Data-Driven Neuro-Optimal Temperature Control of Water–Gas Shift Reaction Using Stable Iterative Adaptive Dynamic Programming

Qinglai Wei, *Member, IEEE*, and Derong Liu, *Fellow, IEEE*

***Abstract*—In this paper, a novel data-driven stable iterative adaptive dynamic programming (ADP) algorithm is developed to solve optimal temperature control problems for water–gas shift (WGS) reaction systems. According to the system data, neural networks (NNs) are used to construct the dynamics of the WGS system and solve the reference control, respectively, where the mathematical model of the WGS system is unnecessary. Considering the reconstruction errors of NNs and the disturbances of the system and control input, a new stable iterative ADP algorithm is developed to obtain the optimal control law. The convergence property is developed to guarantee that the iterative performance index function converges to a finite neighborhood of the optimal performance index function. The stability property is developed to guarantee that each of the iterative control laws can make the tracking error uniformly ultimately bounded (UUB). NNs are developed to implement the stable iterative ADP algorithm. Finally, numerical results are given to illustrate the effectiveness of the developed method.**

***Index Terms*—Adaptive critic designs, adaptive dynamic programming (ADP), approximate dynamic programming, approximation errors, data-driven control, neural networks (NNs), optimal control, reinforcement learning, water–gas shift (WGS).**

I. Introduction

A water–gas shift (WGS) reactor is an essential component in the coal-based chemical industry [1]. The WGS reactor combines carbon monoxide (CO) and water (H_2O) in the reactant stream to produce carbon dioxide (CO_2) and hydrogen (H_2). Proper regulation of the operating temperature is critical to achieving adequate CO conversion during transients [2]. Hence, optimal control of the reaction temperature is a key problem for the WGS reaction process. To describe the dynamics of the WGS reaction process, many discussions focused on WGS modeling approaches [3], [4]. Unfortunately, the established WGS models are generally complex with high nonlinearities. This makes the traditional linearized control method [5]–[7] only effective in the neighborhood of the equilibrium point. When the required operating range is large, the nonlinearities in the system cannot be properly compensated by using a linear model. Therefore, it is necessary to study an optimal control approach for the original nonlinear system [1], [2]. Although optimal control of nonlinear systems has been the focus of the control field in the latest several decades [8]–[16], the optimal controller design for WGS reaction systems (WGS systems in brief) is still challenging, due to the complexity of the WGS reaction process.

Based on function approximators, such as neural networks (NNs), adaptive dynamic programming (ADP), which was proposed by Werbos [17], [18], has played an important role as a way to solve optimal control problems forward in time [19]–[21]. Iterative methods, which are mainly based on policy and value iterations [22], are widely used in ADP to obtain the solution of Hamilton–Jacobi–Bellman (HJB) equation indirectly and have received much attention [23]–[29]. For most previous ADP algorithms, it requires that the system model, the iterative control, and the performance index function can accurately be approximated, which guarantees the convergence property of the proposed algorithms. In real-world implementation of ADP, e.g., for WGS systems, the reconstruction errors by approximators and the disturbances of system states and controls inherently exist. These make the accurate system models, iterative control laws, and performance index functions impossible to obtain accurately. Although in [30] and [31], ADP was explored to design the optimal temperature controller of the WGS system, the effects of approximation errors and disturbances were not considered. Furthermore, the convergence and stability properties were not discussed.

In this paper, for the first time, a new stable iterative ADP algorithm is developed to obtain the optimal control law for the WGS systems, which makes the temperature of the WGS systems track the desired temperature. By employing NNs, the dynamics of the WGS systems and the reference control are established. Via system transformation, the optimal tracking problem is effectively transformed into an optimal regulation problem. Considering the reconstruction errors of NNs and the disturbances, a new stable iterative ADP algorithm is developed to obtain the optimal control law iteratively. We emphasize that the convergence and stability analysis is established to guarantee that the performance index function is convergent to a finite neighborhood of the optimal performance index function and that each of the iterative control laws makes the tracking error

Manuscript received August 31, 2013; revised December 2, 2013; accepted December 25, 2013. Date of publication January 21, 2014; date of current version June 6, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 61034002, Grant 61374105, Grant 61233001, and Grant 61273140, in part by the Beijing Natural Science Foundation under Grant 4132078, and by the Early Career Development Award of The State Key Laboratory of Management Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences.

The authors are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: qinglai.wei@ia.ac.cn; derong.liu@ia.ac.cn).

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/TIE.2014.2301770

uniformly ultimately bounded (UUB). NN implementation of the stable iterative ADP algorithm is presented, and the convergence property of the weights is analyzed. Finally, numerical results and comparisons with a traditional data-driven method are given to show the effectiveness of the developed iterative ADP algorithm.

II. Preliminaries

A. WGS Reaction

The WGS reaction inputs the water gas, which includes CO, CO_2, H_2, and H_2O, into the WGS reactor. The WGS reaction, which is slightly exothermic, converts CO to CO_2 and H_2, as shown in the following:

$$\mathrm{CO} + \mathrm{H_2O} \leftrightarrow \mathrm{CO_2} + \mathrm{H_2} - 41.9\ \mathrm{kJ/mol}. \tag{1}$$

The WGS reaction rate [2] can be described as follows:

$$r_{\mathrm{WGS}} = \rho_{\mathrm{cat}} k_r \exp\left(-\frac{5126\ \mathrm{K}}{T}\right)[\mathrm{CO}]^{0.78}[\mathrm{H_2O}]^{0.15} \times \left(1 - \frac{[\mathrm{CO_2}][\mathrm{H_2}]}{[\mathrm{CO}][\mathrm{H_2O}]K_T}\right) \tag{2}$$

where the rate is in (kmol/m^3/s). The catalyst density $\rho_{\mathrm{cat}} = 1.8 \times 10^{-4}$ kg/m^3. The rate constant is $k_r = 1.32 \times 10^9$ kmol/kg/s. The reaction equilibrium coefficient K_T is given as in [32], which is expressed by

$$K_T = \exp\left(\frac{4577.7\ \mathrm{K}}{T} - 4.33\right).$$

For the WGS reaction (2), we can see that the reaction temperature is a key parameter [1], [2].

Let $P(u(k)) = [\theta_{\mathrm{CO}}(k), \theta_{\mathrm{CO_2}}(k), \theta_{\mathrm{H_2}}(k), \theta_{\mathrm{H_2O}}(k)]^T u(k)$ denote the control input (m^3/s), where θ_{CO}, $\theta_{\mathrm{CO_2}}$, $\theta_{\mathrm{H_2}}$, and $\theta_{\mathrm{H_2O}}$ denote the given percentage compositions of CO, CO_2, H_2, and H_2O, respectively. Generally, the water gas of WGS systems comes from the previous reaction process, such as coal gasification [4]. This means that the composition ratio of the mixed gas is uncontrollable for the WGS systems. Let $x(k)$ denote the temperature of the WGS reactor; then, the WGS system can be expressed as

$$x(k+1) = F\left(x(k), u(k)\right) \tag{3}$$

where $F(\cdot)$ is an unknown system function. Let $x(k) \in \mathbb{R}^n$ and $u(k) \in \mathbb{R}^m$, where $n = 1$ and $m = 1$. Let the desired state be τ. Then, our goal is to design an optimal state-feedback tracking control law $u^*(k) = u^*(x(k))$, which makes the system state track the desired state trajectory.

B. Data-Based Modeling and Properties

Three-layer back-propagation (BP) NNs are introduced to construct the dynamics of the WGS system and to solve the reference control, respectively. Let $\mathcal{L}$ be the number of hidden layer neurons. Let $X \in \mathbb{R}^{\mathcal{N}}$ be the input of NN, and let $Z \in \mathbb{R}^{\mathcal{M}}$ be the output, Then, the function of the BP NNs can be expressed by

$$Z = \hat{F}_N(X, Y, W) = W^T \sigma(YX)$$

where $Y \in \mathbb{R}^{\mathcal{L}\times\mathcal{N}}$ is the input-hidden layer weight matrix, and $W \in \mathbb{R}^{\mathcal{L}\times\mathcal{M}}$ is the hidden-output layer weight matrix. Let σ be a sigmoid activation function [33], [34]. For convenience of analysis, only the hidden-output weight W is updated during the NN training, whereas the input-hidden weight is fixed [23]. Hence, in the following, the NN function is simplified by the expression $\hat{F}_N(X, W) = W^T \sigma_N(X)$, where $\sigma_N(X) = \sigma(YX)$.

Let the number of hidden layer neurons be L_m. Then, the model NN of system (3) can be written as

$$x(k+1) = W_m^{*T} \sigma_m\left(z(k)\right) + \varepsilon_{m1}(k)$$

where $z(k) = [x^T(k), u^T(k)]^T$ and W_m^* denote the NN input and the ideal weight matrix of the model NN, respectively. Let $\sigma_m(z(k)) = \sigma(Y_m z(k))$, where Y_m is an arbitrary weight matrix with a suitable dimension. Let $\|\sigma_m(\cdot)\| \le \sigma_M$ for a constant σ_M, and let $\varepsilon_{m1}(k)$ be the bounded NN reconstruction error, which satisfies $\|\varepsilon_{m1}(k)\| \le \varepsilon_M$ for a constant ε_M. To train the model NN, it requires an array of WGS systems and control data, such as the data from a period of time. The NN model for the system is constructed as

$$\hat{x}(k+1) = \hat{W}_m^T(k) \sigma_m\left(z(k)\right) \tag{4}$$

where $\hat{x}(k)$ is the estimated system-state vector. Let $\hat{W}_m(k)$ be the estimation of the ideal weight matrix W_m^*. Then, we define the system identification error as

$$\tilde{x}(k+1) = \hat{x}(k+1) - x(k+1) = \tilde{W}_m^T(k)\sigma_m\left(z(k)\right) - \varepsilon_{m1}(k)$$

where $\tilde{W}_m(k) = \hat{W}_m(k) - W_m^*$. Let $\phi_m(k) = \tilde{W}_m^T(k)\sigma_m(z(k))$; then, we can obtain

$$\tilde{x}(k+1) = \phi_m(k) - \varepsilon_{m1}(k).$$

The weights are adjusted to minimize the following error function:

$$E_m(k) = \frac{1}{2}\tilde{x}^T(k+1)\tilde{x}(k+1).$$

By a gradient-based adaptation rule [33], [34], the weights are updated as

$$\hat{W}_m(k+1) = \hat{W}_m(k) - l_m \sigma_m\left(z(k)\right)\tilde{x}^T(k+1) \tag{5}$$

where $l_m > 0$ is the learning rate.

Theorem 2.1: Let the model network (4) be used to identify WGS system (3). If there exists a constant $0 < \lambda_m < 1$ that satisfies $\varepsilon_{m1}^T(k)\varepsilon_{m1}(k) \le \lambda_m \tilde{x}^T(k)\tilde{x}(k)$, then the system identification error $\tilde{x}(k)$ is asymptotically stable and the error matrix $\tilde{W}_m(k)$ converges to zero, as $k \to \infty$.

Proof: Consider the following Lyapunov function candidate defined as

$$L\left(\tilde{x}(k), \tilde{W}_m(k)\right) = \tilde{x}^T(k)\tilde{x}(k) + \frac{1}{l_m}\mathrm{tr}\left\{\tilde{W}_m^T(k)\tilde{W}_m(k)\right\}.$$

By taking the difference of the Lyapunov function candidate, we can obtain

$$\begin{aligned}
&\Delta L\left(\tilde{x}(k), \tilde{W}_m(k)\right) \\
&\leq \phi_m^T(k)\phi_m(k) + \varepsilon_{m1}^T(k)\varepsilon_{m1}(k) - \tilde{x}^T(k)\tilde{x}(k) \\
&\quad + 2l_m\sigma_m^T\left(z(k)\right)\sigma_m\left(z(k)\right)\left(\phi_m^T(k)\phi_m(k) + \varepsilon_m^T(k)\varepsilon_m(k)\right) \\
&\leq -\left(1 - 2l_m\sigma_M^2\right)\|\phi_m(k)\|^2 \\
&\quad - \left(1 - \lambda_m\left(1 + 2l_m\sigma_M^2\right)\right)\|\tilde{x}(k)|^2.
\end{aligned}$$

By selecting the learning rate

$$l_m < \min\left\{\frac{1}{2\sigma_M^2}, \frac{1-\lambda_m}{2\lambda_m\sigma_M^2}\right\}$$

we can obtain $\Delta L(\tilde{x}(k),\, \tilde{W}_m(k)) \leq 0$. The proof is completed. ■

Next, we will solve the reference control by NN (u_f network in brief). According to the state equation in (3), we give $x(k)$ and $x(k+1)$ to approximate the reference control function $u_f(k)$, which is expressed as $u_f(k) = F_u(x(k),\, x(k+1))$. We notice that solving $u_f(k)$ needs the data of $x(k+1)$. Hence, it requires to adopt offline or history data to train the u_f network. Let the number of hidden layer neurons be L_u. Let W_u^* be the ideal weight matrix. The NN representation of the u_f network can be written as

$$u_f(k) = W_u^{*T}\sigma_u\left(z_u(k)\right) + \varepsilon_{u1}(k)$$

where $z_u(k) = [x^T(k),\, x^T(k+1)]^T$, and $\varepsilon_{u1}(k)$ is the NN reconstruction error, which satisfies $\|\varepsilon_{u1}(k)\| \leq \varepsilon_{\bar{u}1}$ for a constant $\varepsilon_{\bar{u}1}$. Let $\sigma_u(z_u(k)) = \sigma(Y_u z_u(k))$, where Y_u is an arbitrary weight matrix with a suitable dimension.

The NN reference control is constructed as

$$\hat{u}_f(k) = \hat{F}_u\left(x(k),\, x(k+1)\right) = \hat{W}_u^T(k)\sigma_u\left(z_u(k)\right)$$

where $\hat{u}_f(k)$ is the estimated reference control, and $\hat{W}_u^T(k)$ is the estimated weight matrix. Define the identification error as

$$\tilde{u}_f(k) = \hat{u}_f(k) - u_f(k) = \phi_u(k) - \varepsilon_{u1}(k)$$

where $\phi_u(k) = \tilde{W}_u^T(k)\sigma_u(z_u(k))$, and $\tilde{W}_u(k) = \hat{W}_u(k) - W_u^*$. The weight of the u_f network is adjusted to minimize the error function, i.e.,

$$E_u(k) = \frac{1}{2}\tilde{u}_f^T(k)\tilde{u}_f(k).$$

By gradient-based adaptation rule, the weight is updated as

$$\hat{W}_u(k+1) = \hat{W}_u(k) - l_u\sigma_u\left(z_u(k)\right)\tilde{u}_f^T(k) \tag{6}$$

where $l_u > 0$ is the learning rate.

Theorem 2.2: Let the NN weight of the u_f network be updated by (6). If there exists a constant $0 < \lambda_u < 1$ that satisfies $\phi_u^T(k)\varepsilon_{u1}(k) \leq \lambda_u\phi_u^T(k)\phi_u(k)$, then the error matrix $\tilde{W}_u(k)$ asymptotically converges to zero, as $k \to \infty$.

III. Design of the Neuro-Optimal Temperature Controller

Here, a stable iterative ADP algorithm will be developed to obtain the optimal control law that makes the temperature of the WGS system track the desired one with convergence and stability analysis.

A. System Transformation

For WGS system (3), if we let the desired state be τ, then we can define the tracking error $e(k)$ as

$$e(k) = x(k) - \tau. \tag{7}$$

Let $u_d(k)$ be the corresponding desired reference control (desired control in brief) for the desired state τ. As the system function is unknown, the desired control $u_d(k)$ cannot directly be obtained by WGS system (3). On the other hand, in the real-world WGS systems, the disturbances of the system and control input are both unavoidable. These make the system transformation method with accurate system model [35] difficult to implement. To overcome these difficulties, a system transformation with NN reconstruction errors and disturbances are presented. First, according to the desired state τ, we can obtain $u_d(k) = F_u(\tau,\, \tau)$. Let $\hat{u}_d(k) = \hat{F}_u(\tau,\, \tau) = \hat{W}_u^T(k)\sigma_u(\tau,\, \tau)$ be the output of the u_f network. Let $\varepsilon_{u2}(k)$ be an unknown bounded control disturbance, which satisfies $\|\varepsilon_{u2}(k)\| \leq \varepsilon_{u2}$ for a constant ε_{u2}. Then, we can define the control error $u_e(k)$ as

$$u_e(k) = u(k) - \hat{u}_d(k) - \varepsilon_u(k) \tag{8}$$

where $\varepsilon_u(k) = \varepsilon_{u1}(k) + \varepsilon_{u2}(k)$. As $\varepsilon_{u1}(k)$ and $\varepsilon_{u2}(k)$ are bounded, there exists a constant $\varepsilon_u \geq 0$ that satisfies $\|\varepsilon_u(k)\| \leq \varepsilon_u$. On the other hand, let $\hat{F}(z(k)) = \hat{W}_m^T(k)\sigma_m(z(k))$ be the model NN function. Let $\varepsilon_{m2}(k)$ be an unknown bounded system disturbance, which satisfies $\|\varepsilon_{m2}(k)\| \leq \varepsilon_{m2}$ for a constant ε_{m2}. Then, the tracking error system $e(k+1)$ can be defined as

$$\begin{aligned}
e(k+1) &= \bar{F}\left(e(k), u_e(k)\right) \\
&= \hat{F}\left(\left(e(k)+\tau\right), \left(u_e(k) + \hat{u}_d(k)\right)\right) - \tau \\
&\quad + \nabla\hat{F}(\xi_u)\varepsilon_u + \varepsilon_m(k)
\end{aligned} \tag{9}$$

where $\nabla\hat{F}(\xi_u) = (\partial\hat{F}((e(k)+\tau), \xi_u)/\partial\xi_u)$, where $\xi_u = c_u(u_e(k) + \hat{u}_d(k)) + (1-c_u)(u_e(k) + \hat{u}_d(k) + \varepsilon_u(k))$ and $0 \leq c_u \leq 1$. Let $\varepsilon_m(k) = \varepsilon_{m1}(k) + \varepsilon_{m2}(k)$ and we have $\|\varepsilon_m(k)\| \leq \varepsilon_m$ for a constant ε_m. Let the NN tracking error $\hat{e}(k+1)$ be expressed as

$$\begin{aligned}
\hat{e}(k+1) &= F_e\left(e(k), \hat{u}_e(k)\right) \\
&= \hat{F}\left(\left(e(k)+\tau\right), \left(u_e(k) + \hat{u}_d(k)\right)\right) - \tau.
\end{aligned} \tag{10}$$

Then, we can get $e(k+1) = \hat{e}(k+1) + \varepsilon_e(k)$, where we define $\varepsilon_e(k) = \nabla\hat{F}(\xi_u)\varepsilon_u + \varepsilon_m(k)$ as the system error, which satisfies $\|\varepsilon_e(k)\| \leq \varepsilon_e$ for a constant ε_e.

B. Derivation of the Stable Iterative ADP Algorithm

Here, our goal is to design an optimal control scheme that makes the tracking error $e(k)$ converge to zero. Let $U(e(k),\, u_e(k)) = e^T(k)Qe(k) + u_e^T(k)Ru_e(k)$ be the utility function, where Q and R are both positive definite matrices with suitable dimensions. Define the performance index function as

$$J\left(e(0),\, \underline{u}_e(0)\right) = \sum_{k=0}^{\infty} \left(U\left(e(k),\, u_e(k)\right)\right) \tag{11}$$

where we let $\underline{u}_e(k) = (u_e(k),\, u_e(k+1),\, \ldots)$. The optimal performance index function can be defined as $J^*(e(k)) = \inf_{\underline{u}_e(k)}\{J(e(k),\, \underline{u}_e(k))\}$. According to the principle of optimality, $J^*(e(k))$ satisfies the discrete-time HJB equation as follows:

$$J^*\left(e(k)\right) = \inf_{u_e(k)} \left\{U\left(e(k), u_e(k)\right) + J^*\left(e(k+1)\right)\right\}. \tag{12}$$

Define the laws of optimal controls as $u_e^*(e(k)) = \arg\inf_{u_e(k)} \{U(e(k), u_e(k)) + J^*(e(k+1))\}$. Hence, HJB equation (12) can be written as

$$J^*\left(e(k)\right) = U\left(e(k),\, u_e^*\left(e(k)\right)\right) + J^*\left(e(k+1)\right). \tag{13}$$

Generally, $J^*(e(k))$ is a high nonlinear and nonanalytical function, which is nearly impossible to obtain by solving (13) directly. To overcome this difficulty, a new ADP algorithm is developed to obtain the optimal control law iteratively.

In the developed stable iterative ADP algorithm, the performance index function and control law are updated by iterations, with the iteration index i increasing from 0 to infinity. First, let $\mu(e(k))$ be an arbitrary admissible control law, and let $P(e(k))$ be the corresponding performance index function, which satisfies

$$P\left(e(k)\right) = U\left(e(k),\, \mu\left(e(k)\right)\right) + P\left(e(k+1)\right). \tag{14}$$

Let the initial performance index function $V_0(e(k)) = P(e(k))$. Then, for $i = 0,\, 1, \ldots$, the iterative ADP algorithm will iterate between

$$\hat{v}_i\left(e(k)\right) = \min_{\hat{u}_e(k)} \left\{U\left(e(k), \hat{u}_e(k)\right) + \hat{V}_i\left(\hat{e}(k+1)\right)\right\} + \rho_i\left(e(k)\right) \tag{15}$$

$$\hat{V}_{i+1}\left(e(k)\right) = U\left(e(k), \hat{v}_i\left(e(k)\right)\right) + \hat{V}_i\left(\hat{e}(k+1)\right) + \pi_i\left(e(k)\right) \tag{16}$$

where $\rho_i(e(k))$ and $\pi_i(e(k))$ are iteration errors, and $\hat{u}_e(k) = u(k) - \hat{u}_d(k)$.

From the stable iterative ADP algorithm (15) and (16), we can see that the iterative performance index function $\hat{V}_i(e(k))$ is used to approximate $J^*(e(k))$, and the iterative control law $\hat{v}_i(e(k))$ is used to approximate $u^*(e(k))$. Therefore, when $i \to \infty$, the algorithm should be convergent, which makes $\hat{V}_i(e(k))$ and $\hat{v}_i(e(k))$ converge to the optimal ones. In the following, we will show the properties of the developed iterative ADP algorithm.

C. Properties of the Stable Iterative ADP Algorithm With Approximation Errors and Disturbances

From the iterative ADP algorithm (15) and (16), due to the existence of system errors, iteration errors, and disturbances, the convergence analysis methods for the accurate ADP algorithms are invalid. In this paper, inspired by the work in [25] and [36], a new "error bound"-based convergence and stability analysis will be developed. First, we define a new performance index function, i.e.,

$$\Gamma_i\left(e(k)\right) = \min_{u_e(k)} \left\{U\left(e(k), u_e(k)\right) + \hat{V}_i\left(e(k+1)\right)\right\}. \tag{17}$$

Then, we can derive the following theorem.

Theorem 3.1: For $i = 0, 1, \ldots$, the iterative performance index function $\hat{V}_i(e(k))$ and the iterative control law $\hat{v}_i(e(k))$ are obtained by (15) and (16), respectively. Let $\Gamma_i(e(k))$ be expressed as in (17). Then, there exists a constant $\sigma > 1$ that makes

$$\hat{V}_i\left(e(k)\right) \le \sigma\Gamma_i\left(e(k)\right) \tag{18}$$

hold uniformly.

Proof: For $\forall\, i = 0, 1, \ldots$, if we let

$$\bar{v}_i\left(e(k)\right) = \arg\min_{u_e(k)} \left\{U\left(e(k),\, \hat{u}_e(k)\right) + \hat{V}_i\left(\hat{e}(k+1)\right)\right\} \tag{19}$$

then we have $\bar{v}_i(e(k)) = \hat{v}_i(e(k)) - \rho_i(e(k))$. According to (16), we have

$$\hat{V}_{i+1}\left(e(k)\right) = U(e(k),\, (\bar{v}_i\left(e(k)\right) + \rho_i\left(e(k)\right)) + \pi_i\left(e(k)\right) + V_i\left(F_e\left(e(k),\, (\bar{v}_i\left(e(k)\right) + \rho_i\left(e(k)\right)\right)\right)\right). \tag{20}$$

Let $\nabla U(\xi) = \partial U(e(k), \xi)/\partial\xi$ and $\nabla V_i(\xi) = \partial\hat{V}_i(F_e(e(k), \xi))/\partial\xi$. Let $0 \le c_{U_i} \le 1$, $0 \le c'_{U_i} \le 1$, $0 \le c_{V_i} \le 1$, and $0 \le c'_{V_i} \le 1$ be constants, and let $\xi_{U_i} = c_{U_i}\hat{v}_i(e(k)) + (1 - c_{U_i})\bar{v}_i(e(k))$, $\xi'_{U_i} = c'_{U_i}\hat{u}_e(k) + (1 - c'_{U_i})u_e(k)$

$$\begin{aligned}
\xi_{V_i} &= c_{V_i}\hat{V}_i\left(F_e\left(e(k), \bar{v}_i\left(e(k)\right)\right)\right) + (1 - c_{V_i}) \\
&\quad \times \hat{V}_i\left(F_e\left(e(k), \hat{v}_i\left(e(k)\right)\right)\right) \\
\xi'_{V_i} &= c'_{V_i}\hat{V}_i\left(F_e\left(e(k), \hat{u}_e\left(e(k)\right)\right)\right) + \left(1 - c'_{V_i}\right) \\
&\quad \times \hat{V}_i\left(F_e\left(e(k), u_e\left(e(k)\right)\right)\right).
\end{aligned} \tag{21}$$

Then, the iterative performance index function $\hat{V}_i(e(k))$ can be expressed as

$$\begin{aligned}
&\hat{V}_{i+1}\left(e(k)\right) \\
&= U\left(e(k), (\bar{v}_i\left(e(k)\right) + \rho_i\left(e(k)\right))\right) + \pi_i\left(e(k)\right) \\
&\quad + V_i\left(F_e\left(e(k), (\bar{v}_i\left(e(k)\right) + \rho_i\left(e(k)\right))\right)\right) \\
&= \min_{u_e(k)} \left\{U\left(e(k), u_e(k)\right) + \nabla U\left(\xi'_{U_i}\right)\varepsilon_u(k) \right. \\
&\qquad \left. + V_i\left(e(k+1)\right) + \nabla V_i\left(\xi'_{V_i}\right)\varepsilon_e(k)\right\} + \pi_i\left(e(k)\right) \\
&\quad + \nabla U(\xi_{U_i})\rho_i(e(k) + \nabla V_i(\xi_{V_i})\rho_i\left(e(k)\right).
\end{aligned} \tag{22}$$

As $\nabla U(\xi'_{U_i})$, $\nabla V_i(\xi'_{V_i})$, $\nabla U(\xi_{U_i})$, and $\nabla V_i(\xi_{V_i})$ are upper bounded, if we let $|\nabla U(\xi'_{U_i})\varepsilon_u(k)| \le \bar{\varepsilon}_{U_i}$, $|\nabla V_i(\xi'_{V_i})\varepsilon_e(k)| \le \bar{\varepsilon}_{V_i}$, $|\nabla U(\xi_{U_i})\rho_i(e(k))| \le \varepsilon_{U_i}$, $|\nabla V_i(\xi_{V_i})\rho_i(e(k))| \le \varepsilon_{V_i}$, and

$|\pi_i(e(k))| \leq \varepsilon_{\pi_i}$ for constants $\bar{\varepsilon}_{U_i}$, $\bar{\varepsilon}_{V_i}$, ε_{U_i}, and ε_{V_i}, then we have

$$\hat{V}_{i+1}\left(e(k)\right) \leq \Gamma_{i+1}\left(e(k)\right) + \varepsilon_i \tag{23}$$

where $\varepsilon_i = \bar{\varepsilon}_{U_i} + \bar{\varepsilon}_{V_i} + \varepsilon_{U_i} + \varepsilon_{V_i} + \varepsilon_{\pi_i}$ is finite. Hence, for $\forall\, i = 0, 1, \ldots$, there exists a $\sigma \geq 1$ that satisfies (18). The proof is completed. ■

From Theorem 3.1, we can see that, for $\forall\, i = 0, 1, \ldots$, there must exist a finite $\sigma \geq 1$ that makes (18) hold uniformly. Thus, σ can be seen as a uniform approximation error. Then, we can derive the following theorem.

Theorem 3.2: For $\forall\, i = 0, 1, \ldots$, let $\Gamma_i(e(k))$ be expressed as in (18), where $\sigma \geq 1$ is a constant. Let $0 < \gamma < \infty$ and $1 \leq \delta < \infty$ be both constants that make

$$\begin{aligned} J^*\left(\bar{F}\left(e(k), u_e(k)\right)\right) \leq & \gamma U\left(e(k), u_e(k)\right) \\ V_0\left(e(k)\right) \leq & \delta J^*\left(e(k)\right) \end{aligned} \tag{24}$$

hold uniformly. If constant σ in (18) satisfies

$$\sigma \leq 1 + \frac{\delta - 1}{\gamma\delta} \tag{25}$$

then we have that the iterative performance index function $\hat{V}_i(e(k))$ converges to a finite neighborhood of the optimal performance index function $J^*(e(k))$.

Proof: The theorem can be proven in two steps. First, using mathematical induction, we will prove that, for $\forall\, i = 0, 1, \ldots$, the iterative performance index function $\hat{V}_i(e(k))$ satisfies

$$\begin{aligned} \hat{V}_i\left(e(k)\right) \leq \sigma & \left(1 + \sum_{j=1}^{i} \frac{\gamma^j \sigma^{j-1}(\sigma - 1)}{(\gamma+1)^j} + \frac{\gamma^i \sigma^i(\delta - 1)}{(\gamma+1)^i}\right) \\ & \times J^*\left(e(k)\right). \end{aligned} \tag{26}$$

Let $i = 0$. Then, (26) becomes $\hat{V}_0(e(k)) \leq \sigma\delta J^*(e(k))$. We have the conclusion holds for $i = 0$. Assume that (26) holds for $i = l - 1$ and $l = 1, 2, \ldots$. Then, for $i = l$, we have the equation shown at the bottom of the page, which proves (26). The mathematical induction is completed.

Second, according to (25), we have $(\gamma^j \sigma^{j-1}/(\gamma+1)^j) < 1$; hence, the geometrical series $\{(\gamma^j \sigma^{j-1}(\sigma - 1)/(\gamma+1)^j)\}$ is finite as $i \to \infty$. According to (26), the iterative performance index function $\hat{V}_i(e(k))$ is convergent to a finite neighborhood of the optimal performance index function $J^*(e(k))$. ■

Next, we can derive the stability property.

Theorem 3.3: For $i = 0, 1, \ldots$, let $\hat{V}_i(e(k))$ and $\hat{v}_i(e(k))$ be obtained by (15) and (16), respectively. Then, the tracking error system (9) is UUB under the iterative control law $\hat{v}_i(e(k))$.

Proof: According to (26), for $i = 0, 1, \ldots$, let

$$\chi_i = \sigma\left(1 + \sum_{j=1}^{i} \frac{\gamma^j \sigma^{j-1}(\sigma - 1)}{(\gamma+1)^j} + \frac{\gamma^i \sigma^i(\delta - 1)}{(\gamma+1)^i}\right). \tag{27}$$

Define a new iterative performance index function as $\bar{V}_i(e(k)) = \chi_i J^*(e(k))$, where χ_i is defined as in (27). According to (25), we can get $\chi_{i+1} - \chi_i \leq 0$, which means $\bar{V}_{i+1}(e(k)) \leq \bar{V}_i(e(k))$. Let

$$\begin{aligned} \xi_{V_i} = & c_{\bar{V}_i} \bar{V}_i\left(F_e\left(e(k), \bar{v}_i\left(e(k)\right)\right)\right) \\ & + (1 - c_{\bar{V}_i}) \bar{V}_i\left(F_e\left(e(k), \hat{v}_i\left(e(k)\right)\right)\right) \end{aligned}$$

for $0 \leq c_{\bar{V}_i} \leq 1$. Let $|\nabla(\xi_{\bar{V}_i})\varepsilon_e| \leq \varepsilon_{\bar{V}_i}$ for a constant $\varepsilon_{\bar{V}_i}$, and we can obtain

$$\begin{aligned} \bar{V}_i\left(e(k+1)\right) - \bar{V}_i\left(e(k)\right) \leq & -U\left(e(k), \hat{v}_i\left(e(k)\right)\right) + \nabla(\xi_{\bar{V}_i})\varepsilon_e \\ \leq & -U\left(e(k), \hat{v}_i\left(e(k)\right)\right) + \varepsilon_{\bar{V}_i}. \end{aligned}$$

Define a new state error set $\Omega_e = \{e(k) | U(e(k), \hat{v}_i(e(k))) \leq \varepsilon_{\bar{V}_i}\}$. As $U(e(k), \hat{v}_i(e(k)))$ is a positive definite function, for $\forall\, e(k) \in \Omega_e$, we have that $\|e(k)\|$ is finite, where $\|\cdot\|$ denotes the Euclidean norm. We can define

$$\Gamma = \sup_{x \in \Omega_x} \{\|x\|\}.$$

Define a new set as follows:

$$\Omega_\Gamma = \{x | \|x\| \leq \Gamma, x \in \mathbb{R}^n\}.$$

$$\begin{aligned} \Gamma_l\left(e(k)\right) \leq \min_{u_e(k)} \Bigg\{ & \left(1 + \gamma \sum_{j=1}^{l-1} \frac{\gamma^{j-1}\sigma^{j-1}(\sigma - 1)}{(\gamma+1)^j} + \frac{\gamma^{l-1}\sigma^{l-1}(\sigma\delta - 1)}{(\gamma+1)^l}\right) U\left(e(k), u_e(k)\right) \\ & + \left(\sigma\left(1 + \sum_{j=1}^{l} \frac{\gamma^j \sigma^{j-1}(\sigma - 1)}{(\gamma+1)^j} + \frac{\gamma^l \sigma^l(\delta - 1)}{(\gamma+1)^l}\right) - \left(\sum_{j=1}^{l-1} \frac{\gamma^{j-1}\sigma^{j-1}(\sigma - 1)}{(\gamma+1)^j} + \frac{\gamma^{l-1}\sigma^{l-1}(\sigma\delta - 1)}{(\gamma+1)^l}\right)\right) \\ & \times J^*\left(\bar{F}\left(e(k), u_e(k)\right)\right) \Bigg\} \\ = & \left(1 + \sum_{j=1}^{l} \frac{\gamma^j \sigma^{j-1}(\sigma - 1)}{(\gamma+1)^j} + \frac{\gamma^l \sigma^l(\delta - 1)}{(\gamma+1)^l}\right) J^*\left(e(k)\right) \end{aligned}$$

Define two scalar functions $\alpha(\|e(k)\|)$ and $\beta(\|e(k)\|)$, which satisfy the following two conditions.

1) If $e(k) \in \Omega_\Gamma$, then

$$\alpha\left(\|e(k)\|\right) = \beta\left(\|e(k)\|\right) = \bar{V}_i\left(e(k)\right). \tag{28}$$

2) If $e(k) \in \overline{\Omega}_\Gamma$, where $\overline{\Omega}_\Gamma := \mathbb{R}^n \setminus \Omega_\Gamma$, then $\alpha(\|e(k)\|)$ and $\beta(\|e(k)\|)$ are both monotonically increasing functions and satisfy

$$0 < \alpha\left(\|e(k)\|\right) \le \bar{V}_i\left(e(k)\right) \le \beta\left(\|e(k)\|\right). \tag{29}$$

For an arbitrary constant $\varsigma > \Gamma$, there exists a $\varrho(\varsigma) > \Gamma$ that satisfies $\beta(\varrho) \le \alpha(\varsigma)$. For $T = 1, 2, \ldots$, if $e(k) \in \overline{\Omega}_\Gamma$ and $e(k+T) \in \overline{\Omega}_\Gamma$ hold, then $\bar{V}_i(e(k+T)) - \bar{V}_i(e(k)) \le 0$. Hence, for $\forall\, e(k) \in \overline{\Omega}_\Gamma$ that satisfies $\Gamma \le \|e(k)\| \le \beta(\varrho)$, there exists a $T > 0$ that satisfies

$$\alpha(\varsigma) \ge \beta(\varrho) \ge \bar{V}_i\left(e(k)\right) \ge \bar{V}_i\left(e(k+T)\right) \ge \alpha\left(\|e(k+T)\|\right)$$

which obtains $\varsigma > \|e(k+T)\|$. Therefore, for $\forall\, e(k) \in \overline{\Omega}_\Gamma$, there exist a $T = 1, 2, \ldots$ that makes $\|e(k+T)\| \le \varsigma$ hold. As ς is arbitrary, let $\varsigma \to \Gamma$; then, we can obtain $\|e(k+T)\| \in \Omega_\Gamma$. According to the definition in [37], we have that $e(k)$ is UUB, when $\hat{V}_i(e(k))$ reaches the upper hound $\bar{V}_i(e(k))$.

Next, for $\hat{V}_i(e(k)) \le \bar{V}_i(e(k))$, there exists time instants T_0 and T_1 that make

$$\bar{V}_i\left(e(k)\right) \ge \bar{V}_i\left(e(k+T_0)\right) \ge \hat{V}_i\left(e(k)\right) \ge \bar{V}_i\left(e(k+T_1)\right) \tag{30}$$

hold for $\forall\, e(k)$, $e(k+T_0)$, $e(k+T_1) \in \overline{\Omega}_\Gamma$. Choose $\varsigma_1 > 0$ that satisfies $\hat{V}_i(e(k)) \ge \alpha(\varsigma_1) \ge \bar{V}_i(e(k+T_1))$. Then, there exists $\varrho_1(\varsigma_1) > 0$ that makes $\alpha(\varsigma_1) \ge \beta(\varrho_1) \ge \bar{V}_i(e(k+T_1))$ hold. According to (30), we have

$$\begin{aligned}\alpha(\varsigma) \ge \beta(\varrho) \ge \hat{V}_i\left(e(k)\right) \ge \alpha(\varsigma_1) \ge \beta(\varrho_1) \ge \bar{V}_i\left(e(k+T_1)\right) \\ \ge \alpha\left(\|e(k+T_1)\|\right).\end{aligned} \tag{31}$$

According to the definition of $\alpha(\|e(k)\|)$ and $\beta(\|e(k)\|)$ in (28) and (29), for an arbitrary constant $\varsigma > \Gamma$, we can obtain $\|e(k+T_1)\| \le \varsigma$, which shows that $\hat{v}_i(e(k))$ is a UUB control law for the tracking error system (9). ■

Corollary 3.1: For $i = 0, 1, \ldots$, let $\hat{V}_i(e(k))$ and $\hat{v}_i(e(k))$ be obtained by (15) and (16), respectively. If $U(e(k), \hat{v}_i(e(k))) > \nabla V_i(\xi_{\bar{V}_i})\varepsilon_e(k)$ holds for $\forall\, e(k)$, then the iterative control law $\hat{v}_i(e(k))$ is an asymptotically stable control law for system (9).

IV. NN Implementation for the Optimal Tracking Control Scheme

Here, NNs, including the action network and the critic network, are used to implement the developed stable iterative ADP algorithm. The whole structure diagram is shown in Fig. 1.

For $\forall\, i = 0, 1, \ldots$, the critic network is used to approximate the performance index function in (16). For $j = 0, 1, \ldots$, let the output of the critic network be $\hat{V}_{i+1}^j(e(k)) = W_{ci}^{jT}\sigma_c(e(k))$, where $\sigma_c(e(k)) = \sigma(Y_c e(k))$, with Y_c being an arbitrary matrix with a suitable dimension. The target function can be written as

$$V_{i+1}\left(e(k)\right) = U\left(e(k), \hat{v}_i\left(e(k)\right)\right) + \hat{V}_i\left(\hat{e}(k+1)\right). \tag{32}$$

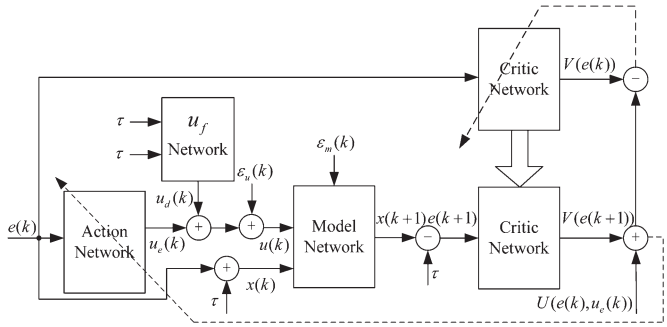

Fig. 1. Structure diagram of the stable iterative ADP algorithm.

Collect an array of tracking errors $\mathcal{E}(k) = \{e^1(k), \ldots, e^p(k)\}$, where p is a large integer. Introduce an iteration index $j = 1, 2, \ldots, p$, and define the error function for the critic network as $\vartheta_{ci}^j(e^j(k)) = \hat{V}_{i+1}^j(e^j(k)) - V_{i+1}(e^j(k))$. The weights of the critic network are updated as [34], [38]

$$W_{ci}^{j+1} = W_{ci}^j - l_c\vartheta_{ci}^j\left(e^j(k)\right)\sigma_c\left(e^j(k)\right) \tag{33}$$

where $\|\sigma_c(e^j(k))\| \le \sigma_C$ for a constant σ_C, and $l_c > 0$ is the learning rate of critic network.

The action network is used to approximate the iterative control law $\bar{v}_i(e(k))$, where $\bar{v}_i(e(k))$ is defined by (19). The output can be formulated as $\hat{v}_i^j(e(k)) = W_{ai}^{jT}\sigma_a(e(k))$, where $\sigma_a(e(k)) = \sigma(Y_a e(k))$. Let Y_a be an arbitrary matrix with a suitable dimension. According to $\mathcal{E}(k)$, we can define the output error of the action network as $\vartheta_{ai}^j(e^j(k)) = \hat{v}_i^j(e^j(k)) - \bar{v}_i(e^j(k))$, $j = 1, 2, \ldots, p$. The weight of the action network can be updated as

$$W_{ai}^{j+1} = W_{ai}^j - l_a\sigma_a\left(e^j(k)\right)\vartheta_{ai}^{jT}\left(e^j(k)\right) \tag{34}$$

where $\|\sigma_a(e^j(k))\| \le \sigma_A$ for a constant σ_A, and $l_a > 0$ is the learning rate of action network. The weight convergence property of the NNs is shown in the following theorem.

Theorem 4.1: For $j = 1, 2, \ldots, p$, let the ideal critic and action network functions be expressed by

$$\begin{aligned}V_{i+1}\left(e^j(k)\right) &= W_{ci}^{*T}\sigma_c\left(e^j(k)\right) + \varepsilon_{ci}\left(e^j(k)\right) \\ \bar{v}_i\left(e^j(k)\right) &= W_{ai}^{*T}\sigma_a\left(e^j(k)\right) + \varepsilon_{ai}\left(e^j(k)\right)\end{aligned}$$

respectively. The critic and action networks are trained by (33) and (34), respectively. Let $\tilde{W}_{ci}^j = W_{ci}^j - W_{ci}^*$ and $\tilde{W}_{ai}^j = W_{ai}^j - W_{ai}^*$. For $\forall\, i = 1, 2, \ldots$, if there exist constants $0 < \lambda_c < 1$ and $0 < \lambda_a < 1$ that satisfy

$$\begin{aligned}\phi_{ci}^{jT}(k)\varepsilon_{ci}\left(e^j(k)\right) &\le \lambda_c\phi_{ci}^{jT}(k)\phi_{ci}^j(k) \\ \phi_{ai}^{jT}(k)\varepsilon_{ai}\left(e^j(k)\right) &\le \lambda_a\phi_{ai}^{jT}(k)\phi_{ai}^j(k)\end{aligned}$$

respectively, where $\phi_{ci}^j(k) = \tilde{W}_{ci}^{jT}\sigma_c(e^j(k))$ and $\phi_{ai}^j(k) = \tilde{W}_{ai}^{jT}\sigma_a(e^j(k))$, then the error matrices $\tilde{W}_{ci}^j$ and $\tilde{W}_{ai}^j$ converge to zero, as $j \to \infty$.

Proof: Consider the following Lyapunov function candidate:

$$L\left(\tilde{W}_{ci}^j, \tilde{W}_{ai}^j\right) = \frac{1}{l_c}\mathrm{tr}\left\{\tilde{W}_{ci}^{jT}\tilde{W}_{ci}^j\right\} + \frac{1}{l_a}\mathrm{tr}\left\{\tilde{W}_{ai}^{jT}\tilde{W}_{ai}^j\right\}.$$

The difference of the Lyapunov function candidate is given by

$$\begin{aligned}
&\Delta L\left(\tilde{W}_{ci}^{j},\tilde{W}_{ai}^{j}\right)\\
&\le -2\left(\phi_{ci}^{jT}(k)\phi_{ci}^{j}(k)-\phi_{ci}^{jT}(k)\varepsilon_{ci}\left(e^{j}(k)\right)\right)\\
&\quad -2\left(\phi_{ai}^{jT}(k)\phi_{ai}^{j}(k)-\phi_{ai}^{jT}(k)\varepsilon_{ai}\left(e^{j}(k)\right)\right)\\
&\quad +l_c\sigma_C^2\left\|\phi_{ci}^{j}(k)-\varepsilon_{ci}\left(e^{j}(k)\right)\right\|^2\\
&\quad +l_a\sigma_A^2\left\|\phi_{ai}^{j}(k)-\varepsilon_{ai}\left(e^{j}(k)\right)\right\|^2\\
&\le \left(-2(1-\lambda_c)+l_c\sigma_C^2(1+\chi_c)\right)\left\|\phi_{ci}^{j}(k)\right\|^2\\
&\quad +\left(-2(1-\lambda_a)+l_a\sigma_A^2(1+\chi_a)\right)\left\|\phi_{ai}^{j}(k)\right\|^2 \qquad (35)
\end{aligned}$$

where we let $\chi_c>0$ and $\chi_a>0$ be constants that satisfy $\varepsilon_{ci}^{T}(e^{j}(k))\varepsilon_{ci}(e^{j}(k))\le\chi_c\phi_{ci}^{jT}(k)\phi_{ci}^{j}(k)$ and $\varepsilon_{ai}^{T}(e^{j}(k))\varepsilon_{ai}(e^{j}(k))\le\chi_a\phi_{ai}^{jT}(k)\phi_{ai}^{j}(k)$, respectively. Selecting l_c and l_a that satisfy $l_c<(2(1-\lambda_c)/\sigma_C^2(1+\chi_c)$ and $l_a<(2(1-\lambda_a)/\sigma_A^2(1+\chi_a)$, respectively, we have $\Delta L(\tilde{W}_{ci}^{j},\tilde{W}_{ai}^{j})\le 0$, $j=1,2,\ldots,p$. Let $j\to\infty$, and we can obtain the conclusion. ■

Based on the given analysis, the whole data-driven stable iterative ADP algorithm for the WGS system can be summarized in Algorithm 1.

Algorithm 1 Data-Driven Stable Iterative ADP Algorithm.

NN modeling and **system transformation:**

1: Collect an array of system data of WGS system (3).
2: Establish model network, where the NN training rule is expressed as in (5).
3: Establish u_f network, where the NN training rule is expressed as in (6).
4: Transform the WGS tracking system (3) into an error regulation system (10).

Stable iterative ADP algorithm:

5: Let $i=0$ and $V_0(e(k))=P(e(k))$, where $P(e(k))$ satisfies (14).
6: Compute the iterative control law $\hat{v}_i(e(k))$ by (15). Update the iterative performance index function $\hat{V}_{i+1}(e(k))$ by (16).
7: If the approximation error σ satisfies (25), then go to Step 8; else, reduce the approximation error σ and go to Step 6.
8: If $|\hat{V}_{i+1}(e(k))-\hat{V}_i(e(k))|\le\zeta$, then go to next step. Otherwise, let $i=i+1$ and go to Step 6.
9: **return** $\hat{v}_i(e(k))$ and $\hat{V}_i(e(k))$.

Remark 4.1: One property should be pointed out. For $\forall\, i=1,2,\ldots$, if we define the approximation error function $\epsilon_i(e(k))$ as

$$\hat{V}_i\left(e(k)\right)=\Gamma_i\left(e(k)\right)+\epsilon_i\left(e(k)\right) \qquad (36)$$

then, according to (23), we have $\epsilon_i(e(k))\le\varepsilon$, where $\varepsilon=\sup\{\varepsilon_i\}$, $i=0,1,\ldots$. According to (18) and (25), we can obtain the following equivalent convergence criterion:

$$\epsilon_i\left(e(k)\right)\le\frac{\hat{V}_i\left(e(k)\right)(\delta-1)}{\gamma\delta+\delta-1}. \qquad (37)$$

From (37), we can see that, if $\|e(k)\|$ is large, then the developed iterative ADP algorithm permits convergence under large approximation errors, and if $\|e(k)\|$ is small, then small approximation errors are required to ensure the convergence of the iterative ADP algorithm. As the approximation errors and disturbances exist, the convergence criterion (37) cannot generally be satisfied for $\forall\, e(k)$. Define a new tracking error set as follows:

$$\Theta_e=\left\{e(k)|\epsilon_i\left(e(k)\right)>\frac{\hat{V}_i\left(e(k)\right)(\delta-1)}{\gamma\delta+\delta-1}\right\}.$$

As $\epsilon_i(e(k))\le\varepsilon$ is finite, if we define $\Upsilon=\sup_{e(k)\in\Theta_e}\{\|e(k)\|\}$, then we have that Υ is finite. Thus, for $\forall\, e(k)\in\overline{\Theta}_e$, $\overline{\Theta}_e:=\mathbb{R}^n\setminus\Theta_e$, we can get that $\hat{V}_i(e(k))$ is convergent, i.e., $\hat{V}_\infty(e(k))=\lim_{i\to\infty}\hat{V}_i(e(k))$.

V. Numerical Analysis

Here, numerical experiments will be studied to show the effectiveness of the developed stable iterative ADP algorithm with approximation errors and disturbances. For WGS system (3), we let the current reaction temperature be $x(0)=273$ °C. Let the desired reaction temperature be $\tau=375$ °C. Observe the volume percentage compositions in the inlet water gas of the WGS system; we obtain $[\theta_{\mathrm{CO}},\theta_{\mathrm{CO_2}},\theta_{\mathrm{H_2}},\theta_{\mathrm{H_2O}}]=[24.39\%,14.71\%,22.79\%,38.11\%]$.

To model WGS system (3), we collect 20 000 input-state data from the real-world WGS operational system. Then, three-layer BP NNs are established with the structures 2–15–1 and 2–15–1 to approximate the WGS system and reference control, respectively. Give the disturbances of the system and control input in Fig. 2(a) and (b), respectively. Let the learning rates of NNs be 0.001, and implement the developed iterative ADP algorithm for 25 iterations. The curve of the admissible approximation errors for the developed ADP algorithm is displayed in Fig. 3. In Fig. 3, we can see that, for different $e(k)$ and iteration index i, it requires different approximation error to guarantee the convergence of the developed iterative ADP algorithm. Let $\bar{\varepsilon}=\max\{\varepsilon_m,\,\varepsilon_u,\,\rho_i,\,\pi_i\}$ be the maximum reconstruction error of NNs for $i=0,1,\ldots,25$. We choose two reconstruction errors $\bar{\varepsilon}$'s, which are 10^{-6} and 10^{-4}, respectively, to train the NNs. The convergence trajectories of the iterative performance index functions are shown in Fig. 2(c) and (d), respectively.

Implement the iterative control law for the WGS system (3). Let the implementation time $T_f=100$. The trajectories of the states and controls are displayed in Fig. 4(a)–(d), respectively. From the numerical results, we can see that, by using the stable iterative ADP algorithm, the iterative control law can make the tracking error system be UUB, which shows the robustness of the developed algorithm. Moreover, we can see that, if we

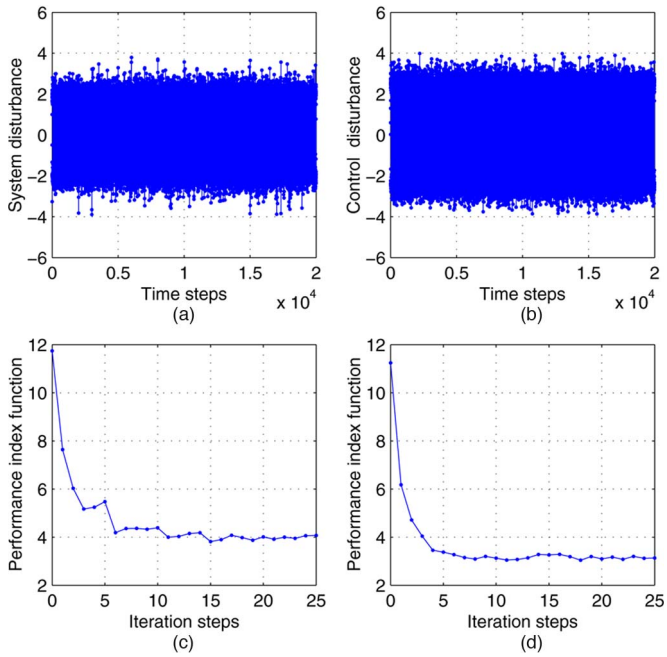

Fig. 2. Disturbances and iterative performance index function. (a) System disturbance. (b) Control disturbance. (c) Iterative performance index function under $\bar{\varepsilon} = 10^{-4}$. (d) Iterative performance index function under $\bar{\varepsilon} = 10^{-6}$.

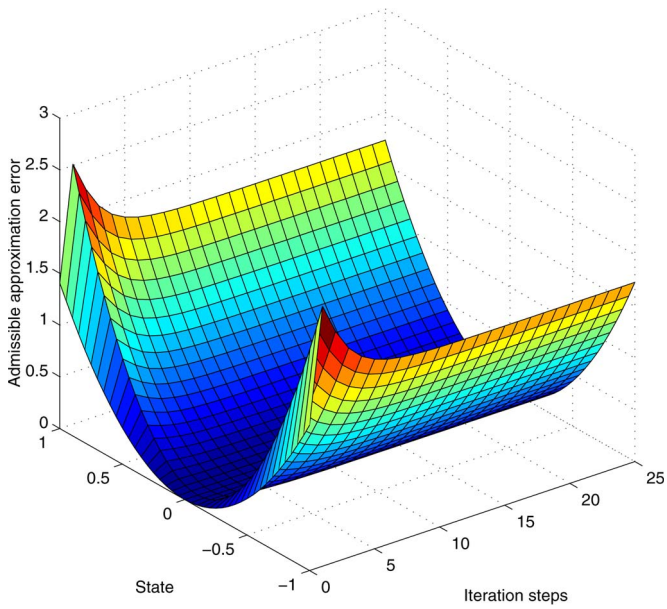

Fig. 3. Curve of the admissible approximation errors.

enhance the training precision of the NNs, such as reduce $\bar{\varepsilon}$ from 10^{-4} to 10^{-6}, then the approximation errors can be reduced, and the system state will be closer to the desired one. The optimal state and control trajectories for $\bar{\varepsilon} = 10^{-6}$ are shown in Fig. 5(a) and (b), respectively. In real-world NN training, the training precision of NNs is generally set to a uniform one. Thus, it is recommended that the developed iterative ADP algorithm is implemented with high training precision that makes the iterative performance index function converge for most of the state space.

On the other hand, to show the effectiveness of the stable iterative ADP algorithm, numerical results by the developed

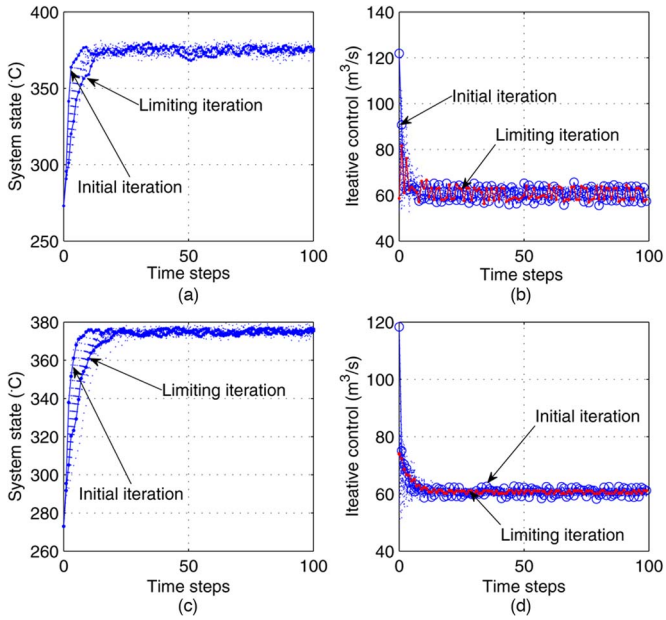

Fig. 4. Iterative trajectories of states and controls for different $\bar{\varepsilon}$'s. (a) State for $\bar{\varepsilon} = 10^{-4}$. (b) Control for $\bar{\varepsilon} = 10^{-4}$. (c) State for $\bar{\varepsilon} = 10^{-6}$. (d) Control for $\bar{\varepsilon} = 10^{-6}$.

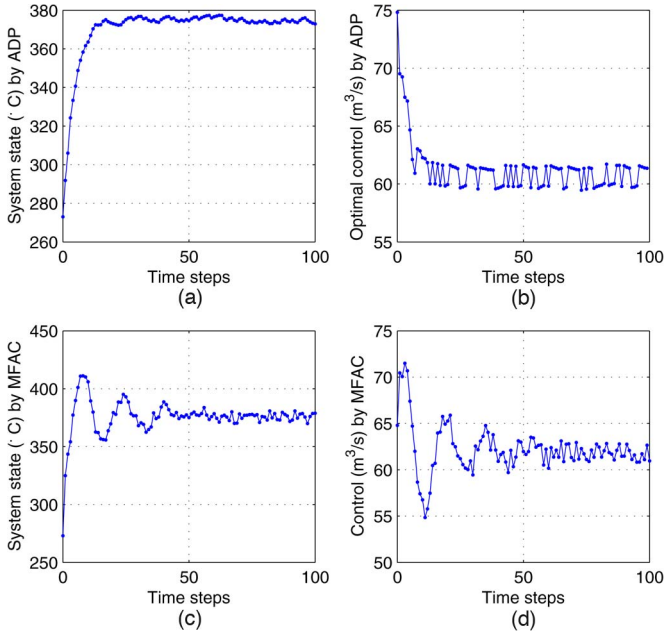

Fig. 5. Comparisions by ADP and MFAC. (a) State trajectory by ADP. (b) Control trajectory by ADP. (c) State trajectory by MFAC. (d) Control trajectory by MFAC.

algorithm will be compared with the ones by the data-driven model-free adaptive control (MFAC) algorithm [39]. According to [39], the controller is designed by

$$u(k) = u(k-1) + \frac{\rho \Phi^T(k)\,(\tau - x(k))}{\lambda + \|\Phi(k)\|^2}$$

$$\Phi(k) = \Phi(k-1) + \frac{\eta(\Delta x(k) - \Phi(k-1)\Delta u(k-1))\,\Delta u(k-1)^2}{\mu + \|\Delta u(k-1)\|}$$

where $\eta = \rho = \mu = 1$, and $\lambda = 0.5$. Let $\Phi(0)$ be initialized by an arbitrary positive definite matrix. The corresponding

state and control trajectories are shown in Fig. 5(c) and (d), respectively. From the numerical results, we can see that, using the developed stable iterative ADP algorithm, it takes 25 time steps to make the system state track the desired one. By MFAC algorithm in [39], it takes 50 iteration steps to make the system state track the desired one. Furthermore, there exist overshoots by the method of [39], although the overshoots are avoided by the developed stable iterative ADP algorithm. These illustrate the effectiveness of the developed algorithm.

VI. Conclusion

In this paper, an effective data-driven stable iterative ADP algorithm has been established to solve optimal temperature control problems for WGS systems. Using the WGS system data, NNs are used to approximate the system model and the reference control, respectively. The stable iterative ADP algorithm is established to obtain the optimal control law where the approximation errors of NNs and the disturbances are both considered. The convergence and stability properties are both analyzed. Finally, numerical results on the WGS system illustrate the effectiveness of the developed algorithm.

References

[1] G.-Y. Kim, J. R. Mayor, and J. Ni, "Parametric study of microreactor design for water gas shift reactor using an integrated reaction and heat exchange model," *Chem. Eng. J.*, vol. 110, no. 1–3, pp. 1–10, Jun. 2005.

[2] T. Baier and G. Kolb, "Temperature control of the water gas shift reaction in microstructured reactors," *Chem. Eng. Sci.*, vol. 62, no. 17, pp. 4602–4611, Sep. 2007.

[3] M. D. Falco, V. Piemonte, and A. Basile, "Performance assessment of water gas shift membrane reactors by a two-dimensional model," *Comput. Aided Chem. Eng.*, vol. 31, pp. 610–614, 2012.

[4] X. Lu and T. Wang, "Water–gas shift modeling in coal gasification in an entrained-flow gasifier–Part 1: Development of methodology and model calibration," *Fuel*, vol. 108, pp. 629–638, Jun. 2013.

[5] G. T. Wright and T. F. Edgar, "Adaptive control of a laboratory water–gas shift reactor with dynamic inlet condition," in *Proc. Amer. Control Conf.*, Pittsburgh, PA, USA, Jun. 1989, pp. 1828–1833.

[6] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, Oct. 2012.

[7] S. Yin, H. Luo, and S. X. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2402–2411, May 2014.

[8] J. Alonso-Martinez, J. Eloy-Garcia, D. Santos-Martin, and S. Arnaltes, "A new variable-frequency optimal direct power control algorithm," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1442–1451, Apr. 2013.

[9] G. Andrikopoulos, G. Nikolakopoulos, I. Arvanitakis, and S. Manesis, "Piecewise affine modeling and constrained optimal control for a pneumatic artificial muscle," *IEEE Trans. Ind. Electron.*, vol. 61, no. 2, pp. 904–916, Feb. 2014.

[10] J. D. Barros, J. F. A. Silva, and E. G. A. Jesus, "Fast-predictive optimal control of NPC multilevel converters," *IEEE Trans. Ind. Electron.*, vol. 60, no. 2, pp. 619–627, Feb. 2013.

[11] T. D. Do, H. H. Choi, and J. W. Jung, "SDRE-based near optimal control system design for PM synchronous motor," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4063–4074, Nov. 2012.

[12] X. Jing and L. Cheng, "An optimal PID control algorithm for training feedforward neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 6, pp. 2273–2283, Jun. 2013.

[13] C. Olalla, I. Queinnec, R. Leyva, and A. E. Aroudi, "Optimal state-feedback control of bilinear DC-DC converters with guaranteed regions of stability," *IEEE Trans. Ind. Electron.*, vol. 59, no. 10, pp. 3868–3880, Oct. 2012.

[14] R. Rathore, H. Holtz, and T. Boller, "Generalized optimal pulsewidth modulation of multilevel inverters for low-switching-frequency control of medium-voltage high-power industrial AC drives," *IEEE Trans. Ind. Electron.*, vol. 60, no. 10, pp. 4215–4224, Oct. 2013.

[15] Y. Ueyama and E. Miyashita, "Optimal feedback control for predicting dynamic stiffness during arm movement," *IEEE Trans. Ind. Electron.*, vol. 61, no. 2, pp. 1044–1052, Feb. 2014.

[16] S. Xiao and Y. Li, "Optimal design, fabrication, and control of an micropositioning stage driven by electromagnetic actuators," *IEEE Trans. Ind. Electron.*, vol. 60, no. 10, pp. 4613–4626, Oct. 2013.

[17] P. J. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *Gen. Syst. Yearbook*, vol. 22, pp. 25–38, 1977.

[18] P. J. Werbos, "A menu of designs for reinforcement learning over time," in *Neural Networks for Control*, W. T. Miller, R. S. Sutton, and P. J. Werbos, Eds. Cambridge, MA, USA: MIT Press, 1991, pp. 67–95.

[19] H. Zhang, F. L. Lewis, and Z. Qu, "Lyapunov, adaptive, and optimal design techniques for cooperative systems on directed communication graphs," *IEEE Trans. Ind. Electron.*, vol. 59, no. 7, pp. 3026–3041, Jul. 2012.

[20] A. Heydari and S. N. Balakrishnan, "Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 1, pp. 145–157, Jan. 2013.

[21] D. Molina, G. K. Venayagamoorthy, J. Liang, and R. G. Harley, "Intelligent local area signals based damping of power system oscillations using virtual generators and approximate dynamic programming," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 498–508, Mar. 2013.

[22] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Control Syst. Mag.*, vol. 32, no. 6, pp. 76–105, 3rd Quart., 2012.

[23] T. Dierks and S. Jagannathan, "Online optimal control of affine nonlinear discrete-time systems with unknown internal dynamics by using time-based policy update," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1118–1129, Jul. 2012.

[24] F. L. Lewis and D. Liu, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Hoboken, NJ, USA: Wiley, 2012.

[25] D. Liu and Q. Wei, "Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 779–789, Apr. 2013.

[26] Q. Wei and D. Liu, "An iterative ϵ-optimal control scheme for a class of discrete-time nonlinear systems with unfixed initial state," *Neural Netw.*, vol. 32, pp. 236–244, Aug. 2012.

[27] Q. Wei and D. Liu, "Numerical adaptive learning control scheme for discrete-time nonlinear systems," *IET Control Theory Appl.*, vol. 7, no. 11, pp. 1472–1486, Jul. 2013.

[28] Q. Wei, D. Wang, and D. Zhang, "Dual iterative adaptive dynamic programming for a class of discrete-time nonlinear systems with time-delays," *Neural Comput. Appl.*, vol. 23, no. 7/8, pp. 1851–1863, Dec. 2013.

[29] H. Zhang, Q. Wei, and D. Liu, "An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games," *Automatica*, vol. 47, no. 1, pp. 207–214, Jan. 2011.

[30] M. Sudhakar, S. Narasimhan, and N. S. Kaisare, "Approximate dynamic programming based control for water gas shift reactor," *Comput. Aided Chem. Eng.*, vol. 31, pp. 340–344, 2012.

[31] Y. Huang, D. Liu, and Q. Wei, "Temperature control in water–gas shift reaction with adaptive dynamic programming," in *Proc. 9th Int. Symp. Neural Netw.*, Shenyang, China, Jul. 2012, pp. 478–487.

[32] J. M. Moe, "Design of water gas shift reactors," *Chem. Eng. Progr.*, vol. 58, no. 3, pp. 33–36, Dec. 1962.

[33] J. Si and Y.-T. Wang, "Online learning control by association and reinforcement," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 264–276, Mar. 2001.

[34] Q. Wei and D. Liu, "A novel iterative θ-adaptive dynamic programming for discrete-time nonlinear systems," *IEEE Trans. Autom. Sci. Eng.*, to be published.

[35] H. Zhang, Q. Wei, and Y. Luo, "A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 937–942, Aug. 2008.

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

[37] H. K. Khalil, *Nonlinear System*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[38] 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, Mar. 2014.

[39] Z. Hou and S. Jin, "Data-driven model-free adaptive control for a class of MIMO nonlinear discrete-time systems," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2173–2188, Dec. 2011.

Qinglai Wei (M'11) received the B.S. degree in automation and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University, Shenyang, China, in 2002, 2005, and 2008, respectively.

From 2009 to 2011, he was a Postdoctoral Fellow with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently an Associate Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interests include neural-network-based control, adaptive dynamic programming, optimal control, nonlinear systems, and their industrial applications.

Derong Liu (S'91–M'94–SM'96–F'05) received the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1994.

From 1993 to 1995, he was a Staff Fellow with the General Motors Research and Development Center, Warren, MI, USA. From 1995 to 1999, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA. In 1999, he joined the University of Illinois at Chicago, Chicago, IL, USA, where he became a Full Professor of electrical and computer engineering and of computer science in 2006. He is the author of 14 books (six research monographs and eight edited volumes).

Dr. Liu currently serves as the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. He received the Faculty Early Career Development Award from the National Science Foundation in 1999, the University Scholar Award from the University of Illinois in 2006, and the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008. He was selected for the "100 Talents Program" by the Chinese Academy of Sciences in 2008.