Adaptive Optimal Tracking Control of Unknown Nonlinear Systems Using System Augmentation

Yongfeng Lv, Jing Na[§], *Member IEEE*, Qinmin Yang, Guido Herrmann, *Senior Member IEEE*

***Abstract*—In this paper, an alternative solution for adaptive optimal tracking control of nonlinear completely unknown systems is proposed. Firstly, an adaptive identifier is used to estimate the unknown system dynamics. Then, a recently developed system augmentation approach is adopted to design the optimal control, where the reference signal is incorporated into the augmented system. Thus, both the feedforward control and feedback control can be obtained simultaneously. Then, a critic neural network (NN) is used to estimate the augmented performance index, and calculate the optimal control action. Thus, the widely used actor NN is not needed. Finally, a new adaptive law recently proposed by the authors is used to online update the NN weight. The closed-loop stability and the convergence of the optimal control are all proved. The feasibility of the suggested approach is demonstrated by a simulation example.**

I. INSTRUCTION

The objective of solving optimal tracking control (OTC) is to design a controller in such a way that the system state or output tracks a given reference in an optimal manner by minimizing a predefined performance index. The direct extension of optimal control schemes used for regulation to solve the OTC problem is not straightforward [1]. For specific continuous-time linear systems, the OTC may be designed by using Riccita equation method [1, 2]. However, only a few results have been suggested for nonlinear systems because it is not trivial to solve the associated Hamilton-Jacobi-Bellman (HJB) equation [3]. Nevertheless, the direct application of dynamic programming (DP) [5] to solve OTC problem also encountered difficulties for high order systems.

Adaptive dynamic programming (ADP) proposed by Werbos [6] has been developed as a feasible method to address the optimal control problems forward-in-time for discrete-time (DT) systems. However, extensions of the ADP methods for continuous-time (CT) systems [7] entail challenges in proving the closed-loop system stability. Moreover, most available ADP results assume that the system dynamics are partially or fully known. To relax these requirements of system dynamics, Zhang et al.[8] used a neural network (NN) identifier to reconstruct unknown drift dynamics, and proposed an adaptive optimal control. We have also suggested a new 'identifier-critic' framework in [4, 9], where the actor NN is not used. This method simplifies the online implementation for systems with unknown dynamics. The convergence of the obtained control to its optimal solution can be rigorously proved by using a new adaptive law in [4, 9]. However, most existing ADP methods for solving OTC problem divide the overall control into feedback and feedforward parts, which are designed separately.

Recently, a new system augmentation method was proposed for designing OTC of nonlinear systems [12], where the generator of the reference to be tracked is augmented into the error system. Thus, the optimal control including the feedback and feedforward parts is obtained in one step, which simplifies the control design and analysis. However, partial system dynamics (e.g. input function) was used in [12].

In this paper, we study the optimal tracking control of nonlinear CT systems with completely unknown dynamics by further improving our previously proposed 'identifier-critic' strategy [3, 4, 9]. However, different to [4, 9], the overall optimal control can be obtained simultaneously by using the the system augmentation method [12]. First, an adaptive identifier as [3, 4, 9] is used to estimate the unknown system dynamics. Then, an augmented system composed of the tracking error dynamics and the desired trajectory is constructed, and a new cost function for the augmented system is suggested. A critic NN is used to *online* approximate the solution of the augmented HJB equation, and to calculate the control action. In this respect, the widely used actor NN is avoided. To online update the weight of identifier NN and critic NN, the adaptive method based on the parameter estimation error that was initially suggested in our paper [13] is used. This direct parameter estimation scheme is clearly different to the widely used ideas of minimizing the residual approximation error in the HJB equation by using the Least-Squares [11] or modified Levenberg-Marquardt algorithms [10]. The stability of the closed-loop system consisting of the identifier and the tracking controller is proved by using Lyapunov theory, and the convergence of the obtained control to a small set around the optimal policy is guaranteed. A simulation example is given to verify the proposed method.

II. PROBLEM FORMULATION

Consider the following affine nonlinear CT system

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

where $x \in \mathbb{R}^n$ is the measurable system state, $u \in \mathbb{R}^m$ is the control input, $f(x) \in \mathbb{R}^n$ is the unknown drift dynamic and $g(x) \in \mathbb{R}^{n\times m}$ is the unknown input dynamic, which are Lipschitz functions.

This work was supported by the Marie Curie Fellowships Project (Grant FP7-PEOPLE-2013-IEF-625531), the NSFC of China (No. 61573174) and the Fundamental Research Funds for the Central Universities.

§ Corresponding to: Jing Na, Email: najing25@163.com.

Jing Na and Yongfeng Lv are with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, 650500, Kunming, China.

Qinmin Yang is with the Department of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China.

Guido Herrmann is with the Department of Mechanical Engineering, University of Bristol, BS8 1TR, UK

The problem is to design an optimal control such that the system state can track a given continuous trajectory x_d, and the following cost function can be minimized

$$V(e(t)) = \int_t^\infty (e^T Q e + u^T R u) d\tau \tag{2}$$

where Q, R are positive definite matrices, $e = x - x_d$ is the tracking error.

The control design to be presented will be conducted in two steps: 1) the unknown system dynamics are estimated in terms of an adaptive identifier; 2) an online optimal control will be obtained to achieve tracking control.

III. IDENTIFICATION OF UNKNOWN DYNAMICS

This section will present an adaptive identifier to estimate the unknown dynamics. Without loss of generality, we assume that the system dynamics are continuous on any compact set Ω, so that $f(x)$ and $g(x)$ can be approximated by NNs [14] as

$$f(x) = \theta\xi(x) + \varepsilon_f, \qquad g(x) = \psi\varsigma(x) + \varepsilon_g \tag{3}$$

where $\theta \in \mathbb{R}^{n\times k_\theta}$ and $\psi \in \mathbb{R}^{n\times k_\psi}$ are the unknown weights, $\xi \in \mathbb{R}^{k_\theta}$ and $\varsigma \in \mathbb{R}^{k_\psi \times m}$ are the regressors, and ε_f and ε_g are the approximation errors.

Assumption 1 [15]**:** The unknown parameters θ, ψ and the approximation errors $\varepsilon_f, \varepsilon_g$ are all bounded.

From (3), system (1) can be rewritten as

$$\dot{x} = W_1^T \phi_1(x,u) + \varepsilon_T \tag{4}$$

where $W_1 = [\theta, \psi]^T \in \mathbb{R}^{d\times n}$ is the unknown parameter matrix, $\phi_1(x,u) = [\xi^T(x), u^T\varsigma^T(x)]^T \in \mathbb{R}^d$ is the regressor vector with $d = k_\theta + k_\psi$, and $\varepsilon_T = \varepsilon_f + \varepsilon_g u$ defines the lumped estimation error.

We define the filtered variables x_f and ϕ_{1f} as

$$\begin{cases} k\dot{x}_f + x_f = x \\ k\dot{\phi}_{1f} + \phi_{1f} = \phi_1 \end{cases} \tag{5}$$

where $k > 0$ is a constant filter parameter.

Furthermore, we define auxiliary matrices $P_1 \in \mathbb{R}^{d\times d}$ and $Q_1 \in \mathbb{R}^{d\times n}$ as

$$\begin{cases} \dot{P}_1 = -\ell_1 P_1 + \phi_{1f}\phi_{1f}^T, & P_1(0) = 0 \\ \dot{Q}_1 = -\ell_1 Q_1 + \phi_{1f}\left[(x - x_f)/k\right]^T, & Q_1(0) = 0 \end{cases} \tag{6}$$

where $\ell_1 > 0$ is a positive constant.

Then the adaptive law for updating $\hat{W}_1$ is provided as

$$\dot{\hat{W}}_1 = -\Gamma_1 M_1 \tag{7}$$

where $\Gamma_1 > 0$ is a constant learning gain, and the matrix $M_1 \in \mathbb{R}^{d\times n}$ containing the parameter estimation error are calculated based on P_1 and Q_1 as

$$M_1 = P_1 \hat{W}_1 - Q_1 \tag{8}$$

As shown in [3, 4, 9], the matrix M_1 can be written as $M_1 = -P_1\tilde{W}_1 + \upsilon_1$, where $\upsilon_1 = -\int_0^t e^{-\ell_1(t-r)}\phi_{1f}(r)\varepsilon_{Tf}^T(r)dr$ with ε_{Tf} being the filtered version of ε_T. Thus, ε_T and υ_1 are all bounded, i.e. $\|\upsilon_1\| \le \varepsilon_{\upsilon 1}$ for positive constant $\varepsilon_{\upsilon 1}$.

Then, we have the following Lemma:

Lemma 1: For system (4) with the adaptive law (7), if the regressor vector ϕ_1 is persistently excited (PE), then

i) If there are no approximation errors (i.e. $\varepsilon_T = 0$), the estimation error $\tilde{W}_1 = W_1 - \hat{W}_1$ exponentially converges to zero.

ii) When there are bounded approximation errors (i.e. $\varepsilon_T \ne 0$), the estimation error $\tilde{W}_1$ converges to a compact set around zero.

We refer to [3, 4, 9] for a similar proof, which will not be provided here due to the limited space.

IV. SYSTEM AUGMENTATION AND OPTIMAL CONTROL DESIGN

In this section, we will propose a optimal tracking control based on the identified system dynamics. For this purpose, system (1) can be further presented as

$$\dot{x} = \hat{\theta}\xi(x) + \hat{\psi}\varsigma(x)u + \varepsilon_N + \varepsilon_T \tag{9}$$

where $\hat{\theta}$ and $\hat{\psi}$ are the estimations of θ and ψ, which can be obtained in the estimated matrix $\hat{W}_1$, and $\varepsilon_N = \tilde{W}_1\phi_1$ is a bounded identifier error because $\tilde{W}_1$ and ϕ_1 are bounded as shown in Theorem 1. Thus, $\|\varepsilon_N\| \le \eta_N$ holds for a positive constant η_N.

Different to available results, e.g. [4,8,9], where the feedforward control and feedback control are designed separately, we will design the control by introducing an augmented system, such that both the feedback and feedforward control are obtained simultaneously by solving an augmented HJB via a critic NN approximation.

A. System augmentation and optimal control design

A new method for solving the OTC problem will be presented by using the system augmentation method. For this purpose, we define the augmented system state as

$$X(t) = \left[e(t)^T \; x_d(t)\right]^T \in \mathbb{R}^{2n} \tag{10}$$

Then an augmented system can be described as

$$\dot{X}(t) = F(X(t)) + G(X(t))u(t) \tag{11}$$

where $F(X) = [f(x) - \dot{x}_d \;\; \dot{x}_d]^T$, $G(X) = [g(x) \;\; 0]^T$ denote the augmented drift and input dynamics. By using the identified system (9), the system can be written as

$$\dot{X}(t) = \hat{F}(X(t)) + \hat{G}(X(t))u(t) + \Im \tag{12}$$

where $\hat{F}(X) = [\hat{\theta}\xi(x) - \dot{x}_d \;\; \dot{x}_d]^T$ and $\hat{G}(X) = [\hat{\psi}\varsigma(x) \;\; 0]^T$ can be obtained based on the identified dynamics, and $\Im = [(\varepsilon_N + \varepsilon_T)^T, 0]^T$ is the identifier error.

For the augmented system (11), the cost function (2) can be represented as

$$V(X(t)) = \int_t^\infty (X^T Q_T X + u^T R_T u) d\tau \tag{13}$$

which is the function of system state X and control u, and $Q_T = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$ and $R_T = \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}$ are also symmetric positive definite matrices with appropriate dimensions.

We find that the optimal tracking control of system (9) can be considered as the optimal regulation of system (11). The problem now is to design an admissible control policy $u(X) \in \mu(X)$ for the augmented system (11) such that the infinite horizon cost (13) is minimized. For this purpose, we define the Hamiltonian of (11) as

$$H(X,u,V_X) = V_X^T\left[F(X(t)) + G(X(t))u(t)\right] + X^T Q_T X + u^T R_T u \tag{14}$$

where $V_X \triangleq \partial V / \partial X$ denotes the partial derivative of the value function V with respect to X.

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

$$V^*(X) = \min_{u\in\Psi(\Omega)}\left(\int_t^\infty (X^T Q_T X + u^{*T} R_T u^*)d\tau\right) \tag{15}$$

and it satisfies the HJB equation

$$0 = \min_{u\in\Psi(\Omega)}\left[H(X,u^*,V^*)\right] \tag{16}$$

Then the optimal control u^* can be derived by solving $\partial H(X,u^*,V^*)/\partial u^* = 0$ from (14) as

$$u^* = -\frac{1}{2}R_T^{-1}G(X)^T\frac{\partial V^*(X)}{\partial X} \tag{17}$$

where V^* is the solution of the HJB equation (16).

To implement the desired control (17), the HJB equation (16) needs to be solved, which is not a trivial task, and the input dynamics $G(X)$ should be known. To remedy these problems, the next section will introduce a practical optimal control.

B. Critic NN and online learning

In this section, an online solution to solve the augmented HJB equation (16) is presented. We will use the value function approximation via a critic NN, which will be used to calculate the optimal control action. Thus, we assume there exists a linearly parameterized NN [4, 8, 17], such that

$$V^*(X) = W_2^T\phi_2(X) + \varepsilon_v \tag{18}$$

and its derivative with respect to X can be given as

$$\frac{\partial V^*(X)}{\partial X} = \nabla\phi_2^T W_2 + \nabla\varepsilon_v \tag{19}$$

where $W_2 \in \mathbb{R}^l$ is the unknown critic NN weight, $\phi_2(X) \in \mathbb{R}^l$ is the basis function vector and ε_v is the approximation error, l is the number of neurons. $\nabla\phi_2 = \partial\phi_2/\partial X$ and $\nabla\varepsilon_v = \partial\varepsilon_v/\partial X$ are the partial derivative of ϕ_2 and ε_v with respect to X, respectively.

Assumption 2 [7]**:** The ideal critic NN weight W_2, rgressor function $\phi_2(\bullet)$ and its derivative $\nabla\phi_2(\bullet)$ are all bounded, i.e. $\|W_2\| \le W_N$, $\|\phi_2\| \le \phi_N$, $\|\nabla\phi_2\| \le \phi_M$; and the errors ε_v and $\nabla\varepsilon_v$ are bounded, e.g. $\|\nabla\varepsilon_v\| \le \phi_\varepsilon$.

By substituting (19) into (17), u^* can be given as

$$u^* = -\frac{1}{2}R_T^{-1}G(X)^T(\nabla\phi_2^T(X)W_2 + \nabla\varepsilon_v) \tag{20}$$

It is noted that the critic NN weight W_2 and the input dynamics $G(X)$ are unknown. Instead, we can use the identified dynamics $\hat{G}(X)$, and the practical critic NN $\hat{V}(X)$ that approximates the ideal cost function $V^*(X)$ as

$$\hat{V}(X) = \hat{W}_2^T\phi_2(X) \tag{21}$$

where $\hat{W}_2$ is the estimation of the critic NN weight W_2, which will be updated online in terms of the adaptive law to be given.

Then from (17) and (21), the practical optimal control u can be obtained as

$$u = -\frac{1}{2}R_T^{-1}\hat{G}(X)^T\frac{\partial\hat{V}(X)}{\partial X} = -\frac{1}{2}R_T^{-1}\hat{G}(X)^T(\nabla\phi_2^T\hat{W}_2) \tag{22}$$

where $\partial\hat{V}(X)/\partial X = \nabla\phi_2^T\hat{W}_2$ is the derivative of the practical critic NN.

The final task to be addressed is to develop an online algorithm to update the weight $\hat{W}_2$, such that $\hat{W}_2$ converges to a small region around its ideal value W_2. Thus, we will present an adaptive law based on the parameter estimation error.

For this purpose, the HJB equation (16) with (20) can be represented as

$$0 = H(X,u,V_X) = W_2^T\nabla\phi\left[\hat{F}(X(t)) + \hat{G}(X(t))u(t)\right] + X^T Q_T X + u^T R_T u + \varepsilon_{HJB} \tag{23}$$

where $\varepsilon_{HJB} = W_2^T\nabla\phi_2\Im + \nabla\varepsilon_v(\hat{F}(X(t)) + \hat{G}(X(t))u(t))$ is the residual HJB error due to the NN approximation errors $\Im$ and $\nabla\varepsilon_v$, which is bounded and can be made arbitrarily small with sufficient NN nodes [7, 16], i.e. $\Im \to 0$ for $k_\theta, k_\psi \to +\infty$ and $\nabla\varepsilon_v \to 0$ for $l \to +\infty$.

We denote $\Xi = \nabla\phi_2[\hat{F}(X) + \hat{G}(X)u]$ and $\Theta = X^T Q_T X + u^T R_T u$, and rewrite the HJB equation (23) as

$$\Theta = -W_2^T\Xi - \varepsilon_{HJB} \tag{24}$$

We can extend the adaptive law based on the parameter estimation error in Section III to '*directly*' estimate W_2. Define the matrix $P_2 \in \mathbb{R}^{l\times l}$ and vector $Q_2 \in \mathbb{R}^l$ as

$$\begin{cases}\dot{P}_2 = -\ell_2 P_2 + \Xi\Xi^T, \ P_2(0) = 0\\ \dot{Q}_2 = -\ell_2 Q_2 + \Xi\Theta, \ Q_2(0) = 0\end{cases} \tag{25}$$

where $\ell_2 > 0$ is a positive constant.

An auxiliary vector $M_2 = P_2\hat{W}_2 + Q_2$ can be calculated based on P_2 and Q_2, which is used to design the adaptive law to update the critic NN weight $\hat{W}_2$ as

$$\dot{\hat{W}}_2 = -\Gamma_2 M_2 \tag{26}$$

Similar to Lemma 1, we can verify that M_2 can be given as $M_2 = -P_2\tilde{W}_2 + \upsilon_2$, where $\upsilon_2 = -\int_0^t e^{-\ell_2(t-r)}\varepsilon_{HJB}(r)\Xi(r)dr$ is a bounded variable , i.e. $\|\upsilon_2\| \le \varepsilon_{\upsilon 2}$.

Thus we have:

Lemma 2: For critic NN (18) with adaptive law (26), if the regressor vector Ξ in (24) id PE, then

i) If there are no approximation errors (i.e. $\varepsilon_{HJB} = 0$), the estimation error $\tilde{W}_2 = W_2 - \hat{W}_2$ exponentially converges to zero.

ii) When there are bounded approximation errors (i.e. $\varepsilon_{HJB} \ne 0$), the estimation error $\tilde{W}_2$ converges to a small compact set around zero.

The proof of Lemma 2 can be conducted following a similar procedure of Lemma 1, and thus will not be given.

C. Stability analysis

This subsection presents the stability analysis. For this purpose, the system dynamics with the proposed optimal control is first studied. By substituting the optimal control (22) into (11), one may have the closed-loop system dynamics as

$$\dot{X} = F(X) + G(X)\left(-\frac{1}{2}R_T^{-1}\hat{G}^T(X)\nabla\phi_2^T\hat{W}_2 + \frac{1}{2}R_T^{-1}G^T(X)(\nabla\phi_2^T W_2 + \nabla\varepsilon_v)\right) + G(X)u^* \quad (27)$$

In this case, we can further verify that $G^T\nabla\phi_2^T W_2 - \hat{G}^T\nabla\phi_2^T\hat{W}_2 = G^T\nabla\phi_2^T\tilde{W}_2 + \tilde{G}^T\nabla\phi_2^T\hat{W}_2$, so that (27) can be rewritten as

$$\dot{X} = F + \frac{1}{2}GR_T^{-1}\left(G^T\nabla\phi_2^T\tilde{W}_2 + \tilde{G}^T\nabla\phi_2^T\hat{W}_2\right) + Gu^* + \frac{1}{2}GR_T^{-1}G^T\nabla\varepsilon_v \quad (28)$$

To facilitate the stability analysis, the following assumption used in the literature (e.g. [7, 16]) is made:

Assumption 3: The dynamics of system (23) fulfill the condition $\|F(X)\| \le b_f\|X\|, \|G(X)\| \le b_g$ for some positive constants $b_f > 0$, $b_g > 0$.

Now, we can summarize the main results of this paper in the following theorem:

Theorem 1: For system (11) with adaptive optimal control (22) and adaptive laws (7) and (26), if the regressor vectors ϕ_1 and Ξ are PE, then

i) If there are no NN approximation errors, the tracking error e and the NN weight errors $\tilde{W}_1$, $\tilde{W}_2$ converge to zero, and the optimal tracking control u in (22) converges to the ideal optimal solution u^*.

ii) When there are NN approximation errors, the system tracking error e and the NN weight errors $\tilde{W}_1$, $\tilde{W}_2$ are all bounded, and the optimal tracking control u in (22) converges to a small set around its optimal solution u^* in (20), i.e. $\|u - u^*\| \le \varepsilon_u$ for a positive constant ε_u.

The proof of Theorem 1 is given in Appendix.

V. SIMULATION

In this section, a numerical simulation is presented to verify the effectiveness of the proposed control method. Consider the following CT system as:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -5x_1 - 0.5x_2^2 + u \end{cases} \quad (29)$$

In the simulation, it is assumed that the knowledge of system dynamics (including the input dynamic $g(x)$ and drift dynamic $f(x)$) in (29) are unavailable. Thus, an identifier will be designed. For this purpose, the unknown parameters $W_1 = [\theta, \psi] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -5 & -0.5 & 1 \end{bmatrix}^T$ associated with the regressor vector $\phi_1(x,u) = [x_2, x_1, x_2^2, u]^T$ can be estimated using the adaptive law (7). The simulation parameters are set as $k = 0.001$, $\ell = 3$, $\Gamma_1 = 3000$. The initial states are given as $x_1(0) = 0, x_2(0) = -1$, the initial identifier weight are $\hat{W}_1(0)=0$. Fig.1 shows the profiles of the identifier parameters with the adaptive law (7), which converge to their true values.

Moreover, to design the optimal control, the given reference to be tracked is given as $x_{1d} = 0.5\sin\sqrt{5}t$, $x_{2d} = 0.5\sqrt{5}\cos\sqrt{5}t$. Then, the augmented system state is defined as $X = [X_1, X_2, X_3, X_4] = [e_1, e_2, x_{1d}, x_{2d}]$, and the regressor vector of the critic NN can be designed as $\phi_2(X) = [X_1^2, X_1X_2, X_1X_3, X_1X_4, X_2^2, X_2X_3, X_2X_4, X_3^2, X_3X_4, X_4^2]^T$. Clearly, a second order polynomial type regressor is used as [4, 8]. The parameters in the control implementation are set as $k_1 = 0.6, k = 0.8$, $\ell_2 = 300$, $\Gamma_2 = 1800$.

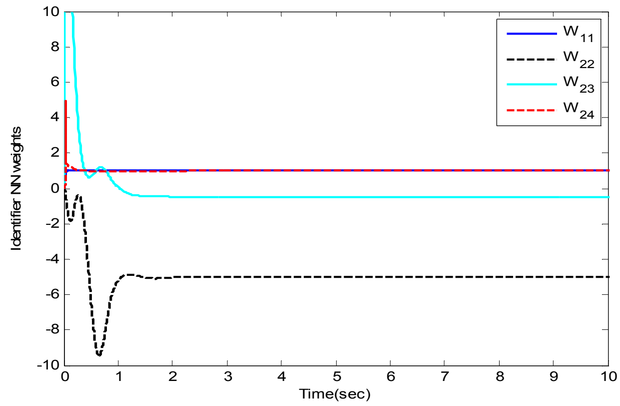

Fig.1. Profiles of Estimated Identifier Weight.

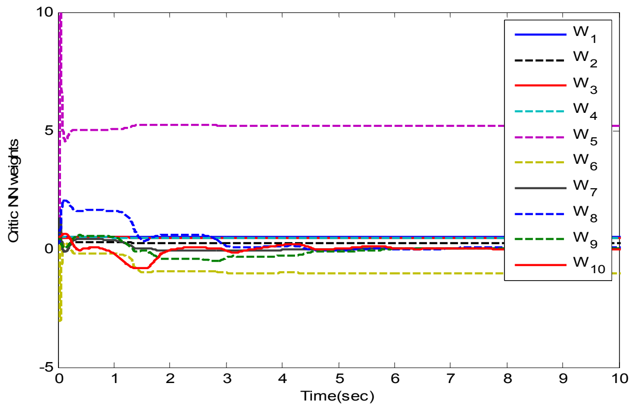

Fig.2. Profiles of Critic NN Weight.

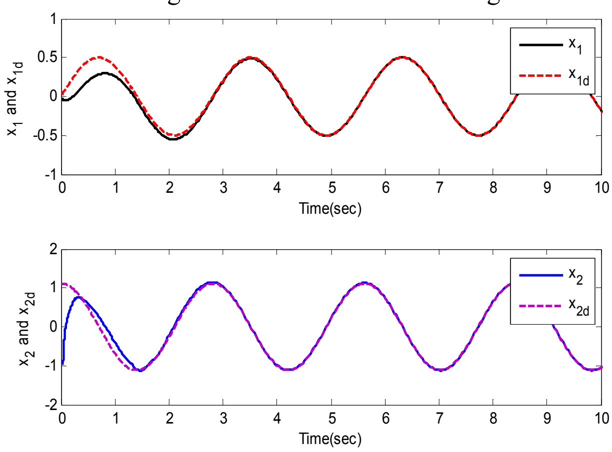

Fig.3. Tracking Control Performances.

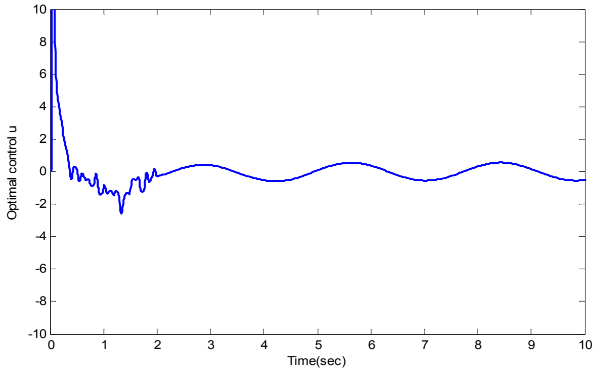

Fig.4. Optimal Control Action.

Fig.2 shows the evolution of the critic NN weight, and Fig.3 provides the tracking response of the controlled system states x_1, x_2, which indicates that satisfactory tracking performance can be obtained with the provided optimal control (22). Moreover, the obtained optimal tracking control action with (22) is provided in Fig.4. These simulation results show the effectiveness of the proposed methods.

VI. CONCLUSION

The 'identifier-critic' based ADP structure recently proposed by the authors is further extended in this paper to solve the optimal tracking problem for completely unknown nonlinear systems. The basic idea is to incorporate a new system augmentation approach into the design of the optimal tracking control, such that two elements of the overall control, i.e. feedforward control and feedback control, can be obtained simultaneously. This result can be taken as a further extension of our previous work [4, 9], where these two control parts are designed separately. Moreover, a new adaptive approach based on the estimated error is employed to update the identifier NN weight and critic NN weight. The convergence of the estimated identifier and critic NN weights and the stability of the closed-loop system are all proved. A numerical simulation is provided to verify the effectiveness of the proposed methods.

Appendix- Proof of Theorem 3

Proof: Consider the Lyapunov function as

$$V = V_1 + V_2 + V_3 + V_4 + V_5 = \frac{1}{2} tr\left(\tilde{W}_1^T \Gamma_1^{-1} \tilde{W}_1\right) + \frac{1}{2} \tilde{W}_2^T \Gamma_2^{-2} \tilde{W}_2 + \Gamma X^T X + K V^* + \Upsilon_1 \upsilon_1^T \upsilon_1 + \Upsilon_2 \upsilon_2^T \upsilon_2 \quad (30)$$

where V^* is the optimal cost function defined in (15) and $K > 0$, $\Gamma > 0$, $\Upsilon_1 > 0, \Upsilon_2 > 0$ are positive constants.

Consider the Young's inequality $ab \le a^2 \eta / 2 + b^2 / 2\eta$ with $\eta > 0$, then we can obtain from (7) and (26) that

$$\begin{aligned} \dot{V}_1 &= -tr\left(\tilde{W}_1^T P \tilde{W}_1\right) + tr\left(\tilde{W}_1^T \upsilon_1\right) \le -\sigma_1 \left\|\tilde{W}_1\right\|^2 + \left\|\tilde{W}_1^T \upsilon_1\right\| \\ &\le -(\sigma_1 - \frac{1}{2\eta}) \left\|\tilde{W}_1\right\|^2 + \frac{\eta \left\|\upsilon_1\right\|^2}{2} \end{aligned} \quad (31)$$

and

$$\begin{aligned} \dot{V}_2 &= -\tilde{W}_2^T P_2 \tilde{W}_2 + \tilde{W}_2^T \upsilon_2 \le -\sigma_2 \left\|\tilde{W}_2\right\|^2 + \left\|\tilde{W}_2^T \upsilon_2\right\| \\ &\le -(\sigma_2 - \frac{1}{2\eta}) \left\|\tilde{W}_2\right\|^2 + \frac{\eta \left\|\upsilon_2\right\|^2}{2} \end{aligned} \quad (32)$$

Moreover, one may calculate $\dot{V}_3$ from (13) and (28) as

$$\begin{aligned} \dot{V}_3 &= 2\Gamma X^T \dot{X} + K(-X^T Q_T X - u^{*T} R_T u^*) \\ &= 2\Gamma X^T \left(F + \frac{1}{2} G R_T^{-1} \left(G^T \nabla \phi_2^T \tilde{W}_2 + \tilde{G}^T \nabla \phi_2^T \hat{W}_2 \right) + G u^* \right. \\ &\quad \left. + \frac{1}{2} G R_T^{-1} G^T \nabla \varepsilon_v \right) + K(-X^T Q_T X - u^{*T} R u^*) \\ &\le -\left[K \lambda_{\min}(Q_T) - 2 b_f \Gamma - \left(\eta b_g^2 \phi_M \lambda_{\max}(R_T^{-1}) + \right.\right. \\ &\quad \left.\left. \eta b_g \phi_M b_w \lambda_{\max}(R_T^{-1}) + 2 \right) \right] \|X\|^2 + \frac{1}{4\eta} \Gamma^2 b_g^2 \phi_M \lambda_{\max}(R_T^{-1}) \left\|\tilde{W}_2\right\|^2 \\ &\quad + \frac{1}{4\eta} \Gamma^2 b_g \phi_M b_w \lambda_{\max}(R_T^{-1}) \left\|\tilde{W}_1\right\|^2 + \frac{1}{4} \Gamma^2 b_g^4 \lambda_{\max}^2(R_T^{-1}) \nabla \varepsilon_v^T \nabla \varepsilon_v \\ &\quad - \left(K \lambda_{\min}(R_T) - \Gamma^2 b_g^2 \right) \left\|u^*\right\|^2 \end{aligned} \quad (33)$$

where $b_w = \left\|\hat{W}_2\right\|$ is a bounded variable.

From (8), it is evident that $\dot{\upsilon}_1 = -\ell_1 \upsilon_1 + \phi_{1f} \varepsilon_{Tf}^T$, so that

$$\begin{aligned} \dot{V}_4 &= 2\Upsilon_1 \upsilon_1^T \dot{\upsilon}_1 = 2\Upsilon_1 \upsilon_1^T \left(-\ell_1 \upsilon_1 + \phi_{1f} \varepsilon_{Tf}^T \right) \\ &\le -(2\Upsilon_1 \ell_1 - \eta) \left\|\upsilon_1\right\|^2 + \frac{1}{\eta} \left\|\Upsilon_1 \phi_{1f} \varepsilon_{Tf}^T\right\|^2 \end{aligned} \quad (34)$$

Moreover, we obtain from (26) that $\dot{\upsilon}_2 = -\ell \upsilon_2 + \Xi \varepsilon_{HJB}$, so that $\dot{V}_5$ can be given as

$$\begin{aligned} \dot{V}_5 &= 2\Upsilon_2 \upsilon_2^T \dot{\upsilon}_2 \\ &= 2\Upsilon_2 \upsilon_2^T \left\{ -\ell_2 \upsilon_2 + \Xi \left[W_2^T \nabla \phi_2 \Im + \nabla \varepsilon_v (F + G u) \right] \right\} \\ &\le -(2\Upsilon_2 \ell_2 - 4\eta) \left\|\upsilon_2\right\|^2 + \frac{1}{\eta} \Upsilon_2^2 W_N^2 \phi_M^2 \|\Xi\|^2 \|\Im\|^2 \\ &\quad + \frac{1}{\eta} \Upsilon_2^2 b_f^2 \phi_\varepsilon^2 \|\Xi\|^2 \|X\|^2 + \frac{1}{4\eta} \Upsilon_2^2 b_g^2 b_w^2 b_\psi^2 \phi_M^2 \lambda_{\max}^2(R_T^{-1}) \|\Xi\|^2 \nabla \varepsilon_v^T \nabla \varepsilon_v \end{aligned}$$

(35)

where $b_\psi = \left\|\hat{\psi} \varsigma\right\|$ is a bounded variable. Consequently, we substitute $\varepsilon_N = \tilde{W}_1 \phi_1$ into (35) and have

$$\begin{aligned} \dot{V} &= \dot{V}_1 + \dot{V}_2 + \dot{V}_3 + \dot{V}_4 + \dot{V}_5 \\ &\le -\left(\sigma_1 - \frac{1}{2\eta} - \frac{1}{4\eta} \Gamma^2 b_g \phi_M b_w \lambda_{\max}(R_T^{-1}) - \frac{1}{\eta} \Upsilon_2^2 W_N^2 \phi_M^2 \|\Xi\|^2 \|\phi_1\|^2 \right) \left\|\tilde{W}_1\right\|^2 \\ &\quad - \left(\sigma_2 - \frac{1}{2\eta} - \frac{1}{4\eta} \Gamma^2 b_g^2 \phi_M \lambda_{\max}(R_T^{-1}) \right) \left\|\tilde{W}_2\right\|^2 \\ &\quad - \left[K \lambda_{\min}(Q_T) - 2\Gamma b_f - \frac{1}{\eta} \Upsilon_2^2 b_f^2 \phi_\varepsilon^2 \|\Xi\|^2 \right. \\ &\quad \left. - \left(\eta b_g^2 \phi_M \lambda_{\max}(R_T^{-1}) + \eta b_g \phi_M b_w \lambda_{\max}(R_T^{-1}) + 2 \right) \right] \|X\|^2 \\ &\quad - \left(2\Upsilon_1 \ell_1 - \frac{3\eta}{2} \right) \left\|\upsilon_1\right\|^2 - \left(2\Upsilon_2 \ell_2 - \frac{9\eta}{2} \right) \left\|\upsilon_2\right\|^2 - \left(K \lambda_{\min}(R_T) - \Gamma^2 b_g^2 \right) \left\|u^*\right\|^2 \\ &\quad + \left(\frac{1}{4} \Gamma^2 b_g^4 \lambda_{\max}^2(R_T^{-1}) + \frac{1}{4\eta} \Upsilon_2^2 b_g^2 b_w^2 b_\psi^2 \phi_M^2 \lambda_{\max}^2(R_T^{-1}) \|\Xi\|^2 \right) \left\|\nabla \varepsilon_v\right\|^2 \\ &\quad + \frac{1}{\eta} \left\|\Upsilon_1 \phi_{1f} \varepsilon_{Tf}^T\right\|^2 + \frac{1}{\eta} \Upsilon_2^2 W_N^2 \phi_M^2 \|\Xi\|^2 \|\Im\|^2 \end{aligned}$$

(36)

Clearly, we can choose the parameters $K, \Gamma, \Upsilon_1, \Upsilon_2, \eta$ fulfilling the following conditions

$$K > (\eta b_g^2 \phi_M \lambda_{\max}(R_T^{-1}) + \eta b_g \phi_M b_w \lambda_{\max}(R_T^{-1}) + 2 + 2\Gamma b_f + \Upsilon_2^2 b_f^2 \phi_\varepsilon^2 \|\Xi\|^2 \frac{1}{\eta}) / \lambda_{\min}(Q)$$

$$\eta > \max\left\{ \left(2 + \Gamma b_g \phi_M b_w \lambda_{\max}(R_T^{-1}) + 4\Upsilon_2^2 W_N^2 \phi_M^2 \|\Xi\|^2 \|\phi_1\|^2 \right) / 4\sigma_1 \right. \\ \left. \left(2 + \Gamma b_g^2 \phi_M \lambda_{\max}(R_T^{-1}) \right) / 4\sigma_2 \right\}$$

$$\Gamma > \sqrt{K \lambda_{\min}(R_T) / b_g^2}, \quad \Upsilon_1 > 3\eta / 4\ell_1, \quad \Upsilon_2 > 9\eta / 4\ell_2 .$$

Then, Eq. (36) can be further presented as

$$\dot{V} \le -a_1 \left\|\tilde{W}_1\right\|^2 - a_2 \left\|\tilde{W}_2\right\|^2 - a_3 \left\|X\right\|^2 - a_4 \left\|\upsilon_1\right\|^2 - a_5 \left\|\upsilon_2\right\|^2 + \gamma \quad (37)$$

where a_1, a_2, a_3, a_4 and a_5 are all positive constants, and

$$\gamma = \left(\frac{1}{4}\Gamma^2 b_g^4 \lambda_{\max}^2 (R_T^{-1}) + \frac{1}{4\eta} \Upsilon_2^2 b_g^2 b_w^2 b_\psi^2 \phi_M^2 \lambda_{\max}^2 (R_T^{-1}) \left\|\Xi\right\|^2 \right) \left\|\nabla \varepsilon_v\right\|^2$$

$$+\frac{1}{\eta}\left\|\Upsilon_1 \phi_{1f} \varepsilon_{Tf}^T\right\|^2 + \frac{1}{\eta}\Upsilon_2^2 W_N^2 \phi_M^2 \left\|\Xi\right\|^2 \left\|\Im\right\|^2$$ is a positive constant.

1) If there are no approximation errors in both the identifier NN and critic NN, i.e. $\varepsilon_T = \nabla \varepsilon_v = 0$ and thus $\varepsilon_N = \Im = \varepsilon_{Tf} = 0$, we know $\gamma = 0$, and then (37) can be rewritten as

$$\dot{V} = -a_1 \left\|\tilde{W}_1\right\|^2 - a_2 \left\|\tilde{W}_2\right\|^2 - a_3 \left\|X\right\|^2 \le 0 \quad (38)$$

According to Lyapunov Theorem, we know $V \to 0$ holds for $t \to +\infty$, such that the estimation error $\tilde{W}_1$, $\tilde{W}_2$ and the tracking error e all converge to zero. In this case, we have $\hat{W}_1 \to W_1$ and $\hat{W}_2 \to W_2$ so that $\hat{\psi}\varsigma(x) \to g(x)$ and $G(X) \to \hat{G}(X)$ holds. Thus, it can be verified that the error between the ideal optimal control u^* in (20) and the proposed approximated optimal control u in (22) can be represented as

$$\begin{aligned} u - u^* &= \frac{1}{2} R_T^{-1} \left(G^T \nabla \phi_2^T \tilde{W}_2 \right) + \frac{1}{2} R_T^{-1} [G - \hat{G}]^T \nabla \phi_2^T W_2 \\ &\quad - \frac{1}{2} R_T^{-1} [G - \hat{G}]^T \nabla \phi_2^T \tilde{W}_2 \end{aligned} \quad (39)$$

so that $\lim_{t \to +\infty} \left\|\hat{u} - u^*\right\| = 0$ is true.

2) When there are bounded approximation errors in the identifier NN and critic NN, it can be shown that $\dot{V}$ is negative if

$$\left\|\tilde{W}_1\right\| > \sqrt{\gamma / a_1}, \quad \left\|\tilde{W}_2\right\| > \sqrt{\gamma / a_2}, \quad \left\|x\right\| > \sqrt{\gamma / a_3},$$
$$\left\|\upsilon_1\right\| > \sqrt{\gamma / a_4}, \quad \left\|\upsilon_2\right\| > \sqrt{\gamma / a_5}$$

which implies that the NN weight errors $\tilde{W}_1$, $\tilde{W}_2$ and the tracking error e are all uniformly ultimately bounded.

To address the convergence of the proposed optimal control, we consider (39) under the NN errors ε_g and $\nabla \varepsilon_v$, and have

$$\begin{aligned} u - u^* &= -\frac{1}{2} R_T^{-1} \hat{G}^T \nabla \phi_2^T \hat{W}_2 + \frac{1}{2} R_T^{-1} G^T \left(\nabla \phi_2^T W_2 + \nabla \varepsilon_v \right) \\ &= \frac{1}{2} R_T^{-1} \left(G^T \nabla \phi_2^T \tilde{W}_2 \right) + \frac{1}{2} R_T^{-1} [G - \hat{G}]^T \nabla \phi_2^T W_2 \\ &\quad - \frac{1}{2} R_T^{-1} [G - \hat{G}]^T \nabla \phi_2^T \tilde{W}_2 + \frac{1}{2} R_T^{-1} G^T \nabla \varepsilon_v \end{aligned} \quad (40)$$

Thus, we can verify that

$$\begin{aligned} \lim_{t \to +\infty} \left\|u - u^*\right\| &\le \frac{1}{2} \lambda_{\max} (R_T^{-1}) \Big[b_g \left(\phi_M \left\|\tilde{W}_2\right\| + \phi_\varepsilon \right) + \phi_M W_N \left(\left\|\tilde{W}_1\right\| + \left\|\varepsilon_g\right\| \right) \\ &\quad + \phi_M \left\|\tilde{W}_2\right\| \left(\left\|\tilde{W}_1\right\| + \left\|\varepsilon_g\right\| \right) \Big] \le \varepsilon_u \end{aligned} \quad (41)$$

where $\varepsilon_u > 0$ is a positive constant. Q.E.D

REFERENCES

[1] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*: Wiley. com, 2012.

[2] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *Automatic Control, IEEE Transactions on,* vol. 59, pp. 3051-3056, 2014.

[3] Y. Lv, J. Na, Q. Yang, X. Wu, and Y. Guo, "Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics," *International Journal of Control,* vol. 89, pp. 99-112, 2016.

[4] J. Na and G. Herrmann, "Online adaptive approximate optimal tracking control with simplified dual approximation structure for continuous-time unknown nonlinear systems," *IEEE/CAA Journal of Automatica Sinica,* vol. 1, pp. 412-422, 2014.

[5] R. Bellman, "Dynamic programming and Lagrange multipliers," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 42, p. 767, 1956.

[6] P. J. Werbos, "A menu of designs for reinforcement learning over time," *Neural networks for control,* pp. 67-95, 1990.

[7] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica,* vol. 41, pp. 779-791, 2005.

[8] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Transactions on Neural Networks,* vol. 22, pp. 2226-2236, 2011.

[9] J. Na, Y. Lv, X. Wu, Y. Guo, and Q. Chen, "Approximate optimal tracking control for continuous-time unknown nonlinear systems," in *Control Conference (CCC), 2014 33rd Chinese*, 2014, pp. 8990-8995.

[10] R. Kamalapurkar, P. Walters, and W. Dixon, "Concurrent learning-based approximate optimal regulation," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, 2013, pp. 6256-6261.

[11] H. Modares, F. L. Lewis, and M. B. Naghibi-Sistani, "Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 24, pp. 1513 - 1525 2013.

[12] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica,* vol. 50, pp. 1780-1792, 2014.

[13] J. Na, G. Herrmann, X. Ren, M. N. Mahyuddin, and P. Barber, "Robust adaptive finite-time parameter estimation and control of nonlinear systems," in *IEEE International Symposium on Intelligent Control (ISIC)*, 2011, pp. 1014-1019.

[14] X. Ren, F. L. Lewis, and J. Zhang, "Neural network compensation control for mechanical systems with disturbances," *Automatica,* vol. 45, pp. 1221-1226, 2009.

[15] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica,* vol. 49, pp. 82-92, 2013.

[16] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica,* vol. 46, pp. 878-888, 2010.

[17] D. Vrabie and F. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks,* vol. 22, pp. 237-246, 2009.