mathematics

MDPI

Article

Robust Tracking Control for Non-Zero-Sum Games of Continuous-Time Uncertain Nonlinear Systems

Chunbin Qin [1], Ziyang Shang [1], Zhongwei Zhang [1], Dehua Zhang [1] and Jishi Zhang [2,*]

[1] School of Artificial Intelligence, Henan University, Zhengzhou 450000, China; qcb@henu.edu.cn (C.Q.); sziyang@henu.edu.cn (Z.S.); zhangzw@henu.edu.cn (Z.Z.); dhuazhang@vip.henu.edu.cn (D.Z.)

[2] School of Software, Henan University, Kaifeng 475000, China

* Correspondence: 10250051@vip.henu.edu.cn

Abstract: In this paper, a new adaptive critic design is proposed to approximate the online Nash equilibrium solution for the robust trajectory tracking control of non-zero-sum (NZS) games for continuous-time uncertain nonlinear systems. First, the augmented system was constructed by combining the tracking error and the reference trajectory. By modifying the cost function, the robust tracking control problem was transformed into an optimal tracking control problem. Based on adaptive dynamic programming (ADP), a single critic neural network (NN) was applied for each player to solve the coupled Hamilton–Jacobi–Bellman (HJB) equations approximately, and the obtained control laws were regarded as the feedback Nash equilibrium. Two additional terms were introduced in the weight update law of each critic NN, which strengthened the weight update process and eliminated the strict requirements for the initial stability control policy. More importantly, in theory, through the Lyapunov theory, the stability of the closed-loop system was guaranteed, and the robust tracking performance was analyzed. Finally, the effectiveness of the proposed scheme was verified by two examples.

Keywords: adaptive dynamic programming (ADP); non-zero-sum (NZS) games; robust trajectory tracking; Hamilton–Jacobi–Bellman (HJB) equation; uncertain nonlinear systems

MSC: 93C10; 93D05; 93D21

Citation: Qin, C.; Shang, Z.; Zhang, Z.; Zhang, D.; Zhang, J. Robust Tracking Control for Non-Zero-Sum Games of Continuous-Time Uncertain Nonlinear Systems. *Mathematics* **2022**, *10*, 1904. https://doi.org/10.3390/math10111904

Academic Editors: Ravi P. Agarwal and Maria Alessandra Ragusa

Received: 28 April 2022
Accepted: 30 May 2022
Published: 2 June 2022

1. Introduction

Control theory has been gradually developed to meet the needs of engineering. In practical engineering, environmental uncertainties, such as noise, temperature, etc., greatly affect the stability of a system, so it is very important to find a control method to solve this problem. In recent years, some methods to deal with disturbance or uncertainty have been proposed, such as sliding mode control, type-2 fuzzy control [1,2], internal model control, and so on. However, in a sense, robust control can be also applied to solve the control problems of uncertain dynamic systems [3,4]. Based on the development of adaptive dynamic programming (ADP) and control algorithms, some methods have been effective in solving robust control problems, including guaranteed cost control [5], the system transformation method [6–8], control schemes for robust stabilization using integral reinforcement learning (IRL) methods [9,10]. These methods mainly embody the ideas of reinforcement learning and adaptive dynamic programming. When studying the optimal control problem using adaptive dynamic programming [11–13], the key is to solve the Hamilton–Jacobi–Bellman (HJB) equation; however, due to the curse of dimensionality, it is almost impossible to solve directly. Combining neural network (NN) approximation methods and ADP ideas, the adaptive critic design has been widely used in robust control [14,15]. Considering the adaptive critic design, the approximate solution of the HJB equation can be attained to cope with robust control problems [5,9,15]. For a system with uncertainties, the upper bound

function of uncertainties is usually given, and then the cost function is modified so that the robust control problem can be transformed into the optimal control problem of a nominal system [15]. It has inspired our processing method for uncertain disturbance. From the above results, it can be seen that the basic regulation problem has been solved.

As the complexity of a system increases, a large class of systems often has multiple controllers, such as immune systems [16] and interconnected systems [17]. Game theory considers individual predictive behavior and practical behavior in a game, and studies optimization strategies, and multi-controller system issues can be well addressed by it [18]. As an important theory in game theory, non-zero-sum (NZS) game theory was first proposed in [19] and it aims to find a set of feedback control strategies to achieve the so-called Nash equilibrium while satisfying the defined performance indicators and guaranteeing the system's stability. In this process, the most important aspect is to solve the coupled HJB equations. Since the coupled HJB equations are difficult to solve directly, many advanced algorithms have been developed. In general, iteration-based algorithms can be used to approximate the solution of HJB equations. A policy-based iteration algorithm was used to solve the system of NZS game problems in [20,21]. Considering that it is difficult to know the specific dynamics of complex systems, in [22], based on the iteration algorithm, the Nash equilibrium was obtained approximately by the data-based IRL, which does not need known system dynamics. As policy iteration requires an initial stable control policy, an off-policy IRL method was given to solve the coupled HJB equations in [23]. Recently, the ADP method has become an effective tool in solving the coupled HJB equations. To solve the NZS game of unknown nonlinear systems, using a generalized fuzzy hyperbolic model, an approximately optimal control scheme based on the ADP method was presented in [24]. Combined with the ADP method and the NN structure, the adaptive critic design was also applied to the NZS game. Based on the structure of an actor?critic NN, an adaptive algorithm was proposed for NZS games in the nonlinear system in [25]. In [26], using experience replay techniques, based on the framework of a single critic NN, the NZS game of the unknown dynamical systems was studied. The method proposed above can effectively solve the NZS game. However, there are few studies on NZS games with uncertain disturbances. Therefore, based on adaptive critic design, the NZS game of nonlinear systems with uncertain perturbations was studied in this work.

Initially, our research for the system was limited to allowing the state of the system to converge to the origin; however, many system controller designs also require the controlled object to track a reference trajectory, especially in noisy and uncertain environments. Usually, this is a very common control problem. Trajectory tracking control problems have been solved by some algorithms in [27–34]. The iterative algorithm can still be effectively applied to trajectory tracking control. In [27], to overcome some shortcomings of the traditional controller, an adaptive iterative algorithm was proposed for the robot trajectory tracking problem. Considering disturbance, an iterative algorithm based on Q-learning was presented to solve the H∞ tracking problem of discrete-time systems in [28], which didn't require system dynamics. In [29], the tracking problem was transformed into the tracking error adjustment problem through system transformation, which was solved by the iterative ADP algorithm. Then, some non-iterative algorithms for tracking problems were proposed in [30–34]. In [30], the optimal tracking control was studied using online approximators, but this method involved the reversibility of the control matrix. To overcome the requirement of invertibility of the control matrix, some new methods were proposed. In [31], based on system transformation, a self-learning optimal control method was used to solve the robust trajectory tracking design of uncertain nonlinear systems. Considering the need for multiple outputs in some systems, the robust tracking control of discrete-time systems with multiple inputs and multiple outputs was studied utilizing the adaptive critic design in [32]. By modifying the cost function and introducing a discount factor, the guaranteed cost tracking problem was transformed into an optimal tracking problem, and by developing a new critic NN the optimal tracking control problem could be addressed without policy iteration in [33]. As with some systems with unmatched perturbation, the

NN-based ADP algorithm was used to obtain the approximately optimal tracking control law of uncertain nonlinear systems with a predefined cost function in [34]. In this paper, based on the critic NN structure and the ADP method, an augmented system was used to solve the tracking control problem for NZS games with perturbation.

The main contributions of this paper are as follows:

(1) An augmented system was constructed by combining the tracking error and the reference trajectory. The robust tracking control problem was transformed into an optimal tracking control problem of the nominal augmented system by modifying the cost functions. This method no longer strictly required the control matrix to be reversible. Moreover, in most cases, robust tracking control is applied to some special systems, but here we considered a general system similar to a spring-mass-damper system [31].

(2) For the NZS game between two players with uncertainties, a newly improved adaptive critic design was proposed to solve the revised coupled HJB equations. Two additional terms were introduced in the critic NN weight design, one was used to ensure that the system could always be in a stable state without the need for the initial stability control policy, and the other was used to analyze the stability of the system.

(3) Compared with the actor–critic NN, each player only used one critic NN to approximate their value function and control policy, which could greatly reduce the amount of calculation. By the Lyapunov theory, the stability of the closed-loop system was proved, and the trajectory tracking performance was analyzed. What is more, the adaptive critic design could be carried out online.

The rest of this paper is arranged as follows. In the second section, the description of the two-player NZS game with uncertain terms and the construction method of the augmented matrix structure are given. Then in the third section, a single critic NN structure is used to approximate the value function for each player, and the approximate feedback Nash equilibrium is then solved. Moreover, the system stability analysis and the tracking performance analysis are given. Finally, the effectiveness of the proposed scheme is verified by two examples.

2. Problem Statement

A class of continuous-time uncertain nonlinear dynamical systems for two-player NZS games is given by

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) + k(x(t))v(t) + \Delta f(x(t)), \tag{1}$$

where $x \in R^n$ is the system state, $u \in R^m$ is the first control input, $v \in R^q$ is the second control input. The known functions $f(\cdot)$, $g(\cdot)$ and $k(\cdot)$ are Lipschitz continuous on a compact set $\Omega \subseteq R^n$ with $f(0)$=0. $\Delta f(x(t)) = M(x)d(x)$ is the unknown perturbation satisfying $\Delta f(0) = 0$. Here, $M(\cdot) \in R^{n\times r}$ is a known function, and $d(\cdot) \in R^r$ is an uncertain function with $d(0) = 0$. One chooses the initial state as $x(0) = x_0$. Let the uncertain term $\Delta f(x(t))$ be bounded by a known function $\lambda_f(x)$, i.e., $\|\Delta f(x)\| \leq \lambda_f(x)$ with $\lambda_f(0) = 0$.

Here, we introduce a system reference trajectory command generator to implement the trajectory tracking, that is

$$\dot{s}(t) = \varphi(s(t)), \tag{2}$$

where $s(t) \in R^n$ denotes the bounded reference trajectory. Let the initial trajectory be $s(0) = s_0$ and $\varphi(s(t))$ is a Lipschitz continuous function with $\varphi(0) = 0$. The tracking error is defined as

$$e_r(t) = x(t) - s(t). \tag{3}$$

Then, the initial error vector is $e_r(0) = e_{r0} = x_0 - s_0$. According to (1)–(3), the tracking error dynamics can be obtained as

$$\dot{e}_r(t) = f(x(t)) - \varphi(s(t)) + g(x(t))u(t) + k(x(t))v(t) + \Delta f(x(t)). \tag{4}$$

Due to $x(t) = e_r(t) + s(t)$, system (4) is written as

$$\begin{aligned}\dot{e}_r(t) = f(e_r(t) + s(t)) - \varphi(s(t)) + g(e_r(t) + s(t))u(t) + k(e_r(t) + s(t))v(t) \\ + \Delta f(e_r(t) + s(t)).\end{aligned} \tag{5}$$

To introduce the augmented system, we define an augmented state vector $\zeta(t) = [e_r^{\mathsf{T}}(t), s^{\mathsf{T}}(t)]^{\mathsf{T}} \in R^{2n}$, and we can choose its initial condition as $\zeta(0) = \zeta_0 = [e_r^{\mathsf{T}}(0), s^{\mathsf{T}}(0)]^{\mathsf{T}} \in R^{2n}$. Combining (2) and (5), the augmented system dynamics is simplified to

$$\dot{\zeta}(t) = \mathcal{F}(\zeta(t)) + \mathcal{G}(\zeta(t))u(t) + \mathcal{K}(\zeta(t))v(t) + \Delta\mathcal{F}(\zeta(t)), \tag{6}$$

where $\mathcal{F}(\cdot)$, $\mathcal{G}(\cdot)$ and $\mathcal{K}(\cdot)$ are new system matrices. What is more, $\Delta\mathcal{F}(\zeta)$ represents the augmented system uncertainty, and they are written in the following specific form:

$$\mathcal{F}(\zeta(t)) = \begin{bmatrix} f(e_r(t) + s(t)) - \varphi(s(t)) \\ \varphi(s(t)) \end{bmatrix}, \tag{7}$$

$$\mathcal{G}(\zeta(t)) = \begin{bmatrix} g(e_r(t) + s(t)) \\ 0_{n\times m} \end{bmatrix}, \tag{8}$$

$$\mathcal{K}(\zeta(t)) = \begin{bmatrix} k(e_r(t) + s(t)) \\ 0_{n\times q} \end{bmatrix}, \tag{9}$$

$$\Delta\mathcal{F}(\zeta(t)) = \begin{bmatrix} \Delta f(e_r(t) + s(t)) \\ 0_{n\times 1} \end{bmatrix}. \tag{10}$$

It's easy to conclude that $\Delta\mathcal{F}(\zeta)$ is upper bounded, and the details are as follows:

$$\|\Delta\mathcal{F}(\zeta)\| = \|\Delta f(e_r + s)\| = \|\Delta f(x)\| \leq \lambda_f(e_r + s) \triangleq \lambda_f(\zeta). \tag{11}$$

In order to better analyze the NZS game with the uncertain perturbation, we decompose the uncertain term $\Delta\mathcal{F}(\zeta)$ into

$$\Delta\mathcal{F}(\zeta) = \Delta\mathcal{F}_1(\zeta) + \Delta\mathcal{F}_2(\zeta) = M_1(\zeta)d_1(\zeta) + M_2(\zeta)d_2(\zeta), \tag{12}$$

where $M_1(\cdot) \in R^{n\times r}$ and $M_2(\cdot) \in R^{n\times r}$ are known functions in the uncertain term. $d_1(\cdot) \in R^r$ and $d_2(\cdot) \in R^r$ are the uncertain functions satisfying $d_1(0) = d_2(0) = 0$. Similarly, two known functions $\lambda_{f1}(\zeta)$ and $\lambda_{f2}(\zeta)$ are the upper bounds of $\Delta\mathcal{F}_1(\zeta)$ and $\Delta\mathcal{F}_2(\zeta)$ with $\lambda_{f1}(0) = \lambda_{f2}(0) = 0$.

Assumption 1. *The control function matrixes $g(x)$ and $k(x)$ are bounded as $\|g(x)\| \leq \lambda_g$ and $\|k(x)\| \leq \lambda_k$ [31], where λ_g and λ_k are positive constants, and hence*

$$\|\mathcal{G}(\zeta)\| = \|g(e_r + s)\| = \|g(x)\| \leq \lambda_g, \tag{13}$$

$$\|\mathcal{K}(\zeta)\| = \|k(e_r + s)\| = \|k(x)\| \leq \lambda_k. \tag{14}$$

By constructing the augmented dynamics (6), the feedback control laws $u(\zeta)$ and $v(\zeta)$ are found to make the state of system move along the reference trajectory. At the same time, the closed-loop system is asymptotically stable under the influence of the uncertain term. Next, we can give the appropriate cost functions to transform the robust control the problem into the optimal control problem for its nominal system.

For the augmented system (6), we focus on the nominal system part

$$\dot{\zeta}(t) = \mathcal{F}(\zeta(t)) + \mathcal{G}(\zeta(t))u(t) + \mathcal{K}(\zeta(t))v(t). \tag{15}$$

The two-player cost functions are

$$\mathcal{J}_1(\zeta_0, u, v) = \int_0^\infty \{\Gamma_1(\zeta(t)) + U_1(\zeta(t), u(t), v(t))\}dt, \tag{16}$$

$$\mathcal{J}_2(\zeta_0, u, v) = \int_0^\infty \{\Gamma_2(\zeta(t)) + U_2(\zeta(t), u(t), v(t))\}dt, \tag{17}$$

where $U_1(\zeta, u, v)$ and $U_2(\zeta, u, v)$ are the the basic parts of utility functions with $U_1(0,0,0) = U_2(0,0,0) = 0$, $U_1(\zeta, u, v) \geq 0$ and $U_2(\zeta, u, v) \geq 0$ for all ζ, u and v. Utility functions are chosen as $U_1(\zeta, u, v) = \zeta^\mathsf{T}\bar{Q}_1\zeta + u^\mathsf{T}R_{11}u + v^\mathsf{T}R_{12}v$ and $U_2(\zeta, u, v) = \zeta^\mathsf{T}\bar{Q}_2\zeta + u^\mathsf{T}R_{21}u + v^\mathsf{T}R_{22}v$, where $\bar{Q}_1 = diag\{Q_1, 0_{n\times n}\}$, $\bar{Q}_2 = diag\{Q_2, 0_{n\times n}\}$, Q_1, Q_2, R_{11}, R_{12}, R_{21} and R_{22} are positive definite matrices. $\Gamma_1(\zeta)$ and $\Gamma_2(\zeta)$ are related to the dynamical uncertainty with $\Gamma_1(\zeta) \geq 0$ and $\Gamma_2(\zeta) \geq 0$. What is more, the feedback controllers required to solve the optimal control problem are admissible. Then, the definition of admissible policies is described below.

Definition 1. *(Admissible policies) Control functions $u(\zeta)$ and $v(\zeta)$ are said to be admissible with respect to (16) and (17) on $\Omega \subseteq R^n$ [26], if $u(\zeta)$ and $v(\zeta)$ are continuous on Ω, $u(0) = v(0) = 0$, $u(\zeta)$ and $v(\zeta)$ stabilize system (15) on Ω, moreover, the cost functions (16) and (17) are finite $\forall\zeta_0 \in \Omega$.*

Given admissible feedback policies $u(\zeta) \in \mathcal{A}(\Omega)$ and $v(\zeta) \in \mathcal{A}(\Omega)$, one can define value functions that correspond to the cost functions as

$$V_1(\zeta(t)) = \int_t^\infty \{\Gamma_1(\zeta(\tau)) + U_1(\zeta(\tau), u(\tau), v(\tau))\}d\tau, \tag{18}$$

$$V_2(\zeta(t)) = \int_t^\infty \{\Gamma_2(\zeta(\tau)) + U_2(\zeta(\tau), u(\tau), v(\tau))\}d\tau, \tag{19}$$

where one can define $\Gamma_1(\zeta)$ and $\Gamma_2(\zeta)$ as

$$\Gamma_1(\zeta) = \lambda_{f1}^2(\zeta) + \frac{1}{4}(\nabla V_1(\zeta))^\mathsf{T} M_1(\zeta) M_1^\mathsf{T}(\zeta)\nabla V_1(\zeta), \tag{20}$$

$$\Gamma_2(\zeta) = \lambda_{f2}^2(\zeta) + \frac{1}{4}(\nabla V_2(\zeta))^\mathsf{T} M_2(\zeta) M_2^\mathsf{T}(\zeta)\nabla V_2(\zeta). \tag{21}$$

In this paper, a 2-tuple of policies $\{u, v\}$ is found to minimize (18) and (19), thus, the optimal value functions V_1^* and V_2^* are defined as

$$V_1^*(\zeta(t)) = \min_{u\in\mathcal{A}(\Omega)} \int_t^\infty \{\Gamma_1(\zeta(\tau)) + U_1(\zeta(\tau), u(\tau), v(\tau))\}d\tau, \tag{22}$$

$$V_2^*(\zeta(t)) = \min_{v\in\mathcal{A}(\Omega)} \int_t^\infty \{\Gamma_2(\zeta(\tau)) + U_2(\zeta(\tau), u(\tau), v(\tau))\}d\tau. \tag{23}$$

In addition, there exists a Nash equilibrium in the NZS game between two players. Next, we give the Nash equilibrium definition.

Definition 2. *(Nash equilibrium policies) A 2-tuple of policies $\{u^*, v^*\}$ with $u, v \in \mathcal{A}(\Omega)$ is said to constitute a Nash equilibrium solution for the two-player game [35], if the following two inequalities are satisfied for all $u, v \in \mathcal{A}(\Omega)$:*

$$\mathcal{J}_1^*(u^*, v^*) \leq \mathcal{J}_1(u, v^*), \tag{24}$$

$$\mathcal{J}_2^*(u^*, v^*) \leq \mathcal{J}_2(u^*, v). \tag{25}$$

Under the admissible feedback policies, if the value functions (18) and (19) are continuously differentiable, their differential equivalents are given by

$$0 = \Gamma_1(\zeta) + U_1(\zeta, u, v) + (\nabla V_1)^{\mathsf{T}}[(\mathcal{F}(\zeta) + \mathcal{G}(\zeta)u(\zeta) + \mathcal{K}(\zeta)v(\zeta)], \tag{26}$$

$$0 = \Gamma_2(\zeta) + U_2(\zeta, u, v) + (\nabla V_2)^{\mathsf{T}}[\mathcal{F}(\zeta) + \mathcal{G}(\zeta)u(\zeta) + \mathcal{K}(\zeta)v(\zeta)], \tag{27}$$

with $V_i(0) = 0$ and $\nabla V_i = \partial V_i / \partial \zeta, i = 1, 2$. Define the Hamiltonian functions

$$\begin{aligned} H_1(\zeta, u(\zeta), v(\zeta), \nabla V_1) =& \Gamma_1(\zeta) + U_1(\zeta, u(\zeta), v(\zeta)) \\ &+ (\nabla V_1)^{\mathsf{T}}[(\mathcal{F}(\zeta) + \mathcal{G}(\zeta)u(\zeta) + \mathcal{K}(\zeta)v(\zeta)], \end{aligned} \tag{28}$$

$$\begin{aligned} H_2(\zeta, u(\zeta), v(\zeta), \nabla V_2) =& \Gamma_2(\zeta) + U_2(\zeta, u(\zeta), v(\zeta)) \\ &+ (\nabla V_2)^{\mathsf{T}}[(\mathcal{F}(\zeta) + \mathcal{G}(\zeta)u(\zeta) + \mathcal{K}(\zeta)v(\zeta)]. \end{aligned} \tag{29}$$

According to the stationarity conditions [36], two players' optimal feedback control policies are given by

$$\frac{\partial H_1}{\partial u} = 0 \Rightarrow u^* = -\frac{1}{2} R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta) \nabla V_1^*, \tag{30}$$

$$\frac{\partial H_2}{\partial v} = 0 \Rightarrow v^* = -\frac{1}{2} R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta) \nabla V_2^*. \tag{31}$$

Combining (26), (27), (30) and (31), one obtains the coupled HJB equations

$$\begin{aligned} 0 =& \Gamma_1(\zeta) + \zeta^{\mathsf{T}} \bar{Q}_1 \zeta + (\nabla V_1^*)^{\mathsf{T}} \mathcal{F}(\zeta) - \frac{1}{2} (\nabla V_1^*)^{\mathsf{T}} \mathcal{G}(\zeta) R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta) (\nabla V_1^*) \\ &- \frac{1}{2} (\nabla V_1^*)^{\mathsf{T}} \mathcal{K}(\zeta) R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta) (\nabla V_2^*) + \frac{1}{4} (\nabla V_1^*)^{\mathsf{T}} \mathcal{G}(\zeta) R_{11}^{-\mathsf{T}} R_{11} R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta) \nabla V_1^* \\ &+ \frac{1}{4} (\nabla V_2^*)^{\mathsf{T}} \mathcal{K}(\zeta) R_{22}^{-\mathsf{T}} R_{12} R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta) \nabla V_2^*, \end{aligned} \tag{32}$$

$$\begin{aligned} 0 =& \Gamma_2(\zeta) + \zeta^{\mathsf{T}} \bar{Q}_2 \zeta + (\nabla V_2^*)^{\mathsf{T}} \mathcal{F}(\zeta) - \frac{1}{2} (\nabla V_2^*)^{\mathsf{T}} \mathcal{G}(\zeta) R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta) (\nabla V_1^*) \\ &- \frac{1}{2} (\nabla V_2^*)^{\mathsf{T}} \mathcal{K}(\zeta) R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta) (\nabla V_2^*) + \frac{1}{4} (\nabla V_1^*)^{\mathsf{T}} \mathcal{G}(\zeta) R_{11}^{-\mathsf{T}} R_{21} R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta) \nabla V_1^* \\ &+ \frac{1}{4} (\nabla V_2^*)^{\mathsf{T}} \mathcal{K}(\zeta) R_{22}^{-\mathsf{T}} R_{22} R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta) \nabla V_2^*, \end{aligned} \tag{33}$$

where $V_1^*(0) = 0$ and $V_2^*(0) = 0$. To simplify the operation, eight non-negative matrices $A_i(\zeta), B_i(\zeta), C_i(\zeta)$ and $D_i(\zeta), i = 1, 2$ are given by

$$A_1(\zeta) = M_1(\zeta) M_1^{\mathsf{T}}(\zeta), \tag{34a}$$

$$A_2(\zeta) = M_2(\zeta) M_2^{\mathsf{T}}(\zeta), \tag{34b}$$

$$B_1(\zeta) = \mathcal{G}(\zeta) R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta), \tag{34c}$$

$$B_2(\zeta) = \mathcal{K}(\zeta) R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta), \tag{34d}$$

$$C_1(\zeta) = \mathcal{K}(\zeta) R_{22}^{-\mathsf{T}} R_{12} R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta), \tag{34e}$$

$$C_2(\zeta) = \mathcal{G}(\zeta) R_{11}^{-\mathsf{T}} R_{21} R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta), \tag{34f}$$

$$D_1(\zeta) = \mathcal{K}(\zeta) R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta), \tag{34g}$$

$$D_2(\zeta) = \mathcal{G}(\zeta) R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta). \tag{34h}$$

We all know that it is difficult to directly solve the coupled HJB equations, so, next, we approximate their solutions using the NN-based adaptive critic design.

3. Robust Trajectory Tracking Design for Non-Zero-Sum Games

This section mainly includes two parts. First, the solution of coupled HJB equations is approximated by the adaptive critic design based on a single NN structure, so that the so-called Nash equilibrium is found. Secondly, the stability of the system is proved and the tracking performance is analyzed via the Lyapunov theory.

3.1. Neural Network Implementation

In order to realize the neural network approximation, we first introduce the Weierstrass high-order approximation theorem [37,38].

Assumption 2. *The solutions to* (26) *and* (27) *are smooth.*

According to Assumption 2, there exist complete independent basis sets $\{\varpi_i(\zeta)\}$ and $\{\mu_i(\zeta)\}$ such that the solutions to (26) and (27) and their gradients are uniformly approximated, that is, there exist coefficients c_i and z_i such that

$$V_1(\zeta) = \sum_{i=1}^{\infty} c_i\varpi_i(\zeta) = \sum_{i=1}^{K} c_i\varpi_i(\zeta) + \sum_{i=K+1}^{\infty} c_i\varpi_i(\zeta), \tag{35}$$

$$V_2(\zeta) = \sum_{i=1}^{\infty} z_i\mu_i(\zeta) = \sum_{i=1}^{K} z_i\mu_i(\zeta) + \sum_{i=K+1}^{\infty} z_i\mu_i(\zeta). \tag{36}$$

Then we have

$$V_1(\zeta) \equiv C_1^{\mathsf{T}}\phi_1(\zeta) + \sum_{i=K+1}^{\infty} c_i\varpi_i(\zeta), \tag{37}$$

$$V_2(\zeta) \equiv Z_1^{\mathsf{T}}\phi_2(\zeta) + \sum_{i=K+1}^{\infty} z_i\mu_i(\zeta), \tag{38}$$

where $\phi_1(\zeta) = [\varpi_1(\zeta), \varpi_2(\zeta) \dots \varpi_K(\zeta)]^{\mathsf{T}}$, $\phi_2(\zeta) = [\mu_1(\zeta), \mu_2(\zeta) \dots \mu_K(\zeta)]^{\mathsf{T}}$, and the last terms in these equations converge uniformly to zero as $K \to \infty$. Next, we give the specific content of the value function approximation.

For the augmented dynamics (15), the value functions are re-expressed as

$$V_1(\zeta) = W_1^{\mathsf{T}}\phi_1(\zeta) + \varepsilon_1, \tag{39}$$

$$V_2(\zeta) = W_2^{\mathsf{T}}\phi_2(\zeta) + \varepsilon_2, \tag{40}$$

where $W_1, W_2 \in R^K$ are ideal weights, $\phi_1(\zeta), \phi_2(\zeta) \in R^K$ are defined as activation function vectors, K is the number of hidden neurons, and ε_1 and ε_2 are the critic NN approximation errors. When $K \to \infty$, ε_1 and ε_2 converge to zero; however, when K is a fixed constant they are bounded.

Assumption 3. *In order to ensure the boundedness, we make the following assumptions, as in [26].*

(1) The critic NN activation functions and their gradients are bounded such as $\|\phi_i\| \le \lambda_{\phi_i}$ *and* $\|\nabla\phi_i\| \le \lambda_{d\phi_i}, i = 1, 2$. λ_{ϕ_i} *and* $\lambda_{d\phi_i}$*are positive constants.*

(2) The critic NN approximation errors and their gradients are bounded by positive constants such that $\|\varepsilon_1\| \le \lambda_{\varepsilon_i}$ *and* $\|\nabla\varepsilon_i\| \le \lambda_{d\varepsilon_i}, i = 1, 2$. λ_{ε_i} *and* $\lambda_{d\varepsilon_i}$ *are positive constants.*

(3) The critic NN weights are upper bounded such that $\|W_i\| \le \bar{W}_i, i = 1, 2$. $\bar{W}_i$ *are positive constants.*

The derivatives of (39) and (40) along with ζ are

$$\nabla V_1(\zeta) = \nabla\phi_1^{\mathsf{T}}(\zeta)W_1 + \nabla\varepsilon_1, \tag{41}$$

$$\nabla V_2(\zeta) = \nabla \phi_2^{\mathsf{T}}(\zeta) W_2 + \nabla \varepsilon_2, \tag{42}$$

where $\nabla \phi_i = \partial \phi_i / \partial \zeta, \nabla \varepsilon_i = \partial \varepsilon_i / \partial \zeta, i = 1, 2$. Noticing (30), (31), (41) and (42), the optimal control laws are written as

$$u^* = -\frac{1}{2} R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta) [\nabla \phi_1^{\mathsf{T}}(\zeta) W_1 + \nabla \varepsilon_1], \tag{43}$$

$$v^* = -\frac{1}{2} R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta) [\nabla \phi_2^{\mathsf{T}}(\zeta) W_2 + \nabla \varepsilon_2]. \tag{44}$$

Then the associated Bellman equations can be derived as

$$\Gamma_1(\zeta) + U_1(\zeta, u, v) + W_1^{\mathsf{T}} \nabla \phi_1(\zeta) [\mathcal{F}(\zeta) + \mathcal{G}(\zeta) u(\zeta) + \mathcal{K}(\zeta) v(\zeta)] = \varepsilon_{b_1}, \tag{45}$$

$$\Gamma_2(\zeta) + U_2(\zeta, u, v) + W_2^{\mathsf{T}} \nabla \phi_2(\zeta) [\mathcal{F}(\zeta) + \mathcal{G}(\zeta) u(\zeta) + \mathcal{K}(\zeta) v(\zeta)] = \varepsilon_{b_2}, \tag{46}$$

where $\varepsilon_{b_i} = -(\nabla \varepsilon_i)^{\mathsf{T}} (\mathcal{F} + \mathcal{G} u + \mathcal{K} v), i = 1, 2$ are the Bellman equation errors. When the number of the critic NN hidden neurons $K \to \infty$, they converge to zero [36]. However, when K is a fixed constant they are bounded by constants such as $\|\varepsilon_{b_i}\| \leq \lambda_{\varepsilon_{b_i}}, i = 1, 2$.

Based on (32), (33), (43) and (44), one obtains

$$\begin{aligned} H_1 =& \zeta^{\mathsf{T}} \bar{Q}_1 \zeta + \lambda_{f1}^2(\zeta) + W_1^{\mathsf{T}} \nabla \phi_1(\zeta) \mathcal{F}(\zeta) + \frac{1}{4} W_1^{\mathsf{T}} \nabla \phi_1(\zeta) A_1(\zeta) \nabla \phi_1^{\mathsf{T}}(\zeta) W_1 \\ & - \frac{1}{4} W_1^{\mathsf{T}} \nabla \phi_1(\zeta) B_1(\zeta) \nabla \phi_1^{\mathsf{T}}(\zeta) W_1 + \frac{1}{4} W_2^{\mathsf{T}} \nabla \phi_2(\zeta) C_1(\zeta) \nabla \phi_2^{\mathsf{T}}(\zeta) W_2 \\ & - \frac{1}{2} W_1^{\mathsf{T}} \nabla \phi_1(\zeta) D_1(\zeta) \nabla \phi_2^{\mathsf{T}}(\zeta) W_2 = \varepsilon_{HJ_1}, \end{aligned} \tag{47}$$

$$\begin{aligned} H_2 =& \zeta^{\mathsf{T}} \bar{Q}_2 \zeta + \lambda_{f2}^2(\zeta) + W_2^{\mathsf{T}} \nabla \phi_2(\zeta) \mathcal{F}(\zeta) + \frac{1}{4} W_2^{\mathsf{T}} \nabla \phi_2(\zeta) A_2(\zeta) \nabla \phi_2^{\mathsf{T}}(\zeta) W_2 \\ & - \frac{1}{4} W_2^{\mathsf{T}} \nabla \phi_2(\zeta) B_2(\zeta) \nabla \phi_2^{\mathsf{T}}(\zeta) W_2 + \frac{1}{4} W_1^{\mathsf{T}} \nabla \phi_1(\zeta) C_2(\zeta) \nabla \phi_1^{\mathsf{T}}(\zeta) W_1 \\ & - \frac{1}{2} W_2^{\mathsf{T}} \nabla \phi_2(\zeta) D_2(\zeta) \nabla \phi_1^{\mathsf{T}}(\zeta) W_1 = \varepsilon_{HJ_2}. \end{aligned} \tag{48}$$

ε_{HJ_1} and ε_{HJ_2} are the coupled HJB equations approximation errors shown in [36]. Without loss of generality, as the number of the critic NN hidden neurons $K \to \infty$, they converge to zero. However, when K is a fixed constant they are bounded by positive constants such that $\|\varepsilon_{HJ_i}\| \leq \lambda_{\varepsilon_{HJ_i}}, i = 1, 2$.

Since the ideal weights W_1 and W_2 are unknown, they are estimated as $\hat{W}_1$ and $\hat{W}_2$, then the weight estimation errors are defined as $\tilde{W}_i = W_i - \hat{W}_i, i = 1, 2$. The estimated value functions are given by

$$\hat{V}_1(\zeta) = \hat{W}_1^{\mathsf{T}} \phi_1(\zeta), \tag{49}$$

$$\hat{V}_2(\zeta) = \hat{W}_2^{\mathsf{T}} \phi_2(\zeta). \tag{50}$$

Meanwhile, the approximate optimal control policies are presented as

$$\hat{u}^* = -\frac{1}{2} R_{11}^{-1} \mathcal{G}^{\mathsf{T}}(\zeta) \nabla \phi_1^{\mathsf{T}}(\zeta) \hat{W}_1, \tag{51}$$

$$\hat{v}^* = -\frac{1}{2} R_{22}^{-1} \mathcal{K}^{\mathsf{T}}(\zeta) \nabla \phi_2^{\mathsf{T}}(\zeta) \hat{W}_2. \tag{52}$$

Based on (32), (33), (51) and (52), the approximate Hamilton functions are

$$\begin{aligned}\hat{H}_1 =&\zeta^{\mathsf{T}}\bar{Q}_1\zeta+\lambda_{f1}^2(\zeta)+\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)\mathcal{F}(\zeta)+\frac{1}{4}\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)A_1(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1\\&-\frac{1}{4}\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)B_1(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1+\frac{1}{4}\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)C_1(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2\\&-\frac{1}{2}\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)D_1(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2\triangleq e_1,\end{aligned}\tag{53}$$

$$\begin{aligned}\hat{H}_2 =&\zeta^{\mathsf{T}}\bar{Q}_2\zeta+\lambda_{f2}^2(\zeta)+\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)\mathcal{F}(\zeta)+\frac{1}{4}\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)A_2(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2\\&-\frac{1}{4}\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)B_2(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2+\frac{1}{4}\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)C_2(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1\\&-\frac{1}{2}\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)D_2(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1\triangleq e_2,\end{aligned}\tag{54}$$

where e_1 and e_2 are the residual errors. The next tasks are to train neural networks and design $\hat{W}_1$ and $\hat{W}_2$ to minimize the target function $E=\frac{1}{2}e_1^{\mathsf{T}}e_1+\frac{1}{2}e_2^{\mathsf{T}}e_2$. Then $\hat{W}_1$ and $\hat{W}_2$ converge to W_1 and W_2.

To overcome the difficulty of finding the initial admissible controllers, the following assumption is given. Furthermore, an additional term is developed to strengthen the learning process of the critic NN.

Assumption 4. *Given the cost functions* (16) *and* (17)*, for the nominal augmented system* (15)*, under the optimal control policies of the two players, we define a continuously differentiable Lyapunov function candidate $J_s(\zeta)$ satisfying*

$$\dot{J}_s(\zeta)=(\nabla J_s(\zeta))^{\mathsf{T}}[\mathcal{F}(\zeta)+\mathcal{G}(\zeta)u^*(\zeta)+\mathcal{K}(\zeta)v^*(\zeta)]<0,\tag{55}$$

where $\nabla J_s(\zeta)=\partial J_s(\zeta)/\partial\zeta$. Suppose there exists a positive definite matrix $\Xi(\zeta)$ such that

$$(\nabla J_s(\zeta))^{\mathsf{T}}[\mathcal{F}(\zeta)+\mathcal{G}(\zeta)u^*(\zeta)+\mathcal{K}(\zeta)v^*(\zeta)]=-(\nabla J_s(\zeta))^{\mathsf{T}}\Xi(\zeta)\nabla J_s(\zeta)\tag{56}$$

holds [5].

Remark 1. *We assume that $\|\mathcal{F}(\zeta)+\mathcal{G}(\zeta)u^*(\zeta)+\mathcal{K}(\zeta)v^*(\zeta)\|\leq\theta\|\nabla J_s(\zeta)\|$, and θ is a positive constant [5]. Hence, we have $\|(\nabla J_s(\zeta))^{\mathsf{T}}\mathcal{F}(\zeta)+\mathcal{G}(\zeta)u^*(\zeta)+\mathcal{K}(\zeta)v^*(\zeta)\|\leq\theta\|\nabla J_s(\zeta)\|^2$. The minimum and maximum eigenvalues of matrix $\Xi(\zeta)$ are λ_m and λ_M, then we obtain*

$$\lambda_m\|\nabla J_s(\zeta)\|^2\leq(\nabla J_s(\zeta))^{\mathsf{T}}\Xi(\zeta)\nabla J_s(\zeta)\leq\lambda_M\|\nabla J_s(\zeta)\|^2.\tag{57}$$

Here, $J_s(\zeta)$ can be selected as $J_s(\zeta)=0.5\zeta^{\mathsf{T}}\zeta$.

Now, based on the normalized gradient descent algorithm, the weights of the critic NN for each player are tuned with two additional terms, that is

$$\begin{aligned}\dot{\hat{W}}_1=&-a\frac{\sigma_{11}}{(1+\sigma_{11}^{\mathsf{T}}\sigma_{11})^2}[\sigma_{11}^{\mathsf{T}}\hat{W}_1+\lambda_{f1}^2(\zeta)+U_1(\zeta,\hat{u},\hat{v})-\frac{1}{4}\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)A_1(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1]\\&+\frac{b}{2}\Pi(\zeta,\hat{u},\hat{v})\nabla\phi_1(\zeta)B_1(\zeta)\nabla J_s(\zeta)+\frac{a\sigma_{11}}{4\times(1+\sigma_{11}^{\mathsf{T}}\sigma_{11})^2}[\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)A_1(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1\\&-\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)B_1(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1-\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)C_1(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2],\end{aligned}\tag{58}$$

$$\begin{aligned}\dot{\hat{W}}_2=&-a\frac{\sigma_{22}}{(1+\sigma_{22}^{\mathsf{T}}\sigma_{22})^2}[\sigma_{22}^{\mathsf{T}}\hat{W}_2+\lambda_{f2}^2(\zeta)+U_2(\zeta,\hat{u},\hat{v})-\frac{1}{4}\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)A_2(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2]\\&+\frac{b}{2}\Pi(\zeta,\hat{u},\hat{v})\nabla\phi_2(\zeta)B_2(\zeta)\nabla J_s(\zeta)+\frac{a\sigma_{22}}{4\times(1+\sigma_{22}^{\mathsf{T}}\sigma_{22})^2}[\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)A_2(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2\\&-\hat{W}_2^{\mathsf{T}}\nabla\phi_2(\zeta)B_2(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\hat{W}_2-\hat{W}_1^{\mathsf{T}}\nabla\phi_1(\zeta)C_2(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\hat{W}_1],\end{aligned}\tag{59}$$

where $a > 0$ is the learning rate of the critic NN and the third term, $b > 0$ is the learning rate of the second term. $\sigma_{ii} = \nabla\phi_i(\mathcal{F}(\zeta) + \mathcal{G}(\zeta)\hat{u} + \mathcal{K}(\zeta)\hat{v}) + \frac{1}{2}\nabla\phi_i(\zeta)M_i(\zeta)M_i^{\mathsf{T}}(\zeta)\nabla\phi_i^{\mathsf{T}}(\zeta)\hat{W}_i$, $i = 1, 2$. In addition, $J_s(\zeta)$ is given in Assumption 3. In (58) and (59), the $\Pi(\zeta, \hat{u}, \hat{v})$ is the additional stabilizing term defined as

$$\Pi(\zeta, \hat{u}, \hat{v}) = \begin{cases} 0, \text{if} \dot{J}_s(\zeta) = (\nabla J_s(\zeta))^{\mathsf{T}}[\mathcal{F}(\zeta) + \mathcal{G}(\zeta)\hat{u}(\zeta) + \mathcal{K}(\zeta)\hat{v}(\zeta)] < 0 \\ 1, \text{else} \end{cases} \tag{60}$$

Remark 2. *The second term introduced guarantees that the system remains stable during the weight update process. When the system is stable, the value of this item is 0. When the system is unstable, this item is activated to reinforce system stability by enhancing the training process. On account of*

$$-\frac{\partial \dot{J}_s(\zeta)}{\partial \hat{W}_1} = -\left(\frac{\partial \hat{u}}{\partial \hat{W}_1}\right)\frac{\dot{J}_s(\zeta)}{\partial \hat{u}} = \frac{1}{2}\nabla\phi_1(\zeta)B_1(\zeta)\nabla J_s(\zeta) \tag{61}$$

and

$$-\frac{\partial \dot{J}_s(\zeta)}{\partial \hat{W}_2} = -\left(\frac{\partial \hat{v}}{\partial \hat{W}_2}\right)\frac{\dot{J}_s(\zeta)}{\partial \hat{v}} = \frac{1}{2}\nabla\phi_2(\zeta)B_2(\zeta)\nabla J_s(\zeta), \tag{62}$$

the additional stability term makes the weights update in the opposite direction of $\dot{J}_s(\zeta)$. If $\dot{J}_s(\zeta) \geq 0$, the reinforced training process can reduce it to a negative value. On the other hand, when the probing noise is needed to satisfy the persistent excitation (PE) condition, the additional stabilizing term can keep the system in a closed-loop stable state, which leads the system to no longer need initial stability control. The third terms given in (58) and (59) are for the next stability analysis.

3.2. Stability Analysis

In this section, we give several theorems and then add some assumptions to prove the stability of the closed-loop nominal augmented system and analyze the tracking performance.

Assumption 5. *Assume that the matrices associated with each player's control input have upper bounds, i.e.$R_{11} \leq R_{11M}$, $R_{12} \leq R_{12M}$, $R_{21} \leq R_{21M}$ and $R_{22} \leq R_{22M}$. Eight non-negative matrices $A_i(\zeta), B_i(\zeta), C_i(\zeta)$ and $D_i(\zeta), i = 1, 2$ are bounded, i.e.$A_i(\zeta) \leq \lambda_{Ai}, B_i(\zeta) \leq \lambda_{Bi}, C_i(\zeta) \leq \lambda_{Ci}$ and $D_i(\zeta) \leq \lambda_{Di}, i = 1, 2$, λ_{Ai}, λ_{Bi}, λ_{Ci} and $\lambda_{Di}, i = 1, 2$ are positive constants. Moreover, $B_1(\zeta)\nabla\varepsilon_1(\zeta) \leq \lambda_2$ and $B_2(\zeta)\nabla\varepsilon_2(\zeta) \leq \lambda_3$. λ_2, λ_3, R_{11M}, R_{12M}, R_{21M} and R_{22M} are positive constants.*

Theorem 1. *For the nominal augmented system (15), a pair of feedback control laws $\{u^*, v^*\}$ are derived by (51) and (52), moreover, the weight vectors of the critic NN are trained by (58) and (59), respectively. Then, we have that the closed-loop system state and the critic NN weights' estimation errors are both uniformly ultimately bounded (UUB).*

Proof. See the Appendix A. □

According to Thereom 1, it is easy to conclude that the feedback control laws converge.

Corollary 1. *The control policies converge to the approximate Nash equilibrium solution of the NZS game.*

Proof of Corollary 1. Based on (43), (44), (51) and (52), we have

$$u^* - \hat{u}^* = -\frac{1}{2}R_{11}^{-1}\mathcal{G}^{\mathsf{T}}(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\tilde{W}_1 - \frac{1}{2}R_{11}^{-1}\mathcal{G}^{\mathsf{T}}(\zeta)\nabla\varepsilon_1(\zeta), \tag{63}$$

$$v^* - \hat{v}^* = -\frac{1}{2}R_{22}^{-1}\mathcal{K}^{\mathsf{T}}(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\tilde{W}_2 - \frac{1}{2}R_{22}^{-1}\mathcal{K}^{\mathsf{T}}(\zeta)\nabla\varepsilon_2(\zeta). \tag{64}$$

According to Theorem 1, Assumption 1 and Assumption 3, we conclude that $\tilde{W}_i$, $i = 1, 2$, the terms $R_{11}^{-1}\mathcal{G}^{\mathsf{T}}(\zeta)\nabla\phi_1^{\mathsf{T}}(\zeta)\tilde{W}_1$, $R_{22}^{-1}\mathcal{K}^{\mathsf{T}}(\zeta)\nabla\phi_2^{\mathsf{T}}(\zeta)\tilde{W}_2$, $R_{11}^{-1}\mathcal{G}^{\mathsf{T}}(\zeta)\nabla\varepsilon_1(\zeta)$ and $R_{22}^{-1}\mathcal{K}^{\mathsf{T}}(\zeta) \times \nabla\varepsilon_2(\zeta)$ are bounded. Furthermore, we have

$$\|u^* - \hat{u}^*\| \leq \frac{1}{2}R_{11M}^{-1}\lambda_g\lambda_{d\phi_1}\mathcal{M} + \frac{1}{2}R_{11M}^{-1}\lambda_g\lambda_{d\varepsilon_1} \triangleq \lambda_u, \tag{65}$$

$$\|v^* - \hat{v}^*\| \leq \frac{1}{2}R_{22M}^{-1}\lambda_k\lambda_{d\phi_2}\mathcal{M} + \frac{1}{2}R_{22M}^{-1}\lambda_k\lambda_{d\varepsilon_2} \triangleq \lambda_v. \tag{66}$$

where λ_u and λ_v are the finite bounds. Therefore, $\|u^* - \hat{u}^*\|$ and $\|v^* - \hat{v}^*\|$ are UUB. This completes the proof. □

In addition to the convergence of system states to the origin, the tracking performance of the system is also an important indicator. Therefore, we put forward Theorem 2 to show that system (1) can track the reference trajectory (2) well, and the proof is given.

Theorem 2. *Given the cost functions* (16) *and* (17)*, for the nominal augmented system* (15)*, the approximate optimal control laws obtained by* (51) *and* (52) *ensure that the tracking error dynamics are UUB.*

Proof. See the Appendix A. □

Remark 3. *In this section, we give an optimal robust tracking control scheme for the NZS game, which can be extended to the N-player NZS game system in theory.*

4. Simulation

4.1. Two-Player Linear Non-Zero-Sum Game

Consider a continuous-time uncertain linear system:

$$\dot{x} = \begin{bmatrix} x_2 \\ -3x_1 - 0.5x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + \begin{bmatrix} 0 \\ 2 \end{bmatrix} v + \begin{bmatrix} \eta_1 x_2 cos x_1 \\ \eta_2 x_1 sin x_2 \end{bmatrix}, \tag{67}$$

where $x = [x_1, x_2]^{\mathsf{T}} \in \mathbb{R}^2$ is the state variable, $u \in \mathbb{R}$ and $v \in \mathbb{R}$ are the control inputs and the uncertain parameters $\eta_1, \eta_2 \in [-1, 1]$. The last term of system (67) is the uncertain term that is bounded by $\lambda_f(\zeta) = \sqrt{x_1^2 + x_2^2}$, then we have $\lambda_{f1}(\zeta) = \sqrt{x_2^2}$ and $\lambda_{f2}(\zeta) = \sqrt{x_1^2}$. Let the initial system state vector be $x_0 = [-1, 1]^{\mathsf{T}}$.

Here, the reference trajectory $s(t)$ is generated by the following system:

$$\dot{s} = \begin{bmatrix} -0.5s_1 - s_2 cos(s_1) \\ 3sin(s_1) - s_2 \end{bmatrix}, \tag{68}$$

where $s = [s_1, s_2]^{\mathsf{T}} \in \mathbb{R}^2$ is the reference state. One lets the initial reference state vector be $s_0 = [0.5, 0.5]^{\mathsf{T}}$.

Defining the tracking error as $e_r = x - s$ so that $\dot{e}_r = \dot{x} - \dot{s}$, let the augmented state vector be $\zeta = [e_r^{\mathsf{T}}, s^{\mathsf{T}}]^{\mathsf{T}}$. Then, we have the augmented system dynamics as follows:

$$\dot{\zeta} = \begin{bmatrix} \zeta_2 + \zeta_4 + 0.5\zeta_3 + \zeta_4 cos(\zeta_3) \\ -3(\zeta_1 + \zeta_3) - 0.5(\zeta_2 + \zeta_4) - 3sin(\zeta_3) + \zeta_4 \\ -0.5\zeta_3 - \zeta_4 cos(\zeta_3) \\ 3sin(\zeta_3) - \zeta_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix} v + \Delta\mathcal{F}(\zeta), \tag{69}$$

where $\zeta = [\zeta_1, \zeta_2, \zeta_3, \zeta_4]^{\mathsf{T}} \in \mathbb{R}^4$ with $\zeta_1 = e_{r1}$, $\zeta_2 = e_{r2}$, $\zeta_3 = s_1$, $\zeta_4 = s_2$, and $\Delta\mathcal{F}(\zeta)$ is the uncertain term of the augmented system. Here, we choose $M_1(\zeta) = [1, 0, 0, 0]^{\mathsf{T}}$ and $M_2(\zeta) = [0, 1, 0, 0]^{\mathsf{T}}$. Meanwhile, the decomposed the uncertain term are respectively $\lambda_{f1}(\zeta) = \sqrt{(\zeta_2 + \zeta_4)^2}$ and $\lambda_{f2}(\zeta) = \sqrt{(\zeta_1 + \zeta_3)^2}$. Therefore, the initial state of

the augmented system is $\zeta_0 = [-1.5, 0.5, 0.5, 0.5]^{\mathsf{T}}$ with the initial tracking error vector $e_{r0} = x_0 - s_0 = [-1.5, 0.5]$.

Select $\bar{Q}_1 = diag\{2I_2, 0_{2\times 2}\}$, $\bar{Q}_2 = diag\{I_2, 0_{2\times 2}\}$, $R_{11} = R_{21} = 1$, $R_{12} = R_{22} = 0.5$, $\eta_1 = 1$ and $\eta_2 = -1$. The critic NN activation functions are chosen as $\phi_1(\zeta) = \phi_2(\zeta) = [\zeta_1^2, \zeta_1\zeta_2, \zeta_1\zeta_3, \zeta_1\zeta_4, \zeta_2^2, \zeta_2\zeta_3, \zeta_2\zeta_4, \zeta_3^2, \zeta_3\zeta_4, \zeta_4^2]^{\mathsf{T}}$. Let the learning rates be $a = 2$ and $b = 0.5$. Moreover, one brings in a probing noise to satisfy the persistence excitation (PE) condition. The state trajectories and reference trajectories are displayed in Figures 1 and 2. After the learning process, Figures 3 and 4 show that the weights of critic NN1 and NN2 converged to $[0.2521, 0.0627, -0.0501, 0.0213, -0.0487, 0.0373, 0.0134, 0.0188, 0.0171, 0.0273]^{\mathsf{T}}$ and $[0.1934, -0.0558, 0.0248, 0.2574, 0.1487, -0.0026, -0.1406, -0.0039, -0.0134, 0.0928]^{\mathsf{T}}$. Since the value of the initial weights was all set as zero, we could conclude that the system did not require the initial stable control policies. The control trajectories for each player are in Figure 5. Figure 6 demonstrates that the tracking errors convergenced to 0, which indicated that system (67) could track the reference trajectory (68) well. To verify the robustness of the method, one could choose $\eta_1 = -0.5$ and $\eta_2 = 0.5$, and then perform the simulation and verification.The tracking error and control input are depicted in Figures 7 and 8, which still demonstrated the desired trajectory tracking performance again.

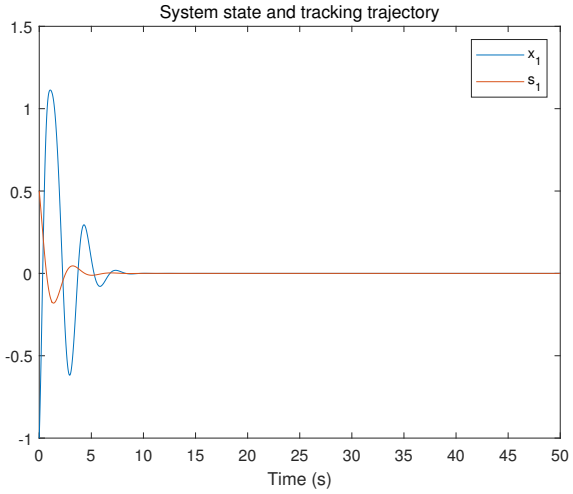

Figure 1. System state x_1 and its tracking trajectory when $\eta_1 = 1$ and $\eta_2 = -1$.

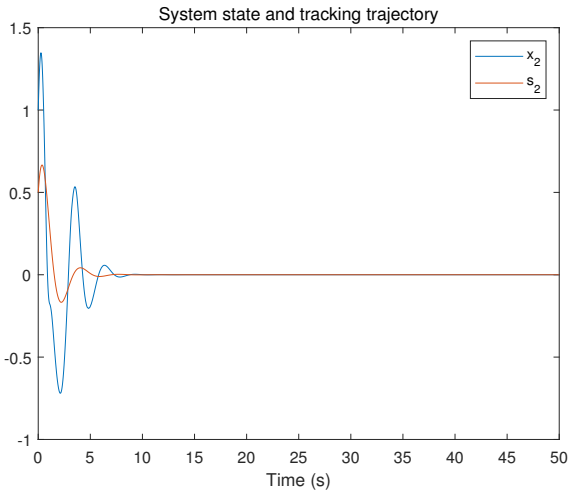

Figure 2. System state x_2 and its tracking trajectory when $\eta_1 = 1$ and $\eta_2 = -1$.

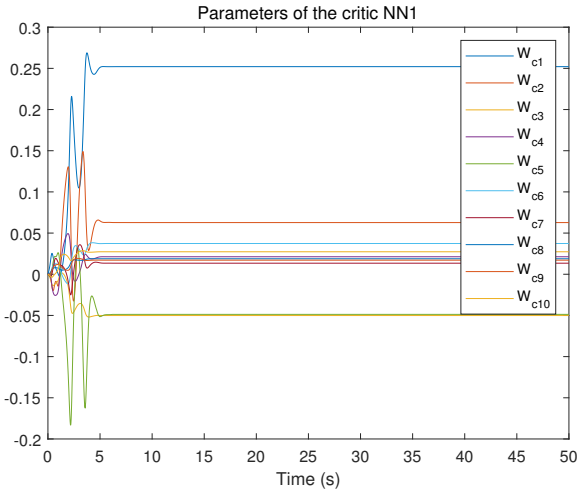

Figure 3. Convergence curves of the critic NN1 weights for player 1.

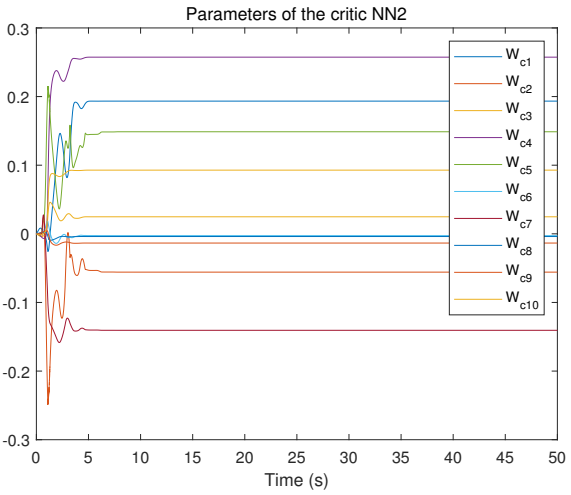

Figure 4. Convergence curves of the critic NN2 weights for player 2.

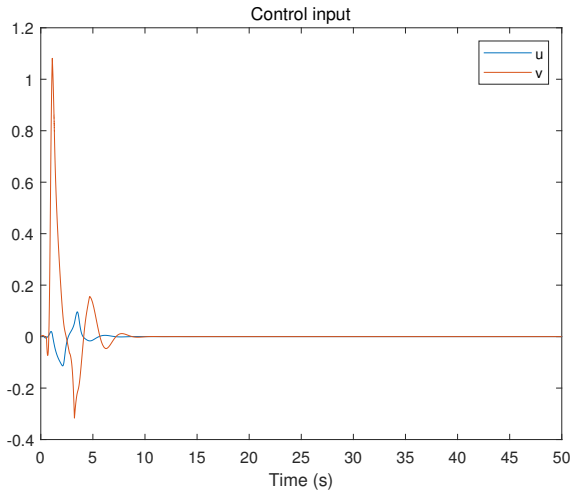

Figure 5. Control trajectories for two players when $\eta_1 = 1$ and $\eta_2 = -1$.

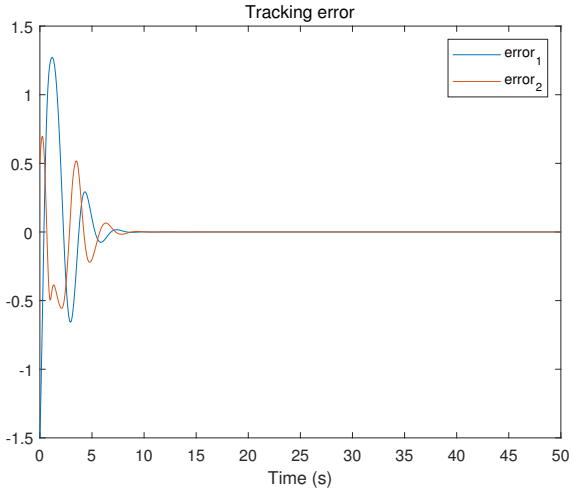

Figure 6. Tracking error trajectories when $\eta_1 = 1$ and $\eta_2 = -1$.

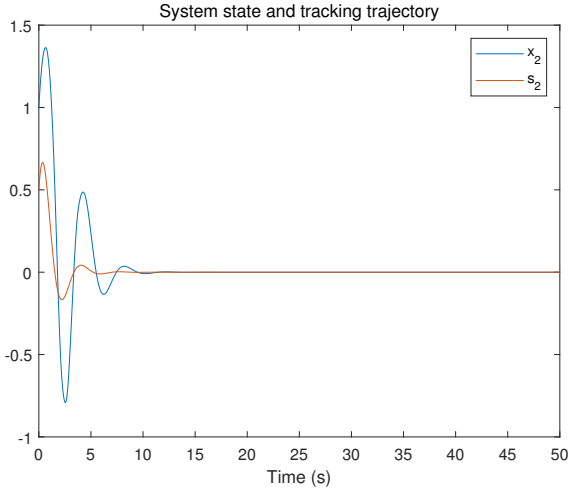

Figure 7. System state x_1 and its tracking trajectory when $\eta_1 = -0.5$ and $\eta_2 = 0.5$.

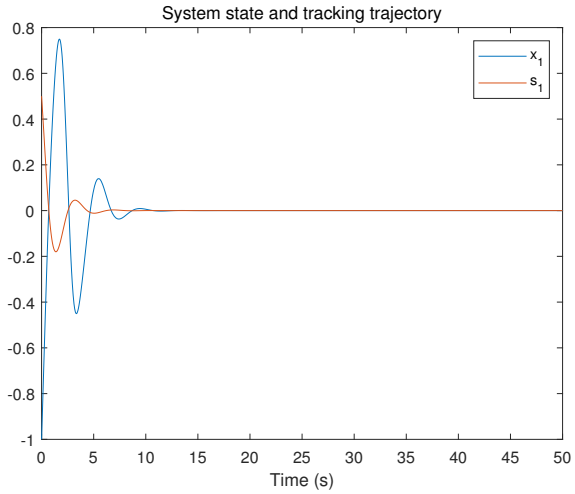

Figure 8. System state x_2 and its tracking trajectory when $\eta_1 = -0.5$ and $\eta_2 = 0.5$.

4.2. Two-Player Nonlinear Non-zero-Sum Game

Consider a continuous-time uncertain nonlinear system:

$$\dot{x} = \begin{bmatrix} x_2 \\ x_2 - 0.5x_1 - 0.25x_2(cos(2x_1) + 2)^2 \\ -0.25x_2(sin(4x_1) + 2)^2 \end{bmatrix} + \begin{bmatrix} 0 \\ cos(2x_1) + 2 \end{bmatrix} u + \begin{bmatrix} 0 \\ sin(4x_1^2) + 2 \end{bmatrix} v + \begin{bmatrix} \eta_1 x_2 cos x_1 sin x_2 \\ \eta_2 x_1 sin x_2^2 \end{bmatrix}. \tag{70}$$

In this example, the reference signal $s(t)$ is derived by

$$\dot{s} = \begin{bmatrix} -s_1 + sin(s_2) \\ -2sin^3(s_1) - 0.5s_2 \end{bmatrix}. \tag{71}$$

The critic NN activation functions, a and b are the same as in the first example. Similarly, the augmented system dynamics are as follows:

$$\dot{\zeta} = \begin{bmatrix} \zeta_2 + \zeta_4 + \zeta_3 - sin(\zeta_4) \\ \zeta_2 + \zeta_4 - 0.5(\zeta_1 + \zeta_3) - 0.25(\zeta_2 + \zeta_4)(cos(2(\zeta_1 + \zeta_3)) + 2)^2 - 0.25(\zeta_2 + \zeta_4) \\ \times (sin(4(\zeta_1 + \zeta_3)) + 2))^2 + 2sin^3(\zeta_3) + 0.5\zeta_4 \\ -\zeta_3 + sin(\zeta_4) \\ -2sin^3(\zeta_3) - 0.5\zeta_4 \end{bmatrix} + \begin{bmatrix} 0 \\ cos(2(\zeta_1 + \zeta_3)) + 2 \\ 0 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ sin(4(\zeta_1 + \zeta_3)^2) + 2 \\ 0 \\ 0 \end{bmatrix} v + \Delta\mathcal{F}(\zeta). \tag{72}$$

Here, we select $M_1(\zeta) = [1, 0, 0, 0]^\mathsf{T}$, $M_2(\zeta) = [0, 1, 0, 0]^\mathsf{T}$, $\lambda_{f1}(\zeta) = \sqrt{(\zeta_2 + \zeta_4)^2}$ and $\lambda_{f2}(\zeta) = \sqrt{(\zeta_1 + \zeta_3)^2}$. Let the initial system state vector be $x_0 = [-0.5, -0.5]^\mathsf{T}$ and the initia reference trajectory vector be $s_0 = [0.5, 0.5]^\mathsf{T}$, then the initial state of the augmented system is $\zeta_0 = [-1, -1, 0.5, 0.5]^\mathsf{T}$.

Select $\bar{Q}_1 = diag\{5I_2, 0_{2\times2}\}$, $\bar{Q}_2 = diag\{2I_2, 0_{2\times2}\}$, $R_{11} = R_{21} = 2$, $R_{12} = R_{22} = 1$, $\eta_1 = -0.2$ and $\eta_2 = 0.2$. The state trajectories and reference trajectories are displayed in Figures 9 and 10. Figures 11 and 12 show that the weights of critic NN1 and NN2 converge to $[0.4582, 0.2514, -0.2907, -0.2567, 0.1455, -0.1353, -0.1050, 0.1527, 0.1321, 0.1112]^\mathsf{T}$ and $[0.2622, 0.0666, -0.0854, -0.0858, 0.0879, -0.0610, -0.0470, 0.0601, 0.0487, 0.0406]^\mathsf{T}$, respectively. It could also be seen that initial stability control policies were not required. The control trajectories for each player are in Figure 13. The tracking errors are displayed in Figure 14, which indicated that system (70) could track the reference trajectory (71) well. These experimental results verified the effectiveness of the proposed method in this paper.

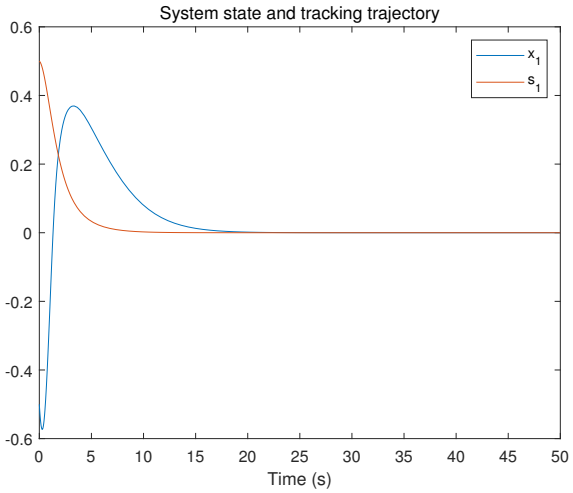

Figure 9. System state x_1 and its tracking trajectory when $\eta_1 = -0.2$ and $\eta_2 = 0.2$.

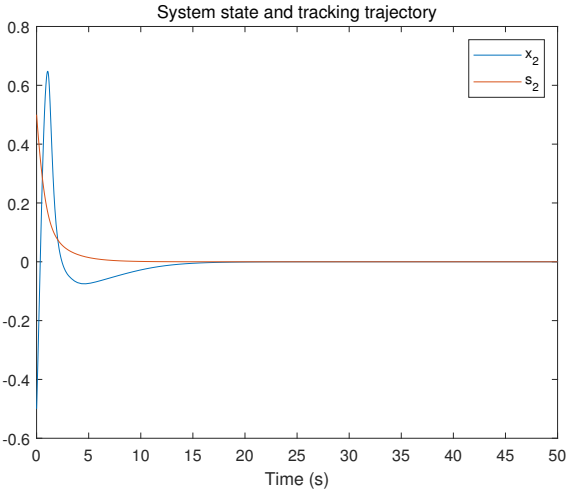

Figure 10. System state x_2 and its tracking trajectory when $\eta_1 = -0.2$ and $\eta_2 = 0.2$.

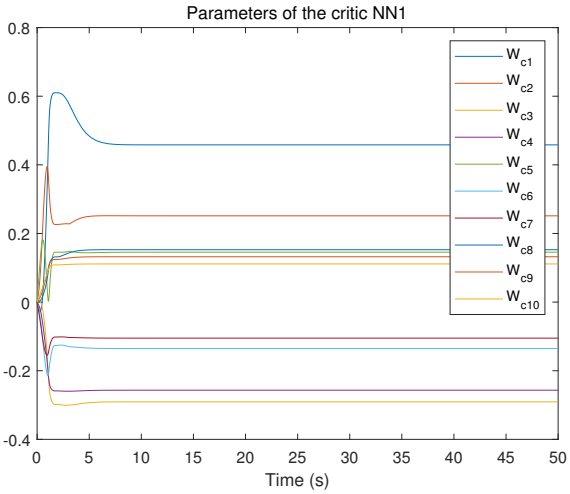

Figure 11. Convergence curves of critic NN1 weights for player 1.

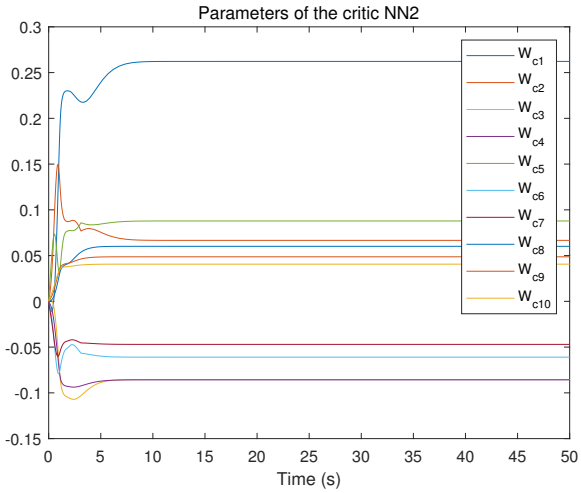

Figure 12. Convergence curves of critic NN2 weights for player 2.

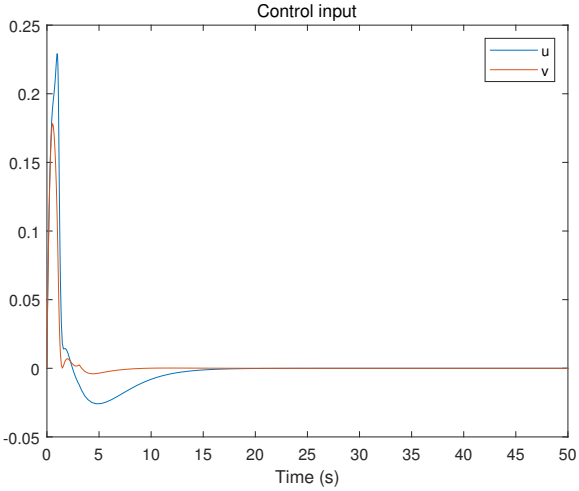

Figure 13. Control trajectories for two players when $\eta_1 = -0.2$ and $\eta_2 = 0.2$.

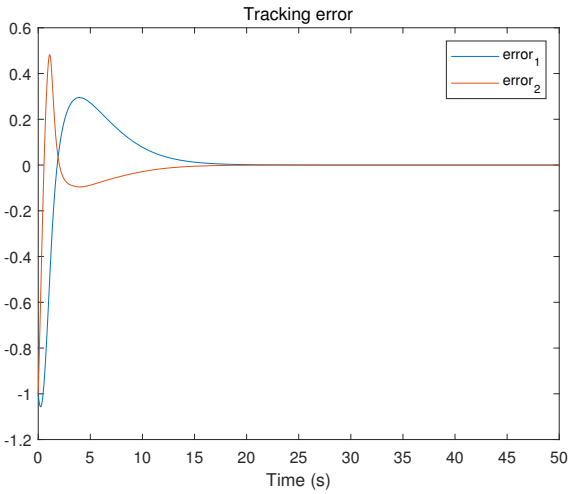

Figure 14. Tracking error trajectories when $\eta_1 = -0.2$ and $\eta_2 = 0.2$.

5. Conclusions

In this paper, an ADP-based robust tracking control design was proposed for the NZS game of nonlinear systems with dynamic uncertainties. Firstly, the tracking error and reference trajectory were used to construct the augmented system. The coupled HJB equations were modified by defining appropriate performance indicators. Then, a new adaptive critic design was proposed to solve the coupled HJB equations. A single-network structure was used to approximate the value function and control policy for each player. By a modified critic NN weights' tuning law, the control policies of the two players converged to the Nash equilibrium of NZS games. What is more, the proof that the system state, tracking error and weight estimation error were UUB was given via the Lyapunov theory. Finally, two simulation results verified the effectiveness of the proposed scheme. We will consider the input constraints and state constraints for this problem in the future.

Author Contributions: C.Q. and Z.S. provided methodology, validation, and writing—original draft preparation; Z.Z. and D.Z.provided conceptualization, writing—review; J.Z. provided supervision; C.Q. provided funding support. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation of China under Grant (U1504615, 61703141), Youth Backbone Teachers in Colleges and Universities of Henan Province 2018GGJS017, and Science and Technology Research Project of the Henan Province 222102240014.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors can confirm that all relevant data are included in the article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Theorem 1. We choose the following Lyapunov function candidate:

$$L = \frac{1}{2a}\tilde{W}_1^{\mathrm{T}} W_1 + \frac{1}{2a}\tilde{W}_2^{\mathrm{T}} W_2 + \frac{b}{a} J_s(\zeta), \tag{A1}$$

where $J_s(\zeta)$ is presented in Assumption 3. Let $\sigma_{ii}^{'} = -\sigma_{ii}, \bar{\sigma}_{ii}^{'} = -\bar{\sigma}_{ii}, m_{sii} = 1 + \sigma_{ii}^{\mathrm{T}}\sigma_{ii}, \bar{\sigma}_{ii} = \sigma_{ii}/m_{sii}, i = 1, 2$, combining (47), (48), (51), (52), (58) and (59), we obtain the weight estimation error dynamics as

$$\begin{aligned}
\dot{\tilde{W}}_1 = & -a\frac{\bar{\sigma}_{11}^{'}}{m_{s11}}(\sigma_{11}^{'\mathrm{T}}\tilde{W}_1 + \frac{1}{4}W_1^{\mathrm{T}}\nabla\phi_1 A_1(\zeta)\nabla\phi_1^{\mathrm{T}}W_1 - \frac{1}{2}\tilde{W}_1^{\mathrm{T}}\nabla\phi_1 A_1(\zeta)\nabla\phi_1^{\mathrm{T}}W_1 \\
& - \frac{1}{4}W_1^{\mathrm{T}}\nabla\phi_1 B_1(\zeta)\nabla\phi_1^{\mathrm{T}}W_1 + \frac{1}{2}\tilde{W}_1^{\mathrm{T}}\nabla\phi_1 B_1(\zeta)\nabla\phi_1^{\mathrm{T}}W_1 - \frac{1}{4}W_2^{\mathrm{T}}\nabla\phi_2 C_1(\zeta)\nabla\phi_2^{\mathrm{T}}W_2 \\
& + \frac{1}{2}\tilde{W}_2^{\mathrm{T}}\nabla\phi_2 C_1(\zeta)\nabla\phi_2^{\mathrm{T}}W_2 - \frac{1}{2}\tilde{W}_2^{\mathrm{T}}\nabla\phi_2 C_1(\zeta)\nabla\phi_1^{\mathrm{T}}W_2 + \frac{1}{2}W_1^{\mathrm{T}}\nabla\phi_1 D_1(\zeta)\nabla\phi_2^{\mathrm{T}}\tilde{W}_2 + \varepsilon_{HJ_1}) \\
& - \frac{b}{2}\Pi\nabla\phi_1 B_1(\zeta)\nabla J_s(\zeta),
\end{aligned} \tag{A2}$$

$$\begin{aligned}
\dot{\tilde{W}}_2 = & -a\frac{\bar{\sigma}_{22}^{'}}{m_{s22}}(\sigma_{22}^{'\mathrm{T}}\tilde{W}_2 + \frac{1}{4}W_2^{\mathrm{T}}\nabla\phi_2 A_2(\zeta)\nabla\phi_2^{\mathrm{T}}W_2 - \frac{1}{2}\tilde{W}_2^{\mathrm{T}}\nabla\phi_2 A_2(\zeta)\nabla\phi_1^{\mathrm{T}}W_2 \\
& - \frac{1}{4}W_2^{\mathrm{T}}\nabla\phi_2 B_2(\zeta)\nabla\phi_2^{\mathrm{T}}W_2 + \frac{1}{2}\tilde{W}_2^{\mathrm{T}}\nabla\phi_2 B_2(\zeta)\nabla\phi_2^{\mathrm{T}}W_2 - \frac{1}{4}W_1^{\mathrm{T}}\nabla\phi_1 C_2(\zeta)\nabla\phi_1^{\mathrm{T}}W_1 \\
& + \frac{1}{2}\tilde{W}_1^{\mathrm{T}}\nabla\phi_1 C_2(\zeta)\nabla\phi_1^{\mathrm{T}}W_1 - \frac{1}{2}\tilde{W}_1^{\mathrm{T}}\nabla\phi_1 C_2(\zeta)\nabla\phi_2^{\mathrm{T}}W_1 + \frac{1}{2}W_2^{\mathrm{T}}\nabla\phi_2 D_2(\zeta)\nabla\phi_1^{\mathrm{T}}\tilde{W}_1 + \varepsilon_{HJ_2}) \\
& - \frac{b}{2}\Pi\nabla\phi_1 B_2(\zeta)\nabla J_s(\zeta).
\end{aligned} \tag{A3}$$

Based on (A2) and (A3), the derivation of L can be rewritten as

$$\begin{aligned}\dot{L} &= \frac{1}{a}\tilde{W}_1^{\mathsf{T}}\dot{\tilde{W}}_1 + \frac{1}{a}\tilde{W}_2^{\mathsf{T}}\dot{\tilde{W}}_2 + \frac{b}{a}(\nabla J_s(\zeta))^{\mathsf{T}}\dot{\zeta} \\ &= -\tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\bar{\sigma}_{11}'^{\mathsf{T}}\tilde{W}_1 - \tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\frac{1}{4m_{s11}}W_1^{\mathsf{T}}\nabla\phi_1 A_1(\zeta)\nabla\phi_1^{\mathsf{T}}W_1 + \tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\frac{1}{2m_{s11}}\tilde{W}_1^{\mathsf{T}}\nabla\phi_1 A_1(\zeta)\nabla\phi_1^{\mathsf{T}} \\ &\times W_1 + \tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\frac{1}{4m_{s11}}W_1^{\mathsf{T}}\nabla\phi_1 B_1(\zeta)\nabla\phi_1^{\mathsf{T}}W_1 - \tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\frac{1}{2m_{s11}}\tilde{W}_1^{\mathsf{T}}\nabla\phi_1 B_1(\zeta)\nabla\phi_1^{\mathsf{T}}W_1 + \tilde{W}_1^{\mathsf{T}} \\ &\times \bar{\sigma}_{11}'\frac{1}{4m_{s11}}W_2^{\mathsf{T}}\nabla\phi_2 C_1(\zeta)\nabla\phi_2^{\mathsf{T}}W_2 - \tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\frac{1}{2m_{s11}}\tilde{W}_2^{\mathsf{T}}\nabla\phi_2 C_1(\zeta)\nabla\phi_1^{\mathsf{T}}W_2 - \tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\frac{1}{2m_{s11}} \\ &\times W_1^{\mathsf{T}}\nabla\phi_1 D_1\nabla\phi_2^{\mathsf{T}}\tilde{W}_2 - \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\bar{\sigma}_{22}'^{\mathsf{T}}\tilde{W}_2 - \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\frac{1}{4m_{s22}}W_2^{\mathsf{T}}\nabla\phi_2 A_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_2 + \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}' \\ &\times \frac{1}{2m_{s22}}\tilde{W}_2^{\mathsf{T}}\nabla\phi_2 A_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_2 + \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\frac{1}{4m_{s22}}W_2^{\mathsf{T}}\nabla\phi_2 B_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_2 - \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\frac{1}{2m_{s22}} \\ &\times \tilde{W}_2^{\mathsf{T}}\nabla\phi_2 B_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_2 + \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\frac{1}{4m_{s22}}W_1^{\mathsf{T}}\nabla\phi_1 C_2(\zeta)\nabla\phi_1^{\mathsf{T}}W_1 - \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\frac{1}{2m_{s22}}\tilde{W}_1^{\mathsf{T}}\nabla\phi_1 \\ &\times C_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_1 - \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\frac{1}{2m_{s22}}W_1^{\mathsf{T}}\nabla\phi_2 D_2\nabla\phi_1^{\mathsf{T}}\tilde{W}_1 - \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}'\frac{1}{m_{s22}}\varepsilon_{HJ_2} - \frac{b}{2}\Pi\tilde{W}_2\nabla\phi_2 \\ &\times B_2(\zeta)\nabla J_s(\zeta) - \tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}'\frac{1}{m_{s11}}\varepsilon_{HJ_1} - \frac{b}{2}\Pi\tilde{W}_1\nabla\phi_1 B_1(\zeta)\nabla J_s(\zeta).\end{aligned} \tag{A4}$$

Defining $p = [\tilde{W}_1^{\mathsf{T}}\bar{\sigma}_{11}', \tilde{W}_2^{\mathsf{T}}\bar{\sigma}_{22}', \tilde{W}_1^{\mathsf{T}}, \tilde{W}_2^{\mathsf{T}}]^{\mathsf{T}}$, the derivation of L can be rewritten as

$$L = -p^{\mathsf{T}}\begin{bmatrix} N_{11} & N_{12} & N_{13} & N_{14} \\ N_{21} & N_{22} & N_{23} & N_{24} \\ N_{31} & N_{32} & N_{33} & N_{34} \\ N_{41} & N_{42} & N_{43} & N_{44} \end{bmatrix} p + p^{\mathsf{T}}\psi \tag{A5}$$

where

$$\begin{aligned} N_{11} = N_{22} &= I, \\ N_{12} = N_{21} = N_{33} = N_{34} = N_{43} &= N_{44} = 0, \\ N_{13} = N_{31}^{\mathsf{T}} &= \frac{W_1^{\mathsf{T}}}{4m_{s11}}(\nabla\phi_1 B_1(\zeta)\nabla\phi_1^{\mathsf{T}} - \nabla\phi_1 A_1(\zeta)\nabla\phi_1^{\mathsf{T}}), \\ N_{14} = N_{41}^{\mathsf{T}} &= \frac{1}{4m_{s11}}W_2^{\mathsf{T}}\nabla\phi_2 C_1(\zeta)\nabla\phi_2^{\mathsf{T}} - \frac{1}{4m_{s11}}W_2^{\mathsf{T}}\nabla\phi_1 C_1(\zeta)\nabla\phi_2^{\mathsf{T}} \\ &\quad + \frac{1}{4m_{s11}}W_1^{\mathsf{T}}\nabla\phi_1 D_1(\zeta)\nabla\phi_2^{\mathsf{T}}, \\ N_{23} = N_{32}^{\mathsf{T}} &= \frac{1}{4m_{s11}}W_1^{\mathsf{T}}\nabla\phi_1 C_2(\zeta)\nabla\phi_1^{\mathsf{T}} - \frac{1}{4m_{s11}}W_1^{\mathsf{T}}\nabla\phi_2 C_2(\zeta)\nabla\phi_1^{\mathsf{T}} \\ &\quad + \frac{1}{4m_{s11}}W_2^{\mathsf{T}}\nabla\phi_2 D_2(\zeta)\nabla\phi_1^{\mathsf{T}}, \\ N_{24} = N_{42}^{\mathsf{T}} &= \frac{W_2^{\mathsf{T}}}{4m_{s22}}(\nabla\phi_2 B_2(\zeta)\nabla\phi_2^{\mathsf{T}} - \nabla\phi_2 A_2(\zeta)\nabla\phi_2^{\mathsf{T}}), \end{aligned}$$

and the vector $\psi = [\psi_1, \psi_2, \psi_3, \psi_4]^{\mathsf{T}}$ is given by

$$\begin{aligned}\psi_1 = &\frac{1}{4m_{s11}}(W_1^{\mathsf{T}}\nabla\phi_1 A_1(\zeta)\nabla\phi_1^{\mathsf{T}}W_1 + W_1^{\mathsf{T}}\nabla\phi_1 B_1(\zeta)\nabla\phi_1^{\mathsf{T}}W_1 \\ &+ W_2^{\mathsf{T}}\nabla\phi_2 C_1(\zeta)\nabla\phi_1^{\mathsf{T}}W_2) - \frac{1}{m_{s11}}\varepsilon_{HJ_1},\end{aligned}$$

$$\psi_2 = \frac{1}{4m_{s22}}(W_2^{\mathsf{T}}\nabla\phi_2 A_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_2 + W_2^{\mathsf{T}}\nabla\phi_2 B_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_2 + W_1^{\mathsf{T}}\nabla\phi_1 C_2(\zeta)\nabla\phi_2^{\mathsf{T}}W_1) - \frac{1}{m_{s22}}\varepsilon_{HJ_2},$$
$$\psi_3 = \psi_4 = 0.$$

According to Assumption 3 and the fact that $\sigma_{ii}, i = 1, 2$ are bounded, we derive that ψ is bounded. Selecting the appropriate parameters such that $N > 0$, one lets $\lambda_{min}(N)$ denote the minimum eigenvalue of N and ψ be bounded by ψ_M, We can conclude that

$$\begin{aligned}\dot{L} \leq & -\lambda_{min}(N)\|p\|^2 + \psi_M\|p\| - \frac{b}{2}\Pi\tilde{W}_1\nabla\phi_1 B_1(\zeta)\nabla J_s(\zeta) \\ & - \frac{b}{2}\Pi\tilde{W}_2\nabla\phi_1 B_2(\zeta)\nabla J_s(\zeta) + \frac{b}{a}(\nabla J_s(\zeta))^{\mathsf{T}}\dot{\zeta}.\end{aligned} \tag{A6}$$

In the following, the cases of $\Pi = 0$ and $\Pi = 1$ will be considered.

Case 1. $\Pi = 0$. *Since* $\nabla J_s(\zeta)^{\mathsf{T}}\dot{\zeta} < 0$, *we have* $-\nabla J_s(\zeta)^{\mathsf{T}}\dot{\zeta} > 0$. *According to the density property of real numbers, there exists a positive constant* λ_1 *such that* $0 < \lambda_1\|\nabla J_s(\zeta)\| \leq -(\nabla J_s(\zeta))^{\mathsf{T}}\dot{\zeta}$ *holds for all* $\zeta \in \Omega$, *i.e.,*$(\nabla J_s(\zeta))^{\mathsf{T}}\dot{\zeta} \leq -\lambda_1\|\nabla J_s(\zeta)\|$. *Hence, the inequality* (A6) *becomes*

$$\dot{L} \leq -\lambda_{min}(N)\|p\|^2 + \psi_M\|p\| - \frac{b}{a}\lambda_1\|\nabla J_s(\zeta)\|. \tag{A7}$$

Therefore, given that the following inequalities

$$\|p\| \geq \frac{\psi_M}{\lambda_{min}(N)} \triangleq \mathcal{M}_1 \tag{A8}$$

or

$$\|\nabla J_s(\zeta)\| \geq \frac{b\psi_M^2}{4a\lambda_{min}(N)\lambda_1} \triangleq \mathcal{N}_1 \tag{A9}$$

hold, we conclude $\dot{L} < 0$.

Case 2. $\Pi = 1$. *Adding and subtracting* $b(\nabla J_s(\zeta))^{\mathsf{T}}B_1(\zeta)\nabla\varepsilon_1(\zeta)/(2a)$ *and* $b(\nabla J_s(\zeta))^{\mathsf{T}}B_2(\zeta)\times\nabla\varepsilon_2(\zeta)/(2a)$ *to the right hand side of* (A6), *meanwhile taking Assumption 1 and Assumption 4 into consideration, we can conclude that*

$$\begin{aligned}\dot{L} \leq & -\lambda_{min}(N)\|p\|^2 + \psi_M\|p\| - \frac{b}{2}\Pi\tilde{W}_1\nabla\phi_1 B_1(\zeta)\nabla J_s(\zeta) - \frac{b}{2}\Pi\tilde{W}_2\nabla\phi_1 B_2(\zeta)\nabla J_s(\zeta) \\ & + \frac{b}{a}(\nabla J_s(\zeta))^{\mathsf{T}}\dot{\zeta} \\ = & -\lambda_{min}(N)\|p\|^2 + \psi_M\|p\| + \frac{b}{a}(\nabla J_s(\zeta))^{\mathsf{T}}(\mathcal{F}(\zeta) + \mathcal{G}(\zeta)u^* + \mathcal{K}(\zeta)v^*) + \frac{b}{2a}(\nabla J_s(\zeta))^{\mathsf{T}} \\ & \times B_1(\zeta)\nabla\varepsilon_1(\zeta) + \frac{b}{2a}(\nabla J_s(\zeta))^{\mathsf{T}}B_2(\zeta)\nabla\varepsilon_2(\zeta) \\ \leq & -\lambda_{min}(N)\|p\|^2 + \psi_M\|p\| - \frac{b}{a}\lambda_m\|\nabla J_s(\zeta)\|^2 + \frac{b}{2a}(\lambda_2 + \lambda_3)\|\nabla J_s(\zeta)\|.\end{aligned} \tag{A10}$$

Therefore, given that the following inequalities

$$\|p\| \geq \sqrt{\frac{\psi_M^2}{4\lambda_{min}^2(N)} + \frac{b(\lambda_2+\lambda_3)^2}{16a\lambda_{min}(N)\lambda_m}} + \frac{\psi_M}{2\lambda_{min}(N)} \triangleq \mathcal{M}_2 \tag{A11}$$

or

$$\|\nabla J_s(\zeta)\| \geq \sqrt{\frac{a\psi_M^2}{4\lambda_{min}(N)\lambda_m} + \frac{(\lambda_2+\lambda_3)^2}{16\lambda_m^2}} + \frac{\lambda_2+\lambda_3}{4\lambda_m} \triangleq \mathcal{N}_2 \tag{A12}$$

hold, we conclude $\dot{L} < 0$.

To summarize, if the inequality $\|p\| > max(\mathcal{M}_1, \mathcal{M}_2) = \mathcal{M}$ or $\|\nabla J_s(\zeta)\| > max(\mathcal{N}_1, \mathcal{N}_2) = \mathcal{N}$ holds, then $\dot{L} < 0$ and we have that system state and the weight estimation errors are UUB. This completes the proof. □

Proof of Theorem 2. We choose the following Lyapunov function candidate:

$$L_1 = V_1 + V_2. \tag{A13}$$

Differentiating L_1 along ζ, we have

$$\begin{aligned}\dot{V}_1 = &-\zeta^{\mathsf{T}}\bar{Q}_1\zeta - [\lambda_{f1}^2(\zeta) + \frac{1}{4}W_1^{\mathsf{T}}\nabla\phi_1 A_1 \nabla\phi_1^{\mathsf{T}} W_1 + \Delta\mathcal{F}(\zeta)^{\mathsf{T}}\Delta\mathcal{F}(\zeta)] - \frac{1}{4}W_1^{\mathsf{T}}\nabla\phi_1 B_1 \nabla\phi_1^{\mathsf{T}} W_1 \\ &+ \frac{1}{2}W_1^{\mathsf{T}}\nabla\phi_1 B_1 \nabla\phi_1^{\mathsf{T}}\tilde{W}_1 - \frac{1}{4}W_2^{\mathsf{T}}\nabla\phi_2 C_1 \nabla\phi_2^{\mathsf{T}} W_2 + \frac{1}{2}W_1^{\mathsf{T}}\nabla\phi_1 D_1 \nabla\phi_2^{\mathsf{T}}\tilde{W}_2 - [\nabla\phi_1^{\mathsf{T}} W_1 \\ &- \Delta\mathcal{F}(\zeta)]^{\mathsf{T}}[\nabla\phi_1^{\mathsf{T}} W_1 - \Delta\mathcal{F}(\zeta)] + \varepsilon_{HJ_1} + \varepsilon_{b1} + \varepsilon_{\mathcal{F}1},\end{aligned} \tag{A14}$$

where $\varepsilon_{\mathcal{F}1} = \nabla\varepsilon_1\Delta\mathcal{F}(\zeta)$, since $\nabla\varepsilon_1$ and $\Delta\mathcal{F}(\zeta)$ are bounded, let $\varepsilon_{\mathcal{F}1} \leq \lambda_{\varepsilon_{\mathcal{F}1}}$. $\dot{V}_2$ is similarly as $\dot{V}_1$, it is not hard to see that

$$\begin{aligned}\dot{L}_1 = \dot{V}_1 + \dot{V}_2 &\leq -\zeta^{\mathsf{T}}(\bar{Q}_1 + \bar{Q}_2)\zeta - q^{\mathsf{T}} Y q + \lambda_4 \\ &\leq -\lambda_{min}(\bar{Q}_1 + \bar{Q}_2)\|\zeta\|^2 - \lambda_{min}(Y)\|q\|^2 + \lambda_4,\end{aligned} \tag{A15}$$

where $\varepsilon_{HJ_1} + \varepsilon_{b1} + \varepsilon_{HJ_2} + \varepsilon_{b2} + \varepsilon_{\mathcal{F}1} + \varepsilon_{\mathcal{F}2} \leq \lambda_{\varepsilon_{HJ_1}} + \lambda_{\varepsilon_{HJ_2}} + \lambda_{\varepsilon_{b1}} + \lambda_{\varepsilon_{b2}} + \lambda_{\varepsilon_{\mathcal{F}1}} + \lambda_{\varepsilon_{\mathcal{F}2}} = \lambda_4$, $q = [W_1^{\mathsf{T}}, W_2^{\mathsf{T}}, \tilde{W}_1, \tilde{W}_2]^{\mathsf{T}}$, $\lambda_{min}(\bar{Q}_1 + \bar{Q}_2)$ and $\lambda_{min}(Y)$ are the minimum eigenvalues of $\bar{Q}_1 + \bar{Q}_2$ and Y, respectively. In the top formula,

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{21} & Y_{22} & Y_{23} & Y_{24} \\ Y_{31} & Y_{32} & Y_{33} & Y_{34} \\ Y_{41} & Y_{42} & Y_{43} & Y_{44} \end{bmatrix}$$

and

$$\begin{aligned} Y_{11} &= \frac{1}{4}\nabla\phi_1 B_1 \nabla\phi_1^{\mathsf{T}} + \frac{1}{4}\nabla\phi_1 C_2 \nabla\phi_1^{\mathsf{T}}, \\ Y_{22} &= \frac{1}{4}\nabla\phi_2 B_2 \nabla\phi_2^{\mathsf{T}} + \frac{1}{4}\nabla\phi_2 C_1 \nabla\phi_2^{\mathsf{T}}, \\ Y_{13} &= Y_{31}^{\mathsf{T}} = -\frac{1}{4}\nabla\phi_1 B_1 \nabla\phi_1^{\mathsf{T}}, \\ Y_{14} &= Y_{41}^{\mathsf{T}} = -\frac{1}{4}\nabla\phi_1 D_1 \nabla\phi_2^{\mathsf{T}}, \\ Y_{23} &= Y_{32}^{\mathsf{T}} = -\frac{1}{4}\nabla\phi_2 D_2 \nabla\phi_1^{\mathsf{T}}, \\ Y_{24} &= Y_{42}^{\mathsf{T}} = -\frac{1}{4}\nabla\phi_2 B_2 \nabla\phi_2^{\mathsf{T}}. \end{aligned}$$

Therefore, if the following inequalities

$$\|\zeta\| \geq \sqrt{\frac{\lambda_4}{\lambda_{min}(\bar{Q}_1 + \bar{Q}_2)}} \triangleq \mathcal{C}_1 \tag{A16}$$

or

$$\|q\| \geq \sqrt{\frac{\lambda_4}{\lambda_{min}(Y)}} \triangleq \mathcal{C}_2 \tag{A17}$$

hold, we obtain $\dot{L}_1 < 0$.

To summarize, if the inequality $\|\zeta\| > \mathcal{C}_1$ or $\|q\| > \mathcal{C}_2$ holds, then $\dot{L}_1 < 0$ and we have that the tracking errors of the closed-loop uncertain augmented system are UUB. This completes the proof. □

References

1. Namadchian, Z.; Zare, A. Stability analysis of dynamic nonlinear interval type-2 TSK fuzzy control systems based on describing function. *Soft Comput.* **2020**, *24*, 14623–14636. [CrossRef]
2. Tavoosi, J.; Suratgar, A.A.; Menhaj, M.B.; Mosavi, A.; Mohammadzadeh, A.; Ranjbar, E. Modeling Renewable Energy Systems by a Self-Evolving Nonlinear Consequent Part Recurrent Type-2 Fuzzy System for Power Prediction. *Sustainability* **2021**, *13*, 3301. [CrossRef]
3. Zhang, H.; Hong, Q.; Yan, H.; Yang, F.; Guo, G. Event-Based Distributed H_∞ Filtering Networks of 2-DOF Quarter-Car Suspension Systems. *IEEE Trans. Ind. Inform.* **2017**, *13*, 312–321. [CrossRef]
4. Li, L.; Xiao, J.; Zhao, Y.; Liu, K.; Peng, X.; Luan, H.; Li, K. Robust position anti-interference control for PMSM servo system with uncertain disturbance. *CES Trans. Electr. Mach. Syst.* **2020**, *4*, 151–160. [CrossRef]
5. Liu, D.; Wang, D.; Wang, F.-Y.; Li, H.; Yang, X. Neural-Network-Based Online HJB Solution for Optimal Robust Guaranteed Cost Control of Continuous-Time Uncertain Nonlinear Systems. *IEEE Trans. Cybern.* **2014**, *44*, 2834–2847. [CrossRef]
6. Zhong, X.; He, H.; Prokhorov, D.V. Robust controller design of continuous-time nonlinear system using neural network. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–8.
7. Sun, J.; Liu, C.; Ye, Q. Robust differential game guidance laws design for uncertain interceptor-target engagement via adaptive dynamic programming. *Int. J. Control* **2017**, *90*, 990–1004. [CrossRef]
8. Yang, X.; Liu, D.; Luo, B.; Li, C. Data-based robust adaptive control for a class of unknown nonlinear constrained-input systems via integral reinforcement learning. *Inf. Sci.* **2016**, *369*, 731–747. [CrossRef]
9. Yang, X.; He, H. Adaptive Critic Designs for Event-Triggered Robust Control of Nonlinear Systems With Unknown Dynamics. *IEEE Trans. Cybern.* **2019**, *49*, 2255–2267. [CrossRef]
10. Wang, X.; Ye, X. Optimal Robust Control of Nonlinear Uncertain System via Off-Policy Integral Reinforcement Learning. In Proceedings of the 2020 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020; pp. 1928–1933.
11. Vamvoudakis, K.G.; Lewis, F.L. Online actor critic algorithm to solve the continuous-time infinite horizon optimal control problem. In Proceedings of the 2009 International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009; pp. 3180–3187.
12. Dierks, T.; Jagannathan, S. 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 (CDC), Atlanta, GA, USA, 15–17 December 2010; pp. 3048–3053.
13. Lv, Y.; Na, J.; Yang, Q.; Wu, X.; Guo, Y. Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics. *Int. J. Control* **2016**, *89*, 99–112. [CrossRef]
14. Wang, D.; He, H.; Liu, D. Adaptive Critic Nonlinear Robust Control: A Survey. *IEEE Trans. Cybern.* **2017**, *47*, 3429–3451. [CrossRef]
15. Wang, D.; Liu, D.; Zhang, Q.; Zhao, D. Data-Based Adaptive Critic Designs for Nonlinear Robust Optimal Control With Uncertain Dynamics. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *46*, 1544–1555. [CrossRef]
16. Sun, J.; Zhang, H.; Yan, Y.; Xu, S.; Fan, X. Optimal Regulation Strategy for Nonzero-Sum Games of the Immune System Using Adaptive Dynamic Programming. *IEEE Trans. Cybern.* **2021**, *47*, 1–10. [CrossRef]
17. Narayanan, V.; Sahoo, A.; Jagannathan, S.; George, K. Approximate Optimal Distributed Control of Nonlinear Interconnected Systems Using Event-Triggered Nonzero-Sum Games. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 1512–1522. [CrossRef]
18. Morris, P. *Introduction to Game Theory*, 1st ed.; Springer: New York, NY, USA, 1994; pp. 115–147.
19. Starr, A.W.; Ho, Y.C. Nonzero-sum differential games. *J. Optim. Theory Appl.* **1969**, *3*, 184–206. [CrossRef]
20. Zhang, H.; Jiang, H.; Luo, C.; Xiao, G. Discrete-Time Nonzero-Sum Games for Multiplayer Using Policy-Iteration-Based Adaptive Dynamic Programming Algorithms. *IEEE Trans. Cybern.* **2017**, *47*, 3331–3340. [CrossRef] [PubMed]
21. Mu, C.; Wang, K.; Sun, C. Policy-Iteration-Based Learning for Nonlinear Player Game Systems with Constrained Inputs. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 6488–6502. [CrossRef]
22. Zhang, Q.; Zhao, D. Data-Based Reinforcement Learning for Nonzero-Sum Games with Unknown Drift Dynamics. *IEEE Trans. Cybern.* **2019**, *49*, 2874–2885. [CrossRef] [PubMed]
23. Song, R.; Lewis, F.L.; Wei, Q. Off-Policy Integral Reinforcement Learning Method to Solve Nonlinear Continuous-Time Multiplayer Nonzero-Sum Games. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 704–713. [CrossRef] [PubMed]
24. Zhang, H.; Su, H.; Zhang, K.; Luo, Y. Event-Triggered Adaptive Dynamic Programming for Non-Zero-Sum Games of Unknown Nonlinear Systems via Generalized Fuzzy Hyperbolic Models. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 2202–2214. [CrossRef]
25. Zhao, Q.; Sun, J.; Wang, G.; Chen, J. Event-Triggered ADP for Nonzero-Sum Games of Unknown Nonlinear Systems. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *27*, 1–9. [CrossRef]
26. Zhao, D.; Zhang, Q.; Wang, D.; Zhu, Y. Experience Replay for Optimal Control of Nonzero-Sum Game Systems with Unknown Dynamics. *IEEE Trans. Cybern.* **2016**, *46*, 854–865. [CrossRef] [PubMed]

27. Zhang, C.; Zhang, Z.. Adaptive Iterative Learning Trajectory Tracking Control of SCARA Robot. In Proceedings of the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 18–20 June 2021; pp. 910–914.
28. Yang, Y.; Wan, Y.; Zhu, J.; Lewis, F.L. H_∞ Tracking Control for Linear Discrete-Time Systems: Model-Free Q-Learning Designs. *IEEE Control. Syst. Lett.* **2021**, *5*, 175–180. [CrossRef]
29. Huang, Y.; Liu, D. Neural-network-based optimal tracking control scheme for a class of unknown discrete-time nonlinear systems using iterative ADP algorithm. *Neurocomputing* **2014**, *125*, 46–56. [CrossRef]
30. Dierks, T.; Jagannathan, S. Non-zero sum games: Optimal tracking control of affine nonlinear discrete-time systems with unknown internal dynamics. In Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference, Shanghai, China, 29 January 2010; pp. 6750–6755.
31. Wang, D.; Mu, C. Adaptive-Critic-Based Robust Trajectory Tracking of Uncertain Dynamics and Its Application to a Spring–Mass–Damper System. *IEEE Trans. Ind. Electron.* **2018**, *65*, 654–663. [CrossRef]
32. Liu, L.; Wang, Z.; Zhang, H. Neural-Network-Based Robust Optimal Tracking Control for MIMO Discrete-Time Systems with Unknown Uncertainty Using Adaptive Critic Design. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 1239–1251. [CrossRef]
33. Yang, X.; Liu, D.; Wei, Q.; Wang, D. Guaranteed cost neural tracking control for a class of uncertain nonlinear systems using adaptive dynamic programming. *Neurocomputing* **2016**, *198*, 80–90. [CrossRef]
34. Mu, C.; Zhang, Y.; Gao, Z.; Sun, C. ADP-Based Robust Tracking Control for a Class of Nonlinear Systems with Unmatched Uncertainties. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 4056–4067. [CrossRef]
35. Başar, T.; Olsder, G.J. *Dynamic Noncooperative Game Theory*, 2nd ed.; Academic Press: Cambridge, MA, USA, 1999.
36. Vamvoudakis, K.G.; Lewis, F.L. Non-zero sum games: Online learning solution of coupled Hamilton-Jacobi and coupled Riccati equations. In Proceedings of the 2011 IEEE International Symposium on Intelligent Control, Denver, CO, USA, 28–30 September 2011; pp. 171–178.
37. Finlayson, B.A. The Method of Weighted Residuals and Variational Principles. *J. Fluid Mech.* **1973**, *57*, 623.
38. Vamvoudakis, K.G.; Lewis, F.L. Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica* **2010**, *46*, 878-888.