Chapter 2
Neural Networks-Based Immune Optimization Regulation Using Adaptive Dynamic Programming

2.1 Introduction

In the fight against cancer, there had been no effective measures before chemotherapy and radiation appeared since there only exist tiny differences between cancer cells and normal cells. Doctors operate to remove solid tumors that have not yet spread, which can not guarantee cancer from recurring. When radiotherapy and chemotherapy have increased side effects, and targeted therapy is not flexible because of its strong pertinence, the scientific research direction began to turn to the human body system. Generally, tumor cells escape from the immune system, not because it fails for the immune system to recognize them or it is not activated, but cancer cells have evolved a way to block the activation of T cells by making a specific binding. Thus, the medical communities have struggled to find a lot of special means for cancer cells to intercept the activation of the T cells, freeing up the immune system. Compared with traditional treatments such as surgery, radiation and chemotherapy, immunotherapy has fewer side effects and better therapeutic effects. However, it is difficult to tackle the transient period of immune agents. Therefore, the hybrid therapy of chemotherapy and immunotherapy is a better choice. As [1], it is hardly sufficient to control tumor growth through neither chemotherapy nor immunotherapy alone, but tumor cells can be eradicated by adopting the combination therapies which is known as biochemotherapy described in [2].

With extensive development of nonlinear dynamic [3, 4], its engineering application scenarios enjoy increasing diversification such as competitive Nash equilibrium problems, especially in the biomedical field. And not coincidentally, game theory has been introduced into the interaction model of tumor cells and immune cells. Both of the chemotherapy and immunotherapy aim at reducing the number of tumor cells. Based on this fact, the collaborative game is formed and one can design adaptive therapy from the view of game theory. Multiple biological interactions constitute complex nonlinear growth process of tumor cells, however, regarding major influence factors of tumor cell populations as research object is the focus. Hunting cells refer to the immune cells participating in removing foreign agents and strengthening

the immune response. temperatures have suggested that cell-mediated anti-tumor immunity contributed to increasing the population of hunting tumor cells to maintain a specific proportion between the resting and the hunting predator cells as 40% in literature [5], which was beneficial for maintenance of the tumor dormant state. The immune regulations vary from individual to individual, but immunotherapy-based optimal regulation plays the role of reducing tumor cells without considering certain circumstances in case of special invocation. Enhanced tumor antigen presentation could effectively stimulate dendritic cells and increase the immunotherapy-based curative effect in [6]. The known "predator-prey" between immune cells and tumor cells leads to cyclic growth and reduction, which can be continue indefinitely or reach an equilibrium saddle point determined by system parameters. Literature [7] investigated nonlinear dynamical model which provided guiding significance for introducing that into cybernetics. As known, system identification or optimal control is of great practical value. As a powerful and effective optimization algorithm, the ADP method can solve the nonlinear optimal control problems well, realizing the most appropriate therapeutic strategy.

Of course, the immune system has the responsibility for restraining tumor growth, but it is hardly to fight out the tumor cells alone. Firstly, ego characteristic of tumor cells compared to normal cells within the body leads to no exclusion and tolerance to tumor cells of the immune system. Secondly, there is no strong defense mechanism itself in fighting with the cancer cells which means the failure of the immune response. Finally, Immune function was observed to be protective through intervention with organic binding agents of CD4 and CD8 cells. Chemotherapy can not only rapidly kill differentiated tumor cells, but also destroy regular cells. This side effect caused by chemotherapy can be lessened through introducing the immunotherapy. Thus the combined therapy of chemotherapy and immunotherapy is more reasonable. Immunotherapies can strengthen the immune system through extra stimulation, on the other hand, improve the ability to recognize foreign entity. Therefore, decelerating the growth rate of tumor cells with minimized dose of chemotherapy and immunotherapeutic drugs is the control objective. Furthermore, optimal control strategy is obtained through ADP method, giving the optimal levels of each treatment regimen through nonzero-sum differential games strategy developed in [8].

Prescribed performance tracking control has been creatively developed as [9], however, there is seldom any literatures focusing on this scope considers mutual relationship among tumor cells, immune cells, chemotherapy and immunotherapy drugs, let alone setting the performance as eventually acquired of optimal therapeutic effect associated with coupling behaviors mentioned above. Retrospect to literatures as [10], the chapter transformed it into multi-player nonzero-sum games problems whose optimal control was obtained by complex decoupling in dealing with Hamilton-Jacobi equation as [11]. Subsequently, online adaptive and off-policy learning algorithms were respectively developed in [12–14]. Of course, the constrained-input was taken into consideration, when it comes to practical applications in [15], even more intensive work on uncertain constraints were in contemplation considered as [16]. As [17], the control policies of the distributed subsystems acted as players, noticeably, the chapter was formulated as a two-players nonzero-sum game including chemotherapy

and immunotherapy. [18] first introduced an updating strategy based on intertask relationships. Synchronously, reciprocal action between the tumor cells and immune cells which could be analogous to interactions between systems in [19, 20].

The unknown nonlinear dynamic is usually implemented by fuzzy control as [21, 22] and neural networks in [18, 23], where the actor network and critic network are adopted for updating control policy at an appropriate time through policy iteration technique as [24–26]. The convergence of model-based policy iteration algorithm is equivalent to that of data-based learning as [27]. Similarly, states of the system and critic error are required to be ultimately uniformly bounded during the process of value iteration, which is guaranteed through event-triggered formation control scheme firstly proposed for all signals of the closed-loop system in literature [28]. According to the iterative value algorithm, the optimum can be obtained through learning continuously [29, 30]. However there is little research on the two-players nonzero-sum game considering tumor cells and immune cells using the proposed value iteration learning.

2.2 Preliminaries

As is known, there exist interaction relationships among the anticancer agent cells, lymphocytes and macrophages that constitute the basic immune system microenvironment, which can be presented as follows. Firstly, T-lymphocytes and cytotoxic macrophages/natural killer cells can effectively damage tumor cells. Secondly, destroyed behaviour of macrophages can also active T-lymphocytes for launching another attack. Meanwhile, the population of T-lymphocytes can be fed through resting cells. Finally, the model is guided by degradation of resting cells and activation of immune cells by natural growth rate. This section gives the nonlinear growth equation which can represent the whole immune response.

$$N_{total} = \frac{\upsilon N_H(t) N_T(t)}{\nu + N_T(t)} \tag{2.1}$$

where $N_H(t)$, $N_T(t)$ denote the number of hunting cells and tumor cells at time t, respectively. υ and ν are positive constants. The changes in quantity caused by the inactivation of the immune cells and the apoptosis of tumor cells are presented as:

$$\frac{dN_T(t)}{dt} = -\sigma_1 N_H(t) N_T(t)$$
$$\frac{dN_H(t)}{dt} = -\sigma_2 N_H(t) N_T(t) \tag{2.2}$$

where σ_1 denotes the loss rate of $N_T(t)$ caused by $N_H(t)$ and σ_2 represents the loss rate of $N_H(t)$ caused by $N_T(t)$. The situations above reflect the competition between tumor cells and the host cells. Then we construct the dynamic equations as follows

Table 2.1 Detailed descriptions of system parameters

Parameter	Estimated value
ι_1	Intrinsic growth rate of $N_T(t)$ ignoring $N_{CD}(t)$
ϱ_1	Reciprocal carrying capacity of $N_T(t)$ irrespective of $N_T(t)$ and $N_{CD}(t)$
$\hbar$	Constant influx rate of $N_H(t)$
σ_1	Rate of loss of $N_T(t)$ for $N_H(t)$
δ_1	Response coefficient to $N_{CD}(t)$ for $N_T(t)$
$\mathfrak{D}$	Per capita decay rate of $N_H(t)$ without regard to $N_T(t)$, $N_{CD}(t)$ and $N_{ID}(t)$
σ_2	Rate of loss of $N_H(t)$ for $N_T(t)$
δ_2	Response coefficient to $N_{CD}(t)$ for $N_H(t)$
υ	Maximum recruitment rate of $N_H(t)$ by ligand-transduced $N_T(t)$
ς	Maximum recruitment rate of $N_H(t)$ by $N_{ID}(t)$
ν	Steepness coefficient of $N_H(t)$ by $N_T(t)$
ϑ	Steepness coefficient of $N_H(t)$ by $N_{ID}(t)$
φ_1	Decay rate of $N_{CD}(t)$
φ_2	Decay rate of $N_{ID}(t)$

$$
\begin{aligned}
\dot{N}_T(t) =\ & \iota_1 N_T(t)(1 - \varrho_1 N_T(t)) - \sigma_1 N_T(t) N_H(t) \\
& - \delta_1 N_{CD}(t) N_T(t) \\
\dot{N}_H(t) =\ & \frac{\upsilon N_H(t) N_T^2(t)}{\nu + N_T^2(t)} + \frac{\varsigma N_H(t) N_{ID}(t)}{\vartheta + N_{ID}(t)} - \sigma_2 N_T(t) N_H(t) \\
& - \mathfrak{D} N_H(t) - \delta_2 N_{CD}(t) N_H(t)
\end{aligned}
\tag{2.3}
$$

where $\mathfrak{D}$ represents the death rate of cells without considering any tumor cells. ι_α ($\alpha = 1, 2$) and ϱ_α denote the per capita growth rates and reciprocal carrying capacities. The descriptions of the other associated parameters are given in Table 2.1.

Consider the given chemotherapy and immunotherapy drugs as $u(t)$ and $v(t)$ at time t, which is regarded as multiple dose administration compared with influence of recombinant human interleukin-11 for injection or recombinant human granulocyte colony-stimulating factor injection. Assume that targeted therapy cannot be achieved through only chemotherapeutic drugs. Then we can obtain that

$$
f_{response}(t) = s_\alpha (1 - e^{-\lambda u(t)})
\tag{2.4}
$$

where s_α is the different response coefficients for distinguishing the change rate of different cells. The mathematical model considering injected drugs is presented as

$$\dot{N}_{CD}(t) = u(t) - \varphi_1 N_{CD}(t)$$

$$\dot{N}_{ID}(t) = v(t) - \varphi_2 N_{ID}(t)$$

$$\dot{N}_T(t) = \iota_1 N_T(t)(1 - \varrho_1 N_T(t)) - \sigma_1 N_T(t) N_H(t)$$
$$- \delta_1 N_{CD}(t) N_T(t) - s_2(1 - e^{-\lambda u(t)})$$

$$\dot{N}_H(t) = \frac{\upsilon N_H(t) N_T^2(t)}{\nu + N_T^2(t)} + \frac{\varsigma N_H(t) N_{ID}(t)}{\vartheta + N_{ID}(t)} - \sigma_2 N_T(t) N_H(t) - \mathfrak{D} N_H(t)$$
$$- \delta_2 N_{CD}(t) N_H(t) - s_1(1 - e^{-\lambda u(t)}) \tag{2.5}$$

where $N_{CD}(t)$ and $N_{ID}(t)$ are concentrations of chemotherapy and immunotherapy. $v(t)$ and $u(t)$ are the doses of chemotherapeutic drug and immunotherapeutic drug. Generally speaking, λ is taken as 1 for the unknown role of cytokines.

Remark 2.1 The model (2.5) describes the relations among the hunting cells, the tumor cells, the concentration of chemotherapy agentia, and the concentration of immunotherapy agentia. From (2.5) we can find both of the hunting cells and the chemotherapy agentia can reduce the number of tumor cells, and the immunotherapy agentia can stimulate the growth of hunting cells. On the other hand, the tumor cells can influence the number of hunting cells. Based on this complicated interactive relationship, we can obtain the optimal object through ADP, that is, minimization of tumor cells while ensuring the number of normal cells at certain time t.

Before proceeding, let $X = [N_T, N_H, N_{CD}, N_{ID}]^T$, then the model (2.5) can be simplified as

$$\dot{X}(t) = f(X) + g(X)u(t) + \kappa(X)v(t) \tag{2.6}$$

where $f(X)$ is the right-hand dynamics of (2.5) excluding the control $u(t)$ and $v(t)$. The matrixes $g(X) = [0, 0, 1, 0]^T$ and $\kappa(X) = [0, 0, 0, 1]^T$.

For system (2.6), the performance index function of the ϵ player can be given as

$$J_\epsilon(X_0) = \int_0^\infty \left(X^T \mathcal{Q}_\epsilon X + u^T \mathcal{R}_{\epsilon 1} u + v^T \mathcal{R}_{\epsilon 2} v \right) d\tau \tag{2.7}$$

where $\mathcal{Q}_\epsilon$ is positive definite matrix, $\mathcal{R}_{\epsilon 1}$ and $\mathcal{R}_{\epsilon 2}$ are symmetric positive matrixes. The corresponding cost functions are presented as:

$$\mathcal{V}_\epsilon(X, u, v) = \int_t^\infty \mathfrak{R}_\epsilon(X, u, v) d\tau \tag{2.8}$$

with the utility function

$$\mathfrak{R}_\epsilon(X, u, v) = X^T \mathcal{Q}_\epsilon X + u^T \mathcal{R}_{\epsilon 1} u + v^T \mathcal{R}_{\epsilon 2} v. \tag{2.9}$$

Definition 2.2 For two-player NZS game of system (2.6), the Nash equilibrium solution is said to be obtained with the control pair (u^*, v^*) which satisfied that,

$$\begin{aligned} \mathcal{V}_\epsilon(u^*, v^*) &\leq \mathcal{V}_\epsilon(u, v^*) \\ \mathcal{V}_\epsilon(u^*, v^*) &\leq \mathcal{V}_\epsilon(u^*, v) \end{aligned} \tag{2.10}$$

for any admissible control policies u and v.

The Hamilton functions can be constructed as:

$$\begin{aligned} \mathrm{H}_\epsilon(X, u, v) = {}& X^T \mathcal{Q}_\epsilon X + u^T \mathcal{R}_{\epsilon 1} u + v^T \mathcal{R}_{\epsilon 2} v \\ & + \nabla \mathcal{V}_\epsilon^T (f(X) + g(X)u(t) + \kappa(X)v(t)) \end{aligned} \tag{2.11}$$

where $\nabla \mathcal{V}_\epsilon$ is the partial derivative of the cost function and $\epsilon = 1, 2$. According to the stationarity conditions at equilibrium points, the optimal control for two players are obtained

$$\begin{aligned} u^* &= -\frac{1}{2} \mathcal{R}_{11}^{-1} g^T(X) \nabla \mathcal{V}_1^* \\ v^* &= -\frac{1}{2} \mathcal{R}_{22}^{-1} \kappa^T(X) \nabla \mathcal{V}_2^* \end{aligned} \tag{2.12}$$

with $\mathcal{V}_1^*$ and $\mathcal{V}_2^*$ being the solutions of coupled HJ equations as

$$\begin{aligned} X^T \mathcal{Q}_1 X &- \frac{1}{4} \nabla \mathcal{V}_1^{*T} g(X) \mathcal{R}_{11}^{-1} g^T(X) \nabla \mathcal{V}_1^* + \nabla \mathcal{V}_1^{*T} f(X) \\ &+ \frac{1}{4} \nabla \mathcal{V}_2^{*T} \kappa(X) \mathcal{R}_{22}^{-1} \mathcal{R}_{12} \mathcal{R}_{22}^{-1} \kappa^T(X) \nabla \mathcal{V}_2^* \\ &- \frac{1}{2} \nabla \mathcal{V}_1^{*T} \kappa(X) \mathcal{R}_{22}^{-1} \kappa^T(X) \nabla \mathcal{V}_2^* = 0, \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} X^T \mathcal{Q}_2 X &- \frac{1}{4} \nabla \mathcal{V}_2^{*T} \kappa(X) \mathcal{R}_{22}^{-1} \kappa^T(X) \nabla \mathcal{V}_2^* + \nabla \mathcal{V}_2^{*T} f(X) \\ &+ \frac{1}{4} \nabla \mathcal{V}_1^{*T} g(X) \mathcal{R}_{11}^{-1} \mathcal{R}_{21} \mathcal{R}_{11}^{-1} g^T(X) \nabla \mathcal{V}_1^* \\ &- \frac{1}{2} \nabla \mathcal{V}_2^{*T} g(X) \mathcal{R}_{11}^{-1} g^T(X) \nabla \mathcal{V}_1^* = 0. \end{aligned} \tag{2.14}$$

Lemma 2.3 *For nonlinear system (2.6), suppose that $\mathcal{V}_1^*$ and $\mathcal{V}_2^*$ satisfy the equations (2.13) and (2.14). Then under the optimal control (2.12), the system is asymptotically stable.*

Proof The proof is omitted since it is similar to that in [31, 32].

By solving the coupled HJ equations (2.13) and (2.14), one can obtain the optimal control as (2.12), which means the Nash equilibrium for the two-player NZS game system is attained. Nevertheless, due to the existence of nonlinear terms and coupled terms, these partial differential equations are uneasy to solve. Since ADP is a powerful approximate learning method, the approximate solutions of (2.13) and (2.14) can be acquired.

2.3 Design of Adaptive Controller

In order to find the optimal control strategy, a critic network is constructed based on neural network firstly. And then optimal value function can be shown as:

$$V_\epsilon^* = (\zeta_\epsilon^*)^T \xi_\epsilon(X) + o_\epsilon, \epsilon = 1, 2, \tag{2.15}$$

where $\zeta_\epsilon^* \in R^{p_\epsilon}$, $\xi_\epsilon \in R^{p_\epsilon}$ and $o_\epsilon \in R$ are the ideal weight vector, activation function and approximation error of the neural network. As it's scarcely possible to get the weight ζ_ϵ^*, we give the approximate version

$$\hat{V}_\epsilon^* = (\hat{\zeta}_\epsilon)^T \xi_\epsilon(X). \tag{2.16}$$

Based on (2.12) and (2.15), we obtain the optimal control as

$$u^* = -\frac{1}{2}\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)((\nabla\xi_1(X))^T\zeta_1^* + \nabla o_1)$$
$$v^* = -\frac{1}{2}\mathcal{R}_{22}^{-1}\kappa^T(X)((\nabla\xi_2(X))^T\zeta_2^* + \nabla o_2) \tag{2.17}$$

Then we further get the approximate control policies as

$$\hat{u} = -\frac{1}{2}\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)(\nabla\xi_1(X)^T\hat{\zeta}_1$$
$$\hat{v} = -\frac{1}{2}\mathcal{R}_{22}^{-1}\kappa^T(X)(\nabla\xi_2(X)^T\hat{\zeta}_2 \tag{2.18}$$

Remark 2.4 For the unknowable nature of ideal weights, the NNs are used to approximate the system dynamics and approximate version as (2.16), aming at minimizing the current estimate of the value functions in (2.15) by selecting policies (2.18) can be obtained with available closed-form expressions.

According to (2.18), the closed-loop system can be rewritten as

$$\dot{X}(t) = f(X) + \boldsymbol{g}(X)\hat{u} + \kappa(X)\hat{v}. \tag{2.19}$$

Furthermore, we can attain the approximate Hamilton as

$$
\begin{aligned}
H_\epsilon(X, \hat{u}, \hat{v}) &= X^T \mathcal{Q}_\epsilon X + \hat{u}^T \mathcal{R}_{\epsilon 1} \hat{u} \\
&\quad + \hat{v}^T \mathcal{R}_{\epsilon 2} \hat{v} + (\hat{\zeta}_\epsilon)^T \nabla \xi_\epsilon(X) \dot{X}(t) \\
&= e_\epsilon(t).
\end{aligned}
\tag{2.20}
$$

To approach the optimal strategy and minimize $e_\epsilon(t)$, the goal of adaptive learning is set to be $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 = 1/2 e_1^2 + 1/2 e_2^2$. Then applying the gradient descent method, we obtain the learning law of critic for player ϵ

$$
\dot{\hat{\zeta}}_\epsilon = -\varrho_\epsilon \frac{1}{(\delta_\epsilon^T \delta_\epsilon + 1)^2} \frac{\partial \mathcal{E}(t)}{\partial \hat{\zeta}_\epsilon} = -\varrho_\epsilon \frac{1}{(\delta_\epsilon^T \delta_\epsilon + 1)^2} \frac{\partial \mathcal{E}_\epsilon(t)}{\partial \hat{\zeta}_\epsilon} = -\varrho_\epsilon \frac{\delta_\epsilon e_\epsilon(t)}{(\delta_\epsilon^T \delta_\epsilon + 1)^2}
\tag{2.21}
$$

where $\delta_\epsilon = \nabla \xi_\epsilon(X) \dot{X}(t)$, and ϱ_ϵ is the positive learning law. Let $\tilde{\zeta}_\epsilon = \zeta_\epsilon^* - \hat{\zeta}_\epsilon$, then we have

$$
\dot{\tilde{\zeta}}_\epsilon = \varrho_\epsilon \frac{\delta_\epsilon \sigma_\epsilon(t)}{(\delta_\epsilon^T \delta_\epsilon + 1)^2} - \varrho_\epsilon \frac{\delta_\epsilon \delta_\epsilon^T \tilde{\zeta}_\epsilon}{(\delta_\epsilon^T \delta_\epsilon + 1)^2} = \varrho_\epsilon \underline{\delta}_\epsilon \sigma_\epsilon(t) - \varrho_\epsilon \bar{\delta}_\epsilon \bar{\delta}_\epsilon^T \tilde{\zeta}_\epsilon,
\tag{2.22}
$$

where $\underline{\delta}_\epsilon = \delta_\epsilon / (\delta_\epsilon^T \delta_\epsilon + 1)^2$, $\bar{\delta}_\epsilon = \delta_\epsilon / (\delta_\epsilon^T \delta_\epsilon + 1)$ and $\sigma_\epsilon(t) = -\nabla o_\epsilon^T(X)(f(X) + g(X)\hat{u} + \kappa(X)\hat{v})$ is the approximate residual error when employing critic neural network [33].

Before presenting the main results of this chapter, two regular assumptions are necessary [34–36].

Assumption 2.1 For $\epsilon = 1, 2$, the signal $\bar{\delta}_\epsilon$ is persistently excited such that the following inequality is satisfied

$$
\varsigma_\epsilon I_{\nu_\epsilon \times \nu_\epsilon} \le \int_t^{t+T} \bar{\delta}_\epsilon \bar{\delta}_\epsilon^T d\varepsilon,
\tag{2.23}
$$

where ν_ϵ denotes the neuro number of the ϵth critic network.

Assumption 2.2 For $\epsilon = 1, 2$, there exist positive constants $\xi_{\epsilon max}$, $o_{\epsilon max}$ and $\sigma_{\epsilon max}$ such that the following inequalities hold, that is, $\|\nabla \xi_\epsilon(X)\| \le \xi_{\epsilon max}$, $\|\nabla o\| \le o_{\epsilon max}$ and $\|\sigma_\epsilon\| \le \sigma_{\epsilon max}$.

Applying the Lyapunov method, the stability in the sense of UUB is demonstrated to be guaranteed by the following theorem.

Theorem 2.5 *For system (2.6), when the weight updating laws of critic networks are given by (2.21), then the UUB properties of the weight estimation error $\tilde{\zeta}_\epsilon$ can be guaranteed by the obtained control policies (2.18).*

Proof Select the Lyapunov function as

$$\mathcal{L} = \frac{1}{2}\varrho_1^{-1}\tilde{\zeta}_1^T\tilde{\zeta}_1^T + \frac{1}{2}\varrho_2^{-1}\tilde{\zeta}_2^T\tilde{\zeta}_2^T. \tag{2.24}$$

Taking the time derivative of (2.24), then we obtain

$$\begin{aligned}
\dot{\mathcal{L}} &= \varrho_1^{-1}\tilde{\zeta}_1^T\dot{\tilde{\zeta}}_1 + \varrho_2^{-1}\tilde{\zeta}_2^T\dot{\tilde{\zeta}}_2 \\
&= \tilde{\zeta}_1^T\left(\underline{\delta}_1\sigma_1(t) - \bar{\delta}_1\bar{\delta}_1^T\tilde{\zeta}_1\right) + \tilde{\zeta}_2^T\left(\underline{\delta}_2\sigma_2(t) - \bar{\delta}_2\bar{\delta}_2^T\tilde{\zeta}_2\right)
\end{aligned} \tag{2.25}$$

According to Young's inequality, we have

$$\tilde{\zeta}_1^T\underline{\delta}_1\sigma_1(t) \le \tilde{\zeta}_1^T\bar{\delta}_1\sigma_1(t) \le \frac{1}{2}\tilde{\zeta}_1^T\bar{\delta}_1\bar{\delta}_1^T\tilde{\zeta}_1 + \frac{1}{2}\sigma_{1max}^2. \tag{2.26}$$

Similarly,

$$\tilde{\zeta}_2^T\underline{\delta}_2\sigma_2(t) \le \frac{1}{2}\tilde{\zeta}_2^T\bar{\delta}_2\bar{\delta}_2^T\tilde{\zeta}_2 + \frac{1}{2}\sigma_{2max}^2. \tag{2.27}$$

Substituting (2.26) and (2.27) into (2.25), we get

$$\dot{\mathcal{L}} \le -\frac{1}{2}\tilde{\zeta}_1^T\bar{\delta}_1\bar{\delta}_1^T\tilde{\zeta}_1 - \frac{1}{2}\tilde{\zeta}_2^T\bar{\delta}_2\bar{\delta}_2^T\tilde{\zeta}_2 + \frac{1}{2}(\sigma_{1max}^2 + \sigma_{2max}^2). \tag{2.28}$$

From (2.28) we can conclude that $\dot{\mathcal{L}} < 0$ when one of the following conditions holds

$$\|\tilde{\zeta}_1\| > \sqrt{\frac{\sigma_{1max}^2 + \sigma_{2max}^2}{\lambda_{min}(\bar{\delta}_1\bar{\delta}_1^T)}}, \tag{2.29}$$

or

$$\|\tilde{\zeta}_2\| > \sqrt{\frac{\sigma_{1max}^2 + \sigma_{2max}^2}{\lambda_{min}(\bar{\delta}_2\bar{\delta}_2^T)}}. \tag{2.30}$$

According to Lyapunov theory, it yields that the weight estimation errors for both critic networks are UUB.

Remark 2.6 The weight matrices are usually updated through certain renewal equations, and from (2.29) and (2.30), we can draw that the approximation weight error will asymptotically converge to zero as $\nu_\epsilon \to \infty$.

Theorem 2.7 *Consider the system (2.6). The weight updating laws for critic networks are given by (2.21). Then the obtained policies (2.18) can force system states X to be UUB.*

Proof In order to discuss the stability of closed-loop system, the derivative of $\mathcal{V} = \mathcal{V}_1^* + \mathcal{V}_2^*$ is considered as

$$
\begin{aligned}
\dot{\mathcal{V}} =& (\nabla \mathcal{V}_1^*)^T (f(X) + \boldsymbol{g}(X)\hat{u} + \kappa(X)\hat{v}) \\
&+ (\nabla \mathcal{V}_2^*)^T (f(X) + \boldsymbol{g}(X)\hat{u} + \kappa(X)\hat{v}).
\end{aligned}
\tag{2.31}
$$

Recalling (2.13) and (2.14), we have

$$
\begin{aligned}
\nabla \mathcal{V}_1^{*T} f(X) =& - X^T \mathcal{Q}_1 X + \frac{1}{4} \nabla \mathcal{V}_1^{*T} \boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)\nabla \mathcal{V}_1^* \\
&- \frac{1}{4} \nabla \mathcal{V}_2^{*T} \kappa(X)\mathcal{R}_{22}^{-1}\mathcal{R}_{12}\mathcal{R}_{22}^{-1}\kappa^T(X)\nabla \mathcal{V}_2^* \\
&+ \frac{1}{2} \nabla \mathcal{V}_1^{*T} \kappa(X)\mathcal{R}_{22}^{-1}\kappa^T(X)\nabla \mathcal{V}_2^*,
\end{aligned}
\tag{2.32}
$$

and

$$
\begin{aligned}
\nabla \mathcal{V}_2^{*T} f(X) =& - X^T \mathcal{Q}_2 X + \frac{1}{4} \nabla \mathcal{V}_2^{*T} \kappa(X)\mathcal{R}_{22}^{-1}\kappa^T(X)\nabla \mathcal{V}_2^* \\
&- \frac{1}{4} \nabla \mathcal{V}_1^{*T} \boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\mathcal{R}_{21}\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)\nabla \mathcal{V}_1^* \\
&+ \frac{1}{2} \nabla \mathcal{V}_2^{*T} \boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)\nabla \mathcal{V}_1^*.
\end{aligned}
\tag{2.33}
$$

For $\epsilon = 1$, we can obtain $\dot{\mathcal{V}}_1^*$ as

$$
\begin{aligned}
\dot{\mathcal{V}}_1^* =& - X^T \mathcal{Q}_1 X - \frac{1}{4} \nabla \mathcal{V}_1^{*T} \boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)\nabla \mathcal{V}_1^* \\
&- \frac{1}{4} \nabla \mathcal{V}_2^{*T} \kappa(X)\mathcal{R}_{22}^{-1}\mathcal{R}_{12}\mathcal{R}_{22}^{-1}\kappa^T(X)\nabla \mathcal{V}_2^* \\
&- \nabla \mathcal{V}_1^{*T} (\boldsymbol{g}(X)(u^* - \hat{u}) + \kappa(X)(v^* - \hat{v})).
\end{aligned}
\tag{2.34}
$$

According to (2.15) and (2.16) we have

$$
\begin{aligned}
\dot{\mathcal{V}}_1^* =& - X^T \mathcal{Q}_1 X - \frac{1}{4} \nabla \mathcal{V}_1^{*T} \boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)\nabla \mathcal{V}_1^* \\
&- \frac{1}{4} \nabla \mathcal{V}_2^{*T} \kappa(X)\mathcal{R}_{22}^{-1}\mathcal{R}_{12}\mathcal{R}_{22}^{-1}\kappa^T(X)\nabla \mathcal{V}_2^* \\
&+ \frac{1}{2} ((\nabla \xi_1(X))^T \zeta_1^* + \nabla o_1)^T \Big(\boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X) \\
&\times ((\nabla \xi_1^T(X))^T \tilde{\zeta}_1 + \nabla o_1) + \kappa(X)\mathcal{R}_{22}^{-1}\kappa^T(X) \\
&\times ((\nabla \xi_2^T(X))^T \tilde{\zeta}_2 + \nabla o_2) \Big).
\end{aligned}
\tag{2.35}
$$

Due to Assumption 2.2 and Theorem 2.5, we obtain that

$$
\begin{aligned}
\dot{\mathcal{V}}_1^* \leq & - X^T \mathcal{Q}_1 X - \frac{1}{4}\nabla\mathcal{V}_1^{*T}\boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)\nabla\mathcal{V}_1^* \\
& - \frac{1}{4}\nabla\mathcal{V}_2^{*T}\kappa(X)\mathcal{R}_{22}^{-1}\mathcal{R}_{12}\mathcal{R}_{22}^{-1}\kappa^T(X)\nabla\mathcal{V}_2^* + \theta_1,
\end{aligned}
\tag{2.36}
$$

where the positive constant θ_1 denotes the bound of the term $\frac{1}{2}((\nabla\xi_1(X))^T\zeta_1^* + \nabla o_1)^T\Big(\boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)((\nabla\xi_1^T(X))^T\tilde{\zeta}_1 + \nabla o_1) + \kappa(X)\mathcal{R}_{22}^{-1}\kappa^T(X)((\nabla\xi_2^T(X))^T\tilde{\zeta}_2 + \nabla o_2)\Big)$. As $\mathcal{R}_{11}$, $\mathcal{R}_{12}$ and $\mathcal{R}_{22}$ are symmetric positive definite, we have

$$
\begin{aligned}
& \frac{1}{4}\nabla\mathcal{V}_2^{*T}\kappa(X)\mathcal{R}_{22}^{-1}\mathcal{R}_{12}\mathcal{R}_{22}^{-1}\kappa^T(X)\nabla\mathcal{V}_2^* \\
& + \frac{1}{4}\nabla\mathcal{V}_1^{*T}\boldsymbol{g}(X)\mathcal{R}_{11}^{-1}\boldsymbol{g}^T(X)\nabla\mathcal{V}_1^* > 0.
\end{aligned}
\tag{2.37}
$$

Furthermore, we attain

$$
\dot{\mathcal{V}}_1^* \leq -X^T\mathcal{Q}_1 X + \theta_1 \leq -\lambda_{min}(\mathcal{Q}_1)\|X\|^2 + \theta_1.
\tag{2.38}
$$

Similarly, for $\epsilon = 2$, it yields that

$$
\dot{\mathcal{V}}_2^* \leq -X^T\mathcal{Q}_2 X + \theta_2 \leq -\lambda_{min}(\mathcal{Q}_2)\|X\|^2 + \theta_2,
\tag{2.39}
$$

where the definition of θ_2 is similar to that of θ_1. Then it can be concluded that $\dot{\mathcal{V}} < 0$ when the following inequality is satisfied

$$
\|X\| > \max\left\{\sqrt{\frac{\theta_1}{\lambda_{min}(\mathcal{Q}_1)}}, \sqrt{\frac{\theta_2}{\lambda_{min}(\mathcal{Q}_2)}}\right\} \triangleq \Theta.
\tag{2.40}
$$

Thus with the proposed control policies (2.18), the system state N is UUB with the bound Θ. This completes the proof.

Remark 2.8 From Theorems 2.5 and 2.7, we can conclude that under the obtained control policies the states of the system X and the critic weight error $\tilde{\zeta}_\epsilon$ are ultimately uniformly bounded.

Remark 2.9 According to the clinical requirements, the specific value of the cost function is finalised. Transformation is implemented from the mathematical mechanism model to the solvable affine model. Subsequently, the chapter solve the optimal control problem that means minimum dose of medicine can realize the best therapeutic effect.

2.4   Simulation and Numerical Experiments

To verify the proposed method in the previous section, a simulation is given as followed.

2.4.1   States Analysis on Tumor Cell Growth

According to clinical medical statistics borrowed from the literature [37], the specific parameters of the dynamic models are presented as Table 2.2.
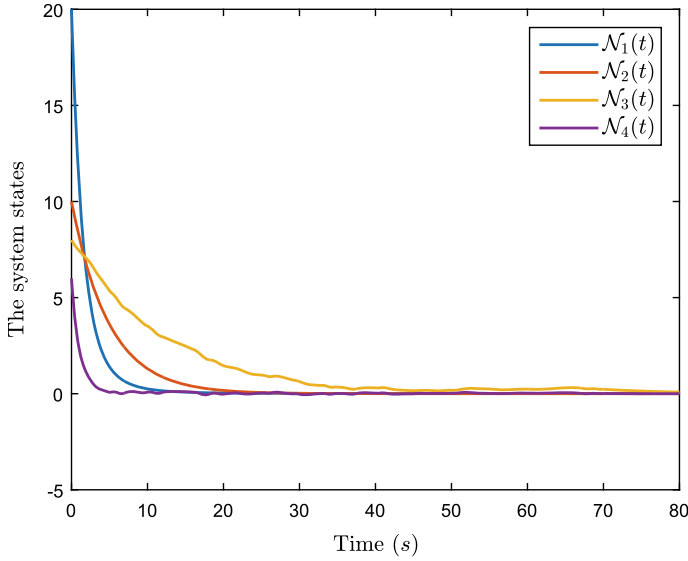
According to (2.5) and Table 2.2, we construct the model (2.41)

$$
\begin{aligned}
\dot{N}_T(t) =\ & 0.00431 N_T(t)(1 - 1.02 \times 10^{-9}) N_T(t)) \\
& - 6.41 \times 10^{-11} N_T(t) N_H(t) \\
& - 0.08 N_{CD}(t) N_T(t) - (1 - e^{-u(t)}) \\
\dot{N}_H(t) =\ & 0.33 + \frac{0.0125 N_H(t) N_T^2(t)}{2.02 \times 10^7 + N_T^2(t)} + \frac{0.125 N_H(t) N_{ID}(t)}{2 \times 10^7 + N_{ID}(t)} \\
& - 3.42 \times 10^{-6} N_T(t) N_H(t) - (1 - e^{-u(t)}) \\
& - 0.204 N_H(t) - 3.42 \times 10^{-6} N_{CD}(t) N_H(t) \\
\dot{N}_{CD}(t) =\ & u(t) - 0.1 N_{CD}(t) \\
\dot{N}_{ID}(t) =\ & v(t) - N_{ID}(t)
\end{aligned}
\tag{2.41}
$$

The initial state of tumor cells $N_1(t)$ and immune cells $N_2(t)$ in a patient and follow a certain chemotherapy and immunotherapy regimen. Correspondingly, $N_3(t)$ and $N_4(t)$ respectively denote the concentrations of chemotherapy and immunotherapy. And we can get the following curves on systems states tumor cells, immune cells, chemotherapy and immunotherapy drugs as shown in Fig. 2.1. Initial value is set as $X_0 = \begin{bmatrix} 20 & 10 & 8 & 6 \end{bmatrix}^T$.

Table 2.2 Concentration variation on immune cells, tumor cells, chemotherapeutic drug and immunoagents

Parameter	Estimated value	Units	Parameter	Estimated value	Units
ι_1	0.00431	day^{-1}	ϱ_1	1.02×10^{-9}	cell^{-1}
σ_1	6.41×10^{-11}	cell^{-1}	δ_1	0.08	day^{-1}
$\mathfrak{D}$	0.204	day^{-1}	σ_2	3.42×10^{-6}	cell^{-1}
δ_2	2×10^{-11}	day^{-1}	υ	0.0125	day^{-1}
ς	0.125	day^{-1}	ν	2.02×10^7	cell^2
ϑ	2×10^7	cell	φ_1	0.1	day^{-1}
φ_2	1	day^{-1}			

Fig. 2.1 The curves of system states

It is obviously that the control policies can stabilize the nonlinear system and make the system states tend to zero which means that the closed-system is stable and the control method is effective. Retrospect the original problem that the key is to minimize cancer cells and reduce therapy toxicity as possible.

2.4.2 Weight Analysis of Control Policies

The weights ζ_ϵ^* of the control policies $u(t)$ and $v(t)$ can be estimated through the value function $\hat{V}_\epsilon^* = (\hat{\zeta}_\epsilon)^T \xi_\epsilon(X)$ in (2.16), and the performance index is shown as (2.6) with $\mathcal{Q}_1 = I_{4 \times 4}$, $\mathcal{Q}_2 = 5\mathcal{Q}_1$, $\mathcal{R}_{11} = \mathcal{R}_{22} = 1$, $\mathcal{R}_{12} = \mathcal{R}_{21} = 2$. The initialize weights are set as $[-0.25, -0.25, -1, -0.25]^T$. The selected activation function is selected as $[\zeta_{11 \to 15}^T, \zeta_{16 \to 18}^T, \zeta_{19 \to 10}^T]$, where $\zeta_{11 \to 15} = [N_1^2(t), N_1(t)N_2(t), N_1(t)N_3(t), N_1(t)N_4(t),$ $N_2^2(t)]$ and $\zeta_{16 \to 18} = [N_2(t)N_3(t), N_2(t)N_4(t), N_3^2(t)]$ and $\zeta_{19 \to 10} = [N_3(t)N_4(t),$ $N_4^2(t)]$

According to Fig. 2.2, we can conclude that the proposed optimal control demonstrated a shorter convergence time than that without taking optimal control, where the former needs only $10s$, but the later may be $38s$, which draws the superiority of the proposed method.

In Fig. 2.3, we can obtain the less doses of the drugs is another advantage compared with that without taking optimal control. Taking comprehensive consideration of Figs. 2.2 and 2.3, we can draw a conclusion that the adopted algorithm can not only
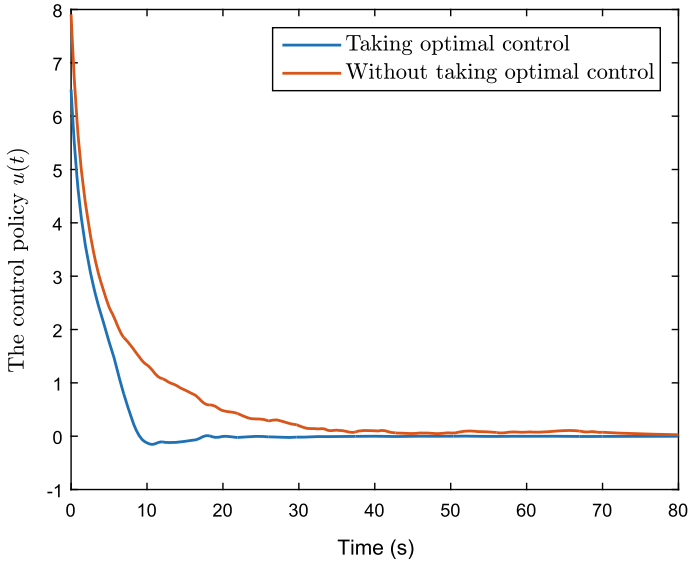
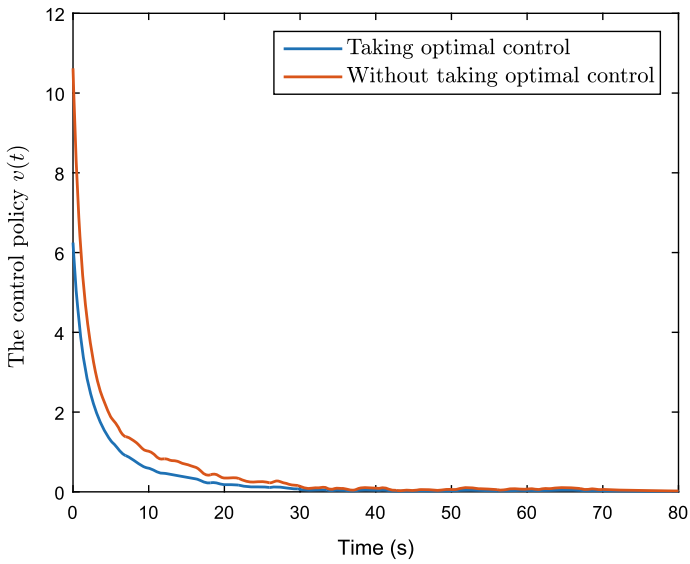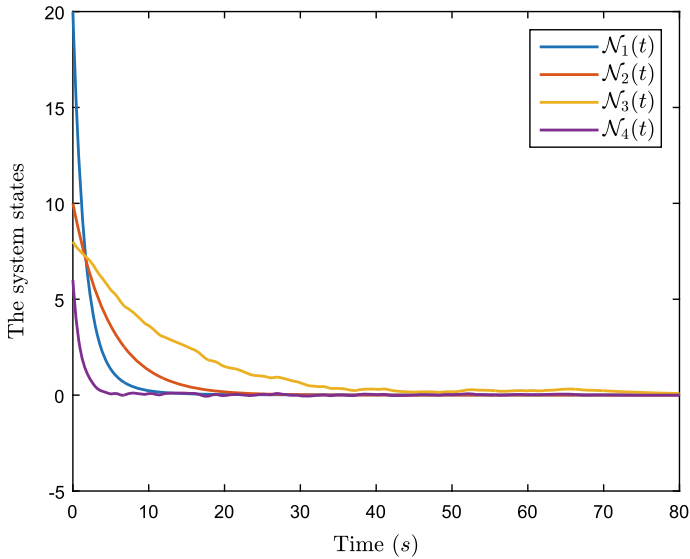**Fig. 2.2** Optimal control policies $u(t)$

Fig. 2.3 Optimal control policies $v(t)$

Fig. 2.4 The curves of system states

decrease the convergence time but also reduce doses of chemotherapy drugs and immune agents, and patients will benefit from for the minimal toxicity and shorter response time.

When the initialize state is set as $[-0.5, -0.1, -1, -0.4]^T$, and the other parameters are unaltered, we give another set of figures as Figs. 2.4, 2.5 and 2.6. In Figs. 2.5 and 2.6, there exist more obvious advantages for the proposed algorithms over that without taking optimal control in response time and control policies,and we can conclude that effectiveness of the control method does not vary in the different initial weights.

2.5 Conclusion

This chapter has introduced adaptive dynamic programming into solving the optimal control policies of tumor cells growth model and realized objective of minimizing tumor cells with the minimum dose of chemotherapeutic and immunotherapeutic drugs. As is known, the negative effect caused by chemotherapy and immunotherapy must be reduced for the reasonable treatment plan extracted from the optimal control behavior. Convergence properties have been proved to be guaranteed through Lyapunov theory. Meanwhile, states of the system and critic error have been demonstrated to be ultimately uniformly bounded. Simulations have been given to verify
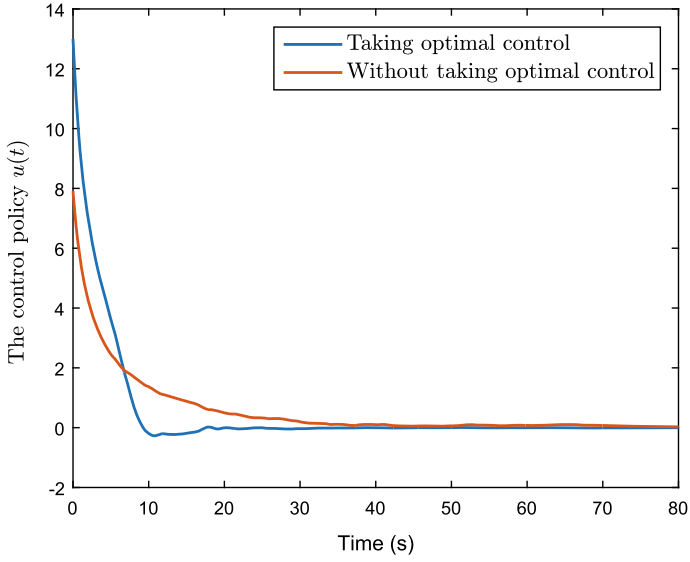
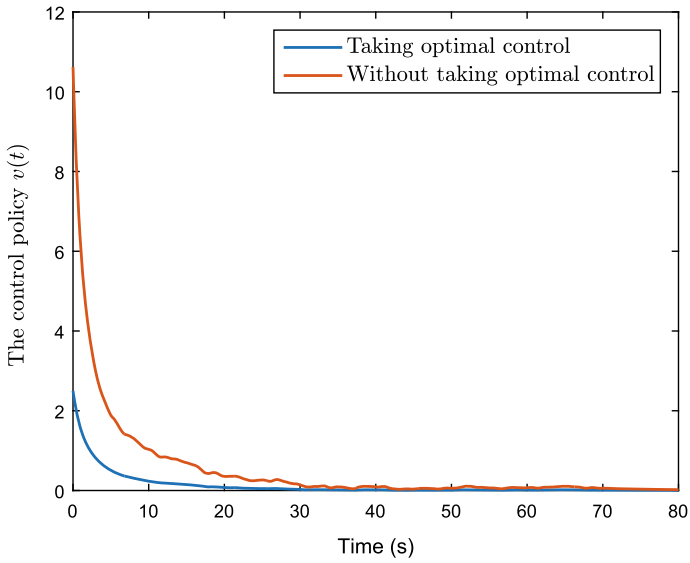**Fig. 2.5** Optimal control policies $u(t)$

Fig. 2.6 Optimal control policies $v(t)$

rationality of the proposed method. In the future work, we will further investigate the medical frontier topics and propose adaptive therapeutic methods to solve these issues by employing ADP approach.

References

1. de Pillis LG, Gu W, Radunskaya AE (2006) Mixed immunotherapy and chemotherapy of tumors: modeling, applications and biological interpretations. J Theor Biol 238(4):841–862
2. Ogunmadeji B, Yusuf TT (2018) Optimal control strategy for improved cancer biochemotherapy outcome. Int J Sci Eng Res 9(12):583–600
3. Liang H, Liu G, Zhang H, Huang T (2021) Neural-network-based event-triggered adaptive control of nonaffine nonlinear multiagent systems with dynamic uncertainties. IEEE Trans Neural Netw Learn Syst 32(5):2239–2250
4. Sun J, Zhang H, Wang Y, Sun S (2022) Fault-tolerant control for stochastic switched IT2 fuzzy uncertain time-delayed nonlinear systems. IEEE Trans Cybernet 52(2):1335–1346
5. Lodhi I, Ahmad I, Uneeb M, Liaquat M (2019) Nonlinear Control for Growth of Cancerous Tumor Cells. IEEE Access 7:177628–177636
6. Wang J, Huang M, Chen S, Luo Y, Shen S, Du X (2021) Nanomedicine-mediated ubiquitination inhibition boosts antitumor immune response via activation of dendritic cells. Nano Res 14:3900–3906
7. Kuznetsov VA, Makalkin IA (1992) Bifurcation analysis of a mathematical model of the interaction of cytotoxic lymphocytes with tumor cells. The effect of immunologic amplification of tumor growth and its interconnection with other anomolous phenomena of oncoimmunology. Biofizika 37(6):1063–1070
8. Zhang H, Cui L, Luo Y (2013) Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP. IEEE Trans Cybernet 43(1):206–216
9. Fan QY, Xu S, Xu B, Qiu J (2022) Simplified prescribed performance tracking control of uncertain nonlinear systems. Sci China Inf Sci 65:189204
10. Starr WA, Ho YC (1969) Nonzero-sum differential games. J Optim Theory Appl 3(2):184–206
11. Su H, Zhang H, Jiang H, Wen Y (2020) Decentralized event-triggered adaptive control of discrete-time nonzero-sum games over wireless sensor-actuator networks with input constraints. IEEE Trans Neural Netw Learn Syst 31(10):4254–4266
12. Wei Q, Zhu L, Song R, Zhang P, Liu D, Xiao J (2022) Model-free adaptive optimal control for unknown nonlinear multiplayer nonzero-sum game. IEEE Trans Neural Netw Learn Syst 33(2):879–892
13. Song R, Lewis FL, Wei Q (2017) Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games. IEEE Trans Neural Netw Learn Syst 28(3):704–713
14. Zhang K, Su R, Zhang H, Tian Y (2021) Adaptive resilient event-triggered control design of autonomous vehicles with an iterative single critic learning framework. IEEE Trans Neural Netw Learn Syst 32(12):5502–5511
15. Yang D, Li T, Xie X, Zhang H (2020) Event-triggered integral sliding-mode control for nonlinear constrained-input systems with disturbances via adaptive dynamic programming. IEEE Trans Syst Man Cybernet: Syst 50(11):4086–4096
16. Wei Q, Li H, Yang X, He H (2021) Continuous-time distributed policy iteration for multicontroller nonlinear systems. IEEE Trans Cybernet 51(5):2372–2383
17. Zhao B, Liu D, Luo C (2020) Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints. IEEE Trans Neural Netw Learn Syst 31(10):4330–4340

18. Zhang T, Su G, Qing C, Xu X, Cai B, Xing X (2021) Hierarchical lifelong learning by sharing representations and integrating hypothesis. IEEE Trans Syst Man Cybernet: Syst 51(2):1004–1014
19. Narayanan V, Sahoo A, Jagannathan S, George K (2019) Approximate optimal distributed control of nonlinear interconnected systems using event-triggered nonzero-sum games. IEEE Trans Neural Netw Learn Syst 30(5):1512–1522
20. Zhong X, He H (2017) An event-triggered ADP control approach for continuous-time system with unknown internal states. IEEE Trans Cybernet 47(3):683–694
21. Sun J, Zhang H, Wang Y, Shi Z (2022) Dissipativity-based fault-tolerant control for stochastic switched systems with time-varying delay and uncertainties. IEEE Trans Cybernet 52(10):10683–10694
22. Zhang K, Zhang H, Mu Y, Liu C (2021) Decentralized tracking optimization control for partially unknown fuzzy interconnected systems via reinforcement learning method. IEEE Trans Fuzzy Syst 29(4):917–926
23. Li T, Yang D, Xie X, Zhang H (2022) Event-triggered control of nonlinear discrete-time system with unknown dynamics based on HDP(λ). IEEE Trans Cybernet 52(7):6046–6058
24. Mu C, Wang K, Ni Z (2022) Adaptive learning and sampled-control for nonlinear game systems using dynamic event-triggering strategy. IEEE Trans Neural Netw Learn Syst 33(9):4437–4450
25. Mu C, Wang K, Qiu T (2022) Dynamic event-triggering neural learning control for partially unknown nonlinear systems. IEEE Trans Cybernet 52(4):2200–2213
26. Zhang H, Qin C, Jiang B, Luo Y (2014) Online adaptive policy learning algorithm for H_∞ state feedback control of unknown affine nonlinear discrete-time systems. IEEE Trans Cybernet 44(12):2706–2718
27. Zhang Q, Zhao D (2019) Data-based reinforcement learning for nonzero-sum games with unknown drift dynamics. IEEE Trans Cybernet 49(8):2874–2885
28. Chen G, Yao D, Zhou Q, Li H, Lu R (2022) Distributed event-triggered formation control of usvs with prescribed performance. J Syst Sci Complex 35(3):820–838
29. Zhang H, Su H, Zhang K, Luo Y (2019) Event-triggered adaptive dynamic programming for non-zero-sum games of unknown nonlinear systems via generalized fuzzy hyperbolic models. IEEE Trans Fuzzy Syst 27(11):2202–2214
30. Wei Q, Song R, Liao Z, Li B, Lewis FL (2020) Discrete-time impulsive adaptive dynamic programming. IEEE Trans Cybernet 50(10):4293–4306
31. Lewis FL, Vrabie DL, Syrmos VL (2012) Optimal control. Wiley, Hoboken, New Jersey, USA
32. Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. SIAM, Philadelphia, PA, USA
33. Vamvoudakis KG (2014) Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems. IEEE/CAA J Autom Sinica 1(3):282–293
34. Vamvoudakis KG, Lewis FL (2010) Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. Automatica 46(5):878–888
35. Wang D, He H, Zhong X, Liu D (2017) Event-driven nonlinear discounted optimal regulation involving a power system application. IEEE Trans Ind Electron 64(10):8177–8186
36. Zhang H, Zhang J, Yang G-H, Luo Y (2015) Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming. IEEE Trans Fuzzy Syst 23(1):152–163
37. Sharma S, Samanta GP (2016) Analysis of the dynamics of a tumor-immune system with chemotherapy and immunotherapy and quadratic optimal control. Differ Equ Dyn Syst 24(2):149–171