Neural-Network-Based Control for Discrete-Time Nonlinear Systems with Input Saturation Under Stochastic Communication Protocol

Xueli Wang, Derui Ding, *Senior Member, IEEE*, Hongli Dong, *Senior Member, IEEE*, and Xian-Ming Zhang, *Senior Member, IEEE*

***Abstract*—In this paper, an adaptive dynamic programming (ADP) strategy is investigated for discrete-time nonlinear systems with unknown nonlinear dynamics subject to input saturation. To save the communication resources between the controller and the actuators, stochastic communication protocols (SCPs) are adopted to schedule the control signal, and therefore the closed-loop system is essentially a protocol-induced switching system. A neural network (NN)-based identifier with a robust term is exploited for approximating the unknown nonlinear system, and a set of switch-based updating rules with an additional tunable parameter of NN weights are developed with the help of the gradient descent. By virtue of a novel Lyapunov function, a sufficient condition is proposed to achieve the stability of both system identification errors and the update dynamics of NN weights. Then, a value iterative ADP algorithm in an offline way is proposed to solve the optimal control of protocol-induced switching systems with saturation constraints, and the convergence is profoundly discussed in light of mathematical induction. Furthermore, an actor-critic NN scheme is developed to approximate the control law and the proposed performance index function in the framework of ADP, and the stability of the closed-loop system is analyzed in view of the Lyapunov theory. Finally, the numerical simulation results are presented to demonstrate the effectiveness of the proposed control scheme.**

***Index Terms*—Adaptive dynamic programming (ADP), constrained inputs, neural network (NN), stochastic communication protocols (SCPs), suboptimal control.**

Manuscript received October 23, 2020; accepted December 6, 2020. This work was supported in part by the Australian Research Council Discovery Early Career Researcher Award (DE200101128), and Australian Research Council (DP190101557). Recommended by Associate Editor Hongyi Li. *(Corresponding author: Derui Ding.)*

Citation: X. L. Wang, D. Ding, H. L. Dong, and X.-M. Zhang, "Neural-network-based control for discrete-time nonlinear systems with input saturation under stochastic communication protocol," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 4, pp. 766–778, Apr. 2021.

X. L. Wang is with the Department of Control Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: xuelywang@163.com).

D. Ding and X.-M. Zhang are with the School of Software and Electrical Engineering, Swinburne University of Technology, Melbourne 3122, Victoria, Australia (e-mail: dding@swin.edu.au; xianmingzhang@swin.edu.au).

H. L. Dong is with the Institute of Complex Systems and Advanced Control, Northeast Petroleum University, Daqing 163318, China (e-mail: shiningdhl@vip.126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2021.1003922

I. Introduction

OPTIMAL control has been one of the main focuses of control fields due to its wide applications in various emerging industrial systems, such as electrical power systems, industrial control systems, and spacecraft attitude control systems [1]–[7]. It is usually equivalent to solve the well-known Hamilton-Jacobi-Bellman (HJB) equation, which is a critical challenge for nonlinear systems [8]. Fortunately, the adaptive dynamic programming (ADP) algorithm, as the most efficient tool, has been developed to perform various suboptimal control issues with known or unknown system dynamics [9]–[11] by virtue of both its ability of effectively approximating correlation functions and the characteristics of iterative forward transfer. The main idea of ADP algorithms is to utilize two function sequences to iteratively approximate the cost and value functions corresponding to the solution of the HJB equation in a forward-in-time manner [12]. It should be pointed out that the value iteration technology developed in [13], [14] is one of the most important iterative ADP algorithms, and its convergence has also been thoroughly discussed in [15]–[17]. Furthermore, some representative algorithms including heuristic dynamic programming (HDP), dual heuristic dynamic programming (DHP), as well as globalized DHP have been proposed and implemented in various control issues benefiting from the famous actor-critic structure, see [18]–[20]. It is noteworthy that the obtained controller is usually a suboptimal one because of the existence of approximation errors of such a structure, and therefore the corresponding control is also regarded as near-optimal control.

In engineering practice, the actuator saturation is very pervasive due mainly to the facility protection or physical limits of the actuators. If the saturation of the actuator is not considered adequately, the performance of the closed-loop system is often severely damaged [21]. As a result, it is of tremendous significance to survey the influence of the input saturation phenomenon. Under the framework of optimal control, a bounded and invertible one-to-one function in a nonquadratic performance functional is usually exploited to evaluate the cost of saturated inputs and the analytical solution of the optimal controller can be obtained although it is still dependent on the cost functional [8], [22], [23]. Inspired by these work, the near-optimal control for various networked control systems has been investigated and some interesting results have been preliminarily reported in the literature, see

[24]–[27], for instance. Near-optimal regulation under the actor-critic framework has been investigated in [26] for discrete-time nonlinear systems subject to quantized effects, where the quantization errors can be eliminated via a dynamic quantizer with adaptive step size. Furthermore, an online policy iteration algorithm has been presented in [28] to learn the optimal solution for a class of unknown constrained-input systems. Obviously, compared with the case without control constraints, the near-optimal control issues subject to constrained-inputs and various network-induced phenomena remain at an infant stage and thus require further research efforts.

On another frontier of research, in the past few years, we have witnessed the persistent development of network technologies, which has been attracting recurring attention on networked control systems [29]. In order to effectively utilize the limited resource or reduce the switching frequency for prolonging the service life of the equipment, only one (or a limited number of) sensor/control node, governed by various protocols, is permitted to get access to the communication network. These protocols include, but not limited to, the round-robin protocol [30], the try-once-discard protocol [31] and the stochastic communication protocol [32], and the event-triggered protocol [33], [34]. There is no doubt that the utilization of these protocols tremendously results in the complexity and the difficulty of both the stability analysis and the design of weight updating rules, which is the main reason why there are sparse results on this topic. Very recently, consensus control with the help of reconstructed dynamics of the local tracking errors has been investigated in [35] for multi-agent systems with event-triggered mechanism and input constraints, where the effect on the local cost from the adopted triggering scheme has been investigated. The critic and actor networks combined with an identifier network have been simultaneously designed in [27] to deal with a constrained-input control issue with unknown drift dynamics and event-triggered communication protocols. Unfortunately, so far, near-optimal control for the discrete-time nonlinear systems subjected to input saturations has not yet been adequately investigated, not to mention the stochastic communication protocol (SCP) is also a concern, which constitutes the motivation of this paper.

The addressed system with unknown nonlinear dynamics is essentially a protocol-induced switching system when SCP is employed to govern the data transmission or update between the controller and the actuator. Usually, SCP can be modeled by a Markov chain and the relative networked control issues can be effectively handled via the switching system theory combined with Lyapunov approaches. It is worth noting that this is a nontrivial topic for optimal control issues due mainly to the challenge of the cost function from such a switch. Recently, two typical approaches have been, respectively, developed in [36] via a combined cost function related to transition probabilities and in [37] via the dynamic programming principle [38]. However, when an identifier is designed to approximate the unknown nonlinear dynamics, there exists a great challenge to disclose the influence on the updating rules of the identifier's weights and the identification errors. Furthermore, the convergence of the designed ADP algorithm and the practical execution with critic and actor networks should be further inspected. As such, motivated by the above discussions, the focus of this paper is to handle the neural networks (NN)-based near-optimal control problem for a discrete-time nonlinear system subject to constrained-inputs and SCPs. This appears to be nontrivial due to the following essential difficulties: 1) how to design an NN-based identifier under SCPs to estimate system dynamics, 2) how to perform the convergence analysis of the ADP algorithm, and 3) how to disclose the performance of the closed-loop system in the framework of critic and actor networks.

In response to the above discussions, this paper is concerned with the near-optimal control problem for a class of discrete-time nonlinear systems with constrained inputs and SCPs, and hence its main contributions are highlighted as follows: 1) an NN-based identifier with a robust term is presented to approximate the unknown nonlinear system, where novel weight updating rules are constructed by virtue of an additional tunable parameter; 2) a set of conditions are derived to check the stability of both identification error dynamics and updated error dynamics of NN weights; 3) the convergence of proposed value iterative ADP algorithm, which solves the optimal control issue of protocol-induced switching systems with saturation constraints in an off-line way, is profoundly discussed in light of mathematical induction; and 4) an actor-critic NN scheme is employed to perform the addressed near-optimal control issue.

The rest of this paper is formulated as follows: the problem formulation and preliminaries are presented in Section II. For the addressed control issue, four subsections are involved in Section III: an NN-based identification with a robust modification term is designed in Section III-A to identify discrete-time systems with unknown nonlinear dynamics; the value iterative ADP algorithm with convergence analysis is developed in Section III-B; the implementation of ADP algorithm with actor-critic networks in Section III-C, and the performance of closed-loop systems is discussed in Section III-D. Furthermore, a numerical example is given in Section IV to demonstrate the effectiveness of the proposed algorithms. Finally, the conclusion is given in Section V.

Notation: The notation used in this paper is standard. N denotes the set of nonnegative integers. $\mathbb{R}^N$ denotes the set of all N-dimensional real matrices. For the matrix Q, Q^T and $\mathrm{tr}\{Q\}$ denote the transpose and the trace of Q, respectively. $\mathrm{diag}\{Q_1, Q_2, \ldots, Q_n\}$ stands for a block-diagonal matrix where the square matrices Q_i are in the corresponding main diagonal blocks. For a vector x, $\| x \|$ denotes the Euclidean norm.

II. Problem Formulation and Preliminaries

In this paper, the investigated networked control system consists of a nonlinear plant, sensors, identifier, controller, as well as actuator. We assume that the system states x_k can be measured directly by sensors and then sent to the controller via shared networks. By using NNs, identifier along with the controller is utilized to realize the approximation of the nonlinear systems based on the received signals. To reduce the communication burden, SCPs are employed to schedule the

information transmission between the controller and the actuator.

A. System Description

Consider the unknown discrete-time nonlinear system with the following form:

$$x_{k+1} = f(x_k) + g(x_k)\bar{u}_k \tag{1}$$

where $x_k \in \mathbb{R}^M$ is the system state directly measured by sensors, $\bar{u}_k = [\bar{u}_{1,k}, \bar{u}_{2,k}, \ldots, \bar{u}_{N,k}]^T \in \mathbb{R}^N$ is the actuator input scheduled by SCPs, and $f(x_k)$ and $g(x_k)$ are, respectively, the unknown nonlinear functions with $f(0) = 0$ and $g(0) = 0$. Assume that $f + gu$ is Lipschitz continuous on a set Ω_x containing the origin. Furthermore, the actuator input $\bar{u}_k$ belongs to $\Omega_u = \{\bar{u}_k \| \bar{u}_{i,k}| \le \bar{u}\}$ where $\bar{u}$ is a positive scalar representing the saturation-level of the actuator.

In light of unknown nonlinearities, an NN-based system identifier via x_k, which will be designed in the next section, needs to be adopted to obtain the ideal control signal $u_k = [u_{1,k}, u_{2,k}, \ldots, u_{N,k}]^T$. For the convenience of analysis, the saturation constraint can be removed from $\bar{u}_k$ into the control signal u_k, that is, $u_k \in \Omega_u$. In what follows, SCP scheduling is performed to reduce the switching frequency and improve the communication burden between the controller and the actuator. To model this process, let us introduce the scheduling signal $\xi_k \in \{1, 2, \ldots, N\}$ to describe the selected element obtaining access to the network at time instant k. Under SCPs, ξ_k, a random variable, is modeled by a Markov chain with the known transition probability

$$p_{ij} = \text{Prob}\{\xi_{k+1} = j | \xi_k = i\} \tag{2}$$

where $p_{ij} \ge 0$ for $\forall i,\ j = \{1, 2, \ldots, N\}$ and $\sum_{j=1}^{N} p_{ij} = 1$. By means of the above variable, the signal $\bar{u}_k$ received by the actuator is expressed as

$$\bar{u}_{i,k} = \begin{cases} u_{i,k}, & \text{if } i = \xi_k \\ \bar{u}_{i,k-1}, & \text{otherwise} \end{cases} \tag{3}$$

where zero-order-holders are utilized in the viewpoint of practical engineering.

The actuator is further denoted as

$$\bar{u}_k = \Phi(\xi_k) u_k \tag{4}$$

with $\Phi(i) = \text{diag}\{\sigma_i^1, \sigma_i^2, \ldots, \sigma_i^N\}$ where $\sigma_i^n \triangleq \sigma(i-n) \in \{0, 1\}$ $(n = 1, 2, \ldots, N)$ is the Kronecker delta function, i.e., $\sigma(i-n)$ is a binary function that equals 1 if $i = n$ and equals 0 otherwise.

Thus, the closed-loop system is as follows:

$$x_{k+1} = f(x_k) + g(x_k)\Phi(\xi_k)u_k := f_{\xi_k}(x_k) + \tilde{g}_{\xi_k}(x_k)u_k. \tag{5}$$

Remark 1: The main idea of SCP is to assign the access privilege for each node in a random manner. The "random switch" behavior of the node scheduling can be usually characterized by a Markov chain, see the corresponding research in [39]. Obviously, the addressed system (5) is essentially a protocol-induced switching system.

B. Design Objective

To quantify the control performance, the associate utility of each scheduling is employed as follows:

$$\begin{aligned} J_i(x_k) &= \sum_{j=k}^{\infty} \ell_i(x_j, u_j) \\ &= \sum_{j=k}^{\infty} (Q_i(x_j) + S_i(u_j)) \end{aligned} \tag{6}$$

where $\ell_i(x_j, u_j)$ is the cost function, in which $Q_i(x_j)$ is positive and usually a quadratic function, and $S_i(u_j)$ is generally a positive nonquadratic function to evaluate the constrained control input. In this paper, $S_i(u_k)$ is selected as

$$\begin{aligned} S_i(u_k) &= \int_0^{u_k} 2\bar{u}(\tanh^{-1}(\frac{\nu}{\bar{u}}))^T R_i d\nu \\ &= \sum_{i=1}^{N} \int_0^{u_{i,k}} 2\bar{u}(\tanh^{-1}(\frac{\nu_i}{\bar{u}}))^T R_i d\nu_i \end{aligned} \tag{7}$$

where $\tanh(\nu/\bar{u})$ stands for the hyperbolic tangent function; $\nu = [\nu_1, \nu_2, \ldots, \nu_N]^T$ is an integral vector; and $R_i \triangleq \text{diag}\{r_{i,1}, r_{i,2}, \ldots, r_{i,N}\}$ is a known positive definite diagonal matrix with appropriate dimension. The operator $\tanh^{-1}(\nu/\bar{u})$ means

$$\tanh^{-1}(\nu/\bar{u}) = [\tanh^{-1}(\nu_1/\bar{u}), \ldots, \tanh^{-1}(\nu_N/\bar{u})]^T.$$

Via the same with the approach in [27], $S_i(u_k)$ can also be expressed as

$$S_i(u_k) = 2\bar{u}u_k^T R_i \tanh^{-1}(\frac{u_k}{\bar{u}}) + \bar{u}^2 \bar{R}_i \ln(\mathbf{1} - \frac{u_k^2}{\bar{u}^2})$$

where $\bar{R}_i \triangleq [r_{i,1}, r_{i,2}, \ldots, r_{i,N}]$.

Remark 2: In the framework of optimal control, the term $S_i(u_k)$ in the associate utility should satisfy the following three conditions: 1) a continuous and positive function for the performance evaluation, 2) a monotonic function for each component, and 3) a derivable function whose derived function should be invertible for the analytic solution of the optimal control law. Obviously, the adopted $S_i(u_k)$ (7) is definitely the best choice for the control u_k subject to input saturation.

In order to disclose the effect from statistical characteristic of SCPs, similarly to the scheme in [36], reconstruct the performance index function (6) by embedding the transition probability matrix as follows:

$$\begin{cases} J_I(x_k) = p_{11}J_1(x_k) + p_{12}J_2(x_k) + \cdots + p_{1N}J_N(x_k) \\ J_I I(x_k) = p_{21}J_1(x_k) + p_{22}J_2(x_k) + \cdots + p_{2N}J_N(x_k) \\ \qquad \vdots \\ J_N(x_k) = p_{N1}J_1(x_k) + p_{N2}J_2(x_k) + \cdots + p_{NN}J_N(x_k). \end{cases}$$

By virtue of the weighted sum technique, a combined performance index is constructed as

$$J(x_k) = \lambda_1 J_I(x_k) + \lambda_2 J_I I(x_k) + \cdots + \lambda_N J_N(x_k) \tag{8}$$

where $\lambda_i > 0$ is the weight vector satisfying $\sum_{i=1}^{N} \lambda_i = 1$.

Define

$$\Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_N]^T$$

$$L(x_k, u_k) = [l_1(x_k, u_k), \ldots, l_N(x_k, u_k)]$$

where $\Gamma_i = \sum_{s=1}^{n} \lambda_s p_{si} > 0$. It follows from (6) and (8) that

$$\begin{aligned} J(x_k) &= \sum_{i=1}^{N} \Gamma_i J_i(x_k) \\ &= \sum_{i=1}^{N} \Gamma_i l_i(x_k, u_k) + \sum_{i=1}^{N} \Gamma_i J(x_{k+1}) \\ &= \Gamma^T L(x_k, u_k) + J(x_{k+1}). \end{aligned} \tag{9}$$

Before proceeding further, let us introduce the following definition.

Definition 1: A law u_k is an admissible control, if u_k is continuous on the compact set $\Omega_u \subseteq \mathbb{R}^N$ and can stabilize the closed-loop system (5) for all $x_0 \in \Omega_x$ and $J(x_k)$ is finite $\forall x_k \in \Omega_x$.

The purpose of this paper is to find a suboptimal control law u_k^* to optimize the combined performance index (9), which consists of the following three aspects:

1) Designing an NN-based identifer to identify the unknown nonlinear dynamics;

2) Developing a value iterative ADP algorithm to solve the optimal control of protocol-induced switching systems with saturation constraints in an off-line way;

3) In light of the obtained value iterative ADP algorithm, proposing an actor-critic NN scheme to perform the addressed near optimal control.

The following assumption is needed in order to reveal the boundedness of developed approximate scheme in sequel.

Assumption 1: The cost function $\ell_i(x,u)$ $(i = 1, 2, \ldots, n_y)$ satisfies the following conditions:

1) $\ell_i(x,u)$ is continuously differentiable on u and its derivative is denoted as $\psi_{\ell,i}(u) := \partial \ell_i(x,u)/\partial u$;

2) The derivative function $\psi_{\ell,i}(u)$ is invertible with its inverse function denoted as $u_i(x) = \psi_{\ell,i}^{-1}(\partial \ell_i(x,u)/\partial u)$;

3) The inverse function satisfies $\|\psi_{\ell,i}^{-1}(x)\|^2 \le \gamma \|x\|^2$ with a known positive constant γ.

Note, that the function $\ell_i(x,u)$ is quite general with examples including 1) $\kappa \sum_{s=1}^{p} \int_0^u \tanh^{-1}(\nu/\kappa) R_i d\nu$ for nonlinear systems with input constraints; and 2) $x^T P_i x + u^T R_i u$ for linear systems.

III. Main Results

Four subsections are embodied in this section, including the design of the NN-based identifer, the iterative ADP algorithm, and the actor-critic NN scheme, as well as the performance analysis of the identification errors, the iterative ADP algorithm and the closed-loop system.

A. Identification of Closed-Loop Systems via NNs

In this paper, an NN-based approximator is utilized to identify discrete-time nonlinear systems without the knowledge of system dynamics to solve the optimal control issue. Specifically, to learn the unknown nonlinear functions, a stable adaptive weight updating law is proposed for tuning the nonlinear identifier, and a robust modification term, a function of estimated error and an additional tunable parameter, are also introduced to guarantee asymptotic stability of the proposed nonlinear identification scheme.

To start the development of NN-based identifier, the system dynamic (5) is rewritten as

$$x_{k+1} = F_{\xi_k}(x_k, u_k) \tag{10}$$

where $F_i(x_k, u_k) \triangleq f(x_k) + \tilde{g}_i(x_k) u_k$.

According to the universal approximation property of NNs, there exists an NN representation of the function $F_{\xi_k}(x_k, u_k)$ on a compact set Ω_x. In this paper, a three-layer NN is considered as the function approximation structure, under which the number of neurons in the hidden layer is r, the weight matrix (a predetermined constant matrix) between the input and hidden layers is denoted by W_1, and the weight matrix between the hidden layer and output layer is denoted as W_{2,ξ_k}, which needs to be estimated during the training process. In this case, the closed-loop system (10) is further described as

$$x_{k+1} = W_{2,\xi_k}^T \phi_x(\omega_k) + \varepsilon_k \tag{11}$$

where $\omega_k = W_1^T [x_k^T \; u_k^T]^T$ is the hidden layer input, $\phi_x(\omega_k)$ is the bounded activation function satisfying $\|\phi_x(\omega_k)\| \le \phi_{x,m}$, and ε_k is the approximation error salifying a general assumption to be provided as follows.

For the NN represented closed-loop system (11), an identifier is designed to estimate the system state, which is described by

$$\hat{x}_{k+1} = (\hat{W}_{2,\xi_k}^k)^T \phi_x(\omega_k) - q_k \tag{12}$$

where $\hat{W}_{2,\xi_k}^k$ denotes the estimation of the ideal weight matrix W_{2,ξ_k}^k, and q_k is a robust term to reduce the approximation error.

Define the identification error and the estimated error of weight matrix as follows:

$$\tilde{x}_k = \hat{x}_k - x_k, \; \tilde{W}_{2,\xi_k}^k = \hat{W}_{2,\xi_k}^k - W_{2,\xi_k}. \tag{13}$$

Then, subtracting (11) from (12) obtains the following identification error dynamics:

$$\tilde{x}_{k+1} = \hat{x}_{k+1} - x_{k+1} = (\tilde{W}_{2,\xi_k}^k)^T \phi_x(\omega_k) - \varepsilon_k - q_k. \tag{14}$$

Considering this error dynamics, the robust term inspired by the work of [40] is constructed as

$$q_k = \frac{\nu_k \tilde{x}_k}{\tilde{x}_k^T \tilde{x}_k + c_2}$$

where $c_2 > 1$ is a given constant, ν_k is an additional tunable parameter to be designed subsequently. Therefore, the system dynamics (14) can be further rewritten as

$$\begin{aligned} \tilde{x}_{k+1} &= (\tilde{W}_{2,\xi_k}^k)^T \phi_x(\omega_k) - \frac{\nu_k \tilde{x}_k}{\tilde{x}_k^T \tilde{x}_k + c_2} - \varepsilon_k \\ &= \Phi_{1,\xi_k}^k - \Phi_{2,\xi_k}^k - \varepsilon_k \end{aligned} \tag{15}$$

where $\Phi_{1,i}^k$ and $\Phi_{2,i}^k$ are introduced for brevity in writing.

For adopted communication protocols, ξ_k, modeled by a Markov chain with the known transition probability, is usually known via the communication coding. To minimize the square residual error $E_{k+1} = (1/2)\tilde{x}_{k+1}^T \tilde{x}_{k+1}$, the tuning law of $\hat{W}_{2,i}^k$ is given as follows:

$$\hat{W}_{2,i}^{k+1} = \begin{cases} \hat{W}_{2,i}^k - \gamma_w \phi_x(\omega_k) \tilde{x}_{k+1}^T, & \text{if } \xi_k = i, \quad \xi_{k-1} = i \\ \hat{W}_{2,i}^k, & \text{otherwise} \end{cases} \tag{16}$$

and the tuning law of additional tunable parameter ν_k introduced as

$$\nu_{k+1} = \alpha_\nu \nu_k + \frac{\gamma_\nu}{\tilde{x}_k^T \tilde{x}_k + c_2} \tilde{x}_{k+1}^T \tilde{x}_k$$
$$= \alpha_\nu \nu_k + \gamma_\nu \Phi_3^k \tilde{x}_{k+1}^T \tilde{x}_k \tag{17}$$

where $\gamma_w > 0$ is the NN learning rate, and $\alpha_\nu > 0$ and $\gamma_\nu > 0$ are the designed parameters.

Remark 3: The proposed updating rule (16) is novel and nontrivial. First, a zero-order holder is adopted to keep the weights of unactivated subsystems. Specifically, it can be found from the second case that the weights are unchanged along with the time k. Second, the update of weights is performed only when two successive schedulings are satisfied in order to avoid the fluctuation of weight updates.

The following assumption and lemma are used to prove the convergence of the error dynamics.

Assumption 2: The NN approximation error ε_k is upper bounded by a function of identification error $\tilde{x}_k$, that is

$$\varepsilon_k^T \varepsilon_k \le \bar{\vartheta} \tilde{x}_k^T \tilde{x}_k \tag{18}$$

where $\bar{\vartheta}$ is a known constant.

Lemma 1: For any positive definite matrix $\Pi \in \mathbb{R}^{n\times n}$, vectors $x, y \in \mathbb{R}^n$ and scalar $a > 0$, the following inequality is true

$$2x^T \Pi y \le a x^T \Pi x + a^{-1} y^T \Pi y. \tag{19}$$

Theorem 1: Let the identifier (12) be used to identify the nonlinear system (10), where the parameter updating laws given in (16) and (17) are used tuning the NN weights and the robust modification term, respectively. The estimation error $\tilde{x}_k$ in (14) is asymptotically stable while the weights $\hat{W}_{2,i}^k$ and the additional tunable parameter ν_k are convergent if the learning rate γ_w satisfies $6\gamma_w \phi_{x,m}^2 \le \theta_1$, and parameters γ_ν and α_ν satisfy $\gamma_\nu = \gamma_w \phi_{x,m}^2$ and

$$\begin{cases} 0 < \theta_1 < \dfrac{1}{2} \\ 0 < \varepsilon < \dfrac{1}{4} \\ 0 < \bar{\vartheta} < 1 \\ \alpha_\nu < \sqrt{\dfrac{7}{8}}. \end{cases} \tag{20}$$

Proof: Consider the following Lyapunov function candidate

$$L^k = L_1^k + \sum_{s=1}^N L_{2,s}^k + L_3^k$$
$$= \tilde{x}_k^T \tilde{x}_k + \frac{1}{\gamma_w} \sum_{s=1}^N \mathrm{tr}\left\{ (\tilde{W}_{2,s}^k)^T \tilde{W}_{2,s}^k \right\} + \frac{1}{\gamma_\nu} \nu_k^2. \tag{21}$$

Taking the first-order difference of L_1^k along with the dynamics (15) yields

$$E\{\Delta L_1^k | \xi_k = i, x_k\}$$
$$\triangleq E\{\tilde{x}_{k+1}^T \tilde{x}_{k+1} | \xi_k = i, x_k\} - \tilde{x}_k^T \tilde{x}_k$$
$$= \sum_{j=1}^N p_{i,j} \tilde{x}_{k+1}^T \tilde{x}_{k+1} - \tilde{x}_k^T \tilde{x}_k$$
$$= (\Phi_{1,i}^k)^T \Phi_{1,i}^k + (\Phi_{2,i}^k)^T \Phi_{2,i}^k + \varepsilon_k^T \varepsilon_k - \tilde{x}_k^T \tilde{x}_k$$
$$- 2(\Phi_{1,i}^k)^T \Phi_{2,i}^k - 2(\Phi_{1,i}^k)^T \varepsilon_k + 2(\Phi_{2,i}^k)^T \varepsilon_k. \tag{22}$$

Similarly, taking the first-order difference of L_2^k along with the dynamics (16) results into

$$\sum_{s=1}^N E\left\{ \Delta L_{2,s}^k | \xi_k = i, x_k \right\}$$
$$\triangleq \sum_{j=1}^N E\Big\{ \frac{p_{i,j}}{\gamma_w} \mathrm{tr}\Big((\tilde{W}_{2,i}^{k+1})^T \tilde{W}_{2,i}^{k+1} \Big) \Big| \xi_k = i, x_k \Big\}$$
$$+ \sum_{s=1, s\ne i}^N \sum_{j=1}^N \frac{p_{i,j}}{\gamma_w} E\Big\{ \mathrm{tr}\Big((\tilde{W}_{2,s}^{k+1})^T \tilde{W}_{2,s}^{k+1} \Big) \Big| \xi_k = i, x_k \Big\}$$
$$- \frac{1}{\gamma_w} \sum_{s=1}^N \mathrm{tr}\Big((\tilde{W}_{2,s}^k)^T \tilde{W}_{2,s}^k \Big)$$
$$= \frac{1}{\gamma_w} E\Big\{ \mathrm{tr}\Big((\tilde{W}_{2,i}^k - \gamma_w \phi_x(\omega_k) \tilde{x}_{k+1}^T)^T$$
$$\times (\tilde{W}_{2,i}^k - \gamma_w \phi_x(\omega_k) \tilde{x}_{k+1}^T) \Big) | \xi_k = i, x_k \Big\}$$
$$- \frac{1}{\gamma_w} \mathrm{tr}\Big((\tilde{W}_{2,i}^k)^T \tilde{W}_{2,i}^k \Big). \tag{23}$$

Noting $\phi_x(\omega_k) \le \phi_{x,m}$, one has

$$\sum_{s=1}^N E\{\Delta L_{2,s}^k | \xi_k = i, x_k\}$$
$$\le -2\Phi_{1,i}^k \tilde{x}_{k+1} + \gamma_w \phi_{x,m}^2 \tilde{x}_{k+1}^T \tilde{x}_{k+1}$$
$$= -2(\Phi_{1,i}^k)^T \Phi_{1,i}^k + 2(\Phi_{1,i}^k)^T \Phi_{2,i}^k + 2(\Phi_{1,i}^k)^T \varepsilon_k$$
$$+ 3\gamma_w \phi_{x,m}^2 \Big((\Phi_{1,i}^k)^T \Phi_{1,i}^k + (\Phi_{2,i}^k)^T \Phi_{2,i}^k + \varepsilon_k^T \varepsilon_k \Big). \tag{24}$$

Furthermore, it is not difficult to calculate that

$$E\left\{ \Delta L_3^k | \xi_k = i, x_k \right\}$$
$$= \frac{1}{\gamma_\nu} E\left\{ \nu_{k+1}^2 | \xi_k = i, x_k \right\} - \frac{1}{\gamma_\nu} \nu_k^2$$
$$= \frac{1}{\gamma_\nu} \Big((\alpha_\nu \nu_k + \gamma_\nu \Phi_3^k \tilde{x}_{k+1}^T \tilde{x}_k)^2 - \nu_k^2 \Big)$$
$$= 2(\Phi_{2,i}^k)^T \tilde{x}_{k+1} + \gamma_\nu (\Phi_3^k \tilde{x}_{k+1}^T \tilde{x}_k)^2$$
$$- \gamma_\nu^{-1} (1 - \alpha_\nu^2) \nu_k^2$$
$$\le -2(\Phi_{2,i}^k)^T \Phi_{2,i}^k + 2(\Phi_{1,i}^k)^T \Phi_{2,i}^k - 2(\Phi_{2,i}^k)^T \varepsilon_k$$
$$+ 3\gamma_\nu (\Phi_3^k)^2 \tilde{x}_k^T \tilde{x}_k \Big((\Phi_{1,i}^k)^T \Phi_{1,i}^k + (\Phi_{2,i}^k)^T \Phi_{2,i}^k$$
$$+ \varepsilon_k^T \varepsilon_k \Big) - \gamma_\nu^{-1} (1 - \alpha_\nu^2) \nu_k^2. \tag{25}$$

Denote the first-order difference of ΔL^k as

$$\Delta L^k = \Delta L_1^k + \sum_{s=1}^N \Delta L_{2,s}^k + \Delta L_3^k. \tag{26}$$

Considering (22), (24) and (25), the equation (26) can be handled as

$$E\left\{ \Delta L^k | \xi_k = i, x_k \right\}$$
$$\le -(\Phi_{1,i}^k)^T \Phi_{1,i}^k - (\Phi_{2,i}^k)^T \Phi_{2,i}^k - \tilde{x}_k^T \tilde{x}_k$$
$$+ \varepsilon_k^T \varepsilon_k + 2\Phi_{2,i}^k (\Phi_{1,i}^k)^T - \gamma_\nu^{-1} (1 - \alpha_\nu^2) \nu_k^2$$
$$+ 3\Big(\gamma_w \phi_{x,m}^2 + \gamma_\nu (\Phi_3^k)^2 \tilde{x}_k^T \tilde{x}_k \Big)$$
$$\times \Big((\Phi_{1,i}^k)^T \Phi_{1,i}^k + (\Phi_{2,i}^k)^T \Phi_{2,i}^k + \varepsilon_k^T \varepsilon_k \Big).$$

Then, considering Assumption 2 and $(\Phi_3^k)^2\tilde{x}_k^T\tilde{x}_k \le \tilde{x}_k^T\tilde{x}_k$, one has

$$\begin{aligned} &E\left\{\Delta L^k|\xi_k = i, x_k\right\} \\ &\quad\le -(1-3\gamma_w\phi_{x,m}^2-3\gamma_v)(\| \Phi_{1,i}^k \|^2 + \| \Phi_{2,i}^k \|^2) \\ &\qquad -(1-\bar{\vartheta}-3\bar{\vartheta}(\gamma_w\phi_{x,m}^2+\gamma_v)) \| \tilde{x}_k \|^2 \\ &\qquad +2 \| \Phi_{1,i}^k \| \| \Phi_{2,i}^k \| -\gamma_v^{-1}(1-\alpha_v^2)\nu_k^2. \end{aligned} \tag{27}$$

Furthermore, noting

$$\|\Phi_{2,i}^k\|^2 = \left\| \frac{\nu_k\tilde{x}_k}{\tilde{x}_k^T\tilde{x}_k + c_2} \right\|^2 \le \nu_k^2 \tag{28}$$

one has

$$\begin{aligned} 2 \| \Phi_{1,i}^k \| \| \Phi_{2,i}^k \| &\le \theta_1 \| \Phi_{1,i}^k \|^2 + \theta_1^{-1} \| \Phi_{2,i}^k \|^2 \\ &\le \theta_1 \| \Phi_{1,i}^k \|^2 + \varepsilon\theta_1^{-1} \| \Phi_{2,i}^k \|^2 \\ &\quad + \theta_1^{-1}(1-\varepsilon)\nu_k^2 \end{aligned}$$

where the scalar ε belongs to (0, 1).

Furthermore, select the parameters as $\gamma_v = \gamma_w\phi_{x,m}^2$, and $6\gamma_w\phi_{x,m}^2 \le \theta_1$. Applying Lemma 1, it follows from (27) that

$$\begin{aligned} &E\{\Delta L^k|\xi_k = i, x_k\} \\ &\le -(1-3\gamma_w\phi_{x,m}^2-3\gamma_v-\theta_1) \| \Phi_i^k \|^2 \\ &\quad -(1-3\gamma_w\phi_{x,m}^2-3\gamma_v-\varepsilon\theta_1^{-1}) \| q_k \|^2 \\ &\quad -\Big(1-\bar{\vartheta}-3\bar{\vartheta}(\gamma_w\phi_{x,m}^2+\gamma_v)) \| \tilde{x}_k \|^2 \\ &\quad -(\gamma_v^{-1}(1-\alpha_v^2)-\theta_1^{-1}(1-\varepsilon)\Big)\nu_k^2 \\ &= -(1-6\gamma_v-\theta_1) \| \Phi_i^k \|^2 -(1-6\gamma_v-\varepsilon\theta_1^{-1}) \| q_k \|^2 \\ &\quad -(1-\bar{\vartheta}-6\bar{\vartheta}\gamma_v) \| \tilde{x}_k \|^2 -(\gamma_v^{-1}(1-\alpha_v^2) \\ &\quad -\theta_1^{-1}(1-\varepsilon))\nu_k^2 \\ &\le -(1-2\theta_1) \| \Phi_i^k \|^2 -(1-\theta_1-\varepsilon\theta_1^{-1}) \| q_k \|^2 \\ &\quad -(1-\bar{\vartheta}(1+\theta_1)) \| \tilde{x}_k \|^2 -\Big((\frac{\theta_1}{6})^{-1}(1-\alpha_v^2) \\ &\quad -\theta_1^{-1}(1-\varepsilon)\Big)\nu_k^2. \end{aligned} \tag{29}$$

Therefore, one has $E\{\Delta L^k|\xi_k = i, x_k\} < 0$ if the following inequalities hold

$$\begin{cases} 1-2\theta_1 > 0 \\ 1-\theta_1-\varepsilon\theta_1^{-1} > 0 \\ 1-\bar{\vartheta}(1+\theta_1) > 0 \\ \left(\dfrac{\theta_1}{6}\right)^{-1}(1-\alpha_v^2)-\theta_1^{-1}(1-\varepsilon) \ge 0 \end{cases} \tag{30}$$

which yields

$$\begin{cases} 0 < \theta_1 < \dfrac{1}{2} \\ \theta_1^2-\theta_1 < \varepsilon < \theta_1-\theta_1^2 \\ \bar{\vartheta} < \dfrac{1}{1+\theta_1} \\ 6(1-\alpha_v^2) > (1-\varepsilon). \end{cases} \tag{31}$$

that is, the inequalities (20). ■

Remark 4: It should be pointed out that the approximate error ε_k of NNs should be dependent on system states and will trend to zero as system states are close to the original point. As such, the feature should be adequately taken into consideration. In this paper, a robust term q_k with an additional tunable parameter ν_k, inspired by the work of [40], is employed to improve the system perform while guaranteeing the asymptotic stability. Furthermore, the update of $\hat{W}_{2,i}^k$ in (16) is affected by the Markov jump, and therefore a novel Lyapunov function candidate is constructed by adding the term $(1/\gamma_w)\sum_{s=1}^N \mathrm{tr}\left\{(\tilde{W}_{2,s}^k)^T\tilde{W}_{2,s}^k\right\}$ to discover the desired condition of identifer dynamics.

B. Design of ADP Algorithm

According to the Bellman's optimality principle, the optimal performance index function $J^*(x_k)$ satisfies the discrete-time HJB equation

$$J^*(x_k) = \min_{u_k}\{\Gamma^T L(x_k, u_k) + J^*(x_{k+1})\} \tag{32}$$

and the corresponding optimal control strategy is given by

$$u_k^* = \arg\min_{u_k}\{\Gamma^T L(x_k, u_k) + J^*(x_{k+1})\}. \tag{33}$$

Assume that the minimum on the right-hand side of (32) exists and is unique. Taking the first-derivative of the right-hand part, the ideal optimal control u_k^* is given by

$$\begin{aligned} u_k^* &= -\bar{u}\tanh\Big(\frac{1}{2\bar{u}}R^{-1}g_i^T(x_k)\nabla J^*(x_{k+1})\Big) \\ &= -\bar{u}\tanh\Big(\frac{1}{2\bar{u}}R^{-1}g_i^T(x_k)\nabla J^*(\Gamma_i(x_k)+g_i(x_k)u_k\Big). \end{aligned} \tag{34}$$

Since the direct solution of the HJB equation for nonlinear systems is computationally intensive, the value iteration algorithm, usually named as an ADP algorithm, needs to be developed in light of the Bellman's principle of optimality. Initializing the value function $J_0(x_k) = \Sigma(x_k)$, one construct the following iterative algorithm:

$$\begin{aligned} u_s(x_k) &= \arg\min_{u_k}\{\Gamma^T L(x_k, u_k) + J_s(x_{k+1})\} \\ &= \arg\min_{u_k}\{\Gamma^T L(x_k, u_k) + J_s(F_i(x_k, u_k)\} \end{aligned} \tag{35}$$

and

$$\begin{aligned} J_{s+1}(x_k) &= \min_{u_k}\{\Gamma^T L(x_k, u_k) + J_s(x_{k+1})\} \\ &= \Gamma^T L(x_k, u_s(x_k)) + J_s(F_i(x_k, u_k) \end{aligned} \tag{36}$$

where s is the iterative step, and $J_s(x_k)$ and $u_s(x_k)$ are used to approximate $J^*(x_k)$ and u_k^*, respectively, as $s \to \infty$.

Inspired by [17], [41], we further demonstrate the convergence of the developed scheme with the help of a "functional bound" method.

Lemma 2: Consider the sequences $J_s(x_k)$ and $u_s(x_k)$ introduced by (36) and (35), respectively. Given the initial value function $J_0(x)$ for $x \in \mathbb{R}^{M+N}$, the value function $J_s(x)$ is a monotonically sequence as s increases. Specifically, if $J_0(x) \le J_1(x)$, the value function $J_s(x)$ is a monotonically nondecreasing sequence as s increases, i.e., $J_s(x) \le J_{s+1}(x)$, otherwise the value function $J_s(x)$ is a monotonically decreasing sequence as s increases, i.e., $J_s(x) > J_{s+1}(x)$.

Proof: Let us first prove the case of $J_0(x) \le J_1(x)$ by mathematical induction. Consider (36) and $J_0(x) \le J_1(x)$, one has

$$\begin{aligned}J_2(x_k) &= \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_1(x_{k+1})\}\\ &\le \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_0(x_{k+1})\}\\ &= J_1(x_k).\end{aligned}$$

Assume that $J_{q-1}(x)\le J_q(x)$ holds when $s=q-1$. Then, for $s=q$, one can conclude that

$$\begin{aligned}J_{q+1}(x_k) &= \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_q(x_{k+1})\}\\ &\le \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_{q-1}(x_{k+1})\}\\ &= J_q(x_k).\end{aligned}$$

Therefore, this case is true. Similarly, one can conclude that $J_s(x)\ge J_{s+1}(x)$ when $J_0(x)\ge J_1(x)$. ■

Theorem 2: Consider the sequences $J_s(x_k)$ and $u_s(x_k)$ introduced in (36) and (35), respectively. If there exist four constants $\underline{\rho}$, $\bar{\rho}$, $\underline{\varrho}$ and $\bar{\varrho}$ satisfying $0<\underline{\rho}\le\bar{\rho}$ and $0\le\underline{\varrho}\le\bar{\varrho}$ such that

$$\underline{\rho}\{\Gamma^T L(x_k,u_k)\}\le J^*(x_k)\le\bar{\rho}\{\Gamma^T L(x_k,u_k)\} \tag{37}$$

and

$$\underline{\varrho}J^*(x_k)\le J_0(x_k)\le\bar{\varrho}J^*(x_k) \tag{38}$$

hold uniformly, then the iterative value function $J_s(x_k)$ converges to the optimal value $J^*(x_k)$ as $s\to\infty$, i.e.,

$$\lim_{s\to\infty}J_s(x_k)=J^*(x_k). \tag{39}$$

Proof: To verify this result, we will first prove the following assertion by using the mathematical induction method.

Assertion: Case I: For parameters $0\le\underline{\varrho}\le\bar{\varrho}<1$, the iterative value function $J_s(x_k)$ satisfies

$$\begin{aligned}&\Big(1+\frac{\underline{\varrho}-1}{(1+\bar{\rho}^{-1})^s}\Big)J^*(x_k)\le J_s(x_k)\\ &\qquad\le\Big(1+\frac{\bar{\varrho}-1}{(1+\underline{\rho}^{-1})^s}\Big)J^*(x_k).\end{aligned} \tag{40}$$

Case II: For parameters $0\le\underline{\varrho}\le 1\le\bar{\varrho}\le\infty$, the iterative value function $J_s(x_k)$ satisfies

$$\begin{aligned}&\Big(1+\frac{\underline{\varrho}-1}{(1+\bar{\rho}^{-1})^s}\Big)J^*(x_k)\le J_s(x_k)\\ &\qquad\le\Big(1+\frac{\bar{\varrho}-1}{(1+\bar{\rho}^{-1})^s}\Big)J^*(x_k).\end{aligned} \tag{41}$$

Case III: For parameters $1\le\underline{\varrho}\le\bar{\varrho}\le\infty$, the iterative value function $J_s(x_k)$ satisfies (40).

Considering the limited space, we only prove the left-hand side of the inequality (40) in Case I and the right-hand side of the inequality (41) in Case II. Furthermore, the proof of Case III is similar to those the first two cases and hence its proof is omitted.

Obviously, the left-hand side of the inequality (40) in Case I holds for $s=0$. Then, combing with the condition (37), one can derive that

$$\begin{aligned}J_1(x_k) &= \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_0(x_{k+1})\}\\ &\ge \min_{u_k}\{\Gamma^T L(x_k,u_k)+\underline{\varrho}J^*(x_{k+1})\}\\ &\ge \min_{u_k}\Big\{\Gamma^T L(x_k,u_k)+\underline{\varrho}J^*(x_{k+1})\\ &\quad+\frac{(\underline{\varrho}-1)}{1+\bar{\rho}}(\bar{\rho}\Gamma^T L(x_k,u_k)-J^*(x_{k+1}))\Big\}\\ &\ge \min_{u_k}\Big\{\Big(1+\frac{\bar{\rho}(\underline{\varrho}-1)}{1+\bar{\rho}}\Big)\Gamma^T L(x_k,u_k)\\ &\quad+\Big(\underline{\varrho}-\frac{\underline{\varrho}-1}{1+\bar{\rho}}\Big)J^*(x_{k+1})\Big\}\\ &= \Big(1+\frac{\underline{\varrho}-1}{1+\bar{\rho}^{-1}}\Big)\min_{u_k}\{\Gamma^T L(x_k,u_k)+J^*(x_{k+1})\}\\ &= \Big(1+\frac{\underline{\varrho}-1}{1+\bar{\rho}^{-1}}\Big)J^*(x_k).\end{aligned}$$

Furthermore, assume that the conclusion holds for $s=q-1$, that is

$$\Big(1+\frac{\underline{\varrho}-1}{(1+\bar{\rho}^{-1})^{q-1}}\Big)J^*(x_k)\le J_{q-1}(x_k).$$

When $s=q$, combining with the condition (37) again, one has

$$\begin{aligned}J_q(x_k) &= \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_{q-1}(x_{k+1})\}\\ &\ge \min_{u_k}\Big\{\Gamma^T L(x_k,u_k)+\Big(1+\frac{\underline{\varrho}-1}{(1+\bar{\rho}^{-1})^{q-1}}\Big)J^*(x_{k+1})\Big\}\\ &\ge \min_{u_k}\Big\{\Gamma^T L(x_k,u_k)+\Big(1+\frac{\underline{\varrho}-1}{(1+\bar{\rho}^{-1})^{q-1}}\Big)J^*(x_{k+1})\\ &\quad+\frac{(\underline{\varrho}-1)(\bar{\rho}\Gamma^T L(x_k,u_k)-J^*(x_{k+1}))}{(1+\bar{\rho})(1+\bar{\rho}^{-1})^{q-1}}\Big\}\\ &= \Big(1+\frac{\underline{\varrho}-1}{(1+\bar{\rho}^{-1})^{q}}\Big)\min_{u_k}\{\Gamma^T L(x_k,u_k)+J^*(x_{k+1})\}\\ &= \Big(1+\frac{\underline{\varrho}-1}{(1+\bar{\rho}^{-1})^{q}}\Big)J^*(x_k).\end{aligned}$$

According to the mathematical induction method, the left-hand side of the inequality (40) holds.

In what follows, let us prove the right-hand side of the inequality (41) in Case II. Obviously, it is not difficult to find that

$$\begin{aligned}J_1(x_k) &= \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_0(x_{k+1})\}\\ &\le \min_{u_k}\{\Gamma^T L(x_k,u_k)+\bar{\varrho}J^*(x_{k+1})\}\\ &\le \min_{u_k}\Big\{\Gamma^T L(x_k,u_k)+\bar{\varrho}J^*(x_{k+1})\\ &\quad+\frac{\bar{\varrho}-1}{1+\bar{\rho}}(\bar{\rho}\Gamma^T L(x_k,u_k)-J^*(x_{k+1}))\Big\}\\ &= \Big(1+\frac{\bar{\varrho}-1}{1+\bar{\rho}^{-1}}\Big)J^*(x_k)\end{aligned}$$

where the term $\bar{\rho}\Gamma^T L(x_k,u_k)-J^*(x_{k+1})$ induced by the condition (37) is added.

Furthermore, assume that the conclusion holds for $s=q-1$, that is

$$J_{q-1}(x_k)\le\Big(1+\frac{\bar{\varrho}-1}{(1+\bar{\rho}^{-1})^{q-1}}\Big)J^*(x_k).$$

When $s=q$, combining with the condition (37), one has

$$
\begin{aligned}
J_q(x_k) &= \min_{u_k}\{\Gamma^T L(x_k,u_k)+J_{q-1}(x_{k+1})\}\\
&\le \min_{u_k}\Big\{\Gamma^T L(x_k,u_k)+\Big(1+\frac{\bar{\varrho}-1}{(1+\bar{\rho}^{-1})^{q-1}}\Big)J^*(x_{k+1})\Big\}\\
&\le \min_{u_k}\Big\{\Gamma^T L(x_k,u_k)+\Big(1+\frac{\bar{\varrho}-1}{(1+\bar{\rho}^{-1})^{q-1}}\Big)J^*(x_{k+1})\\
&\quad+\frac{(\bar{\varrho}-1)(\bar{\rho}\Gamma^T L(x_k,u_k)-J^*(x_{k+1}))}{(1+\bar{\rho})(1+\bar{\rho}^{-1})^{q-1}}\Big\}\\
&=\Big(1+\frac{\bar{\varrho}-1}{(1+\bar{\rho}^{-1})^{q}}\Big)J^*(x_k).
\end{aligned}
$$

In light of the mathematical induction method, the right-hand side of the inequality (41) holds.

Combining the above conclusions, we can obtain that this assertion is true. Finally, letting $s\to\infty$, the convergence (39) is easily derived. ■

Remark 5: The above theorem discloses the convergence of the developed ADP scheme with the help of a "functional bound" method, which comes from [17], [41]. An assertion has been allocated for the convenience of processing. For the practical application, a terminal condition (or a fixed size number s^*) is adopted, the related algorithm (i.e., ADP Algorithm 1) is provided as follows.

Algorithm 1 ADP algorithm

Initialization Value $J_0(x_k)=\Sigma(x_k)$ and error threshold $\varpi>0$. Set $s=0$.

1: **while** $|J_s(x_k)-J_{s-1}(x_k)|>\varpi$ **do**
2: Solve $u_s(x_k)$ according to (35);
Update the value $J_{s+1}(x_k)$ according to (36);
Set $s=s+1$;
3: **end while**
4: **Output** Control strategy $u_{s-1}(x_k)$.

C. Implementation of ADP Algorithm

Due to the unknown $J_s(x_{k+1})$, a approximation structure via NNs is employed to approximate both $J^*(x_k)$ and $u^*(x_k)$. Such a structure consists of a critic network and an actor network, which are all chosen as three-layer feed forward NNs and their implementation process is shown in Fig. 1. In light of the above conception, the optimal value function (32) and the control input (34) can be described by the following critic NN and actor NN:

$$J^*(x_k)=W_{2c}^T\phi_c(W_{1c}^T z_k)+\theta_c(z_k) \tag{42}$$

and

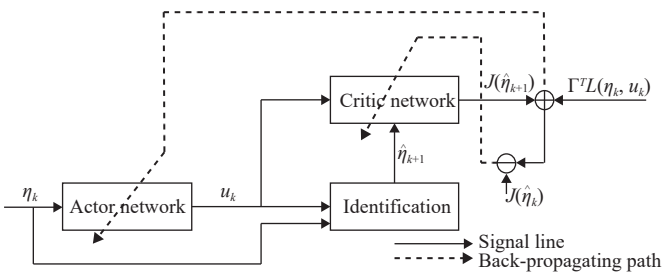


Fig. 1. Neural network structure of the proposed ADP approach.

$$u^*(x_k)=\phi_{2a}(W_{2a}^T\phi_{1a}(W_{1a}^T x_k))+\theta_a(x_k) \tag{43}$$

with $z_k=[x_k^T\ u_k^T]^T$ where W_{2c} and W_{2a} are the ideal wights of designed NNs and bounded, respectively, by two positive scalars $\bar{W}_{2cM}$ and $\bar{W}_{2aM}$, i.e., $\|W_{2c}\|\le W_{2cM}$ and $\|W_{2a}\|\le W_{2aM}$; $\theta_c(z_k)$ and $\theta_a(x_k)$ are the bounded approximation errors, i.e., $\|\theta_c(z_k)\|\le\theta_{cM}$ and $\|\theta_a(x_k)\|\le\theta_{aM}$; W_{1c} and W_{1a} are the known weight matrices of between the input layer and hidden layer; and $\phi_c(\cdot)$, $\phi_{1a}(\cdot)$ and $\phi_{2a}(\cdot)$ are the activation functions satisfying $\|\phi_c(\cdot)\|\le\phi_{c,m}$, $\|\phi_{1a}(\cdot)\|\le\phi_{1a,m}$ and $\|\phi_{2a}(\cdot)\|\le\phi_{2a,m}$.

In order to identify the ideal weight W_{2c}, the following approximation is developed by virtue of the ADP algorithm

$$\hat{J}_s(x_k)=\hat{W}_{2c,s}^T\phi_c(W_{1c}^T z_{s,k}) \tag{44}$$

where $z_{s,k}=[x_k^T\ u_{s,k}^T]^T$.

Taking the above equation into (36), there is usually

$$\hat{J}_s(x_k)\ne\Gamma^T L(x_k,u_{s-1}(x_k))+\hat{J}_{s-1}(x_{k+1})$$

that is

$$\hat{W}_{2c,s}^T\phi(W_{1c}^T z_{s,k})\ne\Gamma^T L(x_k,u_{s-1}(x_k))+\hat{J}_{s-1}(x_{k+1}). \tag{45}$$

Introduce the gap

$$
\begin{aligned}
\Delta J_s(x_k)&=\hat{W}_{2c,s}^T\phi_c(W_{1c}^T z_{s,k})-\hat{J}_{s-1}(x_{k+1})\\
&\quad-\Gamma^T L(x_k,u_{s-1}(x_k))
\end{aligned} \tag{46}
$$

and then define the cost function

$$e_{c,s}=\frac{1}{2}\Delta J_s^2(x_k).$$

Minimizing such a function results in the updating rule of the weights of the critic network

$$
\begin{aligned}
\hat{W}_{2c,s+1}&=\hat{W}_{2c,s}-\varepsilon_c\frac{\partial e_{c,s}}{\partial\hat{W}_{2c,s}}\\
&=\hat{W}_{2c,s}-\varepsilon_c\frac{\partial e_{c,s}(k)}{\partial\tilde{J}_s(x_k)}\frac{\partial(\Delta J_s(x_k))}{\partial\hat{W}_{2c,s}}\\
&=\hat{W}_{2c,s}-\varepsilon_c\phi_c(W_{1c}^T z_{s,k})\Delta J_s^T(x_k)
\end{aligned} \tag{47}
$$

where $\varepsilon_c>0$ is the learning rate of the critic network. The weights of the model network are kept unchanged after finished the training process.

In the actor network, x_k is used as the input, while the control input is approximated by

$$\hat{u}_s(x_k)=\phi_{2a}\Big(\hat{W}_{2a,s}^T\phi_{1a}(W_{1a}^T x_k)\Big). \tag{48}$$

On the other hand, it follows from (34) that the approximated value is also obtained by

$$u_s(x_k)=-\bar{u}\tanh\Big(\frac{1}{2\bar{u}}R^{-1}g_i^T(x_k)\nabla\hat{J}_s^T(x_{k+1})\Big). \tag{49}$$

Denote $\Upsilon_i=(1/2\bar{u})R^{-1}\tilde{g}_i^T(x_k)$ and then introduce the gap

$$
\begin{aligned}
\Delta u_s(x_k)&=\hat{u}_s(x_k)-u_s(x_k)\\
&=\phi_{2a}\Big(\hat{W}_{2a,s}^T\phi_{1a}(W_{1a}^T x_k)\Big)\\
&\quad+\bar{u}\tanh(\Upsilon_i\nabla\phi_c^T(W_{1c}^T z_{s,k+1})\hat{W}_{2c,s}^T).
\end{aligned} \tag{50}
$$

In what follows, define the cost of this gap

$$e_{a,s}=\frac{1}{2}\Delta u_s^T(x_k)\Delta u_s(x_k).$$

By employing the gradient descent approach again to minimize $e_{a,s}$, one has the updating rule of the weights of the actor network

$$\begin{aligned}\hat{W}_{2a,s+1} &= \hat{W}_{2a,s} - \varepsilon_a \frac{\partial e_{a,s}(k)}{\partial \hat{W}_{2a,s}} \\ &= \hat{W}_{2a,s} - \varepsilon_a \frac{\partial e_{a,s}(k)}{\partial \Delta u_s(x_k)} \frac{\partial \Delta u_s(x_k)}{\partial \hat{u}_s(x_k)} \frac{\partial \hat{u}_s(x_k)}{\partial \hat{W}_{2a,s}} \\ &= \hat{W}_{2a,s} - \frac{1}{2}\varepsilon_a \phi_{1a}(W_{1a}^T x_k) \\ &\quad \times \Big(1 - \phi_{2a}^T(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k)) \\ &\quad \times \phi_{2a}(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k))\Big) \Delta u_s^T(x_k) \end{aligned} \tag{51}$$

where $\varepsilon_a > 0$ is the learning rate of the action network.

Defining the estimation errors of weight matrices

$$\tilde{W}_{2c,s} = \hat{W}_{2c,s} - W_{2c},\ \tilde{W}_{2a,s} = \hat{W}_{2a,s} - W_{2a}$$

one has

$$\begin{aligned}\tilde{W}_{2c,s+1} &= \tilde{W}_{2c,s} - \varepsilon_c \phi_c(W_{1c}^T z_{s,k}) \Delta J_s^T(x_k) \\ &= \tilde{W}_{2c,s} - \varepsilon_c \phi_c(W_{1c}^T z_{s,k}) \Big(\hat{W}_{2c,s}^T \phi_c(W_{1c}^T z_{s,k}) \\ &\quad - \hat{J}_{s-1}(x_{k+1}) - \Gamma^T L(x_k, u_{s-1}(x_k))\Big)^T \\ &= \tilde{W}_{2c,s} - \varepsilon_c \phi_c(W_{1c}^T z_{s,k}) \Big(\tilde{W}_{2c,s}^T \phi_c(W_{1c}^T z_{s,k}) \\ &\quad + W_{2c}^T \phi_c(W_{1c}^T z_{s,k}) - \hat{W}_{2c,s-1}^T \phi_c(W_{1c}^T z_{s,k+1}) \\ &\quad - \Gamma^T L(x_k, u_{s-1}(x_k))\Big)^T \end{aligned} \tag{52}$$

and

$$\begin{aligned}\tilde{W}_{2a,s+1} &= \tilde{W}_{2a,s} - \frac{1}{2}\varepsilon_a \phi_{1a}(W_{1a}^T x_k) \\ &\quad \times \Big(1 - \phi_{2a}^T(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k)) \\ &\quad \times \phi_{2a}(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k))\Big) \Delta u_s^T(x_k) \\ &= \tilde{W}_{2a,s} - \frac{1}{2}\varepsilon_a \phi_{1a}(W_{1a}^T x_k) \\ &\quad \times \Big(1 - \phi_{2a}^T(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k)) \\ &\quad \times \phi_{2a}(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k))\Big) \\ &\quad \times \Big(\phi_{2a}(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k)) \\ &\quad + \bar{u} \tanh(\Upsilon \nabla \phi_c^T(W_{1c}^T z_{s,k+1}) \hat{W}_{2c,s}^T)\Big). \end{aligned} \tag{53}$$

D. Performance Analysis

It is easily seen that the estimation errors of weights in actor and critic networks will inevitably affect the performance of the above ADP algorithm. Thus, it is necessary to prove the boundedness of the critic and actor NN weights.

Theorem 3: Consider the discrete-time Markov jump system (MJS) (5), the critic NN (44) and the actor NN (48). Then, for the fixed time k, the weight estimation error $\tilde{W}_{c,s}$ in (52) of the critic NN and the weight estimation error $\tilde{W}_{a,s}$ in (50) of the actor NN are all UUB, if the following conditions for the learning rates are satisfied

$$0 < \varepsilon_c \le \phi_{c,m}^{-2},\ \ 0 < \varepsilon_a \le \phi_{1a,m}^{-2}. \tag{54}$$

Proof: In order to show the boundedness, we introduce a Lyapunov function candidate

$$\begin{aligned} L_{\tilde{W}_s} &= L_{\tilde{W}_{2c,s}} + L_{\tilde{W}_{2a},s} \\ &= \frac{1}{\alpha_c} \mathrm{tr}\{\tilde{W}_{2c,s}^T \tilde{W}_{2c,s}\} + \frac{1}{\alpha_a} \mathrm{tr}\{\tilde{W}_{2a,s}^T \tilde{W}_{2a,s}\}. \end{aligned}$$

In what follows, the proof is similar to the one in literature [42], and therefore its details are omitted, and the corresponding learning rates need to satisfy

$$\begin{aligned} 0 < \varepsilon_c &\le \frac{1}{\| \phi_c(W_{1c}^T z_k) \|^2} \\ 0 < \varepsilon_a &\le \frac{\| \phi_{1a}(W_{1a}^T x_k) \|^{-2}}{1 - \| \phi_{2a}(\hat{W}_{2a,s}^T \phi_{1a}(W_{1a}^T x_k)) \|^2}. \end{aligned}$$

Since the excitation functions $\phi_c(\cdot)$, $\phi_{1a}(\cdot)$ and $\phi_{2a}(\cdot)$ are bounded, the ideal learning rates can be obtained. ■

Assumption 3: The function $\|g(x_k)\|$ in (1) is bounded, and therefore the function $\|g_i(x_k)\|$ is also bounded.

Theorem 4: Let the initial control input be admissible and the initial actor-NN and critic-NN weights be selected from a compact set which includes the ideal weights. The NN weight updating laws (47) and (51) are adopted in an off-line way for the critic network (44) and the actor network (48), and the updating law (50) with (17) is employed in an online way for the identifier (12). Then, the closed-loop system (5) (or (10)) with control law (48) selecting $\hat{u}(x_k) = \phi_{2a}(\hat{W}_{2a,\infty}^T \phi_{1a}(W_{1a}^T x_k))$ is ultimately bounded in mean-square sense if all conditions in Theorems 1 and 3 hold.

Proof: In the framework of identifier-based control, taking the control policy (48) into account, the actual closed-loop system as follows:

$$\begin{aligned} x_{k+1} &= F_{\xi_k}(x_k, \hat{u}_\infty(x_k)) \\ &= F_{\xi_k}(x_k, u^*(x_k)) + g_{\xi_k}(x_k)(\hat{u}_\infty(x_k) - u^*(x_k)) \\ &= F_{\xi_k}(x_k, u^*(x_k)) + g_{\xi_k}(x_k)(\phi_{2a}(\tilde{W}_{2a}^T \psi_k) - \theta_a(x_k)) \end{aligned}$$

where

$$\begin{aligned} \psi_k &= \phi_{1a}(W_{1a}^T x_k) \\ \phi_{2a}(\tilde{W}_{2a}^T \psi_k) &= \phi_{2a}(\hat{W}_{2a,\infty}^T \psi_k) - \phi_{2a}(W_{2a}^T \psi_k). \end{aligned}$$

Obviously, considering the property of activation functions of NNs, one has that $\|\phi_{2a}(\tilde{W}_{2a}^T \psi_k) - \theta_a(x_k)\|$ is bounded. Furthermore, benefiting from Assumption 3, the additional term $g_{\xi_k}(x_k)(\phi_{2a}(\tilde{W}_{2a}^T \psi_k) - \theta_a(x_k))$ is also bounded.

On the other hand, according to the optimal control theory, the policy (43) stabilizes the system (11) (i.e., (10)) on the compact set. With the same approach in [37], it is clear that there exists a constant H^* such that

$$E \| \sum_{j=1}^{N} p_{ij} F_i(x_k, u^*(x_k)) \|^2 \le H^* E \| x_k \|^2 . \tag{55}$$

By virtue of the input-to-state stability or the similar line in [37], one can conclude that the actual closed-loop system is ultimately bounded in mean-square sense. ■

Remark 6: In the above subsections, a set of critic and actor networks are designed to approximate the performance index function sequence $J_s(x_k)$ and the control law sequence $u_s(x_k)$ for the fixed x_k, where the updating rules of NN weights are

derived via the gradient descent. By means of the well-known Lyapunov stability theory, we obtain the conditions on learning rates of neural networks, under which both the weight error dynamics and the closed-loop systems are bounded stable.

Remark 7: In almost all ADP-based suboptimal control issues for the nonlinear systems, NNs are widely utilized to approximate the unknown nonlinear dynamics as well as the actor and critic functions. Such a structure, named the actor-critic structure, provides the capability of forwarding calculation while avoiding the dimensional disaster. Inspired by the idea in [43], a tuning parameter q_k has been employed in the identification of unknown nonlinear systems to adjust the approximate error ε_k. Furthermore, the three-layer feed forward NNs have been adopted to approximate actor and critic functions where the approximation capability is enhanced due to the utilization of a hidden layer.

Remark 8: Up to date, two typical iteration strategies of ADP algorithms are utilized to obtain the desired controller parameter and the associate utility, and they are policy iteration (PI) and value iteration (VI), respectively. One major difference between PI and VI strategies is that PI requires an initial admissible control policy that stabilizes the system states [44]. From a mathematical point of view, the initial admissible control can be regarded as a suboptimal control which requires to solve the nonlinear partial differential equations (PDEs) analytically. To overcome the shortage, a VI-based strategy has been developed in this paper to definitely deal with the control issue with input saturation and communication scheduling.

IV. Illustrative Example

In this section, we use a simulation example to show that the proposed suboptimal control is effective for discrete-time nonlinear systems with input saturation under SCPs.

Consider the following nonlinear system:

$$x_{k+1} = \begin{bmatrix} -0.5x_{1,k} + 0.1x_{2,k} \\ 0.1\sin(x_{1,k})\exp(|x_{2,k}|) + 1.2x_{2,k} \end{bmatrix} + \begin{bmatrix} \bar{u}_{1,k} \\ \bar{u}_{2,k} \end{bmatrix}$$

where $x_{i,k}$ $(i = 1,2)$ stands for the i-th element of vector x_k with the initial value $x_0 = [-0.5, -0.2]^T$, and $\bar{u}_{i,k}$ $(i = 1,2)$ is the actuator input scheduled by SCPs, where the scheduling probabilities are $p_{11} = 0.65$ and $p_{22} = 0.6$. Its initial value is $u_0 = [-0.1, 0.5]^T$ and the saturation level $\bar{u}$ is 5.

In this example, choose three-layer feedforward NNs in the identifier, the critic network and the action network with structures $4-4-2$, $4-2-1$, and $2-2-2$, respectively. Furthermore, select the activation functions as follows

$$\phi_x(*) = \frac{2(e^* - e^{-*})}{e^* + e^{-*}}$$
$$\phi_c(*) = \phi_{1a}(*) = \frac{e^* - e^{-*}}{e^* + e^{-*}}$$
$$\phi_{2a}(*) = \frac{\bar{u}(e^* - e^{-*})}{e^* + e^{-*}}.$$

Thus, the bounds of ϕ_x, ϕ_c, ϕ_{1a} and ϕ_{2a} are $\phi_{x,m} = 2$, $\phi_{c,m} = \phi_{1a,m} = 1$ and $\phi_{2a} = \bar{u}$, respectively.

In virtue of Theorem 1, we can employ the learning rate $\gamma_v = \gamma_w = 0.1$, and parameters $\alpha_v = 0.9$, $c_2 = 1$, and $v_0 = 0.1$ in the tuning law $\hat{W}_{2,i}^k$ and the additional tunable parameter v_k. Furthermore, the weight matrix W_1 between the input and hidden layers is adopted $1.2I$ and the initial weight matrices $\hat{W}_{2,i}^0$ $(i = 1,2)$ in (12) between the hidden layer and output layer are selected as

$$\hat{W}_{2,i}^0 = \begin{bmatrix} -0.50, & 0.1, & 1.0, & 0 \\ 0.02, & 1.2, & 0, & 1.0 \end{bmatrix}^T, \quad i = 1,2.$$

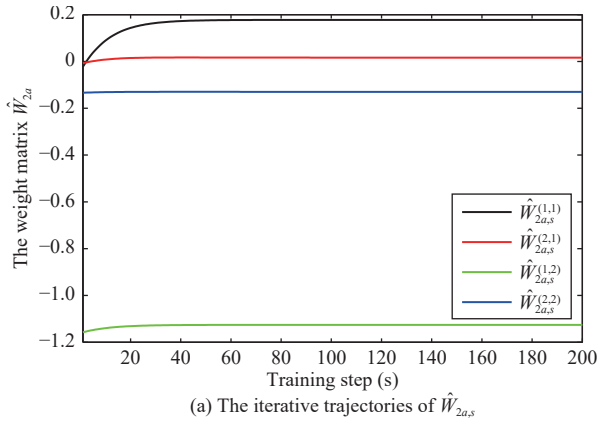
In what follows, we consider the matrices $Q_1 = Q_2 = 0.5I$ and $R_1 = R_2 = 0.2I$ in the cost function (6) and the corresponding scalars $\lambda_1 = 0.4$ and $\lambda_2 = 0.6$ in the combined performance index (8). Furthermore, the parameters in the updating rules (47) and (51) are chosen as $\varepsilon_a = \varepsilon_c = 0.4$ and for the adopted critic-actor network with the help of Theorem 3, the initial weight matrices of this critic-actor network are selected as

$$\hat{W}_{2c,0} = \begin{bmatrix} 1.00, & 1.05 \end{bmatrix}^T, \hat{W}_{2a,0} = \begin{bmatrix} -0.04, & -1.16 \\ -0.01, & -0.134 \end{bmatrix}^T.$$

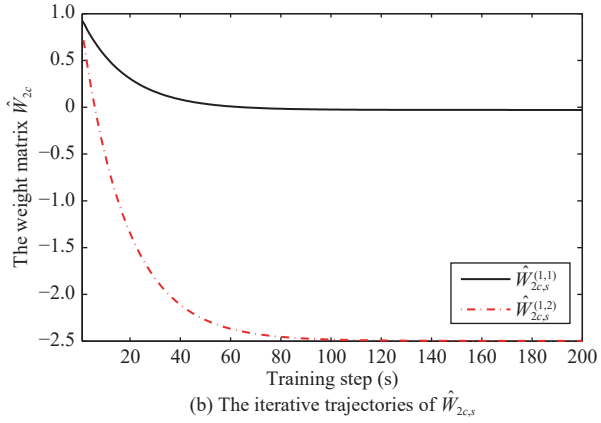
In addition, the weight matrices W_{1a} and W_{1c} between the input and hidden layers are

$$W_{1a} = 0.2I, \ W_{1c} = \begin{bmatrix} 2, & 0, & 0.01, & 0 \\ 0, & 2.5, & 0, & 0.01 \end{bmatrix}^T.$$

Training of weight matrices for critic-actor networks is performed in instant $k = 4$ with 200 steps. After being trained, the weights are kept unchanged. The training process is shown in Fig. 2 and their trajectories are convergent, which verifies the effectiveness of developed ADP scheme.



(a) The iterative trajectories of $\hat{W}_{2a,s}$



(b) The iterative trajectories of $\hat{W}_{2c,s}$

Fig. 2. The iterative trajectories of the weight matrices $\hat{W}_{2a,s}$ and $\hat{W}_{2c,s}$.

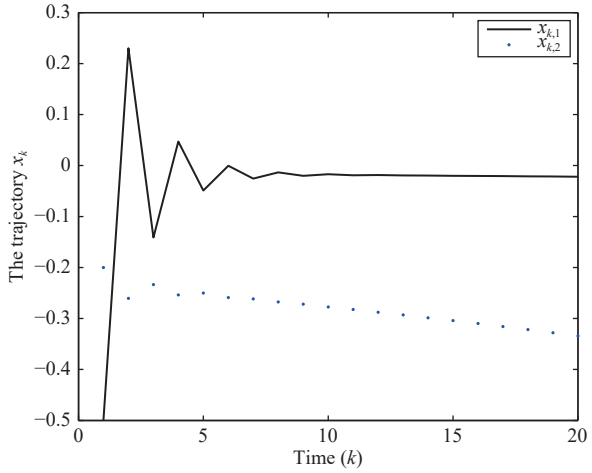


Fig. 3. State trajectories x_k of the open-loop system.

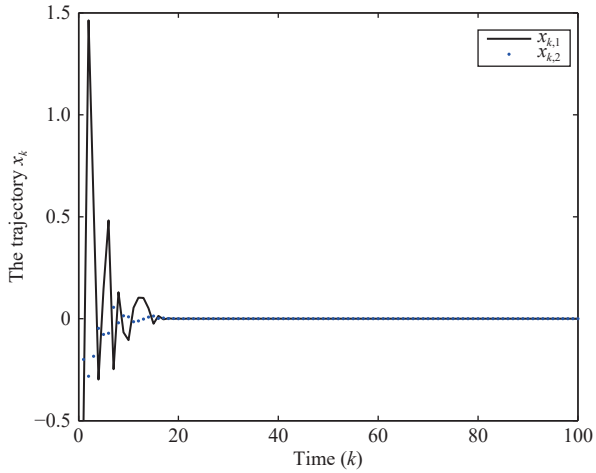


Fig. 4. State trajectories x_k of the closed-loop system.

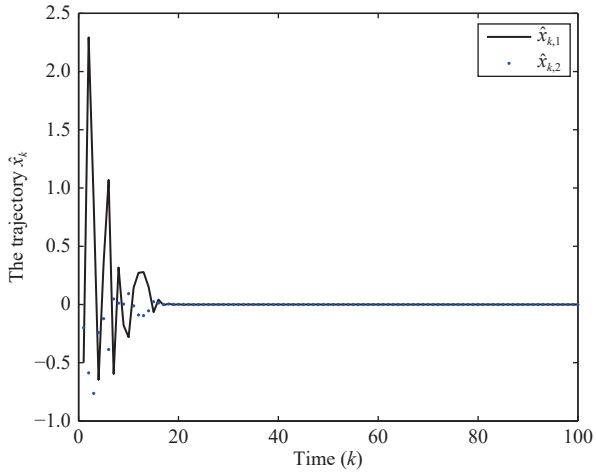


Fig. 5. State trajectories $\hat{x}_k$ of the identifer.

The simulation results are presented in Figs. 3–7. The state trajectories x_k of the open-loop system are depicted in Fig. 3 to reveal that the open-loop system is divergent. With the help of the designed controller, the state trajectories x_k of the closed-loop system and the corresponding trajectories $\hat{x}_k$ of

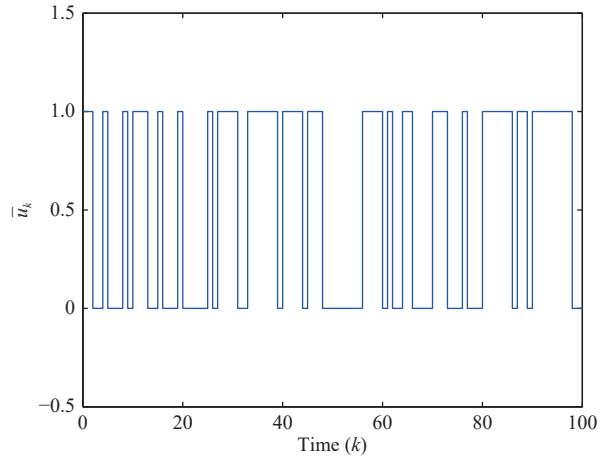


Fig. 6. The scheduling of actuator units.

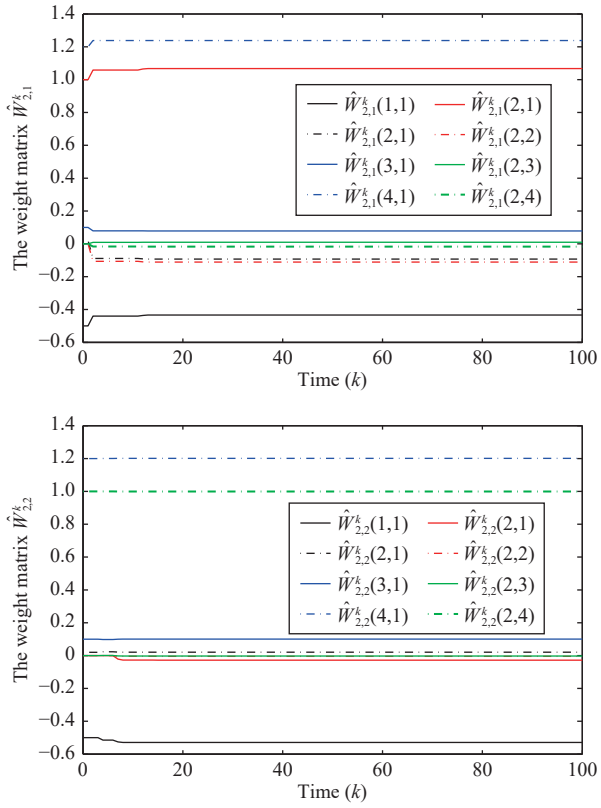


Fig. 7. The iterative trajectories of the weight matrices $\hat{W}^k_{2,i}$.

the identifer are respectively shown in Figs. 4 and 5. For this control issue, the secluded node is shown in Fig. 6, which clearly discloses that the system randomly jumps due to the utilization of different actuator units, and the weight matrices of the identifier $\hat{W}^k_{2i}$ are plotted in Fig. 7, all of which are eventually convergent. It is not difficult to see that the closed-loop system is stable, and therefore the developed control strategy is effective.

V. Conclusions

In this paper, we have developed a suboptimal control strategy in the framework of ADP for a class of unknown

nonlinear discrete-time systems subject to input constraints. An identification with robust term based on a three-layer neural network in which the weight update relies on protocol-induced jumps, has been established to approximate nonlinear systems and the corresponding stability has been provided. Then, the value iterative ADP algorithm has been developed to solve the suboptimal control problem with boundedness analysis, and the convergence of iterative algorithm, as well as the boundedness of the estimation errors for critic and actor NN weights, has been analyzed. Furthermore, an actor-critic NN scheme has been developed to approximate the control law and the proposed performance index function and the stability of the closed-loop systems have been discussed. Finally, the numerical simulation result has been utilized to demonstrate the effectiveness of the proposed control scheme.

REFERENCES

[1] M. Mazouchi, M. B. N. Sistani, and S. K. H. Sani, "A novel distributed optimal adaptive control algorithm for nonlinear multi-agent differential graphical games," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 331–341, Jan. 2018.

[2] Y. J. Liu, L. Tang, S. Tong, C. L. Chen, and D. J. Li, "Reinforcement learning design-based adaptive tracking control with less learning parameters for nonlinear discrete-time MIMO systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 1, pp. 165–176, Jan. 2015.

[3] R. Song and L. Zhu, "Optimal fixed-point tracking control for discrete-time nonlinear systems via ADP," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 3, pp. 657–666, May 2019.

[4] L. Sun, and Z. Zheng, "Disturbance-observer-based robust backstepping attitude stabilization of spacecraft under input saturation and measurement uncertainty," *IEEE Trans. Ind. Electron.*, vol. 64, no. 10, pp. 7994–8002, 2017.

[5] D. Wang, H. He, X. Zhong, and D. Liu, "Event-driven nonlinear discounted optimal regulation involving a power system application," *IEEE Trans. Ind. Electron.*, vol. 64, no. 10, pp. 8177–8186, 2017.

[6] H. Li, Y. Wu and M. Chen, "Adaptive fault-tolerant tracking control for discrete-time multi-agent systems via reinforcement learning algorithm," *IEEE Trans. Cybern.*, to be published. DOI: 10.1109/TCYB.2020.2982168.

[7] T. Wang, H. Gao, and J. Qiu, "A combined adaptive neural network and nonlinear model predictive control for multirate networked industrial process control," *IEEE Trans. Ind. Electron.*, vol. 27, no. 2, pp. 416–425, 2016.

[8] H. Zhang, Y. Luo, and D. Liu, "Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints," *IEEE Trans. Neural Netw.*, vol. 20, no. 9, pp. 1490–1503, 2009.

[9] Z. Shi and Z. Wang, "Optimal control for a class of complex singular system based on adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 1, pp. 188–197, Jan. 2019.

[10] R. Song, Q. Wei, H. Zhang, and F. L. Lewis, "Discrete-time non-zero-sum games with completely unknown dynamics," *IEEE Trans. Cybern.*, vol. 99, pp. 1–15, 2019.

[11] Q. Wei, and D. Liu, "Data-driven neuro-optimal temperature control of waterCgas shift reaction using stable iterative adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6399–6408, 2014.

[12] P. J. Werbos, "Foreword-ADP: the key direction for future research in intelligent control and understanding brain intelligence," *IEEE Trans. Syst. Man*, *Cybern. Part B*, vol. 38, pp. 898–900, 2008.

[13] D. P. Bertsekas and J. N. Tsitsiklis, "*Neuro-Dynamic Programming*," Athena Scientific, USA, Belmont, MA, 1996.

[14] R. S. Sutton and A. G. Barto, "*Reinforcement Learning: An Introduction*", Cambridge, MA, USA: MIT Press, 1998.

[15] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof," *IEEE Trans. Syst. Man*, *Cybern. Part B*, vol. 38, no. 4, pp. 943–949, 2008.

[16] A. Heydari, "Stability analysis of optimal adaptive control under value iteration using a stabilizing initial policy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4522–4527, Sept. 2018.

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

[18] W. B. Powell, "*Approximate Dynamic Programming*," IHoboken, NJ, USA: Wiley, 2007.

[19] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1007, 1997.

[20] X. Zhong, N. Zhen, and H. He, "A theoretical foundation of goal representation heuristic dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst*, vol. 27, no. 12, pp. 2513–2525, 2017.

[21] Y. Yuan, Z. Wang, P. Zhang, and H. Liu, "Near-optimal resilient control strategy design for state-saturated networked systems under stochastic communication protocol," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 1–13, 2018.

[22] 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, no. 5, pp. 779–791, 2005.

[23] X. Yang and B. Zhao, "Optimal neuro-control strategy for nonlinear systems with asymmetric input constraints," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 2, pp. 575–583, Mar. 2020.

[24] D. Liu, X. Yang, D. Wang, and Q. Wei, "Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1372–1385, Jul. 2015.

[25] Y. J. Liu, S. Li, S. Tong, and C. L. P. Chen, "Neural approximation-based adaptive control for a class of nonlinear nonstrict feedback discrete-time systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1531–1541, Jul. 2017.

[26] H. Xu, Q. Zhao, and S. Jagannathan, "Finite-horizon near-optimal output feedback neural network control of quantized nonlinear discrete-time systems with input constraint," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 8, pp. 1776–1788, Aug. 2015.

[27] Y. Zhu, D. Zhao, H. He, and J. Ji, "Event-triggered optimal control for partially unknown constrained-input systems via adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4101–4109, 2017.

[28] H. Modares, F. L. Lewis, and M. Naghibi-Sistani, "Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1513–1525, 2013.

[29] D. Ding, Q. L. Han, X. Ge, and J. Wang, "Secure state estimation and control of cyber-physical systems: A survey," *IEEE Trans. Syst. Man, Cybern.: Syst.*, to be published. DOI: 10.1109/TSMC.2020.3041121.

[30] V. Ugrinovskii and E. Fridman, "A round-robin type protocol for distributed estimation with H_∞ consensus," *Syst. Control Lett.*, vol. 69, pp. 103–110, 2014.

[31] G. Walsh, H. Ye, and L. Bushnell, "Stability analysis of networked control systems," *IEEE Trans. Control Syst. Tech.*, vol. 10, no. 3, pp. 438–446, 2002.

[32] L. Zou, Z. Wang, and H. Gao, "Observer-based H_∞ Control of networked systems with stochastic communication protocol: The finite-horizon case," *Automatica*, vol. 63, pp. 366–373, 2016.

[33] H. Ma, H. Li, R. Lu, and T. Huang, "Adaptive event-triggered control for a class of nonlinear systems with periodic disturbances," *Sci China Inf. Sci.*, vol. 63, no. 5, pp. 157–171, 2020.

[34] Z. Wang, Q. Wei, and D. Liu, "Event-triggered adaptive dynamic programming for discrete-time multi-player games," *Inf. Sci.*, vol. 506, pp. 457–470, Jan. 2020.

[35] D. Ding, Z. Wang, and Q. L. Han, "Neural-network-based consensus control for multi-agent systems with input constraints: The event-

triggered case," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 1–12, 2019.

[36] X. Zhong, H. He, H. Zhang, and Z. Wang, "Optimal control for unknown discrete-time nonlinear Markov jump systems using adaptive dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2141–2155, 2014.

[37] D. Ding, Z. Wang, and Q. L. Han, "Neural-network-based output-feedback control with stochastic communication protocols," *Automatica*, vol. 106, pp. 221–229, Aug. 2019.

[38] N. Azevedo, D. Pinheiro, and G.-W. Weber, "Dynamic programming for a Markov-switching jump-diffusion," *J Comput. Appl. Math.*, vol. 267, no. 6, pp. 1–19, Sep. 2014.

[39] M. C. F. Donkers, W. P. M. H. Heemels, and D. Bernardini, A. Bemporad, and V. Shneer, "Stability analysis of stochastic networked control systems," *Automatica*, vol. 48, no. 4, pp. 917–925, 2012.

[40] T. Dierks, B. T. Thumati, and S. Jagannathan, "Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence," *Neural Networks*, vol. 22, no. 5–6, pp. 851–860, 2009.

[41] B. Lincoln and A. Rantzer, "Relaxing dynamic programming," *IEEE Trans. Autom. Control*, vol. 51, no. 8, pp. 1249–1260, Aug. 2006.

[42] J. Song, Y. Niu, and Y. Zou, "Convergence analysis for an identifier-based adaptive dynamic programming algorithm," In *Proc. the* 34*th Chinese Control Conf.*, 2015.

[43] D. Liu, D. Wang, and X. Yang, "An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs," *Inf. Sci.*, vol. 220, no. 1, pp. 331–342, 2013.

[44] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear Systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 621–634, 2014.

Xueli Wang received the B.Sc. degree from Qufu Normal University, Qufu, China, in 2013. She is currently a Ph.D. candidate in control science and engineering at University of Shanghai for Science and Technology, Shanghai, China. Her current research interests include networked control systems, adaptive dynamic program, and neural networks. She is a very active reviewer for many international journals.

Derui Ding (M'16–SM'20) received both the B.Sc. degree in industry engineering in 2004 and the M.Sc. degree in detection technology and automation equipment in 2007 from Anhui Polytechnic University, Wuhu, China, and the Ph.D. degree in control theory and control engineering in 2014 from Donghua University, Shanghai, China. From July 2007 to December 2014, he was a Teaching Assistant and then a Lecturer in the Department of Mathematics, Anhui Polytechnic University, Wuhu, China. He is currently a Senior Research Fellow with the School of Software and Electrical Engineering, Swinburne University of Technology, Melbourne, Australia. From June 2012 to September 2012, he was a Research Assistant in the Department of Mechanical Engineering, the University of Hong Kong, Hong Kong. From March 2013 to March 2014, he was a Visiting Scholar in the Department of Information Systems and Computing, Brunel University London, UK. His research interests include nonlinear stochastic control and filtering, as well as multi-agent systems and sensor networks. He has published around 80 papers in refereed international journals. He is serving as an Associate Editor for *Neurocomputing* and *IET Control Theory & Applications*. He is also a very active reviewer for many international journals.

Hongli Dong (SM'16) received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2012. From 2009 to 2010, she was a Research Assistant with the Department of Applied Mathematics, City University of Hong Kong, Hong Kong, China. From 2010 to 2011, she was a Research Assistant with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China. From 2011 to 2012, she was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, London, UK. From 2012 to 2014, she was an Alexander von Humboldt Research Fellow with the University of DuisburgEssen, Duisburg, Germany. She is currently a Professor with the Institute of Complex Systems and Advanced Control, Northeast Petroleum University, Daqing, China. She is also the Director with the Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, China. Her current research interests include robust control and networked control systems.

Xian-Ming Zhang (M'16–SM'18) received the M.Sc. degree in applied mathematics and the Ph.D. degree in control theory and engineering from Central South University, Changsha, China, in 1992 and 2006, respectively. In 1992, he joined Central South University, where he was an Associate Professor with the School of Mathematics and Statistics. From 2007 to 2014, he was a Postdoctoral Research Fellow and a Lecturer with the School of Engineering and Technology, Central Queensland University, Rockhampton, QLD, Australia. From 2014 to 2016, he was a Lecturer with the Griffith School of Engineering, Griffith University, Gold Coast, QLD, Australia. In 2016, he joined the Swinburne University of Technology, Melbourne, VIC, Australia, where he is currently an Associate Professor with the School of Software and Electrical Engineering. His current research interests include H_∞ filtering, event-triggered control, networked control systems, neural networks, distributed systems, and time-delay systems. He was a recipient of the National Natural Science Award (Secondclass) in China in 2013, and the Hunan Provincial Natural Science Award (First-class) in Hunan Province in China in 2011, both jointly with Profs. M. Wu and Y. He. He was also a recipient of the *IEEE Transactions on Industrial Informatics* Outstanding Paper Award 2020, the Andrew P. Sage Best Transactions Paper Award 2019, and the *IET Control Theory and Applications* Premium Award 2016. He is acting as an Associate Editor of several international journals, including the *IEEE Transactions on Cybernetics*, *Neural Processing Letters*, *Journal of the Franklin Institute*, *International Journal of Control*, *Automation*, *and Systems*, *Neurocomputing*.