Article

Robust Trajectory Tracking Control for Continuous-Time Nonlinear Systems with State Constraints and Uncertain Disturbances

Chunbin Qin [ID], Xiaopeng Qiao, Jinguang Wang [ID] and Dehua Zhang *

School of Artificial Intelligence, Henan University, Zhengzhou 450000, China; qcb@henu.edu.cn (C.Q.); qxp@henu.edu.cn (X.Q.); wjg@henu.edu.cn (J.W.)
* Correspondence: dhuazhang@vip.henu.edu.cn

Abstract: In this paper, a robust trajectory tracking control method with state constraints and uncertain disturbances on the ground of adaptive dynamic programming (ADP) is proposed for nonlinear systems. Firstly, the augmented system consists of the tracking error and the reference trajectory, and the tracking control problems with uncertain disturbances is described as the problem of robust control adjustment. In addition, considering the nominal system of the augmented system, the guaranteed cost tracking control problem is transformed into the optimal control problem by using the discount coefficient in the nominal system. A new safe Hamilton–Jacobi–Bellman (HJB) equation is proposed by combining the cost function with the control barrier function (CBF), so that the behavior of violating the safety regulations for the system states will be punished. In order to solve the new safe HJB equation, a critic neural network (NN) is used to approximate the solution of the safe HJB equation. According to the Lyapunov stability theory, in the case of state constraints and uncertain disturbances, the system states and the parameters of the critic neural network are guaranteed to be uniformly ultimately bounded (UUB). At the end of this paper, the feasibility of the proposed method is verified by a simulation example.

1. Introduction

With the continuous application of automatic driving technology [1,2] and intelligent robot technology [3,4], the role of the safety-critical system has attracted extensive attention. In the process of designing a controller, safety is the primary consideration compared to other performances. For the control system with strict safety requirements, the CBF was applied to the control system as a tool to achieve the purpose of state constraints.

Reinforcement learning (RL) can be regarded as the technology of strategy learning and evaluation learning. In the actual engineering application, although the phenomenon of the dimension curse exists in dynamic programming, RL can deal with it well, and we also call it adaptive dynamic programming (ADP) [5–7]. Adaptive dynamic programming is an intelligent control method, and it is also an approximate tool to deal with optimal control problems. However, the analytical solution of the Hamilton–Jacobi–Bellman (HJB) equation is generally difficult to obtain; therefore, the adaptive dynamic programming (ADP) can learn the solution of the HJB equation online by the neural network (NN) approximation method [8–10]. At present, a variety of control methods based on ADP have been proposed by researchers to deal with the problem of trajectory tracking and optimal control [11–16].

Adaptive dynamic programming enables the complex nonlinear system to achieve the desired tracking control goal [17–20]. In reference [17], the tracking performance of

continuous-time nonlinear systems was analyzed by considering the influence of input constraints. Due to the influence of the actual situation, a series of uncertain disturbance factors are often considered. Therefore, robust optimal tracking control has become a research hot spot. In reference [18], solving tracking problems for complex nonlinear systems with uncertainty can be more difficult, and the adaptive criticism technique was used to solve the robust tracking control problems of nonlinear systems with random disturbances. Considering the nonlinear system with a continuous-time matching uncertainty, an effective robust tracking control method was adopted, and the discounted coefficient was selected for the nominal augmented error system in references [19,20]. Considering the system with disturbances, H_∞ tracking control was used in control systems with disturbances [21]. In order to reduce the design cost and waste of resources and adjust the accuracy of the control system, a tracking control method based on the event triggering was proposed [22]. Considering the optimal regulation problem, a new non-quadratic discount performance function was proposed in reference [23]. In reference [24], an improved adaptive robust tracking method was proposed for the uncertainty of nonlinear systems and successfully extended to the mass-spring-damper system. The tracking control method proposed above enabled the feasibility of the control strategy and enabled the system to achieve the predetermined control target. However, none of the tracking control methods proposed above consider the state constraints problem.

In references [25–28], different ADP-based methods were proposed to solve various engineering problems. In some specific environments, the control system is often required to have reliable security. The purpose for which the safety system was designed is to find its control strategy by conforming to the safety specifications specified by the physical constraints of the system [29]. The use of the CBF method to solve the safety constraints of systems with strict requirements has attracted extensive attention [30–33]. Let the states displayed by the system converge to the desired equilibrium point; an approximate adjustment method for solving the optimization problem of safety boundary control was proposed, and the cost of violating the safety constraint was directly embedded into the value function [34]. In reference [35], the application of the CBF was introduced, and the verification method and the characteristics of implementation safety in the context of a safety-critical control system are summarized. The discrete-time state constraints problem was described in reference [36], and the HJB equation with the CBF was solved by using the approximate properties of the neural network.

In this paper, a new guaranteed cost robust tracking method with state constraints and uncertain disturbances is proposed. This method can guarantee the convergence of the system error under conditions of uncertain disturbances and state constraints. The discounted coefficient is selected for the nominal augmented system with tracking errors. In addition, the CBF is added to the system to solve the constraint problem of system states. Finally, the approximation property of the critic NN is adopted to deal with the HJB equation. The contributions of this paper are described below:

1. For robust tracking control problems, the CBF is applied to the tracking control system with uncertain disturbances so that the system can still have good tracking performance in the case of state constraints;
2. Combining the traditional adaptive control method with the CBF, the CBF is directly extended to the original system, and the CBF is used as a penalty function to punish unsafe behavior;
3. A new guaranteed cost robust adaptive tracking method with state constraints and uncertain disturbances is proposed to solve the safety HJB equation through the critic NN learning framework, and the critic NN parameters are guaranteed to be uniformly ultimately bounded (UUB) under the influence of state constraints and uncertain disturbances.

The arrangement of other parts of this article is described below: Section 2 states the preliminary knowledge and introduces the relevant contents of the control barrier function. Section 3 describes the selection of discount value functions for the nominal augmented

system and introduces the form of the new cost function after adding the barrier function. Section 4 introduces the learning method of a critic neural network with state constraints and uncertain disturbances. In Section 5, the effectiveness of the proposed method is verified by a simulation example. Finally, some conclusions are summarized in Section 6.

2. Preliminaries

2.1. Problem Statement

Consider the following uncertain nonlinear safety system

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

where $x(t) \in \Phi \subset \mathbb{R}^n$ is the state variable, $u(t) \in \mathcal{U} \subset \mathbb{R}^m$ is the control vector, Φ represents the safe feasible states set, $\mathcal{U}$ represents all admissible input sets, $f(x(t)) \in \mathbb{R}^n$ and $g(x(t)) \in \mathbb{R}^m$ are known functions with $f(0) = 0$, and $\Delta f(x(t)) \in \mathbb{R}^n$ is the unknown perturbation term with $\Delta f(0) = 0$. Here, let the initial state $x(0) = x_0$; we assume that there exists a constant g_M and it satisfies $0 < \|g(x)\| \leq g_M$ for $\forall x \in \mathbb{R}^n$, and $\Delta f(x) = g(x)d(x)$, where $d(x) \in \mathbb{R}^m$ is an unknown perturbation, and the known perturbation function $d_M(x) > 0$ is the boundary of $d(x)$, here, $\|d(x)\| \leq d_M(x)$. In addition, $d(0) = 0$ and $d_M(0) = 0$.

Assumption 1. *Let the reference trajectory of the system (1) be $x_d(t)$, and $x_d(t)$ is a bounded function, which is limited and generated by the command generator $\dot{x}_d(t) = r(x_d(t))$. Meanwhile, the reference trajectory $x_d \in \mathbb{R}^n$ and the command function $r \in \mathbb{R}^n$ are all Lipschitz continuous.*

2.2. Control Barrier Function

The application of the CBF further solves the constraint problem of the system [36]. In a predefined security set, the CBF candidate is always positive and tends to infinity at the defined set boundary. The CBF has a negative derivative at infinity, so the CBF will not reach infinity. If the state of the system is close to the safety boundary, then the condition that the derivative is negative will return the state to the safety set, so that the state displayed by the system will be maintained within the predetermined set. The safe feasible set Φ consists of operational constraints and safety specifications [34],

$$\Phi = \{x \in \mathbb{R}^n | h(x) \geq 0\}, \tag{2}$$

$$\partial\Phi = \{x \in \mathbb{R}^n | h(x) = 0\}, \tag{3}$$

$$Int\Phi = \{x \in \mathbb{R}^n | h(x) > 0\}, \tag{4}$$

where $\partial\Phi$ represents the boundary of the safe feasible set Φ, $Int\Phi$ represents the interior of the set Φ, and h is a continuously differentiable function of x, which is composed of a one-dimensional system constraint range.

The CBF candidate $B(x)$ satisfies all of the following properties,

$$\frac{1}{\alpha_1(h(x))} \leq B(x) \leq \frac{1}{\alpha_2(h(x))}, \forall x \in int\Phi \tag{5}$$

$$\dot{B}(x) \leq \alpha_3(h(x)), \forall x \in int\Phi \tag{6}$$

where $\alpha_1(\cdot)$, $\alpha_2(\cdot)$, and $\alpha_3(\cdot)$ are Lipschitz class $\mathcal{K}$ functions, and $B(x)$ is a control barrier function.

Assumption 2. *Under the condition of uncertain disturbances, to ensure that the states of the system are constrained . We use a logarithmic control barrier function $B_r(x)$, which satisfies the following properties,*

$$\begin{cases} B_r(x) > 0, \forall x \in \Phi, \\ B_r(x) \to \infty, \forall x \in \partial\Phi. \end{cases} \tag{7}$$

Besides, $B_r(x)$ is monotonically decreasing for $\forall\ x \in \Phi$.

Under the condition of satisfying Assumption 2, the expression of the specific logarithmic barrier function can be defined as

$$B_r(x) = -log(\frac{\gamma h(x)}{\gamma h(x) + 1}). \tag{8}$$

In (8), the parameter γ is a constant, and γ also determines the speed at which $B_r(x)$ is limited as it approaches the safety barrier.

Before describing the modified robust tracking method with constraints, we first make the following definitions and assumptions.

Definition 1. *The safety control input set of the nonlinear system (1) is given below*

$$\mathcal{U}_c = \{u \in \mathbb{R}^m | x^u \in int\Phi\}, \tag{9}$$

where x^u is the system state associated with the control strategy u, and $int\Phi$ is the interior of the set defined in (4).

Assumption 3. *The initial condition of the nonlinear system (1) is strictly within Φ; in other words, $x_0 \in int\Phi$. Assume that the initial set of allowed inputs is not empty and satisfies $\mathcal{U}_a = \mathcal{U} \cap \mathcal{U}_c$. In addition, the control strategy $u(x_0) \in \mathcal{U}_a$ exists.*

3. Guaranteed Cost Robust Tracking Design with State Constraints and Uncertain Disturbances

3.1. Modified Robust Adaptive Tracking Control

The augmented system is constructed by combining the tracking error and the reference trajectory. Before describing the modified robust adaptive tracking control, the tracking error is written as $e_x(t) = x(t) - x_d(t)$. According to (1), the tracking error system is derived as

$$\dot{e}_x(t) = f(x_d(t) + e_x(t)) + g(x_d(t) + e_x(t))u(t) + \Delta f(x_d(t) + e_x(t)) - r(x_d(t)), \tag{10}$$

where $r(x_d(t))$ is a Lipschitz continuous function.

By considering the tracking error dynamics (10), the infinite horizon cost function is given below [37]

$$\bar{V}(e_x(t), u) = \int_t^\infty e^{-\alpha(\tau - t)} U(e_x(\tau), u(\tau)) d(\tau), \tag{11}$$

where $\alpha > 0$ is a discount factor, and $U(e_x, u) = e_x^T Q e_x + u^T R u$, both $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$ are symmetric positive definite matrices.

Under the condition of state constraints and uncertain disturbances, the purpose of dealing with the guaranteed cost tracking problem is to find the control input $u = u(e_x(t), x_d(t))$ and a positive real number $\bar{V}^*$; then the tracking error $e_x(t)$ converges to zero. Meanwhile, the cost function described in (11) satisfies $\bar{V} < \bar{V}^*$. It should be pointed out that $\bar{V}^*$ is called a guaranteed cost function, and the control u is called a guaranteed cost control input.

Remark 1. *The discount term $e^{-\alpha(\tau - t)}$ given in (11) is mainly used to ensure that the cost function is $\bar{V} < \infty$ since the control $u(e_x(t), x_d(t))$ contains a part depending on the reference trajectory $x_d(t)$. In the absence of the discount term, $u(e_x(t), x_d(t))$ may make (11) to be unbounded. If the reference trajectory $x_d(t))$ does not converge to zero, the cost function (11) is unbounded without considering the discount term $e^{-\alpha(\tau - t)}$.*

Let $s(t) = [e_x^T(t), x_d^T(t)]^T \in \mathbb{R}^{2n}$, and the augmented system of error dynamics can be given

$$\dot{s}(t) = F(s(t)) + G(s(t))u(t) + \Delta F(s(t)), \tag{12}$$

the specific forms of $F(s(t))$ and $G(s(t))$ can be expressed as

$$F(s(t)) = \begin{bmatrix} f(x_d(t) + e_x(t)) - r(x_d(t)) \\ r(x_d(t)) \end{bmatrix}$$

and

$$G(s(t)) = \begin{bmatrix} g(x_d(t) + e_x(t)) \\ 0 \end{bmatrix},$$

and $\Delta F(s(t)) = G(s(t))d(s(t))$, and because we know that $\|d(x)\| \le d_M(x)$, it is very easy to know that the uncertain disturbance term $\|d(s(t))\| \le d_M(s(t))$ holds, and $d_M(s(t))$ is the boundary of the uncertain disturbance term $d(s(t))$.

Remark 2. *There is a random disturbance term $d(s(t))$ in the augmented system (12) described above, which makes the process of designing the controller difficult. In the following introduction, the augmented error system (12) is equivalent to the optimal control of its nominal system, and the tracking problem with the random disturbance is transformed into an optimization adjustment problem with the discounted value function.*

Considering the existence of the uncertain term $d(s(t))$, the nominal system description of the system (12) is

$$\dot{s}(t) = F(s(t)) + G(s(t))u(t). \tag{13}$$

Inspired by references [29,36], $B_r(x)$ is combined with the nominal augmented system (13), and the modified value function is

$$V(s(t)) = \int_t^\infty e^{-\alpha(\tau - t)}[\rho d_M^2(s(\tau)) + s^T(\tau)Q_T s(\tau) + u^T(\tau)Ru(\tau) + B_r(s(\tau))]d(\tau), \tag{14}$$

where $Q_T = diag\{Q, 0_{n \times n}\}$, $\rho = \lambda_{max}(R)$, the maximum eigenvalue of R can be expressed by $\lambda_{max}(R)$, both $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$ are weighted symmetric positive definite matrices of augmented systems, and $\alpha > 0$ is a discount coefficient.

According to Bellman's principle of the optimal control theory [38], the minimum value of the Hamiltonian of the modified value function (14) of the nominal system (13) is given

$$H_{min}(s, u, V_s) = \rho d_M^2(s) + s^T Q_T s + u^T Ru - \alpha V(s) + B_r(s) + V_s^T(F(s) + G(s)u), \tag{15}$$

where $V_s = \partial V / \partial s$, and the cost function $V^*(s(t))$ can be considered as

$$V^*(s(t)) = \min_u V(s(t)). \tag{16}$$

For the system (13) with the control barrier value function (14), since the equation $\partial H(s, u^*, V_s^*) / \partial u^* = 0$ holds, we can obtain the optimal control input u^* from (15)

$$u^* = -\frac{1}{2}R^{-1}G^T(s)V_s^*, \tag{17}$$

where $V_s^* = \partial V^*(s) / \partial s$, and $V^*(s)$ denotes the optimal value $V(s)$.

3.2. State Constraints Analysis

In the process of designing a robust tracking controller, the CBF as a constraint tool makes the states of the system evolve within the specified constraints, and the system can maintain good performance within the set safety constraints. The CBF provides a constraint tool for safety-critical systems to optimize the performance of other control objectives and clearly explains the priority of security compared to other performance indexes. In order to further describe that the CBF is bounded, it is described below that the boundedness of the CBF is demonstrated by changing the order of the controller.

Lemma 1. *Consider an admissible feedback control strategy $u_1 \in \mathcal{U}_a$; there is the following time-invariant positive definite function Z, which satisfies $Z \in N^1$*

$$\frac{\partial V^T}{\partial x}(F(s) + G(s)u_1) + \rho d_M^2(s) + s^T Q_T s + u_1^T R u_1 - \alpha V(s) + B_r(s) = 0, \tag{18}$$

$$V(s_0, u_1) = Z(s_0, u_1), \tag{19}$$

where V is the value function of the system for all $t \in [0, \infty)$, and the following formula holds

$$V(s, u) = Z(s, u). \tag{20}$$

Proof. Assume $V(s, u_1) > 0$ exists and is continuously differentiable; then, we have

$$V(s(t), u_1) - V(s_0, u_1) = \int_0^t \dot{V}(s(\tau), u_1) d(\tau) = \int_0^t \frac{\partial V}{\partial s}(F + G u_1) d(\tau). \tag{21}$$

Considering (21), there are also

$$Z(s(t), u_1) - Z(s_0, u_1) = -\int_0^t P(s(\tau), u_1) d(\tau), \tag{22}$$

where $P(s, u) = \rho d_M^2(s) + s^T Q_T s + u^T R u - \alpha V(s) + B_r(s)$.
We can derive from (21) and (22)

$$Z(s(t), u_1) - V(s(t), u_1) = \int_0^t \left(-\frac{\partial V}{\partial s}(F + G u_1) - P(s(\tau), u_1)\right) d(\tau) + Z(s_0, u_1) - V(s_0, u_1). \tag{23}$$

Combining (18), (21), and (23), we can obtain

$$Z(s(t), u_1) - V(s(t), u_1) = \int_0^t (P(s(\tau), u_1) - P(s(\tau), u_1)) d(\tau) = 0. \tag{24}$$

Therefore, we can obtain

$$Z(s(t), u_1) = V(s(t), u_1). \tag{25}$$

This completes the proof. □

Lemma 2. *We consider a series of positive definite value functions $V(s, t, u_1)$, $V(s, t, u_2)$,..., and $V(s, t, u_i)$, and the corresponding abbreviations are V_1, V_2, ..., and V_k, which are concerned with the allowable control inputs $u_1(s, t)$, $u_2(s, t)$, ..., and $u_k(s, t) \in \mathcal{U}_a$. Then, the Hamiltonian value defined in (15) satisfies the following conditions*

$$H_{min_1} \leq H_{min_2} \leq \ldots H_{min_i}, \tag{26}$$

and the CBF candidate B_r^k is bounded in the range of $1 < k < i$.

Proof. Assume that $0 \leq k \leq j \leq i$ is satisfied for any j and k, and the condition $H_{min_k} \leq H_{min_j}$ holds; therefore, one has

$$V_j = V_k + V_o, \tag{27}$$

where $V_o = V_o(s(t), u_k)$. According to (17), u^* can be rewritten as

$$u^* = -\frac{1}{2}R^{-1}G^T \nabla V_j,$$

$$H_{min_j} = \nabla V_j^T (F + G(-\frac{1}{2}R^{-1}G^T \nabla V_j)) + T(s) + \frac{1}{4}\nabla V_j^T G R^{-1} G^T \nabla V_j.$$

Considering $T(s) = \rho d_M^2(s) + s^T Q_T s + B_r(s) - \alpha V(s)$. According to (27), one may obtain

$$H_{min_j} = H_{min_k} + \nabla V_o^T (F + G u_k^*) - (u_o^{*T} R u_o^*).$$

According to the above description, since $H_{min_j} - H_{min_k} + (u_o^{*T} R u_o^*) \geq 0$, we can obtain

$$\frac{\partial V_o(s, u_k)}{\partial t} \geq 0.$$

Because $\lim_{t \to \infty} V_o(s(t)) = 0$, the following results are obtained

$$V_o \leq 0,$$

$$V(s, t, u_1) > V(s, t, u_2) > \ldots > V(s, t, u_i).$$

From the above Lemmas 1 and 2, we can obtain

$$Z(s(t), u_k) < Z(s(t), u_1), 1 < k < i. \tag{28}$$

In the above derivation, not only $Z(s(t), u_k)$ is bounded, but also $P(s, u)$ is positive definite, and then B_r^k is also bounded. In other words, in the case of state constraints, the system states will not reach the safety boundary in the process of tracking the reference trajectory. This proves that the CBF is bounded within each moment. □

Theorem 1. *For the performance optimization problem described in (16), let both Assumption 2 and Assumption 3 hold. Through the improvement of control input (17), the security of the tracking state is guaranteed within a certain range for all $t > 0$.*

Proof. Through the introduction to Lemmas 1 and 2 above, the performance functions $Z(s, u_k)$ and candidate function B_r^k are bounded at each moment after the control input (17) is changed. From Assumptions 1 and 2, at the boundary of the constraint range, the value of the barrier function B_r^k will reach infinity; in other words, the CBF remains is bounded at any moment, which ensures that the states of the system never reach the safe boundary. □

In the above introduction, the CBF is directly added to the cost function, which makes the states of the system constrained. This method is applicable to the guaranteed cost robust trajectory tracking control without initial admissible control. The traditional tracking controller usually needs the initial admissible control law. Although the appropriate initial admissible control law is found, the appropriate initial admissible control law may not satisfy the condition of state constraints.

Due to the existence of the discount term $e^{-\alpha(\tau - t)}$ in Equation (11), to guarantee the stability of the closed-loop system in the process of the tracking reference trajectory process, a guaranteed cost adaptive critic NN learning framework is designed. Before proceeding to the next step, we make the following assumption.

Assumpton 4. *Let $J_1(s)$ be a candidate of Lyapunov function and satisfy the condition of $\dot{J}_1(s) = \nabla J_1^T(s)(F(s) + G(s)u^*) < 0$, and $J_1(s)$ is continuously differentiable, where $\nabla J_1(s) = \partial J_1(s)/\partial s$. Assume there exists a symmetric positive definite matrix $\Lambda(s)$, and the condition of expression $(\nabla J_1(s))^T (F(s) + G(s)u^*) = -(\nabla J_1(s))^T \Lambda(s) \nabla J_1(s)$ holds.*

4. Design of Guaranteed Cost Adaptive Critic NN Learning Framework

In this section, the approximation property of the critic NN is used to approximate the solution of the safety HJB Equation (15), a guaranteed cost adaptive critic NN learning framework is proposed, the weight of the critic NN is updated through online the learning scheme, and all the vectors of the critic NN finally are guaranteed to be UUB. Considering the cost function described in (16), we design a critic NN to approximate the cost function $V^*(s(t))$ and its partial derivative

$$V^*(s) = W^T \phi(s) + \epsilon_v(s), \tag{29}$$

$$\nabla V^*(s) = (\nabla \phi(s))^T W + \nabla \epsilon_v(s), \tag{30}$$

where $W \in \mathbb{R}^l$ is the ideal vector of the critic neural network, the activation function of the critic NN can be expressed as $\phi(s) = [\varphi_1 \quad \varphi_2 \quad \varphi_3 \quad \cdots \quad \varphi_l]^T \in \mathbb{R}^l$, l is the number of hidden-layer neurons, $\nabla \phi(s)$ is denoted as the derivative of $\phi(s)$, the approximation error of the critic NN is denoted by $\epsilon_v(s)$, and $\nabla \epsilon_v(s)$ is the derivative of $\epsilon_v(s)$.

Assumption 5. *The vector W of the critic NN is bounded by a positive constant, i.e., $\|W\| < W_M$, the activation function $\phi(s)$ and its derivative $\nabla \phi(s)$, the critic NN error $\epsilon_v(s)$ and its derivative $\nabla \epsilon_v(s)$, are bounded, and satisfy $\|\phi(s)\| < \phi_\varepsilon$, $\|\nabla \phi(s)\| < \phi_{d\varepsilon}$, $\|\epsilon_v(s)\| < \epsilon_\sigma$, and $\|\nabla \epsilon_v(s)\| < \epsilon_{d\sigma}$, where ϕ_ε, $\phi_{d\varepsilon}$, ϵ_σ, and $\epsilon_{d\sigma}$ are positive constants.*

From Equations (15), (16), and (30), the approximate error of the safety HJB form is

$$\rho d_M^2(s) + s^T Q_T s + u^{*T} R u^* - \alpha V^*(s) + B_r(s) + W^T \nabla \phi(s)(F(s) + G(s)u^*) + \epsilon_{v1}(s) = 0, \tag{31}$$

where $\epsilon_{v1}(s) = \nabla \epsilon_v^T(s)(F(s) + G(s)u^*)$.

Considering Equations (17) and (30), we can draw the following conclusion,

$$u^* = -\frac{1}{2} R^{-1} G^T(s)(\nabla \phi(s))^T W + \epsilon_{v2}, \tag{32}$$

where $\epsilon_{v2} = -(1/2)R^{-1}G^T(s)\nabla \epsilon_v$. At the same time, we substitute (32) into (31) and can obtain

$$\rho d_M^2(s) + s^T Q_T s - \alpha W^T \phi + B_r(s) + W^T \nabla \phi F - \frac{1}{4} W^T \nabla \phi \varrho (\nabla \phi)^T W - \omega = 0, \tag{33}$$

where $\varrho = G(s)R^{-1}G(s)$, and $\omega = \alpha \epsilon_v - (\nabla \epsilon_v)^T F + (1/4)(\nabla \epsilon_v)^T \varrho \nabla \epsilon_v + (1/2)W^T \nabla \phi \varrho \nabla \epsilon_v$ is the approximation error.

We do not know the value of the ideal weight W; therefore, by using the critic NN to approximate the cost function $V^*(s)$ as

$$\hat{V}(s) = \hat{W}^T \phi(s), \tag{34}$$

where $\hat{W}$ denotes the estimated value of the ideal vector W, and $\hat{V}$ is the estimated value of the ideal cost function V^*. We can obtain the approximate HJB equation form from Equations (15) and (34)

$$\hat{H}(s, u, \hat{V}_s) = \rho d_M^2(s) + s^T Q_T s + u^T R u - \alpha \hat{V}(s) + B_r(s) + \hat{V}_s^T(F(s) + G(s)u). \tag{35}$$

Based on Equation (34), the control input $\hat{u}(s)$ can be approximated by

$$\hat{u}(s) = -\frac{1}{2} R^{-1} G^T(s)(\nabla \phi(s))^T \hat{W}. \tag{36}$$

Through Equations (31) and (35), we define the HJB equation error caused by the critic NN in the approximation process as

$$\varepsilon = \rho d_M^2(s) + s^T Q_T s - \alpha \hat{W}^T \phi + B_r(s) + \hat{W}^T \nabla \phi F - \frac{1}{4} \hat{W}^T \nabla \phi \varrho (\nabla \phi)^T \hat{W}. \tag{37}$$

The estimation error of the weights of the critic NN is defined as $\tilde{W}$, and we can obtain

$$\tilde{W} = W - \hat{W}. \tag{38}$$

The HJB approximation error can be defined as

$$\varepsilon = -\tilde{W}^T \xi + \frac{1}{4}\tilde{W}^T \nabla\phi \varrho (\nabla\phi)^T \tilde{W} + \omega, \tag{39}$$

where $\xi = \nabla\phi(F(s) + G(s)\hat{u}) - \alpha\phi(s)$. The Lyapunov function candidate $J_1(s)$ is shown in Assumption 4, and we take $\Pi(s, \hat{u})$ as an indicator function and define it as

$$\Pi(s, \hat{u}) = \begin{cases} 0, \text{if } J_1^T(s)(F(s) + G(s)\hat{u}) < 0, \\ 1, \text{else.} \end{cases} \tag{40}$$

We choose $\hat{W}$ to minimize the square residual $E = (1/2)\varepsilon^T \varepsilon$, and then we obtain the minimum value of the HJB approximation error ε. We use the gradient descent method as the critic vector adjustment optimization law

$$\dot{\hat{W}} = -\beta\bar{\xi}(L(s) + Y(s) + B_r(s) + \rho d_M^2(s)) + \frac{\beta}{2}\Pi(s, \hat{u})\nabla\phi \varrho J_1(s) + \beta((K_1\theta^T - K_2) + A(s))\hat{W}, \tag{41}$$

where $\bar{\xi} = \xi/(1 + \xi^T\xi)^2$, $\theta = \xi/(1 + \xi^T\xi)$, $L(s) = \hat{W}^T\nabla\phi F - \alpha\hat{W}^T\phi + s^T Q_T s$, $Y(s) = -(1/4)\hat{W}^T\nabla\phi\varrho(\nabla\phi)^T\hat{W}$, and $A(s) = (1/4)\nabla\phi\varrho(\nabla\phi)^T\hat{W}(\theta/(1 + \xi^T\xi))^T$, and $\beta > 0$ is a learning rate that determines the convergence speed of the critic NN. K_1 and K_2 are two tuning parameters.

From the above description, it is deduced that the weight estimation error is

$$\dot{\tilde{W}} = -\beta\bar{\xi}(\tilde{W}^T\xi + \tilde{Y}(s) - B_r(s)) - \frac{\beta}{2}\Pi(s, \hat{u})\nabla\phi\varrho J_1 - \beta((K_1\theta^T - K_2) - \tilde{A}(s))(W - \tilde{W}), \tag{42}$$

where $\tilde{Y}(s) = -(1/4)\tilde{W}^T\nabla\phi\varrho(\nabla\phi)^T\tilde{W}$, $\tilde{A}(s) = (1/4)\nabla\phi\varrho(\nabla\phi)^T(W - \tilde{W})(\theta/(1 + \xi^T\xi))^T$.

Theorem 2. *Consider the nominal system* (13), *the modified value function* (15), *and the tuning laws* (41). *Only if all the above Assumptions 1–5 hold, then the critic NN error $\tilde{W}$, the system state x, and the control input u^* are guaranteed to be UUB.*

Proof. Analyze the Lyapunov candidate function described below

$$L(t) = V_s(s(t)) + \frac{1}{2}\tilde{W}^T\beta^{-1}\tilde{W}. \tag{43}$$

The result of deriving Equation (43) is shown as

$$\begin{aligned} \dot{L}(t) &= \dot{V}_s(s(t)) + \tilde{W}^T\beta^{-1}\dot{\tilde{W}} \\ &= \dot{L}_V + \dot{L}_W. \end{aligned} \tag{44}$$

The first term $\dot{L}_V$ is

$$\begin{aligned} \dot{L}_V &= \tilde{W}^T\nabla\phi(s)(F(s) + G(s)u) + \nabla\epsilon_v^T(s)(F(s) + G(s)u) \\ &= W^T(\nabla\phi(s)F(s) - \frac{1}{2}D_1\hat{W}) + \epsilon_{v1}(s) \\ &= W^T\nabla\phi(s)F(s) + \frac{1}{2}W^T D_1(W - \hat{W}) - \frac{1}{2}W^T D_1 W + \epsilon_{v1}(s) \\ &= W^T\nabla\phi(s)F(s) + \frac{1}{2}W^T D_1\tilde{W} - \frac{1}{2}W^T D_1(s)W + \epsilon_{v1}(s) \\ &= W^T\sigma + \frac{1}{2}W^T D_1\tilde{W} + \epsilon_{v1}(s), \end{aligned} \tag{45}$$

where $\epsilon_{v1}(s) = \nabla\epsilon_{v1}^T(s)(F(s) - \frac{1}{2}G(s)R^{-1}G(s)^T\nabla\phi(s)\tilde{W})$, $\sigma = \nabla\phi(s)(F(s) + G(s)u)$, and $D_1 = \nabla\phi\varrho(\nabla\phi)^T$.

The second term $\dot{L}_W$ can be obtained by (41)

$$
\begin{aligned}
\dot{L}_W = {} & \tilde{W}^T \beta^{-1} \dot{\tilde{W}} \\
= {} & \tilde{W}^T \beta^{-1} [-\beta \bar{\xi}(\tilde{W}^T \zeta + \tilde{A}(s) - B_r(s)) - \frac{\beta}{2} \Pi(s, \hat{u}) \nabla \phi \varrho J_1 - \beta((K_1 \theta^T - K_2) \\
& - \tilde{A}(s))(W - \tilde{W})] \\
= {} & \tilde{W}^T [-\bar{\xi}(\tilde{W}^T \zeta + \tilde{A}(s) - B_r(s)) - \frac{1}{2} \Pi(s, \hat{u}) \nabla \phi \varrho J_1 - ((K_1 \theta^T - K_2) - \tilde{A}(s))(W - \tilde{W})] \\
= {} & \tilde{W}^T [-\bar{\xi}((\tilde{W}^T \zeta - \omega) - \frac{1}{4} \tilde{W}^T D_1 \tilde{W}) - ((K_1 \theta^T - K_2) - \frac{D_1}{4}(W - \tilde{W})\frac{\theta^T}{m})(W - \tilde{W})] \\
& - \frac{1}{2} \Pi(s, \hat{u}) \tilde{W}^T \nabla \phi \varrho J_1 \\
= {} & \tilde{W}^T [-\theta \tilde{W}^T \theta + \frac{\theta \omega}{m} + \frac{\theta \tilde{W}^T D_1 \tilde{W}}{4}] - \tilde{W}^T [(K_1 \theta^T - K_2)(W - \tilde{W}) \\
& - \frac{D_1}{4}(W - \tilde{W})(\theta/m)^T (W - \tilde{W})] - c,
\end{aligned}
$$
$$(46)$$

where $m = 1 + \xi^T \xi$, $c = -\frac{\Pi(s,\hat{u})\tilde{W}^T \nabla \phi \varrho J_1}{2}$. Further, we can obtain

$$
\begin{aligned}
\dot{L}_W = {} & -\tilde{W}^T \theta \tilde{W}^T \theta + \frac{\tilde{W}^T \theta \omega}{m} + \frac{\tilde{W}^T \theta \tilde{W}^T D_1 \tilde{W}}{4} - \tilde{W}^T (K_1 \theta^T - K_2)(W - \tilde{W}) + \frac{\tilde{W}^T D_1 \theta^T}{4m}(W \\
& - \tilde{W})(W - \tilde{W}) - c \\
= {} & -\tilde{W}^T \theta \tilde{W}^T \theta + \frac{\tilde{W}^T \theta \omega}{m} + \frac{\tilde{W}^T \theta \tilde{W}^T D_1 \tilde{W}}{4} - \tilde{W}^T (K_1 \theta^T - K_2)W + \tilde{W}^T (K_1 \theta^T - K_2)\tilde{W} \\
& + \frac{\tilde{W}^T D_1 W \theta^T}{4m} W - \frac{\tilde{W}^T D_1 \tilde{W} \theta^T}{4m} W - \frac{\tilde{W}^T D_1 W \theta^T}{4m}\tilde{W} + \frac{\tilde{W}^T D_1 \tilde{W} \theta^T}{4m}\tilde{W} - c.
\end{aligned}
$$
$$(47)$$

Taking the sum of the terms $\dot{L}_V$ and $\dot{L}_W$, we obtain

$$
\begin{aligned}
\dot{L}(t) = {} & W^T \sigma + \frac{1}{2} W^T D_1 \tilde{W} - \tilde{W}^T \theta \tilde{W}^T \theta + \frac{\tilde{W}^T \theta \omega}{m} + \frac{\tilde{W}^T \theta \tilde{W}^T D_1 \tilde{W}}{4} - \tilde{W}^T (K_1 \theta^T - K_2)W \\
& + \tilde{W}^T (K_1 \theta^T - K_2)\tilde{W} + \frac{\tilde{W}^T D_1 W \theta^T}{4m} W - \frac{\tilde{W}^T D_1 \tilde{W} \theta^T}{4m} W - \frac{\tilde{W}^T D_1 W \theta^T}{4m}\tilde{W} \\
& + \frac{\tilde{W}^T D_1 \tilde{W} \theta^T}{4m}\tilde{W} - c + \epsilon_{v1}(s).
\end{aligned}
$$
$$(48)$$

Assume that $Z = [\tilde{W}^T \theta, \tilde{W}^T]^T$, then we can obtain

$$
\begin{aligned}
\dot{L}(t) = {} & -Z^T \begin{bmatrix} I & -\frac{W^T D_1}{8m} - \frac{K_1^T}{2} \\ -\frac{D_1 W}{8m} - \frac{K_1}{2} & K_2 - \frac{\theta^T W D_1}{4m} \end{bmatrix} Z \\
& + Z^T \begin{bmatrix} \frac{\omega}{m} \\ \frac{D_1 W \theta^T W}{4m} + K_2 W - K_1 \theta^T W \end{bmatrix} + b + d,
\end{aligned}
$$
$$(49)$$

where

$$
b = c + W^T \sigma + \|\epsilon_{v1}(s)\|,
$$
$$(50)$$

$$
d = \frac{W^T D_1(\tilde{W}\theta^T \tilde{W} - \tilde{W}\theta^T W - W\theta^T \tilde{W})}{4m}.
$$
$$(51)$$

Define

$$
M = \begin{bmatrix} I & -\frac{W^T D_1}{8m} - \frac{K_1^T}{2} \\ -\frac{D_1 W}{8m} - \frac{K_1}{2} & K_2 - \frac{\theta^T W D_1}{4m} \end{bmatrix},
$$
$$(52)$$

$$
a = \begin{bmatrix} \frac{\omega}{m} \\ \frac{D_1 W \theta^T W}{4m} + K_2 W - K_1 \theta^T W \end{bmatrix}.
$$
$$(53)$$

Let the tuning parameters K_1, K_2, and γ be chosen so that $M > 0$, and we obtain

$$\dot{L}(t) < -\|Z\|^2 \sigma_{min}(M) + \|a\| \|Z\| + \mu, \tag{54}$$

where $\mu = b + d$. In summary, the Lyapunov derivative $\dot{L}(t)$ is negative if

$$\|Z\| > \frac{\|a\|}{2\sigma_{min}(M)} + \sqrt{\frac{\mu}{\sigma_{min}(M)} + \frac{a^2}{4\sigma_{min}(M)}}. \tag{55}$$

Based on the Lyapunov theorem [39], as long as the selected appropriately tuning parameters K_1, K_2, and γ make the formula (32) hold, in the case of state constraints and uncertain disturbances, the critic NN weight error $\tilde{W}$, the system state x, and the control input u^* are guaranteed to be UUB, and the nonlinear system (1) is guaranteed to be closed-loop stable in the presence of state constraints and uncertain disturbances. The proof is completed. □

5. Simulation

We consider a spring-mass-damping system with nonlinear properties [22], and the system dynamics of the spring-mass-damper are as follows [24]

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -\frac{K(x_2)}{M} - \frac{C}{M}x_1 + \frac{1}{M}u + px_1 sin(x_2), \end{cases} \tag{56}$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$ and the nonlinear condition $K(x) = x^3$, x_1, and x_2 are the position and velocity, respectively, and u is the force applied to the object. M is the mass of the object. K is the stiffness constant of the spring, and C is the damping. The above system dynamics parameters are $M = 1$ kg and $C = 0.5$ N· s/m. A mismatched disturbance may lead to system instability. Considering that the system still has stable performance under disturbances, the uncertain disturbance matching the system is selected, the uncertain disturbance term $d(x) = px_1 sin(x_2)$, and we assume that $p \in [-1, 1]$ and $d_M(x) = \|x\|$.

In the simulation process, considering that the initial allowable control law is not required, to make the tracking errors of the system converge to zero, the reference trajectory gradually tending to zero is selected, and the following reference trajectory $x_d(t)$ is given

$$\dot{x}_d = \begin{bmatrix} -0.5x_{d1} - x_{d2}cos(x_{d1}) \\ sin(x_{d1}) - x_{d2} \end{bmatrix}, \tag{57}$$

the initial condition is given as $x_d(0) = [0.15, 0.25]^T$, and we set the augmented state vector as $s = [e_x^T, x_d^T]^T$, and then combine (56) with (57), the dynamics of the augmented system can be derived

$$\dot{s} = \begin{bmatrix} s_2 + s_4 + 0.5s_3 + s_4cos(s_3) \\ -(s_2 + s_4)^3 - 0.5(s_1 + s_3) - sin(s_3) + s_4 \\ -0.5s_3 - s_4cos(s_3) \\ sin(s_3) - s_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} (u + d(s)), \tag{58}$$

where $s = [s_1, s_2, s_3, s_4]^T = [e_{x1}, e_{x2}, x_{d1}, x_{d2}]^T$ with $e_{xi} = x_i - x_{di}$.

To constrain the states of the system in augmented system dynamics (58), the control barrier function used is as follows

$$\begin{cases} B_{r1}(s_1 + s_3) = -log(\frac{\gamma h(s_1+s_3)}{\gamma h(s_1+s_3)+1}) \\ B_{r2}(s_2 + s_4) = -log(\frac{\gamma h(s_2+s_4)}{\gamma h(s_2+s_4)+1}). \end{cases} \tag{59}$$

The state constraints of the system are given as $-0.2 \leq x_1 \leq 0.35$ and $-0.15 \leq x_2 \leq 0.4$, and the parameter $\gamma = 0.02$.

To complete the design of robust trajectory tracking control, the modified value function (14) can be specified as

$$V(s(t)) = \int_t^\infty e^{-\alpha(\tau-t)} [\rho d_M^2(s(\tau)) + s^T(\tau) Q_T s(\tau) + u^T(\tau) R u(\tau) + B_{r1}(s(\tau)) + B_{r2}(s(\tau))] d(\tau). \tag{60}$$

Besides, we select the learning rate as $\beta = 1.5$ and the discount factor $\alpha = 0.15$, respectively. In order to deal with the approximate optimal control for the nominal augmented part of (58), we choose $Q_T = diag\{5I_2, 0_{2\times2}\}$ and $R = I$, and I denotes an identity matrix of appropriate dimensions. In this example, the activation function for the critic NN is chosen as $\phi(s) = [s_1^2, s_1 s_2, s_1 s_3, s_1 s_4, s_2^2, s_2 s_3, s_2 s_4, s_3^2, s_3 s_4, s_4^2]^T$. In addition, the weights of the critic NN are denoted as $\hat{W}_c = [W_{c1}, W_{c2}, \ldots, W_{c10}]^T$. The initial value of the state is given as $x(0) = [-0.2, 0.4]^T$, and it is easy to calculate the initial error vector according to $s(0) = x(0) - x_d(0)$, so the initial state of the augmented system is $s(0) = [-0.35, 0.15, 0.15, 0.25]^T$. In order to satisfy the condition of persistency of excitation, an exploration noise $exp(-0.25t) sin^2(t) cos(t)$ is added during the training of the neural network.

The convergence of critic parameters is shown in Figure 1, and the critic parameters after 30 s converge to $\hat{W} = [3.3767, 0.9606, 0.8867, 0.7752, 1.9266, 1.0686, 1.105, 1.067, 1.0992, 1.0898]^T$. Figure 2 shows the control inputs of the system. Figure 3 shows the trajectory of the tracking errors e_{x1} and e_{x2} of the system without state constraints. Figure 4 shows the tracking error of the system under state constraints. Figures 5 and 6 show that the system tracks the reference trajectory without state constraints, and we can see that the system states violate the constraints. Figures 7 and 8 show that the system tracks the desired trajectory with state constraints, and that under the condition of state constraints and uncertain disturbances, the system still maintains good performance. The method described in this paper can ensure the stability of the closed-loop system. In summary, the simulation results display the effectiveness of the proposed method.
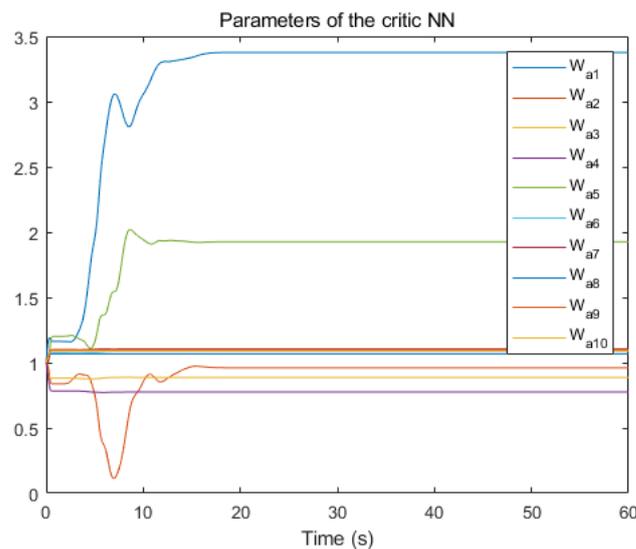

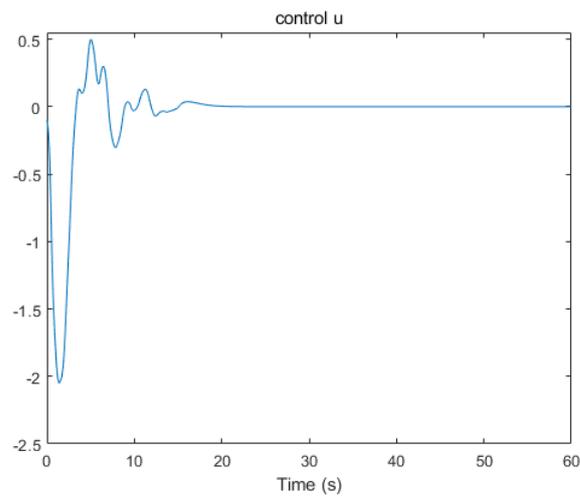**Figure 1.** Convergence of parameters of the critic NN.

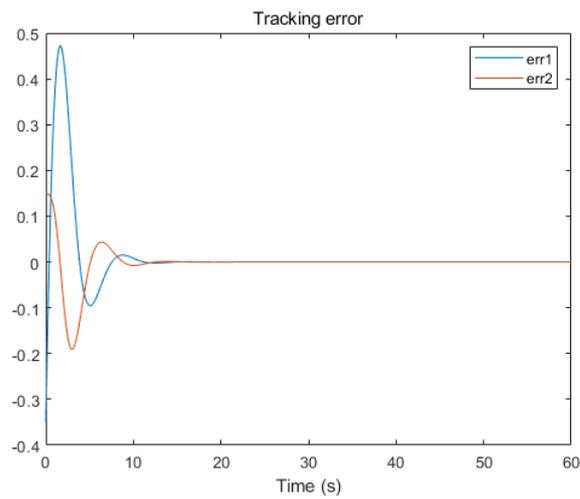**Figure 2.** The control input of the system.

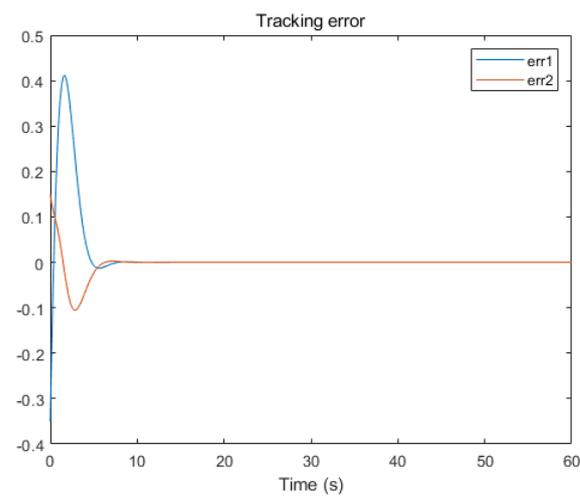**Figure 3.** Tracking error of system without state constraints ($p = 0.8$).

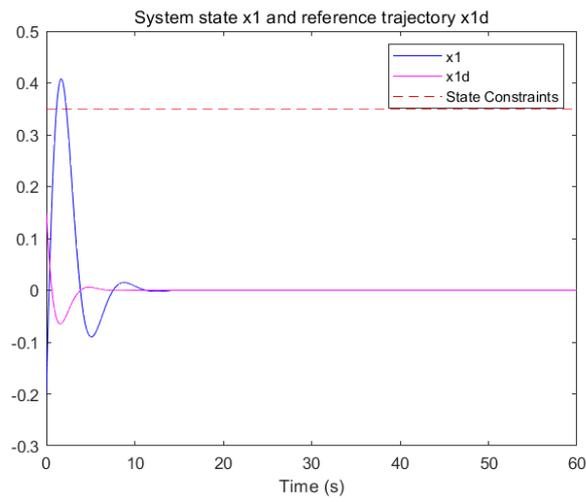**Figure 4.** Tracking error of system with state constraints ($p = 0.8$).

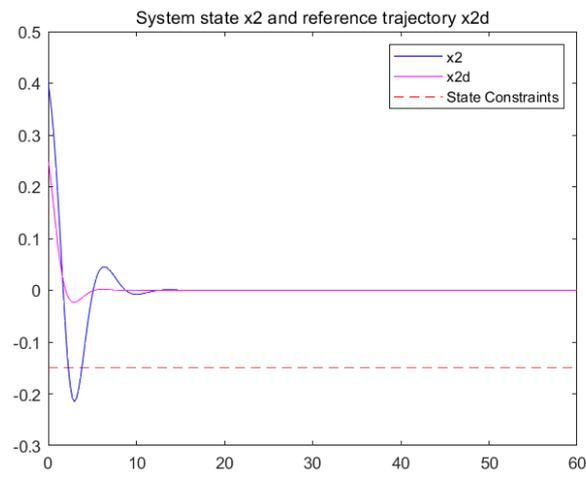**Figure 5.** Trajectory of the system state x_1 without state constraints ($p = 0.8$).

Figure 6. Trajectory of the system state x_2 without state constraints ($p = 0.8$).
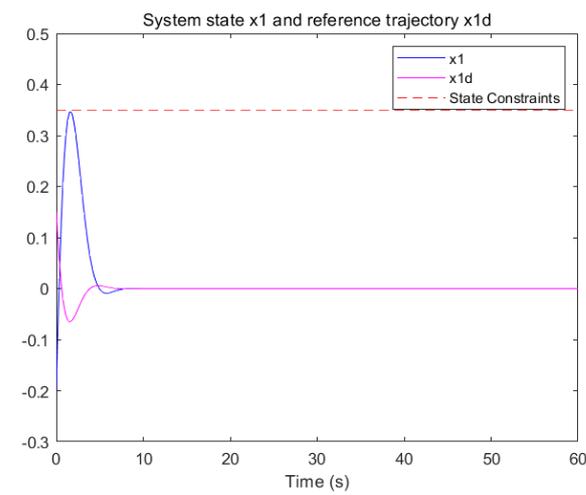

Figure 7. Trajectory of the system state x_1 with state constraints ($p = 0.8$).
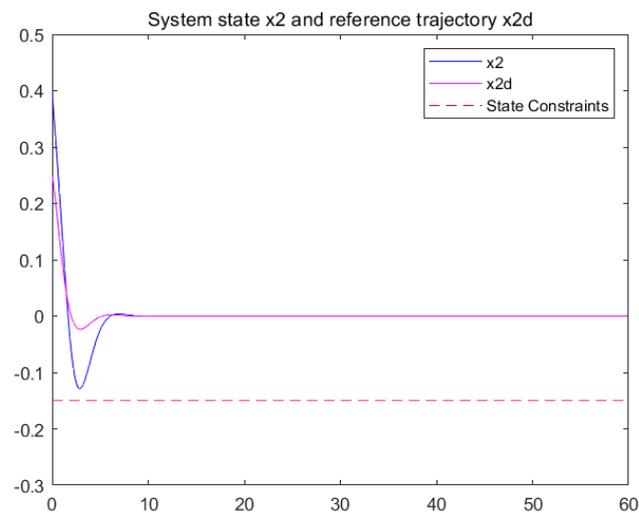
Figure 8. Trajectory of the system state x_2 with state constraints ($p = 0.8$).

6. Conclusions

This paper presented a robust trajectory tracking method for nonlinear systems with state constraints and uncertain disturbances based on adaptive dynamic programming. Firstly, the system error was combined with the reference trajectory to construct the augmented system, and at the same time, the nominal system of the augmented system was considered. In order to overcome the uncertain disturbances of the augmented system, the discount coefficient was introduced into the nominal system, and the CBF was added into the nominal system with the discount coefficient to constrain the states of the system. In addition, cost functions and control strategies were learned by designing a guaranteed cost adaptive critic NN learning framework. Finally, the simulation results demonstrated that the described method can converge the system error within the state constraints. In the next work, we will try to extend the state constraints method to discrete-time tracking control systems and multi-agent systems.

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

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 that they have no conflict of interest. All authors have approved the manuscript and agreed with submission to this journal.

References

1. Xie, R.; Tian, X.; Zhang, Q.; Pan, X. Research on longitudinal control algorithm for intelligent automatic driving. In Proceedings of the 2021 IEEE 4th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, 19–21 November 2021; pp. 664–667.
2. Chen, Y.; Peng, H.; Grizzle, J. Obstacle Avoidance for Low-Speed Autonomous Vehicles with Barrier Function. *IEEE Trans. Control Syst. Technol.* **2018**, *26*, 194–206. [CrossRef]
3. Molnar, T.G.; Cosner, R.K.; Singletary, A.W.; Ubellacker, W.; Ames, A.D. Model-Free Safety-Critical Control for Robotic Systems. *IEEE Robot. Autom. Lett.* **2022**, *7*, 944–951. [CrossRef]

4. Ferraguti, F.; Landi, C.T.; Costi, S.; Bonfe, M.; Fantuzzi, C. Safety barrier functions and multi-camera tracking for human–robot shared environment. *Robot. Auton. Syst.* **2020**, *124*, 103388. [CrossRef]

5. Liu, D.; Xue, S.; Zhao, B.; Luo, B.; Wei, Q. Adaptive Dynamic Programming for Control: A Survey and Recent Advances. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 142–160. [CrossRef]

6. Zhang, H.; Zhang, X.; Luo, Y. An overview of research on adaptive dynamic programming. *Acta Autom. Sin.* **2013**, *39*, 303–311. [CrossRef]

7. Wang, F.Y.; Zhang, H.; Liu, D. Adaptive dynamic programming: An introduction. *IEEE Comput. Intell. Mag.* **2009**, *4*, 39–47. [CrossRef]

8. 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. [CrossRef]

9. Mehraeen, S.; Dierks, T.; Jagannathan, S.; Crow, M.L. Zero-Sum Two-Player Game Theoretic Formulation of Affine Nonlinear Discrete-Time Systems Using Neural Networks. *IEEE Trans. Cybern.* **2013**, *43*, 1641–1655. [CrossRef]

10. Munos, R.; Baird, L.; Moore, A. Gradient descent approaches to neural-net-based solutions of the Hamilton Jacobi Bellman equation. In Proceedings of the International Joint Conference on Neural Networks, Washington, DC, USA, 10–16 July 1999; pp. 2152–2157.

11. Wei, Q.; Liu, D.; Lin, H. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems. *IEEE Trans. Cybern.* **2016**, *46*, 840–853. [CrossRef]

12. Wei, Q.; Liao, Z.; Yang, Z.; Li, B.; Liu, D. Continuous-time time-varying policy iteration. *IEEE Trans. Cybern.* **2020**, *50*, 4958–4971. [CrossRef]

13. Kiumarsi, B.; Lewis, F.L.; Modares, H.; Karimpour, A.; Naghibi, S. Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics. *Automatica* **2014**, *50*, 1167–1175. [CrossRef]

14. Wei, Q.; Song, R.; Yan, P. Data-Driven Zero-Sum Neuro-Optimal Control for a Class of Continuous-Time Unknown Nonlinear Systems With Disturbance Using ADP. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 444–458. [CrossRef] [PubMed]

15. Modares, H.; Lewis F.L. Linear Quadratic Tracking Control of Partially-Unknown Continuous-Time Systems Using Reinforcement Learning. *IEEE Trans. Autom. Control* **2014**, *59*, 3051–3056. [CrossRef]

16. Qin, C.; Zhang, H.; Luo, Y. Online optimal tracking control of continuous-time linear systems with unknown dynamics by using adaptive dynamic programming. *Int. J. Control* **2014**, *87*, 1000–1009. [CrossRef]

17. Cui, L.; Xie, X.; Wang, X.; Luo, Y.; Liu, J. Event-triggered single-network ADP method for constrained optimal tracking control of continuous-time non-linear systems. *Appl. Math. Comput.* **2019**, *352*, 220–234. [CrossRef]

18. Wang, D.; Liu, D.; Zhang, Y.; Li, H. Neural network robust tracking control with adaptive critic framework for uncertain nonlinear systems. *Neural Netw.* **2018**, *97*, 11–18. [CrossRef] [PubMed]

19. Zhao, J.; Na, J.; Gao, G. Robust tracking control of uncertain nonlinear systems with adaptive dynamic programming. *Neurocomputing* **2022**, *471*, 21–30. [CrossRef]

20. Wang, D.; Liu, D.; Li, H.; Ma, H. Adaptive dynamic programming for infinite horizon optimal robust guaranteed cost control of a class of uncertain nonlinear systems. In Proceedings of the American Control Conference, Chicago, IL, USA, 1–3 July 2015; pp. 2900–2905.

21. Modares, H.; Lewis, F.L.; Jiang, Z. H_∞ Tracking Control of Completely Unknown Continuous-Time Systems via Off-Policy Reinforcement Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 2550–2562. [CrossRef]

22. Wang, F.; Gu, Q.; Huang, B.; Wei, Q.; Zhou, T. Adaptive Event-Triggered Near-Optimal Tracking Control for Unknown Continuous-Time Nonlinear Systems. *IEEE Access* **2022**, *10*, 9506–9518. [CrossRef]

23. Modares, H.; Lewis, F.L. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. *Automatica* **2014**, *50*, 1780–1792. [CrossRef]

24. 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]

25. Lu, K.; Liu, C.; Sun, J.; Li, C.; Ma, C. Compensator-based approximate optimal control for affine nonlinear systems with input constraints and unmatched disturbances. *Trans. Inst. Meas. Control* **2020**, *42*, 3024–3034. [CrossRef]

26. Liu, D.; Wang, D.; Wang, F.; 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] [PubMed]

27. Yang, X.; Liu, D.; Wei, Q. Online approximate optimal control for affine nonlinear systems with unknown internal dynamics using adaptive dynamic programming. *IET Control Theory Appl.* **2014**, *8*, 1676–1688. [CrossRef]

28. Zhu, Y.; Zhao, D.; Li, X. Using reinforcement learning techniques to solve continuous-time nonlinear optimal tracking problem without system dynamics. *IET Control Theory Appl.* **2016**, *10*, 1339–1347. [CrossRef]

29. Marvi, Z.; Kiumarsi, B. Safe Off-policy Reinforcement Learning Using Barrier Functions. In Proceedings of the 2020 American Control Conference (ACC), Denver, USA, 1–3 July 2020; pp. 2176–2181.

30. Yang, Y.; Ding, D.; Xiong, H.; Yin, Y.; Wunsch, D. Online barrier-actor-critic learning for H_∞ control with full-state constraints and input saturation. *J. Frankl. Inst.* **2020**, *357*, 3316–3344. [CrossRef]

31. Ames, A.D.; Grizzle, J.W.; Tabuada, P. Control barrier function based quadratic programs with application to adaptive cruise control. In Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, CA, USA, 15–17 December 2014; pp. 6271–6278.

32. Yang, Y.; Vamvoudakis, K.; Modares, H. Safe reinforcement learning for dynamical games. *Int. J. Robust Nonlinear* **2020**, *30*, 3706–3726. [CrossRef]

33. Srinivasan, M.; Coogan, S. Control of Mobile Robots Using Barrier Functions Under Temporal Logic Specifications. *IEEE Trans. Robot.* **2021**, *37*, 363–374. [CrossRef]

34. Cohen, M.; Belta, C. Approximate Optimal Control for Safety-Critical Systems with Control Barrier Functions. In Proceedings of the 59th IEEE Conference on Decision and Control (CDC), Jeju, Korea, 14–18 December 2020; pp. 2062–2067.

35. Ames, A.D.; Coogan, S.; Egerstedt, M.; Notomista, G.; Sreenath, K.; Tabuada, P. Control Barrier Functions: Theory and Applications. In Proceedings of the 18th European Control Conference (ECC), Naples, Italy, 25–28 June 2019; pp. 3420–3431.

36. Xu, J.; Wang, J.; Rao, J.; Zhong, Y.; Wang, H. Adaptive dynamic programming for optimal control of discrete-time nonlinear system with state constraints based on control barrier function. *Int. J. Robust Nonlinear Control* **2022**, *32*, 3408–3424. [CrossRef]

37. 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]

38. Lewis, F.L.; Vrabie, D.L.; Syrmos, V.L. *Optimal Control*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012; pp. 277–279.

39. Lewis, F.L.; Jagannathan, S.; Yesildirek, A. *Neural Network Control of Robot Manipulators and Nonlinear Systems*; Taylor& Francis: London, UK, 1999.