where $m_{[T_1, T_2]} = \min\{th_{\frac{1}{2}} \sqrt{\text{Var}(\mu_b(t); \Phi_j)}, t \in [T_1, T_2]\}$.

(3) And for estimator $\mu_d(t) = \int_0^t e^{\rho(t-\tau)} \epsilon(\tau) d\tau$

$$P(\mu(t) \notin B_j(t) \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\ = P(|\mu_d(t) - E[\mu_d(t); \Phi_j]| \geq h_{\frac{1}{2}} \sqrt{\text{Var}(\mu_d(t); \Phi_j)}, \\ \text{for some } t \in [T - \Delta t, T]; \Phi_j) \\ \leq \frac{e^{-2\rho T} \text{Var}(\mu_d(T); \Phi_j)}{m_{[T-\Delta t, T]}^2},$$

where, in this case, $m_{[T_1, T_2]} = \min\{e^{-\rho t} h_{\frac{1}{2}} \sqrt{\text{Var}(\mu_d(t); \Phi_j)}, t \in [T_1, T_2]\}$.

Proof. See Appendix A. \square

5. Examples

The proposed FI methods are illustrated with the following two examples.

5.1. Unidimensional example

This example illustrates the application of the proposed FI scheme as well as the isolability conditions on unidimensional systems with incipient faults.

Consider the system given by the differential equation:

$$\dot{x}(t) = \sin x(t) + \eta(t) + \beta(t - T_0)\phi_k(t) \\ y(t) = x(t), \quad x(0) = 0,$$

where $\eta(t)$ and $v(t)$ are independent WGN processes with zero mean and autocorrelation functions $R_\eta(t_1, t_2) = \sigma_\eta^2 \delta(t_1 - t_2)$ and $R_v(t_1, t_2) = \sigma_v^2 \delta(t_1 - t_2)$, respectively. The fault in this simulation will occur at time $T_0 = 70$ and it is an incipient fault with the time profile function

$$\beta(t - T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 - e^{-\rho(t-T_0)} & \text{if } t \geq T_0 \end{cases}$$

In this simulation example $\rho = 0.05$.

The diagnostic observer presented in Section 3 will be

$$\dot{\hat{x}}(t) = \lambda(\hat{x}(t) - x(t)) + \sin x(t), \\ \hat{x}(0) = 0,$$

with $\lambda < 0$ (for instance $\lambda = -0.5$ is taken). Then subtracting model and diagnostic observer equations the residual equation

$$\dot{\epsilon}(t) = \lambda\epsilon(t) + \eta(t) + \beta(t - T_0)\phi_k(t), \\ \epsilon(0) = 0,$$

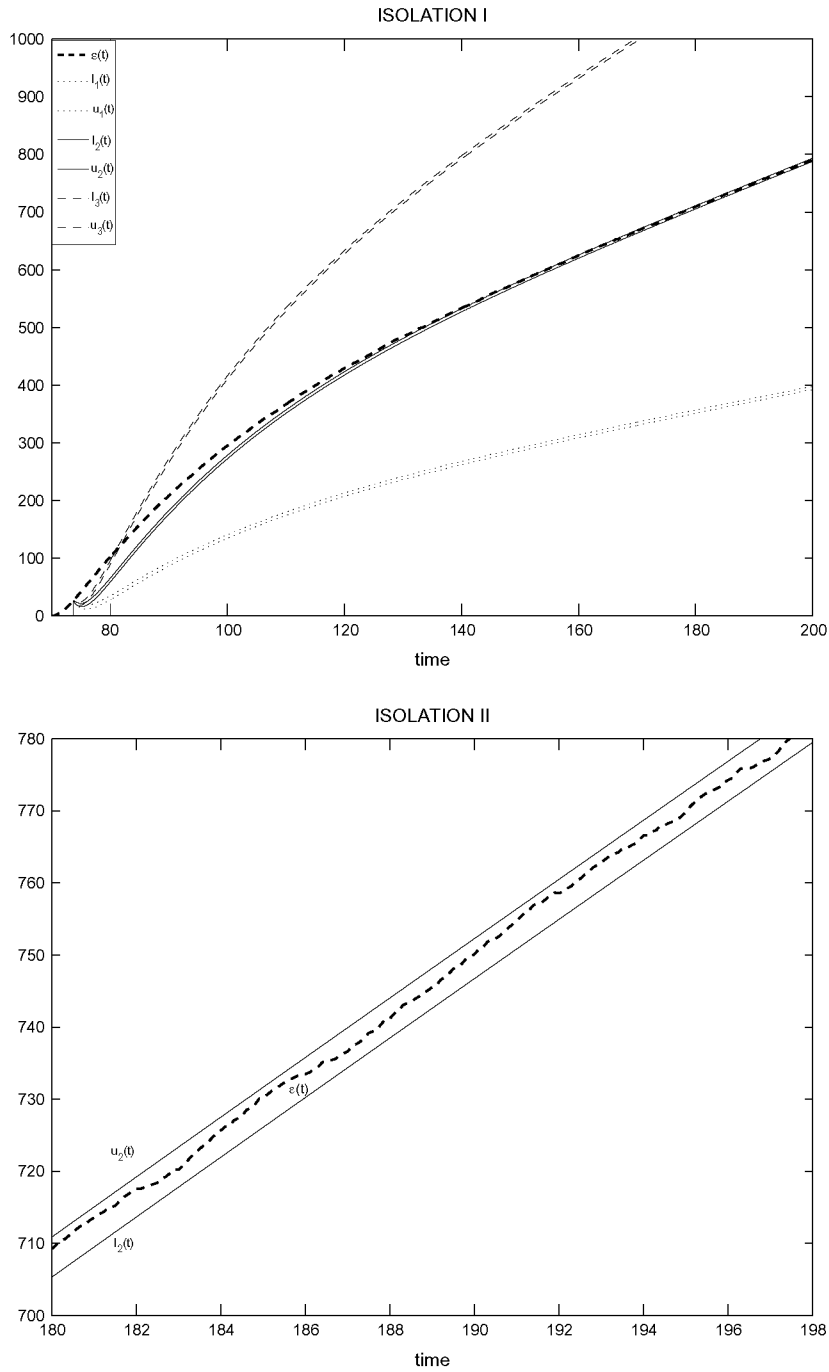


Fig. 1. Isolation: Bank of hypotheses tests. Unidimensional example.

is obtained.

We have simulated the occurrence of fault Φ_2 given by fault function $\phi_2(t) = 2t + v(t)$. Applying one of the Fault Detection schemes proposed in Castillo and Zufria (2009) the fault is detected at time $T_d = 73.67$.

Once the presence of a fault is detected in plant, the next step is to isolate it. Let us establish the following set of possible fault processes

$$\{\phi_1(t) = t + v(t), \phi_2(t) = 2t + v(t), \phi_3(t) = 3t + v(t)\},$$

or equivalently: $\phi_k(t) = kt + v(t)$, $k = 1, 2, 3$, where $v(t)$ is, as said before, a WGN process independent of the WGN process $\eta(t)$.

In order to isolate the fault, the residual process distribution is needed for the cases when each one of the possible faults was happening. In this case the residual is

$$\begin{aligned} \epsilon(t) = & \epsilon_{\{T_d, \Phi_k\}} e^{\lambda(t-T_d)} + \int_{T_d}^t e^{\lambda(t-\tau)} \eta(\tau) d\tau \\ & + \int_{T_d}^t e^{\lambda(t-\tau)} \beta(\tau - T_0) \phi_k(\tau) d\tau, \quad t \geq T_d, \end{aligned}$$

where the value $\epsilon_{\{T_d, \Phi_k\}}$ is the measured value of the residual process at time T_d under assumption the fault ϕ_k is affecting the system. Such residual is a Gaussian random process with mean

$$\begin{aligned}
E[\epsilon(t); \Phi_k] &= \epsilon_{(T_d, \Phi_k)} e^{\lambda(t-T_d)} + \int_{T_d}^t e^{\lambda(t-\tau)} \beta(\tau - T_0) E[\phi_k(\tau)] d\tau \\
&= \epsilon_{(T_d, \Phi_k)} e^{\lambda(t-T_d)} + k \left(\frac{t}{-\lambda} - \frac{1}{(-\lambda)^2} \right) - k e^{\lambda(t-T_d)} \left(\frac{T_d}{-\lambda} - \frac{1}{(-\lambda)^2} \right) \\
&\quad - k e^{-\rho(t-T_0)} \left(\frac{t}{-(\lambda+\rho)} - \frac{1}{(\lambda+\rho)^2} \right) - k e^{\lambda(t-T_d)-\rho(T_d-T_0)} \\
&\quad \left(\frac{1}{(\lambda+\rho)^2} - \frac{T_d}{-(\lambda+\rho)} \right).
\end{aligned}$$

Observe that for high values of t , namely asymptotically,

$$E[\epsilon(t); \Phi_k] \approx k \left(\frac{t}{-\lambda} - \frac{1}{(-\lambda)^2} \right).$$

Note that since

$$0 \leq \beta(t - T_0) E[\phi_j(t) - \phi_k(t)] \xrightarrow{t \rightarrow +\infty} +\infty, \quad k < j, \quad k, j = 1, 2, 3,$$

the isolability sufficient condition (see Theorem 4.1) is clearly satisfied, so that we have a guarantee that the isolation approach presented in Section 4.3 is able to isolate any fault ϕ_k from the rest of faults in the list.

Since $\eta(t)$ and $v(t)$ are independent and zero mean processes then:

$$\begin{aligned}
\text{Cov}_{\eta+\beta\phi}(t_1, t_2) &= \text{Cov}_{\eta}(t_1, t_2) + \beta(t_1 - T_0) \text{Cov}_{\phi}(t_1, t_2) \beta(t_2 - T_0) \\
&= R_{\eta}(t_1, t_2) + \beta(t_1 - T_0) R_v(t_1, t_2) \beta(t_2 - T_0).
\end{aligned}$$

Therefore the variance of $\epsilon(t)$, depending on Φ_k is given by

$$\begin{aligned}
\text{Var}(\epsilon(t); \Phi_k) &= \int_{T_d}^t \int_{T_d}^t e^{\lambda(t-\tau_1)} R_{\eta}(\tau_1, \tau_2) e^{\lambda(t-\tau_2)} d\tau_2 d\tau_1 \\
&\quad + \int_{T_d}^t \int_{T_d}^t e^{\lambda(t-\tau_1)} (1 - e^{-\rho(\tau_1-T_0)}) R_v(\tau_1, \tau_2) \\
&\quad \times (1 - e^{-\rho(\tau_2-T_0)}) e^{\lambda(t-\tau_2)} d\tau_2 d\tau_1 \\
&= -\frac{\sigma_{\eta}^2}{2\lambda} (1 - e^{2\lambda(t-T_d)}) - \frac{\sigma_v^2}{2\lambda} (1 - e^{2\lambda(t-T_d)}) \\
&\quad + 2 \frac{\sigma_v^2}{2\lambda + \rho} (e^{-\rho(t-T_0)} - e^{2\lambda(t-T_d)-\rho(T_d-T_0)}) \\
&\quad - \frac{\sigma_v^2}{2(\lambda + \rho)} (e^{-2\rho(t-T_0)} - e^{2\lambda(t-T_d)-2\rho(T_d-T_0)}) \\
&\xrightarrow{t \rightarrow \infty} -\frac{\sigma_{\eta}^2}{2\lambda} - \frac{\sigma_v^2}{2\lambda}.
\end{aligned}$$

The preceding expressions are valid when the diagnostic observer design parameter $\lambda \notin \{-\rho, \frac{-\rho}{2}\}$, where ρ is the fault time profile parameter. This is the case for this particular example because $|\lambda| \gg \rho$. Since the fault occurrence time T_0 is unknown, it is substituted in the given expressions by the detection time T_d ; therefore, for the sake of accuracy, it is important to consider a quick detection method.

The approximations in this example must take into account the time the estimator takes in reaching its steady state, for instance in computing $\hat{T}_{iso}^{j,k}$, an estimation of such time should be added to represent a real upperbound of the isolation time.

The acceptance regions corresponding to a collection of tests on the residual mean (one for each fault) are the basis of the isolation scheme. The acceptance region of the hypothesis testing scheme corresponding to the k th fault, considering the raw estimator $\mu(t) = \epsilon(t)$ and the test size $\gamma = 0.05$, results in

$$\begin{aligned}
(l_k, u_k) &= \left(E[\epsilon(t); \Phi_k] - 1.96 \sqrt{\text{Var}(\epsilon(t); \Phi_k)}, E[\epsilon(t); \Phi_k] \right. \\
&\quad \left. + 1.96 \sqrt{\text{Var}(\epsilon(t); \Phi_k)} \right).
\end{aligned}$$

In this example, tests acceptance regions are simulated in steady state, so that their intersection from T_d is empty; hence, the isolation time will be upperbounded by $T_d + \Delta t + T_{ss}$, being T_{ss} an approximation of the time the estimator takes in reaching its steady state.

Fig. 1 shows, in two different scales, the time evolution of the estimator realization and the acceptance regions. As expected, the residual realization $\epsilon(t)$, once it gets its steady state, remains between the test limits corresponding to Φ_2 , and outside from the limits determined by the rest of tests, as it can be seen in Fig. 1 (up). Notice that initially the error crosses the acceptance region of the test corresponding to fault Φ_1 . It is necessary to wait until about $T_{iso} = 180$, the instant of time in which it is clear that the error remains uniquely in the acceptance region associated with Φ_2 (see Fig. 1 (down)).

5.2. Bidimensional example

This example illustrates the application of the proposed multi-dimensional FI approach, based on the SM-distance, to a bidimensional system.

The example consists of a circuit with a nonlinear resistor, an inductor and a capacitor joined in parallel to a noisy resistor (see Fig. 2). We consider that the circuit is in space or belongs to a nuclear plant, so that it could suffer two types of fault as a consequence of a radiation.

The system equations are:

$$\begin{aligned}
\frac{di(t)}{dt} &= -\frac{1}{L} V_C(t) - \frac{R}{L} i(t) - \frac{\tilde{R}}{L} i^3(t) + \frac{1}{L} V_0 + \frac{1}{L} \eta(t) + \beta_{k1}(t - T_0) \phi_{k1}(t) \\
\frac{dV_C(t)}{dt} &= \frac{1}{C} i(t) + \frac{1}{C} v(t) + \beta_{k2}(t - T_0) \phi_{k2}(t) \\
y(t) &= \begin{pmatrix} i(t) \\ V_C(t) \end{pmatrix}, \quad y(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad 0 \leq t.
\end{aligned}$$

At time $T_0 = 50$ (known for the simulation but obviously unknown for FDI tasks) a fault Φ_k occurs in the system; we consider that such fault can be one of the two $\{\Phi_1, \Phi_2\}$, given by the profile functions

$$\beta_{11}(t - T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 - e^{-\zeta(t-T_0)} & \text{if } t \geq T_0 \end{cases}$$

$$\beta_{22}(t - T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 & \text{if } t \geq T_0 \end{cases}$$

and the fault functions

$$\left\{ \phi_1(t) = \begin{pmatrix} -\frac{V_0}{L} \\ 0 \end{pmatrix}, \quad \phi_2(t) = \begin{pmatrix} 0 \\ M((t - T_0) - \frac{1}{C})v(t) \end{pmatrix} \right\}.$$

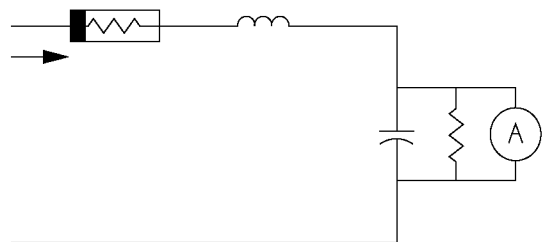


Fig. 2. Bidimensional example. Circuit.

The random processes $\eta(t)$ and $v(t)$ are mutually independent, and they are distributed as WGN, with zero mean and autocorrelation function

$$R_{WGN}(t_1, t_2) = \sigma_{WGN}^2 \delta(t_1 - t_2),$$

where, for the sake of simplicity, $\sigma_\eta^2 = \sigma_v^2 = 1$. Also, $R = L = C = M = 1$ (in their corresponding units), $V_0 = 7$ and $\varsigma = 0.01$ were chosen for the simulation,

Fault Φ_1 represents an incipient fault in the circuit provoked by the effect of radiation whose consequences show as a decrease in the initial constant voltage generated by the source voltage under healthy conditions. Abrupt fault Φ_2 could be caused due to a change in temperature around the capacitor, changing so the intensity of the thermal noise generated in the capacitor noisy resistor (from a constant value to a function evolving with time).

Following the indications of Section 3 we take the diagnostic observer,

$$\begin{aligned} \frac{d\hat{i}(t)}{dt} &= -\frac{1}{L}V_C(t) - \frac{R}{L}\hat{i}(t) - \frac{\tilde{R}}{L}\hat{i}^3(t) + \frac{1}{L}V_0 + \lambda_1(\hat{i}(t) - i(t)) \\ \frac{d\hat{V}_C(t)}{dt} &= \frac{1}{C}\hat{i}(t) + \lambda_2(\hat{V}_C(t) - V_C(t)) \\ \hat{y}(t) &= \begin{pmatrix} \hat{i}(t) \\ \hat{V}_C(t) \end{pmatrix}, \quad \hat{y}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad 0 \leq t. \end{aligned}$$

with $\lambda_1 = \lambda_2 = -0.5$.

Subtracting model and observer, the equations defining the residual vector given by the components

$$\epsilon_1(t) = i(t) - \hat{i}(t), \quad \epsilon_2(t) = V_C(t) - \hat{V}_C(t)$$

become

$$\begin{aligned} \frac{d\epsilon_1(t)}{dt} &= \lambda_1\epsilon_1(t) + \frac{1}{L}\eta(t) + \beta_{k1}(t - T_0)\phi_{k1}(t) \\ \frac{d\epsilon_2(t)}{dt} &= \lambda_2\epsilon_2(t) + \frac{1}{C}v(t) + \beta_{k2}(t - T_0)\phi_{k2}(t) \\ \epsilon(t) &= \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{pmatrix}, \quad \epsilon(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad 0 \leq t. \end{aligned}$$

This system is linear and uncoupled, so that each one of its differential equations can be solved separately.

Under the given assumptions on the stochastic processes $\eta(t)$ and $v(t)$, the residual $\epsilon(t)$ will be a bivariate Gaussian random process (totally determined by its mean vector and covariance matrix) and it can be written as

$$\epsilon(t) = \begin{pmatrix} \int_0^t e^{\lambda_1(t-\tau)} \left(\frac{1}{L}\eta(\tau) + \beta_{k1}(\tau - T_0)\phi_{k1}(\tau) \right) d\tau \\ \int_0^t e^{\lambda_2(t-\tau)} \left(\frac{1}{C}v(\tau) + \beta_{k2}(\tau - T_0)\phi_{k2}(\tau) \right) d\tau \end{pmatrix}$$

For the sake of a compromise between performance and efficiency we have chosen the mean estimator

$$\mu(t) = \begin{pmatrix} \frac{1}{T} \int_{t-T}^t \epsilon_1(\tau) d\tau \\ \frac{1}{T} \int_{t-T}^t \epsilon_2(\tau) d\tau \end{pmatrix},$$

where the window size T has been set to 10 for simulations. This process inherits the Gaussianity of the residual process due to the linearity of the integral. Therefore, in order to draw the proposed FI scheme, that is to determine the SM-distance corresponding to each possible fault, we determine the mean vector and the inverse of the covariance matrix for the estimator process:

$$E[\mu(t); \Phi_k] = \begin{pmatrix} \frac{1}{T} \int_{t-T}^t E[\epsilon_1(\tau); \Phi_k] d\tau \\ \frac{1}{T} \int_{t-T}^t E[\epsilon_2(\tau); \Phi_k] d\tau \end{pmatrix},$$

Table 1
Isolation times.

Fault Φ_1	$T_d = 57$ $T_d = 64.7$	$T_{iso} = 73$ $T_{iso} = 105$
Fault Φ_2	$T_d = 55$ $T_d = 67.4$	$T_{iso} = 67$ $T_{iso} = 78$

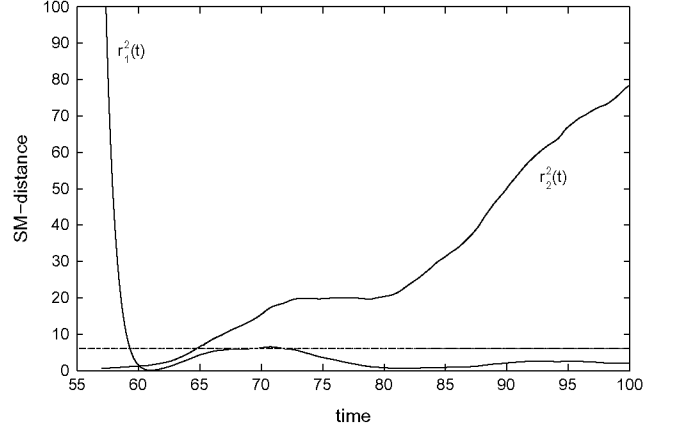


Fig. 3. Isolation: Bidimensional example. Fault Φ_1 . $T_d = 57$.

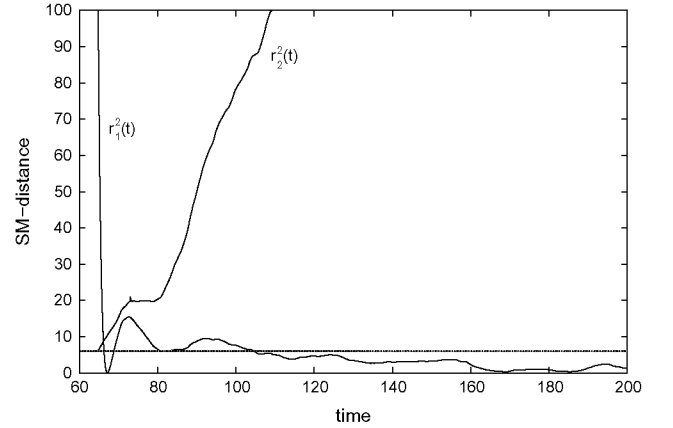


Fig. 4. Isolation: Bidimensional example. Fault Φ_1 . $T_d = 64.7$.

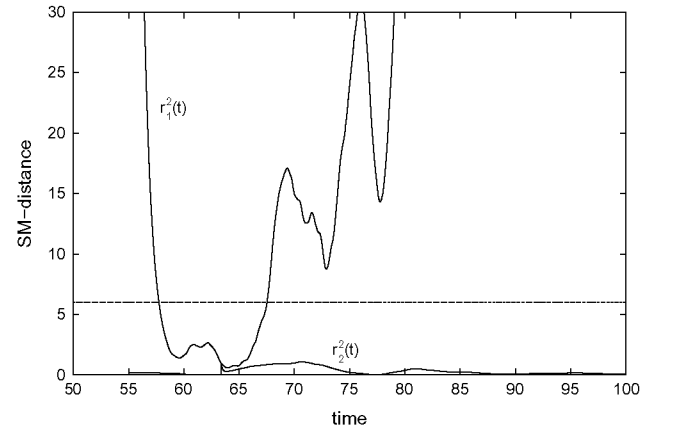


Fig. 5. Isolation: Bidimensional example. Fault Φ_2 . $T_d = 55$.

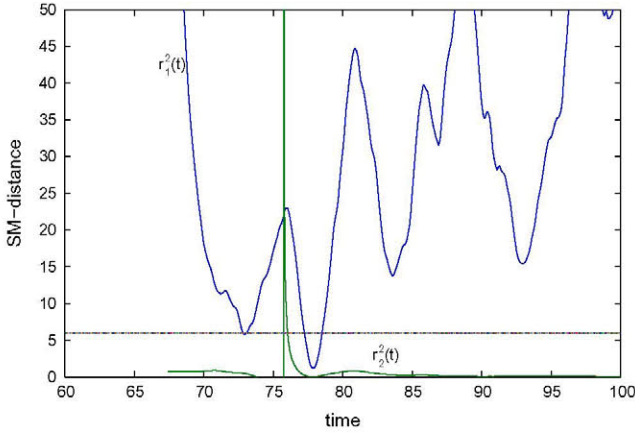


Fig. 6. Isolation: Bidimensional example. Fault Φ_2 . $T_d = 67.4$.

$$\begin{aligned} & \text{Cov}(\mu_{i_1}(t_1), \mu_{i_2}(t_2); \Phi_k) \\ &= \frac{1}{T^2} \int_{t_1-T}^{t_1} \int_{t_2-T}^{t_2} \text{Cov}(\epsilon_{i_1}(\tau_1), \epsilon_{i_2}(\tau_2)) d\tau_2 d\tau_1 \quad i_1, i_2 \in \{1, 2\}, \end{aligned}$$

$$\text{Cov}(\mu; \Phi_k)(t, t) = \begin{pmatrix} \text{Cov}(\mu_1(t), \mu_1(t); \Phi_k) & \text{Cov}(\mu_1(t), \mu_2(t); \Phi_k) \\ \text{Cov}(\mu_2(t), \mu_1(t); \Phi_k) & \text{Cov}(\mu_2(t), \mu_2(t); \Phi_k) \end{pmatrix},$$

to finally determine the SM-distance

$$r_k^2(t) = (\mu(t) - E[\mu(t); \Phi_k])' \text{Cov}^{-1}(\mu; \Phi_k)(t, t) (\mu(t) - E[\mu(t); \Phi_k]).$$

(The corresponding specific expressions are computed in Appendix B.)

5.2.1. Simulation results

In order to determine the regions defined in (5) (see Section 4.3) where k_γ is such that $B_k(t)$ fulfills the condition (6), the test size has been predefined to $\gamma = 0.05$, being so $k_\gamma = -2\ln(0.05) \approx 6$.

This example has been simulated for two detection methods with different response time, and the isolation method proposed in this work, using the SM-distance. The results are summarized in the following Table 1 and Figs. 3–6.

Note that the isolation method works even in case of large detection times; besides, as expected, this approach takes more time in isolating the incipient fault Φ_1 , specially for the case of slow detection.

Also, as a consequence of approximating $T_0 = 50$ by T_d (different in each case) the computation of the SM-distance suffers a transient phase, as it can be seen in Figs. 3–6.

6. Conclusion

In this paper the Fault Isolation task for a class of nonlinear stochastic dynamical systems has been addressed. A new fault isolation stochastic method has been proposed and framed in the context of statistical classification theory. Based on a model-based analytical redundancy scheme, a residual signal is obtained via a diagnostic observer, for its posterior analysis. Then the method determines the fault affecting the system through the application of a bank of hypotheses tests in parallel, one for each possible fault. This approach allows for the derivation of a rigorous isolability sufficient condition.

The method has been illustrated in two simulation examples corresponding respectively to unidimensional and bidimensional systems, where the isolation of both incipient and abrupt faults has been accomplished.

Acknowledgements

This work has been partially supported by projects MTM2010-15102 of Ministerio de Ciencia e Innovación, CCG10-UPM/ESP-5236 of Comunidad de Madrid/UPM, and Ayuda Q10 0930-144 of the Universidad Politécnica de Madrid (UPM), Spain.

Appendix A. Proof of Theorem 4.2

We address the simplest mean estimator

$$\mu(t) = \epsilon(t).$$

When a fault Φ is affecting the system, such estimator can be written as the product of a deterministic function and a stochastic process

$$\begin{aligned} \epsilon(t) &= \int_0^t e^{\lambda(t-\tau)} (\eta(\tau) + \beta(\tau - T_0)\phi(\tau)) d\tau \\ &= e^{\lambda t} \int_0^t e^{-\lambda\tau} (\eta(\tau) + \beta(\tau - T_0)\phi(\tau)) d\tau, \end{aligned}$$

by naming

$$Y(t) = \int_0^t e^{-\lambda\tau} (\eta(\tau) + \beta(\tau - T_0)\phi(\tau)) d\tau$$

the estimator can be rewritten as

$$\mu(t) = e^{\lambda t} Y(t).$$

Due to (14) and (15) we first prove that process $Z(t) = Y(t) - E[Y(t)]$ is a **martingale** (Larson and Shubert, 1979). As a consequence the probability of missed isolation in the interval $[T - \Delta t, T]$ will be bounded.

$Z(t)$ can be expressed as

$$Z(t) = \int_0^t e^{-\lambda\tau} (\eta(\tau) + \beta(\tau - T_0)\phi(\tau) - E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau)]) d\tau.$$

Let \mathcal{F}_t be the σ -algebra generated by $\{(\eta(\tau), \phi(\tau)), 0 \leq \tau \leq t\}$ (Billingsley, 1995). So,

$$\begin{aligned} & E[Z(t + \Delta t) / \mathcal{F}_t] \\ &= E \left[\int_0^t e^{-\lambda\tau} (\eta(\tau) + \beta(\tau - T_0)\phi(\tau) - E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau)]) d\tau \right. \\ & \quad \left. + \int_t^{t+\Delta t} e^{-\lambda\tau} (\eta(\tau) + \beta(\tau - T_0)\phi(\tau) - E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau)]) d\tau / \mathcal{F}_t \right] \\ &= E \left[\int_0^t e^{-\lambda\tau} (\eta(\tau) + \beta(\tau - T_0)\phi(\tau) - E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau)]) d\tau / \mathcal{F}_t \right] \\ & \quad + \int_t^{t+\Delta t} e^{-\lambda\tau} E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau) - E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau)] / \mathcal{F}_t] d\tau \stackrel{(14)}{=} Z(t) \\ & \quad + \int_t^{t+\Delta t} e^{-\lambda\tau} E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau) - E[\eta(\tau) + \beta(\tau - T_0)\phi(\tau)]] d\tau \\ &= Z(t) + 0 = Z(t), \end{aligned}$$

that is, $Z(t)$ is a **martingale** respect to the σ -algebra \mathcal{F}_t .

Note that since $Z(\cdot)$ depends on $\eta(\cdot)$ and on $\phi(\cdot)$, then the associated σ -algebra satisfies

$$\sigma(\{Z(\tau), 0 \leq \tau \leq t\}) \subseteq \mathcal{F}_t,$$

and by the martingales properties, if $Z(t)$ is a martingale with respect to \mathcal{F}_t then it is also a martingale with respect to a smaller σ -algebra (Billingsley, 1995), that is $Z(t)$ is a **martingale** with respect to $\sigma(\{Z(\tau), 0 \leq \tau \leq t\})$.

Coming back to the probability of missed isolation of fault Φ_j in the interval $[T - \Delta t, T]$ for a unidimensional system, applying the isolation approach described above with the raw estimator, can be upperbounded as follows,

$$\begin{aligned}
& P(\mu(t) \notin B_j(t) \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\
&= \underset{\text{unidim.}}{P\left(\frac{|\mu(t) - E[\mu(t); \Phi_j]|}{\sqrt{\text{Var}(\mu(t); \Phi_j)}} > h_{\frac{1}{2}} \text{ for some } t \in [T - \Delta t, T]; \Phi_j\right)} \\
&= P(|Y(t) - E[Y(t); \Phi_j]| > e^{-\lambda t} h_{\frac{1}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)} \text{ for some } t \in [T - \Delta t, T]; \Phi_j).
\end{aligned}$$

$$E[\mu(t); \Phi_1] = \begin{pmatrix} \frac{-V_0}{L} \frac{1}{T} \left[\frac{T}{\lambda_1} + (1 - e^{-\lambda_1 T}) \left(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_1(\lambda_1 + \zeta_1)} \right) e^{\lambda_1(t-T_0)} - \frac{1}{(\lambda_1 + \zeta_1)\zeta_1} (1 - e^{\zeta_1 T}) e^{-\zeta_1(t-T_0)} \right] \\ 0 \end{pmatrix}$$

$$\text{Cov}(\mu; \Phi_1)(t, t) = \begin{pmatrix} \frac{1}{L^2 T^2} \frac{-\sigma_{\eta}^2}{2\lambda_1^3} (2e^{\lambda_1 T} - e^{2\lambda_1 t} + 2e^{\lambda_1(2t-T)} - e^{2\lambda_1(t-T)} - 2\lambda_1 T - 2) & 0 \\ 0 & \frac{1}{C^2 T^2} \frac{-\sigma_{\xi}^2}{2\lambda_2^3} (2e^{\lambda_2 T} - e^{2\lambda_2 t} + 2e^{\lambda_2(2t-T)} - e^{2\lambda_2(t-T)} - 2\lambda_2 T - 2) \end{pmatrix}$$

Let us define the minimum

$$m_{[T-\Delta t, T]} = \min \left\{ e^{-\lambda t} h_{\frac{1}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)}, \quad t \in [T - \Delta t, T] \right\}$$

(note that it exists thanks to the continuity of the involved functions, and compactness of the interval).

Then,

$$\begin{aligned}
& P(|Y(t) - E[Y(t); \Phi_j]| > e^{-\lambda t} h_{\frac{1}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)} \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\
&\leq P(|Y(t) - E[Y(t); \Phi_j]| \geq m_{[T-\Delta t, T]} \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\
&= P(\max\{|Y(t) - E[Y(t); \Phi_j]|, t \in [T - \Delta t, T]\} \geq m_{[T-\Delta t, T]}; \Phi_j).
\end{aligned}$$

Besides that, since the uncertainty $\eta(t)$ and the fault process $\phi(t)$ are both assumed to be a linear combination of a MS continuous process and WGN, applying results from Larson and Shubert (1979) we can conclude that the process

$$Y(t) = \int_0^t e^{-\lambda \tau} \eta(\tau) + \beta(\tau - T_0) \phi(\tau) d\tau$$

is sample continuous. Consequently, the mean $E[Y(t)]$ is also a continuous function, and so the random process $Z(t) = Y(t) - E[Y(t)]$ is **sample continuous**.

Being $Z(t) = Y(t) - E[Y(t)]$ a martingale and a sample continuous process, we can apply a consequence of a basic martingale inequality (Castillo, 2006) to get a bound for the missed detection probability in the interval $[T - \Delta t, T]$. Namely,

$$P(\max\{|Z(t)|, t \in [T - \Delta t, T]\} \geq m_{[T-\Delta t, T]}) \leq \frac{E(Z^2(T))}{m_{[T-\Delta t, T]}^2} = \frac{\text{Var}_Y(T)}{m_{[T-\Delta t, T]}^2}.$$

Therefore and taking into account that

$$\text{Var}_{\mu}(t) = e^{2\lambda t} \text{Var}_Y(t),$$

we can conclude that the probability of missed detection in the interval $[T - \Delta t, T]$ is bounded as follows

$$\begin{aligned}
& P(\mu(t) \notin B_j(t) \text{ for some } t \in [T - \Delta t, T]; \Phi_j) \\
&= \underset{\text{unidim.}}{P\left(|\mu(t) - E[\mu(t); \Phi_j]| \geq h_{\frac{1}{2}} \sqrt{\text{Var}(\mu(t); \Phi_j)}, \right.} \\
&\quad \left. \text{for some } t \in [T - \Delta t, T]; \Phi_j\right) \leq \frac{e^{-2\lambda T} \text{Var}(\mu(T); \Phi_j)}{m_{[T-\Delta t, T]}^2}.
\end{aligned}$$

This procedure can be easily extended to estimators μ_b and μ_d , by noting that $\mu_b(t) = \frac{1}{T} Y_b(t)$ and $\mu_d(t) = e^{\rho t} \int_0^t e^{-\rho \tau} \epsilon(\tau) d\tau = e^{\rho t} Y_d(t)$, and applying an equivalent reasoning to $Y_b(t)$ and $Y_d(t)$. This procedure can be easily extended to estimators μ_b and μ_d , by noting that $\mu_b(t) = \frac{1}{T} Y_b(t)$ and $\mu_d(t) = e^{\rho t} \int_0^t e^{-\rho \tau} \epsilon(\tau) d\tau = e^{\rho t} Y_d(t)$, and applying an equivalent reasoning to $Y_b(t)$ and $Y_d(t)$.

Appendix B. Bidimensional example appendix

B.1. Estimator distribution assuming occurrence of fault Φ_1 : mean vector, variances and squared Mahalanobis distance

$$\begin{aligned}
r_1^2(t) &= (\mu(t) - E[\mu(t); \Phi_1])' \text{Cov}^{-1}(\mu; \Phi_1)(t, t) (\mu(t) - E[\mu(t); \Phi_1]) \\
&= \frac{\left[\frac{1}{T} \int_t^t \epsilon_1(\tau) d\tau - \frac{1}{T} \int_t^t \int_{T_0}^{\tau} e^{\lambda_1(\tau - \xi)} \beta_{11}(\xi - T_0) \frac{V_0}{L} d\xi d\tau \right]^2}{\frac{1}{L^2 T^2} \frac{-\sigma_{\eta}^2}{2\lambda_1^3} (2e^{\lambda_1 T} - e^{2\lambda_1 t} + 2e^{\lambda_1(2t-T)} - e^{2\lambda_1(t-T)} - 2\lambda_1 T - 2)} \\
&\quad + \frac{\left[\frac{1}{T} \int_t^t \epsilon_2(\tau) d\tau \right]^2}{\frac{1}{C^2 T^2} \frac{-\sigma_{\xi}^2}{2\lambda_2^3} (2e^{\lambda_2 T} - e^{2\lambda_2 t} + 2e^{\lambda_2(2t-T)} - e^{2\lambda_2(t-T)} - 2\lambda_2 T - 2)}
\end{aligned}$$

B.2. Estimator distribution assuming occurrence of fault Φ_2 : mean vector, covariances and squared Mahalanobis distance

$$E[\mu(t); \Phi_2] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\text{Cov}(\mu_1(t), \mu_1(t); \Phi_2) = \frac{1}{L^2 T^2} \frac{-\sigma_{\eta}^2}{2\lambda_1^3} (2e^{\lambda_1 T} - e^{2\lambda_1 t} + 2e^{\lambda_1(2t-T)} - e^{2\lambda_1(t-T)} - 2\lambda_1 T - 2)$$

$$\begin{aligned}
\text{Cov}(\mu_2(t), \mu_2(t); \Phi_2) &= \frac{1}{C^2 T^2} \frac{-\sigma_{\xi}^2}{2\lambda_2^3} (2e^{\lambda_2 T} - e^{2\lambda_2 t} + 2e^{\lambda_2(2t-T)} - e^{2\lambda_2(t-T)} - 2\lambda_2 T - 2) \\
&\quad + \frac{M^2 - \sigma_v^2}{T^2 \lambda_2^2} \left[\frac{(t - T - T_0 - \frac{1}{C})^2 (e^{\lambda_2 T} - 1)}{\lambda_2} \right. \\
&\quad \left. + \frac{2(e^{\lambda_2 T} - 1)(t - T - T_0 - \frac{1}{C})}{\lambda_2^2} - \left(t - T_0 - \frac{1}{C} \right)^2 \right. \\
&\quad \left. T + \left(T^2 + \frac{2T}{\lambda_2} \right) \left(t - T_0 - \frac{1}{C} \right) - \frac{T^3}{3} + \frac{T^2}{\lambda_2^2} + \frac{2(e^{\lambda_2 T} - 1)}{\lambda_2^3} \right] \\
&\quad + \frac{M^2 - \sigma_v^2}{T^2 \lambda_2^3} \left[-\frac{1}{C^2} + \left(\frac{1}{\lambda_2} + \frac{2}{CM} \right) \left(\frac{1}{C} + \frac{1}{-2\lambda_2} \right) \right] \\
&\quad \times e^{2\lambda_2 T_0} (e^{\lambda_2 t} - e^{\lambda_2(t-T)})^2 \\
&\quad + \left(\frac{2M}{C} + \frac{M^2}{\lambda_2} \right) \frac{2}{T^2} \frac{-\sigma_v^2}{2\lambda_2^2} \left[t \left(\frac{1}{-\lambda_2} + \frac{e^{\lambda_2 T}}{\lambda_2} - T \right) \right. \\
&\quad \left. + e^{\lambda_2 T} \left(\frac{T + T_0 + \frac{1}{C} - \frac{1}{2\lambda_2}}{-\lambda_2} + \frac{1}{(-\lambda_2)^2} \right) + \frac{T^2}{2} \right. \\
&\quad \left. + \left(T_0 + \frac{1}{C} - \frac{1}{2\lambda_2} \right) \left(T + \frac{1}{\lambda_2} \right) - \frac{1}{(-\lambda_2)^2} \right]
\end{aligned}$$

$$\begin{aligned}
r_2^2(t) &= (\mu(t) - E[\mu(t); \Phi_2])' \text{Cov}^{-1}(\mu; \Phi_2)(t, t) (\mu(t) - E[\mu(t); \Phi_2]) \\
&= \frac{\left[\frac{1}{T} \int_{t-T}^t \epsilon_1(\tau) d\tau \right]^2}{\text{Cov}(\mu_1(t), \mu_1(t); \Phi_2)} + \frac{\left[\frac{1}{T} \int_{t-T}^t \epsilon_2(\tau) d\tau \right]^2}{\text{Cov}(\mu_2(t), \mu_2(t); \Phi_2)}
\end{aligned}$$

References

- Alcorta-García, E., & Frank, P. M. (1997). Deterministic nonlinear observer-based approaches to fault diagnosis: A survey. *IFAC Control Engineering Practice*, 5, 663–670.
- Arató, M. (1982). Linear stochastic systems with constant coefficients. In *A statistical approach. Lecture notes in control and information sciences* (Vol. 45). Berlin-Heidelberg-New York: Springer Verlag.
- Aström, K. J., Albertos, P., Isidori, A., Schaufelberger, W., & Sanz, R. (Eds.). (2001). *Control of complex systems*. Berlin-Heidelberg-New York: Springer Verlag.
- Basseville, M., & Nikiforov, I. V. (1993). *Detection of abrupt changes: Theory and application*. Englewood Cliffs, NJ: Prentice Hall.
- Billingsley, P. (1995). *Probability and measure*. New York, NY: Wiley-Interscience.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Singapore: Springer.
- Blanke, M., Kinnaert, M., Lunze, J., & Staroswiecki, M. (2003). *Diagnosis and fault-tolerant control*. Berlin: Springer Verlag.
- Castillo Á. (2006). *Fault detection and isolation via continuous time statistics*. Ph.D. Thesis, E.T.S. Ingenieros Industriales (Universidad Politécnica de Madrid).
- Castillo, Á., Zufiria, P. J., Polycarpou, M. M., Previdi, F., & Parisini, T. (2003). Fault detection and isolation scheme in continuous time nonlinear stochastic systems. In *Proceedings of the 5th IFAC symposium on fault detection, supervision and safety of technical processes SAFEPROCESS 2003* (pp. 651–656).
- Castillo, Á., & Zufiria, P. J. (2009). Fault detection schemes for continuous-time stochastic dynamical systems. *IEEE Transactions on Automatic Control*, 54(8), 1820–1836.
- Chen, J., & Patton, R. J. (1999). *Robust model-based fault diagnosis for dynamic systems*. London: Kluwer Academic Publishers.
- Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault detection and diagnosis in industrial systems. Advanced books in control and signal processing*. London: Springer Verlag.
- Darkhovski, B. S., & Staroswiecki, M. (2003). The fault detection problem for dynamic systems: General results. In *Proceedings of the 5th IFAC symposium on fault detection, supervision and safety of technical processes SAFEPROCESS 2003* (pp. 193–198). Washington, DC, USA.
- De Persis, C., & Isidori, A. (2001). A geometric approach to nonlinear fault detection and isolation. *IEEE Transactions on Automatic Control*, 46, 853–865.
- Ding, S. X. (2008). *Model-based fault diagnosis techniques. design schemes, algorithms and tools*. Berlin: Springer Verlag.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York, NY: John Wiley and Sons, Inc..
- Frank, P. M. (1996). Analytical and qualitative model-based fault diagnosis – A survey and some new results. *European Journal of Control*, 1(2), 26–28.
- Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. San Diego, CA: Academic Press.
- Gertler, J. J. (1998). *Fault detection and diagnosis in engineering systems*. New York, NY: Marcel Dekker.
- Iserman, R. (2006). *Fault-diagnosis systems. An introduction to fault detection and fault tolerance*. Berlin: Springer Verlag.
- Jiang, B., Cocquemot, V., & Christophe, C. (2002). Fault diagnosis using sliding mode observer for nonlinear systems. In *Proceedings of IFAC'02 15th triennial world congress*. Barcelona, Spain.
- Kailath, T., & Poor, H. V. (1998). Detection of stochastic processes. *IEEE Transactions on Information Theory*, 44(6), 2230–2259.
- Korbicz, J., Kościelny, J. M., Kowalczyk, Z., & Cholewa, W. (Eds.). (2004). *Fault diagnosis. models, artificial intelligence, applications*. Berlin-Heidelberg-New York: Springer Verlag.
- Larson, H. M., & Shubert, B. O. (1979). *Probabilistic models in engineering sciences. Random noise, signals and dynamic systems* (Vol. II). John Wiley and Sons.
- Larson, H. M., & Shubert, B. O. (1979). *Probabilistic models in engineering sciences. Random variables and stochastic processes* (Vol. I). John Wiley and Sons.
- Li, L., & Zhou, D. (2004). Fast and robust fault diagnosis for a class of nonlinear systems: detectability analysis. *Computers and Chemical Engineering*, 28, 2635–2646.
- Mattone, R., & De Luca, A. (2006). Relaxed fault detection and isolation: An application to a nonlinear case study. *Automatica*, 42, 109–116.
- Münz, U., & Zufiria, P. J. (2005). Parametric fault diagnosis in stochastic dynamical systems. In *Proceedings of the 19th CEDYA*.
- Münz, U., & Zufiria, P. J. (2009). Diagnosis of unknown parametric faults in nonlinear stochastic dynamical systems. *International Journal of Control*, 82(4), 603–619.
- Nise, N. N. (2011). *Control systems engineering* (6th ed.). NJ, USA: John Wiley and Sons, Inc..
- Palade, V., Bocaniala, C. D., & Jain, L. C. (Eds.). (2006). *Computational intelligence in fault diagnosis*. London: Springer Verlag.
- Polycarpou, M. M., & Helmicki, A. (1995). Automated fault detection and accommodation: A learning systems approach. *IEEE Transaction on Systems, Man, and Cybernetics*, 25(11), 1447–1458.
- Polycarpou, M. M., & Trunov, A. B. (2000). Learning approach to nonlinear fault diagnosis: Detectability analysis. *IEEE Transactions on Automatic Control*, 45, 806–812.
- Polycarpou, M. M., & Vemuri, A. T. (1995). Learning methodology for failure detection and accommodation. *IEEE Control Systems*, 16–24.
- Poor, H. V., & Hadjiliadis, O. (2009). *Quickest detection*. Cambridge: Cambridge University Press.
- Reble, M., Münz, U., & Allgower, F. (2007). Diagnosis of parametric faults in multivariable nonlinear systems. In *Proceedings of the 46th IEEE conference on decision and control conference 2007* (pp. 336–371).
- Simani, S., Fantuzzi, C., & Patton, R. J. (2003). *Model-based fault diagnosis in dynamic systems using identification techniques*. London: Springer Verlag.
- Tafazoli, S., & Sun, X. (2006). Hybrid system state tracking and fault detection using particle filters. *IEEE Transactions on Control Systems and Technology*, 14, 1078–1087.
- Trunov, A. B., & Polycarpou, M. M. (2000). Automated fault diagnosis in nonlinear multivariable systems using a learning methodology. *IEEE Transaction on Neural Networks*, 11(1), 91–101.
- Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003). A review of process fault detection and diagnosis. Part i: Quantitative model-based methods. *Computer and Chemical Engineering*, 27, 293–311.
- Willsky, A. S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica*, 12, 601–611.
- Willsky, A. S., & Jones, H. L. (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, 21(February), 08–112.
- Witczak, M., Korbicz, J., & Patton, R. (2002). A bounded-error approach to designing unknown input observers. In *Proceedings of IFAC'02 15th triennial world congress*. Barcelona, Spain.
- Xu, A., & Zhang, Q. (2004). Nonlinear system fault diagnosis based on adaptive estimation. *Automatica*, 40, 1181–1193.
- Zhang, Q., Campillo, F., Crou, F., & LeGland, F. (2005). Nonlinear system fault detection and isolation based on bootstrap particle filters. In *Proceedings of the 44th IEEE conference on decision and control and the European control conference 2005* (pp. 3821–3826).
- Zhang, X., Parisini, T., & Polycarpou, M. M. (2004). Adaptive fault-tolerant control of nonlinear uncertain systems: An information-based diagnostic approach. *IEEE Transactions on Automatic Control*, 49(8), 1259–1273.
- Zhang, X., Parisini, T., & Polycarpou, M. M. (2005). Sensor bias fault isolation in a class of nonlinear systems. *IEEE Transactions on Automatic Control*, 50(3), 370–376.
- Zhang, X., Polycarpou, M. M., & Parisini, T. (2002). A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. *IEEE Transactions on Automatic Control*, 47, 576–593.
- Zhang, X., Polycarpou, M. M., & Parisini, T. (2008). Design and analysis of a fault isolation scheme for a class of uncertain nonlinear systems. *Annual Reviews in Control*, 32, 107–121.
- Zufiria, P. J. (2009). A formulation for fault detection in stochastic continuous-time dynamical systems. *International Journal of Computer Mathematics*, 86(10-11), 1778–1797.
- Zufiria, P. J. (2012). A mathematical framework for new fault detection schemes in nonlinear stochastic continuous-time dynamical systems. *Applied Mathematics and Computation*, 218(23), 11391–11403.

Ángela Castillo was born in Albacete (Spain) on February 23, 1975. She received a M.Sc. degree in Mathematics from the *Universidad Complutense de Madrid*, in 1998. She received the Ph.D. degree in Applied Mathematics from the *Universidad Politécnica de Madrid* in 2006. She joined the *Matemática Aplicada a las T.T. I. Department (E.T.S.I. Telecomunicación-UPM)* in 2001 where she is presently Associate Professor. She teaches Mathematics foundations and her field of research is Fault Diagnosis.

Pedro J. Zufiria received the Ingeniero de Telecomunicación degree by the Technical University of Madrid (UPM) in 1986, the M.Sc. in mechanical engineering, M.Sc. in electrical engineering, and Ph.D. degrees from the University of Southern California (USC) in 1989. He also received the Doctor Ingeniero de Telecomunicación degree by the Ministry of Education and Science (MEC) in 1991 and the Licenciado en Ciencias Matemáticas degree by the Complutense University of Madrid (UCM) in 1997. He is Professor in the Telecommunications Engineering School (ETSIT) of the UPM, being also the scientific responsible of the Dynamical Systems, Learning and Control Group. His research interests focus on the control and fault diagnosis of dynamical systems, as well as the study of learning paradigms for applications in network science and data processing.