Taylor & Francis
Taylor & Francis Group

Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Online adaptive optimal tracking control for model-free nonlinear systems via a dynamic neural network

Yuming Yin, Zhiiun Fu & Yan Lu

To cite this article: Yuming Yin, Zhiiun Fu & Yan Lu (2023) Online adaptive optimal tracking control for model-free nonlinear systems via a dynamic neural network, Automatika, 64:3, 431-440, DOI: 10.1080/00051144.2023.2170058

To link to this article: https://doi.org/10.1080/00051144.2023.2170058

Published online: 13 Feb 2023.

AUTOMATIKA
2023, VOL. 64, NO. 3, 431–440
https://doi.org/10.1080/00051144.2023.2170058

REGULAR PAPER

OPEN ACCESS Check for updates

Online adaptive optimal tracking control for model-free nonlinear systems via a dynamic neural network

Yuming Yin[a], Zhiiun Fu[b] and Yan Lu[b]

[a]College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, People's Republic of China; [b]College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, People's Republic of China

ABSTRACT
This paper presents an online adaptive approximate solution for the optimal tracking control problem of model-free nonlinear systems. Firstly, a dynamic neural network identifier with properly designed weights updating laws is developed to identify the unknown dynamics. Then an adaptive optimal tracking control policy consisting of two terms is proposed, i.e. a steady-state control term is established to ensure the desired tracking performance at the steady state, and an optimal control term is proposed to ensure the optimal tracking error dynamics optimally. The composite Lyapunov method is used to analyse the stability of the closed-loop system. Two simulation examples are presented to demonstrate the effectiveness of the proposed method.

ARTICLE HISTORY
Received 14 June 2021
Accepted 13 January 2023

KEYWORDS
Dynamic neural network; nonlinear systems; nonlinear identifier; adaptive control; optimal control

1. Introduction

The basic idea of the classical adaptive control is to update the model parameter and control law directly or indirectly, such that the control error can be minimized. However, it is generally not optimal. On the other side, the main drawback of the classical optimal control approach lies in that the system dynamics must be precisely known for solving the Hamilton-Jacobi-Bellman (HJB) equation in an off-line manner [1]. Hence, by merging the knowledge from adaptive control and optimal control, the adaptive optimal control approach has been developed during the past decade and a survey of this research can be found in [2–4].

To develop an online adaptive optimal control, Werbos [5] introduced the general actor-critic (AC) framework for adaptive optimal control. The critic neural network (NN) approximates the evaluation function, mapping states to an estimated measure of the value function, whereas the NN approximates an optimal control law and generates the actions or control signals. Since then, various modifications to adaptive optimal control algorithms have been proposed as model-based methods (heuristic dynamic programming – HDP [6] and dual heuristic programming-DHP [7]) and model-free methods (action-dependent heuristic dynamic programming – ADHDP [8] and Q learning [9]). However, most of the previous works on adaptive optimal control have focused on discrete-time systems. The extensions of these adaptive optimal control research to continuous-time systems pose challenges in proving stability, convergence and ensuring the online updating law with model free [10].

Discretizinging the continuous time system is generally not accurate, especially for the high-dimensional systems that prohibit the learning process. Hence, the online policy iteration-based algorithms are proposed to solve the linear [11] and nonlinear [12] continuous-time infinite horizon optimal control problems, which involve synchronous adaptive of both actor and critic NN. Furthermore, ref. [10] extended the idea in refs. [11,12] by designing a novel AC-identifier architecture to approximate the HJB equation without the knowledge of system drift dynamics, but the knowledge of the input dynamics is required. The recent research in [13] cancels this requirement by using the experience iteration technique. Based on ref. [10], a simply identifier-critic structure-based optimal control method is proposed in [14,15], where just a critic NN is used to approximate the solution of the HJB equation and to calculate the optimal control action. In [16], an optimal control method for nonzero-sum differential games of continuous-time nonlinear systems is designed directly from the critic NN instead of the action-critic dual network, which greatly simplifies the algorithm architecture.

Most of the existing adaptive optimal research studies mainly focus on dealing with regulation problems rather than trajectory tracking problems. The combined consideration of two aspects can ensure not only the realization of trajectory tracking and stabilization but also satisfying the prescribed performance index (such as minimization of the trajectory error, fuel consumption, etc.). In [17] a new data-based iterative optimal learning control scheme is

CONTACT Zhiiun Fu fuzhijun2008@163.com College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, People's Republic of China

developed to solve a coal gasification optimal tracking control problem in the discrete-time domain. For continuous-time systems, linear quadratic tracking control of partially-unknown systems using reinforcement learning is present in [18] and a nonlinear approximately optimal trajectory tracking method with exact model information is developed in [19]. To relax the requirement of an explicit model, a steady-state control conjunction with an optimal control for nonlinear continuous-time systems is developed in [20], which stabilizes the error dynamics in an optimal way.

Most of the above-mentioned adaptive optimal control method is based on the affine nonlinear system, to the best of our knowledge, only [21] addressed the adaptive optimal control of unknown non-affine nonlinear systems in the discrete-time domain and [22] introduces an adaptive recursive control for the model-based non-affine nonlinear continuous system. The optimal control of an unknown non-affine nonlinear continuous-time system is still a challenging task, which is the motivation of this paper.

The main contributions of this paper are listed as follows.

(1) The optimal tracking control of unknown non-affine nonlinear systems based on the critic identifier architecture is first proposed in this paper. Model-free property is achieved by a neuro identifier in conjunction with the novel updating laws for both the weights and the linear part matrix which is usually assumed to be a known Hurwitz matrix for the conventional black-box nonlinear system identification.
(2) Adaptive optimal tracking control policy consisting of two terms is proposed, i.e. a steady-state control term is established to ensure the desired tracking performance at the steady state, and an optimal control term is proposed to ensure the optimal tracking error dynamics. Online solution of the optimal control term is obtained directly by a single critic NN to approximate the optimal cost function of the HJB equation instead of the conventional action-critic dual network, which greatly reduces complexity and saves calculation time. A novel learning law driven by filtered parameter error is proposed for critic NN. The stability of the entire closed-loop system is proved by the properly designed composite Lyapunov method.

The main organization of the paper is as follows. The problem formulation is given in Section 2. The DNN identifier is designed in Section 3. Then, the optimal control strategy, based on the critic-identifier architecture, is present in Section 4. Two simulation examples are presented to verify the proposed scheme in Section 5 and the conclusion is drawn in Section 6.

2. Problem formulation

Consider the following non-affine nonlinear continuous-time systems

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

where $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T \in R^n$ is the state vector, $u(t) = (u_1(t), u_2(t), \ldots, u_m(t))^T \in R^m$ is the control input vector and $f(\cdot)$ is an unknown continuous nonlinear smooth function for $x(t)$ and $u(t)$.

The objective of the optimal tracking control problem is to design an optimal controller (1) to ensure that the state vector $x(t)$ tracks the specified trajectory $x_r(t)$ and minimize the infinite horizon performance cost function as follows:

$$V(e(t)) = \int_t^\infty r(e(\tau), u_e(e(\tau))d\tau \tag{2}$$

where the tracking error is defined as $e(t) = x(t) - x_r(t)$, the utility function with symmetric positive definite matrices Q and R is defined as $r(e(t), u(t)) = e^T(t)Qe(t) + u^T(t)Ru(t)$.

From the basic optimal control theory, we define the Hamiltonian of (1) as

$$H(e, u_e, V) = V_e^T[f(x(t), u(t))] + e^T Qe + u^T(t)Ru(t) \tag{3}$$

where $V_e \triangleq \frac{\partial V}{\partial x}$ denotes the partial derivative of the cost function $V(e(t))$ with respect to $e(t)$.

The optimal cost function $V^*(e(t))$ is given as

$$V^*(e(t)) = \min \min_{u \in \psi(\Omega)} \int_t^\infty r(e(\tau), u_e(e(\tau))d\tau \tag{4}$$

and it satisfies the HJB equation

$$H(e, u^*, V^*) = V_e^{*T}[f(x(t), u^*(t))] + e^T(t)Qe(t) + u^{*T}(t)Ru^*(t) \tag{5}$$

where the control u is defined to be admissible for (2) on a compact set $\Omega \in R^n$, denoted by $u \in \psi(\Omega)$.

Theoretically, the optimal control for nonlinear system (1) can be obtained from Equations (4) and (5). However, optimal control cannot be obtained in practical systems due to two reasons: 1). The optimal cost function $V^*(e(t))$ should be obtained by solving the HJB equation (5). However, it is usually difficult to solve the high-order nonlinear partial differential equation (PDE) for general nonlinear systems via analytical methods. Moreover, the unknown nonlinear dynamic $f(\cdot)$ makes the solution unavailable for HJB Equation (2). The idea of optimal control $u^*(t)$ cannot be derived by solving $\frac{\partial H(e,u^*,V^*)}{\partial u^*} = 0$ due to the unavailability of $V^*(e(t))$.

In this paper, we develop a critic-identifier to solve the optimal control of an unknown non-affine nonlinear continuous-time system, all the learning processes can be updated online.

3. Adaptive model-free identifier

We employ the following dynamic neural network (DNN) model to approximate the nonlinear dynamic system (1)

$$\dot{\hat{x}}(t) = A\hat{x}(t) + W_1\sigma(V_1[\hat{x}(t)]) + W_2\phi(V_2[\hat{x}(t)])u(t) \tag{6}$$

where $\hat{x}(t) \in R^n$ is the state of the DNN, $W_1 \in R^{n\times m}$, $W_2 \in R^{m\times n}$ are the weights in the output layers, $W_1 \in R^{n\times m}$, $W_2 \in R^{m\times n}$ are the weights in the hidden layer, $A \in R^{n\times n}$ is the matrix for the linear part of NNs, $u(t) = (u_1(t), u_2(t), \ldots, u_k(t), 0, \ldots, 0)^T \in R^m$ is the control input, the active function $\sigma(\cdot)$ (as well as $\phi(\cdot)$) is the sigmoidal vector function which is defined as $\sigma(\cdot) = a/(1 + e^{-bx}) - c$, where a, b and c are constants.

Remark 3.1: If we define $W = [W_1, W_2]$, $\Xi = \{\sigma(V_1[\hat{x}(t)]), \varphi(V_2[\hat{x}(t)])u(t)\}$, then (6) can be written as $\dot{\hat{x}}(t) = A\hat{x}(t) + W\Xi$. It has been proved in [23] that DNN with the form $\dot{\hat{x}}(t) = A\hat{x}(t) + W\Xi$ can approximate the nonlinear system (1) to any degree of accuracy if the hidden layer V is large enough. Here, to simplify the analysis process, we consider the simplest structure (i.e. $m = n, V = I, \phi(\cdot) = I$).

Then the nonlinear system (1) can be modelled by the DNN as follows:

$$\dot{x}(t) = A^*x(t) + W_1{}^*\sigma(x(t)) + W_2{}^*u(t) + \xi_1 \tag{7}$$

where A^*, $W_1{}^*$, $W_2{}^*$ are the nominal unknown matrices and $W_1{}^*$, $W_2{}^*$ are bounded as $W_1{}^*\Lambda_1{}^{-1}W_1{}^{*T} \le \overline{W_1}$, $W_2{}^*\Lambda_2{}^{-1}W_2{}^{*T} \le \overline{W_2}$ ($\Lambda_1{}^{-1}$, $\Lambda_2{}^{-1}$ are any positive definite symmetric matrices), and ξ_1 is regarded as the modelling error or disturbance and is assumed to be bounded.

Assumption 3.1: The identification error is defined by $\Delta x = x(t) - \hat{x}(t)$. The difference in the activation function $\tilde{\sigma} = \sigma(x(t)) - \sigma(\hat{x}(t))$ satisfies the generalized Lipshitz condition $\tilde{\sigma}^T\Lambda\tilde{\sigma} < [\Delta x]^T D[\Delta x] = \Delta x^T D\Delta x$, and $D = D^T > 0$ is the known normalizing matrices.

Then from (6) and (7), we can obtain the error dynamic equation

$$\Delta\dot{x} = A^*\Delta x + \tilde{A}\hat{x}(t) + W_1{}^*\tilde{\sigma} + \tilde{W}_1\sigma\hat{x}(t) + \tilde{W}_2u(t) + \xi_1 \tag{8}$$

where $\tilde{A} = A^* - A$ $\tilde{W}_1 = W_1{}^* - W_1$, $\tilde{W}_2 = W_2{}^* - W_2$,

Lemma 3.1 ([24]): *$A \in \Re^{n\times n}$ is a Hurwitz matrix, $R, Q \in \Re^{n\times n}$ $R = R^T > 0, Q = Q^T > 0$ if $(A, R^{1/2})$ is controllable, $(A, Q^{1/2})$ is observable and $A^TR^{-1}A - Q \ge \frac{1}{4}(A^TR^{-1} - R^{-1}A)R(A^TR^{-1} - R^{-1}A)$ is satisfied, the algebraic Riccati equation $A^TX + XA + XRX + Q = 0$ has a unique positive definite solution $X = X^T > 0$.*

Theorem 3.1: *Consider the identification scheme* (6) *for* (1)*, the following updating law*

$$\dot{A} = -k_1\Delta x\hat{x}^T, \dot{W}_1 = -k_2\Delta x\sigma_1{}^T(\hat{x}), \quad \dot{W}_2 = -k_3\Delta xu^T \tag{9}$$

where k_1, k_2 and k_3 are positive constants, can guarantee the following stability properties:

(1) *For a precise identifier case i.e. $\xi_1 = 0$*

$$\hat{W}_{1,2}, \hat{A} \in L_\infty, \Delta x \in L_2 \cap L_\infty, \lim_{t\to\infty}\Delta x = 0.$$

(2) *For bounded modelling error and disturbances i.e. $\xi_1 \le \overline{\xi_1}$*

$$\Delta x, \hat{W}_{1,2}, \hat{A} \in L_\infty.$$

Proof: Consider the Lyapunov function candidate

$$L_I = \Delta x^TP\Delta x + \frac{1}{k_2}tr\{\tilde{W}_1^TP_x\tilde{W}_1\} + \frac{1}{k_3}tr\{\tilde{W}_2^TP_x\tilde{W}_2\} + \frac{1}{k_1}tr\{\tilde{A}^TP_x\tilde{A}\} \tag{10}$$

■

Hence, differentiating (11) and using (8) yield

$$\begin{aligned}\dot{L}_I = {}& \Delta x^T(A^{*T}P + PA^*)\Delta x + 2\Delta x^TP\tilde{A}\hat{x} \\ & + 2\Delta x^TP\tilde{W}_1\sigma(\hat{x}) + 2\Delta x^TP\tilde{W}_2u \\ & + 2\Delta x^TPW_1^*\tilde{\sigma} + 2\Delta x^TP\xi_1 + \frac{2}{k_1}tr\{\dot{\tilde{A}}^TP\tilde{A}\} \\ & + \frac{2}{k_2}tr\{\dot{\tilde{W}}_1^TP\tilde{W}_1\} + \frac{2}{k_3}tr\{\dot{\tilde{W}}_2^TP\tilde{W}_2\}\end{aligned} \tag{11}$$

By using the updating laws (9) and taking the facts $\dot{\tilde{A}} = -\dot{A}$, $\dot{\tilde{W}}_{1,2} = -\dot{W}_{1,2}$, into consideration, then (11) becomes

$$\begin{aligned}\dot{L}_I = {}& \Delta x^T(A^{*T}P + PA^*)\Delta x \\ & + 2\Delta x^TPW_1^*\tilde{\sigma} + 2\Delta x^TP\xi_1\end{aligned} \tag{12}$$

Using the following matrix inequality

$$X^TY + (X^TY)^T \le X^T\Lambda^{-1}X + Y^T\Lambda Y \tag{13}$$

where $X, Y \in R^{j\times k}$ are any matrices and $\Lambda \in R^{j\times k}$ is any positive definite matrix. From Assumption 3.1, one obtains

$$2\Delta x^TPW_1^*\tilde{\sigma} \le \Delta x^TPW_1^*\Lambda^{-1}W_1^*P\Delta x + \tilde{\sigma}^T\Lambda\tilde{\sigma}$$

$$\leq \Delta x^T P \bar{W}_1 P \Delta x + \Delta x^T D \Delta x$$
$$2\Delta x^T P \xi_1 \leq \Delta x^T P \Lambda_\xi^{-1} P \Delta x + \xi_1^T \Lambda_\xi^{-1} \xi_1 \quad (14)$$

Then substituting (14) into (12) obtains

$$\dot{L}_I \leq \Delta x^T (A^{*T} P + P A^* + P \bar{W}_1 P + D + Q_0) \Delta x - \Delta x^T Q_0 \Delta x + \Delta x^T P \Lambda_\xi P \Delta x + \Delta \xi_1^T \Lambda_\xi^{-1} \Delta \xi_1 \quad (15)$$

By defining $R = \overline{W}_1, Q = D + Q_o$, then if we can select proper Q_o so that Q satisfies the conditions in Lemma 3.1, there exists matrix P satisfying the equation $A^{*T}P + PA^* + PRP + Q = 0$.

Hence (15) becomes

$$\dot{L}_I \leq -\Delta x^T Q_0 \Delta x + \Delta x^T P \Lambda_\xi^{-1} P \Delta x + \Delta \xi_1^T \Lambda_\xi^{-1} \Delta \xi_1 \quad (16)$$

Case 1: For precise identifier case i.e. $\xi_1 = 0$, (16) becomes

$$\dot{L}_I \leq -\Delta x^T Q_0 \Delta x \leq -\lambda_{\min}(Q_x) \|\Delta x\|_{Q_x}^2 \leq 0 \quad (17)$$

From (17) we get $\Delta x, \hat{W}_{1,2}, \hat{A} \in L_\infty$. Furthermore, from the error dynamics (8) we have $\dot{\Delta} x \in L_\infty$. By integrating (17) on both sides from 0 to ∞, we have $\int_0^\infty [-\lambda_{\min}(Q_x)||\Delta x||_{Q_x}^2] \leq [V_1(0) - V_1(\infty)] < \infty$, which implies that $\Delta x \in L_2$. Since $\Delta x \in L_2 \cap L_\infty$ and $\dot{\Delta} x \in L_\infty$, using Barbalat's Lemma we have $\lim_{t\to\infty} \Delta x = 0$.

Case 2: For bounded modelling error and disturbances i.e. $\xi_1 \leq \bar{\xi}_1$. Equation (16) can be represented as

$$\dot{L}_I \leq -\Delta x^T Q_0 \Delta x + \Delta x^T P \Lambda_\xi^{-1} P \Delta x + \Delta \xi_1^T \Lambda_\xi^{-1} \Delta \xi_1 \leq -\alpha(\|\Delta x\|) + \beta(\|\xi_1\|) \quad (18)$$

where $\alpha(\|\Delta x\|) = (\lambda_{\min}(Q_0) - \lambda_{\max}(P\Lambda_\xi P))\|\Delta x\|^2$ $\beta(||\xi_1||) = \lambda_{\max}(\Lambda_\xi^{-1})||\xi_1||^2$.

Since $\alpha_x, \beta_x, \alpha_y, \beta_y$ are K_∞ functions, L_I is the ISS-Lyapunov function. Using Theorem 3.1 in [24], the dynamics of the identification error (8) is input to state stable, which implies $\Delta x, W_{1,2}, A \in L_\infty$. This completes the proof of Theorem 3.1.

4. Optimal control design

In this section, adaptive optimal control is designed based on the DNN identifier. From Section 3, we know that a nonlinear system (1) can be represented by DNN with the updating law (9) as follows:

$$\dot{x} = A\hat{x} + W_1 \sigma(\hat{x}) + W_2 u + \xi_1 \quad (19)$$

where the model error ξ_1 is still assumed to be bounded $\xi_1 \leq \bar{\xi}_1$. Δx and $W_{1,2}$ are bound as Theorem 3.1.

Then (19) can be further rewritten as

$$\dot{x} = Ax + W_1 \sigma(\hat{x}) + W_2 u + \xi_2 \quad (20)$$

where $\xi_2 = \xi_1 + A\hat{x} - Ax = \xi_1 - A\Delta x$. For bounded ξ_1 and Δx, ξ_2 is bounded as well i.e. $\xi_2 \leq \bar{\xi}_2$.

To achieve optimal tracking control, the control action u is designed as $u = u_r + u_e$ where u_r is the steady-state control which ensures that the tracking error is at the steady state, and u_e is the adaptive optimal control which is used to minimize the infinite horizon performance index function optimally. u_r should be designed to compensate for the nonlinear dynamic in (20). Hence, let u_r be

$$u_r = W_2^+ [\dot{x}_d - Ax - W_1 \sigma(x) - Ke] \quad (21)$$

where $e = x - x_r$ denotes the state tracking error, K is the feedback gain and W_2^+ denotes the generalized inverse of W_2.

From (20) and (21), the error dynamic equation becomes

$$\dot{e} = -Ke + W_2 u_e + \xi_2 \quad (22)$$

In this case, the tracking problem with (20) is transferred to the regulator problem of (22). The adaptive optimal control u_e is designed to stabilize (22) optimally. Hence rewrite the infinite horizon performance cost function (2) as

$$V(e(t)) = \int_t^\infty r(e(\tau), u_e(e(\tau)) d\tau \quad (23)$$

where $r(e, u_e) = e^T Q e + u_e^T R u_e$ is the utility function with the optimal control u_e.

According to the optimal regulator problem design in [25], an admissible control policy u_e should be designed to ensure that the infinite horizon cost function (23) related to (22) is minimized. So, design the Hamiltonian of (22) as

$$H(e, u_e, V) = V_e^T [-Ke + W_2^+ u_e + \xi_2] + e^T Q e + u_e^T R u_e \quad (24)$$

where $V_e = \frac{\partial V(e)}{\partial e}$ is the partial derivative of the value function with respect to e.

Then we define the optimal cost function as

$$V^*(e(t)) = \min_{u_e \in \psi(\Omega)} \left(\int_t^\infty r(e(\tau), u_e(e(\tau)) d\tau \right) \quad (25)$$

and it satisfies the following HJB equation

$$\min_{u_e \in \psi(\Omega)} [H(e, u_e^*, V^*)] = 0 \quad (26)$$

The last optimal control value u_e^* for (22) can be obtained by solving $\frac{\partial H(e, u_e^*, V^*)}{\partial u_e^*} = 0$ from (24)

$$u_e^* = -\frac{1}{2} R^{-1} [W_2]^T \frac{\partial V^*(e)}{\partial e} \quad (27)$$

where $V^*(e)$ is the solution of the HJB equation (26).

From (27), we can learn that the optimal control value u_e^* is based on the optimal value function $V^*(e)$. However, it is difficult to solve the nonlinear partial differential HJB equation (26) to obtain $V^*(e)$. The usual method is to get the approximate solution via a critic NN as [4,5,25]. A single-layer NN will be used to approximate the optimal value function

$$V^*(e) = \mathrm{W}_3^{*\mathrm{T}}\phi\psi(e) + \xi_3 \tag{28}$$

and its derivative is

$$\frac{\partial V^*(e)}{\partial e} = \nabla\psi^T(e)\mathrm{W}_3^* + \nabla\xi_3 \tag{29}$$

where $\mathrm{W}_3^* \in R^{\mathrm{I}}$ is the nominal weight vector, $\psi(e) \in R^I$ is the active function and ξ_3 is the approximation error, I represents the number of neurons. $\nabla\psi(e) = \frac{\partial\psi(e)}{\partial e}$ and $\nabla\xi_3 = \frac{\partial\xi_3}{\partial e}$ are the partial derivatives of $\psi(e)$ and ξ_3 with respect to e, respectively.

Assumption 4.2: The nominal weight vector W_3^*, the active function $\psi(e)$ and its derivative $\nabla\psi(e)$ are all bound, i.e $||\mathrm{W}_3^*|| \le \bar{W}_3, \|\psi(e)\| \le \bar{\psi}_1, \|\nabla\psi(e)\| \le \bar{\psi}_2$ $||\nabla\xi_3|| \le \bar{\psi}_3$.

Then substituting (28) with (27), one obtains

$$u_e^* = -\frac{1}{2}R^{-1}W_2^T(\nabla\phi^T(e)\mathrm{W}_3^* + \nabla\xi_3) \tag{30}$$

The critic NN is approximated as

$$V(e) = \mathrm{W}_3^{\mathrm{T}}\phi(e) \tag{31}$$

where W_3 is the estimation of the nominal W_3^*.

Then the approximate optimal control can be obtained from (30) and (31)

$$u_e = -\frac{1}{2}R^{-1}W_2^T\nabla\phi^T(e)\mathrm{W}_3 \tag{32}$$

Remark 4.2: The available adaptive optimal control method is usually based on the dual NN architecture, where the critic NN and action NN are employed to approximate the optimal cost function and optimal control policy, respectively. The complicated structure and computational burden make it difficult for practical implantation. In the following, we will calculate the optimal control action directly from the critic NN instead of the action-critic dual network.

Substituting (28) with (24), one obtains

$$\begin{aligned} 0 = {} & \mathrm{W}_3^{*\mathrm{T}}\nabla\phi(e)[-Ke + W_2^+u_e] \\ & + e^TQe + u_e^TRu_e + \xi_{HJB} \end{aligned} \tag{33}$$

where $\xi_{HJB} = \mathrm{W}_3^{*\mathrm{T}}\phi(e)\xi_2 + \nabla\xi_3[-Ke + W_2^+u_e + \xi_2]$ is the residual HJB equation error due to the DNN identifier error ξ_2 and NN approximation error $\nabla\xi_3$.

Then (33) can be written as the general identification form as

$$Y = -\mathrm{W}_3^{*\mathrm{T}}X - \xi_{HJB} \tag{34}$$

where $X = \nabla\phi(e)[-Ke + W_2^+u_e]$, $Y = e^TQe + u_e^TRu_e$.

According to the least square method learning rules, one can get the estimation of nominal W_3^* as $\mathrm{W}_3 = -(XX^T)^{-1}XY^T$ in the case of residual HJB equation error equals zero. However, ξ_{HJB} is not always zero and it is also difficult to finish the subsequent closed-loop stability analysis based on the least square method. Inspired by [14,26], we develop a novel robust estimation method of W_3^*. The following equation is used to identify (34)

$$Y = -\mathrm{W}_3^{\mathrm{T}}X - \xi_{HJB1} \tag{35}$$

where ξ_{HJB} can be assumed to be the model error and unknown disturbance.

For (35), the filtered version of Y is defined as

$$\dot{z} = \tau z + Y, \quad z(0) = 0 \tag{36}$$

where $\dot{L}_o = \tilde{W}_3^T\mu^{-1}\dot{\tilde{W}}_3 = -E(t)\tilde{W}_3^T\tilde{W}_3 + \tilde{W}_3^T\varsigma_f \le -\sigma||\tilde{W}_3||^2 - ||\tilde{W}_3||\bar{\varsigma}_f$ is a positive constant,and z is an auxiliary variable.

We further define the auxiliary variables z_f, Y_f, X_f and ξ_{HJB1f} as

$$\begin{cases} \eta\dot{z}_f + z_f = z, z_f(0) = 0 \\ \eta\dot{Y}_f + Y_f = Y, Y_f(0) = 0 \\ \eta\dot{X}_f + X_f = X, X_f(0) = 0 \\ \eta\dot{\xi}_{HJB1f} + \xi_{HJB1f} = \xi_{HJB1}, \xi_{HJB1f}(0) = 0 \end{cases} \tag{37}$$

where η is a filter parameter. It should be noted that the fictitious filtered variable ξ_{HJB1f} is just used for analysis.

Then we get

$$Y_f = -\mathrm{W}_3^{\mathrm{T}}X_f - \xi_{HJB1f} \tag{38}$$

$$\dot{z}_f = -\tau z_f + Y_f \tag{39}$$

From the first equation in (36), one obtains

$$\dot{z}_f = (z - z_f)/\eta \tag{40}$$

According to (38), (39) and (40), we have

$$(z - z_f)/\eta + \tau z_f = -\mathrm{W}_3^{\mathrm{T}}X_f - \xi_{HJB1f} \tag{41}$$

Furthermore, we define the auxiliary regression matrix $E \in R^{l\times l}$ and vector $F \in R^l$ as

$$\begin{cases} \dot{E}(t) = -\eta E(t) + X_fX_f^T, \quad E(0) = 0 \\ \dot{F}(t) = -\eta F(t) + X_f[(z - z_f)/\eta + \tau z_f]F(0) = 0 \end{cases} \tag{42}$$

where η is a positive constant as defined in (37).

The solution of (42) is derived as

$$\begin{cases} E(t) = \int_0^t e^{-\eta(t-r)} X(r) X^T(r) dr \\ F(t) = \int_0^t e^{-\eta(t-r)} X(r)[(z(r) - z_f(r))/\eta \\ \qquad + \tau z_f(r)] dr \end{cases} \tag{43}$$

Finally, we denote a vector M as

$$M = E(t) W_3 + F(t) \tag{44}$$

The adaptive law for updating W_3 is provided by

$$\dot{W}_3 = -\mu M \tag{45}$$

where μ is the learning gain.

Theorem 4.2: *For system* (34) *with the updating law* (44) *then the value function weight error* $\tilde{W}_3 = W_3^* - W_3$ *converges to a compact set around zero.*

Proof: The Lyapunov function is selected as

$$L_o = \frac{1}{2} \tilde{W}_3^T \mu^{-1} \tilde{W}_3 \tag{46}$$

■

Then, by substituting (42) with (44), one obtains

$$M = E(t) W_3 + F(t) = -E(t) \tilde{W}_3 + \varsigma_f \tag{47}$$

where $\varsigma_f = -\int_0^t e^{-\eta(t-r)} X_f \xi_{HJB1f} dr$ is bounded as $||s_f|| \leq \bar{\zeta}_f$.

It can be seen from [26] that the persistently excited (PE) for X can make the matrix defined in (43) is positive define, i.e. $\lambda_{\min}(E) > \sigma > 0$ Then according $\dot{\tilde{W}}_3 = -\dot{W}_3$, the derivative of (46) is calculated as

$$\begin{aligned} \dot{L}_o &= W_3^T \mu^{-1} \dot{\tilde{W}}_3 = -E(t) \tilde{W}_3^T \tilde{W}_3 + \tilde{W}_3^T \varsigma_f \\ &\leq -||\tilde{W}_3||(\sigma ||\tilde{W}_3|| - \bar{\zeta}_f) \end{aligned} \tag{48}$$

Then $\tilde{W}_3$ converges into the compact set $\Omega : \{||\tilde{W}_3|| \leq \bar{\zeta}_f/\sigma\}$

Theorem 4.3: *For system* (1) *with an adaptive optimal control u signal* (21) *and* (32) *and adaptive laws* (9) *and* (45)*, the tracking error e is uniformly ultimately bound, and the optimal control* u_e *in* (32) *converges to a small bound around its ideal optimal solution* u_e^* *in* (30).

Proof: Design the Lyapunov function as

$$L = L_I + L_o + L_c$$

where L_I can be expressed as (10) and the time derivative of (18) satisfies the following inequality

$$\dot{L}_I \leq -(\lambda_{\min}(Q_0) - \lambda_{\max}(P\Lambda_\xi P))\|\Delta x\|^2 + \lambda_{\max}(\Lambda_\xi^{-1})||\xi_1||^2 \tag{49}$$

■

L_o is defined as (46) and its derivation is obtained from (48) such that

$$\begin{aligned} \dot{L}_o &= \tilde{W}_3^T \mu^{-1} \dot{\tilde{W}}_3 = -E(t) \tilde{W}_3^T \tilde{W}_3 + \tilde{W}_3^T \varsigma_f \\ &\leq -\sigma ||\tilde{W}_3||^2 - ||\tilde{W}_3|| \bar{\zeta}_f \end{aligned} \tag{50}$$

From the basic inequality $ab \leq a^2\delta/2 + b^2/2\delta$ with $\delta > 0$, we can rewrite (50) as

$$\dot{L}_o \leq -\left(\sigma - \frac{1}{2\delta}\right) ||\tilde{W}_3||^2 + \frac{\delta \bar{\zeta}_f^2}{2} \tag{51}$$

L_C is defined as

$$L_C = \Gamma e^T e + \kappa V^*(e) \tag{52}$$

where $V^*(e)$ is the optimal cost function defined in (25) and $\Gamma, \kappa > 0$ are positive constants.

Substituting (32) with (22), one obtains

$$\begin{aligned} \dot{e} &= -Ke + W_2(-1/2R^{-1} W_2^T \nabla\varphi^T \mathrm{W}_3) + \xi_2 \\ &= -Ke + 1/2 W_2 R^{-1} W_2^T \nabla\varphi^T \tilde{\mathrm{W}}_3 + W_2 u_e^* \\ &\quad + 1/2 W_2 R^{-1} W_2^T \nabla\varphi^T \nabla\xi_3 + \xi_2 \end{aligned} \tag{53}$$

Then time derivation of (52) can be deduced from (28) and (53) as

$$\begin{aligned} \dot{L}_C &= 2\Gamma e^T \dot{e} + \kappa(-e^T Q e - u_e^{*T} R u_e^*) \\ &= 2\Gamma e^T (-Ke + \frac{1}{2} W_2 R^{-1} W_2^T \nabla\varphi^T \tilde{\mathrm{W}}_3 + W_2 u_e^* \\ &\quad + \frac{1}{2} W_2 R^{-1} W_2^T \nabla\varphi^T \nabla\xi_3 + \xi_2) \\ &\quad + \kappa(-e^T Q e - u_e^{*T} R u_e^*) \\ &\leq -[\Gamma K + \kappa\lambda_{\min}(Q) - \Gamma(||W_2^T R^{-1} W_2 \nabla\varphi|| \\ &\quad + ||W_2^T R^{-1} W_2^T|| + 2)]||e||^2 \\ &\quad + \frac{1}{4}\Gamma(||W_2^T R^{-1} W_2 \nabla\varphi||||\tilde{\mathrm{W}}_3||^2 \\ &\quad - [\kappa\lambda_{\min}(R) - \Gamma||W_2||^2]||u_e^*||^2 \\ &\quad + \frac{1}{2}\Gamma||W_2^T R^{-1} W_2^T||\nabla\xi_3^{\mathrm{T}} \nabla\xi_3 + \Gamma\xi_2^T \xi_2 \end{aligned} \tag{54}$$

Then from (49), (50) and (54), the time derivative of L is $\dot{L} = \dot{L}_I + \dot{L}_o + \dot{L}_c$ and satisfied the following inequality

$$\begin{aligned} \dot{L} \leq & -(\lambda_{\min}(Q_0) - \lambda_{\max}(P\Lambda_\xi P))\|\Delta x\|^2 \\ & - [\Gamma K + \kappa\lambda_{\min}(Q) - \Gamma(||W_2^T R^{-1} W_2^T||(\|\nabla\phi\| \\ & + 1) + 2)]\|\|^2 - [\kappa\lambda_{\min}(R) - \Gamma||W_2||^2]||u_e^*||^2 \\ & - \left[\sigma - \frac{1}{2\delta} - \frac{1}{4}\Gamma(||W_2^T R^{-1} W_2 \nabla\phi||]||\tilde{W}_3||^2 \right. \\ & \lambda_{\max}(\Lambda_\xi^{-1})||\xi_1||^2 + \frac{1}{2}\Gamma||W_2^T R^{-1} W_2^T||\nabla\xi_3^{\mathrm{T}} \nabla\xi_3 \end{aligned}$$

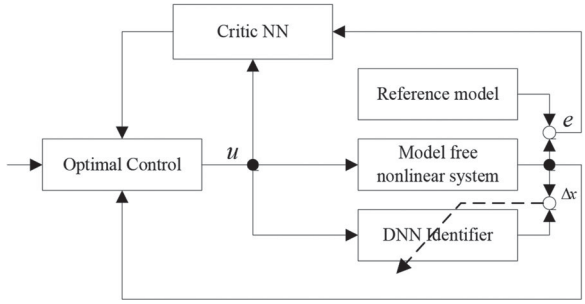

Figure 1. Structural diagram of the control scheme.

$$+\Gamma\xi_2^T\xi_2+\frac{\delta\bar{\xi}_f^2}{2} \tag{55}$$

If we can choose the appropriate parameters to satisfy the following condition

$$\lambda_{\min}(Q_0)>\lambda_{\max}(P\Lambda_\xi P)),\Gamma<\frac{4\sigma\delta-2}{\delta||W_2^TR^{-1}W_2\nabla\phi||}$$
$$\kappa>\max\left\{\frac{\Gamma||W_2||^2}{\lambda_{\min}(R)},\right.$$
$$\left.\frac{\Gamma(||W_2^TR^{-1}W_2^T||(\|\nabla\phi\|+1)+2)}{\lambda_{\min}(Q)}\right\} \tag{56}$$

Then (55) can be further represented as

$$\dot{L}\le-h_1\|\Delta x\|^2-h_2||\tilde{W}_3||^2-h_3\|e\|^2+\vartheta \tag{57}$$

where $h_1=\lambda_{\min}(Q_0)-\lambda_{\max}(P\Lambda_\xi P)$ $h_2=\sigma-\frac{1}{2\delta}-\frac{1}{4}\Gamma(||W_2^TR^{-1}W_2\nabla\phi||)h_3=\Gamma K+\kappa\lambda_{\min}(Q)-\Gamma(||W_2^TR^{-1}W_2^T||(\|\nabla\phi\|+1)+2\,\vartheta=\lambda_{\max}(\Lambda_\xi^{-1})||\xi_1||^2+\frac{1}{2}\Gamma||W_2^TR^{-1}W_2^T||\nabla\xi_3^T\nabla\xi_3+\Gamma\xi_2^T\xi_2+\frac{\delta\bar{\xi}_f^2}{2}$ are all positive constants from condition (56).

Then $\dot{L}<0$ if

$$\|\Delta x\|>\sqrt{\vartheta/h_1},||\tilde{W}_3||>\sqrt{\vartheta/h_2},\|e\|>\sqrt{\vartheta/h_3} \tag{58}$$

which means the identification error $\|\Delta x\|$, the tracking error e and NN weights error $||\tilde{W}_3||$ are all bound.

Moreover, we have

$$\hat{u}_e-u_e^*=\frac{1}{2}R^{-1}W_2^T\nabla\phi^T\tilde{W}_3+\frac{1}{2}R^{-1}W_2^T\nabla\xi_3 \tag{59}$$

When $t\to\infty$, the upper bound of (59) is

$$\lim_{t\to\infty}||\hat{u}_e-u_e^*||$$
$$\le\frac{1}{2}||R^{-1}W_2^T||(||\nabla\phi^T||||\tilde{W}_3||+\nabla\bar{\xi}_3)\le\zeta \tag{60}$$

where ζ depends on the DNN identification approximation error and the critic NN weight error $\tilde{W}_3$.

The structure diagram of the control scheme is illustrated in Figure 1.

A summary of the ADP-based optimal tracking control algorithm is as follows

(1) Select the proper initial values of active functions $\sigma(\cdot)$ and $\phi(\cdot)$ in Equation (6) and updating gains k_1, k_2, k_3 in Equation (9) for the identifier. $\sigma(\cdot)$ is usually selected as the sigmoidal function $\sigma(\cdot)=a/(1+e^{-bx})-c$ where a, b and c are the designed constants. ψ $(\cdot)$ is selected as ψ $(\cdot)=$ I. α,β and γ are tuned online according to equations (9). Hence, there is no need to select the initial weight values of α,β and γ. Meanwhile, select the proper function $\phi(\cdot)$ in Equation (31) and the updating gain μ in Equation (45) for the critic NN ϕ $(\cdot)$ is usually selected as a smooth function consisting of a different combination between state tracking errors.
(2) The inputs/outputs data of an unknown non-affine nonlinear system (1) is used to train the identifier.
(3) Adaptive optimal tracking control law consisting of the steady-state control law in an equation and the optimal control law in Equation (32) is obtained based on the first two steps.

5. Simulations

We consider the following two examples to illustrate the theoretical results in this section.

Example 5.1: Considering the following non-affine nonlinear system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}-x_1+x_2\\ -0.5x_1-0.5x_2(1-(\cos(2x_1+2)^2)\end{bmatrix}$$
$$+\begin{bmatrix}u_1\\ (\cos(2x_1+2))+\sin(u_2)\end{bmatrix} \tag{61}$$

The matrices Q and R of the performance index function are chosen as identify matrices. The control objective is to make the state x_1 and x_2 follow the desired trajectory $x_{1r}=\sin t$ and $x_{2r}=\cos(t)-\sin(t)$. First, a DNN identifier (6) with the updating law (9) is used to identify the non-affine nonlinear system. Parameters are selected as $k_1=k_2=k_3=1$, active function is selected as $\sigma(\cdot)=2/(1+e^{-2x})-0.5$.

The identification error is shown in Figure 2. We can see that the proposed identifier can model the non-affine nonlinear system accurately. Then, with the identified model, the adaptive optimal tracking controller is implemented for the unknown non-affine nonlinear continuous system (61). Define the trajectory error as $e_1=x_1-x_{r1},e_2=x_2-x_{r2}$. The activation function of critic NN is selected as $\phi=[e_1^2,e_1e_2,e_2^2]$. The adaptive gain of the critic NN is selected as $\mu=100$, and the steady control gain is selected as $K=1200$. Figures 3 and 4 represent the trajectory tracking, and the convergence property for the weight of the critic NN is shown

in Figure 5, which demonstrates that the proposed adaptive optimal tracking controller can ensure satisfactory tracking performance for an unknown non-affine nonlinear continue system.

Example 5.2: The classical 2-DOF single-track vehicle model, as shown in Figure 6, is commonly used in AFS/DYC control design [27]. The parameter notations are shown in Table 1.

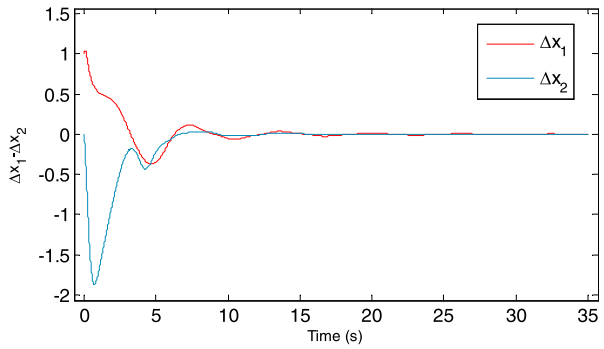

Figure 2. State identification error.

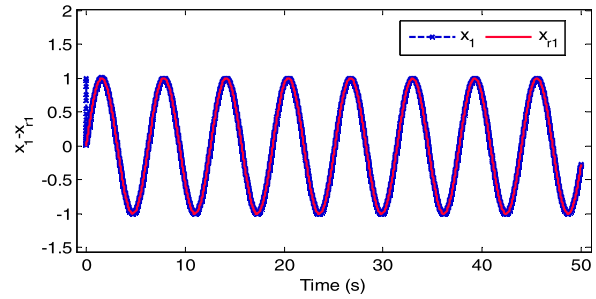

Figure 3. State tracking for x_1.

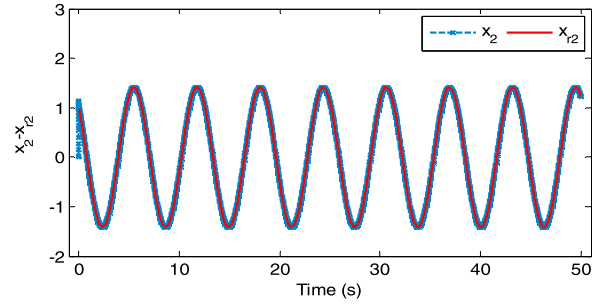

Figure 4. State tracking for x_2.

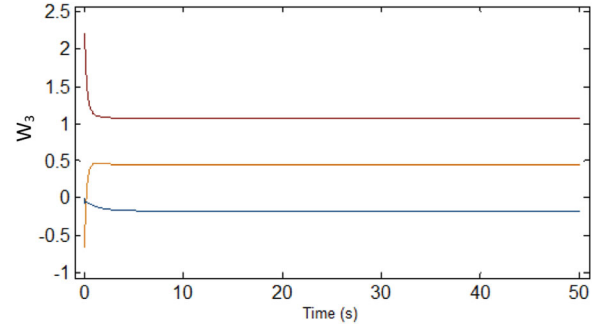

Figure 5. Convergence property for the critic NN weight $x = [\beta \quad \gamma]$,.

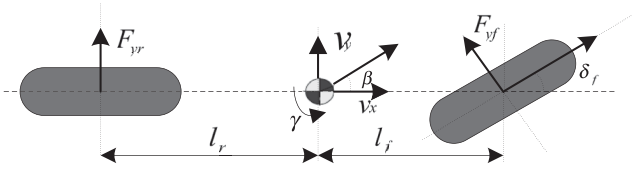

Figure 6. Single-track vehicle model.

Table 1. Description of vehicle model parameters.

Parameters	Description
M	Vehicle mass
I_z	Yaw inertia
l_f,l_r	Distance from CG to the front axle and rear axle
C_{f0}, C_{r0}	Corner stiffness of the front and rear wheel
C_x	Longitudinal tire stiffness
H	CG height
v_x,v_y	Vehicle longitudinal and lateral speed
$RMS = \sqrt{\sum_{i=1}^{n} e^2(i)/n}$	Vehicle longitudinal and lateral acceleration
F_{xi}, F_{yi}	Longitudinal and lateral force, i = 1,2,3,4.
F_{xwi}, F_{ywi}	Longitudinal and lateral tire force
F_{zi}	The normal force of the ith wheel
g	Gravity acceleration
R_w,J_w	Wheel rolling radius, the moment of inertia
ω_{Wi}	Wheel angular speed
T_{wi}	Active brake torque
T	Wheel track width
μ	The friction coefficient between tire and road
γ	Yaw rate about the z-axis
α_i, σ_i	The ith wheel slip angle, slip ratio
δ_f	Steer angle of the front wheel

The mathematical model of Figure 6 considering the uncertainty parameters is expressed as follows:

$$\dot{x} = (A + \Delta A)x + (B + \Delta B)u + E\delta_f$$
$$y = Cx \tag{62}$$

where $x = [\beta \ \gamma]$ β is the side slip angel, γ is the yaw rate; $u = \begin{bmatrix} \delta_c \\ M_c \end{bmatrix}$, δ_c is the active steer angle, M_c is the corrective yaw moment and δ_f is the driver steer input

$$A = \begin{bmatrix} -2\dfrac{C_r + C_f}{mv_x} & -1 - 2\dfrac{C_f l_f - C_r l_r}{mv_x^2} \\ -2\dfrac{C_f l_f - C_r l_r}{I_z} & -2\dfrac{C_f l_f^2 + C_r l_r^2}{I_z v_x} \end{bmatrix},$$

$$B = \begin{bmatrix} \dfrac{2C_f}{mv_x} & 0 \\ \dfrac{2C_f l_f}{I_z} & \dfrac{1}{I_z} \end{bmatrix}, E = \begin{bmatrix} \dfrac{2C_f}{mv_x} \\ \dfrac{2C_f l_f}{I_z} \end{bmatrix}, \Delta A = DFE_1,$$

$$\Delta B = DFE_2, F = \begin{bmatrix} \rho_f & 0 \\ 0 & \rho_r \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} -\dfrac{2C_f \Delta_f}{mv_x} & \dfrac{2C_r \Delta_r}{mv_x} \\ -\dfrac{2C_f \Delta_f l_f}{I_z} & -\dfrac{2C_r \Delta_r l_r}{I_z v_x} \end{bmatrix},$$

$$E_1 = \begin{bmatrix} 1 & \frac{l_f}{v_x} \\ -1 & \frac{l_r}{v_x} \end{bmatrix}, E_2 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}.$$

The main object of vehicle stability control is to design the proper controller to make the actual vehicle yaw rate and sideslip to follow the desired responses. The reference model is usually selected as

$$\dot{x}_r = A_r x_r + E_r \delta_f \tag{63}$$

where $x_r = \begin{bmatrix} \beta_r \\ \gamma_r \end{bmatrix}, A_r = \begin{bmatrix} -\frac{1}{\tau_\beta} & 0 \\ 0 & -\frac{1}{\tau_r} \end{bmatrix}$

$$E_r = \begin{bmatrix} 1 - \frac{ml_f}{2(l_f+l_r)l_r C_r} v_x^2 \\ \hline 1 + \frac{m}{(l_f+l_r)} \left(\frac{l_f}{2C_r} - \frac{l_r}{2C_f} \right) v_x^2 \\ \frac{v_x}{l_f+l_r} \\ \hline 1 + \frac{m}{(l_f+l_r)} \left(\frac{l_f+l_r)}{2C_r} - \frac{l_r}{2C_f} \right) v_x^2 \end{bmatrix}$$

τ_r, τ_β are the designed time constants of raw rate and sideslip angle, respectively.

With the assumption that the variation and uncertainty of tire cornering stiffness can be described as

$$\begin{cases} C_f = C_{f0}(1 + \Delta_f \rho_f), ||\rho_f|| \leq 1 \\ C_r = C_{r0}(1 + \Delta_r \rho_r), ||\rho_r|| \leq 1 \end{cases} \tag{64}$$

where C_{f0}, C_{r0} and C_f, C_r are the nominal and actual cornering stiffness of the front and rear tires respectively, C_{f0}, C_{r0} are the deviation magnitude, ρ_f, ρ_r are perturbations.

Simulation parameters of the vehicle system are selected as $m = 1704$kg, $C_f = 63224$N/rad, $C_r = 84680$ N/rad, $I_z = 3048$ kg m^2, $l_f = 1.135$ m and $l_r = 1.555$ m. A 28-degree step steer manoeuvre with an initial speed (of 80 km/h) is simulated to verify the proposed method. The time-varying parameters of C_f and C_r are obtained from (64) by selecting Δ_f, Δ_r as constant 0.5 and ρ_f, ρ_r as band-limited white noise with the amplitude ± 0.01. As shown in Figures 7 and 8, the proposed method still demonstrates strong robustness and self-adaptive performance, i.e. less tracking error for yaw rate and sideslip angle, when encountering time-varying cornering stiffness in step steer manoeuvre.

To show the identification performance of the proposed algorithm, the performance index-Root Mean Square (RMS) for the state's error has been adopted for comparison.

$$RMS = \sqrt{\sum_{i=1}^{n} e^2(i)/n} \tag{65}$$

where n is the number of the simulation steps, $e(i)$ is the corresponding state response at the ith step.

The RMS values of the side slip angle and yaw rate are 0.915×10^{-4} and 3.173×10^{-4}, respectively.

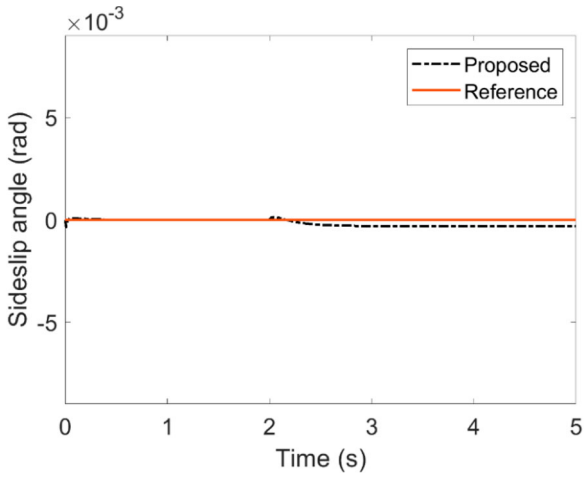

Figure 7. Side-slip angle.

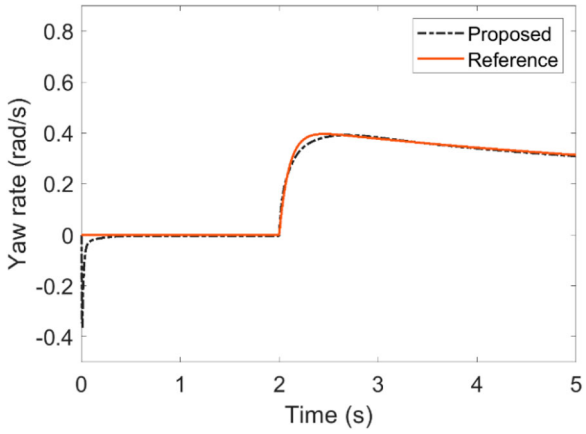

Figure 8. Yaw rate.

6. Conclusions

In this paper, we develop an adaptive optimal controller with a critic-identifier structure to solve the trajectory tracking problem for model uncertain non-affine nonlinear continuous-time system. First, a model-free DNN identifier is designed to reconstruct the unknown dynamic. Then, based on the identification model, an adaptive optimal controller is presented, which can realize the trajectory tracking and stabilize the error dynamic optimally. In addition, a critic NN is introduced to approximate the optimal value function, and a novel robust tuning law is established to update the critic NN weight. The stability of the closed-loop system is proved by the Lyapunov approach. Simulation results of two examples are presented to verify the validity of the proposed approach.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Natural Science Foundation of China (NSFC) under Grant 62073298, Key Research

and Development Projects of Henan Province in 2022 under Grant 221111240200, Special Application for Key Scientific and Technological Project of Henan Province [Grant Number JDG20220037].

References

[1] Powell B. Approximate dynamic programming: solving the curses of dimensionality. New Jersey: Wiley-Blackwell; 2007.
[2] Lebedev D, Margellos K, Goulart P. Convexity and feedback in approximate dynamic programming for delivery time slot pricing. IEEE Trans Control Syst Technol. 2022;30(2):893–900.
[3] Zhang H, Liu D, Luo Y, et al. Adaptive dynamic programming for control: algorithms and stability. London: Springer; 2013.
[4] Lewis FL, Liu D, Editors. Approximate dynamic programming and reinforcement learning for feedback control. Hoboken (NJ): Wiley; 2013.
[5] Werbos P. Approximate dynamic programming for real-time control and neural modeling. handbook of intelligent control, neural, fuzzy, and adaptive approaches. New York (NY): Van Nostrand Reinhold; 1992.
[6] Miller W, Sutton R, Werbos P. Neural networks for control. Cambridge (MA): MIT Press; 1990.
[7] Fairbank M, Alonso E, Prokhorov D. Simple and fast calculation of the second-order gradients for globalized dual heuristic dynamic programming in neural networks. IEEE Trans Neural Netw Learn Syst. 2012;23(10):1671–1676.
[8] Zhu L, Modares H, Peen G, et al. Adaptive suboptimal output-feedback control for linear systems using integral reinforcement learning. IEEE Trans Control Syst Technol. 2015;23(1):264–273.
[9] Wei Q, Liu D, Shi G. A novel dual iterative q-learning method for optimal battery management in smart residential environments. IEEE Trans Ind Electron. 2015;62(4):2509–2518.
[10] Bhasin S, Kamalapurkar R, Johnson M, et al. A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. Automatica (Oxf). 2013;49(1):82–92.
[11] Vrabie D, Pastravanu O, Abu-Khalaf M, et al. Adaptive optimal control for continuous-time linear systems based on policy iteration. Automatica (Oxf). 2009;45:477–484.
[12] Vamvoudakis K, Lewis F. Online actor critic algorithm to solve the continuous-time infinite horizon optimal control problem. Proc Int Joint Conf Neural Netw. 2009;46:3180–3187.
[13] Modares H, Lewis F, Naghibi-Sistani M. Adaptive optimal control of unknown constrained-input dystems using policy iteration and neural networks. IEEE Trans Neural Netw Learn Syst. 2013;24(10):1513–1525.
[14] Na J, Lv Y, Wu X, et al. Approximate optimal tracking control for continuous-time unknown nonlinear systems). Nan Jing, China, Proceedings of the 33rd Chinese control conference; 2014; p. 8990–8995.
[15] Lv Y, Na J, Yang Q, et al. Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics. Int J Control. 2016;89(1):99–112.
[16] Zhang H, Cui L, Luo Y. Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP. IEEE Trans Cybern. 2013;43(1):2168–2267.
[17] Wei Q, Liu D. Adaptive dynamic programming for optimal tracking control of unknown nonlinear systems with application to coal gasification. IEEE Trans Autom Sci Eng. 2014;11(4):1020–1036.
[18] Modares H, Lewis F. Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning. IEEE Trans Autom Control. 2014;59(11):3051–3056.
[19] Kamalapurkar R, Dinhb H, Bhasin S, et al. Approximate optimal trajectory tracking for continuous-time nonlinear systems. Automatica (Oxf). 2015;51:40–48.
[20] Lv Y, Ren X, Na J. Adaptive optimal tracking controls of unknown multi-input systems based on nonzero-sum game theory. J Franklin Inst. 2019;22(12):2226–2236.
[21] Zhang X, Zhang H, Sun Q, et al. Adaptive dynamic programming-based optimal control of unknown non-affine nonlinear discrete-time systems with proof of convergence. Neurocomputing. 2012;91:48–55.
[22] Wang H, Tian Y. Non-affine nonlinear systems adaptive optimal trajectory tracking controller design and application. Stud Inf Control. 2015;24(1):5–11.
[23] Li X, Yu W. Dynamic system identification via recurrent multilayer perceptrons. Inf Sci (Ny). 2002;147:45–63.
[24] Poznyak A, Yu W, Sanchez E, et al. Nonlinear adaptive trajectory tracking using dynamic neural networks. IEEE Trans Neural Netw. 1999;10(6):1402–1411.
[25] Abu-Khalaf M, Lewis F. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. Automatica (Oxf). 2005;41(5):779–791.
[26] Na J, Yang J, Wu X, et al. Robust adaptive parameter estimation of sinusoidal signals. Automatica (Oxf). 2015;53:376–384.
[27] Yang X, Wang Z, Peng W. Coordinated control of AFS and DYC for vehicle handling and stability based on optimal guaranteed cost theory. Veh Syst Dyn. 2009;47(1):57–79.