# A Robust Detection and Isolation Scheme for Abrupt and Incipient Faults in Nonlinear Systems

Xiaodong Zhang, Marios M. Polycarpou, and Thomas Parisini

***Abstract*—This paper presents a robust fault diagnosis scheme for abrupt and incipient faults in nonlinear uncertain dynamic systems. A detection and approximation estimator is used for online health monitoring. Once a fault is detected, a bank of isolation estimators is activated for the purpose of fault isolation. A key design issue of the proposed fault isolation scheme is the adaptive residual threshold associated with each isolation estimator. A fault that has occurred can be isolated if the residual associated with the matched isolation estimator remains below its corresponding adaptive threshold, whereas at least one of the components of the residuals associated with all the other estimators exceeds its threshold at some finite time. Based on the class of nonlinear uncertain systems under consideration, an isolation decision scheme is devised and fault isolability conditions are given, characterizing the class of nonlinear faults that are isolable by the robust fault isolation scheme. The nonconservativeness of the fault isolability conditions is illustrated by deriving a subclass of nonlinear systems and of faults for which these conditions are also necessary for fault isolability. Moreover, the analysis of the proposed fault isolation scheme provides rigorous analytical results concerning the fault isolation time. Two simulation examples are given to show the effectiveness of the fault diagnosis methodology.**

***Index Terms*—Fault detection and approximation, fault isolation, nonlinear adaptive estimator, nonlinear uncertain systems.**

## NOMENCLATURE

| | |
|---|---|
| $f$ | Nominal model dynamics. |
| $\eta$ | Modeling uncertainty. |
| $\bar{\eta}_i$ | Known bound on the $i$th component of the modeling uncertainty. |
| $\phi$ | Fault vector function. |
| $\mathcal{B}$ | Fault time-profile matrix function. |
| $\alpha_i$ | Incipient-fault evolution rate in the $i$th state equation. |
| $\bar{\alpha}_i$ | Known lower bound on $\alpha_i$. |
| $\mathcal{F}$ | Class of faults. |
| $\theta_i^s$ | Parameter vector associated with the $s$th fault affecting the $i$th state equation. |
| $\hat{\theta}_i^s$ | Estimate of the parameter vector $\theta_i^s$. |
| $\Theta_i^s$ | Known compact set to which $\theta_i^s$ belongs. |
| $g_i^s$ | Known vector field associated with the $s$th fault affecting the $i$th state equation. |
| $\Lambda^s$ | Diagonal matrix of the poles associated with the $s$th estimator. |
| $\hat{\phi}$ | Online fault approximation model. |
| $\epsilon_i^s$ | $i$th component of the state estimation error associated with the $s$th estimator. |
| $\bar{\epsilon}_i^0$ | $i$th component of the dead-zone threshold associated with the fault detection and approximation estimator. |
| $\mu_i^s$ | $i$th component of the adaptive threshold associated with the $s$th estimator. |
| $\xi_i^s$ | $i$th component of the fault approximation error associated with the $s$th estimator in the case that fault $s$ occurs. |
| $\kappa_i^s$ | Computable bound on the $i$th component of the parameter vector estimation error in the case of a matched fault function. |
| $h_i^{sr}$ | Fault mismatch function between the $s$th and $r$th faults in the case of an incipient fault. |
| $\tilde{h}_i^{sr}$ | Fault mismatch function between the $s$th and $r$th faults in the case of an abrupt fault. |
| $T_d$ | Absolute fault detection time. |
| $t_d$ | Fault detection time. |
| $T_{\mathbf{isol}}^s$ | Absolute fault isolation time associated with the $s$th fault. |
| $t_{\mathbf{isol}}^s$ | Fault isolation time associated with the $s$th fault. |
| $\bar{t}_{\mathbf{isol}}^s$ | Maximum fault isolation time associated with the $s$th fault. |

Manuscript received December 12, 1999; revised July 30, 2000, March 2, 2001, and August 2, 2001. Recommended by Associate Editor M. Krstic. This work was supported in part by the EU RTN Project "DAMADICS," in part by the MURST Project "Identification and Control of Industrial Systems," and in part by the Italian Space Agency.

X. Zhang is with Intelligent Automation, Inc., Rockville, MD 20855 USA (e-mail: xzhang@i-a-i.com).

M. M. Polycarpou is with the Department of Electrical and Computer Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221-0030 USA, and also with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus (e-mail: mpolycar@ececs.uc.edu).

T. Parisini is with the Department of Electrical, Electronic and Computer Engineering, University of Trieste, 34127 Trieste, Italy (e-mail: parisini@univ.trieste.it).

Publisher Item Identifier S 0018-9286(02)03737-6.

## I. INTRODUCTION

A FAULT diagnosis procedure is typically divided into three tasks: i) *fault detection* indicates the occurrence of a fault in a monitored system; ii) *fault isolation* establishes the type and/or location of the fault; and iii) *fault identification* determines the magnitude of the fault. After a fault has been detected and diagnosed, in some applications it is required that the fault be self-corrected, usually through controller reconfiguration. This is usually referred to as *fault accommodation*. The design and analysis of fault detection and isolation (FDI) algorithms using the model-based analytical redundancy approach have received significant attention in the literature (see, for example, the survey papers by Frank [9], Gertler [13], and Isermann [20] and the books by Patton *et al.* [36], Gertler [15], and Chen and Patton [3]).

The objective of this paper is the design and analysis of a fault isolation scheme for nonlinear uncertain systems. Unlike the fault detection problem, which has been extensively investigated in the literature, the fault isolation problem has received less attention, especially in the case of *nonlinear uncertain systems*. Some of the approaches that have been examined for fault isolation in linear systems include the utilization of *structured residuals* and *fixed directional residuals* [15], which can be generated by observer-based methods or parity relations. For example, the unknown input observer approach [9], [38] and the eigenstructure assignment method [34] have been used to generate structured residuals for fault isolation in linear systems, whereas fault detection filters [29], [32], [50] have been used for fixed directional residuals. Structured and directional residuals can also be generated via parity relations for fault isolation [14], [16]. The equivalence between diagnostic observers and parity equations is discussed in [14].

In recent years, there has been considerable research activity aimed at the design and analysis of fault diagnosis schemes specific for nonlinear systems [3], [12], [26]. Several researchers have developed nonlinear fault diagnosis schemes based on nonlinear observer approaches. In [11], the unknown input observer approach has been extended to include nonlinear terms. A class of nonlinear systems that has attracted a lot of attention is that of systems with bilinear dynamics [23], [53], [55]. Some studies have attempted to extend the parity relations approach to nonlinear systems [25], [27]. Recently, there has been significant activity and some exciting results [18], [37] have been obtained in addressing the FDI problem in the case of nonlinear systems in which the structured modeling uncertainty and faults can be decoupled. Adaptive and online approximation approaches to nonlinear fault diagnosis have also been developed [8], [39], [44]–[47], [54]. These techniques are based on the idea of online adaptation and approximation of the fault function. One of the tools that have been widely used is represented by an *online approximation model*, which is usually in the form of a neural network, a fuzzy logic system, etc. Despite these promising approaches to addressing the problem of fault diagnosis in a nonlinear framework, there have not been many analytical results on fault isolation, especially in the case of unstructured modeling uncertainty and nonlinear faults, which cannot be exactly decoupled from each other.

In this paper, we present a fault detection and isolation architecture for nonlinear uncertain dynamic systems, and provide a rigorous analysis of the performance properties of the related isolation scheme. The class of faults considered is allowed to be nonlinear with respect to the state and input, and includes both abrupt and incipient faults. We consider a class of nonlinear systems with full-state measurements and the presence of possibly nonlinear and unstructured modeling uncertainty. The proposed FDI scheme consists of a bank of nonlinear adaptive estimators. One of them is the *fault detection and approximation estimator*, whereas the others are used for fault isolation (each associated with a specific type of fault). Under normal operating conditions, only the *detection and approximation estimator* is used to monitor the process for any fault. Once a fault is detected, the *fault isolation estimators* are activated, and the fault detection and approximation estimator adopts the mode of approximating the fault, by using online approximation methods.

The main contributions of this research are the design of a fault isolation scheme as the key part of a diagnosis architecture based on a nonlinear framework justified by practical considerations, and the analysis of the proposed isolation scheme in terms of derivation of adaptive threshold functions, fault isolability conditions, and fault isolation time. The residual of each fault isolation estimator is associated with an *adaptive threshold*, which can be implemented online by using linear filtering methods. The case of the occurrence of a particular fault is excluded if at least one of the residual components of the corresponding isolation estimator exceeds its threshold in a finite time. Fault isolation is achieved when all faults but one are excluded. Under the imposed assumptions, an incorrect isolation decision is precluded. However, two faults may be nonisolable if the two fault functions are not "sufficiently different." This concept is formalized by the definition of the so-called *fault mismatch function*.

The presented fault isolation analysis consists of three parts: i) derivation of adaptive thresholds; ii) investigation of fault isolability conditions; and iii) computation of the fault isolation time. The derived adaptive thresholds ensure that an incorrect isolation decision will be avoided. This is achieved by selecting an adaptive threshold for each possible fault such that the residual associated with the isolation estimator that matches the occurred fault is guaranteed to remain below its threshold. In the design of adaptive thresholds, there is always a tradeoff between false alarms and missed faults. The analysis of fault isolability conditions characterizes (in nonclosed form) the class of faults that can be isolated by the isolation scheme. This class is rigorously characterized by the *fault mismatch function*, which intuitively provides a measure of the difference between two faults. The nonconservativeness of fault isolability conditions is illustrated by the derivation of a subclass of nonlinear systems and faults for which these conditions are also necessary for fault isolability. The *fault isolation time* is defined as the length of the time interval between the detection of a fault and the determination of its type. For the proposed fault isolation scheme, an upper bound on the fault isolation time is derived. The design scheme and the analytical results are described through the use of two nonlinear simulation examples. The first deals with a simple second-order nonlinear system, whereas the second example refers to the well-known FDI benchmark problem concerning a three-tank system.

The paper is organized as follows. Section II defines the classes of nonlinear systems and faults to be investigated. The design of the proposed FDI scheme, including the derivation of adaptive thresholds, is described in Section III. Section IV analyzes the fault isolability conditions on the robust fault isolation scheme. In Section V, the fault isolation time is addressed. Finally, the FDI scheme design and the analytical results are illustrated by two simulation examples in Section VI, and Section VII contains some concluding remarks.

## II. Problem Formulation

In this section, we formulate the classes of nonlinear systems and faults to be investigated, and discuss the practical motivation of the proposed formulation.

### A. Nominal Plant, Uncertainty, and Fault Representation

Let us consider a general multivariable nonlinear dynamic system described by the differential equation

$$\dot{x} = f(x,u) + \eta(x,u,t) + \mathcal{B}(t-T_0)\phi(x,u) \qquad (1)$$

where $x \in \Re^n$ is the state vector of the system, $u \in \Re^m$ is the input vector, $f, \phi$: $\Re^n \times \Re^m \mapsto \Re^n$, and $\eta$: $\Re^n \times \Re^m \times \Re^+ \mapsto \Re^n$ are smooth vector fields, and $\mathcal{B}(t-T_0)$ is a matrix function representing the time profiles of the faults, where $T_0$ denotes the unknown fault occurrence time. The vector fields $f$, $\eta$, and $\phi$ represent the dynamics of the nominal model, the modeling uncertainty, and the change in the system dynamics due to a fault, respectively. For the sake of well-posedness of (1), the following assumption will be made.

*Assumption 1:* The system states and controls remain bounded before and after the occurrence of a fault, i.e., there exists some stability region $\mathcal{D} \subset \Re^n \times \Re^m$, such that $(x(t), u(t)) \in \mathcal{D}, \forall\, t \geq 0$.

*Remark 2.1:* It is worth noting that the reason for introducing such a uniform boundedness assumption is just a formal one. In general, this paper deals with the design and analysis of a detection and isolation scheme based on the measurements of $x(t)$ and $u(t)$. Since no fault accommodation is considered in the paper, the feedback controller must be such that the measurable signals $x(t)$ and $u(t)$ remain bounded for all $t \geq 0$ (i.e., before and after the occurrence of a fault). However, it is important to note that the proposed FDI design is not dependent on the structure of the controller. Actually, as will be clear later on, the proposed fault diagnosis scheme will make use of $x(t)$ and $u(t)$ to yield the detection and isolation decisions, but it will not influence at all the dynamic behavior of system (1).

The modeling uncertainty, represented by the vector field $\eta$, includes external disturbances as well as modeling errors. In the fault-diagnosis literature, efforts to enhance the robustness of FDI schemes can be made either at the residual generation stage by using *decoupling techniques* or at the decision making stage by using *adaptive thresholds*. In the first approach, the modeling uncertainty is often assumed to be structured, i.e., to be of the form $\eta = Ew(t)$, where $E$ is a *known* (or approximately known) and not necessarily constant distribution matrix, and $w$ denotes an unknown function of time. This structured model of uncertainty allows the use of linear and nonlinear state transformations to exactly decouple faults from unknown inputs [23], [37], [41], [52]. In the cases where such a decoupling framework can be achieved, it provides powerful methods for developing FDI algorithms. However, if the modeling uncertainty is *unstructured*, decoupling faults from modeling uncertainty is not possible and this justifies the use of adaptive thresholds to obtain robustness at the residual-evaluation stage. In the adaptive threshold approach [3], [7], [10], modeling uncertainty can be unstructured but has to be bounded by some suitable constant or function. This bound is used to derive thresholds for distinguishing between the effect of a fault and the effect of modeling uncertainty [7], [45]–[47]. Another important approach that has been extensively used to represent modeling uncertainty in fault diagnosis is the formulation of the problem in a stochastic framework [2], [30]. The FDI scheme presented in this work is based on the adaptivethreshold approach.

As regards modeling uncertainty, the following assumption will be used throughout the paper.

*Assumption 2:* The modeling uncertainty represented by the vector field $\eta$ in (1) is unstructured and possibly a unknown nonlinear function of $x$, $u$, and $t$, but it is bounded by some known functional, i.e.,

$$|\eta_i(x,u,t)| \leq \bar{\eta}_i(x,u,t), \quad \forall\,(x,u) \in \bar{\mathcal{D}}, \ \forall\, t \geq 0 \qquad (2)$$

where, for each $i = 1,\ldots,n$, the bounding function $\bar{\eta}_i(x,u,t) > 0$ is known, integrable, and bounded for all $(x,u)$ in some compact region of interest $\bar{\mathcal{D}} \supseteq \mathcal{D}$ and for all $t \geq 0$.

*Remark 2.2:* The above assumption provides nonuniform bounding functions on the modeling uncertainty $\eta(x,u,t)$ in some compact region $\bar{\mathcal{D}} \supseteq \mathcal{D}$, where $\mathcal{D}$ is defined in Assumption 1. It is worth noting that a simpler, though more restrictive assumption, would be to assume that $\eta$ is globally uniformly bounded, i.e., $|\eta_i(x,u,t)| \leq \bar{\eta}_i, \forall\,(x,u) \in \Re^n \times \Re^m$ and $\forall\ t \geq 0$, where $\bar{\eta}_i$ is a constant bound. It is important to emphasize that by allowing each $\bar{\eta}_i$ to be a function of $x$, $u$, and $t$, the above formulation provides a framework for *nonuniform bounds*, thus enhancing the achievable fault sensitivity and decreasing the detection and isolation times. For example, in many practical applications the nominal model is obtained by small-signal linearization techniques (around a nominal operating point or trajectory). In this case, $\eta(x,u,t)$ may represent the residual nonlinear terms, which are typically small for $(x,u)$ close to the operating point but can be large elsewhere. If nonuniform bounds $\bar{\eta}_i(x,u,t)$ are not known, the designer can consider the worst-case scenario and use uniform *constant* bounds $\bar{\eta}_i$ as a special case.

As to the faults affecting the nominal system modes, from a qualitative viewpoint, the term $\mathcal{B}(t-T_0)\phi(x,u)$ represents the deviation in the system dynamics due to a fault. The matrix $\mathcal{B}(t-T_0)$ characterizes the time profile of a fault that occurs at some *unknown* time $T_0$, and $\phi(x,u)$ denotes the nonlinear fault function. This characterization allows both additive and multiplicative faults (since $\phi$ is a function of $x$ and $u$) [15], and even more general nonlinear faults. We let the fault time profile $\mathcal{B}(\cdot)$ be a diagonal matrix of the form

$$\mathcal{B}(t-T_0) \triangleq \operatorname{diag}\left[\beta_1(t-T_0),\ldots,\beta_n(t-T_0)\right]$$

where $\beta_i$: $\Re \mapsto \Re$ is a function representing the time profile of a fault affecting the $i$-state equation, for $i = 1,\ldots,n$. More specifically, we consider faults with time profiles modeled by

$$\beta_i(t-T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 - e^{-\alpha_i(t-T_0)} & \text{if } t \geq T_0 \end{cases} \qquad (3)$$

where the scalar $\alpha_i > 0$ denotes the unknown fault evolution rate. Small values of $\alpha_i$ characterize slowly developing faults, also known as *incipient faults*. For large values of $\alpha_i$, the time profile $\beta_i$ approaches a step function, which models *abrupt faults*. The main difficulty in dealing with incipient faults is that their small effects on the residuals can be hidden as if they are due to modeling uncertainty. The incipient-fault time profile described by (3) has been considered in [6], [40], [44]

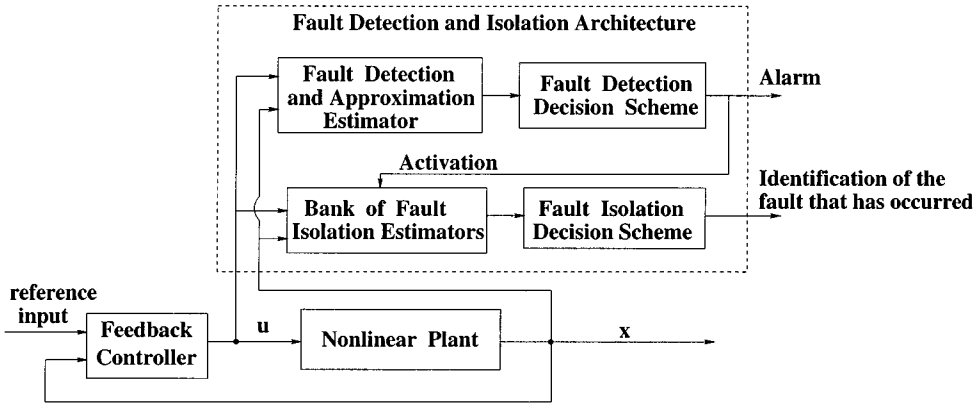


Fig. 1. Architecture of the fault detection and isolation scheme.

to develop a learning-based fault detection methodology. In this paper, we first consider the fault isolation problem in the general case of incipient faults, and then examine the special case of abrupt faults.

Note that the fault time profile given by (3) only reflects the developing speed of the fault, while all its other basic features are captured by the function $\phi(x,u)$ described below. For isolation purposes, we assume that there are $N$ types of possible nonlinear fault functions; specifically, $\phi(x,u)$ belongs to a finite set of functions given by

$$\mathcal{F} \triangleq \left\{\phi^1(x,u),\ldots,\phi^N(x,u)\right\}. \tag{4}$$

Each fault function $\phi^s$, $s=1,\ldots,N$, is described by

$$\phi^s(x,u) \triangleq \left[(\theta_1^s)^T g_1^s(x,u),\ldots,(\theta_n^s)^T g_n^s(x,u)\right]^T \tag{5}$$

where $\theta_i^s$, $i=1,\ldots,n$, is an unknown $q_i^s$-dimensional parameter vector assumed to belong to a known compact set $\Theta_i^s$ (i.e., $\theta_i^s \in \Theta_i^s \subset \Re^{q_i^s}$), and $g_i^s: \Re^n \times \Re^m \mapsto \Re^{q_i^s}$ is a known smooth vector field.

This representation characterizes a general class of nonlinear faults where the nonlinear vector field $g_i^s$ represents the functional structure of the $s$th fault affecting the $i$th state equation, whereas the unknown parameter vector $\theta_i^s$ characterizes the "magnitude" of the fault in the $i$th state equation. The dimension $q_i^s$ of each parameter vector $\theta_i^s$ is determined by both the type of fault and the specific state component considered. In the case where the fault function $\phi(x,u)$ is completely unknown (i.e., $\phi(x,u)$ does not belong to $\mathcal{F}$), the fault approximation estimator designed in Section III by approximation methods can be used to reconstruct online the unknown fault function.

As discussed in [13], most practical faults are nonlinear functions of the system state $x$ and/or input $u$. For example, the magnitude of a leak in a thermal system or in a chemical process is, in general, a nonlinear function of the pressure and the temperature. Such failure representation characteristics are captured in (1) by allowing the deviation $\phi$ to be a nonlinear function of $x$ and $u$. Moreover, it is worth noting that the above formulation allows parametric faults [15] and, in addition, other types of nonlinear faults such as the ones that cause the nominal plant model to change from $f(x,u)$ to another new nonlinear function.

*Remark 2.3:* In many engineering applications, the full-state measurement assumption in the above formulation may result in quite a critical and possibly limiting requirement. The removal of this assumption requires the use of nonlinear observers, which, in general, impose additional restrictions on the class of nonlinear systems and the type of allowable faults [45]. Note, however, that several nonlinear control design methods require full-state measurements for the design of the feedback controller. Such methods include feedback linearization [22], backstepping and adaptive backstepping methods [28], input-to-state stability (ISS) control design [42], and robust nonlinear control using nonlinear damping [5]. Therefore, the nonlinear fault-isolation design method developed in this paper can be applied to such a class of feedback control systems.

*Remark 2.4:* Typically, a robust feedback control system may "hide" the occurrence of a fault, especially a small, incipient one. While, in some cases, it is desirable to automatically accommodate a small fault by using the robustness of the controller, in most situations small faults may prognosticate future larger faults that can result in catastrophic consequences, unless they are detected and accommodated early. This problem of robust feedback control "hiding" or "desensitizing" fault effects has been recognized by several researchers (see, for example, [17], [35], and [51]). Allowing the fault function $\phi$ in the above formulation to depend explicitly on $u$ provides a suitable method for detecting faults, even if the control input $u$ has been adjusted to reduce the effect of the fault on the tracking error. Another approach that has been proposed to address this problem is based on designing the fault diagnosis scheme and the feedback controller simultaneously [31], [43].

## III. Fault Detection and Isolation Architecture

A bank of $N+1$ nonlinear adaptive estimators are used in the proposed FDI scheme, where $N$ is the number of nonlinear faults of the fault class $\mathcal{F}$ described in Section II. One of the nonlinear adaptive estimators is the *fault detection and approximation estimator* (FDAE) used to detect faults. The remaining ones are *fault isolation estimators* (FIEs) that are used for isolation purposes only after a fault has been detected. Each FIE corresponds to a particular type of fault of the class $\mathcal{F}$. A block diagram representation of the overall architecture is shown in Fig. 1.

Under normal operating conditions (without faults), the FDAE is the only estimator monitoring the system. Once a fault is detected, the bank of FIEs is activated and the FDAE adopts the mode of approximating the fault function. The case where none of the isolation estimators matches the fault that has occurred (to some reasonable degree) corresponds to the occurrence of a new and unknown type of fault, and the approximated fault model can then be used to update the fault class $\mathcal{F}$ and also the bank of isolation estimators. The fault model generated by either an isolation estimator (in the case of a match) or the detection/approximation estimator can be used for fault diagnosis and possibly fault accommodation.

In Sections III-A and III-B, the structures and the adaptation mechanisms for the FDAE and the bank of FIEs will be described.

### A. FDAE

Based on the system representation (1), the FDAE is chosen as follows:

$$\dot{\hat{x}}^0 = -\Lambda^0(\hat{x}^0 - x) + f(x,u) + \hat{\phi}(x,u,\hat{\theta}^0) \tag{6}$$

where $\hat{x}^0 \in \Re^n$ is the estimated state vector, $\hat{\phi}\colon \Re^n \times \Re^m \times \Re^p \mapsto \Re^n$ is an online approximation model, $\hat{\theta}^0 \in \Re^p$ represents a vector of adjustable weights of the online approximator, and $\Lambda^0 = \operatorname{diag}(\lambda_1^0, \ldots, \lambda_n^0)$, where $-\lambda_i^0 < 0$ is the $i$th estimator pole. The initial weight vector, $\hat{\theta}^0(0)$ is chosen such that $\hat{\phi}\left(x,u,\hat{\theta}^0(0)\right) = 0, \forall\, (x,u) \in \mathcal{D}$, which corresponds to the case where the system is in "healthy" (no fault) condition.

A key component of the nonlinear adaptive estimator described by (6) is the *online approximator*, denoted by $\hat{\phi}$, which can be described as follows: the $i$th component $\hat{\phi}_i$ of the function $\hat{\phi}$ has the structure

$$\hat{\phi}_i(x,u,\hat{\theta}^0) = \sum_{j=1}^{\nu} c_{ij}\varphi_j(x,u,\sigma_j) \quad c_{ij} \in \Re\ \ \sigma_j \in \Re^k \tag{7}$$

where $\varphi_j(\cdot,\cdot,\cdot)$ are given parametrized basis functions and $c_{ij}$ and the components of $\sigma_j$ are the parameters to be determined, i.e., $\hat{\theta}^0 \triangleq \operatorname{col}(c_{ij}, \sigma_j\colon i = 1,\ldots,n, \quad j = 1,\ldots,\nu)$. In the presence of a fault, $\hat{\phi}$ provides the adaptive structure for approximating online the unknown fault function. This is achieved by adapting the weight vector $\hat{\theta}^0(t)$ which has the effect of changing the input/output behavior of the approximator. The term "online approximator" is used to represent nonlinear multivariable approximation models with adjustable parameters or weights, such as neural networks, fuzzy logic networks, polynomials, spline functions, wavelet networks, etc. In the last few years, several online approximation models have been studied in the context of intelligent systems and control [33], [49], [56]. Some of the properties of online approximators, like linear parametrization and "curse of dimensionality" [1], [57], and localization [48], also play a crucial role when such approximators are used, in this paper, as estimators of fault functions. Although a comparison of different online approximation models would reveal some interesting issues (see [57] for an extensive treatment of the approximation properties relevant to rather a large class of approximation models), in this paper, we simply consider the general class of sufficiently smooth parametrized functions represented by (7) as online approximators.

The next step in the construction of the FDAE is the design of the learning algorithm for updating the weights $\hat{\theta}^0$. Let $\epsilon^0(t) \triangleq x(t) - \hat{x}^0(t)$ be the state estimation error. Using techniques from adaptive control (Lyapunov synthesis method) [19], the learning algorithm of the online approximator is chosen as follows:

$$\dot{\hat{\theta}}^0 = \mathcal{P}_{\Theta^0}\left\{\Gamma^0 Z^T D[\epsilon^0]\right\} \tag{8}$$

where the *projection operator* $\mathcal{P}$ restricts the parameter estimation vector $\hat{\theta}^0$ to a predefined compact and convex region $\Theta^0 \subset \Re^p$, $\Gamma^0 = \Gamma^{0T} \in \Re^{p\times p}$ is a symmetric positive definite learning rate matrix, and $Z\colon \Re^n \times \Re^m \times \Re^p \mapsto \Re^{n\times p}$ denotes the gradient matrix of the online approximator with respect to its adjustable weights, i.e., $Z \triangleq \partial\hat{\phi}(x,u,\hat{\theta}^0)/\partial\hat{\theta}^0$. The *dead-zone* operator $D[\cdot]$ is defined as

$$D[\epsilon^0(t)] \triangleq \begin{cases} 0 & \text{if } |\epsilon_i^0(t)| \le \bar{\epsilon}_i^0(t), \quad i = 1,\ldots,n \\ \epsilon^0(t) & \text{otherwise} \end{cases} \tag{9}$$

where $\bar{\epsilon}_i^0(t)$ is a suitable threshold function that will be specified later on.

The presence of modeling errors (denoted by $\eta(x,u,t)$ in the state equation) causes a nonzero state estimation error $\epsilon^0(t)$, even in the absence of a fault. The dead-zone operator $D[\cdot]$ prevents adaptation of the approximator weights when the modulus of every estimation error component $\epsilon_i^0(t)$ is below its corresponding threshold $\bar{\epsilon}_i^0(t)$, thereby preventing any false alarms. The decision on the occurrence of a fault (detection) is made when the modulus of at least one of the estimation error components $\epsilon_i^0(t)$ exceeds its corresponding threshold $\bar{\epsilon}_i^0(t)$. More precisely, the *absolute fault detection time* $T_d$ is defined as the first instant of time such that $|\epsilon_i^0(t)| > \bar{\epsilon}_i^0(t)$, for $t \ge T_0$, for some $i$, that is

$$T_d \triangleq \inf \bigcup_{i=1}^{n} \left\{t \ge T_0\colon |\epsilon_i^0(t)| > \bar{\epsilon}_i^0(t)\right\}. \tag{10}$$

The *fault detection time* $t_d$ is defined as the difference between the absolute fault detection time $T_d$ and fault occurrence time $T_0$, i.e., $t_d \triangleq T_d - T_0$.

The time-varying dead-zone threshold $\bar{\epsilon}_i^0(t)$ need to be sufficiently large to prevent false alarms. To this end, we choose $\bar{\epsilon}_i^0(t)$ as

$$\bar{\epsilon}_i^0(t) \triangleq \int_0^t e^{-\lambda_i^0(t-\tau)}\bar{\eta}_i(x(\tau),u(\tau),\tau)d\tau \tag{11}$$

which can be easily implemented as the output of a linear filter (with the transfer function $1/(s+\lambda_i^0)$ and under zero initial conditions) whose input is given by $\bar{\eta}_i(t) = \bar{\eta}_i(x(t),u(t),t)$. Note that, as long as $\bar{\eta}_i$ is bounded, the output of the stable filter remains bounded as well.

In the absence of any faults and with the initial weights of the online approximator such that $\hat{\phi}(x,u,\hat{\theta}^0(0)) = 0$, by (1) and (6) it can be easily verified that each component $\epsilon_i^0(t)$ of the state estimation error satisfies

$$\begin{aligned} |\epsilon_i^0(t)| &= \left|\int_0^t e^{-\lambda_i^0(t-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right| \\ &\le \int_0^t e^{-\lambda_i^0(t-\tau)}\bar{\eta}_i(x(\tau),u(\tau),\tau)d\tau = \bar{\epsilon}_i^0(t). \end{aligned} \tag{12}$$

Therefore, the *robustness* of the detection scheme, i.e., the ability to avoid any false alarms in the presence of modeling uncertainty, is guaranteed. In the special case of uniform (constant) bounds $\bar{\eta}_i$ on the modeling uncertainty, the dead-zone

threshold is given by $\bar{\epsilon}_i^0(t) = \bar{\eta}_i/\lambda_i^0\left(1-e^{-\lambda_i^0 t}\right)$. The dead-zone can be further simplified to a *constant* threshold $\bar{\epsilon}_i^0 = \bar{\eta}_i/\lambda_i^0$ by taking a uniform upper bound over time.

As is well known in the fault diagnosis literature, there is an inherent tradeoff between robustness and fault-detectability. The detectability property of the nonlinear fault diagnosis scheme described by (6) and (8) was rigorously investigated for the special case of a constant bound $\bar{\eta}_i$ in a previous work [40]. For completeness of the presentation, this detectability result is also stated in the following theorem (the proof can be found in [40]).

*Theorem 3.1:* Consider the nonlinear fault diagnosis scheme described by (6) and (8).

a) If there exists an interval of time $[t_1, t_2]$, over which $t_2 > t_1 \geq T_0$, such that at least one component $\phi_i(x,u)$ of the fault vector $\phi(x,u)$ satisfies the condition

$$\left|\int_{t_1}^{t_2} e^{-\lambda_i^0(t_2-\tau)}\left(1-e^{-\alpha_i(\tau-T_0)}\right)\phi_i(x(\tau),u(\tau))d\tau\right| > \frac{2\bar{\eta}_i}{\lambda_i^0} \tag{13}$$

then a fault will be detected, that is, $\left|\epsilon_i^0(t_2)\right| > \bar{\epsilon}_i^0$.

b) For any positive constants $\alpha_i$, $\lambda_i^0$ and for any $t_1 \geq T_0$, there does exist a time $t_2 \geq t_1$ such that if at least one component $\phi_i(x,u)$ of the fault vector $\phi(x,u)$ satisfies the condition

$$|\phi_i(x(t),u(t))| > 2\bar{\eta}_i, \quad \forall\, t \in [t_1, t_2]$$

then a fault will be detected, that is, $\left|\epsilon_i^0(t_2)\right| > \bar{\epsilon}_i^0$.

The first part of the above theorem shows that, if at least one component $\phi_i(x,u)$ of the fault vector function $\phi(x,u)$ satisfies (13) over some time interval $[t_1, t_2]$, then a fault will be detected at $t = t_2$, thus triggering the learning algorithm. Intuitively, condition (13) includes the case where the fault function $\phi_i(x(t),u(t))$ changes its sign over time. The second part of the above theorem shows that, if there is no change of sign and the magnitude of the fault function $\phi_i$ is greater than $2\bar{\eta}_i$ for a sufficiently long time, then a fault will be detected.

In general, after the detection of a fault (i.e., for $t \geq T_d$), the dead-zone becomes unnecessary during the approximation phase and can therefore be disabled. The projection operator $\mathcal{P}$ is required during the approximation phase in order to guarantee the stability of the learning algorithm in the presence of approximation errors, which may be caused by the inability of the online approximator to match the fault function exactly. Moreover, some stability properties of the above FDAE (with a *constant* dead-zone threshold $\bar{\epsilon}_i^0$), e.g., the boundedness of the state and parameter estimates and the convergence of the estimator error to a neighborhood of zero in the presence of modeling uncertainty, have been analytically studied in [6].

### B. Fault Isolation Estimators and Decision Scheme

After a fault has been detected, the isolation scheme is activated (see Fig. 1). Specifically, the following $N$ nonlinear adaptive estimators are used as isolation estimators:

$$\begin{aligned}\dot{\hat{x}}^s &= -\Lambda^s(\hat{x}^s - x) + f(x,u) + \hat{\phi}^s(x,u,\hat{\theta}^s)\\ \hat{\phi}^s(x,u,\hat{\theta}^s) &= \left[\left(\hat{\theta}_1^s\right)^T g_1^s(x,u),\ldots,\left(\hat{\theta}_n^s\right)^T g_n^s(x,u)\right]^T\end{aligned} \tag{14}$$

where $\hat{\theta}_i^s \in \Re^{q_i^s}$, for $i = 1,\ldots,n$, $s = 1,\ldots,N$, is the estimate of the fault parameter vector in the $i$th state variable. Moreover, $\Lambda^s = \mathrm{diag}\,(\lambda_1^s,\ldots,\lambda_n^s)$, where $-\lambda_i^s < 0$ are design constants representing the estimator pole locations. For notational simplicity and without loss of generality, in this paper we assume that $\lambda_i^s = \lambda_i$, for all $s = 1,\ldots,N$.

The design of FIEs is similar to the design of the FDAE. Each isolation estimator corresponds to one of the possible types of nonlinear faults belonging to the fault class $\mathcal{F}$. The adaptation in the isolation estimators arises due to the unknown parameter vector $\theta_i^s$. The adaptive law for updating each $\hat{\theta}_i^s$ is derived by using the Lyapunov synthesis approach, with the projection operator restricting $\hat{\theta}_i^s$ to the corresponding known set $\Theta_i^s$. Specifically, if we let $\epsilon_i^s \triangleq x_i - \hat{x}_i^s$ be the $i$th component of the state estimation error vector of the $s$th estimator, then the learning algorithm is chosen as:

$$\dot{\hat{\theta}}_i^s = \mathcal{P}_{\Theta_i^s}\left\{\Gamma_i^s g_i^s(x,u)\epsilon_i^s\right\} \tag{15}$$

where $\Gamma_i^s = \Gamma_i^{s\,T} > 0$ is a symmetric, positive–definite learning rate matrix. Note that, since the isolation estimators are activated only after the detection of a fault, there is no need to use the dead-zone on the state estimation error. In addition to the state estimation error of each isolation estimator, the parameter estimate $\hat{\theta}_i^s$ also provides useful information for fault isolation purposes. However, it is important to stress that it cannot be guaranteed that for the actual fault the parameter estimate $\hat{\theta}_i^s$ converges to the true value $\theta_i^s$, unless we assume persistency of excitation [19], a condition which, in general, is too restrictive (in this paper, we do *not* assume persistency of excitation).

The fault-isolation decision scheme is based on the following intuitive principle: if the $s$th fault occurs at some time $T_0$ and is detected at time $T_d$, then a set of adaptive thresholds $\{\mu_i^s(t), i = 1,\ldots,n\}$ can be designed such that the $i$th component of the state estimation error associated with the $s$th estimator satisfies $|\epsilon_i^s(t)| \leq \mu_i^s(t)$, for all $t \geq T_d$. Consequently, for each $s = 1,\ldots,N$, such a set of thresholds $\{\mu_i^s(t), i = 1,\ldots,n\}$ can be designed for the $s$th fault isolation estimator. In the fault isolation procedure, if for a particular isolation estimator $s$ and some $i = 1,\ldots,n$, its state estimation error satisfies $|\epsilon_i^s(t)| > \mu_i^s(t)$ for some $t > T_d$, then the possibility that the fault $s$ may have occurred can be excluded. Using this intuitive idea, the following fault isolation decision scheme can be devised.

*Fault isolation decision scheme:* If, for each $r \in \{1,\ldots,N\}\backslash\{s\}$, there exist some finite time $t^r > T_d$ and some $i \in \{1,\ldots,n\}$ such that $|\epsilon_i^r(t^r)| > \mu_i^r(t^r)$, then the occurrence of the fault $s$ is deduced. The absolute fault isolation time is defined as $T_{\mathbf{isol}}^s \triangleq \max\{t^r, r \in \{1,\ldots,N\}\backslash\{s\}\}$ and the fault isolation time $t_{\mathbf{isol}}^s$ is defined as the difference between $T_{\mathbf{isol}}^s$ and the absolute fault detection time $T_d$, i.e., $t_{\mathbf{isol}}^s \triangleq T_{\mathbf{isol}}^s - T_d$.

In order to gain a deeper insight into the above-stated fault isolation decision scheme, we refer to Fig. 2. For the sake of simplicity, a scalar case is considered (i.e., the index $i$ is dropped). Moreover, without loss of generality, we assume that the class $\mathcal{F}$ is made up of three different kinds of faults (i.e., $N = 3$) and that Fault 1 occurs at time $T_0$. After detection of the occurrence of a fault at time instant $T_d$ (see (10)), the FIEs are activated and the time-instants $t^2$ and $t^3$ are determined. Accordingly, Fault 1

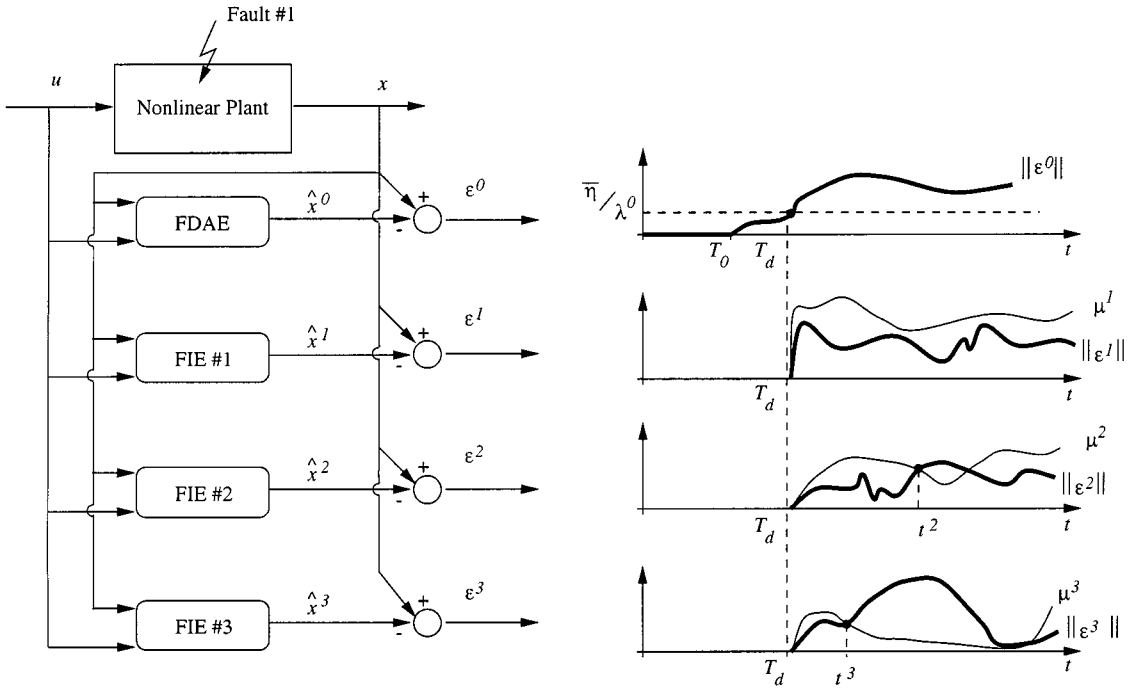


Fig. 2. Example of application of the fault decision scheme to a scalar three-fault case. A faulty situation is detected at time-instant $T_d$, and fault number 1 is isolated at time $T^1_{\text{isol}} = t^2$.

is isolated at time $T^1_{\text{isol}} = t^2$. In the situation presented in Fig. 2, a constant bound $\bar{\epsilon}^0 = \bar{\eta}/\lambda^0$ is considered for the sake of simplicity [see the discussion after (12)].

*Remark 3.1:* In the fault-diagnosis literature, one can find several types of observer schemes. For example, within the fault isolation framework, the *dedicated observer scheme* (DOS) proposed by Clark and the *generalized observer scheme* (GOS) presented by Frank are typically used [4], [9], [36]. In both schemes, the FDI architectures consist of $N$ observers, where $N$ is the number of faults under consideration. In the DOS, the $s$th residual is designed to be sensitive only to the $s$th fault, $s = 1, \ldots, N$, but decoupled from all other faults. In the special case where the DOS can be designed, this scheme permits a single detection and a single isolation of $N$ faults, even if they occur simultaneously. A more commonly used scheme is the *generalized observer scheme*, where the $s$th residual is sensitive to all faults but the $s$th one. The decision function of the GOS is as follows: if the $s$th residual is zero (or below a certain threshold) and all the remaining residuals are nonzero (or above their corresponding thresholds), then a decision on the occurrence of the $s$th fault is made. Therefore, the above-stated fault isolation scheme falls within the GOS architectural framework.

Clearly, a basic role in the above fault isolation scheme is played by the adaptive thresholds $\mu^s_i(t)$. In this respect, we now proceed to compute nonconservative thresholds associated with the residual of each fault isolation estimator in the general case of incipient faults (in the following analysis, we denote by $T_d$ the absolute fault detection time given by (10)). The following lemma provides a bounding function for the state estimator error of the $s$th isolation estimator in the case where the incipient fault $s$ occurs. Later on, the bounding function will be used to derive adaptive thresholds for the fault isolation scheme.

*Lemma 3.1:* If the incipient fault $s$ occurs, then for all $t \geq T_d$ and for all $i \in \{1, \ldots, n\}$, the $i$th component of the state estimation error of the $s$th isolation estimator satisfies the following inequality:

$$|\epsilon^s_i(t)| \leq \int_{T_d}^{t} e^{-\lambda_i(t-\tau)} \times \left[ |\xi^s_i(\tau)| + \bar{\eta}_i\left(x(\tau), u(\tau), \tau\right) + e^{-\alpha_i(\tau - T_0)} \left| \left(\hat{\theta}^s_i(\tau)\right)^T g^s_i(x(\tau), u(\tau)) \right| \right] d\tau + |\epsilon^s_i(T_d)|\, e^{-\lambda_i(t-T_d)} \tag{16}$$

where

$$\xi^s_i(t) \triangleq \left(\theta^s_i - \hat{\theta}^s_i(t)\right)^T g^s_i(x(t), u(t)) \tag{17}$$

represents the fault function estimation error in the case of a matched fault.

*Proof:* On the basis of (1) and (14), in the presence of the fault $s$, the $i$th component of the error dynamics of the $s$th isolation estimator for $t \geq T_d$ is given by

$$\begin{aligned} \dot{\epsilon}^s_i(t) =& -\lambda_i \epsilon^s_i(t) + \left(1 - e^{-\alpha_i(t-T_0)}\right) (\theta^s_i)^T g^s_i(x(t), u(t)) \\ & - \left(\hat{\theta}^s_i(t)\right)^T g^s_i(x(t), u(t)) + \eta_i(x(t), u(t), t) \\ =& -\lambda_i \epsilon^s_i(t) - e^{-\alpha_i(t-T_0)} \left(\hat{\theta}^s_i(t)\right)^T g^s_i(x(t), u(t)) \\ & + \left(1 - e^{-\alpha_i(t-T_0)}\right) \left(\theta^s_i - \hat{\theta}^s_i(t)\right)^T g^s_i(x(t), u(t)) \\ & + \eta_i(x(t), u(t), t). \end{aligned}$$

Therefore, the solution to the previous differential equation is

$$\begin{aligned}\epsilon_i^s(t) = &\int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\left(1-e^{-\alpha_i(\tau-T_0)}\right)\xi_i^s(\tau)d\tau \\ &- \int_{T_d}^{t} e^{-\lambda_i(t-\tau)}e^{-\alpha_i(\tau-T_0)} \\ &\qquad \left(\hat{\theta}_i^s(\tau)\right)^T \cdot g_i^s(x(\tau),u(\tau))d\tau \\ &+ \int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau \\ &+ \epsilon_i^s(T_d)e^{-\lambda_i(t-T_d)}.\end{aligned}$$

where $\xi_i^s(t)$ is defined in (17). By taking norms, we have

$$\begin{aligned}|\epsilon_i^s(t)| \le &\int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\left(1-e^{-\alpha_i(\tau-T_0)}\right)|\xi_i^s(\tau)|\,d\tau \\ &+ \int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\left|\eta_i\left(x(\tau),u(\tau),\tau\right)\right|d\tau \\ &+ |\epsilon_i^s(T_d)|\,e^{-\lambda_i(t-T_d)} + \int_{T_d}^{t} e^{-\lambda_i(t-\tau)} \\ &\cdot e^{-\alpha_i(\tau-T_0)}\left|\left(\hat{\theta}_i^s(\tau)\right)^T g_i^s\left(x(\tau),u(\tau)\right)\right|d\tau.\end{aligned}$$

Note that $0 \le 1-e^{-\alpha_i(t-T_0)} \le 1$. Then, we obtain

$$\begin{aligned}|\epsilon_i^s(t)| \le &\int_{T_d}^{t} e^{-\lambda_i(t-\tau)} \\ &\cdot \left[|\xi_i^s(\tau)| + e^{-\alpha_i(\tau-T_0)}\right. \\ &\quad \left.\cdot\left|\left(\hat{\theta}_i^s(\tau)\right)^T g_i^s\left(x(\tau),u(\tau)\right)\right|\right]d\tau \\ &+ \int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\left|\eta_i\left(x(\tau),u(\tau),\tau\right)\right|d\tau \\ &+ |\epsilon_i^s(T_d)|\,e^{-\lambda_i(t-T_d)}.\end{aligned}$$

Equation (16) follows directly from (2), thus concluding the proof. ■

Although Lemma 3.1 provides an upper bound on the state estimation error of the $s$th estimator, it cannot be directly used as a threshold function for fault isolation because in (16) the fault approximation error $\xi_i^s(t)$, the fault evolution rate $\alpha_i$ and the fault occurrence time $T_0$ are unknown. However, as the estimate $\hat{\theta}_i^s(t)$ belongs to the known compact parameter set $\Theta_i^s$, we have $\left|\theta_i^s - \hat{\theta}_i^s(t)\right| \le \kappa_i^s(t)$ for a suitable $\kappa_i^s(t)$ dependent on the geometric properties of the set $\Theta_i^s$. For instance, letting the parameter set $\Theta_i^s$ be a hypersphere (or the smallest hypersphere containing the set of all possible $\hat{\theta}_i^s(t)$) with center $O_i^s$ and radius $R_i^s$, it follows immediately that $\kappa_i^s(t) = R_i^s + \left|\hat{\theta}_i^s(t) - O_i^s\right|$ and

$$\begin{aligned}|\xi_i^s(t)| &= \left|\left(\theta_i^s - \hat{\theta}_i^s(t)\right)^T g_i^s(x(t),u(t))\right| \\ &\le \kappa_i^s(t)\left|g_i^s\left(x(t),u(t)\right)\right|. \end{aligned} \tag{18}$$

Moreover, we assume that, for the incipient fault time profile given by (3), the unknown fault evolution rate satisfies $\alpha_i > \bar{\alpha}_i$, for $i = 1,\ldots,n$, where $\bar{\alpha}_i$ denotes a known lower bound on the unknown fault evolution rate $\alpha_i$. In a sense, $\bar{\alpha}_i$ can be interpreted as a tuning parameter that can be set by exploiting some *a priori* knowledge of the fault developing dynamics. If no specific knowledge of the fault evolution rate is available, it is always possible to make a cautious (and possibly conservative) choice of a suitably small $\bar{\alpha}_i$. Note that $e^{-\alpha_i(t-T_0)}$ decreases with respect to $\alpha_i$ and $(t-T_0)$. In addition, as $T_d \ge T_0$, it follows that:

$$e^{-\alpha_i(t-T_0)} \le e^{-\bar{\alpha}_i(t-T_d)}. \tag{19}$$

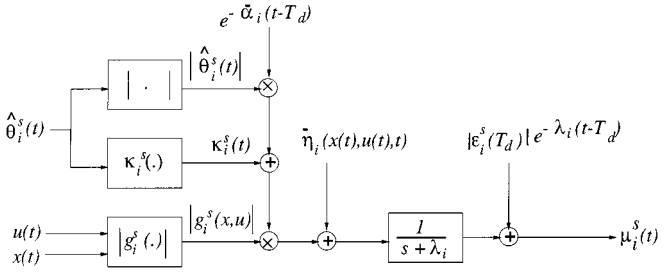


Fig. 3. A block diagram of the algorithm for generating online the adaptive threshold $\mu_i^s(t)$.

Hence, based on (16), (18), and (19), the following threshold functions for fault isolation are chosen:

$$\begin{aligned}\mu_i^s(t) = &\int_{T_d}^{t} e^{-\lambda_i(t-\tau)} \\ &\cdot \left[\left(\kappa_i^s(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^s(\tau)\right|\right)\right. \\ &\quad \left.\cdot\left|g_i^s(x(\tau),u(\tau))\right| + \bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau \\ &+ |\epsilon_i^s(T_d)|\,e^{-\lambda_i(t-T_d)}.\end{aligned} \tag{20}$$

The bound described by (20) represents an adaptive threshold, which, as discussed in [3], [7], [10], has obvious advantages over a fixed threshold. The adaptive threshold can be easily implemented online, as shown in Fig. 3. Specifically, the first term of the threshold can be implemented as the output of a linear filter (with the transfer function $1/(s+\lambda_i)$) with the input given by $\left(\kappa_i^s(t) + e^{-\bar{\alpha}_i(t-T_d)}|\hat{\theta}_i^s(t)|\right)|g_i^s(x(t),u(t))| + \bar{\eta}_i\left(x(t),u(t),t\right)$ and under zero initial conditions.

Let us now address the special case of abrupt faults. As described above, large values of the fault evolution rate $\alpha_i$ in (3) represent abrupt faults. Specifically, we consider abrupt faults whose time profiles are modeled by a step function, i.e.,

$$\beta_i(t-T_0) = \begin{cases} 0 & \text{if } t < T_0 \\ 1 & \text{if } t \ge T_0 \end{cases} \tag{21}$$

where $T_0$ is the occurrence time of the fault. Then using (20) in the special case where $\bar{\alpha}_i$ approaches infinity, the following adaptive thresholds for abrupt-fault isolation are chosen:

$$\begin{aligned}\mu_i^s(t) = &|\epsilon_i^s(T_d)|\,e^{-\lambda_i(t-T_d)} + \int_{T_d}^{t} e^{-\lambda_i(t-\tau)} \\ &\cdot\left[\kappa_i^s(\tau)\cdot\left|g_i^s(x(\tau),u(\tau))\right| + \bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau.\end{aligned} \tag{22}$$

Again, the adaptive threshold described by (22) can be easily implemented as the output of a linear filter (with the transfer function $1/(s+\lambda_i)$) with the input $\kappa_i^s(t)\left|g_i^s(x(t),u(t))\right| + \bar{\eta}_i\left(x(t),u(t),t\right)$ and under zero initial conditions.

## IV. Fault Isolability Condition

In this section, we analyze the fault isolability condition on the proposed FDI scheme; the condition deals with the fault sensitivity property and characterizes the class of faults that can be isolated by the robust fault-isolation algorithm. Moreover, the nonconservativeness of the isolability condition is illustrated by the derivation of a subclass of nonlinear systems and a subclass of faults, for which this condition is also necessary for fault isolability. First, the general case of incipient faults is investigated.

Intuitively, faults are easier to isolate if they are sufficiently "mutually different" in terms of a suitable measure. In the following analysis, we introduce a *fault mismatch function* in the form:

$$\begin{aligned} h_i^{sr}(t) \triangleq & (1-e^{-\alpha_i(t-T_0)})\,(\theta_i^s)^T\, g_i^s\,(x(t),u(t)) \\ & -\left(\hat{\theta}_i^r(t)\right)^T g_i^r\,(x(t),u(t))\,, \\ & r,s=1,\ldots,n, r\neq s \end{aligned} \tag{23}$$

which can be interpreted as the difference between the actual $s$th fault function in the $i$th state equation, represented by $(1-e^{-\alpha_i(t-T_0)})(\theta_i^s)^T g_i^s(x,u)$, and the estimated fault function $(\hat{\theta}_i^r)^T g_i^r(x,u)$ associated with any other isolation estimator $r$ whose structure does not match the actual fault $s$. Before stating a theoretical result on the isolability of incipient faults, we need the following definition.

*Definition 1:* A fault is isolable if the fault isolation scheme described in Section III is able to make a correct decision in a finite time.

The following theorem characterizes the class of incipient nonlinear faults that are isolable by the proposed FDI scheme according to Definition 1.

*Theorem 4.1:* Consider the fault isolation scheme described by (14), (15) and (20). The incipient fault $s$ is isolable if for each $r\in\{1,\ldots,N\}\backslash\{s\}$ there exist some time $t^r>T_d$ and some $i\in\{1,\ldots,n\}$ such that the $i$th component $h_i^{sr}(t)$ of the fault mismatch function satisfies the following inequality:

$$\begin{aligned} & \left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}h_i^{sr}(\tau)d\tau\right| \\ & > 2\left|\epsilon_i^r(T_d)\right| e^{-\lambda_i(t^r-T_d)} + \int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)} \\ & \quad \cdot\left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right)\right. \\ & \quad \left.\cdot\left|g_i^r(x(\tau),u(\tau))\right|+\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau \\ & \quad +\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right|. \end{aligned} \tag{24}$$

*Proof:* Based on (1) and (14), in the presence of the fault $s$, the $i$th component of the error dynamics associated with the estimator $r$ is given by

$$\dot{\epsilon}_i^r=-\lambda_i\epsilon_i^r+\eta_i(x,u,t)+h_i^{sr}(t)$$

where $h_i^{sr}(t)$ is the fault mismatch function defined in (23). Therefore, the solution of the above differential equation for $t\geq T_d$ is

$$\begin{aligned} \epsilon_i^r(t) = & \int_{T_d}^{t} e^{-\lambda_i(t-\tau)}h_i^{sr}(\tau)d\tau+\epsilon_i^r(T_d)e^{-\lambda_i(t-T_d)} \\ & +\int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau. \end{aligned} \tag{25}$$

By using the triangle inequality, we obtain

$$\begin{aligned} \left|\epsilon_i^r(t)\right| \geq & \left|\int_{T_d}^{t} e^{-\lambda_i(t-\tau)}h_i^{sr}(\tau)d\tau\right|-\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i(t-T_d)} \\ & -\left|\int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right|. \end{aligned} \tag{26}$$

We recall that the threshold for the state estimation error of the $r$th estimator is

$$\begin{aligned} \mu_i^r(t) = & \int_{T_d}^{t} e^{-\lambda_i(t-\tau)}\cdot\left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right)\right. \\ & \left.\cdot\left|g_i^r(x(\tau),u(\tau))\right|+\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau \\ & +\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i(t-T_d)}. \end{aligned}$$

Therefore, if (24) is fulfilled, the occurrence of the fault $r$ is excluded at time $t^r$, i.e., $|\epsilon_i^r(t^r)|>\mu_i^r(t^r)$. If this is satisfied for each $r\in\{1,\ldots,N\}\backslash\{s\}$, then the $s$th fault can be isolated, thus concluding the proof. ■

*Remark 4.1:* According to the above theorem, if, for each $r\in\{1,\ldots,N\}\backslash\{s\}$, at least one of the components of the fault mismatch function $h^{sr}(t)$ satisfies condition (24) for some $t^r>T_d$, then the corresponding $i$th residual component associated with the isolation estimator $r$ will exceed its threshold at $t^r$, i.e., $|\epsilon_i^r(t^r)|>\mu_i^r(t^r)$, hence excluding the occurrence of the $r$th fault. Therefore, the above theorem characterizes in nonclosed form the class of nonlinear faults that are isolable by the proposed FDI scheme.

*Remark 4.2:* Based on the bound $\bar{\eta}_i(x(t),u(t),t)$ on the modeling uncertainty, we can easily obtain a more practical version of (24) as follows:

$$\begin{aligned} & \left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}h_i^{sr}(\tau)d\tau\right| \\ & > 2\left|\epsilon_i^r(T_d)\right| e^{-\lambda_i(t^r-T_d)} + \int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)} \\ & \quad \cdot\left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right)\right. \\ & \quad \left.\cdot\left|g_i^r(x(\tau),u(\tau))\right|+2\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau. \end{aligned} \tag{27}$$

Note that all the quantities on the right-hand side of inequality (27) are now known. Therefore, given a particular fault, its fault isolability can be checked by condition (27) and, as a consequence, the class of isolable faults can be approximately determined by a suitable numerical algorithm.

From a qualitative point of view, the fault isolability condition describes an interplay between the fault mismatch function

$h_i^{sr}$ on the one hand and the maximum fault approximation error in the case of a match, the modeling uncertainty and the initial conditions on the other hand. It should be noted that (24) is obtained in the worst-case scenario. In other words, in general, (24) is a sufficient condition for fault isolability. However, among all possible fault scenarios, there exist some cases such that (24) is also necessary for fault isolability, as stated by Theorem 4.2.

*Theorem 4.2:* Consider the fault isolation scheme defined by (14), (15), and (20). Inequality (24) is also necessary for fault isolability, if the following conditions are satisfied:

$$\begin{aligned}&\operatorname{sign}\left[\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}h_i^{sr}(\tau)d\tau\right]\\&\quad=-\operatorname{sign}\left[\epsilon_i^r(T_d)\right]\\&\quad=-\operatorname{sign}\left[\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right]\end{aligned}\quad(28)$$

where $t^r$ is defined in Theorem 4.1, for $r\in\{1,\ldots,N\}\backslash\{s\}$.

*Proof:* In the proof of Theorem 4.1, suppose that (26) becomes an equation at time $t^r$, i.e.,

$$\begin{aligned}\left|\epsilon_i^r(t^r)\right|=&\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}h_i^{sr}(\tau)d\tau\right|-e^{-\lambda_i(t^r-T_d)}\left|\epsilon_i^r(T_d)\right|\\&-\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right|.\end{aligned}\quad(29)$$

In this case, if the fault isolability condition (24) is not satisfied, that is

$$\begin{aligned}&\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}h_i^{sr}(\tau)d\tau\right|\\&\quad\le 2\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i(t^r-T_d)}+\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\\&\qquad\cdot\left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right)\right.\\&\qquad\left.\cdot\left|g_i^r(x(\tau),u(\tau))\right|+\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau\\&\qquad+\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right|\end{aligned}$$

then, by (29), we obtain

$$\begin{aligned}\left|\epsilon_i^r(t^r)\right|\le&\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\\&\cdot\left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right)\right.\\&\left.\cdot\left|g_i^r(x(\tau),u(\tau))\right|+\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau\\&+\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i(t^r-T_d)}\\=&\mu_i^s(t^r).\end{aligned}$$

Therefore, the fault cannot be isolated at $t^r$. From the above analysis, it follows that inequality (24) is also a necessary condition for fault isolability if (26) is an equation. According to (25), this needs the sign condition given by (28) and

$$\begin{aligned}\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}h_i^{sr}(\tau)d\tau\right|\ge&\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i(t^r-T_d)}\\&+\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right|.\end{aligned}$$

Clearly, the above inequality is always guaranteed by (24), thus concluding the proof. ■

*Remark 4.3:* Theorem 4.2 characterizes a subclass of nonlinear uncertain systems and a subclass of nonlinear faults for which the fault isolability condition described by (24) is both sufficient and necessary for fault isolability. The conditions given in the theorem are existence ones, and are included only to gain a more theoretical insight into the nonconservativeness of Theorem 4.1. In other words, the fault isolability condition given in Theorem 4.1 is *not* conservative in the sense that, among all the possible nonlinear systems and faults under consideration, there does exist a case in which a fault will not be isolated by the proposed FDI scheme, unless condition (24) is satisfied.

*A Special Case—Abrupt Nonlinear Faults:* The analysis developed so far for the case of general incipient faults can be specialized to the important case of abrupt faults. Specifically, in order to investigate the fault isolability properties in the abrupt-fault case, we redefine the fault mismatch function as

$$\tilde{h}_i^{sr}(t)\triangleq(\theta_i^s)^T g_i^s(x(t),u(t))-\left(\hat{\theta}_i^r(t)\right)^T g_i^r(x(t),u(t))\quad(30)$$

which represents the difference between the actual fault function $(\theta_i^s)^T g_i^s(x,u)$ and the estimated fault function $\left(\hat{\theta}_i^r(t)\right)g_i^r(x,u)$ associated with the estimator $r$ whose structure does not match the actual fault $s$. Then, from (24), in the special case where $\bar{\alpha}_i$ approaches infinity, the following result follows immediately.

*Corollary 4.1:* Consider the fault isolation scheme described by (14), (15) and (22). The abrupt fault $s$ is isolable if, for each $r\in\{1,\ldots,N\}\backslash\{s\}$, there exist some time $t^r>T_d$ and some $i\in\{1,\ldots,n\}$ such that the $i$th component $\tilde{h}_i^{sr}(t)$ of the fault mismatch function satisfies the following inequality:

$$\begin{aligned}&\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\tilde{h}_i^{sr}(\tau)d\tau\right|\\&\quad>2\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i(t^r-T_d)}+\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\\&\qquad\cdot\left[\kappa_i^r(\tau)\left|g_i^r(x(\tau),u(\tau))\right|+\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau\\&\qquad+\left|\int_{T_d}^{t^r} e^{-\lambda_i(t^r-\tau)}\eta_i(x(\tau),u(\tau),\tau)d\tau\right|.\end{aligned}$$

## V. Fault Isolation Time

One of the most important performance criteria in fault diagnosis is *fault isolation time*, which refers to the time taken by the fault isolation scheme to identify a fault that has occurred [15]. However, in the literature, there exist very few analytical results on fault isolation time. In this section, we derive an analytical upper bound on the incipient-fault isolation time, which is defined as the length of time between the detection and the isolation of a fault. Specifically, we have the following result.

*Theorem 5.1:* Consider the fault isolation scheme described by (14), (15), and (20). For each $r \in \{1,\dots,N\}\backslash\{s\}$, assume that there exist a time interval $[T_d+t_1^r, T_d+t_2^r]$, an index $i \in \{1,\dots,n\}$, and a scalar $\delta_i^{sr} > 0$ such that, for all $t \in [T_d+t_1^r, T_d+t_2^r]$

$$|h_i^{sr}(t)| \geq \left(\kappa_i^r(t) + e^{-\bar{\alpha}_i(t-T_d)}\left|\hat{\theta}_i^r(t)\right|\right)|g_i^r(x(t),u(t))| + 2\bar{\eta}_i(x(t),u(t),t) + \delta_i^{sr} \quad (31)$$

where $t_1^r \geq 0$, $t_2^r > t_1^r + D^r(t_1^r)$, and $D^r(t_1^r)$ is a time period given by

$$D^r(t_1^r) = \frac{1}{\lambda_i}\ln\Bigg[1 + \frac{1}{\delta_i^{sr}}\Bigg(2\lambda_i \int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \cdot \left[\left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) + |g_i^r(x(\tau),u(\tau))| + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau + 4\lambda_i|\epsilon_i^r(T_d)|\cdot e^{-\lambda_i t_1^r}\Bigg)\Bigg]. \quad (32)$$

Then, the maximum fault-isolation time for the incipient fault $s$ is given by

$$\bar{t}_{\mathbf{isol}}^s = \max_{r\in\{1,\dots,N\}\backslash\{s\}}\left\{t_1^r + D^r(t_1^r)\right\}. \quad (33)$$

*Proof:* In order to compute the fault isolation time, we adopt a more practical version of the fault isolability condition given by (27), whose right-hand side is based on known quantities. Specifically, for a given $r \in \{1,\dots,N\}\backslash\{s\}$, consider a time instant $t \in [t_1^r, t_2^r]$ such that

$$\left|\int_{T_d}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right| > 2|\epsilon_i^r(T_d)|e^{-\lambda_i t} + \int_{T_d}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)} \cdot \left[\left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot |g_i^r(x(\tau),u(\tau))| + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau. \quad (34)$$

From the inequality

$$\left|\int_{T_d}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right| \geq \left|\int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)} \cdot h_i^{sr}(\tau)d\tau\right| - \left|\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right|$$

it follows that a sufficient condition for (34) to be satisfied is given by

$$\left|\int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right| > 2|\epsilon_i^r(T_d)|e^{-\lambda_i t} + \int_{T_d}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)} \cdot \left[\left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot |g_i^r(x(\tau),u(\tau))| + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau + \left|\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right|.$$

The aforementioned inequality can be rewritten as

$$\left|\int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right| > 2|\epsilon_i^r(T_d)|e^{-\lambda_i t} + \int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)} \cdot \left[\left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot |g_i^r(x(\tau),u(\tau))| + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau + e^{-\lambda_i(t-t_1^r)} \cdot \left(\left|\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)}h_i^{sr}(\tau)d\tau\right| + \int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \cdot \left[\left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot |g_i^r(x(\tau),u(\tau))| + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau\right). \quad (35)$$

Now, consider a time instant $t_1^r$ such that

$$\left|\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)}h_i^{sr}(\tau)d\tau\right| \leq 2|\epsilon_i^r(T_d)|e^{-\lambda_i t_1^r} + \int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \left[\left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot |g_i^r(x(\tau),u(\tau))| + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau. \quad (36)$$

Note that the previous definition of time-instant $t_1^r$ stems from assuming that $|\epsilon_i^r(T_d+t_1^r)| \leq |\mu_i^r(T_d+t_1^r)|$. Otherwise, $|\epsilon_i^r(T_d+t_1^r)| > |\mu_i^r(T_d+t_1^r)|$ and the possibility of the occurrence of the fault $r$ would already be excluded. Hence, from (36), it follows that (35) is satisfied if

$$\left|\int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right| > 2|\epsilon_i^r(T_d)|e^{-\lambda_i t} + \int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)} \cdot \left[\left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot |g_i^r(x(\tau),u(\tau))| + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau + e^{-\lambda_i(t-t_1^r)} \cdot \left(2\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)}\left[|g_i^r(x(\tau),u(\tau))| \cdot \left(\kappa_i^r(\tau) + e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) + 2\bar{\eta}_i(x(\tau),u(\tau),\tau)\right]d\tau + 2|\epsilon_i^r(T_d)|e^{-\lambda_i t_1^r}\right). \quad (37)$$

Then, under (31), we obtain

$$\begin{aligned}\left|\int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)}h_i^{sr}(\tau)d\tau\right| &= \int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)}\left|h_i^{sr}(\tau)\right|d\tau \\ &\geq \int_{T_d+t_1^r}^{T_d+t} e^{-\lambda_i(T_d+t-\tau)} \cdot \left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot \left|g_i^r(x(\tau),u(\tau))\right|+2\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau \\ &\quad + \frac{\delta_i^{sr}}{\lambda_i}\left(1-e^{-\lambda_i(t-t_1^r)}\right). \end{aligned} \tag{38}$$

By combining (37) with (38), we have

$$\begin{aligned}\frac{\delta_i^{sr}}{\lambda_i}\left(1-e^{-\lambda_i(t-t_1^r)}\right) >& 2\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i t}+e^{-\lambda_i(t-t_1^r)} \cdot \left(2\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \cdot \left[2\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right) + \left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot \left|g_i^r(x(\tau),u(\tau))\right|\right]d\tau \right. \\ & \left. + 2\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i t_1^r}\right). \end{aligned}$$

The previous inequality can be simplified as

$$\begin{aligned}\frac{\delta_i^{sr}}{\lambda_i}\left(1-e^{-\lambda_i(t-t_1^r)}\right) >& 4\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i t}+2e^{-\lambda_i(t-t_1^r)} \cdot \int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \cdot \left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot \left|g_i^r(x(\tau),u(\tau))\right| + 2\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau. \end{aligned}$$

Note that the left-hand side of the aforementioned inequality is an increasing function of $t$, whereas the right-hand side is a decreasing function of $t$. Therefore, the fault isolation time can be obtained by solving the following equation for $t$:

$$\begin{aligned}\frac{\delta_i^{sr}}{\lambda_i}\left(1-e^{-\lambda_i(t-t_1^r)}\right) =& 4\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i t}+2e^{-\lambda_i(t-t_1^r)} \cdot \int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \cdot \left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot \left|g_i^r(x(\tau),u(\tau))\right| + 2\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau. \end{aligned}$$

By some algebraic manipulation, we obtain

$$\begin{aligned}t =& t_1^r+\frac{1}{\lambda_i}\ln\left[1+\frac{1}{\delta_i^{sr}}\left(2\lambda_i\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \cdot \left[\left(\kappa_i^r(\tau)+e^{-\bar{\alpha}_i(\tau-T_d)}\left|\hat{\theta}_i^r(\tau)\right|\right) \cdot \left|g_i^r(x(\tau),u(\tau))\right|+2\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right]d\tau \right. \right. \\ & \left. \left. + 4\lambda_i\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i t_1^r}\right)\right]. \end{aligned}$$

The proof is completed by letting $D^r\left(t_1^r\right) \triangleq t-t_1^r$. ■

*Remark 5.1:* By the previous theorem, if the fault mismatch component $h_i^{sr}(t)$ is sufficiently large for some time period $[T_d+t_1^r, T_d+t_2^r]$, which, in turn, is longer than the time period $D^r\left(t_1^r\right)$ given by (32), then the possibility of the occurrence of the fault $r$ is excluded at time $T_d+t_1^r+D^r\left(t_1^r\right)$. Note that $D^r\left(t_1^r\right)$ can be easily computed by linear filtering techniques. Specifically, the integration term in (32) can be implemented as the output of a linear filter (with the transfer function $1/(s+\lambda_i)$) with the input $\left(\kappa_i^r(t)+e^{-\bar{\alpha}_i(t-T_d)}\left|\hat{\theta}_i^r(t)\right|\right)\left|g_i^r(x(t),u(t))\right| + 2\bar{\eta}_i(x(t),u(t),t)$ and under zero initial conditions. In addition, this theorem describes a relationship between the fault isolation time and the magnitude of the fault mismatch function $h_i^{sr}(t)$, which is represented by the maximum positive constant $\delta_i^{sr}$ satisfying (31). More specifically, (32) shows that the time period $D^r\left(t_1^r\right)$ decreases with respect to $\delta_i^{sr}$. In other terms, we obtained analytical evidence for the intuitive fact that the larger the fault mismatch function, for a sufficiently long period of time, the earlier a fault can be isolated.

*Remark 5.2:* In addition to providing an upper bound $\bar{t}_{\mathbf{isol}}^s$ on the isolation time of the incipient fault $s$, the above theorem also gives a relationship between this upper bound and the fault evolution rate $\alpha_i$. Specifically, (32) and (33) show that the maximum fault isolation time $\bar{t}_{\mathbf{isol}}^s$ decreases with respect to $\bar{\alpha}_i$, which means that the faster a fault evolves, the earlier it can be isolated.

As in the incipient-fault case, the following results provide an estimate of the abrupt-fault isolation time.

*Corollary 5.1:* Consider the fault isolation scheme described by (14), (15) and (22). For each $r \in \{1,\ldots,N\}\backslash\{s\}$, assume that there exist a time interval $[T_d+t_1^r, T_d+t_2^r]$, an index $i \in \{1,\ldots,n\}$, and a scalar $\delta_i^{sr}>0$ such that

$$\left|\tilde{h}_i^{sr}(t)\right| \geq \kappa_i^r(t)\left|g_i^r(x(t),u(t))\right|+2\bar{\eta}_i\left(x(t),u(t),t\right)+\delta_i^{sr}, \quad \forall\, t \in [T_d+t_1^r, T_d+t_2^r]$$

where $t_1^r \geq 0$, $t_2^r > t_1^r+\tilde{D}^r\left(t_1^r\right)$, and $\tilde{D}^r(t_1^r)$ is a time period given by

$$\begin{aligned}\tilde{D}^r\left(t_1^r\right) =& \frac{1}{\lambda_i}\ln\left[1+\frac{1}{\delta_i^{sr}}\left(2\lambda_i\int_{T_d}^{T_d+t_1^r} e^{-\lambda_i(T_d+t_1^r-\tau)} \cdot \left(\kappa_i^r(\tau)\left|g_i^r(x(\tau),u(\tau))\right| + 2\bar{\eta}_i\left(x(\tau),u(\tau),\tau\right)\right)d\tau \right. \right. \\ & \left. \left. + 4\lambda_i\left|\epsilon_i^r(T_d)\right|e^{-\lambda_i t_1^r}\right)\right]. \end{aligned}$$

Then, the maximum fault isolation time for the abrupt fault $s$ is given by

$$\bar{t}^s_{\text{isol}} = \max_{r\in\{1,\ldots,N\}\setminus\{s\}} \left\{ t^r_1 + \check{D}^r\left(t^r_1\right)\right\}.$$

*Proof:* Consider inequalities (31) and (32) in the special case where $\bar{\alpha}_i$ approaches infinity. Then, the above results can be immediately obtained. ■

## VI. Simulation Results

We now present two examples to illustrate the effectiveness of the proposed FDI methodology. The first example is based on a simple nonlinear system, and aims at showing a complete application of the analytical results presented in the paper. The second example addresses the well-known three-tank benchmark problem in FDI [21]. This application is particularly important in order to point out both the practical significance of the FDI problem statement in terms of faults with known functional structures, and the applicability of the proposed FDI architecture to a *feedback controlled system*.

### A. Van Der Pol Oscillator Example

In this section, we use the proposed FDI scheme to detect and isolate incipient faults in a simple nonlinear second-order dynamic system, i.e., the Van der Pol oscillator, which is described by

$$\ddot{y} + 2\omega\zeta(cy^2 - 1)\dot{y} + \omega^2 y = u + \beta(t - T_0)\phi(y)$$

where $\omega, \zeta, c$ are positive constants, $\beta$ represents the time profile of a fault and $\phi$ is the change in the system due to the fault. Specifically, we consider two types of faults: $\phi^1 = \theta^1\sin(y)$ and $\phi^2 = \theta^2\cos(y)$, where $\theta^1 \in \Theta^1 = [-1.5, 1.5]$ and $\theta^2 \in \Theta^2 = [-1.5, 1.5]$. We assume that the unknown incipient-fault evolution rate $\alpha$ defined in (3) satisfies: $\alpha \geq \bar{\alpha} = 0.1$. The modeling uncertainty is unstructured and assumed to be some inaccuracy in the value of $\zeta$. Therefore, the state equations for the nominal system are

$$\dot{x} = \begin{bmatrix} x_2 \\ 2\omega\zeta\left(1 - c(x_1)^2\right)x_2 - \omega^2 x_1 + u \end{bmatrix}$$

where $x \triangleq [x_1, x_2]^T = [y, \dot{y}]^T$ denotes the state vector. Moreover, the class of faults is described as

$$\begin{aligned}\mathcal{F} &= \left\{ \begin{bmatrix} 0 \\ \phi^1 \end{bmatrix}, \begin{bmatrix} 0 \\ \phi^2 \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} 0 \\ \theta^1\sin(x_1) \end{bmatrix}, \begin{bmatrix} 0 \\ \theta^2\cos(x_1) \end{bmatrix} \right\}.\end{aligned}$$

By using the methodology described in Section III-B, a bank of two isolation estimators is designed

$$\begin{aligned}\dot{\hat{x}}^1 &= \begin{bmatrix} x_2 \\ 2\omega\zeta\left(1 - c(x_1)^2\right)x_2 - \omega^2 x_1 + u \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 \\ \hat{\theta}^1\sin(x_1) \end{bmatrix} - \lambda \begin{bmatrix} \hat{x}^1_1 - x_1 \\ \hat{x}^1_2 - x_2 \end{bmatrix}, \\ \dot{\hat{x}}^2 &= \begin{bmatrix} x_2 \\ 2\omega\zeta(1 - c\left(x_1\right)^2)x_2 - \omega^2 x_1 + u \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 \\ \hat{\theta}^2\cos(x_1) \end{bmatrix} - \lambda \begin{bmatrix} \hat{x}^2_1 - x_1 \\ \hat{x}^2_2 - x_2 \end{bmatrix}\end{aligned}$$

where $\hat{x}^1 = [\hat{x}^1_1\ \hat{x}^1_2]^T$ and $\hat{x}^2 = [\hat{x}^2_1\ \hat{x}^2_2]^T$ denote the estimated state vectors associated with estimator 1 and estimator 2, respectively, and $-\lambda < 0$ is the filter pole location; $\hat{\theta}^1$ and $\hat{\theta}^2$ are the adjustable parameters. For the FDAE, the online approximator is implemented as a continuous radial basis function (RBF) neural network with eleven fixed centers evenly distributed over the interval $[-2, 2]$. As described in Section III-A, the stability and fault-detectability properties of the FDAE have been investigated in [6], [40]. Note that, in this example, faults are only possible in the state component $x_2$; therefore, for the sake of notational simplicity, the state index $i$ is dropped.

We perform the simulation with the following nominal system parameters: $\omega = 0.9$, $\zeta = 0.6$, $c = 0.95$. The control input is set to $u(t) = 0$. The modeling uncertainty is assumed to arise out of a 5% inaccuracy in the value of $\zeta$. It is also assumed that the uncertainty in $\zeta$ is at most 10%, which gives a nonuniform bound on the modeling uncertainty as $\bar{\eta} = |0.2\omega\zeta(1 - c(x_1)^2)x_2|$. The bounding function $\bar{\eta}$ is clearly bounded in any compact region of the state space. Moreover, we set $\lambda = 5$ and $\gamma = 1$ for the isolation estimators.

Fig. 4 shows the simulation results when an incipient fault of type 1, with $\theta^1 = 0.75$ and the fault evolution rate $\alpha = 0.2$, occurs at $t = 10$ s. The evolution of the actual fault function $\phi(y)$ (solid line) and the output of the neural network approximator $\hat{\phi}(y, \hat{\theta})$ (dash-dotted line) associated with the FDAE estimator are shown in Fig. 4(a). The state estimation error (solid line) of the FDAE and its corresponding dead-zone threshold (dash-dotted line) are shown in Fig. 4(b). As we can see, the fault is detected at approximately $T_d = 11.5$ s. Moreover, in Fig. 4(c) and (d), the residuals $\epsilon(t)$ (solid lines) and their corresponding thresholds $\mu(t)$ (dash-dotted lines), associated with each isolation estimator, are shown. It can be seen that the residual of estimator 1 always remains below its threshold, whereas the residual of estimator 2 exceeds its threshold at approximately $T^1_{\text{isol}} = 12.8$ s, thus allowing the isolation of fault 1.

Concerning the fault isolation time, the time-behavior of $|h^{12}(t)| - (\kappa^2(t) + e^{-\bar{\alpha}_i(t-T_d)}|\hat{\theta}^2(t)|)|g^2(x,u)| - 2\bar{\eta}$ (see inequality (31)) is shown in Fig. 4(e). Moreover, Fig. 4(f) shows the time period $D^2\left(t^2_1\right)$ (described by (32)) corresponding to each $t^2_1$ in the case where $\delta^{12} = 0.1$. From Fig. 4(f), we can see that $D^2\left(t^2_1\right)$ is approximately 0.5 s when $T_d + t^2_1 = 17$ s. In other words, according to Theorem 5.1, if there exists an interval of time (longer than 0.5 s and with $T_d + t^2_1 = 17$ s as the starting point) over which the condition $|h^{12}(t)| - (\kappa^2(t) + e^{-\bar{\alpha}_i(t-T_d)}|\hat{\theta}^2(t)|)|g^2(x,u)| - 2\bar{\eta} \geq \delta^{12}$ is satisfied, then the maximum fault isolation time is $\bar{t}^1_{\text{isol}} = t^2_1 + D^2(t^2_1) = (17 - 11.5)$ s $+ 0.5$ s $= 6.0$ s, i.e., the absolute fault isolation time $\bar{T}^1_{\text{isol}} = 11.5$ s $+ 6.0$ s $= 17.5$ s. In Fig. 4(e) we can see that this condition is satisfied for all $t \in [17\text{ s}, 18.2\text{ s}]$. Therefore, $\bar{T}^1_{\text{isol}} = 17.5$ s is a valid upper bound on the absolute fault isolation time.

An analogous example is shown in Fig. 5, corresponding to the occurrence of an incipient fault of type 2, with $\theta^2 = 0.9$ and the fault evolution rate $\alpha = 0.2$, occurs at $t = 10$ s. In this case, too, the fault isolation turns out to be successful. In Fig. 5(e) and (f), with $\delta^{21} = 0.2$, an upper bound on the absolute fault isolation time can be similarly computed as: $\bar{T}^2_{\text{isol}} = 25.3$ s $+$ $0.48$ s $= 25.78$ s.

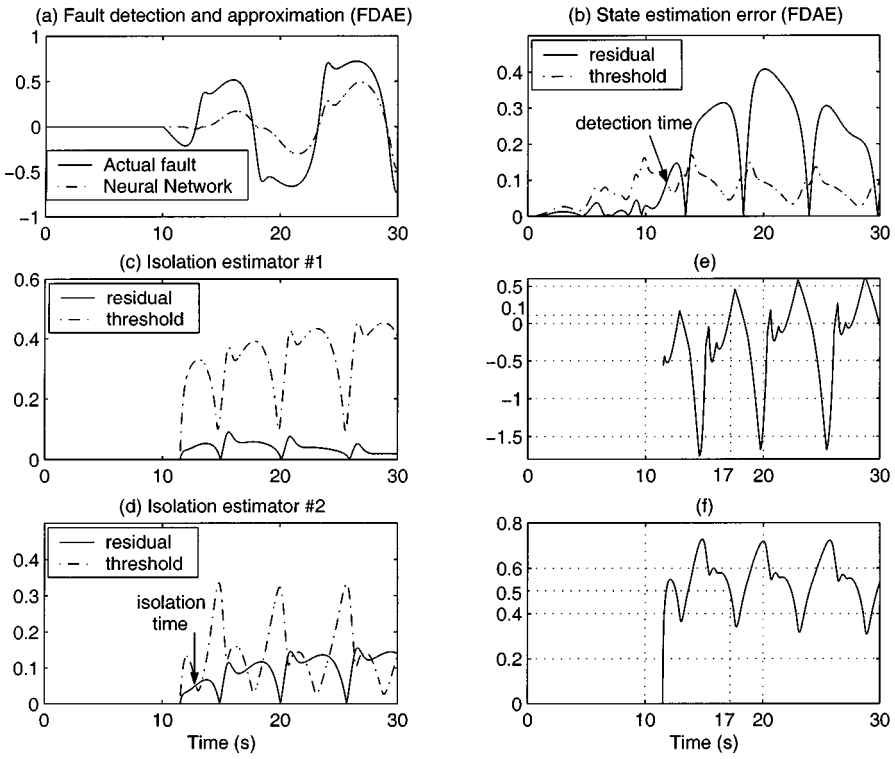


Fig. 4. (a) Time-behaviors of the fault function (solid line) and the neural-network output (dash-dotted line) associated with the FDAE estimator. (b) Time-behaviors of the state estimation error (solid line) associated with the FDAE and the dead-zone threshold (dash-dotted line) (the fault detection time instant is shown by an arrow). (c) and (d) Time-behaviors of the state estimation errors (solid lines) and the thresholds (dash-dotted lines) associated with the two isolation estimators (the fault isolation time instant is shown by a vertical arrow). (e) Time-behaviors of $|h^{12}(t)| - (\kappa^2(t) + e^{-\bar{\alpha}_i(t-T_d)}|\hat{\theta}^2(t)|)|g^2(x,u)| - 2\bar{\eta}$ associated with estimator 2; (f) The time period $D^2(t_1^2)$ for each $t_1^2$ (derived from (32) with $\delta^{12} = 0.10$).

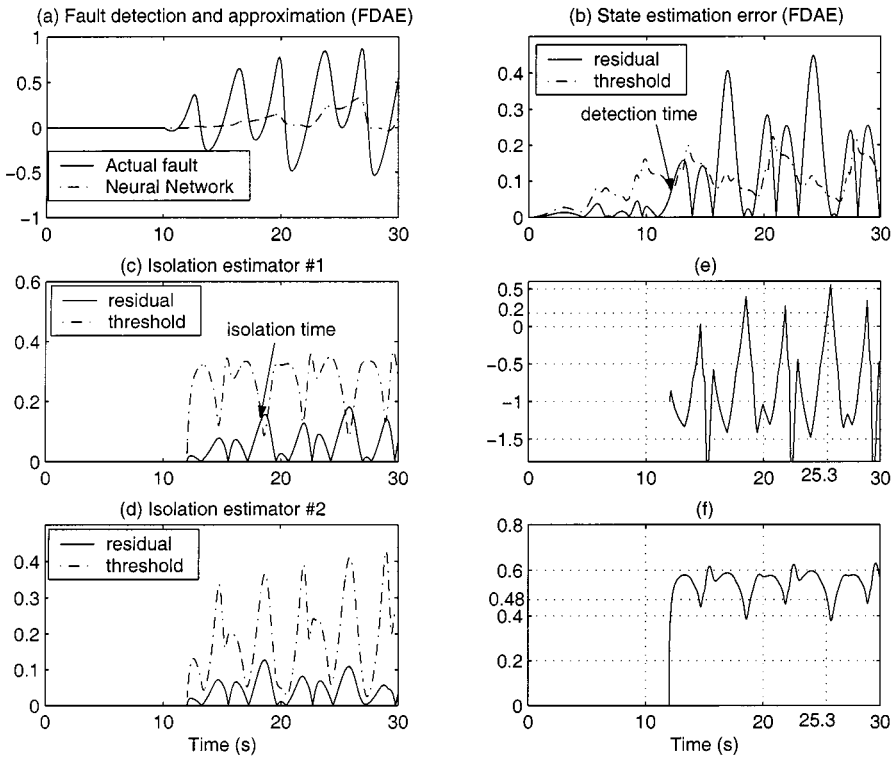


Fig. 5. (a) Time-behaviors of the fault function (solid line) and the neural-network output (dash-dotted line) associated with the FDAE estimator. (b) Time-behaviors of the state estimation error (solid line) associated with the FDAE and the dead-zone threshold (dash-dotted line) (the fault detection time instant is shown by an arrow).(c) and (d) Time-behaviors of the state estimation errors (solid lines) and the thresholds (dash-dotted lines) associated with the two isolation estimators (the fault isolation time instant is shown by an arrow). (e) Time-behaviors of $|h^{21}(t)| - (\kappa^1(t) + e^{-\bar{\alpha}_i(t-T_d)}|\hat{\theta}^1(t)|)|g^1(x,u)| - 2\bar{\eta}$ associated with estimator 1; (f) The time period $D^1(t_1^1)$ for each $t_1^1$ (derived from (32) with $\delta^{12} = 0.20$).

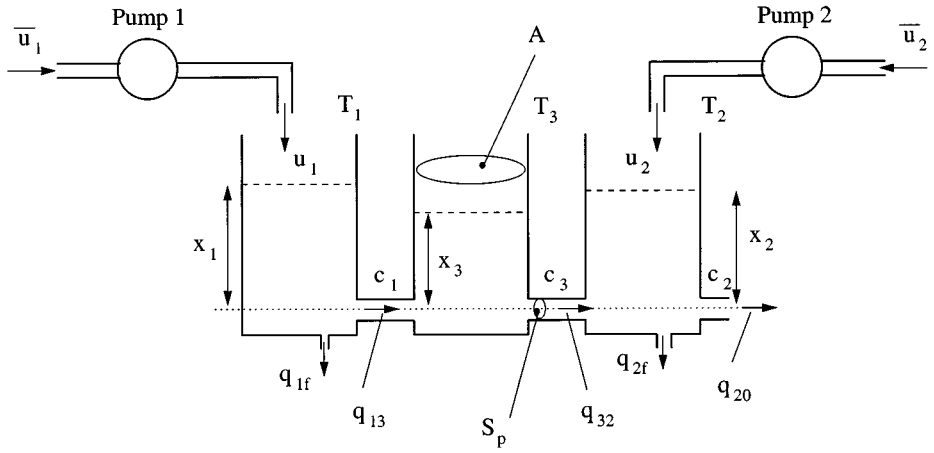


Fig. 6. The three-tank system.

### B. The Three-Tank System Example

Let us consider the *controlled* three-tank system depicted in Fig. 6 (the reader is referred to [21] and to the invited session [24] for several interesting issues regarding this well-known benchmark for FDI). The three tanks $T_1, T_2$, and $T_3$ are identical and are cylindrical in shape with a cross section $A = 0.0154\,\mathrm{m}^2$. The cross section of the connection pipes is $S_p = 5 \cdot 10^{-5}\ \mathrm{m}^2$, and the liquid levels in the three tanks are denoted by $x_1$, $x_2$, and $x_3$, respectively ($0 \le x_i \le 0.69$ m, $i = 1, 2, 3$). The supplying flow rates coming from pumps 1 and 2 are denoted by $u_1$ and $u_2$, respectively ($0 \le u_j \le 10^{-4}\ \mathrm{m}^3/\mathrm{s}$, $j = 1, 2$). $q_{13}$ and $q_{32}$ represent the flow rates between tanks 1 and 3 and between tanks 3 and 2, respectively, and $q_{20}$ is the outflow rate.

By using balance equations and Torricelli's rule, we obtain the state equations shown at the bottom of the page, where $x \triangleq [x_1 \quad x_2 \quad x_3]^T$ denotes the state vector, $u \triangleq [u_1 \quad u_2]^T$ denotes the control vector, and $f_i$ denotes the $i$th component of the vector function $f(\cdot)$. Moreover, $c_1 = 1$, $c_2 = 0.8$, and $c_3 = 1$ denote nondimensional outflow coefficients, and $g$ is the gravity acceleration.

We consider the case of abrupt faults (the case of incipient faults is completely analogous and is not addressed here for the sake of brevity). The class $\mathcal{F}$ of nonlinear faults is defined by the following two types of faults possibly acting on each of the two tanks $T_1$ and $T_2$.

1) **Leakage in tank** $T_1$. We assume that the leak is circular in shape and of unknown radius $r_1$. Then, denoting by $q_{1f}$ the outflow rate of the unknown-size leak in tank $T_1$, we have $q_{1f} = c_1\pi(r_1)^2\sqrt{2gx_1}$.
2) **Actuator fault in pump 1**. We consider a simple multiplicative actuator fault in pump 1 by letting $u_1 = \bar{u}_1 + (a_1 - 1)\bar{u}_1$, where $\bar{u}_1$ is the supply flow rate in the nonfault case, and $a_1$ is the parameter characterizing the magnitude of the fault. For $a_1 = 1$, we have the nonfault situation in pump 1, whereas $a_1 = 0$ implies that the pump is completely faulty, in the sense that there is no flow.
3) **Leakage in tank** $T_2$. Analogously to the case of a leakage in tank $T_1$, we have $q_{2f} = c_2\pi(r_2)^2\sqrt{2gx_2}$.
4) **Actuator fault in pump 2**. Analogously to the case of a fault in pump 1, we have $u_2 = \bar{u}_2 + (a_2 - 1)\bar{u}_2$.

The fault class $\mathcal{F}$ can now be written as

$$\mathcal{F} = \left\{ \begin{bmatrix} \theta_1^1 g_1^1(x,u) \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta_1^2 g_1^2(x,u) \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \theta_2^3 g_2^3(x,u) \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \theta_2^4 g_2^4(x,u) \\ 0 \end{bmatrix} \right\}$$

where $\theta_1^1 \triangleq (r_1)^2$, $g_1^1(x,u) \triangleq c_1\pi\sqrt{2gx_1}$, $\theta_1^2 \triangleq a_1 - 1$, $g_1^2(x,u) \triangleq \bar{u}_1$, $\theta_2^3 \triangleq (r_2)^2$, $g_2^3(x,u) \triangleq c_2\pi\sqrt{2gx_2}$, $\theta_2^4 \triangleq$

$$\begin{aligned}
\dot{x}_1 &= f_1(x,u) \\
&= \frac{\left(-c_1 S_p \mathrm{sign}(x_1 - x_3)\sqrt{2g|x_1 - x_3|} + u_1\right)}{A} \\
\dot{x}_2 &= f_2(x,u) \\
&= \frac{\left(-c_3 S_p \mathrm{sign}(x_2 - x_3)\sqrt{2g|x_2 - x_3|} - c_2 S_p\sqrt{2gx_2} + u_2\right)}{A} \\
\dot{x}_3 &= f_3(x,u) \\
&= \frac{\left(c_1 S_p \mathrm{sign}(x_1 - x_3)\sqrt{2g|x_1 - x_3|} - c_3 S_p \mathrm{sign}(x_3 - x_2)\sqrt{2g|x_3 - x_2|}\right)}{A}
\end{aligned}$$

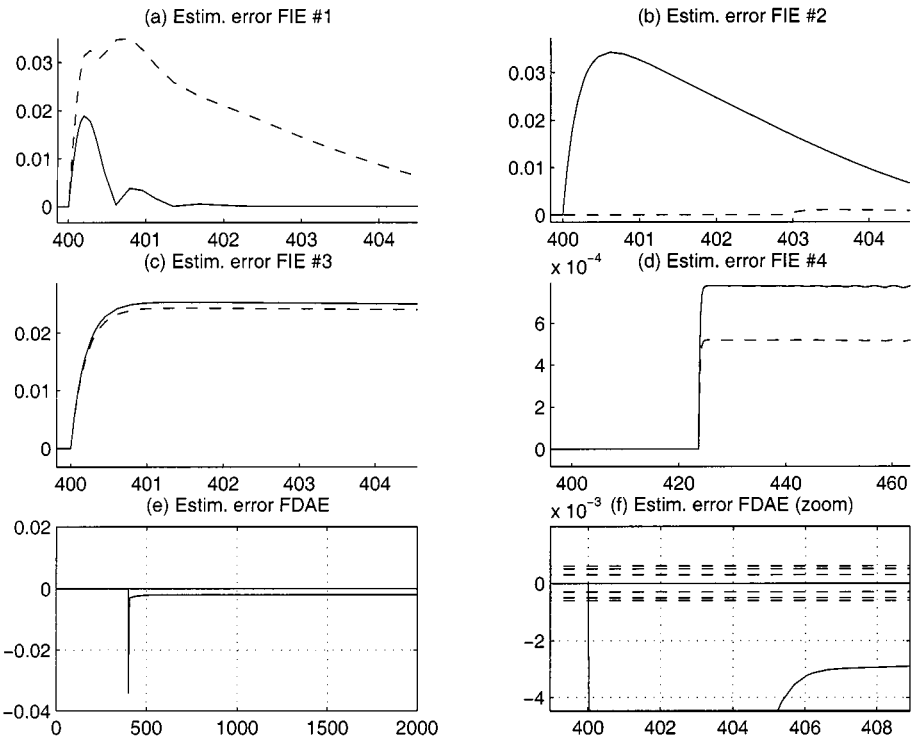


Fig. 7. (a)–(d) Time-behaviors of the estimation errors in the first state associated with the four isolation estimators. (e) Time-behavior of the estimation errors associated with the FDAE. (f) Same as for (e) but plotted in enlarged form; the dashed lines represent the dead-zone thresholds.

$a_2 - 1$, and $g_2^4(x,u) \triangleq \bar{u}_2$. Therefore, the state equations for the three-tank system can be put into the general form (1).

With regards to modeling uncertainties, a 4% inaccuracy in the cross section $S_p$ of the connection pipe has been considered. Moreover, after simulating the whole system under several operating conditions, we have obtained $\bar{\eta}_1 = 3 \cdot 10^{-4}$, $\bar{\eta}_2 = 5 \cdot 10^{-4}$, and $\bar{\eta}_3 = 6 \cdot 10^{-4}$ (for simplicity, uniform bounds on the modeling error are used). In order to guarantee fault isolability (see (24), Theorem 4.1), after a suitable offline simulation procedure, the following parameter sets have been defined: $\Theta_1^1 = [1.49 \cdot 10^{-5}, 4.9 \cdot 10^{-4}]$, $\Theta_2^3 = [1.88 \cdot 10^{-5}, 4.9 \cdot 10^{-4}]$, $\Theta_1^2 = [0.05, 1]$, and $\Theta_2^4 = [0.1, 1]$. A bank of four isolation estimators has been implemented according to the scheme depicted in Fig. 2. We have set $\Lambda^1 = \Lambda^2 = \Lambda^3 = \Lambda^4 = \mathrm{diag}(5,5,5)$, $\gamma^1 = \gamma^3 = 10^{-4}$, and $\gamma^2 = \gamma^4 = 3 \cdot 10^3$. For the detection/approximation estimator, the online approximator has been implemented as a feedforward neural approximator with one-hidden layer of five neurons and a linear output layer. The dead-zone has been computed on the basis of the uncertainty bounds described above.

As an illustrative example (an exhaustive simulation study is clearly beyond the scope of the paper), Fig. 7 shows the simulation results obtained when fault 1, with $\theta_1^1 = 10^{-4}$, occurred at time $T_0 = 400$ s. The estimation errors in the first state component $x_1$ associated with each of the four FIEs are shown in Fig. 7(a)–(d), respectively. Moreover, in Fig. 7(e) the state estimation error of the FDAE is presented. Finally, Fig. 7(e) is replotted in enlarged form in Fig. 7(f); the dead-zone thresholds are also plotted (dashed lines). Fig. 7(f) allows one to appreciate the absolute fault detection time $T_d$ (time instant when one of the state estimation errors crosses its corresponding threshold due to uncertainty). As can be noticed from Fig. 7(a)–(d), only the state estimation error of the first estimator always remains below its threshold, whereas the estimation errors of the other three estimators exceed their corresponding thresholds immediately after $T_d$, thus allowing the isolation of fault no. 1. In this specific case, the absolute fault isolation time $T_{\mathbf{isol}}^1$ is approximately equal to the absolute detection time $T_d$.

## VII. Conclusion

Presently, there is great industrial interest in automated fault-diagnosis methodologies. This is fueled by two main factors. First, the cost of a failure in a dynamic system can be tremendous. Secondly, modern engineering systems are becoming more automated and complex, thus making it almost impossible to manually monitor the health condition of a system, except for very simple faults. Fault isolation is one of the key tasks of fault diagnosis as it provides the user with information about the type of fault; this can be a significant step toward correcting the fault (either online or offline).

In this paper, we have designed and analyzed a robust fault detection and isolation scheme for nonlinear uncertain dynamic systems. The analysis has addressed both abrupt and incipient developing faults. The proposed architecture consists of a bank of nonlinear adaptive estimators, one of which is used for the detection and the approximation of a fault, whereas the rest (one for each fault type) are used for online fault isolation. The fault-isolation decision scheme is based on adaptive threshold functions that are derived to guarantee no false isolation decision. We have also investigated the fault isolability conditions on the developed FDI scheme, and derived the class of faults

that can be isolated, and that is defined in terms of the so-called fault mismatch functions. The nonconservativeness of the fault isolability conditions is characterized by a subclass of nonlinear uncertain systems and a subclass of nonlinear faults for which these conditions are sufficient and necessary for fault isolability. Moreover, an analytical upper bound on the fault isolation time has been obtained. Finally, two simulation examples have been given to illustrate the effectiveness of the proposed FDI scheme.

## REFERENCES


[1] A. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Trans. Inform. Theory*, vol. 39, pp. 930–945, May 1993.

[2] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[3] J. Chen and R. J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Boston, MA: Kluwer, 1999.

[4] R. N. Clark, "Instrument fault detection," *IEEE Trans. Aero. Eletron. Syst.*, vol. AES-14, pp. 456–465, May 1978.

[5] M. Corless and G. Leitmann, "Continuous state feedback guaranteeing uniform ultimate boundedness for uncertain dynamical systems," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 1139–1144, Oct. 1981.

[6] M. A. Demetriou and M. M. Polycarpou, "Incipient fault diagnosis of dynamical systems using online approximators," *IEEE Trans. Automat. Contr.*, vol. 43, pp. 1612–1617, Nov. 1998.

[7] A. Emami-Naeini, M. M. Akhter, and S. M. Rock, "Effect of model uncertainty on failure detection: The threshold selector," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 1106–1115, Dec. 1988.

[8] J. A. Farrell, T. Berger, and B. D. Appleby, "Using learning techniques to accommodate unanticipated faults," *IEEE Control Syst. Mag.*, vol. 13, pp. 40–49, 1993.

[9] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—A survey and some new results," *Automatica*, vol. 26, pp. 459–474, 1990.

[10] ——, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *IFAC J. Process Control*, vol. 7, pp. 403–424, 1997.

[11] P. M. Frank and R. Seliger, "Fault detection and isolation in automatic processes," in *Control Dynamic Systems*, C. Leondes, Ed. New York: Academic, 1991, pp. 241–287.

[12] E. A. Garciá and P. M. Frank, "Deterministic nonlinear observer-based approaches to fault diagnosis: A survey," *IFAC Control Eng. Prac.*, vol. 5, pp. 663–670, 1997.

[13] J. J. Gertler, "Survey of model-based failure detection and isolation in complex plants," *IEEE Control Syst. Mag.*, vol. 8, pp. 3–11, Dec. 1988.

[14] ——, "Analytical redundancy methods in fault detection and isolation," in *Proc. IFAC Symp. Fault Detection, Supervision Safety Technical Processes*, 1991, pp. 9–21.

[15] ——, *Fault Detection and Diagnosis in Engineering Systems*. New York: Marcel Dekker, 1998.

[16] J. J. Gertler and M. M. Kunwer, "Optimal residual decoupling for robust fault diagnosis," *Int. J. Control*, vol. 61, no. 2, pp. 395–421, 1995.

[17] G. Guglielmi, T. Parisini, and G. Rossi, "Fault diagnosis and neural networks: A power plant application," *IFAC Control Eng. Prac.*, vol. 3, pp. 601–620, 1995.

[18] H. Hammouri, M. Kinnaert, and E. H. E. Yaagoubi, "Oberver-based approach to fault detection and isolation for nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1879–1884, Oct. 1999.

[19] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[20] R. Isermann, "Process fault detection based on modeling and estimation methods: A survey," *Automatica*, vol. 20, pp. 387–404, 1984.

[21] ——, "Supervision, fault detection and fault diagnosis methods—An introduction," *IFAC Control Eng. Prac.*, vol. 5, pp. 639–652, 1997.

[22] A. Isidori, *Nonlinear Control Systems*, 2 ed. Berlin, Germany: Springer-Verlag, 1995.

[23] M. Kinnaert, "Robust fault detection based on observers for bilinear systems," *Automatica*, vol. 35, pp. 1829–1842, 1999.

[24] B. Koppen-Seliger and S. X. Ding, "Fault detection and isolation of a three tank benchmark," in Proc. European Control Conference, 1999.

[25] V. Krishnaswami and G. Rizzoni, "Nonlinear parity equation residual generation for fault detection and isolation," in *Proc. IFAC Symp. Fault Detection, Supervision Safety Technical Processes*, 1994, pp. 317–322.

[26] ——, "A survey of observer based residual generation for FDI," in *Proc. IFAC Symp. Fault Detection, Supervision Safety Technical Processes*, 1994, pp. 34–39.

[27] ——, "Robust residual generation for nonlinear system fault detection and isolation," in *Proc. IFAC Symp. Fault Detection, Supervision Safety Technical Processes*, 1997, pp. 163–168.

[28] M. Krstic, I. Kanellakopoulos, and P. Kokotovic, *Nonlinear and Adaptive Control Design*. New York: Wiley, 1995.

[29] B. Liu and J. Si, "Fault isolation filter design for linear time-invariant systems," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 704–707, May 1997.

[30] K. A. Loparo, M. R. Buchner, and K. S. Vasudeva, "Leak detection in an experimental heat exchanger process: A multiple model approach," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 167–177, Feb. 1991.

[31] C. N. Nett, C. A. Jacobson, and A. T. Miller, "An integrated approach to controls and diagnostics: The 4-parameter controller," in *Proc. Amer. Control Conf.*, 1988, pp. 824–835.

[32] J. Park and G. Rizzoni, "An eigenstructure assignment algorithm for the design of fault detection filters," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 1521–1524, July 1994.

[33] K. M. Passino and S. Yurkovich, *Fuzzy Control*. Menlo Park, CA: Addison-Wesley, 1997.

[34] R. J. Patton, "Robust fault detection using eigenstructure assignment," in *Proc. 12th IMACS World Congress Mathematical Modeling Scientific Computation*, 1988, pp. 431–433.

[35] ——, "Fault-tolerant control: The 1997 situation (survey)," in *Proc. IFAC Symp. Fault Detection, Supervision Safety Technical Processes*, 1997, pp. 1029–1052.

[36] R. J. Patton, P. M. Frank, and R. N. Clark, Eds., *Fault Diagnosis in Dynamic Systems: Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[37] C. De Persis and A. Isidori, "On the problem of residual generation for fault detection in nonlinear systems and some related facts," in Proc. European Control Conference, 1999.

[38] M. S. Phatak and N. Viswanadham, "Actuator fault detection and isolation in linear systems," *Int. J. Syst. Sci.*, vol. 19, pp. 2593–2603, 1988.

[39] M. M. Polycarpou and A. J. Helmicki, "Automated fault detection and accommodation: A learning systems approach," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, pp. 1447–1458, Nov. 1995.

[40] M. M. Polycarpou and A. B. Trunov, "Learning approach to nonlinear fault diagnosis: Detectability analysis," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 806–812, Apr. 2000.

[41] R. Seliger and P. M. Frank, "Robust component fault detection and isolation in nonlinear dynamic systems using nonlinear unknown input observers," in *Proc. IFAC Symp. Fault Detection, Supervision Safety Technical Processes*, 1991, pp. 313–318.

[42] E. D. Sontag, "Smooth stabilization implies coprime factorization," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 435–443, Apr. 1989.

[43] J. Stoustrup and M. J. Grimble, "Integrating control and fault diagnosis; a separation result," in *Proc. IFAC Symp. Fault Detection, Supervision Safety Technical Processes*, 1997, pp. 313–318.

[44] A. B. Trunov and M. M. Polycarpou, "Automated fault diagnosis in nonlinear multivariable systems using a learning methodology," *IEEE Trans. Neural Networks*, vol. 11, pp. 91–101, Feb. 2000.

[45] A. T. Vemuri and M. M. Polycarpou, "Robust nonlinear fault diagnosis in input–output systems," *Int. J. Control*, vol. 68, no. 2, pp. 343–360, 1997.

[46] H. Wang and S. Daley, "Actuator fault diagnosis: An adaptive observer-based technique," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 1073–1078, July 1996.

[47] H. Wang, Z. J. Huang, and S. Daley, "On the use of adaptive updating rules for actuator and sensor fault diagnosis," *Automatica*, vol. 33, pp. 217–225, 1997.

[48] S. Weaver, L. Baird, and M. Polycarpou, "An analytical framework for local feedforward networks," *IEEE Trans. Neural Networks*, vol. 9, pp. 473–482, May 1998.

[49] *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds., Van Nostrand and Reinhold, New York, 1992.

[50] J. E. White and J. L. Speyer, "Detection filter design: Spectral theory and algorithm," *IEEE Trans. Automat. Contr.*, vol. AC-32, pp. 593–603, July 1987.

[51] N. E. Wu, "Failure sensitizing reconfigurable control design," in *Proc. 31st IEEE Conf. Decision Control*, 1992, pp. 44–49.

[52] J. Wünnenberg, "Observer-Based Fault Detection in Dynamic Systems ," Ph.D. dissertation, Universitaet Duisburg, Duisburg, Germany, 1990.
[53] H. L. Yang and M. Saif, "State observation, failure detection and isolation (FDI) in bilinear systems," *Int. J. Control*, vol. 67, pp. 901–920, 1997.
[54] G. Yen and L. Ho, "Fault tolerant control: An intelligent sliding model control strategy," in *Proc. Amer. Control Conf.*, 2000, pp. 4204–4208.
[55] D. L. Yu and D. N. Shields, "A bilinear fault detection filter," *Int. J. Control*, vol. 68, pp. 417–430, 1997.
[56] R. Zbikowski and K. J. Hunt, Eds., *Neural Adaptive Control Technology*, Singapore: World Scientific, 1996.
[57] R. Zoppoli, M. Sanguineti, and T. Parisini, "Approximating networks and extended Ritz method for the solution of functional optimization problems," *J. Optim. Theory Appl.*, vol. 112, no. 2, pp. 403–439, 2002.

**Xiaodong Zhang** was born in Jiangsu, China, in 1973. He received the B.S. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, the M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shangai, China, and the Ph.D. degree in electrical engineering from the University of Cincinnati, Cincinnati, OH, in 1994, 1997, and 2001, respectively.

Since December 2001, he has been a Senior Research Engineer at Intelligent Automation, Inc., Rockville, MD. His research interests include intelligent systems and fault diagnosis of nonlinear systems.

**Marios M. Polycarpou** (S'87–M'93–SM'98) received the B.A. (Cum Laude) degree in computer science and the B.Sc. (Cum Laude) degree in electrical engineering from Rice University, Houston, TX, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1987, 1989, and 1992, respectively.

In 1992, he joined the Department of Electrical and Computer Engineering and Computer Science, University of Cincinnati, Cincinnati, OH, where he is currently an Associate Professor. He teaches and conducts research in the areas of systems and control, adaptive and intelligent systems, neural network learning, fault diagnosis, and cooperative control.

Dr. Polycarpou is an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is Vice President, Conferences, of the IEEE Neural Network Society.

**Thomas Parisini** (S'87–M'88–SM'99) was born in Genoa, Italy, in 1963. He received the "Laurea" (*Cum Laude*) and Ph.D. degrees in electronic engineering from the University of Genoa, Italy, in 1988 and 1993, respectively.

From 1988 to 1995, he was with DIST, University of Genoa. In 1995, he joined DEEI, University of Trieste, Italy, as an Assistant Professor, and in 1998, he joined DEI, Politecnico di Milano, Italy, as Associate Professor. In 2001, he was appointed Full Professor and Danieli Endowed Chair of Automation Engineering at DEEI, University of Trieste. He is currently serving as an Associate Editor of *Automatica*, *The International Journal of Control*, and as the Subject Editor on Intelligent Control of *The International Journal of Adaptive Control and Signal Processing*. His research interests include neural-network approximations for optimal control and filtering problems, fault diagnosis for nonlinear systems, and hybrid control systems.

Dr. Parisini is the Chairman of the IEEE Control Systems Society Technical Committee on Intelligent Control, a member of the Board of Governors of the IEEE Control Systems Society, and an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.