Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7953358, 9 pages
https://doi.org/10.1155/2022/7953358

Research Article

Optimal Tracking Control for a Discrete Time Nonlinear Nuclear Power System

Zhenhua Luan,[1,2] Mengxuan Wang,[3] Yuzhen Zhang,[4] Qinglai Wei,[4] Tianmin Zhou,[4] Zhiwu Guo,[1] and Jun Ling[5]

[1]*State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, China Nuclear Power Engineering Co. Ltd, Shenzhen 518172, Guangdong, China*
[2]*State Key Lab of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, China*
[3]*Harbin University of Science and Technology, Harbin, China*
[4]*The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[5]*Shenzhen Urban Public Safety and Technology Institute, Shenzhen 518046, Guangdong, China*

Correspondence should be addressed to Yuzhen Zhang; zhangyuzhen@ia.ac.cn

Received 19 October 2021; Revised 25 May 2022; Accepted 26 May 2022; Published 22 June 2022

Academic Editor: Junyong Zhai

Recently, increasing attention has been paid to nuclear power control with the appeals of clean energy and demands of power regulation to integrate into the power grid. However, a nuclear power system is a discrete-time (DT) nonlinear and complicated system, where the parameters entangle with intrinsic states. Furthermore, the need for huge computational ability due to the high-level order property in the nuclear reactor model causes many difficulties in the power control of nuclear industries. In this study, a new scheme of optimal tracking control for DT nonlinear nuclear power systems is provided to accomplish the power control of a 2500-MW pressurized water reactor (PWR) nuclear power plant. The proposed approach based on the value iteration method is a novel algorithm in the human intelligence community, which has a basic actor-critic structure with neural networks (NNs). The new approach has some modifications, where the cost function is redefined by leveraging the higher-order polynomial to substitute neural networks in the entire actor critic architecture. Simulation results of the 2500-MW PWR nuclear power plant are given to demonstrate the effectiveness of the developed method.

1. Introduction

Considering the issues of environmental deterioration, e.g., air pollution due to excessive fossil fuel consumption, it is significant that humans develop clean energy technology to ease this situation. Nuclear energy is almost the most rapidly developing clean power to provide power to the power grid. However, currently adopted control strategies have problems such as the intrinsic nonlinearity of nuclear reactor systems and varying parameters following the power level. In fact, over decades of development for nuclear power industry technology and control policy, many outstanding researchers have made excellent progress in this field.

Since the last century, one mature control strategy called PID control policy has deeply affected the power control in the nuclear industry [1, 2]. With the advancement in control technology, some model predictive control (MPC) and multimodel adaptive control theories focused on local linearization to approximate the nonlinearity of nuclear power systems and have also been applied in this area [3–5]. The control algorithm extensively applied in nuclear power control is fuzzy control or its combinations with other control theories to address different demands. Wu leveraged the parallel distribution compensation (PDC) method-based T-S fuzzy control to restrict the nonlinearity of a nuclear power system [6], and Eliasi designed an appropriate

controller for the UTSG water level in nuclear power plants using a fuzzy control policy and MPC algorithm [7]. Many researchers have used different methods to tackle various problems. For example, Gang employed a radial basis function neural network (RBFN) to guarantee the correctness in identifying the nuclear steam generator process dynamics [8]. Wang applied the adaptive control method and guaranteed a cost control method in nuclear power control problems [9].

The aforementioned approaches mostly relate to the linearization procedure, which largely omits the numerical error toward a nonlinear model. The intrinsic nonlinearity of the high-order nuclear model is ignored. To better satisfy the demands of tracking control problems, it is necessary to propose a new control strategy in this area. With the development in the intelligent control community, researchers are pursuing reinforcement learning (RL) algorithms to solve nonlinear problems in practice. Adaptive dynamic programming (ADP), which was proposed by Werbos, plays an important role in reinforcement learning-based control policy [10–13], and it is well known as a self-learning optimal control policy. The well-studied iteration method is the policy iteration (PI) algorithm and value iteration (VI) algorithm. Among the iteration methods, the value iteration algorithm is one type of the most crucial iterative ADP algorithms. It has been studied in many types of research [14–16]. To find the optimal control policy of discrete-time affine nonlinear systems, Al-Tamimi and Lewis used heuristic dynamic programming (HDP) to fulfill the design purpose [17]. Wei. *Q*. [18] proposed a new value iteration method, which mainly focuses on optimal control for DT nonlinear systems. This study also provided detailed proof of the iterative control policy and illustrated that the value function was monotonically nondecreasing, which implies that it will converge to the extremum.

To satisfy the demands of industrial systems, optimal tracking control ADP methods have been deeply investigated. There are also ADP techniques [19–21] to obtain solutions of optimal tracking problems with various system dynamics, such as partially unknown system models or completely unknown system models. Related optimal tracking control techniques have been applied in many industrial plants in recent years [22–26].

In this study, a value-iteration optimal tracking control method is developed for DT nonlinear systems. The main contributions of this study are summarized as follows:

(1) Compared with the traditional control methods dealing with DT nonlinear models of the 2500 MW pressurized water reactor (PWR) systems [1, 2, 4, 5], a self-learning optimal tracking controller is designed to satisfy complex nonlinear behaviors of the 2500 MW PWR nuclear system

(2) The developed value-iteration method guarantees the control law converges to a near-optimal control solution and the admissibility of iterative control laws is analyzed

In this study, our major work is to design an optimal tracking controller for a 2500 MW PWR nuclear power plant by combining the properties of the value iteration and actor critic algorithm. The 2500-MW PWR nuclear power plant is introduced in Section 2, and the discrete definition is given. In Section 3, the details of this proposed algorithm are thoroughly described. The implementation of the proposed method and simulation works are provided in Section 4. Finally, the conclusions are drawn in Section 5.

2. Nonlinear 2500-MW PWR Nuclear Power Plant

The famous nuclear system model is based on Mann's model without xenon poisoning, which consists of a core full lump and two coolant lumps. The discrete version and its transformation are also given in this section.

2.1. Nonlinear 2500-MW PWR Nuclear Power Plant. This fifth-order nonlinear PWR model includes the point kinetics equations, six delayed neutron groups, two equations for the lumped coolant outlet temperature and average fuel temperature, and the reactive equation of the control rod [27, 28]. Multiple sets of delayed neutron point reactor dynamic equations are described as follows:

$$\begin{cases} \dfrac{dn(t)}{dt} = \dfrac{\rho-\beta}{\Lambda} n(t) + \displaystyle\sum_{i=1}^{6} \lambda_i c_i(t), & \dfrac{dc_i(t)}{dt} = \dfrac{\beta_i}{\Lambda} n(t) - \lambda_i c_i(t). \end{cases} \tag{1}$$

To reduce the computational work caused by six delayed neutron point reactor dynamic equations, the simple method is to use single delayed neutron point kinetics equations to approximate multiple sets of delayed neutron point reactor dynamic equations [29]. Thus, the entire PWR model is summarized as follows:

$$\begin{aligned} \frac{dn_r}{dt} &= \frac{\rho-\beta}{\Lambda} n_r + \frac{\beta}{\Lambda} c_r, \\ \frac{dc_r}{dt} &= \lambda n_r - \lambda c_r, \\ \frac{dT_f}{dt} &= \frac{f_f P_{0a}}{\mu_f} n_r - \frac{O}{\mu_f} T_f + \frac{O}{2\mu_f} T_l + \frac{O}{2\mu_f} T_e, \\ \frac{dT_l}{dt} &= \frac{\left(1-f_f\right) P_{0a}}{\mu_c} n_r + \frac{O}{\mu_c} T_f - \frac{(2M+O)}{2\mu_c} T_l + \frac{(2M-O)}{2\mu_c} T_{,e} \\ \frac{d\rho_r}{dt} &= G_r Z_r, \\ \rho &= \rho_r + \alpha_f \left(T_f - T_{f0}\right) + \frac{\alpha_c \left(T_l - T_{l0}\right)}{2} + \frac{\alpha_c \left(T_e - T_{e0}\right)}{2}, \end{aligned} \tag{2}$$

where n_r is the neutron density relative to the initial equilibrium density, %, c_r is the delay neutron density relative to its initial equilibrium density, %, T_f is the average fuel temperature, °C, T_l is the coolant temperature at the core outlet, °C, ρ_r is reactivity contributed by the control rod

movement, and Z_r is the speed of the control rod. The remaining specifications are illustrated in Nomenclature [30]. T_e always approximates T_{e0} in the PWR model. The state n_r can be described as a percentage factor of the full power level, since the reactor power is expressed as $P(t) = P_{0a}n_r$.

In addition, five parameters vary with n_r, which causes severe instability of the nuclear reactor power model and increases the control complexity. The remaining parameters and specific relation are shown in Tables 1 and 2. With the lifting (lowering) load of the PWR model, the varying parameters will lead to a sharp difference in the solutions of the dynamic model. Thus, the model will be uncontrollable, and the solutions of this dynamic model may become divergent. Linearizing nuclear systems to realize various control goals has been a common method in traditional control policies in the past few years. However, there should be a new approach in nonlinear systems to solve these problems.

2.2. System Discretization and Transformation. The optimal tracking control problem can be considered minimizing the real dynamic trajectory with the desired trajectory. Depending on the model, we let $x = [x_1, x_2, x_3, x_4, x_5]^T$; thus, the control-oriented nuclear power system can be defined as

$$\dot{x} = f(x) + g(x)u, \tag{3}$$

where $x_1 = n_r, x_2 = c_r, x_3 = T_f, x_4 = T_l$, and $x_5 = \rho_r$; thus, $f(x)$ and $g(x)$ can be derived as

$$f(x) = \begin{bmatrix} \left(\dfrac{x_5 + \alpha_f\left(x_3 - T_{f0}\right) - \beta}{\Lambda} + \dfrac{\alpha_c\left(T_l - T_{l0}\right)}{2\Lambda}\right)x_1 + \dfrac{\beta}{\Lambda}x_2 \\ \lambda x_1 - \lambda x_2 \\ \dfrac{f_f P_{0a}}{\mu_f}x_1 - \dfrac{O}{\mu_f}x_3 + \dfrac{O}{2\mu_f}x_4 + \dfrac{O}{2\mu_f}T_e \\ \dfrac{\left(1 - f_f\right)P_{0a}}{\mu_c}x_1 + \dfrac{O}{\mu_c}x_3 - \dfrac{(2M+O)}{2\mu_c}x_4 + \dfrac{(2M-O)}{2\mu_c}T_e \\ 0 \end{bmatrix} \tag{4}$$

$$g(x) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ Gr \end{bmatrix}. \tag{5}$$

As the above descriptions show, there is only one controllable signal Z_r, i.e., the speed of the control rod. We let control variable $u = Z_r$.

According to the definition, $\dot{x_k} = x_{k+1} - x_k/\Delta T$, and we have the discretization version of the PWR power model:

$$x_{k+1} = f(x_k) + g(x_k)u_k. \tag{6}$$

Basically, we define the optimal tracking control problem as obtaining an optimal control strategy so that the system can track the reference state, and the desired trajectory x_d can be expressed as

$$x_{d,k+1} = f\left(x_{d,k}\right) + g\left(x_{d,k}\right)u_{d,k}. \tag{7}$$

The tracking error of the state is defined as $\varepsilon_k = x_k - x_{d,k}$. Thus, the dynamic error system can be described as follows:

$$\varepsilon_{k+1} = f(x_k) + g(x_k)u_k - x_{d,k}. \tag{8}$$

Additionally, we must confirm an initial steady-state control policy u_d, and the error of the controller can be defined as $u_{\varepsilon,k} = u_k - u_{d,k}$, so the dynamic tracking error system can also be written as

$$\varepsilon_{k+1} = f_\varepsilon(\varepsilon_k) + g_\varepsilon(\varepsilon_k)u_{\varepsilon,k}, \tag{9}$$

where $f_\varepsilon(\varepsilon_k) = f(\varepsilon_k + x_{d,k}) - x_{d,k}$ and $g_\varepsilon(\varepsilon_k) = g(\varepsilon_k + x_{d,k})$.

We define the utility function as follows:

$$U\left(\varepsilon_k, u_{\varepsilon,k}\right) = \varepsilon_k^T Q\varepsilon_k + u_{\varepsilon,k}^T Ru_{\varepsilon,k}. \tag{10}$$

Thus, the tracking error cost function is written as

$$\mathfrak{J}\left(\varepsilon_k, u_{\varepsilon,k}\right) = \sum_k^\infty U\left(\varepsilon_k, u_{\varepsilon,k}\right). \tag{11}$$

From the principle of optimality, the DT Hamiltonian function is derived as

$$H\left(\varepsilon_k, u_{\varepsilon,k}, \nabla\mathfrak{J}_{\varepsilon,k}\right) = U\left(\varepsilon_k, u_{\varepsilon,k}\right) + \mathfrak{J}_{\varepsilon,k+1} - \mathfrak{J}_{\varepsilon,k}. \tag{12}$$

Thus, the HJB equation can be written as

$$\min_{u_\varepsilon}\left[H\left(\varepsilon_k, u_{\varepsilon,k}, \nabla\mathfrak{J}_{\varepsilon,k}^*\right)\right] = 0. \tag{13}$$

Then, the optimal tracking control law for the error system is derived:

$$u_{\varepsilon,k}^* = \frac{1}{2}R^{-1}g_\varepsilon^T(\varepsilon_k)\frac{\partial\mathfrak{J}_\varepsilon(\varepsilon_{k+1})}{\partial\varepsilon_{k+1}}. \tag{14}$$

Finally, we obtain the standard optimal control law as

$$u_k^* = u_{d,k} + u_{\varepsilon,k}^*.. \tag{15}$$

For linear systems, the HJB equation is reduced to the Riccati equation. However, due to the nonlinearity of the nuclear power system, it is extremely intractable to solve the HJB equation (13) for the nonlinear system. Thus, the value iteration (VI) method based on the actor critic NN algorithm will be adopted to find an approximate optimal solution of the HJB equation (13), which implies that nothing is required about the knowledge of the model drifts or the command generator.

Table 1: Nomenclature.

Nomenclature	
G_r	Reactivity worth of the control rod bank
P_{0a}	Nominal core power (MW)
Λ	Mean neutron lifetime (s)
β	The total fraction of effective the ith group of delayed neutrons
O	The heat transfer coefficient between fuel and coolant ($MW \cdot s/^\circ C$)
M	Heat capacity of mass flux of coolant ($MW \cdot s/\ ^\circ C$)
α_c	The reactivity coefficient of coolant temperature
α_f	The reactivity coefficient of fuel temperature.
μ_c	Heat capacity of coolant ($MW \cdot s/^\circ C$)
μ_f	Heat capacity of fuel ($MW \cdot s/^\circ C$)
λ	Equivalent single-group delayed neutron precursor nuclear decay constant
ρ	Core reactivity
T_e	Average inlet temperature of coolant ($^\circ C$)
T_{e0}	Initial equilibrium inlet temperature of coolant ($^\circ C$)
T_{l0}	Initial equilibrium outlet temperature of coolant ($^\circ C$).
T_{f0}	Initial equilibrium fuel temperature ($^\circ C$).

3. Algorithm Analysis

The detailed convergence properties of this proposed value iteration algorithm are illustrated in this section, and the details of actor critic NN will also be discussed.

3.1. Analysis of the Value Iteration Algorithm for Tracking Control of Nonlinear Systems. Considering the nonaffine nonlinear system, for an infinite time optimal tracking problem, the goal is to obtain an optimal controller such that the state x_k tracks the specified reference trajectory $x_{d,k}$.

Remark 1. For many nonlinear systems, there is a feedback control $u_{d,k}$ that makes (9) work. For example, with regard to DT nonlinear systems (7) with invertible $g(x_{d,k})$, the desired control $u_{d,k}$ can be derived as

$$u_{d,k} = g^{-1}\left(x_{d,k}\right)\left(x_{d,k+1} - f\left(x_{d,k}\right)\right). \tag{16}$$

From equation (11) of Section 2.2, the quadratic cost function of tracking errors ε_k is defined as

$$\mathfrak{F}\left(\varepsilon_0, \underline{u_{\varepsilon,0}}\right) = \sum_{k=0}^{\infty} \left\{\varepsilon_k Q \varepsilon_k + u_{\varepsilon,k}^T R u_{\varepsilon,k}\right\}, \tag{17}$$

where $\underline{u_{\varepsilon,0}} = (u_{\varepsilon,0}, u_{\varepsilon,1}, \ldots)$ and $Q(\cdot)$ and $R(\cdot)$ are positive definite functions.

To obtain an optimum tracking control law that tracks the reference state $x_{d,k}$ and minimizes the tracking error cost function (18), we can redefine the optimal tracking error cost function as follows:

$$\mathfrak{F}^*\left(\varepsilon_k\right) = \inf_{\underline{u_{\varepsilon,k}}}\left\{\mathfrak{F}\left(\varepsilon_k, \underline{u_{\varepsilon,k}}\right)\right\}. \tag{18}$$

Based on Bellman's principle of optimality, $\mathfrak{F}^*(\varepsilon_k)$ satisfies Bellman equation:

$$\mathfrak{F}^*\left(\varepsilon_k\right) = \min_{\varepsilon_k}\left\{U\left(\varepsilon_k, u_{\varepsilon,k}\right) + \mathfrak{F}^*\left(\varepsilon_{k+1}\right)\right\}. \tag{19}$$

Then, the optimal tracking control law is obtained by

$$u^*\left(\varepsilon_k\right) = \arg\min_{u\varepsilon,k}\left\{U\left(\varepsilon_k, u_{\varepsilon,k}\right) + \mathfrak{F}^*\left(\varepsilon_k\right)\right\}. \tag{20}$$

Given the above formulation, we can derive the tracking error performance index function as follows:

$$\mathfrak{F}^*\left(\varepsilon_k\right) = U\left(\varepsilon_k, u_{\varepsilon,k}^*\right) + \mathfrak{F}^*\left(\varepsilon_{k+1}\right). \tag{21}$$

Let $\varphi(\varepsilon_k)$ be a positive definite function for $\varepsilon_k \in \mathbb{R}^5$, and the initial tracking value function is

$$\mathfrak{F}_0\left(\varepsilon_k\right) = \varphi\left(\varepsilon_k\right). \tag{22}$$

The optimal control law $v_0(\varepsilon_k)$ can be obtained by

$$v_0\left(\varepsilon_k\right) = \arg\min_{u_{\varepsilon_k,k}}\left\{U\left(\varepsilon_k, u_{\varepsilon,k}\right) + \mathfrak{F}_0\left(\varepsilon_{k+1}\right)\right\}, \tag{23}$$

where $\mathfrak{F}_0(\varepsilon_{k+1}) = \varphi(\varepsilon_k)$. For $i = 1, 2, 3\ldots$, in this iterative value function algorithm, the value function is updated through

$$\mathfrak{F}_i\left(\varepsilon_k\right) = \min_{u_{\varepsilon,k}}\left\{U\left(\varepsilon_k, u_{\varepsilon,k} + \mathfrak{F}_{i-1}\left(\varepsilon_{k+1}\right)\right)\right\}, \tag{24}$$

and the control policy is improved by

$$\pi_i\left(\varepsilon_k\right) = \arg\min_{u_{\varepsilon,k}}\left\{U\left(\varepsilon_k, u_{\varepsilon,k}\right) + \mathfrak{F}_{i-1}\left(\varepsilon_{k+1}\right)\right\}. \tag{25}$$

Theorem 1. *For the tracking error cost function $\mathfrak{F}_i(\varepsilon_k)$ and control law $\pi_i(\varepsilon_k)$ obtained by (22)–(25), we have α, β, γ, and η satisfying $0 < \eta \le \gamma < \infty$ and $0 \le \alpha \le \beta < 1$, respectively. If $\forall \varepsilon_k$, we have*

$$\eta U\left(\varepsilon_k, \pi_k\right) \le \mathfrak{F}^*\left(\varepsilon_{k+1}\right) \le \gamma U\left(\varepsilon_k, \pi_k\right), \tag{26}$$

$$\alpha \mathfrak{F}^*\left(\varepsilon_k\right) \le \mathfrak{F}_0\left(\varepsilon_k\right) \le \beta \mathfrak{F}^*\left(\varepsilon_k\right), \tag{27}$$

are satisfied uniformly; thus, the iterative value function $\mathfrak{F}_i(\varepsilon_k)$ satisfies

$$\left(1 + \frac{\alpha - 1}{\left(1 + \gamma^{-1}\right)^i}\right)\mathfrak{F}^*\left(\varepsilon_k\right) \le \mathfrak{F}_i\left(\varepsilon_k\right) \le \left(1 + \frac{\beta - 1}{\left(1 + \eta^{-1}\right)^i}\right)\mathfrak{F}^*\left(\varepsilon_k\right). \tag{28}$$

Table 2: Parameters of the PWR nuclear reactor.

Parameters	Value	Parameters	Value
β	0.0065	μ_f	26.3
λ	0.15 (s^{-1})	f_f	0.98
Λ	0.00002(s)	T_{e0}	290°C
G_r	0.0145	T_{f0}	673.8°C
P_0	2500(MW)	T_{l0}	302.2°C
μ_c	$160n_r/9 + 54.022$	O	$5nr/3 + 4.933$
α_f	$(n_r - 4.24) \times 10^{-5}$	M	$28.0n_r + 74.0$
α_c	$(-4.0n_r - 17.3) \times 10^{-5}$		

Theorem 2. *For cost function $\mathfrak{I}_i(\varepsilon_k)$ and control law $\pi_i(\varepsilon_k)$ obtained by (22)–(25), we have α, β, γ, and η satisfying $1 \le \alpha \le \beta < \infty$. If $\forall \varepsilon_k$, indexes (26) and (27) hold uniformly; then, the value function $J_i(\varepsilon_k)$ satisfies equation (28)*

The proof is thoroughly provided in [17].

Corollary 1. *For $i = 0, 1, \ldots$, $\pi_i(\varepsilon_k)$ and $V_i(\varepsilon_k)$ are obtained by (22)–(25). Let α, β, γ, and η be constants that satisfy $0 < \eta \le \gamma < \infty$ and $0 \le \alpha \le \beta < \infty$, respectively. If $\forall \varepsilon_k$, inequalities (26) and (27) hold uniformly. Then, the iterative value function $\mathfrak{I}_i(\varepsilon_k)$ converges to the optimal cost function $\mathfrak{I}^*(\varepsilon_k)$, i.e.,*

$$\lim_{i \longrightarrow \infty} \mathfrak{I}_i(\varepsilon_k) = \mathfrak{I}^*(\varepsilon_k). \tag{29}$$

Based on the aforementioned analysis, we can conclude that the iterative tracking error performance index function will converge to the optimum as $i \longrightarrow \infty$, which is independent of the initial value function $\varphi(\varepsilon_k)$.

According to Lyapunov stability principle, $\mathfrak{I}_i(\varepsilon_k)$ is a Lyapunov function. Since the utility function $U(\varepsilon_k, \pi_k)$ is a positive definite function and $\mathfrak{I}_i(0) = 0$, it should be noticed that $\mathfrak{I}_i(\varepsilon_k)$ is a positive definite function as well. We let

$$\begin{aligned} &\mathfrak{I}_i(\varepsilon_{k+1}) - \mathfrak{I}_i(\varepsilon_k) \le \mathfrak{I}_i(\varepsilon_{k+1}) - \mathfrak{I}_{i+1}(\varepsilon_k), \\ &= -U(\varepsilon_k, \pi_k) \le 0, \end{aligned} \tag{30}$$

where the error tracking control law π_k is admissible.

3.2. Actor Critic NN Implementation of the Value Iteration Algorithm for DT Nonlinear Systems. The actor critic NN has been employed in various fields to approximate the cost function and optimal controller. For example, the optimal tracking applied on partially unknown DT nonlinear systems [31] and rigorous proof for this method are provided. The actor critic structure also has good performance in the tracking control problem for continuous-time nonlinear systems [32, 33].

With regard to optimal tracking control algorithms of DT nonlinear systems, it is quite difficult to directly obtain solutions by solving (13). Thus, the actor critic network structure with the flowchart of the nuclear power system is given in Figure 1, which describes the inner procedures of the method.

While the tracking error system is fed with a specific initial state and desired trajectory, the error will be calculated by a utility function. Simultaneously, the critic network will

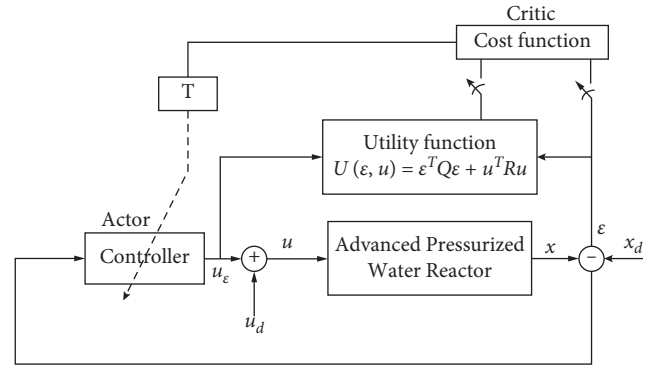

Figure 1: Actor-critic structure of the optimal tracking error system.

be trained in a way that minimizes the utility function. Under the training process, the actor network will behave like an optimal controller. To avoid overfitting, a specified threshold is given at the beginning of the training procedure so that it can be stopped in time.

Inspired by Abu-Khalaf [34], an optimal control algorithm with a high-order polynomial was proposed to substitute the neural unit. This technique is introduced in this actor critic NN structure to obtain a better approximate effect.

To solve (24) and (25), we let the tracking error cost function $\widehat{V}_i$ be approximated by a critic NN:

$$\widehat{V}_i(\varepsilon) = \sum_{j=1}^{L} w_c^j \Theta^j(\varepsilon) = W_{Vi} \Theta(\varepsilon), \tag{31}$$

where we have the approximate activation function $\Theta(\varepsilon) \triangleq [\Theta^1(\varepsilon_k), \Theta^2(\varepsilon_k), \ldots, \Theta^L(\varepsilon_k)]^{\mathrm{T}}$, the weight vector $W_{V_i} = [w_c^1, w_c^2, \ldots, w_c^L]$, and L is the number of neural units in the hidden structure of the critic NN.

Then, the iterative formulation can be obtained as follows:

$$W_{V_{i+1}} \Theta(\varepsilon)^{i+1} = U\left(\varepsilon_k, u_{\varepsilon,k}\right) + W_{Vi} \Theta^i(\varepsilon). \tag{32}$$

For each sample ε_k related to x_k, we formulate the definition as follows:

$$W_{V_{i+1}} = \left(Z^T Z\right)^{-1} Z^T \zeta^i, \tag{33}$$

where $Z = \Theta(\varepsilon)^i$ and $\zeta^i = U(\varepsilon_k, u_{\varepsilon,k}) + W_{Vi} \Theta^i(\varepsilon)$. Then, the weights of critic network are obtained. We let $u_i(\varepsilon)$ be approximated by an actor NN:

$$\widehat{u}_i(\varepsilon) = \sum_{j=1}^{M} w_a^j \delta^j(\varepsilon) = W_{ui} \delta(\varepsilon), \tag{34}$$

where we have the approximate activation function $\delta(\varepsilon) \triangleq [\delta^1(\varepsilon_k), \delta^2(\varepsilon_k), \ldots, \delta^M(\varepsilon_k)]^{\mathrm{T}}$ and weight vector $W_{ui} = [w_a^1, w_a^2, \ldots, w_a^M]$. M is the number of neural units in actor NN.

According to (24) and (30), we will tune the weights of the critic NN at each iteration of this VI algorithm. Our goal is to minimize the residual error between each $\widehat{V}_i(\varepsilon)$ to obtain a new target function as follows:

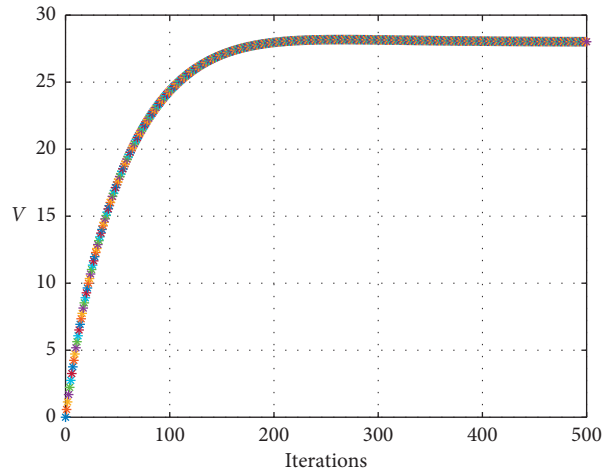

Figure 2: Performance index function of the tracking error.

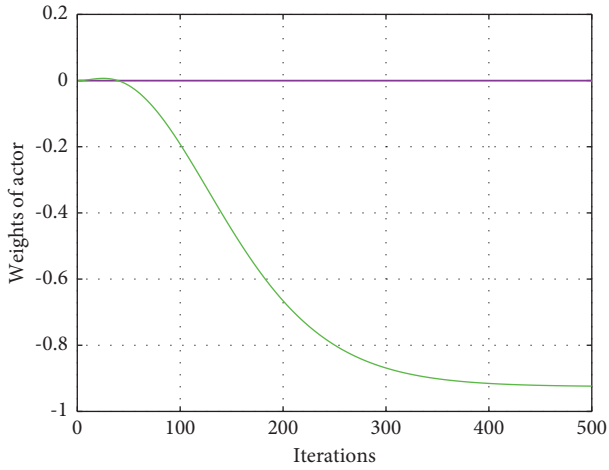

Figure 3: Weights of the critic network.

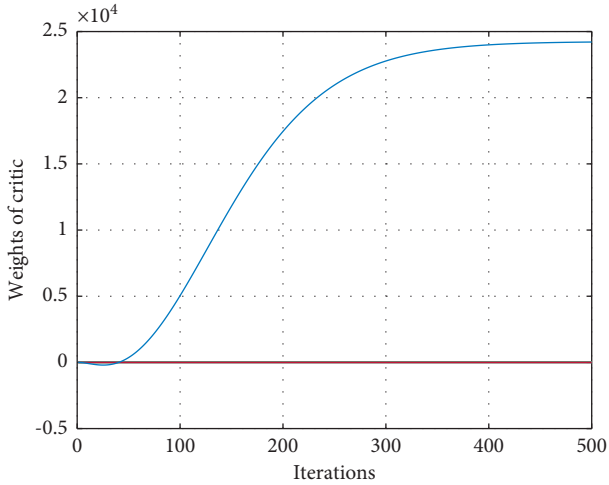

Figure 4: Weights of the actor network.

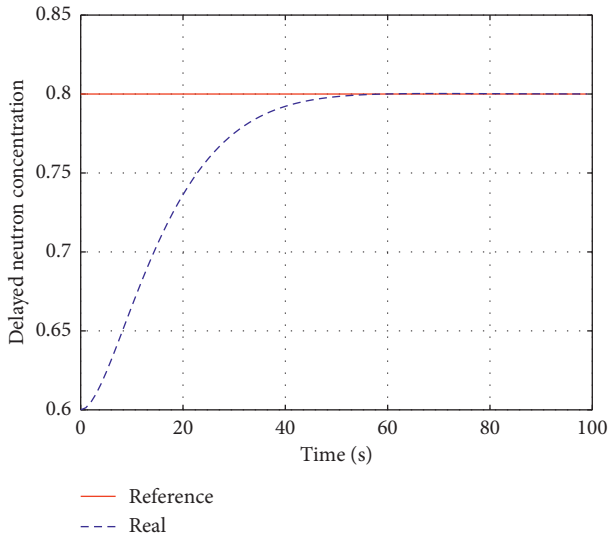

Figure 5: Power-level load-varying curve of this PWR power plant.

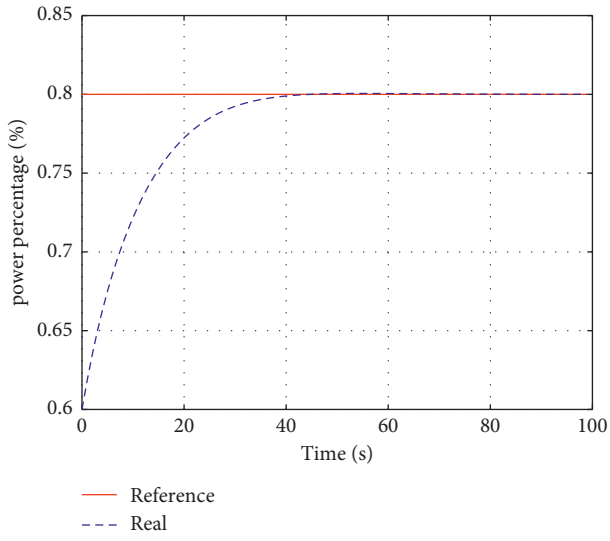

Figure 6: Delay neutron density relative curve of this PWR power plant.

$$d(\varepsilon_k, \varepsilon_{k+1}, W_{ui}, W_{Vi}) = \varepsilon_k^{\mathrm{T}} Q \varepsilon_k + \hat{u}_i^{\mathrm{T}}(\varepsilon_k) R \hat{u}_i(\varepsilon_k) + \hat{V}_i(\varepsilon_{k+1}),$$
$$= \varepsilon_k^{\mathrm{T}} Q \varepsilon_k + \hat{u}_i^{\mathrm{T}}(\varepsilon_k) R \hat{u}_i(\varepsilon_k),$$
$$+ \hat{V}_i(\varepsilon_{k+1}). \tag{35}$$

Similarly, the actor NN is applied to evaluate and approximate the optimal tracking control policy. We will tune the weights of action NN to solve (20) at each iteration of this VI algorithm. According to $\hat{u}_i(\varepsilon_k, W_{ui})$, from (33), we can rewrite (14) as

$$W_{ui} = \arg\min_{\hat{u}_i} \varepsilon_k^{\mathrm{T}} Q \varepsilon_k + \hat{u}_i^{\mathrm{T}}(\varepsilon_k, \delta) R \hat{u}_i(\varepsilon_k, \delta) + \hat{V}_i(\varepsilon_{k+1}), \tag{36}$$

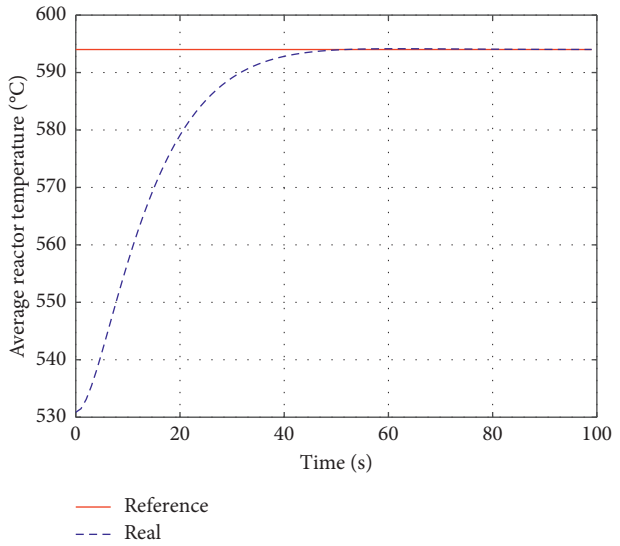

Figure 7: Average temperature of the reactor core of this PWR power plant.

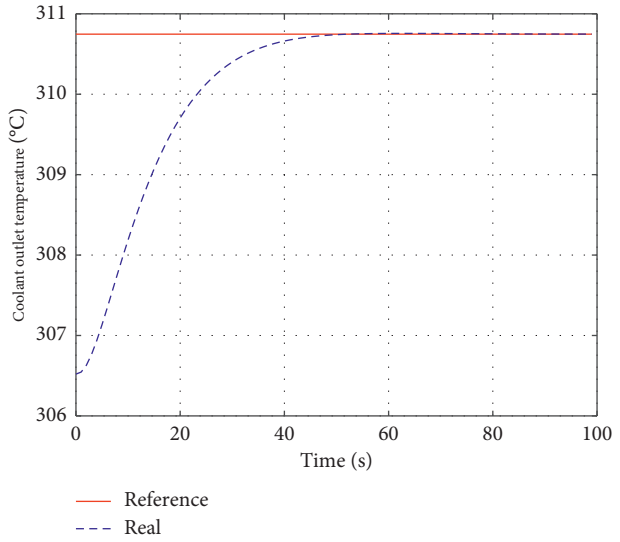

Figure 8: Average temperature of the coolant outlet of this PWR power plant.

where $\varepsilon_{k+1} = f_\varepsilon(\varepsilon_k) + g_\varepsilon(\varepsilon_k)\widehat{u}_i(\varepsilon_k, \delta)$ and δ are updated by the same method as the weights of the critic NN. For getting the approximation of the weights of actor critic NNs, the least square (LS) method is utilized to solve the weights of NNs.

4. Numerical Simulations

In this section, numerical results are given to demonstrate the validity of this value iteration optimal tracking method. Experimental simulations of the performance index and weights of actor critic NNs are provided. This developed method is an offline policy with an initial random control policy.

4.1. Actor Critic NN's Implementation of the Value Iteration Algorithm. It is generally known that NNs can be leveraged

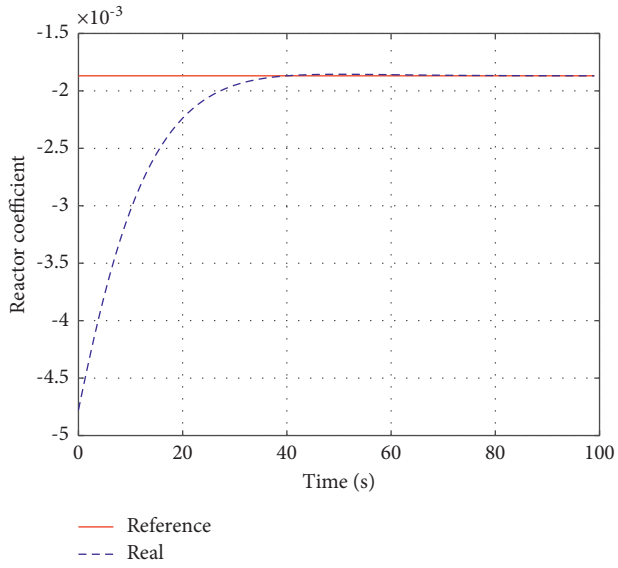

Figure 9: Reactor coefficient of the control rod of this PWR power plant.

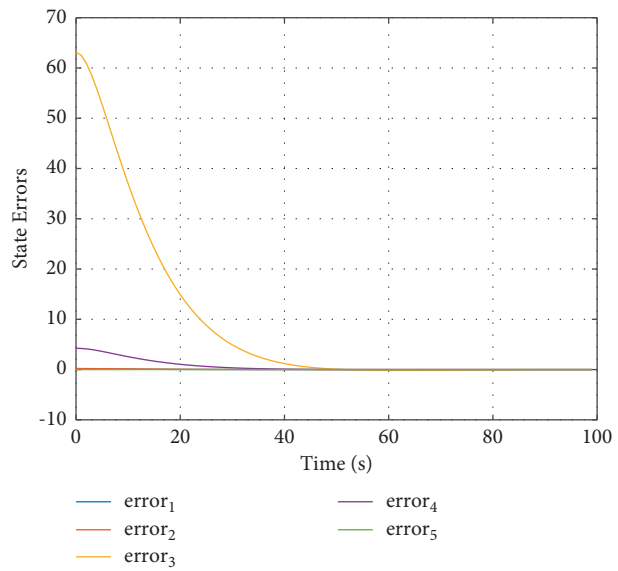

Figure 10: Five tracking errors of this PWR power system.

to approximate any functions on prescribed compact sets. We choose an error compact set to train this actor critic NN to obtain an offline tracking policy. As predefined in Section 3, the critic NN is approximated as $\Im(\varepsilon) = W_{cL}\phi(e)$ with 15 neurons $(L = 1, 2, \ldots, 15)$, and the weights are $W_{cL} = [W_{c1}, W_{c2}, \ldots, W_{c15}]$. The actor NN is chosen with 5 neurons, and the weights are $W_{aL} = [W_{a1}, W_{a2}, W_{a3}, W_{a4}, W_{a5}]$.

At the beginning of this algorithm execution, generally, the matrices are $Q = 10I_{5\times5}$ and $R = 0.1I_{1\times1}$, where I is the identity matrix. The tracking error compact sets are randomly set as the difference between the initial state and the desired trajectory. We chose the control period as 1s

per iteration, and the total iteration epoch was 50. Simultaneously, we chose the %60 initial working point with $x_{60} = [0.6, 0.6, 530.8572, 306.5198, -0.0047]^T$ as state inputs and the %80 power working point with $x_{80} = [0.8, 0.8, 594.0060, 310.7468, -0.0018]^T$ as the desired tracking trajectory. Under the proposed value iteration algorithm, the tracking error performance index value, as shown in Figure 2, is monotonically nondecreasing and converges to $\mathfrak{I}_\varepsilon^*$, which is identical to our analysis results above.

Additionally, when the tracking error performance index function converges, the training process of the actor-critic NN will stop. As shown in Figure 3 and 4, the weights of the actor-critic NN converge to a steady solution, which implies that a perfect approximator of the optimal tracking controller is obtained.

4.2. Application on the PWR Nuclear Power System. In this section, we apply the calculated control law on this DT nonlinear PWR power model, and the implementation results are shown in Figure 5–10.

Generally, the optimal tracking control of PWR power plants focuses on power-level adjustment, but there is high interference among different states in this nuclear power model. Thus, it is necessary to track all states in the PWR power model, and the tracking objects should guarantee the stability of each state and safety of nuclear plants.

As shown in Figures 5–9, we give a %20 step increase signal to this PWR power system. These 5 figures demonstrate that the 5 states catch the desired trajectory in less than 50 *s*. We use a 5th-order DT nonlinear PWR power system in this study, which implies that Xenon poison is without consideration. Although the PWR power model is quite simplified, all states of this model are difficult to track.

As shown in Figure 5, the power level tracks the desired states without overshoot and oscillation. The average temperature of the reactor core and coolant outlet approximate the desired temperature; compared with the reference temperature, the maximum deviation is 0.031°C and 0.002°C, respectively. In addition, the reactor coefficient of the control rod tracks the desired curve in less than 40 *s*. Based the aforementioned results, we also can see that our tracking method applied this nuclear system has no steady-state error and less regulation time. As shown in Figure 10, the tracking errors progressively converge to 0 with the proposal of this value iteration optimal tracking method.

5. Conclusion

It is well known that optimal tracking power-level control for DT nonlinear nuclear power systems is crucial for both regular operation and safety problems, and manual control is inefficient. However, the intrinsic nonlinearity and parameters that vary with the states cause difficulties in power-level control, and there are tracking issues.

In this study, a value iteration-based actor critic NN algorithm is designed to obtain an optimal tracking control policy for the DT nonlinear nuclear power plant. The proposed algorithm performs well in tracking states, as shown in the simulation results, and can also swiftly calculate the optimal control law. Thus, we formulate the tracking control problem as HJB equations to solve it.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Open Project Program of State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment (no. K-A2020.406), National Key R&D Program of China (no. 2021YFE0206100), National Natural Science Foundation of China (no. 62073321), National Defense Basic Scientific Research Program (no. JCKY2019203C029), and Science andTechnology Development Fund, Macau SAR (0015/2020/AMJ).

References

[1] J. I. Choi, J. E. Meyer, and D. D. Lanning, "Automatic controller for steam generator water level during low power operation," *Nuclear Engineering and Design*, vol. 117, no. 3, pp. 263–274, 1989.

[2] Kim, L. Meyer, and Bernard, "Design and Evaluation of Model-Based Compensators for the Control of Steam Generator Level," in *Proceedings of the American Control Conference*, pp. 2055–2060, San Francisco, CA, USA, 1993.

[3] T. Yun, H. Su-xia, L. Chong, and Z. Fu-yu, "An improved implicit multiple model predictive control used for movable nuclear power plant," *Nuclear Engineering and Design*, vol. 240, no. 10, pp. 3582–3585, 2010.

[4] X. Liu and M. Wang, "Nonlinear fuzzy model predictive control for a pwr nuclear power plant," *Mathematical Problems in Engineering*, vol. 2014, no. 2, pp. 1–10, Article ID 908526, 2014.

[5] Y. Zhang, Q. Li, W. Zhang, and Y. Yang, "Survey of multimodel adaptive control theory and its applications," *Chinese Journal of Engineering*, vol. 42, no. 2, pp. 135–143, 2020.

[6] L. Wutainhao and Wangjunling, "Fuzzy generalized predictive control of the nuclear reactor power," *Nuclear Science & Engineering*, vol. 036, no. 003, pp. 299–305, 2016.

[7] H. Eliasi, H. Davilu, and M. B. Menhaj, "Adaptive fuzzy model based predictive control of nuclear steam generators," *Nuclear Engineering and Design*, vol. 237, no. 6, pp. 668–676, 2007.

[8] Z. Gang, C. Xin, and Y. Weicheng, "Identification of dynamics for nuclear steam generator water level process using rbf neural networks," in *Proceedings of the 2007 8th International Conference on Electronic Measurement and Instruments*, Xian, China, Auguest 2007.

[9] Li Wangjunling and Luanxiuchun, "Load-following adaptive guaranteed cost control of pressurized water reactors," *Control Theory & Applications*, vol. 34, no. 09, pp. 105–110, 2017.

[10] P. Werbos, "Neural networks for control and system identification," in *Proceedings of the IEEE Conference on Decision and Control*, Tampa, USA, December 1989.

[11] Sutton and S. Richard, *Neural Networks for Control*, MIT Press, Cambridge, USA, 1990.

[12] S. Leven and J Wiley, "The roots of backpropagation: from ordered derivatives to neural networks and political forecasting," *Neural Networks*, vol. 9, no. 3, pp. 543-544, 1996.

[13] S. Cao, L. Sun, J. Jiang, and Z. Zuo, "Reinforcement learning-based fixed-time trajectory tracking control for uncertain robotic manipulators with input saturation," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2021, in Press.

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

[15] H. Huaguang Zhang, R. Ruizhuo Song, Q. Qinglai Wei, and T. Tieyan Zhang, "Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 1851–1862, 2011.

[16] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont, USA, 2005.

[17] A. Al-Tamimi and F. Lewis, "Discrete-time nonlinear hjb solution using approximate dynamic programming: convergence proof," in *Approximate Dynamic Programming and Reinforcement Learning*vol. 38, no. 4, IEEE International Symposium on, 2008.

[18] Q. Wei, D. Liu, and H. Lin, "Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 840–853, 2016.

[19] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 3051–3056, 2014.

[20] H. Modares, F. L. Lewis, and Z.-P. Jiang, "$ {H}_{ {\infty }}$ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2550–2562, 2015.

[21] X. Yang, D. Liu, Q. Wei, and D. Wang, "Guaranteed cost neural tracking control for a class of uncertain nonlinear systems using adaptive dynamic programming," *Neurocomputing*, vol. 198, pp. 80–90, 2016.

[22] Q. Wei and D. Liu, "Adaptive dynamic programming for optimal tracking control of unknown nonlinear systems with application to coal gasification," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 4, pp. 1020–1036, 2014.

[23] J. Lu, Q. Wei, and F. Y. Wang, "Parallel control for optimal tracking via adaptive dynamic programming," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 6, pp. 1662–1674, 2020.

[24] J. Lu, Q. Wei, Z. Wang, T. Zhou, and F.-Y. Wang, "Event-triggered optimal control for discrete-time multi-player non-zero-sum games using parallel control," *Information Sciences*, vol. 584, pp. 519–535, 2022.

[25] T. Sardarmehni and X. Song, "Sub-optimal tracking in switched systems with fixed final time and fixed mode sequence using reinforcement learning," *Neurocomputing*, vol. 420, pp. 197–209, 2021.

[26] T. Lala, D. P. Chirla, and M. B. Radac, "Model reference tracking control solutions for a visual servo system based on a virtual state from unknown dynamics," *Energies*, vol. 15, no. 1, p. 267, 2021.

[27] A. Ben Abdennour, R. M. Edwards, and K. Y. Lee, "Lqg/ltr robust control of nuclear reactors with improved temperature performance," *IEEE Transactions on Nuclear Science*, vol. 6, no. 39, p. 9, 1992.

[28] J. Wan and F. Zhao, "Design of a two-degree-of-freedom controller for nuclear reactor power control of pressurized water reactor," *Annals of Nuclear Energy*, vol. 144, Article ID 107583, 2020.

[29] J. Wan, P. Wang, S. Wu, and F. Zhao, "Conventional controller design for the reactor power control system of the advanced small pressurized water reactor," *Nuclear Technology*, vol. 198, no. 1, pp. 26–42, 2017.

[30] E. Hatami, N. Vosoughi, and H. Salarieh, "Design of a fault tolerated intelligent control system for load following operation in a nuclear power plant," *International Journal of Electrical Power & Energy Systems*, vol. 78, no. Jun, pp. 864–872, 2016.

[31] B. Kiumarsi and F. L. Lewis, "Actor-critic-based optimal tracking for partially unknown nonlinear discrete-time systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 1, pp. 140–151, 2015.

[32] D. Vrabie and F. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks*, vol. 22, no. 3, pp. 237–246, 2009.

[33] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica*, vol. 50, no. 7, pp. 1780–1792, 2014.

[34] 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.