Proceedings of the 2006 American Control Conference
Minneapolis, Minnesota, USA, June 14-16, 2006

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

Rastko R. Selmic, Marios M. Polycarpou, and Thomas Parisini

Abstract—**This paper describes actuator fault identification in unknown, input-affine, nonlinear systems using neural networks. Neural net tuning algorithms have been derived and identifier have been developed using the Lyapunov approach. The paper defines and analyses the fault dynamics i.e., the dynamical properties of a failure process. A rigorous detectability condition is given for actuator faults in nonlinear systems relating the actuator desired input signal and neural net-based observer sensitivity. Sufficient conditions are given in terms of the input signal and related actuator fault such that a fault can be detected. Simulation results are presented to illustrate the detectability criteria and fault detection in nonlinear systems.**

I. Introduction

EARLY detection of faults in complex electromechanical systems with many actuators is not only essential for prevention of cascading catastrophic failures, but also for enhanced system performance and availability. Advanced failure detection systems improve the overall system performance and environmental safety. 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 [3]. The main research effort has been concentrated in combining rule-based fault detection methods with other intelligent techniques such as neural networks (NNs) and fuzzy logic [1], [4], [17], [18].

Model-based fault detection methods compare estimated models to a nominal system model [16]. The error between the two models is a measure of the deviation between estimated and nominal models. The main deficiency in these approaches is the requirement of an accurate system model that is not available in most of practical applications. NNs and fuzzy logic are artificial intelligence tools that can overcome these restrictive requirements. The main advantage of NNs is their nonlinear approximation property [5], [6] that provides the ability to learn anomalies and failures from actual monitoring data. If detected, NNs may classify them without knowing detailed system models [16], [17], [18], [19]. Statistical analysis based on the Lyapunov function for fault tolerant control systems is described in [9], [13]. Fault tolerant control system design is presented in [11].

Recent work in adaptive control enabled an interesting application in actuator failure detection [20], 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 [21], [22], Zhang et al. developed combined detection, isolation, and compensation techniques of faults in nonlinear systems. It is assumed that the nonlinear system is known to some degree and that faults belong to a known class of faults. Groundwork on multiple models/controllers in adaptive control was provided by Narendra's work [15] and this tool can effectively be used in actuator failure detection and compensation [12].

This paper addresses the problem of actuator fault detection in *unknown*, input-affine nonlinear systems. While papers [21], [22] address fault detection, isolation, and compensation in nonlinear systems, 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 relating the input control signal, neural net identifier parameters, and detectability of the faults. Results in [20] describe robust adaptive compensators for actuators faults and do not deal with detectability and isolation of faults. Polycarpou and Trunov [17] developed fault detectability conditions for nonlinear systems with adaptive learning methods. Here we extend the results presented in [17] by considering the actuator fault detection problem, by referring to a different class of nonlinear systems, and by use of neural networks for systems identification. Indeed, the following fundamental questions in actuator fault detection are addressed: What kind of actuator faults can be detected? Under what conditions faults are detectable when identified using neural networks-based methodology? If faults are not presently detectable, how neural net identifier parameters need to be adjusted in order to detect the faults?

Manuscript received September 20, 2005. This work was supported in part by the NSF EPSCoR Pfund Grant #32-0967-58159 and the Research Promotion Foundation of Cyprus.

R. R. Selmic is with Department of Electrical Engineering, Louisiana Tech University, LA 71272, USA (phone: 318-257-4641; fax: 318-257-4922; e-mail: rselmic@latech.edu).

M. M. Polycarpou is with Department of Electrical and Computer Engineering, University of Cyprus, Cyprus (mpolycar@ucy.ac.cy).

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

1-4244-0210-7/06/$20.00 ©2006 IEEE

II. PROBLEM FORMULATION

In this section we formulate the problem of actuator fault detection, fault observability, and fault dynamics where the fault is considered as a time-varying and evolving process in actuators with nonlinear systems.

Let $\Re$ denote real numbers, $\Re^n$ denote the real *n* vector, and $\Re^{m\times n}$ denote the real $m\times n$ matrices. Consider a single-input single-output nonlinear, input-affine system given by

$$\begin{aligned} \dot{x} &= f(x) + g(x)u + d(t) \\ y &= h(x) \end{aligned} \quad (1)$$

where $x = [x_1, x_2, \cdots, x_n]^T$ is a system state and *f*, *g*: $\Re^n \to \Re^n$, *h*: $\Re^n \to \Re$ are unknown, smooth functions, *u* is the input control signal, *y* is the output, and *d* represents a system disturbance. It is assumed that the disturbance is bounded such that

$$\|d(t)\| \le D_B \ . \quad (2)$$

It is also assumed that the full state measurement of the system (1) is available.

The actuator of the nonlinear system (1) generates the control signal *u* based on desired actuator signal *v*. We consider here the Tao's actuator fault model [20] given by

$$u(t) = v(t) + \gamma(\bar{u} - v(t)) \ , \quad (3)$$

where the fault value of the actuator input is $\bar{u}$. The case when $\gamma = 0$ models the actuator without a fault; $\gamma = 1$ models the actuator with a fault. In case of an actuator fault, from equation (3) one has that $u(t) = \bar{u}$. In the analysis, we will consider the actuator fault value as a time-varying function. Then, the actuator model is given by

$$u(t) = v(t) + \gamma(\bar{u}(t) - v(t)) \ . \quad (4)$$

Combined, the nonlinear system and the actuator fault model is

$$\dot{x} = f(x) + g(x)\left[v(t) + \gamma(\bar{u}(t) - v(t))\right] + d(t) \ . \quad (5)$$

The combined model (5) is equivalent to nonlinear systems fault model in [22]

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

where $\beta(t-T)$ represents the time profile of a fault occurrence and g_1 is a fault function. Model (5) emphasizes actuator faults while the general model (6) considers faults of a whole nonlinear system.

It is assumed that the actuator is healthy at the beginning of operation. This is a reasonable assumption since the only way to have successful learning/identification phase is to assume that the system is healthy initially.

III. BACKGROUND

Consider a NN consisting of two layers of tunable weights. The output of such NN is given by

$$y = W^T \mu(V^T x + v_0) \ , \quad (7)$$

$x \in \Re^n$, $y \in \Re^m$. Many well-known results indicate that any sufficiently smooth function can be approximated arbitrary closely on a compact set using a two-layer NN with appropriate weights [2], [5]. The size of the NN determines performance and convergence of the tracking error. A common approach is NN selection is to start with the smaller network, and gradually increase it until satisfactory performance is achieved. The first layer weights *V* are selected off-line randomly and will not be tuned. Such NN is then linearly parameterized. The NN *universal approximation property* says that any continuous function can be approximated arbitrarily well using a linear combination of sigmoidal functions, namely

$$f(x) = W^T \sigma(x) + \varepsilon(x) \ , \quad (8)$$

where the $\varepsilon(x)$ is the NN approximation error. The reconstruction error is bounded on a compact set *S* by $\|\varepsilon(x)\| < \varepsilon_N$ [7].

IV. ACTUATOR FAULT DETECTION, FAULT DYNAMICS AND DETECTABILITY

A neural net (NN) learning is considered as an identification method for the nonlinear system (1). A NN system observer is given by [19]

$$\dot{\hat{x}} = -A\hat{x} + Ax + \hat{W}_f^T \sigma_f(x) + \hat{W}_g^T \sigma_g(x) v(t) \ , \quad (9)$$

where the matrix *A* is Hurwitz and two NNs are used to approximate nonlinear functions *f*(*x*) and *g*(*x*)

$$f(x) = W_f^T \sigma_f(x) + \varepsilon_f(x) \quad (10)$$

$$g(x) = W_g^T \sigma_g(x) + \varepsilon_g(x) \ , \quad (11)$$

and W_f, W_g are some ideal target NN weights, and $\varepsilon_f(x)$, $\varepsilon_g(x)$ are NN approximation errors.

Assumption 1. NN Weights Bounds
Ideal neural net weights are bounded such that $\|W_f\|_F \le W_{fM}$, $\|W_g\|_F \le W_{gM}$ with W_{fM} and W_{gM} known bounds.

In this paper we consider unknown nonlinear systems but we do require some a priori knowledge about the system i.e., ideal NN weight bounds. The structure of the NN actuator fault observer is then given in Figure 1.

The quality measure of the actuator operation is given by the state observer error between the system states and the NN observer as

$$e = x - \hat{x} \quad (12)$$

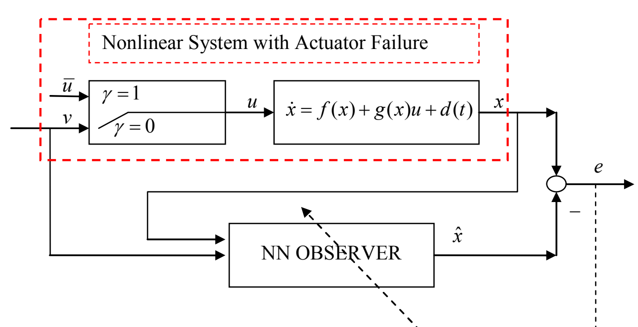

Figure 1. NN system observer – fault identifier.

For a healthy system, a nonlinear system and observer dynamics are given by

$$\dot{e} = f(x) + g(x)v + d(t) + A\hat{x} - Ax - \hat{W}_f^T \sigma_f(x) - \hat{W}_g^T \sigma_g(x)v(t) \quad (13)$$

$$\dot{e} = -Ae + \tilde{W}_f^T \sigma_f(x) + \varepsilon_f(x) + \tilde{W}_g^T \sigma_g(x)v(t) + \varepsilon_g(x)v + d(t), \quad (14)$$

where $\tilde{W}_f = W_f - \hat{W}_f$, $\tilde{W}_g = W_g - \hat{W}_g$.

We assume that the system has already been stabilized, meaning that control signals are already bounded. In this paper we do not consider the controller of the system (1); we concentrate on actuator fault detection, fault detectability conditions, and detectability time. The following theorem provides NN tuning laws and a bound on the state observer error using *e*-modification type of adaptation [8], [14].

Theorem 1. Stable NN Observer Tuning Law

Given the nonlinear system (1), and the NN observer (9), let the estimated NN weights be provided by the NN tuning algorithm

$$\dot{\hat{W}}_f = C_f \sigma_f(x) e^T - kC_f \|e\| \hat{W}_f \quad (15)$$

$$\dot{\hat{W}}_g = C_g \sigma_g(x) v(t) e^T - kC_g \|e\| \hat{W}_g \quad (16)$$

with any constant matrices $C_f = C_f^T > 0$, $C_g = C_g^T > 0$, and a small design parameter *k*. Then the state observer error *e* and the NN weight estimation errors $\tilde{W}_f$, $\tilde{W}_g$ are uniformly ultimately bounded. The convergence region for the state observer error *e* can be reduced by increasing the gain *A*.

Proof: Select the Lyapunov function as

$$L = \frac{1}{2}e^T e + \frac{1}{2}tr\left[\tilde{W}_f^T C_f^{-1} \tilde{W}_f\right] + \frac{1}{2}tr\left[\tilde{W}_g^T C_g^{-1} \tilde{W}_g\right]. \quad (17)$$

Then its derivative is given by

$$\dot{L} = e^T \dot{e} + tr\left[\tilde{W}_f^T C_f^{-1} \dot{\tilde{W}}_f\right] + tr\left[\tilde{W}_g^T C_g^{-1} \dot{\tilde{W}}_g\right] \quad (18)$$

$$\dot{L} = -e^T Ae + e^T \tilde{W}_f^T \sigma_f(x) + e^T \varepsilon_f(x) + e^T \tilde{W}_g^T \sigma_g(x)v(t) + e^T \varepsilon_g(x)v + e^T d(t) + tr\left[\tilde{W}_f^T C_f^{-1} \dot{\tilde{W}}_f\right] + tr\left[\tilde{W}_g^T C_g^{-1} \dot{\tilde{W}}_g\right]$$

$$\dot{L} = -e^T Ae + tr\left[\tilde{W}_f^T (C_f^{-1} \dot{\tilde{W}}_f + \sigma_f(x) e^T)\right] + tr\left[\tilde{W}_g^T (C_g^{-1} \dot{\tilde{W}}_g + \sigma_g(x) v(t) e^T)\right] + e^T \varepsilon_f(x) + e^T \varepsilon_g(x)v + e^T d(t)$$

Applying tuning law equations yields

$$\dot{L} = -e^T Ae + tr\left[k\tilde{W}_f^T \|e\| \hat{W}_f\right] + tr\left[k\tilde{W}_g^T \|e\| \hat{W}_g\right] + e^T \varepsilon_f(x) + e^T \varepsilon_g(x)v + e^T d(t)$$

$$\dot{L} = -e^T Ae + k\|e\| tr\left[\tilde{W}_f^T \hat{W}_f\right] + k\|e\| tr\left[\tilde{W}_g^T \hat{W}_g\right] + e^T \varepsilon_f(x) + e^T \varepsilon_g(x)v + e^T d(t)$$

$$\dot{L} = -e^T Ae + k\|e\| tr\left[\tilde{W}_f^T (W_f - \tilde{W}_f)\right] + k\|e\| tr\left[\tilde{W}_g^T (W_g - \tilde{W}_g)\right] + e^T \varepsilon_f(x) + e^T \varepsilon_g(x)v + e^T d(t)$$

Since the matrix *A* is Hurwitz, it follows

$$\dot{L} \le -\lambda_{\min} \|e\|^2 + k\|e\| \|\tilde{W}_f\| (W_{fM} - \|\tilde{W}_f\|) + k\|e\| \|\tilde{W}_g\| (W_{gM} - \|\tilde{W}_g\|) + e^T \varepsilon_f(x) + e^T \varepsilon_g(x)v + e^T d(t)$$

where $\lambda_{\min}$ is a minimum eigenvalue of *A*. Assuming that the disturbance is bounded (2) and that the system has already been stabilized meaning that the control signal is bounded, $\|v(t)\| \le V_B$, yields to

$$\dot{L} \le -\lambda_{\min} \|e\|^2 + k\|e\| \|\tilde{W}_f\| (W_{fM} - \|\tilde{W}_f\|) + k\|e\| \|\tilde{W}_g\| (W_{gM} - \|\tilde{W}_g\|) + \|e\| \left[\varepsilon_{fM} + \varepsilon_{gM} V_B + D_B\right]$$

$$\dot{L} \le \|e\| \left\{ -\lambda_{\min} \|e\| + k\|\tilde{W}_f\| (W_{fM} - \|\tilde{W}_f\|) + k\|\tilde{W}_g\| (W_{gM} - \|\tilde{W}_g\|) + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B \right\}$$

$$\dot{L} \le \|e\| \left\{ -\lambda_{\min} \|e\| + k\|\tilde{W}_f\| (W_{fM} - \|\tilde{W}_f\|) + k\|\tilde{W}_g\| (W_{gM} - \|\tilde{W}_g\|) + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B \right\}$$

$$\dot{L} \le \|e\| \{ -\lambda_{\min} \|e\| - k\left(\|\tilde{W}_f\| - \frac{1}{2}W_{fM}\right)^2 + \frac{1}{4}W_{fM}^2 - k\left(\|\tilde{W}_g\| - \frac{1}{2}W_{gM}\right)^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B \}$$

Therefore, $\dot{L}$ is negative semi-definite providing that

$$-\lambda_{\min} \|e\| + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B < 0$$

or

$$-k\left(\|\tilde{W}_f\| - \frac{1}{2}W_{fM}\right)^2 + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B < 0$$

or

$$-k\left(\|\tilde{W}_g\| - \frac{1}{2}W_{gM}\right)^2 + \frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B < 0 \cdot$$

Equivalently, the $\dot{L}$ is negative semi-definite as long as

$$\|e\| > \frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B}{\lambda_{\min}} \quad (19)$$

or

$$\|\tilde{W}_f\| > \sqrt{\frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B}{k}} + \frac{1}{2}W_{fM} \quad (20)$$

or

$$\|\tilde{W}_g\| > \sqrt{\frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B}{k}} + \frac{1}{2}W_{gM} \quad (21)$$

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

$$e_B = \frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B}{\lambda_{\min}} \quad (22)$$

$$\tilde{W}_{fB} = \sqrt{\frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B}{k}} + \frac{1}{2}W_{fM} \quad (23)$$

$$\tilde{W}_{gB} = \sqrt{\frac{\frac{1}{4}W_{fM}^2 + \frac{1}{4}W_{gM}^2 + \varepsilon_{fM} + \varepsilon_{gM} V_B + D_B}{k}} + \frac{1}{2}W_{gM} \quad (24)$$

□

The equations (22), (23), and (24) provide the explicit bounds on the observer state error and NN weights approximation errors. Note that (22) represents the bound on the observer state error under healthy system assumption. It is also a conservative bound since it includes the NN weights bounds.

Even though increase in the matrix gain A can reduce the convergence region for the state observer error, it is not recommended to have the gain too large since that will amplify noise and disturbance effects.

The NN tuning algorithms for nonlinear system identifier (15), (16) are similar to Lewis' NN robotic control tuning algorithms [8], [9].

Proposition 1. 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. Once the state observer error has reached steady state value, the NN weights estimation errors will be limited by the following algorithm modification

$$\dot{\tilde{W}}_f = \begin{cases} -C_f \sigma_f(x) e^T + kS\|e\|\hat{W}_f \text{ , for } \tilde{W}_f \le \tilde{W}_{fB} \\ 0 \text{ , for } \tilde{W}_f > \tilde{W}_{fB} \end{cases} \quad (25)$$

$$\dot{\tilde{W}}_g = \begin{cases} -C_g \sigma_g(x) v(t) e^T + kT\|e\|\hat{W}_g \text{ , for } \tilde{W}_g \le \tilde{W}_{gB} \\ 0 \text{ , for } \tilde{W}_g > \tilde{W}_{gB} \end{cases} \quad (26)$$

Assumption 2.

An actuator fault in nonlinear system (1) has occurred after the learning phase has been completed.

The Assumption 2 is a natural requirement that the system must be healthy during the learning phase in order to be able to detect the potential fault. Any differences, anomalies, and/or failures in a system dynamics can be detected later during the operation phase.

If the system fails suddenly, the state observer error is given by

$$e_1 = x - \hat{x} \quad (27)$$

$$\begin{aligned} \dot{e}_1 &= f(x) + g(x)\bar{u}(t) + g(x)v(t) - g(x)v(t) + d(t) \\ &\quad + A\hat{x} - Ax - \hat{W}_{f1}^T \sigma_f(x) - \hat{W}_{g1}^T \sigma_g(x) v(t) \end{aligned} \quad . \quad (28)$$

Modified closed-loop error dynamics due to actuator fault is given by

$$\begin{aligned} \dot{e}_1 &= -Ae_1 + \tilde{W}_{f1}^T \sigma_f(x) + \varepsilon_f(x) + \tilde{W}_{g1}^T \sigma_g(x) v(t) + \varepsilon_g(x) v(t) \\ &\quad + d(t) + g(x)(\bar{u}(t) - v(t)) \end{aligned}$$

where the state x will have a different value compared with the healthy system. The complete system dynamics (nonlinear system and NN observer) is given by the following system of coupled differential equations

$$\begin{aligned} \dot{e}_1 &= -Ae_1 + \tilde{W}_{f1}^T \sigma_f(\bar{x}) + \varepsilon_f(\bar{x}) + \tilde{W}_{g1}^T \sigma_g(\bar{x}) v(t) \\ &\quad + \varepsilon_g(\bar{x}) v(t) + d(t) + g(\bar{x})(\bar{u}(t) - v(t)) \end{aligned} \quad (29)$$

$$\dot{\tilde{W}}_{f1} = \begin{cases} -C_f \sigma_f(\bar{x}) e_1^T + kC_f\|e_1\|\hat{W}_{f1} \text{ , for } \tilde{W}_{f1} \le \tilde{W}_{fB} \\ 0 \text{ , for } \tilde{W}_{f1} > \tilde{W}_{fB} \end{cases} \quad (30)$$

$$\dot{\tilde{W}}_g = \begin{cases} -C_g \sigma_g(\bar{x}) v(t) e_1^T + kC_g\|e_1\|\hat{W}_{g1} \text{ , for } \tilde{W}_{g1} \le \tilde{W}_{gB} \\ 0 \text{ , for } \tilde{W}_{g1} > \tilde{W}_{gB} \end{cases} \quad (31)$$

and $\bar{x}$ denotes the state x when there is a fault in the actuator.

After the initial time period when the system has been failure-free, it is important to study under what conditions we will be able to detect the failure or anomaly in the system, 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.

Note that the bound on the failure-free state observer error is given by

$$\|e\| \le e_B \text{ ,} \quad (32)$$

where e_B is given by (22).

It is important to consider the difference in behavior of the state observer error when the actuator is healthy and when there is an actuator fault. This difference in state observer error behavior has its own dynamics and we define this as a fault dynamics.

Failure itself as a process has its own dynamics. The crucial point is to examine the error between the failure-free observer tracking error and observer error when there is an actuator fault. This *failure dynamics* is given by

$$\begin{aligned} \dot{\tilde{e}} &= -A\tilde{e} + \tilde{W}_f^T \sigma_f(x) - \tilde{W}_{f1}^T \sigma_f(\bar{x}) + \varepsilon_f(x) - \varepsilon_f(\bar{x}) + \tilde{W}_g^T \sigma_g(x) v(t) \\ &\quad - \tilde{W}_{g1}^T \sigma_g(\bar{x}) v(t) + \varepsilon_g(x) v(t) + \varepsilon_g(\bar{x}) v(t) - g(\bar{x})(\bar{u}(t) - v(t)) \end{aligned} \quad (33)$$

where

$$\tilde{e} = e - e_1 \text{ .} \quad (34)$$

The system (33) can be represented in the following format

$$\dot{\tilde{e}} = h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, x, \bar{x}, t) - g(\bar{x})(\bar{u}(t) - v(t)) \quad (35)$$

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

$$\begin{aligned} h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, x, \bar{x}, t) &= -A\tilde{e} + \tilde{W}_f^T \sigma_f(x) - \tilde{W}_{f1}^T \sigma_f(\bar{x}) + \varepsilon_f(x) \\ &\quad - \varepsilon_f(\bar{x}) + \tilde{W}_g^T \sigma_g(x) v(t) - \tilde{W}_{g1}^T \sigma_g(\bar{x}) v(t) + \varepsilon_g(x) v(t) + \varepsilon_g(\bar{x}) v(t) \end{aligned}$$

Consider the system (33) and assume that an actuator fault occurs at $t=t_0$. The solution of the system (33) is given by

$$\tilde{e}(t) = \tilde{e}(t_0) + \int_{t_0}^{t} \left[h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, \tilde{W}_g^T, \tilde{W}_{g1}^T, x, \bar{x}, \tau) - g(\bar{x})(\bar{u} - v) \right] d\tau$$

Since the fault occurred at $t=t_0$, then $\tilde{e}(t_0) = 0$. It follows that

$$\tilde{e}(t) = \int_{t_0}^{t} \left[h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, \tilde{W}_g^T, \tilde{W}_{g1}^T, x, \bar{x}, \tau) - g(\bar{x})(\bar{u} - v) \right] d\tau \quad (36)$$

$$\tilde{e}(t) = \int_{t_0}^{t} h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, \tilde{W}_g^T, \tilde{W}_{g1}^T, x, \bar{x}, \tau) d\tau - \int_{t_0}^{t} g(\bar{x})(\bar{u} - v) d\tau$$

Taking the norm of the fault error signal yields

$$\|\tilde{e}(t)\| \ge \left\| \int_{t_0}^{t} g(\bar{x})(\bar{u} - v) d\tau \right\| - \left\| \int_{t_0}^{t} h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, \tilde{W}_g^T, \tilde{W}_{g1}^T, x, \bar{x}, \tau) d\tau \right\|$$

.

Condition for the fault detectability is given by

$$\left\| \int_{t_0}^{t} g(\bar{x})(\bar{u}(\tau) - v(\tau)) d\tau \right\| - \left\| \int_{t_0}^{t} h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, \tilde{W}_g^T, \tilde{W}_{g1}^T, x, \bar{x}, \tau) d\tau \right\| \ge 2e_B$$

or equivalently

$$\left\| \int_{t_0}^{t} g(\bar{x})(\bar{u}(\tau) - v(\tau)) d\tau \right\| \ge 2e_B + \left\| \int_{t_0}^{t} h(\tilde{e}, \tilde{W}_f^T, \tilde{W}_{f1}^T, \tilde{W}_g^T, \tilde{W}_{g1}^T, x, \bar{x}, \tau) d\tau \right\| \quad (37)$$

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

Theorem 2. Detectability of Actuator Faults
Given Assumptions 1-2 and a NN observer from Theorem 1, a fault in the system actuator can be detected after time interval T if the following condition is satisfied

$$\left\| \int_{t_0}^{t_0+T} g(\bar{x})(\bar{u}(\tau)-v(\tau))d\tau \right\| \ge 2e_B + T\,C_1 \tag{38}$$

where $C_1 = e_B\lambda_{\max} + \left(2\tilde{W}_{fB} + 2\tilde{W}_{gB} + 2\varepsilon_M\right)\left(V_B + 1\right)$.

The above condition provides rigorous justification for an intuitive concept of fault detectability which says that for the fault to be detected there should be "enough" difference 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.

Proof:
Using some norms inequalities one has

$$\begin{aligned} &2e_B + \left\| \int_{t_0}^{t} h(\tilde{e},\tilde{W}_f^T,\tilde{W}_{f1}^T,\tilde{W}_g^T,\tilde{W}_{g1}^T,x,\bar{x},\tau)d\tau \right\| \le \\ &2e_B + \int_{t_0}^{t} \left\| h(\tilde{e},\tilde{W}_f^T,\tilde{W}_{f1}^T,\tilde{W}_g^T,\tilde{W}_{g1}^T,x,\bar{x},\tau) \right\| d\tau \end{aligned} \tag{39}$$

$$\begin{aligned} &2e_B + \left\| \int_{t_0}^{t} h(\tilde{e},\tilde{W}_f^T,\tilde{W}_{f1}^T,\tilde{W}_g^T,\tilde{W}_{g1}^T,x,\bar{x},\tau)d\tau \right\| \le \\ &2e_B + \int_{t_0}^{t} \left[\|e\|\|A\|_F + \left(\|\tilde{W}_f\| + \|\tilde{W}_{f1}\| + \|\tilde{W}_g\| + \|\tilde{W}_{g1}\| + \|\varepsilon_f\| + \|\varepsilon_{f1}\| \right)\left(\|v(t)\| + 1\right) \right] d\tau \\ &\le 2e_B + (t-t_0)\left[\|e\|\|A\|_F + \left(\|\tilde{W}_f\| + \|\tilde{W}_{f1}\| + \|\tilde{W}_g\| + \|\tilde{W}_{g1}\| + \|\varepsilon_f\| + \|\varepsilon_{f1}\| \right)\left(\|v(t)\| + 1\right) \right] \\ &\le 2e_B + (t-t_0)\left[e_B\lambda_{\max} + \left(2\tilde{W}_{fB} + 2\tilde{W}_{gB} + 2\varepsilon_M\right)\left(V_B + 1\right) \right] \end{aligned} \tag{40}$$

Therefore, the sufficient condition for the fault to be detected at time t is given by

$$\begin{aligned} &\left\| \int_{t_0}^{t} g(\bar{x})(\bar{u}(\tau)-v(\tau))d\tau \right\| \ge \\ &2e_B + (t-t_0)\left[e_B\lambda_{\max} + \left(2\tilde{W}_{fB} + 2\tilde{W}_{gB} + 2\varepsilon_M\right)\left(V_B + 1\right) \right] \end{aligned} \tag{41}$$

□

Expression (41) is a mathematical condition which tells us when and how the fault in the nonlinear system can be detected. The result in Theorem 2 says that $\bar{u}(t)$ and the control signal $v(t)$ need to be sufficiently different over 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 control signal is accidentally at, or close to, that failed position, then the fault does not really have any effect on system states as Theorem 2 shows, and therefore, can not be detected.

V. SIMULATION EXAMPLE

The simulation was performed to evaluate performance of the proposed NN fault identifier and also to verify results from Theorem 2. We consider 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

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -5x_1^3 - 2x_2 + (x_1+1)u(t)\ , \\ y &= x_1 \end{aligned} \tag{42}$$

where the fault model is given by

$$u(t) = v(t) + \gamma(\bar{u}(t) - v(t))\ . \tag{43}$$

The NN observer given by Theorem 1 consists of two NNs.

In this paper, both NNs have 2, 30, 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 [7]. 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 chosen as k=0.0001, C_f=20, C_g=20.

The state observer matrix is A=diag{30,30}. An ideal control signal $v(t)$ is given by

$$v(t) = 10\sin(t)\ . \tag{44}$$

The system is first assumed healthy. Figure 2 and Figure 3 show observer state errors $e_1(t)$, $e_2(t)$, and norm of the error signal.

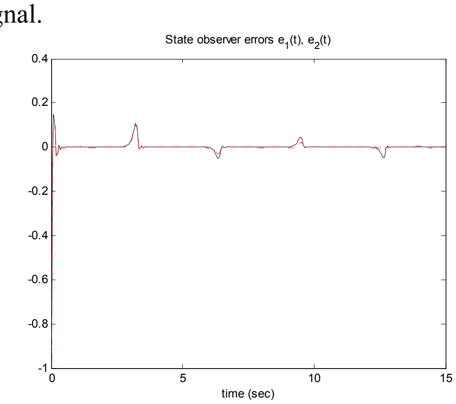

Figure 2. System state observer errors $e_1(t)$ (full line) and $e_2(t)$ (dotted line).

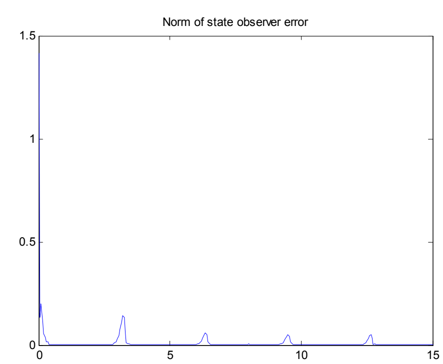

Figure 3. Norm of the error $e(t)$.

NN weights bounds and approximation error bounds are estimated as $W_{fM} = W_{gM} = 3$, $\varepsilon_{fM} = \varepsilon_{gM} = 0.1$. Actual simulation or experimental results can also be used to estimate the above parameters. Then, the error bound is estimated as $e_B \approx 0.2$. We now assume that there is a fault at *t*=5sec where $\bar{u} = 18$. Simulation results are given in Figures 4-5.

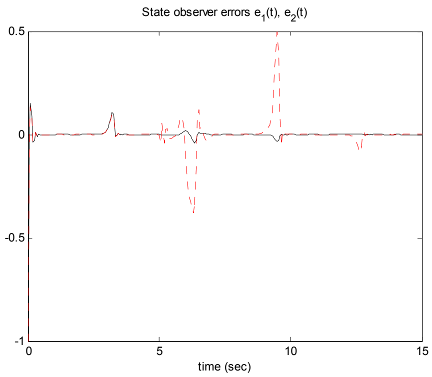

Figure 4. Actuator fault at *t*=5sec.; system state observer errors $e_1(t)$ (full line) and $e_2(t)$ (dotted line).

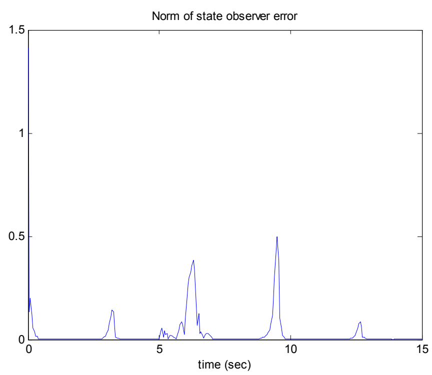

Figure 5. Norm of the error *e*(*t*).

It is clear that the actuator fault will be detected after 2 seconds. If there is an actuator fault where $\bar{u} = 6$ for example, then simulation showed that the fault can not be detected with the same observer "sensitivity".

VI. References

[1] J. D. Boskovic, S. Li, and R. K. Mehra, "Intelligent control of spacecraft in the presence of actuator failures," *In 38th IEEE Conference on Decision and Control*, Phoenix, AZ, 1999, pp. 4472-4477.

[2] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Contr. Signals, Syst.*, vol. 2, no. 4, pp. 303- 314, 1989.

[3] R. W. Cue and D. E. Muir, "Engine performance monitoring and troubleshooting techniques for the CF-18 aircraft," *ASME Journal of Engineering for Gas Turbines and Power*, 113. 11-19, January 1991.

[4] J. E. Dickerson and J. A. Dickerson, "Fuzzy network profiling for intrusion detection," *In NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society*, Atlanta, 2000, pp. 301-306.

[5] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, 2. 183-192.

[6] K. Hornik, M. Stinchombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, 2. 359-366, 1989.

[7] B. Igelnik and Y. H. Pao, "Stochastic Choice of Basis Functions in Adaptive Function Approximation and the Functional-Link Net," *IEEE Trans. Neural Networks*, vol. 6, no. 6, pp. 1320-1329, Nov. 1995.

[8] F. L. Lewis, S. Jagannathan, and A. Yesildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*, Taylor and Francis, Philadelphia, PA, 1999.

[9] F. L. Lewis, J. Campos, and R. R. Selmic, *Neuro-Fuzzy Control of Industrial Systems with Actuator Nonlinearities,* SIAM Press, Philadelphia, PA, 2002.

[10] M. Mahmoud, J. Jiang, and Y. Zhang, "Analysis of the stochastic stability for fault tolerant control systems," *In 38th IEEE Conference on Decision and Control*, Phoenix, AZ, 1999.

[11] V. D. Maron, F. Hamelin, and H. Noura, A fault tolerant design against major actuator failures: application to a three-tank system," *In 38th IEEE Conference on Decisistion and Control*, Phoenix, AZ, 1999.

[12] P. S. Maybeck and R. D. Stevens, "Reconfigurable flight control via multiple model adaptive control methods," *IEEE Transactions on Aerospace Electronic Systems*, 27 (3). 470-480, 1991.

[13] R. K. Mehra, C. Rago, and J. D. Boskovic, "Adaptive fault-tolerant reconfigurable control of a tailless advanced fighter aircraft using multiple model stochastic control," *Tenth Yale Workshop on Adaptive and Learning Systems*, Yale University. 165-172.

[14] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive Systems*, Prentice Hall, New Jersey, 1989.

[15] K. S. Narendra and J. Balekrishnan, "Adaptive control using multiple models," *IEEE Transactions on Automatic Control*, 42 (2). 171-187, 1997.

[16] M. M. Polycarpou, "An on-line approximation approach to fault monitoring, diagnosis, and accommodation," *SAE Transactions-Journal of Aerospace*, 103. 371-380, 1994.

[17] M. M. Polycarpou and A. B. Trunov, "Learning approach to nonlinear fault diagnosis: detectability analysis," *IEEE Transactions on Automatic Control*, 45(4). 806-812, 2000.

[18] K. Rahnamai, P. Arabshahi, T. Y. Yan, T. Pham, and S. G. Finley, "An intelligent fault detection and isolation architecture for antenna arrays," *The Telecommunications and Data Acquisition Progress Report*, NASA.

[19] R. R. Selmic and F. L. Lewis, "Multimodel neural networks identification and failure detection of nonlinear systems," *Proc. 40th IEEE Conference on Decision and Control*, pp. 3128-3133, Orlando, FL, Dec. 2001.

[20] G. Tao, S. Chen, X. Tang, and S. M. Joshi, *Adaptive Control of Systems with Actuator Failures*, Springer-Verlag, London 2004.

[21] X. Zhang, M. M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *IEEE Transactions on Automatic Control*, 47(4). pp. 576-593, April 2002.

[22] X. Zhang, T. Parisini, and M. M. Polycarpou, "Adaptive fault-tolerant control of nonlinear uncertain systems: an information-based diagnostic approach," *IEEE Transactions on Automatic Control*, 48(8). pp. 1259-1274, Aug. 2004.