Information Sciences 328 (2016) 435–454

Contents lists available at ScienceDirect

Information Sciences

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

Online approximate solution of HJI equation for unknown constrained-input nonlinear continuous-time systems☆

Xiong Yang, Derong Liu*, Hongwen Ma, Yancai Xu

The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 26 December 2014
Revised 30 April 2015
Accepted 1 September 2015
Available online 9 September 2015

Keywords:
Adaptive dynamic programming
Hamilton–Jacobi–Isaacs equation
Input constraint
Neural network
Optimal control
Reinforcement learning

ABSTRACT

This paper is concerned with the approximate solution of Hamilton–Jacobi–Isaacs (HJI) equation for constrained-input nonlinear continuous-time systems with unknown dynamics. We develop a novel online adaptive dynamic programming-based algorithm to learn the solution of the HJI equation. The present algorithm is implemented via an identifier-critic architecture, which consists of two neural networks (NNs): an identifier NN is applied to estimate the unknown system dynamics and a critic NN is constructed to obtain the approximate solution of the HJI equation. An advantage of the proposed architecture is that the identifier NN and the critic NN are tuned simultaneously. With introducing two additional terms, namely, the stabilizing term and the robustifying term to update the critic NN, the initial stabilizing control is no longer required. Meanwhile, the developed critic tuning rule not only ensures convergence of the critic to the optimal saddle point but also guarantees stability of the closed-loop system. Moreover, the uniform ultimate boundedness of the weights of the identifier NN and the critic NN are proved by using Lyapunov's direct method. Finally, to illustrate the effectiveness and applicability of the developed approach, two simulation examples are provided.

1. Introduction

Over the past several decades, H_∞ optimal control problems for nonlinear systems have attracted intensive attention. Many remarkable results have been obtained in this filed [3–5,10,30,31], especially the results reported in [4,31]. In [4], Basar and Bernhard showed that the H_∞ optimal control problem is equivalent to the minimax optimization problem, which is termed as two-player zero-sum games where the controller is a minimizing player and the exogenous disturbance is a maximizing one. In [31], by using the theory of dissipative systems, van der Schaft transformed the H_∞ optimal control problem to the L_2-gain optimal control problem. Nevertheless, the bottleneck for applying theories of the H_∞ optimal control in practice still exists. This is mainly because the solutions of two-player zero-sum games and L_2-gain optimal control problems are often required to solve the Hamilton–Jacobi–Isaacs (HJI) equations. It is well-known that Hamilton–Jacobi–Isaacs (HJI) equations for nonlinear systems are actually nonlinear first-order partial differential equations (PDEs), which are difficult or impossible to solve by analytical approaches.

☆ This work was supported in part by the National Natural Science Foundation of China under Grants 61034002, 61233001, 61273140, 61304086, and 61374105, in part by Beijing Natural Science Foundation under Grant 4132078, and in part by the Early Career Development Award of the State Key Laboratory of Management and Control for Complex Systems (SKLMCCS).

* Corresponding author. Tel.: +86-10-82544761; fax: +86-10-82544799.
E-mail addresses: xiong.yang@ia.ac.cn (X. Yang), derong.liu@ia.ac.cn, derongliu@gmail.com (D. Liu), mahongwen2012@ia.ac.cn (H. Ma), yancai.xu@ia.ac.cn (Y. Xu).

http://dx.doi.org/10.1016/j.ins.2015.09.001
0020-0255/

Since accurate solutions of HJI equations are intractable to obtain, an increasing number of researchers pay their attentions to deriving approximate solutions of this kind of equations. In the past few years, adaptive dynamic programming (ADP) methods have been successfully used to solve HJI equations. The ADP approach was first introduced by Werbos [36]. After that, various ADP methods were proposed (see surveys [15,35]). A distinct feature of the ADP approach is that it employs neural networks (NNs) to derive the approximate optimal control forward in time. Due to this feature, the curse of dimensionality can be avoided while applying ADP approaches to solve the Hamilton–Jacobi–Bellman/HJI equations [27]. In light of this advantage, ADP methods have been extensively utilized to solve HJI equations.

For discrete-time nonlinear systems, Mehraeen et al. [20] presented an offline ADP-based iterative approach to solve the HJI equation for zero-sum two-player games. By using the proposed method and Taylor series expansion, a sufficient condition for the convergence to the saddle point is obtained. After that, Liu et al. [16] developed a greedy iterative ADP algorithm to solve the HJI equations associated with zero-sum two-player games. Based on the algorithm, three NNs referred to as action NN, critic NN and disturbance NN can approximate the optimal control, the optimal value and the worst disturbance, respectively. Later, Zhang et al. [45] proposed an online ADP-based algorithm to learn the solution of the HJI equation for a class of H_∞ control problems. By the algorithm given in [45], the prior knowledge of the nonlinear system is not required.

For continuous-time (CT) nonlinear systems, the HJI equations are often approximately solved by using reinforcement learning (RL), which is considered as a special case of ADP approaches by Werbos [37]. Abu-Khalaf et al. [2] introduced an offline RL-based algorithm to give the approximate solution of the HJI equation for constrained-input nonlinear systems. After that, Luo et al. [19] proposed an off-policy RL method to solve the HJI equation of H_∞ control problems. Differing from [2], the algorithm given in [19] generated the system data by arbitrary policies rather than evaluating policies. Recently, Vamvoudakis and Lewis [33] introduced an online RL-based algorithm to solve the HJI equation for two-player zero-sum games. By using the algorithm, the actor, critic and disturbance NNs were tuned simultaneously. Distinct from the above online RL-based algorithm, Dierks and Jagannathan [7] developed a single online approximator-based scheme to solve the HJI equation. Based on the algorithm, only a single critic NN is employed to learn the solution of the HJI equation and the initial stabilizing control is not required. It should be mentioned that, prior knowledge of system dynamics is required to be available in [7,33]. After that, Luo and Wu [38] presented a simultaneous policy update algorithm to solve the HJI equation arising in nonlinear H_∞ control problems. By the proposed algorithm, the internal dynamics of nonlinear system is not required. Later, Liu et al. [17] employed the simultaneous policy update algorithm to obtain the approximate solution of the HJI equation for multi-player nonzero-sums with completely unknown dynamics. More recently, Johnson et al.[12] developed a projection algorithm to give the approximate solution of the coupled HJI equations for uncertain nonlinear CT systems.

To the best of authors' knowledge, there are still no ADP-based algorithms proposed to solve the HJI equation for constrained-input nonlinear CT systems with unknown dynamics. In this paper, we develop a novel online ADP-based algorithm to learn the solution of the HJI equation for unknown constrained-input nonlinear CT systems. The present algorithm is implemented via an identifier-critic architecture, which consists of two neural networks (NNs): an identifier NN is utilized to estimate the unknown system dynamics and a critic NN is constructed to obtain the approximate solution of the HJI equation. An advantage of the present architecture is that the identifier NN and the critic NN are tuned simultaneously. With introducing two additional terms, namely, the stabilizing term and the robustifying term to update the critic NN, no initial stabilizing control is required. Meanwhile, the developed critic tuning rule not only ensures convergence of the critic to the optimal saddle point but also guarantees stability of the closed-loop system. In addition, Lyapunov's direct method is utilized to demonstrate the uniform ultimate boundedness of the weights of the identifier NN and the critic NN.

It is significant to point out that, though our methodology in this work is in a similar spirit as [7], this paper extends the work of [7] to give an online approximate solution of the HJI equation for constrained-input nonlinear CT systems with unknown dynamics. Solving the HJI equation of unknown constrained-input nonlinear CT systems is more intractable than those with the knowledge of system dynamics regardless of control constraints.

The rest of the paper is organized as follows. Section 2 provides preliminaries of H_∞ optimal control problems for constrained-input nonlinear CT systems. Section 3 presents the design of identifier NNs for unknown controlled systems with stability proof. Section 4 develops a single critic NN to approximate the solution of the HJI equation. Section 5 shows the stability analysis. Section 6 presents two numerical examples to verify the effectiveness of the developed method. Finally, Section 7 gives several concluding remarks and potential future extensions.

Notations: $\mathbb{R}$ represents the set of all real numbers. $\mathbb{R}^m$ denotes the Euclidean space of all real m-vectors. $\mathbb{R}^{n\times m}$ denotes the space of all $n \times m$ real matrices. I_n represents the $n \times n$ identity matrix. T is the transposition symbol. C^m represents the class of functions having continuous mth derivative. When $\bar{\xi} = [\bar{\xi}_1, \ldots, \bar{\xi}_m]^\mathsf{T} \in \mathbb{R}^m$, $\|\bar{\xi}\| = \left(\sum_{i=1}^m |\bar{\xi}_i|^2\right)^{1/2}$ denotes the Euclidean norm of $\bar{\xi}$. When $A \in \mathbb{R}^{m\times m}$, $\|A\| = \left(\lambda_{\max}(A^\mathsf{T}A)\right)^{1/2}$ denotes the 2-norm of A, where $\lambda_{\max}(A^\mathsf{T}A)$ represents the maximum eigenvalue of $A^\mathsf{T}A$.

2. Preliminaries and problem statement

Consider the nonlinear CT system described by

$$\begin{aligned} \dot{x} &= f(x) + g(x)u + k(x)\omega, \\ z &= h(x) + p(x)u, \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathfrak{U} \subset \mathbb{R}^m$ is the control input, $\mathfrak{U} = \{u \in \mathbb{R}^m : |u_i| \le \kappa, i = 1, \ldots, m\}$, and $\kappa > 0$ is the saturating bound. $\omega(t) \in \mathbb{R}^{q_1}$ is the exogenous disturbance, and $z(t) \in \mathbb{R}^{q_2}$ is the fictitious output, $f(x) \in \mathbb{R}^n$, $g(x) \in \mathbb{R}^{n\times m}$, $k(x) \in \mathbb{R}^{n\times q_1}$, $h(x) \in \mathbb{R}^{q_2}$, $p(x) \in \mathbb{R}^{q_2\times m}$ with $f(0) = 0$, and $x = 0$ is the equilibrium point of the system.

Assumption 1. $f(x)$, $g(x)$, and $k(x)$ are unknown smooth functions defined on $\mathbb{R}^n$. $\omega(t) \in L_2[0, \infty)$, and it implies that there exists a constant $\omega_M > 0$ such that $\|\omega(t)\| \le \omega_M$. In addition, $h^{\mathsf{T}}(x)p(x) = 0$ and $p^{\mathsf{T}}(x)p(x) = I$ for every $x(t) \in \mathbb{R}^n$.

The objective for general H_∞ optimal control problems is to find a state feedback control $u(x)$ such that system (1) is locally asymptotically stable (when $\omega(t) = 0$), and there exists L_2-gain less than or equal to γ, that is,

$$\int_0^\infty \|z(t)\|^2 \mathrm{d}t = \int_0^\infty \left(h^{\mathsf{T}}h + \|u\|^2\right)\mathrm{d}t \le \gamma^2 \int_0^\infty \|\omega(t)\|^2 \mathrm{d}t,$$

where $\gamma > 0$ is a prescribed level of the disturbance attenuation. Noticing that u is constrained (i.e., $u \in \mathfrak{U}$) and motivated by the work of [2,23], this problem can be transformed to solve the zero-sum game

$$V^*(x_0) = \min_u \max_\omega \int_0^\infty \left(h^{\mathsf{T}}h + Y(u) - \gamma^2\|\omega\|^2\right)\mathrm{d}t, \tag{2}$$

where

$$Y(u) = 2\kappa \int_0^u \tanh^{-1}(\upsilon/\kappa)\mathrm{d}\upsilon = 2\kappa \sum_{i=1}^m \int_0^{u_i} \tanh^{-1}(\upsilon_i/\kappa)\mathrm{d}\upsilon_i.$$

By (2), the value function for system (1) is given as

$$V(x(t)) = \int_t^\infty \left(h^{\mathsf{T}}(x(s))h(x(s)) + 2\kappa \int_0^{u(s)} \tanh^{-1}(\upsilon/\kappa)\mathrm{d}\upsilon - \gamma^2\|\omega(s)\|^2\right)\mathrm{d}s. \tag{3}$$

According to [23], if $V(x(t)) \in C^1$, then the Hamiltonian for the control u, the disturbance ω, and the value function $V(x)$ can be defined as

$$H(x, V_x, u, \omega) = V_x^{\mathsf{T}}\left(f(x) + g(x)u + k(x)\omega\right) + h^{\mathsf{T}}(x)h(x) + 2\kappa \int_0^u \tanh^{-1}(\upsilon/\kappa)\mathrm{d}\upsilon - \gamma^2\|\omega\|^2, \tag{4}$$

where $V_x \in \mathbb{R}^n$ represents the partial derivative of $V(x)$ with respect to x.

The optimal value $V^*(x_0)$ given in (2) can be obtained by solving the equation

$$\min_u \max_\omega H\left(x, V_x^*, u, \omega\right) = 0. \tag{5}$$

Remark 1. If the saddle point does not exist for the two player zero-sum game, there might be many solutions to (5) [4,46]. In this sense, it is intractable to derive the optimal value $V^*(x_0)$. To avoid this case, similar to [23,33], we require that the following condition holds

$$\min_u \max_\omega H\left(x, V_x^*, u, \omega\right) = \max_\omega \min_u H\left(x, V_x^*, u, \omega\right), \tag{6}$$

which guarantees the existence of the saddle point. Then, (5) has a unique solution. Actually, as shown in [1], (6) is valid when the optimal control u^* and the worst disturbance ω^* are obtained.

Combining (4) with (5), we obtain the optimal control and the worst disturbance, respectively, as

$$u^*(x) = -\kappa \tanh\left(\frac{1}{2\kappa}g^{\mathsf{T}}(x)V_x^*\right), \tag{7}$$

$$\omega^*(x) = \frac{1}{2\gamma^2}k^{\mathsf{T}}(x)V_x^*. \tag{8}$$

Substituting (7) and (8) into (5), we derive the HJI equation for the nonlinear system as

$$\begin{aligned} &V_x^{*\mathsf{T}}f(x) - 2\kappa^2\mathfrak{A}^{\mathsf{T}}(x)\tanh\left(\mathfrak{A}(x)\right) + h^{\mathsf{T}}(x)h(x) \\ &\quad + \frac{1}{4\gamma^2}V_x^{*\mathsf{T}}k(x)k^{\mathsf{T}}(x)V_x^* + 2\kappa \int_0^{-\kappa\tanh(\mathfrak{A}(x))} \tanh^{-\mathsf{T}}(\upsilon/\kappa)\mathrm{d}\upsilon = 0, \end{aligned} \tag{9}$$

where $\mathfrak{A}(x) = \frac{1}{2\kappa}g^{\mathsf{T}}(x)V_x^*$.

Denote $\mathfrak{A}(x) = \left[\mathfrak{A}_1(x), \ldots, \mathfrak{A}_m(x)\right]^{\mathsf{T}} \in \mathbb{R}^m$ with $\mathfrak{A}_i(x) \in \mathbb{R}$, $i = 1, \ldots, m$. Notice that

$$2\kappa \int_0^{-\kappa \tanh(\mathfrak{A}(x))} \tanh^{-\mathsf{T}}(\upsilon/\kappa)\mathrm{d}\upsilon = 2\kappa \sum_{i=1}^{m} \int_0^{-\kappa \tanh(\mathfrak{A}_i(x))} \tanh^{-\mathsf{T}}(\upsilon_i/\kappa)\mathrm{d}\upsilon_i$$
$$= 2\kappa^2 \mathfrak{A}^{\mathsf{T}}(x)\tanh(\mathfrak{A}(x)) + \kappa^2 \sum_{i=1}^{m} \ln\left[1 - \tanh^2(\mathfrak{A}_i(x))\right].$$

Then, the HJI equation (9) becomes

$$V_x^{*\mathsf{T}} f(x) + \kappa^2 \sum_{i=1}^{m} \ln\left[1 - \tanh^2(\mathfrak{A}_i(x))\right] + h^{\mathsf{T}}(x)h(x) + \frac{1}{4\gamma^2} V_x^{*\mathsf{T}} k(x) k^{\mathsf{T}}(x) V_x^* = 0. \tag{10}$$

From (10), one can find that it is actually a nonlinear PDE with respect to $V^*(x)$. It is difficult to solve by analytical approaches. In this paper, we shall develop an online ADP-based algorithm to approximately solve (10). The present algorithm is implemented without using policy iteration and value iteration. Only a single critic NN is constructed to derive the approximate solution of (10). In addition, differing from [2,23], the prior knowledge of $f(x)$, $g(x)$, and $k(x)$ given in (10) are all completely unavailable. Therefore, to solve (10), we need first to get the knowledge of $f(x)$, $g(x)$, and $k(x)$. In what follows, a dynamic NN is constructed to identify the unknown system dynamics.

3. Identifier design via dynamic NNs

According to [42], the first equation of system (1) can be represented by a dynamic NN as

$$\dot{x} = Ax + W_f^{\mathsf{T}} \phi(x) + W_g^{\mathsf{T}} \rho(x) u + W_k^{\mathsf{T}} \ell(x) \omega + \varepsilon(x), \tag{11}$$

where $A \in \mathbb{R}^{n\times n}$ is a Hurwitz matrix, $W_f \in \mathbb{R}^{n\times n}$, $W_g \in \mathbb{R}^{n\times n}$, and $W_k \in \mathbb{R}^{n\times n}$ are ideal NN weight matrices, and $\varepsilon(x) \in \mathbb{R}^n$ is the NN function reconstruction error. The vector function $\phi(x) \in \mathbb{R}^n$ is assumed to be n-dimensional with the elements increasing monotonically. The matrix function $\rho(x) \in \mathbb{R}^{n\times m}$ is assumed to be $\rho(x) = \left[\rho_1(\zeta_1^{\mathsf{T}} x), \ldots, \rho_n(\zeta_n^{\mathsf{T}} x)\right]^{\mathsf{T}}$, where $\zeta_i \in \mathbb{R}^{n\times m}$ is a constant matrix and $\rho_i(\cdot)$ is a bounded nondecreasing function. In addition, $\ell(x) \in \mathbb{R}^{n\times q_1}$ is set to be $\ell(x) = \left[\ell_1(\varsigma_1^{\mathsf{T}} x), \ldots, \ell_n(\varsigma_n^{\mathsf{T}} x)\right]^{\mathsf{T}}$, where $\varsigma_i \in \mathbb{R}^{n\times q_1}$ is a constant matrix and $\ell_i(\cdot)$ is a bounded nondecreasing function. The typical presentations of $\phi(x)$, $\rho(x)$ and $\ell(x)$ are sigmoid functions, such as $\tanh(x)$. In this paper, $\phi(x)$, $\rho(x)$ and $\ell(x)$ are selected to be sigmoid functions. Noticing the property of sigmoid functions, we assume that, for arbitrary $\xi_1, \xi_2 \in \mathbb{R}^n$, the following inequality holds:

$$\|\mathscr{A}(\xi_1) - \mathscr{A}(\xi_2)\| \le \lambda_{\mathscr{A}} \|\xi_1 - \xi_2\|, \tag{12}$$

where $\mathscr{A}$ denotes ϕ, ρ and ℓ, respectively, and $\lambda_{\mathscr{A}}$ ($\mathscr{A} = \phi, \rho, \ell$) are known positive constants.

In this paper, we employ the dynamic NN identifier to approximate the first equation of system (1) as

$$\dot{\hat{x}} = A\hat{x} + \hat{W}_f^{\mathsf{T}} \phi(\hat{x}) + \hat{W}_g^{\mathsf{T}} \rho(\hat{x}) u + \hat{W}_k^{\mathsf{T}} \ell(\hat{x}) \omega + \nu, \tag{13}$$

where $\hat{x} \in \mathbb{R}^n$ is the dynamic NN state, $\hat{W}_f \in \mathbb{R}^{n\times n}$, $\hat{W}_g \in \mathbb{R}^{n\times n}$, and $\hat{W}_k \in \mathbb{R}^{n\times n}$ are dynamic NN weight estimates, and $\nu = \eta\tilde{x}$ with the design parameter $\eta > 0$ and the identification error $\tilde{x} \triangleq x - \hat{x}$.

By using (11) and (13), the identification error dynamics can be derived as

$$\dot{\tilde{x}} = A\tilde{x} + W_f^{\mathsf{T}} \tilde{\phi} + W_g^{\mathsf{T}} \tilde{\rho} u + W_k^{\mathsf{T}} \tilde{\ell} \omega - \eta\tilde{x}$$
$$+ \tilde{W}_f^{\mathsf{T}} \phi(\hat{x}) + \tilde{W}_g^{\mathsf{T}} \rho(\hat{x}) u + \tilde{W}_k^{\mathsf{T}} \ell(\hat{x}) \omega + \varepsilon(x), \tag{14}$$

where $\tilde{W}_f = W_f - \hat{W}_f$, $\tilde{W}_g = W_g - \hat{W}_g$, $\tilde{W}_k = W_k - \hat{W}_k$, $\tilde{\phi} = \phi(x) - \phi(\hat{x})$, $\tilde{\rho} = \rho(x) - \rho(\hat{x})$, and $\tilde{\ell} = \ell(x) - \ell(\hat{x})$.

Before proceeding further, we provide some assumptions and facts. These assumptions are common techniques, which have been used in [13,14,28,42,44].

Assumption 2. The ideal dynamic NN weight matrices W_f, W_g, and W_k satisfy

$$W_f^{\mathsf{T}} W_f \le \Lambda_1, \quad W_g^{\mathsf{T}} W_g \le \Lambda_2, \quad W_k^{\mathsf{T}} W_k \le \Lambda_3,$$

where Λ_i ($i = 1, 2, 3$) are prior known positive definite matrices.

Assumption 3. The NN function reconstruction error $\varepsilon(x)$ is bounded; that is, there exists a known constant $b_\varepsilon > 0$ such that $\|\varepsilon(x)\| < b_\varepsilon$.

Fact 1. Since A is a Hurwitz matrix, there exists a positive-definite symmetric matrix $P \in \mathbb{R}^{n\times n}$ satisfying the Lyapunov equation

$$A^{\mathsf{T}} P + PA = -\beta I_n,$$

where $\beta > 0$ is a design parameter.

Fact 2. Let the symmetric matrix P be positive definite. Then, $\tilde{x}^{\mathsf{T}}P\tilde{x}$ satisfies

$$\lambda_{\min}(P)\|\tilde{x}\|^2 \le \tilde{x}^{\mathsf{T}}P\tilde{x} \le \lambda_{\max}(P)\|\tilde{x}\|^2,$$

where $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$ represent the minimum eigenvalue and the maximum eigenvalue of P, respectively.

Theorem 1. *Let Assumptions 1–3 hold. If the dynamic NN weight estimates $\hat{W}_f$, $\hat{W}_g$, and $\hat{W}_k$ are updated as*

$$\dot{\hat{W}}_f = \Gamma_1\phi(\hat{x})\tilde{x}^{\mathsf{T}}P, \quad \dot{\hat{W}}_g = \Gamma_2\rho(\hat{x})u\tilde{x}^{\mathsf{T}}P, \quad \dot{\hat{W}}_k = \Gamma_3\ell(\hat{x})\omega\tilde{x}^{\mathsf{T}}P, \tag{15}$$

where Γ_i $(i = 1, 2, 3)$ are given positive-definite symmetric matrices, then, the identifier developed in (13) can ensure that the identification error $\tilde{x}(t)$ converges to the compact set

$$\Omega_{\tilde{x}} = \left\{\tilde{x} : \|\tilde{x}\| \le \frac{b_\varepsilon}{\sqrt{\beta + 2\eta\lambda_{\min}(P) - \mu}}\right\}, \tag{16}$$

where $\mu > 0$ is a constant to be determined later (see (22) in the proof). In addition, the weight estimation errors $\tilde{W}_f$, $\tilde{W}_g$, and $\tilde{W}_k$ are all guaranteed to be uniformly ultimately bounded (UUB).

Proof. Consider the Lyapunov function candidate

$$J(t) = \underbrace{\frac{1}{2}\tilde{x}^{\mathsf{T}}P\tilde{x}}_{J_1(t)} + \underbrace{\frac{1}{2}\mathrm{tr}\left(\tilde{W}_f^{\mathsf{T}}\Gamma_1^{-1}\tilde{W}_f + \tilde{W}_g^{\mathsf{T}}\Gamma_2^{-1}\tilde{W}_g + \tilde{W}_k^{\mathsf{T}}\Gamma_3^{-1}\tilde{W}_k\right)}_{J_2(t)}. \tag{17}$$

Taking the time derivative of $J_1(t)$ and using (14), we have

$$\begin{aligned}\dot{J}_1(t) = {} & \frac{1}{2}\tilde{x}^{\mathsf{T}}(A^{\mathsf{T}}P + PA)\tilde{x} + \tilde{x}^{\mathsf{T}}PW_f^{\mathsf{T}}\tilde{\phi} + \tilde{x}^{\mathsf{T}}PW_g^{\mathsf{T}}\tilde{\rho}u + \tilde{x}^{\mathsf{T}}PW_k^{\mathsf{T}}\tilde{\ell}\omega \\ & - \eta\tilde{x}^{\mathsf{T}}P\tilde{x} + \tilde{x}^{\mathsf{T}}P\tilde{W}_f^{\mathsf{T}}\phi(\hat{x}) + \tilde{x}^{\mathsf{T}}P\tilde{W}_g^{\mathsf{T}}\rho(\hat{x})u + \tilde{x}^{\mathsf{T}}P\tilde{W}_k^{\mathsf{T}}\ell(\hat{x})\omega + \tilde{x}^{\mathsf{T}}P\varepsilon(x).\end{aligned} \tag{18}$$

Denote $y^{\mathsf{T}} = \tilde{x}^{\mathsf{T}}P$. Applying Cauchy–Schwarz inequality $a^{\mathsf{T}}b \le \frac{1}{2}a^{\mathsf{T}}a + \frac{1}{2}b^{\mathsf{T}}b$ to $\tilde{x}^{\mathsf{T}}PW_f^{\mathsf{T}}\tilde{\phi}$ and $\tilde{x}^{\mathsf{T}}P\varepsilon(x)$, and by using (12) and Assumptions 2 and 3, we obtain

$$\begin{aligned}&\tilde{x}^{\mathsf{T}}PW_f^{\mathsf{T}}\tilde{\phi} = y^{\mathsf{T}}W_f^{\mathsf{T}}\tilde{\phi} \le \frac{1}{2}\tilde{y}^{\mathsf{T}}W_f^{\mathsf{T}}W_f\tilde{y} + \frac{1}{2}\tilde{\phi}^{\mathsf{T}}\tilde{\phi} \le \frac{1}{2}\tilde{y}^{\mathsf{T}}\Lambda_1\tilde{y} + \frac{\lambda_\phi^2}{2}\tilde{x}^{\mathsf{T}}\tilde{x}, \\ &\tilde{x}^{\mathsf{T}}P\varepsilon(x) \le \frac{1}{2}\tilde{x}^{\mathsf{T}}P^2\tilde{x} + \frac{b_\varepsilon^2}{2}.\end{aligned} \tag{19}$$

Similarly, noticing that u and ω are bounded, i.e., $\|u\| \le \left(\sum_{i=1}^m \alpha_i^2\right)^{\frac{1}{2}} \triangleq \alpha$ and $\|\omega(t)\| \le \omega_M$, we have

$$\begin{aligned}&\tilde{x}^{\mathsf{T}}PW_g^{\mathsf{T}}\tilde{\rho}u = y^{\mathsf{T}}W_g^{\mathsf{T}}\tilde{\rho}u \le \frac{1}{2}\tilde{y}^{\mathsf{T}}W_g^{\mathsf{T}}W_g\tilde{y} + \frac{1}{2}u^{\mathsf{T}}\tilde{\rho}^{\mathsf{T}}\tilde{\rho}u \le \frac{1}{2}\tilde{y}^{\mathsf{T}}\Lambda_2\tilde{y} + \frac{(\alpha\lambda_\rho)^2}{2}\tilde{x}^{\mathsf{T}}\tilde{x}, \\ &\tilde{x}^{\mathsf{T}}PW_k^{\mathsf{T}}\tilde{\ell}\omega = y^{\mathsf{T}}W_k^{\mathsf{T}}\tilde{\ell}\omega \le \frac{1}{2}\tilde{y}^{\mathsf{T}}W_k^{\mathsf{T}}W_k\tilde{y} + \frac{1}{2}\omega^{\mathsf{T}}\tilde{\ell}^{\mathsf{T}}\tilde{\ell}\omega \le \frac{1}{2}\tilde{y}^{\mathsf{T}}\Lambda_3\tilde{y} + \frac{(\omega_M\lambda_\ell)^2}{2}\tilde{x}^{\mathsf{T}}\tilde{x}.\end{aligned} \tag{20}$$

Substituting (19) and (20) into (18) and using Facts 1 and 2, we derive

$$\begin{aligned}\dot{J}_1(t) \le {} & \frac{1}{2}\left[\lambda_\phi^2 + \alpha^2\lambda_\rho^2 + \omega_M^2\lambda_\ell^2 - \beta\right]\|\tilde{x}\|^2 + \frac{1}{2}\sum_{i=1}^3 \tilde{x}^{\mathsf{T}}(P\Lambda_iP)\tilde{x} - \eta\tilde{x}^{\mathsf{T}}P\tilde{x} \\ & + \frac{1}{2}\tilde{x}^{\mathsf{T}}P^2\tilde{x} + \tilde{x}^{\mathsf{T}}P\tilde{W}_f^{\mathsf{T}}\phi(\hat{x}) + \tilde{x}^{\mathsf{T}}P\tilde{W}_g^{\mathsf{T}}\rho(\hat{x})u + \tilde{x}^{\mathsf{T}}P\tilde{W}_k^{\mathsf{T}}\ell(\hat{x})\omega + \frac{b_\varepsilon^2}{2} \\ \le {} & -\frac{1}{2}\left(\beta + 2\eta\lambda_{\min}(P) - \mu\right)\|\tilde{x}\|^2 + \tilde{x}^{\mathsf{T}}P\tilde{W}_f^{\mathsf{T}}\phi(\hat{x}) + \tilde{x}^{\mathsf{T}}P\tilde{W}_g^{\mathsf{T}}\rho(\hat{x})u + \tilde{x}^{\mathsf{T}}P\tilde{W}_k^{\mathsf{T}}\ell(\hat{x})\omega + \frac{b_\varepsilon^2}{2},\end{aligned} \tag{21}$$

where

$$\mu = \sum_{i=1}^3 \lambda_{\max}(P\Lambda_iP) + \lambda_{\max}^2(P) + \lambda_\phi^2 + \alpha^2\lambda_\rho^2 + \omega_M^2\lambda_\ell^2. \tag{22}$$

On the other hand, taking the time derivative of $J_2(t)$ and using (15), we have

$$\dot{J}_2(t) = -\mathrm{tr}\left(\tilde{W}_f^{\mathsf{T}}\phi(\hat{x})\tilde{x}^{\mathsf{T}}P + \tilde{W}_g^{\mathsf{T}}\rho(\hat{x})u\tilde{x}^{\mathsf{T}}P + \tilde{W}_k^{\mathsf{T}}\ell(\hat{x})\omega\tilde{x}^{\mathsf{T}}P\right). \tag{23}$$

Observe that $\mathrm{tr}(XY) = \mathrm{tr}(YX) = YX$ for every $X \in \mathbb{R}^{n\times 1}, Y \in \mathbb{R}^{1\times n}$. Therefore, (23) can be represented as

$$\dot{J}_2(t) = -\tilde{x}^{\mathsf{T}}P\tilde{W}_f^{\mathsf{T}}\phi(\hat{x}) - \tilde{x}^{\mathsf{T}}P\tilde{W}_g^{\mathsf{T}}\rho(\hat{x})u - \tilde{x}^{\mathsf{T}}P\tilde{W}_k^{\mathsf{T}}\ell(\hat{x})\omega. \tag{24}$$

Combining (17), (21) and (24), we obtain

$$\dot{J}(t) \le -\frac{1}{2}\left(\beta + 2\eta\lambda_{\min}(P) - \mu\right)\|\tilde{x}\|^2 + \frac{b_\varepsilon^2}{2}. \tag{25}$$

Select proper parameters β and η such that $\beta + 2\eta\lambda_{\min}(P) - \mu > 0$. Then, (25) yields $\dot{J}(t) < 0$ as long as the following inequality holds:

$$\|\tilde{x}\| > \frac{b_\varepsilon}{\sqrt{\beta + 2\eta\lambda_{\min}(P) - \mu}},$$

where μ is given in (22). According to the standard Lyapunov extension theorem [13], this verifies that the identification error $\tilde{x}(t)$ converges to $\Omega_{\tilde{x}}$ defined as in (16), and it also demonstrates the uniform ultimate boundedness of the weight estimation errors $\tilde{W}_f$, $\tilde{W}_g$, and $\tilde{W}_k$. □

Remark 2. Though the identifier NN (13) and the weight update law (15) share similar features with [6,11], a significant difference is that, in our case, we do not use the projection algorithm. In addition, it should be mentioned that parameters η and β can be selected sufficiently large such that $\beta + 2\eta\lambda_{\min}(P) - \mu > 0$. In this sense, $\Omega_{\tilde{x}}$ given in (16) can be kept small enough by properly selecting parameters.

From Remark 2, we know that the identifier NN can approximate the first equation of system (1) within a sufficiently small compact set $\Omega_{\tilde{x}}$. Hence, in what follows we replace the first equation of system (1) with (13). Meanwhile, we replace the actual state $x(t)$ with the estimated state $\hat{x}(t)$. Then, system (1) can be rewritten as

$$\begin{aligned} \dot{\hat{x}} &= \bar{f}(\hat{x}) + \bar{g}(\hat{x})u + \bar{k}(\hat{x})\omega, \\ \hat{z} &= h(\hat{x}) + p(\hat{x})u, \end{aligned} \tag{26}$$

where $\bar{f}(\hat{x}) = A\hat{x} + \eta\tilde{x} + \hat{W}_f^{\mathsf{T}}\phi(\hat{x})$, $\bar{g}(\hat{x}) = \hat{W}_g^{\mathsf{T}}\rho(\hat{x})$, $\bar{k}(\hat{x}) = \hat{W}_k^{\mathsf{T}}\ell(\hat{x})$.

The value function (3) can be expressed as

$$V(\hat{x}(t)) = \int_t^\infty \left(h^{\mathsf{T}}(\hat{x}(s))h(\hat{x}(s)) + 2\kappa \int_0^{u(s)} \tanh^{-1}(\upsilon/\kappa)\mathrm{d}\upsilon - \gamma^2\|\omega(s)\|^2 \right)\mathrm{d}s. \tag{27}$$

Meanwhile, the optimal control (7) and the worst disturbance (8) become

$$u^*(\hat{x}) = -\kappa \tanh\left(\frac{1}{2\kappa}\bar{g}^{\mathsf{T}}(\hat{x})V_{\hat{x}}^*\right), \tag{28}$$

$$\omega^*(\hat{x}) = \frac{1}{2\gamma^2}\bar{k}^{\mathsf{T}}(\hat{x})V_{\hat{x}}^*. \tag{29}$$

Then the HJI equation (10) is developed as

$$V_{\hat{x}}^{*\mathsf{T}}\bar{f}(\hat{x}) + \kappa^2 \sum_{i=1}^{m} \ln\left[1 - \tanh^2(\bar{\mathfrak{A}}_i(\hat{x}))\right] + h^{\mathsf{T}}(\hat{x})h(\hat{x}) + \frac{1}{4\gamma^2}V_x^{*\mathsf{T}}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})V_{\hat{x}}^* = 0, \tag{30}$$

where $\bar{\mathfrak{A}}(\hat{x}) = \frac{1}{2\kappa}\bar{g}^{\mathsf{T}}(\hat{x})V_{\hat{x}}^*$, and $\bar{\mathfrak{A}}(\hat{x}) = \left[\bar{\mathfrak{A}}_1(\hat{x}), \ldots, \bar{\mathfrak{A}}_m(\hat{x})\right]^{\mathsf{T}} \in \mathbb{R}^m$ with $\bar{\mathfrak{A}}_i(\hat{x}) \in \mathbb{R}$, $i = 1, \ldots, m$, and $V_{\hat{x}}^* \in \mathbb{R}^n$ denotes the partial derivative of $V^*(\hat{x})$ with respect to $\hat{x}$.

4. Approximate solution of the HJI equation via a single critic NN

According to the universal approximation property of NNs, the value function $V^*(\hat{x})$ given in (30) can be represented by a single-layer NN on a compact set Ω as

$$V^*(\hat{x}) = W_c^{\mathsf{T}}\sigma(\hat{x}) + \varepsilon_c(\hat{x}),$$

where $W_c \in \mathbb{R}^{N_0}$ is the ideal NN weight vector, $\sigma(\hat{x}) = [\sigma_1(\hat{x}), \sigma_2(\hat{x}), \ldots, \sigma_{N_0}(\hat{x})]^{\mathsf{T}} \in \mathbb{R}^{N_0}$ is the activation function with $\sigma_j(\hat{x}) \in C^1$ and $\sigma_j(0) = 0$, the set $\{\sigma_j(\hat{x})\}_1^{N_0}$ is often selected to be linearly independent, N_0 is the number of neurons, and $\varepsilon_c(\hat{x})$ is the NN function reconstruction error. Meanwhile, the derivative of $V^*(\hat{x})$ with respect to $\hat{x}$ is derived as

$$V_{\hat{x}}^* = \nabla\sigma^{\mathsf{T}}(\hat{x})W_c + \nabla\varepsilon_c, \tag{31}$$

where $\nabla\sigma(\hat{x}) = \partial\sigma(\hat{x})/\partial\hat{x}$, and it satisfies $\nabla\sigma(0) = 0$.

Substituting (31) into (30), we have

$$\begin{aligned} &W_c^{\mathsf{T}}\nabla\sigma\bar{f}(\hat{x}) + \frac{1}{4\gamma^2}W_c^{\mathsf{T}}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}W_c + h^{\mathsf{T}}(\hat{x})h(\hat{x}) \\ &\quad + \mathcal{G}(\nabla\varepsilon_c) + \kappa^2 \sum_{i=1}^{m} \ln\left[1 - \tanh^2\left(\Phi_{1i}(\hat{x}) + \Psi_i(\hat{x})\right)\right] = 0, \end{aligned} \tag{32}$$

where $\Psi(\hat{x})$, $\Phi_1(\hat{x})$ and $\mathcal{G}(\nabla\varepsilon_c)$ are given, respectively, as

$$\begin{aligned}\Psi(\hat{x}) &= \frac{1}{2\kappa}\bar{g}^{\mathsf{T}}(\hat{x})\nabla\varepsilon_c,\\ \Phi_1(\hat{x}) &= \frac{1}{2\kappa}\bar{g}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}W_c,\\ \mathcal{G}(\nabla\varepsilon_c) &= \nabla\varepsilon_c^{\mathsf{T}}\left(\bar{f}(\hat{x}) + \frac{1}{2\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}W_c + \frac{1}{4\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\varepsilon_c\right),\end{aligned} \tag{33}$$

and $\Phi_1(\hat{x}) = \left[\Phi_{11}(\hat{x}), \ldots, \Phi_{1m}(\hat{x})\right]^{\mathsf{T}} \in \mathbb{R}^m$ with $\Phi_{1i}(\hat{x}) \in \mathbb{R}$, and $\Psi(\hat{x}) = \left[\Psi_1(\hat{x}), \ldots, \Psi_m(\hat{x})\right]^{\mathsf{T}} \in \mathbb{R}^m$ with $\Psi_i(\hat{x}) \in \mathbb{R}$, $i = 1, \ldots, m$.

By [23], (32) can be represented as

$$\begin{aligned}&W_c^{\mathsf{T}}\nabla\sigma\bar{f}(\hat{x}) + \frac{1}{4\gamma^2}W_c^{\mathsf{T}}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}W_c\\ &\quad + h^{\mathsf{T}}(\hat{x})h(\hat{x}) + \kappa^2\sum_{i=1}^{m}\ln\left[1-\tanh^2\left(\Phi_{1i}(\hat{x})\right)\right] + \varepsilon_{\mathrm{HJI}} = 0,\end{aligned} \tag{34}$$

where $\varepsilon_{\mathrm{HJI}}$ is the HJI approximation error [2,23].

Remark 3. It was shown in [2,23] that $\varepsilon_{\mathrm{HJI}}$ converges to zero as the number of neurons N_0 goes to infinity. In other words, for given $\varepsilon_h > 0$, there exists a positive N_h (depending only on ε_h) such that $N_0 > N_h$ implies $\|\varepsilon_{\mathrm{HJI}}\| \le \varepsilon_h$. More specifically, one can select a sufficiently large number of neurons N_0 to keep $\varepsilon_{\mathrm{HJI}}$ small.

By using (31) and the mean-value theorem [29], the optimal control (28) and the worst disturbance (29) are, respectively, developed as

$$u^*(\hat{x}) = -\kappa\tanh\left(\Phi_1(\hat{x})\right) + \varepsilon_{u^*}, \tag{35}$$

$$\omega^*(\hat{x}) = \frac{1}{2\gamma^2}\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}W_c + \varepsilon_{\omega^*}, \tag{36}$$

where $\Phi_1(\hat{x})$ is given in (33), $\varepsilon_{u^*} = -\frac{1}{2}\left(\underline{\mathbf{1}} - \tanh^2(a)\right)\bar{g}^{\mathsf{T}}(\hat{x})\nabla\varepsilon_c$ with $\underline{\mathbf{1}} = [1, \ldots, 1]^{\mathsf{T}} \in \mathbb{R}^m$, $a \in \mathbb{R}^m$ chosen between $\Phi_1(\hat{x})$ and $\bar{\mathfrak{A}}(\hat{x})$, and $\varepsilon_{\omega^*} = \frac{1}{2\gamma^2}\bar{k}^{\mathsf{T}}(\hat{x})\nabla\varepsilon_c$.

Since the ideal critic NN weight W_c is typically unknown, (35) cannot be implemented. Hence, we use a critic NN to approximate the value function $V^*(\hat{x})$ as

$$\hat{V}(\hat{x}) = \hat{W}_c^{\mathsf{T}}\sigma(\hat{x}), \tag{37}$$

where $\hat{W}_c$ is the estimate of W_c. The estimation error for the critic NN weights is defined as $\tilde{W}_c = W_c - \hat{W}_c$.

By using (37), the estimated values of the optimal control (28) and the worst disturbance (29) are

$$\hat{u}(\hat{x}) = -\kappa\tanh\left(\frac{1}{2\kappa}\bar{g}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\hat{W}_c\right), \tag{38}$$

$$\hat{\omega}(\hat{x}) = \frac{1}{2\gamma^2}\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\hat{W}_c. \tag{39}$$

Combining (4), (26), (37)–(39), we obtain the approximate Hamiltonian as

$$\begin{aligned}H(\hat{x}, \hat{W}_c) &= \hat{W}_c^{\mathsf{T}}\nabla\sigma\bar{f}(\hat{x}) + \frac{1}{4\gamma^2}\hat{W}_c^{\mathsf{T}}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\hat{W}_c\\ &\quad + h^{\mathsf{T}}(\hat{x})h(\hat{x}) + \kappa^2\sum_{i=1}^{m}\ln\left[1-\tanh^2\left(\Phi_{2i}(\hat{x})\right)\right] \triangleq e,\end{aligned} \tag{40}$$

where $\Phi_2(\hat{x}) = \frac{1}{2\kappa}\bar{g}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\hat{W}_c$, and $\Phi_2(\hat{x}) = \left[\Phi_{21}(\hat{x}), \ldots, \Phi_{2m}(\hat{x})\right]^{\mathsf{T}} \in \mathbb{R}^m$ with $\Phi_{2i}(\hat{x}) \in \mathbb{R}$, $i = 1, \ldots, m$.

By using (34), (40) can be derived as

$$e = \kappa^2\sum_{i=1}^{m}\left[\mathscr{B}(\Phi_{2i}) - \mathscr{B}(\Phi_{1i})\right] - \tilde{W}_c^{\mathsf{T}}\nabla\sigma\left(\bar{f}(\hat{x}) + \bar{k}(\hat{x})\hat{\omega}\right) - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\sigma^{\mathsf{T}}\tilde{W}_c - \varepsilon_{\mathrm{HJI}}, \tag{41}$$

where $\mathscr{B}(\Phi_{\iota i}) = \ln\left[1-\tanh^2\left(\Phi_{\iota i}(\hat{x})\right)\right]$, $\iota = 1, 2$, and $i = 1, \ldots, m$.

For every $\Phi_{\iota i}(\hat{x}) \in \mathbb{R}$, $\mathscr{B}(\Phi_{\iota i})$ can be rewritten as [40]

$$\mathscr{B}(\Phi_{\iota i}) = \ln 4 - 2\Phi_{\iota i}(\hat{x})\mathrm{sgn}(\Phi_{\iota i}(\hat{x})) - 2\ln\left[1+\exp\left(-2\Phi_{\iota i}(\hat{x})\mathrm{sgn}(\Phi_{\iota i}(\hat{x}))\right)\right],$$

where $\mathrm{sgn}(\Phi_{\iota i}(\hat{x})) \in \mathbb{R}^m$ is a sign function [29]. Then, we have

$$\sum_{i=1}^{m} \mathscr{B}(\Phi_{\iota i}) = m\ln 4 - 2\Phi_{\iota}^{\mathsf{T}}(\hat{x})\mathrm{sgn}(\Phi_{\iota}(\hat{x})) - 2\sum_{i=1}^{m} \ln\left[1+\exp\left(-2\Phi_{\iota i}(\hat{x})\mathrm{sgn}(\Phi_{\iota i}(\hat{x}))\right)\right]. \tag{42}$$

Combining (41) with (42), we obtain

$$\begin{aligned} e &= 2\kappa^2\left[\Phi_1^{\mathsf{T}}(\hat{x})\mathrm{sgn}(\Phi_1(\hat{x})) - \Phi_2^{\mathsf{T}}(\hat{x})\mathrm{sgn}(\Phi_2(\hat{x}))\right] + \kappa^2\Delta_\Phi \\ &\quad - \tilde{W}_c^{\mathsf{T}}\nabla\sigma\left(\bar{f}(\hat{x}) + \bar{k}(\hat{x})\hat{\omega}\right) - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\sigma^{\mathsf{T}}\tilde{W}_c - \varepsilon_{\mathrm{HJI}} \\ &= \kappa\left[W_c^{\mathsf{T}}\nabla\sigma\bar{g}(\hat{x})\mathrm{sgn}(\Phi_1(\hat{x})) - \hat{W}_c^{\mathsf{T}}\nabla\sigma\bar{g}(\hat{x})\mathrm{sgn}(\Phi_2(\hat{x}))\right] + \kappa^2\Delta_\Phi \\ &\quad - \tilde{W}_c^{\mathsf{T}}\nabla\sigma\left(\bar{f}(\hat{x}) + \bar{k}(\hat{x})\hat{\omega}\right) - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\sigma^{\mathsf{T}}\tilde{W}_c - \varepsilon_{\mathrm{HJI}} \\ &= -\tilde{W}_c^{\mathsf{T}}\left[\nabla\sigma\left(\bar{f}(\hat{x}) + \bar{k}(\hat{x})\hat{\omega}\right) - \kappa\nabla\sigma\bar{g}(\hat{x})\mathrm{sgn}(\Phi_2(\hat{x}))\right] - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\sigma^{\mathsf{T}}\tilde{W}_c + \delta(\hat{x}), \end{aligned} \tag{43}$$

where

$$\Delta_\Phi = 2\sum_{i=1}^{m}\ln\frac{1+\exp\left[-2\Phi_{1i}(\hat{x})\mathrm{sgn}(\Phi_{1i}(\hat{x}))\right]}{1+\exp\left[-2\Phi_{2i}(\hat{x})\mathrm{sgn}(\Phi_{2i}(\hat{x}))\right]},$$

$$\delta(x) = \kappa W_c^{\mathsf{T}}\nabla\sigma\bar{g}(\hat{x})\left[\mathrm{sgn}(\Phi_1(\hat{x})) - \mathrm{sgn}(\Phi_2(\hat{x}))\right] + \kappa^2\Delta_\Phi - \varepsilon_{\mathrm{HJI}}.$$

To get the minimum value of e, it is desired to choose $\hat{W}_c$ to minimize the squared residual error $E = \frac{1}{2}e^{\mathsf{T}}e$. The traditional way for deriving such a $\hat{W}_c$ is to employ the gradient descent method. By using the approach, the weight tuning law for the critic NN is often given as

$$\dot{\hat{W}}_{c_{\mathrm{tra}}} = -\frac{l}{(1+\psi^{\mathsf{T}}\psi)^2}\frac{\partial E}{\partial \hat{W}_c} = -\frac{l\psi}{(1+\psi^{\mathsf{T}}\psi)^2}e, \tag{44}$$

where $\hat{W}_{c_{\mathrm{tra}}}$ denotes the critic NN weight, i.e., $\hat{W}_{c_{\mathrm{tra}}} = \hat{W}_c$, $l > 0$ is a design constant, $\psi = \nabla\sigma\left(\bar{f}(\hat{x}) + \bar{g}(\hat{x})\hat{u} + \bar{k}(\hat{x})\hat{\omega}\right)$, and $(1+\psi^{\mathsf{T}}\psi)^2$ is employed for normalization.

Two points about the tuning rule (44) should be mentioned. That is,

(i) By using (44), an initial stabilizing control for system (26) is often required. However, such a control law is generally hard to obtain when system (26) contains high-order terms. More importantly, if the initial control for system (26) is unstable, then the tuning law (44) might not guarantee stability of the closed-loop system during the learning process of the critic NN [7].

(ii) To ensure the weights of the critic NN converge to the actual optimal values, an exploration signal is often added to the input to keep the persistence of excitation (PE) of $\psi/(1+\psi^{\mathsf{T}}\psi)$ while utilizing (44). Nevertheless, there is no general approach proposed to give such an exploration signal. The provided exploration signal might give rise to instability of the closed-loop system. Therefore, when the exploration signal is added, it is necessary to check the stability of the closed-loop system.

In light of (*i*) and (*ii*), the weight update law for the critic NN shall be redefined. Before proceeding, we provide another assumption, which has been used in [7,25,39,41,43].

Assumption 4. $L_1(\hat{x})$ is a continuously differentiable radially unbounded Lyapunov function candidate such that $\dot{L}_1(\hat{x}) = L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*\right) < 0$ with $L_{1\hat{x}}$ the partial derivative of $L_1(\hat{x})$ with respect to $\hat{x}$. Moreover, there exists a symmetric positive-definite matrix $Q(\hat{x}) \in \mathbb{R}^{n\times n}$ defined on Ω such that

$$L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*\right) = -L_{1\hat{x}}^{\mathsf{T}}Q(\hat{x})L_{1\hat{x}}. \tag{45}$$

Remark 4. $\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*$ is often assumed to be bounded by a positive constant on a compact set Ω [14,17,23,32]. That is, for every $\hat{x} \in \Omega$, there exists a constant $b_1 > 0$ such that $\|\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*\| \le b_1$. To relax the condition, in this paper, we assume that $\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*$ is bounded by a function with respect to $\hat{x}$. Because $L_{1\hat{x}}$ is the function with respect to $\hat{x}$, without loss of generality, we assume that $\|\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*\| \le b_2\|L_{1\hat{x}}\|$ $(b_2 > 0)$. In this sense, we have $\|L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*\right)\| \le b_2\|L_{1\hat{x}}\|^2$. Observing that $L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) + \bar{g}(\hat{x})u^* + \bar{k}(\hat{x})\omega^*\right) < 0$, one shall find that (45) defined as in Assumption 4 is reasonable. In addition, it should be mentioned that $L_1(\hat{x})$ is usually derived through properly selecting functions, such as polynomials.

Based on Assumption 4 and aforementioned analyses, in this paper, we develop a novel weight update law for the critic NN as

$$\begin{aligned}\dot{\hat{W}}_c = &-l\bar{\psi}\left(\hat{W}_c^{\mathsf{T}}\nabla\sigma\bar{f}(\hat{x}) + \frac{1}{4\gamma^2}\hat{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\hat{W}_c + h^{\mathsf{T}}(\hat{x})h(\hat{x}) + \kappa^2\sum_{i=1}^{m}\ln\left[1-\tanh^2\left(\Phi_{2i}(\hat{x})\right)\right]\right)\\ &+\frac{l}{2}\Pi(\hat{x},\hat{u},\hat{\omega})\nabla\sigma\left(\bar{g}(\hat{x})\left[I_m - \mathscr{C}(\Phi_2(\hat{x}))\right]\bar{g}^{\mathsf{T}}(\hat{x}) - \frac{1}{\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\right)L_{1\hat{x}}\\ &+l\left(\kappa\nabla\sigma\bar{g}(\hat{x})\left[\tanh\left(\Phi_2(\hat{x})\right) - \mathrm{sgn}(\Phi_2(\hat{x}))\right]\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c - \frac{1}{4\gamma^2}\mathfrak{B}(\hat{x})\hat{W}_c\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c - \left(K_2 - K_1\varphi^{\mathsf{T}}\right)\hat{W}_c\right),\end{aligned} \tag{46}$$

where $\bar{\psi} = \psi/m_s^2$, $\varphi = \psi/m_s$, $m_s = 1+\psi^{\mathsf{T}}\psi$, $\mathscr{C}(\Phi_2(\hat{x})) = \mathrm{diag}\{\tanh^2(\Phi_{2i}(\hat{x}))\}$, $i = 1,\ldots,m$, $\mathfrak{B}(\hat{x}) = \nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}$, $L_{1\hat{x}}$ is defined as in Assumption 4, K_1 and K_2 are given parameter matrices with suitable dimensions, and $\Pi(\hat{x},\hat{u},\hat{\omega})$ is an indicator function given by

$$\Pi(\hat{x},\hat{u},\hat{\omega}) = \begin{cases}0, & \text{if } L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) + \bar{g}(\hat{x})\hat{u} + \bar{k}(\hat{x})\hat{\omega}\right) < 0,\\ 1, & \text{otherwise.}\end{cases} \tag{47}$$

Remark 5. Compared with (44), a distinct feature of (46) is that it contains two additional terms. The second term given in (46) is utilized to guarantee the stability of the closed-loop system during the NN learning process. To explain it clearly, we denote the derivative of the Lyapunov function candidate for system (26) with the control (38) and the disturbance (39) as

$$\Theta = L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) - \kappa\bar{g}(\hat{x})\tanh\left(\Phi_2(\hat{x})\right) + \frac{1}{2\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\hat{W}_c\right).$$

If the closed-loop system is unstable, then we obtain $\Theta > 0$. To keep the closed-loop system stable, we need make $\Theta < 0$. Using the gradient descent method, we have

$$\begin{aligned}-l\frac{\partial\Theta}{\partial\hat{W}_c} &= -l\frac{\partial\left[L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) - \kappa\bar{g}(\hat{x})\tanh\left(\Phi_2(\hat{x})\right)\right)\right]}{\partial\hat{W}_c} - \frac{l}{2\gamma^2}\frac{\partial\left[L_{1\hat{x}}^{\mathsf{T}}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\hat{W}_c\right]}{\partial\hat{W}_c}\\ &= l\left(\frac{\partial\Phi_2(\hat{x})}{\partial\hat{W}_c}\right)^{\mathsf{T}}\cdot\frac{\partial\left[\kappa L_{1\hat{x}}^{\mathsf{T}}\bar{g}(\hat{x})\tanh\left(\Phi_2(\hat{x})\right)\right]}{\partial\Phi_2(\hat{x})} - \frac{l}{2\gamma^2}\nabla\sigma\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})L_{1\hat{x}}\\ &= \frac{l}{2}\nabla\sigma\left(\bar{g}(\hat{x})\left[I_m - \mathscr{C}(\Phi_2(\hat{x}))\right]\bar{g}^{\mathsf{T}}(\hat{x}) - \frac{1}{\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\right)L_{1\hat{x}},\end{aligned} \tag{48}$$

where $\mathscr{C}(\Phi_2(\hat{x})) = \mathrm{diag}\{\tanh^2(\Phi_{2i}(\hat{x}))\}$, $i = 1,\ldots,m$. Eq.(48) indicates the reason that we employ the second term in (46). Actually, by $\Pi(\hat{x},\hat{u},\hat{\omega})$ given in (47), we find that if there exists $\Theta < 0$ (that is, the closed-loop system is stable), then $\Pi(\hat{x},\hat{u},\hat{\omega}) = 0$ and the second term given in (46) disappears. If the closed-loop system is unstable, then $\Pi(\hat{x},\hat{u},\hat{\omega}) = 1$ and the second term given in (46) (i.e., (48)) works. By (46), it makes no requirement of the initial stabilizing control for system (26). The property will be verified in the subsequent numerical simulation. The third term given in (46) is a robustifying term, which is used for stability analysis in the subsequent discussion.

Remark 6. Observing the expression of (46), if selecting proper parameter matrices K_i $(i = 1, 2)$ such that $K_2 = K_1\psi^{\mathsf{T}}$, one shall find that $\dot{\hat{W}}_c = 0$ when $\hat{x} = 0$. In this sense, $\hat{V}(\hat{x})$ will no longer be updated. Nevertheless, the optimal control might not be derived at the finite time t_f which makes $\hat{x}(t_f) = 0$. To avoid this case, the exploration signal is often added to the control input, that is, the PE condition is required. Interestingly, the second term given in (46) can be used to check the stability of the closed-loop system when the exploration signal is added.

By ψ given in (44) and using (38), we get $\nabla\sigma\left(\bar{f}(\hat{x}) + \bar{k}(\hat{x})\hat{\omega}\right) = \psi + \kappa\nabla\sigma\bar{g}(\hat{x})\tanh\left(\Phi_2(\hat{x})\right)$. Then, noticing that $\tilde{W}_c = W_c - \hat{W}_c$ and utilizing (40), (43), and (46), we have

$$\begin{aligned}\dot{\tilde{W}}_c = &\,l\frac{\varphi}{m_s}\left(-\tilde{W}_c^{\mathsf{T}}\psi + \kappa\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{g}(\hat{x})\mathcal{F}(\hat{x}) - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\tilde{W}_c + \delta(\hat{x})\right)\\ &-\frac{l}{2}\Pi(\hat{x},\hat{u},\hat{\omega})\nabla\sigma\left(\bar{g}(\hat{x})\left[I_m - \mathscr{C}(\Phi_2(\hat{x}))\right]\bar{g}^{\mathsf{T}}(\hat{x}) - \frac{1}{\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\right)L_{1\hat{x}}\\ &+l\left(\kappa\nabla\sigma\bar{g}(\hat{x})\mathcal{F}(\hat{x})\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c + \frac{1}{4\gamma^2}\mathfrak{B}(\hat{x})\hat{W}_c\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c + \left(K_2 - K_1\varphi^{\mathsf{T}}\right)\hat{W}_c\right),\end{aligned} \tag{49}$$

where $\mathcal{F}(\hat{x}) = \mathrm{sgn}(\Phi_2(\hat{x})) - \tanh\left(\Phi_2(\hat{x})\right)$.

5. Stability analysis

Before demonstrating the main theorems, we present several required assumptions. These assumptions have been used in [18,21,22,24,32,34].

Assumption 5. The ideal NN weight W_c is bounded by a known positive constant W_{cM}, i.e., $\|W_c\| \le W_{cM}$. There exist known constants $b_{\varepsilon_c} > 0$ and $b_{\varepsilon_c\hat{x}} > 0$ such that $\|\varepsilon_c(\hat{x})\| < b_{\varepsilon_c}$, $\|\nabla\varepsilon_c(\hat{x})\| < b_{\varepsilon_c\hat{x}}$ for every $\hat{x} \in \Omega$. In addition, there exist known constants $b_{\varepsilon_{u^*}} > 0$ and $b_{\varepsilon_{\omega^*}} > 0$ such that $\|\varepsilon_{u^*}\| \le b_{\varepsilon_{u^*}}$, $\|\varepsilon_{\omega^*}\| \le b_{\varepsilon_{\omega^*}}$ for every $\hat{x} \in \Omega$.

Assumption 6. There exist known constants $b_\sigma > 0$ and $b_{\sigma\hat{x}} > 0$ such that $\|\sigma(\hat{x})\| \le b_\sigma$ and $\|\nabla\sigma(\hat{x})\| \le b_{\sigma\hat{x}}$ for every $\hat{x} \in \Omega$.

From Theorem 1, we know that $\hat{W}_g$ and $\hat{W}_k$ are bounded. Noticing that $\rho(\hat{x})$ and $\ell(\hat{x})$ are bounded functions over Ω, therefore, we can obtain that $\bar{g}(\hat{x})$ and $\bar{k}(\hat{x})$ are bounded over Ω. Accordingly, we give another assumption as follows.

Assumption 7. There exist known constants $\bar{g}_M > 0$ and $\bar{k}_M > 0$ such that $\|\bar{g}(\hat{x})\| \le \bar{g}_M$ and $\|\bar{k}(\hat{x})\| \le \bar{k}_M$ for every $\hat{x} \in \Omega$.

Let $G(\Phi_\iota) = \tanh(\Phi_\iota(\hat{x}))$, $\iota = 1, 2$. By employing Taylor series expansion, we have

$$\begin{aligned} G(\Phi_1) &= G(\Phi_2) + \frac{\partial G(\Phi_2)}{\partial \Phi_2}\big(\Phi_1(\hat{x}) - \Phi_2(\hat{x})\big) + O\big((\Phi_1(\hat{x}) - \Phi_2(\hat{x}))^2\big) \\ &= G(\Phi_2) + \frac{1}{2\kappa}\big[I_m - \mathscr{C}(\Phi_2(\hat{x}))\big]g^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c + O\big((\Phi_1(\hat{x}) - \Phi_2(\hat{x}))^2\big), \end{aligned} \tag{50}$$

where $\mathscr{C}(\Phi_2(\hat{x})) = \mathrm{diag}\{\tanh^2(\Phi_{2i}(\hat{x}))\}$, $i = 1, \ldots, m$, and $O\big((\Phi_1(\hat{x}) - \Phi_2(\hat{x}))^2\big)$ is the high-order terms of the Taylor series [29]. Then, we derive

$$O\big((\Phi_1(\hat{x}) - \Phi_2(\hat{x}))^2\big) = G(\Phi_1) - G(\Phi_2) + \frac{1}{2\kappa}\big[\mathscr{C}(\Phi_2(\hat{x})) - I_m\big]g^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c. \tag{51}$$

Lemma 1. *For hyperbolic function* tanh , *the high-order term in the Taylor series is bounded as*

$$\big\|O\big((\Phi_1(\hat{x}) - \Phi_2(\hat{x}))^2\big)\big\| \le 2\sqrt{m} + (1/\kappa)\bar{g}_M b_{\sigma\hat{x}}\big\|\tilde{W}_c\big\|. \tag{52}$$

Proof. From (51), we have

$$\begin{aligned} \big\|O\big((\Phi_1(\hat{x}) - \Phi_2(\hat{x}))^2\big)\big\| &\le \|G(\Phi_1) - G(\Phi_2)\| + \frac{1}{2\kappa}\big\|\big[\mathscr{C}(\Phi_2(\hat{x})) - I_m\big]g^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c\big\| \\ &\le \|G(\Phi_1)\| + \|G(\Phi_2)\| + \frac{1}{2\kappa}\big\|\mathscr{C}(\Phi_2(\hat{x})) - I_m\big\|\big\|g(\hat{x})\big\|\|\nabla\sigma\|\big\|\tilde{W}_c\big\|. \end{aligned} \tag{53}$$

Notice that $\|G(\Phi_\iota)\| = \big(\sum_{i=1}^m |\tanh(\Phi_{\iota i})|^2\big)^{1/2} \le \sqrt{m}$, $\iota = 1, 2$, $\|\mathscr{C}(\Phi_2(\hat{x})) - I_m\| \le 2$. Then, by Assumptions 6 and 7 and from (53), we can obtain (52). □

Theorem 2. *Given the input-affine dynamics described by (26) with associated HJI equation (30). Let Assumptions 4–7 hold and take the control input and disturbance input for system (26) as given in (38) and (39), respectively. Meanwhile, let weight update law for the identifier NN be (15), and let the weight tuning rule for the critic NN be (46). Then, the function $L_{1\hat{x}}$ and the critic NN weight estimation error $\tilde{W}_c$ are guaranteed to be UUB.*

Proof. Consider the Lyapunov function candidate

$$L(t) = L_1(\hat{x}(t)) + \frac{1}{2}\tilde{W}_c^{\mathsf{T}} l^{-1}\tilde{W}_c, \tag{54}$$

where $L_1(\hat{x}(t))$ is given in Assumption 4. Taking the time derivative of (54), we have

$$\dot{L}(t) = L_{1\hat{x}}^{\mathsf{T}}\big(\bar{f}(\hat{x}) + \bar{g}(\hat{x})\hat{u} + \bar{k}(\hat{x})\hat{\omega}\big) + \dot{\tilde{W}}_c^{\mathsf{T}} l^{-1}\tilde{W}_c. \tag{55}$$

Using (49), the second term of (55) can be represented as

$$\dot{\tilde{W}}_c^{\mathsf{T}} l^{-1}\tilde{W}_c = \sum_{i=1}^{3}\mathfrak{N}_i - \frac{1}{2}\Pi(\hat{x}, \hat{u}, \hat{\omega})L_{1\hat{x}}^{\mathsf{T}}\left(\bar{g}(\hat{x})\big[I_m - \mathscr{C}(\Phi_2(\hat{x}))\big]\bar{g}^{\mathsf{T}}(\hat{x}) - \frac{1}{\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\right)\nabla\sigma^{\mathsf{T}}\tilde{W}_c, \tag{56}$$

where

$$\begin{aligned} \mathfrak{N}_1 &= \left(-\tilde{W}_c^{\mathsf{T}}\psi + \kappa\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{g}(\hat{x})\mathcal{F}(\hat{x}) - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\tilde{W}_c + \delta(\hat{x})\right)\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c \\ &\quad + \kappa\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{g}(\hat{x})\mathcal{F}(\hat{x})\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c \\ &= -\tilde{W}_c^{\mathsf{T}}\varphi\varphi^{\mathsf{T}}\tilde{W}_c - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\tilde{W}_c\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c + \kappa\tilde{W}_c^{\mathsf{T}}\nabla\sigma\bar{g}(\hat{x})\mathcal{F}(\hat{x})\frac{\varphi^{\mathsf{T}}}{m_s}W_c + \delta(\hat{x})\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c, \end{aligned}$$

$$\begin{aligned}\mathfrak{N}_2 &= \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\hat{W}_c\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c\\ &= \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\tilde{W}_c\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c + \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})W_c\frac{\varphi^{\mathsf{T}}}{m_s}W_c\\ &\quad - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})W_c\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\tilde{W}_c\frac{\varphi^{\mathsf{T}}}{m_s}W_c,\end{aligned}$$

$$\begin{aligned}\mathfrak{N}_3 &= \tilde{W}_c^{\mathsf{T}}\left(K_2 - K_1\varphi^{\mathsf{T}}\right)(W_c - \tilde{W}_c)\\ &= \tilde{W}_c^{\mathsf{T}}K_2W_c - \tilde{W}_c^{\mathsf{T}}K_2\tilde{W}_c - \tilde{W}_c^{\mathsf{T}}K_1\varphi^{\mathsf{T}}W_c + \tilde{W}_c^{\mathsf{T}}K_1\varphi^{\mathsf{T}}\tilde{W}_c.\end{aligned}$$

From (56), we obtain

$$\begin{aligned}\dot{\tilde{W}}_c^{\mathsf{T}}l^{-1}\tilde{W}_c =& -\tilde{W}_c^{\mathsf{T}}\varphi\varphi^{\mathsf{T}}\tilde{W}_c - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})\tilde{W}_c\frac{\varphi^{\mathsf{T}}}{m_s}W_c - \tilde{W}_c^{\mathsf{T}}K_2\tilde{W}_c - \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})W_c\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c\\ &+ \tilde{W}_c^{\mathsf{T}}K_1\varphi^{\mathsf{T}}\tilde{W}_c + \frac{1}{4\gamma^2}\tilde{W}_c^{\mathsf{T}}\mathfrak{B}(\hat{x})W_c\frac{\varphi^{\mathsf{T}}}{m_s}W_c + \tilde{W}_c^{\mathsf{T}}\bar{\mathfrak{D}}(\hat{x}) + \delta(\hat{x})\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c\\ &- \frac{1}{2}\Pi(\hat{x},\hat{u},\hat{\omega})L_{1\hat{x}}^{\mathsf{T}}\left(\bar{g}(\hat{x})\left[I_m - \mathscr{C}(\Phi_2(\hat{x}))\right]\bar{g}^{\mathsf{T}}(\hat{x}) - \frac{1}{\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\right)\nabla\sigma^{\mathsf{T}}\tilde{W}_c,\end{aligned} \tag{57}$$

where $\bar{\mathfrak{D}}(\hat{x}) = \left(\kappa\nabla\sigma\bar{g}(\hat{x})\mathcal{F}(\hat{x})\frac{\varphi^{\mathsf{T}}}{m_s} + K_2 - K_1\varphi^{\mathsf{T}}\right)W_c$.

Let $\mathcal{Y}^{\mathsf{T}} = \left[\tilde{W}_c^{\mathsf{T}}\varphi, \tilde{W}_c^{\mathsf{T}}\right]$. Then, (57) can be rewritten as

$$\begin{aligned}\dot{\tilde{W}}_c^{\mathsf{T}}l^{-1}\tilde{W}_c =& -\mathcal{Y}^{\mathsf{T}}M\mathcal{Y} + \mathcal{Y}^{\mathsf{T}}N\\ &- \frac{1}{2}\Pi(\hat{x},\hat{u},\hat{\omega})L_{1\hat{x}}^{\mathsf{T}}\left(\bar{g}(\hat{x})\left[I_m - \mathscr{C}(\Phi_2(\hat{x}))\right]\bar{g}^{\mathsf{T}}(\hat{x}) - \frac{1}{\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\right)\nabla\sigma^{\mathsf{T}}\tilde{W}_c,\end{aligned} \tag{58}$$

where

$$M = \begin{bmatrix} I & \left(\frac{1}{8m_s\gamma^2}\mathfrak{B}(\hat{x})W_c - \frac{1}{2}K_1\right)^{\mathsf{T}}\\ \frac{1}{8m_s\gamma^2}\mathfrak{B}(\hat{x})W_c - \frac{1}{2}K_1 & K_2 + \frac{1}{4m_s\gamma^2}\varphi^{\mathsf{T}}W_c\mathfrak{B}(\hat{x})\end{bmatrix}, \qquad N = \begin{bmatrix} \frac{1}{m_s}\delta(\hat{x})\\ \frac{1}{4m_s\gamma^2}\mathfrak{B}(\hat{x})W_c\varphi^{\mathsf{T}}W_c + \bar{\mathfrak{D}}(\hat{x})\end{bmatrix}.$$

Substituting (58) into (55) and choosing K_i $(i = 1, 2)$ such that the matrix M is positive definite, we have

$$\begin{aligned}\dot{L}(t) \le& L_{1\hat{x}}^{\mathsf{T}}\left(\bar{f}(\hat{x}) + \bar{g}(\hat{x})\hat{u} + \bar{k}(\hat{x})\hat{\omega}\right) - \lambda_{\min}(M)\|\mathcal{Y}\|^2 + \zeta_N\|\mathcal{Y}\|\\ &- \frac{1}{2}\Pi(\hat{x},\hat{u},\hat{\omega})L_{1\hat{x}}^{\mathsf{T}}\left(\bar{g}(\hat{x})\left[I_m - \mathscr{C}(\Phi_2(\hat{x}))\right]\bar{g}^{\mathsf{T}}(\hat{x}) - \frac{1}{\gamma^2}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\right)\nabla\sigma^{\mathsf{T}}\tilde{W}_c,\end{aligned} \tag{59}$$

where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of M, and ζ_N is the upper bound of $\|N\|$, i.e., $\|N\| \le \zeta_N$.

Based on the definition of $\Pi(\hat{x},\hat{u},\hat{\omega})$ given in (47), we divide (59) into the following two cases for discussion:

Case 1: $\Pi(\hat{x},\hat{u},\hat{\omega}) = 0$. In this case, we derive that the first term in (59) is negative via (47). Since $\|\hat{x}\| > 0$ is guaranteed by adding the PE signal, we can obtain that there exists a constant τ such that $0 < \tau < \|\hat{x}\|$ implies $L_{1\hat{x}}^{\mathsf{T}}\dot{\hat{x}} < -\|L_{1\hat{x}}\|\tau < 0$ by using *dense property* of $\mathbb{R}$ [29]. Then, noticing that $\dot{\hat{x}} = \bar{f}(\hat{x}) + \bar{g}(\hat{x})\hat{u} + \bar{k}(\hat{x})\hat{\omega}$, (59) is developed as

$$\begin{aligned}\dot{L}(t) &\le L_{1\hat{x}}^{\mathsf{T}}\dot{\hat{x}} - \lambda_{\min}(M)\|\mathcal{Y}\|^2 + \zeta_N\|\mathcal{Y}\|\\ &< -\|L_{1\hat{x}}\|\tau - \lambda_{\min}(M)\left(\|\mathcal{Y}\| - \frac{\zeta_N}{2\lambda_{\min}(M)}\right)^2 + \frac{\zeta_N^2}{4\lambda_{\min}(M)}.\end{aligned} \tag{60}$$

Thus, (60) yields $\dot{L}(t) < 0$ as long as one of the following conditions holds:

$$\|L_{1\hat{x}}\| > \frac{\zeta_N^2}{4\tau\lambda_{\min}(M)} \triangleq \mathcal{B}_1, \quad \text{or} \quad \|\mathcal{Y}\| > \frac{\zeta_N}{\lambda_{\min}(M)}. \tag{61}$$

Noticing that $\|\mathcal{Y}\| \le \sqrt{1 + \|\varphi\|^2}\|\tilde{W}_c\|$ and $\|\varphi\| \le 1/2$, we derive $\|\mathcal{Y}\| \le (\sqrt{5}/2)\|\tilde{W}_c\|$. Then, from (61), we have

$$\|\tilde{W}_c\| > \frac{2\zeta_N}{\sqrt{5}\lambda_{\min}(M)} \triangleq \mathcal{B}_2.$$

Case 2: $\Pi(\hat{x},\hat{u},\hat{\omega})=1$. In this circumstance, the first term in (59) is nonnegative. It implies that the control given in (38) might not stabilize system (1). Then, (59) becomes

$$\dot{L}(t)\le-\lambda_{\min}(M)\|\mathcal{Y}\|^2+\zeta_N\|\mathcal{Y}\|+\mathfrak{R}_1+\mathfrak{R}_2, \tag{62}$$

where

$$\mathfrak{R}_1=L_{1\hat{x}}^{\mathsf{T}}\big(\bar{f}(\hat{x})+\bar{g}(\hat{x})\hat{u}\big)-\frac{1}{2}L_{1\hat{x}}^{\mathsf{T}}\bar{g}(\hat{x})\big[I_m-\mathscr{C}(\Phi_2(\hat{x}))\big]\bar{g}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c, \tag{63}$$

$$\mathfrak{R}_2=L_{1\hat{x}}^{\mathsf{T}}\bar{k}(\hat{x})\hat{\omega}+\frac{1}{2\gamma^2}L_{1\hat{x}}^{\mathsf{T}}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c. \tag{64}$$

From (50), we have

$$\tanh(\Phi_2(\hat{x}))+\frac{1}{2\kappa}\big[I_m-\mathscr{C}(\Phi_2(\hat{x}))\big]g^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c=\tanh(\Phi_1(\hat{x}))-O\big((\Phi_1(\hat{x})-\Phi_2(\hat{x}))^2\big).$$

Then, by utilizing (35) and (38), $\mathfrak{R}_1$ given in (63) is developed as

$$\begin{aligned}\mathfrak{R}_1&=L_{1\hat{x}}^{\mathsf{T}}\bar{f}(\hat{x})-\kappa L_{1\hat{x}}^{\mathsf{T}}\bar{g}(\hat{x})\left(\tanh(\Phi_2(\hat{x}))+\frac{1}{2\kappa}\big[I_m-\mathscr{C}(\Phi_2(\hat{x}))\big]g^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c\right)\\&=L_{1\hat{x}}^{\mathsf{T}}\big(\bar{f}(\hat{x})-\kappa\bar{g}(\hat{x})\tanh(\Phi_1(\hat{x}))\big)+\kappa L_{1\hat{x}}^{\mathsf{T}}\bar{g}(\hat{x})O\big((\Phi_1(\hat{x})-\Phi_2(\hat{x}))^2\big)\\&=L_{1\hat{x}}^{\mathsf{T}}\big(\bar{f}(\hat{x})+\bar{g}(\hat{x})u^*\big)-L_{1\hat{x}}^{\mathsf{T}}\bar{g}(\hat{x})\varepsilon_{u^*}+\kappa L_{1\hat{x}}^{\mathsf{T}}\bar{g}(\hat{x})O\big((\Phi_1(\hat{x})-\Phi_2(\hat{x}))^2\big).\end{aligned} \tag{65}$$

Similarly, by using (36) and (39), $\mathfrak{R}_2$ given in (64) becomes

$$\mathfrak{R}_2=\frac{1}{2\gamma^2}L_{1\hat{x}}^{\mathsf{T}}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\hat{W}_c+\frac{1}{2\gamma^2}L_{1\hat{x}}^{\mathsf{T}}\bar{k}(\hat{x})\bar{k}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c=L_{1\hat{x}}^{\mathsf{T}}\bar{k}(\hat{x})\omega^*-L_{1\hat{x}}^{\mathsf{T}}\bar{k}(\hat{x})\varepsilon_{\omega^*}. \tag{66}$$

Combining (65) with (66), and by Assumption 4 and Lemma 1, we obtain

$$\begin{aligned}\mathfrak{R}_1+\mathfrak{R}_2&=L_{1\hat{x}}^{\mathsf{T}}\big(\bar{f}(\hat{x})+\bar{g}(\hat{x})u^*+\bar{k}(\hat{x})\omega^*\big)-L_{1\hat{x}}^{\mathsf{T}}\big(\bar{g}(\hat{x})\varepsilon_{u^*}+\bar{k}(\hat{x})\varepsilon_{\omega^*}\big)\\&\quad+\kappa L_{1\hat{x}}^{\mathsf{T}}\bar{g}(\hat{x})O\big((\Phi_1(\hat{x})-\Phi_2(\hat{x}))^2\big)\\&\le-\lambda_{\min}(Q(\hat{x}))\|L_{1\hat{x}}\|^2+c_1\|L_{1\hat{x}}\|+b_{\sigma\hat{x}}\bar{g}_M^2\|L_{1\hat{x}}\|\|\tilde{W}_c\|,\end{aligned} \tag{67}$$

where $c_1=(2\kappa\sqrt{m}+b_{\varepsilon_{u^*}})\bar{g}_M+b_{\varepsilon_{\omega^*}}\bar{k}_M$.

Let $\theta_i\in(0,1)\,(i=1,2)$, and $\theta_1+\theta_2=1$. Then, from (62) and (67), we have

$$\begin{aligned}\dot{L}(t)&\le-\theta_1\lambda_{\min}(Q(\hat{x}))\|L_{1\hat{x}}\|^2+c_1\|L_{1\hat{x}}\|-\theta_2\lambda_{\min}(Q(\hat{x}))\left(\|L_{1\hat{x}}\|-\frac{b_{\sigma\hat{x}}\bar{g}_M^2\|\tilde{W}_c\|}{2\theta_2\lambda_{\min}(Q(\hat{x}))}\right)^2\\&\quad+\frac{b_{\sigma\hat{x}}^2\bar{g}_M^4\|\tilde{W}_c\|^2}{4\theta_2\lambda_{\min}(Q(\hat{x}))}-\lambda_{\min}(M)\|\mathcal{Y}\|^2+\zeta_N\|\mathcal{Y}\|\\&\le-\theta_1\lambda_{\min}(Q(\hat{x}))\|L_{1\hat{x}}\|^2+c_1\|L_{1\hat{x}}\|+\frac{b_{\sigma\hat{x}}^2\bar{g}_M^4\|\tilde{W}_c\|^2}{4\theta_2\lambda_{\min}(Q(\hat{x}))}-\lambda_{\min}(M)\|\mathcal{Y}\|^2+\zeta_N\|\mathcal{Y}\|.\end{aligned} \tag{68}$$

By the definition of $\mathcal{Y}$, we obtain $\|\tilde{W}_c\|^2\le\|\mathcal{Y}\|^2$. Then, we can develop (68) as

$$\begin{aligned}\dot{L}(t)&\le-\theta_1\lambda_{\min}(Q(\hat{x}))\left(\|L_{1\hat{x}}\|-\frac{c_1}{2\theta_1\lambda_{\min}(Q(\hat{x}))}\right)^2+\frac{c_1^2}{4\theta_1\lambda_{\min}(Q(\hat{x}))}\\&\quad-\left(\lambda_{\min}(M)-\frac{b_{\sigma\hat{x}}^2\bar{g}_M^4}{4\theta_2\lambda_{\min}(Q(\hat{x}))}\right)\|\mathcal{Y}\|^2+\zeta_N\|\mathcal{Y}\|\\&=-\theta_1\lambda_{\min}(Q(\hat{x}))\left(\|L_{1\hat{x}}\|-\frac{c_1}{2\theta_1\lambda_{\min}(Q(\hat{x}))}\right)^2-\frac{c_2}{4\theta_2\lambda_{\min}(Q(\hat{x}))}\left(\|\mathcal{Y}\|-\frac{2\theta_2\lambda_{\min}(Q(\hat{x}))\zeta_N}{c_2}\right)^2\\&\quad+\frac{c_1^2}{4\theta_1\lambda_{\min}(Q(\hat{x}))}+\frac{\theta_2\lambda_{\min}(Q(\hat{x}))\zeta_N^2}{c_2},\end{aligned} \tag{69}$$

where $c_2=4\theta_2\lambda_{\min}(M)\lambda_{\min}(Q(\hat{x}))-b_{\sigma\hat{x}}^2\bar{g}_M^4$. Observe that c_2 depends on the parameters θ_2, $\lambda_{\min}(Q(\hat{x}))$, and K_i $(i=1,2)$. Therefore, c_2 can be kept positive by properly selecting these parameters.

For convenience, we denote

$$\mathfrak{T}=\frac{c_1^2}{4\theta_1\lambda_{\min}(Q(\hat{x}))}+\frac{\theta_2\lambda_{\min}(Q(\hat{x}))\zeta_N^2}{c_2}.$$

Then, (69) implies $\dot{L}(t) < 0$ as long as one of the following conditions holds:

$$\|L_{1x}\| > \frac{c_1}{2\theta_1\lambda_{\min}(Q(\hat{x}))} + \sqrt{\frac{\mathfrak{T}}{\theta_1\lambda_{\min}(Q(\hat{x}))}} \triangleq \mathcal{B}_1',$$

or

$$\|\mathcal{Y}\| > \frac{2\theta_2\lambda_{\min}(Q(\hat{x}))\zeta_N}{c_2} + 2\sqrt{\frac{\theta_2\lambda_{\min}(Q(\hat{x}))\mathfrak{T}}{c_2}}. \tag{70}$$

Observe that $\|\mathcal{Y}\| \le (\sqrt{5}/2)\|\tilde{W}_c\|$. Then, (70) yields

$$\|\tilde{W}_c\| > \frac{4\theta_2\lambda_{\min}(Q(\hat{x}))\zeta_N}{\sqrt{5}c_2} + 4\sqrt{\frac{\theta_2\lambda_{\min}(Q(\hat{x}))\mathfrak{T}}{5c_2}} \triangleq \mathcal{B}_2'.$$

Combining *Cases 1* and *2* and using the standard Lyapunov extension theorem [13], we obtain that the function $L_{1\hat{x}}$ is UUB with ultimate bound $\mathcal{B}_1$ (or $\mathcal{B}_1'$) and the critic NN weight estimation error $\tilde{W}_c$ is UUB with ultimate bound $\mathcal{B}_2$ (or $\mathcal{B}_2'$). □

Remark 7. $L_1(\hat{x})$ given in Assumption 4 is often obtained by selecting polynomials. Therefore, $L_{1\hat{x}}$ is also a polynomial with respect to $\hat{x}$. Since Theorem 2 has verified that $L_{1\hat{x}}$ is UUB, we can obtain that the trajectory of the closed-loop system is UUB.

The following theorem is established to show that the estimated control $\hat{u}$ given in (38) and the disturbance $\hat{\omega}$ given in (39) can approximate the optimal control u^* and the worst disturbance ω^* within finite bounds, respectively.

Theorem 3. *Consider system (26) with associated HJI equation (30). Suppose Assumptions 4–7 hold and take the control input and the disturbance input for system (26) as given in (38) and (39), respectively. Meanwhile, let weight update laws for the identifier NN and the critic NN be described by (15) and (46), respectively. Then, the estimated control $\hat{u}$ and the estimated disturbance $\hat{\omega}$ can be close to the optimal control u^* and the worst disturbance ω^* within finite bounds, respectively. In addition, $\hat{V}(\hat{x})$ converges to the optimal cost function $V^*(\hat{x})$ within a small bound $\mathfrak{L}$ (given in (71)).*

Proof. By (35), (38), and using Assumptions 5–7 and Lemma 1, we have

$$\begin{aligned}\|\hat{u} - u^*\| &= \left\|\kappa\left[\tanh(\Phi_1(\hat{x})) - \tanh(\Phi_2(\hat{x}))\right] - \varepsilon_{u^*}\right\| \\ &= \left\|\frac{1}{2}\left[I_m - \mathscr{C}(\Phi_2(\hat{x}))\right]\bar{g}^{\mathsf{T}}(\hat{x})\nabla\sigma^{\mathsf{T}}\tilde{W}_c + \kappa O\left((\Phi_1(\hat{x}) - \Phi_2(\hat{x}))^2\right) - \varepsilon_{u^*}\right\| \\ &\le 2b_{\sigma\hat{x}}\bar{g}_M\left\|\tilde{W}_c\right\| + 2\kappa\sqrt{m} + b_{\varepsilon_{u^*}}.\end{aligned}$$

From Theorem 2, we know that $\tilde{W}_c$ is UUB with ultimate bound $\mathcal{B}_2$ (or $\mathcal{B}_2'$). Denote $\Xi = \max\{\mathcal{B}_2, \mathcal{B}_2'\}$. Then, we derive

$$\|\hat{u} - u^*\| \le 2b_{\sigma\hat{x}}\bar{g}_M\Xi + 2\kappa\sqrt{m} + b_{\varepsilon_{u^*}}.$$

Similarly, using Assumptions 5–7, we obtain

$$\begin{aligned}&\|\hat{\omega} - \omega^*\| \le \frac{1}{2\gamma^2}\bar{k}_M b_{\sigma\hat{x}}\Xi + b_{\varepsilon_{\omega^*}}, \\ &\|\hat{V} - V^*\| \le b_\sigma\Xi + b_{\varepsilon_c} \triangleq \mathfrak{L}.\end{aligned} \tag{71}$$

Remark 8. Noticing the expressions of $\mathcal{B}_2$ and $\mathcal{B}_2'$, we can find that Ξ can be kept very small by selecting proper parameters (e.g., $\lambda_{\min}(M)$ is large enough). In addition, as pointed out in [8,9], if the number of neurons N_0 goes to infinity, there exist $\varepsilon_c \to 0$ and $\nabla\varepsilon_c \to 0$. That is, b_{ε_c} can be kept arbitrarily small. Therefore, $\mathfrak{L}$ given in (71) can be made very small.

6. Simulation results

In this section, two examples are provided to illustrate the effectiveness of the developed theoretical results.

6.1. Example 1

Consider the CT linear system given by

$$\dot{x} = \begin{bmatrix} -1.01887 & 0.90506 & -0.00215 \\ 0.82225 & -1.07741 & -0.17555 \\ 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \omega \tag{72}$$

with the state $x = [x_1, x_2, x_3]^{\mathsf{T}} \in \mathbb{R}^3$, and the control $u \in \mathfrak{U} = \{u \in \mathbb{R} : |u| \le 1\}$. The nonquadratic function is given by

$$\|z\|^2 = x_1^2 + x_2^2 + x_3^2 + 2\kappa\int_0^u \tanh^{-1}(\upsilon/\kappa)\mathrm{d}\upsilon.$$

It is desired to solve the H_∞ optimal control problem with $\gamma = 5$. The prior knowledge of system (72) is assumed to be unavailable. To obtain the knowledge of system (72), the identifier NN (13) is employed. The identifier gains are selected as

$$A = \begin{bmatrix} -2 & 1 & 0 \\ 0 & -0.5 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 1 & 0.5 \\ 0.3 & 0.5 & 1 \end{bmatrix}, \quad \Gamma_2 = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 1 & 0.3 \\ 0.1 & 0.3 & 1 \end{bmatrix}, \quad \Gamma_3 = \begin{bmatrix} 1 & 0.2 & 0.5 \\ 0.2 & 1 & 0.1 \\ 0.5 & 0.1 & 1 \end{bmatrix},$$

and $\beta = 2$, $\eta = 60$. $\phi(x)$, $\rho_i(\zeta_i^{\mathsf{T}} x)$, and $\ell_j(\varsigma_j^{\mathsf{T}} x)$ are chosen as hyperbolic tangent functions $\tanh(x)$, $\tanh(\zeta_i^{\mathsf{T}} x)$, and $\tanh(\varsigma_j^{\mathsf{T}} \hat{x})$, respectively. ζ_i and ς_j $(i, j = 1, 2, 3)$ are selected randomly within an interval of $[-1, 1]$ and held constant. Meanwhile, the initial weights $\hat{W}_f$, $\hat{W}_g$, and $\hat{W}_k$ are all chosen randomly within the interval of $[-1, 1]$.

The gains for the critic NN are given as $l = 0.95$ and $\kappa = 1$. The activation function for the critic NN is chosen with $N_0 = 6$ neurons as

$$\sigma(x) = \left[x_1^2, x_2^2, x_3^2, x_1 x_2, x_1 x_3, x_2 x_3\right]^{\mathsf{T}},$$

and the weight of the critic NN is denoted as $\hat{W}_c = \left[\hat{W}_{c1}, \hat{W}_{c2}, \ldots, \hat{W}_{c6}\right]^{\mathsf{T}}$.

Remark 9. It should be emphasized that, the number of neurons required for any particular application is still an open problem. Choosing the proper number of neurons for NNs is more of an art than science [26]. In this example, the number of neurons is obtained by computer simulations. We find that selecting 6 neurons in the hidden layer for the critic NN can lead to satisfactory simulation results.

The initial state is $x_0 = [3, -0.5, 0.5]^{\mathsf{T}}$ (Note: x_0 can be selected arbitrarily in $D_1 = \{2 \le x_1 \le 3; -0.5 \le x_2 \le 0.5; -0.5 \le x_3 \le 0.5\}$. For simplicity of discussion, we assume $x_0 = [3, -0.5, 0.5]^{\mathsf{T}}$). Meanwhile, the initial weights for the critic NN are chosen randomly within an interval of $[0, 2]$. In this sense, by using (38), we can find that there is no a special requirement imposed on the initial control; that is, no initial stabilizing control is required. Since system (72) is linear, we choose $L_1(x) = 0.5 x^{\mathsf{T}} x$. To guarantee the PE condition, a small exploratory signal $n(t) = 8e^{(-0.2t)}[\sin(t)\cos(t) + \sin^3(2t)\cos(0.2t) + \sin^5(1.2t)]$ is added to the control $u(t)$ for the first 24 sec.

The computer simulation results are shown in Figs. 1–5. Fig. 1 illustrates the system identification error. Fig. 2 presents the convergence of the critic NN weight matrix, where it converges to $[1.4376, 0.1573, 0.5998, 0.3897, -0.7959, 0.7393]^{\mathsf{T}}$. By Theorem 3, the critic NN is considered to arrive at the approximate optimal value. Then the approximate solution of the HJI equation (10) can be computed with (37) and the nearly optimal control policy is obtained via (38). The disturbance signal is given as $\omega(t) = 3r(t)e^{-0.2t}\cos(t)$ with $r(t)$ randomly chosen within an interval $[0, 1]$. Fig. 3 indicates the state trajectories of the closed-loop system when system (72) is at rest and experiencing the disturbance $\omega(t)$. Fig. 4 presents the control input for the closed-loop system. Define the ratio of the disturbance attenuation as

$$\gamma_d = \left(\frac{\int_0^\infty \left(h^{\mathsf{T}} h + \|u\|^2\right) \mathrm{d}t}{\int_0^\infty \|\omega(t)\|^2 \mathrm{d}t} \right)^{1/2}. \tag{73}$$

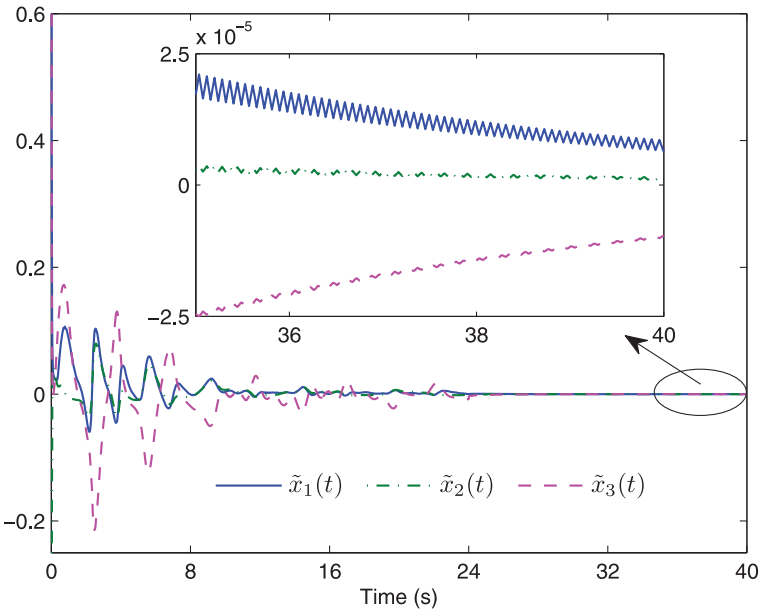

Fig. 1. System identification error $\tilde{x}(t)$ in Example 1.

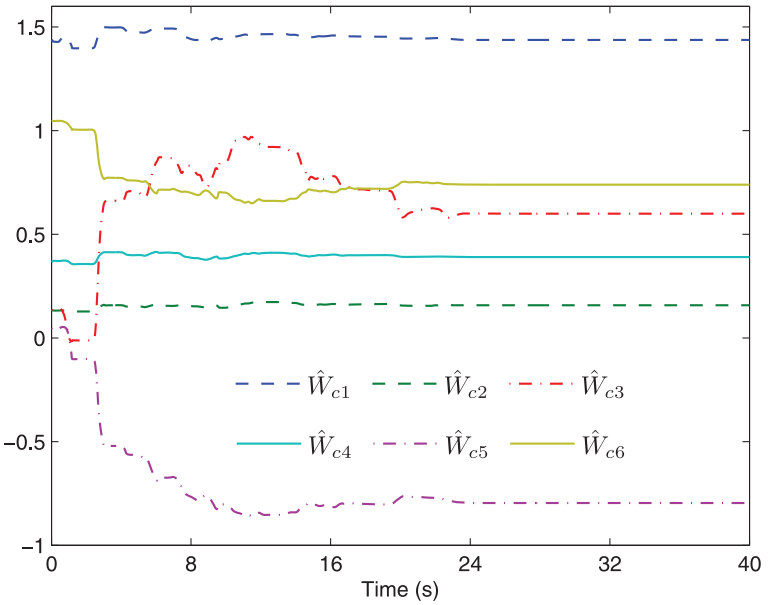

Fig. 2. Convergence of the critic NN weight $\hat{W}_c$ in Example 1.

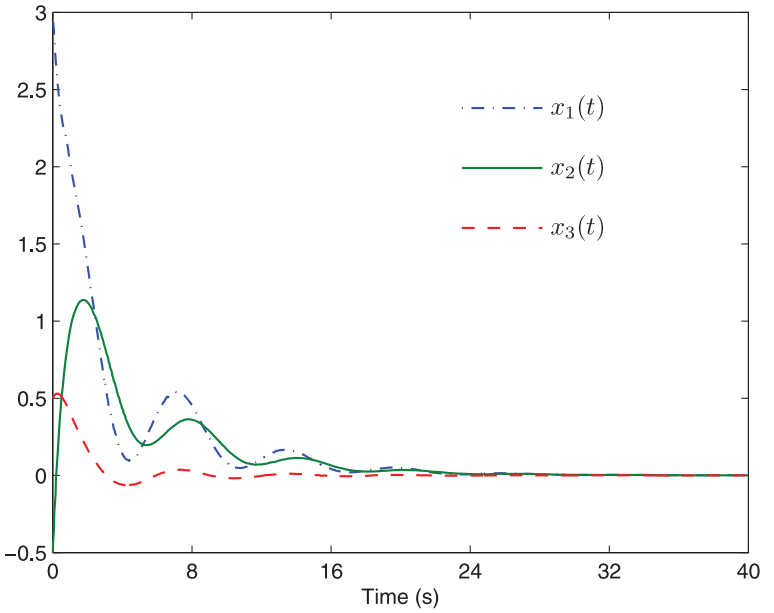

Fig. 3. State trajectories $x_i(t)(i = 1, 2, 3)$ of the closed-loop system in Example 1.

Fig. 5 illustrates the evolution of γ_d, where it converges to $1.6935(< \gamma = 5)$. Therefore, the obtained control policy can achieve a prescribed L_2-gain performance level γ for the closed-loop system.

6.2. Example 2

Consider the CT nonlinear system [33] given by

$$\dot{x} = f(x) + g(x)u + k(x)\omega, \tag{74}$$

where

$$f(x) = \begin{bmatrix} -x_1 + x_2 \\ -x_1^3 - x_2^3 + 0.25x_2(\cos(2x_1) + 2)^2 - 0.25x_2\gamma^{-2}\left(\sin(4x_1^2) + 2\right)^2 \end{bmatrix},$$

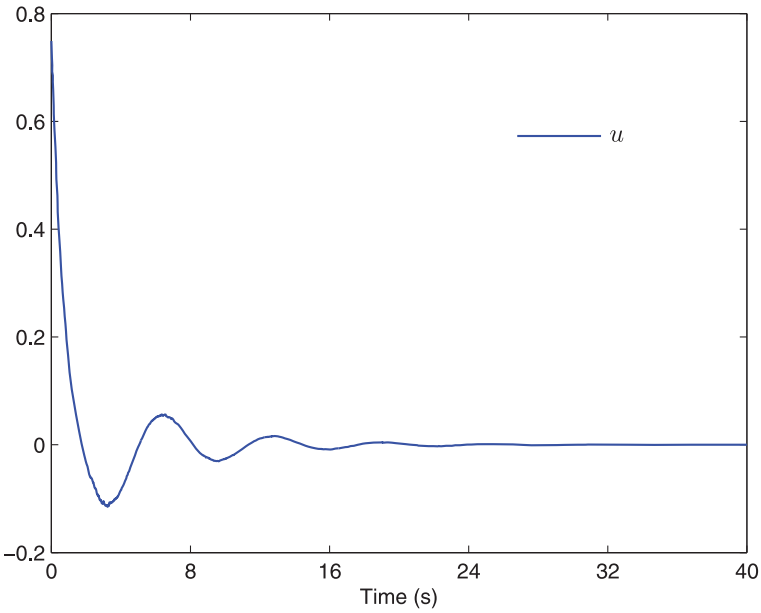

Fig. 4. Control input u of the closed-loop system in Example 1.

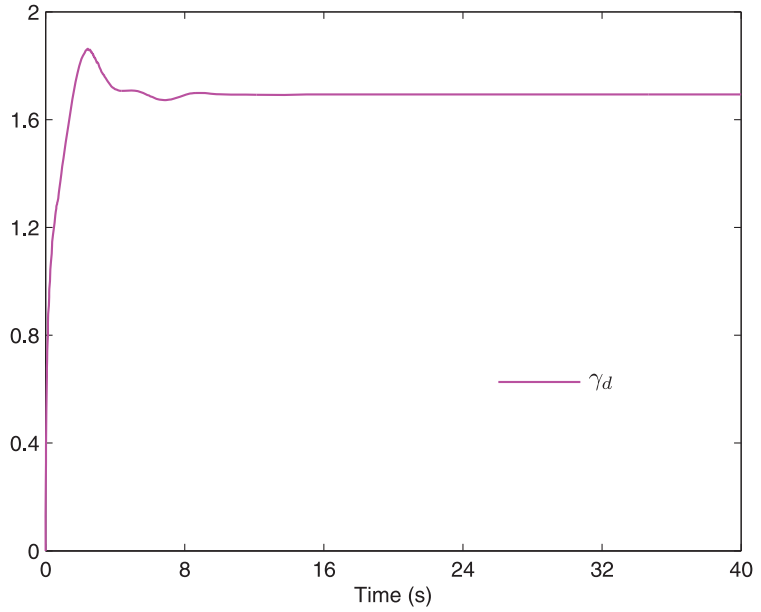

Fig. 5. Evolution of γ_d in Example 1.

$$g(x) = \begin{bmatrix} 0 \\ \cos(2x_1 + 2) \end{bmatrix}, \qquad k(x) = \begin{bmatrix} 0 \\ \sin(4x_1^2) + 2 \end{bmatrix}.$$

with the state $x = [x_1, x_2]^{\mathsf{T}} \in \mathbb{R}^2$, and the control $u \in \mathfrak{U} = \{u \in \mathbb{R} : |u| \leq 1\}$. The nonquadratic function is given by

$$\|z\|^2 = x_1^2 + x_2^2 + 2\kappa \int_0^u \tanh^{-1}(\upsilon/\kappa) \mathrm{d}\upsilon.$$

It is desired to solve the H_∞ optimal control problem with $\gamma = 4$. The prior knowledge of system (74) is assumed to be unknown. The dynamic NN (13) is utilized to identify system (74). The identifier gains are chosen as

$$A = \begin{bmatrix} -2 & 0.5 \\ 0 & -1 \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}, \quad \Gamma_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \quad \Gamma_3 = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix},$$

and $\beta = 2$, $\eta = 20$. $\phi(x)$, $\rho_i(\zeta_i^{\mathrm{T}}x)$, and $\ell_j(\varsigma_j^{\mathrm{T}}x)$ are hyperbolic tangent functions $\tanh(x)$, $\tanh(\zeta_i^{\mathrm{T}}x)$, and $\tanh(\varsigma_j^{\mathrm{T}}x)$, respectively. ζ_i and ς_j $(i, j = 1, 2)$ are selected randomly within an interval of $[-1, 1]$ and kept constant. The initial weights $\hat{W}_f$, $\hat{W}_g$, and $\hat{W}_k$ are all chosen randomly within the interval of $[-1, 1]$. The gains for the critic NN are given as $l = 0.8$ and $\kappa = 1$. The activation function for the critic NN is chosen with $N_0 = 8$ neurons as

$$\sigma(x) = \left[x_1^2, x_2^2, x_1x_2, x_1^4, x_2^4, x_1^3x_2, x_1^2x_2^2, x_1x_2^3\right]^{\mathrm{T}},$$

and the weight of the critic NN is denoted as $\hat{W}_c = \left[\hat{W}_{c1}, \hat{W}_{c2}, \ldots, \hat{W}_{c8}\right]^{\mathrm{T}}$. Similar to Example 1, the number of neurons is obtained by computer simulations.

The initial system state is $x_0 = [1.5, -0.5]^{\mathrm{T}}$ (Note: x_0 can be selected arbitrarily in $D_2 = \{1 \le x_1 \le 2; -0.5 \le x_2 \le 0.5\}$. For simplicity of discussion, we assume $x_0 = [1.5, -0.5]^{\mathrm{T}}$). Meanwhile, the initial weights for the critic NN are chosen randomly within an interval of $[-0.5, 0.5]$. In this sense, it implies that no initial stabilizing control is required. Due to the expression of system (74),

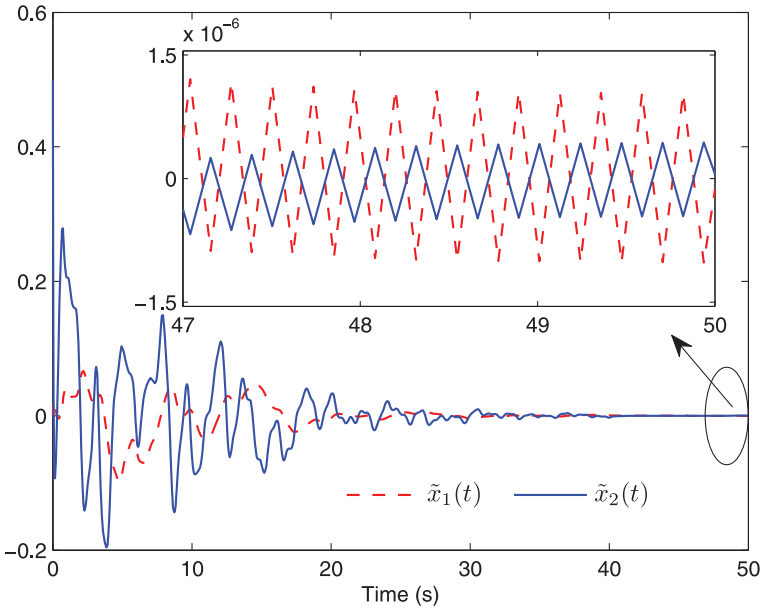

Fig. 6. System identification error $\tilde{x}(t)$ in Example 2.

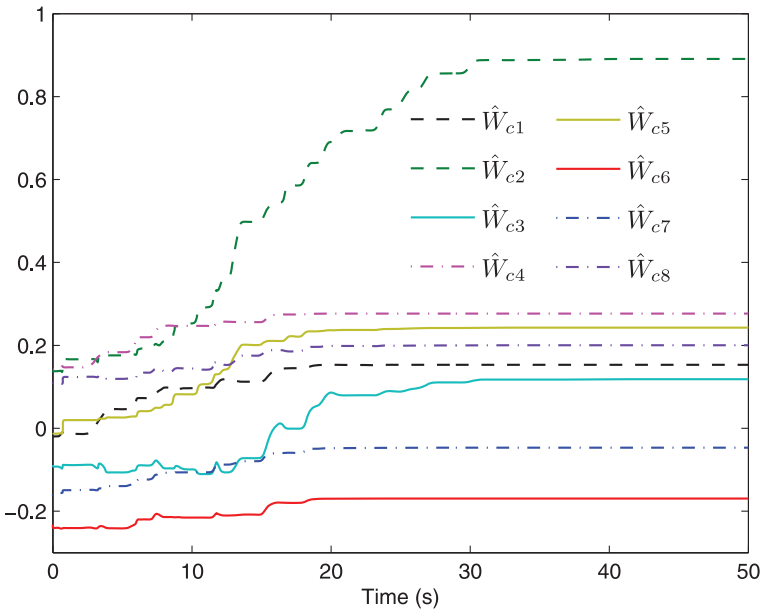

Fig. 7. Convergence of the critic NN weight $\hat{W}_c$ in Example 2.

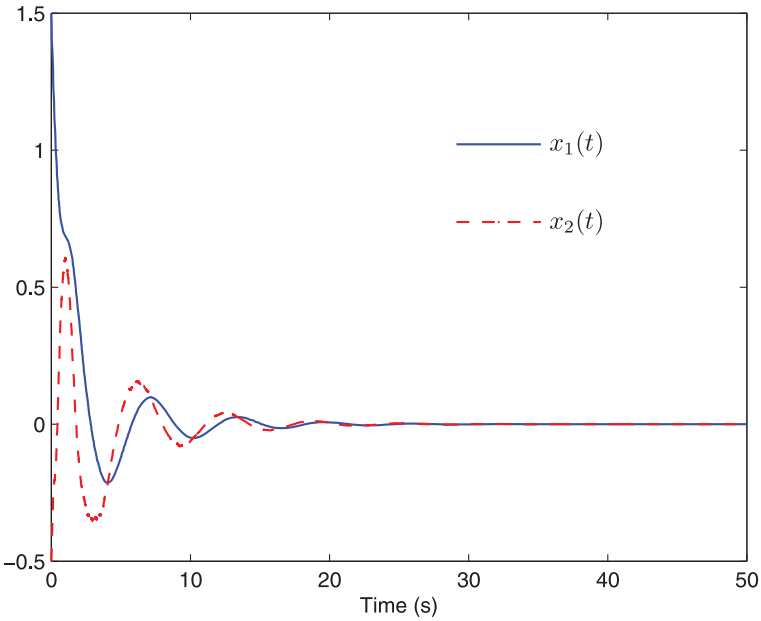

Fig. 8. State trajectories $x_i(t)(i = 1, 2)$ of the closed-loop system in Example 2.

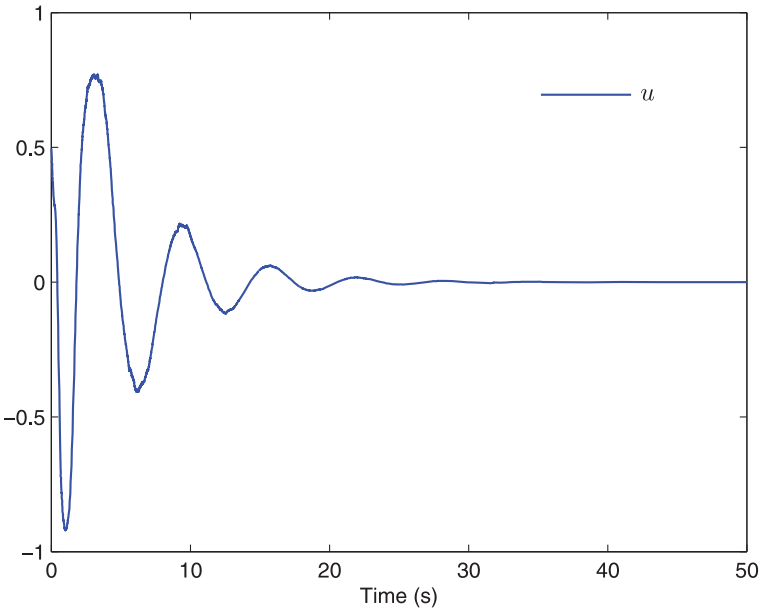

Fig. 9. Control input u of the closed-loop system in Example 2.

we choose $L_1(x) = 0.25(x^{\mathsf{T}}x)x^{\mathsf{T}}x$. To guarantee the PE condition, a small exploratory signal $n(t) = 2.4e^{(-0.1t)}[\sin^2(t)\cos(t) + \sin^2(-1.2t)\cos(0.5t) + \cos(2.4t)\sin^3(2.4t) + \sin^5(t)]$ is added to the control $u(t)$ for the first 40 sec.

The computer simulation results are shown in Figs. 6–10. Fig. 6 presents the system identification error. Fig. 7 shows the convergence of the critic NN weight matrix, where it converges to $[0.1531, 0.8908, 0.1183, 0.2765, 0.2426, -0.1694, -0.0468, 0.2002]^{\mathsf{T}}$. By Theorem 3, the critic NN is considered to reach the approximate optimal value. Then the approximate solution of the HJI equation (10) can be computed with (37), and the nearly optimal control policy is derived via (38). The disturbance signal $\omega(t)$ is the same as in Example 1. Fig. 8 shows the state trajectories of the closed-loop system when system (74) is at rest and experiencing the disturbance $\omega(t)$. Fig. 9 presents the control input for the closed-loop system. The ratio of the disturbance attenuation γ_d is defined as (73). Fig. 10 indicates the evolution of γ_d, where it converges to 0.9986 ($< \gamma = 4$). Accordingly, the developed control law can achieve a prescribed L_2-gain performance level γ for the closed-loop system.

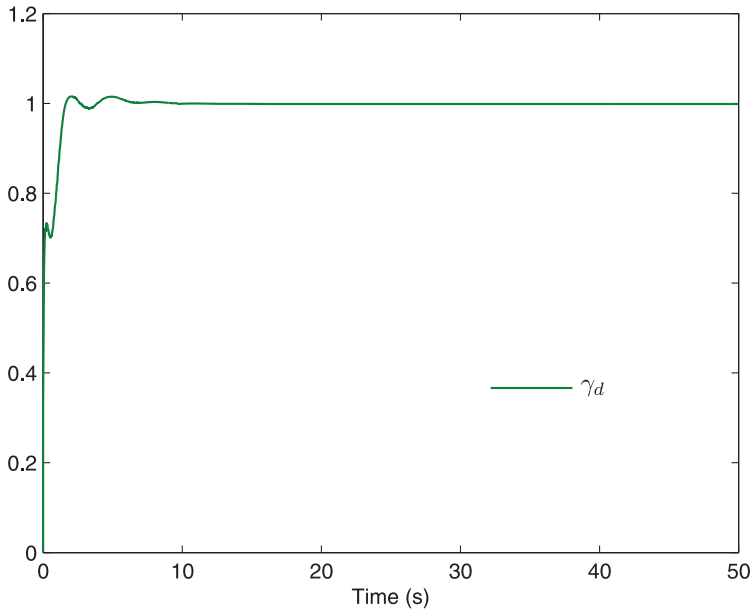

Fig. 10. Evolution of γ_d in Example 2.

7. Conclusions

In this paper, we have presented a new ADP-based algorithm which solves the HJI equation for constrained-input affine nonlinear CT systems in the presence of unknown dynamics. The algorithm employs an identifier-critic architecture. Based on the present algorithm, the identifer NN and the critic NN are tuned simultaneously. Meanwhile, no initial stabilizing control is required. A limitation of the present algorithm is that the system state is required to be available. In our future work, we shall remove this condition. Furthermore, due to the output of nonaffine nonlinear systems depending nonlinearly on the control input, it will be more intractable to obtain the solutions of HJI equations for nonaffine nonlinear systems than affine nonlinear systems. Therefore, how to develop efficient online learning algorithms to solve HJI equations for nonaffine nonlinear systems is also a direction of our future work.

References

[1] M. Abu-Khalaf, F.L. Lewis, J. Huang, Policy iterations on the Hamilton–Jacobi–Isaacs equation for state feedback control with input saturation, IEEE Trans. Autom. Control 51 (2006) 1989–1995.
[2] M. Abu-Khalaf, F.L. Lewis, J. Huang, Neurodynamic programming and zero-sum games for constrained control systems, IEEE Trans. Neural Netw. 19 (2008) 1243–1252.
[3] M. Aliyu, Nonlinear H_∞ Control, Hamiltonian Systems and Hamilton–Jacobi Equations, CRC Press, Boca Raton, FL, 2011.
[4] T. Basar, P. Bernhard, H_∞ Optimal Control and Related Minimax Design Problems, Birkhäuser, Boston, USA, 1995.
[5] R.W. Beard, T.W. Mclain, Successive Galerkin approximation algorithms for nonlinear optimal and robust control, Int. J. Control 71 (1998) 717–743.
[6] S. Bhasin, R. Kamalapurkar, M. Johnson, K.G. Vamvoudakis, F.L. Lewis, W.E. Dixon, A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems, Automatica 49 (2013) 82–92.
[7] T. Dierks, S. Jagannathan, Optimal control of affine nonlinear continuous-time systems using an online Hamilton–Jacobi–Isaacs formulation, in: Proceedings of the 49th IEEE Conference on Decision and Control, Atlanta, GA, USA, 2010, pp. 3048–3053.
[8] B.A. Finlayson, The Method of Weighted Residuals and Variational Principles, Academic Press, New York, 1972.
[9] K. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Netw. 2 (1989) 359–366.
[10] J. Huang, An algorithm to solve the discrete HJI equation arising in the L_2 gain optimization problem, Int. J. Control 72 (1999) 49–57.
[11] M. Johnson, S. Bhasin, W. Dixon, Nonlinear two-player zero-sum game approximate solution using a policy iteration algorithm, in: Proceedings of IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, USA, 2011, pp. 142–147.
[12] M. Johnson, R. Kamalapurkar, S. Bhasin, W.E. Dixon , Approximate N-player nonzero-sum game solution for an uncertain continuous nonlinear system, IEEE Trans. Neural Netw. Learn. Syst. 26 (8) (2015) 1645–1658.
[13] F.L. Lewis, S. Jagannathan, A. Yesildirak, Neural Network Control of Robot Manipulators and Nonlinear Systems, Taylor & Francis, London, UK, 1999.
[14] F.L. Lewis, D. Liu, Reinforcement Learning and Approximate Dynamic Programming for Feedback Control, Wiley-IEEE Press, Hoboken, New Jersey, 2013.
[15] F.L. Lewis, D. Vrabie, Reinforcement learning and adaptive dynamic programming for feedback control, IEEE Circuits Syst. Mag. 9 (2009) 32–50.
[16] 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.
[17] D. Liu, H. Li, D. Wang, Online synchronous approximate optimal learning algorithm for multiplayer nonzero-sum games with unknown dynamics, IEEE Trans. Syst. Man Cybern. Syst. 44 (2014) 1015–1027.
[18] D. Liu, D. Wang, X. Yang, An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs, Inf. Sci. 220 (2013) 331–342.
[19] B. Luo, H.N. Wu, T.W. Huang, Off-policy reinforcement learning for H_∞ control design, IEEE Trans. Cybern. 45 (2015) 65–76.

[20] S. Mehraeen, T. Dierks, S. Jagannathan, M.L. Crow, Zero-sum two-player game theoretic formulation of affine nonlinear discrete-time systems using neural networks, IEEE Trans. Cybern. 43 (2013) 1641–1655.
[21] 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 Netw. Learn. Syst 24 (2013) 1513–1525.
[22] H. Modares, F.L. Lewis, M.B. Naghibi-Sistani, Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems, Automatica 50 (2014) 193–202.
[23] H. Modares, F.L. Lewis, M.B.N. Sistani, Online solution of nonquadratic two-player zero-sum games arising in the H_∞ control of constrained input systems, Int. J. Adapt. Control Signal Process. 28 (2014) 232–254.
[24] Z. Ni, H. He, J. Wen, Adaptive learning in tracking control based on the dual critic network design, IEEE Trans. Neural Netw. Learn. Syst. 24 (2013) 913–928.
[25] D. Nodland, H. Zargarzadeh, J. Sarangapani, Neural network-based optimal adaptive output feedback control of a helicopter UAV, IEEE Trans. Neural Netw. Learn. Syst. 24 (2013) 1061–1073.
[26] R. Padhi, N. Unnikrishnan, X. Wang, S. Balakrishnan, A single network adaptive critic (SNAC) architecture for optimal control synthesis for a class of nonlinear systems, Neural Netw. 19 (2006) 1648–1660.
[27] W.B. Powell, Approximate Dynamic Programming: Solving the Curses of Dimensionality, John Wiley & Sons, Hoboken, New Jersey, 2007.
[28] J. Rubio, W. Yu, Stability analysis of nonlinear system identification via delayed neural networks, IEEE Trans. Circuits Syst. II: Expr. Br. 54 (2007) 161–165.
[29] W. Rudin, Principles of Mathematical Analysis, McGraw-Hill Publishing Co., USA, 1976.
[30] M. Sassano, A. Astolfi, Dynamic approximate solutions of the HJ inequality and of the HJB equation for input-affine nonlinear systems, IEEE Trans. Autom. Control 57 (2012) 2490–2503.
[31] A.J. van der Schaft, L_2-gain and Passivity Techniques in Nonlinear Control, Springer, London, 2000.
[32] K.G. Vamvoudakis, F.L. Lewis, Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton–Jacobi equations, Automatica 47 (2011) 1556–1569.
[33] K.G. Vamvoudakis, F.L. Lewis, Online solution of nonlinear two-player zero-sum games using synchronous policy iteration, Int. J. Robust Nonlinear Control 22 (2012) 1460–1483.
[34] D. Wang, D. Liu, H. Li, H. Ma, Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming, Inf. Sci. 282 (2014) 167–179.
[35] F.Y. Wang, H. Zhang, D. Liu, Adaptive dynamic programming: an introduction, IEEE Comput. Intell. Mag. 4 (2009) 39–47.
[36] P. Werbos, Beyond regression: new tools for prediction and analysis in the behavioral sciences, Harvard University, USA, 1974 (Ph.d. thesis).
[37] P.J. Werbos, Intelligence in the brain: a theory of how it works and how to build it, Neural Netw. 22 (2009) 200–212.
[38] H.N. Wu, B. Luo, Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 1884–1895.
[39] X. Yang, D. Liu, Y. Huang, Neural-network-based online optimal control for uncertain nonlinear continuous-time systems with control constraints, IET Control Theory Appl. 7 (2013) 2037–2047.
[40] X. Yang, D. Liu, D. Wang, Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints, Int. J. Control 87 (2014) 553–566.
[41] X. Yang, D. Liu, Q. Wei, Online approximate optimal control for affine nonlinear systems with unknown internal dynamics using adaptive dynamic programming, IET Control Theory Appl. 8 (2014) 1676–1688.
[42] W. Yu, Recent Advances in Intelligent Control Systems, Springer, London, UK, 2009.
[43] H. Zhang, L. Cui, Y. Luo, Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP, IEEE Trans. Cybern. 43 (2013) 206–216.
[44] H. Zhang, L. Cui, X. Zhang, Y. Luo, Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method, IEEE Trans. Neural Netw. 22 (2011) 2226–2236.
[45] H. Zhang, C. Qin, B. Jiang, Y. Luo, Online adaptive policy learning algorithm for H_∞ state feedback control of unknown affine nonlinear discrete-time systems, IEEE Trans. Cybern. 44 (2014) 2706–2718.
[46] H. Zhang, Q. Wei, D. Liu, An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games, Automatica 47 (2011) 207–214.