Information Sciences 384 (2017) 21–33

Contents lists available at ScienceDirect

Information Sciences

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

# Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems[☆]

Bo Zhao[a], Derong Liu[b,*], Yuanchun Li[c]

[a] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[c] Department of Control Science and Engineering, Changchun University of Technology, Changchun 130012, China

ARTICLE INFO

*Article history:*
Received 29 October 2015
Revised 7 December 2016
Accepted 13 December 2016
Available online 14 December 2016

*Keywords:*
Adaptive dynamic programming
Fault tolerant control
Policy iteration
Nonlinear systems
Fault observer
Neural network

ABSTRACT

This paper develops a novel fault tolerant control (FTC) scheme for a class of nonlinear systems with actuator failures based on adaptive dynamic programming (ADP). The estimated actuator failure from a fault observer is utilized to construct an improved performance index function that reflects the failure, regulation and control simultaneously. By employing a proper performance index function, the FTC problem can be transformed into an optimal control problem. By using policy iteration, the Hamilton–Jacobi–Bellman equation can be solved by constructing a critic neural network. Then, the approximated optimal controller can be derived directly. The closed-loop system is guaranteed to be uniformly ultimately bounded via the Lyapunov stability theorem. The effectiveness of the developed FTC scheme is demonstrated by two simulation examples. The significant contribution of the proposed strategy lies in that the well-known ADP method is extended to solving the FTC problem.



## 1. Introduction

Modern industries are becoming increasingly complex and large-scale to satisfy the requirement of the improving production efficiency [40]. Consequently, the demands for reliability and safety of sophisticated control systems are urgent. As we know, various components such as actuators, sensors and processors may undergo abrupt failures individually or simultaneously during operation, which may lead to serious damages [28]. Among all kinds of malfunctions, actuator failures significantly account for the degradation of control performance. Hence, it is important to develop fault tolerant control (FTC) methods to deal with such failures and maintain acceptable system performance.

In recent decades, many FTC approaches have been developed to solve the aforementioned problems through different theories and methods. In early stages of research, hardware redundancy based FTC was achieved by installing some backup components, which increased the weight, volume and cost. In contrast to hardware designs, analytical redundancy based FTC was distinguished for its low cost and flexible structures, and therefore, it attracted many researchers' attention. In general,

☆ This work was supported in part by the National Natural Science Foundation of China under Grants 61233001, 61273140, 61304086, 61374105, 61374051, 61533017, 61503377, 61503379, 61603387 and U1501251, and in part by the Scientific and Technological Development Plan Project in Jilin Province of China under Grants 20150520112JH and 20160414033GH.

* Corresponding author.
*E-mail addresses:* zhaobo@ia.ac.cn (B. Zhao), derong@ustb.edu.cn (D. Liu), liyc@ccut.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.ins.2016.12.016

analytical redundancy based FTC schemes could be categorized into passive approaches and active ones. By passive design, Zhou et al. [50] proposed an architecture that included two parts, i.e., the feedback control system was solely controlled by the performance controller, and the model uncertainties and external disturbances were handled by the robustness of the controller. Wang et al. [33] investigated a robust fault-tolerant $H_\infty$ control of active suspension systems with finite-frequency constraints. The scheme considered actuator faults, suspension deflection and actuator saturation simultaneously. In fact, the insensitivity of passive FTCs depends on their robustness. Thus, they are quite limited in handling large failures. On the contrary, active FTC methods provide stronger fault tolerant capability. Zhang and Jiang [46] and Hwang et al. [10] gave some excellent reviews on fault reconfiguration methods. It should be mentioned that the observer technique [16], which is active in reconfiguring control, plays an important role in achieving active fault tolerance. For example, Cristofaro et al. [5] reconfigured the control law by using an unknown input observer for fault detection and isolation (FDI) in overactuated systems. The presented FTC was independent of fault estimation. Tong et al. [28] investigated an observer-based adaptive decentralized fuzzy FTC scheme, which could be applied to a class of nonlinear large-scale systems with actuator failures and without satisfying the matching condition. Shi et al. [25] proposed a descriptor sliding mode observer approach using quantized signals for a class of Markovian jump systems against actuator faults. By injecting a quantizer density related parameter to the sliding mode input term, the observer possessed the ability to compensate the quantization effects. Fault accommodation strategy is another way to achieve the goal of active FTC. Yoo [43] investigated a time-delay independent fault detection and accommodation scheme, where an approximation-based fault accommodation design was activated to compensate for multiple time-delay faults after the faults were detected. Based on the fault compensation technique, Wang et al. [34] proposed an adaptive failure compensation control scheme with the nonlinear damping and parameter projection techniques for parametric strict feedback nonlinear systems. In this method, the control module and parameter estimator module were designed separately.

As an effective tool to deal with optimal control problems in nonlinear systems, adaptive/approximate dynamic programming (ADP) [12,32] is an useful approximation method proposed by Werbos [39] to solve Hamilton–Jacobi–Bellman (HJB) equations. ADP has many synonyms such as neuron-dynamic programming [3], neural dynamic programming [26], adaptive critic designs [24] and reinforcement learning [13]. Great efforts have been made on ADP and related research fields in the past decade. Many excellent results have significantly promoted the development of relevant disciplines. For ADP in optimal control, value iteration (VI) [14,38] and policy iteration (PI) [21,22] are fundamental algorithms. Al-Tamimi et al. [1] and Liu et al. [20] solved the HJB equation by using the VI technique. Zhong et al. [49] and Dong et al. [7] proposed a novel event-triggered ADP method to solve the optimal control problem of nonlinear systems. Compared with the traditional ADP design with a fixed sampling period, the method can reduce the computational cost and transmission load. From these studies, we can conclude that VI can avoid the requirement of initial stabilizing control, but it cannot guarantee the stability of the system. In fact, only the converged optimal control law can be used to control nonlinear systems [21]. In contrast to VI algorithms, PI algorithms, which consist of policy evaluation and policy improvement, could avoid this shortcoming because they begin with a stable control. Many researchers have investigated PI algorithms to solve optimal control for discrete-time linear systems [4,22], nonlinear systems with saturating actuators [2], and partially model-free systems [29]. Recently, Wang et al. [31] transformed the robust control problem into an optimal control problem, and established a robust control based on an online PI algorithm for nonlinear systems with uncertainties. Song et al. [27] redesigned an off-policy actor-critic structure to compensate the disturbances for unknown systems. Modares et al. [23] presented an online PI algorithm to obtain the optimal control for completely unknown constrained input systems. Liu et al. [19] developed a decentralized control strategy to stabilize the nonlinear interconnected system using an online learning optimal approach. Zhang et al. [44] proposed a data-driven robust approximated optimal tracking control scheme for unknown nonlinear systems by using the ADP method. It was the first solution to the tracking problem of unknown general nonlinear systems based on ADP. Throughout the literature on ADP, many excellent works have been carried out on various classes of nonlinear systems, such as continuous-time systems [31], discrete-time systems [35,36], nonlinear systems with uncertainties [27] and constraints [2], unknown nonlinear systems [15] and data-driven systems [37]. Due to the advantages of ADP methods in solving optimal control problem of nonlinear systems, they have been implemented in many real applications, such as transportation [47], temperature control systems [37], cooperative control [45], smart grid systems [42], multimachine power systems [11] and zero-sum game problems [17,18].

For the FTC problem, only a few results were presented based on ADP methods. Feng et al. [8] proposed a reconfigurable fault-tolerant deflection routing algorithm based on reinforcement learning for network on chip (NoC). It reconfigured the routing table through Q-learning and used 2-hop fault information to make efficient routing decisions to avoid faults. He et al. [9] developed a reinforcement-learning-based fast algorithm for proactive network fault management. Tradeoffs were made by a partially observable Markov decision process between the collected and transformed data, as well as between the speed and accuracy of fault detection and diagnosis. Zhu et al. [51] presented a novel approach to automate recovery policy generation with reinforcement learning techniques. The method could learn a new locally optimal policy to outperform the original one. Zhao et al. [48] established an online fault compensation control scheme based on PI algorithm for affine nonlinear systems with actuator failures. Yen et al. [6,41] proposed a supervisor which made use of two quality indices to perform fault detection, identification and isolation based on the knowledge stored in a dynamic model bank (DMB), which could reduce the reconfiguration time of the globalized dual heuristic programming (GDHP) controller, but they did not give a systematic FTC strategy.

Motivated by the above ADP and observer techniques, in this paper, a fault tolerant controller based on fault observer and ADP algorithm is developed for nonlinear systems with actuator failures. The estimated failure from a fault observer is utilized to construct the performance index function, which reflects the actuator failure, regulation and control. Hence, the FTC problem is transformed into an optimal control problem. A PI algorithm is employed to solve the HJB equation by constructing a critic neural network. Based on the Lyapunov stability theorem, the closed-loop system with actuator failures can be guaranteed to be uniformly ultimately bounded (UUB).

The main contribution of this paper can be summarized as follows. (i) By designing a fault observer, the estimated unknown actuator failure can be employed to construct the improved performance index function, which reflects the actuator failure, regulation and control. Thus, the FTC problem is transformed into an optimal control problem. (ii) The FTC can be derived depending only on the critic neural network. The training of the action neural network which was commonly implemented is no longer required. (iii) The well-known ADP method is extended to solve FTC problems. It implies that this is a novel approach against actuator failures in operated systems.

The rest of this paper is organized as follows. In Section 2, we present the problem statement. In Section 3, with the help of the estimated actuator fault from the fault observer, the performance index function is constructed to transform the FTC problem into an optimal control problem, and then the approximated fault tolerant controller is derived directly. In Section 4, two simulation examples are provided to demonstrate the effectiveness of the proposed scheme. In Section 5, the conclusion is drawn.

## 2. Problem statement

Consider the following continuous-time nonlinear system with an actuator failure:

$$\dot{x} = f(x) + g(x)(u - u_a) \tag{1}$$

where $x \in \mathbb{R}^n$ is the system state vector, $u \in \mathbb{R}^m$ is the control input vector, $f(\cdot)$ and $g(\cdot)$ are locally Lipchitz and differentiable in their arguments with $f(0) = 0$ and $u_a(t) \in \mathbb{R}^m$ is an additive actuator failure. Here, let $x(0) = x_0$ be the initial state.

**Assumption 1.** The actuator failure $u_a$ is unknown but norm-bounded as $\|u_a\| \le \delta_1 < +\infty$, where $\delta_1$ is a positive constant.

**Remark 1.** In real applications, the occurrence of actuator failures is random, and it is reasonable to assume that the unknown actuator failure $u_a$ is bounded.

Since $f(\cdot)$ and $g(\cdot)$ are locally Lipchitz continuous on a set $\Omega \subset \mathbb{R}^n$, the system (1) with $u_a = 0$ is controllable. Thus, it is desired to find the feedback control policy $u(x)$ which minimizes the infinite horizon performance index function as

$$J(x_0) = \int_0^{\infty} \Big(\rho \hat{u}_a^{\mathsf{T}}(\tau)\hat{u}_a(\tau) + U(x(\tau), u(\tau))\Big)\mathrm{d}\tau, \tag{2}$$

where $\rho$ is a positive constant, $U(x, u) = x^{\mathsf{T}}Qx + u^{\mathsf{T}}Ru$ is the utility function, $U(0, 0) = 0$, and $U(x, u) \ge 0$ for all $x$ and $u$, $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are positive definite matrices, and $\hat{u}_a \in \mathbb{R}^m$ is the estimation of the actuator failure $u_a$. We can see that (2) appropriately reflects the actuator failure, regulation and control simultaneously.

To obtain an acceptable control performance after the actuator failure occurs, the designed feedback control must be admissible. Therefore, before the algorithm is presented, the definition of admissible control is introduced [1].

**Definition 1.** For system (1) with $u_a = 0$, a control policy $u(x)$ is said to be admissible, if $u(x)$ is continuous on a set $\Omega \subset \mathbb{R}^n$, $u(0) = 0$, $u(x)$ stabilizes the system, and $J(x_0)$ in (2) is finite for all $x_0 \in \Omega$.

For any admissible control policy $\mu \in \Psi(\Omega)$, where $\Psi(\Omega)$ denotes the set of admissible control, if the performance index function

$$V(x_0) = \int_0^{\infty} \Big(\rho \hat{u}_a^{\mathsf{T}}(\tau)\hat{u}_a(\tau) + U(x(\tau), \mu(\tau))\Big)\mathrm{d}\tau \tag{3}$$

is continuously differentiable, then the infinitesimal version of (3) is the so-called nonlinear Lyapunov equation

$$0 = \rho \hat{u}_a^{\mathsf{T}}\hat{u}_a + U(x, \mu) + (\nabla V(x))^{\mathsf{T}}(f(x) + g(x)\mu) \tag{4}$$

with $V(0) = 0$, and the term $\nabla V(x)$ denotes the partial derivative of $V(x)$ with respect to $x$, i.e., $\nabla V(x) = \frac{\partial V(x)}{\partial x}$.

Define the Hamiltonian of the problem and the optimal performance index function as

$$H(x, \mu, \nabla V(x)) = \rho \hat{u}_a^{\mathsf{T}}\hat{u}_a + U(x, \mu) + (\nabla V(x))^{\mathsf{T}}(f(x) + g(x)\mu)$$

and

$$J^*(x_0) = \min_{\mu\in\Psi(\Omega)} \int_0^{\infty} \Big(\rho \hat{u}_a^{\mathsf{T}}\hat{u}_a + U(x(\tau), \mu(\tau))\Big)\mathrm{d}\tau \tag{5}$$

respectively. Let $J^*(x)$ be the optimal performance index function, then

$$0 = \min_{\mu\in\Psi(\Omega)} H(x, \mu, \nabla J^*(x)), \tag{6}$$

where $\nabla J^*(x) = \frac{\partial J^*(x)}{\partial x}$. If the solution $J^*(x)$ exists and is continuously differentiable, the optimal control can be expressed as

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla J^*(x). \tag{7}$$

In general, if the system is fault-free (i.e., $u_a = 0$), the solution of (6) can be approximated with the PI technique (see Algorithm 1).

**Algorithm 1** Online policy iteration.

**Step 1** Let $i = 0$, begin with an initial admissible control policy $\mu^{(0)}(x)$, and select a small positive constant $\varepsilon$;
**Step 2** Let $i > 0$, based on the control policy $\mu^{(i)}(x)$, solve $V^{(i+1)}$ from

$$0 = \rho\hat{u}_a^{\mathsf{T}}\hat{u}_a + U(x, \mu^{(i)}) + \left(\nabla V^{(i+1)}(x)\right)^{\mathsf{T}}(f(x) + g(x)\mu^{(i)}).$$

**Step 3** Update the control policy by

$$\mu^{(i+1)} = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla V^{(i)}(x).$$

**Step 4** If $i > 0$ and $\left\|V^{(i+1)}(x) - V^{(i)}(x)\right\| \le \varepsilon$, stop and obtain the approximate optimal control; else, let $i = i + 1$ and return to Step 2.

By simple transformation, (7) implies

$$(\nabla J^*(x))^{\mathsf{T}}g(x) = -2(u^*(x))^{\mathsf{T}}R. \tag{8}$$

## 3. Fault tolerant controller design via observer based ADP

### 3.1. Problem transformation

For system (1), a feedback control policy $u(x)$ should be presented to deal with the FTC problem, such that the closed-loop system can be guaranteed to be UUB for all possible actuator failures $u_a$. In order to achieve this objective, we will transform the FTC problem into designing an optimal controller for the fault-free system, i.e., $u_a = 0$, with a proper performance index function.

**Assumption 2.** The actuator fault estimation error $e_a = u_a - \hat{u}_a$ is norm-bounded as $\|e_a\| \le \delta_2$, where $\delta_2$ is a positive constant.

**Theorem 1.** *Consider system (1) with $u_a = 0$, with Assumptions 1 and 2, and the control policy (7), the continuously differentiable function $J^*(x)$ is a Lyapunov function with the conditions $\rho \ge \lambda_{\min}(R)$ and $\|x\| \ge \frac{\delta_a}{\lambda_{\min}(Q)}$ satisfied. Furthermore, assume that $J^*(x)$ is a solution to the HJB Eq. (6). Thus, the optimal control policy (7) can guarantee the closed-loop nonlinear system with an actuator failure (1) to be UUB, i.e., $u^*(x)$ in (7) is a solution to the FTC problem.*

**Proof.** In order to prove that $u^*(x)$ in (7) is a solution to the FTC problem, we prove that $J^*(x)$ is a Lyapunov function. According to (5), we can see that $J^*(x) > 0$ for all $x \neq 0$ and $J^*(0) = 0$. It means that $J^*(x)$ is a positive definite function. Thus, its time derivative is

$$\begin{aligned}
\dot{J}^*(x) &= (\nabla J^*(x))^{\mathsf{T}}\dot{x} \\
&= (\nabla J^*(x))^{\mathsf{T}}(f(x) + g(x)(u^* - u_a)) \\
&= (\nabla J^*(x))^{\mathsf{T}}(f(x) + g(x)u^*) - (\nabla J^*(x))^{\mathsf{T}}g(x)u_a.
\end{aligned} \tag{9}$$

From (6), we have

$$(\nabla J^*(x))^{\mathsf{T}}(f(x) + g(x)u^*) = -\hat{u}_a^{\mathsf{T}}\hat{u}_a - U(x, u^*).$$

Hence, (9) becomes

$$\dot{J}^*(x) = -\hat{u}_a^{\mathsf{T}}\hat{u}_a - U(x, u^*) - (\nabla J^*(x))^{\mathsf{T}}g(x)u_a.$$

Noticing (8), we have

$$\begin{aligned}
\dot{J}^*(x) &= -\rho\hat{u}_a^{\mathsf{T}}\hat{u}_a - U(x, u^*) + 2u^{*\mathsf{T}}Ru_a \\
&= -\rho\hat{u}_a^{\mathsf{T}}\hat{u}_a - x^{\mathsf{T}}Qx - u^{*\mathsf{T}}Ru^* + 2u^{*\mathsf{T}}Ru_a \\
&\le -\rho\hat{u}_a^{\mathsf{T}}\hat{u}_a - x^{\mathsf{T}}Qx - u^{*\mathsf{T}}Ru^* + u^{*\mathsf{T}}Ru^* + u_a^{\mathsf{T}}Ru_a \\
&\le -\rho\hat{u}_a^{\mathsf{T}}\hat{u}_a - x^{\mathsf{T}}Qx + \lambda_{\min}(R)\|u_a\|^2
\end{aligned}$$

$$
\begin{aligned}
&= -\big(\rho - \lambda_{\min}(R)\big)\hat{u}_a^{\mathrm{T}}\hat{u}_a - x^{\mathrm{T}}Qx + \lambda_{\min}(R)\left(\|u_a\|^2 - \left\|\hat{u}_a\right\|^2\right) \\
&= -\big(\rho - \lambda_{\min}(R)\big)\hat{u}_a^{\mathrm{T}}\hat{u}_a - x^{\mathrm{T}}Qx + \lambda_{\min}(R)\big(\|u_a\| + \left\|\hat{u}_a\right\|\big)\big(\|u_a\| - \left\|\hat{u}_a\right\|\big) \\
&= -\big(\rho - \lambda_{\min}(R)\big)\hat{u}_a^{\mathrm{T}}\hat{u}_a - x^{\mathrm{T}}Qx + \lambda_{\min}(R)\big(\|u_a\| + \left\|\hat{u}_a - u_a + u_a\right\|\big)\big(\left\|u_a - \hat{u}_a\right\|\big) \\
&\le -\big(\rho - \lambda_{\min}(R)\big)\hat{u}_a^{\mathrm{T}}\hat{u}_a - x^{\mathrm{T}}Qx + \lambda_{\min}(R)\big(2\|u_a\| + \left\|\hat{u}_a - u_a\right\|\big)\big(\left\|u_a - \hat{u}_a\right\|\big).
\end{aligned} \tag{10}
$$

From Assumptions 1 and 2, we can obtain

$$\dot{J}^*(x) \le -\big(\rho - \lambda_{\min}(R)\big)\hat{u}_a^{\mathrm{T}}\hat{u}_a - x^{\mathrm{T}}Qx + \lambda_{\min}(R)\big(2\delta_1 + \delta_2\big)\delta_2.$$

Let $\delta_a = \lambda_{\min}(R)\big(2\delta_1 + \delta_2\big)\delta_2$. Thus

$$\dot{J}^*(x) \le -\big(\rho - \lambda_{\min}(R)\big)\hat{u}_a^{\mathrm{T}}\hat{u}_a - \lambda_{\min}(Q)\|x\|^2 + \delta_a.$$

Hence, we can conclude that $\dot{J}^*(x) \le 0$, i.e., $J^*(x)$ is a Lyapunov function, if the following conditions hold:

$$\begin{cases} \rho \ge \lambda_{\min}(R) \\ \|x\| \ge \frac{\delta_a}{\lambda_{\min}(Q)}. \end{cases}$$

This indicates that $x(t)$ will converge to a small neighborhood surrounding the initial position. This ensures the equivalence of problem transformation. This completes the proof. □

In light of Theorem 1, the optimal control policy (7) is derived to handle the FTC problem for system (1), which suffers from the actuator failure. However, it is difficult to find the solution to the HJB function and derive $\hat{u}_a$ to construct the performance index function (3).

### *3.2. Fault observer design*

For the system with actuator failure (1), we can develop a fault observer as

$$\dot{\hat{x}} = \hat{f}(x) + \hat{g}(x)(u - \hat{u}_a) + L_1(x - \hat{x}) \tag{11}$$

where $\hat{x}$ is the observation of the system state $x$, $L_1$ is the positive definite observer gain matrix, and $\hat{u}_a$ is the estimated actuator failure which can be updated by the following adaptive law:

$$\dot{\hat{u}}_a = -L_2\hat{g}^{\mathrm{T}}(x)e_o \tag{12}$$

where $L_2$ is a positive definite matrix, and $e_o = x - \hat{x}$ is the state observation error. Combining (1) with (11), we have

$$
\begin{aligned}
\dot{e}_o &= e_f + e_g u - g(x)u_a + \hat{g}(x)\hat{u}_a - L_1 e_o \\
&= e_f + e_g(u - u_a) - \hat{g}(x)(u_a - \hat{u}_a) - L_1 e_o
\end{aligned} \tag{13}
$$

where $e_f = f(x) - \hat{f}(x)$ and $e_g = g(x) - \hat{g}(x)$ are the observation errors of nonlinear terms $f(x)$ and $g(x)$, respectively. Define $w = e_f + e_g(u - u_a)$.

**Assumption 3.** $w$ is norm-bounded as $\|w\| \le \delta_3$, where $\delta_3$ is a positive constant.

**Theorem 2.** *For the system with actuator failure (1) with Assumptions 1 and 3, the fault observation error can be guaranteed to be UUB through the developed fault observer (11) with the adaptive law (12).*

**Proof.** Select the Lyapunov function candidate as

$$\Sigma_1 = \frac{1}{2}e_o^{\mathrm{T}}e_o + \frac{1}{2}\tilde{u}_a^{\mathrm{T}}L_2^{-1}\tilde{u}_a \tag{14}$$

where $\tilde{u}_a = u_a - \hat{u}_a$ is the estimation error of the actuator failure.

Substituting (13) into the time derivative of (14), one can obtain

$$
\begin{aligned}
\dot{\Sigma}_1 &= e_o^{\mathrm{T}}\dot{e}_o - \dot{\hat{u}}_a^{\mathrm{T}}L_2^{-1}\tilde{u}_a \\
&= e_o^{\mathrm{T}}\big(e_f + e_g(u - u_a) - \hat{g}(x)(u_a - \hat{u}_a) - L_1 e_o\big) - \dot{\hat{u}}_a^{\mathrm{T}}L_2^{-1}\tilde{u}_a \\
&\le \delta_3\|e_o\| - e_o^{\mathrm{T}}\hat{g}(x)\tilde{u}_a - \lambda_{\min}(L_1)\|e_o\|^2 - \dot{\hat{u}}_a^{\mathrm{T}}L_2^{-1}\tilde{u}_a \\
&= -(\lambda_{\min}(L_1)\|e_o\| - \delta_3)\|e_o\| - \big(e_o^{\mathrm{T}}\hat{g}(x) + \dot{\hat{u}}_a^{\mathrm{T}}L_2^{-1}\big)\tilde{u}_a.
\end{aligned} \tag{15}
$$

Substituting the adaptive law (12) into (15), it becomes

$$\dot{\Sigma}_1 = -(\lambda_{\min}(L_1)\|e_o\| - \delta_3)\|e_o\|.$$

This indicates $\dot{\Sigma}_1 < 0$ as long as $e_o$ lies outside the compact set $\|e_o\| \le \frac{\delta_3}{\lambda_{\min}(L_1)}$. According to the Lyapunov stability theorem, the fault observation error is UUB. This completes the proof. □

### 3.3. Online PI algorithm

In this subsection, we will introduce an online PI algorithm to overcome the difficulty in solving the HJB equation. The PI algorithm consists of policy evaluation based on (4) and policy improvement based on (7), and its iteration process can be described as follows.

This algorithm will converge to the optimal performance index function and optimal control policy, i.e., $V^{(i)}(x) \to J^*(x)$ and $\mu^{(i)}(x) \to u^*(x)$ as $i \to \infty$ [30].

### 3.4. Neural network implementation

In this section, a single-layer neural network $V(x)$ is employed to approximate the assumed differentiable performance index function on the compact set $\Omega$ as

$$V(x) = W_c^{\mathsf{T}} \sigma(x) + \varepsilon_c(x) \tag{16}$$

where $W_c \in \mathbb{R}^l$ is the ideal weight vector, $\sigma(x) \in \mathbb{R}^l$ is the activation function, $l$ is the number of neurons in the hidden layer, and $\varepsilon_c(x)$ is the approximation error of the neural network. Then, the gradient of (16) along with $x$ is

$$\nabla V(x) = (\nabla\sigma(x))^{\mathsf{T}} W_c + \nabla\varepsilon_c(x) \tag{17}$$

where $\nabla\sigma(x) = \frac{\partial\sigma(x)}{\partial x} \in \mathrm{R}^{l\times n}$ and $\nabla\varepsilon_c(x)$ are the gradients of the activation function and the approximation error, respectively.

Substituting (17) into (4), one can obtain

$$0 = \rho \hat{u}_a^{\mathsf{T}} \hat{u}_a + U(x, \mu) + \big((\nabla\sigma(x))^{\mathsf{T}} W_c + \nabla\varepsilon_c(x)\big)\dot{x}.$$

Thus, the Hamiltonian can be expressed as

$$H(x, \mu, W_c) = \rho \hat{u}_a^{\mathsf{T}} \hat{u}_a + U(x, \mu) + W_c^{\mathsf{T}} \nabla\sigma(x)\dot{x} = -\nabla\varepsilon_c(x)\dot{x} \triangleq e_{cH} \tag{18}$$

where $e_{cH}$ is the residual error due to the neural network approximation.

Since the ideal weight of vector $W_c$ is unknown, the critic neural network is employed to approximate $V(x)$ as

$$\hat{V}(x) = \hat{W}_c^{\mathsf{T}} \sigma_c(x).$$

Then, the gradient of $\hat{V}(x)$ can be expressed as

$$\nabla\hat{V}(x) = (\nabla\sigma(x))^{\mathsf{T}} \hat{W}_c.$$

Thus, the approximate Hamiltonian can be obtained as

$$H(x, \mu, \hat{W}_c) = \rho \hat{u}_a^{\mathsf{T}} \hat{u}_a + U(x, \mu) + \hat{W}_c^{\mathsf{T}} \nabla\sigma(x)\dot{x} \triangleq e_c.$$

Let $\theta = \nabla\sigma_c(x)\dot{x}$, and define the weight approximation error as $\tilde{W}_c = W_c - \hat{W}_c$. By (18) and (20), one has

$$e_c = e_{cH} - \tilde{W}_c^{\mathsf{T}} \theta.$$

The weight approximation error can be updated as

$$\dot{\tilde{W}}_c = -\dot{\hat{W}}_c = l(e_{cH} - \tilde{W}_c^{\mathsf{T}} \theta)\theta. \tag{19}$$

In order to tune the critic neural network weight vector $\hat{W}_c$, the objective function $E_c = \frac{1}{2} e_c^{\mathsf{T}} e_c$ should be minimized by the normalized gradient algorithm. $\hat{W}_c$ should be updated as

$$\dot{\hat{W}}_c = -l e_c \theta \tag{20}$$

where $l > 0$ is the learning rate of the critic neural network.

Hence, according to (7) and (16), the ideal control policy can be described as

$$\mu(x) = -\frac{1}{2} R^{-1} g^{\mathsf{T}}(x) \big((\nabla\sigma(x))^{\mathsf{T}} W_c + \nabla\varepsilon_c(x)\big).$$

It can be approximated as

$$\hat{\mu}(x) = -\frac{1}{2} R^{-1} g^{\mathsf{T}}(x) (\nabla\sigma(x))^{\mathsf{T}} \hat{W}_c. \tag{21}$$

The structural diagram of the observer based ADP scheme for FTC is depicted in Fig. 1.

**Remark 2.** From the approximate control policy (21), we can observe that it can be derived depending only on the critic neural network, whose weight vector can be updated by (20). Meanwhile, the training of the action neural network is no longer required. Hence, it is feasible for the implementation and computation.

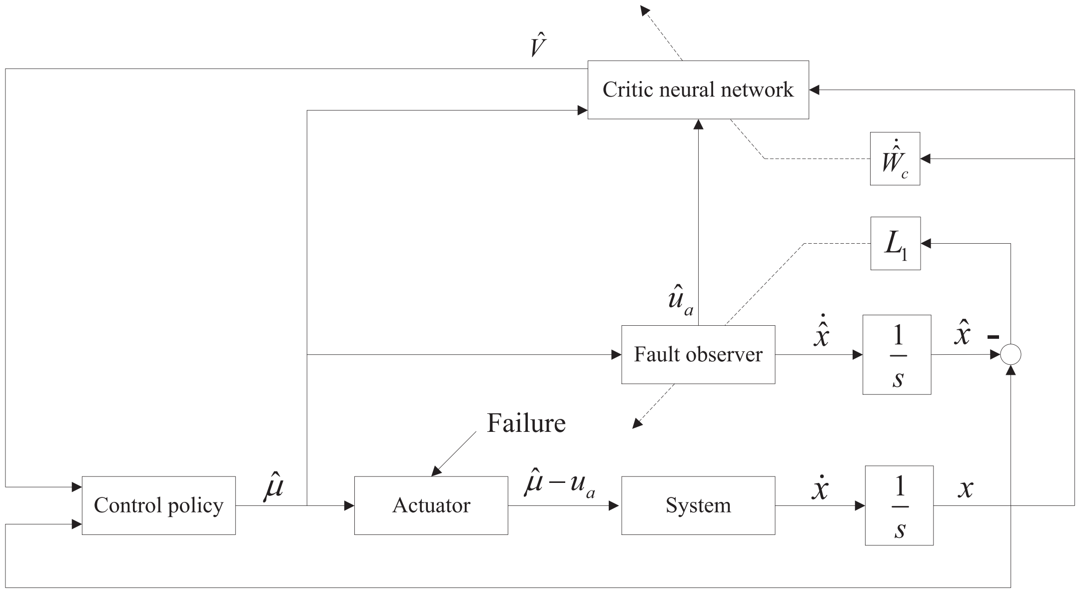


**Fig. 1.** The structural diagram of observer based ADP scheme for FTC.

**Theorem 3.** *Considering the nonlinear system (1) without an actuator failure, if the weights of the critic neural network are updated by (19), then the weight approximation error is UUB.*

**Proof.** Select the Lyapunov function candidate as

$$\Sigma_2 = \frac{1}{2l}\tilde{W}_c^{\mathsf{T}}\tilde{W}_c.$$

Its time derivative is

$$\begin{aligned}\dot{\Sigma}_2 &= \frac{1}{l}\tilde{W}_c^{\mathsf{T}}\dot{\tilde{W}}_c \\ &= \tilde{W}_c^{\mathsf{T}}\left(e_{cH} - \tilde{W}_c^{\mathsf{T}}\theta\right)\theta \\ &= \tilde{W}_c^{\mathsf{T}}e_{cH}\theta - \left\|\tilde{W}_c^{\mathsf{T}}\theta\right\|^2 \\ &\le \frac{1}{2}e_{cH}^2 - \frac{1}{2}\left\|\tilde{W}_c^{\mathsf{T}}\theta\right\|^2.\end{aligned}$$

Hence, $\dot{\Sigma}_2 < 0$ if $\tilde{W}_c$ lies outside the compact set $\|\tilde{W}_c\| \le \|\frac{e_{cH}}{\theta_M}\|$ with assumption $\|\theta\| \le \theta_M$, where $\theta_M$ is a positive constant. According to the Lyapunov stability theorem, the weight approximation error is UUB. This completes the proof. □

### 3.5. Stability analysis

**Theorem 4.** *Assume that the neural network based HJB solution to the optimal control problem exists. For the considered system (1), the approximate FTC policy (21) can guarantee the closed-loop system UUB with the performance index function (2).*

**Proof.** Select the Lyapunov function candidate as

$$\Sigma_3 = \frac{1}{2}x^{\mathsf{T}}x + J^*(x).$$

Its time derivative is

$$\begin{aligned}\dot{\Sigma}_3 &= x^{\mathsf{T}}\dot{x} + (\nabla J^*(x))^{\mathsf{T}}\dot{x} \\ &= x^{\mathsf{T}}(f(x) + g(x)(\mu - u_a)) + (\nabla J^*(x))^{\mathsf{T}}(f(x) + g(x)(\mu - u_a)) \\ &= x^{\mathsf{T}}f(x) + x^{\mathsf{T}}g(x)\mu - x^{\mathsf{T}}g(x)u_a - (\nabla J^*(x))^{\mathsf{T}}g(x)u_a + (\nabla J^*(x))^{\mathsf{T}}(f(x) + g(x)\mu).\end{aligned} \tag{22}$$

According to (4), (22) becomes

$$\dot{\Sigma}_3 = x^{\mathsf{T}}f(x) + x^{\mathsf{T}}g(x)\mu - x^{\mathsf{T}}g(x)u_a - (\nabla J^*(x))^{\mathsf{T}}g(x)u_a - \rho\hat{u}_a^{\mathsf{T}}\hat{u}_a - x^{\mathsf{T}}Qx - \mu^{\mathsf{T}}R\mu.$$

As $f(x)$ is locally Lipchitz, there exists a positive constant $D_f$ such that $\|f(x)\| \le D_f \|x\|$. Assume that $\|g(x)\| \le D_g$. By Young's inequality, we can obtain

$$\begin{aligned}\dot{\Sigma}_3 &\le D_f\|x\|^2 + \frac{1}{2}\|x\|^2 + \frac{1}{2}D_g^2\|\mu\|^2 + \frac{1}{2}\|u_a\|^2 + \frac{1}{2}D_g^2\|x\|^2 - (\nabla J^*(x))^{\mathsf{T}}g(x)u_a - \rho\hat{u}_a^{\mathsf{T}}\hat{u}_a \\ &\quad - \lambda_{\min}(Q)\|x\|^2 - \lambda_{\min}(R)\|\mu\|^2 \\ &= -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}D_g^2 - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}D_g^2\right)\|\mu\|^2 + \frac{1}{2}\|u_a\|^2 \\ &\quad - (\nabla J^*(x))^{\mathsf{T}}g(x)u_a - \rho\hat{u}_a^{\mathsf{T}}\hat{u}_a. \end{aligned} \tag{23}$$

Substituting (8) into (23), it follows

$$\begin{aligned}\dot{\Sigma}_3 &\le -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}D_g^2 - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}D_g^2\right)\|\mu\|^2 + \frac{1}{2}\|u_a\|^2 + 2\mu^{\mathsf{T}}Ru_a - \rho\hat{u}_a^{\mathsf{T}}\hat{u}_a \\ &\le -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}D_g^2 - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}D_g^2 - R^2\right)\|\mu\|^2 + \frac{3}{2}\|u_a\|^2 - \rho\hat{u}_a^{\mathsf{T}}\hat{u}_a \\ &= -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}D_g^2 - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}D_g^2 - R^2\right)\|\mu\|^2 \\ &\quad - \left(\rho - \frac{3}{2}\right)\hat{u}_a^{\mathsf{T}}\hat{u}_a + \frac{3}{2}\left(\|u_a\|^2 - \hat{u}_a^{\mathsf{T}}\hat{u}_a\right). \end{aligned}$$

Noticing (9), it follows

$$\begin{aligned}\dot{\Sigma}_3 &\le -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}D_g^2 - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}D_g^2 - R^2\right)\|\mu\|^2 - \left(\rho - \frac{3}{2}\right)\hat{u}_a^{\mathsf{T}}\hat{u}_a + \frac{3}{2}(2\delta_1 + \delta_2)\delta_2 \\ &= -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}D_g^2 - \frac{1}{2}\right)\|x\|^2 - \left(\lambda_{\min}(R) - \frac{1}{2}D_g^2 - R^2\right)\|\mu\|^2 - \left(\rho - \frac{3}{2}\right)\hat{u}_a^{\mathsf{T}}\hat{u}_a + \delta_b \end{aligned}$$

where $\delta_b = \frac{3}{2}(2\delta_1+\delta_2)\delta_2$. Hence, we can observe that $\dot{\Sigma}_3 \le 0$ when $x(t)$ lies outside the compact set $\|x\| \le \frac{\delta_b}{\lambda_{\min}(Q)-D_f-D_g^2-\frac{1}{2}}$ if the following conditions hold:

$$\begin{cases}\lambda_{\min}(Q) \ge D_f + \frac{1}{2}D_g^2 + \frac{1}{2} \\ \lambda_{\min}(R) - R^2 \ge \frac{1}{2}D_g^2 \\ \rho \ge \frac{3}{2}.\end{cases}$$

Therefore, the state trajectories of the closed-loop system under the FTC input are UUB. This completes the proof. □

## 4. Simulation studies

In order to show the effectiveness of the proposed FTC based on ADP, two simulation examples are given in this section.

**Example 1.** A torsional pendulum system is employed to examine the control performance of the proposed FTC scheme [21]. The system dynamics of the torsional pendulum with an actuator failure is as follows:

$$\begin{cases}\dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega \\ J\dfrac{\mathrm{d}\omega}{\mathrm{d}t} = (u - u_a) - Mgl\sin\theta - f_d\dfrac{\mathrm{d}\theta}{\mathrm{d}t}.\end{cases}$$

where $M = \frac{1}{3}$ kg and $l = \frac{2}{3}$ m are the mass and length of the pendulum bar, respectively. The angle $\theta$ and the angular velocity $\omega$ are the system states. $u_a(t) \in \mathbb{R}$ is an unknown actuator failure, which we choose as

$$u_a(t) = \begin{cases}0, & 0 \le t \le 30\,\mathrm{s} \\ 0.3 + 0.5\sin\left(\frac{3t}{2\pi}\right), & 30\,\mathrm{s} < t \le 60\,\mathrm{s}.\end{cases}$$

Let $J = \frac{4}{3}Ml^2$ and $f_d = 0.2$ be the rotary inertia and frictional factor, respectively. Let $g = 9.8$ m/s$^2$ be the gravitational acceleration.

Define $x = [x_1, x_2]^{\mathsf{T}} = [\theta, \omega]^{\mathsf{T}} \in \mathbb{R}^2$. Let the initial state of the torsional pendulum system and the observed state be $x_0 = [1, -1]^{\mathsf{T}}$ and $\hat{x}_0 = [2, -2]^{\mathsf{T}}$, and the admissible control be $u = [-0.4, -0.8]x$. In this simulation, the performance index function is approximated by a critic neural network, whose weight vector is denoted as $\hat{W}_c = [\hat{W}_{c1}, \hat{W}_{c2}, \hat{W}_{c3}]^{\mathsf{T}}$, and its initial value is $\hat{W}_{c0} = [5, 3, 7]^{\mathsf{T}}$. The activation function of the critic neural network is chosen as $\sigma_c(x) = [x_1^2,\ x_1x_2,\ x_2^2]$. Let $Q = I_2$ and $R = 0.05I_1$, where $I_n$ denotes the $n \times n$ identity matrix, the fault observer gain be $L_1 = 20I_2$, the learning rate of the critic neural network and actuator failure be $l_1 = 0.01$ and $L_2 = 50I_1$, and the gain in performance index function (3) be $\rho = 8$, respectively. The initial value of the actuator failure is chosen as $u_{a0} = 1$.

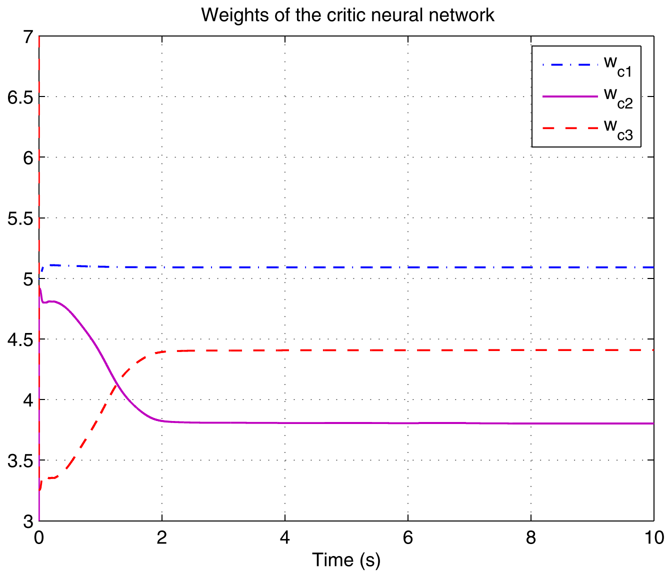


**Fig. 2.** The weights of the critic neural network.

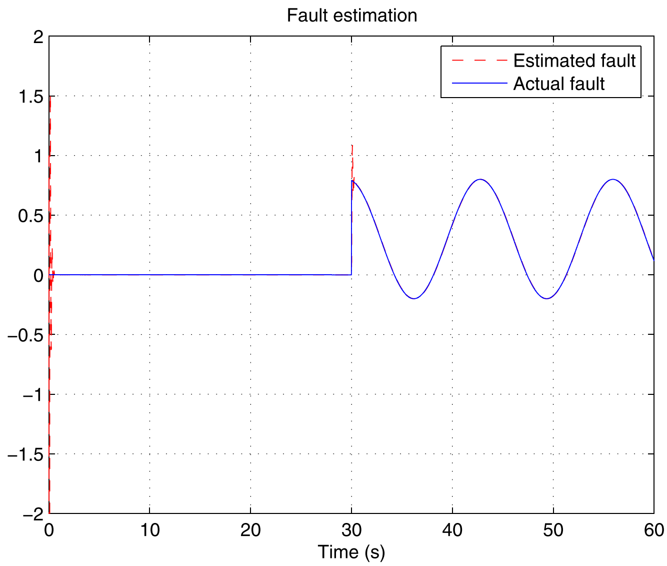


**Fig. 3.** The estimated fault based on the fault observer.

The simulation results are illustrated in Figs. 2–5. We can see in Fig. 2, the weights of the critic neural network converge to $[5.091011,\ 3.802569,\ 4.409337]^{\mathsf{T}}$ with the proposed algorithm. As the estimated value of actuator failure $\hat{u}_a$ is required in the constructed performance index function (3), Fig. 3 shows the precise estimation performance of the actuator failure, which is the key point in the proposed FTC method. As shown in Fig. 4, the converged system states under the ADP based FTC are shown for the system suffers from an actuator failure. Fig. 5 illustrates the control input. We can observe that after the failure occurs at $t = 30\,\mathrm{s}$, the control input presents a change against the failure, so that an acceptable control performance can be derived. Therefore, the simulation results demonstrate that the proposed FTC scheme is effective.

**Example 2.** Consider the following nonlinear system with an actuator failure:

$$\dot{x} = \begin{bmatrix} x_2 - x_1 \\ -0.5x_1 - 0.5x_2 + 0.5x_2(\cos(2x_1) + 2)^2 \\ x_4 - x_3 \\ -x_3 - 0.5x_4 + 0.5x_4x_3^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \cos(2x_1) + 2 & 0 \\ 0 & 0 \\ 0 & x_3 \end{bmatrix}(u - u_a)$$

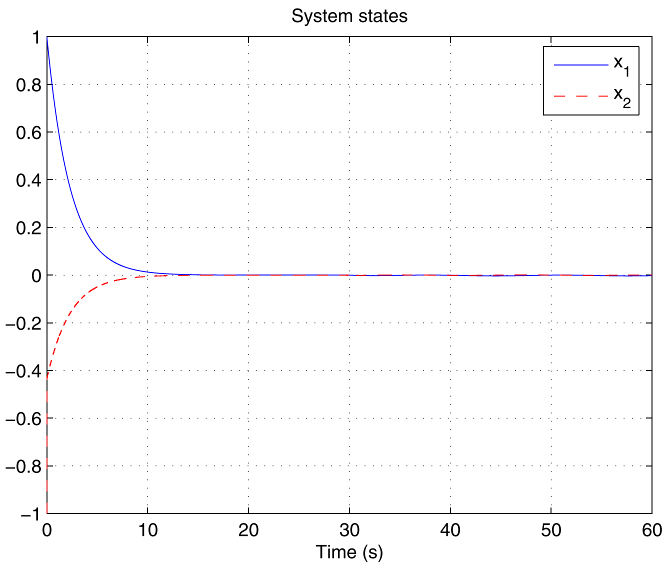


**Fig. 4.** The system states under the ADP based FTC input.

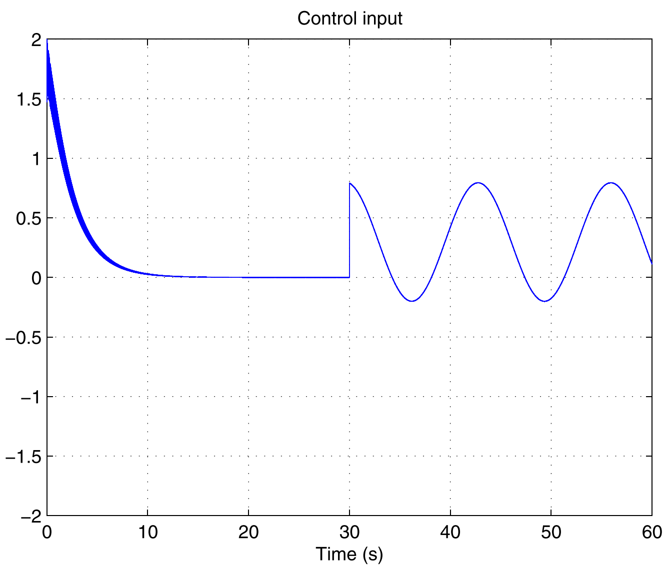


**Fig. 5.** The system input of the proposed FTC scheme.

where $x = [x_1, x_2, x_3, x_4]^\mathsf{T} \in \mathbb{R}^4$ and $u = [u_1, u_2]^\mathsf{T} \in \mathbb{R}^2$ are the state and control input variables, respectively. The term $u_a = [u_{a1}, u_{a2}]^\mathsf{T} \in \mathbb{R}^2$ reflects the unknown actuator failure, which we choose as

$$u_a(t) = \begin{cases} [0, 0]^\mathsf{T}, & 0 \le t \le 10\,\mathrm{s} \\ [6, 0]^\mathsf{T}, & 10\,\mathrm{s} < t \le 20\,\mathrm{s} \end{cases}$$

for the purpose of simulation.

Let the initial system state be $x_0 = [1, -1, 1, -1]^\mathsf{T}$, the initial observed state be $\hat{x}_0 = [2, -2, 2, -2]^\mathsf{T}$, and the initial admissible control policy be $u = [-0.2, -0.4, -0.6, -0.8]x$, respectively. We employ a critic neural network to approximate the performance index function, and its weight vector is denoted as $\hat{W}_c = \left[\hat{W}_{c1}, \hat{W}_{c2}, \ldots, \hat{W}_{c10}\right]^\mathsf{T}$, whose initial value is $\hat{W}_{c0} = [20, 25, 80, 10, 90, 15, 60, 50, 35, 75]^\mathsf{T}$. The activation function of the critic neural network is chosen as $\sigma_c(x) = \left[x_1^2,\ x_1x_2,\ x_1x_3,\ x_1x_4,\ x_2^2,\ x_2x_3,\ x_2x_4,\ x_3^2,\ x_3x_4,\ x_4^2\right]$. Let $Q = I_4$, $R = 0.05I_2$, the fault observer gain be $L_1 = 20I_4$, the learning rate of the critic neural network and actuator failure be $l_1 = 0.5$ and $L_2 = 50I_2$, and the gain in performance index function be $\rho = 5$, respectively. The initial value of the actuator failure vector is chosen as $u_{a0} = [0, 0]^\mathsf{T}$.

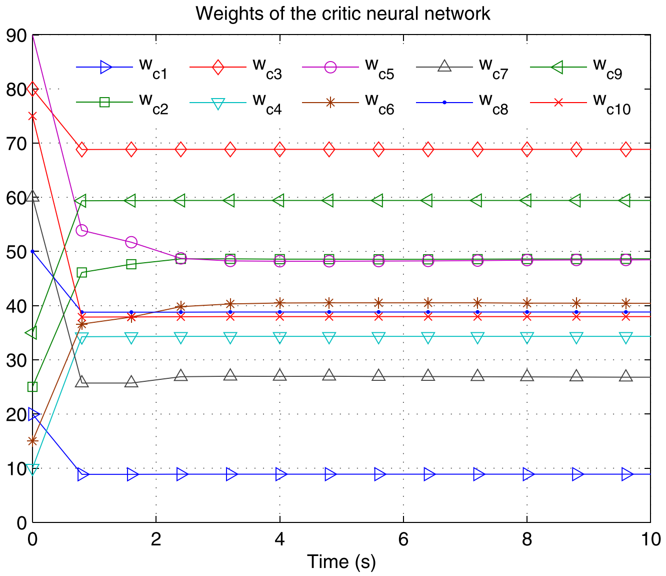


**Fig. 6.** The weights of the critic neural network.

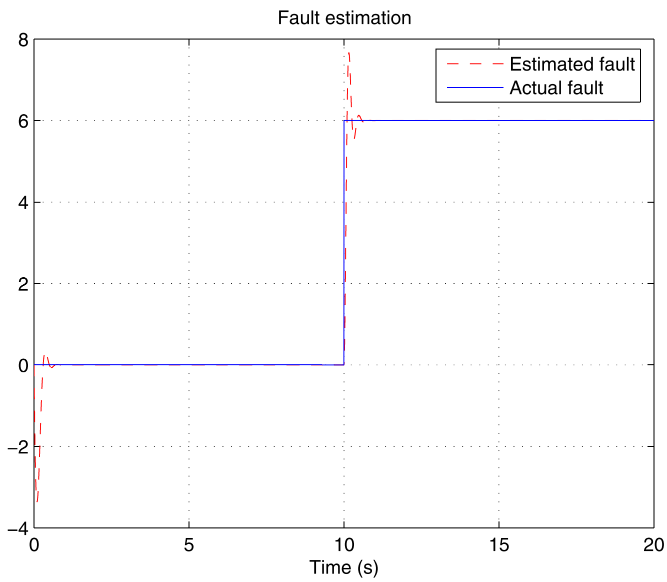


**Fig. 7.** The estimated fault based on the fault observer.

By using the proposed algorithm, the weights of the critic neural network converge to $[8.883293, 48.669910, 68.846855, 34.323722, 48.565825, 40.415194, 26.800737, 38.819658, 59.425341, 37.962929]^T$. The simulation results are shown as Figs. 6–9, and from these figures, we can conclude similar results as in Example 1. Therefore, we can declare the effectiveness of the developed ADP based FTC scheme.

## 5. Conclusions

An observer based ADP algorithm for the FTC problem of a class of nonlinear systems with actuator failures is developed in this paper. With the help of the estimated failure from the fault observer, a novel performance index function is constructed to account for the system failure. Thus, the FTC problem can be transformed into an optimal control problem. A critic neural network is constructed to solve the improved HJB equation online, and the approximated optimal controller can be directly derived. Based on the Lyapunov stability theorem, the closed-loop system is guaranteed to be UUB. Two numerical simulations are provided to reinforce the theoretical results.

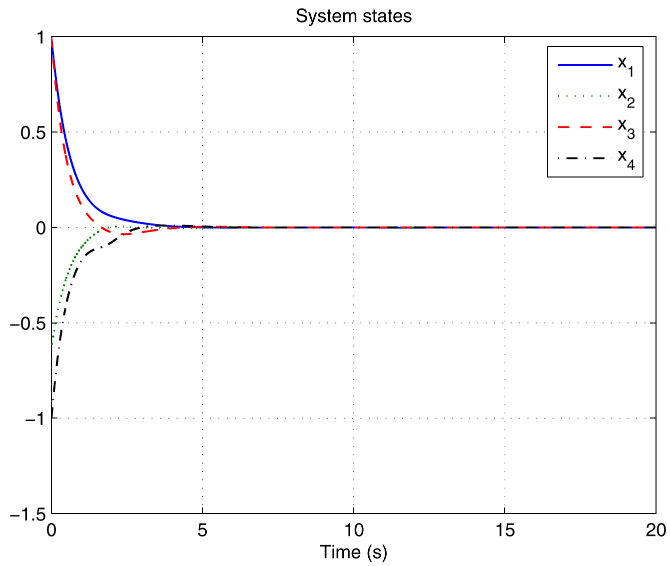


**Fig. 8.** The system states under the ADP based FTC input.

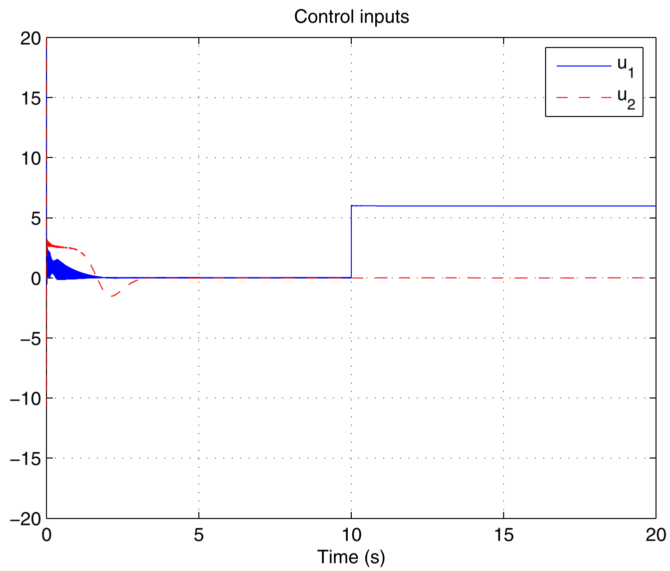


**Fig. 9.** The system input of the proposed FTC scheme.

## References


[1] A. Al-Tamimi, F.L. Lewis, M. Abu-Khalaf, Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof, IEEE Trans. Syst. Man Cybern. Part B Cybern. 38 (4) (2008) 943–949.

[2] M. Abu-Khalaf, F.L. Lewis, Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach, Automatica 41 (5) (2005) 779–791.

[3] D.P. Bertsekas, M.L. Homer, D. Logan, S.D. Patek, N.R. Sandell, Missile defense and interceptor allocation by neuro-dynamic programming, IEEE Trans. Syst. Man Cybern. Part A Syst. Humans 30 (1) (2000) 42–51.

[4] S.J. Bradtke, B.E. Ydestie, A.G. Barto, Adaptive linear quadratic control using policy iteration, in: Proceedings of American Control Conference, Baltimore, MD, 1994, pp. 3475–3476.

[5] A. Cristofaro, T.A. Johansen, Fault tolerant control allocation using unknown input observers, Automatica 50(7) (2014) 1891–1897.

[6] P. DeLima, G. Yen, Accommodating controller malfunctions through fault tolerant control architecture, IEEE Trans. Aerosp. Electron. Syst. 43 (2) (2007) 706–722.

[7] L. Dong, X. Zhong, C. Sun, H. He, Adaptive event-triggered control based on heuristic dynamic programming for nonlinear discrete-time systems, IEEE Trans. Neural Networks Learn. Syst. (2016), doi:10.1109/TNNLS.2016.2541020.

[8] C. Feng, Z. Lu, A. Jantsch, J. Li, M. Zhang, A reconfigurable fault-tolerant deflection routing algorithm based on reinforcement learning for network-on-chip, in: Proceedings of the Third International Workshop on Network on Chip Architectures, 2010, pp. 11–16.

[9] Q. He, M.A. Shayman, Using reinforcement learning for pro-active network fault management, in: Proceedings of International Conference on Communication Technology, 2000, pp. 515–521.

[10] I. Hwang, S. Kim, Y. Kim, C.E. Seah, A survey of fault detection, isolation, and reconfiguration methods, IEEE Trans. Control Syst. Technol. 18 (3) (2010) 636–653.
[11] Y. Jiang, Z. Jiang, Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems, IEEE Trans. Circuits-II 59 (10) (2012) 693–697.
[12] F. Lewis, D. Liu, Reinforcement Learning and Approximate Dynamic Programming for Feedback Control, Wiley-IEEE Press, 2013.
[13] F.L. Lewis, D. Vrabie, K.G. Vamvoudakis, Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers, IEEE Control Syst. 32(6) (2012) 76–105.
[14] H. Li, D. Liu, Optimal control for discrete-time affine non-linear systems using general value iteration, IET Control Theory Appl. 6 (18) (2012) 2725–2736.
[15] H. Li, D. Liu, D. Wang, Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics, IEEE Trans. Autom. Sci. Eng. 11 (3) (2014) 706–714.
[16] F. Li, L. Wu, P. Shi, C.C. Lim, State estimation and sliding mode control for semi-Markovian jump systems with mismatched uncertainties, Automatica 51 (2015) 385–393.
[17] D. Liu, H. Li, D. Wang, Neural-network-based zero-sum game for discrete-time nonlinear systems via iterative adaptive dynamic programming algorithm, Neurocomputing 110 (2013) 92–100.
[18] D. Liu, H. Li, D. Wang, Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics, IEEE Trans. Syst. Man Cybern. A Syst. 44 (8) (2014) 1015–1027.
[19] D. Liu, D. Wang, H. Li, Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach, IEEE Trans. Neural Networks Learn. Sys. 25 (2) (2014) 418–428.
[20] D. Liu, D. Wang, D. Zhao, Q. Wei, N. Jin, Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming, IEEE Trans. Autom. Sci. Eng. 9 (3) (2012) 628–634.
[21] D. Liu, Q. Wei, Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems, IEEE Trans. Neural Networks Learn. Syst. 25 (3) (2014) 621–634.
[22] D. Liu, Q. Wei, P. Yan, Generalized policy iteration adaptive dynamic programming for discrete-time nonlinear systems, IEEE Trans. Syst. Man Cybern. A Syst. 45 (12) (2015) 1577–1591.
[23] H. Modares, F.L. Lewis, M.B. Naghibi-Sistani, Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks, IEEE Trans. Neural Networks Learn. Syst. 24 (10) (2013) 1513–1525.
[24] D.V. Prokhorov, D.C. Wunsch, Adaptive critic designs, IEEE Trans. Neural Networks 8 (5) (1997) 997–1007.
[25] P. Shi, M. Liu, L.X. Zhang, Fault-tolerant sliding-mode-observer synthesis of markovian jump systems using quantized measurements, IEEE Trans. Ind. Electron. 62 (9) (2015) 5910–5918.
[26] J. Si, Y.T. Wang, On-line learning control by association and reinforcement, IEEE Trans. Neural Networks 12 (2) (2000) 221–226.
[27] R. Song, F.L. Lewis, Q. Wei, H. Zhang, Off-policy actor-critic structure for optimal control of unknown systems with disturbances, IEEE Trans. Cybern. 46 (5) (2015) 1041–1050.
[28] S. Tong, B. Huo, Y. Li, Observer-based adaptive decentralized fuzzy fault-tolerant control of nonlinear large-scale systems with actuator failures, IEEE Trans. Fuzzy Syst. 22 (1) (2014) 1–15.
[29] D. Vrabie, F. Lewis, Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems, Neural Networks 22 (3) (2009) 237–246.
[30] K.G. Vamvoudakis, F.L. Lewis, Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem, Automatica 46 (5) (2010) 878–888.
[31] D. Wang, D. Liu, H. Li, Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems, IEEE Trans. Autom. Sci. Eng. 11 (2) (2014) 627–632.
[32] F.Y. Wang, H. Zhang, D. Liu, Adaptive dynamic programming: an introduction, IEEE Comput. Intell. Mag. 4 (2) (2009) 39–47.
[33] R. Wang, H. Jing, H.R. Karimi, N. Chen, Robust fault-tolerant $h\infty$ control of active suspension systems with finite-frequency constraint, Mech. Syst. Signal Process. 62 (2015) 341–355.
[34] W. Wang, C. Wen, Adaptive compensation for infinite number of actuator failures or faults, Automatica 47 (10) (2011) 2197–2210.
[35] Q. Wei, D. Liu, A novel $\theta$-iterative-adaptive dynamic programming for discrete-time nonlinear systems, IEEE Trans. Autom. Sci. Eng. 11 (4) (2014) 1176–1190.
[36] Q. Wei, D. Liu, Numerical adaptive learning control scheme for discrete-time non-linear systems, IET Control Theory Appl. 7 (11) (2013) 1472–1486.
[37] Q. Wei, D. Liu, Data-driven neuro-optimal temperature control of water-gas shift reaction using stable iterative adaptive dynamic programming, IEEE Trans. Ind. Electron. 61 (11) (2014) 6399–6408.
[38] Q. Wei, F.Y. Wang, D. Liu, et al., Finite-approximation-error-based discrete-time iterative adaptive dynamic programming, IEEE Trans. Cybern. 44 (12) (2014) 2820–2833.
[39] P. Werbos, Approximate dynamic programming for real-time control and neural modeling, in: Handbook of Intelligent Control Neural Fuzzy & Adaptive Approaches, 1992.
[40] L. Wu, R. Yang, P. Shi, X. Su, Stability analysis and stabilization of 2-d switched systems under arbitrary and restricted switchings, Automatica 59 (2015) 206–215.
[41] G.G. Yen, P.G. DeLima, Improving the performance of globalized dual heuristic programming for fault tolerant control through an online learning supervisor, IEEE Trans. Autom. Sci. Eng. 2 (2) (2005) 121–131.
[42] R. Yu, W. Zhong, S. Xie, Y. Zhang, Y. Zhang, QoS differential scheduling in cognitive-radio-based smart grid networks: an adaptive dynamic programming approach, IEEE Trans. Neural Networks Learn. Syst. (2015), doi:10.1109/TNNLS.2015.2411673.
[43] S.J. Yoo, Fault detection and accommodation of a class of nonlinear systems with unknown multiple time-delayed faults, Automatica 50 (1) (2014) 255–261.
[44] H. Zhang, L. Cui, X. Zhang, et al., Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method, IEEE Trans. Neural Networks 22 (12) (2011) 2226–2236.
[45] H. Zhang, J. Zhang, G.H. Yang, Y. Luo, Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming, IEEE Trans. Fuzzy Syst. 23 (1) (2015) 152–163.
[46] Y. Zhang, J. Jiang, Bibliographical review on reconfigurable fault-tolerant control systems, Annu. Rev. Control 32 (2) (2008) 229–252.
[47] D. Zhao, Y. Dai, Z. Zhang, Computational intelligence in urban traffic signal control: a survey, IEEE Trans. Syst. Man Cybern. C 42 (4) (2012) 485–494.
[48] B. Zhao, D. Liu, Y. Li, Online fault compensation control based on policy iteration algorithm for a class of affine non-linear systems with actuator failures, IET Control Theory Appl. 10 (15) (2016) 1816–1823.
[49] X. Zhong, H. He, An event-triggered ADP control approach for continuous-time system with unknown internal states, IEEE Trans. Cybern. (2016), doi:10.1109/TCYB.2016.2523878.
[50] K. Zhou, Z. Ren, A new controller architecture for high performance, robust, and fault-tolerant control, IEEE Trans. Autom. Control 46 (10) (2001) 1613–1618.
[51] Q. Zhu, C. Yuan, A reinforcement learning approach to automatic error recovery, in: Proceeding of IEEE 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, 2007, pp. 729–738.