Automatica 47 (2011) 207–214

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games[☆]

Huaguang Zhang[a,*], Qinglai Wei[b], Derong Liu[b]

[a] School of Information Science and Engineering, Northeastern University, Shenyang 110004, PR China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Article history:
Available online 4 December 2010

Keywords:
Adaptive critic designs
Adaptive dynamic programming
Approximate dynamic programming
Neural network
Zero-sum differential games

ABSTRACT

In this paper, a new iterative adaptive dynamic programming (ADP) method is proposed to solve a class of continuous-time nonlinear two-person zero-sum differential games. The idea is to use the ADP technique to obtain the optimal control pair iteratively which makes the performance index function reach the saddle point of the zero-sum differential games. If the saddle point does not exist, the mixed optimal control pair is obtained to make the performance index function reach the mixed optimum. Stability analysis of the nonlinear systems is presented and the convergence property of the performance index function is also proved. Two simulation examples are given to illustrate the performance of the proposed method.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Nowadays, zero-sum differential game theory has been widely applied to decision making problems. Many control schemes are presented in order to reach some form of optimality, e.g., obtaining the saddle point. Hence, in most papers, the existence of the saddle point is proposed before obtaining the saddle point (Abu-Khalaf, Lewis, & Huang, 2006; Chang, Hu, Fu, & Marcus, 2010; Jain & Watrous, 2009; Jiang, 2009; Shi, 2002; Wang, 2008). In the real world, however, the existence conditions of the saddle point are so difficult to satisfy that many applications of the zero-sum differential games are restricted to linear systems (Engwerda, 2008; Jimenez-Lizarraga, Basin, & Alcorta-Garcia, 2009). On the other hand, for many zero-sum differential games, especially in nonlinear cases, the saddle point of the game does not exist, which means that we can only obtain the mixed optimal solution of the games. Meanwhile, the mixed trajectory method is also very difficult to apply (Basar & Olsder, 1982). The main difficulty of the mixed trajectory method is that the optimal probability distribution is too hard to obtain in the whole real space. Furthermore, the mixed optimal solution is hardly obtained once the control schemes are determined. In most cases (i.e., in engineering case), however, the saddle point or mixed optimal solution of the zero-sum differential games has to be achieved by a deterministic optimal control scheme or by a mixed optimal control scheme. Therefore, how to obtain the saddle point without the complex existence conditions of the saddle point and how to obtain the mixed optimal solution under a deterministic mixed optimal control scheme when the saddle point does not exist are important research topics. These motivate our research.

Adaptive dynamic programming (ADP) is a powerful method for solving optimal control problems and has been paid much attention in recent years (Wang, Zhang, & Liu, 2009; Zhang, Wei, & Luo, 2008). ADP has also been applied to solve zero-sum games (Abu-Khalaf, Lewis, & Huang, 2008; Al-Tamimi, Lewis, & Abu-Khalaf, 2007). While in these papers, ADP was only used to solve zero-sum games for the situation that the saddle point exists under the complex existence conditions, so the applications are very limited. To the best of our knowledge, there are no ADP methods for solving zero-sum differential games under the condition that the saddle point does not exist.

In this paper, we propose a new iterative ADP method which is effective both for the situations that the saddle point exists or does not exist. For the situation that the saddle point exists, the existence conditions of the saddle point are avoided. The performance index function can reach the saddle point using the

[☆] This work was supported in part by the National Natural Science Foundation of China under Grants 60821063, 60621001, 60728307, 61034005, 60904037, 60921061 and 61034002, in part by the Beijing Natural Science Foundation under grant 4102061, in part by the Program for Changjiang Scholars and Innovative Research Groups of China under Grant No. 60521003, and in part by the National High Technology Research and Development Program of China under Grant No. 2006AA04Z183. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Kok Lay Teo under the direction of Editor Ian R. Petersen.

[*] Corresponding author. Tel.: +86 24 83687762; fax: +86 24 83687762.
E-mail addresses: hgzhang@ieee.org (H. Zhang), qinglai.wei@ia.ac.cn (Q. Wei), derong.liu@ia.ac.cn (D. Liu).

0005-1098/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2010.10.033

proposed iterative ADP method. For the situation that the saddle point does not exist, it is emphasized that for the first time, the mixed optimal performance index function is obtained under a deterministic mixed optimal control scheme, using the proposed iterative ADP algorithm.

2. Preliminaries and assumptions

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

$$\dot{x} = f(x, u, w) = a(x) + b(x)u + c(x)w, \tag{1}$$

where $x \in R^n$, $u \in R^k$, $w \in R^m$, and the initial condition $x(0) = x_0$ is given. The performance index function is the following generalized quadratic form (see Basar & Bernhard, 1995, p. 7) given by

$$V(x(0), u, w) = \int_0^\infty l(x, u, w)\mathrm{d}t, \tag{2}$$

where $l(x, u, w) = x^TAx + u^TBu + w^TCw + 2u^TDw + 2x^TEu + 2x^TFw$. The matrices A, B, C, D, E, F are with suitable dimensions and $A \geq 0$, $B > 0$, $C < 0$. According to the situation of two players we have the following definitions. Let $\overline{V}(x) := \inf_u \sup_w V(x, u, w)$ be the upper performance index function and $\underline{V}(x) := \sup_w \inf_u V(x, u, w)$ be the lower performance index function with the obvious inequality $\overline{V}(x) \geq \underline{V}(x)$ (Basar & Bernhard, 1995). Define the optimal control pairs to be $(\overline{u}, \overline{w})$ and $(\underline{u}, \underline{w})$ for upper and lower performance index functions, respectively. Then, we have $\overline{V}(x) = V(x, \overline{u}, \overline{w})$ and $\underline{V}(x) = V(x, \underline{u}, \underline{w})$. If both $\overline{V}(x)$ and $\underline{V}(x)$ exist and

$$\overline{V}(x) = \underline{V}(x) = V^*(x) \tag{3}$$

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 1 (*Bardi & Capuzzo-Dolcetta, 1997*). *If the nonlinear system (1) is controllable and the upper performance index function and the lower performance index function both exist, then $\overline{V}(x)$ is a solution of the following upper Hamilton–Jacobi–Isaacs (HJI) equation*

$$\inf_u \sup_w \{\overline{V}_t + \overline{V}_x^T f(x, u, w) + l(x, u, w)\} = 0, \tag{4}$$

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

$$\sup_w \inf_u \{\underline{V}_t + \underline{V}_x^T f(x, u, w) + l(x, u, w)\} = 0, \tag{5}$$

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

3. Iterative adaptive dynamic programming method for zero-sum differential games

As the HJI Eqs. (4) and (5) cannot be solved in general, in this section, a new iterative ADP method for zero-sum differential games is proposed.

3.1. Derivation of the iterative ADP method

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

Theorem 1. *Let $(\overline{u}, \overline{w})$ be the optimal control pair for $\overline{V}(x)$ and $(\underline{u}, \underline{w})$ be the optimal control pair for $\underline{V}(x)$. Then, there exist control pairs $(\overline{u}, w)$ and $(u, \underline{w})$ which make $V^o(x) = V(x, \overline{u}, w) = V(x, u, \underline{w})$. Furthermore, if the saddle point exists, then $V^o(x) = V^*(x)$.*

Proof. According to the definition of $\overline{V}(x)$, we have $V(x, \overline{u}, w) \leq V(x, \overline{u}, \overline{w})$. As $V^o(x)$ is mixed optimal performance index function, we also have $V^o(x) \leq V(x, \overline{u}, \overline{w})$ (Owen, 1982). As system (1) is controllable and w is continuous on R^m, there exists a control pair $(\overline{u}, w)$ which makes $V^o(x) = V(x, \overline{u}, w)$. On the other hand, we have $V^o(x) \geq V(x, \underline{u}, \underline{w})$. We also have $V(x, u, \underline{w}) \geq V(x, \underline{u}, \underline{w})$. As u is continuous on R^k, there exists control pair $(u, \underline{w})$ which makes $V^o(x) = V(x, u, \underline{w})$. If the saddle point exists, we have (3) holds. On the other hand, $\underline{V}(x) \leq V^o(x) \leq \overline{V}(x)$. Then clearly $V^o(x) = V^*(x)$. □

If Eq. (3) holds, we have a saddle point; if not, we adopt the mixed trajectory method 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 performance index function as $E(V(x)) = \min_{P_{Ii}} \max_{P_{IIj}} \sum_{i=1}^2 \sum_{j=1}^2 P_{Ii}L_{ij}P_{IIj}$, where we let $L_{11} = V(x, \overline{u}, \overline{w})$, $L_{12} = V(x, (\underline{u} + \gamma_u), \underline{w})$, $L_{21} = V(x, \underline{u}, \underline{w})$ and $L_{22} = V(x, \overline{u}, (\overline{w} + \gamma_w))$. 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 performance index function for N times, we can obtain $E_1(V(x)), E_2(V(x)), \ldots, E_N(V(x))$. Then, the mixed optimal performance index function can be written as

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

Remark 1. In the classical mixed trajectory method (Basar & Olsder, 1982), the whole control sets R^k and R^m should be searched under some distribution functions. As there are no constraints for both controls, we can 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 as 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 3). Similar works can also be seen in Al-Tamimi et al. (2007) and Wei, Zhang, and Dai (2009).

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 proposed in this paper. Let $\alpha = (V^o(x) - \underline{V}(x))/(\overline{V}(x) - \underline{V}(x))$. Then $V^o(x)$ can be written as $V^o(x) = \alpha\overline{V}(x) + (1-\alpha)\underline{V}(x)$. Let $l^o(x, \overline{u}, \overline{w}, \underline{u}, \underline{w}) = \alpha l(x, \overline{u}, \overline{w}) + (1-\alpha)l(x, \underline{u}, \underline{w})$. Then we have $V^o(x(0)) = \int_0^\infty l^o\mathrm{d}t$. According to Theorem 1, the mixed optimal control pair can be obtained by regulating the control $\overline{w}$ in the control pair $(\overline{u}, \overline{w})$ that minimizes the error between $\mathcal{V}(x)$ and $V^o(x)$, where the performance index function $\mathcal{V}(x)$ is defined as $\mathcal{V}(x(0)) = V(x(0), \overline{u}, w) = \int_0^\infty l(x, \overline{u}, w)\mathrm{d}t$ and $\underline{V}(x(0)) \leq \mathcal{V}(x(0)) \leq \overline{V}(x(0))$. Define $\widetilde{V}(x(0)) = \int_0^\infty \widetilde{l}(x, w)\mathrm{d}x$, 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 $\min_w(\widetilde{V}(x))^2$. According to the principle of optimality, when $\widetilde{V}(x) \geq 0$ we have the following Hamilton–Jacobi–Bellman (HJB) equation

$$\begin{aligned}\mathrm{HJB}(\widetilde{V}(x), w) &:= \min_{w}\{\widetilde{V}_t(x) + \widetilde{V}_x f(x, u, w) + \widetilde{l}(x, w)\}\\ &= 0. \end{aligned} \tag{6}$$

For $\widetilde{V}(x) < 0$, we have $-\widetilde{V}(x) = -(\mathcal{V}(x) - V^o(x)) > 0$, and we can also obtain the same HJB equation as (6).

3.2. The iterative ADP algorithm

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

Step 1. Initialize the algorithm with a stabilizing control pair $(u^{(0)}, w^{(0)})$ and the performance index function $V^{(0)}$. Choose the computation precision $\zeta > 0$. Set $i = 0$.

Step 2. For the upper performance index function, let

$$\overline{V}^{(i)}(x(0)) = \int_0^\infty l(x, \overline{u}^{(i+1)}, \overline{w}^{(i+1)})\mathrm{d}t, \tag{7}$$

where the iterative optimal control pair is formulated as

$$\begin{aligned}\overline{u}^{(i+1)} &= -\frac{1}{2}(B - DC^{-1}D^T)^{-1}\big(2(E^T - DC^{-1}F^T)x\\ &\quad + (b^T(x) - DC^{-1}c^T(x))\overline{V}_x^{(i)}\big),\end{aligned} \tag{8}$$

and

$$\overline{w}^{(i+1)} = -\frac{1}{2}C^{-1}\big(2D^T\overline{u}^{(i+1)} + 2F^Tx + c^T(x)\overline{V}_x^{(i)}\big), \tag{9}$$

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

Step 3. If $\left|\overline{V}^{(i+1)}(x(0)) - \overline{V}^{(i)}(x(0))\right| < \zeta$, let $\overline{u} = \overline{u}^{(i)}$, $\overline{w} = \overline{w}^{(i)}$ and $\overline{V}(x) = \overline{V}^{(i+1)}(x)$, set $i = 0$ and go to Step 4. Else, set $i = i + 1$ and go to Step 2.

Step 4. For the lower performance index function, let

$$\underline{V}^{(i)}(x(0)) = \int_0^\infty l(x, \underline{u}^{(i+1)}, \underline{w}^{(i+1)})\mathrm{d}t \tag{10}$$

where the iterative optimal control pair is formulated as

$$\underline{u}^{(i+1)} = -\frac{1}{2}B^{-1}(2D\underline{w}^{(i+1)} + 2E^Tx + b^T(x)\underline{V}_x^{(i)}), \tag{11}$$

and

$$\begin{aligned}\underline{w}^{(i+1)} &= -\frac{1}{2}(C - D^TBD)^{-1}(2(F^T - D^TB^{-1}E)x\\ &\quad + (c^T(x) - D^TB^{-1}b^T(x))\underline{V}_x^{(i)}),\end{aligned} \tag{12}$$

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

Step 5. If $\left|\underline{V}^{(i+1)}(x(0)) - \underline{V}^{(i)}(x(0))\right| < \zeta$, let $\underline{u} = \underline{u}^{(i)}$, $\underline{w} = \underline{w}^{(i)}$ and $\underline{V}(x) = \underline{V}^{(i+1)}(x)$. Set $i = 0$ and go to Step 6. Else, set $i = i + 1$ and go to Step 4.

Step 6. If $\left|\overline{V}(x(0)) - \underline{V}(x(0))\right| < \zeta$, stop, and the saddle point is achieved. Else, set $i = 0$ and go to the next step.

Step 7. Regulate the control w for the upper performance index function and let

$$\begin{aligned}\widetilde{V}^{(i+1)}(x(0)) &= \mathcal{V}^{(i+1)}(x(0)) - V^o(x(0))\\ &= \int_0^\infty \widetilde{l}(x, \overline{u}, w^{(i)})\mathrm{d}t.\end{aligned} \tag{13}$$

The iterative optimal control is formulated as

$$w^{(i)} = -\frac{1}{2}C^{-1}(2D^T\overline{u} + 2F^Tx + c^T(x)\widetilde{V}_x^{(i+1)}) \tag{14}$$

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

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

3.3. Properties of the iterative ADP algorithm

In this subsection, results are presented to show the stability and convergence of the proposed iterative ADP algorithm.

Theorem 2. *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 (1) asymptotically stable.*

Proof. According to (7), for $\forall\, t$, taking the derivative of $\overline{V}^{(i)}(x)$, we have

$$\frac{\mathrm{d}\overline{V}^{(i)}(x)}{\mathrm{d}t} = \overline{V}_x^{(i)T}\left(a(x) + b(x)\overline{u}^{(i+1)} + c(x)\overline{w}^{(i+1)}\right). \tag{15}$$

From the HJI equation we have

$$0 = \overline{V}_x^{(i)T}f(x, \overline{u}^{(i)}, \overline{w}^{(i)}) + l(x, \overline{u}^{(i)}, \overline{w}^{(i)}). \tag{16}$$

Combining (15) and (16), we get

$$\begin{aligned}\frac{\mathrm{d}\overline{V}^{(i)}(x)}{\mathrm{d}t} &= \overline{V}_x^{(i)T}(b(x) - c(x)C^{-1}D^T)(\overline{u}^{(i+1)} - \overline{u}^{(i)})\\ &\quad - x^TAx - \overline{u}^{(i)T}(B - DC^{-1}D^T)\overline{u}^{(i)}\\ &\quad - \frac{1}{4}\overline{V}_x^{(i)T}c(x)C^{-1}c^T(x)\overline{V}_x^{(i)}\\ &\quad - 2x^T(E - FC^{-1}D^T)\overline{u}^{(i+1)} + x^TFC^{-1}F^Tx.\end{aligned} \tag{17}$$

According to (8), we have

$$\begin{aligned}\frac{\mathrm{d}\overline{V}^{(i)}(x)}{\mathrm{d}t} &= -(\overline{u}^{(i+1)} - \overline{u}^{(i)})^T(B - DC^{-1}D^T)\\ &\quad \times(\overline{u}^{(i+1)} - \overline{u}^{(i)}) - l(x, \overline{u}^{(i+1)}, \overline{w}^{(i+1)})\\ &\leq 0.\end{aligned} \tag{18}$$

So, $\overline{V}^{(i)}(x)$ is a Lyapunov function (Cloutier, 1997). 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

$$\alpha(\varepsilon) \geq \beta(\delta) \geq \overline{V}^{(i)}(x(t_0)) \geq \overline{V}^{(i)}(x(t)) \geq \alpha(\|x\|). \tag{19}$$

Then we have that system (1) is asymptotically stable (details can be seen in Liao, Wang, and Yu (2007)). □

Theorem 3. *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 (1) asymptotically stable.*

Corollary 1. *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 (1) asymptotically stable.*

Proof. As $\underline{V}^{(i)}(x) \le \overline{V}^{(i)}(x)$ and $l(x, \underline{u}^{(i)}, \underline{w}^{(i)}) \ge 0$, we have $0 \le \underline{V}^{(i)}(x) \le \overline{V}^{(i)}(x)$. From Theorem 2, we know that for $\forall t_0$, there exist two functions $\alpha(\|x\|)$ and $\beta(\|x\|)$ which belong to class $\mathcal{K}$ and satisfy (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

$$\overline{V}^{(i)}(x(t_0)) \ge \overline{V}^{(i)}(x(t_1)) \ge \underline{V}^{(i)}(x(t_0)) \ge \overline{V}^{(i)}(x(t_2)). \tag{20}$$

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

$$\begin{aligned}\underline{V}^{(i)}(x(t_0)) &\ge \alpha(\varepsilon_1) \ge \beta(\delta_1) \ge \overline{V}^{(i)}(x(t_2)) \ge \overline{V}^{(i)}(x(t)) \\ &\ge \underline{V}^{(i)}(x(t)) \ge \alpha(\|x\|).\end{aligned} \tag{21}$$

According to (19), we have

$$\begin{aligned}\alpha(\varepsilon) &\ge \beta(\delta) \ge \underline{V}^{(i)}(x(t_0)) \ge \alpha(\varepsilon_1) \ge \beta(\delta_1) \ge \underline{V}^{(i)}(x(t)) \\ &\ge \alpha(\|x\|).\end{aligned} \tag{22}$$

Since $\alpha(\|x\|)$ belongs to class $\mathcal{K}$, we can obtain $\|x\| \le \varepsilon$. Then we can conclude that the system (1) is asymptotically stable. □

Corollary 2. *If for $\forall i \ge 0$, $\mathrm{HJI}(\overline{V}^{(i)}(x), \overline{u}^{(i)}, \overline{w}^{(i)}) = 0$ 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 (1) asymptotically stable.*

Theorem 4. *If for $\forall i \ge 0$, $\mathrm{HJI}(\overline{V}^{(i)}(x), \overline{u}^{(i)}, \overline{w}^{(i)}) = 0$ 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 (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 $l(x, \overline{u}^{(i)}, \overline{w}^{(i)}) \ge 0$ 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 $n = 0, 1, \ldots$.

For $t \in [t_0, t_1)$, we have $l(x, \overline{u}^{(i)}, \overline{w}^{(i)}) \ge 0$ and $\int_{t_0}^{t_1} l(x, \overline{u}^{(i)}, \overline{w}^{(i)})\mathrm{d}t \ge 0$. According to Theorem 2, we have $\|x(t_0)\| \ge \|x(t)\| \ge \|x(t_1)\|$. 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 2, we have $\|x(t_1)\| > \|x(t)\| > \|x(t_2)\|$. So, we can obtain $\|x(t_0)\| \ge \|x(t)\| > \|x(t_2)\|$, for $\forall t \in [t_0, t_2)$.

Then, using the mathematical induction, for $\forall t$, we have $\|x(t')\| \le \|x(t)\|$, where $t' \in [t, \infty)$. So, we can conclude that the system (1) is asymptotically stable and the proof is completed. □

Theorem 5. *If for $\forall i \ge 0$, $\mathrm{HJI}(\underline{V}^{(i)}(x), \underline{u}^{(i)}, \underline{w}^{(i)}) = 0$ 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 (1) asymptotically stable.*

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

Proposition 1. *If for $\forall i \ge 0$, $\mathrm{HJI}(\overline{V}^{(i)}(x), \overline{u}^{(i)}, \overline{w}^{(i)}) = 0$ holds, then the control pairs $(\overline{u}^{(i)}, \overline{w}^{(i)})$ make the upper performance index function satisfy*

$$\lim_{i\to\infty} \overline{V}^{(i)}(x) = \overline{V}(x).$$

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

$$\begin{aligned}\frac{\mathrm{d}\overline{V}^{(i+1)}(x)}{\mathrm{d}t} &= -(x^TAx + \overline{u}^{(i+1)T}(B - DC^{-1}D^T)\overline{u}^{(i+1)} \\ &\quad + \frac{1}{4}\overline{V}_x^{(i)T}c(x)C^{-1}c^T(x)\overline{V}_x^{(i)} \\ &\quad + 2x^T(E - FC^{-1}D^T)\overline{u}^{(i+1)} - x^TFC^{-1}F^Tx).\end{aligned} \tag{23}$$

According to (18), we can obtain

$$\begin{aligned}\frac{\mathrm{d}(\overline{V}^{(i+1)}(x) - \overline{V}^{(i)}(x))}{\mathrm{d}t} &= \frac{\mathrm{d}\overline{V}^{(i+1)}(x)}{\mathrm{d}t} - \frac{\mathrm{d}\overline{V}^{(i)}(x)}{\mathrm{d}t} \\ &= (\overline{u}^{(i+1)} - \overline{u}^{(i)})^T(B - DC^{-1}D^T) \\ &\quad \times (\overline{u}^{(i+1)} - \overline{u}^{(i)}) > 0.\end{aligned} \tag{24}$$

Since the system (1) is asymptotically stable, its state x converges to zero, and so does $\overline{V}^{(i+1)}(x) - \overline{V}^{(i)}(x)$. Since $\mathrm{d}\left(\overline{V}^{(i+1)}(x) - \overline{V}^{(i)}(x)\right)/\mathrm{d}t \ge 0$ on these trajectories, it implies that $\overline{V}^{(i+1)}(x) - \overline{V}^{(i)}(x) \le 0$; that is $\overline{V}^{(i+1)}(x) \le \overline{V}^{(i)}(x)$. As such, $\overline{V}^{(i)}(x)$ 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 $\overline{w}^* = \arg\max_w \left\{\int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{(i)}(x(\hat{t}))\right\}$. Then, according to the principle of optimality (Basar & Olsder, 1982), we have

$$\begin{aligned}\overline{V}^{(i)}(x) &\le \sup_w \left\{\int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{(i)}(x(\hat{t}))\right\} \\ &= \int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{(i)}(x(\hat{t})).\end{aligned} \tag{25}$$

Since $\overline{V}^{(i+1)}(x) \le \overline{V}^{(i)}(x)$, we have $\overline{V}^{(\infty)}(x) \le \int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{(i)}(x(\hat{t}))$. Let $i \to \infty$, we can obtain $\overline{V}^{(\infty)}(x) \le \int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{(\infty)}(x(\hat{t}))$. So, we have $\overline{V}^{(\infty)}(x) \le \inf_u \sup_w \left\{\int_t^{\hat{t}} l(x,u,w)\mathrm{d}t + \overline{V}^{(i)}(x(\hat{t}))\right\}$.

Let $\epsilon > 0$ be an arbitrary positive number. Since the upper performance index function is nonincreasing and convergent, there exists a positive integer i such that $\overline{V}^{(i)}(x) - \epsilon \le \overline{V}^{(\infty)}(x) \le \overline{V}^{(i)}(x)$. Let $\overline{u}^* = \arg\min_u \left\{\int_t^{\hat{t}} l(x,u,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{(i)}(x(\hat{t}))\right\}$. Then, we can 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

$$\begin{aligned}\overline{V}^{(\infty)}(x) &\ge \int_t^{\hat{t}} l(x,\overline{u}^*,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{(i)}(x(\hat{t})) - \epsilon \\ &\ge \int_t^{\hat{t}} l(x,\overline{u}^*,\overline{w}^*)\mathrm{d}\tau + \overline{V}^{(\infty)}(x(\hat{t})) - \epsilon \\ &= \inf_u \sup_w \left\{\int_t^{\hat{t}} l(x,u,w)\mathrm{d}\tau + \overline{V}^{(\infty)}(x(\hat{t}))\right\} - \epsilon.\end{aligned} \tag{26}$$

Since ϵ is arbitrary, we have

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

Therefore, we can 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 $\overline{V}^{(\infty)}(x) = \inf_u \sup_w V(x,u,w) = \overline{V}(x)$. □

Proposition 2. *If for $\forall i \ge 0$, $\mathrm{HJI}(\underline{V}^{(i)}(x), \underline{u}^{(i)}, \underline{w}^{(i)}) = 0$ holds, then the control pairs $(\underline{u}^{(i)}, \underline{w}^{(i)})$ make the lower performance index function $\underline{V}^{(i)}(x) \to \underline{V}(x)$ as $i \to \infty$.*

Theorem 6. *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 $\overline{V}^{(i)}(x) \to V^*(x)$ and $\underline{V}^{(i)}(x) \to V^*(x)$, respectively, as $i \to \infty$.*

Proof. For the upper performance index function, according to Proposition 1, we have $\overline{V}^{(i)}(x) \to \overline{V}(x)$ under the control pairs $(\overline{u}^{(i)}, \overline{w}^{(i)})$ as $i \to \infty$. So the optimal control pair for upper performance index function satisfies $\overline{V}(x) = V(x, \overline{u}, \overline{w}) = \inf_u \sup_w V(x, u, w)$. On the other hand, there exists an optimal control pair (u^*, w^*) to make the performance index reach the saddle point. According to the property of the saddle point (Basar & Olsder, 1982; Owen, 1982), the optimal control pair (u^*, w^*) satisfies $V^*(x) = V(x, u^*, w^*) = \inf_u \sup_w V(x, u, w)$. So, we have $\overline{V}^{(i)}(x) \to V^*(x)$ under the control pair $(\overline{u}^{(i)}, \overline{w}^{(i)})$ as $i \to \infty$. Similarly, we can derive $\underline{V}^{(i)}(x) \to V^*(x)$ under the control pairs $(\underline{u}^{(i)}, \underline{w}^{(i)})$ as $i \to \infty$. □

Remark 2. From the proofs we can see that the complex existence conditions of the saddle point in Abu-Khalaf et al. (2006, 2008), Bianchini, Genesio, Parenti, and Tesi (2004) and Tsiotras, Corless, and Rotea (1998), are not necessary. If the saddle point exists, the iterative performance index functions can converge to the saddle point using the proposed iterative ADP algorithm. This is an important contribution of the proposed algorithm in this paper.

We emphasize that when the saddle point does not exist, the mixed optimal solution can be obtained effectively using the iterative ADP algorithm. The following propositions can be proved using similar steps as the proofs of Theorem 2 and Proposition 1.

Proposition 3. *If $\bar{u} \in R^k$, $w^{(i)} \in R^m$ and the utility function is $\tilde{l}(x, w^{(i)}) = l(x, \bar{u}, w^{(i)}) - l^o(x, \overline{u}, \overline{w}, \underline{u}, \underline{w})$, and $w^{(i)}$ is expressed in (14), then the control pairs $(\bar{u}, w^{(i)})$ make the system (1) asymptotically stable.*

Proposition 4. *If $\bar{u} \in R^k$, $w^{(i)} \in R^m$ and for $\forall t$, the utility function $\tilde{l}(x, w^{(i)}) \geq 0$, then the control pairs $(\bar{u}, w^{(i)})$ make $\widetilde{V}^{(i)}(x)$ a nonincreasing convergent sequence as $i \to \infty$.*

Proposition 5. *If $\bar{u} \in R^k$, $w^{(i)} \in R^m$ and for $\forall t$, the utility function $\tilde{l}(x, w^{(i)}) < 0$, then the control pairs $(\bar{u}, w^{(i)})$ make $\widetilde{V}^{(i)}(x)$ a nondecreasing convergent sequence as $i \to \infty$.*

Theorem 7. *If $\bar{u} \in R^k$, $w^{(i)} \in R^m$ and $\tilde{l}(x, w^{(i)})$ is the utility function, then the control pairs $(\bar{u}, w^{(i)})$ 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 4, 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 5, we have $\widetilde{V}^{(i+1)}(x) > \widetilde{V}^{(i)}(x)$. Then, for $\forall t_0$, we have

$$\begin{aligned}\left\|\widetilde{V}^{(i+1)}(x(t_0))\right\| &= \left\|\int_{t_0}^{t_1} \tilde{l}(x, w^{(i)}) \mathrm{d}t\right\| + \left\|\int_{t_1}^{t_2} \tilde{l}(x, w^{(i)}) \mathrm{d}t\right\| \\ &\quad + \cdots + \left\|\int_{t_m}^{t_{(m+1)}} \tilde{l}(x, w^{(i)}) \mathrm{d}t\right\| + \cdots \\ &< \left\|\widetilde{V}^{(i)}(x(t_0))\right\|. \end{aligned} \tag{27}$$

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

Theorem 8. *If $\bar{u} \in R^k$, $w^{(i)} \in R^m$ and $\tilde{l}(x, w^{(i)})$ is the utility function, then the control pairs $(\bar{u}, w^{(i)})$ make $\mathcal{V}^{(i)}(x) \to V^o(x)$ as $i \to \infty$.*

Proof. It is proved by contradiction. Suppose that the control pairs $(\bar{u}, w^{(i)})$ make the performance index function $\mathcal{V}^{(i)}(x)$ converge to $\mathcal{V}'(x)$ and $\mathcal{V}'(x) \neq V^o(x)$.

According to Theorem 7, based on the principle of optimality, as $i \to \infty$ we have the HJB equation $\mathrm{HJB}(\widetilde{V}(x), w) = 0$. From the assumptions we know that $|\mathcal{V}^{(i)}(x) - V^o(x)| \neq 0$ as $i \to \infty$. From Theorem 1, we know that there exists a control pair $(\bar{u}, w')$ that makes $V(x, \bar{u}, w') = V^o(x)$ which minimizes the performance index function $\widetilde{V}(x)$. According to the principle of optimality, we also have the HJB equation $\mathrm{HJB}(\widetilde{V}(x), w') = 0$. It is a contradiction. So the assumption does not hold. Thus, we have $\mathcal{V}^{(i)}(x) \to V^o(x)$ as $i \to \infty$. □

Remark 3. For the situation where the saddle point does not exist, the methods in Abu-Khalaf et al. (2006, 2008) and Bianchini et al. (2004) are all invalid. While using our iterative ADP method, the iterative performance index function reached the mixed optimal performance index function $V^o(x)$ under the deterministic control pair. This is another important contribution of the proposed algorithm in this paper. Therefore, we emphasize that the proposed iterative ADP method is more effective.

4. Simulation study

In the first example, we will show that using the proposed iterative ADP method, the saddle point of the game can be obtained where the existence conditions of the saddle point are avoided and we will compare the results with Abu-Khalaf et al. (2008).

Example 1. The dynamics of the benchmark nonlinear plant can be expressed by system (1), where

$$\begin{aligned} a(x) &= \begin{bmatrix} x_2 & \dfrac{-x_1 + \varepsilon x_4^2 \sin x_3}{1 - \varepsilon^2 \cos^2 x_3} & x_4 & \dfrac{\varepsilon \cos x_3 (x_1 - \varepsilon x_4^2 \sin x_3)}{1 - \varepsilon^2 \cos^2 x_3} \end{bmatrix}^T \\ b(x) &= \begin{bmatrix} 0 & \dfrac{-\varepsilon \cos x_3}{1 - \varepsilon^2 \cos^2 x_3} & 0 & \dfrac{1}{1 - \varepsilon^2 \cos^2 x_3} \end{bmatrix}^T \\ c(x) &= \begin{bmatrix} 0 & \dfrac{1}{1 - \varepsilon^2 \cos^2 x_3} & 0 & \dfrac{-\varepsilon \cos x_3}{1 - \varepsilon^2 \cos^2 x_3} \end{bmatrix}^T \end{aligned} \tag{28}$$

and $\varepsilon = 0.2$. The initial state is given as $x(0) = [1, 1, 1, 1]^T$. The performance index function is defined by (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, in this paper, we choose a three-layer neural network as the critic network with the structure of 4-8-1. The structures of the u and w for the upper performance index function are 4-8-1 and 5-8-1; while 5-8-1 and 4-8-1 for the lower performance index function. The initial weights are all randomly chosen in $[-0.1, 0.1]$. Then, for each i, the critic network and the action network 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 performance index function is shown in Fig. 1(a). We can see that the saddle point of the game exists. Then, we apply the controller to the benchmark system and run for $T_f = 60$ s. The optimal control trajectories are shown in Fig. 1(b). The corresponding state trajectories are shown in Fig. 1(c) and (d), respectively.

Remark 4. The simulation results illustrate the effectiveness of the proposed iterative ADP algorithm. If the saddle point exists, the iterative control pairs $(\overline{u}^{(i)}, \overline{w}^{(i)})$ and $(\underline{u}^{i}, \underline{w}^{(i)})$ make the iterative performance index functions reach the saddle point while the existence conditions of the saddle point are avoided.

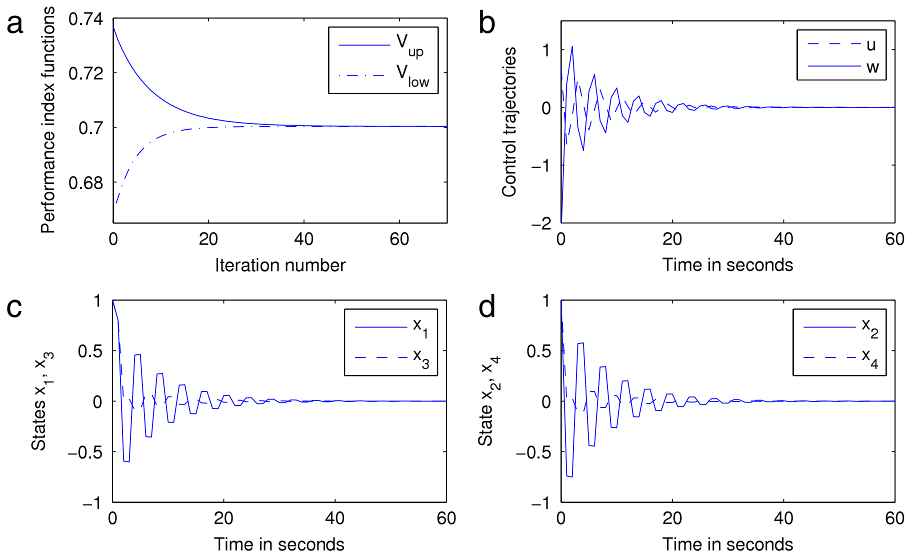


Fig. 1. Simulation results for Example 1. (a) Trajectories of the upper and lower performance index functions. (b) Trajectories of the controls. (c) Trajectories of state x_1 and x_3. (d) Trajectories of state x_2 and x_4.

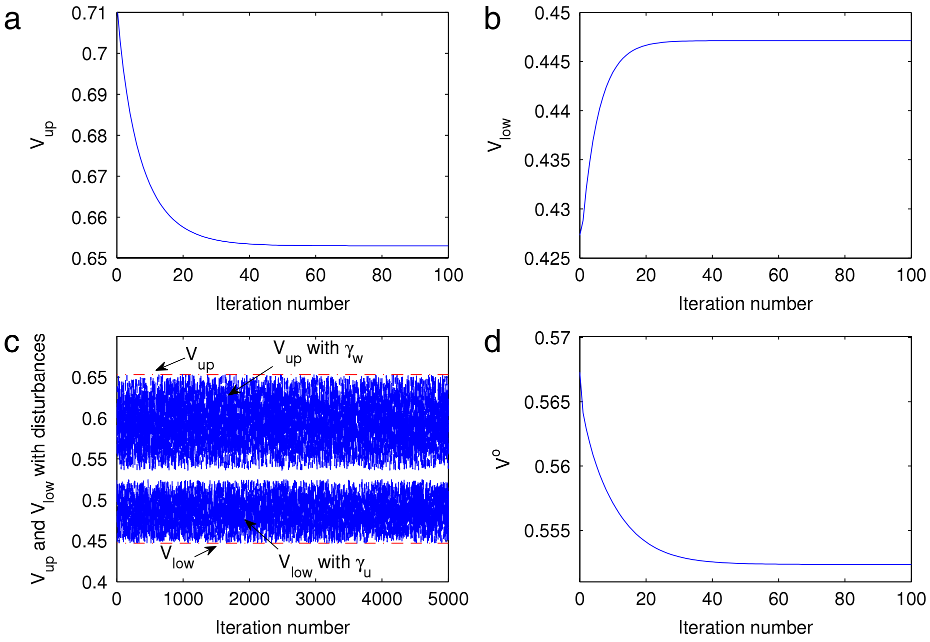


Fig. 2. Performance index function trajectories. (a) Trajectory of the upper performance index function. (b) Trajectory of the lower performance index function. (c) Performance index functions with disturbances. (d) Trajectory of the mixed optimal performance index function.

Example 2. In this example, we just change the utility function to $l(x, u, w) = x_1^2 + 0.1x_2^2 + 0.1x_3^2 + 0.1x_4^2 + \|u\|^2 - \gamma^2\|w\|^2 - 0.1uw + 0.1x^T u + 0.1x^T w$ and all other conditions are the same as the ones in Example 1. We can obtain $\overline{V}(x(0)) = 0.65297$ and $\underline{V}(x(0)) = 0.44713$, with trajectories shown in Fig. 2(a) and (b), respectively. Obviously, the saddle point does not exist. Thus, the method in Abu-Khalaf et al. (2008) is invalid. Using the mixed trajectory method proposed in this paper, we choose the Gaussian noise $\gamma_u(0, 0.05^2)$ and $\gamma_w(0, 0.05^2)$. Let $N = 5000$ times. The performance index function trajectories are shown in Fig. 2(c). Then, we obtain the value of the mixed optimal performance index function $V^o(x(0)) = 0.55235$ and then $\alpha = 0.5112$. Regulate the control w and obtain the trajectory of mixed optimal performance index function displayed in Fig. 2(d). The state trajectories are shown in Fig. 3(a) and (b), respectively. The corresponding control trajectories are shown in Fig. 3(c) and (d), respectively.

5. Conclusion

In this paper, an iterative ADP algorithm is developed that is effective for both the situations that the saddle point exists or does not exist. For the situation that the saddle point exists, the contribution of the algorithm is that existence conditions of the saddle point are avoided and the saddle point can directly be obtained. As there is no ADP method for the situation that the saddle point does not exist, we emphasize that for the first time the

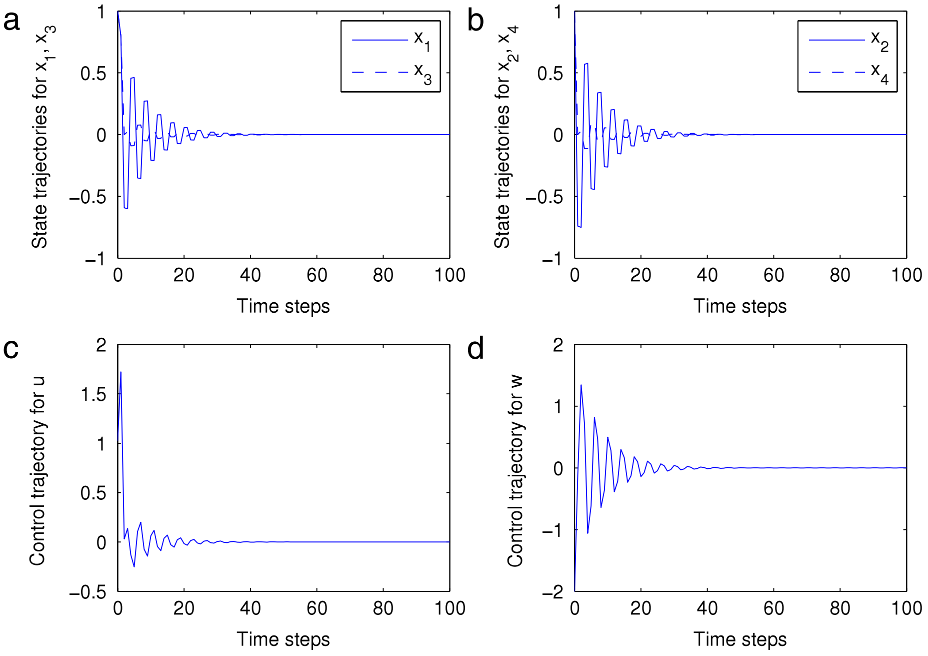


Fig. 3. State and control trajectories. (a) Trajectories of states x_1 and x_3. (b) Trajectories of states x_2 and x_4. (c) Trajectory of control u. (d) Trajectory of control w.

mixed optimal solution is obtained under the deterministic mixed optimal control pairs. The stability convergence properties are also guaranteed. This is another contribution that we emphasize.

References

Abu-Khalaf, M., Lewis, F. L., & Huang, J. (2006). Policy iterations on the Hamilton–Jacobi–Isaacs equation for state feedback control with input saturation. *IEEE Transactions on Automatic Control*, *51*(12), 1989–1995.

Abu-Khalaf, M., Lewis, F. L., & Huang, J. (2008). Neurodynamic programming and zero-sum games for constrained control systems. *IEEE Transactions on Neural Networks*, *19*(7), 1243–1252.

Al-Tamimi, A., Lewis, F. L., & Abu-Khalaf, M. (2007). Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control. *Automatica*, *43*(3), 473–481.

Bardi, M., & Capuzzo-Dolcetta, I. (1997). *Optimal control and viscosity solutions of Hamilton–Jacobi–Bellman equations*. Boston: Birkhäuser Press.

Basar, T., & Bernhard, P. (1995). *H∞ optimal control and related minimax design problems*. Boston: Birkhäuser Press.

Basar, T., & Olsder, G. J. (1982). *Dynamic noncooperative game theory*. New York: Academic Press.

Bianchini, G., Genesio, R., Parenti, A., & Tesi, A. (2004). Global H_∞ controllers for a class of nonlinear systems. *IEEE Transactions on Automatic Control*, *49*(2), 244–249.

Chang, H. S., Hu, J., Fu, M. C., & Marcus, S. I. (2010). Adaptive adversarial multi-armed bandit approach to two-person zero-sum Markov games. *IEEE Transactions on Automatic Control*, *55*(2), 463–468.

Cloutier, J. R. (1997). State-dependent Riccati equation techniques: an overview. In *Proceeding of the American control conference. Albuquerque, NM. June* (pp. 932–936).

Engwerda, J. (2008). Uniqueness conditions for the affine open-loop linear quadratic differential game. *Automatica*, *44*(2), 504–511.

Jain, R., & Watrous, J. (2009). Parallel approximation of non-interactive zero-sum quantum games. In *Proceeding of 24th annual IEEE conference on computational complexity. Paris, France. July* (pp. 243–253).

Jiang, D. (2009). Strictly pure Nash equilibria and expected equilibria in a symmetrical 0-1 game and its dual game. *ICIC Express Letters*, *3*(3), 295–300.

Jimenez-Lizarraga, M., Basin, M., & Alcorta-Garcia, A. M. (2009). Equilibrium in linear quadratic stochastic games with unknown parameters. *ICIC Express Letters*, *3*(2), 107–114.

Liao, X., Wang, L., & Yu, P. (2007). *Stability of dynamical systems*. Amsterdam, The Netherland: Elsevier Press.

Owen, G. (1982). *Game theory*. New York: Acadamic Press.

Shi, P. (2002). Limited Hamilton–Jacobi–Isaacs equations for singularly perturbed zero-sum dynamic (discrete time) games. *SIAM Journal on Control and Optimization*, *41*(3), 826–850.

Tsiotras, P., Corless, M., & Rotea, M. (1998). An L_2 disturbance attenuations solution to the nonlinear benchmark problem. *International Journal of Robust and Nonlinear Control*, *8*(2), 311–330.

Wang, X. (2008). Numerical solution of optimal control for scaled systems by hybrid functions. *International Journal of Innovative Computing, Information and Control*, *4*(4), 849–856.

Wang, F. Y., Zhang, H., & Liu, D. (2009). Adaptive dynamic programming: an introduction. *IEEE Computational Intelligence Magazine*, *4*(2), 39–47.

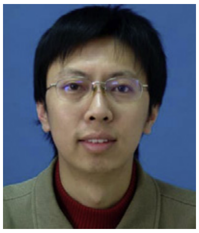
Wei, Q. L., Zhang, H. G., & Dai, J. (2009). Model-free multiobjective approximate dynamic programming for discrete-time nonlinear systems with general performance index functions. *Neurocomputing*, *72*(7–9), 1839–1848.

Zhang, H. G., Wei, Q. L., & Luo, Y. H. (2008). A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *38*(4), 937–942.

Huaguang Zhang received the B.S. and M.S. degrees in control engineering from Northeastern Electric Power University, Jilin, China, in 1982 and 1985, respectively, and the Ph.D. degree in thermal power engineering and automation from Southeast University, Nanjing, China, in 1991.

Dr. Zhang joined the Department of Automatic Control, Northeastern University, Shenyang, China, in 1992, as a Postdoctoral Fellow. Since 1994, he has been a Professor and the Head of the Electric Automation Institute, Northeastern University. He has authored and coauthored over 200 journal and conference papers and four monographs. He holds nine patents. His main research interests are neural-network-based control, fuzzy control, chaos control, nonlinear control, signal processing, and their industrial applications.

Dr. Zhang is an Associate Editor of Automatica and Neurocomputing. He was Program Chair for the 2007 International Symposium on Neural Networks (Nanjing, China). He is currently an Associate Editor of the IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics and the IEEE Transactions on Fuzzy Systems. He is Program Chair for the 2009 IEEE International Conference on Automation and Logistics (Shengyang, China). He was awarded the "Excellent Youth Science Foundation Award" by the National Natural Science Foundation of China in 2003, he was named the Changjiang Scholar by China Education Ministry in 2005, and he received the National Science and Technology Invention Award (Second Grade) from the Chinese Government in 2007, among several other awards.



Qinglai Wei received the B.S. degree in automation, the M.S. degree in control theory and control engineering, and the Ph.D. degree in control theory and control engineering, from the Northeastern University, Shenyang, China, in 2002, 2005, and 2008, respectively. He is currently a postdoctoral fellow with the Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include neural-networks-based control, nonlinear control, adaptive dynamic programming, and their industrial applications.

Derong Liu received the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, in 1994.

Dr. Liu was a Staff Fellow with General Motors Research and Development Center, Warren, MI, from 1993 to 1995. He was an Assistant Professor in the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, from 1995 to 1999. He joined the University of Illinois at Chicago in 1999, where he became a Full Professor of Electrical and Computer Engineering and of Computer Science in 2006. He was selected for the "100 Talents Program" by the Chinese Academy of Sciences in 2008.

Dr. Liu was an Associate Editor of the IEEE Transactions on Circuits and Systems-Part I: Fundamental Theory and Applications (1997–1999), the IEEE Transactions on Signal Processing (2001–2003), the IEEE Transactions on Neural Networks (2004–2009), the IEEE Computational Intelligence Magazine (2006–2009), and the IEEE Circuits and Systems Magazine (2008–2009), and the Letters Editor of the IEEE Transactions on Neural Networks (2006–2008). Currently, he is the Editor-in-Chief of the IEEE Transactions on Neural Networks and an Associate Editor of the IEEE Transactions on Control Systems Technology. He received the Michael J. Birck Fellowship from the University of Notre Dame (1990), the Harvey N. Davis Distinguished Teaching Award from Stevens Institute of Technology (1997), the Faculty Early Career Development (CAREER) Award from the National Science Foundation (1999), the University Scholar Award from University of Illinois (2006), and the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China (2008).