



Rui Luo<sup>1</sup>, Zhinan Peng<sup>1,2,\*</sup> and Jiangping Hu<sup>1,3</sup>

- <sup>1</sup> School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; nicole9922@163.com (R.L.); hujp@uestc.edu.cn (J.H.)
- <sup>2</sup> Institute of Electronic and Information Engineering, University of Electronic Science and Technology of China, Dongguan 523808, China
- <sup>3</sup> Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China
- \* Correspondence: zhinanpeng@uestc.edu.cn

Abstract: This paper presents an adaptive learning structure based on neural networks (NNs) to solve the optimal robust control problem for nonlinear continuous-time systems with unknown dynamics and disturbances. First, a system identifier is introduced to approximate the unknown system matrices and disturbances with the help of NNs and parameter estimation techniques. To obtain the optimal solution of the optimal robust control problem, a critic learning control structure is proposed to compute the approximate controller. Unlike existing identifier-critic NNs learning control methods, novel adaptive tuning laws based on Kreisselmeier's regressor extension and mixing technique are designed to estimate the unknown parameters of the two NNs under relaxed persistence of excitation conditions. Furthermore, theoretical analysis is also given to prove the significant relaxation of the proposed convergence conditions. Finally, effectiveness of the proposed learning approach is demonstrated via a simulation study.

**Keywords:** optimal robust control; nonlinear systems; parameter estimation; neural networks learning; relaxed PE conditions

# 1. Introduction

In the past several decades, much attention has been given to  $H_{\infty}$  control problems, wherein the aim is to eliminate the influence of disturbance on the system.  $H_{\infty}$  control mainly focuses on designing a robust controller to regulate and stabilize the system. In practice, we should not only focus on the control performance, but also consider the optimization of the system [1,2]. Therefore, optimal  $H_{\infty}$  control problems will always be a hot research topic.

Adaptive dynamic programming (ADP), as one of the optimal control methods, has emerged as a powerful tool through which to deal with the optimal control problems of all kinds of dynamic systems [3]. The ADP framework combines dynamic programming and neural network approximation, and it has strong learning and adaptive ability. In this sense, ADP has rapidly developed in the control community in recent years. Generally speaking, the core of controller designs mainly concentrates on solving a Hamilton–Jacobi–Bellman (HJB) equation for nonlinear systems or an algebraic Riccati equation for linear systems [4]. Unfortunately, the HJB equation contains nonlinear, partial differential parts, which are difficult to solve directly [5]. Therefore, many efforts have been made for finding approximate solutions to the HJB equation using iterative or learning methods. Regarding the case of iterative methods, the ADP can be classed into two categories: value iteration (VI) [6,7] and policy iteration (PI) [8,9]. Regarding the case of learning-based methods, neural network (NN) approximation is generally utilized to learn the optimal or suboptimal solutions to the HJB equation. The standard learning frameworks include the following: actor–critic



Citation: Luo, R.; Peng, Z.; Hu, J. Optimal Robust Control of Nonlinear Systems with Unknown Dynamics via NN Learning with Relaxed Excitation. *Entropy* **2024**, *26*, 72. https://doi.org/ 10.3390/e26010072

Academic Editor: Tao Zhao

Received: 11 December 2023 Revised: 4 January 2024 Accepted: 10 January 2024 Published: 14 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). NNs and only-critic NNs. However, the abovementioned pieces of literature require partial or full model information in the controller design loop. To avoid relying on system models, many data-driven or model-free methods have been developed for improving the existing ADP frameworks, that is, data-driven RL [7], integral RL (IRL) [10,11], and system identification-based ADP methods [12–14].

More recently, excellent development has been realized with the use of ADP for the robust controller designs of optimal  $H_{\infty}$  control problems [15–17]. The main way through which to solve optimal  $H_{\infty}$  control problems is to model such problems as a two-player zero-sum game (min–max optimization problem), where the controller and the disturbance are viewed as players that try to find a controller to minimize the performance index function in worst-case disturbance conditions [18,19]. However, the disadvantage of zero-sum games is in judging the existence of the saddle point, which is generally difficult to judged. In order to overcome this issue, an indirect method motivated by [20] was developed by formulating an optimal regulation for a nominal system with new designs of the cost/value function [21]. For instance, Yang et al. proposed an event-triggered robust control strategy for nonlinear systems [22] using the indirect method. Xue et al. studied a tracking control problem for partial continuous-time systems with uncertainties and constraints [23] by transforming the robust control problem into an optimal regulation of nominal systems.

However, the existing results on  $H_{\infty}$  optimal control designs have two main characteristics: (1) their controller designs are based on the assumption that the complete or partial knowledge of the system dynamics are known in advance; however, (2) to address this issue, some system identification methods have been proposed, such as the identifier–criticor identifier–actor–critic-based designs of  $H_{\infty}$  optimal control. However, it is generally required that the persistence of excitation (PE) condition must be satisfied to ensure the learning performance of the weight updating of neural networks, which is difficult to check online in practice [18,19,23]. Therefore, how to weaken the PE condition is also the research motivation of this paper.

From the abovementioned observations and considerations, in this paper, we propose a novel online parameter estimation method based on an identifier–critic learning control framework for the  $H_{\infty}$  optimal control of nonlinear systems that have unknown dynamics with relaxed PE conditions. The contributions of our work can be summarized as follows:

- A new online identifier-critic learning control framework with a relaxed PE condition is proposed to address robust control for unknown continuous-time systems subject to unknown disturbances. To reconstruct the information of the system dynamics, neural networks combined with the linear regressor method are established to approximate the unknown system dynamics and disturbances.
- The approach in this paper is different from the existing weight adaption laws [18,19,23], where the PE condition is needed to ensure the learning performance of the NN's weight parameters. However, such a condition is difficult to check online, and a general way through which to satisfy this condition is to add external noise to the controller, which may lead to the instability of the system. To overcome this issue, a Kreisselmeier regressor extension and mixing (KREM)-based weight adaption law is designed for identifier–critic NNs with new convergence conditions.
- Weak PE properties of new convergence conditions are analyzed rigorously compared to traditional PE conditions. Moreover, the theoretical results indicate that the closedloop system's stability and the convergence of identifier–critic learning are guaranteed.

The remainder of this article is organized as follows. In Section 2, some preliminaries are introduced and the optimal robust control problem of nonlinear continuous-time systems is given. Then, a system identifier design with a relaxed PE condition is constructed in Section 3. Section 4 gives the critic NN design for robust control under a relaxed PE condition. Theoretical analyses of the weak PE properties under new convergence conditions and the stability of the closed-loop systems are given in Section 5. The simulation results are provided in Section 6. Some conclusions are summarized in Section 7.

## 2. Preliminaries and Problem Formulation

In this section, some notation and definitions are first introduced. Then, the optimal robust control problem of the nonlinear continuous-time systems is described.

#### 2.1. Preliminaries

To facilitate readability, some notations are listed.

- $\lambda(\cdot)$  Eigenvalue of a matrix
- $\{\cdot\}^*$  Adjoint matrix
- *I<sub>n</sub>* Identity matrix
- $tr(\cdot)$  Trace of a matrix
- $\lambda_M(\cdot)$  Maximum eigenvalues
- $\lambda_m(\cdot)$  Minimum eigenvalues

The following definitions will be used in the sequel.

**Definition 1** (*Persistence of Excitation* [24]). *A bounded signal*  $\psi(t)$  *is said to be PE, if there exist positive constants T and*  $\delta_1$  *such that* 

$$\int_t^{t+T} \psi(r) \psi^T(r) dr \ge \delta_1 I.$$

For clarity, we indicate that  $\psi(t)$  satisfies the PE condition using the notation  $\psi(t) \in PE$ ; otherwise,  $\psi(t) \notin PE$ .

**Definition 2** (Uniformly Ultimately Bounded [24]). The time function x(t) is said to be uniformly ultimately bounded (UUB) on a compact set  $\Omega_x$ , if, for all  $x(t_0) = x_0 \in \Omega_x$ , there exists a  $\delta_2 > 0$  and a number  $T(\delta_2, x_0)$  such that  $||x(t)|| < \delta_2$  for all  $t \ge t_0 + T$ .

### 2.2. Problem Formulation

Consider the nonlinear continuous-time (NCT) systems with disturbances described by the following dynamics:

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

where  $x(t) \in \mathbb{R}^n$  and  $u(t) \in \mathbb{R}^m$  denote the system state and control input, respectively.  $d(t) \in \mathbb{R}^q$  represents the external disturbance. The terms  $f(x) \in \mathbb{R}^n$ ,  $g(x) \in \mathbb{R}^{n \times m}$ , and  $G(x) \in \mathbb{R}^{n \times q}$  are the drift dynamics, input dynamics, and disturbance injection dynamics, respectively. In this study, f(x), g(x), and G(x) are assumed to be unknown. Furthermore, it is assumed that f(x), g(x), and G(x) are Lipschitz continuous with f(0) = 0, and that the system (1) is stabilizing and controllable.

The goal of this study is to solve an  $H_{\infty}$  control problem for the system (1). This problem can be equivalently transformed into a two-player zero-sum game, where the control input u(t) acts as the minimizing player and the disturbance d(t) acts as the maximizing player. The solution to the  $H_{\infty}$  control problem corresponds to a saddle point in the game, which stabilizes the equilibrium of the two-player zero-sum game.

Define the infinite-horizon performance index function as

$$V(x,u,d) = \int_{t}^{\infty} \left( x^{T} Q x + u^{T} R u - \kappa^{2} d^{T} d \right) d\tau,$$
(2)

where  $\kappa > 0$ , V(0) = 0, and Q and R are symmetric positive-definite matrices with appropriate dimensions. Let  $u^*$  be the optimal control input and  $d^*$  be the worst disturbance. Our objective is to find the saddle point  $(u^*, d^*)$  that optimizes the performance index (2), which can be more precisely clarified by the following inequality:

$$V(u^{\star},d) \le V(u^{\star},d^{\star}) \le V(u,d^{\star}).$$
(3)

We then define the optimal performance index function  $V^*$  as follows:

$$V^{\star}(x,u,d) = \min_{u} \max_{d} \int_{t}^{\infty} \left( x^{T}Qx + u^{T}Ru - \kappa^{2}d^{T}d \right) d\tau.$$
(4)

The Hamiltonian of system (1) can be written as

$$H(V_x, x, u, d) = V_x^T[f(x) + g(x)u + G(x)d] + x^TQx + u^TRu - \kappa^2 d^Td,$$
(5)

where  $V_x = \partial V / \partial x \in \mathbb{R}^n$ . The Hamilton–Jacobi–Isaacs (HJI) equation related to this game has the form

$$\min_{u} \max_{d} H(V_x^{\star}, x, u, d) = 0, \tag{6}$$

where  $V_x^* = \partial V^* / \partial x \in \mathbb{R}^n$ . Based on the stationarity condition, the  $H_\infty$  control pair  $(u^*, d^*)$  for (1) has the following form:

$$u^{\star} = -\frac{1}{2}R^{-1}g^{T}(x)V_{x}^{\star}(x), \tag{7}$$

$$d^{\star} = \frac{1}{2\kappa^2} G^T(x) V_x^{\star}(x).$$
(8)

Thus, according to (7) and (8), the HJI Equation (6) can be rewritten as

$$x^{T}Qx + V_{x}^{\star T}f(x) - \frac{1}{4}V_{x}^{\star T}g(x)R^{-1}g^{T}(x)V_{x}^{\star} + \frac{1}{4\kappa^{2}}V_{x}^{\star T}G(x)G^{T}(x)V_{x}^{\star} = 0.$$
(9)

Indeed, the HJI Equation (9) represents a highly nonlinear partial differential equation (PDE) and requires complete system information for its resolution. To address these challenges, a new IC framework with relaxed PE conditions will be proposed in the following sections. Furthermore, new adaptive update laws for the identifier and critic NNs are provided with the help of the KREM technique. The block diagram of the proposed control system is shown in Figure 1, and detailed theoretical analysis will be presented in subsequent sections.



Figure 1. Schematic of the proposed control system.

# 3. System Identifier Design with Relaxed PE Condition

In this section, an NN-based identifier is utilized to reconstruct the unknown system dynamics in (1). The KREM technique is introduced to adjust the identifier weights under

relaxed PE conditions. We assume that the unknown system dynamics f(x), g(x), and G(x) in (1) are continuous functions defined on compact sets. The NN-based identifier is designed as follows:

$$f(x) = W_f \theta_f(x) + \epsilon_f, \tag{10}$$

$$g(x) = W_g \theta_g(x) + \epsilon_g, \tag{11}$$

$$G(x) = W_G \theta_G(x) + \epsilon_G, \tag{12}$$

where  $W_f \in \mathcal{R}^{n \times d_f}$ ,  $W_g \in \mathcal{R}^{n \times d_g}$  and  $W_G \in \mathcal{R}^{n \times d_G}$  are the ideal NN weights;  $\theta_f(x) \in \mathcal{R}^{d_f}$ ,  $\theta_g(x) \in \mathcal{R}^{d_g \times m}$  and  $\theta_G(x) \in \mathcal{R}^{d_G \times q}$  are the basis functions; and  $\epsilon_f \in \mathcal{R}^n$ ,  $\epsilon_g \in \mathcal{R}^{n \times m}$  and  $\epsilon_G \in \mathcal{R}^{n \times q}$  are the reconstruction errors. Then, according to the Weierstrass theorem and the statements in [10], the approximation errors  $\epsilon_f$ ,  $\epsilon_g$ , and  $\epsilon_G$  can be shown to approach zero as the number of NN neurons  $d_f$ ,  $d_g$ , and  $d_G$  increases to infinity.

Before proceeding, it is essential to establish the following underlying assumption.

# Assumption 1.

- (1) The basis functions  $\theta_f(x)$ ,  $\theta_g(x)$  and  $\theta_G(x)$  are bounded, that is,  $\|\theta_f(x)\| \le b_{\theta_f}$ ,  $\|\theta_g(x)\| \le b_{\theta_G}$ , respectively.
- (2) The reconstruction errors  $\varepsilon_f$ ,  $\varepsilon_g$  and  $\varepsilon_G$  are bounded, that is,  $\|\varepsilon_f\| \leq b_{\varepsilon_f}$ ,  $\|\varepsilon_g\| \leq b_{\varepsilon_g}$ ,  $\|\varepsilon_G\| \leq b_{\varepsilon_G}$ , respectively.

Using (10)–(12), the system (1) can be rewritten as

$$\dot{x} = W_I^I \theta_I(x, u) + \epsilon_T, \tag{13}$$

where  $W_I = [W_f, W_g, W_G]^T \in \mathcal{R}^{d \times n}$  is the augmented weight matrix with  $d = d_f + d_g + d_G$ , and  $\theta_I(x, u) = [\theta_f^T(x), u^T \theta_g^T(x), d^T \theta_G^T(x)]^T \in \mathcal{R}^d$  is the augmented regressor vector.  $\epsilon_T = \epsilon_f + \epsilon_g u + \epsilon_G d \in \mathcal{R}^n$  is the model approximation error.

Note that  $\dot{x}$  and  $W_I$  are unknown. Therefore, we define the filtered variables  $x_f$  and  $\theta_{If}$  as

$$\begin{cases} \rho \dot{x}_f + x_f = x, \quad x_f(0) = 0\\ \rho \dot{\theta}_{If} + \theta_{If} = \theta_I, \quad \theta_{If}(0) = 0 \end{cases}$$
(14)

where  $\rho \in \mathcal{R} > 0$  is the filter coefficient. From Equations (13) and (14), we can deduce that

$$\dot{x}_f = \frac{x - x_f}{\rho} = W_I^T \theta_{If} + \epsilon_{Tf}, \tag{15}$$

where  $\epsilon_{Tf}$  denotes the filtered version of  $\epsilon_T$  as  $\rho \dot{\epsilon}_{Tf} + \epsilon_{Tf} = \epsilon_T$ . Clearly, (15) is a linear regression equation (LRE), where  $\dot{x}_f$  and  $\theta_{If}$  can be calculated from (14). In the following, we describe how the KREM technique is applied to estimate  $W_I$  by using the measured information  $\dot{x}_f$  and  $\theta_{If}$ .

To approximate the unknown weights  $W_I$  in (15) such that the estimated weights  $\hat{W}_I$  converge to their true values under a relaxed PE condition, we aim to construct an extended LRE (E-LRE) based on (15). We define the matrices  $P_I \in \mathbb{R}^{d \times d}$  and  $Q_I \in \mathbb{R}^{d \times n}$  as follows:

$$\begin{cases} P_I = \mathcal{H}_I[\theta_{If}\theta_{If}^T], P_I(0) = 0\\ Q_I = \mathcal{H}_I[\theta_{If}(\frac{x - x_f}{\rho})^T], Q_I(0) = 0 \end{cases}$$
(16)

where

$$\mathcal{H}_I = \frac{1}{p+l_I}[s](t)$$

$$\begin{cases} P_{I} = \int_{0}^{t} e^{-l_{I}(t-\tau)} \theta_{If}(\tau) \theta_{If}^{T}(\tau) d(\tau), \\ Q_{I} = \int_{0}^{t} e^{-l_{I}(t-\tau)} \theta_{If}(\tau) (\frac{x(\tau) - x_{f}(\tau)}{\rho})^{T} d(\tau). \end{cases}$$
(17)

Note that it can be verified that  $P_I$  and  $Q_I$  are bounded for any given bounded  $\theta_I$  and x due to the appropriate choice of  $l_I$ . Thus, an E-LRE is obtained

$$Q_{I}(t) = P_{I}(t)W_{I} + v_{I},$$
(18)

where  $v_I = \int_0^t e^{-l_I(t-\tau)} \theta_{If}(\tau) \epsilon_{Tf}^T(\tau) d(\tau)$ .

To construct an identifier weight error dynamics that achieves better convergence properties, we define the variables  $Q_I(t) \in \mathbb{R}^{d \times n}$ ,  $\mathcal{P}_I \in \mathbb{R}^{d \times d}$ , and  $\mathcal{V}_I \in \mathbb{R}^{d \times n}$  as follows:

$$\begin{cases}
\mathcal{Q}_I = P_I^* Q_I, \\
\mathcal{P}_I = P_I^* P_I, \\
\mathcal{V}_I = P_I^* v_I.
\end{cases}$$
(19)

Then Equation (18) becomes

$$Q_I(t) = \mathcal{P}_I(t)W_I + \mathcal{V}_I.$$
(20)

Note that for any square matrix  $M \in \mathcal{R}^{q \times q}$ , we have  $M^*M = |M|I_q$ , even if M is not full rank. Thus,  $\mathcal{P}_I = |P_I|I_d \in \mathcal{R}^{d \times d}$ . Moreover,  $\mathcal{P}_I$  is a scalar diagonal matrix, where (20) can be decoupled into a series of scalar LREs:

$$Q_{I(i,j)}(t) = |P_I|(t)W_{I(i,j)} + \mathcal{V}_{I(i,j)}, i = 1, \dots, d, j = 1, \dots, n,$$
(21)

where  $Q_{I(i,j)}$  and  $W_{I(i,j)}$  indicate the *i*th row and *j*th column of  $Q_I$  and  $W_I$ , respectively.

Then, the estimation algorithm for the unknown identifier NN weights can be designed based on (21) as follows:

$$\hat{W}_{I(i,j)} = -\gamma_1 |P_I| [|P_I| \hat{W}_{I(i,j)} - Q_{I(i,j)}],$$
(22)

where  $\gamma_1 \in \mathcal{R} > 0$  presents the adaptive learning gain.

The convergence of identifier (22) can be given as follows.

**Theorem 1.** Consider the system (13) with the online update law (22); if  $|P_I| \in PE$ , then

- (*i*) for  $\epsilon_T = 0$ , the estimator error  $\tilde{W}_{I(i,j)}$  converges to zero exponentially;
- (ii) for  $\epsilon_T \neq 0$ , the estimator error  $\tilde{W}_{I(i,j)}$  converges to a compact set around zero.

**Proof.** If  $|P_I| \in PE$ , according to Definition 1 we have  $\int_t^{t+T} |P_I|^2 dr \ge \delta_I > 0$ . Defining the estimation error  $\tilde{W}_{I(i,j)} = \hat{W}_{I(i,j)} - W_{I(i,j)}$ , i = 1, ..., d, j = 1, ..., n. Due to (21) and (22), the identifier weight error dynamics can be obtained

$$\tilde{W}_{I(i,j)} = -\gamma_1 |P_I|^2 \tilde{W}_{I(i,j)} + \gamma_1 |P_I| \mathcal{V}_{I(i,j)}.$$
(23)

Considering the Lyapunov function  $V_I = 0.5\gamma_1^{-1}\tilde{W}_{I(i,j)}^2$ , the derivation of  $V_I$  can be calculated as

$$\dot{V}_{I} = \frac{1}{\gamma_{1}} \tilde{W}_{I(i,j)} \dot{\tilde{W}}_{I(i,j)} 
= -|P_{I}|^{2} \tilde{W}_{I(i,j)}^{2} + |P_{I}| \tilde{W}_{I(i,j)} \mathcal{V}_{I(i,j)}.$$
(24)

In fact, when  $\epsilon_T = 0$ , (24) can be rewritten as

$$\dot{V}_I = -|P_I|^2 \tilde{W}_{I(i,j)}^2 < -\mu_I V_I,$$
(25)

where  $\mu_I = 2\gamma_1 \delta_I > 0$ . According to the Lyapunov theorem, the weight estimation error  $\tilde{W}_{I(i,j)}$  exponentially converges to zero.

When  $\epsilon_T \neq 0$ , (24) can be further presented as

$$\begin{split} \dot{V}_{I} &= -|P_{I}|^{2} \tilde{W}_{I(i,j)}^{2} + |P_{I}| \tilde{W}_{I(i,j)} \mathcal{V}_{I(i,j)} \\ &= -[|P_{I}|^{2} \tilde{W}_{I(i,j)} - |P_{I}| \mathcal{V}_{I(i,j)}] \tilde{W}_{I(i,j)}. \end{split}$$
(26)

According to Assumption 1,  $|P_I|\mathcal{V}_{I(i,j)}$  is bounded, denoted as  $||P_I|\mathcal{V}_{I(i,j)}|| < b_{P_I\mathcal{V}_I}$ . Then,

$$\dot{V}_{I} \leq -[|P_{I}|^{2}||\tilde{W}_{I(i,j)}|| - b_{P_{I}\mathcal{V}_{I}}]||\tilde{W}_{I(i,j)}||.$$
(27)

According to the extended Lyapunov theorem, the estimation error  $\hat{W}_{I(i,j)}$  uniformly ultimately converges to a compact set  $\{\tilde{W}_{I(i,j)}|||\tilde{W}_{I(i,j)}|| \leq b_{P_I \mathcal{V}_I}/p_I^2\}$ .  $\Box$ 

**Remark 1.** In [12], the update law for the unknown weight  $W_I$  was designed based on (18), while the PE condition (i.e.,  $\theta_I \in PE$ ) was required to ensure convergence. However, satisfying the PE condition is generally challenging. In Theorem 1, we provide a new convergence condition  $|P_I| \in PE$ . Notably, this new condition is significantly superior to the conventional PE condition for two reasons. (1) We theoretically prove that  $|P_I| \in PE$  is much weaker than  $\theta_I \in PE$ , as detailed in Section 5. (2)  $|P_I|$  is directly related to the determinant of the matrix  $P_I(t)$ . Therefore, checking  $|P_I| \in PE$  online becomes feasible by calculating the determinant of  $P_I(t)$ . In contrast, assessing the standard PE condition directly online is not possible [18,19,23].

Based on the above analysis, the unknown information f(x), g(x), and G(x) can be estimated using (13) and (22). This allows for the reconstruction of the completely unknown system dynamics. In order to obtain the optimal  $H_{\infty}$  control pair, the critic NN will be introduced to learn the solution of the HJB equation in the subsequent section.

## 4. Critic NN Design for $H_{\infty}$ Control under Relaxed PE Condition

In this section, the performance index will be approximated via a critic NN to obtain the optimal  $H_{\infty}$  control pair. The KREM algorithm will be continually utilized to design the update law of critic NN under the relaxed PE condition. Firstly, based on the above identifier, the system (1) can be represented as

$$\dot{x} = \hat{W}_f \theta_f(x) + \hat{W}_g \theta_g(x) u + \hat{W}_G \theta_G(x) d(t) + \epsilon_I + \epsilon_T,$$
(28)

where  $\hat{W}_f$ ,  $\hat{W}_g$  and  $\hat{W}_G$  are the estimated values of  $W_f$ ,  $W_g$  and  $W_G$ , respectively.  $\epsilon_I = \tilde{W}_I \theta_I$  denotes the identifier error. And, the Hamiltonian (5) can be further written as

$$H = V_x^T [\hat{W}_f \theta_f(x) + \hat{W}_g \theta_g(x)u + \hat{W}_G \theta_G(x)d(t) + \epsilon_I + \epsilon_T] + x^T Q x + u^T R u - \kappa^2 d^T d.$$
<sup>(29)</sup>

Then, the HJI Equation (6) becomes

$$0 = \min_{u} \max_{d} [H(V_x^*, x, u^*, d^*)]$$
  
=  $V_x^{*T} [\hat{W}_f \theta_f(x) + \hat{W}_g \theta_g(x) u^* + \hat{W}_G \theta_G(x) d^*(t) + \epsilon_I + \epsilon_T] + x^T Q x + u^{*T} R u^* - \kappa^2 d^{*T} d^*.$  (30)

Therefore, based on (30), the  $H_{\infty}$  control pair  $(u^*, d^*)$  for the estimated system (28) can be expressed as follows:

$$u^{\star} = -\frac{1}{2} R^{-1} [\hat{W}_g \theta_g]^T V_x^{\star}, \qquad (31)$$

$$d^{\star} = \frac{1}{2\kappa^2} [\hat{W}_G \theta_G]^T V_x^{\star}(x). \tag{32}$$

Since the HJI Equation (30) is a nonlinear PDE, similar to (6), we utilize a critic NN to estimate  $V^*(x)$  and its gradient  $V^*_x(x)$  as follows:

$$V^{\star}(x) = W_c^T \theta_c(x) + \epsilon_v, \tag{33}$$

$$V_x^{\star}(x) = \nabla \theta_c^T(x) W_c + \nabla \epsilon_v, \tag{34}$$

where  $W_c \in \mathbb{R}^l$  is the unknown constant weight.  $\theta_c(x) \in \mathbb{R}^l$  represents the independent basis function with  $\nabla \theta_c(x) = \partial \theta_c / \partial x$ . *l* is the number of neurons. The approximation error is presented as  $\epsilon_v$  with  $\nabla \epsilon_v = \partial \epsilon_v / \partial x$ . Note that as the number of independent basis functions increases, both the approximation errors and their gradients can approach zero. Before proceeding, the following assumption is needed.

Assumption 2.

- (1) The ideal critic NN's weight  $W_c$  is bounded, that is,  $||W_c|| < b_{W_c}$ .
- (2) The basis functions  $\theta_c(x)$  and its gradients  $\nabla \theta_c(x)$  are bounded, that is,  $\|\theta_c\| \le b_{\theta_c}$ ,  $\|\nabla \theta_c\| \le b_{\nabla \theta_c}$ .
- (3) The approximator reconstruction error  $\epsilon_v$  and its gradients  $\nabla \epsilon_v$  are bounded, that is,  $\|\epsilon_v\| \le b_{\epsilon_v}$ ,  $\|\nabla \epsilon_v\| \le b_{\nabla \epsilon_v}$ .

Since the ideal critic NN weights  $W_c$  are unknown, take  $\hat{W}_c$  as the estimated value of  $W_c$  and  $\hat{V}$  as the estimated value of V, where the practical critic NN is given by

<u>л</u>т

$$V(x) = W_c^T \theta_c(x),$$
  

$$\hat{V}_x(x) = \nabla \theta_c^T(x) \hat{W}_c.$$
(35)

The estimated  $H_{\infty}$  control pair  $\hat{u}$  and  $\hat{d}$  can be obtained as

$$\hat{u} = -\frac{1}{2}R^{-1}[\hat{W}_{g}\theta_{g}]^{T}\hat{V}_{x} = -\frac{1}{2}R^{-1}[\hat{W}_{g}\theta_{g}]^{T}\nabla\theta_{c}^{T}\hat{W}_{c},$$
(36)

$$\hat{d} = \frac{1}{2\kappa^2} [\hat{W}_G \theta_G]^T \hat{V}_x = \frac{1}{2\kappa^2} [\hat{W}_G \theta_G]^T \nabla \theta_c^T \hat{W}_c.$$
(37)

To online estimate the unknown weights of the critic NN using KREM technology, we aim to construct a linear equation according to (30) and (34) as

$$\epsilon_{HJI} + x^T Q x + \hat{u}^T R \hat{u} - \kappa^2 \hat{d}^T \hat{d} + W_c^T \nabla \theta_c \hat{W}_f \theta_f(x) + W_c^T \nabla \theta_c \hat{W}_g \theta_g(x) \hat{u} + W_c^T \nabla \theta_c \hat{W}_G \theta_G(x) \hat{d} = 0,$$
(38)

where  $\epsilon_{HJI} = W_c^T \nabla \theta_c(\epsilon_I + \epsilon_T) + \nabla \epsilon_v^T (\hat{W}_f \theta_f + \hat{W}_g \theta_g \hat{u} + \hat{W}_G \theta_G \hat{d} + \epsilon_I + \epsilon_T)$  is a bounded residual HJI equation error. Let  $\Theta = \nabla \theta_c [\hat{W}_f \theta_f + \hat{W}_g \theta_g \hat{u} + \hat{W}_G \theta_G \hat{d}]$  and  $\Sigma = x^T Q x + \hat{u}^T R \hat{u} - \kappa^2 \hat{d}^T \hat{d}$ , where a linear equation is obtained as follows:

$$\Sigma = -W_c^T \Theta - \epsilon_{HII}.$$
(39)

Similar to the previous section, we define the filtered regressor matrix  $P_c \in \mathcal{R}^{l \times l}$  and the vector  $Q_c \in \mathcal{R}^l$  as follows:

$$\begin{cases} P_c = \mathcal{H}_c[\Theta\Theta^T], \ P_c(0) = 0\\ Q_c = \mathcal{H}_c[\Theta\Sigma], \ Q_c(0) = 0 \end{cases}$$
(40)

where

$$\mathcal{H}_c = \frac{1}{p+l_c}[s](t),$$

and  $l_c > 0$  is the forgetting factor. Then, the solution of (40) can be deduced as

$$\begin{cases} P_c = \int_0^t e^{-l_c(t-\tau)} \Theta \Theta^T d\tau, \\ Q_c = \int_0^t e^{-l_c(t-\tau)} \Theta \Sigma d\tau. \end{cases}$$
(41)

From (39) and (41), an E-LRE related to  $P_c$  and  $Q_c$  is obtained

$$Q_{c}(t) = -P_{c}(t)W_{c} - v_{c},$$
(42)

where  $v_c = \int_0^t e^{-l_c(t-\tau)} \Theta(\tau) \epsilon_{HJI}^T(\tau) d\tau$  is bounded. To estimate the unknown parameter  $W_c$  in (42) under a relaxed PE condition, define the variables  $Q_c(t) \in \mathcal{R}^l$ ,  $\mathcal{P}_c \in \mathcal{R}^{l \times l}$ , and  $\mathcal{V}_c \in \mathcal{R}^l$  as

$$\begin{array}{l}
\mathcal{Q}_c = P_c^* \mathcal{Q}_c, \\
\mathcal{P}_c = P_c^* \mathcal{P}_c, \\
\mathcal{V}_c = P_c^* v_c.
\end{array}$$
(43)

Then Equation (42) becomes

$$Q_c(t) = -\mathcal{P}_c(t)W_c - \mathcal{V}_c. \tag{44}$$

Note that  $\mathcal{P}_c = |P_c|I_l$ . Since  $\mathcal{P}_c$  is a scalar matrix, a series of scalar LREs is obtained as

$$Q_{c(i)}(t) = -|P_c|(t)W_{c(i)} - \mathcal{V}_{c(i)}, i = 1, \dots, l,$$
(45)

where  $Q_{c(i)}$ ,  $W_{c(i)}$  and  $V_{c(i)}$  indicate the *i*th rows of  $Q_c$ ,  $W_c$ , and  $V_c$ , respectively.

Driven by the parameter error based on (45), the critic unknown weight  $W_{c(i)}$  is designed as

$$\hat{W}_{c(i)} = -\gamma_2 |P_c| [|P_c| \hat{W}_{c(i)} + Q_{c(i)}], \tag{46}$$

where  $\gamma_2 \in \mathcal{R} > 0$  presents the adaptive learning gain.

The convergence condition for the proposed critic NN adaptive law is provided in Theorem 2.

**Theorem 2.** For adaptive law (46) of critic NN with the regressor matrix  $\mathcal{P}_c$  in (44); if  $|P_c| \in PE$ , then

(i) for  $\epsilon_{HII} = 0$ , the estimator error  $\tilde{W}_{c(i)}$  converges to zero exponentially;

(ii) for  $\epsilon_{HII} \neq 0$ , the estimator error  $\tilde{W}_{c(i)}$  converges to a compact set around zero;

**Proof.** Defining the estimation error  $\tilde{W}_{c(i)} = \hat{W}_{c(i)} - W_{c(i)}$ , i = 1, ..., l. The proofs presented in Theorem 1 can be extended to establish similar results in the current context. Note that the Lyapunov function  $V_c$  here is chosen as  $0.5\gamma_2^{-1}\tilde{W}_{c(i)}^2$ .  $\Box$ 

**Remark 2.** According to Theorem 2, a new convergence condition for the estimation error of the critic neural network weights, denoted as  $\tilde{W}_c$ , is provided. This condition does not rely on the conventional parameter estimation (PE) condition, i.e.,  $\Theta \in PE$ . In this paper, the additional exploration signal is not required to guarantee  $\Theta \in PE$ . Instead, the satisfaction of  $|P_c| \in PE$  can be achieved by adjusting the forgetting factor  $l_c$ . It is worth noting that the new convergence condition is associated with the matrix  $P_c$ , and it can be verified online by calculating the determinant of  $P_c$ . The proof of the weak PE property for the new convergence condition will be presented in the following section.

**Remark 3.** The convergence analysis of  $\tilde{W}_{I(i,j)}$  and  $\tilde{W}_{c(i)}$  are provided in Theorem 1 and Theorem 2, respectively. In fact, we can derive the convergence of  $\tilde{W}_I$  and  $\tilde{W}_c$  using simple matrix operations, which will be omitted in this paper.

Till now, the identifier–critic learning-based framework for  $H_{\infty}$  optimal control under the relaxed PE condition is given. For clarity, the design details of the proposed method are shown in Algorithm 1, which can be considered the pseudocode for the simulation part.

#### **Algorithm 1** Identifier–critic learning-based $H_{\infty}$ optimal control algorithm

- 1: Initialization
- 2: Initialize system parameters: x(0), Q, R and running time T;
- 3: Set the identifier and critic filter operators:  $\mathcal{H}_I$  and  $\mathcal{H}_c$ ;
- 4: Set the basis functions of identifier and critic NNs:  $\theta_I(x, u)$  and  $\theta_c(x)$ ;
- 5: Initialize and set the filter operator parameters:  $\rho$ ,  $l_I$ ,  $l_c$ ,  $x_f(0) = 0$ ,  $\theta_{If}(0) = 0$  and  $\epsilon_{If}(0) = 0$ ;
- 6: Initialize identifier NNs parameters:  $\gamma_1 > 0$ ,  $\hat{W}_I^{initial} \in (0, 1]$ ;
- 7: Initialize critic NNs parameters:  $\gamma_2 > 0$ ,  $\hat{W}_c^{initial} \in (0, 1]$ ;
- 8: Initialize the control pair by (36) and (37);
- 9: while  $t \leq T$  do
- 10: Calculate the filter processing of the identifier NNs by (14);
- 11: Calculate the dynamic regressor extension (DRE) of the identifier NNs by (15);
- 12: Calculate the regressor "mixing" of the identifier NNs by (18);
- 13: Update the weight parameters of the identifier NNs  $\hat{W}_{I(i,j)}$  by (20);

$$\hat{W}_{I(i,j)} = -\gamma_1 |P_I| [|P_I| \hat{W}_{I(i,j)} - Q_{I(i,j)}];$$

- 14: Compute the approximated HJB equation by (39);
- 15: Calculate the dynamic regressor extension (DRE) of the critic NNs by (40);
- 16: Calculate the regressor "mixing" of the critic NNs by (42);
- 17: Update the weight parameters of the critic NNs  $W_{c(i)}$  by (46);

$$\hat{W}_{c(i)} = -\gamma_2 |P_c| [|P_c| \hat{W}_{c(i)} + Q_{c(i)}];$$

- 18: Update the control pair by (36) and (37);
- 19: Update the system states x by (28);

20: end while

#### 5. Stability and Convergence Analysis

In this section, we present the main results, which include the theoretical analysis of weak PE properties under new convergence conditions proposed in Theorems 1 and 2. Furthermore, we provide a stability result for the closed-loop system under the proposed online learning optimal control method.

To facilitate the analysis, the following assumption is made.

**Assumption 3.** The system dynamics in (1) satisfy  $||f(x)|| \le b_f ||x||$ ,  $||g(x)|| \le b_g$  and  $||G(x)|| \le b_G$ , where  $b_f > 0$ ,  $b_g > 0$  and  $b_G > 0$ .

#### 5.1. Weak PE Properties of New Convergence Conditions

As shown in Theorem 1, Theorem 2 and Remark 3, the convergence of  $W_I$  and  $W_c$  is established without the restrictive PE condition, i.e.,  $\theta_I \in PE$  and  $\Theta \in PE$ . These new convergence conditions can be easily checked online, as mentioned in Remarks 1 and 2. Furthermore, we will analyze the superiority of the new convergence conditions compared to the conventional PE condition from a theoretical standpoint.

**Theorem 3.** *Consider the system* (13) *with the online identifier NN adaptive law* (22) *and critic NN adaptive law* (46),

(*i*) The convergence condition of estimation error  $\tilde{W}_I$  in Theorem 1, that is,  $|P_c| \in PE$ , is weaker than  $\theta_I \in PE$  in the following precise sense

$$\theta_I(t) \in PE \Rightarrow |P_c| \in PE, \tag{47}$$

$$|P_c| \in PE \Rightarrow \theta_I(t) \in PE; \tag{48}$$

(ii) The convergence condition of estimation error  $\tilde{W}_c$  in Theorem 2, that is,  $|P_c| \in PE$ , is weaker than  $\Theta \in PE$  in the following precise sense

$$\Theta \in PE \Rightarrow |P_c| \in PE, \tag{49}$$

$$|P_c| \in PE \Rightarrow \Theta \in PE. \tag{50}$$

**Proof.** For (*i*), suppose that  $\theta_I(t)$  in (13) is PE, indicating that  $\theta_{If}(t) \in PE$  [25]. From Definition 1, we have

$$\int_{t}^{t+\tau} \theta_{If}(r) \theta_{If}^{T}(r) dr \ge \delta I \Leftrightarrow \int_{t-\tau}^{t} \theta_{If}(r) \theta_{If}^{T}(r) dr \ge \delta I \text{ for } t > \tau > 0.$$
(51)

Moreover, since  $e^{-\beta_I(t-r)} \ge e^{-\beta_I \tau} > 0$  with  $r \in [t - \tau, t]$ , the following inequality holds

$$\int_{t-\tau}^{t} e^{-\beta_{I}(t-r)} \theta_{If}^{T}(r) \theta_{If}(r) dr \ge \int_{t-\tau}^{t} e^{-\beta_{I}\tau} \theta_{If}^{T}(r) \theta_{If}(r) dr \ge e^{-\beta_{I}\tau} \delta I.$$
(52)

Furthermore, for  $t > \tau > 0$ , we also have

$$\int_{0}^{t} e^{-\beta_{I}(t-r)} \theta_{If}^{T}(r) \theta_{If}(r) dr > \int_{t-\tau}^{t} e^{-\beta_{I}(t-r)} \theta_{If}^{T}(r) \theta_{If}(r) dr.$$
(53)

From (17), (52) and (53), we conclude that

$$P_{I} = \int_{0}^{t} e^{-\beta_{I}(t-r)} \theta_{If}^{T}(r) \theta_{If}(r) dr > e^{-\beta_{I}\tau} \int_{t-\tau}^{t} \theta_{If}^{T}(r) \theta_{If}(r) dr \ge e^{-\beta_{I}\tau} \delta I.$$
(54)

Hence, the matrix  $P_I$  in (16) is positive definite, that is,  $\lambda_i(P_I) > 0$ , i = 1, ..., d. Considering that the determinant of a matrix is equal to the product of all its eigenvalues,

that is, 
$$|P_I| = \lambda_1(P_I)\lambda_2(P_I)\dots\lambda_d(P_I)$$
, we obtain  $\lambda_i(P_I) > 0 \Rightarrow \prod_{i=1}^{u} \lambda_i(P_I) > 0 \Rightarrow |P_I| > 0$ .  
Thus, (47) is true.

The proof of (48) is established by the following:

$$|P_{I}| \in PE \Leftrightarrow \int_{0}^{t} |P_{I}|^{2}(\tau)d\tau > 0$$
  
$$\Leftrightarrow \int_{0}^{t} \prod_{i=1}^{d} \lambda_{i}^{2}(P_{I}) > 0$$
  
$$\Rightarrow \lambda_{i}(P_{I}) > 0, i = 1, \dots, d$$
  
$$\Leftrightarrow P_{I} > 0$$
  
$$\Leftrightarrow P_{I} \in PE.$$
(55)

For (*ii*), the proof process can be referred to in (*i*). This finishes the proof.  $\Box$ 

## 5.2. Stability and Convergence Analysis

The stability result for the closed-loop system under the proposed online learning optimal control method will be presented in the following theorem.

**Theorem 4.** Let Assumptions 1 and 2 hold. Considering system (1) with the identifier weight tuning law given by (22), the  $H_{\infty}$  control pair are computed by (36) and (37), respectively. The critic NN weight tuning laws are updated by (46). If  $|P_I| \in PE$  and  $|P_c| \in PE$ , then the closed-loop system, system identifier estimation error  $\tilde{W}_I$ , and critic estimation error  $\tilde{W}_c$  are uniformly ultimately bounded (UUB). Moreover, the approximated  $H_{\infty}$  control pair given by (36) and (37) are close to the optimal control pair  $u^*$  and  $d^*$  within a small region  $b_u$  and  $b_d$ , that is,  $||\hat{u} - u^*|| \leq b_u$  and  $||\hat{d} - d^*|| \leq b_d$ , where  $b_u$  and  $b_d$  are positive constants.

**Proof.** We consider the Lyapunov function as follows:

$$\mathcal{J}(t) = \frac{1}{2} tr\{\tilde{W}_{I}^{T}(t)\gamma_{1}^{-1}\tilde{W}_{I}(t)\} + \frac{1}{2}\tilde{W}_{c}^{T}(t)\gamma_{2}^{-1}\tilde{W}_{c}(t) + \gamma_{3}x^{T}x + \gamma_{4}V^{\star}(x) 
+ \gamma_{5}tr\{\mathcal{V}_{I}^{T}\mathcal{V}_{I}\} + \gamma_{6}\mathcal{V}_{c}^{T}\mathcal{V}_{c} 
= \mathcal{J}_{1} + \mathcal{J}_{2} + \mathcal{J}_{3} + \mathcal{J}_{4} + \mathcal{J}_{5} + \mathcal{J}_{6},$$
(56)

where  $\gamma_3$ ,  $\gamma_4$ ,  $\gamma_5$  and  $\gamma_6$  are positive constants.

By applying matrix operations, we can obtain the following:

$$\begin{split} \dot{\tilde{W}}_I &= -\gamma_1 |P_I| [|P_I| \tilde{W}_I - \mathcal{V}_I], \\ \dot{\tilde{W}}_c &= -\gamma_2 |P_c| [|P_c| \tilde{W}_c - \mathcal{V}_c]. \end{split}$$
(57)

According to Definition 1,  $|P_I| \in PE$  and  $|P_c| \in PE$  imply that  $\int_t^{t+T} |P_I|^2 dr \ge \delta_I > 0$ and  $\int_t^{t+T} |P_c|^2 dr \ge \delta_c > 0$ . Substituting (19), (43), and using Young's inequality  $ab \le a^2\eta/2 + b^2/2\eta$  with  $\eta > 0$ , we have

$$\dot{\mathcal{J}}_{1} = tr \left\{ -\tilde{W}_{I}^{T} |P_{I}|^{2} \tilde{W}_{I} + |P_{I}| \tilde{W}_{I}^{T} \mathcal{V}_{I} \right\} \leq -\left(\delta_{I} - \frac{1}{2\eta}\right) \left\| \tilde{W}_{I} \right\|^{2} + \frac{\eta}{2} b_{P_{I}^{*}}^{2} \left\| |P_{I}| v_{I} \right\|^{2}, \tag{58}$$

$$\dot{\mathcal{J}}_{2} = -\tilde{W}_{c}^{T}|P_{c}|^{2}\tilde{W}_{c} + |P_{c}|\tilde{W}_{c}^{T}\mathcal{V}_{c} \leq -\left(\delta_{c} - \frac{1}{2\eta}\right)\left\|\tilde{W}_{c}\right\|^{2} + \frac{\eta}{2}b_{P_{c}^{*}}^{2}\left\||P_{c}|v_{c}\right\|^{2}.$$
(59)

where  $\|P_I^*\| \leq b_{P_I^*}$ ,  $\|P_c^*\| \leq b_{P_c^*}$ . For  $\mathcal{J}_3$  and  $\mathcal{J}_4$ ,

$$\begin{aligned} \dot{\mathcal{J}}_{3} + \dot{\mathcal{J}}_{4} &= 2\gamma_{3}x^{T}\dot{x} + \gamma_{4}\dot{V}^{\star}(x) \\ &= 2\gamma_{3}x^{T} \Big[ f(x) + g(x)\hat{u} + G(x)\hat{d} - g(x)u^{\star} + g(x)u^{\star} - G(x)d^{\star} + G(x)d^{\star} \Big] \\ &+ \gamma_{4}(-x^{T}Qx - u^{\star T}Ru^{\star} + \kappa^{2}d^{\star T}d^{\star}) \\ &= 2\gamma_{3}x^{T} \Big[ f(x) + g(x)\Big( -\frac{1}{2}R^{-1}\hat{g}^{T}(x)\nabla\theta_{c}^{T}(x)\hat{W}_{c} + \frac{1}{2}R^{-1}g^{T}(x)(\nabla\theta_{c}^{T}(x)W_{c} - (60)) \\ &+ \nabla\epsilon_{v}\Big) \Big) + g(x)u^{\star} + G(x)\Big( \frac{1}{2\kappa^{2}}\hat{G}^{T}(x)\nabla\theta_{c}^{T}(x)\hat{W}_{c} - \frac{1}{2\kappa^{2}}G^{T}(x)(\nabla\theta_{c}^{T}(x)W_{c} - \nabla\theta_{c}^{T}(x)) \\ &+ \nabla\epsilon_{v}\Big) \Big) + G(x)d^{\star} \Big] + \gamma_{4}(-x^{T}Qx - u^{\star T}Ru^{\star} + \kappa^{2}d^{\star T}d^{\star}). \end{aligned}$$

Since  $g^T \nabla \theta_c^T W_c - \hat{g}^T \nabla \theta_c^T \hat{W}_c = g^T \nabla \theta_c^T \tilde{W}_c + \tilde{g}^T \nabla \theta_c^T \hat{W}_c$ , and  $-G^T \nabla \theta_c^T W_c + \hat{G}^T \nabla \theta_c^T \hat{W}_c = -G^T \nabla \theta_c^T \tilde{W}_c - \tilde{G}^T \nabla \theta_c^T \hat{W}_c$ , (60) can be rewritten as

$$\begin{aligned} \dot{\mathcal{J}}_{3} + \dot{\mathcal{J}}_{4} =& 2\gamma_{3}x^{T} \Big[ f(x) + g(x) \Big( \frac{1}{2} R^{-1} g^{T}(x) \nabla \theta_{c}^{T}(x) \tilde{W}_{c} + \frac{1}{2} R^{-1} \tilde{g}^{T}(x) \nabla \theta_{c}^{T}(x) \hat{W}_{c} \\ &+ \frac{1}{2} R^{-1} g^{T}(x) \nabla \varepsilon_{v} \Big) + g(x) u^{\star} + G(x) \Big( - \frac{1}{2\kappa^{2}} G^{T}(x) \nabla \theta_{c}^{T}(x) \tilde{W}_{c} \\ &- \frac{1}{2\kappa^{2}} \tilde{G}^{T}(x) \nabla \theta_{c}^{T}(x) \hat{W}_{c} - \frac{1}{2\kappa^{2}} G^{T}(x) \nabla \varepsilon_{v} \Big) + G(x) d^{\star} \Big] \\ &+ \gamma_{4} (-x^{T} Qx - u^{\star T} Ru^{\star} + \kappa^{2} d^{\star T} d^{\star}) \\ \leq &- \Big[ \gamma_{4} \lambda_{m}(Q) - 2\gamma_{3} b_{f} - 4\eta \Big] \|x\|^{2} + \Big[ \frac{1}{2\eta} \gamma_{3}^{2} b_{g}^{2} b_{\omega}^{2} b_{\nabla \theta_{c}}^{2} \lambda_{M}^{2} (R^{-1}) \\ &+ \frac{1}{2\eta \kappa^{4}} \gamma_{3}^{2} b_{G}^{2} b_{\omega}^{2} b_{\nabla \theta_{c}}^{2} \Big] \|\tilde{W}_{I}\|^{2} + \Big[ \frac{1}{2\eta} \gamma_{3}^{2} b_{g}^{4} b_{\nabla \theta_{c}}^{2} \lambda_{M}^{2} (R^{-1}) + \frac{1}{2\eta \kappa^{4}} \gamma_{3}^{2} b_{G}^{4} b_{\nabla \theta_{c}}^{2} \Big] \|\tilde{W}_{c}\|^{2} \\ &- \Big[ \gamma_{4} \lambda_{m}(R) - 2\gamma_{3}^{2} b_{g}^{2} / \eta \Big] \|u^{\star}\|^{2} + \Big[ \frac{1}{2\eta} \gamma_{3}^{2} b_{g}^{4} \lambda_{M}^{2} (R^{-1}) + \frac{\gamma_{3}^{2}}{\kappa^{4}} b_{G}^{4} \Big] \|\nabla \varepsilon_{v}\|^{2} \\ &+ \Big[ \gamma_{4} \kappa^{2} + 2\gamma_{3}^{2} b_{G}^{2} / \eta \Big] \|d^{\star}\|^{2}, \end{aligned}$$

where  $b_{\omega} = \|\hat{W}_{c}\|$  is a bounded variable. Recall that  $\mathcal{V}_{I} = P_{I}^{*}v_{I}$  and  $\dot{v}_{I} = -l_{I}v_{I} + \theta_{If}\epsilon_{Tf}^{T}$ , thus

$$\begin{aligned} \dot{\mathcal{J}}_{5} \leq & 2\gamma_{5}b_{P_{I}^{*}}^{2} \left\| v_{I}^{T} \dot{v}_{I} \right\| = 2\gamma_{5}b_{P_{I}^{*}}^{2} \left\| v_{I}^{T} [-l_{I}v_{I} + \theta_{If}\epsilon_{Tf}^{T}] \right\| \\ \leq & - \left( 2\gamma_{5}b_{P_{I}^{*}}^{2}l_{I} - b_{P_{I}^{*}}^{2}\eta \right) \|v_{I}\|^{2} + \frac{1}{\eta}\gamma_{5}^{2}b_{P_{I}^{*}}^{2} \left\| \theta_{If}\epsilon_{Tf}^{T} \right\|^{2}. \end{aligned}$$

$$(62)$$

Since  $\dot{v}_c = -l_c v_c + \Theta \epsilon_{HJI}^T$ . Hence, the last term of (56) can be given as

$$\begin{aligned} \dot{\mathcal{J}}_{6} &\leq 2\gamma_{6}b_{P_{c}^{*}}^{2} \left\| v_{c}^{T}\dot{v}_{c} \right\| \\ &= 2\gamma_{6}b_{P_{c}^{*}}^{2} \left\| v_{c}^{T} \left[ -l_{c}v_{c} + \Theta\epsilon_{HJI}^{T} \right] \right\| \\ &\leq - \left( 2\gamma_{6}l_{c}b_{P_{c}^{*}}^{2} - 5b_{P_{c}^{*}}^{2}\eta \right) \|v_{c}\|^{2} + \frac{1}{\eta}\gamma_{6}^{2}b_{P_{c}^{*}}^{2}b_{\nabla\theta_{c}}^{2}\|\Theta\|^{2} \|\epsilon_{T}\|^{2} \\ &+ \frac{1}{\eta}\gamma_{6}^{2}b_{P_{c}^{*}}^{2}b_{\nabla\theta_{c}}^{2}b_{W_{c}}^{2} \|\theta_{I}\|^{2} \|\Theta\|^{2} \|\tilde{W}_{I}\|^{2} + \frac{1}{\eta}\gamma_{6}^{2}b_{P_{c}^{*}}^{2}b_{f}^{2}b_{\nabla\epsilon_{v}}^{2}\|\Theta\|^{2} \|x\|^{2} \\ &+ \frac{1}{4\eta}\gamma_{6}^{2}b_{P_{c}^{*}}^{2}b_{g}^{2}b_{\omega^{1}}^{2}b_{\nabla\theta_{c}}^{2}b_{\omega}^{2}\lambda_{M}^{2}(R^{-1}) \|\Theta\|^{2} \|\nabla\epsilon_{v}\|^{2} \\ &+ \frac{1}{4\eta\kappa^{4}}\gamma_{6}^{2}b_{P_{c}^{*}}^{2}b_{G}^{2}b_{\omega^{2}}^{2}b_{\nabla\theta_{c}}^{2}b_{\omega}^{2} \|\Theta\|^{2} \|\nabla\epsilon_{v}\|^{2}. \end{aligned}$$

where  $b_{\omega 1} = \|\hat{W}_g \theta_g\|$  and  $b_{\omega 2} = \|\hat{W}_G \theta_G\|$  are bounded variables. Consequently, we substitute (58), (59), and (61)–(63) into (56); thus, we have

$$\begin{split} \dot{\mathcal{J}}(t) &= \dot{\mathcal{J}}_{1} + \dot{\mathcal{J}}_{2} + \dot{\mathcal{J}}_{3} + \dot{\mathcal{J}}_{4} + \dot{\mathcal{J}}_{5} + \dot{\mathcal{J}}_{6} \\ &\leq -\left(\delta_{I} - \frac{1}{2\eta} - \frac{1}{2\eta}\gamma_{3}^{2}b_{g}^{2}b_{\omega}^{2}b_{\nabla\theta_{c}}^{2}\lambda_{M}^{2}(R^{-1}) - \frac{1}{2\eta\kappa^{4}}\gamma_{3}^{2}b_{G}^{2}b_{\omega}^{2}b_{\nabla\theta_{c}}^{2} \\ &- \frac{1}{\eta}\gamma_{6}^{2}b_{P_{c}}^{2}b_{\nabla\theta_{c}}^{2}b_{W_{c}}^{2}\|\theta_{I}\|^{2}\|\Theta\|^{2}\right)\|\tilde{W}_{I}\|^{2} - \left(\gamma_{4}\lambda_{m}(R) - \frac{2}{\eta}\gamma_{3}^{2}b_{g}^{2}\right)\|u^{\star}\|^{2} \\ &- \left(\delta_{c} - \frac{1}{2\eta} - \frac{1}{2\eta}\gamma_{3}^{2}b_{g}^{4}b_{\nabla\theta_{c}}^{2}\lambda_{M}^{2}(R^{-1}) - \frac{1}{2\eta\kappa^{4}}\gamma_{3}^{2}b_{G}^{4}b_{\nabla\theta_{c}}^{2}\right)\|\tilde{W}_{c}\|^{2} \\ &- \left(\gamma_{4}\lambda_{m}(Q) - 2\gamma_{3}b_{f} - 4\eta - \frac{1}{\eta}\gamma_{6}^{2}b_{P_{c}}^{2}b_{f}^{2}b_{\nabla\epsilon_{v}}^{2}\|\Theta\|^{2}\right)\|x\|^{2} + \frac{1}{\eta}\gamma_{5}^{2}b_{P_{I}}^{2}\left\|\phi_{If}\epsilon_{Tf}^{T}\right\|^{2} \\ &- \left(2\gamma_{5}b_{P_{I}}^{2}I_{I} - b_{P_{I}}^{2}\eta - \frac{\eta}{2}b_{P_{I}}^{2}|P_{I}|^{2}\right)\|v_{I}\|^{2} - \left(2\gamma_{6}I_{c}b_{P_{c}}^{2} - 5b_{P_{c}}^{2}\eta - \frac{\eta}{2}b_{P_{I}}^{2}|P_{I}|^{2}\right)\|v_{c}\|^{2} \\ &+ \left(\frac{1}{2\eta}\gamma_{3}^{2}b_{g}^{4}\lambda_{M}^{2}(R^{-1}) + \frac{\gamma_{3}^{2}}{\kappa^{4}}b_{G}^{4} + \frac{1}{4\eta}\gamma_{6}^{2}b_{P_{c}}^{2}b_{g}^{2}b_{\omega}^{2}b_{\omega}^{2}b_{\omega}^{2}b_{\omega}^{2}b_{\omega}^{2}h_{\omega}^{2}(R^{-1})\|\Theta\|^{2} \\ &+ \frac{1}{4\eta\kappa^{4}}\gamma_{6}^{2}b_{P_{c}}^{2}b_{G}^{2}b_{\omega}^{2}b_{\omega}^{2}b_{\omega}^{2}\|\Theta\|^{2}\right)\|\nabla\epsilon_{v}\|^{2} + \frac{1}{\eta}\gamma_{6}^{2}b_{P_{c}}^{2}b_{\omega}^{2}b_{\omega}^{2}b_{\omega}^{2}b_{\omega}^{2}h_{\omega}^{2}\|\Theta\|^{2}\|\varepsilon_{T}\|^{2}. \end{split}$$

We choose the parameters  $\gamma_3$ ,  $\gamma_4$ ,  $\gamma_5$ ,  $\gamma_6$  and  $\eta$ , fulfilling the following conditions

$$\begin{split} \eta &> \max\left\{ \left(\kappa^{4} + \kappa^{4} \gamma_{3}^{2} b_{g}^{2} b_{\nabla \theta_{c}}^{2} \lambda_{M}^{2} (R^{-1}) + \gamma_{3}^{2} b_{G}^{2} b_{\nabla \theta_{c}}^{2} b_{\nabla \theta_{c}}^{2} b_{P_{c}^{*}}^{2} b_{\nabla \theta_{c}}^{2} b_{W_{c}}^{2} \\ & \|\theta_{I}\|^{2} \|\Theta\|^{2} \right) / 2\kappa^{4} \delta_{I}, \left(\kappa^{4} + \kappa^{4} \gamma_{3}^{2} b_{g}^{4} b_{\nabla \theta_{c}}^{2} \lambda_{M}^{2} (R^{-1}) + \kappa^{4} \gamma_{3}^{2} b_{G}^{4} b_{\nabla \theta_{c}}^{2} \right) / 2\kappa^{4} \delta_{c} \right\}, \\ \gamma_{3} &< \sqrt{\gamma_{4} \eta \lambda_{m}(R)} / 2b_{g}^{2}, \\ \gamma_{4} &> \left\{ 2\gamma_{3} b_{f} - 4\eta - \frac{1}{\eta} \gamma_{6}^{2} b_{P_{c}^{*}}^{2} b_{f}^{2} b_{\nabla e_{v}}^{2} \|\Theta\|^{2} \right\} / \lambda_{m}(Q), \\ \gamma_{5} &> (\eta + \frac{\eta}{2} |P_{I}|^{2}) / 2l_{I}, \\ \gamma_{6} &> (5\eta - \frac{\eta}{2} |P_{c}|^{2}) / 2l_{c}. \end{split}$$

Then, (64) can be further presented as

$$\dot{\mathcal{J}}(t) \le -k_1 \|\tilde{W}_I\|^2 - k_2 \|\tilde{W}_c\|^2 - k_3 \|x\|^2 - k_4 \|v_I\|^2 - k_5 \|v_c\|^2 + b_{\gamma},$$
(65)

where  $k_1$ ,  $k_2$ ,  $k_3$ ,  $k_4$ ,  $k_5$  and  $b_\gamma$  are positive constants

$$\begin{split} k_{1} &= \delta_{I} - \frac{1}{2\eta} - \frac{1}{2\eta} \gamma_{3}^{2} b_{g}^{2} b_{\omega}^{2} b_{\nabla \theta_{c}}^{2} \lambda_{M}^{2} (R^{-1}) - \frac{1}{2\eta \kappa^{4}} \gamma_{3}^{2} b_{G}^{2} b_{\omega}^{2} b_{\nabla \theta_{c}}^{2} - \frac{1}{\eta} \gamma_{6}^{2} b_{P_{c}}^{2} b_{\nabla \theta_{c}}^{2} b_{W_{c}}^{2} \|\theta_{I}\|^{2} \|\Theta\|^{2}, \\ k_{2} &= \delta_{c} - \frac{1}{2\eta} - \frac{1}{2\eta} \gamma_{3}^{2} b_{g}^{4} b_{\nabla \theta_{c}}^{2} \lambda_{M}^{2} (R^{-1}) - \frac{1}{2\eta \kappa^{4}} \gamma_{3}^{2} b_{G}^{4} b_{\nabla \theta_{c}}^{2}, \\ k_{3} &= \gamma_{4} \lambda_{m} (Q) - 2\gamma_{3} b_{f} - 4\eta - \frac{1}{\eta} \gamma_{6}^{2} b_{P_{c}}^{2} b_{f}^{2} b_{\nabla e_{c}}^{2} \|\Theta\|^{2}, \\ k_{4} &= 2\gamma_{5} b_{P_{I}}^{2} l_{I} - b_{P_{I}}^{2} \eta - \frac{\eta}{2} b_{P_{I}}^{2} |P_{I}|^{2}, \quad k_{5} &= 2\gamma_{6} l_{c} b_{P_{c}}^{2} - 5 b_{P_{c}}^{2} \eta - \frac{\eta}{2} b_{P_{c}}^{2} |P_{c}|^{2}, \\ b_{\gamma} &= \left(\frac{1}{2\eta} \gamma_{3}^{2} b_{g}^{4} \lambda_{M}^{2} (R^{-1}) + \frac{\gamma_{3}^{2}}{\kappa^{4}} b_{G}^{4} + \frac{1}{4\eta} \gamma_{6}^{2} b_{P_{c}}^{2} b_{g}^{2} b_{\omega}^{2} b_{\omega}^{2} \lambda_{M}^{2} (R^{-1}) \|\Theta\|^{2} + \frac{1}{\eta} \gamma_{5}^{2} b_{P_{I}}^{2} \|\phi_{If} \epsilon_{Tf}^{T}\|^{2} \\ &+ \frac{1}{4\eta \kappa^{4}} \gamma_{6}^{2} b_{P_{c}}^{2} b_{G}^{2} b_{\omega}^{2} b_{\omega}^{2} b_{\omega}^{2} \|\Theta\|^{2} \right) \|\nabla \epsilon_{v}\|^{2} + \frac{1}{\eta} \gamma_{6}^{2} b_{P_{c}}^{2} b_{\Theta c}^{2} \|\Theta\|^{2} \|\Theta\|^{2} \|\epsilon_{T}\|^{2}. \end{split}$$

Thus,  $\dot{\mathcal{J}}(t)$  is negative if

$$\|\tilde{W}_{I}\| > \sqrt{b_{\gamma}/k_{1}}, \|\tilde{W}_{c}\| > \sqrt{b_{\gamma}/k_{2}}, \|x\| > \sqrt{b_{\gamma}/k_{3}}, \|\Xi_{I}\| > \sqrt{b_{\gamma}/k_{4}}, \|\Xi_{c}\| > \sqrt{b_{\gamma}/k_{5}},$$

which implies that the NN weight estimation errors  $\tilde{W}_I$ ,  $\tilde{W}_c$  and the system state *x* are all UUB.

Lastly, the error between the proposed  $H_{\infty}$  control pair and the ideal one are written as

$$\begin{split} \hat{u} - u^{\star} &= -\frac{1}{2} R^{-1} \left[ \hat{W}_{g} \theta_{g}(x) \right]^{T} \nabla \theta_{c}^{T}(x) \hat{W}_{c} + \frac{1}{2} R^{-1} g^{T} \left( \nabla \theta_{c}^{T}(x) W_{c} + \nabla \epsilon_{v} \right) \\ &= \frac{1}{2} R^{-1} g^{T} \nabla \theta_{c}^{T}(x) \tilde{W}_{c} + \frac{1}{2} R^{-1} \left[ g - \hat{W}_{g} \theta_{g}(x) \right]^{T} \nabla \theta_{c}^{T}(x) W_{c} \\ &- \frac{1}{2} R^{-1} \left[ g - \hat{W}_{g} \theta_{g}(x) \right]^{T} \nabla \theta_{c}^{T}(x) \tilde{W}_{c} + \frac{1}{2} R^{-1} g^{T} \nabla \epsilon_{v}, \\ \hat{d} - d^{\star} &= \frac{1}{2 \kappa^{2}} \left[ \hat{W}_{G} \theta_{G}(x) \right]^{T} \nabla \theta_{c}^{T}(x) \hat{W}_{c} - \frac{1}{2 \kappa^{2}} G^{T} \left( \nabla \theta_{c}^{T}(x) W_{c} + \nabla \epsilon_{v} \right) \\ &= - \frac{1}{2 \kappa^{2}} G^{T} \nabla \theta_{c}^{T}(x) \tilde{W}_{c} + \frac{1}{2 \kappa^{2}} \left[ G - \hat{W}_{G} \theta_{G}(x) \right]^{T} \nabla \theta_{c}^{T}(x) W_{c} \\ &+ \frac{1}{2 \kappa^{2}} \left[ G - \hat{W}_{G} \theta_{G}(x) \right]^{T} \nabla \theta_{c}^{T}(x) \tilde{W}_{c} - \frac{1}{2 \kappa^{2}} \left[ \hat{W}_{G} \theta_{G}(x) \right]^{T} \nabla \epsilon_{v}, \end{split}$$

which further implies the following fact

$$\begin{split} \lim_{t \to +\infty} \|\hat{u} - u^{\star}\| &\leq \frac{1}{2} \lambda_M(R^{-1}) \left\{ b_g \left( b_{\nabla \theta_c} \| \tilde{W}_c \| + b_{\nabla \varepsilon_v} \right) + b_{\nabla \theta_c} b_{W_c} \left( \| \tilde{W}_I \| + \| b_g \| \right) \right. \\ &+ b_{\nabla \theta_c} \| \tilde{W}_c \| \left( \| \tilde{W}_I \| + \| b_g \| \right) \right\} \leq b_u, \\ \lim_{t \to +\infty} \left\| \hat{d} - d^{\star} \right\| &\leq \frac{1}{2\kappa^2} \left\{ b_G \left( b_{\nabla \theta_c} \| \tilde{W}_c \| + b_{\nabla \varepsilon_v} \right) + b_{\nabla \theta_c} b_{W_c} \left( \| \tilde{W}_I \| + \| b_G \| \right) \right. \\ &+ b_{\nabla \theta_c} \| \tilde{W}_c \| \left( \| \tilde{W}_I \| + \| b_G \| \right) \right\} \leq b_d, \end{split}$$

where  $b_u > 0$  and  $b_d > 0$  are constants determined by the identifier NN estimation error  $\tilde{W}_I$  and the critic NN estimation error  $\tilde{W}_c$ . It proves that the approximate  $H_{\infty}$  control pair can converge to a set around the optimal solution.

This completes the proof.  $\Box$ 

# 6. Numerical Simulation

This section aims to verify the effectiveness of the proposed KREM-based IC learning approach for optimal robust control. We consider the following NCT system [12]

$$\dot{x} = f(x) + g(x)u + G(x)d,$$
 (66)

where 
$$f(x) = \begin{bmatrix} -x_1 + x_2 \\ -0.5x_1 - 0.5x_2(1 - (\cos(2x_1) + 2)^2) \end{bmatrix}$$
,  $g(x) = \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix}$ ,  $G(x) = \begin{bmatrix} 0 \\ \sin(4x_1) + 2 \end{bmatrix}$ .

We choose the regressor of identifier NN as

$$\theta_I(x, u) = [x_1, x_2, x_2(1 - (\cos(2x_1) + 2)^2), u\cos(2x_1), u, d\sin(4x_1), d]^T,$$

with the unknown identifier weight matrix given by

$$W_I = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ -0.5 & 0 & -0.5 & 1 & 2 & 1 & 2 \end{bmatrix}.$$

The activation function in (33) for the critic NN is selected as

$$\theta_c(x, u) = [x_1^2, x_1 x_2, x_2^2]^T.$$

The ideal critic NN weights were  $W_c = [0.5, 0, 1]^T$ .

In this numerical example, several other parameters are set as follows: the initial values of the system states are  $x_1(0) = 3$  and  $x_2(0) = -1$ .  $Q = I_2$  and R = 1. The filter coefficients are  $\rho = 0.001$ ,  $l_I = 0.1$ ,  $l_c = 20$ ,  $\gamma_1 = 800$ ,  $\gamma_2 = 200 diag\{0.3, 1, 1\}$ . It is important to note that in this simulation, there is no need to add noise to the control input u(t) to ensure the PE condition. This condition is often necessary for many existing ADP-based control methods to ensure that  $\theta_I(t) \in PE$  and  $\Theta(t) \in PE$ .

For comparison, we consider the Kreisselmeier's Regressor Extension (KRE) based identifier-critic network framework [12] for the system (66). Figures 2 and 3 display the convergence of the identifier NN weights and the critic NN weights, respectively, under our KREM-based optimal robust control method and the KRE-based control method [12]. As illustrated in Figure 2, the KREM-based ADP method proposed in this paper exhibits faster convergence compared to the KRE-based ADP method. Furthermore, it demonstrates element-wise monotonicity, thus preventing oscillations and peaking in the learning curve. The trajectories of the approximate control input  $\hat{u}$  and the estimated disturbance  $\hat{d}$  are presented in Figures 4 and 5, respectively. By applying the optimal  $H_{\infty}$  control pair, the system states are stabilized, as depicted in Figure 6.



**Figure 2.** Comparison of the convergence of identifier NN's weights  $\hat{W}_{l}$ : (**a**) KREM-based method; (**b**) KRE-based method in [12].



**Figure 3.** Comparison of the convergence of critic NN's weights  $\hat{W}_c$ : (a) KREM-based method; (b) KRE-based method in [12].



**Figure 4.** Evolution of the approximate control input  $\hat{u}$ .



**Figure 5.** Disturbance action *d*.



**Figure 6.** Trajectories of the system states  $x = [x_1, x_2]^T$ .

# 7. Conclusions

This paper presents a novel adaptive learning approach using neural networks (NNs) to address the problem of optimal robust control for nonlinear continuous-time systems with unknown dynamics. The approach involves employing a system identifier that utilizes NNs and parameter estimation techniques to approximate the unknown system matrices and disturbances. Additionally, a critic NNs learning structure is introduced to obtain an approximate controller that corresponds to the optimal control problem. Unlike existing identifier-critic NNs learning control methods, this approach incorporates adaptive tuning laws based on a regressor extension and mixing technique. These laws facilitate the learning of unknown parameters in the two NNs under relaxed persistence of excitation conditions. The convergence conditions of the proposed approach have been theoretically demonstrated. Finally, the effectiveness of the proposed learning control approach is validated via a simulation study.

**Author Contributions:** Methodology, R.L.; Validation, R.L.; Formal analysis, R.L.; Investigation, R.L. and Z.P.; Writing—original draft, R.L.; Writing—review & editing, Z.P. and J.H.; Supervision, Z.P. and J.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 62203089 and Grant 62103084, in part by the Project funded by China Postdoctoral Science Foundation under Grant 2021M700695, in part by the Sichuan Science and Technology Program, China under Grant 2022NSFSC0890, Grant 2022NSFSC0865, and Grant 2021YFS0016, and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515110135.

Institutional Review Board Statement: Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**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

- Luo, R.; Peng, Z.; Hu, J. On model identification based optimal control and it's applications to multi-agent learning and control. *Mathematics* 2023, 11, 906. [CrossRef]
- Luo, B.; Wu, H.N.; Huang, T. Off-policy reinforcement learning for H<sub>∞</sub> control design. *IEEE Trans. Cybern.* 2014, 45, 65–76. [CrossRef] [PubMed]
- 3. Werbos, P. Approximate Dynamic Programming for Realtime Control and Neural Modelling; White, D.A., Sofge, D.A., Eds.; Van Nostrand: New York, NY, USA, 1992.
- Lewis, F.L.; Vrabie, D. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits Syst. Mag.* 2009, 9, 32–50. [CrossRef]
- 5. Vamvoudakis, K.G.; Lewis, F.L. Online actor–critic algorithm to solve the continuoustime infinite horizon optimal control problem. *Automatica* **2010**, *46*, 878–888. [CrossRef]
- 6. Wei, Q.; Liu, D.; Lin, H. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems. *IEEE Trans. Cybern.* **2015**, *46*, 840–853. [CrossRef]
- Peng, Z.; Zhao, Y.; Hu, J.; Luo, R.; Ghosh, B.K.; Nguang, S.K. Input-output data-based output antisynchronization control of multi-agent systems using reinforcement learning approach. *IEEE Trans. Ind. Inform.* 2021, 17, 7359–7367. [CrossRef]
- Peng, Z.; Zhao, Y.; Hu, J.; Ghosh, B.K. Data-driven optimal tracking control of discrete-time multi-agent systems with two-stage policy iteration algorithm. *Inf. Sci.* 2019, 481, 189–202. [CrossRef]
- 9. Zhang, H.; Jiang, H.; Luo, C.; Xiao, G. Discrete-time nonzero-sum games for multiplayer using policy-iteration-based adaptive dynamic programming algorithms. IEEE transactions on cybernetics. *IEEE Trans. Cybern.* **2016**, *47*, 3331–3340. [CrossRef]
- 10. Modares, H.; Lewis, F.L. Optimal tracking control of nonlinear partiallyunknown constrained input systems using integral reinforcement learning. *Automatica* 2014, *50*, 1780–1792. [CrossRef]
- 11. 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]
- Lv, Y.; Na, J.; Ren, X. Online H<sub>∞</sub> control for completely unknown nonlinear systems via an identifier-critic-based ADP structure. *Int. J. Control Autom.* 2019, 92, 100–111. [CrossRef]
- 13. Luo, R.; Peng, Z.; Hu, J.; Ghosh, B.K. Adaptive optimal control of affine nonlinear systems via identifier-critic neural network approximation with relaxed PE conditions. *Neural Netw.* **2023**, *167*, 588–600. [CrossRef] [PubMed]

- 14. Luo, R.; Tan, W.; Peng, Z.; Zhang, J.; Hu, J.; Ghosh, B.K. Optimal consensus control for multi-agent systems with unknown dynamics and states of leader: A distributed KREM learning method. *IEEE Trans. Circuits Syst. II Express Briefs* 2023. [CrossRef]
- 15. 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.* **2015**, *27*, 444–458. [CrossRef] [PubMed]
- Wang, D.; Zhou, Z.; Liu, A.; Qiao, J. Event-triggered robust adaptive critic control for nonlinear disturbed systems. *Nonlinear Dyn.* 2023, 111, 19963–19977. [CrossRef]
- 17. Zhao, J.; Na, J.; Gao, G. Adaptive dynamic programming based robust control of nonlinear systems with unmatched uncertainties. *Neurocomputing* **2020**, *395*, 56–65. [CrossRef]
- 18. Vamvoudakis, K.G.; Lewis, F.L. Online solution of nonlinear two-player zero-sum games using synchronous policy iteration. *Int. J. Robust Nonlinear Control* **2012**, 22, 1460–1483. [CrossRef]
- 19. Peng, Z.; Ji, H.; Zou, C.; Kuang, Y.; Cheng, H.; Shi, K.; Ghosh, B.K. Optimal H∞ tracking control of nonlinear systems with zero-equilibrium-free via novel adaptive critic designs. *Neural Netw.* **2023**, *164*, 105–114. [CrossRef]
- 20. Lin, F.; Brandt, R.D. An optimal control approach to robust control of robot manipulators. *IEEE Trans. Robot. Automat.* **1998**, *14*, 69–77.
- 21. Yang, X.; He, H.; Zhong, X. Adaptive dynamic programming for robust regulation and its application to power systems. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5722–5732. [CrossRef]
- 22. Yang, X.; He, H. Adaptive critic designs for event-triggered robust control of nonlinear systems with unknown dynamics. *IEEE Trans. Cybern.* **2018**, *49*, 2255–2267. [CrossRef] [PubMed]
- 23. Xue, S.; Luo, B.; Liu, D.; Gao, Y. Event-triggered ADP for tracking control of partially unknown constrained uncertain systems. *IEEE Trans. Cybern.* **2021**, *52*, 9001–9012. [CrossRef] [PubMed]
- 24. Lewis, F.W.; Jagannathan, S.; Yesildirak, A. Neural Network Control of Robot Manipulators and Non-Linear Systems; Taylor & Francis: London, UK, 1999.
- 25. Boyd, S.; Sastry, S. Adaptive Control: Stability, Convergence and Robustness; Prentice-Hall: Englewood Cliffs, NJ, USA, 1989.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.