

An Integrated Learning and Filtering Approach for Fault Diagnosis of a Class of Nonlinear Dynamical Systems

Christodoulos Keliris, *Member, IEEE*, Marios M. Polycarpou, *Fellow, IEEE*, and Thomas Parisini, *Fellow, IEEE*

Abstract—This paper develops an integrated filtering and adaptive approximation-based approach for fault diagnosis of process and sensor faults in a class of continuous-time nonlinear systems with modeling uncertainties and measurement noise. The proposed approach integrates learning with filtering techniques to derive tight detection thresholds, which is accomplished in two ways: 1) by learning the modeling uncertainty through adaptive approximation methods and 2) by using filtering for dampening measurement noise. Upon the detection of a fault, two estimation models, one for process and the other for sensor faults, are initiated in order to identify the type of fault. Each estimation model utilizes learning to estimate the potential fault that has occurred, and adaptive isolation thresholds for each estimation model are designed. The fault type is deduced based on an exclusion-based logic, and fault detectability and identification conditions are rigorously derived, characterizing quantitatively the class of faults that can be detected and identified by the proposed scheme. Finally, simulation results are used to demonstrate the effectiveness of the proposed approach.

Index Terms—Adaptive estimation, fault detection, fault diagnosis, learning systems.

I. INTRODUCTION

DUE to the increased complexity of systems in the modern era, there is a further need for methods that would ensure a robust, resilient, and reliable operation, in order to avoid potential failures that could impose a significant economic, social, and health damage. Especially, in today's environment where many large-scale systems are interconnected and interdependent, such as electric power systems, communication, and water networks, a potential failure could trigger a domino effect of failing systems. Therefore, prompt fault detection, isolation, and identification is a crucial feature that remains at the forefront of the technological evolution.

Manuscript received March 30, 2015; revised August 31, 2015; accepted November 1, 2015. This work was supported in part by the European Research Council (ERC) within the ERC Advanced Grant through the FAULT-ADAPTIVE Program and in part by the Engineering and Physical Sciences Research Council through the STABLE-NET Program under Grant EP/L014343/1.

C. Keliris and M. M. Polycarpou are with the KIOS Research Center for Intelligent Systems and Networks, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus (e-mail: keliris.chris@gmail.com; mpolyar@ucy.ac.cy).

T. Parisini is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K., and also with the Department of Engineering and Architecture, University of Trieste, Trieste 34127, Italy (e-mail: t.parisini@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2015.2504418

The importance of the fault detection and isolation (FDI) problem was acknowledged several decades back and is now well established with various approaches. Several survey papers [1]–[3] and books [4]–[7] detail the main methods that have been proposed over the years. The research conducted for fault diagnosis in linear systems is well established (the interested reader may refer to the aforementioned survey papers and books), whereas for nonlinear systems, the work is more limited and focused on specific classes of nonlinear systems by using various methods, such as observer methods [8]–[13], differential-geometric approaches [14], change-detection approaches [15], and learning methods [16]–[18]. These approaches were based on a centralized framework, whereas lately, due to advances in communications and distributed sensing, several fault diagnosis techniques for hierarchical, decentralized, and distributed systems have emerged [19]–[30]. In general, though, there are some issues that are usually overseen or dealt independently such as: 1) the presence of measurement noise; 2) the presence of nonlinear modeling uncertainties; and 3) the identification of the fault type (i.e., process or sensor fault). For instance, in the research literature, the modeling uncertainty has been addressed by considering either a known constant or adaptive threshold [17], [31], and measurement noise was treated similarly [26], [32]. A filtering approach to attenuate the effects of measurement noise in order to enhance fault detectability was recently presented in [27] and contained a rigorous detectability analysis. The aforementioned issues significantly affect the performance of fault diagnosis schemes, and therefore, the main objective of this paper is to address them collectively in a unified methodology by integrating learning and filtering methods.

Moreover, in the research literature, fault diagnosis approaches are typically developed targeting a specific fault type, for example, either a process or a sensor fault. However, this treatment may not operate satisfactory in real world applications, since each approach is only valid for the corresponding fault type. For instance, by considering fault diagnosis schemes only for the process faults, the degraded performance or erroneous measurements from faulty sensors are ignored and may be mistakenly considered as process faults, endangering system stability. Similarly, in fault diagnosis schemes that only consider the sensor faults, the crucial element of potential faults in the plant process is overlooked. As a result, dealing with only one fault type imposes the danger of false alarms due to monitoring the specific fault type and ignoring

the other, something that results in unnecessary component replacement and increased maintenance costs.

The research conducted on the fault diagnosis problem that deals with process and sensor faults simultaneously is limited [29], [33]–[38]. In [33], a framework for dealing with joint diagnosis of process and sensor faults is proposed by using a principal component analysis, whereas in [34], the sensor and process fault detection problem is addressed using multisensor data fusion techniques. In the context of analytical redundancy methods, Talebi *et al.* [37] propose a recurrent neural-network-based fault detection scheme for nonlinear systems, which employs two nonlinear-in-parameters neural networks to isolate actuator and sensor faults. In [35], a learning approach is proposed that is able to determine the type of the fault that has occurred (process or sensor) by considering that only a single fault occurs and by utilizing adaptive approximation methods to construct suitable fault isolation estimators. Moreover, the work in [36] follows along similar concepts of [35] for a more general class of systems. In the case of discrete-time systems, Thumati and Halligan [38] propose a nonlinear observer-based fault diagnostics scheme, dealing with process and sensor faults for nonlinear systems, which consists of an artificial immune system as an online approximator that identifies the fault type by monitoring the outputs' magnitude of two online approximators designed for estimating the system state and output. In [29], a distributed detection scheme for process and sensor faults for a special class of input–output interconnected systems under continuous-time is proposed by using nonlinear observer design that only deals with the fault detection issue.

This paper deals with the aforementioned challenges and proposes an integrated learning and filtering approach for a fault diagnosis approach of process and sensor faults in a class of continuous-time nonlinear systems with modeling uncertainty and in the presence of measurement noise. The primary objective and main contribution of this paper are the design of a unified fault diagnosis approach by: 1) integrating learning and filtering techniques for obtaining tight detection thresholds that guarantee no false alarms and, thus, enhancing fault detectability and 2) identifying the fault type (process or sensor) when a fault is detected and generating an estimation of the fault. The first objective is achieved in two ways: 1) adaptive approximation methods are used for learning the modeling uncertainty (so that the learned modeling uncertainty function is used in the design of the residual signals) and 2) by using filtering to attenuate the effect of measurement noise on the diagnosis thresholds. Adaptive approximation methods have been used in the area of fault diagnosis for learning the modeling uncertainties and the fault function for fault isolation purposes [16], [17], [39], [40]. Therefore, the novelty in this paper is that, both tasks, learning and filtering, are integrated in a unified framework and intertwined through the recent filtering approach in [27], which is decomposed in a two stage filtering process in order to derive the required signals for the adaptive approximation and for the residual derivation. Preliminary results of the first objective were presented in [41], which is now significantly extended to formulate the unified fault diagnosis approach for process

and sensor faults proposed in this paper. In order to achieve the second objective (fault type identification), two estimation models are constructed, one for process and one for sensor faults, which utilize adaptive approximation methods to learn the potential faults, and then, the fault type is determined based on an exclusion-based logic. The main differences of this paper with respect to the other papers [35], [36] that deal with the FDI problem for process and sensor faults are: 1) in the other works, the process faults were considered to belong to a class of known functions with unknown magnitudes, whereas in this work, we do not impose this restriction; instead, we exploit the sheer learning potential of the learning method to approximate any function that represents the process fault and 2) in the other works, only a single-sensor fault is considered and can be identified, whereas in this work, multiple sensor faults are allowed. Finally, fault detectability and identification conditions are derived for both process and sensor faults that provide an implicit characterization of the faults that can be detected and identified by the proposed scheme.

This paper is organized as follows. In Section II, the problem formulation is given. In Section III, the detailed design of the fault detection scheme by combining adaptive approximation with filtering is presented in detail. In Section IV, the detectability conditions for process and sensor faults that characterize the class of detectable faults are derived. The details of the fault identification procedure are given in Section V. In Section VI, the fault-type identification conditions are derived. In Section VII, a simulation example demonstrating the effectiveness of the scheme is presented. Finally, the conclusions are drawn in Section VIII.

II. PROBLEM FORMULATION

Consider a nonlinear dynamic system

$$\Sigma : \begin{cases} \dot{x}(t) = Ax(t) + f(x(t), u(t)) + \eta(x(t), u(t)) \\ \quad + \beta^x(t - T_0^x)\phi(x(t), u(t)) \\ y(t) = x(t) + \zeta(t) + \beta^y(t - T_0^y)\sigma(t) \end{cases} \quad (1)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $y \in \mathbb{R}^n$ are the state, input, and measured output vectors, respectively, the matrix $A \in \mathbb{R}^{n \times n}$ and the function $f : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ are the known (nominal) function dynamics, and $\eta : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ is the modeling uncertainty associated with the nominal function. The vector $\zeta \in \mathcal{D}_\zeta \subset \mathbb{R}^n$ (\mathcal{D}_ζ is a compact set) represents the unknown measurement noise. The term $\beta^x(t - T_0^x)\phi(x, u)$ characterizes the time-varying process fault function dynamics affecting the system. More specifically, the term $\phi : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ represents the unknown process fault function, and the term $\beta^x(t - T_0^x) : \mathbb{R} \mapsto \mathbb{R}^+$ models the time evolution of the process fault, where T_0^x is the unknown time of the process fault occurrence. The time profile $\beta^x(t - T_0^x)$ can be used to model both abrupt and incipient faults. Specifically, prior to the fault occurrence, the time profile is considered to be zero, i.e., $\beta^x(t - T_0^x) = 0$, for all $t < T_0^x$, and after the fault occurrence, it is monotonically increasing to one as $t \rightarrow \infty$ (in the case of incipient fault, since in the case of abrupt faults, the time profile takes the form of a step function). The term $\beta^y(t - T_0^y)\sigma(t)$ characterizes the time-varying magnitude of the sensor faults. More specifically, each component $\sigma^{(k)}(t)$,

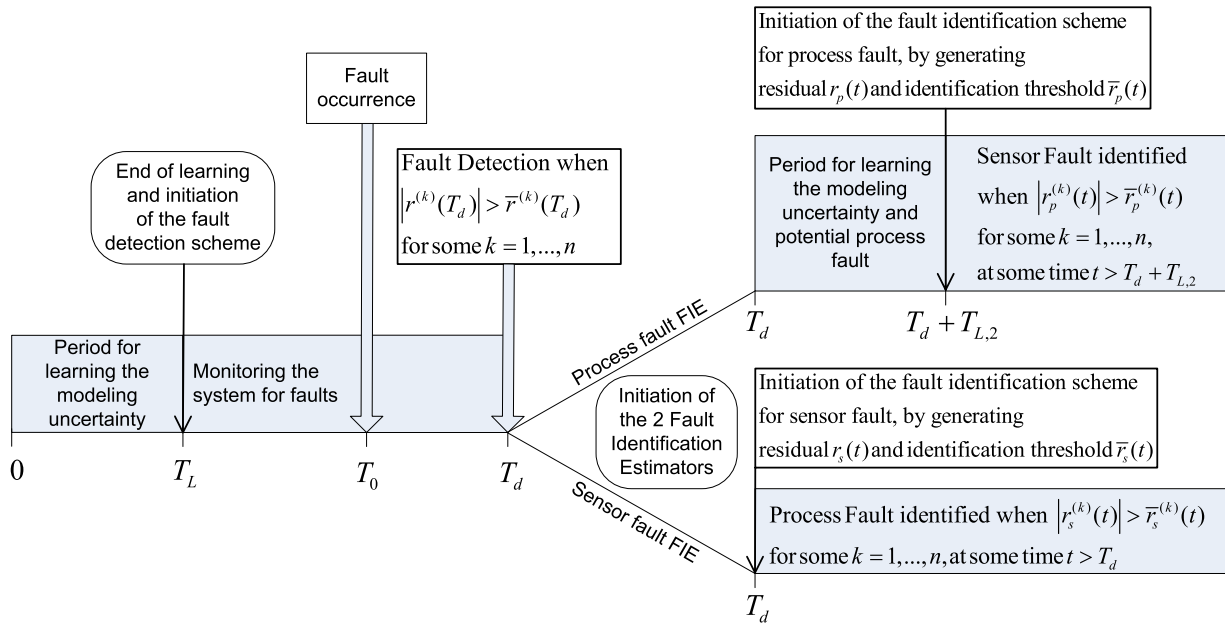


Fig. 1. Diagram describing the general procedure of the fault detection and identification scheme. The time T_L indicates the learning time allowed for learning the modeling uncertainty, T_0 indicates the time of the fault occurrence (process or sensor), T_d indicates the fault detection time, and $T_{L,2}$ indicates the learning time allowed for learning the combined effect of the modeling uncertainty with the potential process fault.

$k = 1, 2, \dots, n$, of the sensor fault vector $\sigma : \mathbb{R}^+ \mapsto \mathbb{R}^n$, represents the (possibly time varying) bias due to a sensor fault that occurs in the k th sensor at time T_0^y . In this paper, we only consider the case of abrupt sensor faults

$$\beta^y(t - T_0^y) = \begin{cases} 0, & \text{if } t < T_0^y \\ 1, & \text{if } t \geq T_0^y. \end{cases} \quad (3)$$

Please note that without loss of generality, in this paper, all the sensor faults are considered to occur at time T_0^y for simplicity. In addition, we consider the occurrence of a single type of fault: either a process fault occurring at time T_0^x or single/multiple sensor fault(s) occurring at time T_0^y .

The objective is twofold: 1) to exploit the filtering framework, recently developed in [27], in order to exploit the noise suppression properties of filtering and to integrate learning methods for approximating the modeling uncertainty, so that tight detection thresholds are obtained, thus enhancing fault detectability and 2) when a fault is detected, to determine whether it is a process or sensor fault and provide an estimation of the fault function (i.e., $\hat{\phi}, \hat{\sigma}$). The formulation is independent of the controller used, which may either be a generic one achieving some desired control objectives or one that incorporates an advanced fault accommodation scheme. The learning of the overall uncertainty function η is based on adaptive approximation methods and allows the use of the learned function dynamics (indicated by $\hat{\eta}$) in a suitable residual signal to cancel out the true function η and, hence, aid in the derivation of a tighter detection threshold.

After the detection of a fault, two fault identification estimators (FIEs) are initiated to identify the type of the fault that has occurred and provide an estimation of the fault. Each FIE corresponds to a fault type (i.e., process or sensor), and by utilizing adaptive approximation methods, the fault that has occurred is being learned, while, at the same time, residual

and isolation thresholds are derived that are used for the identification of the fault type. The fault type identification is based on an exclusion logic, in the sense that when a particular residual generated from an FIE exceeds its corresponding isolation threshold, then the particular fault type that the FIE corresponds to is excluded. Hence, it is guaranteed that the fault type is the other one, for which all the residuals of its FIE model remain below their corresponding isolation thresholds, and the fault estimation is provided by the latter FIE model. Fig. 1 shows the general procedure of the scheme developed in this paper.

In the sequel, $|\cdot|$ indicates the absolute value of a scalar function, the Euclidean 2-norm for vectors, and the matrix norm induced by the 2-norm for matrices. In addition, the notation $z_f(t) = H(s)[z(t)]$ denotes the output $z_f(t)$ of any signal $z(t)$, which is passed through a filter with a transfer function $H(s)$. In other words, $z_f(t)$ is the output of a linear system represented by the transfer function $H(s)$ with $z(t)$ as input. In terms of more rigorous notation, let $h(t)$ be the impulse response associated with $H(s)$, i.e., $h(t) \triangleq \mathcal{L}^{-1}[H(s)]$, where \mathcal{L}^{-1} is the inverse Laplace transform operator. Then, $z_f(t) = H(s)[z(t)] = \int_0^t h(\tau)z(t-\tau) d\tau$. Moreover, the notation $z^{(k)}$ denotes the k th component of the vector z or the k th row of z in case it is a matrix. The following assumptions are used throughout this paper.

Assumption 1: The state variable x and the local input u remain bounded in some compact region of interest $\mathcal{D} = \mathcal{D}_x \times \mathcal{D}_u \subset \mathbb{R}^n \times \mathbb{R}^m$, before and after the occurrence of a fault.

Assumption 2: The k th component of the function $f(x, u)$ satisfies the Lipschitz condition

$$|f^{(k)}(x_1, u) - f^{(k)}(x_2, u)| \leq \lambda_{f_k} |x_1 - x_2| \quad (4)$$

for $x_1, x_2 \in \mathcal{D}_x$, $u \in \mathcal{D}_u$, where λ_{f_k} is the known Lipschitz constant. Hence, for the vector function $f(x, u)$,

the Lipschitz condition becomes $|f(x_1, u) - f(x_2, u)| \leq \lambda_f |x_1 - x_2|$, where λ_f is the Lipschitz constant that can be considered as $\lambda_f \triangleq \sum_{k=1}^n \lambda_{f_k}$.

Assumption 3: The rate of change of the sensor bias $\sigma^{(k)}(t)$ is uniformly bounded as follows:

$$|\dot{\sigma}^{(k)}(t)| \leq \psi, \quad k = 1, \dots, n \quad (5)$$

where ψ is a known positive scalar.

Assumption 4: The measurement noise is bounded, i.e., $|\zeta^{(k)}(t)| \leq \bar{\zeta}^{(k)}$, where $\bar{\zeta}^{(k)}$ is a known constant. Hence, $|\zeta(t)| \leq \bar{\zeta}_d \triangleq |\bar{\zeta}|$.

Assumption 1 is required for well-posedness, since, in this paper, we address the fault detection problem, not the control design and fault accommodation problem. Assumptions 2 and 3 are required for the design of the FIE model for the sensor faults. Assumption 4 is required in order to distinguish the effects of sensor faults.

In Section III, the details of the proposed scheme for the tasks of learning the modeling uncertainty and designing suitable fault detection thresholds are given.

III. FAULT DETECTION

First, we derive suitable detection thresholds by computing uncertainty bounds associated with the fault-free operation of the system.

A. Filtering

Each measured variable $y^{(k)}(t)$ is filtered by an induced filter $H(s)$ that is implemented as a series of two filters $H_1(s)$ and $H_2(s)$, for reasons that will become apparent in the sequel, such that $H(s) = H_1(s)H_2(s)$. The filter $H(s)$ is of p th order with a strictly proper transfer function $H(s) = sH_p(s)$, where

$$H_p(s) = \frac{d_{p-2}s^{p-2} + d_{p-3}s^{p-3} + \dots + d_0}{s^p + c_{p-1}s^{p-1} + \dots + c_1s + c_0}. \quad (6)$$

At first, the selection of the filter $H(s)$ is made, so that the effect of the measurement noise is dampened. Then, the filters $H_1(s)$ and $H_2(s)$ are obtained by introducing the design constant $\alpha > 0$ (which is required for the learning task) and are given by

$$H_1(s) = \frac{s}{s + \alpha} \quad (7)$$

$$H_2(s) = \frac{(s + \alpha)(d_{p-2}s^{p-2} + d_{p-3}s^{p-3} + \dots + d_0)}{s^p + c_{p-1}s^{p-1} + \dots + c_1s + c_0}. \quad (8)$$

In the following, the initial conditions of the filters $H_1(s)$ and $H_2(s)$ are considered to be zero.

It must be stressed that, although four filters have been introduced above (namely, $H(s)$, $H_p(s)$, $H_1(s)$, $H_2(s)$), only $H(s)$ is required to be selected, so that the noise attenuation is achieved, and then, the other three filters can be easily deduced from the particular selection of $H(s)$.

The filters $H_1(s)$ and $H_2(s)$ [and hence $H(s)$ and $H_p(s)$] are asymptotically stable and, therefore, BIBO stable. Hence, for bounded measurement noise $\zeta(t)$, the filtered measurement noise $\epsilon_\zeta(t) \triangleq H(s)[\zeta(t)]$ is bounded as follows:

$$|\epsilon_\zeta^{(k)}(t)| \leq \bar{\epsilon}_\zeta^{(k)}(t) \quad k = 1, 2, \dots, n$$

where $\bar{\epsilon}_\zeta^{(k)}$ are bounding functions that are computable since $\zeta \in \mathcal{D}_\zeta$.

In general, each measured variable $y^{(k)}(t)$ can be filtered by a different induced filter $H(s)$ with different design constants α . In this paper, without loss of generality, we consider $H(s)$ and α to be the same for all the output variables in order to simplify the notation and presentation.

The two-step filtering process using filters $H_1(s)$ and $H_2(s)$ is employed, instead of directly filtering each measurement with $H(s)$, in order to integrate the filtering with the learning task and be able to derive the required signals for both tasks (approximation error for learning and residual for fault detection). It must be noted that adaptive approximation methods require that the filters used are strictly positive real (SPR) [42], [43]. Therefore, the filter $H(s)$, which is selected for dampening the measurement noise, may not be SPR and, hence, not suitable for the learning task. With the decomposition of $H(s)$ into the two filters $H_1(s)$ and $H_2(s)$, the output of the filter $H_1(s)$ (which is SPR) can be used for the learning task, and the output of the filter $H_2(s)$ can be used for the fault detection task. Therefore, the decoupling of the two tasks into two separate and independent ones is achieved.

The choice of a particular type of filter to be used is application dependent, and it is made according to the available *a priori* knowledge on the noise properties. Usually, measurement noise is constituted by high-frequency components, and therefore, the use of low-pass filter for dampening noise is well justified. On other occasions, one may want to focus the fault detectability on a prescribed frequency band of the measurement signals and, hence, choose the filter accordingly. The particular selection criteria for choosing a suitable filter and its tradeoffs are out of the scope of this paper, and the reader is referred to the continuous-time case in our recent work [27], where a rigorous investigation of the filtering impact (according to the poles location and filters order) on the detection time is presented.

B. Adaptive Approximation

An adaptive approximator is used for learning the modeling uncertainty η and providing an estimate $\hat{\eta}$. To simplify the notation in the following analysis, (1) is rewritten as:

$$\begin{aligned} \dot{x}(t) &= g(x(t), u(t)) + \eta(x(t), u(t)) \\ &\quad + \beta^x(t - T_0^x)\phi(x(t), u(t)) \end{aligned} \quad (9)$$

where $g(x(t), u(t)) \triangleq Ax(t) + f(x(t), u(t))$.

Based on (9), an estimation model $\hat{x}(t)$ for $x(t)$ under fault-free operation is generated as follows:

$$\dot{\hat{x}}(t) = g(y(t), u(t)) + \hat{\eta}(y(t), u(t), \hat{\theta}(t)) \quad (10)$$

with initial condition $\hat{x}(0) = y(0)$, $\hat{\eta}$ denotes the output of an adaptive approximator structure and $\hat{\theta} \in \mathbb{R}^q$ is a set of adjustable parameters in vector form. More details regarding the adaptive approximator design will be given in the sequel.

The signal, which is used in the adaptive law for adjusting the parameter vector $\hat{\theta}$, is given by

$$\epsilon(t) \triangleq H_1(s)[y(t) - \hat{x}(t)] \quad (11)$$

and by using (2), (7), and the Laplace differentiation property $s[x(t)] = \dot{x}(t) + x(0)\delta(t)$ (where $\delta(t)$ is the Dirac delta function), we obtain

$$\begin{aligned}\epsilon(t) &= \frac{s}{s+\alpha}[x(t) + \zeta(t) - \hat{x}(t)] \\ &= \frac{1}{s+\alpha}[\dot{x}(t) + x(0)\delta(t)] + \frac{s}{s+\alpha}[\zeta(t)] \\ &\quad - \frac{1}{s+\alpha}[\dot{\hat{x}}(t) + \hat{x}(0)\delta(t)].\end{aligned}\quad (12)$$

By using (1) and (10), (12) becomes

$$\begin{aligned}\epsilon(t) &= \frac{1}{s+\alpha}[\Delta g(t) + \Delta \eta(t) - \alpha \zeta(t)] \\ &\quad - e^{-\alpha t} \zeta(0) + \zeta(t)\end{aligned}\quad (13)$$

where

$$\Delta g(t) \triangleq g(x(t), u(t)) - g(y(t), u(t)) \quad (14)$$

$$\Delta \eta(t) \triangleq \eta(x(t), u(t)) - \hat{\eta}(y(t), u(t), \hat{\theta}(t)). \quad (15)$$

As it can be seen from (13), $\epsilon(t)$ is comprised of the functional error $\Delta \eta(t)$ and some additional terms, which arise due to the measurement noise $\zeta(t)$. Therefore, the signal $\epsilon(t)$ provides a measure of $\Delta \eta(t)$, and by an appropriate selection of α , the scaled quantity $\alpha \epsilon(t)$ (which is measurable) provides a good approximation of $\Delta \eta(t)$ (which is unknown) over the lower frequency range (depending on the pole of the filter $H_1(s)$). Hence, $\epsilon(t)$ can be used in the adaptive law for adjusting the parameter vector $\hat{\theta}$ of the approximator [16].

In this paper, we use a linearly parameterized approximator $\hat{\eta}(y, u, \hat{\theta})$, so that each component $k = 1, 2, \dots, n$ is given by $\hat{\eta}^{(k)}(y, u, \hat{\theta}_k) = \omega_k^\top(y, u) \hat{\theta}_k$, where $\omega_k : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^{q_k}$ is a vector composed of smooth functions independent of $\hat{\theta}_k$, and $\hat{\theta}_k \in \mathbb{R}^{q_k}$ is a set of adjustable parameters in vector form. In compact form, we have

$$\hat{\eta}(y, u, \hat{\theta}) = \Omega(y, u) \hat{\theta} \quad (16)$$

where $\Omega : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^{n \times q}$ is the diagonal block matrix $\Omega(y, u) = \omega_1^\top(y, u) \oplus \omega_2^\top(y, u) \oplus \dots \oplus \omega_n^\top(y, u)$, where \oplus indicates the direct sum operator and $\hat{\theta} \in \mathbb{R}^q$ ($q = \sum_{k=1}^n q_k$) is the column vector $\hat{\theta} = [\hat{\theta}_1^\top, \dots, \hat{\theta}_n^\top]^\top$.

Based on (13) and by using techniques from adaptive control (Lyapunov synthesis approach [42]), $\hat{\theta}$ is updated according to the following law:

$$\dot{\hat{\theta}}(t) = \mathcal{P}_p(\Gamma^x \Omega^\top(t) \epsilon(t)) \quad (17)$$

where $\Gamma^x \in \mathbb{R}^{q \times q}$ is a symmetric and positive definite learning rate matrix, and \mathcal{P}_p is a projection operator that restricts $\hat{\theta}$ in a predefined compact and convex set $\hat{\Theta}_p \in \mathcal{R}^q$.

The initial weight vector is chosen as $\hat{\theta}(0) = 0$, so that $\hat{\eta}(y, u, \hat{\theta}(0)) = 0$ that corresponds to the case in which the dynamics of the estimator are only described in terms of the known dynamics f .

Remark 1: The signal $\epsilon(t)$ is used in the adaptive law for adjusting the parameter vector $\hat{\theta}$, since it provides a measure of the uncertainty $\Delta \eta(t)$. It is important to note that, from (13), a suitable bound for $|\epsilon(t)|$ can be derived, and hence, $\epsilon(t)$ can also be used as a residual signal for the task of fault detection.

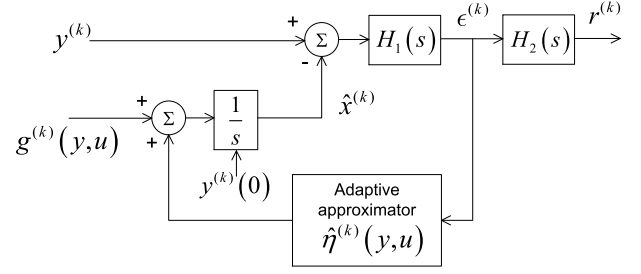


Fig. 2. Residual generation.

In this case, though $\epsilon(t)$ is generated by using the filter $H_1(s)$, which, although being suitable for the learning task (due to the SPR property), it has a proper transfer function that allows the noise $\zeta(t)$ to pass through unaffected. As a result, the bound on $\epsilon(t)$ may be too conservative for practical purposes, and hence, it may not be exceeded, thus leading to missed faults. Therefore, the fault detection scheme described in the sequel relies on a general strictly proper transfer function $H(s)$ (through the series of $H_1(s)$ and $H_2(s)$) that possesses better noise dampening characteristics and allows the derivation of tight detection thresholds.

C. Residual Generation

The residual signal $r(t)$ to be used for fault detection in this paper is given by

$$r(t) \triangleq H_2(s)[\epsilon(t)] \quad (18)$$

where $H_2(s)$ is given by (8) and $\epsilon(t)$ is given by (11).

By using (11), the residual (18) becomes

$$\begin{aligned}r(t) &= H(s)[y(t) - \hat{x}(t)] \\ &= sH_p(s)[x(t) - \hat{x}(t)] + H(s)[\zeta(t)] \\ &\quad + H(s)[\beta^y(t - T_0^y)\sigma(t)] \\ &= H_p(s)[\dot{x}(t) - \dot{\hat{x}}(t) + (x(0) - \hat{x}(0))\delta(t)] + \epsilon_\zeta(t) \\ &\quad + H(s)[\beta^y(t - T_0^y)\sigma(t)] \\ &= H_p(s)[\Delta g(t) + \Delta \eta(t) + \beta^x(t - T_0^x)\phi(x(t), u(t))] \\ &\quad + H(s)[\beta^y(t - T_0^y)\sigma(t)] - h_p(t)\zeta(0) + \epsilon_\zeta(t)\end{aligned}\quad (19)$$

where $h_p(t)$ is the impulse response associated with $H_p(s)$.

Under fault-free operation, the residual (19) is written as

$$r(t) = H_p(s)[\Delta g(t) + \Delta \eta(t)] - h_p(t)\zeta(0) + \epsilon_\zeta(t). \quad (20)$$

Fig. 2 shows the residual generation procedure according to (7), (8), (10), (11), and (18).

D. Detection Threshold

In this section, the derivation of suitable detection threshold signals that guarantee no-false alarms is presented. In order to exploit the benefits from the learning process, it is considered that the learning period is conducted during the time interval $[0, T_L]$ in which it is assumed that no faults occur. In other words, at time T_L , the update of the parameter vector $\hat{\theta}(t)$ stops, and the fault detection scheme is initiated. The training time T_L is selected by the designer and should be sufficiently large to allow learning of the modeling

uncertainty η . The detection decision of a fault in the system is made when $|r^{(k)}(t)| > \bar{r}^{(k)}(t)$ at some time $t > T_L$, for at least one component $k = 1, 2, \dots, n$, where $\bar{r}^{(k)}(t)$ is the detection threshold, to be designed in the sequel.

For the derivation of the detection threshold, we consider that after time T_L , the functional discrepancy between the unknown function η and its approximator $\hat{\eta}$ satisfies the following assumption.

Assumption 5: After the end of the learning phase at time T_L , the error between the modeling uncertainty η and the adaptive approximator $\hat{\eta}$ is bounded as follows:

$$|\eta^{(k)}(x(t), u(t)) - \hat{\eta}_L^{(k)}(y(t), u(t))| \leq \bar{\eta}_L^{(k)}(y(t), u(t))$$

for all $k = 1, 2, \dots, n$, $t > T_L$ and for all $(x, u) \in \mathcal{D}$, where $\bar{\eta}_L^{(k)}$ is a known bounding function and $\hat{\eta}_L^{(k)}(y, u)$ is the functional approximator using the parameter vector obtained at the end of the learning phase, i.e., $\hat{\eta}_L(y, u) \triangleq \Omega(y, u)\hat{\theta}(T_L)$.

The bound $\bar{\eta}_L^{(k)}$ can be computed by applying a set membership technique, as indicated in [40] and the references therein. In addition, the bound $\bar{\eta}_L^{(k)}$ can take a constant value for a given compact region according to the universal approximation property of the RBF networks [43]. Specifically, consider the case of the linearly parameterized network consisting by radial basis functions (RBFs) described by $f_{nn}(Z) = W^T S(Z) = \sum_{i=1}^Q w_i s_i(Z)$, where $Z \in \Omega_Z \subset \mathcal{Q}^q$ is the input vector, $W = [w_1, \dots, w_Q]^T$ is the weight vector, Q is the number of nodes of the network, and $S(Z) = [s_1(Z), \dots, s_Q(Z)]^T$ is the vector of RBFs. It has been shown in [44] and [45] that for any continuous function $f(Z) : \Omega_Z \rightarrow \mathcal{R}$, where $\Omega_Z \subset \mathcal{R}^q$ is a compact set, and for the linear in parameters RBF network $W^T S(Z)$ with a sufficiently large node number Q , then there exists an ideal constant weight vector W^* , such that the approximation error $\epsilon(Z) \triangleq f(Z) - (W^*)^T S(Z)$ satisfies $|\epsilon(Z)| \leq \epsilon^*$ for all $Z \in \Omega_Z$. Hence, by selecting a sufficiently large set of basis functions, the approximation error can be made arbitrarily small. The main issue, though, is selecting an appropriate time T_L that may pose a challenge for learning the modeling uncertainty. However, the key point for fault detection enhancement is achieved as long as the bound $\bar{\eta}_L^{(k)}(y, u)$ (with the use of learning) is smaller than the bound on $|\eta^{(k)}(x, u)|$ that would be used instead if no learning was used.

Therefore, $\Delta\eta(t)$ from (15), for $t > T_L$, becomes $\Delta\eta(t) = \Delta\eta_L(t)$, where

$$\Delta\eta_L(t) \triangleq \eta(x(t), u(t)) - \hat{\eta}_L(y(t), u(t)) \quad (21)$$

and therefore, the k th component of the residual (20) becomes

$$r^{(k)}(t) = H_p(s)[\Delta\eta_L^{(k)}(t)] + \epsilon_{\Delta g}^{(k)}(t) - h_p(t)\zeta^{(k)}(0) + \epsilon_{\zeta}^{(k)}(t) \quad (22)$$

where $\epsilon_{\Delta g}^{(k)}(t) \triangleq H_p(s)[g^{(k)}(x(t), u(t)) - g^{(k)}(x(t) + \zeta(t), u(t))]$ is the filtered form of the function discrepancy term that arises due to the measurement noise. Note that $\epsilon_{\Delta g}^{(k)}(t) = H_p(s)[\Delta g^{(k)}(t)]$ under sensor fault-free operation [due to the use of measurements in $\Delta g(t)$ given by (14)]. For the derivation of the detection threshold, we make the following assumption.

Assumption 6: The filtered function mismatch term $\epsilon_{\Delta g}^{(k)}(t)$ (under sensor fault-free operation) is bounded by a computable positive function $\bar{\epsilon}_{\Delta g}^{(k)}(t)$, i.e., for all $t > 0$

$$|\epsilon_{\Delta g}^{(k)}(t)| \leq \bar{\epsilon}_{\Delta g}^{(k)}(t), \quad k = 1, 2, \dots, n.$$

Assumption 6 is based on the observation that filtering dampens the error effect of measurement noise present in the function mismatch term $\Delta g(t)$. A suitable selection of the bound $\bar{\epsilon}_{\Delta g}^{(k)}$ can be derived through the use of simulations (i.e., Monte Carlo) by filtering the function mismatch term using the nominal function and the available noise characteristics (note that the measurement noise is assumed to take values in a compact set). Alternatively, one may exploit the Lipschitz property of the function $f^{(k)}(\cdot)$ and derive a suitable bound as $\bar{\epsilon}_{\Delta g}^{(k)} = (|A^{(k)}| + \lambda_{f_k})\bar{H}_p(s)[\bar{\zeta}_d]$, where $A^{(k)}$ indicates the k th row of A and $\bar{H}_p(s)$ is a transfer function with an impulse response $\bar{h}_p(t)$, such that $|h_p(t)| \leq \bar{h}_p(t)$ for all $t > 0$. However, this bound is more conservative, since the filtering benefits are not exploited.

By using the triangle inequality, the residual (22) satisfies

$$|r^{(k)}(t)| \leq |H_p(s)[\Delta\eta^{(k)}(t)]| + |\epsilon_{\Delta g}^{(k)}(t)| + |h_p(t)\zeta^{(k)}(0)| + |\epsilon_{\zeta}^{(k)}(t)|.$$

Therefore, by using the well-known result that the impulse response $h_p(t)$ of a strictly proper and asymptotically stable transfer function $H_p(s)$ decays exponentially, i.e., $|h_p(t)| \leq \nu e^{-\zeta t}$ for some $\nu > 0$, $\zeta > 0$, for all $t > 0$ [46], and by using Assumptions 5 and 6, the detection threshold for $t > T_L$ is given by

$$\bar{r}^{(k)}(t) \triangleq \bar{H}_p(s)[\bar{\eta}_L^{(k)}(y(t), u(t))] + \bar{\epsilon}_{\Delta g}^{(k)}(t) + |h_p(t)|\bar{\zeta}^{(k)} + \bar{\epsilon}_{\zeta}^{(k)}(t) \quad (23)$$

where $\bar{H}_p(s)$ is a filter with transfer function $\bar{H}_p(s) = \nu/(s + \zeta)$. Note that, even a conservative bound $\bar{\zeta}^{(k)}$ only affects the threshold during the initial transient, since $h_p(t)$ is exponentially decaying.

Remark 2: In earlier works [17], [27], [39], it was considered that the modeling uncertainty $\eta^{(k)}$ was bounded by an appropriate bounding function $\bar{\eta}^{(k)}$, which could lead to conservative detection thresholds. In this paper, we exploit the learned modeling uncertainty to derive less conservative thresholds through $\bar{\eta}_L^{(k)}$, which can be smaller than $\bar{\eta}^{(k)}$, in addition to noise attenuation through filtering and, therefore, enhance fault detectability. For instance, in this works' setting, if we did not utilize learning or filtering, but instead considered that the modeling uncertainty satisfies $|\eta^{(k)}(x, u)| \leq \bar{\eta}^{(k)}(y, u)$ for all $k = 1, \dots, n$, then the estimator would simply be $\hat{x} = g(y, u)$, the residual would directly be given by $\epsilon(t) = H_1(s)[y(t) - \hat{x}(t)]$, and a suitable threshold in this case would be $\bar{\epsilon}^{(k)}(t) = 1/(s + \alpha)[\bar{\Delta g}^{(k)}(t) + \bar{\eta}^{(k)}(y(t), u(t)) + \alpha\bar{\zeta}^{(k)}(t)] + e^{-\alpha t}\bar{\zeta}^{(k)}(t) + \bar{\zeta}^{(k)}(t)$, where $\bar{\Delta g}^{(k)}(t)$ is a bound on the mismatch function $\Delta g^{(k)}(t)$. Therefore, the threshold in this case may be more conservative due to the bounds used (on the modeling uncertainty, the noise, and the mismatch function $\Delta g^{(k)}(t)$), leading to missed detections. This is further illustrated in the simulation results (Section VII).

Remark 3: Qualitatively, if no filtering is used, then in order to guarantee no-false alarms, bounds on the noise magnitude and mismatch functions must be used, which can be rather conservative leading to conservative detection thresholds. By using filtering, the effect of noise is attenuated, allowing to set tighter bounds on the filtered noise and the filtered mismatch functions and, thus, leading to tighter detection thresholds. Even though the residual signal is lowered due to the use of filtering, the detection threshold in this case is tighter—relatively to the residual—allowing the easier detection of a fault in comparison with the case in which no filtering is used. Of course, the filter selection plays a crucial role. Further details regarding the performance enhancement of the filtering scheme are given in our earlier work [27]. Note that, if the fault that has occurred (process or sensor) is not sufficiently large, so that the residual $r(t)$ crosses the detection threshold $\bar{r}(t)$, then the fault remains undetected (missed alarm). In general, this is not uncommon in fault diagnosis schemes, because it is difficult to strike a balance between conservative detection thresholds and avoiding the presence of false alarms. This is a key reason that we propose the integrated learning and filtering approach in order to achieve tight detection thresholds that guarantee no-false alarms.

IV. FAULT DETECTABILITY ANALYSIS

The design and analysis of the fault detection scheme in Sections II and III were based on the derivation of suitable thresholds $\bar{r}^{(k)}(t)$, such that in the absence of any fault, the residual signals $r^{(k)}(t)$ are bounded by $\bar{r}^{(k)}(t)$ for $t > T_L$. In this section, fault detectability conditions for the aforementioned fault detection scheme are derived. The fault detectability analysis provides an implicit characterization, in nonclosed form, of the class of faults detectable by the proposed scheme.

Theorem 1: Consider the nonlinear system (1) and (2) with the adaptive approximation structure described in (7), (10), (11), (16), and (17), and the fault detection scheme described in (8), (18), and (23).

- 1) *Process Fault Detectability:* A process fault occurring at time $T_0^x > T_L$ is detectable at time $T_d > T_0^x$ if the following condition is satisfied for some $k = 1, 2, \dots, n$:

$$|H_p(s)[\beta^x(T_d - T_0^x)\phi^{(k)}(x(T_d), u(T_d))]| > 2\bar{r}^{(k)}(T_d). \quad (24)$$

- 2) *Sensor Fault Detectability:* A sensor fault occurring at time $T_0^y > T_L$ is detectable at time $T_d > T_0^y$ if the following condition is satisfied for some $k = 1, 2, \dots, n$:

$$|H(s)[\sigma^{(k)}(T_d)] + H_p(s)[g^{(k)}(x(T_d) + \zeta(T_d), u(T_d)) - g^{(k)}(x(T_d) + \zeta(T_d) + \sigma(T_d), u(T_d))]| > 2\bar{r}^{(k)}(T_d). \quad (25)$$

Proof:

- 1) In the presence of a process fault that occurs at some time $T_0^x > T_L$, the residual from (19) for $t > T_0^x$ becomes

$$r(t) = H_p(s)[\Delta g(t) + \Delta \eta_L(t) + \beta^x(t - T_0^x)\phi(x(t), u(t))] - h_p(t)\zeta(0) + \epsilon_\zeta(t). \quad (26)$$

By using the triangle inequality, the k th component of (26) satisfies

$$\begin{aligned} |r^{(k)}(t)| &\geq -|H_p(s)[\Delta \eta_L^{(k)}(t)]| - |\epsilon_{\Delta g}^{(k)}(t)| - |h_p(t)||\zeta^{(k)}(0)| \\ &\quad - |\epsilon_\zeta^{(k)}(t)| + |H_p(s)[\beta^x(t - T_0^x)\phi^{(k)}(x(t), u(t))]| \\ &\geq -\bar{H}_p(s)[\bar{\eta}_L^{(k)}(y(t), u(t))] - \bar{\epsilon}_{\Delta g}^{(k)}(t) - |h_p(t)||\zeta^{(k)}(0)| \\ &\quad - \bar{\epsilon}_\zeta^{(k)}(t) + |H_p(s)[\beta^x(t - T_0^x)\phi^{(k)}(x(t), u(t))]| \\ &\geq -\bar{r}^{(k)}(t) + |H_p(s)[\beta^x(t - T_0^x)\phi^{(k)}(x(t), u(t))]|. \end{aligned}$$

Therefore, for fault detection, the inequality $|r^{(k)}(t)| > \bar{r}^{(k)}(t)$ must hold at some time $t = T_d$ for some $k = 1, 2, \dots, n$, so the final fault detectability condition in (24) is obtained.

- 2) In the presence of a sensor fault that occurs at some time $T_0^y > T_L$, the residual from (19) by using (14), and by adding and subtracting $g(x(t) + \zeta(t), u(t))$ for $t > T_0^y$ becomes

$$\begin{aligned} r(t) &= H_p(s)[g(x(t), u(t)) - g(x(t) + \zeta(t), u(t))] \\ &\quad + g(x(t) + \zeta(t), u(t)) \\ &\quad - g(x(t) + \zeta(t) + \sigma(t), u(t)) \\ &\quad + H_p(s)[\Delta \eta_L(t)] - h_p(t)\zeta(0) + \epsilon_\zeta(t) + H(s)[\sigma(t)]. \end{aligned} \quad (27)$$

By using the triangle inequality and Assumptions 5 and 6, the k th component of (27) satisfies

$$\begin{aligned} |r^{(k)}(t)| &\geq |H(s)[\sigma^{(k)}(t)] + H_p(s)[g^{(k)}(x(t) + \zeta(t), u(t)) \\ &\quad - g^{(k)}(x(t) + \zeta(t) + \sigma(t), u(t))]| \\ &\quad - |\epsilon_{\Delta g}^{(k)}(t)| - |H_p(s)[\Delta \eta_L^{(k)}(t)]| - |h_p(t)||\zeta^{(k)}(0)| - |\epsilon_\zeta^{(k)}(t)| \\ &\geq |H(s)[\sigma^{(k)}(t)] + H_p(s)[g^{(k)}(x(t) + \zeta(t), u(t)) \\ &\quad - g^{(k)}(x(t) + \zeta(t) + \sigma(t), u(t))]| \\ &\quad - \bar{\epsilon}_{\Delta g}^{(k)}(t) - \bar{H}_p(s)[\bar{\eta}_L^{(k)}(y(t), u(t))] - |h_p(t)||\zeta^{(k)}(0)| - \bar{\epsilon}_\zeta^{(k)}(t) \\ &\geq |H(s)[\sigma^{(k)}(t)] + H_p(s)[g^{(k)}(x(t) + \zeta(t), u(t)) \\ &\quad - g^{(k)}(x(t) + \zeta(t) + \sigma(t), u(t))]| \\ &\quad - \bar{r}^{(k)}(t). \end{aligned}$$

Therefore, for fault detection, the inequality $|r^{(k)}(t)| > \bar{r}^{(k)}(t)$ must hold at some time $t = T_d$ for some $k = 1, 2, \dots, n$, so the final fault detectability condition in (25) is obtained. ■

The above theorem provides sufficient conditions for the implicit characterization of a class of faults that can be detected by the proposed fault detection scheme. As seen from the detectability conditions, the detection of the fault depends on the filtered fault functions. This demonstrates the crucial importance of the filter selection, which allows the derivation of tight detection thresholds. As a result, some filter selections may lead to better fault detection performance than others. A rigorous investigation of the filtering impact on the detection time is presented in [27]. In addition, note that the detectability conditions, which are obtained under worst case conditions,

are only sufficient (but not necessary), and hence, the classes of the detectable faults may be significantly larger.

V. FAULT TYPE IDENTIFICATION

In this section, we proceed to the task of identification of the fault type and the estimation of the magnitude of the fault that has occurred. When a fault is detected at time T_d , we initiate two distinct estimation models in order to identify the potential fault that has occurred. The first estimation model considers that a process fault has occurred and relies on the same logic for learning the modeling uncertainty by resuming learning to allow learning of the process fault as well. The second estimation model considers the occurrence of a sensor fault and makes use of learning to estimate adaptively the potential sensor fault(s) magnitude that occurred. The fault type identification methodology relies on an exclusion-based logic, i.e., suitable threshold signals in both estimation models are obtained, and when they are exceeded, the particular fault type is excluded, and hence, the occurrence of the other fault type is guaranteed. Furthermore, the use of learning in both estimation models provides an estimation of the fault that has occurred, once the fault type identification has been concluded.

A. Process Fault

In this section, we consider the case of a process fault. When a fault is detected at time T_d , we use the same logic that was used to learn the modeling uncertainty, with the only difference that this time the task is to learn the combined effect of the modeling uncertainty coupled with the potential process fault function. Essentially, we use the same estimation model as before indicated by (10), (11), (18), and resume learning. For clarity, we rewrite the equations as follows:

$$\dot{\hat{x}}_p(t) = g(y(t), u(t)) + \hat{\eta}(y(t), u(t), \hat{\theta}(t)) \quad (28)$$

$$\epsilon_p(t) \triangleq H_1(s)[y(t) - \hat{x}_p(t)] \quad (29)$$

$$r_p(t) \triangleq H_2(s)[\epsilon_p(t)] \quad (30)$$

with initial conditions $\hat{x}_p(T_d) = y(T_d)$ and $\hat{\theta}(T_d) = \hat{\theta}(T_L)$ (resume learning). As before, we allow a training time $T_{L,2}$ to learn the combined effect of the modeling uncertainty and the potential process fault, i.e., the training is conducted during the interval $[T_d, T_d + T_{L,2}]$, and make the following assumption, similar to Assumption 5.

Assumption 7: In the case of a process fault, after the end of the learning phase at time $T_d + T_{L,2}$, the error between the combined modeling uncertainty and fault function $\eta(\cdot) + \beta^x(\cdot)\phi(\cdot)$ and the adaptive approximator $\hat{\eta}$, i.e., $\Delta\eta_{L,2}(t) \triangleq \eta(x(t), u(t)) + \beta^x(t - T_0^x)\phi(x(t), u(t)) - \hat{\eta}_{L,2}(y(t), u(t))$, is bounded as follows:

$$|\Delta\eta_{L,2}^{(k)}(t)| \leq \bar{\eta}_{L,2}^{(k)}(y(t), u(t))$$

for all $k = 1, 2, \dots, n$, $t > T_d + T_{L,2}$ and for all $(x, u) \in \mathcal{D}$, where $\bar{\eta}_{L,2}^{(k)}$ is a known bounding function and $\hat{\eta}_{L,2}^{(k)}(y, u)$ is the functional approximator using the parameter vector obtained at the end of the second stage of the learning phase, i.e., $\hat{\eta}_{L,2}(y, u) \triangleq \Omega(y, u)\hat{\theta}(T_d + T_{L,2})$.

Therefore, in the case of a process fault, by using (28)–(30), r_p satisfies the following equation for $t > T_d + T_{L,2}$ [derived similar to (20)]:

$$r_p(t) = H_p(s)[\Delta\eta_{L,2}(t)] + H_p(s)[\Delta g(t)] - h_p(t)\zeta(T_d) + \epsilon_\zeta(t). \quad (31)$$

Hence, by using the triangle inequality and by using Assumption 6 (note that it is valid under sensor fault-free operation) and Assumption 7, the k th component of r_p from (31) satisfies

$$|r_p^{(k)}(t)| \leq |H_p(s)[\Delta\eta_{L,2}^{(k)}(t)]| + |\epsilon_{\Delta g}^{(k)}(t)| + |h_p(t)|\zeta^{(k)}(T_d) + |\epsilon_\zeta^{(k)}(t)|$$

and hence, $|r_p^{(k)}(t)| \leq \bar{r}_p^{(k)}(t)$, where $\bar{r}_p^{(k)}(t)$ is the process isolation threshold given by

$$\bar{r}_p^{(k)}(t) \triangleq \bar{H}_p(s)[\bar{\eta}_{L,2}^{(k)}(y(t), u(t))] + \bar{\epsilon}_{\Delta g}^{(k)}(t) + |h_p(t)|\bar{\zeta}^{(k)} + \bar{\epsilon}_\zeta^{(k)}(t). \quad (32)$$

Fault Identification Logic: When a process fault occurs and it is detected at time T_d , then the residual $r_p(t)$ given by (28)–(30) satisfies $|r_p^{(k)}(t)| \leq \bar{r}_p^{(k)}(t)$ for all $k = 1, \dots, n$ and for all $t > T_d + T_{L,2}$. Hence, if at some time $t > T_d + T_{L,2}$, we have $|r_p^{(k)}(t)| > \bar{r}_p^{(k)}(t)$ for some $k = 1, \dots, n$, then the possibility of a process fault is excluded, and it is guaranteed that a sensor fault has occurred [single or multiple sensor fault(s)]. The estimation of the sensor fault is given by $\hat{\sigma}(t)$ (to be designed in the sequel) from which the faulty sensors can be identified as the components that deviate significantly from zero.

Remark 4: As it is the case for the time T_L , which is allowed for learning the modeling uncertainty, the training time $T_{L,2}$ should be sufficiently large to allow learning of the combined effect of the modeling uncertainty with the potential process fault. Note that the fault could be incipient and still be developing at the end of the learning period, but given that it has been detected at time T_d (therefore, it has been developing for time $T_d - T_0^x$ which was sufficient for the fault to be detected), it is considered that with the additional time $T_{L,2}$, the error at the end of the learning period will be within the bound stated with Assumption 7.

B. Sensor Faults

In this section, we consider the occurrence of a sensor fault. When a fault is detected at time T_d , the following estimation model is initiated to learn the potential sensor fault that occurred:

$$\dot{\hat{x}}_s(t) = A\hat{x}_s(t) + f(\hat{x}_s(t), u(t)) + \hat{\eta}_L(y(t), u(t)) + \Lambda\epsilon_s^y(t) \quad (33)$$

$$\hat{y}_s(t) = \hat{x}_s(t) + \hat{\sigma}(t) \quad (34)$$

$$\hat{\sigma}(t) = \mathcal{P}_\sigma(\Gamma^y\epsilon_s^y(t)) \quad (35)$$

where $\epsilon_s^y(t)$ denotes the output estimation error $\epsilon_s^y(t) \triangleq y(t) - \hat{y}_s(t)$, Λ is a design matrix that is selected, so that $A_0 \triangleq A - \Lambda$ is Hurwitz, $\Gamma^y \in \mathbb{R}^{n \times n}$ is a symmetric and positive definite learning rate matrix, and $\hat{\sigma}(t) \in \mathbb{R}^n$ denotes

the sensor fault estimation. Finally, the initial conditions are $\hat{x}_s(T_d) = 0$ and $\hat{\sigma}(T_d) = 0$. In the following, we denote the state estimation error as $\epsilon_s^x(t) \triangleq x(t) - \hat{x}_s(t)$. The projection operator \mathcal{P}_σ restricts the estimation vector $\hat{\sigma}(t)$ in a predefined and convex region $\Theta_\sigma \in \mathbb{R}^n$ in order to guarantee the stability of the learning algorithm in the presence of noise and modeling uncertainty. In this paper, Θ_σ is considered to be a zero-origin hypersphere of radius M_σ , and hence, the adaptive law for $\hat{\sigma}$ can be expressed as [32], [47]

$$\dot{\hat{\sigma}}(t) = \Gamma^y \epsilon_s^y(t) - \chi^* \Gamma^y \frac{\hat{\sigma}(t) \hat{\sigma}^\top(t)}{\hat{\sigma}^\top(t) \Gamma^y \hat{\sigma}(t)} \Gamma^y \epsilon_s^y(t) \quad (36)$$

where χ^* denotes the indicator function

$$\chi^* = \begin{cases} 1, & \text{if } |\hat{\sigma}(t)| = M_\sigma \text{ and } \hat{\sigma}^\top(t) \Gamma^y \epsilon_s^y(t) > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (37)$$

The following main result summarizes the properties of the aforementioned adaptive estimation scheme.

Theorem 2: In the case of a sensor fault $\sigma(t)$ that occurs at time T_0^y and is detected at time T_d , the adaptive nonlinear estimation scheme described by (33)–(35) under Assumptions 1–3 and 5, with the positive constants μ , ρ selected such that $|e^{A_0 t}| \leq \mu e^{-\rho t}$ for all $t > 0$ and $\rho - \mu \lambda_f > 0$, guarantees that the following holds.

- 1) The state estimation error $\epsilon_s^x(t)$, the output estimation error $\epsilon_s^y(t)$, and the sensor fault estimation vector $\hat{\sigma}(t)$ are uniformly bounded, i.e., $\epsilon_s^x(t)$, $\epsilon_s^y(t)$, $\hat{\sigma}(t) \in L_\infty$. Specifically, a bound on $\epsilon_s^x(t)$ such that $|\epsilon_s^x(t)| \leq \bar{\epsilon}_s^x(t)$ for all $t > T_d$ is given by (48), and a bound for $\epsilon_s^y(t)$ such that $|\epsilon_s^y(t)| \leq \bar{\epsilon}_s^y(t)$ for all $t > T_d$ is given by (50).
- 2) There exist a positive constant q and a bounded function $v(t)$, such that for all finite $t > T_d$, the output estimation error $\epsilon_s^y(t)$ satisfies

$$\int_{T_d}^t |\epsilon_s^y(\tau)|^2 d\tau \leq q + 4 \int_{T_d}^t |v(\tau)|^2 d\tau. \quad (38)$$

- 3) In the absence of measurement noise (i.e., $\zeta(t) = 0$), if the bounded function $v(t)$ is square integrable (i.e., $v(t) \in L_2$), then $\lim_{t \rightarrow \infty} \epsilon_s^y(t) = 0$.

Proof:

1) When a sensor fault occurs, by using (1) (note that $\phi(\cdot) = 0$) and (33), the state estimation error $\epsilon_s^x(t) = x(t) - \hat{x}_s(t)$ satisfies

$$\begin{aligned} \dot{\epsilon}_s^x(t) &= \dot{x}(t) - \dot{\hat{x}}_s(t) \\ &= Ax(t) + f(x(t), u(t)) + \eta(x(t), u(t)) - A\hat{x}_s(t) \\ &\quad - f(\hat{x}_s(t), u(t)) - \bar{\eta}_L(y(t), u(t)) - \Lambda(y(t) - \hat{y}_s(t)) \\ &= A\epsilon_s^x(t) + \Delta f_s(t) + \Delta \eta_L(t) \\ &\quad - \Lambda(x(t) + \zeta(t) + \sigma(t) - \hat{x}_s(t) - \hat{\sigma}(t)) \\ &= A_0 \epsilon_s^x(t) + \Lambda \tilde{\sigma}(t) - \Lambda \zeta(t) + \Delta f_s(t) + \Delta \eta_L(t) \end{aligned} \quad (39)$$

where

$$\Delta f_s(t) \triangleq f(x(t), u(t)) - f(\hat{x}_s(t), u(t)) \quad (41)$$

$$\tilde{\sigma}(t) \triangleq \hat{\sigma}(t) - \sigma(t) \quad (42)$$

and $\Delta \eta_L(t)$ is given by (21) (note that a fault is considered to occur after the end of training at time T_L). By solving (40), we obtain

$$\begin{aligned} \epsilon_s^x(t) &= e^{A_0(t-T_d)} \epsilon_s^x(T_d) + \int_{T_d}^t e^{A_0(t-\tau)} \\ &\quad \times [\Lambda \tilde{\sigma}(\tau) - \Lambda \zeta(\tau) + \Delta f_s(\tau) + \Delta \eta_L(\tau)] d\tau. \end{aligned} \quad (43)$$

Using the triangle inequality, (43) satisfies

$$\begin{aligned} |\epsilon_s^x(t)| &\leq |e^{A_0(t-T_d)}| |\epsilon_s^x(T_d)| \\ &\quad + \int_{T_d}^t |e^{A_0(t-\tau)}| [|\Lambda| |\tilde{\sigma}(\tau)| + |\Lambda| |\zeta(\tau)| + |\Delta f_s(\tau)| \\ &\quad + |\Delta \eta_L(\tau)|] d\tau. \end{aligned} \quad (44)$$

Since A_0 is Hurwitz, there exist positive constants μ and ρ , such that $|e^{A_0 t}| \leq \mu e^{-\rho t}$ for all $t > 0$ [42]. Moreover, by using the initial conditions $\hat{x}_s(T_d) = 0$, we obtain $|\epsilon_s^x(T_d)| = |x(T_d) - \hat{x}_s(T_d)| = |x(T_d)| \leq \bar{x}_d$, where \bar{x}_d is a suitable bound for the region of the state operation. By using Assumption 2, $|\Delta f_s(t)| \leq \lambda_f |x(t) - \hat{x}_s(t)| = \lambda_f |\epsilon_s^x(t)|$, and by using Assumption 5, we have that $|\Delta \eta_L(t)| \leq |\bar{\eta}_L(y(t), u(t))|$. Regarding the term $\tilde{\sigma}(t) = \hat{\sigma}(t) - \sigma(t)$, it is generally unknown but, since $\hat{\sigma}(t)$ belongs to the known compact set Θ_σ , we have that $|\hat{\sigma}(t) - \sigma(t)| \leq \kappa(t)$ for a suitable $\kappa(t)$ depending on the geometric properties of the set Θ_σ [17], [32]. More specifically, in our case where the compact set Θ_σ is an origin-based hypersphere of radius M_σ , the function $\kappa(t)$ is given by $\kappa(t) = M_\sigma + |\hat{\sigma}(t)|$. Hence, (44) becomes

$$\begin{aligned} |\epsilon_s^x(t)| &\leq \mu e^{-\rho(t-T_d)} \bar{x}_d \\ &\quad + \int_{T_d}^t \mu e^{-\rho(t-\tau)} [|\Lambda| \kappa(\tau) + |\Lambda| \bar{\zeta}_d + \lambda_f |\epsilon_s^x(\tau)| \\ &\quad + |\bar{\eta}_L(y(\tau), u(\tau))|] d\tau \end{aligned} \quad (45)$$

which can be written as

$$|\epsilon_s^x(t)| \leq E_0(t) + \mu \lambda_f e^{-\rho t} \int_{T_d}^t e^{\rho \tau} |\epsilon_s^x(\tau)| d\tau \quad (46)$$

where

$$\begin{aligned} E_0(t) &\triangleq \mu e^{-\rho(t-T_d)} \bar{x}_d + \int_{T_d}^t \mu e^{-\rho(t-\tau)} \\ &\quad \times [|\Lambda| \kappa(\tau) + |\Lambda| \bar{\zeta}_d + |\bar{\eta}_L(y(\tau), u(\tau))|] d\tau. \end{aligned} \quad (47)$$

By applying the Bellman–Gronwall lemma [42] to (46), it results in $|\epsilon_s^x(t)| \leq \bar{\epsilon}_s^x(t)$ where

$$\bar{\epsilon}_s^x(t) \triangleq E_0(t) + \mu \lambda_f \int_{T_d}^t E_0(w) e^{-(\rho - \mu \lambda_f)(t-w)} dw \quad (48)$$

where the positive constants μ and ρ must be selected, such that $\rho - \mu \lambda_f > 0$ to guarantee that $\epsilon_s^x(t)$ remains bounded, and hence, $\epsilon_s^x(t) \in L_\infty$. Moreover, since $|\tilde{\sigma}(t)| = |\hat{\sigma}(t) - \sigma(t)| \leq \kappa(t)$ due to the projection operator, we deduce that $\tilde{\sigma}(t) \in L_\infty$, $\hat{\sigma}(t) \in L_\infty$. Finally, by using (2) and (34), $\epsilon_s^y(t) = y(t) - \hat{y}_s(t)$ becomes

$$\epsilon_s^y(t) = \epsilon_s^x(t) + \zeta(t) - \tilde{\sigma}(t). \quad (49)$$

Using the triangle inequality, (49) satisfies

$$\begin{aligned} |\epsilon_s^y(t)| &\leq |\epsilon_s^x(t)| + |\zeta(t)| + |\tilde{\sigma}(t)| \\ &\leq \bar{\epsilon}_s^y(t) \triangleq \bar{\epsilon}_s^x(t) + \bar{\zeta}_d + \kappa(t) \end{aligned} \quad (50)$$

and hence, $\epsilon_s^y(t)$ is bounded, i.e., $\epsilon_s^y(t) \in L_\infty$.

2) From (43), $\epsilon_s^x(t)$ can be written as

$$\epsilon_s^x(t) = v_1(t) + v_2(t) \quad (51)$$

where $v_1(t)$ and $v_2(t)$ are the solutions of

$$\begin{aligned} \dot{v}_1(t) &= -\Lambda v_1(t) + \Lambda \tilde{\sigma}(t) - \Lambda \zeta(t) + \Delta f_s(t) + \Delta \eta_L(t) \\ \dot{v}_2(t) &= -\Lambda v_2(t) \end{aligned}$$

with initial conditions $v_1(T_d) = 0$ and $v_2(T_d) = \epsilon_s^x(T_d) = x(T_d)$ (note that $\hat{x}_s(T_d) = 0$).

By using (51), (49) becomes

$$\epsilon_s^y(t) = v_1(t) + v_2(t) + \zeta(t) - \tilde{\sigma}(t). \quad (52)$$

Let the Lyapunov function candidate be $V(t) = (1/2)\tilde{\sigma}^\top(t)(\Gamma^y)^{-1}\tilde{\sigma}(t) + \int_t^\infty |v_2(\tau)|^2 d\tau$. The time derivative of $V(t)$ is given by

$$\begin{aligned} \dot{V}(t) &= \tilde{\sigma}^\top(t)(\Gamma^y)^{-1}\dot{\tilde{\sigma}}(t) - |v_2(t)|^2 \\ &= \tilde{\sigma}^\top(t)(\Gamma^y)^{-1}(\dot{\tilde{\sigma}}(t) - \dot{\sigma}(t)) - |v_2(t)|^2 \end{aligned}$$

and by using (36), it becomes

$$\begin{aligned} \dot{V}(t) &= \tilde{\sigma}^\top(t)\epsilon_s^y(t) - \tilde{\sigma}^\top(t)\chi^* \frac{\hat{\sigma}(t)\hat{\sigma}^\top(t)}{\hat{\sigma}^\top(t)\Gamma^y\hat{\sigma}(t)}\Gamma^y\epsilon_s^y(t) \\ &\quad - \tilde{\sigma}^\top(t)(\Gamma^y)^{-1}\dot{\sigma}(t) - |v_2(t)|^2. \end{aligned} \quad (53)$$

Following the same logic as in [47], it can be shown that the second term in (53) satisfies:

$$\tilde{\sigma}^\top(t)\chi^* \frac{\hat{\sigma}(t)\hat{\sigma}^\top(t)}{\hat{\sigma}^\top(t)\Gamma^y\hat{\sigma}(t)}\Gamma^y\epsilon_s^y(t) \geq 0$$

and hence, (53) can be written as

$$\dot{V}(t) \leq \tilde{\sigma}^\top(t)\epsilon_s^y(t) - \tilde{\sigma}^\top(t)(\Gamma^y)^{-1}\dot{\sigma}(t) - |v_2(t)|^2. \quad (54)$$

By using $\tilde{\sigma}(t) = -\epsilon_s^y(t) + v_1(t) + v_2(t) + \zeta(t)$ [from (52)] and by completing the squares, (54) becomes

$$\begin{aligned} \dot{V}(t) &\leq (\epsilon_s^y(t))^\top (-\epsilon_s^y(t) + v_1(t) + v_2(t) + \zeta(t)) \\ &\quad - \tilde{\sigma}^\top(t)(\Gamma^y)^{-1}\dot{\sigma}(t) - |v_2(t)|^2 \\ &\leq -|\epsilon_s^y(t)|^2 + |\epsilon_s^y(t)|\bar{\zeta}_d + |\epsilon_s^y(t)||v_1(t)| + |\epsilon_s^y(t)||v_2(t)| \\ &\quad - |v_2(t)|^2 + |\tilde{\sigma}(t)||(\Gamma^y)^{-1}|\dot{\sigma}(t)| \\ &\leq -\frac{|\epsilon_s^y(t)|^2}{4} + \bar{\zeta}_d^2 + |v_1(t)|^2 + |\tilde{\sigma}(t)||(\Gamma^y)^{-1}|\dot{\sigma}(t)|. \end{aligned} \quad (55)$$

Let $v(t) \triangleq (|v_1(t)|^2 + \bar{\zeta}_d^2 + |\tilde{\sigma}(t)||(\Gamma^y)^{-1}|\dot{\sigma}(t)|)^{(1/2)}$ (note that v is bounded). Then, by integrating (55) from time T_d to t we obtain

$$\begin{aligned} \int_{T_d}^t |\epsilon_s^y(\tau)|^2 d\tau &\leq 4(V(T_d) - V(t)) \\ &\quad + 4 \int_{T_d}^t (|v_1(\tau)|^2 + \bar{\zeta}_d^2 + |\tilde{\sigma}(\tau)||(\Gamma^y)^{-1}|\dot{\sigma}(\tau)|) d\tau \\ &\leq q + 4 \int_{T_d}^t |v(\tau)|^2 d\tau \end{aligned} \quad (56)$$

where $q \triangleq \sup_{t \geq T_d} [4(V(T_d) - V(t))]$ is a positive constant (since $V(t)$ is uniformly bounded).

3) Since $\epsilon_s^y(t) \in L_\infty$ as it was shown in part (1) of the proof, then from (35), we deduce that $\hat{\sigma}(t) \in L_\infty$. Moreover, all terms in the right-hand side of (40) were shown to be bounded, and therefore, $\dot{\epsilon}_s^x(t) \in L_\infty$. Moreover, in the absence of noise (i.e., $\zeta(t) = 0$), from (49), we have that $\dot{\epsilon}_s^y(t) = \dot{\epsilon}_s^x(t) + \dot{\sigma}(t) - \dot{\hat{\sigma}}(t)$, and since $\dot{\epsilon}_s^x(t) \in L_\infty$, $\dot{\sigma}(t) \in L_\infty$ (from Assumption 3), and $\dot{\hat{\sigma}}(t) \in L_\infty$, we obtain that $\dot{\epsilon}_s^y(t) \in L_\infty$.

Therefore, if the bounded function $v(t)$ is square integrable, i.e., $v(t) \in L_2$, then (56) can be valid for $t \rightarrow \infty$, leading to $\epsilon_s^y(t) \in L_2$. Since $\epsilon_s^y(t) \in L_\infty$, $\dot{\epsilon}_s^y(t) \in L_\infty$, and $\epsilon_s^y(t) \in L_2$, then according to Barbalat's Lemma, $\lim_{t \rightarrow \infty} \epsilon_s^y(t) = 0$. ■

Now, we exploit the findings of Theorem 2 to derive suitable bounds to aid in the determination of the fault type. More specifically, we proceed to derive a threshold for the case of a sensor fault, so that when it is exceeded then it is guaranteed that a process fault has occurred. Two such residuals are given. The first one is the output estimation error $\epsilon_s^y(t) = y(t) - \hat{y}_s(t)$, since it is bounded by $\bar{\epsilon}_s^y(t)$ given by (50) for all $t > T_d$ in the case of a sensor fault. The second residual $r_s(t)$ exploits both the use of filtering and some signals from the designed adaptive scheme, to cancel out some terms in the residual and allow a potentially tighter threshold. Therefore, the residual $r_s(t)$ in this case is obtained by

$$r_s(t) \triangleq H(s)[\epsilon_s^y(t) + w(t)] \quad (57)$$

$$\dot{w}(t) = \Lambda(y(t) - \hat{y}_s(t)) + \dot{\hat{\sigma}}(t) \quad (58)$$

with initial condition $w(T_d) = 0$. For simplicity, in the sequel, we omit the initial condition terms (now at time T_d) that are exponentially decaying to zero, since they are multiplied with $h_p(t)$, and do not affect substantially the threshold derivation. Therefore, by using (49), (57) can be written

$$\begin{aligned} r_s(t) &= H(s)[\epsilon_s^x(t) + \zeta(t) - \tilde{\sigma}(t) + w(t)] \\ &= H_p(s)[\dot{\epsilon}_s^x(t) - \dot{\tilde{\sigma}}(t) + \dot{w}(t)] + \epsilon_\zeta(t) \end{aligned} \quad (59)$$

and by using (39) and (58), (59) becomes

$$\begin{aligned} r_s(t) &= H_p(s)[A\epsilon_s^x(t) + \Delta f_s(t) + \Delta \eta_L(t) - \Lambda(y(t) - \hat{y}_s(t)) \\ &\quad - (\dot{\hat{\sigma}}(t) - \dot{\sigma}(t)) + \Lambda(y(t) - \hat{y}_s(t)) + \dot{\hat{\sigma}}(t)] + \epsilon_\zeta(t) \\ &= H_p(s)[A\epsilon_s^x(t) + \Delta f_s(t) + \Delta \eta_L(t) + \dot{\sigma}(t)] + \epsilon_\zeta(t). \end{aligned} \quad (60)$$

Thus, by using the triangle inequality, the k th component of (60) satisfies

$$\begin{aligned} |r_s^{(k)}(t)| &\leq \bar{H}_p(s)[|A^{(k)}|\epsilon_s^x(t) + |\Delta f_s^{(k)}(t)| + |\Delta \eta_L^{(k)}(t)| \\ &\quad + |\dot{\sigma}^{(k)}(t)|] + |\epsilon_\zeta^{(k)}(t)|. \end{aligned} \quad (61)$$

Using Assumption 2, 3, and 5, a suitable threshold $\bar{r}_s^{(k)}(t)$, such that $|r_s^{(k)}(t)| \leq \bar{r}_s^{(k)}(t)$ for all $t > T_d$ and for all $k = 1, \dots, n$ is obtained as

$$\begin{aligned} \bar{r}_s^{(k)}(t) &\triangleq \bar{H}_p(s)[(|A^{(k)}| + \lambda_{f_k})\bar{\epsilon}_s^x(t) + \bar{\eta}_L^{(k)}(y(t), u(t)) + \psi] \\ &\quad + \bar{\epsilon}_\zeta^{(k)}(t). \end{aligned} \quad (62)$$

Fault Identification Logic: When a sensor fault occurs and is detected at time T_d , then $|\epsilon_s^y(t)| \leq \bar{\epsilon}_s^y(t)$ and $|r_s^{(k)}(t)| \leq \bar{r}_s^{(k)}(t)$

for all $k = 1, \dots, n$ and for all $t > T_d$. Therefore, if at some time $t > T_d$, we have $|\epsilon_s^y(t)| > \bar{\epsilon}_s^y(t)$ or $|r_s^{(k)}(t)| > \bar{r}_s^{(k)}(t)$ for some $k = 1, \dots, n$, then the possibility of a sensor fault is excluded, and it is guaranteed that a process fault has occurred. The estimation of the process fault (including the modeling uncertainty) is given by $\hat{\eta}_{L,2}$ for $t > T_d + T_{L,2}$.

VI. FAULT TYPE IDENTIFICATION ANALYSIS

In this section, fault type identification conditions for the aforementioned fault diagnosis scheme are derived. This analysis, similar to the fault detectability analysis given in Section IV, constitutes a theoretical result that characterizes quantitatively and, in nonclosed form, the classes of faults that can be identified (process or sensor) by the proposed scheme.

Theorem 3: Consider the nonlinear system (1) and (2) with the adaptive approximation structure described in (7), (10), (11), (16), and (17) along with the two estimation models: for process faults given by (28)–(30) and for sensor faults given by (33)–(35) and their corresponding isolation thresholds given by (32) ($\bar{r}_p(t)$) and (62) ($\bar{r}_s(t)$).

- 1) *Process Fault Identification:* A process fault type is identified if one of the following conditions (63) or (65) is satisfied for some $k = 1, 2, \dots, n$ at some time $t > T_d$ (the possibility of a sensor fault is excluded):

$$|\epsilon_{s,2}^x(t)| > \bar{\epsilon}_s^y(t) + \bar{\xi}_d + |\hat{\sigma}(t)| \quad (63)$$

where

$$\begin{aligned} \epsilon_{s,2}^x(t) \triangleq & e^{A_0(t-T_d)}x(T_d) + \int_{T_d}^t e^{A_0(t-\tau)} \\ & \times [\Lambda\hat{\sigma}(\tau) - \Lambda\zeta(\tau) + \Delta f_s(\tau) + \Delta\eta_L(\tau) \\ & + \beta^x(\tau - T_0^x)\phi(x(\tau), u(\tau))] d\tau \quad (64) \end{aligned}$$

or

$$\begin{aligned} |H_p(s)[\beta^x(t - T_0^x)\phi^{(k)}(x(t), u(t))] &| > \bar{r}_s^{(k)}(t) \\ + |H_p(s)[A^{(k)}\epsilon_s^x(t) + \Delta f_s^{(k)}(t) + \Delta\eta_L^{(k)}(t)] &| + \epsilon_\xi^{(k)}(t). \quad (65) \end{aligned}$$

- 2) *Sensor Fault Identification:* A sensor fault type is identified if the following condition is satisfied for some $k = 1, 2, \dots, n$ at some time $t > T_d + T_{L,2}$ (the possibility of a process fault is excluded):

$$\begin{aligned} |H(s)[\sigma^{(k)}(t)] + H_p(s)[g^{(k)}(x(t) + \zeta(t), u(t)) \\ - g^{(k)}(x(t) + \zeta(t) + \sigma(t), u(t))] &| > 2\bar{r}_p^{(k)}(t). \quad (66) \end{aligned}$$

Proof:

- 1) When a fault is detected at time T_d , one of the estimators initiated for identifying the fault type is the one for sensor faults given in Section V-B, which satisfies $|\epsilon_s^y(t)| \leq \bar{\epsilon}_s^y(t)$ and $|r_s^{(k)}(t)| \leq \bar{r}_s^{(k)}(t)$ for all $k = 1, \dots, n$ and for all $t > T_d$, if a sensor fault has occurred. Therefore, when a process fault occurs and the fault is detected, the state estimation error $\epsilon_s^x(t) = x(t) - \hat{x}_s(t)$ (now indicated as $\epsilon_{s,2}^x(t)$) is given by (64) (derived similar to (43) by using $\sigma(t) = 0$).

By using the triangle inequality, the output estimation error $\epsilon_s^y(t)$ from (49) satisfies

$$\begin{aligned} |\epsilon_s^y(t)| &\geq |\epsilon_{s,2}^x(t)| - |\zeta(t)| - |\hat{\sigma}(t)| \\ &\geq |\epsilon_{s,2}^x(t)| - \bar{\xi}_d - |\hat{\sigma}(t)|. \quad (67) \end{aligned}$$

For process fault identification, the inequality $|\epsilon_s^y(t)| > \bar{\epsilon}_s^y(t)$ (possibility of a sensor fault is excluded) must hold at some time $t > T_d$, so the sensor fault identification condition in (63) is obtained.

The second process fault identification condition given by (65) is obtained as follows. When a process fault occurs and the fault is detected, the residual $r_s(t)$ from (60) becomes

$$\begin{aligned} r_s(t) = H_p(s)[A\epsilon_s^x(t) + \Delta f_s(t) + \Delta\eta_L(t) \\ + \beta^x(t - T_0^x)\phi(x(t), u(t))] + \epsilon_\xi(t). \quad (68) \end{aligned}$$

Using the triangle inequality, the k th component of (68) satisfies

$$\begin{aligned} |r_s^{(k)}(t)| &\geq |H_p(s)[\beta^x(t - T_0^x)\phi^{(k)}(x(t), u(t))] \\ &\quad - |H_p(s)[A^{(k)}\epsilon_s^x(t) + \Delta f_s^{(k)}(t) + \Delta\eta_L^{(k)}(t)] \\ &\quad + \epsilon_\xi^{(k)}(t). \quad (69) \end{aligned}$$

For process fault determination, the inequality $|r_s^{(k)}(t)| > \bar{r}_s^{(k)}(t)$ for some $k = 1, \dots, n$ (possibility of a sensor fault is excluded) must hold at some time $t > T_d$, so the sensor fault identification condition in (65) is obtained.

- 2) When a fault is detected at time T_d , one of the estimators initiated for identifying the fault type is the one for process faults given in Section V-A, which satisfies $|r_p^{(k)}(t)| \leq \bar{r}_p^{(k)}(t)$ for all $k = 1, \dots, n$ and $t > T_d + T_{L,2}$ if a process fault has occurred. Therefore, when a sensor fault occurs and the fault is detected, the residual $r_p(t)$ from (31), by using (14) and by adding and subtracting $g(x(t) + \zeta(t), u(t))$, becomes

$$\begin{aligned} r_p(t) = H_p(s)[g(x(t), u(t)) \\ - g(x(t) + \zeta(t), u(t)) + g(x(t) + \zeta(t), u(t)) \\ - g(x(t) + \zeta(t) + \sigma(t), u(t))] + H(s)[\sigma(t)] \\ + H_p(s)[\Delta\eta_{L,2}(t)] - h_p(t)\zeta(T_d) + \epsilon_\xi(t). \quad (70) \end{aligned}$$

By using the triangle inequality and Assumptions 6 and 7, the k th component of (70) satisfies

$$\begin{aligned} |r_p^{(k)}(t)| &\geq |H(s)[\sigma^{(k)}(t)] + H_p(s)[g^{(k)}(x(t) + \zeta(t), u(t)) \\ &\quad - g^{(k)}(x(t) + \zeta(t) + \sigma(t), u(t))] \\ &\quad - |\epsilon_{\Delta g}^{(k)}(t)| - |H_p(s)[\Delta\eta_{L,2}^{(k)}(t)] \\ &\quad - |h_p(t)||\zeta^{(k)}(T_d)| - |\epsilon_\xi^{(k)}(t)| \\ &\geq |H(s)[\sigma^{(k)}(t)] + H_p(s)[g^{(k)}(x(t) + \zeta(t), u(t)) \\ &\quad - g^{(k)}(x(t) + \zeta(t) + \sigma(t), u(t))] \\ &\quad - \bar{\epsilon}_{\Delta g}^{(k)}(t) - \bar{H}_p(s)[\bar{\eta}_{L,2}^{(k)}(y(t), u(t))] - |h_p(t)|\bar{\zeta}^{(k)} - \bar{\epsilon}_\xi^{(k)}(t) \\ &\geq |H(s)[\sigma^{(k)}(t)] + H_p(s)[g^{(k)}(x(t) + \zeta(t), u(t)) \\ &\quad - g^{(k)}(x(t) + \zeta(t) + \sigma(t), u(t))] \\ &\quad - \bar{r}_p^{(k)}(t) \end{aligned}$$

and therefore, for sensor fault identification, the inequality $|r_p^{(k)}(t)| > \bar{r}_p^{(k)}(t)$ (possibility of a process fault is excluded) must hold at some time $t > T_d + T_{L,2}$ for some $k = 1, 2, \dots, n$, so the final sensor fault identification condition in (66) is obtained. ■

As in the case of the fault detectability conditions given in Theorem 1, the fault type identification conditions (63), (65), and (66) stated in Theorem 3 are only sufficient (but not necessary), and hence, the classes of the identifiable faults can be larger.

Remark 5: The identification residuals $r_p(t)$ and $r_s(t)$ utilize filtering and the learned (potential) process or sensor fault in order to obtain smaller residuals, so that tighter identification threshold $\bar{r}_p(t)$ and $\bar{r}_s(t)$ can be obtained, respectively. The implementation of these thresholds relies on the bounds stated in Assumptions 5–7, and as a result, crude knowledge of these bounds may lead to conservative identification thresholds, making it hard to be crossed by their respective residual and, thus, not concluding the fault type. Therefore, the selection of proper bounds required by Assumptions 5–7 is of crucial importance in order to ease the fault type identification task. Note that, in the case in which the fault type identification residuals remain below their corresponding identification thresholds, the fault type cannot be identified.

VII. SIMULATION RESULTS

In this section, we apply the proposed approach to an example of a single-link robotic arm with a revolute elastic joint with motion dynamics described by [32], [40]

$$\begin{aligned} J_l \ddot{q}_l + F_l \dot{q}_l + k(q_l - q_m) + (m + \Delta m)gh \sin(q_l) &= 0 \\ J_m \ddot{q}_m + F_m \dot{q}_m - k(q_l - q_m) &= k_\tau u \end{aligned}$$

where q_l is the angular position of the link, q_m is the angular position of the motor, J_l and J_m are the link and motor inertia, respectively, F_l and F_m are the viscous coefficients of the link and motor respectively, k is the elastic constant, m is the link mass, Δm is the link mass inaccuracy, g is the gravity constant, h is the center of the mass, k_τ is the amplifier gain, and u is the torque input delivered by the motor. The values of the parameters in SI units are: $J_l = 4.5$, $J_m = 1$, $F_l = 0.5$, $F_m = 1$, $k = 2$, $m = 4$, $\Delta m = 0.05m$, $g = 9.8$, $h = 0.5$, $k_\tau = 1$, and $u = 2 \sin(0.25t)$. By selecting $x_1 = \dot{q}_m$, $x_2 = q_m$, $x_3 = \dot{q}_l$ and $x_4 = q_l$, the nonlinear uncertain model becomes

$$\dot{x}(t) = Ax(t) + f(x_4(t), u(t)) + \eta(x_4(t)) \quad (71)$$

where $x(t) = [x_1(t), x_2(t), x_3(t), x_4(t)]^\top$ is the state vector

$$A = \begin{bmatrix} -\frac{F_m}{J_m} & -\frac{k}{J_m} & 0 & \frac{k}{J_m} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{k}{J_l} & -\frac{F_l}{J_l} & -\frac{k}{J_l} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$f(x_4(t), u(t)) = \begin{bmatrix} \frac{k_\tau u(t)}{J_m} \\ 0 \\ -\frac{mgh}{J_l} \sin(x_4(t)) \\ 0 \end{bmatrix}$$

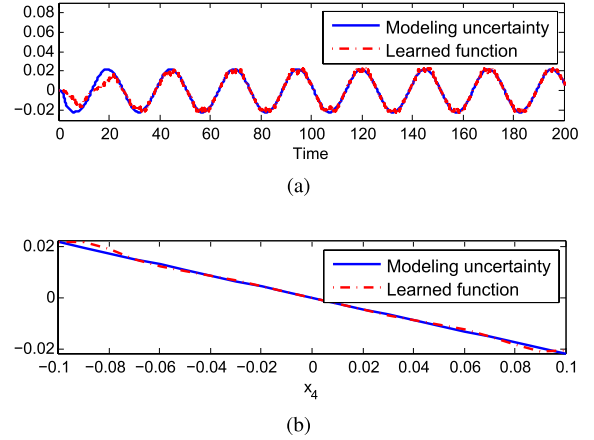


Fig. 3. (a) Learning of the modeling uncertainty. (b) Learned modeling uncertainty after the end of learning.

and $\eta(x_4(t)) = [0, 0, -(\Delta mgh/J_l) \sin(x_4(t)), 0]^\top$. The initial conditions of the plant are set as $x(0) = 0$. The measurement noise $\zeta^{(k)}(t)$ is generated from a uniform distribution with 5% uncertainty of the state. The nominal nonlinear function $f(x_4(t), u(t))$ satisfies the Lipschitz condition in $x_4(t) \in \mathbb{R}$ for all u, t with a Lipschitz constant $\lambda_f = mgh/J_l$. Note that, componentwise the function $f^{(k)}$ has a Lipschitz constant λ_{f_k} , $k = 1, \dots, 4$ with $\lambda_{f_3} = mgh/J_l$ and $\lambda_{f_1} = \lambda_{f_2} = \lambda_{f_4} = 0$.

The filter used for dampening noise is $H(s) = sH_p(s)$, where $H_p(s) = 50^2/((s+10)(s+50)^2)$, and by using $\alpha = 10$, the required implementation filters are given by $H_1(s) = s/(s+10)$ and $H_2(s) = 50^2/(s+50)^2$. Note that $H_p(s)$ has a nonnegative impulse response, and therefore, the filter required for the threshold implementation is $\hat{H}_p(s) = H_p(s)$. For simplicity, the bounds on the filtered noise used are found from the simulation and are $\bar{\epsilon}_\zeta^{(1)} = 0.01$, $\bar{\epsilon}_\zeta^{(2)} = 0.04$, $\bar{\epsilon}_\zeta^{(3)} = 0.002$, and $\bar{\epsilon}_\zeta^{(4)} = 0.006$. Similarly, the bounds on the filtered mismatch function used are $\bar{\epsilon}_{\Delta g}^{(1)} = 0.005$, $\bar{\epsilon}_{\Delta g}^{(2)} = 0.0005$, $\bar{\epsilon}_{\Delta g}^{(3)} = 0.002$, and $\bar{\epsilon}_{\Delta g}^{(4)} = 0.001$.

The objective is to apply the proposed fault diagnosis framework to achieve the following tasks: 1) learn the modeling uncertainty η ; 2) detect any faults; and 3) identification of the type of the fault (process or sensor) and fault estimation. In addition, the enhancement of the fault detectability by using the integrated filtering and learning approach is also demonstrated.

- 1) *Learning the Modeling Uncertainty:* This task is based on Sections III-A and III-B. The training time to learn the modeling uncertainty is selected as $T_L = 200$ s. We use a RBF network with 30 centers distributed evenly in the region of interest $x^{(4)} = [-0.25, 0.25]$ and constant variance $\sigma^2 = 0.02^2$ for each basis function. It is also considered that the radius of the hypersphere for the parameter adaptation is $M_p = 1$ and the constant matrix used for learning is $\Gamma^x = 2\mathbb{I}$, where \mathbb{I} is the identity matrix. The results of the learning process are shown in Fig. 3, in which the modeling uncertainty $\eta^{(3)}$ and its estimation $\hat{\eta}_L^{(3)}$ are shown. As it can be seen,

the learning procedure is effective and the modeling uncertainty is learned quickly.

- 2) *Fault Detection*: This task is based on Sections III-C and III-D. After the end of the learning period at T_L , the fault detection scheme is initiated by implementing the residual and the detection threshold according to (18) and (23), respectively, using a bound for the learned modeling uncertainty $\tilde{\eta}_L = [0, 0, 0.05, 0]^\top$.
- 3) *Fault Type Identification and Estimation*: After the detection of the fault, the fault identification and estimation are initiated according to Section V-A to investigate the process fault occurrence and Section V-B to investigate the sensor fault occurrence. For the investigation of a process fault, we essentially resume the learning process for the modeling uncertainty in order to learn the combined effect of the modeling uncertainty and the potential process fault. The training time is selected as $T_{L,2} = 200$ s, and the learning is conducted during the time interval $[T_d, T_d + T_{L,2}]$. The matrix used for learning is changed to $\Gamma^x = 20\mathbb{I}$ to allow learning of potential faster changes, and the bound used for the learned modeling uncertainty combined with the process fault is the same as before, i.e., $\tilde{\eta}_{L,2} = \tilde{\eta}_L$. After the end of the training, if a residual $r_p^{(k)}$ (30) exceeds its corresponding process fault isolation threshold $\tilde{r}_s^{(k)}$ (32) for some $k = 1, \dots, 4$, then the occurrence of a process fault is excluded, and it is guaranteed that a sensor fault has occurred. Once a process fault has been excluded, then the sensor fault estimation is given by the $\hat{\sigma}(t)$ that contains the estimation of each sensor fault.

For the investigation of a sensor fault, we implement the estimation model (33)–(35). The gain matrix Λ is selected through the pole placement, so that the eigenvalues of A_0 are located at $[-9.1, -9.8, -10.5, -11.2]$. The residual in this case is implemented according to (57) and (58) and the sensor isolation thresholds according to (48) and (62), by using $\mu = 1$, $\rho = 7.5$, $M_\sigma = 0.1$, and $\Gamma^y = 0.1$. If at some time $t > T_d$, a residual $r_s^{(k)}$ (57) exceeds its corresponding sensor fault isolation threshold $\tilde{r}_s^{(k)}$ (62) for some $k = 1, \dots, 4$, then the possibility of a sensor fault is excluded, and it is guaranteed that a process fault has occurred. The estimation of the process fault including the modeling uncertainty is given by $\hat{\eta}_{L,2}$ for $t > T_d + T_{L,2}$ (after the end of learning).

In the sequel, we will examine two cases: the first one considers the occurrence of a process fault and the second one considers the occurrence of a sensor fault.

Case 1 (Process Fault): The process fault is considered to be a fault that results in the reduction of the mass of the link, which occurs abruptly at time $T_0^x = 225$ s. Specifically, the fault function is given by $\phi(x(t), u(t)) = [0, 0, -\theta_x((m + \Delta m)gh/J_l) \sin(x^{(4)}(t)), 0]^\top$, where $\theta_x \in [-1, 0]$ represents the percentage change of the mass, which in this example is considered $\theta_x = -0.5$. At first, in order to demonstrate effectiveness of the proposed scheme, we obtain the fault detection results without the use of filtering or learning. These results, which are shown in Fig. 4, are obtained by using suitable bounds on the noise magnitude and

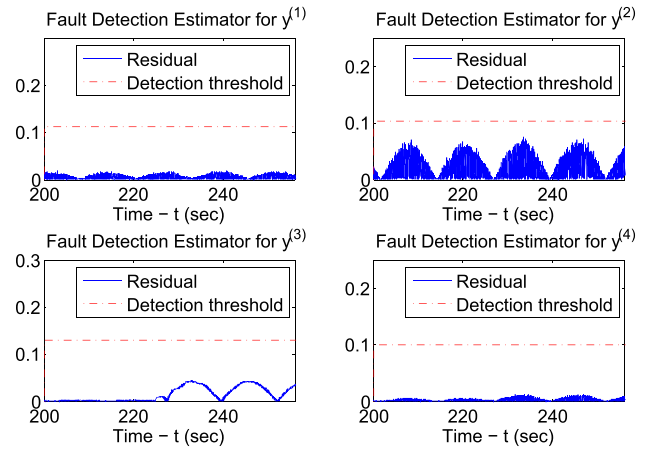


Fig. 4. Process fault detection without the use of learning or filtering.

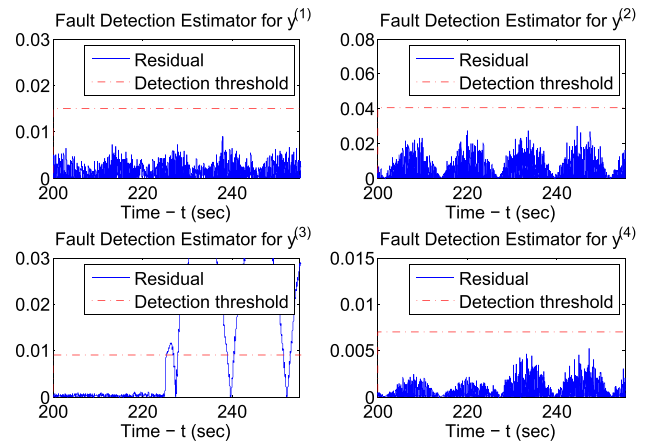


Fig. 5. Process fault detection with the integrated filtering and learning approach.

modeling uncertainty in order to guarantee no-false alarms as described in Remark 2. As it can be seen from Fig. 4, the detection thresholds are too conservative to be crossed by their respective residuals, and hence, the fault is not detected. On the other hand, by using the integrated filtering and learning approach proposed in this paper, we have successful fault detection, as it can be seen from Fig. 5 that shows the residual $r^{(k)}$ and corresponding detection threshold $\tilde{r}^{(k)}$ for each measurement $y^{(k)}$, $k = 1, \dots, 4$. Specifically, the residual exceeds its threshold in the case of the estimator that monitors $y^{(3)}$ at about $T_d = 225.5$ s, and hence, the fault is detected.

After the fault is detected, the process and sensor fault estimation models are activated, and their results are shown in Fig. 6(a) and (b), respectively. As it can be seen from Fig. 6(b), in FIE for $y^{(4)}$, the residual exceeds its sensor identification threshold at ~ 241 s, and hence, the case of a sensor fault is excluded, and the occurrence of a process fault is concluded. In addition, note that in Fig. 6(a), all the residuals remain below their corresponding process fault identification thresholds after the end of the learning. The process fault, including the modeling uncertainty $\eta^{(3)} + \phi^{(3)}$ and its estimation $\hat{\eta}_{L,2}^{(3)}$, is shown in Fig. 7, where it is seen that the function has been learned very well.

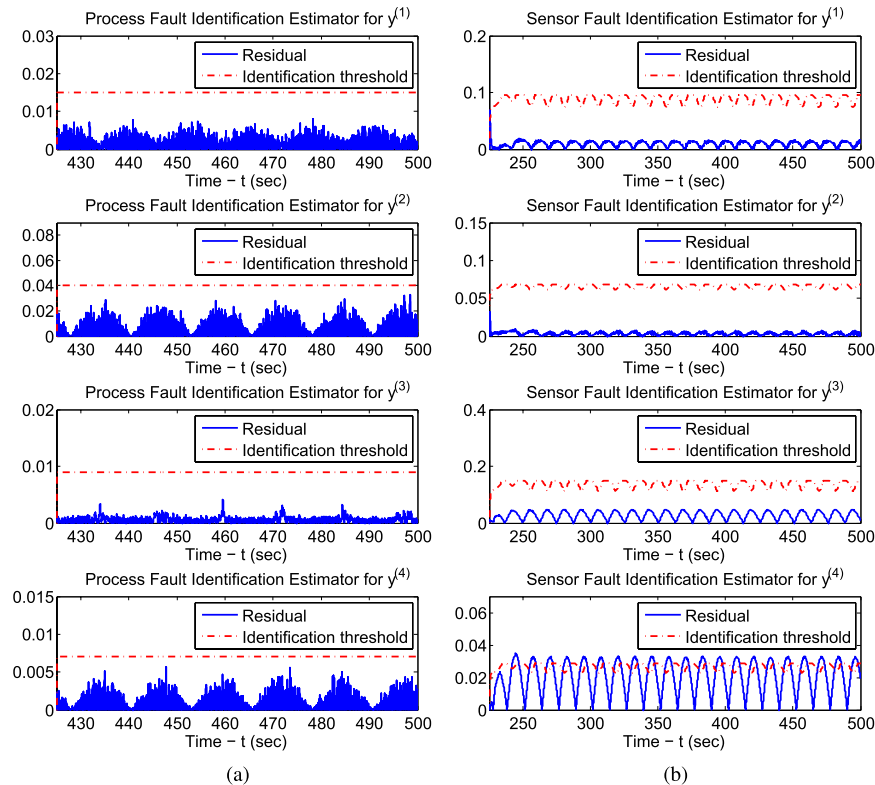


Fig. 6. FIEs for (a) process fault and (b) sensor fault in the occurrence of a process fault.

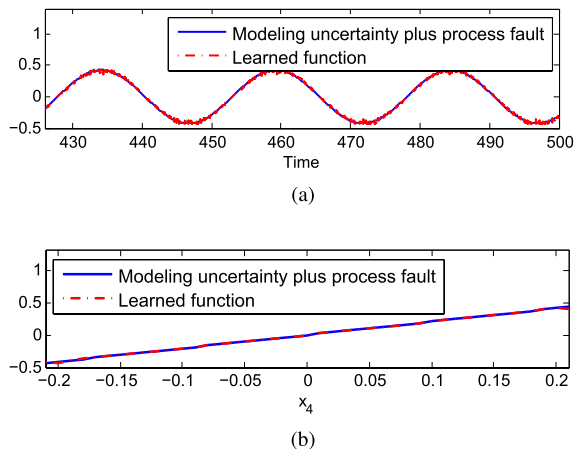


Fig. 7. (a) Learning of process fault including modeling uncertainty. (b) Learned process fault including modeling uncertainty after the end of learning.

Case 2 (Sensor Fault): At time $T_0^y = 225$ s, two sensors (2 and 3) become faulty by measuring the state variables $x^{(2)}$ and $x^{(3)}$ with a constant sensor bias $\sigma^{(2)} = 0.07$ and $\sigma^{(3)} = -0.07$, respectively (in addition to noise). Fig. 8 shows the residual $r^{(k)}$ and corresponding detection threshold $\bar{r}^{(k)}$ for each measurement $y^{(k)}$, $k = 1, \dots, 4$. As it can be seen from Fig. 8, the residual exceeds its threshold in the case of the estimators that monitor $y^{(1)}$ and $y^{(4)}$. The fault is detected the first time the threshold is exceeded at time $T_d = 226.5$ s.

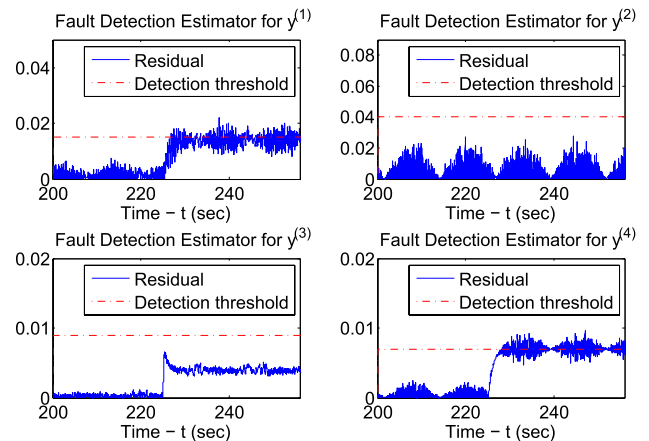


Fig. 8. Sensor fault detection.

After the fault is detected, the process and sensor fault estimation models are activated, and their results are shown in Fig. 9(a) and (b), respectively. As it can be seen from Fig. 9(b), all the residuals remain below their sensor identification thresholds, whereas in Fig. 9(a), the residuals of the FIEs that monitor $y^{(1)}$ and $y^{(4)}$ exceed their corresponding process fault identification thresholds immediately after the end of the training phase at time $T_d + T_{L,2} = 425$ s. Therefore, the case of a process fault is excluded, and the occurrence of a sensor fault is concluded. The estimation of the sensor fault $\hat{\sigma}(t)$ is shown in Fig. 10, in which it can be seen that the sensor fault is estimated correctly.

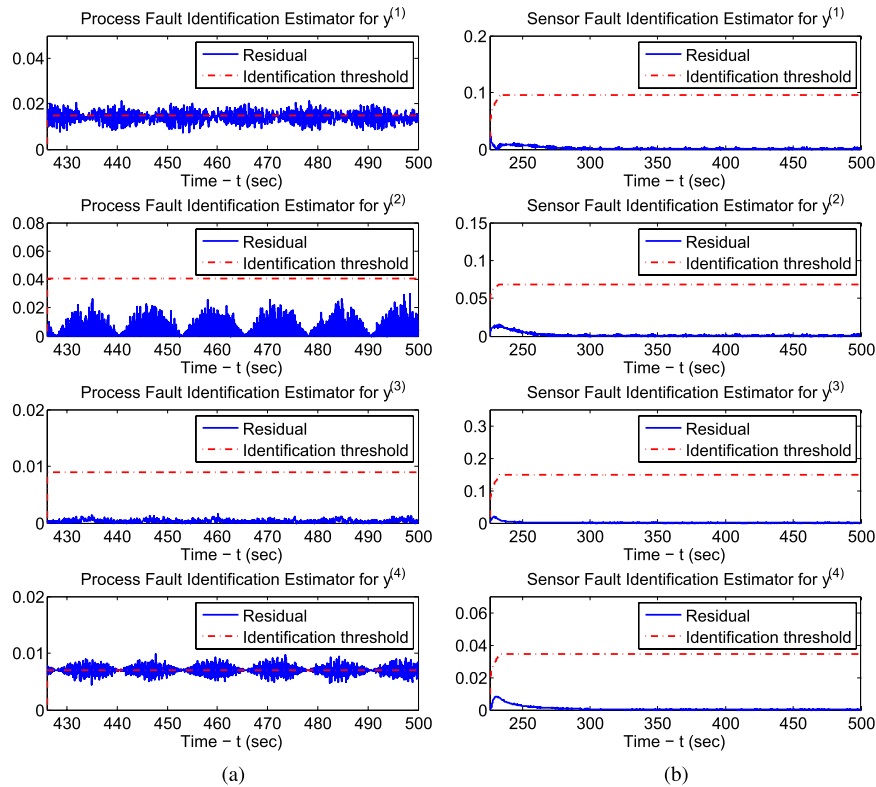


Fig. 9. FIEs for (a) process fault and (b) sensor fault in the occurrence of a sensor fault.

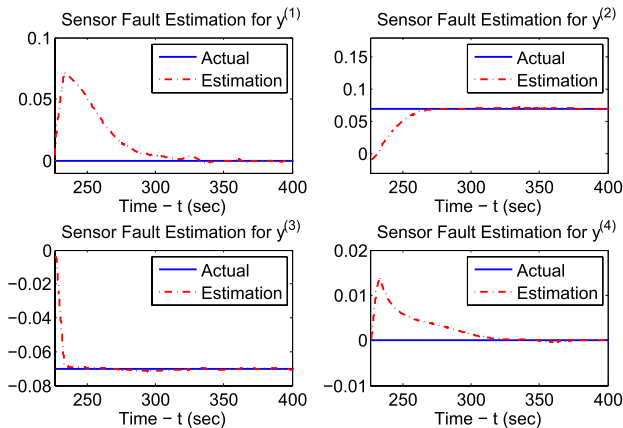


Fig. 10. Learned sensor fault.

VIII. CONCLUSIONS

In this paper, a unified fault diagnosis approach for process and sensor faults in a class of continuous-time, nonlinear systems with modeling uncertainties, and measurement noise is presented, which integrates learning with filtering techniques for obtaining tight detection thresholds. The scheme exploits the filtering framework to deal with the learning and noise attenuation tasks in a two-stage filtering process for designing the required signals both tasks. By using the learned function in the design of the residual signal and exploiting the noise attenuation of a general class of filters, enhanced fault detectability is achieved. When a fault is detected, two estimation models, one for process and one for sensor faults, are initiated in order to learn the potential fault that

has occurred. Through the design of suitable residual and identification thresholds, the fault type can be determined, and the estimation of the fault is obtained.

Future research efforts will be devoted in designing decentralized and distributed fault diagnosis scheme for distributed systems in which the fault effects can propagate among subsystems.

REFERENCES

- [1] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, Mar. 2003.
- [3] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 3, pp. 636–653, May 2010.
- [4] R. Patton and P. M. Frank, *Fault Diagnosis in Dynamic Systems: Theory and Application*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.
- [5] J. Chen and R. J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Norwell, MA, USA: Kluwer, 1999.
- [6] R. Isermann, *Fault-Diagnosis Systems: An Introduction From Fault Detection to Fault Tolerance*, 1st ed. New York, USA, Springer, 2005.
- [7] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*, 2nd ed. New York, USA, Springer-Verlag, 2010.
- [8] E. A. García and P. M. Frank, "Deterministic nonlinear observer-based approaches to fault diagnosis: A survey," *Control Eng. Pract.*, vol. 5, no. 5, pp. 663–670, 1997.
- [9] H. Hammouri, M. Kinnaert, and E. H. El Yaagoubi, "Observer-based approach to fault detection and isolation for nonlinear systems," *IEEE Trans. Autom. Control*, vol. 44, no. 10, pp. 1879–1884, Oct. 1999.
- [10] Q. Zhang, M. Basseville, and A. Benveniste, "Fault detection and isolation in nonlinear dynamic systems: A combined input-output and local approach," *Automatica*, vol. 34, no. 11, pp. 1359–1373, 1998.

- [11] R. Rajamani and A. Ganguli, "Sensor fault diagnostics for a class of nonlinear systems using linear matrix inequalities," *Int. J. Control*, vol. 77, no. 10, pp. 920–930, 2004.
- [12] W. Chen and M. Saif, "Observer-based strategies for actuator fault detection, isolation and estimation for certain class of uncertain nonlinear systems," *IET Control Theory Appl.*, vol. 1, no. 6, pp. 1672–1680, Nov. 2007.
- [13] S. Narasimhan, P. Vachhani, and R. Rengaswamy, "New nonlinear residual feedback observer for fault diagnosis in nonlinear systems," *Automatica*, vol. 44, no. 9, pp. 2222–2229, Sep. 2008.
- [14] C. De Persis and A. Isidori, "A geometric approach to nonlinear fault detection and isolation," *IEEE Trans. Autom. Control*, vol. 46, no. 6, pp. 853–865, Jun. 2001.
- [15] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [16] M. M. Polycarpou and A. J. Helmicki, "Automated fault detection and accommodation: A learning systems approach," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 11, pp. 1447–1458, Nov. 1995.
- [17] X. Zhang, M. M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *IEEE Trans. Autom. Control*, vol. 47, no. 4, pp. 576–593, Apr. 2002.
- [18] H. Ferdowsi, S. Jagannathan, and M. Zawodniok, "An online outlier identification and removal scheme for improving fault detection performance," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 908–919, May 2014.
- [19] R. J. Patton, C. Kambhampati, A. Casavola, P. Zhang, S. Ding, and D. Sauter, "A generic strategy for fault-tolerance in control systems distributed over a network," *Eur. J. Control*, vol. 13, nos. 2–3, pp. 280–296, 2007.
- [20] S. Klinkhieo and R. J. Patton, "A two-level approach to fault-tolerant control of distributed systems based on the sliding mode," in *Proc. 7th IFAC Symp. Fault Detect., Supervision, Safety Tech. Processes*, Barcelona, Spain, 2009, pp. 1043–1048.
- [21] W. Li, W. Gui, Y. Xie, and S. X. Ding, "Decentralized fault detection system design for large-scale interconnected systems," in *Proc. 7th IFAC Symp. Fault Detect., Supervision, Safety Tech. Processes*, Barcelona, Spain, 2009, pp. 816–821.
- [22] N. Léchevin and C. A. Rabbath, "Decentralized detection of a class of non-abrupt faults with application to formations of unmanned airships," *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 2, pp. 484–493, Mar. 2009.
- [23] X. Zhang, M. M. Polycarpou, and T. Parisini, "Decentralized fault detection in a class of large-scale nonlinear uncertain systems," in *Proc. 48th IEEE Conf. Decision Control, 28th Chin. Control Conf.*, Dec. 2009, pp. 6988–6993.
- [24] S. Stanković, N. Ilić, Ž. Djurović, M. Stanković, and K. H. Johansson, "Consensus based overlapping decentralized fault detection and isolation," in *Proc. Conf. Control Fault-Tolerant Syst. (SysTol)*, Oct. 2010, pp. 570–575.
- [25] X. Zhang, "Decentralized fault detection for a class of large-scale nonlinear uncertain systems," in *Proc. Amer. Control Conf.*, Jun./Jul. 2010, pp. 5650–5655.
- [26] R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 275–290, Feb. 2012.
- [27] C. Keliris, M. M. Polycarpou, and T. Parisini, "A distributed fault detection filtering approach for a class of interconnected continuous-time nonlinear systems," *IEEE Trans. Autom. Control*, vol. 58, no. 8, pp. 2032–2047, Aug. 2013.
- [28] H. Ferdowsi, D. L. Raja, and S. Jagannathan, "A decentralized fault prognosis scheme for nonlinear interconnected discrete-time systems," in *Proc. Amer. Control Conf.*, Jun. 2012, pp. 5900–5905.
- [29] Q. Zhang and X. Zhang, "A distributed detection scheme for process faults and sensor faults in a class of interconnected nonlinear uncertain systems," in *Proc. IEEE 51st Conf. Decision Control*, Dec. 2012, pp. 586–591.
- [30] Q. Zhang and X. Zhang, "Distributed sensor fault diagnosis in a class of interconnected nonlinear uncertain systems," *Annu. Rev. Control*, vol. 37, no. 1, pp. 170–179, 2013.
- [31] A. B. Trunov and M. M. Polycarpou, "Automated fault diagnosis in nonlinear multivariable systems using a learning methodology," *IEEE Trans. Neural Netw.*, vol. 11, no. 1, pp. 91–101, Jan. 2000.
- [32] X. Zhang, T. Parisini, and M. M. Polycarpou, "Sensor bias fault isolation in a class of nonlinear systems," *IEEE Trans. Autom. Control*, vol. 50, no. 3, pp. 370–376, Mar. 2005.
- [33] R. Dunia and S. J. Qin, "Joint diagnosis of process and sensor faults using principal component analysis," *Control Eng. Pract.*, vol. 6, no. 4, pp. 457–469, 1998.
- [34] K. Salahshoor, M. Mosallaei, and M. Bayat, "Centralized and decentralized process and sensor fault monitoring using data fusion based on adaptive extended Kalman filter algorithm," *Measurement*, vol. 41, no. 10, pp. 1059–1076, 2008.
- [35] X. Zhang, M. M. Polycarpou, and T. Parisini, "Design and analysis of a fault isolation scheme for a class of uncertain nonlinear systems," *Annu. Rev. Control*, vol. 32, no. 1, pp. 107–121, 2008.
- [36] X. Zhang, Q. Zhang, and N. Sonti, "Diagnosis of process faults and sensor faults in a class of nonlinear uncertain systems," *J. Syst. Eng. Electron.*, vol. 22, no. 1, pp. 22–32, Feb. 2011.
- [37] H. A. Talebi, K. Khorasani, and S. Tafazoli, "A recurrent neural-network-based sensor and actuator fault detection and isolation for nonlinear systems with application to the satellite's attitude control subsystem," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 45–60, Jan. 2009.
- [38] B. T. Thumati and G. R. Halligan, "A novel fault diagnostics and prediction scheme using a nonlinear observer with artificial immune system as an online approximator," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 3, pp. 569–578, May 2013.
- [39] R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault diagnosis with overlapping decompositions: An adaptive approximation approach," *IEEE Trans. Autom. Control*, vol. 54, no. 4, pp. 794–799, Apr. 2009.
- [40] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Adaptive approximation for multiple sensor fault detection and isolation of nonlinear uncertain systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 137–153, Jan. 2014.
- [41] C. Keliris, M. M. Polycarpou, and T. Parisini, "A distributed fault diagnosis approach utilizing adaptive approximation for a class of interconnected continuous-time nonlinear systems," in *Proc. IEEE 53rd Annu. Conf. Control Decision Conf.*, Dec. 2014, pp. 6536–6541.
- [42] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.
- [43] J. A. Farrell and M. M. Polycarpou, *Adaptive Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. New York, NY, USA: Wiley, 2006.
- [44] M. Powell, "The theory of radial basis function approximation in 1990," in *Advances in Numerical Analysis: Wavelets, Subdivision Algorithms, and Radial Basis Functions*, vol. 2, W. Light, Ed. Oxford, U.K.: Oxford Univ. Press, 1992, pp. 105–210.
- [45] T. Chen, C. Wang, and D. J. Hill, "Rapid oscillation fault detection and isolation for distributed systems via deterministic learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 6, pp. 1187–1199, Jun. 2014.
- [46] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*, 1st ed. New York, NY, USA: Academic, 1975.
- [47] 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.



Christodoulos Keliris (M'15) received the Diploma (Hons.) degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2007, the M.Sc. (Hons.) degree in finance from Imperial College London, London, U.K., in 2008, and the Ph.D. degree in electrical engineering from the University of Cyprus, Nicosia, Cyprus, in 2015.

He was a Researcher in various European research and operational programmes. His current research interests include fault diagnosis for nonlinear systems, nonlinear control theory, adaptive learning, and intelligent systems.

Dr. Keliris is a Reviewer for various conferences and journals. His dissertation pricing barrier options received the Best M.Sc. Finance Dissertation Prize.



Marios M. Polycarpou (F'06) received the bachelor's degrees in computer science and electrical engineering from Rice University, Houston, TX, USA, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1989 and 1992, respectively.

He is currently a Professor of Electrical and Computer Engineering and the Director of the KIOS Research Center for Intelligent Systems and Networks with the University of Cyprus, Nicosia,

Cyprus. He has authored over 280 articles in refereed journals, edited books and refereed conference proceedings, and co-authored seven books. He holds six patents. His current research interests include intelligent systems and networks, adaptive and cooperative control systems, computational intelligence, fault diagnosis, and distributed agents.

Dr. Polycarpou was a recipient of the IEEE Neural Networks Pioneer Award for 2016. He has participated in more than 60 research projects/grants, funded by several agencies and industry in Europe and the U.S., including the prestigious European Research Council Advanced Grant. He has served as the President of the IEEE Computational Intelligence Society from 2012 to 2013. He has served as the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2004 to 2010.



Thomas Parisini (F'11) received the Ph.D. degree in electronic engineering and computer science from the University of Genoa, Genoa, Italy, in 1993.

He holds the Chair of Industrial Control with Imperial College London, London, U.K. Since 2001, he has been the Danieli Endowed Chair of Automation Engineering with the University of Trieste, Trieste, Italy. He has authored or co-authored over 250 research papers in archival journals, book chapters, and international conference proceedings. His current research interests include neural-network approximations for optimal control problems, fault diagnosis for nonlinear and distributed systems, and nonlinear model predictive control systems.

Dr. Parisini was a co-recipient of the IFAC Best Application Paper Prize of the *Journal of Process Control* (Elsevier) from 2011 to 2013 and of the 2004 Outstanding Paper Award of the IEEE TRANSACTIONS ON NEURAL NETWORKS. In 2012, he received a prestigious ABB Research Grant dealing with energy-autonomous sensor networks for self-monitoring industrial environments. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY and is currently serving as an Associate Editor of the *International Journal of Control*. He served as an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the IEEE TRANSACTIONS ON NEURAL NETWORKS, *Automatica*, and the *International Journal of Robust and Nonlinear Control*.