Multi-H∞ Controls for Unknown Inputs-interference Nonlinear System with Reinforcement Learning

Yongfeng Lv, *Member, IEEE*, Jing Na, *Member,IEEE,* Xiaowei Zhao, Yingbo Huang, Xuemei Ren

***Abstract*—This paper studies the multi-H∞ controls for the inputs-interference nonlinear systems via adaptive dynamic programming (ADP) method, which allows for multiple inputs to have the individual selfish component of the strategy to resist weighted interference. In this line, the ADP scheme is used to learn the Nash-optimization solutions of the inputs-interference nonlinear system such that multiple H∞ performance indices can reach the defined Nash equilibrium. Firstly, the inputs-interference nonlinear system is given and the Nash equilibrium is defined. An adaptive neural network (NN) observer is introduced to identify the inputs-interference nonlinear dynamics. Then, the critic NNs are used to learn the multiple H∞ performance indices. A novel adaptive law is designed to update the critic NN weights by minimizing the Hamiltonian-Jacobi-Isaacs (HJI) equation, which can be used to directly calculate the multi-H∞ controls effectively by using input-output data, such that the actor structure is avoided. Moreover, the control system stability and updated parameter convergence are proved. Finally, two numerical examples are simulated to verify the proposed ADP scheme for the inputs-interference nonlinear system.**

***Index Terms*—H∞ control, adaptive dynamic programming, multi-input system, neural networks, nonlinear system.**

I. INTRODUCTION

Reinforcement learning (RL) refers to actors or agents modify their actions based on rewards and punishments in response to their actions in the environment. One typical RL method is known as the critic-actor structure, where two neuronlike adaptive elements are used to solve a difficult learning control problem as presented in [1]. Lewis extended the actor-critic based RL method, and designed an optimal feedback control in [2], which is named as adaptive dynamic programming (ADP). It has been shown that ADP is an effective intelligent method to solve the optimal control of nonlinear systems because 'dimension disaster' in the nonlinear Hamiltonian-Jacobi-Bellman (HJB) equation is avoided [3], and it has been used to solve the practical engineering problem in [4]. The event-triggered tracking control problem was also addressed by using a single-network based ADP in [5]. However, previous work only considered the general nonlinear or linear system. In recent years, ADP has been also used in several multi-input systems, such as the nonzero-sum (NZS) game with constraint input in [6] and zero-sum game in [7]. The NZS optimal problems are to obtain a series of optimal strategies of nonlinear or linear games, such that the game system can track the predetermined target and the performance index of each input can reach the optimal value under the Nash equilibrium as presented in [8]. The linear NZS game is resolved with the multiple coupled Riccati equations, but the optimal problem of the nonlinear system is a more intractable issue because of the nonlinear characteristics. As an intelligent learning algorithm, ADP has provided a novel and effective method for NZS and zero-sum games.

There have been some research work on the zero-sum games in [9], [10], where two players compete with each other to reach a Nash equilibrium. This situation is similar to the H∞ control problem because both of them have the minimax value functions. However, zero-game is a more comprehensive theory than H∞ control. In this case, one policy is obtained to try to resist interference such that the value function converges to zero or is minimized. A model-free Q-learning scheme was proposed in [11] to study the solution of the linear zero-sum system, where the system knowledge is avoided. Wei et al. applied the ADP structure into the two-player zero-sum game in [12], where the saddle-point is not necessary and the stability of the performance index is analyzed. An adaptive critic structure with unknown system information is then developed for the discrete-time (DT) zero-sum games in [13], where only the measured data is required for searching the saddle point. An integral ADP is presented to online determine the zero-sum Nash equilibrium in [14], where the offline learning capability is enhanced. Vamvoudakis et al. in [15] solved the bounded L_2-gain problem with the zero-sum game theory using an online synchronous ADP algorithm. Liu et al. in [16] proposed an iterative ADP for DT zero-sum affine nonlinear systems, where

Manuscript received April 18, 2021; revised September 23, 2021; accepted November 17, 2021. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62103296, 61922037, 61873115, 62003153 and 61973036; and in part by the UK Engineering and Physical Sciences Research Council under grant EP/S001905/1. (*Corresponding authors: Xiaowei Zhao; Jing Na.*)

Yongfeng Lv is with the Intelligent Control and Smart Energy (ICSE), the College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China, and also with the School of Engineering, University of Warwick, Coventry CV47AL, UK (e-mail: lvyilian1989@foxmail.com).

Jing Na and Yingbo Huang are with the Faculty of Mechanical & Electrical Engineering, Kunming University of Science & Technology, Kunming 650500, China (e-mail: najing25@163.com; Yingbo_Huang@126.com).

Xiaowei Zhao is with the Intelligent Control and Smart Energy (ICSE), the School of Engineering, University of Warwick, Coventry CV47AL, UK (e-mail: Xiaowei.Zhao@warwick.ac.uk).

Xuemei Ren is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: xmren@bit.edu.cn).

three NNs are presented to approximate the action, the disturbance, and the performance index, respectively. All these reveal that ADP has been widely used to solve the H∞ control or zero-sum games recently.

On the contrary, in the NZS game or optimal controls of multi-input system, multiple inputs try to cooperate with each other and have a selfish policy to optimize the performance index such that a saddle point can be obtained for all the value functions [17]. Vamvoudakis et al. applied an online iterative ADP method to solve the optimal strategy of cooperative game for multi-input system in [18], where the solutions of both the coupled Riccati equations and coupled Hamilton–Jacobi equations are studied, so that the Nash equilibrium is acquired. Song applied the integral reinforcement learning (RL) method with an asynchronous algorithm to study the strategy of nonlinear multi-player in [19]. A synchronous iterative RL method is introduced for the NZS game in [20]. Then, the finite-horizon optimal method is proposed to learn the nonzero-sum game solution with input constraints and partially known system dynamics in [6]. An online ADP method in [21] is developed for unknown NZS nonlinear games. ADP is an effective algorithm for solving the NZS games or optimal control problem of multi-input system.

It is obvious that two different situations in the game theory have been widely studied in the literature [22], [23]: two-person zero-sum competition game and multi-person NZS cooperation game [24], [25]. However, the multi-input model including both game situations has not been fully considered in the previous work, which can be used to solve multiple H∞ controls of multi-input system with disturbance. Moreover, that allows for inputs to have an individual selfish average component of the strategy to resist interference. Nevertheless, there is a Nash equilibrium point between the multiple H∞ performance indicators of multiple inputs. This multiple H∞ control theory can solve the optimal problem of the multi-driven servo system with disturbance that can be applied to the large radars, gantry planer, large shearer and other industry equipment. When the multi-motor driven load system suffers to the disturbance, the proposed multi-H∞ controls can provide an effective and robust control for the better performance as presented in [26].

To address this issue, this paper considers an inputs-interference nonlinear system, which allows for multiple inputs to have the individual selfish component of the strategy to resist disturbance. It is the extension and development of zero-sum and nonzero-sum game theory, and can provide an effective solution to the H∞ controls of the multi-input system with interference. This type of system can be widely found in engineering and economic fields. However, due to the coupling relationship, it is very difficult to use multiple inputs to resist the interference such that the multi-input system with disturbance can be stabilized in an optimal manner. The RL-based ADP algorithm presented in [26], [27] has been widely used to solve the optimal problem of multi-input system [6], [13], [28]. Therefore, this scheme can also provide an effective way for the optimal solution to the proposed the inputs-interference nonlinear system, such that the obtained inputs can make multiple H∞ performance indicators reach an equilibrium point, and satisfy the global optimization. Moreover, we prove that the proposed ADP-based inputs-interference nonlinear model is stable with Lyapunov theory.

The contribution of the paper can be summarized as

1) An inputs-interference nonlinear system is considered, where the multiple inputs cooperate with each other to resist interference.
2) A weighted average method is proposed to realize the balanced resistance of each input to the interference according to its own dynamics.
3) Multi-H∞ controls of the inputs-interference system are studied via the ADP algorithm, such that the system can be stabilized in an optimal manner with the disturbance.

The paper is structured as follows. To address the multi-H∞ control issue, Section II firstly proposes the inputs-interference nonlinear model and defines the Nash equilibrium of multiple H∞ performance indices. Section III introduces an adaptive NN identifier to approximate the completely unknown inputs-interference nonlinear model. Section IV develops the ADP learning scheme to solve the coupled HJI equations, and obtained the optimal policies with the worst interference. Section V proves the stability of inputs-interference nonlinear model. Section VI shows numerical results. Section VII gives the conclusions.

II. H∞ Property Definition For Inputs-interference Nonlinear System

The studied multi-input nonlinear system with the interference is given as

$$\dot{x} = f(x) + \sum_{j=1}^{N} g_j(x)u_j + k(x)\omega \qquad (1)$$

where $x \in \mathbb{R}^n$ is the system state, $u_j \in \mathbb{R}^m$ ($j \in N$) are the policy inputs, $\omega \in \mathbb{R}^q$ is the unknown bounded input interference, which can be caused by the system load, interference, or other external unknown factors. $N \geq 2$ is a positive integer. $f(x) \in \mathbb{R}^n$, $g_j(x) \in \mathbb{R}^{n\times m}$ and $k(x) \in \mathbb{R}^{n\times q}$ are dynamic information. Supposed that $f(0) = 0$, $f(x)$, $g_j(x)$, $k(x)$ are Lipschitz continuous, and multi-input model (1) is controllable such that there exist continuous policies on Ω to asymptotically stabilize the system.

The multi-input systems in practical engineering are usually subjected to external interference. In this system (1), the multiple inputs $\{u_1,\ldots,u_j,\ldots,u_N\}$ cooperate to resist the interference ω . However, how to use multiple inputs to resist the interference is a very intractable problem. Because H∞ policy has the capability to solve the disturbance in the system, it can be used for the inputs-interference nonlinear system. H∞ policies $\left\{u_1^*,\cdots,u_i^*,\cdots,u_N^*,\omega^*\right\}$ for the multi-input system seek for minimizing the value functions, simultaneously attenuating the worst interference, which can be regarded as a zero-sum game. Simultaneously, the multi-H∞ cost functions should reach a Nash equilibrium as the nonzero-sum game. Thus, the following definition is given.

Definition 1 (L_2-gain): There are some policies $\{u_1,\cdots,u_N\}$ for the multi-input system (1) that can make $L_2-gain \leq \gamma$ with a positive constant γ if

$$\int_t^\infty \left\| z_i(x(\tau),u_1(\tau),\ldots u_N(\tau)) \right\|^2 d\tau \le \gamma^2 \int_t^\infty \left\| \omega(\tau) \right\|^2 d\tau ,\ i \in \mathbb{N} \quad (2)$$

where $\left\| z_i(x(\tau),u_1(\tau),\ldots,u_N(\tau)) \right\|^2 = x^T Q_i x + \sum_{j=1}^N u_j^T R_{ij} u_j$ with symmetric positive definite matrices Q_i and R_{ij}, and $\omega(\tau) \in L_2[t,+\infty)$.

To realize the balanced resistance of each input to the interference, we decouple the interference as

$$\omega(\tau) = \omega_1(\tau) + \ldots + \omega_i(\tau) + \ldots + \omega_N(\tau) ,\ i \in \mathbb{N} \quad (3)$$

where $\omega_i(\tau)$ will be defined in the following content, $\mathbb{N}$ is the positive integer set. The decoupled interference can guarantee that every input resists part of the external interference such that all the H∞ performance indices reach a Nash equilibrium.

Then, the infinite horizon cost functions of H∞ control for the inputs-interference system associated with each input is given as

$$J_i(x(\tau)) = \int_t^\infty r_i(x,u_1,\cdots,u_i,\cdots,u_N,\omega) d\tau ,\ i \in \mathbb{N} \quad (4)$$

where $r_i(x,u_1,\cdots,u_N,\omega) = x^T Q_i x + \sum_{j=1}^N u_j^T R_{ij} u_j - \gamma^2 \left\| \omega \right\|^2$ is the utility function. The H∞ policies $\{u_1,\cdots,u_i,\cdots,u_N\}$ will be obtained to minimize the cost functions, while ω is obtained to maximize them. It can be formularized as

$$J_i^*(x(\tau)) = \min_{u_i} \max_{\omega} \int_t^\infty r_i(x(\tau),u_1(\tau),\cdots,u_i(\tau),\cdots,u_N(\tau),\omega(\tau)) d\tau . \quad (5)$$

For system (1) and cost function (4), $\{u_1,\cdots,u_i,\cdots,u_N\}$ are the admissible policies.

The optimal policy sequence will be calculated such that all the H∞ indices of the inputs-interference system can be optimized and the equilibrium $\{u_1^*,\cdots,u_i^*,\cdots,u_N^*,\omega^*\}$ can be researched. The optimal policies $\{u_1,\cdots,u_N\}$ will minimize the cost functions (4) while the interference $\omega(x)$ will maximize it. Thus, the optimal cost functions of each input are described as

$$V_i^*(x) = \min_{u_i} \max_{\omega} \int_t^\infty r_i(x,u_1,\cdots,u_i,\cdots,u_N,\omega) d\tau ,\ i \in \mathbb{N} . \quad (6)$$

The N-tuple of $\{V_1^*(x),\cdots,V_i^*(x),\cdots,V_N^*(x)\}$ are known as an outcome of the H∞ policies for the inputs-interference system. The Nash equilibrium policies $\{u_1^*,\cdots,u_N^*,\omega^*\}$ of the inputs-interference system are defined as

Definition 2 (Nash Equilibrium) [29]: In the inputs-interference nonlinear system, the Nash equilibrium $\{u_1^*,\cdots,u_i^*,\cdots,u_N^*,\omega^*\}$ is obtained, if $\{u_1^*,\cdots,u_i^*,\cdots,u_N^*,\omega^*\}$ satisfy the following inequalities

$$\begin{aligned} V_i(u_1^*,\cdots,u_i^*,\cdots,u_N^*,\omega) &\le V_i(u_1^*,\cdots,u_i^*,\cdots,u_N^*,\omega^*) \\ &\le V_i(u_1^*,\cdots,u_i,\cdots,u_N^*,\omega^*) ,\ i \in \mathbb{N} . \end{aligned} \quad (7)$$

It should be stressed that inequality (7) contains the information of both zero-sum and nonzero-sum games. On the Equilibrium, the optimal inputs $\{u_1^*,\cdots,u_i^*,\cdots,u_N^*\}$ are the solutions to minimize the cost functions $V_i^*(x)$, the worst interference is the solution to maximize the cost functions $V_i^*(x)$. After Nash equilibrium $\{u_1^*,\cdots,u_i^*,\cdots,u_N^*,\omega^*\}$ is obtained, other interference ω will lead to a better cost function performance than ω^*. To obtain the optimal policy sequence $\{u_1^*,\cdots,u_N^*,\omega^*\}$ of the H∞ indices, the differential equation associated with each cost function of the inputs-interference nonlinear system is defined as

$$\begin{aligned} H_i\left(x,\nabla V_i,u_1,\cdots,u_N,\omega_i^a\right) &= r_i\left(x,u_1,\cdots,u_N,\omega_i^a\right) \\ &+ \left(\nabla V_i\right)^T \left(f(x) + \sum_{j=1}^N g_j(x) u_j + k(x)\omega_i^a \right),\ i \in \mathbb{N} \end{aligned} \quad (8)$$

where $\nabla V_i = \partial V_i / \partial x$, and ω_i^a is the interference policy with respect to the *i*th input, which is obtained by the decoupled equation (3). Thus, the Hamiltonian-Jacobi-Isaacs (HJI) function of each input is given by

$$\begin{aligned} 0 = H_i\left(x,\nabla V_i,u_1^*,\cdots,u_N^*,\omega_i^{a*}\right) &= r_i\left(x,u_1^*,\cdots,u_N^*,\omega_i^{a*}\right) \\ &+ \left(\nabla V_i^*\right)^T \left(f(x) + \sum_{j=1}^N g_j(x) u_j^* + k(x)\omega_i^{a*} \right),\ i \in \mathbb{N} . \end{aligned} \quad (9)$$

It should be noted that the term $\sum_{j=1}^N g_j(x) u_j^*$ contains the *i*th input dynamic $g_i(x)u_i^*$ such that *i*th optimal policy u_i^* and ω_i^{a*} can be calculated with the HJI equation (9). From (9), it can be obtained that

$$\frac{\partial H_i}{\partial u_i^*} = 0 \Rightarrow u_i^* = -\frac{1}{2} R_{ii}^{-1} g_i^T(x) \nabla V_i^* ,\ i \in \mathbb{N} \quad (10)$$

$$\frac{\partial H_i}{\partial \omega_i^{a*}} = 0 \Rightarrow \omega_i^{a*} = \frac{1}{2\gamma^2} k^T(x) \nabla V_i^* ,\ i \in \mathbb{N} . \quad (11)$$

It is denoted that

$$\omega_i^* = \frac{\left\| g_i(x) \right\|}{\sum_{j=1}^N \left\| g_j(x) \right\|} \omega_i^{a*}(\tau) = \frac{1}{2\gamma^2} \frac{\left\| g_i(x) \right\|}{\sum_{j=1}^N \left\| g_j(x) \right\|} k^T(x) \nabla V_i^* . \quad (12)$$

From (3), it can be easily obtained that $\omega^*(\tau) = \omega_1^*(\tau) + \ldots + \omega_i^*(\tau) + \ldots + \omega_N^*(\tau)$ is the equilibrium worst interference policy of the cost function with H∞ property. We introduce a weighted method to obtain the worst disturbance of the inputs-interference nonlinear system such that the multiple H∞ policies get the saddle point.

Remark 1: Although multi-input optimal policies have been addressed in some previous work, the H∞ controls with multiple performance indices are not considered. Because of the coupling relationship, it is very difficult to use multiple inputs to resist the interference such that the multi-input system with disturbance can be stabilized in an optimal manner. This paper proposes the H∞ property for the multi-input nonlinear system and solves its optimal problem such that each input can resist the interference averagely. We use the RL-based NN approximation to approximate the H∞ value functions associated with each input. Each optimal policy will minimize the value function itself and all the *N* optimal inputs reach an equilibrium point. Simultaneously, we obtain the worst interference policy related to each policy by using a weighting method, such that the coupled inputs can result in a Nash equilibrium, and satisfy the global optimization.

Remark 2: Before the inputs obtain the optimal control strategies, the interference is regarded as an input such that it can be calculated to be the worst one with the proposed method in this paper. After that, the parameters of the optimal strategies

are adjusted, the worst policy of interference is obtained, and the Nash equilibrium of the performance indices is reached. When the system is subjected to other interference after that, it will get better system performance and value functions.

III. Adaptive Estimation Network

This section adopts the adaptive NNs to identify the unknown nonlinear multi-input system with the interference, where the system dynamic $f(x)$, input dynamics $g_j(x)$ and $k(x)$ are approximated separately. We defined the identifier networks on a compact set Ω as follows.

$$f(x) = \theta\xi(x) + \varepsilon_f \tag{13}$$

$$g_j(x) = \psi_j \varsigma_j(x) + \varepsilon_{gj} \tag{14}$$

$$k(x) = \hbar\zeta(x) + \varepsilon_k \tag{15}$$

where $\theta \in \mathbb{R}^{n\times k_\theta}$, $\psi_j \in \mathbb{R}^{n\times k_{\psi j}}$ and $\hbar \in \mathbb{R}^{n\times k_h}$ are the ideal weights. $\xi(x) \in \mathbb{R}^{k_\theta}$, $\varsigma_j(x) \in \mathbb{R}^{k_{\psi j}\times m}$ and $\zeta(x) \in R^{k_h\times q}$ are the activation functions. ε_f, ε_{gj} and ε_k are the NN errors. According to the NN property in [30], [31], it can be known that the approximation errors ε_f, ε_{gj} and ε_k will vanish as the NN neurons k_θ, $k_{\psi j}, k_h$ increase.

The nonlinear multi-input model (1) with interference is reconstructed as

$$\dot{x} = \theta\xi(x) + \sum_{j=1}^{N}\psi_j\varsigma_j(x)u_j + \hbar\zeta(x)\omega + \varepsilon_f + \sum_{j=1}^{N}\varepsilon_{g_j}u_j + \varepsilon_k\omega. \tag{16}$$

It can be written as a compact form.

$$\dot{x} = W^T\phi(x,u_1,\ldots,u_N,\omega) + \varepsilon_T \tag{17}$$

where $W = [\theta,\psi_1,\ldots,\psi_N,\hbar]^T \in \mathbb{R}^{b\times n}$, $\phi(x,u_1,\ldots,u_N,\omega) = [\xi^T(x),\ u_1^T\varsigma_1^T(x),\ldots,u_N^T\varsigma_N^T(x)\ ,\omega^T\zeta^T(x)]^T \in \mathbb{R}^{b\times n}$ with $b = k_\theta + k_{\psi 1} + \ldots + k_{\psi N} + k_\omega$, and $\varepsilon_T = \varepsilon_f + \varepsilon_{g1}u_1 + \ldots + \varepsilon_{gN}u_N + \varepsilon_k\omega$ is augmented approximation error.

Based on (17), one can claim that the nonlinear multi-input model (1) can be accurately approximated provided that the precise NN weights W can be yielded. Therefore, a novel estimation law suggested in [32]-[34] will be employed here to update the NN weight parameter W in equation (17). To achieve this purpose, a low-order low-pass filer $(\bullet)_f = 1/(ks+1)(\bullet)$ is introduced to obtain the smooth signal of the multi-input model. Thus, we define the following filtered variables as in [35].

$$\begin{cases} k\dot{x}_f + x_f = x \\ k\dot{\phi}_f + \phi_f = \phi \end{cases} \tag{18}$$

where $k > 0$ is a constant of the adopted low-pass filter. From (17) and (18), we can obtain (ignoring the exponentially vanishing term stemming from the non-zero initial condition) that

$$W^T\phi_f + \varepsilon_{Tf} = \frac{x - x_f}{k} \tag{19}$$

where ε_{Tf} is the filtered form with the low-pass filter $k\dot{\varepsilon}_{Tf} + \varepsilon_{Tf} = \varepsilon_T$, which is a bounded variable. Then, to design the adaptive law, it can be obtained that

$$\phi_f\phi_f^T W + \phi_f\varepsilon_{Tf}^T = \phi_f[\frac{x - x_f}{k}]^T \tag{20}$$

We define two auxiliary filtered matrices $M \in \mathbb{R}^{d\times d}$ and $N \in \mathbb{R}^{d\times n}$ as

$$\begin{cases} \dot{M} = -\kappa M + \phi_f\phi_f^T, & M(0) = 0 \\ \dot{N} = -\kappa N + \phi_f[\frac{x - x_f}{k}]^T, & N(0) = 0 \end{cases} \tag{21}$$

where $\kappa > 0$ is a filter factor. M and N are the filtered variables of $\phi_f\phi_f^T$ and $\phi_f[\frac{x - x_f}{k}]^T$ by using another filter $(\bullet)_f = 1/(s+\kappa)(\bullet)$.

Then, one can derive the solution of (21) as

$$\begin{cases} M(t) = \int_0^t e^{-\kappa(t-r)}\phi_f(r)\phi_f^T(r)dr \\ N(t) = \int_0^t e^{--\kappa(t-r)}\phi_f(r)\left[\dfrac{x(r) - x_f(r)}{k}\right]^T dr \end{cases}. \tag{22}$$

The terms κM, κN are involved in (21) can be taken as the forgetting factors; these forgetting factors can guarantee the boundedness of M and N given in (22) as presented in [33]. From (20) and (21), we design another auxiliary weight error matrix as

$$\Xi = M\hat{W} - N \tag{23}$$

where $\hat{W}$ is the estimation NN weights. It can be derived that

$$N = MW^T - \upsilon \tag{24}$$

with $\upsilon(t) = -\int_0^t e^{-\kappa(t-r)}\phi_f(r)\varepsilon_{Tf}^T(r)dr$, which satisfies $\|\upsilon\| \le \varepsilon_\upsilon$ with $\varepsilon_\upsilon > 0$. It should be noted that ε_{Tf} will vanish when the neural nodes converge to infinity, i.e. $\upsilon \to 0$ for $k_\theta, k_{\psi j}, k_\omega \to \infty$.

It can be known Ξ contains the information of W as in [36]. Thus, we design the adaptive law as

$$\dot{\hat{W}} = -\Gamma\Xi \tag{25}$$

where $\Gamma > 0$ is the learning scalar. $\hat{W}$ is the estimation NN weight and $\tilde{W} = W - \hat{W}$ is the NN estimation error. For tuning the parameters in the adaptive law, we can increase the learning gain Γ to enhance the convergence rate of adaptive law, while too large gains may make the system output oscillating. Moreover, the forgetting factors κM, κN are adopted to retain the boundedness of M, N, and thus κ is generally set as a small positive constant. Moreover, as proved in [23], [32], the NN weights in the identifier converge to a compact set around zero.

Thus, the adaptive approximation of the unknown system (1) can be further presented as

$$\dot{x} = \hat{\theta}\xi(x) + \sum_{j=1}^{N}\hat{\psi}_j\varsigma_j(x)u_j + \hat{\hbar}\zeta(x)\omega + \varepsilon_f + \sum_{j=1}^{N}\varepsilon_{g_j}u_j + \varepsilon_k\omega + \varepsilon_N \tag{26}$$

where $\hat{\theta}$, $\hat{\phi}$ and $\hat{\hbar}$ are the approximations of θ, ϕ and $\hbar$, respectively. These estimations can be obtained by the estimated NN weight $\hat{W}$. $\varepsilon_N = \tilde{W}\phi$ is the approximation NN error. Thus, the approximation system can be represented as

$$\dot{\hat{x}} = \hat{f}(x) + \sum_{j=1}^{N} \hat{g}_j(x) u_j + \hat{k}(x)\omega \tag{27}$$

where $\hat{x}$, $\hat{f}(x)$, $\hat{g}_j(x)$ and $\hat{k}(x)$ are the approximated dynamics.

Remark 3: In the practical engineering, the accurate model is usually difficult to obtain. Based on the measured input-input data, this paper uses NNs to identify the unknown multi-input system with disturbance. It should be noted that the interference is regarded as an input at the beginning adjustment of the multi-H∞ controls. In case that the disturbance is a load or other certain system, the initial signal should be added and its dynamic $k(x)$ should be identified to obtain the worst one. In the other case that the disturbance is uncertain, the disturbance dynamic is equal to one, i.e. $k(x) = 1$ in multi-input system (1). Thus, it is unnecessary to identify the $k(x)$, and the identified system (27) can be presented as $\dot{\hat{x}} = \hat{f}(x) + \sum_{j=1}^{N} \hat{g}_j(x) u_j + \omega$.

Remark 4: It should be noted that the disturbance cannot be identified separately in the studied system. In this multi-input system with interference, the disturbance is regarded as an input that needs to be calculated with the ADP scheme. Once the multi-H∞ controls $\{u_1^*, \cdots, u_i^*, \cdots, u_N^*, \omega^*\}$ are obtained and the learning gains are tuned, ω^* is the obtained worst disturbance for the performance index $V_i(x)$. When the system surfers from other disturbance, the performance index satisfies $\int_t^{\infty} \|z_i(x(\tau), u_1(\tau), \ldots u_N(\tau))\|^2 d\tau \le \gamma^2 \int_t^{\infty} \|\omega(\tau)\|^2 d\tau$, which means the system performance will be better than the worst disturbance ω^*.

IV. H∞ Policies Design with Game Theory

The multi-input system with interference has been identified by an adaptive NN structure. Because the H∞ optimal value function is difficult to obtain directly, this section uses a value function approximation (VFA) to solve this issue such that H∞ policies (10) and (11) can be calculated. To realize this purpose, the coupled HJI equations are constructed as

$$\begin{aligned} 0 &= \min_{u_i \in \Psi(\Omega)} \max_{\omega \in \Psi(\Omega)} [H_i(x, \nabla V_i, u_1, \cdots, u_N, \omega_i^{a^*})] \\ &= r_i(x, u_1, \cdots, u_N, \omega_i^{a^*}) \\ &\quad + (\nabla V_i^*)^T (\hat{\theta}^T \xi(x) + \sum_{j=1}^{N} \hat{\psi}_j^T \varsigma_j(x) u_j \\ &\quad + \hat{h}^T \zeta(x) \omega^{a^*} + \varepsilon_f + \sum_{j=1}^{N} \varepsilon_{g_j} u_j + \varepsilon_k \omega^{a^*} + \varepsilon_N). \end{aligned} \tag{28}$$

Then, the H∞ policies of multi-input system with interference can be obtained that

$$\frac{\partial H_i}{\partial u_i} = 0 \Rightarrow u_i^* = -\frac{1}{2} R_{ii}^{-1} \hat{\psi}_i \varsigma_i(x) \nabla V_i^*, \ i \in \mathbb{N} \tag{29}$$

$$\frac{\partial H_i}{\partial \omega^{a^*}} = 0 \Rightarrow \omega_i^{a^*} = \frac{1}{2\gamma^2} \hat{h}^T \zeta(x) \nabla V_i^*, \ i \in \mathbb{N}. \tag{30}$$

From (3), (11) and (12), the weighted worst policy of interference can be obtained as

$$\begin{aligned} \omega^*(t) &= \sum_{i=1}^{N} \omega_i^*(t) \\ &= \sum_{i=1}^{N} \left\{ \frac{1}{2\gamma^2} \left[\frac{\|\hat{\psi}_i \varsigma_i(x)\|}{\sum_{j=1}^{N} \|\hat{\psi}_j \varsigma_j(x)\|} \right] [\hat{h}^T \zeta(x) \nabla V_i^*] \right\}. \end{aligned} \tag{31}$$

This equation uses a weighted algorithm to calculate the worst policy of interference, such that all the policies can reach the Nash equilibrium and the multi-input system with the interference can be stabilized in a nearly-optimal manner. Note that ∇V_i^* cannot be calculated directly because of the nonlinearity and the dimensionality curse [37]-[40]. Hence, we use the NN to approximate it, which can be represented as

$$V_i^*(x) = W_{ci}^T \phi_{ci}(x) + \varepsilon_{ci} \tag{32}$$

with $W_{ci} = [W_{ci1}, \cdots, W_{ciK}] \in \mathbb{R}^K$ and $\varphi_{ci}(x) = [\varphi_{ci1}(x), \cdots, \varphi_{ciK}(x)] \in \mathbb{R}^K$. K is the neuron number, and ε_{ci} denotes the approximated error. In this paper, we use NNs to approximate the optimal value function. Hence, the used ADP scheme belongs to HDP algorithm. Its time derivative is obtained as

$$\nabla V_{ci}^* = \nabla \phi_{ci}^T W_{ci} + \nabla \varepsilon_{ci} \tag{33}$$

where $\nabla \phi_{ci} = \partial \phi_{ci} / \partial x$ and $\nabla \varepsilon_{ci} = \partial \varepsilon_{ci} / \partial x$ are the partial derivatives of ϕ_{ci} and ε_{ci} . $W_{ci} \in \mathbb{R}^l$ is ideal critic weight parameters. $\phi_{ci}(x) \in \mathbb{R}^{l \times n}$ is the activation function. ε_{ci} is the critic network approximated error, l represents the neuron number. It is denoted that $\hat{W}_{ci}$ is the estimation of W_{ci}^*, and $\tilde{W}_{ci} = W_{ci} - \hat{W}_{ci}$ is the estimation error. Then, the approximation of derivative value function is written as

$$\nabla \hat{V}_{ci} = \nabla \phi_{ci}^T \hat{W}_{ci} \tag{34}$$

Finally, the approximation H∞ policies of each index can be derived by

$$\hat{u}_i = -\frac{1}{2} R_{ii}^{-1} \hat{\psi}_i \varsigma_i(x) \nabla \phi_{ci}^T \hat{W}_{ci}, \ i \in \mathbb{N} \tag{35}$$

$$\hat{\omega}_i^a = \frac{1}{2\gamma^2} \hat{h}^T \zeta(x) \nabla \phi_{ci}^T \hat{W}_{ci}, \ i \in \mathbb{N}. \tag{36}$$

The obtained policies can ensure that the *i*th H∞ performance for a prescribed attenuation level γ .

Finally, the approximation weighted worst disturbance can be calculated from (12) and (31), which can guarantee that all the H∞ indices reach a Nash equilibrium as in Definition 2.

$$\begin{aligned} \hat{\omega}(t) &= \sum_{i=1}^{N} \omega_i(t) \\ &= \sum_{i=1}^{N} \left\{ \frac{1}{2\gamma^2} \left[\frac{\|\hat{\psi}_i \varsigma_i(x)\|}{\sum_{j=1}^{N} \|\hat{\psi}_j \varsigma_j(x)\|} \right] [\hat{h}^T \zeta(x) \nabla \phi_{ci}^T \hat{W}_{ci}] \right\}. \end{aligned} \tag{37}$$

We design an adaptive law to update the estimation $\hat{W}_{ci}$ based on the HJI equation. Then, the approximation HJI equations (28) with (33) can be represented as

$$\begin{aligned} &0 = r_i(x, u_1, \cdots, u_N, \omega_i^{a^*}) + \\ &(\nabla \phi_{ci}^T \hat{W}_{ci})^T (\hat{\theta}^T \xi(x) + \sum_{j=1}^{N} \hat{\psi}_j^T \varsigma_j(x) u_j + \hat{h}^T \zeta(x) \omega^{a^*}) + \varepsilon_{Hi} \end{aligned} \tag{38}$$

where $\varepsilon_{Hi} = (\nabla \phi_{ci}^T W_{ci})(\varepsilon_f + \sum_{j=1}^{N} \varepsilon_{g_j} u_j + \varepsilon_k \omega^{a^*}) + (\nabla \varepsilon_{ci})(\hat{\theta}^T \xi(x) + \sum_{j=1}^{N} \hat{\psi}_j^T \varsigma_j(x) u_j + \hat{h}^T \zeta(x) \omega^{a^*} + \varepsilon_f + \sum_{j=1}^{N} \varepsilon_{g_j} u_j + \varepsilon_k \omega^{a^*} + \varepsilon_N)$ is the HJI equation error. It is assumed that the HJI residual

error in (38) is bounded with $\|\varepsilon_{Hi}\| \le b_{Hi}$ for some positive constant b_{Hi} as in [15]. It should be stressed that the HJI equation (38) contains the information of approximated multi-input system with interference, such that the learning algorithm designed with this equation includes the system information and optimal theory. To facilitate the estimation algorithm, it is denoted that $\varpi_i = \nabla\phi_{ci}^T[\hat{\theta}^T\xi(x) + \sum_{j=1}^N \hat{\psi}_j^T\varsigma_j(x)u_j + \hat{\hbar}^T\zeta(x)\omega^{a^*}]$, and $\Theta_i = x^TQ_ix + \sum_{j=1}^N u_j^TR_{ij}u_j - \gamma^2\|\omega_i^{a^*}\|^2$. Then, the HJI equation is simplified by a linearization form as

$$\Theta_i + W_{ci}^T\varpi_i = -\varepsilon_{Hi}. \tag{39}$$

Note that equation (39) contains the information of the identified system and the utility function. We multiply both sides by ϖ_i, it can be obtained that

$$\varpi_i\Theta_i + \varpi_i\varpi_i^TW_{ci}^T = -\varpi_i\varepsilon_{Hi}. \tag{40}$$

Equation (40) will be used to construct the adaptive law of critic NN W_{ci}. To this end, two filtered matrices $M_{ci} \in \mathbb{R}^{l\times l}$ and $N_{ci} \in \mathbb{R}^l$ are given as

$$\begin{cases} \dot{M}_{ci} = -\kappa_{ci}M_{ci} + \varpi_i\varpi_i^T,\ P_{ci}(0) = 0 \\ \dot{N}_{ci} = -\kappa_{ci}N_{ci} + \varpi_i\Theta_i,\quad Q_{ci}(0) = 0 \end{cases} \tag{41}$$

where κ_{ci} is a filter factor. From (40), we define another matrix as

$$\Xi_{ci} = M_{ci}W_{ci} + N_{ci}. \tag{42}$$

Noted that Ξ_{ci} is the filtered variable of $-\varpi_i\varepsilon_{Hi}$ in (40). Finally, the adaptive law is designed as

$$\dot{\hat{W}}_{ci} = -\Gamma_{ci}\Xi_{ci} \tag{43}$$

where $\Gamma_{ci} > 0$ is the learning gain. Moreover, one can obtain that

$$\Xi_{ci} = M_{ci}\hat{W}_{ci} + N_{ci} = -M_{ci}\tilde{W}_{ci} + \upsilon_{ci} \tag{44}$$

where $\tilde{W}_{ci} = W_{ci} - \hat{W}_{ci}$. We can conclude that Ξ_{ci} is obtained from vector $\tilde{W}_{ci}$ and υ_{ci}, and $\upsilon_{ci} = -\int_0^t e^{-\kappa_{ci}(t-r)}\varepsilon_{Hi}(r)\varpi^T(r)dr$ is bound with $\|\upsilon_{ci}\| \le \varepsilon_{ci}$, $\varepsilon_{ci} > 0$. The critic NN weights converge to a compact set around zero, which has been proved in the [32] and will not be presented in this paper.

Remark 5: In this paper, the NN weights are updated online synchronously. The identifier NN weights converge to their true values based on the input-output data as presented in Section III. Then, the critic NN weights are updated based on the HJ equations (28), which depend on the identified dynamics. Thus, the critic NN weights will converge to the true values after the identifier NN weights achieve convergence, as shown in the simulations. Moreover, all the critic NN weights are synchronously updated such that all the performances can come to a Nash equilibrium $\{V_1^*, \cdots, V_N^*\}$, which is known as the static game.

Remark 6: The general ADP method uses the actor-critic structure, where two NNs are used to derive the optimal control of nonlinear systems as in [27]. Moreover, the gradient algorithm is used to minimize the HJB equation in general ADP schemes to update the NN weights. Different to the general ADP methods, a newly developed adaptation algorithm driven by the NN weight errors are used to update the NN weights in the proposed ADP structure, such that the actor can be avoided and only a single-critic NN is used in this paper.

V. Stability Analysis

From the above estimation results and the obtained H∞ policies, this section will analyze the stability of the inputs-interference model. Substituting the H∞ policies (35) and (37) into the system (1), it can be obtained that

$$\begin{aligned} \dot{x} &= f(x) + \sum_{j=1}^N \hat{g}_j(x)u_j + k(x)\omega \\ &= f(x) + \sum_{j=1}^N \hat{g}_j(x)\{-\frac{1}{2}R_{ii}^{-1}[g_j(x)]^T\nabla\phi_{ci}^TW_{ci}\} \\ &\quad + k(x)\{\frac{1}{2\gamma^2}[\hat{k}(x)]^T\sum_{j=1}^N[(\frac{\|\psi_i\varsigma_i(x)\|}{\sum_{j=1}^N\|\psi_j\varsigma_j(x)\|})(\nabla\phi_{ci}^TW_{ci})]\} \end{aligned} \tag{45}$$

where $\hat{g}_j(x) = \hat{\psi}_j\varsigma_j$ and $\hat{k}(x) = \hat{\hbar}\zeta(x)$ are obtained from (25).

Lemma 1 [33]: The matrices M and M_{ci} are positive definite, i.e. $\lambda_{\min}(M) > \sigma > 0$, and $\lambda_{\min}(M_{ci}) > \sigma_{ci} > 0$ for positive constants σ and σ_{ci} in case the activation vectors ϕ and ϕ_{ci} are persistently excited (PE).

Assumption 1 [41]: The NN weights W and W_{ci}, activation function ϕ, ϕ_{ci} and the derivative $\nabla\phi_{ci}$ are bounded, i.e. $\|W\| \le W_d$, $\|\phi\| \le \phi_d$, $\|W_{ci}\| \le W_{Ni}$, $\|\phi_{ci}\| \le \phi_{Ni}$, $\|\nabla\phi_{ci}\| \le \phi_{Mi}$; the NN error ε_v and $\nabla\varepsilon_v$ are bounded, i.e. $\|\varepsilon_v\| \le b_\varepsilon$, $\|\nabla\varepsilon_v\| \le b_{\varepsilon v}$.

Assumption 2 [42], [43]: The multi-input system dynamics satisfy the conditions that $\|f(x)\| \le b_f\|x\|$, $\|g_j(x)\| \le b_{gj}$, $\|k(x)\| \le b_k$ for positive constants $b_f > 0$, $b_{gj} > 0$, $b_k > 0$.

To show the conclusions of the proposed optimal solutions for the multi-input system with interference, the following Theorem is presented to show the convergence of the identifier and the derived optimal control actions, and control system stability.

Theorem 1: For the multi-input system (1) with the interference, by using the H∞ policies (35), (36) and (37), the adaptive laws (25) and (43), if the regressor vectors ϕ and ϕ_{ci} are PE, then

1) The NN weight errors $\tilde{W}$ and $\tilde{W}_{ci}$ are uniformly ultimately bounded (UUB);
2) H∞ policies go closely to the truth values u_i^* and ω^*;
3) The closed-loop inputs-interference model is stable.

Proof: Consider the Lyapunov function as

$$\begin{aligned} L &= L_1 + L_2 + L_3 + L_4 + L_5 \\ &= \frac{1}{2}tr\left(\tilde{W}^T\Gamma^{-1}\tilde{W}\right) + \frac{1}{2}\sum_{i=1}^N\tilde{W}_{ci}^T\Gamma_{ci}^{-1}\tilde{W}_{ci} + x^Tx \\ &\quad + \sum_{i=1}^N K_iV_i^* + \Upsilon\upsilon^T\upsilon + \sum_{i=1}^N\Upsilon_{ci}\upsilon_{ci}^T\upsilon_{ci} \end{aligned} \tag{46}$$

where $L_1 = \frac{1}{2}tr\left(\tilde{W}^T\Gamma^{-1}\tilde{W}\right)$, $L_2 = \sum_{i=1}^N L_{ci} = \sum_{i=1}^N\frac{(\tilde{W}_{ci}^T\Gamma_{ci}^{-1}\tilde{W}_{ci})}{2}$, $i \in \mathbb{N}$, $L_3 = x^Tx + \sum_{i=1}^N K_iV_i^*$, $L_4 = \Upsilon\upsilon^T\upsilon$, $L_5 = \sum_{i=1}^N\Upsilon_{ci}\upsilon_{ci}^T\upsilon_{ci}$

with positive constants $K_i > 0$, $\Gamma_{ci} > 0$, $\Upsilon > 0$, $\Upsilon_{ci} > 0$.

From (23) and (24), it can be obtained that $\Xi = M\hat{W} - MW + \upsilon = -M\tilde{W} + \upsilon$. Using Young's inequality $ab \le a^2\eta/2 + b^2/2\eta$ with $\eta > 0$, from (25) we obtain

$$\dot{L}_1 = -tr\left(\tilde{W}^T M\tilde{W}\right) + tr\left(\tilde{W}^T\upsilon\right) \le -\sigma\left\|\tilde{W}\right\|^2 + \left\|\tilde{W}^T\upsilon\right\| \le -(\sigma - \frac{1}{2\eta})\left\|\tilde{W}\right\|^2 + \frac{\eta\|\upsilon\|^2}{2}. \quad (47)$$

Furthermore, it can be concluded that

$$\begin{aligned}\dot{L}_2 &= \sum_{i=1}^N \dot{L}_{ci} = \sum_{i=1}^N \tilde{W}_{ci}^T\Gamma_{ci}^{-1}\dot{\tilde{W}}_{ci} \\ &= \sum_{i=1}^N (-\tilde{W}_{ci}^T M_{ci}\tilde{W}_{ci} + \tilde{W}_{ci}^T\upsilon_{ci}) \\ &= \sum_{i=1}^N (-\sigma_{ci}\left\|\tilde{W}_{ci}\right\|^2 + \tilde{W}_{ci}^T\upsilon_{ci}) \\ &\le -\sum_{i=1}^N (\sigma_{2i} - \frac{1}{2\eta_i})\left\|\tilde{W}_{ci}\right\|^2 + \sum_{i=1}^N \frac{\eta_i\|\upsilon_{ci}\|^2}{2}, i \in \mathbb{N}\end{aligned} \quad (48)$$

where σ and σ_{2i} are the positive constants. The inequalities (47) and (48) imply that both $\tilde{W}$ and $\tilde{W}_{ci}$ are UUB with appropriate η and η_i. Moreover, consider the inequality $\pm ab \le a^2\eta/2 + b^2/2\eta$, from (6) and (45) one may write $\dot{L}_3$ as

$$\begin{aligned}\dot{L}_3 &= 2x^T\dot{x} - K_i(x^TQ_ix + \sum_{j=1}^N u_j^TR_{ij}u_j - \gamma^2\left\|\omega_i^a\right\|^2) \\ &= 2x^T\{f(x) + \sum_{j=1}^N g_j(x)u_j + k(x)\{\sum_{j=1}^N [(\frac{\|g_i(x)\|}{\sum_{j=1}^N\|g_j(x)\|})\omega_i^a]\}\} \\ &\quad - K_i(x^TQ_ix + \sum_{j=1}^N u_j^TR_{ij}u_j - \gamma^2\left\|\omega_i^a\right\|^2) \\ &\le -\{K_i\lambda_m(Q_i) - 2b_f + \sum_{j=1}^N b_{gj} + b_k\sum_{j=1}^N(\frac{b_{gi}}{\sum_{j=1}^N b_{gj}})\}\|x\|^2 \\ &\quad -\sum_{j=1}^N[\lambda_m(R_{ij}) - b_{gj}]\left\|u_j\right\|^2 + b_k\sum_{j=1}^N[(\frac{b_{gi}}{\sum_{j=1}^N b_{gj}}) + \gamma^2]\left\|\omega_i^a\right\|^2.\end{aligned} \quad (49)$$

From (24), we obtain $\dot{\upsilon} = -\kappa\upsilon + \phi_f\varepsilon_{Tf}^T$. Then, it has

$$\dot{L}_4 = 2\Upsilon\upsilon^T\dot{\upsilon} = 2\Upsilon\upsilon^T\left(-\kappa\upsilon + \phi_f\varepsilon_{Tf}^T\right) \le -(2\Upsilon\kappa - \eta)\|\upsilon\|^2 + \frac{1}{\eta}\left\|\Upsilon\phi_f\varepsilon_{Tf}^T\right\|^2. \quad (50)$$

Moreover, we conclude from (44) that $\dot{\upsilon}_{ci} = -\kappa_{ci}\upsilon_{ci} + \varpi\varepsilon_{Hi}^T$, so that

$$\dot{L}_5 = 2\Upsilon_{ci}\upsilon_{ci}^T\dot{\upsilon}_{ci} = 2\Upsilon\upsilon^T\left(-\kappa_{ci}\upsilon_{ci} + \varpi\varepsilon_{Hi}^T\right) \le -(2\Upsilon_{ci}\kappa_{ci} - \eta)\|\upsilon_{ci}\|^2 + \frac{1}{\eta}\left\|\Upsilon_{ci}\varpi\varepsilon_{Hi}^T\right\|^2 \quad (51)$$

Finally, it is able to obtain that

$$\dot{L} = \dot{L}_1 + \dot{L}_2 + \dot{L}_3 + \dot{L}_4 + \dot{L}_5 \le -a_1\left\|\tilde{W}\right\|^2 - \sum_{i=1}^N a_{2i}\left\|\tilde{W}_{ci}\right\|^2 - a_3\|x\|^2 - a_4\|\upsilon\|^2 - a_{5i}\|\upsilon_{ci}\|^2 - \sum_{j=1}^N[a_{6j}]\left\|u_j\right\|^2 + \partial \quad (52)$$

where $a_1 = \sigma - \frac{1}{2\eta}$, $a_{2i} = \sigma_{2i} - \frac{1}{2\eta_i}$,

$$a_3 = \{K_i\lambda_m(Q_i) - 2b_f + \sum_{j=1}^N b_{gj} + b_k\sum_{j=1}^N(\frac{b_{gi}}{\sum_{j=1}^N b_{gj}})\},$$

$a_4 = 2\Upsilon\kappa - \eta$, $a_{5i} = 2\Upsilon_{ci}\kappa_{ci} - \eta$, $a_{6j} = \lambda_m(R_{ij}) - b_{gj}$,

$$\partial = b_k\sum_{j=1}^N[(\frac{b_{gj}}{\sum_{j=1}^N b_{gj}}) + \gamma^2]\left\|\omega_i^a\right\|^2 + \frac{\eta\|\upsilon\|^2}{2} + \sum_{i=1}^N\frac{\eta_i\|\upsilon_{ci}\|^2}{2} + \frac{1}{\eta}\left\|\Upsilon\phi_f\varepsilon_{Tf}^T\right\|^2 + \frac{1}{\eta}\left\|\Upsilon_{ci}\varpi\varepsilon_{Hi}^T\right\|^2.$$

From the above equation, the upper bound variable ∂ contains the variables ω_i^a, $\upsilon(t)$, υ_{ci}, $\phi_f\varepsilon_{Tf}$ and $\varpi\varepsilon_{Hi}$, it is easily known that when the NN neurons $k_\theta, k_{\psi j}, k_h, k$ increase, the NN approximation errors ε_{Tf}, ε_f, ε_{g_j}, ε_k and $\nabla\varepsilon_{ci}$ will converge to zero, such that the related terms ε_{Hi}, $\upsilon(t)$, υ_{ci}, $\phi_f\varepsilon_{Tf}$ and $\phi_f\varepsilon_{Tf}$ will converge to zero. Thus, as the neurons increase, the bound will converge to $\partial_0 = b_k\sum_{j=1}^N[(\frac{b_{gj}}{\sum_{j=1}^N b_{gj}}) + \gamma^2]\left\|\omega_i^a\right\|^2$, which is a term regarding to the ideal worst-case interference ω_i^a. When the parameters K, Υ, Υ_{ci}, η, η_i satisfy

$$K > \left[2b_f - \sum_{j=1}^N b_{gj} - b_k\sum_{j=1}^N(\frac{b_{gj}}{\sum_{j=1}^N b_{gj}})\right]/\lambda_{\min}(Q), \quad \eta > \frac{1}{2\sigma},$$

$$\eta_i > \frac{1}{2\sigma_i}, \quad \Upsilon > \frac{\eta}{2\kappa}, \quad \Upsilon_{ci} > \frac{\eta}{2\kappa_{ci}},$$

and the NN weights $\tilde{W}$, $\tilde{W}_{ci}$, the system state x locate the outside of the following compact sets

$$\left\{\tilde{W} \mid \left\|\tilde{W}\right\| \le \partial/a_1\right\}, \left\{\tilde{W}_{ci} \mid \left\|\tilde{W}_{ci}\right\| \le \partial/a_{2i}\right\}, \{x \mid \|x\| \le \partial/a_3\}, \{\upsilon \mid \|\upsilon\| \le \partial/a_4\}, \{\upsilon_{ci} \mid \|\upsilon_{ci}\| \le \partial/a_{5i}\} \quad (53)$$

then $\dot{L} < 0$ holds. Thus, from Lyapunov theory as in [44], it can be known that the NN weights $\tilde{W}$, $\tilde{W}_{ci}$, the system state x, the residual errors υ and υ_{ci} are UUB.

Moreover, H∞ policy errors will be analyzed. The following equation can be obtained as

$$\begin{aligned}u_i - u_i^* &= -\frac{1}{2}R_{ii}^{-1}[\hat{\psi}_i\varsigma_i(x)]^T\nabla\phi_{ci}^T\hat{W}_{ci} \\ &\quad + \frac{1}{2}R_{ii}^{-1}[g_i(x)]^T\left(\nabla\phi_{ci}^TW_{ci} + \nabla\varepsilon_{ci}\right) \\ &= \frac{1}{2}R_{ii}^{-1}\left([\hat{\psi}_i\varsigma_i(x)]^T\nabla\phi_{ci}^T\tilde{W}_{ci}\right) \\ &\quad + \frac{1}{2}R_{ii}^{-1}[g_i(x) - \hat{\psi}_i\varsigma_i(x)]^T\nabla\phi_{ci}^T\hat{W}_{ci} \\ &\quad + \frac{1}{2}R_{ii}^{-1}[g_i(x) - \hat{\psi}_i\varsigma_i(x)]^T\nabla\phi_{ci}^T\tilde{W}_{ci} \\ &\quad + \frac{1}{2}R^{-1}[g_i(x)]^T\nabla\varepsilon_{ci}\end{aligned} \quad (54)$$

$$\omega-\omega^*=\frac{1}{2\gamma^2}[\hat{h}^T\zeta(x)]^T\sum_{j=1}^{N}[(\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|})\nabla\phi_{ci}^T\hat{W}_{ci}]$$
$$-\frac{1}{2\gamma^2}[k(x)]^T\sum_{j=1}^{N}[(\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|})(\nabla\phi_{ci}^TW_{ci}+\nabla\varepsilon_{ci})]$$
$$=-\frac{1}{2\gamma^2}\left([\hat{h}^T\zeta(x)]^T\sum_{j=1}^{N}[(\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|})\nabla\phi_{ci}^T\tilde{W}_{ci}]\right)$$
$$+\frac{1}{2\gamma^2}[k(x)-\hat{h}^T\zeta(x)]^T\sum_{j=1}^{N}[(\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|})(\nabla\phi_{ci}^T\hat{W}_{ci})]$$
$$+\frac{1}{2\gamma^2}[k(x)-\hat{h}^T\zeta(x)]^T\sum_{j=1}^{N}[(\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|})(\nabla\phi_{ci}^T\tilde{W}_{ci})]$$
$$-\frac{1}{2\gamma^2}[k(x)]^T\sum_{j=1}^{N}[(\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|})(\nabla\varepsilon_{ci})]. \quad (55)$$

The previous proof shows that the approximated system dynamics are bounded, the estimation errors $\tilde{\psi}_i$ and $\tilde{h}$ are bounded. From equation (54) and (55), one can indicate that

$$\lim_{t\to+\infty}\|\hat{u}_i-u_i^*\|\le\frac{1}{2}\lambda_{\max}(R_{ii}^{-1})[b_{gi}(\phi_{Ni}\|\tilde{W}_{ci}\|+b_{\varepsilon\upsilon})$$
$$+\phi_{Ni}\|\tilde{\psi}_i\varsigma_i(x)\|\|\hat{W}_{Ni}\|]\le\varepsilon_{ui} \quad (56)$$

$$\lim_{t\to+\infty}\|\omega-\omega^*\|\le\frac{1}{2\gamma^2}\{b_k\sum_{j=1}^{N}[\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|}(\phi_{Ni}\|\tilde{W}_{ci}\|+b_{\varepsilon\upsilon})]$$
$$+\phi_{Ni}\sum_{j=1}^{N}[\frac{\|g_i(x)\|}{\sum_{j=1}^{N}\|g_j(x)\|}]\|\hat{W}_{ci}\|\|\tilde{h}^T\zeta(x)\|\}\le\varepsilon_\omega \quad (57)$$

with $\varepsilon_{ui}>0$, $\varepsilon_\omega>0$ from the identifier NN error and value function approximation structure. That completes the proof. It is noted that similar proofs can be found in [45], but where the identifier NN errors are not considered in the control convergence analysis. This paper considers both identifier errors and control design errors in the proof.

VI. Simulation Results

This section validates the proposed H∞ policies for the multi-input system with interference via two numerical examples, where no prior knowledge of the system is known and only the input-output data is needed. The first linear system is adopted to evaluate the convergence of the proposed learning algorithm, while the second nonlinear example is dedicated to showcasing the applicability of the ADP-based H∞ policies for the proposed inputs-interference model.

A. Liner example of two-input with interference

Consider the following two-input linear affine system with an interference

$$\dot{x}=-\frac{3}{4}x+u_1+2u_2-\omega. \quad (58)$$

The H∞ performance indices for the two-input system are defined as

$$J_1=\int_0^\infty\left(2x^2+2u_1^2+2u_2^2-2*4^2\omega_1^{a2}\right)dt,$$

$$J_2=\int_0^\infty\left(x^2+u_1^2+u_2^2-4^2\omega_2^{a2}\right)dt. \quad (59)$$

These value functions show that two inputs work together to resist the interference, and every policy minimizes the value function itself. When the system information (58) is not known, the adaptive NN identifier (17) and adaptive law (25) are used to approximate the unknown system model. The bounded activation function is $\phi(x,u_1,u_2,\omega)=[\varphi_f,u_1^T\varphi_{g1},u_2^T\varphi_{g2},\omega^T\varphi_k]^T$ $=[x,u_1,u_2,\omega]^T$. The initial values are set as $x(0)=1$, $W(0)=[0.5\ 0.5\ 0.5\ 0.5]^T$, $W_{c1}(0)=[0\ \ 0\ \ 0]^T$, $W_{c2}(0)=[0.5\ \ 0\ \ 0]^T$. The learning gains in (25) are given by $k=0.001$, $\kappa=1$, $\Gamma=450I$, and we set $\phi_{c1}(x)=\phi_{c2}(x)=[x^2\quad x^4\quad x^6]^T$. Fig. 1 is the weight parameters in observer, it can be shown that $W=[W_f,W_{g1},W_{g2},W_k]^T=[-3/4,1,2,-1]^T$, which go closely to the ideal values. We select the activation function as the high-order neural network function for the critic NN and preliminary information of system for the identifier.

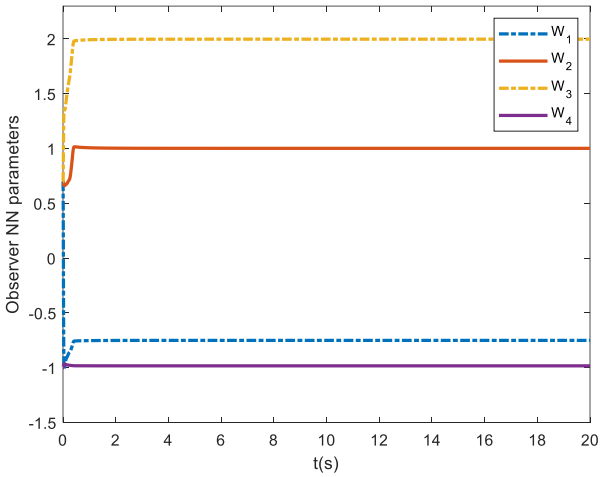


Fig. 1. Parameter convergence of NN observer.

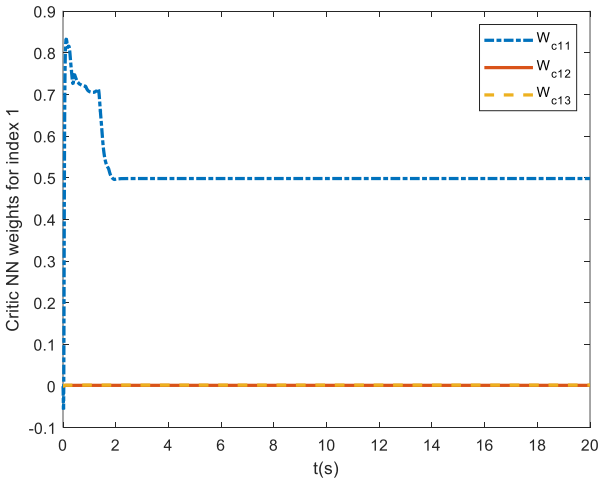


Fig. 2. Weights for index 1 with two inputs and interference.

The adaptive law gains of critic NN are given by $\kappa_{ci}=10$, $\Gamma_{ci}=300diag([0.1,1,1])$, $i=1,2$. Fig. 2 shows the NN approximation weights of input performance index 1, Fig. 3 is the convergence profile of performance index 2, both them

converge to some constant parameters, which indicates that the outcomes $\{V_1^*, V_2^*\}$ of the optimal value functions of two inputs with interference are obtained. Fig. 4 represents the system trajectory x, which is stabilized to zero in a short time. The H∞ policies and the worst policy of interference are presented in Fig. 5.

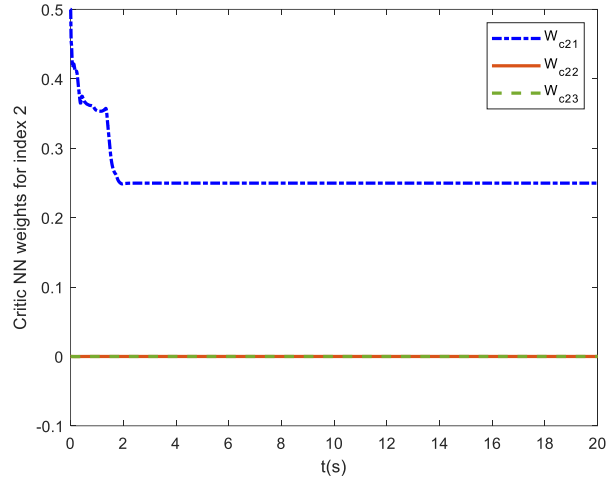


Fig. 3. Weights for index 2 with two inputs and interference.

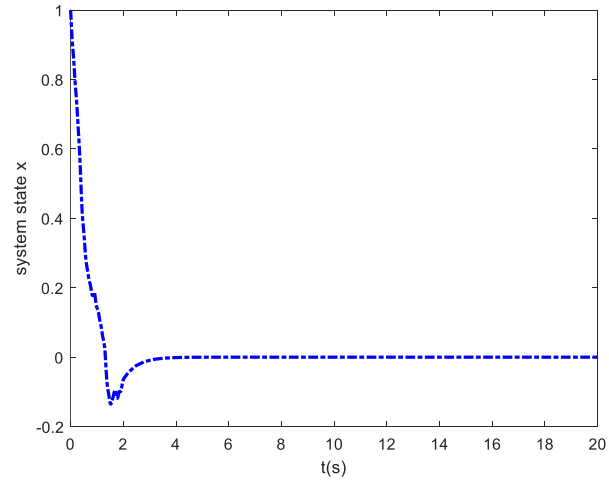


Fig. 4. System state *x*.

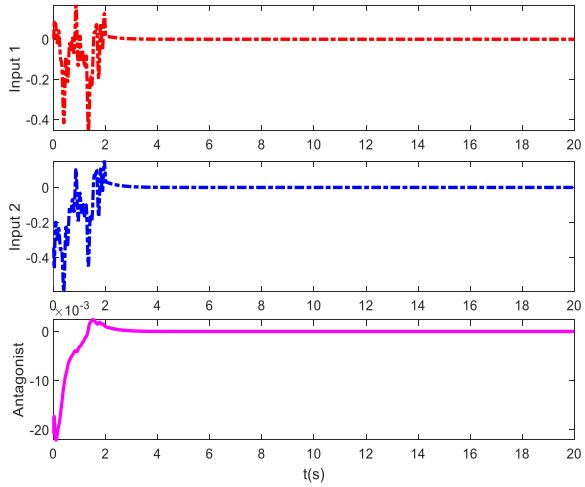


Fig. 5. Obtained policies for each input and interference.

After this simulation result is obtained, the learning gains in the control are tuned such that other interference will make better performance of cost function. In this game situation, inputs u_1 and u_2 cooperate. They cooperate to resist interference ω. These simulations show that the proposed methods can observe the unknown linear system of two-input with interference, and the H∞ indices can be obtained.

B. Nonlinear example of two-input with interference

A two-input nonlinear example with interference is presented to illustrate the effectiveness of the proposed Nash conclusions. The inputs-interference model [15] is given as

$$\dot{x} = f(x) + \sum_{i=1}^{2} g_i(x)u_i + k(x)\omega,\ x \in \mathbb{R}^2 \quad (60)$$

where

$$f(x) = \begin{bmatrix} x_2 \\ -x_2 - 0.5x_1 - 0.25x_2(\cos(2x_1)+2)^2 + 0.25x_2(\sin(4x_1)+2)^2 \end{bmatrix},$$

$$g_1(x) = \begin{bmatrix} 0 \\ \cos(2x_1)+2 \end{bmatrix},\quad g_2(x) = \begin{bmatrix} 0 \\ \sin(4x_1^2)+2 \end{bmatrix},$$

$$k(x) = \begin{bmatrix} 0 \\ \sin(4x_1)+2 \end{bmatrix}.$$

As in our previous work [46], $f(x)$, $g_1(x)$, $g_2(x)$ and $k(x)$ are unknown. Define the initial states as $x_1(0) = 1$, $x_2(0) = -1$, the RL-based NN initial values are given as $\hat{W}_{c1}(0)=[0.25,0,0.25]^T$, $\hat{W}_{c2}(0) = [0.1,0.1,0.1]^T$. The parameters of the adaptive identifier are given by $k = 0.001$, $\kappa = 2$, $\Gamma = 500$ with the regressor vector

$$\phi(x,u_1,u_2,\omega) = [\varphi_f,\ u_1^T\varphi_{g1}(x), u_2^T\varphi_{g2}, \omega^T\varphi_k]^T$$

$$= \begin{bmatrix} x_2 & 0 & 0 & 0 & 0 \\ x_2 & x_1 & x_2(\cos(2x_1)+2)^2 & x_2(\sin(4x_1)+2)^2 & u_1(\cos(2x_1)+2) \\ 0 & 0 & & & \\ u_2(\sin(4x_1^2)+2) & \omega(\sin(4x_1)+2) & & & \end{bmatrix}^T.$$

Fig.6 shows the observer NN parameter convergence of $\hat{W}_d = [W_f, W_{g1}, W_{g2}, W_k]^T$, where the updated parameters converge to their true values, and the two-input unknown nonlinear system with interference is identified well. Besides, there exists an overshoot in Fig. 6, which is produced by the high-order terms in the activation function. However, the estimated weights in the high-order network are steady and converge to their true values in a short time.

For the critic NN, we set $\kappa_{ci} = 2$, $\Gamma_{ci} = diag([1,0.1,5])$, $i = 1,2$, $\gamma_1 = \gamma_2 = 8$, $Q_1 = 2I$, $Q_2 = I$, $R_{11} = R_{12} = 2I$ and $R_{21} = R_{22} = I$ as shown in [15, 20]. Fig. 7 is the critic NN weight convergence of H∞ index 1 and Fig. 8 presents the weight of the index 2, which show that the updated NN weights are all convergent, and the H∞ indices of two-input nonlinear systems with interference are obtained. Fig. 9 presents the system state trajectories of the interference model, which converge to zero with small overshoot in a short time. The most appropriate policies of each input and interference are shown in Fig. 10, where it should be noted that inputs 1 and 2 work together to resist the interference. The optimal inputs 1 and 2

are obtained to minimize the cost function, while the worst interference is to maximize the one. After the sequence $\{u_1^*, u_2^*, \omega^*\}$ is obtained, the other ω will satisfy (2), the performance will be better.

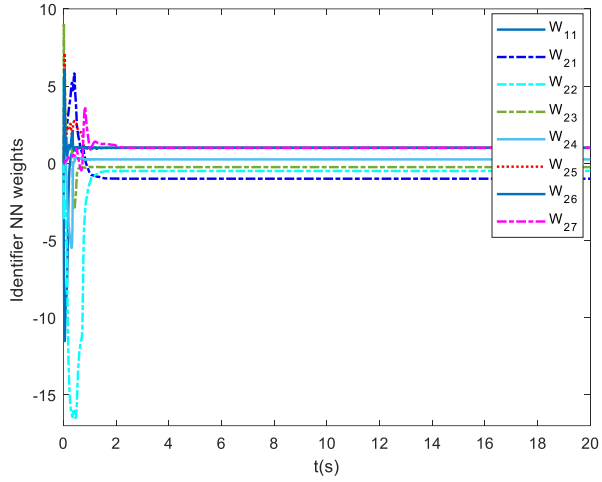


Fig. 6. Observer NN parameters.

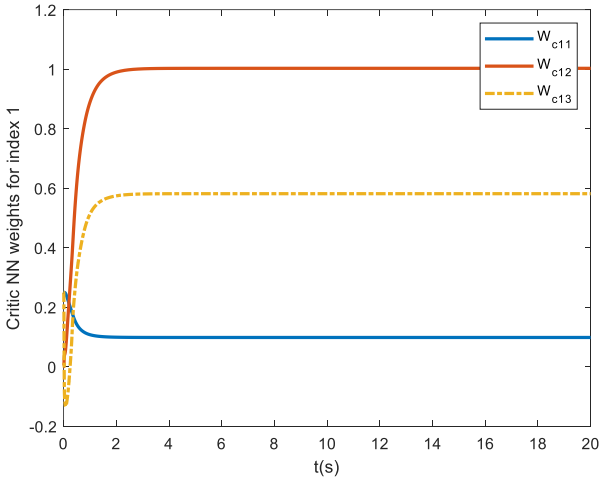


Fig. 7. Critic NN weight convergence of H∞ index 1.

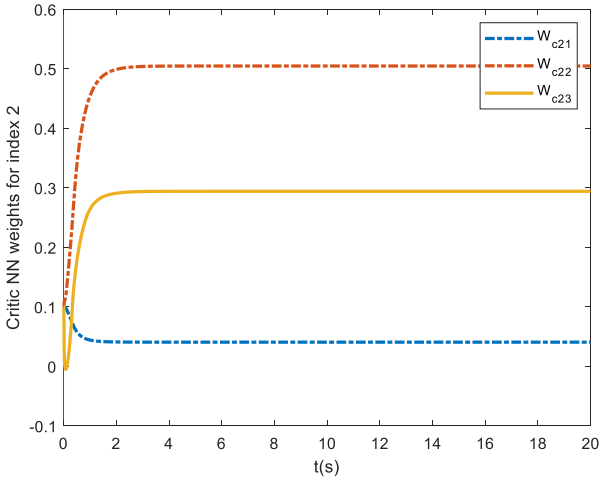


Fig. 8. Critic NN weight convergence of H∞ index 2.

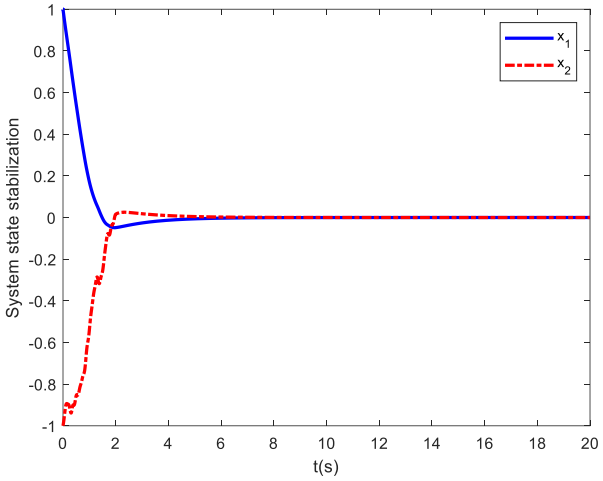


Fig. 9. System state trajectories.

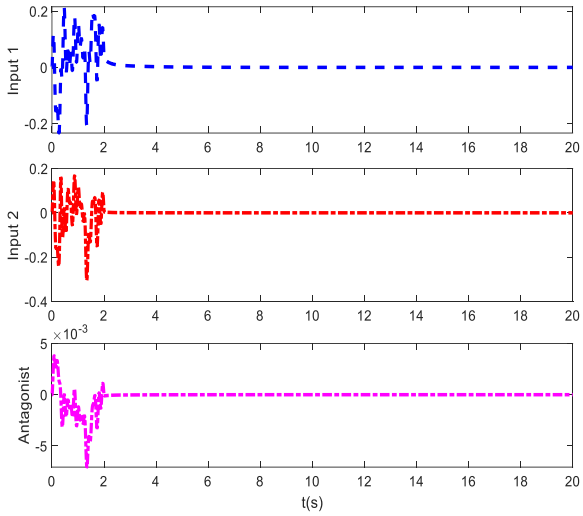


Fig. 10. Policies of each input and interference.

All performances of the simulation including the NN observer, the state trajectories, the index NN weight convergent, and the policies illustrate that the ADP scheme can provide an effective approximate optimal policy to make H∞ indices of the unknown inputs-interference nonlinear game system reach the defined Nash equilibrium. Simultaneously, the linear simulation results are given to show that the proposed H∞ controls can not only solve the nonlinear system, but also suit the linear system. The simulation results show that the proposed methods are more universal in practical engineering.

Moreover, to show the advantage of the adaptive algorithms in this paper over conventional gradient learning algorithms for the ADP synthesis, a comparison simulation is presented with the adaptive law and control design method in [45]. The results are shown as in Figs. 11-12. Fig. 11 shows the actor NN weights, which cannot converge to the true values. Fig. 12 is the system state performance. Compared with Fig. 9, it can be shown that the adaptive law used in this paper can achieve better performance with no overshoot. The comparison results show that the adaptive law in this paper can realize accurate convergence results and better state performance.

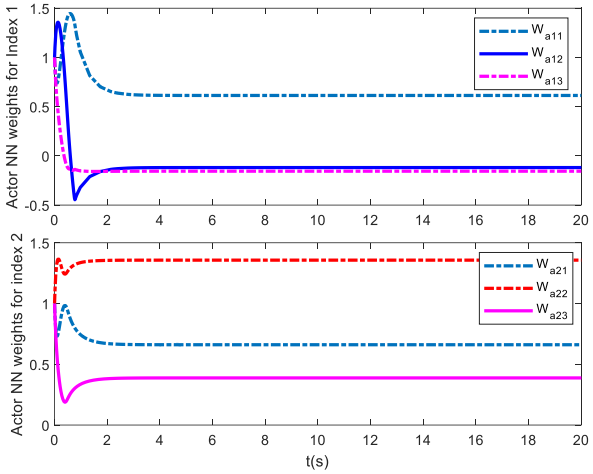


Fig. 11 Actor NN weights with adaptive law in [45].

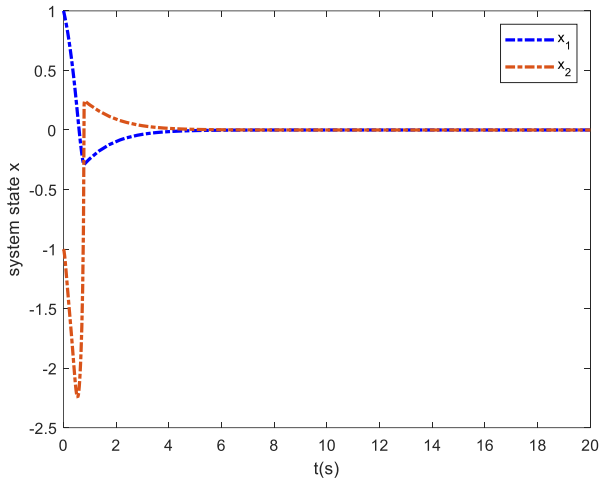


Fig. 12 System states with adaptive law in [45].

VII. Conclusion

This paper considered an inputs-interference nonlinear system model, where the Nash equilibrium of the model was defined. A reinforcement learning (RL) based approximate dynamic programming (ADP) structure was used to study Nash-optimization point of the proposed inputs-interference nonlinear system. The unknown system dynamics were first identified by the parameter-estimation-based NN algorithm with the input-output data. Moreover, the critic NNs are adopted to synthesis the H∞ controls. All the weight parameters throughout the paper were updated with a new estimation algorithm. The parameter convergence and the uniformly ultimate boundedness of the controlled system were both proved. Finally, two examples showed the efficacy of the proposed H∞ controls for the studied inputs-interference nonlinear model and algorithm. In our future work, we will further extend this control method and learning algorithm to practical multi-motor driven system.

References

[1] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE transactions on systems, man, cybernetics*, no. 5, pp. 834-846, Sept. 1983.

[2] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE circuits and systems magazine*, vol. 9, no. 3, pp. 32-50, Aug. 2009.

[3] J. Zhao, J. Na, and G. Gao, "Adaptive dynamic programming based robust control of nonlinear systems with unmatched uncertainties," *Neurocomputing*, vol. 395, pp. 56-65, Jun. 2020.

[4] J. Na, B. Wang, G. Li, S. Zhan, and W. He, "Nonlinear constrained optimal control of wave energy converters with adaptive dynamic programming," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 7904-7915, Oct. 2018.

[5] N. Xu, B. Niu, H. Wang, X. Huo, and X. Zhao, "Single-network ADP for solving optimal event-triggered tracking control problem of completely unknown nonlinear systems," *International Journal of Intelligent Systems*, May. 2021. Doi.org/10.1002/int.22491.

[6] X. Cui, H. Zhang, Y. Luo, and P. Zu, "Online finite-horizon optimal learning algorithm for nonzero-sum games with partially unknown dynamics and constrained inputs," *Neurocomputing*, vol. 185, pp. 37-44, Apr. 2016.

[7] Y. Fu, J. Fu, and T. Chai, "Robust adaptive dynamic programming of two-player zero-sum games for continuous-time linear systems," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 12, pp. 3314-3319, Dec. 2015.

[8] J. Nash, "Non-cooperative games," *Annals of mathematics*, vol. 54, no.2, pp. 286-295, Sep. 1951.

[9] Q. Wei, D. Liu, Q. Lin, and R. Song, "Adaptive dynamic programming for discrete-time zero-sum games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 957-969, Apr. 2018.

[10] Q. Wei, R. Song, and P. Yan, "Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 2, pp. 444-458, Feb. 2016.

[11] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control," *Automatica*, vol. 43, no. 3, pp. 473-481, Mar. 2007.

[12] Q. Wei and H. Zhang, "A new approach to solve a class of continuous-time nonlinear quadratic zero-sum game using ADP," in *2008 IEEE International Conference on Networking, Sensing and Control. (ICNSC)* Sanya, China, May. 2008, pp. 507-512: IEEE.

[13] L. Cui, H. Zhang, X. Zhang, and Y. Luo, "Data-based adaptive critic design for discrete-time zero-sum games using output feedback," in *2011 IEEE Symposium on Adaptive Dynamic Programming And Reinforcement Learning (ADPRL)*, Jul. 2011, pp. 190-195: IEEE.

[14] D. Vrabie and F. Lewis, "Adaptive dynamic programming for online solution of a zero-sum differential game," *Journal of Control Theory and Applications*, vol. 9, no. 3, pp. 353-360, Jul. 2011.

[15] K. G. Vamvoudakis and F. L. Lewis, "Online solution of nonlinear two-player zero-sum games using synchronous policy iteration," *International Journal of Robust and Nonlinear Control*, vol. 22, no. 13, pp. 1460-1483, Jul. 2012.

[16] D. Liu, H. Li, and D. Wang, "Neural-network-based zero-sum game for discrete-time nonlinear systems via iterative adaptive dynamic programming algorithm," *Neurocomputing*, vol. 110, pp. 92-100, Jun. 2013.

[17] Q. Zhang, D. Zhao, and Y. Zhu, "Data-driven adaptive dynamic programming for continuous-time fully cooperative games with partially constrained inputs," *Neurocomputing*, vol. 238, pp. 377-386, May. 2017.

[18] K. G. Vamvoudakis and F. L. Lewis, "Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton–Jacobi equations," *Automatica*, vol. 47, no. 8, pp. 1556-1569, Aug. 2011.

[19] R. Song, F. L. Lewis, and Q. Wei, "Off-Policy Integral Reinforcement Learning Method to Solve Nonlinear Continuous-Time Multiplayer Nonzero-Sum Games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 704-713, Jul. 2017.

[20] D. Liu, H. Li, and D. Wang, "Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 8, pp. 1015-1027, Jan. 2014.

[21] D. Zhao, Q. Zhang, D. Wang, and Y. Zhu, "Experience replay for optimal control of nonzero-sum game systems with unknown dynamics," *IEEE transactions on cybernetics*, vol. 46, no. 3, pp. 854-865, Mar. 2016.

[22] P. Morris, *Introduction to game theory*. Springer Science & Business Media, 2012.

[23] Y. Lv, X. Ren, and J. Na, "Adaptive optimal tracking controls of unknown multi-input systems based on nonzero-sum game theory," *Journal of the*

Franklin Institute, vol. 356, no. 15, pp. 8255-8277, Oct. 2019.
[24] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.
[25] H. Abou-Kandil, G. Freiling, and G. Jank, "Necessary conditions for constant solutions of coupled Riccati equations in Nash games," *Systems & Control Letters*, vol. 21, no. 4, pp. 295-306, Oct. 1993.
[26] Y. Lv, X. Ren, and J. Na, "Online Nash-optimization tracking control of multi-motor driven load system with simplified RL scheme," *ISA transactions*, vol. 98, pp. 251-262, Mar. 2020.
[27] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878-888, May. 2010.
[28] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 943-949, Aug. 2008.
[29] P. J. Reny, "On the existence of pure and mixed strategy Nash equilibria in discontinuous games," *Econometrica*, vol. 67, no. 5, pp. 1029-1056, Dec. 1999.
[30] N. Xu, X. Zhao, G. Zong, and Y. Wang, "Adaptive control design for uncertain switched nonstrict-feedback nonlinear systems to achieve asymptotic tracking performance," *Applied Mathematics Computation*, vol. 408, pp. 126344, Nov. 2021.
[31] Y.-J. Liu, W. Zhao, L. Liu, D. Li, S. Tong, and C. P. Chen, "Adaptive neural network control for a class of nonlinear systems with function constraints on states," *IEEE Transactions on Neural Networks Learning Systems*, Sep. 2021.Doi. 10.1109/TNNLS.2021.3107600.
[32] Y. Lv, J. Na, Q. Yang, X. Wu, and Y. Guo, "Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics," *International Journal of Control*, vol. 89, no. 1, pp. 99-112, Jul. 2016.
[33] J. Na, M. N. Mahyuddin, G. Herrmann, X. Ren, and P. Barber, "Robust adaptive finite-time parameter estimation and control for robotic systems," *International Journal of Robust and Nonlinear Control*, vol. 25, no. 16, pp. 3045-3071, Sep. 2015.
[34] J. Na and G. Herrmann, "Online adaptive approximate optimal tracking control with simplified dual approximation structure for continuous-time unknown nonlinear systems," *IEEE/CAA Journal of Automatica Sinica*, vol. 1, no. 4, pp. 412-422, Oct. 2014.
[35] M. Krstic, I. Kanellakopoulos, and P. V. Kokotovic, *Nonlinear and adaptive control design*. Wiley New York, 1995.
[36] J. Na, S. Wang, Y.-J. Liu, Y. Huang, and X. Ren, "Finite-time convergence adaptive neural network control for nonlinear servo systems," *IEEE transactions on cybernetics*, vol. 50, no. 6, pp. 2568-2579, Jun. 2019.
[37] X. Yang, H. He, and X. Zhong, "Approximate dynamic programming for nonlinear-constrained optimizations," *IEEE Transactions on Cybernetics*, vol. 51, no. 5, pp. 2419-2432, Jul. 2019.
[38] C. Mu, Z. Ni, C. Sun, and H. He, "Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems," *IEEE transactions on cybernetics*, vol. 47, no. 6, pp. 1460-1470, Apr. 2016.
[39] S. Xue, B. Luo, and D. Liu, "Event-triggered adaptive dynamic programming for zero-sum game of partially unknown continuous-time nonlinear systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 9, pp. 3189-3199, Jul. 2018.
[40] H. Zhang, D. Yue, C. Dou, W. Zhao, and X. Xie, "Data-driven distributed optimal consensus control for unknown multiagent systems with input-delay," *IEEE transactions on cybernetics*, vol. 49, no. 6, pp. 2095-2105, Apr. 2018.
[41] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779-791, May. 2005.
[42] T. Hanselmann, L. Noakes, and A. Zaknich, "Continuous-time adaptive critics," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 631-647, May. 2007.
[43] H. Modares, F. L. Lewis, and M. B. Naghibi-Sistani, "Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 10, pp. 1513-1525, Oct. 2013.
[44] L. Liu, Y.-J. Liu, A. Chen, S. Tong, and C. P. Chen, "Integral barrier Lyapunov function-based adaptive control for switched nonlinear systems," *Science China Information Sciences*, vol. 63, no. 3, pp. 1-14, Feb. 2020.
[45] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2226-2236, Dec. 2011.
[46] Y. Lv, X. Ren, and J. Na, "Online optimal solutions for multi-player nonzero-sum game with completely unknown dynamics," *Neurocomputing*, vol. 283, pp. 87-97, Mar. 2018.

Yongfeng Lv (Member, IEEE) received the B.S. and M.S. degrees in Mechatronic Engineering from Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China, in 2012 and 2016, and the Ph.D. degree in Control Science and Engineering with School of Automation, Beijing Institute of Technology, Beijing, China, in 2020.

From 2012 to 2013, he served as an electromechanical technician in Shanxi Coal Transportation and Marketing Group, Linfen, Shanxi. Currently, he is a Lecturer with the College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, China, and also a research fellow with School of Engineering, University of Warwick, Coventry, UK. His current research interests include intelligent control, adaptive dynamic programming, smart energy systems and servo systems.

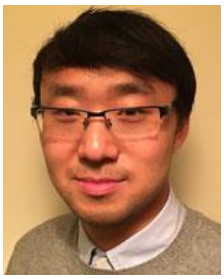
Jing Na (Member, IEEE) received the B.Eng. and Ph.D. degrees in automation and control from the School of Automation, Beijing Institute of Technology, Beijing, China, in 2004 and 2010, respectively.

From 2011 to 2013, he was a Monaco/ITER Postdoctoral Fellow with ITER Organization, Saint-Paul- lez-Durance, France. From 2015 to 2017, he was a Marie Curie Fellow with the Department of Mechanical Engineering, University of Bristol, Bristol, U.K. Since 2010, he has been with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China, where he became a Professor in 2013. He has co-authored one monograph and more than 100 international journal and conference papers. His current research interests include intelligent control, adaptive parameter estimation, nonlinear control and applications for robotics, vehicle systems, and wave energy convertor.

Dr. Na has been awarded the Best Application Paper Award of IFAC ICONS 2013 and the Hsue-Shen Tsien Paper Award in 2017. He is currently an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, Neurocomputing, and has served as the Organization Committee Chair of DDCLS 2019 and International Program Committee Chair of ICMIC 2017.

Xiaowei Zhao received the Ph.D. degree in control theory from Imperial College London, London, U.K., in 2010.

He was a Post-Doctoral Researcher with the University of Oxford, Oxford, U.K., for three years before joining the University of Warwick, Coventry, U.K., in 2013. He is currently Professor of control engineering and an EPSRC Fellow with the School of Engineering, University of Warwick. His main research areas are control theory and machine learning with applications in offshore renewable energy systems, local smart energy systems, and autonomous systems**.**



Yingbo Huang (S'15) received the B.S. degree from Lanzhou City University, China, in 2013, and the Ph.D. degree from the faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China, in 2019. He is currently a Lecturer of mechanical and electrical engineering with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China. His current research interests include adaptive control and transient

performance improvement of nonlinear systems with application to vehicle suspension systems.

Xuemei Ren received the B.S. degree in applied mathematics from Shandong University, Shandong, China, in 1989, and the M.S. and Ph.D. degrees in control engineering from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1992 and 1995, respectively. She has been a Professor with the School of Automation, Beijing Institute of Technology, Beijing, China,

since 2002. From 2001 to 2002, and 2005 to 2005, she visited the Department of Electrical Engineering, Hong Kong Polytechnic University, Hong Kong, China. From 2006 to 2007, she visited the Automation and Robotics Research Institute, University of Texas at Arlington, Arlington, USA, as a Visiting Scholar. She has published over 100 academic papers. Her current research interests include nonlinear systems, intelligent control, neural network control, reinforcement learning and multi-driven servo systems.