Chapter 9
Nonlinear Games for a Class of Continuous-Time Systems Based on ADP

9.1 Introduction

Game theory is concerned with the study of decision making in situations where two or more rational opponents are involved under conditions of conflicting interests. This has been widely investigated by many authors [5, 7, 8, 12, 13]. Though the nonlinear optimal solution in term of Hamilton–Jacobi–Bellman equation is hard to obtain directly [4], it is still fortunate that there is only one controller or decision maker. In the previous chapter, we have studied discrete-time zero-sum games based on the ADP method. In this chapter, we will consider continuous-time games.

For zero-sum differential games, the existence of the saddle point is proposed before obtaining the saddle point in much of the literature [1, 6, 11]. In many real world applications, however, the saddle point of a game may not exist, which means that we can only obtain the mixed optimal solution of the game. In Sect. 9.2, we will study how to obtain the saddle point without complex existence conditions of the saddle point and how to obtain the mixed optimal solution when the saddle point does not exist based on the ADP method for a class of affine nonlinear zero-sum games. Note that many applications of practical zero-sum games have nonaffine control input. In Sect. 9.3, we will focus on finite horizon zero-sum games for a class of nonaffine nonlinear systems.

The non-zero-sum differential games theory also has a number of potential applications in control engineering, economics and the military field [9]. For zero-sum differential games, two players work on a cost functional together and minimax it. However, for non-zero-sum games, the control objective is to find a set of policies that guarantee the stability of the system and minimize the individual performance function to yield a Nash equilibrium. In Sect. 9.4, non-zero-sum differential games will be studied using a single network ADP.

H. Zhang et al., *Adaptive Dynamic Programming for Control*,
Communications and Control Engineering, DOI 10.1007/978-1-4471-4757-2_9,

9.2 Infinite Horizon Zero-Sum Games for a Class of Affine Nonlinear Systems

In this section, the nonlinear infinite horizon zero-sum differential games is studied. We propose a new iterative ADP method which is effective for both the situation that the saddle point does and does not exist. For the situation that the saddle point exists, the existence conditions of the saddle point are avoided. The value function can reach the saddle point using the present iterative ADP method. For the situation that the saddle point does not exist, the mixed optimal value function is obtained under a deterministic mixed optimal control scheme, using the present iterative ADP algorithm.

9.2.1 Problem Formulation

Consider the following two-person zero-sum differential games. The system is described by the continuous-time affine nonlinear equation

$$\dot{x}(t) = f(x(t), u(t), w(t)) = f(x(t)) + g(x(t))u(t) + k(x(t))w(t), \tag{9.1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^k$, $w(t) \in \mathbb{R}^m$, and the initial condition $x(0) = x_0$ is given.

The cost functional is the generalized quadratic form given by

$$J(x(0), u, w) = \int_0^\infty l(x, u, w)\mathrm{d}t, \tag{9.2}$$

where $l(x, u, w) = x^{\mathrm{T}}Ax + u^{\mathrm{T}}Bu + w^{\mathrm{T}}Cw + 2u^{\mathrm{T}}Dw + 2x^{\mathrm{T}}Eu + 2x^{\mathrm{T}}Fw$. The matrices A, B, C, D, E, and F have suitable dimensions and $A \geq 0$, $B > 0$, and $C < 0$. According to the situation of two players we have the following definitions. Let $\overline{J}(x) := \inf_u \sup_w J(x, u, w)$ be the *upper value function* and $\underline{J}(x) := \sup_w \inf_u J(x, u, w)$ be the *lower value function* with the obvious inequality $\overline{J}(x) \geq \underline{J}(x)$. Define the optimal control pairs to be $(\overline{u}, \overline{w})$ and $(\underline{u}, \underline{w})$ for upper and lower value functions, respectively. Then, we have $\overline{J}(x) = J(x, \overline{u}, \overline{w})$ and $\underline{J}(x) = J(x, \underline{u}, \underline{w})$.

If both $\overline{J}(x)$ and $\underline{J}(x)$ exist and

$$\overline{J}(x) = \underline{J}(x) = J^*(x) \tag{9.3}$$

holds, we say that the saddle point exists and the corresponding optimal control pair is denoted by (u^*, w^*).

We have the following lemma.

Lemma 9.1 *If the nonlinear system* (9.1) *is controllable and both the upper value function and lower value function exist, then* $\overline{J}(x)$ *is a solution of the following*

upper Hamilton–Jacobi–Isaacs (*HJI*) *equation*:

$$\inf_u \sup_w \{\overline{J}_t + \overline{J}_x^{\mathrm{T}} f(x,u,w) + l(x,u,w)\} = 0, \tag{9.4}$$

which is denoted by $\mathrm{HJI}(\overline{J}(x), \overline{u}, \overline{w}) = 0$ *and* $\underline{J}(x)$ *is a solution of the following lower HJI equation*:

$$\sup_w \inf_u \{\underline{J}_t + \underline{J}_x^{\mathrm{T}} f(x,u,w) + l(x,u,w)\} = 0, \tag{9.5}$$

which is denoted by $\mathrm{HJI}(\underline{J}(x), \underline{u}, \underline{w}) = 0$.

9.2.2 Zero-Sum Differential Games Based on Iterative ADP Algorithm

As the HJI equations (9.4) and (9.5) cannot be solved in general, in the following, a new iterative ADP method for zero-sum differential games is developed.

9.2.2.1 Derivation of the Iterative ADP Method

The goal of the present iterative ADP method is to obtain the saddle point. As the saddle point may not exist, this motivates us to obtain the mixed optimal value function $J^o(x)$ where $\underline{J}(x) \le J^o(x) \le \overline{J}(x)$.

Theorem 9.2 (cf. [15]) *Let* $(\overline{u}, \overline{w})$ *be the optimal control pair for* $\overline{J}(x)$ *and* $(\underline{u}, \underline{w})$ *be the optimal control pair for* $\underline{J}(x)$. *Then, there exist control pairs* $(\overline{u}, w)$ *and* $(u, \underline{w})$ *which lead to* $J^o(x) = J(x, \overline{u}, w) = J(x, u, \underline{w})$. *Furthermore, if the saddle point exists, then* $J^o(x) = J^*(x)$.

Proof According to the definition of $\overline{J}(x)$, we have $J(x, \overline{u}, w) \le J(x, \overline{u}, \overline{w})$. As $J^o(x)$ is a mixed optimal value function, we also have $J^o(x) \le J(x, \overline{u}, \overline{w})$. As the system (9.1) is controllable and w is continuous on $\mathbb{R}^m$, there exists a control pair $(\overline{u}, w)$ which makes $J^o(x) = J(x, \overline{u}, w)$. On the other hand, we have $J^o(x) \ge J(x, \underline{u}, \underline{w})$. We also have $J(x, u, \underline{w}) \ge J(x, \underline{u}, \underline{w})$. As u is continuous on $\mathbb{R}^k$, there exists a control pair $(u, \underline{w})$ which makes $J^o(x) = J(x, u, \underline{w})$. If the saddle point exists, we have (9.3). On the other hand, $\underline{J}(x) \le J^o(x) \le \overline{J}(x)$. Then, clearly $J^o(x) = J^*(x)$. □

If (9.3) holds, we have a saddle point; if not, we adopt a mixed trajectory to obtain the mixed optimal solution of the game. To apply the mixed trajectory method, the game matrix is necessary under the trajectory sets of the control pair (u, w). Small Gaussian noises $\gamma_u \in R^k$ and $\gamma_w \in R^m$ are introduced that are added to the optimal

control $\underline{u}$ and $\overline{w}$, respectively, where $\gamma_u^i(0, \sigma_i^2)$, $i = 1, \ldots, k$, and $\gamma_w^j(0, \sigma_j^2)$, $j = 1, \ldots, m$, are zero-mean Gaussian noises with variances σ_i^2 and σ_j^2, respectively.

We define the expected value function as

$E(J(x)) = \min_{P_{Ii}} \max_{P_{IIj}} \sum_{i=1}^{2} \sum_{j=1}^{2} P_{Ii} L_{ij} P_{IIj}$, where we let $L_{11} = J(x, \overline{u}, \overline{w})$, $L_{12} = J(x, (\underline{u} + \gamma_u), \underline{w})$, $L_{21} = J(x, \underline{u}, \underline{w})$ and $L_{22} = J(x, \overline{u}, (\overline{w} + \gamma_w))$. Let $\sum_{i=1}^{2} P_{Ii} = 1$ and $P_{Ii} > 0$. Let $\sum_{j=1}^{2} P_{IIj} = 1$ and $P_{IIj} > 0$. Next, let N be a large enough positive integer. Calculating the expected value function N times, we can obtain $E_1(J(x)), E_2(J(x)), \ldots, E_N(J(x))$. Then, the mixed optimal value function can be written as

$$J^o(x) = E(E_i(J(x))) = \frac{1}{N} \sum_{i=1}^{N} E_i(J(x)).$$

Remark 9.3 In the classical mixed trajectory method, the whole control sets $\mathbb{R}^k$ and $\mathbb{R}^m$ should be searched under some distribution functions. As there are no constraints for both controls, we see that there exist controls that cause the system to be unstable. This is not permitted for real-world control systems. Thus, it is impossible to search the whole control sets and we can only search the local area around the stable controls which guarantees stability of the system. This is the reason why the small Gaussian noises γ_u and γ_w are introduced. So the meaning of the Gaussian noises can be seen in terms of the local stable area of the control pairs. A proposition will be given to show that the control pair chosen in the local area is stable (see Proposition 9.14). Similar work can also be found in [3, 14].

We can see that the mixed optimal solution is a mathematically expected value which means that it cannot be obtained in reality once the trajectories are determined. For most practical optimal control problems, however, the expected optimal solution (or mixed optimal solution) has to be achieved. To overcome this difficulty, a new method is developed in this section. Let $\alpha = (J^o(x) - \underline{J}(x))/(\overline{J}(x) - \underline{J}(x))$. Then, $J^o(x)$ can be written as $J^o(x) = \alpha \overline{J}(x) + (1 - \alpha)\underline{J}(x)$. Let $l^o(x, \overline{u}, \overline{w}, \underline{u}, \underline{w}) = \alpha l(x, \overline{u}, \overline{w}) + (1 - \alpha) l(x, \underline{u}, \underline{w})$. We have $J^o(x(0)) = \int_0^\infty l^o \mathrm{d}t$. According to Theorem 9.2, the mixed optimal control pair can be obtained by regulating the control w in the control pair $(\overline{u}, \overline{w})$ that minimizes the error between $\mathcal{J}(x)$ and $J^o(x)$ where the value function $\mathcal{J}(x)$ is defined as $\mathcal{J}(x(0)) = J(x(0), \overline{u}, w) = \int_0^\infty l(x, \overline{u}, w)\mathrm{d}t$ and $\underline{J}(x(0)) \le \mathcal{J}(x(0)) \le \overline{J}(x(0))$.

Define $\widetilde{J}(x(0)) = \int_0^\infty \widetilde{l}(x, w) dx$, where $\widetilde{l}(x, w) = l(x, \overline{u}, w) - l^o(x, \overline{u}, \overline{w}, \underline{u}, \underline{w})$. Then, the problem can be described as $\min_w (\widetilde{J}(x))^2$.

According to the principle of optimality, when $\widetilde{J}(x) \ge 0$ we have the following HJB equation:

$$\mathrm{HJB}(\widetilde{J}(x), w) := \min_w \{\widetilde{J}_t(x) + \widetilde{J}_x f(x, u, w) + \widetilde{l}(x, w)\} = 0. \tag{9.6}$$

For $\widetilde{J}(x) < 0$, we have $-\widetilde{J}(x) = -(\mathcal{J}(x) - J^o(x)) > 0$, and we can obtain the same HJB equation as (9.6).

9.2.2.2 The Iterative ADP Algorithm

Given the above preparation, we now formulate the iterative ADP algorithm for zero-sum differential games as follows:

1. Initialize the algorithm with a stabilizing control pair $(u^{[0]}, w^{[0]})$, and the value function is $V^{[0]}$. Choose the computation precision $\zeta > 0$. Set $i = 0$.
2. For the upper value function, let

$$\overline{V}^{[i]}(x(0)) = \int_0^\infty l(x, \overline{u}^{[i+1]}, \overline{w}^{[i+1]})\mathrm{d}t, \tag{9.7}$$

where the iterative optimal control pair is formulated as

$$\begin{aligned}\overline{u}^{[i+1]} = &-\frac{1}{2}(B - DC^{-1}D^{\mathrm{T}})^{-1}\big(2(k^{\mathrm{T}} - DC^{-1}F^{\mathrm{T}})x \\ &+ (g^{\mathrm{T}}(x) - DC^{-1}k^{\mathrm{T}}(x))\overline{V}_x^{[i]}\big),\end{aligned} \tag{9.8}$$

and

$$\overline{w}^{[i+1]} = -\frac{1}{2}C^{-1}\big(2D^{\mathrm{T}}\overline{u}^{[i+1]} + 2F^{\mathrm{T}}x + k^{\mathrm{T}}(x)\overline{V}_x^{[i]}\big). \tag{9.9}$$

$(\overline{u}^{[i]}, \overline{w}^{[i]})$ satisfies the HJI equation $\mathrm{HJI}(\overline{V}^{[i]}(x), \overline{u}^{[i]}, \overline{w}^{[i]}) = 0$, and $\overline{V}_x^{[i]} = d\overline{V}^{[i]}(x)/dx$.

3. If $|\overline{V}^{[i+1]}(x(0)) - \overline{V}^{[i]}(x(0))| < \zeta$, let $\overline{u} = \overline{u}^{[i]}$, $\overline{w} = \overline{w}^{[i]}$ and $\overline{J}(x) = \overline{V}^{[i+1]}(x)$. Set $i = 0$ and go to Step 4. Else, set $i = i + 1$ and go to Step 2.
4. For the lower value function, let

$$\underline{V}^{[i]}(x(0)) = \int_0^\infty l(x, \underline{u}^{[i+1]}, \underline{w}^{[i+1]})\mathrm{d}t, \tag{9.10}$$

where the iterative optimal control pair is formulated as

$$\underline{u}^{[i+1]} = -\frac{1}{2}g^{-1}(2D\underline{w}^{[i+1]} + 2k^{\mathrm{T}}x + g^{\mathrm{T}}(x)\underline{V}_x^{[i]}), \tag{9.11}$$

and

$$\begin{aligned}\underline{w}^{[i+1]} = &-\frac{1}{2}(C - D^{\mathrm{T}}BD)^{-1}(2(F^{\mathrm{T}} - D^{\mathrm{T}}g^{-1}E)x \\ &+ (k^{\mathrm{T}}(x) - D^{\mathrm{T}}g^{-1}g^{\mathrm{T}}(x))\underline{V}_x^{[i]}).\end{aligned} \tag{9.12}$$

$(\underline{u}^{[i]}, \underline{w}^{[i]})$ satisfies the HJI equation $\mathrm{HJI}(\underline{V}^{[i]}(x), \underline{u}^{[i]}, \underline{w}^{[i]}) = 0$, and $\underline{V}_x^{[i]} = d\underline{V}^{[i]}(x)/dx$.

5. If $|\underline{V}^{[i+1]}(x(0)) - \underline{V}^{[i]}(x(0))| < \zeta$, let $\underline{u} = \underline{u}^{[i]}$, $\underline{w} = \underline{w}^{[i]}$ and $\underline{J}(x) = \underline{V}^{[i+1]}(x)$. Set $i = 0$ and go to Step 6. Else, set $i = i + 1$ and go to Step 4.
6. If $|\overline{J}(x(0)) - \underline{J}(x(0))| < \zeta$, stop, and the saddle point is achieved. Else set $i = 0$ and go to the next step.

7. Regulate the control w for the upper value function and let

$$\widetilde{J}^{[i+1]}(x(0)) = \mathcal{V}^{[i+1]}(x(0)) - J^o(x(0)) = \int_0^\infty \widetilde{l}(x, \overline{u}, w^{[i]})\mathrm{d}t. \tag{9.13}$$

The iterative optimal control is formulated as

$$w^{[i]} = -\frac{1}{2}C^{-1}(2D^{\mathrm{T}}\overline{u} + 2F^{\mathrm{T}}x + k^{\mathrm{T}}(x)\widetilde{V}_x^{[i+1]}), \tag{9.14}$$

where $\widetilde{V}_x^{[i]} = d\widetilde{V}^{[i]}(x)/dx$.

8. If $|\mathcal{V}^{[i+1]}(x(0)) - J^o(x(0))| < \zeta$, stop. Else, set $i = i + 1$ and go to Step 7.

9.2.2.3 Properties of the Iterative ADP Algorithm

In this part, some results are presented to show the stability and convergence of the present iterative ADP algorithm.

Theorem 9.4 (cf. [15]) *If for* $\forall i \geq 0$, $\mathrm{HJI}(\overline{V}^{[i]}(x), \overline{u}^{[i]}, \overline{w}^{[i]}) = 0$ *holds, and for* $\forall t$, $l(x, \overline{u}^{[i]}, \overline{w}^{[i]}) \geq 0$, *then the control pairs* $(\overline{u}^{[i]}, \overline{w}^{[i]})$ *make system* (9.1) *asymptotically stable.*

Proof According to (9.7), for $\forall t$, taking the derivative of $\overline{V}^{[i]}(x)$, we have

$$\frac{\mathrm{d}\overline{V}^{[i]}(x)}{\mathrm{d}t} = \overline{V}_x^{[i]\mathrm{T}}\left(f(x) + g(x)\overline{u}^{[i+1]} + k(x)\overline{w}^{[i+1]}\right). \tag{9.15}$$

From the HJI equation we have

$$0 = \overline{V}_x^{[i]\mathrm{T}} f(x, \overline{u}^{[i]}, \overline{w}^{[i]}) + l(x, \overline{u}^{[i]}, \overline{w}^{[i]}). \tag{9.16}$$

Combining (9.15) and (9.16), we get

$$\begin{aligned}\frac{\mathrm{d}\overline{V}^{[i]}(x)}{\mathrm{d}t} =& \overline{V}_x^{[i]\mathrm{T}}(g(x) - k(x)C^{-1}D^{\mathrm{T}})(\overline{u}^{[i+1]} - \overline{u}^{[i]}) \\ & - x^{\mathrm{T}}Ax - \overline{u}^{[i]\mathrm{T}}(B - DC^{-1}D^{\mathrm{T}})\overline{u}^{[i]} - \frac{1}{4}\overline{V}_x^{[i]\mathrm{T}}k(x)C^{-1}k^{\mathrm{T}}(x)\overline{V}_x^{[i]} \\ & - 2x^{\mathrm{T}}(E - FC^{-1}D^{\mathrm{T}})\overline{u}^{[i+1]} + x^{\mathrm{T}}FC^{-1}F^{\mathrm{T}}x.\end{aligned} \tag{9.17}$$

According to (9.8) we have

$$\frac{\mathrm{d}\overline{V}^{[i]}(x)}{\mathrm{d}t} = -(\overline{u}^{[i+1]} - \overline{u}^{[i]})^{\mathrm{T}}(B - DC^{-1}D^{\mathrm{T}})$$

$$\times (\overline{u}^{[i+1]} - \overline{u}^{[i]}) - l(x, \overline{u}^{[i+1]}, \overline{w}^{(i+1)})$$
$$\leq 0. \tag{9.18}$$

So, $\overline{V}^{[i]}(x)$ is a Lyapunov function. Let $\varepsilon > 0$ and $\|x(t_0)\| < \delta(\varepsilon)$. Then, there exist two functions $\alpha(\|x\|)$ and $\beta(\|x\|)$ which belong to class $\mathcal{K}$ and satisfy

$$\alpha(\varepsilon) \geq \beta(\delta) \geq \overline{V}^{[i]}(x(t_0)) \geq \overline{V}^{[i]}(x(t)) \geq \alpha(\|x\|). \tag{9.19}$$

Therefore, system (9.1) is asymptotically stable. □

Theorem 9.5 (cf. [15]) *If for* $\forall\, i \geq 0$, $\mathrm{HJI}(\underline{V}^{[i]}(x), \underline{u}^{[i]}, \underline{w}^{[i]}) = 0$ *holds, and for* $\forall\, t$, $l(x, \underline{u}^{[i]}, \underline{w}^{[i]}) < 0$, *then the control pairs* $(\underline{u}^{[i]}, \underline{w}^{[i]})$ *make system* (9.1) *asymptotically stable.*

Corollary 9.6 *If for* $\forall i \geq 0$, $\mathrm{HJI}(\underline{V}^{[i]}(x), \underline{u}^{[i]}, \underline{w}^{[i]}) = 0$ *holds, and for* $\forall t$, $l(x, \underline{u}^{[i]}, \underline{w}^{[i]}) \geq 0$, *then the control pairs* $(\underline{u}^{[i]}, \underline{w}^{[i]})$ *make system* (9.1) *asymptotically stable.*

Proof As $\underline{V}^{[i]}(x) \leq \overline{V}^{[i]}(x)$ and $l(x, \underline{u}^{[i]}, \underline{w}^{[i]}) \geq 0$, we have $0 \leq \underline{V}^{[i]}(x) \leq \overline{V}^{[i]}(x)$.

From Theorem 9.4, we know that for $\forall t_0$, there exist two functions $\alpha(\|x\|)$ and $\beta(\|x\|)$ which belong to class $\mathcal{K}$ and satisfy (9.19).

As $\overline{V}^{[i]}(x) \to 0$, there exist time instants t_1 and t_2 (without loss of generality, let $t_0 < t_1 < t_2$) that satisfy

$$\overline{V}^{[i]}(x(t_0)) \geq \overline{V}^{[i]}(x(t_1)) \geq \underline{V}^{[i]}(x(t_0)) \geq \overline{V}^{[i]}(x(t_2)). \tag{9.20}$$

Choose $\varepsilon_1 > 0$ that satisfies $\underline{V}^{[i]}(x(t_0)) \geq \alpha(\varepsilon_1) \geq \overline{V}^{[i]}(x(t_2))$. Then, there exists $\delta_1(\varepsilon_1) > 0$ that makes $\alpha(\varepsilon_1) \geq \beta(\delta_1) \geq \overline{V}^{[i]}(x(t_2))$. Then, we obtain

$$\underline{V}^{[i]}(x(t_0)) \geq \alpha(\varepsilon_1) \geq \beta(\delta_1) \geq \overline{V}^{[i]}(x(t_2)) \geq \overline{V}^{[i]}(x(t)) \geq \underline{V}^{[i]}(x(t)) \geq \alpha(\|x\|). \tag{9.21}$$

According to (9.19), we have

$$\alpha(\varepsilon) \geq \beta(\delta) \geq \underline{V}^{[i]}(x(t_0)) \geq \alpha(\varepsilon_1) \geq \beta(\delta_1) \geq \underline{V}^{[i]}(x(t)) \geq \alpha(\|x\|). \tag{9.22}$$

Since $\alpha(\|x\|)$ belongs to class $\mathcal{K}$, we obtain $\|x\| \leq \varepsilon$.

Therefore, we conclude that the system (9.1) is asymptotically stable. □

Corollary 9.7 *If for* $\forall i \geq 0$, $\mathrm{HJI}(\overline{V}^{[i]}(x), \overline{u}^{[i]}, \overline{w}^{[i]}) = 0$ *holds, and for* $\forall t$, $l(x, \overline{u}^{[i]}, \overline{w}^{[i]}) < 0$, *then the control pairs* $(\overline{u}^{[i]}, \overline{w}^{[i]})$ *make system* (9.1) *asymptotically stable.*

Theorem 9.8 (cf. [15]) *If for* $\forall i \geq 0$, $\mathrm{HJI}(\overline{V}^{[i]}(x), \overline{u}^{[i]}, \overline{w}^{[i]}) = 0$ *holds, and* $l(x, \overline{u}^{[i]}, \overline{w}^{[i]})$ *is the utility function, then the control pairs* $(\overline{u}^{[i]}, \overline{w}^{[i]})$ *make system* (9.1) *asymptotically stable.*

Proof For the time sequence $t_0 < t_1 < t_2 < \cdots < t_m < t_{m+1} < \cdots$, without loss of generality, we assume $l(x, \overline{u}^{[i]}, \overline{w}^{[i]}) \geq 0$ in $[t_{2n}, t_{(2n+1)})$ and $l(x, \overline{u}^{[i]}, \overline{w}^{[i]}) < 0$ in $[t_{2n+1}, t_{(2(n+1))})$ where $n = 0, 1, \ldots$.

Then, for $t \in [t_0, t_1)$ we have $l(x, \overline{u}^{[i]}, \overline{w}^{[i]}) \geq 0$ and $\int_{t_0}^{t_1} l(x, \overline{u}^{[i]}, \overline{w}^{[i]})\mathrm{d}t \geq 0$. According to Theorem 9.4, we have $\|x(t_0)\| \geq \|x(t)\| \geq \|x(t_1)\|$.

For $t \in [t_1, t_2)$ we have $l(x, \overline{u}^{[i]}, \overline{w}^{[i]}) < 0$ and $\int_{t_1}^{t_2} l(x, \overline{u}^{[i]}, \overline{w}^{[i]})\mathrm{d}t < 0$. According to Corollary 9.7, we have $\|x(t_1)\| > \|x(t)\| > \|x(t_2)\|$. So we obtain $\|x(t_0)\| \geq \|x(t)\| > \|x(t_2)\|$, for $\forall t \in [t_0, t_2)$.

Using mathematical induction, for $\forall t$, we have $\|x(t')\| \leq \|x(t)\|$ where $t' \in [t, \infty)$. So we conclude that the system (9.1) is asymptotically stable, and the proof is completed. □

Theorem 9.9 (cf. [15]) *If for* $\forall i \geq 0$, $\mathrm{HJI}(\underline{V}^{[i]}(x), \underline{u}^{[i]}, \underline{w}^{[i]}) = 0$ *holds, and* $l(x, \underline{u}^{[i]}, \underline{w}^{[i]})$ *is the utility function, then the control pairs* $(\underline{u}^{[i]}, \underline{w}^{[i]})$ *make system* (9.1) *asymptotically stable.*

Next, we will give the convergence proof of the iterative ADP algorithm.

Proposition 9.10 *If for* $\forall i \geq 0$, $\mathrm{HJI}(\overline{V}^{[i]}(x), \overline{u}^{[i]}, \overline{w}^{[i]}) = 0$ *holds, then the control pairs* $(\overline{u}^{[i]}, \overline{w}^{[i]})$ *make the upper value function* $\overline{V}^{[i]}(x) \to \bar{J}(x)$ *as* $i \to \infty$.

Proof According to $HJI(\overline{V}^{[i]}(x), \overline{u}^{[i]}, \overline{w}^{[i]}) = 0$, we obtain $\mathrm{d}\overline{V}^{[i+1]}(x)/\mathrm{d}t$ by replacing the index "i" by the index "$i+1$":

$$\begin{aligned}\frac{\mathrm{d}\overline{V}^{[i+1]}(x)}{\mathrm{d}t} = &-(x^{\mathrm{T}}Ax + \overline{u}^{(i+1)\mathrm{T}}(B - DC^{-1}D^{\mathrm{T}})\overline{u}^{[i+1]} \\ &+ \frac{1}{4}\overline{V}_x^{[i]T}k(x)C^{-1}k^{\mathrm{T}}(x)\overline{V}_x^{[i]} + 2x^{\mathrm{T}}(E - FC^{-1}D^{\mathrm{T}})\overline{u}^{[i+1]} \\ &- x^{\mathrm{T}}FC^{-1}F^{\mathrm{T}}x). \end{aligned} \tag{9.23}$$

According to (9.18), we obtain

$$\begin{aligned}\frac{\mathrm{d}(\overline{V}^{[i+1]}(x) - \overline{V}^{[i]}(x))}{\mathrm{d}t} &= \frac{\mathrm{d}\overline{V}^{[i+1]}(x)}{\mathrm{d}t} - \frac{\mathrm{d}\overline{V}^{[i]}(x)}{\mathrm{d}t} \\ &= (\overline{u}^{[i+1]} - \overline{u}^{[i]})^{\mathrm{T}}(B - DC^{-1}D^{\mathrm{T}}(\overline{u}^{[i+1]} - \overline{u}^{[i]}) \\ &> 0. \end{aligned} \tag{9.24}$$

Since the system (9.1) is asymptotically stable, its state trajectories x converge to zero, and so does $\overline{V}^{[i+1]}(x) - \overline{V}^{[i]}(x)$. Since $d(\overline{V}^{[i+1]}(x) - \overline{V}^{[i]}(x))/\mathrm{d}t \geq 0$

on these trajectories, it implies that $\overline{V}^{[i+1]}(x) - \overline{V}^{[i]}(x) \leq 0$; that is $\overline{V}^{[i+1]}(x) \leq \overline{V}^{[i]}(x)$. Thus, $\overline{V}^{[i]}(x)$ is convergent as $i \to \infty$.

Next, we define $\lim_{i\to\infty} \overline{V}^{[i]}(x) = \overline{V}^{[\infty]}(x)$.

For $\forall i$, let $\overline{w}^* = \arg\max_w\{\int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{[i]}(x(\hat{t}))\}$. Then, according to the principle of optimality, we have

$$\begin{aligned}\overline{V}^{[i]}(x) &\leq \sup_w \left\{ \int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{[i]}(x(\hat{t})) \right\} \\ &= \int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{[i]}(x(\hat{t})). \end{aligned} \tag{9.25}$$

Since $\overline{V}^{[i+1]}(x) \leq \overline{V}^{[i]}(x)$, we have $\overline{V}^{[\infty]}(x) \leq \int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{[i]}(x(\hat{t}))$.

Letting $i \to \infty$, we obtain $\overline{V}^{[\infty]}(x) \leq \int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{[\infty]}(x(\hat{t}))$. So, we have $\overline{V}^{[\infty]}(x) \leq \inf_u \sup_w\{\int_t^{\hat{t}} l(x,u,w)\mathrm{d}t + \overline{V}^{[i]}(x(\hat{t}))\}$.

Let $\epsilon > 0$ be an arbitrary positive number. Since the upper value function is non-increasing and convergent, there exists a positive integer i such that $\overline{V}^{[i]}(x) - \epsilon \leq \overline{V}^{[\infty]}(x) \leq \overline{V}^{[i]}(x)$.

Let $\overline{u}^* = \arg\min_u\{\int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{[i]}(x(\hat{t}))\}$. Then we get $\overline{V}^{[i]}(x) = \int_t^{\hat{t}} l(x,\overline{u}^*,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{[i]}(x(\hat{t}))$.

Thus, we have

$$\begin{aligned}\overline{V}^{[\infty]}(x) &\geq \int_t^{\hat{t}} l(x,\overline{u}^*,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{[i]}(x(\hat{t})) - \epsilon \\ &\geq \int_t^{\hat{t}} l(x,\overline{u}^*,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{[\infty]}(x(\hat{t})) - \epsilon \\ &= \inf_u \sup_w \left\{ \int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{[\infty]}(x(\hat{t})) \right\} - \epsilon. \end{aligned} \tag{9.26}$$

Since ϵ is arbitrary, we have

$$\overline{V}^{[\infty]}(x) \geq \inf_u \sup_w \left\{ \int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{[\infty]}(x(\hat{t})) \right\}.$$

Therefore, we obtain

$$\overline{V}^{[\infty]}(x) = \inf_u \sup_w \left\{ \int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{[\infty]}(x(\hat{t})) \right\}.$$

Let $\hat{t} \to \infty$, we have

$$\overline{V}^{[\infty]}(x) = \inf_u \sup_w J(x,u,w) = \overline{J}(x).$$

□

Proposition 9.11 *If for* $\forall i \geq 0$, $\mathrm{HJI}(\underline{V}^{[i]}(x), \underline{u}^{[i]}, \underline{w}^{[i]}) = 0$ *holds, then the control pairs* $(\underline{u}^{[i]}, \underline{w}^{[i]})$ *make the lower value function* $\underline{V}^{[i]}(x) \to \underline{J}(x)$ *as* $i \to \infty$.

Theorem 9.12 (cf. [15]) *If the saddle point of the zero-sum differential game exists, then the control pairs* $(\overline{u}^{[i]}, \overline{w}^{[i]})$ *and* $(\underline{u}^{[i]}, \underline{w}^{[i]})$ *make* $\overline{V}^{[i]}(x) \to J^*(x)$ *and* $\underline{V}^{[i]}(x) \to J^*(x)$, *respectively, as* $i \to \infty$.

Proof For the upper value function, according to Proposition 9.10, we have $\overline{V}^{[i]}(x) \to \overline{J}(x)$ under the control pairs $(\overline{u}^{[i]}, \overline{w}^{[i]})$ as $i \to \infty$. So the optimal control pair for the upper value function satisfies $\overline{J}(x) = J(x, \overline{u}, \overline{w}) = \inf_u \sup_w J(x, u, w)$.

On the other hand, there exists an optimal control pair (u^*, w^*) making the value reach the saddle point. According to the property of the saddle point, the optimal control pair (u^*, w^*) satisfies $J^*(x) = J(x, u^*, w^*) = \inf_u \sup_w J(x, u, w)$.

So, we have $\overline{V}^{[i]}(x) \to J^*(x)$ under the control pair $(\overline{u}^{[i]}, \overline{w}^{[i]})$ as $i \to \infty$. Similarly, we can derive $\underline{V}^{[i]}(x) \to J^*(x)$ under the control pairs $(\underline{u}^{[i]}, \underline{w}^{[i]})$ as $i \to \infty$. □

Remark 9.13 From the proofs we see that the complex existence conditions of the saddle point in [1, 2] are not necessary. If the saddle point exists, the iterative value functions can converge to the saddle point using the present iterative ADP algorithm.

In the following part, we emphasize that when the saddle point does not exist, the mixed optimal solution can be obtained effectively using the iterative ADP algorithm.

Proposition 9.14 *If* $\overline{u} \in \mathbb{R}^k$, $w^{[i]} \in \mathbb{R}^m$ *and the utility function is* $\tilde{l}(x, w^{[i]}) = l(x, \overline{u}, w^{[i]}) - l^o(x, \overline{u}, \overline{w}, \underline{u}, \underline{w})$, *and* $w^{[i]}$ *is expressed in* (9.14), *then the control pairs* $(\overline{u}, w^{[i]})$ *make the system* (9.1) *asymptotically stable.*

Proposition 9.15 *If* $\overline{u} \in \mathbb{R}^k$, $w^{[i]} \in \mathbb{R}^m$ *and for* $\forall t$, *the utility function* $\tilde{l}(x, w^{[i]}) \geq 0$, *then the control pairs* $(\overline{u}, w^{[i]})$ *make* $\widetilde{V}^{[i]}(x)$ *a nonincreasing convergent sequence as* $i \to \infty$.

Proposition 9.16 *If* $\overline{u} \in \mathbb{R}^k$, $w^{[i]} \in \mathbb{R}^m$ *and for* $\forall t$, *the utility function* $\tilde{l}(x, w^{[i]}) < 0$, *then the control pairs* $(\overline{u}, w^{[i]})$ *make* $\tilde{V}^{[i]}(x)$ *a nondecreasing convergent sequence as* $i \to \infty$.

Theorem 9.17 (cf. [15]) *If* $\overline{u} \in \mathbb{R}^k$, $w^{[i]} \in \mathbb{R}^m$, *and* $\tilde{l}(x, w^{[i]})$ *is the utility function, then the control pairs* $(\overline{u}, w^{[i]})$ *make* $\widetilde{V}^{[i]}(x)$ *convergent as* $i \to \infty$.

Proof For the time sequence $t_0 < t_1 < t_2 < \cdots < t_m < t_{m+1} < \cdots$, without loss of generality, we suppose $\tilde{l}(x, w^{[i]}) \geq 0$ in $[t_{2n}, t_{2n+1})$ and $\tilde{l}(x, w^{[i]}) < 0$ in $[t_{2n+1}, t_{2(n+1)})$, where $n = 0, 1, \ldots$.

For $t \in [t_{2n}, t_{2n+1})$ we have $\tilde{l}(x, w^{[i]}) \geq 0$ and $\int_{t_0}^{t_1} \tilde{l}(x, w^{[i]})\mathrm{d}t \geq 0$. According to Proposition 9.15, we have $\widetilde{V}^{[i+1]}(x) \leq \widetilde{V}^{[i]}(x)$. For $t \in [t_{2n+1}, t_{2(n+1)})$ we have $\tilde{l}(x, w^{[i]}) < 0$ and $\int_{t_1}^{t_2} \tilde{l}(x, w^{[i]})\mathrm{d}t < 0$. According to Proposition 9.16 we have $\widetilde{V}^{[i+1]}(x) > \widetilde{V}^{[i]}(x)$. Then, for $\forall t_0$, we have

$$\begin{aligned}\left\|\widetilde{V}^{[i+1]}(x(t_0))\right\| &= \left\|\int_{t_0}^{t_1} \tilde{l}(x, w^{[i]})\mathrm{d}t\right\| + \left\|\int_{t_1}^{t_2} \tilde{l}(x, w^{[i]})\mathrm{d}t\right\| \\ &\quad + \ldots + \left\|\int_{t_m}^{t_{(m+1)}} \tilde{l}(x, w^{[i]})\mathrm{d}t\right\| + \ldots \\ &< \left\|\widetilde{V}^{[i]}(x(t_0))\right\|. \end{aligned} \tag{9.27}$$

So, $\widetilde{V}^{[i]}(x)$ is convergent as $i \to \infty$. □

Theorem 9.18 (cf. [15]) *If $\overline{u} \in R^k$, $w^{[i]} \in R^m$, and $\tilde{l}(x, w^{[i]})$ is the utility function, then the control pairs $(\overline{u}, w^{[i]})$ make $\mathcal{V}^{[i]}(x) \to J^o(x)$ as $i \to \infty$.*

Proof It is proved by contradiction. Suppose that the control pair $(\overline{u}, w^{[i]})$ makes the value function $\mathcal{V}^{[i]}(x)$ converge to $\mathcal{J}'(x)$ and $\mathcal{J}'(x) \neq J^o(x)$.

According to Theorem 9.17, based on the principle of optimality, as $i \to \infty$ we have the HJB equation $\mathrm{HJB}(\widetilde{J}(x), w) = 0$.

From the assumptions we know that $|\mathcal{V}^{[i]}(x) - J^o(x)| \neq 0$ as $i \to \infty$. From Theorem 9.5, we know that there exists a control pair $(\overline{u}, w')$ that makes $J(x, \overline{u}, w') = J^o(x)$, which minimizes the performance index function $\widetilde{J}(x)$. According to the principle of optimality, we also have the HJB equation $\mathrm{HJB}(\widetilde{J}(x), w') = 0$.

It is a contradiction. So the assumption does not hold. Thus, we have $\mathcal{V}^{[i]}(x) \to J^o(x)$ as $i \to \infty$. □

Remark 9.19 For the situation where the saddle point does not exist, the methods in [1, 2] are all invalid. Using our iterative ADP method, the iterative value function reaches the mixed optimal value function $J^o(x)$ under the deterministic control pair. Therefore, we emphasize that the present iterative ADP method is more effective.

9.2.3 Simulations

Example 9.20 The dynamics of the benchmark nonlinear plant can be expressed by system (9.1) where

$$f(x) = \left[\begin{array}{cccc} x_2 & \dfrac{-x_1+\varepsilon x_4^2 \sin x_3}{1-\varepsilon^2\cos^2 x_3} & x_4 & \dfrac{\varepsilon \cos x_3 (x_1-\varepsilon x_4^2 \sin x_3)}{1-\varepsilon^2\cos^2 x_3} \end{array}\right]^{\mathrm{T}},$$

$$g(x) = \left[\begin{array}{cccc} 0 & \dfrac{-\varepsilon \cos x_3}{1-\varepsilon^2\cos^2 x_3} & 0 & \dfrac{1}{1-\varepsilon^2\cos^2 x_3} \end{array}\right]^{\mathrm{T}},$$

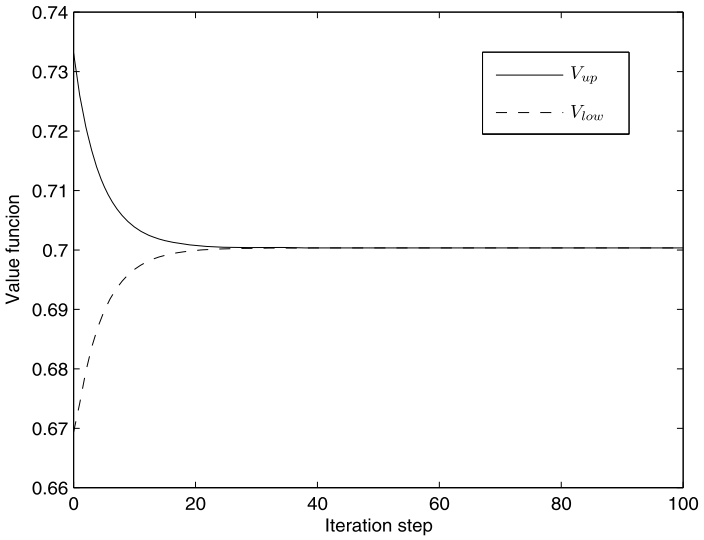

Fig. 9.1 Trajectories of upper and lower value function

$$k(x)=\left[0\ \ \frac{1}{1-\varepsilon^2\cos^2 x_3}\ \ 0\ \ \frac{-\varepsilon\cos x_3}{1-\varepsilon^2\cos^2 x_3}\right]^{\mathrm{T}}, \tag{9.28}$$

and $\varepsilon = 0.2$. The initial state is given as $x(0) = [1, 1, 1, 1]^{\mathrm{T}}$. The cost functional is defined by (9.2) where the utility function is expressed as $l(x, u, w) = x_1^2 + 0.1x_2^2 + 0.1x_3^2 + 0.1x_4^2 + \|u\|^2 - \gamma^2\|w\|^2$ and $\gamma^2 = 10$.

Any differential structure can be used to implement the iterative ADP method. For facilitating the implementation of the algorithm, we choose three-layer neural networks as the critic networks with the structure of 4–8–1. The structures of the u and w for the upper value function are 4–8–1 and 5–8–1; while they are 5–8–1 and 4–8–1 for the lower one. The initial weights are all randomly chosen in $[-0.1,\ 0.1]$. Then, for each i, the critic network and the action networks are trained for 1000 time steps so that the given accuracy $\zeta = 10^{-6}$ is reached. Let the learning rate $\eta = 0.01$. The iterative ADP method runs for $i = 70$ times and the convergence trajectory of the value function is shown in Fig. 9.1. We can see that the saddle point of the game exists. Then, we apply the controller to the benchmark system and run for $T_f = 60$ seconds. The optimal control trajectories are shown in Fig. 9.2. The corresponding state trajectories are shown in Figs. 9.3 and 9.4, respectively.

Remark 9.21 The simulation results illustrate the effectiveness of the present iterative ADP algorithm. If the saddle point exists, the iterative control pairs $(\overline{u}^{[i]}, \overline{w}^{[i]})$ and $(\underline{u}^{[i]}, \underline{w}^{[i]})$ can make the iterative value functions reach the saddle point, while the existence conditions of the saddle point are avoided.

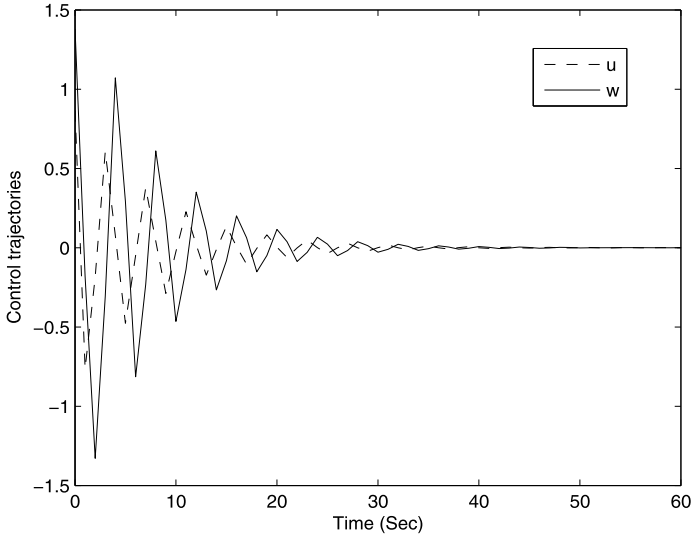

Fig. 9.2 Trajectories of the controls

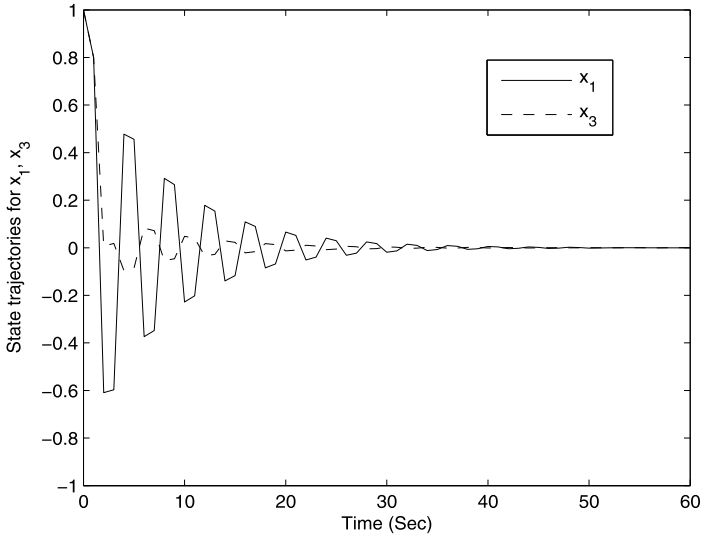

Fig. 9.3 Trajectories of state x_1 and x_3

Example 9.22 In this example, we just change the utility function to

$$l(x,u,w) = x_1^2 + 0.1x_2^2 + 0.1x_3^2 + 0.1x_4^2 + \|u\|^2 - \gamma^2\|w\|^2 - 0.1uw + 0.1x^{\mathrm{T}}u + 0.1x^{\mathrm{T}}w,$$

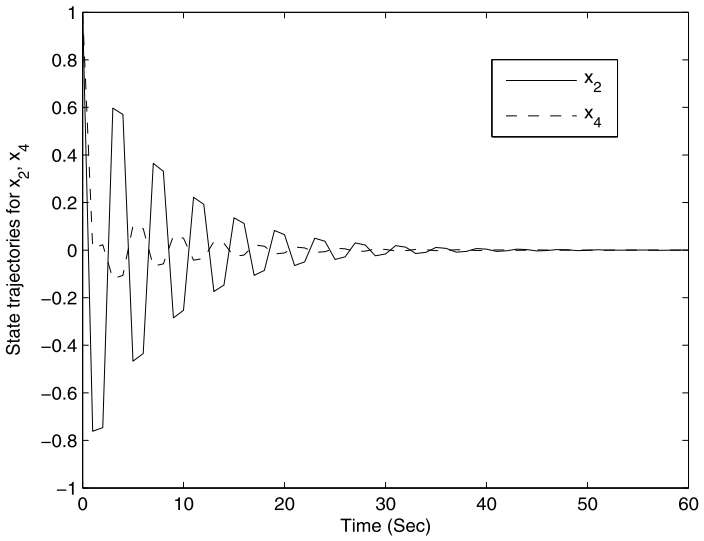

Fig. 9.4 Trajectories of state x_2 and x_4

and all other conditions are the same as the ones in Example 9.20. We obtain $\overline{J}(x(0)) = 0.65297$ and $\underline{J}(x(0)) = 0.44713$, with trajectories shown in Figs. 9.5(a) and (b), respectively. Obviously, the saddle point does not exist. Thus, the method in [1] is invalid. Using the present mixed trajectory method, we choose the Gaussian noises $\gamma_u(0, 0.05^2)$ and $\gamma_w(0, 0.05^2)$. Let $N = 5000$ times. The value function trajectories are shown in Fig. 9.5(c). Then, we obtain the value of the mixed optimal value function $J^o(x(0)) = 0.55235$ and then $\alpha = 0.5936$. Regulating the control w to obtain the trajectory of the mixed optimal value function displayed in Fig. 9.5. The state trajectories are shown in Figs. 9.6(a) and 9.7, respectively. The corresponding control trajectories are shown in Figs. 9.8 and 9.9, respectively.

9.3 Finite Horizon Zero-Sum Games for a Class of Nonlinear Systems

In this section, a new iterative approach is derived to solve optimal policies of finite horizon quadratic zero-sum games for a class of continuous-time nonaffine nonlinear system. Through the iterative algorithm between two sequences, which are a sequence of state trajectories of linear quadratic zero-sum games and a sequence of corresponding Riccati differential equations, the optimal policies for nonaffine nonlinear zero-sum games are given. Under very mild conditions of local Lipschitz continuity, the convergence of approximating linear time-varying sequences is proved.

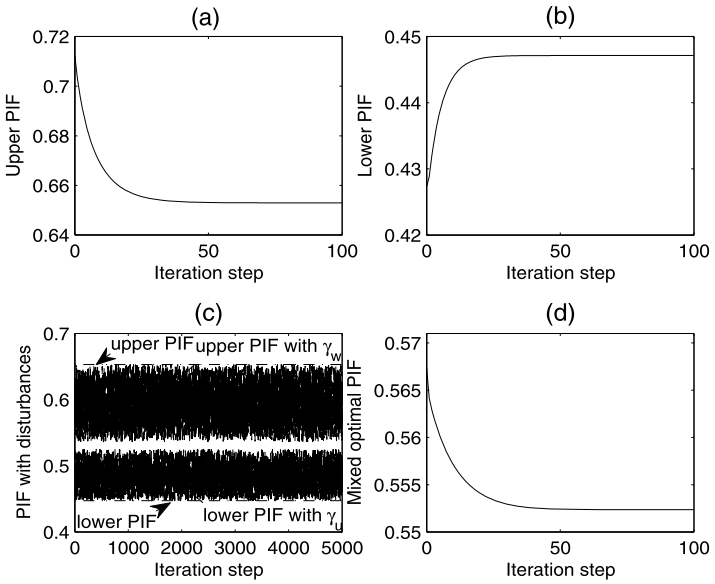

Fig. 9.5 Performance index function trajectories. (**a**) Trajectory of upper value function. (**b**) Trajectory of lower value function. (**c**) Performance index functions with disturbances. (**d**) Trajectory of the mixed optimal performance index function

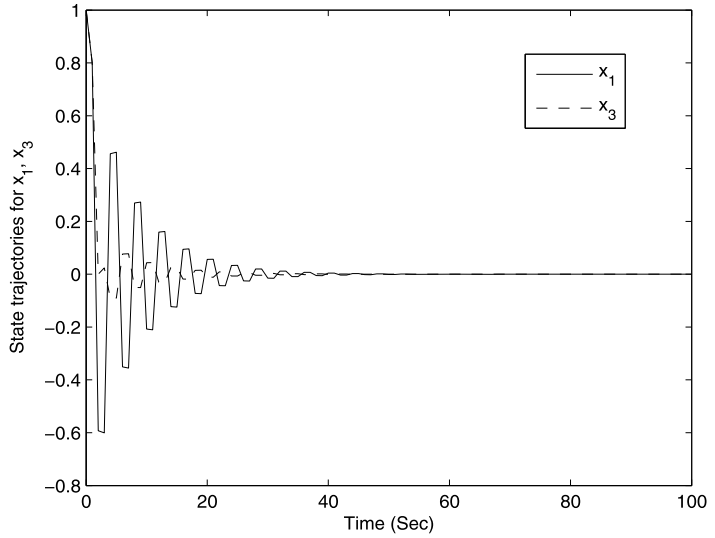

Fig. 9.6 Trajectories of state x_1 and x_3

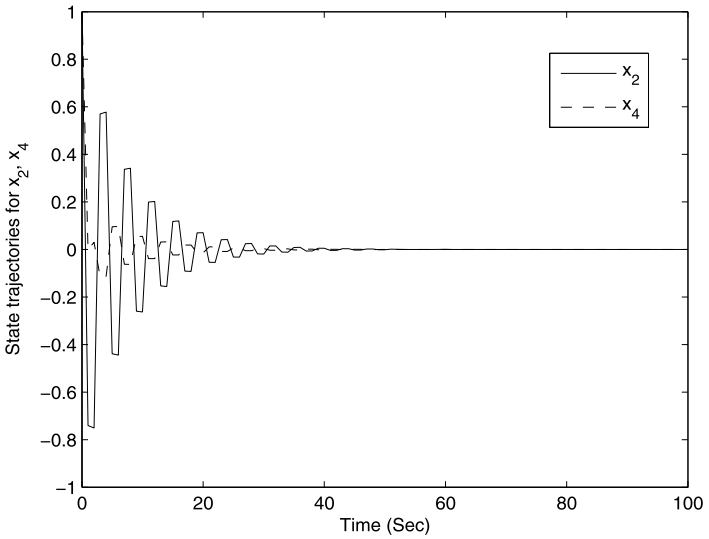

Fig. 9.7 Trajectories of state x_2 and x_4

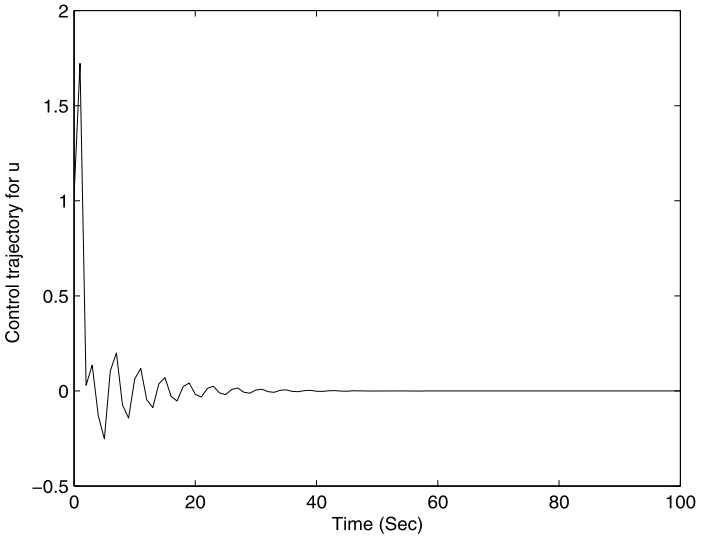

Fig. 9.8 Trajectory of control u

9.3.1 Problem Formulation

Consider a continuous-time nonaffine nonlinear zero-sum game described by the state equation

$$\dot{x}(t) = f(x(t), u(t), w(t)),\ x(t_0) = x_0 \tag{9.29}$$

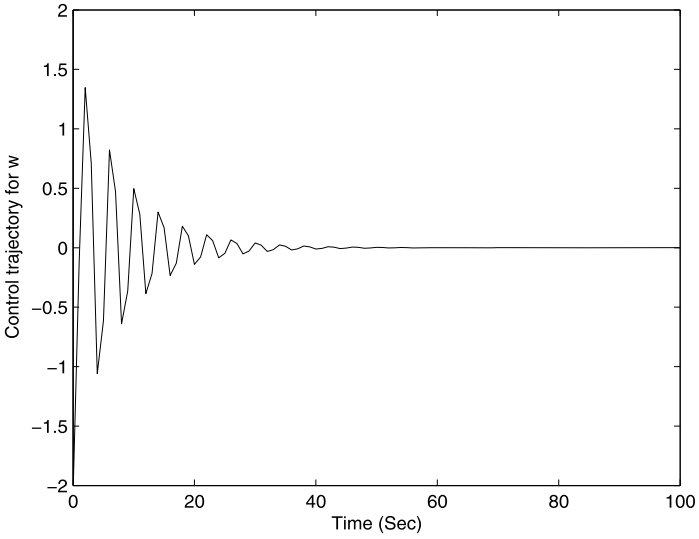

Fig. 9.9 Trajectory of control w

with the finite horizon cost functional

$$\begin{aligned} J(x_0, u, w) = &\frac{1}{2}x^{\mathrm{T}}(t_f)F(x(t_f))x(t_f) \\ &+ \frac{1}{2}\int_{t_0}^{t_f} \left[x^{\mathrm{T}}(t)Q(x(t))x(t) + u^{\mathrm{T}}(t)R(x(t))u(t)\right. \\ &\left. - w^{\mathrm{T}}(t)S(x(t))w(t)\right]\mathrm{d}t, \end{aligned} \tag{9.30}$$

where $x(t) \in \mathbb{R}^n$ is the state, $x(t_0) \in \mathbb{R}^n$ is the initial state, t_f is the terminal time, the control input $u(t)$ takes values in a convex and compact set $U \subset \mathbb{R}^{m_1}$, and $w(t)$ takes values in a convex and compact set $W \subset \mathbb{R}^{m_2}$. $u(t)$ seeks to minimize the cost functional $J(x_0, u, w)$, while $w(t)$ seeks to maximize it. The state-dependent weight matrices $F(x(t))$, $Q(x(t))$, $R(x(t))$, $S(x(t))$ are with suitable dimensions and $F(x(t)) \geq 0$, $Q(x(t)) \geq 0$, $R(x(t)) > 0$, $S(x(t)) > 0$. In this section, $x(t)$, $u(t)$, and $w(t)$ sometimes are described by x, u, and w for brevity. Our objective is to find the optimal policies for the above nonaffine nonlinear zero-sum games.

In the nonaffine nonlinear zero-sum game problem, nonlinear functions are implicit function with respect to controller input. It is very hard to obtain the optimal policies satisfying (9.29) and (9.30). For practical purposes one may just as well be interested in finding a near-optimal or an approximate optimal policy. Therefore, we present an iterative algorithm to deal with this problem. Nonaffine nonlinear zero-sum games are transformed into an equivalent sequence of linear quadratic zero-sum games which can use the linear quadratic zero-sum game theory directly.

9.3.2 Finite Horizon Optimal Control of Nonaffine Nonlinear Zero-Sum Games

Using a factored form to represent the system (9.29), we get

$$\begin{aligned}\dot{x}(t) &= f(x(t))x(t) + g(x(t), u(t))u(t) + k(x(t), w(t))w(t),\\ x(t_0) &= x_0,\end{aligned} \tag{9.31}$$

where $f: \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is a nonlinear matrix-valued function of x, $g: \mathbb{R}^n \times \mathbb{R}^{m_1} \to \mathbb{R}^{n\times m_1}$ is a nonlinear matrix-valued function of both the state x and control input u, and $k: \mathbb{R}^n \times \mathbb{R}^{m_2} \to \mathbb{R}^{n\times m_2}$ is a nonlinear matrix-valued function of both the state x and control input w.

We use the following sequence of linear time-varying differential equations to approximate the state equation (9.31):

$$\begin{aligned}\dot{x}_i(t) &= f(x_{i-1}(t))x_i(t) + g(x_{i-1}(t), u_{i-1}(t))u_i(t) + k(x_{i-1}(t), w_{i-1}(t))w_i(t),\\ x_i(t_0) &= x_0,\ i \geq 0,\end{aligned} \tag{9.32}$$

with the corresponding cost functional

$$\begin{aligned}V_i(x_0, u, w) = &\frac{1}{2}x_i^T(t_f)F(x_{i-1}(t_f))x_i(t_f)\\ &+ \frac{1}{2}\int_{t_0}^{t_f}\Big[x_i^T(t)Q(x_{i-1}(t))x_i(t) + u_i^T(t)R(x_{i-1}(t))u_i(t)\\ &- w_i^T(t)S(x_{i-1}(t))w_i(t)\Big]\mathrm{d}t,\ i \geq 0,\end{aligned} \tag{9.33}$$

where the superscript i represents the iteration index. For the first approximation, $i = 0$, we assume that the initial values $x_{i-1}(t) = x_0$, $u_{i-1}(t) = 0$, and $w_{i-1}(t) = 0$. Obviously, for the ith iteration, $f(x_{i-1}(t))$, $g(x_{i-1}(t), u_{i-1}(t))$, $k(x_{i-1}(t), w_{i-1}(t))$, $F(x_{i-1}(t_f))$, $Q(x_{i-1}(t))$, $R(x_{i-1}(t))$, and $S(x_{i-1}(t))$ are time-varying functions which do not depend on $x_i(t)$, $u_i(t)$, and $w_i(t)$. Hence, each approximation problem in (9.32) and (9.33) is a linear quadratic zero-sum game problem which can be solved by the existing classical linear quadratic zero-sum game theory.

The corresponding Riccati differential equation of each linear quadratic zero-sum game can be expressed as

$$\begin{aligned}\dot{P}_i(t) = &-Q(x_{i-1}(t)) - P_i(t)f(x_{i-1}(t)) - f^{\mathrm{T}}(x_{i-1}(t))P_i(t)\\ &+ P_i(t)\big[g(x_{i-1}(t), u_{i-1}(t))R^{-1}(x_{i-1}(t))\\ &\times g^{\mathrm{T}}(x_{i-1}(t), u_{i-1}(t)) - k(x_{i-1}(t), w_{i-1}(t))\\ &\times S^{-1}(x_{i-1}(t))k^{\mathrm{T}}(x^{[i-1]})(t), w_{i-1}(t))\big]P_i(t),\\ P_i(t_f) = &F(x_{i-1}(t_f)),\ i \geq 0,\end{aligned} \tag{9.34}$$

where $P_i \in \mathbb{R}^{n\times n}$ is a real, symmetric and nonnegative definite matrix.

Assumption 9.23 It is assumed that $S(x_{i-1}(t)) > \hat{S}_i$, where the threshold value $\hat{S}_i$ is defined as $\hat{S}_i = \inf\{S_i(t) > 0,$ and (9.34) does not have a conjugate point on $[0, t_f]\}$.

If Assumption 9.23 is satisfied, the game admits the optimal policies given by

$$\begin{aligned} u_i(t) &= -R^{-1}(x_{i-1}(t))g^{\mathrm{T}}(x_{i-1}(t), u_{i-1}(t))P_i(t)x_i(t), \\ w_i(t) &= S^{-1}(x_{i-1}(t))k^{\mathrm{T}}(x_{i-1}(t), w_{i-1}(t))P_i(t)x_i(t),\ i \geq 0, \end{aligned} \tag{9.35}$$

where $x_i(t)$ is the corresponding optimal state trajectory, generated by

$$\begin{aligned} \dot{x}_i(t) =& \big[f(x_{i-1}(t)) - g(x_{i-1}(t), u_{i-1}(t))R^{-1}(x_{i-1}(t)) \\ &\times g^{\mathrm{T}}(x_{i-1}(t), u_{i-1}(t))P_i(t) + k(x_{i-1}(t), w_{i-1}(t)) \\ &\times S^{-1}(x_{i-1}(t))k^{\mathrm{T}}(x_{i-1}(t), w_{i-1}(t))P_i(t)\big]x_i(t), \\ x_i(t_0) =& x_0. \end{aligned} \tag{9.36}$$

By using the iteration between sequences (9.34) and (9.36) sequently, the limit of the solution of the approximating sequence (9.32) will converge to the unique solution of system (9.29), and the sequences of optimal policies (9.35) will converge, too. The convergence of iterative algorithm will be analyzed in the next section. Notice that the factored form in (9.31) does not need to be unique. The approximating linear time-varying sequences will converge whatever the representation of $f(x(t))$, $g(x(t), u(t))$, and $k(x(t), w(t))$.

Remark 9.24 For the fixed finite interval $[t_0, t_f]$, if $S(x_{i-1}(t)) > \hat{S}_i$, the Riccati differential equation (9.34) has a conjugate point on $[t_0, t_f]$. It means that $V_i(x_0, u, w)$ is strictly concave in w. Otherwise, since $V_i(x_0, u, w)$ is quadratic and $R(t) > 0$, $F(t) \geq 0$, $Q(t) \geq 0$, it follows that $V_i(x_0, u, w)$ is strictly convex in u. Hence, for linear quadratic zero-sum games (9.32) with the performance index function (9.34) there exists a unique saddle point; they are the optimal policies.

The convergence of the algorithm described above requires the following:

1. The sequence $\{x_i(t)\}$ converges on $C([t_0, t_f]; \mathbb{R}^n)$, which means that the limit of the solution of approximating sequence (9.32) converges to the unique solution of system (9.29).
2. The sequences of optimal policies $\{u_i(t)\}$ and $\{w_i(t)\}$ converge on $C([t_0, t_f]; \mathbb{R}^{m1})$ and $C([t_0, t_f]; \mathbb{R}^{m2})$, respectively.

For simplicity, the approximating sequence (9.32) is rewritten as

$$\begin{aligned} &\dot{x}_i(t) = f(x_{i-1}(t))x_i(t) + G(x_{i-1}(t), u_{i-1}(t))x_i(t) + K(x_{i-1}(t), w_{i-1}(t))x_i(t), \\ &x_i(t_0) = x_0,\ i \geq 0, \end{aligned} \tag{9.37}$$

where

$$G(x_{i-1}, u_{i-1}) \triangleq -g(x_{i-1}(t), u_{i-1}(t))R^{-1}(x_{i-1}(t))^{\mathrm{T}}(x_{i-1}(t), u_{i-1}(t))P_i(t),$$

$$K(x_{i-1}, w_{i-1}) \triangleq k(x_{i-1}(t), w_{i-1}(t))S^{-1}(x_{i-1}(t))k^{\mathrm{T}}(x_{i-1}(t), w_{i-1}(t))P_i(t).$$

The optimal policies for zero-sum games are rewritten as

$$\begin{aligned} u_i(t) &= M(x_{i-1}(t), u_{i-1}(t))x_i(t), \\ w_i(t) &= N(x_{i-1}(t), w_{i-1}(t))x_i(t), \end{aligned} \tag{9.38}$$

where

$$M(x_{i-1}, u_{i-1}) \triangleq -R^{-1}(x_{i-1}(t))g^{\mathrm{T}}(x_{i-1}(t), u_{i-1}(t))P_i(t),$$

$$N(x_{i-1}, w_{i-1}) \triangleq S^{-1}(x_{i-1}(t))k^{\mathrm{T}}(x_{i-1}(t), w_{i-1}(t))P_i(t).$$

Assumption 9.25 $g(x,u)$, $k(x,w)$, $R^{-1}(x)$, $S^{-1}(x)$, $F(x)$ and $Q(x)$ are bounded and Lipschitz continuous in their arguments x, u, and w, thus satisfying:

(C1) $\|g(x,u)\| \le b$, $\|k(x,u)\| \le e$
(C2) $\|R^{-1}(x)\| \le r$, $\|S^{-1}(x)\| \le s$
(C3) $\|F(x)\| \le f$, $\|Q(x)\| \le q$

for $\forall x \in \mathbb{R}^n$, $\forall u \in \mathbb{R}^{m_1}$, $\forall w \in \mathbb{R}^{m_2}$, and for finite positive numbers b, e, r, s, f, and q.

Define $\Phi_{i-1}(t, t_0)$ as the transition matrix generated by $f_{i-1}(t)$. It is well known that

$$\|\Phi_{i-1}(t, t_0)\| \le \exp\left[\int_{t_0}^{t} \mu(f(x_{i-1}(\tau)))\mathrm{d}\tau\right], \ t \ge t_0, \tag{9.39}$$

where $\mu(f)$ is the measure of matrix f, $\mu(f) = \lim_{h\to 0+} \frac{\|I+hf\|-1}{h}$. We use the following lemma to get an estimate for $\Phi_{i-1}(t, t_0) - \Phi_{i-2}(t, t_0)$.

The following lemma is relevant for the solution of the Riccati differential equation (9.34), which is the basis for proving the convergence.

Lemma 9.26 *Let Assumption 9.25 hold; the solution of the Riccati differential equation (9.34) satisfies*:

1. $P_i(t)$ *is Lipschitz continuous.*
2. $P_i(t)$ *is bounded, if the linear time-varying system (9.32) is controllable.*

Proof First, let us prove that $P_i(t)$ is Lipschitz continuous. We transform (9.34) into the form of a matrix differential equation:

$$\begin{bmatrix} \dot{\lambda}_i(t) \\ \dot{X}_i(t) \end{bmatrix} = \begin{bmatrix} -f(x_{i-1}(t)) & -Q(x_{i-1}(t)) \\ \Xi & f(x_{i-1}(t)) \end{bmatrix} \begin{bmatrix} \lambda_i(t) \\ X_i(t) \end{bmatrix}, \begin{bmatrix} \lambda_i(t_f) \\ X_i(t_f) \end{bmatrix} = \begin{bmatrix} F(t_f) \\ I \end{bmatrix},$$

where

$$\Xi = g(x_{i-1}(t), u_{i-1}(t))R^{-1}(x_{i-1}(t))g^{\mathrm{T}}(x_{i-1}(t), u_{i-1}(t)) - k(x_{i-1}(t), w_{i-1}(t))S^{-1}(x_{i-1}(t))k^{\mathrm{T}}(x_{i-1}(t), w_{i-1}(t)).$$

Thus, the solution $P_i(t)$ of the Riccati differential equations (9.34) becomes

$$P_i(t) = \lambda_i(t)\,(X_i(t))^{-1}. \tag{9.40}$$

If Assumption 9.25 is satisfied, such that $f(x)$, $g(x,u)$, $k(x,w)$, $R^{-1}(x)$, $S^{-1}(x)$, $F(x)$, and $Q(x)$ are Lipschitz continuous, then $X_i(t)$ and $\lambda_i(t)$ are Lipschitz continuous. Furthermore, it is easy to verify that $(X_i(t))^{-1}$ also satisfies the Lipschitz condition. Hence, $P_i(t)$ is Lipschitz continuous.

Next, we prove that $P_i(t)$ is bounded.

If the linear time varying system (9.32) is controllable, there must exist $\hat{u}_i(t)$, $\hat{w}_i(t)$ such that $x(t_1) = 0$ at $t = t_1$. We define $\bar{u}_i(t)$, $\bar{w}_i(t)$ as

$$\bar{u}_i(t) = \begin{cases} \hat{u}_i(t), & t \in [0, t_1) \\ 0, & t \in [t_1, \infty) \end{cases}$$

$$\bar{w}_i(t) = \begin{cases} \hat{w}_i(t) = S^{-1}(x_{i-1}(t))k^{\mathrm{T}}(x_{i-1}(t), w_{i-1}(t))P_i(t)x_i(t), & t \in [0, t_1) \\ 0, & t \in [t_1, \infty) \end{cases}$$

where $\hat{u}_i(t)$ is any control policy making $x(t_1) = 0$, $\hat{w}_i(t)$ is defined as the optimal policy. We have $t \geq t_1$, and we let $\bar{u}_i(t)$ and $\bar{w}_i(t)$ be 0, the state $x(t)$ will still hold at 0.

The optimal cost functional $V_i^*(x_0, u, w)$ described as

$$V_i^*(x_0,u,w) = \frac{1}{2}x_i^{\mathrm{T}}(t_f)F(x_{i-1}(t_f))x_i(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\big[x_i^{\mathrm{T}}(t)Q(x_{i-1}(t))x_i(t) + u_i^{*\mathrm{T}}(t)R(x_{i-1}(t))u_i^*(t) - w_i^{*\mathrm{T}}(t)S(x_{i-1}(t))w_i^*(t)\big]\mathrm{d}t, \tag{9.41}$$

where $u_i^*(t)$ and $w_i^*(t)$ are the optimal policies. $V_i^*(x_0, u, w)$ is minimized by $u^*(t)$ and maximized by $w_i^*(t)$.

For the linear system, $V_i^*(x_0, u, w)$ can be expressed as $V_i^*(x_0, u, w) = 1/(2x_i^{\mathrm{T}}(t)P_i(t)x_i(t))$. Since $x_i(t)$ is arbitrary, if $V_i^*(x_0, u, w)$ is bounded, then $P_i(t)$ is bounded. Next, we discuss the boundedness of $V_i^*(x_0, u, w)$ in two cases:

Case 1: $t_1 < t_f$; we have

$$V_i^*(x_0,u,w) \leq \frac{1}{2}x_i^{\mathrm{T}}(t_f)F(x_{i-1}(t_f))x_i(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\big[x_i^{\mathrm{T}}(t)Q(x_{i-1}(t))x_i(t) + \hat{u}_i^{\mathrm{T}}(t)R(x_{i-1}(t))\hat{u}_i(t) - w_i^{*\mathrm{T}}(t)S(x_{i-1}(t))w_i^*(t)\big]\mathrm{d}t$$

$$
\begin{aligned}
&= \frac{1}{2}\int_{t_0}^{t_1} \big[x_i^{\mathrm{T}}(t)Q(x_{i-1}(t))x_i(t) \\
&\quad + \hat{u}_i^{\mathrm{T}}(t)R(x_{i-1}(t))\hat{u}_i(t) - w_i^{*\mathrm{T}}(t)S(x_{i-1}(t))w_i^*(t)\big]\mathrm{d}t \\
&= V_{t_1 i}(x) \\
&< \infty. \qquad (9.42)
\end{aligned}
$$

Case 2: $t_1 \geq t_f$; we have

$$
\begin{aligned}
V_i^*(x_0,u,w) &\leq \frac{1}{2}x_i^{\mathrm{T}}(t_f)F(x_{i-1}(t_f))x_i(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \big[x_i^{\mathrm{T}}(t)Q(x_{i-1}(t))x_i(t) \\
&\quad + \hat{u}_i^{\mathrm{T}}(t)R(x_{i-1}(t))\hat{u}_i(t) - w_i^{*\mathrm{T}}(t)S(x_{i-1}(t))w_i^*(t)\big]\mathrm{d}t \\
&\leq \frac{1}{2}\int_{t_0}^{\infty} \big[x_i^{\mathrm{T}}(t)Q(x_{i-1}(t))x_i(t) \\
&\quad + \bar{u}_i^{\mathrm{T}}(t)R(x_{i-1}(t))\bar{u}_i(t) - \bar{w}_i^{\mathrm{T}}(t)S(x_{i-1}(t))\bar{w}_i(t)\big]\mathrm{d}t \\
&= \frac{1}{2}\int_{t_0}^{t_1} \big[x_i^{\mathrm{T}}(t)Q(x_{i-1}(t))x_i(t) \\
&\quad + \hat{u}_i^{\mathrm{T}}(t)R(x_{i-1}(t))\hat{u}_i(t) - w_i^{*\mathrm{T}}(t)S(x_{i-1}(t))w_i^*(t)\big]\mathrm{d}t \\
&= V_{t_1 i}(x) \\
&< \infty. \qquad (9.43)
\end{aligned}
$$

From (9.42) and (9.43), we know that $V_i^*(x)$ has an upper bound, independent of t_f. Hence, $P_i(t)$ is bounded. □

According to Lemma 9.26, $P_i(t)$ is bounded and Lipschitz continuous. If Assumption 9.25 is satisfied, then $M(x,u)$, $N(x,w)$, $G(x,w)$, and $K(x,w)$ are bounded and Lipschitz continuous in their arguments, thus satisfying:

(C4) $\|M(x,u)\| \leq \delta_1$, $\|N(x,w)\| \leq \sigma_1$,
(C5) $\|M(x_1,u_1) - M(x_2,u_2)\| \leq \delta_2\|x_1 - x_2\| + \delta_3\|u_1 - u_2\|$, $\|N(x_1,w_1) - N(x_2,w_2)\| \leq \sigma_2\|x_1 - x_2\| + \sigma_3\|w_1 - w_2\|$,
(C6) $\|G(x,u)\| \leq \zeta_1$, $\|K(x,w)\| \leq \xi_1$,
(C7) $\|G(x_1,u_1) - G(x_2,u_2)\| \leq \zeta_2\|x_1 - x_2\| + \zeta_3\|u_1 - u_2\|$, $\|K(x_1,w_1) - K(x_2,w_2)\| \leq \xi_2\|x_1 - x_2\| + \xi_3\|w_1 - w_2\|$,

$\forall x \in \mathbb{R}^n$, $\forall u \in \mathbb{R}^{m_1}$, $\forall w \in \mathbb{R}^{m_2}$, and for finite positive numbers δ_j, σ_j, ζ_j, ξ_j, $j = 1,2,3$.

Theorem 9.27 (cf. [16]) *Consider the system* (9.29) *of nonaffine nonlinear zero-sum games with the cost functional* (9.30), *the approximating sequences* (9.32) *and* (9.33) *can be introduced. We have* $F(x(t)) \geq 0$, $Q(x(t)) \geq 0$, $R(x(t)) > 0$, *and the terminal time* t_f *is specified. Let Assumption* 9.25, *and Assumptions* (A1) *and* (A2)

hold and $S(x(t)) > \tilde{S}$, *for small enough* t_f *or* x_0; *then the limit of the solution of the approximating sequence* (9.32) *converges to the unique solution of system* (9.29) *on* $C([t_0, t_f]; \mathbb{R}^n)$. *Meanwhile, the approximating sequences of optimal policies given by* (9.35) *also converge on* $C([t_0, t_f]; \mathbb{R}^{m_1})$ *and* $C([t_0, t_f]; \mathbb{R}^{m_2})$, *if*

$$\|\Psi(t)\| < 1, \tag{9.44}$$

where

$$\Psi(t) = \begin{bmatrix} \psi_1 & \psi_2 & \psi_3 \\ \psi_4 & \psi_5 & \psi_6 \\ \psi_7 & \psi_8 & \psi_9 \end{bmatrix},$$

$$\psi_1(t) = \frac{\left\{\left[\frac{\zeta_2+\xi_2}{\zeta_1+\xi_1} + \alpha(t-t_0)\right] e^{(\zeta_1+\xi_1)(t-t_0)} - \frac{\zeta_2+\xi_2}{\zeta_1+\xi_1}\right\}}{\left(1 + \frac{\zeta_1+\xi_1}{\mu_0}\left(1 - e^{\mu_0(t-t_0)}\right)\right)} \|x_0\| e^{\mu_0(t-t_0)},$$

$$\psi_2(t) = \frac{\frac{\zeta_3}{\zeta_1+\xi_1} \|x_0\| e^{\mu_0(t-t_0)}(e^{(\zeta_1+\xi_1)(t-t_0)} - 1)}{\left(1 + \frac{\zeta_1+\xi_1}{\mu_0}\left(1 - e^{\mu_0(t-t_0)}\right)\right)},$$

$$\psi_3(t) = \frac{\frac{\xi_3}{\zeta_1+\xi_1} \|x_0\| e^{\mu_0(t-t_0)}(e^{(\zeta_1+\xi_1)(t-t_0)} - 1)}{\left(1 + \frac{\zeta_1+\xi_1}{\mu_0}\left(1 - e^{\mu_0(t-t_0)}\right)\right)},$$

$$\psi_4(t) = \delta_1\psi_1(t) + \delta_2 \|x_0\| e^{(\mu_0+\zeta_1+\xi_1)(t-t_0)},$$

$$\psi_5(t) = \delta_1\psi_2(t) + \delta_3 \|x_0\| e^{(\mu_0+\zeta_1+\xi_1)(t-t_0)},$$

$$\psi_6(t) = \delta_1\psi_3(t),$$

$$\psi_7(t) = \sigma_1\psi_1(t) + \sigma_2 \|x_0\| e^{(\mu_0+\zeta_1+\xi_1)(t-t_0)},$$

$$\psi_8(t) = \sigma_1\psi_2(t),$$

$$\psi_9(t) = \sigma_1\psi_3(t) + \sigma_3 \|x_0\| e^{(\mu_0+\zeta_1+\xi_1)(t-t_0)},$$

$$\tilde{S} = \max\{\hat{S}_i\}.$$

Proof The approximating sequence (9.37) is an nonhomogeneous differential equation, whose solution can be given by

$$x_i(t) = \Phi_{i-1}(t, t_0)x_i(t_0) + \int_{t_0}^{t} \Phi_{i-1}(t, s)\left[G(x_{i-1}(s), u_{i-1}(s)) + K(x_{i-1}(s), w_{i-1}(s))\right] x_i(s)\mathrm{d}s. \tag{9.45}$$

Then,

$$\|x_i(t)\| \le \|\Phi_{i-1}(t, t_0)\| \|x_i(t_0)\| + \int_{t_0}^{t} \|\Phi_{i-1}(t, s)\| \left[\|G(x_{i-1}, u_{i-1})\| \right.$$

$$+ \|K(x_{i-1}, w_{i-1})\| \big] \|x_i(s)\| \, \mathrm{d}s. \tag{9.46}$$

According to inequality (9.39) and assuming (C6) to hold, we obtain

$$e^{-\mu_0 t} \|x_i(t)\| \leq e^{-\mu_0 t_0} \|x_0\| + \int_{t_0}^{t} (\zeta_1 + \xi_1) e^{-\mu_0 s} \|x_i(s)\| \, \mathrm{d}s. \tag{9.47}$$

On the basis of Gronwall–Bellman's inequality

$$\|x_i(t)\| \leq \|x_0\| \, e^{(\mu_0 + \zeta_1 + \xi_1)(t - t_0)}, \tag{9.48}$$

which is bounded by a small time interval $t \in [t_0, t_f]$ or small x_0.

From (9.45) we have

$$\begin{aligned}
x_i(t) - x_{i-1}(t) = {} & \big[\Phi_{i-1}(t, t_0) - \Phi_{i-2}(t, t_0)\big] x_0 \\
& + \int_{t_0}^{t} \Phi_{i-1}(t, s) G(x_{i-1}, u_{i-1}) \big[x_i(s) - x_{i-1}(s)\big] \mathrm{d}s \\
& + \int_{t_0}^{t} \Phi_{i-1}(t, s) K(x_{i-1}, w_{i-1}) \big[x_i(s) - x_{i-1}(s)\big] \mathrm{d}s \\
& + \int_{t_0}^{t} \Phi_{i-1}(t, s) \big[G(x_{i-1}, u_{i-1}) - G(x_{i-2}, u_{i-2})\big] x_{i-1}(s) \mathrm{d}s \\
& + \int_{t_0}^{t} \Phi_{i-1}(t, s) \big[K(x_{i-1}, w_{i-1}) - K(x_{i-2}, w_{i-2})\big] x_{i-1}(s) \mathrm{d}s \\
& + \int_{t_0}^{t} \big[\Phi_{i-1}(t, s) - \Phi_{i-2}(t, s)\big] G(x_{i-2}, u_{i-2}) x_{i-1}(s) \mathrm{d}s \\
& + \int_{t_0}^{t} \big[\Phi_{i-1}(t, s) - \Phi_{i-2}(t, s)\big] K(x_{i-2}, w_{i-2}) x_{i-1}(s) \mathrm{d}s.
\end{aligned} \tag{9.49}$$

Consider the supremum to both sides of (9.49) and let

$$\begin{aligned}
\beta_i(t) &= \sup_{s \in [t_0, t]} \|x_i(s) - x_{i-1}(s)\|, \\
\gamma_i(t) &= \sup_{s \in [t_0, t]} \|u_i(s) - u_{i-1}(s)\|, \\
\eta_i(t) &= \sup_{s \in [t_0, t]} \|w_i(s) - w_{i-1}(s)\|.
\end{aligned}$$

By using (9.39), (C6), and (C7), we get

$$\beta_i(t) \leq \alpha \|x_0\| \, e^{\mu_0 (t - t_0)} (t - t_0) \beta_{i-1}(t) + (\zeta_1 + \xi_1) \int_{t_0}^{t} e^{\mu_0 (t - s)} \beta_i(s) \mathrm{d}s$$

$$
\begin{aligned}
&+ \|x_0\| e^{\mu_0(t-t_0)} \int_{t_0}^{t} e^{(\zeta_1+\xi_1)(s-t_0)} \left[\zeta_2 \beta_{i-1}(s) + \zeta_3 \gamma_{i-1}(s)\right] \mathrm{d}s \\
&+ \|x_0\| e^{\mu_0(t-t_0)} \int_{t_0}^{t} e^{(\zeta_1+\xi_1)(s-t_0)} \left[\xi_2 \beta_{i-1}(s) + \xi_3 \eta_{i-1}(s)\right] \mathrm{d}s \\
&+ \alpha\zeta_1 \|x_0\| e^{\mu_0(t-t_0)} \int_{t_0}^{t} e^{(\zeta_1+\xi_1)(s-t_0)} (t-s) \beta_{i-1}(s) \mathrm{d}s \\
&+ \alpha\xi_1 \|x_0\| e^{\mu_0(t-t_0)} \int_{t_0}^{t} e^{(\zeta_1+\xi_1)(s-t_0)} (t-s) \beta_{i-1}(s) \mathrm{d}s. \qquad (9.50)
\end{aligned}
$$

Combining similar terms, we have

$$
\beta_i(t) \le \psi_1(t)\beta_{i-1}(t) + \psi_2(t)\gamma_{i-1}(t) + \psi_3(t)\eta_{i-1}(t), \qquad (9.51)
$$

where $\psi_1(t)$ through $\psi_3(t)$ are described in (9.44).

Similarly, from (9.38), we get

$$
\begin{aligned}
u_i(t) - u_{i-1}(t) &= M(x_{i-1}, u_{i-1}) \left[x_i(t) - x_{i-1}(t)\right] \\
&\quad + \left[M(x_{i-1}, u_{i-1}) - M(x_{i-2}, u_{i-2})\right] x_{i-1}(t) \\
w_i(t) - w_{i-1}(t) &= N(x_{i-1}, w_{i-1}) \left[x_i(t) - x_{i-1}(t)\right] \\
&\quad + \left[N(x_{i-1}, w_{i-1}) - N(x_{i-2}, w_{i-2})\right] x_{i-1}(t). \qquad (9.52)
\end{aligned}
$$

According to (C4), (C5), and (9.48), we have

$$
\begin{aligned}
\gamma_i(t) &\le \psi_4(t)\beta_{i-1}(t) + \psi_5(t)\gamma_{i-1}(t) + \psi_6(t)\eta_{i-1}(t) \\
\eta_i(t) &\le \psi_7(t)\beta_{i-1}(t) + \psi_8(t)\gamma_{i-1}(t) + \psi_9(t)\eta_{i-1}(t), \qquad (9.53)
\end{aligned}
$$

where $\psi_4(t)$ through $\psi_9(t)$ are shown in (9.44).

Then, combining (9.51) and (9.53), we have

$$
\Theta_i(t) \le \Psi(t)\Theta_{i-1}(t), \qquad (9.54)
$$

where $\Theta_i(t) = \begin{bmatrix} \beta_i(t) \\ \gamma_i(t) \\ \eta_i(t) \end{bmatrix}$ and $\Psi(t) = \begin{bmatrix} \psi_1 & \psi_2 & \psi_3 \\ \psi_4 & \psi_5 & \psi_6 \\ \psi_7 & \psi_8 & \psi_9 \end{bmatrix}$.

By induction, Θ_i satisfies

$$
\Theta_i(t) \le \Psi^{i-1}(t)\Theta^{[1]}(t), \qquad (9.55)
$$

which implies that we have $x_i(t)$, $u_i(t)$ and Cauchy sequences in Banach spaces $C([t_0, t_f]; \mathbb{R}^n)$, $C([t_0, t_f]; \mathbb{R}^n)$, $C([t_0, t_f]; \mathbb{R}^{m_1})$, and $C([t_0, t_f]; \mathbb{R}^{m_2})$, respectively. If $\{x_i(t)\}$ converges on $C([t_0, t_f]; \mathbb{R}^n)$, and the sequences of optimal policies $\{u_i\}$ and $\{w_i\}$ also converge on $C([t_0, t_f]; \mathbb{R}^{m_1})$ and $C([t_0, t_f]; \mathbb{R}^{m_2})$ on $[t_0, t_f]$.

It means that $x_{i-1}(t) = x_i(t)$, $u_{i-1}(t) = u_i(t)$, $w_{i-1}(t) = w_i(t)$ when $i \to \infty$. Hence, the system (9.29) has a unique solution on $[t_0, t_f]$, which is given by the limit of the solution of approximating sequence (9.32). □

Based on the iterative algorithm described in Theorem 9.27, the design procedure of optimal policies for nonlinear nonaffine zero-sum games is summarized as follows:

1. Give x_0, maximum iteration times $i_{\max}$ and approximation accuracy ε.
2. Use a factored form to represent the system as (9.31).
3. Set $i = 0$. Let $x_{i-1}(t) = x_0$, $u_{i-1}(t) = 0$ and $w_{i-1}(t) = 0$. Compute the corresponding matrix-valued functions $f(x_0)$, $g(x_0, 0)$, $k(x_0, 0)$, $F(x_0)$, $Q(x_0)$, $R(x_0)$, and $S(x_0)$.
4. Compute $x^{[0]}(t)$ and $P^{[0]}(t)$ according to differential equations (9.34) and (9.36) with $x(t_0) = x_0$, $P(t_f) = F(x_f)$.
5. Set $i = i + 1$. Compute the corresponding matrix-valued functions $f(x_{i-1}(t))$, $g(x_{i-1}(t), u_{i-1}(t))$, $k(x_{i-1}(t), w_{i-1}(t))$, $Q(x_{i-1}(t))$, $R(x_{i-1}(t))$, $F(x_{i-1}(t_f))$, and $S(x_{i-1}(t))$.
6. Compute $x_i(t)$ and $P_i(t)$ by (9.34) and (9.36) with $x(t_0) = x_0$, $P(t_f) = F(x_{tf})$.
7. If $\|x_i(t) - x_{i-1}(t)\| < \varepsilon$, go to Step 9); otherwise, go to Step 8.
8. If $i > i_{\max}$, then go to Step 9; else, go to Step 5.
9. Stop.

9.3.3 Simulations

Example 9.28 We now show the power of our iterative algorithm for finding optimal policies for nonaffine nonlinear zero-sum games.

In the following, we introduce an example of a control system that has the form (9.29) with control input $u(t)$, subject to a disturbance $w(t)$ and a cost functional $V(x_0, u, w)$. The control input $u(t)$ is required to minimize the cost functional $V(x_0, u, w)$. If the disturbance has a great effect on the system, the single disturbance $w(t)$ has to maximize the cost functional $V(x_0, u, w)$. The conflicting design can guarantee the optimality and strong robustness of the system at the same time. This is a zero-sum game problem, which can be described by the state equations

$$\begin{aligned}\dot{x}_1(t) &= -2x_1(t) + x_2^2(t) - x_1(t)u(t) + u^2(t) - 3x(t)w(t) + 5w^2(t),\\ \dot{x}_2(t) &= 5x_1^2(t) - 2x_2(t) + x_2^2(t) + u^2(t) + w^2(t).\end{aligned} \tag{9.56}$$

Define the finite horizon cost functional to be of the form (9.30), where $F = 0.01\, I_{2\times 2}$, $Q = 0.01\, I_{2\times 2}$, $R = 1$ and $S = 1$, where I is an identity matrix. Clearly, (9.56) is not affine in $u(t)$ and $w(t)$, it has the control nonaffine nonlinear structure. Therefore, we represent the system (9.56) in the factored form $f(x(t))x(t)$, $g(x(t), u(t))u(t)$ and $k(x(t), w(t))w(t)$, which, given the wide selection of possible representations, have been chosen as

$$f(x(t)) = \begin{bmatrix} 2 & x_2(t) \\ 5x_1(t) & -2 + x_2(t) \end{bmatrix}, \quad g(x(t), u(t)) = \begin{bmatrix} x_1(t) + u(t) \\ u(t) \end{bmatrix},$$

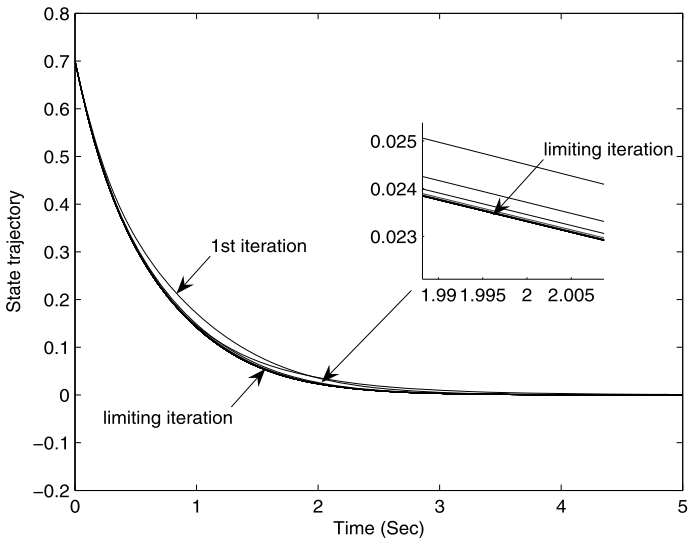

Fig. 9.10 The state trajectory $x_1(t)$ of each iteration

$$k(x(t), w(t)) = \begin{bmatrix} -3x_1(t) + 5w(t) \\ w(t) \end{bmatrix}. \tag{9.57}$$

The optimal policies designs given by Theorem 9.27 can now be applied to (9.31) with the dynamics (9.57).

The initial state vectors are chosen as $x_0 = [0.6, 0]^{\mathrm{T}}$ and the terminal time is set to $t_f = 5$. Let us define the required error norm between the solutions of the linear time-vary differential equations by $\|x_i(t) - x_{i-1}(t)\| < \varepsilon = 0.005$, which needs to be satisfied if convergence is to be achieved. The factorization is given by (9.57). Implementing the present iterative algorithm, it just needs six sequences to satisfy the required bound, $\|x^{[6]}(t) - x^{[5]}(t)\| = 0.0032$. With increasing of number of times of iterations, the approximation error will reduce obviously. When the iteration number $i = 25$, the approximation error is just 5.1205×10^{-10}.

Define the maximum iteration times $i_{\max} = 25$. Figure 9.10 represents the convergence trajectories of the state trajectory of each linear quadratic zero-sum game. It can be seen that the sequence is obviously convergent. The magnifications of the state trajectories are given in the figure, which shows that the error will be smaller as the number of times of iteration becomes bigger. The trajectories of control input $u(t)$ and disturbance input $w(t)$ of each iteration are also convergent, which is shown in Figs. 9.11 and 9.12. The approximate optimal policies $u^*(t)$ and $w^*(t)$ are obtained by the last iteration. Substituting the approximate optimal policies $u^*(t)$ and $w^*(t)$ into the system of zero-sum games (9.56), we get the state trajectory. The norm of the error between this state trajectory and the state trajectory of the last iteration is just 0.0019, which proves that the approximating iterative approach developed in this section is highly effective.

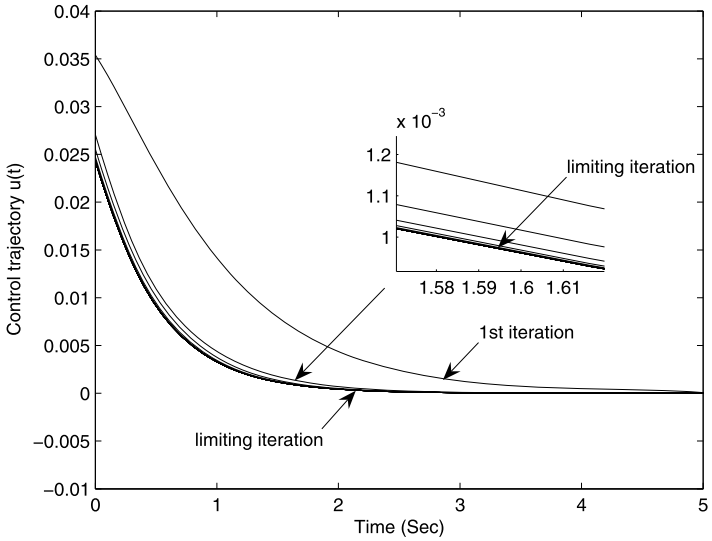

Fig. 9.11 The trajectory $u(t)$ of each iteration

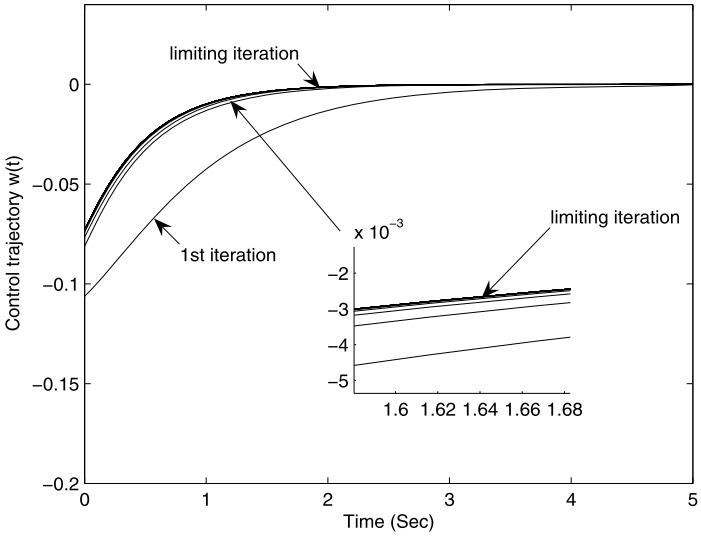

Fig. 9.12 The trajectory $w(t)$ of each iteration

9.4 Non-Zero-Sum Games for a Class of Nonlinear Systems Based on ADP

In this section, a near-optimal control scheme is developed for the non-zero-sum differential games of continuous-time nonlinear systems. The single network ADP

is utilized to obtain the optimal control policies which make the cost functions reach the Nash equilibrium of non-zero-sum differential games, where only one critic network is used for each player, instead of the action-critic dual network used in a typical ADP architecture. Furthermore, novel weight tuning laws for critic neural networks are developed, which not only ensure the Nash equilibrium to be reached, but also guarantee the stability of the system. No initial stabilizing control policy is required for each player. Moreover, Lyapunov theory is utilized to demonstrate the uniform ultimate boundedness of the closed-loop system.

9.4.1 Problem Formulation of Non-Zero-Sum Games

Consider the following continuous-time nonlinear systems:

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) + k(x(t))w(t), \tag{9.58}$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ and $d(t) \in \mathbb{R}^q$ are the control input vectors. Assume that $f(0) = 0$ and that $f(x)$, $g(x)$, $k(x)$ are locally Lipschitz.

The cost functional associated with u is defined as

$$J_1(x, u, w) = \int_t^\infty r_1(x(\tau), u(\tau), w(\tau))\mathrm{d}\tau, \tag{9.59}$$

where $r_1(x, u, w) = Q_1(x) + u^{\mathrm{T}} R_{11} u + w^{\mathrm{T}} R_{12} w$, $Q_1(x) \geq 0$ is the penalty on the states, $R_{11} \in \mathbb{R}^{m\times m}$ is a positive definite matrix, and $R_{12} \in \mathbb{R}^{q\times q}$ is a positive semidefinite matrix.

The cost functional associated with w is defined as

$$J_2(x, u, w) = \int_t^\infty r_2(x(\tau), u(\tau), w(\tau))\mathrm{d}\tau, \tag{9.60}$$

where $r_2(x, u, w) = Q_2(x) + u^{\mathrm{T}} R_{21} u + w^{\mathrm{T}} R_{22} w$, $Q_2(x) \geq 0$ is the penalty on the states, $R_{21} \in \mathbb{R}^{m\times m}$ is a positive semidefinite matrix, and $R_{22} \in \mathbb{R}^{q\times q}$ is a positive definite matrix.

For the above non-zero-sum differential games, the two feedback control policies u and w are chosen by player 1 and player 2, respectively, where player 1 tries to minimize the cost functional (9.59), while player 2 attempts to minimize the cost functional (9.60).

Definition 9.29 $u = \mu_1(x)$ and $w = \mu_2(x)$ are defined as admissible with respect to (9.59) and (9.60) on $\Omega \in \mathbb{R}^n$, denoted by $\mu_1 \in \psi(\Omega)$ and $\mu_2 \in \psi(\Omega)$, respectively, if $\mu_1(x)$ and $\mu_2(x)$ are continuous on Ω, $\mu_1(0) = 0$ and $\mu_2(0) = 0$, $\mu_1(x)$ and $\mu_2(x)$ stabilize (9.58) on Ω, and (9.59) and (9.60) are finite, $\forall x_0 \in \Omega$.

Definition 9.30 The policy set (u^*, w^*) is a Nash equilibrium policy set if the inequalities

$$\begin{aligned} J_1(u^*, w^*) &\leq J_1(u, w^*), \\ J_2(u^*, w^*) &\leq J_2(u^*, w) \end{aligned} \tag{9.61}$$

hold for any admissible control policies u and w.

Next, define the Hamilton functions for the cost functionals (9.59) and (9.60) with associated admissible control input u and w, respectively, as follows:

$$\begin{aligned} H_1(x, u, w) = {} & Q_1(x) + u^{\mathrm{T}} R_{11} u + w^{\mathrm{T}} R_{12} w \\ & + \nabla J_1^{\mathrm{T}} (f(x) + g(x)u + k(x)w), \end{aligned} \tag{9.62}$$

$$\begin{aligned} H_2(x, u, w) = {} & Q_2(x) + u^{\mathrm{T}} R_{21} u + w^{\mathrm{T}} R_{22} w \\ & + \nabla J_2^{\mathrm{T}} (f(x) + g(x)u + k(x)w), \end{aligned} \tag{9.63}$$

where ∇J_i is the partial derivative of the cost function $J_i(x, u, w)$ with respect to x, $i = 1, 2$.

According to the stationarity conditions of optimization, we have

$$\begin{aligned} \partial H_1(x, u, w)/\partial u &= 0, \\ \partial H_2(x, u, w)/\partial w &= 0. \end{aligned}$$

Therefore, the associated optimal feedback control policies u^* and w^* are found and revealed to be

$$u^* = -\frac{1}{2} R_{11}^{-1} g^{\mathrm{T}}(x) \nabla J_1, \tag{9.64}$$

$$w^* = -\frac{1}{2} R_{22}^{-1} k^{\mathrm{T}}(x) \nabla J_2. \tag{9.65}$$

The optimal feedback control policies u^* and w^* provide a Nash equilibrium for the non-zero-sum differential games among all the feedback control policies.

Considering $H_1(x, u^*, w^*) = 0$ and $H_2(x, u^*, w^*) = 0$, and substituting the optimal feedback control policy (9.64) and (9.65) into the Hamilton functions (9.62) and (9.63), we have

$$\begin{aligned} Q_1(x) - {} & \frac{1}{4} \nabla J_1^{\mathrm{T}} g(x) R_{11}^{-1} g^{\mathrm{T}}(x) \nabla J_1 + \nabla J_1^{\mathrm{T}} f(x) \\ & + \frac{1}{4} \nabla J_2^{\mathrm{T}} k(x) R_{22}^{-1} R_{12} R_{22}^{-1} k^{\mathrm{T}}(x) \nabla J_2 \\ & - \frac{1}{2} \nabla J_1^{\mathrm{T}} k(x) R_{22}^{-1} k^{\mathrm{T}}(x) \nabla J_2 = 0, \end{aligned} \tag{9.66}$$

$$Q_2(x)-\frac{1}{4}\nabla J_2^{\mathrm{T}}k(x)R_{22}^{-1}k^{\mathrm{T}}(x)\nabla J_2+\nabla J_2^{\mathrm{T}}f(x)$$
$$+\frac{1}{4}\nabla J_1^{\mathrm{T}}g(x)R_{11}^{-1}R_{21}R_{11}^{-1}g^{\mathrm{T}}(x)\nabla J_1$$
$$-\frac{1}{2}\nabla J_2^{\mathrm{T}}g(x)R_{11}^{-1}g^{\mathrm{T}}(x)\nabla J_1=0. \tag{9.67}$$

If the coupled HJ equations (9.66) and (9.67) can be solved for the optimal value functions $J_1(x,u^*,w^*)$ and $J_2(x,u^*,w^*)$, the optimal control can then be implemented by using (9.64) and (9.65). However, these equations are generally difficult or impossible to solve due to their inherently nonlinear nature. To overcome this difficulty, a near-optimal control scheme is developed to learn the solution of coupled HJ equations online using a single network ADP in order to obtain the optimal control policies.

Before presenting the near-optimal control scheme, the following lemma is required.

Lemma 9.31 *Given the system* (9.58) *with associated cost functionals* (9.59) *and* (9.60) *and the optimal feedback control policies* (9.64) *and* (9.65). *For player* i, $i=1,2$, *let* $L_i(x)$ *be a continuously differentiable, radially unbounded Lyapunov candidate such that* $\dot{L}_i=\nabla L_i^{\mathrm{T}}\dot{x}=\nabla L_i^{\mathrm{T}}(f(x)+g(x)u^*+k(x)w^*)<0$, *with* ∇L_i *being the partial derivative of* $L_i(x)$ *with respect to* x. *Moreover, let* $\bar{Q}_i(x)\in\mathbb{R}^{n\times n}$ *be a positive definite matrix satisfying* $\|\bar{Q}_i(x)\|=0$ *if and only if* $\|x\|=0$ *and* $\bar{Q}_{i\min}\le\|\bar{Q}_i(x)\|\le\bar{Q}_{i\max}$ *for* $\|\chi_{\min}\|\le\|x\|\le\chi_{\max}$ *with positive constants* $\bar{Q}_{i\min}$, $\bar{Q}_{i\max}$, $\chi_{\min}$, $\chi_{\max}$. *In addition, let* $\bar{Q}_i(x)$ *satisfy* $\lim_{x\to\infty}\bar{Q}_i(x)=\infty$ *as well as*

$$\nabla J_i^{*\mathrm{T}}\bar{Q}_i(x)\nabla L_i=r_i(x,u^*,w^*). \tag{9.68}$$

Then the following relation holds:

$$\nabla L_i^{\mathrm{T}}(f(x)+g(x)u^*+k(x)w^*)=-\nabla L_i^{\mathrm{T}}\bar{Q}_i(x)\nabla L_i. \tag{9.69}$$

Proof When the optimal control u^* and w^* in (9.64) and (9.65) are applied to the nonlinear system (9.58), the value function $J_i(x,u^*,w^*)$ becomes a Lyapunov function, $i=1,2$. Then, for $i=1,2$, differentiating the value function $J_i(x,u^*,w^*)$ with respect to t, we have

$$\dot{J}_i^*=\nabla J_i^{*\mathrm{T}}(f(x)+g(x)u^*+k(x)w^*)$$
$$=-r_i(x,u^*,w^*). \tag{9.70}$$

Using (9.68), (9.70) can be rewritten as

$$(f(x)+g(x)u^*+k(x)w^*)=-(\nabla J_i^*\nabla J_i^{*\mathrm{T}})^{-1}\nabla J_i^*r_i(x,u^*,w^*)$$
$$=-(\nabla J_i^*\nabla J_i^{*\mathrm{T}})^{-1}\nabla J_i^*\nabla J_i^{*\mathrm{T}}\bar{Q}_i(x)\nabla L_i$$
$$=-\bar{Q}_i(x)\nabla L_i. \tag{9.71}$$

Next, multiplying both sides of (9.71) by ∇L_i^{T}, (9.69) can be obtained. This completes the proof. □

9.4.2 Optimal Control of Nonlinear Non-Zero-Sum Games Based on ADP

To begin the development, we rewrite the cost functions (9.59) and (9.60) by NNs as

$$J_1(x) = W_{c1}^{\mathrm{T}}\phi_1(x) + \varepsilon_1, \tag{9.72}$$

$$J_2(x) = W_{c2}^{\mathrm{T}}\phi_2(x) + \varepsilon_2, \tag{9.73}$$

where W_i, $\phi_i(x)$, and ε_i are the critic NN ideal constant weights, the critic NN activation function vector and the NN approximation error for player i, $i = 1, 2$, respectively.

The derivative of the cost functions with respect to x can be derived as

$$\nabla J_1 = \nabla\phi_1^{\mathrm{T}} W_{c1} + \nabla\varepsilon_1, \tag{9.74}$$

$$\nabla J_2 = \nabla\phi_2^{\mathrm{T}} W_{c2} + \nabla\varepsilon_2, \tag{9.75}$$

where $\nabla\phi_i \triangleq \partial\phi_i(x)/\partial x$, $\nabla\varepsilon_i \triangleq \partial\varepsilon_i/\partial x$, $i = 1, 2$.

Using (9.74) and (9.75), the optimal feedback control policies (9.64) and (9.65) can be rewritten as

$$u^* = -\frac{1}{2}R_{11}^{-1}g^{\mathrm{T}}(x)\nabla\phi_1^{\mathrm{T}}W_{c1} - \frac{1}{2}R_{11}^{-1}g^{\mathrm{T}}(x)\nabla\varepsilon_1, \tag{9.76}$$

$$w^* = -\frac{1}{2}R_{22}^{-1}k^{\mathrm{T}}(x)\nabla\phi_2^{\mathrm{T}}W_{c2} - \frac{1}{2}R_{22}^{-1}k^{\mathrm{T}}(x)\nabla\varepsilon_2, \tag{9.77}$$

and the coupled HJ equations (9.66) and (9.67) can be rewritten as

$$\begin{aligned} Q_1(x) - &\frac{1}{4}W_{c1}^{\mathrm{T}}\nabla\phi_1 D_1 \nabla\phi_1^{\mathrm{T}} W_{c1} + W_{c1}^{\mathrm{T}}\nabla\phi_1 f(x) \\ &+ \frac{1}{4}W_{c2}^{\mathrm{T}}\nabla\phi_2 S_2 \nabla{\phi_2}^{\mathrm{T}} W_{c2} - \frac{1}{2}W_{c1}^{\mathrm{T}}\nabla\phi_1 D_2 \nabla\phi_2^{\mathrm{T}} W_{c2} - \varepsilon_{\mathrm{HJ1}} = 0, \end{aligned} \tag{9.78}$$

$$\begin{aligned} Q_2(x) - &\frac{1}{4}W_{c2}^{\mathrm{T}}\nabla\phi_2 D_2 \nabla\phi_2^{\mathrm{T}} W_{c2} + W_{c2}^{\mathrm{T}}\nabla\phi_2 f(x) \\ &+ \frac{1}{4}W_{c1}^{\mathrm{T}}\nabla\phi_1 S_1 \nabla{\phi_1}^{\mathrm{T}} W_{c1} - \frac{1}{2}W_{c2}^{\mathrm{T}}\nabla\phi_2 D_1 \nabla\phi_1^{\mathrm{T}} W_{c1} - \varepsilon_{\mathrm{HJ2}} = 0, \end{aligned} \tag{9.79}$$

where

$$\begin{aligned} D_1 &= g(x)R_{11}^{-1}g^{\mathrm{T}}(x), \\ D_2 &= k(x)R_{22}^{-1}k^{\mathrm{T}}(x), \\ S_1 &= g(x)R_{11}^{-1}R_{21}R_{11}^{-1}g^{\mathrm{T}}(x), \\ S_2 &= k(x)R_{22}^{-1}R_{12}R_{22}^{-1}k^{\mathrm{T}}(x). \end{aligned} \tag{9.80}$$

The residual error due to the NN approximation for player 1 is

$$\begin{aligned} \varepsilon_{\mathrm{HJ1}} = &-\nabla\varepsilon_1^{\mathrm{T}}\Big(f(x) - \frac{1}{2}D_1(\nabla\phi_1^{\mathrm{T}}W_{c1} + \nabla\varepsilon_1) \\ &- \frac{1}{2}D_2(\nabla\phi_2^{\mathrm{T}}W_{c2} + \nabla\varepsilon_2)\Big) - \frac{1}{4}\nabla\varepsilon_1^{\mathrm{T}}D_1\nabla\varepsilon_1 + \frac{1}{2}W_{c1}^{\mathrm{T}}\nabla\phi_1 D_2\nabla\varepsilon_2 \\ &- \frac{1}{2}\nabla\varepsilon_2^{\mathrm{T}}S_2\nabla\phi_2^{\mathrm{T}}W_{c2} - \frac{1}{4}\nabla\varepsilon_2^{\mathrm{T}}S_2\nabla\varepsilon_2. \end{aligned} \tag{9.81}$$

The residual error due to the NN approximation for player 2 is

$$\begin{aligned} \varepsilon_{\mathrm{HJ2}} = &-\nabla\varepsilon_2^{\mathrm{T}}\Big(f(x) - \frac{1}{2}D_1(\nabla\phi_1^{\mathrm{T}}W_{c1} + \nabla\varepsilon_1) \\ &- \frac{1}{2}D_2(\nabla\phi_2^{\mathrm{T}}W_{c2} + \nabla\varepsilon_2)\Big) - \frac{1}{4}\nabla\varepsilon_2^{\mathrm{T}}D_2\nabla\varepsilon_2 + \frac{1}{2}W_{c2}^{\mathrm{T}}\nabla\phi_2 D_1\nabla\varepsilon_1 \\ &- \frac{1}{2}\nabla\varepsilon_2^{\mathrm{T}}S_1\nabla\phi_1^{\mathrm{T}}W_{c2} - \frac{1}{4}\nabla\varepsilon_1^{\mathrm{T}}S_1\nabla\varepsilon_1. \end{aligned} \tag{9.82}$$

Let $\hat{W}_{c1}$ and $\hat{W}_{c2}$ be the estimates of W_{c1} and W_{c2}, respectively. Then we have the estimates of $V_1(x)$ and $V_2(x)$ as follows:

$$\hat{J}_1(x) = \hat{W}_{c1}^{\mathrm{T}}\phi_1(x), \tag{9.83}$$

$$\hat{J}_2(x) = \hat{W}_{c2}^{\mathrm{T}}\phi_2(x). \tag{9.84}$$

Substituting (9.83) and (9.84) into (9.64) and (9.65), respectively, the estimates of optimal control policies can be written as

$$\hat{u} = -\frac{1}{2}R_{11}^{-1}g^{\mathrm{T}}(x)\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}, \tag{9.85}$$

$$\hat{w} = -\frac{1}{2}R_{22}^{-1}k^{\mathrm{T}}(x)\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}. \tag{9.86}$$

Applying (9.85) and (9.86) to the system (9.58), we have the closed-loop system dynamics as follows:

$$\dot{x} = f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}. \tag{9.87}$$

Substituting (9.83) and (9.84) into (9.62) and (9.63), respectively, the approximate Hamilton functions can be derived as follows:

$$\begin{aligned} H_1(x,\hat{W}_{c1},\hat{W}_{c2}) &= Q_1(x) - \frac{1}{4}\hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1} + \hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 f(x) \\ &\quad + \frac{1}{4}\hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 S_2 \nabla\phi_2^{\mathrm{T}}\hat{W}_{c2} - \frac{1}{2}\hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_2 \nabla\phi_2^{\mathrm{T}}\hat{W}_{c2} \\ &= e_1, \end{aligned} \tag{9.88}$$

$$\begin{aligned} H_2(x,\hat{W}_{c1},\hat{W}_{c2}) &= Q_2(x) - \frac{1}{4}\hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2 \nabla\phi_2^{\mathrm{T}}\hat{W}_{c2} + \hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 f(x) \\ &\quad + \frac{1}{4}\hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 S_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1} - \frac{1}{2}\hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1} \\ &= e_2. \end{aligned} \tag{9.89}$$

It is desired to select $\hat{W}_{c1}$ and $\hat{W}_{c2}$ to minimize the squared residual error $E = e_1^{\mathrm{T}}e_1/2 + e_2^{\mathrm{T}}e_2/2$. Then we have $\hat{W}_{c1} \to W_{c1}$, $\hat{W}_{c2} \to W_{c2}$, and $e_1 \to \varepsilon_{\mathrm{HJ1}}$, $e_2 \to \varepsilon_{\mathrm{HJ2}}$. In other words, the Nash equilibrium of the non-zero-sum differential games of continuous-time nonlinear system (9.58) can be obtained. However, tuning the critic NN weights to minimize the squared residual error E alone does not ensure the stability of the nonlinear system (9.58) during the learning process of critic NNs. Therefore, we propose the novel weight tuning laws of critic NNs for two players, which cannot only minimize the squared residual error E but also guarantee the stability of the system as follows:

$$\begin{aligned} \dot{\hat{W}}_1 = &-\alpha_1 \frac{\bar{\sigma}_1}{m_{s_1}}\Bigg(Q_1(x) - \frac{1}{4}\hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1} \\ &+ \hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 f(x) + \frac{1}{4}\hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 S_2 \nabla\phi_2^{\mathrm{T}}\hat{W}_{c2} - \frac{1}{2}\hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_2 \nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}\Bigg) \\ &+ \frac{\alpha_1}{4}\nabla\phi_1 D_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}\hat{W}_{c1} + \frac{\alpha_2}{4}\nabla\phi_1 S_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}\hat{W}_{c2} \\ &+ \Sigma(x,\hat{u},\hat{w})\left(\frac{\alpha_1\nabla\phi_1 D_1\nabla L_1}{2} + \frac{\alpha_1\nabla\phi_1 D_1\nabla L_2}{2}\right) \\ &- \alpha_1(F_1\hat{W}_{c1} - F_2\bar{\sigma}_1^{\mathrm{T}}\hat{W}_{c1}), \end{aligned} \tag{9.90}$$

$$\begin{aligned} \dot{\hat{W}}_2 = &-\alpha_2 \frac{\bar{\sigma}_2}{m_{s_2}}\Bigg(Q_2(x) - \frac{1}{4}\hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2 \nabla\phi_2^{\mathrm{T}}\hat{W}_{c2} \\ &+ \hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 f(x) + \frac{1}{4}\hat{W}_{c1}^{\mathrm{T}}\nabla\phi_1 S_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1} - \frac{1}{2}\hat{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_1 \nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}\Bigg) \end{aligned}$$

$$+\frac{\alpha_2}{4}\nabla\phi_2 D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}\hat{W}_{c2}+\frac{\alpha_2}{4}\nabla\phi_2 S_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}\hat{W}_{c1}$$
$$+\Sigma(x,\hat{u},\hat{w})\left(\frac{\alpha_2\nabla\phi_2 D_2\nabla L_2}{2}+\frac{\alpha_2\nabla\phi_2 D_2\nabla L_1}{2}\right)$$
$$-\alpha_2(F_3\hat{W}_{c2}-F_4\bar{\sigma}_2^{\mathrm{T}}\hat{W}_{c2}), \tag{9.91}$$

where $\bar{\sigma}_i=\hat{\sigma}_i/(\hat{\sigma}_i^{\mathrm{T}}\hat{\sigma}_i+1)$, $\hat{\sigma}_i=\nabla\phi_i(f(x)-D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}/2-D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}/2)$, $m_{s_i}=\hat{\sigma}_i^{\mathrm{T}}\hat{\sigma}_i+1$, $\alpha_i>0$ is the adaptive gain, ∇L_i is described in Lemma 9.31, $i=1,2$. F_1, F_2, F_3, and F_4 are design parameters. The operator $\Sigma(x,\hat{u},\hat{w})$ is given by

$$\Sigma(x,\hat{u},\hat{w})=\begin{cases}0 \text{ if } \quad \nabla L_1\dot{x}\le 0 \text{ and } \nabla L_2\dot{x}\le 0,\\ 1 \text{ else},\end{cases} \tag{9.92}$$

where $\dot{x}$ is given as (9.87).

Remark 9.32 The first terms in (9.90) and (9.91) are utilized to minimize the squared residual error E and derived by using a normalized gradient descent algorithm. The other terms are utilized to guarantee the stability of the closed-loop system while the critic NNs learn the optimal cost functions and are derived by following Lyapunov stability analysis. The operator $\Sigma(x,\hat{u},\hat{w})$ is selected based on the Lyapunov's sufficient condition for stability, which means that the state x is stable if $L_i(x)>0$ and $\nabla L_i\dot{x}<0$ for player i, $i=1,2$. When the system (9.58) is stable, the operator $\Sigma(x,\hat{u},\hat{w})=0$ and it will not take effect. When the system (9.58) is unstable, the operator $\Sigma(x,\hat{u},\hat{w})=1$ and it will be activated. Therefore, no initial stabilizing control policies are needed due to the introduction of the operator $\Sigma(x,\hat{u},\hat{w})$.

Remark 9.33 From (9.88) and (9.89), it can be seen that the approximate Hamilton functions $H_1(x,\hat{W}_{c1},\hat{W}_{c2})=e_1=0$ and $H_2(x,\hat{W}_{c1},\hat{W}_{c2})=e_2=0$ when $x=0$. For this case, the tuning laws of critic NN weights for two players (9.90) and (9.91) cannot achieve the purpose of optimization anymore. This can be considered as a persistency of the requirement of excitation for the system states. Therefore, the system states must be persistently excited enough for minimizing the squared residual errors E to drive the critic NN weights toward their ideal values. In order to satisfy the persistent excitation condition, probing noise is added to the control input.

Define the weight estimation errors of critic NNs for two players to be $\tilde{W}_{c1}=W_{c1}-\hat{W}_{c1}$ and $\tilde{W}_{c2}=W_{c2}-\hat{W}_{c2}$, respectively. From (9.78) and (9.79), we observe that

$$Q_1(x)=\frac{1}{4}W_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla\phi_1^{\mathrm{T}}W_{c1}-W_{c1}^{\mathrm{T}}\nabla\phi_1 f(x)-\frac{1}{4}W_{c2}^{\mathrm{T}}\nabla\phi_2 S_2\nabla\phi_2{}^{\mathrm{T}}W_{c2}$$
$$+\frac{1}{2}W_{c1}^{\mathrm{T}}\nabla\phi_1 D_2\nabla\phi_2^{\mathrm{T}}W_{c2}+\varepsilon_{\mathrm{HJ1}}, \tag{9.93}$$

$$Q_2(x) = \frac{1}{4} W_{c2}^{\mathrm{T}} \nabla\phi_2 D_2 \nabla\phi_2^{\mathrm{T}} W_{c2} - W_{c2}^{\mathrm{T}} \nabla\phi_2 f(x) - \frac{1}{4} W_{c1}^{\mathrm{T}} \nabla\phi_1 S_1 \nabla\phi_1{}^{\mathrm{T}} W_{c1} + \frac{1}{2} W_{c2}^{\mathrm{T}} \nabla\phi_2 D_1 \nabla\phi_1^{\mathrm{T}} W_{c1} + \varepsilon_{\mathrm{HJ2}}. \tag{9.94}$$

Combining (9.90) with (9.93), we have

$$\begin{aligned}\dot{\tilde{W}}_1 =& \alpha_1 \frac{\bar{\sigma}_1}{m_{s_1}} \left[-\tilde{W}_{c1}^{\mathrm{T}} \hat{\sigma}_1 + \frac{1}{4} \tilde{W}_{c1}^{\mathrm{T}} \nabla\phi_1 D_1 \nabla\phi_1^{\mathrm{T}} \tilde{W}_{c1} + \frac{1}{2} W_{c1}^{\mathrm{T}} \nabla\phi_1 D_2 \nabla\phi_2^{\mathrm{T}} \tilde{W}_{c2} \right. \\ & \left. - \frac{1}{2} \tilde{W}_{c2}^{\mathrm{T}} \nabla\phi_2 S_2 \nabla\phi_2^{\mathrm{T}} W_{c2} + \frac{1}{4} \tilde{W}_{c2}^{\mathrm{T}} \nabla\phi_2 S_2 \nabla\phi_2^{\mathrm{T}} \tilde{W}_{c2} + \varepsilon_{\mathrm{HJ1}} \right] \\ & - \frac{\alpha_1}{4} \nabla\phi_1 D_1 \nabla\phi_1^{\mathrm{T}} \hat{W}_{c1} \frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}} \hat{W}_{c1} - \frac{\alpha_2}{4} \nabla\phi_1 S_1 \nabla\phi_1^{\mathrm{T}} \hat{W}_{c1} \frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}} \hat{W}_{c2} \\ & - \Sigma(x, \hat{u}, \hat{w}) \left(\frac{\alpha_1 \nabla\phi_1 D_1 \nabla L_1}{2} + \frac{\alpha_1 \nabla\phi_1 D_1 \nabla L_2}{2} \right) \\ & + \alpha_1 (F_1 \hat{W}_{c1} - F_2 \bar{\sigma}_1^{\mathrm{T}} \hat{W}_{c1}). \end{aligned} \tag{9.95}$$

Similarly, combining (9.91) with (9.94), we have

$$\begin{aligned}\dot{\tilde{W}}_2 =& \alpha_2 \frac{\bar{\sigma}_2}{m_{s_2}} \left[-\tilde{W}_{c2}^{\mathrm{T}} \hat{\sigma}_2 + \frac{1}{4} \tilde{W}_{c2}^{\mathrm{T}} \nabla\phi_2 D_2 \nabla\phi_2^{\mathrm{T}} \tilde{W}_{c2} + \frac{1}{2} W_{c2}^{\mathrm{T}} \nabla\phi_2 D_1 \nabla\phi_1^{\mathrm{T}} \tilde{W}_{c1} \right. \\ & \left. - \frac{1}{2} \tilde{W}_{c1}^{\mathrm{T}} \nabla\phi_1 S_1 \nabla\phi_1^{\mathrm{T}} W_{c1} + \frac{1}{4} \tilde{W}_{c1}^{\mathrm{T}} \nabla\phi_1 S_1 \nabla\phi_1^{\mathrm{T}} \tilde{W}_{c1} + \varepsilon_{\mathrm{HJ2}} \right] \\ & - \frac{\alpha_2}{4} \nabla\phi_2 D_2 \nabla\phi_2^{\mathrm{T}} \hat{W}_{c2} \frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}} \hat{W}_{c2} - \frac{\alpha_2}{4} \nabla\phi_2 S_2 \nabla\phi_2^{\mathrm{T}} \hat{W}_{c2} \frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}} \hat{W}_{c1} \\ & - \Sigma(x, \hat{u}, \hat{w}) \left(\frac{\alpha_2 \nabla\phi_2 D_2 \nabla L_2}{2} + \frac{\alpha_2 \nabla\phi_2 D_2 \nabla L_1}{2} \right) \\ & + \alpha_2 (F_3 \hat{W}_{c2} - F_4 \bar{\sigma}_2^{\mathrm{T}} \hat{W}_{c2}). \end{aligned} \tag{9.96}$$

In the following, the stability analysis will be performed. First, the following assumption is made, which can reasonably be satisfied under the current problem settings.

Assumption 9.34

(a) $g(\cdot)$ and $k(\cdot)$ are upper bounded, i.e., $\|g(\cdot)\| \leq g_M$ and $\|k(\cdot)\| \leq k_M$ with g_M and k_M being positive constants.
(b) The critic NN approximation errors and their gradients are upper bounded so that $\|\varepsilon_i\| \leq \varepsilon_{iM}$ and $\|\nabla\varepsilon_i\| \leq \varepsilon_{idM}$ with ε_{iM} and ε_{idM} being positive constants, $i = 1, 2$.
(c) The critic NN activation function vectors are upper bounded, so that $\|\phi_i\| \leq \phi_{iM}$ and $\|\nabla\phi_i\| \leq \phi_{idM}$, with ϕ_{iM} and ϕ_{idM} being positive constants, $i = 1, 2$.

(d) The critic NN weights are upper bounded so that $\|W_i\| \leq W_{iM}$ with W_{iM} being positive constant, $i = 1, 2$. The residual errors $\varepsilon_{\mathrm{HJi}}$ are upper bounded, so that $\|\varepsilon_{\mathrm{HJi}}\| \leq \varepsilon_{\mathrm{HJ}iM}$ with $\varepsilon_{\mathrm{HJ}iM}$ being positive constant, $i = 1, 2$.

Now we are ready to prove the following theorem.

Theorem 9.35 (cf. [17]) *Consider the system given by* (9.58). *Let the control input be provided by* (9.85) *and* (9.86), *and the critic NN weight tuning laws be given by* (9.90) *and* (9.91). *Then, the system state x and the weight estimation errors of critic NNs $\tilde{W}_{c1}$ and $\tilde{W}_{c2}$ are uniformly ultimately bounded* (*UUB*). *Furthermore, the obtained control input $\hat{u}$ and $\hat{w}$ in* (9.85) *and* (9.86) *are proved to converge to the Nash equilibrium policy of the non-zero-sum differential games approximately, i.e., $\hat{u}$ and $\hat{w}$ are closed for the optimal control input u^* and w^* with bounds ϵ_u and ϵ_w, respectively.*

Proof Choose the following Lyapunov function candidate:

$$L = L_1(x) + L_2(x) + \frac{1}{2}\tilde{W}_{c1}^{\mathrm{T}}\alpha_1^{-1}\tilde{W}_{c1} + \frac{1}{2}\tilde{W}_{c2}^{\mathrm{T}}\alpha_2^{-1}\tilde{W}_{c2}, \tag{9.97}$$

where $L_1(x)$ and $L_2(x)$ are given by Lemma 9.31.

The derivative of the Lyapunov function candidate (9.97) along the system (9.87) is computed as

$$\begin{aligned}\dot{L} =& \nabla L_1^{\mathrm{T}}\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}\right)\\ &+ \nabla L_2^{\mathrm{T}}\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}\right)\\ &+ \tilde{W}_{c1}^{\mathrm{T}}\alpha_1^{-1}\dot{\tilde{W}}_1 + \tilde{W}_{c2}^{\mathrm{T}}\alpha_2^{-1}\dot{\tilde{W}}_2.\end{aligned} \tag{9.98}$$

Then, substituting (9.95) and (9.96) into (9.98), we have

$$\begin{aligned}\dot{L} =& \nabla L_1^{\mathrm{T}}\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}\right)\\ &+ \nabla L_2^{\mathrm{T}}\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}\right)\\ &+ \tilde{W}_{c1}^{\mathrm{T}}\bar{\sigma}_1\left(-\bar{\sigma}_1^{\mathrm{T}}\tilde{W}_{c1} + \frac{\varepsilon_{\mathrm{HJ1}}}{m_{s_1}}\right) + \tilde{W}_{c2}^{\mathrm{T}}\bar{\sigma}_2\left(-\bar{\sigma}_2^{\mathrm{T}}\tilde{W}_{c2} + \frac{\varepsilon_{\mathrm{HJ2}}}{m_{s_2}}\right)\\ &+ \frac{1}{4}\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla\phi_1^{\mathrm{T}}W_{c1}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}\tilde{W}_{c1} - \frac{1}{4}\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla\phi_1^{\mathrm{T}}W_{c1}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}W_{c1}\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{4}\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla\phi_1^{\mathrm{T}}\tilde{W}_{c1}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}W_{c1}\\
&+\frac{1}{4}\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla\phi_2^{\mathrm{T}}W_{c2}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}\tilde{W}_{c2}-\frac{1}{4}\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla\phi_2^{\mathrm{T}}W_{c2}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}W_{c2}\\
&+\frac{1}{4}\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla\phi_2^{\mathrm{T}}\tilde{W}_{c2}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}W_{c2}\\
&+\frac{1}{2}\tilde{W}_{c1}^{\mathrm{T}}\frac{\bar{\sigma}_1}{m_{s_1}}W_{c1}^{\mathrm{T}}\nabla\phi_1 D_2\nabla\phi_2^{\mathrm{T}}\tilde{W}_{c2}-\frac{1}{2}\tilde{W}_{c1}^{\mathrm{T}}\frac{\bar{\sigma}_1}{m_{s_1}}\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 S_2\nabla\phi_2^{\mathrm{T}}W_{c2}\\
&+\frac{1}{2}\tilde{W}_{c2}^{\mathrm{T}}\frac{\bar{\sigma}_2}{m_{s_2}}W_{c2}^{\mathrm{T}}\nabla\phi_2 D_1\nabla\phi_1^{\mathrm{T}}\tilde{W}_{c1}-\frac{1}{2}\tilde{W}_{c2}^{\mathrm{T}}\frac{\bar{\sigma}_2}{m_{s_2}}\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 S_1\nabla\phi_1^{\mathrm{T}}W_{c1}\\
&+\frac{1}{4}\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 S_1\nabla\phi_1^{\mathrm{T}}W_{c1}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}\tilde{W}_{c2}-\frac{1}{4}\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 S_1\nabla\phi_1^{\mathrm{T}}W_{c1}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}W_{c2}\\
&+\frac{1}{4}\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 S_1\nabla\phi_1^{\mathrm{T}}\tilde{W}_{c1}\frac{\bar{\sigma}_2^{\mathrm{T}}}{m_{s_2}}W_{c2}\\
&+\frac{1}{4}\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 S_2\nabla\phi_2^{\mathrm{T}}W_{c2}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}\tilde{W}_{c1}-\frac{1}{4}\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 S_2\nabla\phi_2^{\mathrm{T}}W_{c2}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}W_{c1}\\
&+\frac{1}{4}\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 S_2\nabla\phi_2^{\mathrm{T}}\tilde{W}_{c2}\frac{\bar{\sigma}_1^{\mathrm{T}}}{m_{s_1}}W_{c1}\\
&-\Sigma(x,\hat{u},\hat{w})\left(\frac{\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla L_1}{2}+\frac{\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla L_2}{2}\right)\\
&-\Sigma(x,\hat{u},\hat{w})\left(\frac{\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla L_2}{2}+\frac{\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla L_1}{2}\right)\\
&+\tilde{W}_{c1}^{\mathrm{T}}F_1\hat{W}_{c1}-\tilde{W}_{c1}^{\mathrm{T}}F_2\bar{\sigma}_1^{\mathrm{T}}\hat{W}_{c1}\\
&+\tilde{W}_{c2}^{\mathrm{T}}F_3\hat{W}_{c2}-\tilde{W}_{c2}^{\mathrm{T}}F_4\bar{\sigma}_2^{\mathrm{T}}\hat{W}_{c2}.
\end{aligned}
\tag{9.99}
$$

In (9.99), the last two terms can be rewritten as

$$
\begin{aligned}
&\tilde{W}_{c1}^{\mathrm{T}}F_1\hat{W}_{c1}-\tilde{W}_{c1}^{\mathrm{T}}F_2\bar{\sigma}_1^{\mathrm{T}}\hat{W}_{c1}\\
&\quad=\tilde{W}_{c1}^{\mathrm{T}}F_1W_{c1}-\tilde{W}_{c1}^{\mathrm{T}}F_2\bar{\sigma}_1^{\mathrm{T}}\tilde{W}_{c1}-\tilde{W}_{c1}^{\mathrm{T}}F_2\bar{\sigma}_1^{\mathrm{T}}W_{c1}-\tilde{W}_{c1}^{\mathrm{T}}F_2\bar{\sigma}_1^{\mathrm{T}}\tilde{W}_{c1}\\
&\qquad+\tilde{W}_{c2}^{\mathrm{T}}F_3\hat{W}_{c2}-\tilde{W}_{c2}^{\mathrm{T}}F_4\bar{\sigma}_2^{\mathrm{T}}\hat{W}_{c2}\\
&\quad=\tilde{W}_{c2}^{\mathrm{T}}F_3W_{c2}-\tilde{W}_{c2}^{\mathrm{T}}F_3\tilde{W}_{c2}-\tilde{W}_{c2}^{\mathrm{T}}F_4\bar{\sigma}_2^{\mathrm{T}}W_{c2}-\tilde{W}_{c2}^{\mathrm{T}}F_4\bar{\sigma}_2^{\mathrm{T}}\tilde{W}_{c2}.
\end{aligned}
\tag{9.100}
$$

Define $z = [\bar{\sigma}_1^{\mathrm{T}}\tilde{W}_{c1}, \bar{\sigma}_2^{\mathrm{T}}\tilde{W}_{c2}, \tilde{W}_{c1}, \tilde{W}_{c2}]^{\mathrm{T}}$; then (9.99) can be rewritten as

$$\begin{aligned}\dot{L} = &- z^{\mathrm{T}}\begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} z + z^{\mathrm{T}}\delta \\ &+ \nabla L_1^{\mathrm{T}}\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}\right) \\ &+ \nabla L_2^{\mathrm{T}}\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2})}{2}\right) \\ &- \Sigma(x,\hat{u},\hat{w})\left(\frac{\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla L_1}{2} + \frac{\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla L_2}{2}\right) \\ &- \Sigma(x,\hat{u},\hat{w})\left(\frac{\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla L_2}{2} + \frac{\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla L_1}{2}\right), \end{aligned} \tag{9.101}$$

where the components of the matrix M are given by

$$\begin{aligned} M_{11} &= M_{22} = I, \\ M_{12} &= M_{21}^{\mathrm{T}} = 0, \\ M_{13} &= M_{31}^{\mathrm{T}} = -\frac{1}{4m_{s_1}}\nabla\phi_1 D_1\nabla\phi_1^{\mathrm{T}}W_{c1} - \frac{F_2}{2}, \\ M_{14} &= M_{41}^{\mathrm{T}} = -\frac{1}{4m_{s_1}}\nabla\phi_2 D_2\nabla\phi_1^{\mathrm{T}}W_{c1} + \frac{1}{8}\nabla\phi_2 S_2\nabla\phi_2^{\mathrm{T}}W_{c2}, \\ M_{23} &= M_{32}^{\mathrm{T}} = -\frac{1}{4m_{s_2}}\nabla\phi_1 D_1\nabla\phi_2^{\mathrm{T}}W_{c2} + \frac{1}{8}\nabla\phi_1 S_1\nabla\phi_1^{\mathrm{T}}W_{c1}, \\ M_{24} &= M_{42}^{\mathrm{T}} = -\frac{1}{4m_{s_2}}\nabla\phi_2 D_2\nabla\phi_2^{\mathrm{T}}W_{c2} - \frac{F_4}{2}, \\ M_{33} &= -\frac{1}{4m_{s_2}}\nabla\phi_1 S_1\nabla\phi_1^{\mathrm{T}}W_{c2}\bar{\sigma}_2^{\mathrm{T}} + F_1, \\ M_{34} &= M_{43}^{\mathrm{T}} = 0, \\ M_{44} &= -\frac{1}{4m_{s_1}}\nabla\phi_2 S_2\nabla\phi_2^{\mathrm{T}}W_{c1}\bar{\sigma}_1^{\mathrm{T}} + F_3, \end{aligned}$$

and the components of the vector $\delta = [d_1\ d_2\ d_3\ d_4]^{\mathrm{T}}$ are given as

$$d_1 = \frac{\varepsilon_{\mathrm{HJ1}}}{m_{s_1}},$$

$$
\begin{aligned}
d_2 &= \frac{\varepsilon_{\mathrm{HJ2}}}{m_{s_2}},\\
d_3 &= -\frac{1}{4m_{s_1}}\nabla\phi_1 D_1 \nabla\phi_1^{\mathrm{T}} W_{c1}\bar{\sigma}_1^{\mathrm{T}} W_{c1}\\
&\quad -\frac{1}{4m_{s_2}}\nabla\phi_1 S_1 \nabla\phi_1^{\mathrm{T}} W_{c1}\bar{\sigma}_2^{\mathrm{T}} W_{c2} + F_1 W_{c1} - F_2\bar{\sigma}_1^{\mathrm{T}} W_{c1},\\
d_4 &= -\frac{1}{4m_{s_2}}\nabla\phi_2 D_2 \nabla\phi_2^{\mathrm{T}} W_{c2}\bar{\sigma}_2^{\mathrm{T}} W_{c2}\\
&\quad -\frac{1}{4m_{s_1}}\nabla\phi_2 S_2 \nabla\phi_2^{\mathrm{T}} W_{c2}\bar{\sigma}_1^{\mathrm{T}} W_{c1} + F_3 W_{c2} - F_4\bar{\sigma}_2^{\mathrm{T}} W_{c2}.
\end{aligned}
$$

According to Assumption 9.34 and observing the facts that $\bar{\sigma}_1 < 1$ and $\bar{\sigma}_2 < 1$, it can be concluded that δ is bounded by δ_M. Let the parameters F_1, F_2, F_3, and F_4 be chosen such that $M > 0$. Then, taking the upper bounds of (9.101) reveals

$$
\begin{aligned}
\dot{L} \le{}& \nabla L_1\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}\right)\\
&+ \nabla L_2\left(f(x) - \frac{D_1\nabla\phi_1^{\mathrm{T}}\hat{W}_{c1}}{2} - \frac{D_2\nabla\phi_2^{\mathrm{T}}\hat{W}_{c2}}{2}\right)\\
&- \|z\|^2\sigma_{\min}(M) + \|z\|\delta_M\\
&- \Sigma(x,\hat{u},\hat{w})\left(\frac{\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla L_1}{2} + \frac{\tilde{W}_{c1}^{\mathrm{T}}\nabla\phi_1 D_1\nabla L_2}{2}\right)\\
&- \Sigma(x,\hat{u},\hat{w})\left(\frac{\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla L_2}{2} + \frac{\tilde{W}_{c2}^{\mathrm{T}}\nabla\phi_2 D_2\nabla L_1}{2}\right).
\end{aligned}
\tag{9.102}
$$

Now, the cases of $\Sigma(x,\hat{u},\hat{w}) = 0$ and $\Sigma(x,\hat{u},\hat{w}) = 1$ will be considered.

(1) When $\Sigma(x,\hat{u},\hat{w}) = 0$, the first two terms are less than zero. Noting that $\|x\| > 0$ as guaranteed by the persistent excitation condition and using the operator defined in (9.92), it can be ensured that there exists a constant $\dot{x}_{\min}$ satisfying $0 < \dot{x}_{\min} < \|\dot{x}\|$. Then (9.102) becomes

$$
\begin{aligned}
\dot{L} &\le -\dot{x}_{\min}(\|\nabla L_1\| + \|\nabla L_2\|) - \|z\|^2\sigma_{\min}(M) + \|z\|\delta_M\\
&= -\dot{x}_{\min}(\|\nabla L_1\| + \|\nabla L_2\|) - \sigma_{\min}(M)\left(\|z\| - \frac{\delta_M}{2\sigma_{\min}(M)}\right)^2\\
&\quad + \frac{\delta_M^2}{4\sigma_{\min}(M)}.
\end{aligned}
\tag{9.103}
$$

Given that the following inequalities:

$$\|\nabla L_1\| \geq \frac{\delta_M^2}{4\sigma_{\min}(M)\dot{x}_{\min}} \triangleq B_{\nabla L_1}, \tag{9.104}$$

or

$$\|\nabla L_2\| \geq \frac{\delta_M^2}{4\sigma_{\min}(M)\dot{x}_{\min}} \triangleq B_{\nabla L_2}, \tag{9.105}$$

or

$$\|z\| \geq \frac{\delta_M}{\sigma_{\min}(M)} \triangleq B_z \tag{9.106}$$

hold, then $\dot{L} < 0$. Therefore, using Lyapunov theory, it can be concluded that $\|\nabla L_1\|$, $\|\nabla L_2\|$ and $\|z\|$ are UUB.

(2) When $\Sigma(x, \hat{u}, \hat{w}) = 1$, it implies that the feedback control input (9.85) and (9.86) may not stabilize the system (9.58). Adding and subtracting $\nabla L_1^{\mathrm{T}} D_1 \varepsilon_1/2 + \nabla L_2^{\mathrm{T}} D_2 \varepsilon_2/2$ to the right hand side of (9.102), and using (9.64), (9.65), and (9.80), we have

$$\begin{aligned}\dot{L} \leq & \nabla L_1^{\mathrm{T}}(f(x) + g(x)u^* + k(x)w^*) + \nabla L_2^{\mathrm{T}}(f(x) + g(x)u^* + k(x)w^*) \\ & + \frac{1}{2}\nabla L_1 D_1 \nabla \varepsilon_1 + \frac{1}{2}\nabla L_2 D_2 \nabla \varepsilon_2 + \frac{1}{2}\nabla L_1 D_2 \nabla \varepsilon_2 + \frac{1}{2}\nabla L_2 D_1 \nabla \varepsilon_1 \\ & - \sigma_{\min}(M)\left(\|z\| - \frac{\delta_M}{2\sigma_{\min}(M)}\right)^2 + \frac{\delta_M^2}{4\sigma_{\min}(M)}.\end{aligned} \tag{9.107}$$

According to Assumption 9.34, D_i is bounded by D_{iM}, where D_{iM} is a known constant, $i = 1, 2$. Using Lemma 9.31 and recalling the boundedness of $\nabla\varepsilon_1$, $\nabla\varepsilon_2$, and δ, (9.107) can be rewritten as

$$\begin{aligned}\dot{L} \leq & -\bar{Q}_{1\min}\|\nabla L_1\|^2 - \bar{Q}_{2\min}\|\nabla L_2\|^2 + \frac{1}{2}\|\nabla L_1\| D_{1M}\varepsilon_{1dM} \\ & + \frac{1}{2}\|\nabla L_2\| D_{2M}\varepsilon_{2dM} + \frac{1}{2}\|\nabla L_1\| D_{2M}\varepsilon_{2dM} \\ & + \frac{1}{2}\|\nabla L_2\| D_{1M}\varepsilon_{1dM} - \sigma_{\min}(M)\left(\|z\| - \frac{\delta_M}{2\sigma_{\min}(M)}\right)^2 + \frac{\delta_M^2}{4\sigma_{\min}(M)} \\ \leq & -\frac{1}{2}\bar{Q}_{1\min}\|\nabla L_1\|^2 - \frac{1}{2}\bar{Q}_{2\min}\|\nabla L_2\|^2 - \sigma_{\min}(M)\left(\|z\| - \frac{\delta_M}{2\sigma_{\min}(M)}\right)^2 + \eta,\end{aligned} \tag{9.108}$$

where

$$\eta = \frac{D_{1M}^2\varepsilon_{1dM}^2}{4\bar{Q}_{1\min}} + \frac{D_{2M}^2\varepsilon_{2dM}^2}{4\bar{Q}_{2\min}} + \frac{D_{2M}^2\varepsilon_{2dM}^2}{4\bar{Q}_{1\min}} + \frac{D_{1M}^2\varepsilon_{1dM}^2}{4\bar{Q}_{2\min}} + \frac{\delta_M^2}{4\sigma_{\min}(M)}.$$

Given that the following inequalities:

$$\|\nabla L_1\| > \sqrt{\frac{2\eta}{\bar{Q}_{1\min}}} \triangleq B'_{\nabla L_1}, \tag{9.109}$$

or

$$\|\nabla L_2\| > \sqrt{\frac{2\eta}{\bar{Q}_{2\min}}} \triangleq B'_{\nabla L_2}, \tag{9.110}$$

or

$$\|z\| > \sqrt{\frac{\eta}{\sigma_{\min}(M)}} + \frac{\delta_M}{2\sigma_{\min}(M)} \triangleq B'_z \tag{9.111}$$

hold, then $\dot{L} < 0$. Therefore, using Lyapunov theory, it can be concluded that $\|\nabla L_1\|$, $\|\nabla L_2\|$, and $\|z\|$ are UUB.

In summary, for the cases $\Sigma(x, \hat{u}, \hat{w}) = 0$ and $\Sigma(x, \hat{u}, \hat{w}) = 1$, if inequalities $\|\nabla L_1\| > \max(B_{\nabla L_1}, B'_{\nabla L_1}) \triangleq \bar{B}_{\nabla L_1}$, or $\|\nabla L_2\| > \max(B_{\nabla L_2}, B'_{\nabla L_2}) \triangleq \bar{B}_{\nabla L_2}$ or $\|z\| > \max(B_z, B'_z) \triangleq \bar{B}_z$ hold, then $\dot{L} < 0$. Therefore, we can conclude that $\|\nabla L_1\|$, $\|\nabla L_2\|$ and $\|z\|$ are bounded by $\bar{B}_{\nabla L_1}$, $\bar{B}_{\nabla L_2}$, and $\bar{B}_z$, respectively. According to Lemma 9.31, the Lyapunov candidates ∇L_1 and ∇L_2 are radially unbounded and continuously differentiable. Therefore, the boundedness of $\|\nabla L_1\|$ and $\|\nabla L_2\|$ implies the boundedness of $\|x\|$. Specifically, $\|x\|$ is bounded by $\bar{B}_x = \max(B_{1x}, B_{2x})$, where B_{1x} and B_{2x} are determined by $\bar{B}_{\nabla L_1}$ and $\bar{B}_{\nabla L_2}$, respectively. Besides, note that if any component of z exceeds the bound, i.e., $\|\tilde{W}_{c1}\| > \bar{B}_z$ or $\|\tilde{W}_{c2}\| > \bar{B}_z$ or $\|\bar{\sigma}_1^{\mathrm{T}} \tilde{W}_{c1}\| > \bar{B}_z$ or $\|\bar{\sigma}_2^{\mathrm{T}} \tilde{W}_{c2}\| > \bar{B}_z$, the $\|z\|$ are bounded by $\bar{B}_z$, which implies that the critic NN weight estimation errors $\|\tilde{W}_{c1}\|$ and $\|\tilde{W}_{c2}\|$ are also bounded by B_z.

Next, we will prove $\|\hat{u} - u^*\| \leq \epsilon_u$ and $\|\hat{w} - w^*\| \leq \epsilon_w$. From (9.64) and (9.85) and recalling the boundedness of $\|\nabla\phi_1\|$ and $\|\tilde{W}_{c1}\|$, we have

$$\begin{aligned} \|\hat{u} - u^*\| &\leq \left\| -\frac{1}{2} R_{11}^{-1} g^{\mathrm{T}} \nabla\phi_1^{\mathrm{T}} \tilde{W}_{c1} \right\| \\ &\leq \lambda_{\max}(R_{11}^{-1}) \nabla\phi_{1M} \bar{B}_z \\ &\triangleq \epsilon_u. \end{aligned} \tag{9.112}$$

Similarly, from (9.65) and (9.86) and recalling the boundedness of $\|\nabla\phi_2\|$ and $\|\tilde{W}_{c2}\|$, we obtain $\|\hat{w} - w^*\| \leq \epsilon_w$.

This completes the proof. □

Remark 9.36 In [10], each player needs two NNs consisting of a critic NN and an action NN to implement the online learning algorithm. By contrast with [10], only one critic NN is required for each player, the action NN is eliminated, resulting in a simpler architecture, and less computational burden.

Remark 9.37 In Remark 3 of [10] one pointed out that the NN weights can be initialized randomly but non-zero. That is because the method proposed in [10] requires initial stabilizing control policies for guaranteeing the stability of the system. By contrast, no initial stabilizing control policies are needed by adding an operator, which is selected by the Lyapunov's sufficiency condition for stability, on the critic NN weight tuning law for each player in this subsection.

9.4.3 Simulations

Example 9.38 An example is provided to demonstrate the effectiveness of the present control scheme.

Consider the affine nonlinear system as follows:

$$\dot{x} = f(x) + g(x)u + k(x)w, \tag{9.113}$$

where

$$f(x) = \begin{bmatrix} x_2 - 2x_1 \\ -x_2 - 0.5x_1 + 0.25x_2(\cos(2x_1 + 2))^2 + 0.25x_2(\sin(4x_1^2) + 2)^2 \end{bmatrix}, \tag{9.114}$$

$$g(x) = \begin{bmatrix} 0 \\ \cos(2x_1 + 2) \end{bmatrix}, \quad k(x) = \begin{bmatrix} 0 \\ \sin(4x_1^2) + 2) \end{bmatrix}. \tag{9.115}$$

The cost functionals for player 1 and player 2 are defined by (9.59) and (9.60), respectively, where $Q_1(x) = 2x^{\mathrm{T}}x$, $R_{11} = R_{12} = 2I$, $Q_2(x) = x^{\mathrm{T}}x$, $R_{21} = R_{22} = 2I$, and I denotes an identity matrix of appropriate dimensions.

For player 1, the optimal cost function is $V_1^*(x) = 0.25x_1^2 + x_2^2$. For player 2, the optimal cost function is $V_2^*(x) = 0.25x_1^2 + 0.5x_2^2$. The activation functions of critic NNs of two players are selected as $\phi_1 = \phi_2 = [x_1^2, x_1x_2, x_2^2]^{\mathrm{T}}$. Then, the optimal values of the critic NN weights for player 1 are $W_{c1} = [0.5, 0, 1]^{\mathrm{T}}$. The optimal values of the critic NN weights for player 2 are $W_{c2} = [0.25, 0, 0.5]^{\mathrm{T}}$. The estimates of the critic NN weights for two players are denoted $\hat{W}_{c1} = [W_{11}, W_{12}, W_{13}]^{\mathrm{T}}$ and $\hat{W}_{c2} = [W_{21}, W_{22}, W_{23}]^{\mathrm{T}}$, respectively. The adaptive gains for the critic NNs are selected as $a_1 = 1$ and $a_2 = 1$, and the design parameters are selected as $F_1 = F_2 = F_3 = F_4 = 10I$. All NN weights are initialized to zero, which means that no initial stabilizing control policies are needed for implementing the present control scheme. The system state is initialized as $[0.5, 0.2]^{\mathrm{T}}$. To maintain the excitation condition, probing noise is added to the control input for the first 250 s.

After simulation, the trajectories of the system states are shown in Fig. 9.13. The convergence trajectories of the critic NN weights for player 1 are shown in Fig. 9.14, from which we see that the critic NN weights for player 1 finally converge to $[0.4490, 0.0280, 0.9777]^{\mathrm{T}}$. The convergence trajectories of the critic NN weights for player 2 are shown in Fig. 9.15, from which we see that the critic NN weights for

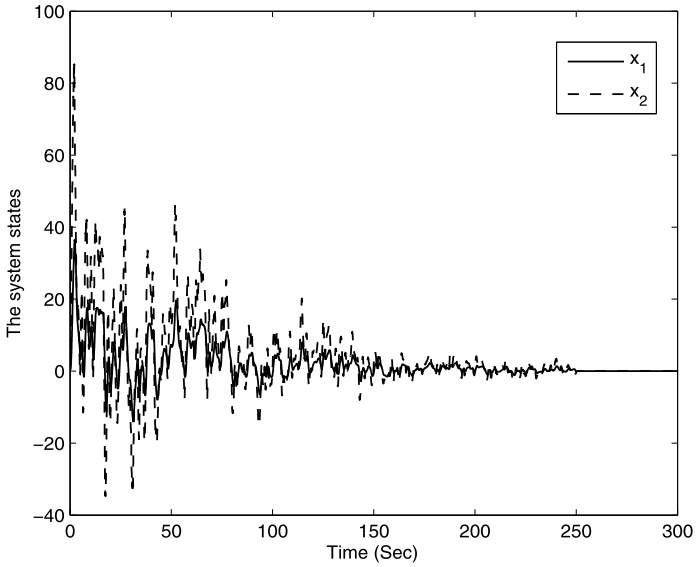

Fig. 9.13 The trajectories of system states

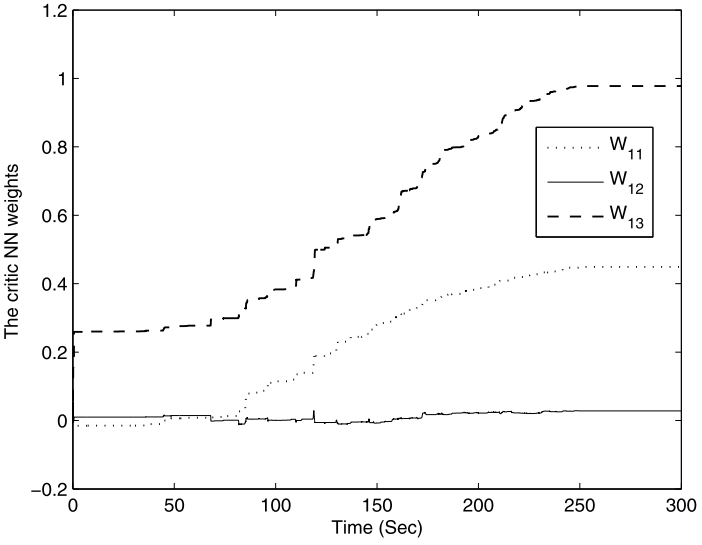

Fig. 9.14 The convergence trajectories of critic NN weights for player 1

player 2 finally converge to $[0.1974, 0.0403, 0.4945]^{\mathrm{T}}$. The convergence trajectory of $e_u = \hat{u} - u^*$ is shown in Fig. 9.16. The convergence trajectory of $e_w = \hat{w} - w^*$ is shown in Fig. 9.17. From Fig. 9.16, we see that the error between the estimated control $\hat{u}$ and the optimal control u^* for player 1 is close to zero when $t = 230$ s.

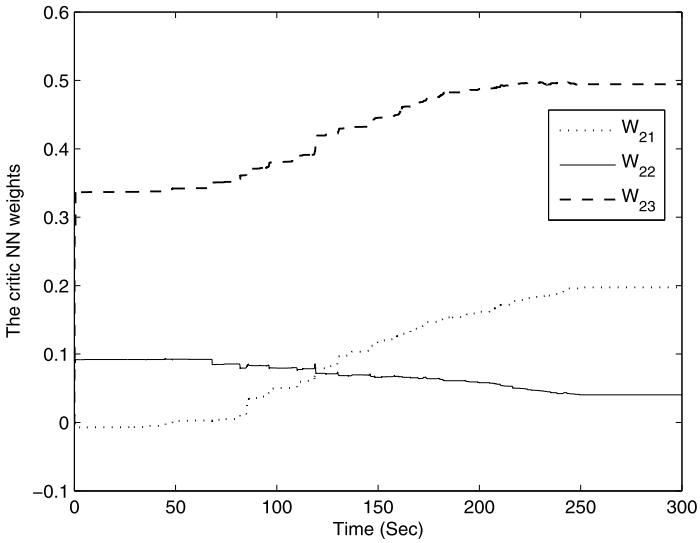

Fig. 9.15 The convergence trajectories of critic NN weights for player 2

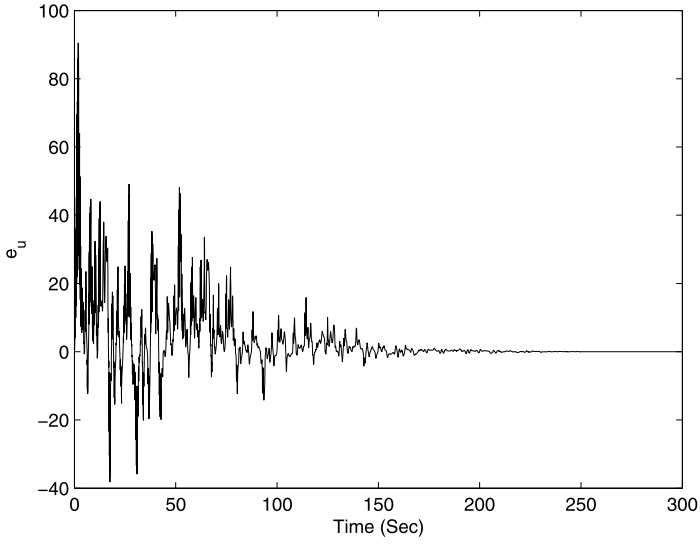

Fig. 9.16 The convergence trajectory of e_u

Similarly, it can been seen from Fig. 9.17 that the estimated control $\hat{w}$ and the optimal control w^* for player 2 are also close to zero when $t = 180$ s. Simulation results reveal that the present control scheme can make the critic NN learn the optimal cost function for each player and meanwhile guarantees stability of the closed-loop system.

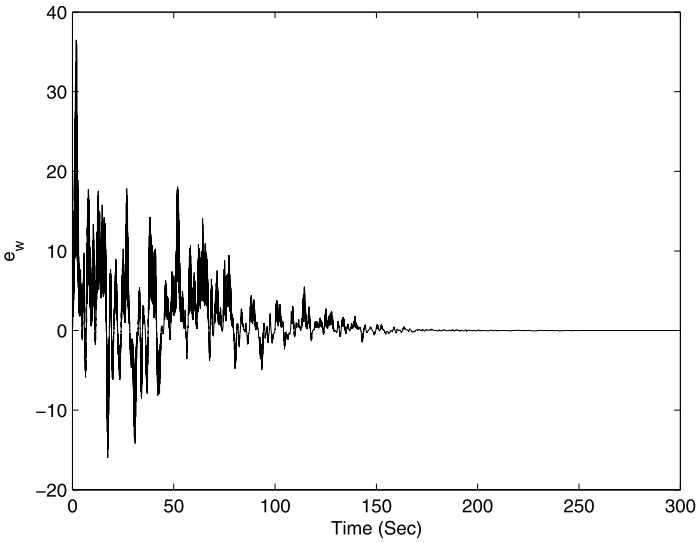

Fig. 9.17 The convergence trajectory of e_w

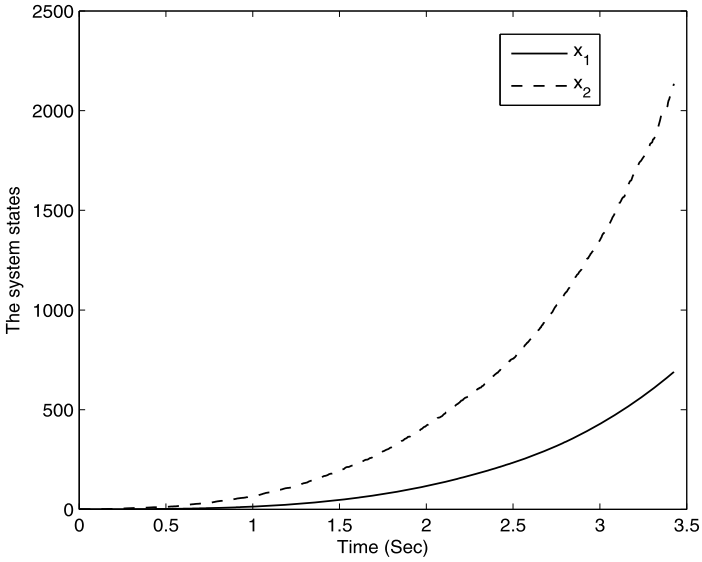

Fig. 9.18 The trajectories of system states obtained by the method in [10] with initial NN weights selected being zero

In order to compare with [10], we use the method proposed in [10] to solve the non-zero-sum games of system (9.113) where all NN weights are initialized to be zero, then obtain the trajectories of system states as shown in Fig. 9.18. It is shown

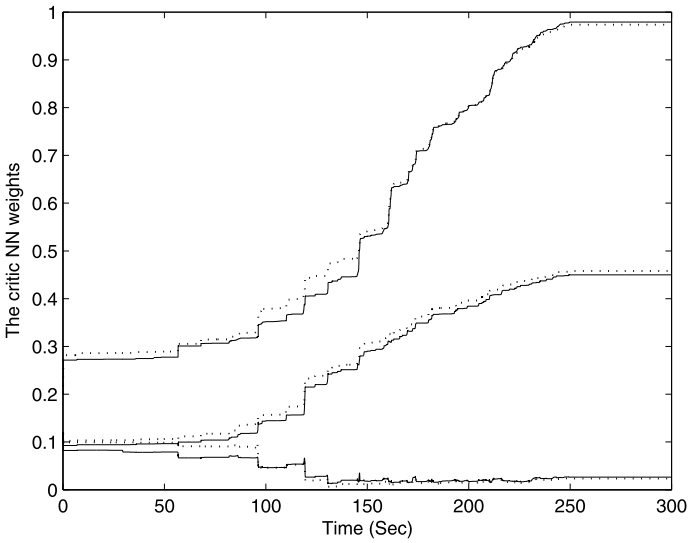

Fig. 9.19 The convergence trajectories of critic NN weights for player 1 (*solid line*: the method in [10]), *dashed line*: our method)

that the system is unstable, which implies that the method in [10] requires initial stabilizing control policies for guaranteeing the stability of the system. By contrast, the present method does not need the initial stabilizing control policies.

As pointed out earlier, one of the main advantages of the single ADP approach is that it results in less computational burden and eliminates the approximation error resulting from the action NNs. To demonstrate this quantitatively, we apply the method in [10] and our method to the system (9.113) with the same initial condition. Figures 9.19 and 9.20 show the convergence trajectories of the critic NN weights for player 1 and player 2, where the solid line and the dashed line represent the results from the method in [10] and our method, respectively. For the convenience of comparison, we define an evaluation function by $\text{PER(i)} = \sum_{k=1}^{N} \|\tilde{W}_i(k)\|$, $i = 1, 2$, which means that the sum of the norm of the critic NN weights error during running time, where N is the number of sample points. The evaluation functions of the critic NN estimation errors as well as the time taken by the method in [10] and our method are calculated and shown in Table 9.1. It clearly indicates that the present method takes less time and obtains a smaller approximation error than [10].

9.5 Summary

In this chapter, we investigated the problem of continuous-time differential games based on ADP. In Sect. 9.2, we developed a new iterative ADP method to obtain

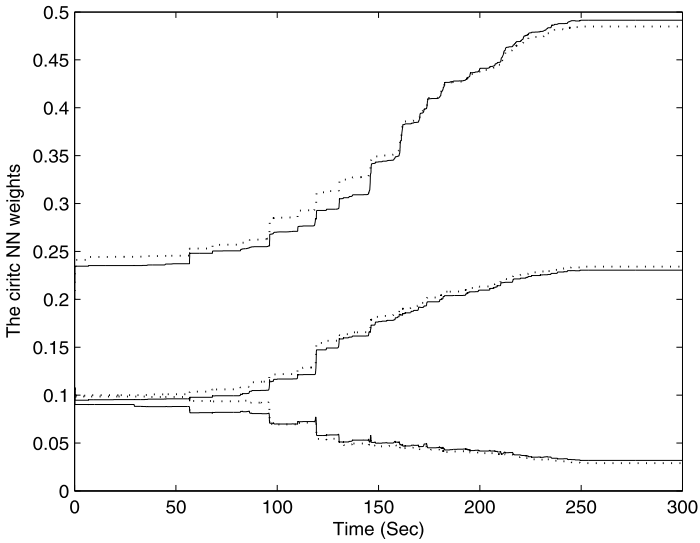

Fig. 9.20 The convergence trajectories of critic NN weights for player 2 (*solid line*: the method in [10]), *dashed line*: our method)

Table 9.1 Critic NN estimation errors and calculation time

Methods	PER(1)	PER(2)	Time
[10]	78.2312	29.4590	184.2024 s
Our method	70.2541	26.4152	111.1236 s

the optimal control pair or the mixed optimal control pair for a class of affine nonlinear zero-sum differential games. In Sect. 9.3, finite horizon zero-sum games for nonaffine nonlinear systems were studied. Then, in Sect. 9.4, the case of non-zero-sum differential games was studied using a single network ADP. Several numerical simulations showed that the present methods are effective.

References

1. Abu-Khalaf M, Lewis FL, Huang J (2006) Policy iterations on the Hamilton–Jacobi–Isaacs equation for H-infinity state feedback control with input saturation. IEEE Trans Autom Control 51:1989–1995
2. Abu-Khalaf M, Lewis FL, Huang J (2008) Neurodynamic programming and zero-sum games for constrained control systems. IEEE Trans Neural Netw 19:1243–1252
3. Al-Tamimi A, Lewis FL, Abu-Khalaf M (2007) Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control. Automatica 43:473–481
4. Bardi M, Capuzzo-Dolcetta I (1997) Optimal control and viscosity solutions of Hamilton–Jacobi–Bellman equations. Birkhäuser, Germany

5. Birkhäuser (1995) H-infinity optimal control and related minimax design problems: a dynamical game approach. Birkhäuser, Berlin
6. Chang HS, Hu J, Fu MC (2010) Adaptive adversial multi-armed bandit approach to two-person zero-sum Markov games. IEEE Trans Autom Control 55:463–468
7. Chen BS, Tseng CS, Uang HJ (2002) Fuzzy differential games for nonlinear stochastic systems: suboptimal approach. IEEE Trans Fuzzy Syst 10:222–233
8. Laraki R, Solan E (2005) The value of zero-sum stopping games in continuous time. SIAM J Control Optim 43:1913–1922
9. Starr AW, Ho YC (1967) Nonzero-sum differential games. J Optim Theory Appl 3:184–206
10. Vamvoudakisand KG, Lewis FL (2011) Multi-player non-zero-sum games: online adaptive learning solution of coupled Hamilton–Jacobi equations. Automatica. doi:10.1016/j.automatica.2011.03.005
11. Wang X (2008) Numerical solution of optimal control for scaled systems by hybrid functions. Int J Innov Comput Inf Control 4:849–856
12. Wei QL, Zhang HG, Liu DR (2008) A new approach to solve a class of continuous-time nonlinear quadratic zero-sum game using ADP. In: Proceedings of IEEE international conference on networking, sensing and control, Sanya, China, pp 507–512
13. Wei QL, Zhang HG, Cui LL (2009) Data-based optimal control for discrete-time zero-sum games of 2-D systems using adaptive critic designs. Acta Autom Sin 35:682–692
14. Wei QL, Zhang HG, Dai J (2009) Model-free multiobjective approximate dynamic programming for discrete-time nonlinear systems with general performance index functions. Neurocomputing 7–9:1839–1848
15. Zhang HG, Wei QL, Liu DR (2011) An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games. Automatica 47:207–214
16. Zhang X, Zhang HG, Wang XY (2011) A new iteration approach to solve a class of finite-horizon continuous-time nonaffine nonlinear zero-sum game. Int J Innov Comput Inf Control 7:597–608
17. Zhang HG, Cui LL, Luo YH (2012) Near-optimal control for non-zero-sum differential games of continuous-time nonlinear systems using single network ADP. IEEE Trans Syst Man Cybern, Part B, Cybern. doi:10.1109/TSMCB.2012.2203336