Automatica 46 (2010) 290–299

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

# Fault diagnosis of a class of nonlinear uncertain systems with Lipschitz nonlinearities using adaptive estimation☆

Xiaodong Zhang[a], Marios M. Polycarpou[b], Thomas Parisini[c,*]

[a] Department of Electrical Engineering, Wright State University, Dayton, OH 45435, USA

[b] Department of Electrical and Computer Engineering, KIOS Research Center, University of Cyprus, Nicosia 1678, Cyprus

[c] Department of Electrical, Electronic and Computer Engineering, University of Trieste, 34127 Trieste, Italy

ARTICLE INFO

*Article history:*
Received 9 April 2009
Received in revised form
26 August 2009
Accepted 5 November 2009
Available online 22 December 2009

*Keywords:*
Fault detection
Fault isolation
Adaptive systems
Nonlinear systems
Uncertain dynamic systems

ABSTRACT

This paper presents a fault detection and isolation (FDI) scheme for a class of Lipschitz nonlinear systems with nonlinear and unstructured modeling uncertainty. This significantly extends previous results by considering a more general class of system nonlinearities which are modeled as functions of the system input and partially measurable state variables. A new FDI method is developed using adaptive estimation techniques. The FDI architecture consists of a fault detection estimator and a bank of fault isolation estimators. The fault detectability and isolability conditions, characterizing the class of faults that are detectable and isolable by the proposed scheme, are rigorously established. The fault isolability condition is derived via the so-called fault mismatch functions, which are defined to characterize the mutual difference between pairs of possible faults. A simulation example of a single-link flexible joint robot is used to illustrate the effectiveness of the proposed scheme.



## 1. Introduction

In recent years, there has been significant research activity in the design and analysis of fault diagnosis and accommodation schemes (Blanke, Kinnaert, Lunze, & Staroswiecki, 2006; Chen & Patton, 1999; Frank, 1990; Gertler, 1998; Isermann, 2006). Considerable effort has been devoted to fault diagnosis of nonlinear systems under various kinds of assumptions and fault scenarios (see, for instance Floquet, Barbot, & Perruquetti, 2004; Garciá & Frank, 1997; Hammouri, Kinnaert, & El Yaagoubi, 1999; Krishnaswami & Rizzoni, 1997; Mhaskar, McFall, Gani, Christofides, & Davis, 2008; De Persis & Isidori, 2001; Wang, Huang, & Daley, 1997; Yan & Edwards, 2007; Zhang, Polycarpou, & Parisini, 2002, and the references cited therein).

The idea of using adaptive and learning techniques in fault diagnosis and accommodation has been presented in Jiang, Staroswiechi, and Cocquempot (2004), Kabore and Wang (2001), Polycarpou and Helmicki (1995), Tang, Tao, and Joshi (2007), Tao, Joshi, and Ma (2001), Vemuri and Polycarpou (1997), Wang et al. (1997), Xu and Zhang (2004), Zhang, Polycarpou, and Parisini (2001), Zhang, Parisini, and Polycarpou (2005). In previous papers, Vemuri and Polycarpou (1997), Zhang et al. (2005), Zhang et al. (2001), the authors developed an adaptive approximation based fault diagnosis methodology for a class of nonlinear systems in which the known nonlinearity is represented as a function of measurable system signals (i.e., system input and output variables). It is assumed that there exists a diffeomorphism that can transform more general nonlinear systems into the class of systems under consideration.

In this paper, we extend the previous results by considering a more general class of nonlinear systems without the need of a diffeomorphism. More specifically, the known nonlinearity under consideration is modeled as a nonlinear function of the system input and state variables and satisfies a Lipschitz condition. With partially measurable states and the possible presence of unstructured modeling uncertainty, the design of fault diagnosis methods for such Lipschitz nonlinear systems is a challenging problem. Several researchers have investigated this problem and presented some interesting results under the framework of *structured* modeling uncertainty and faults. For instance, Vijayaraghavan, Rajamani, and Bokor (2007) addressed the sensor fault diagnosis problem by assuming the absence of modeling uncertainty, and in Chen and Saif (2007), Yan and Edwards (2007) sliding mode

☆ The material in this paper was partially presented at 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Barcelona, Spain, June 30–July 3, 2009. This paper was recommended for publication in revised form by Associate Editor Michael A. Henson under the direction of Editor Frank Allgöwer.

* Corresponding author. Tel.: +39 040 5587138, +39 335 8294017; fax: +39 040 5583460.

*E-mail addresses:* xiaodong.zhang@wright.edu (X. Zhang), mpolycar@ucy.ac.cy (M.M. Polycarpou), t.parisini@paperplaza.net (T. Parisini).

0005-1098/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2009.11.014

observer based fault estimation methods were proposed under certain assumptions on the distribution matrices of the structured modeling uncertainty and faults. However, in many practical systems, the modeling uncertainty and the fault function are often *unstructured*, which makes it difficult to achieve robustness of the fault diagnosis methodology.

In this research work, we develop a new adaptive approximation based fault diagnosis scheme for a class of Lipschitz nonlinear systems with possibly nonlinear and *unstructured* modeling uncertainty. Unstructured modeling uncertainty refers to the case where the modeling uncertainty function appears possibly in all state equations without being pre-multiplied by a known distribution matrix that satisfies certain conditions. The objective is to detect any faults as soon as possible and to determine whether a particular fault type in a partially known nonlinear fault set has occurred. The FDI architecture consists of a fault detection estimator (FDE) and a bank of fault isolation estimators (FIEs). The occurrence of a fault is detected if at least one component of the output estimation error generated by the FDE exceeds the corresponding adaptive threshold, which is derived analytically. Each FIE is designed based on the functional structure of a particular fault. The adaptive thresholds for fault isolation are designed, such that, if a particular fault occurs, then each component of the residual vector generated by the corresponding FIE always remains below its adaptive threshold, therefore avoiding false alarms. Some preliminary results of this research work have been presented in Zhang, Polycarpou, and Parisini (2009).

The fault diagnosis method is presented with a rigorous analytical framework aimed at characterizing the properties the diagnostic scheme. The analysis of the fault diagnosis scheme focuses on: (i) determining adaptive thresholds for fault detection and isolation; (ii) deriving fault detectability conditions characterizing the class of faults that can be detected; (iii) deriving fault isolability conditions characterizing the class of faults that can be isolated by the proposed method. The fault isolability condition is rigorously established based on the so-called *fault mismatch function*, which provides a suitable measure of the mutual "difference" between possible faults (see Zhang et al. (2002)).

The paper is organized as follows. In Section 2, the problem of nonlinear fault diagnosis is formulated. The FDI architecture and the fault detection scheme are presented in Section 3. In Section 4, the derivation of adaptive thresholds for robust fault isolation is given, while the fault isolability condition is analyzed in Section 5. Section 6 describes a simulation example of a single-link robotic arm with a revolute elastic joint, illustrating the effectiveness of the robust FDI scheme. Finally, Section 7 presents some concluding remarks.

## 2. Problem formulation

Consider a class of nonlinear multi-input–multi-output (MIMO) dynamic systems described by

$$\begin{aligned} \dot{x} &= Ax + \zeta(x,u) + \varphi(x,u,t) + \beta(t-T_0)D\phi(y,u) \\ y &= Cx \end{aligned} \tag{1}$$

where $x \in \Re^n$ is the system state vector, $u \in \Re^m$ is the input vector, $y \in \Re^p$ is the output vector ($n \geq p$), $\zeta : \Re^n \times \Re^m \mapsto \Re^n$, $\varphi : \Re^n \times \Re^m \times \Re^+ \mapsto \Re^n$, $\phi : \Re^p \times \Re^m \mapsto \Re^q$ are smooth vector fields. The constant matrices $D \in \Re^{n\times q}$ and $C \in \Re^{p\times n}$ with $q \leq p$ are of full rank, and $(A, C)$ is an observable pair. The model given by

$$\begin{aligned} \dot{x}_N &= Ax_N + \zeta(x_N,u) \\ y_N &= Cx_N \end{aligned}$$

is the *known nominal* system model. The vector field $\varphi$ appearing in (1) represents the modeling uncertainty, and $\beta(t-T_0)D\phi(y,u)$ denotes the changes in the system dynamics due to the occurrence of a fault (Vemuri & Polycarpou, 1997). Specifically, $\beta(t-T_0)$ is the time profile of a fault which occurs at some unknown time $T_0$, $\phi(y,u)$ represents the nonlinear fault function, and $D$ is a fault distribution matrix.

In this paper, we only consider the case of *abrupt* (sudden) faults (for some results on incipient faults, see Demetriou and Polycarpou (1998), Zhang et al. (2002)); therefore, $\beta(\cdot)$ takes on the form of a step function. Moreover, the design and analysis in this paper is based on the assumption that only *a single fault* occurs.

**Assumption 1.** The fault distribution matrix $D$ in (1) satisfies

- rank(CD) = q
- invariant zeros of (A,D,C) lie in the left half plane.

As described in Edwards, Spurgeon, and Patton (2000), under Assumption 1, there exists a linear transformation of coordinates $z = Tx = [z_1^\top\ z_2^\top]^\top$ with $z_1 \in \Re^{(n-p)}$ and $z_2 \in \Re^p$, such that

- $TAT^{-1} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix}$, where the matrix $\mathcal{A}_{11} \in \Re^{(n-p)\times(n-p)}$ is Hurwitz stable.
- $TD = \begin{bmatrix} 0 \\ D_2 \end{bmatrix}$, where $D_2 \in \Re^{p\times q}$.
- $CT^{-1} = [0\ I_p]$, where $I_p \in \Re^{p\times p}$ is an identity matrix.

Therefore, in the new coordinate system, the system (1) is described by

$$\begin{aligned} \dot{z}_1 &= \mathcal{A}_{11}z_1 + \mathcal{A}_{12}z_2 + \rho_1(z,u) + \eta_1(z,u,t) \\ \dot{z}_2 &= \mathcal{A}_{21}z_1 + \mathcal{A}_{22}z_2 + \rho_2(z,u) + \eta_2(z,u,t) \\ &\quad + \beta(t-T_0)D_2\phi(y,u) \\ y &= z_2, \end{aligned} \tag{2}$$

where $\begin{bmatrix} \rho_1(z,u) \\ \rho_2(z,u) \end{bmatrix} = T\zeta(T^{-1}z,u)$ and $\begin{bmatrix} \eta_1(z,u,t) \\ \eta_2(z,u,t) \end{bmatrix} = T\varphi(T^{-1}z,u,t)$. With a more general structure of the nonlinear fault model, system (2) can be extended to

$$\begin{aligned} \dot{z}_1 &= \mathcal{A}_{11}z_1 + \mathcal{A}_{12}z_2 + \rho_1(z,u) + \eta_1(z,u,t) \\ \dot{z}_2 &= \mathcal{A}_{21}z_1 + \mathcal{A}_{22}z_2 + \rho_2(z,u) + \eta_2(z,u,t) \\ &\quad + \beta(t-T_0)f(y,u) \\ y &= \bar{C}z_2, \end{aligned} \tag{3}$$

where $f : \Re^p \times \Re^m \mapsto \Re^p$ is a smooth vector field representing the nonlinear fault function under consideration, and $\bar{C} \in \Re^{p\times p}$ is a nonsingular matrix. Clearly, (2) is a special case of (3) with $f(y,u) = D_2\phi(y,u)$ and $\bar{C} = I_p$.

It is assumed that there are $N$ types of possible faults in the fault set $\mathcal{F}$; specifically, the unknown fault function $f(y,u)$ in (3) belongs to a finite set of fault types given by

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

Each fault type $f^s$, $s = 1, \ldots, N$, is in the form of

$$f^s(y,u) \triangleq \left[ (\theta_1^s(t))^\top g_1^s(y,u), \ldots, (\theta_p^s(t))^\top g_p^s(y,u) \right]^\top, \tag{5}$$

where each $\theta_i^s(t)$, $i = 1, \ldots, p$, is an unknown parameter vector assumed to belong to a known corresponding compact and convex set $\Theta_i^s$ (i.e., $\theta_i^s(t) \in \Theta_i^s \subset \Re^{l_i^s}$, $\forall t \geq 0$) and $g_i^s$: $\Re^p \times \Re^m \mapsto \Re^{l_i^s}$, is a known smooth vector field. As discussed in Zhang et al. (2001), Zhang et al. (2002), the fault model described by (4) and (5) characterizes a general class of nonlinear faults where the vector field $g_i^s$ represents the functional structure of the $s$th fault affecting the $i$th state in $z_2$, while the unknown parameter vector $\theta_i^s(t)$ characterizes the time-varying "magnitude" of the fault. Throughout the paper, the following assumptions are used:

**Assumption 2.** The functions $\eta_1$ and $\eta_2$ in (3), representing the unstructured modeling uncertainty, are possibly unknown nonlinear functions of $z$, $u$, and $t$, but bounded, i.e., $\forall(z, y, u) \in \mathcal{Z} \times \mathcal{Y} \times \mathcal{U}, \forall t \geq 0$,

$$|\eta_1(z, u, t)| \leq \bar{\eta}_1, \qquad |\eta_2(z, u, t)| \leq \bar{\eta}_2(y, u, t), \tag{6}$$

where the constant bound $\bar{\eta}_1$ and the bounding function $\bar{\eta}_2(y, u, t)$ are known, and $\bar{\eta}_2$ is uniformly bounded in $\mathcal{Y} \times \mathcal{U} \times \Re^+$. Additionally, $\mathcal{Z} \subset \Re^n$, $\mathcal{U} \subset \Re^m$, and $\mathcal{Y} \subset \Re^p$ are compact sets of admissible state variables, inputs, and outputs, respectively.

**Assumption 3.** The system state vector $z$ remains bounded before and after the occurrence of a fault, i.e., $z(t) \in L_\infty, \forall t \geq 0$.

**Assumption 4.** The rate of change of each fault parameter vector $\theta_i^s(t)$ in (5) $(s = 1, \ldots, N)$ is uniformly bounded, i.e., $|\dot{\theta}^s(t)| \leq \alpha^s$ for all $t \geq 0$, where $\theta^s(t) \triangleq \left[\left(\theta_1^s(t)\right)^\top, \ldots, \left(\theta_p^s(t)\right)^\top\right]^\top$, and $\alpha^s$ is a known constant.

**Assumption 5.** The known nonlinear terms $\rho_1(z, u)$ and $\rho_2(z, u)$ in (3) are uniformly Lipschitz in $u \in \mathcal{U}$, i.e., $\forall z, \hat{z} \in \mathcal{Z}$,

$$|\rho_1(z, u) - \rho_1(\hat{z}, u)| \leq \gamma_1 |z - \hat{z}| \tag{7}$$

$$|\rho_2(z, u) - \rho_2(\hat{z}, u)| \leq \gamma_2 |z - \hat{z}|, \tag{8}$$

where $\gamma_1$ and $\gamma_2$ are the known Lipschitz constants for $\rho_1(z, u)$ and $\rho_2(z, u)$, respectively.

Assumption 2 characterizes the class of modeling uncertainties under consideration. The bounds on the *unstructured* modeling uncertainties are needed in order to be able to distinguish between the effects of faults and modeling uncertainty (see Emami-Naeini, Akhter, and Rock (1998), Zhang et al. (2001), Zhang et al. (2002)). In this paper, $\bar{\eta}_1$ is assumed to be a constant to facilitate the derivation of adaptive thresholds for fault detection and isolation. It is worth noting that the *unstructured* modeling uncertainty considered in this paper is more general than the *structured* uncertainty considered for Lipschitz nonlinear systems in the fault diagnosis literature (e.g., Chen and Saif (2007), Vijayaraghavan et al. (2007), Yan and Edwards (2007)), which, to achieve robustness, additionally assumes that certain rank conditions are satisfied by the distribution matrix of the *structured* modeling uncertainty. On the other hand, the utilization of structured uncertainty with additional assumptions on the distribution matrix may allow the design of FDI schemes that completely decouple the fault from modeling uncertainty.

Assumption 3 requires the boundedness of the state variables before and after the occurrence of a fault. Hence, it is assumed that the feedback control system is capable of retaining the boundedness of the state variables even in the presence of a fault. This is a technical assumption required for well-posedness since the FDI design that we consider does not influence the closed-loop dynamics and stability. It is important to note that the proposed FDI design does not depend on the structure of the controller.

In Assumption 4, known bounds on the rates of change of $\theta^s(t)$ are assumed. In practice, the rate bounds $\alpha^s$ can be set by exploiting some *a priori* knowledge on the fault developing dynamics. In the case of a constant fault magnitude, we simply set $\alpha^s = 0$. Note that the fault time profile is represented by the function $\beta(t - T_0)$ in (3), while $\theta^s(t)$ only represents the (possibly time-varying) fault magnitude.

**Remark 1.** The known nominal system model in (3) is similar to the model used in Yan and Edwards (2007). The objective of this paper is to develop a robust FDI scheme using adaptive approximation techniques, while Yan and Edwards (2007) presented a fault estimation method using a sliding mode observer, which requires additional assumptions on the distribution matrices of the modeling uncertainty terms $\eta_1$ and $\eta_2$ as well as the fault function $f$ in (3). Moreover, in our previous work (Vemuri & Polycarpou, 1997; Zhang et al., 2005, 2001), the nonlinear term $\zeta$ in (1) is modeled as $\zeta(y, u)$ instead of $\zeta(x, u)$. With the nonlinearity being a function of partially measurable state vector $x$ and in the presence of possibly *unstructured* modeling uncertainty, the design and analysis of adaptive fault diagnostic estimators clearly become more challenging.

## 3. Fault detection and isolation architecture

The fault detection and isolation architecture is based on a bank of $N + 1$ nonlinear adaptive estimators, where $N$ is the number of different nonlinear fault types in the fault set $\mathcal{F}$ (see (4)). One of the nonlinear adaptive estimators is the FDE used for detecting the occurrence of any faults, while the remaining $N$ nonlinear adaptive estimators are FIEs, which are activated after fault detection for the purpose of determining the particular type of fault that has occurred (see Zhang et al. (2002) for the general scheme).

### 3.1. Fault detection scheme

Based on the system model given by (3), the FDE is chosen as follows:

$$\begin{aligned}
\dot{\hat{z}}_1 &= \mathcal{A}_{11}\hat{z}_1 + \mathcal{A}_{12}\bar{C}^{-1}y + \rho_1(\hat{z}, u), \\
\dot{\hat{z}}_2 &= \mathcal{A}_{21}\hat{z}_1 + \mathcal{A}_{22}\hat{z}_2 + \rho_2(\hat{z}, u) + L(y - \hat{y}), \\
\hat{y} &= \bar{C}\hat{z}_2,
\end{aligned} \tag{9}$$

where $\hat{z}_1, \hat{z}_2$, and $\hat{y}$ denote the estimated state and output variables, respectively, $L \in \Re^{p\times p}$ is a design gain matrix, and $\hat{z} \triangleq [\ (\hat{z}_1)^\top \quad (\bar{C}^{-1}y)^\top\ ]^\top$. The initial conditions are $\hat{z}_1(0) = 0$ and $\hat{z}_2(0) = 0$. Let $\tilde{z}_1 \triangleq z_1 - \hat{z}_1$ and $\tilde{z}_2 \triangleq z_2 - \hat{z}_2$ denote the state estimation errors, and $\tilde{y} \triangleq y - \hat{y}$ denote the output estimation error. Then, before fault occurrence (i.e., for $t < T_0$), we have

$$\dot{\tilde{z}}_1 = \mathcal{A}_{11}\tilde{z}_1 + \rho_1(z, u) - \rho_1(\hat{z}, u) + \eta_1 \tag{10}$$

$$\dot{\tilde{z}}_2 = \bar{\mathcal{A}}_{22}\tilde{z}_2 + \mathcal{A}_{21}\tilde{z}_1 + \rho_2(z, u) - \rho_2(\hat{z}, u) + \eta_2 \tag{11}$$

$$\tilde{y} = \bar{C}(z_2 - \hat{z}_2) = \bar{C}\tilde{z}_2, \tag{12}$$

where $\bar{\mathcal{A}}_{22} \triangleq \mathcal{A}_{22} - L\bar{C}$. Note that, since $\bar{C}$ is nonsingular, we can always choose $L$ to make $\bar{\mathcal{A}}_{22}$ stable.

In the analysis of the state estimation error $\tilde{z}_1(t)$, we will need the following technical results:

**Lemma 1** (*Bellman–Gronwall Lemma (Ioannou & Sun, 1996)*). *Let $t_0$ be a given time instant and $c_0$, $c_1$, $c_2$, $\lambda$ be nonnegative constants, and $\kappa(t)$ a nonnegative piecewise continuous function of time. If $h(t)$ satisfies the inequality*

$$h(t) \leq c_0 e^{-\lambda(t-t_0)} + c_1 + c_2 \int_{t_0}^{t} e^{-\lambda(t-\tau)}\kappa(\tau)h(\tau)d\tau, \quad \forall t \geq t_0,$$

*then*

$$h(t) \leq (c_0 + c_1)e^{-\lambda(t-t_0)}e^{c_2\int_{t_0}^{t}\kappa(s)ds} + c_1\lambda\int_{t_0}^{t} e^{-\lambda(t-\tau)}e^{c_2\int_{\tau}^{t}\kappa(s)ds}d\tau, \quad \forall t \geq t_0.$$

**Lemma 2.** *Consider the system described by (3) and the fault detection estimator described by (9). Let $k_0$ and $\lambda_0$ be positive constants chosen such that $\|e^{\mathcal{A}_{11}t}\| \leq k_0 e^{-\lambda_0 t}$. Assume that $\lambda_0 > k_0\gamma_1$, where $\gamma_1$ is the Lipschitz constant given in (7). Then, for $0 \leq t < T_0$, the state estimation error $\tilde{z}_1(t)$ satisfies:*

$$|\tilde{z}_1(t)| \leq \frac{k_0\bar{\eta}_1}{\lambda_0 - k_0\gamma_1} + \left(k_0\omega_1 - \frac{k_0\bar{\eta}_1}{\lambda_0 - k_0\gamma_1}\right)e^{-(\lambda_0 - k_0\gamma_1)t}, \tag{13}$$

*where $\omega_1$ is a constant bound for $|z_1(0)|$, which will be defined later on.*

**Proof.** From (10), we have

$$\tilde{z}_1(t) = \int_0^t e^{\mathcal{A}_{11}(t-\tau)} \left[\rho_1(z,u) - \rho_1(\hat{z},u) + \eta_1(z,u,\tau)\right] d\tau + e^{\mathcal{A}_{11}t}\tilde{z}_1(0).$$

By using (6) and (7) and by applying the triangle inequality, we obtain

$$|\tilde{z}_1(t)| \le \frac{k_0\bar{\eta}_1}{\lambda_0} + \gamma_1 \int_0^t k_0 e^{-\lambda_0(t-\tau)}|z(\tau) - \hat{z}(\tau)|\, d\tau + k_0\left(\omega_1 - \frac{\bar{\eta}_1}{\lambda_0}\right) e^{-\lambda_0 t}, \tag{14}$$

where $k_0$ and $\lambda_0$ are positive constants chosen such that $\|e^{\mathcal{A}_{11}t}\| \le k_0 e^{-\lambda_0 t}$ (since $\mathcal{A}_{11}$ is stable, such constants $k_0$ and $\lambda_0$ always exist (Ioannou & Sun, 1996)), and $\omega_1$ is a (possibly conservative) constant bound for $|z_1(0)|$, such that $|\tilde{z}_1(0)| = |z_1(0)| \le \omega_1$ (note that $\hat{z}_1(0) = 0$). Based on Assumption 3, such a constant bound $\omega_1$ can always be chosen.

By substituting

$$|z(\tau) - \hat{z}(\tau)| = |\tilde{z}_1(\tau)|, \tag{15}$$

into (14), we obtain

$$|\tilde{z}_1(t)| \le \frac{k_0\bar{\eta}_1}{\lambda_0} + k_0\gamma_1 \int_0^t e^{-\lambda_0(t-\tau)}|\tilde{z}_1(\tau)|\, d\tau + k_0\left(\omega_1 - \frac{\bar{\eta}_1}{\lambda_0}\right) e^{-\lambda_0 t}. \tag{16}$$

Now, by applying Lemma 1 to (16) with $c_0 = k_0\left(\omega_1 - \frac{\bar{\eta}_1}{\lambda_0}\right)$ (we can always choose $\omega_1$ to ensure that $c_0 \ge 0$), $c_1 = \frac{k_0\bar{\eta}_1}{\lambda_0}$, $c_2 = k_0\gamma_1$, and $\kappa(t) = 1$, the proof of (13) can be immediately concluded. □

Next, we analyze each component of the output estimation error, i.e., $\tilde{y}_j(t) \triangleq \bar{C}_j\tilde{z}_2(t)$, $j = 1, \ldots, p$, where $\bar{C}_j$ is the $j$th row vector of matrix $\bar{C}$.

By applying the triangle inequality and using (8), (11), (15) and (6), it can be shown that

$$|\tilde{y}_j(t)| \le k_j \int_0^t e^{-\lambda_j(t-\tau)}\left[(\|\mathcal{A}_{21}\| + \gamma_2)\, |\tilde{z}_1(\tau)| + \bar{\eta}_2\right] d\tau + k_j\omega_2 e^{-\lambda_j t}, \tag{17}$$

where $k_j$ and $\lambda_j$ are positive constants chosen such that $|\bar{C}_j e^{\bar{\mathcal{A}}_{22}t}| \le k_j e^{-\lambda_j t}$ (since $\bar{\mathcal{A}}_{22}$ is stable, constants $k_j$ and $\lambda_j$ satisfying the above inequality always exist (Ioannou & Sun, 1996)), and $\omega_2$ is a (possibly conservative) bound for $|z_2(0)|$, such that $|\tilde{z}_2(0)| = |z_2(0)| \le \omega_2$ (note that $\hat{z}_2(0) = 0$). By Assumption 3, we can always find such a constant bound $\omega_2$. Based on (13) and (17), we obtain

$$|\tilde{y}_j(t)| < k_j \int_0^t e^{-\lambda_j(t-\tau)}\left[(\|\mathcal{A}_{21}\| + \gamma_2)\, \chi(\tau) + \bar{\eta}_2\right] d\tau + k_j\omega_2 e^{-\lambda_j t}, \tag{18}$$

where

$$\chi(t) \triangleq \frac{k_0\bar{\eta}_1}{\lambda_0 - k_0\gamma_1} + \left(k_0\,\omega_1 - \frac{k_0\bar{\eta}_1}{\lambda_0 - k_0\gamma_1}\right) e^{-(\lambda_0 - k_0\gamma_1)t}. \tag{19}$$

According to (18), the decision scheme for fault detection is as follows:

**Fault Detection Decision Scheme.** *The decision on the occurrence of a fault (detection) is made when the modulus of at least one component of the output estimation error (i.e., $\tilde{y}_j(t)$) exceeds its corresponding threshold $\nu_j(t)$ given by*

$$\nu_j(t) \triangleq k_j \int_0^t e^{-\lambda_j(t-\tau)} \left[(\|\mathcal{A}_{21}\| + \gamma_2)\, \chi(\tau) + \bar{\eta}_2\right] d\tau + k_j\omega_2 e^{-\lambda_j t}. \tag{20}$$

*The fault detection time $T_d$ is defined as the first time instant such that $\left|\tilde{y}_j(T_d)\right| > \nu_j(T_d)$, for some $T_d \ge T_0$ and some $j \in \{1, \ldots, p\}$, that is,*

$$T_d \triangleq \inf \bigcup_{j=1}^{p} \left\{t \ge 0\colon \left|\tilde{y}_j(t)\right| > \nu_j(t)\right\}.$$

The above design and analysis is summarized by the following result:

**Theorem 1** (*Robustness*). *For the nonlinear system (3), the fault detection decision scheme, characterized by the fault detection estimator (9) and adaptive thresholds (20), guarantees that there will be no false alarms before fault occurrence (i.e., for $t \le T_0$).*

**Remark 2.** The adaptive thresholds $\nu_j(t)$ given by (20) can be easily implemented using linear filtering techniques (Zhang et al., 2001, 2002). Additionally, $\omega_1$ and $\omega_2$ in (19) and (20) are (possibly conservative) bounds for the unknown initial conditions $z_1(0)$ and $z_2(0)$. However, since the effect of these bounds decreases exponentially (i.e., they are multiplied by $e^{-(\lambda_0 - k_0\gamma_1)t}$ and $e^{-\lambda_j t}$, respectively), the use of such bounds will not affect significantly the performance of the fault detection algorithm.

As is well known in the fault diagnosis literature (Blanke et al., 2006; Chen & Patton, 1999; Frank, 1990; Gertler, 1998; Isermann, 2006), there is an inherent tradeoff between robustness and fault sensitivity. The following theorem characterizes implicitly the class of faults that are detectable by the proposed FDI scheme:

**Theorem 2** (*Fault Detectability*). *For the nonlinear system (3) with the fault detection decision scheme defined by the fault detection estimator (9) and the adaptive thresholds (20), if there exist some time instant $T_d > T_0$ and some $j \in \{1, \ldots, p\}$, such that the fault function $f(y,u)$ satisfies*

$$\left| \int_{T_0}^{T_d} \bar{C}_j e^{\bar{\mathcal{A}}_{22}(T_d-\tau)} f(y(\tau), u(\tau))d\tau \right| > 2k_j \int_{T_0}^{T_d} e^{-\lambda_j(T_d-\tau)}\left[\left(\|\mathcal{A}_{21}\| + \gamma_2\right)\chi(\tau) + \bar{\eta}_2(y,u,\tau)\right]d\tau + \left[k_j|\tilde{z}_2(T_0)| + \nu_j(T_0)\right]e^{-\lambda_j(T_d-T_0)}, \tag{21}$$

*then the fault will be detected at time $t = T_d$, i.e., $|\tilde{y}_j(T_d)| > \nu_j(T_d)$.*

**Proof.** In the presence of a fault (i.e., for $t \ge T_0$), based on (3) and (9), the dynamics of the state estimation errors $\tilde{z}_1 \triangleq z_1 - \hat{z}_1$ and $\tilde{z}_2 \triangleq z_2 - \hat{z}_2$ satisfies

$$\dot{\tilde{z}}_1 = \mathcal{A}_{11}\tilde{z}_1 + \rho_1(z,u) - \rho_1(\hat{z},u) + \eta_1(z,u,t) \tag{22}$$

$$\dot{\tilde{z}}_2 = \bar{\mathcal{A}}_{22}\tilde{z}_2 + \mathcal{A}_{21}\tilde{z}_1 + \rho_2(z,u) - \rho_2(\hat{z},u) + \eta_2(z,u,t) + f(y,u). \tag{23}$$

Therefore, each component of the output estimation error, $\tilde{y}_j(t) = y_j(t) - \hat{y}_j(t)$, $j = 1, \ldots, p$, is given by

$$\tilde{y}_j(t) = \int_{T_0}^t \bar{C}_j e^{\bar{\mathcal{A}}_{22}(t-\tau)} \left[\rho_2(z(\tau),u(\tau)) - \rho_2(\hat{z}(\tau),u(\tau))\right] d\tau + \int_{T_0}^t \bar{C}_j e^{\bar{\mathcal{A}}_{22}(t-\tau)}[\mathcal{A}_{21}\tilde{z}_1(\tau) + \eta_2(z(\tau),u(\tau),\tau)]d\tau + \int_{T_0}^t \bar{C}_j e^{\bar{\mathcal{A}}_{22}(t-\tau)} f(y,u)d\tau + \bar{C}_j e^{\bar{\mathcal{A}}_{22}(t-T_0)}\tilde{z}_2(T_0).$$

From (22), we have $|\tilde{z}_1(t)| \le \chi(t)$ (by using similar arguments as in the derivation of the fault detection threshold (20)); therefore, by applying the triangular inequality and based on (6), (8) and (15), it follows that

$$|\tilde{y}_j(t)| \ge \left| \int_{T_0}^{t} \bar{C}_j e^{\bar{A}_{22}(t-\tau)} f(y,u) d\tau \right| - k_j |\tilde{z}_2(T_0)| e^{-\lambda_j (t-T_0)} - k_j \int_{T_0}^{t} e^{-\lambda_j(t-\tau)} \bar{\eta}_2(y,u,\tau) d\tau - k_j \int_{T_0}^{t} e^{-\lambda_j(t-\tau)} \left[ \|\mathcal{A}_{21}\| + \gamma_2 \right] \chi(\tau) d\tau. \quad (24)$$

Since $\int_0^t = \int_0^{T_0} + \int_{T_0}^t$, the detection threshold $\nu_j(t)$ given by (20) for $t > T_0$ can be rewritten as

$$\nu_j(t) = k_j \int_{T_0}^{t} e^{-\lambda_j(t-\tau)} \left[ (\|\mathcal{A}_{21}\| + \gamma_2)\, \chi(\tau) + \bar{\eta}_2(y,u,\tau) \right] d\tau + e^{-\lambda_j(t-T_0)} \nu_j(T_0). \quad (25)$$

Based on (24) and (25), if there exists $T_d > T_0$, such that condition (21) is satisfied, then we can conclude that $|\tilde{y}_i(T_d)| > \nu_j(T_d)$, i.e., the fault is detected at time $t = T_d$. □

**Remark 3.** Noting that the integral on the left-hand side of (21) represents the filtered fault function, in qualitative terms, the fault detectability theorem states that, if the magnitude of the filtered fault function on the time interval $[T_0, T_d]$ becomes sufficiently large, then the fault can be detected.

### 3.2. Fault isolation estimators and decision scheme

Now, assume that a fault is detected at some time $T_d$; accordingly, by following the general approach presented in Zhang et al. (2002), at time $t = T_d$ the FIEs are activated. Each FIE corresponds to one potential fault type. Specifically, the following $N$ nonlinear adaptive estimators are used as isolation estimators: for each $s = 1, \ldots, N$, we have

$$\begin{aligned} \dot{\hat{z}}_1^s &= \mathcal{A}_{11}\hat{z}_1^s + \mathcal{A}_{12}\bar{C}^{-1}y + \rho_1(\hat{z}^s, u), \quad \hat{z}_1^s(T_d) = 0 \\ \dot{\hat{z}}_2^s &= \mathcal{A}_{21}\hat{z}_1^s + \mathcal{A}_{22}\hat{z}_2^s + \rho_2(\hat{z}^s, u) + L^s(y - \hat{y}^s) \\ &\quad + \hat{f}^s(y, u, \hat{\theta}^s) + \Omega^s \dot{\hat{\theta}}^s, \quad \hat{z}_2^s(T_d) = 0 \\ \dot{\Omega}^s &= \bar{A}_{22}\Omega^s + G^s(y,u), \quad \Omega^s(T_d) = 0 \\ \hat{y}^s &= \bar{C}\hat{z}_2^s, \end{aligned} \quad (26)$$

where $\hat{z}_1^s$, $\hat{z}_2^s$, and $\hat{y}^s$ denote the estimated state and output variables, respectively, $L^s \in \Re^{p\times p}$ is a design gain matrix (for the simplicity of presentation and without loss of generality, we let $L^s = L$), $\hat{z}^s \triangleq [\ (\hat{z}_1^s)^\top \ \ (\bar{C}^{-1}y)^\top\ ]^\top$, $\hat{f}^s(y,u,\hat{\theta}^s) = [(\hat{\theta}_1^s)^\top g_1^s(y,u), \ldots, (\hat{\theta}_p^s)^\top g_p^s(y,u)]$, and $\hat{\theta}_i^s \in \Re^{l_i^s}$, for $i = 1, \ldots, p$, is the estimate of the fault parameter vector provided by the $s$th isolation estimator. It is noted that, according to (5), the fault approximation model $\hat{f}^s$ is linear in the adjustable weights $\hat{\theta}^s$. Consequently, the gradient matrix $G^s \triangleq \partial \hat{f}^s(y,u,\hat{\theta}^s)/\partial\hat{\theta}^s = \mathrm{diag}\,[(g_1^s)^\top, \ldots, (g_p^s)^\top]$ does not depend on $\hat{\theta}^s$.

The adaptation in the isolation estimators arises due to the unknown parameter vector $\theta^s \triangleq \left[ \left(\theta_1^s\right)^\top, \ldots, \left(\theta_n^s\right)^\top \right]^\top$. The adaptive law for adjusting $\hat{\theta}^s$ is derived using the Lyapunov synthesis approach (see for example Ioannou and Sun (1996)). Specifically, the learning algorithm is chosen as follows

$$\dot{\hat{\theta}}^s = \mathcal{P}_{\Theta^s} \left\{ \Gamma \Omega^{s\top} \bar{C}^\top \tilde{y}^s \right\}, \quad (27)$$

where $\tilde{y}^s(t) \triangleq y(t) - \hat{y}^s(t)$ denotes the output estimation error of the $s$th estimator, $\Gamma > 0$ is a symmetric, positive-definite learning rate matrix, and $\mathcal{P}_{\Theta^s}$ is the projection operator restricting $\hat{\theta}^s$ to the corresponding *known* set $\Theta^s$ (in order to guarantee stability of the learning algorithm in the presence of modeling uncertainty (Farrell & Polycarpou, 2006; Ioannou & Sun, 1996)). The stability and learning properties of the FIEs are characterized by the following result:

**Theorem 3.** *In the presence of faults, the nonlinear adaptive fault isolation scheme described by (26) and (27) guarantees that, for each fault isolation estimation s, $s \in \{1, \ldots, N\}$,*

- *the estimate variables $\hat{z}_1^s(t)$, $\hat{z}_2^s(t)$, and $\hat{\theta}^s(t)$ are uniformly bounded;*
- *there exist a positive constant $\bar{\kappa}$ and a bounded function $\bar{\xi}^s(t)$ such that, for all finite $t > T_d$, the output estimation error satisfies*

$$\int_{T_d}^{t} |\tilde{y}^s(t)|^2 dt \le \bar{\kappa} + 2\int_{T_d}^{t} |\bar{\xi}^s(t)|^2 dt. \quad (28)$$

**Proof.** See Appendix A. □

**Remark 4.** Theorem 3 ensures the boundedness of all signals in the FIEs. Moreover, the performance measure given by (28) shows that the ability of the FIEs to learn the fault function is limited by the extended $L_2$ norm of $\bar{\xi}^s(t)$ (defined in Appendix A), which corresponds to the modeling uncertainties $\eta_1$ and $\eta_2$, the fault function approximation error, and unknown initial conditions.

The fault isolation decision scheme is based on the following intuitive principle (see Zhang et al. (2002)): if fault $s$ occurs at time $T_0$ and is detected at time $T_d$, then a set of adaptive threshold functions $\{\mu_j^s(t), j = 1, \ldots, p\}$ can be designed for the $s$th FIE, such that the $j$th component of its output estimation error satisfies $|\tilde{y}_j^s(t)| \le \mu_j^s(t)$, for all $t > T_d$. Consequently, for each $s = 1, \ldots, N$, such a set of adaptive thresholds $\{\mu_j^s(t), j = 1, \ldots, p\}$ can be associated with the output estimation error of the $s$th FIE. In the fault isolation procedure, if, for a particular isolation estimator $s$, there exists some $j \in \{1, \ldots, p\}$, such that the $j$th component of its output estimation error satisfies $|\tilde{y}_j^s(t)| > \mu_j^s(t)$ for some finite time $t > T_d$, then the possibility of the occurrence of fault $s$ can be excluded. Based on this intuitive idea, the following fault isolation decision scheme is devised:

**Fault Isolation Decision Scheme.** *If, for each $r \in \{1, \ldots, N\} \setminus \{s\}$, there exist some finite time $t^r > T_d$ and some $j \in \{1, \ldots, p\}$, such that $|\tilde{y}_j^r(t^r)| > \mu_j^r(t^r)$, then the occurrence of fault s is concluded.*

## 4. Adaptive thresholds for fault isolation

The threshold functions $\mu_j^s(t)$ clearly play a key role in the fault isolation decision scheme. The following lemma provides a bounding function for the output estimation error of the $s$th FIE in the case that fault $s$ occurs.

**Lemma 3.** *If fault s occurs, where $s \in \{1, \ldots, N\}$, then for all $t > T_d$, the jth component of the output estimation error of the sth isolation estimator satisfies the following inequality:*

$$|\tilde{y}_j^s(t)| \le k_j \int_{T_d}^{t} e^{-\lambda_j(t-\tau)} \left[ (\|\mathcal{A}_{21}\| + \gamma_2)\chi^s(\tau) \right] d\tau + k_j \int_{T_d}^{t} e^{-\lambda_j(t-\tau)} \left[ \bar{\eta}_2(y,u,\tau) + \alpha^s \|\Omega^s\| \right] d\tau + |(\bar{C}_j \Omega^s)^\top| \, |\tilde{\theta}^s| + k_j \omega_2 e^{-\lambda_j(t-T_d)}, \quad (29)$$

*where*

$$\chi^s(t) \triangleq \left(k_0\omega_1 - \frac{k_0\bar{\eta}_1}{\lambda_0 - k_0\gamma_1}\right) e^{-(\lambda_0 - k_0\gamma_1)(t-T_d)} + \frac{k_0\bar{\eta}_1}{\lambda_0 - k_0\gamma_1}, \quad (30)$$

*and* $\tilde{\theta}^s(t) \triangleq \hat{\theta}^s(t) - \theta^s(t)$ *represents the fault parameter vector estimation error.*

**Proof.** Denote the state estimation error of the $s$th isolation estimator by $\tilde{z}_1^s(t) \triangleq z_1(t) - \hat{z}_1^s(t)$ and $\tilde{z}_2^s(t) \triangleq z_2(t) - \hat{z}_2^s(t)$. By using (26) and (3), in the presence of fault $s$, the state estimation error of the $s$th estimator, for $t > T_d$, satisfies

$$\dot{\tilde{z}}_1^s = \mathcal{A}_{11}\tilde{z}_1^s + \rho_1(z,u) - \rho_1(\hat{z}^s,u) + \eta_1(z,u,t) \quad (31)$$

$$\dot{\tilde{z}}_2^s = \bar{\mathcal{A}}_{22}\tilde{z}_2^s + \mathcal{A}_{21}\tilde{z}_1^s + \rho_2(z,u) - \rho_2(\hat{z}^s,u) + \eta_2(z,u,t) + f(y,u) - \hat{f}^s(y,u,\hat{\theta}^s) - \Omega^s\dot{\hat{\theta}}^s. \quad (32)$$

By substituting $f(y,u) = G^s\theta^s$ and $\hat{f}^s = G^s\hat{\theta}^s$ into (32) and by using $G^s = \dot{\Omega}^s - \bar{\mathcal{A}}_{22}\Omega^s$ (see (26)), we have

$$\dot{\tilde{z}}_2^s = \bar{\mathcal{A}}_{22}\left(\tilde{z}_2^s + \Omega^s\tilde{\theta}^s\right) + \mathcal{A}_{21}\tilde{z}_1^s + \rho_2(z,u) - \rho_2(\hat{z}^s,u) + \eta_2(z,u,t) - \frac{\mathrm{d}}{\mathrm{d}t}(\Omega^s\tilde{\theta}^s) - \Omega^s\dot{\theta}^s(t).$$

By letting $\bar{z}_2^s \triangleq \tilde{z}_2^s + \Omega^s\tilde{\theta}^s$, the above equation can be rewritten as

$$\dot{\bar{z}}_2^s = \bar{\mathcal{A}}_{22}\bar{z}_2^s + \mathcal{A}_{21}\tilde{z}_1^s + \rho_2(z,u) - \rho_2(\hat{z}^s,u) + \eta_2(z,u,t) - \Omega^s\dot{\theta}^s(t). \quad (33)$$

By defining each component of the output estimation error $\tilde{y}_j^s(t) \triangleq y_j(t) - \hat{y}_j^s(t), j = 1,\dots p$, and using (26) and (3), we have

$$\tilde{y}_j^s(t) = \bar{C}_j\tilde{z}_2^s(t) = \bar{C}_j\left(\bar{z}_2^s(t) - \Omega^s\tilde{\theta}^s\right). \quad (34)$$

Now, based on (34) and on the solution of (33), as well as Assumptions 2, 4 and 5, after some algebraic manipulations, it can be shown that

$$|\tilde{y}_j^s(t)| \le k_j\int_{T_d}^t e^{-\lambda_j(t-\tau)}\left[\|\mathcal{A}_{21}\||\tilde{z}_1^s(\tau)| + \bar{\eta}_2(y,u,\tau)\right]\mathrm{d}\tau + k_j\int_{T_d}^t e^{-\lambda_j(t-\tau)}\left[\gamma_2|z(\tau) - \hat{z}^s(\tau)| + \alpha^s\|\Omega^s\|\right]\mathrm{d}\tau + k_je^{-\lambda_j(t-T_d)}|\bar{z}_2^s(T_d)| + |(\bar{C}_j\Omega^s)^\top||\tilde{\theta}^s|.$$

Note that (31) is similar to (10). Thus, based on (13) and (30), inequality (29) follows directly from the initial conditions $\hat{z}_2^s(T_d) = 0$, $\Omega^s(T_d) = 0$, and $|z_2^s(T_d)| \le \omega_2$. □

Although Lemma 3 provides an upper bound on the output estimation error of the $s$th estimator, the right-hand side of (29) cannot be directly used as a threshold function for fault isolation because $\tilde{\theta}^s(t)$ is not available (note that parameter convergence is not guaranteed since we do not make the restrictive assumption of persistency of excitation of signals (Farrell & Polycarpou, 2006; Ioannou & Sun, 1996)). However, since the estimate $\hat{\theta}^s$ belongs to the known compact set $\Theta^s$, we have $\left|\theta^s - \hat{\theta}^s(t)\right| \le \kappa^s(t)$ for a suitable $\kappa^s(t)$ depending on the geometric properties of set $\Theta^s$ (see Zhang et al., 2001, 2002). Hence, based on the above discussions, the following threshold function is chosen:

$$\mu_j^s(t) = k_j\int_{T_d}^t e^{-\lambda_j(t-\tau)}\left[(\|\mathcal{A}_{21}\| + \gamma_2)\chi^s(\tau) + \bar{\eta}_2(y,u,\tau) + \alpha^s\|\Omega^s\|\right]\mathrm{d}\tau + |(\bar{C}_j\Omega^s)^\top|\kappa^s(t) + k_j\omega_2e^{-\lambda_j(t-T_d)}, \quad (35)$$

which again can be easily implemented on-line using linear filtering techniques (Zhang et al., 2001, 2002).

## 5. Fault isolability analysis

For our purpose, a fault is said to be *isolable* if the fault isolation scheme is able to reach a correct decision in finite time. Intuitively (and following the general approach outlined in Zhang et al. (2002)), faults are isolable if they are *mutually different* according to a certain measure quantifying the difference in the effects that different faults have on measurable outputs and on the estimated quantities in the isolation scheme. To quantify this concept, we introduce the *fault mismatch function* between the $s$th fault and the $r$th fault:

$$h_j^{sr}(t) \triangleq \bar{C}_j\left(\Omega^s\theta^s - \Omega^r\hat{\theta}^r\right), \quad (36)$$

where $r, s = 1,\dots,N, r \ne s$. From a qualitative point of view, $h_j^{sr}(t)$ can be interpreted as a filtered version of the difference between the actual fault function $G^s\theta^s$ and its estimate $G^r\hat{\theta}^r$ associated with isolation estimator $r$ whose structure does not match the actual fault $s$. Recalling that each FIE corresponds to one of the nonlinear faults in the fault class. Consequently, if fault $s$ occurs, its estimate $G^r\hat{\theta}^r$ associated with FIE $r$ is determined by the structure of FIE $r$, which in turn is determined by fault $r$. Therefore, the fault mismatch function $h^{sr}(t)$, defined as the ability of FIE $r$ to match fault $s$, provides a measure of the difference between fault $s$ and fault $r$.

The following theorem characterizes the class of isolable faults:

**Theorem 4.** *Consider the fault isolation scheme described by* (26) *and* (35)*. Suppose that a fault $s$, $s = 1,\dots,N$, occurring at time $t = T_0$ is detected at time $t = T_d$. Then fault $s$ is isolable if, for each $r \in \{1,\dots,N\}\setminus\{s\}$, there exist some time $t^r > T_d$ and some $j \in \{1,\dots,p\}$, such that the fault mismatch function $h_j^{sr}(t^r)$ satisfies the following inequality:*

$$|h_j^{sr}(t)| \ge 2k_j\int_{T_d}^t e^{-\lambda_j(t-\tau)}\left[(\|\mathcal{A}_{21}\| + \gamma_2)\chi^r(\tau) + \bar{\eta}_2\right]\mathrm{d}\tau + k_j\int_{T_d}^t e^{-\lambda_j(t-\tau)}\left[\alpha^s\|\Omega^s\| + \alpha^r\|\Omega^r\|\right]\mathrm{d}\tau + |(\bar{C}_j\Omega^r)^\top|\kappa^r(t) + 2\omega_2k_je^{-\lambda_j(t-T_d)}. \quad (37)$$

**Proof.** See Appendix B. □

**Remark 5.** According to the above theorem, if, for each $r \in \{1,\dots,N\}\setminus\{s\}$, the fault mismatch function $h_j^{sr}(t^r)$ satisfies condition (37) for some time $t^r > 0$, then the $j$th component of the output estimation error associated with the $r$th isolation estimation would exceed its corresponding adaptive threshold at time $t = t^r$, i.e., $|\tilde{y}_j^r(t^r)| > \mu_j^r(t^r)$, hence excluding the occurrence of fault $r$. Therefore, Theorem 4 characterizes (in a non-closed form) the class of nonlinear faults that are isolable by the proposed robust FDI scheme.

## 6. Simulation results

Consider a single-link robotic arm with a revolute elastic joint rotating in a vertical plane whose motion equations are given by Raghavan and Hedrick (1994), Spong and Vidyasagar (1989)

$$J_l\ddot{q}_1 + k(q_1 - q_2) + mgh\sin q_1 = 0$$

$$J_m\ddot{q}_2 + F_m\dot{q}_2 - k(q_1 - q_2) = k_\tau u,$$

where $q_1$ and $q_2$ are the angular positions of the link and the motor, respectively. The link inertia $J_1 = 9.3\times10^{-3}$ kg m$^2$, motor inertia $J_m = 3.7\times10^{-3}$ kg m$^2$, torsional spring constant $k = 1.8\times10^{-1}$ Nm/rad, link mass $m = 2.1\times10^{-1}$ kg, link length $h = 0.15$ m, amplifier gain $k_\tau = 8\times10^{-2}$ Nm/V, viscous friction coefficients

$F_m = 4.6 \times 10^{-2}$ Nm/V, and the gravity constant $g = 9.8$. The control $u$ is the torque by the motor.

By choosing $x_1 = q_1, x_2 = \dot{q}_1, x_3 = q_2, x_4 = \dot{q}_2$, and assuming that the motor position, motor velocity, and the sum of link position and link velocity are measured (see, e.g. Yan and Edwards (2007)), a state space model of the system can be obtained. Additionally, by using a linear transformation of coordinates $z = [\, z_1^\top \;\; z_2^\top \,]^\top = Tx$ with $T = [-50\ 0\ 0\ 0;\ -0.1\ -0.1\ 0\ 0;\ 0\ 0\ 1\ 0;\ 0\ 0\ 0\ 1]$, the state space model in the new coordinate system is

$$\begin{bmatrix} \dot{z}_1 \\ \hline \dot{z}_2 \end{bmatrix} = \left[\begin{array}{c|ccc} -1 & 500 & 0 & 0 \\ \hline -0.04 & 1 & -1.94 & 0 \\ 0 & 0 & 0 & 1.00 \\ -0.97 & 0 & -48.65 & -12.43 \end{array}\right] \begin{bmatrix} z_1 \\ \hline z_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \hline -3.32\sin(z_1/50) \\ 0 \\ 21.62u \end{bmatrix} + \eta + \beta f(y,u)$$

$$y = \begin{bmatrix} -10 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} z_2.$$

Note that the effects of modeling uncertainty $\eta$ and faults $f$ have been included in the above model. Specifically, in this simulation example, the modeling uncertainty $\varphi$ is assumed to be up to 10% inaccuracy in the amplifier gain $k_\tau$, which gives $\eta = [\, 0 \;\; 0 \;\; 0 \;\; \eta_2 \,]^\top$ with $\bar{\eta}_2 = \frac{0.1k_\tau}{J_m}|u(t)|$. The following two types of faults are considered:

- *Actuator fault.* We consider a simple multiplicative actuator fault by letting $u = \bar{u} + \theta^1\bar{u}$, where $\bar{u}$ is the nominal control input in the non-fault case and $\theta^1 \in [\, -1 \;\; 0 \,]$ is the parameter characterizing the magnitude of the fault. Note that the case $\theta^1 = 0$ represents the normal operation condition (no fault), while $\theta^1 = -1$ corresponds to the complete failure of the actuator. Therefore, the actuator fault can be described by $f^1(y,u) \triangleq [\, 0 \;\; 0 \;\; 0 \;\; \theta^1 g^1(u) \,]^\top$, where $g^1(u) = k_\tau u/J_m$ and $\theta^1 \in [-1, 0]$.
- *A fault leading to extra abnormal friction in the motor.* Specifically, as a result of the fault, the viscous friction constant $F_m$ increases from 0.046 Nm/V up to 0.146 Nm/V. Then, the fault function is in the form of $f^2(y,u) \triangleq [\, 0 \;\; 0 \;\; 0 \;\; \theta^2 g^2(y) \,]^\top$, where $g^2(y) = 0.1y_3/J_m$, and $\theta^2 \in [-1, 0]$ represents the significance of extra friction.

The above model is clearly in the form of (3). The nonlinear term $-3.32\sin(z_1/50)$ has a global Lipschitz constant of 0.0664.

Based on the fault detection and isolation scheme described in Sections 3 and 4, a fault detection estimator and two fault isolation estimators are constructed. The initial conditions of the plant are chosen to be $x(0) = 0$, and the input to the system is given by $u(t) = 2\sin(2t)$. We set the observer gain matrix $L = [-0.2500\ -1.9355\ 0;\ 0\ 1.7000\ 1.0000;\ 0\ -48.6486\ -10.5324]$, so that the poles of matrix $\bar{A}_{22}$ are located at $-1.5$, $-1.7$, and $-1.9$, respectively. Consequently, the related constants are chosen to be $k_0 = k_1 = k_2 = k_3 = 1$, $\lambda_0 = 1$, $\lambda_1 = 1.2$, $\lambda_2 = 1.4$, and $\lambda_3 = 1.9$. The learning rate of the adaptive algorithm for fault parameter estimation in the FIEs is set to 1.

Figs. 1 and 2 show the simulation results when an actuator fault, with $\theta^1 = -0.4$, occurs at $T_0 = 5$ s. Specifically, the fault detection residual (solid and red line) and its threshold (dashed and blue line) associated with $y_3$ are shown in Fig. 1. As can be seen, the fault is detected almost immediately at approximately $T_d = 5.1$ s. Then, the two FIEs are activated to determine the particular fault type that has occurred. Selected fault isolation residuals and their corresponding thresholds generated by FIE #1 and FIE #2, respectively, are shown in Fig. 2. Note that, for FIE #2, only the residual and threshold associated with $y_3$ are shown, since this is sufficient to exclude the possibility of occurrence of $f^2$ for fault isolation. It can be seen that the residual associated with $y_3$ generated by FIE #2 exceeds its threshold (shown in the top left plot in Fig. 2), while all three residual components generated by the FIE #1 always remain below their thresholds (shown in the remaining plots of Fig. 2), thus indicating the occurrence of an actuator fault (i.e, fault 1). The fault is isolated at approximately 6.2 s. Analogously, Figs. 3 and 4 show the simulation results when a fault of type 2 with $\theta^2 = -0.5$, occurs at $T_0 = 5$ s. Again, the fault is successfully detected and isolated.

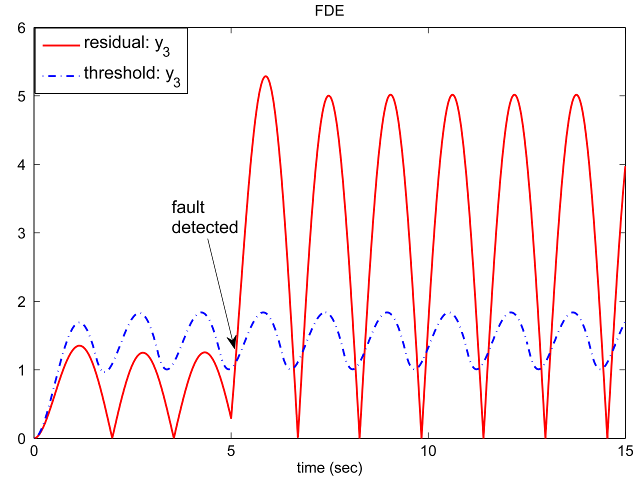


**Fig. 1.** The case of fault 1: Fault detection residual (solid line) and its threshold (dashed line) associated with $y_3$.

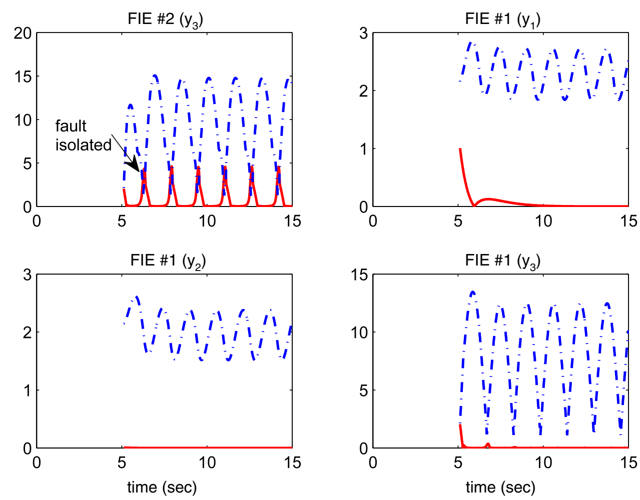


**Fig. 2.** The case of fault 1: Fault isolation residuals (solid line) and their thresholds (dashed line) generated by FIE 2 (only the pair associated with $y_3$ are shown) and FIE 1, respectively.

## 7. Concluding remarks

In this paper, a robust fault diagnosis scheme for a class of Lipschitz uncertain nonlinear system is presented. The robustness, fault detectability and isolability are enhanced via the appropriately designed adaptive thresholds in the diagnostic decision-making stage. Fault detectability and isolability conditions