European Journal of Control (2009)1:29–44
© 2009 EUCA
DOI:10.3166/EJC.15.29–44

European Journal of Control

Actuator Fault Detection in Nonlinear Uncertain Systems Using Neural On-line Approximation Models

Rastko R. Selmic[1, *], Marios M. Polycarpou[2, **], and Thomas Parisini[3, ***]

[1]Department of Electrical Engineering, Louisiana Tech University, Louisiana USA; [2]Department of Electrical and Computer Engineering, University of Cyprus, Cyprus; [3]Department of Electrical, Electronic, and Computer Engineering, University of Trieste, Italy

This paper presents a methodology for actuator fault detection in unknown, input-affine, nonlinear systems using neural networks. Both state feedback and output feedback cases are considered. Neural net tuning algorithms are derived and fault identifiers are developed using the Lyapunov approach. The paper studies properties of the fault dynamics, the dynamics of a fault evolution process. The actuator-fault dynamics are analyzed and a rigorous detectability condition is given for actuator faults relating the actuator desired input signal, neural net-based observer sensitivity, and detectability time. Moreover, the issue of fault propagation through the system dynamics towards the measurable output is addressed and specific conditions under which such faults can be detected are proposed. Simulation results are presented to illustrate the effectiveness of the proposed technique.

Keywords: Fault detection, actuator, nonlinear systems, neural networks

1. Introduction

Early detection of faults in complex electromechanical systems with multiple actuators is essential not only for the prevention of cascading catastrophic failures, but also for enhanced system performance, availability and environmental safety. Early results in stochastic dynamical systems, and in particular linear stochastic models are given in [35]. A traditional technique for fault detection is to build a library of possible faults, based on experience, manufacturing data, and maintenance experience. In the case of an aircraft, for example, there are software packages for such fault detection [6]. Significant research effort has been concentrated on combining rule-based fault detection methods with other intelligent systems techniques such as neural networks (NNs) and fuzzy logic [1], [10], [28], [30].

Model-based fault detection methods compare estimated models to a nominal system model [27]. The error between the two models is a measure of the deviation between estimated and nominal models, which provides an indication of whether or not a fault has occurred in the system. The main drawback of model-based approaches is the requirement of an accurate system model, which may not be available in many practical applications. NNs and fuzzy logic systems are computational intelligence models that can overcome some of these restrictive requirements. The main advantage of NNs is their nonlinear approximation and adaptive learning properties [12], [16] that provide the capability to learn abnormalities and failures from actual monitoring data. If a fault is detected, NNs may classify or identify it without having detailed system models [27], [28], [30], [31]. Statistical analysis

*Correspondence to: R. Selmic, E-mail: rselmic@latech.edu
**E-mail: mpolycar@ucy.ac.cy
***E-mail: parisini@units.it

Received 6 December 2007; Accepted 9 September 2008
Recommended by Z.P. Jiang, A.J. Van der Schaft

based on the Lyapunov function for fault-tolerant control systems is described in [22] and a fault-tolerant control system design is presented in [23]. An optimal stochastic fault detection filter was presented in [4] where an optimal filter gain for an observer is found such that fault detection error residual is affected mostly by target faults, and equivalent decentralized approach in [9] with application to a platoon of cars.

Recent work in adaptive control enabled an interesting application in actuator failure detection [34], in which Tao *et al.* developed a catalogue of adaptive controllers and compensators for systems with unknown actuator failures and with unknown system parameters. They provided actuator failure compensators for linear and nonlinear systems based on adaptive control techniques. In [37], [38], Zhang *et al.* developed combined detection, isolation, and compensation techniques of faults in nonlinear systems. More specifically, in these papers it is assumed that the nonlinear system model is partially known and that faults belong to a known class of faults. They developed fault detection methods and showed how to reconfigure the controller once the fault has been detected. Groundwork on multiple models/controllers in adaptive control was provided by Narendra's work [25] and this tool can effectively be used in actuator failure detection and compensation.

In the case when the system states are not fully available for measurement, as it is considered in this paper, it is generally required to develop an observer using output measurements. Output feedback control schemes that are based on NN-based observers are given in [3], [5], [13], [14], [15], [17], [36].

Much has been written on intelligent control using neural networks (NNs). With the universal approximation property and learning capability [12], [16], NNs have proven to be a powerful tool to control complex dynamic nonlinear systems with functional uncertainty [11]. The common control strategies with regards to NN are direct adaptive NN control method with guaranteed stability [8], indirect adaptive NN control based on identification [26], and static and dynamic inverse NN control [20]. In general, NNs are used to estimate the unknown nonlinear dynamics and/or functions and to compensate for them. Unlike the standard adaptive control schemes, NNs can also cope with nonlinear systems that cannot be linearly parameterizable. Many researchers [7], [19], [21], [29] have used NNs to synthesize the feedback linearization-based controllers using Lyapunov theory to guarantee the overall system stability, command following, and disturbance rejection.

This paper addresses the problem of actuator fault detection in *unknown*, input-affine nonlinear systems. The papers [36], [38] address fault detection, isolation, and compensation in nonlinear systems and they deal with partially known nonlinear systems. We use neural network to identify the *unknown*, nonlinear system dynamics while the actuator is healthy and later to detect actuator faults. Rigorous detectability conditions are given that relate the input control signal, neural net identifier parameters, and detectability of the faults. Results in [34] describe robust adaptive compensators for actuators faults and do not deal with detectability and isolation of faults. Polycarpou and Trunov [28] developed fault detectability conditions for nonlinear systems with adaptive learning methods.

In the paper, we extend the results presented in [28] in the following way: we consider the actuator fault detection problem, we refer to a different class of nonlinear systems, we consider both state feedback and output feedback cases, and we use neural networks learning algorithms for system identification. Indeed, the following fundamental questions in actuator fault detection are addressed: What kind of actuator faults can be detected? Under what conditions are faults detectable? If faults are not presently detectable, how should neural net identifier parameters be adjusted in order to detect the faults?

The paper is organized as follows. Section II formulates the problem, while Section III presents an actuator fault detection methodology and detectability analysis for the case of full state measurement. Section IV presents the corresponding design and analysis of fault detection for the output feedback case. Section V presents a simulation example to illustrate the design and theory developed in the paper, while Section VI presents some concluding remarks.

2. Problem Formulation

Let us consider a single-input single-output nonlinear system in observer canonical form given by

$$\begin{aligned} \dot{x} &= Ax + b[f(x) + g(x)u + d(t)] \\ y &= h^T x \end{aligned} \tag{1}$$

where $x = [x_1, x_2, \cdots, x_n]^T$ denotes the state vector, $f,\ g$: $\Re^n \to \Re$ are *unknown* smooth vector fields, $A \in \Re^{nxn}$, $b \in \Re^n$, $h \in \Re^n$, u is the output from the actuator, y is the output, and d represents a system disturbance. The control signal $u(t)$ is generated by the actuator, and the actuator input signal is $v(t)$. In case of a healthy system $u(t) = v(t)$; however, actuator performance may be affected by faults and actuator output might not follow the input signal. More specifically, we consider actuator fault model proposed by Tao *et al.* (see [34]), that is

$$u(t) = v(t) + \gamma(t)(\bar{u} - v(t)), \tag{2}$$

where the bounded actuator fault value is denoted by $\bar{u}$. The case $\gamma = 0$ models the actuator under healthy mode of behavior, whereas the case $\gamma = 1$ refers to the faulty in which case $u(t) = \bar{u}$. In the analysis, we will consider the general case where the actuator fault value turns out to be a time-varying function. We consider that the actuator is healthy at the beginning of operation.

We assume that the disturbance is uniformly bounded, that is:

$$\|d(t)\| \leq D_B, t \geq 0, \tag{3}$$

where D_B is a suitable positive scalar. It is assumed that the pair $(A,\ h^T)$ is observable. In the case of an observable canonical form the matrices are given by

$$A = \begin{bmatrix} -a_{n-1} & 1 & 0 & \dots & 0 \\ -a_{n-2} & 0 & 1 & \dots & 0 \\ & & \vdots & & \\ -a_1 & 0 & 0 & \cdots & 1 \\ -a_0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

$$b = \begin{bmatrix} b_{n-1} \\ b_{n-2} \\ \vdots \\ b_1 \\ b_0 \end{bmatrix}, h = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}. \tag{4}$$

Hence, it follows that

$$\begin{aligned} \dot{x} &= Ax + b\{f(x) + g(x)[v(t) \\ &\quad + \gamma(\bar{u}(t) - v(t))] + d(t)\} \\ y &= h^T x. \end{aligned} \tag{5}$$

It is worth noting that the combined system and actuator fault model in (5) can be viewed as a special case (emphasizing actuator faults) of the general model for faulty systems considered in [38], namely

$$\dot{x} = f(x) + \beta(t - T_0)g_1(x, v) + d(t), \tag{6}$$

where $\beta(t - T_0)$ represents the time profile of a fault occurring at some unknown time T_0 , and g_1 is a fault function.

The problem to be addressed in this paper is to develop an actuator fault detection method for both state and output feedback cases, and study actuator faults detectability conditions.

The following notation will be used in the paper. Given $A = [a_{ij}]$, $B \in \Re^{mxn}$, the Frobenius norm is defined by

$$\|A\|_F^2 = tr(A^T A) = \sum_{i,j} a_{ij}^2, \tag{7}$$

where $tr(\ )$ denotes the trace. The associated inner product is $\langle A, B\rangle_F = tr(A^T B)$. The Frobenius norm is compatible with the 2-norm so that $\|Ax\|_2 \leq \|A\|_F \|x\|_2$.

For a square matrix Y, with $\underline{\lambda}_Y$ and $\overline{\lambda}_Y$ we annotate the smallest and largest real part of eigenvalues of matrix Y, i.e. $\underline{\lambda}_Y = \mathrm{Re}(\lambda_{1Y}) \leq \mathrm{Re}(\lambda_{2Y}) \leq \ldots \leq \mathrm{Re}(\lambda_{nY}) = \overline{\lambda}_Y$.

3. Actuator Fault Detection, Fault Dynamics and Detectability: Measurable State Case

In this section, the proposed actuator fault detection (FD) approach will be described and analyzed with reference to the special case where the full state is available for measurement, whereas in the next section, the general case will be addressed. The FD approach combines a NN-based identification technique to approximate the unknown system's dynamics with a nonlinear observer-based method. More specifically, let us consider a NN estimator of system (1) given by (see [32])

$$\dot{\hat{x}} = Ax + \hat{W}_f^T \sigma_f(x) + \hat{W}_g^T \sigma_g(x) v(t) + T\hat{x} - Tx, \tag{8}$$

where the matrix T is chosen to be Hurwitz. For any symmetric positive definite matrix Q there exist symmetric positive definite matrix P such that

$$T^T P + PT = -Q. \tag{9}$$

We focus on a problem of actuator fault detection, while fault accommodation is not considered in this paper. In more precise terms, we assume that there exists a feedback control such that the state and control variables remain bounded before and after the occurrence of a possible actuator fault. Thus, we make the following assumption.

Assumption 1 (Stable Nominal System): *The closed-loop system is stable such that the control signal $v(t)$ is uniformly bounded before and after the occurrence of the fault, that is*

$$\|v(t)\| \leq V_B, \quad t \geq 0, \tag{10}$$

where V_B is a suitable positive scalar, and the system state belongs to a compact set $x \in X_B$.

In this paper we do not consider fault-tolerant control systems and fault compensators. For techniques how to compensate for actuator faults see for example [34].

Two-layers neural networks (NNs) are used as online neural approximation models where only the output layer is tunable. However, the approach can be used with any other type of observer that can be applied for nonlinear systems. Such NNs are linearly parameterized and can be represented by

$$y_{NN} = W^T \sigma(x_{NN}), \tag{11}$$

where the output layer weights are collected into W, $x_{NN} \in \Re^p$ denotes the NN input, $y_{NN} \in \Re^m$ denotes the NN output, and $\sigma(\cdot) \in \Re^L$ is the NN activation function, where L is a number of hidden layer nodes. Many well-known results indicate that any sufficiently smooth function can be approximated arbitrary closely on a compact set using such NN with appropriate weights [2]. An adaptive estimate of the ideal NN weights W will be denoted by $\hat{W}$.

Two NNs are introduced for the purpose of approximating the nonlinear functions $f_1(x) = bf(x)$ and $g_1(x) = bg(x)$. Hence, we can write

$$f_1(x) = W_f^T \sigma_f(x) + \varepsilon_f(x) \tag{12}$$

$$g_1(x) = W_g^T \sigma_g(x) + \varepsilon_g(x), \tag{13}$$

where W_f, W_g are some ideal *unknown* target NN weights, $\varepsilon_f(x)$, $\varepsilon_g(x)$ are the corresponding NN residual approximation errors, and f_1, g_1: are unknown smooth functions. The approximation errors are bounded on a compact set S by $\|\varepsilon_f(x)\| \le \varepsilon_{fM}$ and $\|\varepsilon_g(x)\| \le \varepsilon_{gM}$ for all $x\, \varepsilon\, s$. Corresponding numbers of hidden layer nodes are L_f and L_g.

Assumption 2 (NN Approximation Set): *Neural net approximation is valid over domain of interest, $S \supset X_B$ and bounds on ideal neural net weights W_{fM}, and W_{gM} are known.*

A controller design is not a topic of this paper and the nominal system is assumed to be stable. Therefore, for NN approximation to be valid, it is enough to assume that approximation region covers the compact set X_B. It is important to note that, while the system dynamics is in general unknown, Assumption 2 implies that some a priori knowledge about the system has to be assumed, i.e., knowledge about uniform bounds on the ideal NN weights. In practice, as will be clear later on in the paper, some rough estimate of the ideal NN weights will be needed in order to be able to compute bounds on the NN observer error during the system's healthy mode of behavior.

The NN identification error is now introduced:

$$e = x - \hat{x}. \tag{14}$$

It follows that

$$\dot{e} = Te + f_1(x) + g_1(x)u + bd(t) - \hat{W}_f^T \sigma_f(x) - \hat{W}_g^T \sigma_g(x)u(t) \tag{15}$$

$$\dot{e} = Te + \tilde{W}_f^T \sigma_f(x) + \varepsilon_f(x) + \tilde{W}_g^T \sigma_g(x)u(t) + \varepsilon_g(x)u + bd(t), \tag{16}$$

where $\tilde{W}_f$ and $\tilde{W}_g$ represent NN error weights defined as the difference between ideal NN weights and actual NN weights, i.e., $\tilde{W}_f = W_f - \hat{W}_f$ and $\tilde{W}_g = W_g - \hat{W}_g$.

The structure of the NN actuator fault observer is given in Fig. 1. The observer consists of two NNs that adaptively approximate the system unknown functions $f_1(x)$ and $g_1(x)$. Dashed arrows indicate that NN weights are tuned based on the identification error between measured and obtained system states. The tuning algorithm is derived based on Lyapunov approach with details given in the next theorem.

The following theorem provides NN tuning laws and a bound on the state observer error using *e*-modification type of adaptation (Narendra's seminal work [24]). While the system is healthy, the main goal is to approximate the unknown system dynamics maintaining a stable behavior of the observed error. Note that the NN tuning equations for the nonlinear system identifier are similar to Lewis' NN robotic control tuning algorithms [21].

Theorem 1 (Stable NN Observer Tuning Law): *Given the nonlinear system (1), and the NN observer (8), Assumptions 1–2, let the estimated NN weights be provided by the NN tuning algorithm*

$$\dot{\hat{W}}_f = C_f \sigma_f(x) e^T P - kC_f \|e\| \hat{W}_f \tag{17}$$

$$\dot{\hat{W}}_g = C_g \sigma_g(x) v(t) e^T P - kC_g \|e\| \hat{W}_g \tag{18}$$

with any constant matrices $C_f = C_f^T > 0$, $C_g = C_g^T > 0$, and a design parameter k. Then the state observer error e and the NN weight estimation errors $\tilde{W}_f$ and $\tilde{W}_g$ are uniformly ultimately bounded.

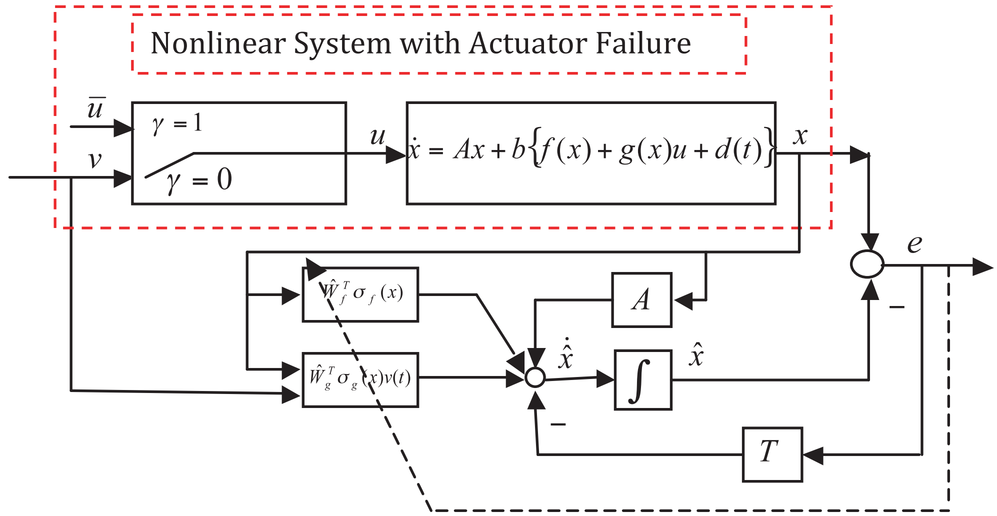

Fig. 1. NN system observer – fault identifier.

Proof: Select the candidate Lyapunov function as

$$L = \frac{1}{2}e^T Pe + \frac{1}{2}tr\left[\tilde{W}_f^T C_f^{-1}\tilde{W}_f\right] + \frac{1}{2}tr\left[\tilde{W}_g^T C_g^{-1}\tilde{W}_g\right]. \tag{19}$$

Then, its time-derivative is given by

$$\dot{L} = \frac{1}{2}\dot{e}^T Pe + \frac{1}{2}e^T P\dot{e} + tr\left[\tilde{W}_f^T C_f^{-1}\dot{\tilde{W}}_f\right] + tr\left[\tilde{W}_g^T C_g^{-1}\dot{\tilde{W}}_g\right] \tag{20}$$

$$\dot{L} = -\frac{1}{2}e^T Qe + e^T P\tilde{W}_f^T\sigma_f(x) + e^T P\varepsilon_f(x) + e^T P\tilde{W}_g^T\sigma_g(x)v(t) + e^T P\varepsilon_g(x)v + e^T Pbd(t) + tr\left[\tilde{W}_f^T C_f^{-1}\dot{\tilde{W}}_f\right] + tr\left[\tilde{W}_g^T C_g^{-1}\dot{\tilde{W}}_g\right] \tag{21}$$

$$\dot{L} = -\frac{1}{2}e^T Qe + tr\left[\tilde{W}_f^T(C_f^{-1}\dot{\tilde{W}}_f(x)e^T P\right] + tr\left[\tilde{W}_g^T(C_g^{-1}\dot{\tilde{W}}_g + \sigma_g(x)v(t)e^T P)\right] + e^T P\varepsilon_f(x) + e^T P\varepsilon_g(x)v + e^T Pbd(t). \tag{22}$$

By applying the tuning law equations (17) and (18), we have

$$\dot{L} = -\frac{1}{2}e^T Qe + tr\left[k\tilde{W}_f^T\|e\|\hat{W}_f\right] + tr\left[k\tilde{W}_g^T\|e\|\hat{W}_g\right] + e^T P\varepsilon_f(x) + e^T P\varepsilon_g(x)v + e^T Pbd(t) \tag{23}$$

$$\dot{L} = -\frac{1}{2}e^T Qe + k\|e\|tr\left[\tilde{W}_f^T\hat{W}_f\right] + k\|e\|tr\left[\tilde{W}_g^T\hat{W}_g\right] + e^T P\varepsilon_f(x) + e^T P\varepsilon_g(x)v + e^T Pbd(t) \tag{24}$$

$$\dot{L} = -\frac{1}{2}e^T Qe + k\|e\|tr\left[\tilde{W}_f^T(W_f - \tilde{W}_f)\right] + k\|e\|tr\left[\tilde{W}_g^T(W_g - \tilde{W}_g)\right] + e^T P\varepsilon_f(x) + e^T P\varepsilon_g(x)v + e^T Pbd(t). \tag{25}$$

Since Q is symmetric positive definite, then (25) yields

$$\dot{L} \le -\frac{1}{2}\underline{\lambda}_Q\|e\|^2 + k\|e\|\|\tilde{W}_f\|(W_{fM} - \|\tilde{W}_f\|) + k\|e\|\|\tilde{W}_g\|(W_{gM} - \|\tilde{W}_g\|) + e^T P\varepsilon_f(x) + e^T P\varepsilon_g(x)v + e^T Pbd(t). \tag{26}$$

Owing to the bounds on the disturbance and on the control variable, we obtain

$$\dot{L} \le -\frac{1}{2}\underline{\lambda}_Q\|e\|^2 + k\|e\|\|\tilde{W}_f\|(W_{fM} - \|\tilde{W}_f\|) + k\|e\|\|\tilde{W}_g\|(W_{gM} - \|\tilde{W}_g\|) + \|e\|\,\|P\|\left[\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B\right], \tag{27}$$

where $\bar{b}$ is the largest component of the vector b.

$$\dot{L} \le \|e\|\Big[-\frac{1}{2}\underline{\lambda}_Q\|e\| + k\|\tilde{W}_f\|(W_{fM} - \|\tilde{W}_f\|) + k\|\tilde{W}_g\|(W_{gM} - \|\tilde{W}_g\|) + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)\Big] \tag{28}$$

$$\dot{L} \le \|e\|\Big[-\frac{1}{2}\underline{\lambda}_Q\|e\| + k\|\tilde{W}_f\|(W_{fM} - \|\tilde{W}_f\|) + k\|\tilde{W}_g\|(W_{gM} - \|\tilde{W}_g\|) + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)\Big] \tag{29}$$

$$\dot{L} \le \|e\|\Big[-\frac{1}{2}\underline{\lambda}_Q\|e\| - k\Big(\|\tilde{W}_f\| - \frac{1}{2}W_{fM}\Big)^2 + \frac{k}{4}W_{fM}^2 - k\Big(\|\tilde{W}_g\| - \frac{1}{2}W_{gM}\Big)^2 + \frac{k}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)\Big] \tag{30}$$

Therefore, L is negative semi-definite if

$$-\frac{1}{2}\underline{\lambda}_Q\|e\| + \frac{k}{4}W_{fM}^2 + k\frac{}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B) < 0 \tag{31}$$

or

$$-k\Big(\|\tilde{W}_f\| - \frac{1}{2}W_{fM}\Big)^2 + \frac{k}{4}W_{fM}^2 + \frac{k}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B) < 0 \tag{32}$$

or

$$-k\Big(\|\tilde{W}_g\| - \frac{1}{2}W_{gM}\Big)^2 + \frac{k}{4}W_{fM}^2 + \frac{k}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B) < 0. \tag{33}$$

Equivalently, the L is negative semi-definite as long as

$$\|e\| > \frac{\frac{k}{2}W_{fM}^2 + \frac{k}{2}W_{gM}^2 + 2\|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{\underline{\lambda}_Q} \tag{34}$$

or

$$\|\tilde{W}_f\| > \sqrt{\frac{\frac{k}{4}W_{fM}^2 + \frac{k}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{k}} + \frac{1}{2}W_{fM} \tag{35}$$

or

$$\|\tilde{W}_g\| > \sqrt{\frac{\frac{k}{4}W_{fM}^2 + \frac{k}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{k}} + \frac{1}{2}W_{gM}. \tag{36}$$

The bounds on the tracking error and NN weights error are then given by

$$e_B = \frac{\frac{k}{2}W_{fM}^2 + \frac{k}{2}W_{gM}^2 + 2\|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{\underline{\lambda}_Q} \tag{37}$$

$$\tilde{W}_{fB} = \sqrt{\frac{\frac{k}{4}W_{fM}^2 + \frac{k}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{k}} + \frac{1}{2}W_{fM} \tag{38}$$

$$\tilde{W}_{gB} = \sqrt{\frac{\frac{k}{4}W_{fM}^2 + \frac{k}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{k}} + \frac{1}{2}W_{gM} \tag{39}$$

thus ending the proof.

It is worth noting that (37), (38), and (39) provide explicit bounds (though conservative as they depend on the bounds on the NN weights) on the observer state error and NN weights approximation errors when the system undergoes healthy mode of behavior. The bound (37) depends on $W_{fM,}$ and W_{gM} due to NN approximation property of unknown functions in the system model and shows that the convergence region can be scaled with the unknown functions in the system model.

Theorem 1 guarantees boundedness of the NN identification error and NN weights if $u(t) = v(t)$. Moreover, we are interested in observing the identification error when an actuator fault occurs. Because of that, NN learning algorithm has been modified to prevent NN weights to become unbounded once the fault occurs.

Note that the bounds on *actual* NN weights can be derived from (38) and (39) as

$$\|\hat{W}_f\| - \|W_f\| \le \|\tilde{W}_f\|, \quad \|\hat{W}_g\| - \|W_g\| \le \|\tilde{W}_g\| \tag{40}$$

$$\|\hat{W}_f\| \le \sqrt{\frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{k}} + \frac{3}{2}W_{fM} = \hat{W}_{fB} \tag{41}$$

$$\|\hat{W}_g\| \le \sqrt{\frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \|P\|(\varepsilon_{fM} + \varepsilon_{gM}V_B + \bar{b}D_B)}{k}} + \frac{3}{2}W_{gM} = \hat{W}_{gB}. \tag{42}$$

Modified NN Weights Tuning Law: We have already discussed the assumption that the NN will be tuned during an initial (healthy) phase of the system operation. Theorem 1 guarantees NN weight boundedness while the system is healthy. In case of an actuator fault, the NN weights will be limited by the following algorithm modification that is equivalent to saturation-based tuning laws

$$\dot{\hat{W}}_f = \begin{cases} C_f\sigma_f(x)e^TP - kS\|e\|\hat{W}_f\,, \; for\ \hat{W}_f \le \hat{W}_{fB} \\ 0\,, \; for\ \hat{W}_f > \hat{W}_{fB} \end{cases} \tag{43}$$

$$\dot{\hat{W}}_g = \begin{cases} C_g\sigma_g(x)v(t)e^TP - kT\|e\|\hat{W}_g\,, for\ \hat{W}_g \le \hat{W}_{gB} \\ 0\,, for\ \hat{W}_g > \hat{W}_{gB} \end{cases} \tag{44}$$

Assumption 3: *An actuator fault in nonlinear system (1) has occurred after the NN identification error is settled below the bounds given by Theorem 1.*

Remark 1: *Assumption 3 describes a natural requirement that the system must be healthy until NN errors settle below the bounds given by Theorem 1 in order to be able to detect the potential fault. The identification is while assuming that the actuator is healthy, i.e. $u(t) = v(t)$.*

Let us now consider when there is a sudden actuator fault. The closed-loop error dynamics is given by

$$\begin{aligned}\dot{e} = {} & bf(x) + bg(x)\bar{u}(t) + bg(x)v(t) \\ & - bg(x)v(t) + bd(t) \\ & + Tx - T\hat{x} - \hat{W}_f^T\sigma_f(x) - \hat{W}_g^T\sigma_g(x)v(t),\end{aligned} \tag{45}$$

$$\begin{aligned}\dot{e} = {} & Te + \tilde{W}_f^T\sigma_f(x) + \varepsilon_f(x) + \tilde{W}_g^T\sigma_g(x)v(t) \\ & + \varepsilon_g(x)v(t) \\ & + bd(t) + bg(x)(\bar{u}(t) - v(t)).\end{aligned} \tag{46}$$

After the initial time period when the system has been failure-free, it is important to study under what conditions the failure or anomaly in the system can be detected, how much time it will take to possibly detect the fault, what are the conditions on the applied input signal for fault to be detectable, etc. [28]. Note that the bound on the failure-free state observer error is given by

$$\|e\| \le e_B, \tag{47}$$

where e_B is given by (37), and that the system is considered faulty if $\|e\| > e_B$.

The system (46) can be represented in the following format

$$\dot{e} = Te + h(\tilde{W}_f^T, \tilde{W}_g^T, x, t) + bg(x)(\bar{u}(t) - v(t)), \tag{48}$$

where the function $h(\,)$ is given by

$$\begin{aligned}h(X, Y, x, t) = {} & X\sigma_f(x) + \varepsilon_f(x) + Y\sigma_g(x)v(t) \\ & + \varepsilon_g(x)v(t) + bd(t).\end{aligned} \tag{49}$$

Consider the system (48) and assume that an actuator fault occurs at $t = t_0$. The solution of the system (48) is given by

$$e(t) = \exp(T(t - t_0))e(t_0) + \int_{t_0}^{t} \exp(T(t - \tau)) \times \left[h(\tilde{W}_f^T, \tilde{W}_g^T, x, \tau) + bg(x)(\bar{u} - v)\right]d\tau \tag{50}$$

Taking the norm of the fault error signal yields

$$\|e(t)\| = \left\| \exp(T(t - t_0))e(t_0) + \int_{t_0}^{t} \exp(T(t - \tau)) \times \left[h(\tilde{W}_f^T, \tilde{W}_g^T, x, \tau) + bg(x)(\bar{u} - v)\right]d\tau \right\|. \tag{51}$$

The next lemma summarizes these results and provides a condition for actuator fault detectability.

Lemma 1 (Detectability of Actuator Faults): *Given Assumptions 1–3 and a NN observer from Theorem 1, a fault in the system actuator can be detected after time interval T_d if the following condition is satisfied*

$$\left\| \int_{t_0}^{t_0+T_d} \exp(T(t_0+T_d-\tau))bg(x)(\bar{u}-v)d\tau \right\| > 2e_B + \frac{C_1}{|\bar{\lambda}_T|}\left[1-\exp(\bar{\lambda}_T T_d)\right] \tag{52}$$

where $C_1 = \tilde{W}_{fB}L_f + \varepsilon_{fM} + \tilde{W}_{gB}V_B L_g + \varepsilon_{gM}V_B + \bar{b}D_B$.

Proof: Using $\|e(t_0)\| \leq e_B$, from (51) if follows that the fault detectability condition is given by

$$\left\| \int_{t_0}^{t} \exp(T(t-\tau))\left[h(\tilde{W}_f^T, \tilde{W}_g^T, x, \tau) + bg(x)(\bar{u}-v)\right]d\tau \right\| > 2e_B \tag{53}$$

$$\left\| \int_{t_0}^{t} \exp(T(t-\tau))bg(x)(\bar{u}-v)d\tau \right\| > 2e_B + \left\| \int_{t_0}^{t} \exp(T(t-\tau))h(\tilde{W}_f^T, \tilde{W}_g^T, x, \tau)d\tau \right\| \tag{54}$$

$$\left\| \int_{t_0}^{t} \exp(T(t-\tau))bg(x)(\bar{u}-v)d\tau \right\| > 2e_B + \frac{C_1}{|\bar{\lambda}_T|}\left[1-\exp(\bar{\lambda}_T(t-t_0))\right], \tag{55}$$

with a constant $C_1 = \tilde{W}_{fB}L_f + \varepsilon_{fM} + \tilde{W}_{gB}V_B L_g + \varepsilon_{gM}V_B + \bar{b}D_B$.

The condition (52) provides rigorous justification of fault detectability based on the intuitive concept that for the fault to be detected there should be "sufficient" discrepancy between the desired input signal and failed actuator values. The result also relates the time before fault has been detected and the NN estimator parameters. The result in Lemma 1 states that $\bar{u}(t)$ and the control signal $v(t)$ need to be sufficiently different over a time period of interest for a fault to be detected in the system dynamics (5). For instance, if a robot actuator fails and gets stuck in one position, the user would be able to detect the problem only if the control signal sufficiently differs from the failed value. If the failed actuator is stuck and the control signal is coincidentally at, or close to, that failed position, then the fault may not have any significant effect on system states as Theorem 1 shows, and therefore, may not be detected.

Note that Lemma 1 requires knowledge of the function $g(x)$. Therefore, the result has more of a theoretical value explaining when and which fault can be detected. If a user knows an approximate value of the function $g(x)$ then Lemma 1 can be used to approximate fault detectability time as a function of observer parameters and actuator control signals. The NN approximation of $g(x)$ can also be used for a sufficient fault detectability condition [32]

$$\left\| \int_{t_0}^{t_0+T_d} \exp(T(t_0+T_d-\tau))b\hat{W}_g^T\sigma_g(\bar{u}-v)d\tau \right\| > (W_{gM}+\varepsilon_{gM}) \int_{t_0}^{t_0+T_d} \exp(T(t_0+T_d-\tau)) \|b(\bar{u}-v)\|d\tau + 2e_B + \frac{C_1}{|\bar{\lambda}_T|}\left[1-\exp(\bar{\lambda}_T T_d)\right] \tag{56}$$

The next section discusses the case when a system state is not available for measurement. The fault propagates through dynamics of the system and can be detected by observing the system output.

4. Actuator Fault Detection, Fault Dynamics and Detectability: Output Measurement Case

A neural net learning algorithm is considered as an identification method for the nonlinear system (1). More specifically, a NN system observer is given by (see [31])

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + b\left[\hat{W}_f^T\sigma_f(\hat{x}) + \hat{W}_g^T\sigma_g(\hat{x})v(t) + r_0(t)\right] + T(y - h^T\hat{x}) \\ \hat{y} &= h^T\hat{x} \end{aligned} \tag{57}$$

where $\hat{x}$ and $\hat{y}$ are estimates of the state vector x and the system output y respectively. Two NNs are used to approximate the nonlinear functions $f(x)$ and $g(x)$:

$$f(x) = W_f^T \sigma_f(x) + \varepsilon_f(x) \tag{58}$$

$$g(x) = W_g^T \sigma_g(x) + \varepsilon_g(x), \tag{59}$$

where W_f, W_g are some ideal target NN weights, and $\varepsilon_f(x)$, $\varepsilon_g(x)$ are NN approximation errors. In (57), the robustifying term $r_0(t)$ is added to accommodate the residual functional approximation error and the unknown disturbances. The observer gain matrix T is chosen so that $A_s = A - Th^T$ is stable. As in the previous case, we need the standard assumption of bounded ideal NN weights.

The structure of the NN actuator fault observer is then given in Fig. 2.

We assume that the system controller has been designed and that control signals are already bounded (see Assumption 2). The error between the real system output and the NN observer is given by $\tilde{y} = y - \hat{y}$, and the identification error between real states and observer states is $\tilde{x} = x - \hat{x}$. Assuming that the initial mode of behavior of the system is healthy, the error dynamics is given by:

$$\begin{aligned} \dot{\tilde{x}} &= Ax + b\Big[W_f^T\sigma_f(x) + W_g^T\sigma_g(x)v(t) \\ &\quad + \varepsilon_f(x) + \varepsilon_g(x)v(t) + d(t)\Big] \\ &\quad - A\hat{x} - b\Big[\hat{W}_f^T\sigma_f(\hat{x}) + \hat{W}_g^T\sigma_g(\hat{x})v(t) \\ &\quad + r_0(t)\Big] - T(y - h^T\hat{x}) \\ \tilde{y} &= h^T\tilde{x}. \end{aligned} \tag{60}$$

Lemma 2: *The system error dynamics (60) can be written in the following form*

$$\begin{aligned} \dot{\tilde{x}} &= A_s\tilde{x} + b\Big[\tilde{W}_f^T\sigma_f(\hat{x}) + \tilde{W}_g^T\sigma_g(\hat{x})v(t) \\ &\quad + c_2(t) - r_0(t)\Big] \\ \tilde{y} &= h^T\tilde{x} \end{aligned} \tag{61}$$

or, equivalently,

$$\begin{aligned} \tilde{y} &= H(s)\Big[\tilde{W}_f^T\sigma_f(\hat{x}) + \tilde{W}_g^T\sigma_g(\hat{x})v(t) \\ &\quad + c_2(t) - r_0(t)\Big], \end{aligned} \tag{62}$$

where $H(s)$ is the transfer function associated with the triple (A_s, b, h), and the signal $c_2(t)$ is bounded by a constant C_2:

$$\begin{aligned} \|c_2(t)\| &\le W_{fM} + W_{gM}V_B + \varepsilon_{fM} + \varepsilon_{gM}V_B \\ &\quad + D_B = C_2. \end{aligned} \tag{63}$$

Proof: The system error dynamics is equivalent to

$$\begin{aligned} \dot{\tilde{x}} &= A\tilde{x} + b\Big[W_f^T\sigma_f(x) + W_g^T\sigma_g(x)v(t) + \varepsilon_f(x) \\ &\quad + \varepsilon_g(x)v(t) + d(t)\Big] \\ &\quad - b\Big[\hat{W}_f^T\sigma_f(\hat{x}) + \hat{W}_g^T\sigma_g(\hat{x})v(t) + r_0(t)\Big] \\ &\quad - T(y - h^T\hat{x}) \\ \tilde{y} &= h^T\tilde{x} \end{aligned} \tag{64}$$

$$\begin{aligned} \dot{\tilde{x}} &= A_s\tilde{x} + b\{\tilde{W}_f^T\sigma_f(\hat{x}) + W_f^T\tilde{\sigma}_f + \tilde{W}_g^T\sigma_g(\hat{x})v(t) \\ &\quad + W_g^T\tilde{\sigma}_g v(t) + \varepsilon_f(x) + \varepsilon_g(x)v(t) + d(t) - r_0(t)\} \\ \tilde{y} &= h^T\tilde{x}. \end{aligned} \tag{65}$$

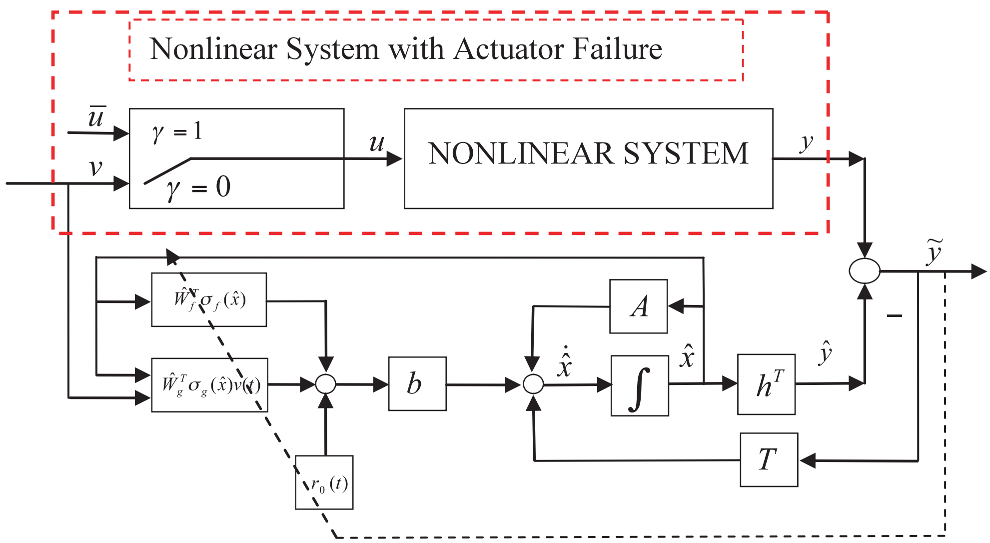

Fig. 2. NN system observer and fault identifier for output feedback case.

The system error dynamics (65) can be written in the following form

$$\dot{\tilde{x}} = A_s\tilde{x} + b\Big[\tilde{W}_f^T\sigma_f(\hat{x}) + \tilde{W}_g^T\sigma_g(\hat{x})v(t) + c_2(t) - r_0(t)\Big]$$
$$\tilde{y} = h^T\tilde{x}, \tag{66}$$

where the signal $c_2(t)$ is given by

$$c_2(t) = W_f^T\tilde{\sigma}_f + W_g^T\tilde{\sigma}_g v(t) + \varepsilon_f(x) + \varepsilon_g(x)v(t) + d(t). \tag{67}$$

This function is bounded

$$\|c_2(t)\| \le C_2, \tag{68}$$

where C_2 is given by

$$C_2 = W_{fM} + W_{gM}V_B + \varepsilon_{fM} + \varepsilon_{gM}V_B + D_B, \tag{69}$$

and W_{fM}, and W_{gM} are maximum NN weights used in networks approximating functions $f(x)$ and $g(x)$.

The NN error weights are given by $\tilde{W}_f = W_f - \hat{W}_f$, $\tilde{W}_g = W_g - \hat{W}_g$, and the error in activation functions is given by $\tilde{\sigma}_f = \sigma_f(x) - \sigma_f(\hat{x})$, and $\tilde{\sigma}_g = \sigma_g(x) - \sigma_g(\hat{x})$.

Remark 2: *The transfer function $H(s)$ might be strictly positive real (SPR). If it is not SPR, then the dynamics can be written in the form*

$$\tilde{y} = H(s)H_L(s)H_L^{-1}(s)\Big[\tilde{W}_f^T\sigma_f(\hat{x}) + \tilde{W}_g^T\sigma_g(\hat{x})v(t) + c_2(t) - r_0(t)\Big], \tag{70}$$

where $H_L^{-1}(s)$ is a proper transfer function with stable poles such that $H(s)H_L(s)$ is a SPR transfer function [17]. In the following analysis variables with "-" indicate the signal filtered by $H_L^{-1}(s)$.

The following theorem provides NN tuning laws and a bound on the state observer error using e-modification type of adaptation [21], [24].

Theorem 2 (Stable NN Observer Tuning Law): *Given the nonlinear system (1), and the NN observer (57), let the error dynamics (66) be SPR and choose the robustifying term as*

$$r_o = k_r\frac{\tilde{y}}{|\tilde{y}|} \tag{71}$$

with a scalar k_r. Let the estimated NN weights be provided by the NN tuning algorithm

$$\dot{\hat{W}}_f = C_f\bar{\sigma}_f(\hat{x})\tilde{y} - kC_f|\tilde{y}|\hat{W}_f \tag{72}$$

$$\dot{\hat{W}}_g = C_g\bar{\sigma}_g(\hat{x})v(t)\tilde{y} - kC_g|\tilde{y}|\hat{W}_g \tag{73}$$

with any constant, symmetric matrices $C_f = C_f^T > 0$, $C_g = C_g^T > 0$, and a suitably small design parameter k. Then the state observer error $\tilde{x}$ and the NN weight estimation errors $\tilde{W}_f$ and $\tilde{W}_g$ are uniformly ultimately bounded.

Proof: Select the Lyapunov function as

$$L = \frac{1}{2}\tilde{x}^TP\tilde{x} + \frac{1}{2}tr\Big[\tilde{W}_f^TC_f^{-1}\tilde{W}_f\Big] + \frac{1}{2}tr\Big[\tilde{W}_g^TC_g^{-1}\tilde{W}_g\Big]. \tag{74}$$

Then its time-derivative is given by

$$\dot{L} = \frac{1}{2}(\dot{\tilde{x}}^TP\tilde{x} + \tilde{x}^TP\dot{\tilde{x}}) + tr\Big[\tilde{W}_f^TC_f^{-1}\dot{\tilde{W}}_f\Big] + tr\Big[\tilde{W}_g^TC_g^{-1}\dot{\tilde{W}}_g\Big]. \tag{75}$$

Using a Kalman-Yakubovich Lemma [33] there exist positive definite matrices P and Q such that

$$A_s^TP + PA_s = -Q$$
$$Pb = h. \tag{76}$$

Therefore, one has

$$\dot{L} = -\frac{1}{2}\tilde{x}^TQ\tilde{x} + \tilde{x}^TPb\tilde{W}_f^T\bar{\sigma}_f(\hat{x}) + \tilde{x}^TPb\tilde{W}_g^T\bar{\sigma}_g(\hat{x})v(t) + \tilde{x}^TPbc_2(t) - \tilde{x}^TPbr_0(t) + tr\Big[\tilde{W}_f^TC_f^{-1}\dot{\tilde{W}}_f\Big] + tr\Big[\tilde{W}_g^TC_g^{-1}\dot{\tilde{W}}_g\Big] \tag{77}$$

$$\dot{L} = -\frac{1}{2}\tilde{x}^TQ\tilde{x} + \tilde{x}^TPbc_2(t) - \tilde{x}^TPbr_0(t) + tr\Big[\tilde{W}_f^T(C_f^{-1}\dot{\tilde{W}}_f + \bar{\sigma}_f(\hat{x})(Pb)^T\tilde{x})\Big] + tr\Big[\tilde{W}_g^T(C_g^{-1}\dot{\tilde{W}}_g + \bar{\sigma}_g(\hat{x})v(t)(Pb)^T\tilde{x})\Big] \tag{78}$$

$$\dot{L} = -\frac{1}{2}\tilde{x}^TQ\tilde{x} + \tilde{x}^TPbc_2(t) - \tilde{x}^TPbr_0(t) + tr\Big[\tilde{W}_f^T(C_f^{-1}\dot{\tilde{W}}_f + \bar{\sigma}_f(\hat{x})\tilde{y})\Big] + tr\Big[\tilde{W}_g^T(C_g^{-1}\dot{\tilde{W}}_g + \bar{\sigma}_g(\hat{x})v(t)\tilde{y})\Big]. \tag{79}$$

Applying the proposed tuning law yields

$$\dot{L} = -\frac{1}{2}\tilde{x}^T Q \tilde{x} + \tilde{x}^T Pbc_2(t) - \tilde{x}^T Pbr_0(t) + k|\tilde{y}|tr\left[\tilde{W}_f^T (W_f - \tilde{W}_f)\right] + k|\tilde{y}|tr\left[\tilde{W}_g^T (W_g - \tilde{W}_g)\right]. \quad (80)$$

The time-derivative of the Lyapunov satisfies

$$\dot{L} \le -\frac{1}{2}\lambda_Q \|\tilde{x}\|^2 + \tilde{y}C_2 - \tilde{y}r_0(t) + k|\tilde{y}|\|\tilde{W}_f\|\left(W_{fM} - \|\tilde{W}_f\|\right) + k|\tilde{y}|\|\tilde{W}_g\|\left(W_{gM} - \|\tilde{W}_g\|\right). \quad (81)$$

Knowing that $|\tilde{y}| \le h_{\max}\|\tilde{x}\|$ where $h_{\max}$ is the maximum element of the vector h, we have

$$\dot{L} \le -\frac{1}{2}\lambda_Q \|\tilde{x}\|^2 + |\tilde{y}|C_2 - \tilde{y}r_0(t) + kh_{\max}\|\tilde{x}\|\|\tilde{W}_f\|\left(W_{fM} - \|\tilde{W}_f\|\right) + kh_{\max}\|\tilde{x}\|\|\tilde{W}_g\|\left(W_{gM} - \|\tilde{W}_g\|\right) \quad (82)$$

$$\dot{L} \le -\frac{1}{2}\lambda_Q \|\tilde{x}\|^2 + \|\tilde{x}\|C_2 - \tilde{y}r_0(t) - kh_{\max}\|\tilde{x}\|\left(\|\tilde{W}_f\| - \frac{1}{2}W_{fM}\right)^2 + kh_{\max}\|\tilde{x}\|\frac{1}{4}W_{fM}^2 - kh_{\max}\|\tilde{x}\|\left(\|\tilde{W}_g\| - \frac{1}{2}W_{gM}\right)^2 + kh_{\max}\|\tilde{x}\|\frac{1}{4}W_{gM}^2 \quad (83)$$

$$\dot{L} \le \|\tilde{x}\|\left\{-\frac{1}{2}\lambda_Q\|\tilde{x}\| + C_2 - k_r - kh_{\max}\left(\|\tilde{W}_f\| - \frac{1}{2}W_{fM}\right)^2 + kh_{\max}\frac{1}{4}W_{fM}^2 - kh_{\max}\left(\|\tilde{W}_g\| - \frac{1}{2}W_{gM}\right)^2 + kh_{\max}\frac{1}{4}W_{gM}^2\right\} \quad (84)$$

Therefore, L is negative semi-definite provided that

$$\|\tilde{x}\| > \frac{2C_2 - 2k_r + \frac{1}{2}kh_{\max}W_{fM}^2 + \frac{1}{2}kh_{\max}W_{gM}^2}{\lambda_Q}, \quad (85)$$

or

$$\|\tilde{W}_f\| > \sqrt{\frac{C_2 - k_r}{kh_{\max}} + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2} + \frac{1}{2}W_{fM}, \quad (86)$$

or

$$\|\tilde{W}_g\| > \sqrt{\frac{C_2 - k_r}{kh_{\max}} + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2} + \frac{1}{2}W_{gM}. \quad (87)$$

Thus, the bounds on the observer error and NN weights error are given by

$$\tilde{x}_B = \frac{2C_2 - 2k_r + \frac{1}{2}kh_{\max}W_{fM}^2 + \frac{1}{2}kh_{\max}W_{gM}^2}{\lambda_Q} \quad (88)$$

$$\tilde{W}_{fB} = \sqrt{\frac{C_2 - k_r}{kh_{\max}} + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2} + \frac{1}{2}W_{fM} \quad (89)$$

$$\tilde{W}_{gB} = \sqrt{\frac{C_2 - k_r}{kh_{\max}} + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2} + \frac{1}{2}W_{gM}. \quad (90)$$

Equations (88), (89), (90) provide explicit bounds on the observer error and NN weights approximation errors. Note that (88) represents the bound on the observer error under healthy system assumption. It is also a conservative bound since it includes the NN weights bounds. The NN tuning algorithms for nonlinear system identifier (72), (73) are similar to Lewis' NN robotic control tuning algorithms found in references [20], [21]. The conservative estimate on the bounds of the observer output error is given by $\tilde{y}_B = h_{\max}\tilde{x}_B$.

Remark 3: *If the transfer function $H(s)$ is not SPR, then signals in the tuning law are filtered by $H_L^{-1}(s)$. Stability analysis is then carried out in the same fashion but in terms of another state-space realization of the $H(s)H_L(s)$. In this case a similar bound on $\tilde{y}_B$ can also be found [17].*

Similarly as in (41), (42) the actual NN weight bounds are given by

$$\|\hat{W}_f\| \le \sqrt{\frac{C_2 - k_r}{kh_{\max}} + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2} + \frac{3}{2}W_{fM} = \hat{W}_{fB} \quad (91)$$

$$\|\hat{W}_g\| \le \sqrt{\frac{C_2 - k_r}{kh_{\max}} + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2} + \frac{3}{2}W_{gM} = \hat{W}_{gB}. \quad (92)$$

We have already discussed the assumption that the NN will be tuned during an initial (healthy) phase of the system operation. Theorem 2 guarantees NN weight boundedness while the system is healthy. The NN weights estimation errors will be limited by the following algorithm modification

$$\hat{W}_f = \begin{cases} C_f\bar{\sigma}_f(\hat{x})\tilde{y} - kC_f|\tilde{y}|\hat{W}_f\,,\ for\ \hat{W}_f \le \hat{W}_{fB} \\ 0\,,\ for\ \hat{W}_f > \hat{W}_{fB} \end{cases} \quad (93)$$

$$\hat{W}_g = \begin{cases} C_g\bar{\sigma}_g(\hat{x})v(t)\tilde{y} - kC_g|\tilde{y}|\hat{W}_g\,,\ for\ \hat{W}_g \le \hat{W}_{gB} \\ 0\,,\ for\ \hat{W}_g > \hat{W}_{gB}. \end{cases} \quad (94)$$

If the system actuator fails suddenly, the closed-loop error dynamics is given by

$$\begin{aligned} \tilde{x} &= A_s\tilde{x} + b\{\tilde{W}_f^T\sigma_f(\hat{x}) + \tilde{W}_g^T\sigma_g(\hat{x})v(t) + c_2(t) \\ &\quad -r_0(t) + g(x)(\bar{u}(t) - v(t))\} \\ \tilde{y} &= h^T\tilde{x}. \end{aligned} \quad (95)$$

After the initial time period when the actuator has been failure-free, similar to the case where the state is available for measurement, it is important to study under what conditions we will be able to detect the failure or anomaly in the system, how long it will take to possibly detect the fault, what are the conditions on the applied input signal at the actuator for the fault to be detectable, etc. The error dynamics can be written as

$$\begin{aligned} \tilde{x} &= A_s\tilde{x} + \psi(\tilde{W}_f^T, \tilde{W}_g^T, \hat{x}, t) + bg(x)(\bar{u}(t) - v(t)) \\ \tilde{y} &= h^T\tilde{x}, \end{aligned} \quad (96)$$

where the function $\psi(.)$ is given by

$$\psi(X, Y, \hat{x}, t) = b\{X\sigma_f(\hat{x}) + Y\sigma_g(\hat{x})v(t) + c_2(t) - r_0(t)\}. \quad (97)$$

Assuming that an actuator fault occurs at time instant $t = t_0$, the solution is given by

$$\begin{aligned} \tilde{x}(t) &= \exp(A_s(t - t_0))\tilde{x}(t_0) \\ &\quad + \int_{t_0}^{t} \exp(A_s(t - \tau))\Big[\psi(\tilde{W}_f^T, \tilde{W}_g^T, \hat{x}, \tau) \\ &\quad + bg(x)(\bar{u}(\tau) - v(\tau))\Big]d\tau \\ \tilde{y} &= h^T\tilde{x} \end{aligned} \quad (98)$$

It follows that

$$\begin{aligned} \|\tilde{x}(t)\| &= \Big\| \exp(A_s(t - t_0))\tilde{x}(t_0) \\ &\quad + \int_{t_0}^{t} \exp(A_s(t - \tau))\Big[\psi(\tilde{W}_f^T, \tilde{W}_g^T, \hat{x}, \tau) \\ &\quad + bg(x)(\bar{u}(\tau) - v(\tau))]d\tau \Big\| \\ \tilde{y} &= h^T\tilde{x} \end{aligned} \quad (99)$$

The next lemma summarizes these results and provides rigorous conditions for actuator fault detectability.

Lemma 3 (Detectability of Actuator Faults): *Given Assumptions 1–3 and a NN observer from Theorem 2, a fault in the system actuator can be detected after time interval T_d if the following condition is satisfied*

$$\left\| \int_{t_0}^{t_0+T_d} \exp(A_s(t_0 + T_d - \tau))bg(x)(\bar{u}(\tau) - v(\tau))d\tau \right\| > 2\tilde{x}_B + \frac{C_3}{|\bar{\lambda}_{A_s}|}\left[1 - \exp(\bar{\lambda}_{A_s}T_d)\right]. \quad (100)$$

where $C_3 = b_{\max}(\tilde{W}_{fB}L_f + \tilde{W}_{gB}V_BL_g + C_2 + k_r)$.

Proof: Similarly as in (53), since $\|\tilde{x}(t_0)\| \le \tilde{x}_B$, fault detectability condition is given by

$$\left\| \int_{t_0}^{t} \exp(A_s(t - \tau))\Big[\psi(\tilde{W}_f^T, \tilde{W}_g^T, \hat{x}, \tau) + bg(x)(\bar{u}(\tau) - v(\tau))\Big]d\tau \right\| > 2\tilde{x}_B. \quad (101)$$

$$\left\| \int_{t_0}^{t} \exp(A_s(t - \tau))bg(x)(\bar{u}(\tau) - v(\tau))d\tau \right\| > 2\tilde{x}_B + \left\| \int_{t_0}^{t} \exp(A_s(t - \tau))\psi(\tilde{W}_f^T, \tilde{W}_g^T, \hat{x}, \tau)d\tau \right\| \quad (102)$$

$$\left\| \int_{t_0}^{t} \exp(A_s(t-\tau))bg(x)(\bar{u}(\tau)-v(\tau))d\tau \right\| > 2\tilde{x}_B + \frac{C_3}{|\bar{\lambda}_{A_s}|}\left[1-\exp(\bar{\lambda}_{A_s}(t-t_0))\right] \quad (103)$$

The above condition provides rigorous justification of fault detectability based on the intuitive concept that for the fault to be detected there should be "sufficient" discrepancy between the desired input signal and failed actuator values. The result also relates the fault value, fault detectability time, and identifier parameters.

Equivalently as in Section III, the next inequality relates observer parameters, i.e. NN weights, with fault detectability and the actuator control signal [32]. A fault in the system actuator can be detected after time interval T_d if the following condition is satisfied

$$\left\| \int_{t_0}^{t_0+T_d} \exp(A_s(t_0+T_d-\tau))b\hat{W}_g^T\sigma_g(\bar{u}-v)d\tau \right\| > \left(W_{gM}+\varepsilon_{gM}\right)\int_{t_0}^{t_0+T_d} \exp(A_s(t_0+T_d-\tau)) \times \|b(\bar{u}-v)\|d\tau + 2\tilde{x}_B + \frac{C_3}{|\bar{\lambda}_{A_s}|}\left[1-\exp(\bar{\lambda}_{A_s}T_d)\right] \quad (104)$$

5. Simulation Examples

The simulation was carried out to illustrate the performance of the proposed NN fault identifier and also to verify detectability conditions. A state feedback case is considered in a nonlinear system with unknown functions $f(x)$ and $g(x)$. The numerical simulation program was written in visual C++ and Matlab. The integration method is the fourth order Runge-Kutta algorithm with an integration time step interval of 0.001.

We consider the second order nonlinear system given by

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -5x_1^3 - 2x_2 + (x_1+1)u(t) \\ y &= x_1, \end{aligned} \quad (105)$$

where the fault model is given by

$$u(t) = v(t) + \gamma(\bar{u}(t) - v(t)). \quad (106)$$

The NN observer given by Theorem 1 consists of two NNs. In this paper, both NNs have 2, 20, and 2 neurons at the input, hidden, and output layers, respectively. Standard sigmoid activation function is used. For both NNs, the first-layer weights are uniformly randomly distributed between -1 and 1 [18]. The threshold weights for the first layer are uniformly randomly distributed between -20 and 20. The second layer weights W are initialized to zero for both neural networks. Neural network tuning parameters are given by $k = 0.0001$, $C_f = 20$, $C_g = 20$.

The state observer matrix is $T = \mathrm{diag}\{-50,-50\}$. A control signal $v(t)$ is given by $v(t) = 10\sin(t)$. The system is first assumed healthy. Fig. 3 and Fig. 4 show the state variables and observer states.

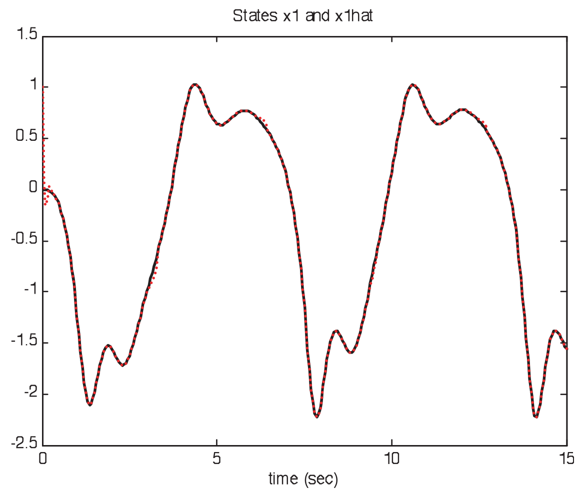

Fig. 3. System state $x_1(t)$ (full line) and observer state $\hat{x}_1(t)$ (dotted line).

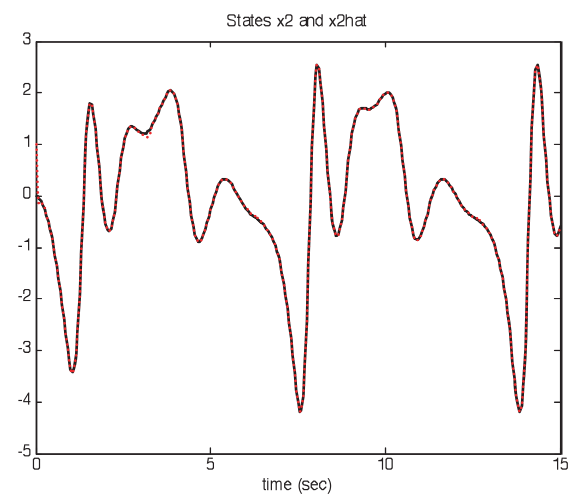

Fig. 4. System state $x_2(t)$ (full line) and observer state $\hat{x}_2(t)$ (dotted line).

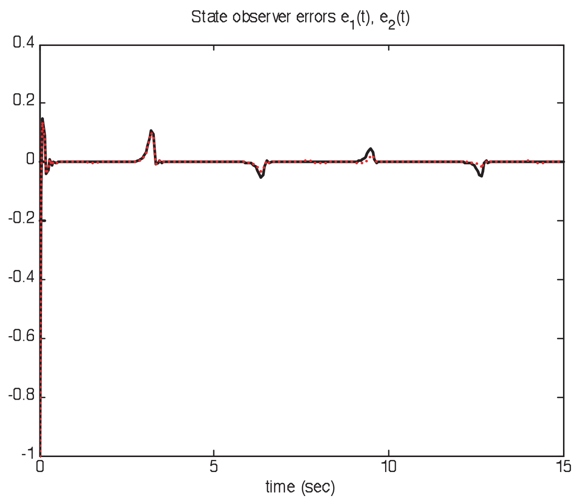

Fig. 5. System state observer errors $e_1(t)$ (full line) and $e_2(t)$ (dotted line).

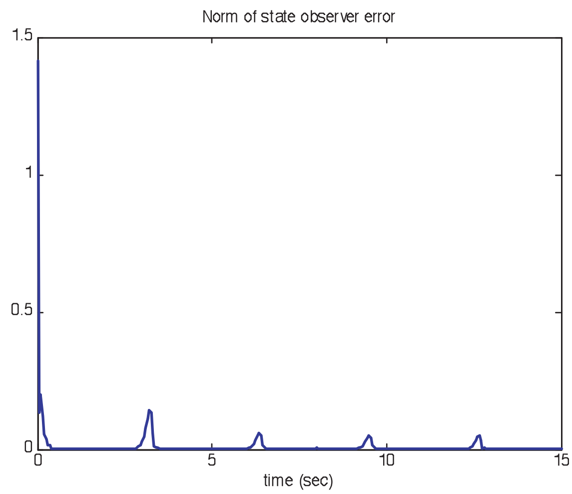

Fig. 6. Norm of the error $e(t)$.

Figs. 5 and 6 show the observer state errors $e_1(t)$, $e_2(t)$, and norm of the error signal.

NN weights bounds and approximation error bounds are estimated as $W_{fM} = W_{gM} = 3$, $\varepsilon_{fM} = \varepsilon_{gM} = 0.1$ based on simulated NN. Then the error bound is estimated as $e_B \approx 0.2$. We now assume that there is a fault at $t = 5\,\text{sec}$ where $\bar{u} = 30$. The approximated values used in inequality (52) are $C_1 \approx 5$ and the integral on the left side is estimated to 0.9. Knowing the matrix T, approximated terms in sufficient fault detectability condition lead to $0.9 > 0.4 + \frac{5}{50}\left[1 - e^{-100}\right]$.

Simulation results are given in Figs. 7 and 8.

It is clear that the actuator fault will be detected after approximately 2 seconds. If there is an actuator fault where $\bar{u} = 7$ then the fault cannot be detected with the same observer "sensitivity" as can be seen in Figs. 9 and 10.

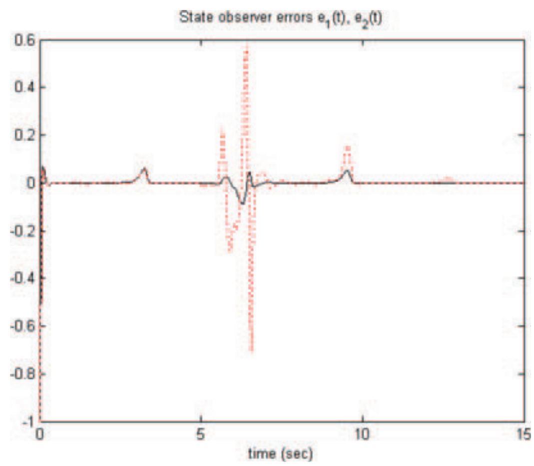

Fig. 7. System state observer errors $e_1(t)$ (full line) and $e_2(t)$ (dotted line) with an actuator fault occurring at $t = 5\,\text{sec}$.

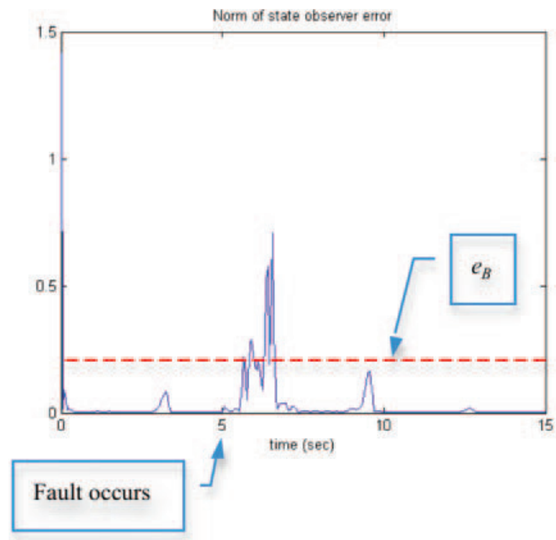

Fig. 8. Norm of the error $e(t)$ with an actuator fault occurring at $t = 5\,\text{sec}$.

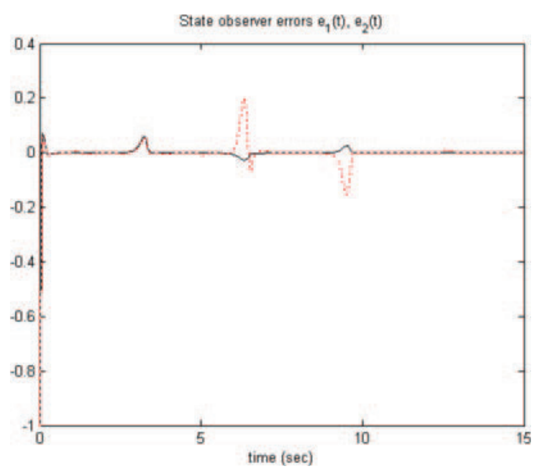

Fig. 9. System state observer errors $e_1(t)$ (full line) and $e_2(t)$ (dotted line) with an actuator fault occurring at $t = 5\,\text{sec}$.

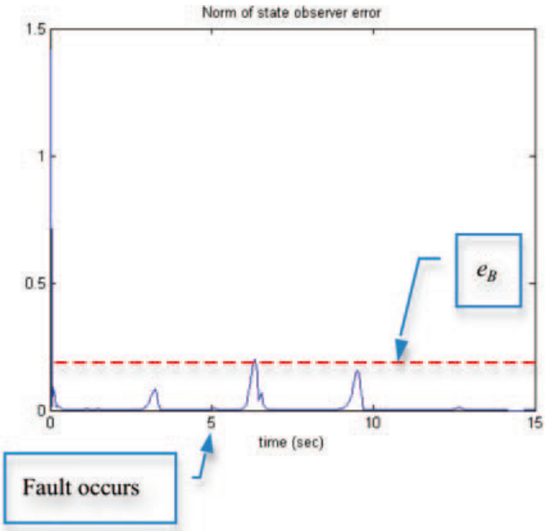

Fig. 10. Norm of the error $e(t)$ with an actuator fault occurring at $t = 5$ sec.

6. Concluding Remarks

We have shown how neural net-based system can be used in actuator fault detection in unknown, nonlinear, input-affine systems. Stable neural net tuning laws are given and estimate on the state observer error is provided using Lyapunov approach. Both state feedback and output feedback cases were considered. Sufficient conditions for actuator fault detectability are given that relate neural net and observer parameters, actuator desired input signal, and fault detectability time. Simulation results show that this method can be effective in actuator fault detection and that identified sensitivity can be adjusted according to the specific application and need.

An open research question is to combine active actuator failure detection methods in case condition (53) is not satisfied.

References

1. Boskovic JD, Li S, Mehra RK. Intelligent control of spacecraft in the presence of actuator failures. Proceedings of 38th IEEE Conference on Decision and Control, Phoenix, AZ, 1999, pp. 4472–4477
2. Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst* 1989; 2(4): 303–314
3. Calise A, Hovakimyan N, Idan M. Adaptive output feedback control of nonlinear systems using neural networks. *Autom Spec Issue Neural Netw Feedback Control* 2001; 37(8): 1201–1211
4. Chen RH, Mingori DL, Speyer JL. Optimal stochastic fault detection filter. *Autom* 2003; 39: 377–390
5. Choi J, Farrell J. Observer-based backstepping control using on-line approximation. *Proceedings of American Control Conference*, June 2000, pp. 3646–3650
6. Cue RW, Muir DE. Engine performance monitoring and troubleshooting techniques for the CF-18 aircraft. *ASME J Eng Gas Turbines Power* 1991; 113(1): 11–19
7. Chen F-C, Khalil HK. Adaptive control of nonlinear systems using neural networks. *Int J Control* 1992; 55(6): 1299–1317
8. Choi Y-K, Lee MJ, Kim S. Design and implementation of adaptive neural-network compensator for control systems. *IEEE Trans Ind Electron* 2001; 48(2): 416–423
9. Chung WH, Speyer JL, Chen RH. A decentralized fault detection filter. *J Dyn Syst Meas Control* 2001; 123: 237–247
10. Dickerson JE, Dickerson JA. Fuzzy network profiling for intrusion detection. In NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society, Atlanta, 2000, 301–306
11. Farrell J, Polycarpou M, *Adaptive Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*, Hoboken, NJ: Wiley, 2006
12. Funahashi K. On the approximate realization of continuous mappings by neural networks. *Neural Netw* 1989; 2: 183–192.
13. Gao W, Selmic RR. Adaptive neural network output feedback control of nonlinear systems with actuator saturation. Proceedings of 44th IEEE Conference on Decision and Control, Seville, Spain, Dec. 2005
14. Ge SS, Hang CC, Zhang T. Adaptive neural network control of nonlinear systems by state and output feedback. *IEEE Trans Systems Man Cybern B* 1999; 29(6): 818–828
15. Gong JQ, Yao B. Output feedback neural network adaptive robust control with application to linear motor drive system. *J Dyn Syst Meas Control* 2006; 128: 227–235
16. Hornik K, Stinchombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw* 1989; 2: 359–366
17. Kim YH, Lewis FL, Abdallah CT. Nonlinear observer design using dynamic recurrent neural networks. Proceedings of 36th Conference on Decision and Control, Kobe, Japan 1996
18. Igelnik B, Pao YH. Stochastic Choice of Basis Functions in Adaptive Function Approximation and the Functional-Link Net. *IEEE Trans Neural Netw* 1995; 6(6): 1320–1329
19. Lin S, Goldenberg AA. Neural-network control of mobile manipulators. *IEEE Trans Neural Netw* 2001; 12(5): 1121–1133
20. Lewis FL, Campos J, Selmic RR. *Neuro-Fuzzy Control of Industrial Systems with Actuator Nonlinearities*. SIAM Press: Philadelphia, PA, 2002
21. Lewis FL, Jagannathan S, Yesildirek A. *Neural Network Control of Robot Manipulators and Nonlinear Systems*, Taylor and Francis: Philadelphia, PA, 1999
22. Mahmoud M, Jiang J, Zhang Y. Analysis of the stochastic stability for fault tolerant control systems. Proceedings of 38th IEEE Conference on Decision and Control, Phoenix, AZ, 1999
23. Maron VD, Hamelin F, Noura H, A fault tolerant design against major actuator failures: application to a

three-tank system. Proceedings of 38th IEEE Conference on Decision and Control, Phoenix, AZ, 1999
24. Narendra KS, Annaswamy AM. *Stable Adaptive Systems*. Prentice Hall: New Jersey, 1989
25. Narendra KS, Balekrishnan J. Adaptive control using multiple models. *IEEE Trans Autom Control* 1997; 42(2): 171–187
26. Narendra KS. Adaptive control using neural networks. *Neural Networks for Control*. WT Miller, RS Sutton, PJ Werbos (eds). Cambridge: MIT Press, 1991, pp. 115–142
27. Polycarpou M, Helmicki A. Automated fault detection and accommodation: a learning system approach. *IEEE Trans Syst, Man Cybern* 1995; 25(11): 1447–1458
28. Polycarpou MM, Trunov AB. Learning approach to nonlinear fault diagnosis: detectability analysis. *IEEE Trans Autom Control* 2000; 45(4): 806–812
29. Polycarpou MM. Stable adaptive neural control scheme for nonlinear systems. *IEEE Trans Autom Control* 1996; 41(3): 447–451
30. Rahnamai K, Arabshahi P, Yan TY, Pham T, Finley SG. An intelligent fault detection and isolation architecture for antenna arrays. *The Telecommunications and Data Acquisition Progress Report*, NASA
31. Selmic RR, Lewis FL. Multimodel neural networks identification and failure detection of nonlinear systems. Proceedings of 40th IEEE Conference on Decision and Control, 3128–3133, Orlando, FL, Dec. 2001
32. Selmic R, Polycarpou M, Parisini T. Output feedback actuator fault detection in nonlinear systems using neural networks. Proceedings of European Control Conference 2007, Kos, Greece, July 2007
33. Slotine J-JE, Li W. *Applied Nonlinear Control*. Prentice Hall: New Jersey, 1991
34. Tao G, Chen S, Tang X, Joshi SM. *Adaptive Control of Systems with Actuator Failures*. Springer-Verlag: London, 2004
35. Willsky AS. A survey of design methods for failure detection in dynamic systems. *Automatica* 1976; 12: 601–611
36. Yang B, Lin W. Robust output feedback stabilization of uncertain nonlinear systems with uncontrollable and unobservable linearization. *IEEE Trans Autom Control* 2005; 50(5): 619–630
37. Zhang X, Polycarpou MM, Parisini T. A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. *IEEE Trans Autom Control* 2002; 47(4): 576–593
38. Zhang X, Parisini T, Polycarpou MM. Adaptive fault-tolerant control of nonlinear uncertain systems: an information-based diagnostic approach. *IEEE Trans Autom Control* 2004; 48(8): 1259–1274