Integral Reinforcement Learning-Based H_∞ Tracking Control for Uncertain Linear Systems and Its Application

Rongsheng Xia(✉), Jiacheng Wu, and Hao Shen

College of Electrical and Information Engineering, Anhui University of Technology,
Ma'anshan 243032, China
xrsh12ujs@126.com

Abstract. In this paper, a robust optimal tracking strategy is presented for linear system with systems uncertainty and bounded disturbance. Firstly, an integral sliding mode control policy is designed to guarantee system trajectories tend to a defined sliding mode surface and the influence of system uncertainty is eliminated. Then the robust tracking control problem of original system is transformed into the H_∞ control problem of an auxiliary error system. Furthermore, an off-policy integral reinforcement learning (IRL) algorithm based H_∞ controller is designed, where the optimal tracking performance is guaranteed under the adverse effect of external disturbance. Finally, simulation test for near space vehicle (NSV) attitude model is introduced to verify the effectiveness of the proposed strategy.

Keywords: Off-policy IRL · Integral sliding mode control · H_∞ control · Zero-sum game theory

1 Introduction

Nowadays, the robust control method has received considerable attention from industrial and academic areas [1]. As far as we know, there are many effective methods to deal with the uncertainty. Such as disturbance observer-based (DO) control [2] and integral sliding mode control (ISMC) [3]. Compared to DO method, ISMC method can deal with the system uncertainty which only requires to be bounded. In [4], the authors investigated ISMC controller design issue for fuzzy semi-Markov systems. In [5], a robust fault-tolerant controller was designed for robot manipulators by using ISMC. In addition, for the purposed of improving control performance, optimal control theory can be widely used [6,7]. In [6], a novel tracking strategy using adaptive dynamic programming (ADP) algorithm was proposed for linear system with unknown dynamics. In [7], a novel value iteration based algorithm was proposed to solve the H_∞ control of linear system. The core mission of optimal control problem for linear system is to solve the algebraic Riccati equation (ARE), and reinforcement learning

Z. Ren et al. (Eds.): *Proceedings of 2021 5th Chinese Conference on Swarm Intelligence and Cooperative Control*, LNEE 934, pp. 1017–1025, 2023.
https://doi.org/10.1007/978-981-19-3998-3_97

(RL) technique can effectively handle this issue [8]. In [9], a novel RL scheme based on incremental learning approach was proposed for continuous-time linear system. In order to obviate the requirement of system dynamics, integral RL (IRL) method was proposed [10]. For linear system with input delay, an IRL-based model free optimal control method was proposed, and only the input and output of system datas were used [11].

Inspired by the above content, in this paper, a composite H_∞tracking control scheme is designed for continuous-time linear system with system uncertainty and bounded disturbance by using ISMC and off-policy IRL-based control methods. The sliding mode controller is designed to eliminate the effect of unknown uncertainty. The developed IRL control method is used to obtain the optimal tracking performance under the adverse effect of external disturbance. Furthermore, we introduce a NSV attitude model to show the effectiveness of the proposed control scheme.

2 Problem Description

In this paper, we consider the following uncertain system:

$$\left\{ \begin{array}{l} \dot{x}\left(t\right) = Ax\left(t\right) + Bu\left(t\right) + E\varpi(x) + D\varsigma(t) \\ y\left(t\right) = Cx\left(t\right) \end{array} \right. \tag{1}$$

where $x(t) = [x_1(t), \cdots, x_n(t)]^T \in \Re^n$ denotes the system state, $y\left(t\right) \in \Re^p$, $\varpi(x) \in \Re^v$ and $\varsigma(t) \in \Re^q$ represent system output, unknown system uncertainty and external disturbance, respectively. $A \in \Re^{n\times n}$, $B \in \Re^{n\times m}$, $C \in \Re^{p\times n}$, $E \in \Re^{n\times v}$ and $D \in \Re^{n\times q}$ are known system matrices. The external disturbance is assumed to belong to $L_2\left[0, \infty\right)$. The system uncertain $\varpi(x)$ is bounded and satisfies $\|\varpi(x)\| \le \varpi_m$.

The desired reference trajectory is generated by

$$\left\{ \begin{array}{l} \dot{x}_r\left(t\right) = A_r x_r\left(t\right) \\ y_r\left(t\right) = C_r x_r\left(t\right) \end{array} \right. \tag{2}$$

where $x_r\left(t\right) \in \Re^{n_r}$ and $y_r\left(t\right) \in \Re^p$ are system state and output of reference trajectory system. A_r and C_r are constant matrices. Furthermore, the following tracking error can be defined as $e(t) = y(t) - y_r(t)$

Here, we introduce a new error variable as

$$z\left(t\right) = x\left(t\right) - Gx_r\left(t\right) \tag{3}$$

where $z\left(t\right) \in \Re^n$, $G \in \Re^{n\times n_r}$ is the constant matrix satisfying $AG + BH = GA_r$ and $CG = C_r$. $H \in \Re^{m\times n_r}$ is the constant matrix, which is employed to model match. Furthermore, one can deduce that $e(t) = Cz(t)$.

Then, combining (1), (2) and (3), we can obtain

$$\dot{z}\left(t\right) = Ax\left(t\right) + Bu\left(t\right) + E\varpi(x) + D\varsigma(t) - GA_r x_r\left(t\right) \tag{4}$$

The control input is designed as $u\left(t\right) = u_a\left(t\right) + u_o\left(t\right)$, where $u_a\left(t\right)$ is an integral sliding mode control policy to eliminate the influence of the system uncertainty, and $u_o\left(t\right)$ is an off-policy IRL-based H_∞ control policy to guarantee the optimal tracking performance.

3 Controller Design

In this section, we will present the porposed control method including ISMC and of-policy IRL-based H_∞ control design. Moreover, the structure of the proposed control method is shown in Fig. 1.

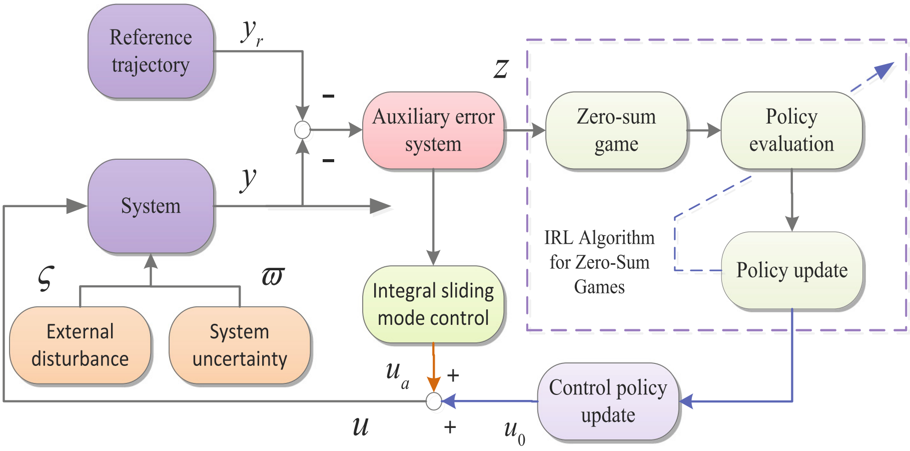


Fig. 1. Estimation results of the unknown disturbance D

3.1 Integral Sliding Mode Control Design

In this paper, we select the following integral sliding mode surface

$$\mathcal{S}\left(z,t\right) = \mathit{\Gamma}[z(t) - z(0) - \int_0^t \left(Az + Bu_o + D\varsigma\right) \mathrm{d}\tau] \tag{5}$$

where $\mathit{\Gamma}$ is a positive matrix to be designed, which satisfies $\mathit{\Gamma}B$ is invertible. Furthermore, the integral sliding mode control policy can be designed as

$$u_a(t) = -\mathit{\Upsilon}\left(\mathit{\Gamma}B\right)^{-1} \mathrm{Sgn}\left(\mathcal{S}\right) - B^{-1}\left(AGx_r(t) - GA_r x_r(t)\right) \tag{6}$$

where $\mathrm{Sgn}\left(\mathcal{S}\right) = \left[\mathrm{sgn}\left(\mathcal{S}_1\right) \ldots \mathrm{sgn}\left(\mathcal{S}_n\right)\right]^T$, and $\mathrm{sgn}(\cdot)$ is a sign function. $\mathit{\Upsilon}$ is positive matrix to be designed.

Theorem 1. *Considering system (4), the integral sliding mode surface and the integral sliding mode control policy are designed as (5)–(6), respectively. Then, integral sliding surface is uniformly asymptotically stable by selecting suitable* $\mathit{\Upsilon}$ *and* $\mathit{\Gamma}$*.*

Proof. The Lyapunov function is selected as follows

$$V(t) = \frac{1}{2}\mathcal{S}^T\mathcal{S} \tag{7}$$

Taking derivative of $V\left(t\right)$ with respect to t, one can obtain that

$$\begin{aligned}\dot{V}\left(t\right) &= \mathcal{S}^T\mathit{\Gamma}[Ax(t) + Bu_a(t) + E\varpi(x) - GA_r x_r(t) - Az(t)]\\ &= \mathcal{S}^T\mathit{\Gamma}[-\mathit{\Upsilon}\mathit{\Gamma}^{-1}\mathrm{Sgn}(\mathcal{S}) + E\varpi(x)]\\ &= -\mathit{\Upsilon}\mathcal{S}^T\mathrm{Sgn}(\mathcal{S}) + \mathcal{S}^T\mathit{\Gamma}E\varpi(x)\\ &\le -\lambda_{\min}\left(\mathit{\Upsilon}\right)\left\|\mathcal{S}\right\| + \mathit{\Gamma}E\varpi_m\left\|\mathcal{S}\right\|\\ &\le -(\lambda_{\min}\left(\mathit{\Upsilon}\right) - \mathit{\Gamma}E\varpi_m)\left\|\mathcal{S}\right\|\end{aligned} \tag{8}$$

By selecting suitable matrixes $\mathit{\Upsilon}$ and $\mathit{\Gamma}$ such that $\lambda_{\min}\left(\mathit{\Upsilon}\right) > \mathit{\Gamma}E\varpi_m$, then, we have $\dot{V}\left(t\right) < 0$, which means that sliding mode surface is uniformly asymptotically stable.

3.2 Off-Policy IRL-Based H_∞ Control Design

Consider the following auxiliary error system

$$\dot{z}\left(t\right) = Az\left(t\right) + Bu_o\left(t\right) + D\varsigma\left(t\right) \tag{9}$$

The corresponding infinite horizon performance index is

$$\mathcal{J}\left(z, u_o, \varsigma\right) = \int_0^\infty (z^TQz + u_o^TRu_o - \varphi^2\varsigma^T\varsigma)\mathrm{d}\tau \tag{10}$$

where $Q = Q^T \ge 0$, $R = R^T > 0$ denote the state and control performance weights, respectively. φ is a constant, which satisfies $\varphi \ge \varphi^*$, φ^* is the smallest L_2 gain. We consider $\varsigma\left(t\right)$ as opponent's policy. The aim is to find a control policy (u_0, ς) to make system (9) is stable and meets a H_∞ performance.

Furthermore, the H_∞ control issue is equivalent to following zero-sum game problem

$$\begin{aligned}\mathcal{V}^*\left(z\right) &= \min_{u_o}\max_{\varsigma}\mathcal{J}\left(z, u_o, \varsigma\right)\\ &= \min_{u}\max_{\varsigma}\int_0^\infty (z^TQz + u_o^TRu_o - \varphi^2\varsigma^T\varsigma)\mathrm{d}\tau\end{aligned} \tag{11}$$

where $\mathcal{V}^*\left(z_0\right)$ is the optimal value function. Control policy and disturbance policy are considered as two hostile players, where control policy desires to minimize the performance index while disturbance policy aims to damage it. Furthermore, we denote control policy $u_o(t) = -Kz(t)$ and disturbance policy $\varsigma(t) = K_w z(t)$, respectively. Then, the value function can be expressed as

$$\mathcal{V}\left(z\right) = z^T(t)Pz(t) \tag{12}$$

Moreover, we can obtain the following algebraic Riccati equation

$$A^T P^* + P^* A + Q - P^* B R^{-1} B^T P^* + \varphi^{-2} P^* D^T D P^* = 0 \tag{13}$$

the saddle point of zero-sum game is

$$\begin{aligned} u_o^*(t) &= -Kz(t) = -R^{-1} B^T P^* z(t) \\ \varsigma^*(t) &= K_w z(t) = \varphi^{-2} D^T P^* z(t) \end{aligned} \tag{14}$$

Then, system (9) can be rewritten as

$$\dot{z}\,(t) = \tilde{A} z(t) + B\,(u_0\,(t) + Kz(t)) + D\,(\varsigma\,(t) - K_w z(t)) \tag{15}$$

where $\tilde{A} = A - BK + DK_w$.

Furthermore, we can obtain that

$$\begin{aligned} z^T(t+T)P_i z(t+T) - z^T(t)P_i z(t) = &- \int_t^{t+T} z^T Q_i z \mathrm{d}\tau \\ &+ 2\int_t^{t+T} [(u_o + K_i z(t))^T R_i K_i z(t)] \mathrm{d}\tau \\ &-2\varphi^2 \int_t^{t+T} [(K_{wi} z(t) - \varsigma)^T K_i z(t)] \mathrm{d}\tau \end{aligned} \tag{16}$$

Then, the left-hand of (9) can be rewritten as

$$z^T(t+T)P_i z(t+T) - z^T(t)P_i z(t) = \tilde{P}_i^T [z^T(t+T) \otimes z^T(t+T) - z^T(t) \otimes z(t)] \tag{17}$$

where

$$\begin{aligned} \tilde{P}_i &= [P_{i11}, 2P_{i12}, \cdots, 2P_{i1n}, P_{i22}, 2P_{i23}, \cdots, P_{inn}]^T \\ z^T \otimes z^T &= \left[z_1^2, z_1 z_2, \cdots, z_1 z_n, z_2^2, z_2 z_3, \cdots, z_n\right]^T \end{aligned}$$

Similarly, we can deduce

$$\begin{aligned} z^T Q z &= (z^T \otimes z^T) vec(Q) \\ (u_o + K_i z)^T R K_i z &= [(z^T \otimes z^T)(I_n \otimes K_i^T R) \\ &\quad + (z^T \otimes u^T)(I_n \otimes R)] vec(K_i) \\ \varphi^2 (K_{wi} z(t) - \varsigma)^T K_i z(t) &= [(z^T \otimes z^T)(I_n \otimes \varphi^2 K_{wi}^T) \\ &\quad - (z^T \otimes \varsigma^T) \left(\varphi^2 I\right)] vec(K_{wi}) \end{aligned} \tag{18}$$

From (17) and (18), (16) can be represented as

$$\mathit{\Pi}_i \times \begin{bmatrix} \tilde{P}_i \\ vec\,(K_{i+1}) \\ vec\,(K_{wi+1}) \end{bmatrix} = \mathit{\Omega}_i$$

where $\varOmega_i = -\gamma_{zz} vec\left(Q\right)$ and

$$\begin{aligned}
\varPi_i &= \begin{bmatrix} \pounds_{zz} \\ -2[\gamma_{zz}(I_n \otimes K_i^T R) + \pi_{zu_0}(I_n \otimes R)] \\ 2[\gamma_{zz}(I_n \otimes K_{wi}^T \varphi^2) + \phi_{zz}(\varphi^2 I_n)] \end{bmatrix}^T \\
\pounds_{zz} &= \left[\tilde{z}\left(t_1\right) - \tilde{z}\left(t_0\right) \; \tilde{z}\left(t_2\right) - \tilde{z}\left(t_1\right) \cdots \tilde{z}\left(t_l\right) - \tilde{z}\left(t_{l-1}\right)\right]^T \\
\gamma_{zz} &= \left[\int_{t_0}^{t_1} z \otimes z \mathrm{d}\tau \; \int_{t_1}^{t_2} z \otimes z \mathrm{d}\tau \cdots \int_{t_{l-1}}^{t_l} z \otimes z \mathrm{d}\tau\right]^T \\
\pi_{zu_o} &= \left[\int_{t_0}^{t_1} z \otimes u_o \mathrm{d}\tau \; \int_{t_1}^{t_2} z \otimes u_o \mathrm{d}\tau \cdots \int_{t_{l-1}}^{t_l} z \otimes u_o \mathrm{d}\tau\right]^T \\
\phi_{z\varsigma} &= \left[\int_{t_0}^{t_1} z \otimes \varsigma \mathrm{d}\tau \; \int_{t_1}^{t_2} z \otimes \varsigma \mathrm{d}\tau \cdots \int_{t_{l-1}}^{t_l} z \otimes \varsigma \mathrm{d}\tau\right]^T
\end{aligned}$$

Furthermore, we have

$$\begin{bmatrix} \tilde{P}_i \\ vec\left(K_{i+1}\right) \\ vec\left(K_{wi+1}\right) \end{bmatrix} = \left(\varPi_i^T \varPi_i\right)^{-1} \varPi_i^T \varOmega_i$$

Then, the online implementation of off-policy IRL-based H_∞ control method is presented in Algorithm 1. Moreover, the stability analysis of the system (9) can be reference to [10].

Algorithm 1: Off-Policy IRL-Based Control Algorithm.

```
Input: Measure $z\left(t\right), u_o\left(t\right)$ and $\varsigma\left(t\right)$
Step I (Data collection): Collect data of $z\left(t\right), u_0\left(t\right)$ and $\varsigma\left(t\right)$ for sufficiently
  large uniformly sampled time instants, and construct the following matrices.
where $\tilde{z} \triangleq z^T \otimes z^T \triangleq \left[z_1^2 \; z_1 z_2 \cdots z_1 z_n \; z_2^2 \cdots z_n^2\right]^T$
Step II (Gain update): Solve $K$,$K_w$ and $P$ iteratively from the following
  equality
    $\begin{bmatrix} \pounds_{zz} \\ -2[\gamma_{zz}(I_n \otimes K_i^T R) + \pi_{zu_0}(I_n \otimes R)] \\ 2[\gamma_{zz}(I_n \otimes K_{wi}^T \varphi) + \phi_{zz}(\varphi^2 I_n)] \end{bmatrix}^T \times \begin{bmatrix} \tilde{P}_i \\ vec\left(K_{i+1}\right) \\ vec\left(K_{wi+1}\right) \end{bmatrix}$
    $= -\gamma_{zz} vec\left(Q_i\right)$
  where $vev(\cdot)$ is a vectorization map from a matrix into a column vector.
Step III (Computation terminated): Stop if $\|P_{i+1} - P_i\| \leq \varepsilon$, where $\varepsilon$ is a
  given constant. Otherwise, set $P_i \leftarrow P_{i+1}$, $K_i \leftarrow K_{i+1}$, $K_{wi} \leftarrow K_{wi+1}$ and
  repeat Step II.
Step IV (Policy update): If $K$, $K_w$ and $P$ converge, apply control policy
  $u_o = -Kz$ to the system.
```

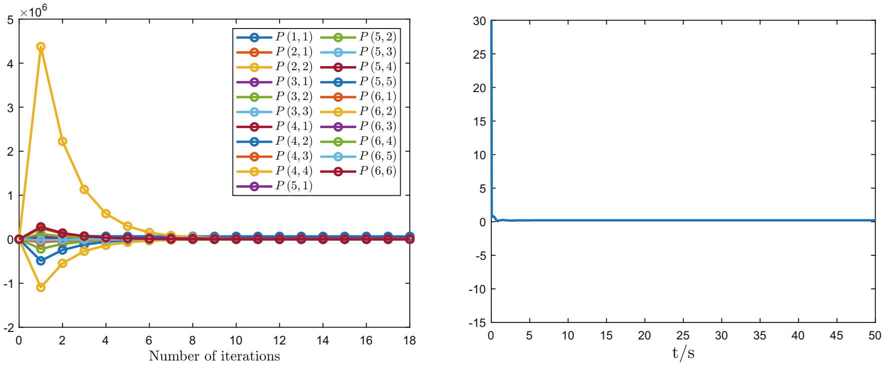
4 Simulation Results

In this section, simulation studies are employed to verified the effectiveness of the proposed method. The nonlinear attitude mode of NSV is linearized at equilibrium point $x_0 = [-0.0005, 0.0001, 0.2, 0, -0.1872, 0.0007]^T$, such the linear attitude mode of NSV is obtained.

$$\dot{x} = Ax + Bu$$

where $x = [\alpha, \beta, \mu, p, q, r]^T$ is system state vector, which are attitude angles and angle rates. $u = [\delta_e, \delta_a, \delta_r, \delta_x, \delta_y, \delta_z]^T$ denotes control input vector. The specific information of NSV mode and matrices A, B can reference to [12]. And

$$D = E = \begin{bmatrix} 0.1 \; 0.4 \; 0.1 \; 0.2 \; 0.1 \; 0.2 \end{bmatrix}^T$$
$$\varpi(x) = 0.01\sin(x_1) + 0.05x_2^2\cos(x_3)\varsigma(t) = 0.01e^{-0.1t}\sin(0.1t)$$



(a) The parameters of matrix P.

(b) The sliding surface function.

Fig. 2. Convergence of matrix P and sliding surface function.

The reference attitude angles are selected as

$$\dot{x}_r = 0, \alpha_r = 0, \beta_r = -0.8, \gamma_r = 0.65$$

For algorithm 1, the parameters are chosen as follows: $Q = 10^4 I$, $R = I$. From $t = 0\,\mathrm{s}$ to $t = 2\,\mathrm{s}$, the following exploration noise is employed as system input

$$\bar{e} = 100\sum_{c=1}^{100}\sin(w_c t)$$

where $c = 1, ..., 100$, and w_c are selected from $[-500, 500]$. Moreover, the weighting matrices are $\mathit{\Pi} = I, Q = 10^4 I$, $R = I$, and $\varphi = 1.5$, $\mathit{\Gamma} = 2.2$, $\mathit{\Upsilon} = 0.01$. Furthermore, by using Algorithm 1, the control gain K can be obtained. The convergence process of P matrix element and sliding surface function are shown

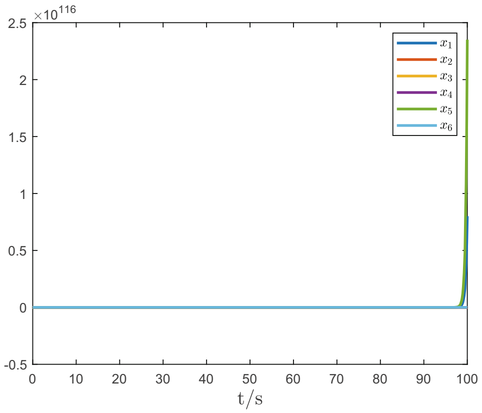


Fig. 3. The state responses of the open-closed system.

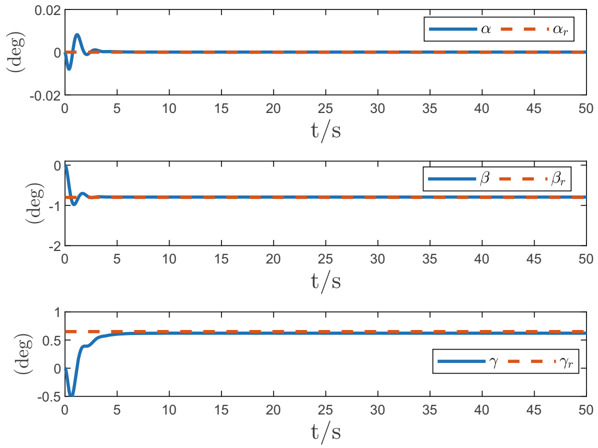


Fig. 4. The responses of the attitude angles.

in Fig. 2. From Fig. 3, it can be observed that system is unstable without the control input. Then, it can be seen from Fig. 4 that actual angles can well track the desired signals in a short time, which means that the proposed control method is effective.

$$K = 10^2 \times \begin{bmatrix} 3.7617 & -0.4746 & -0.6895 & -0.7123 & 0.9036 & -0.5118 \\ 5.1117 & 1.2282 & 1.2564 & 1.2707 & 2.5080 & 1.007 \\ -2.5063 & -2.0072 & -2.7337 & -2.7688 & -2.6176 & -2.8739 \\ -0.4758 & -0.5955 & -0.6830 & -0.6959 & -0.5633 & -0.5337 \\ -0.0510 & -0.0437 & -0.0583 & -0.0592 & -0.054305 & -0.0584 \\ 1.1612 & 0.0907 & 0.0640 & 0.0627 & 0.4401 & 0.0545 \end{bmatrix}$$

5 Conclusions

In this paper, a composite H_∞tracking control scheme is designed for continuous-time linear systems with system uncertainty and bounded disturbance. Firstly, the integral sliding mode controller has been applied to deal with unknown system uncertainty. In addition, an off-policy IRL has been provided for solving the two-player zero-sum game problem of H_∞ control. Finally, the simulation results for NSV attitude control show the effectiveness of the proposed method. In our future work, we will extend the results to nonzero-sum games for practical system.

References

1. Zhao, Z., He, X., Ren, Z., Wen, G.: Boundary adaptive robust control of a flexible riser system with input nonlinearities. IEEE Trans. Syst. Man Cybern. Syst. **49**(10), 1971–1980 (2018)
2. Chen, W.H., Ding, K., Lu, X.: Disturbance-observer-based control design for a class of uncertain systems with intermittent measurement. J. Franklin Inst. **354**(13), 5266–5279 (2017)
3. Pan, Y., Yang, C., Pan, L., Yu, H.: Integral sliding mode control: performance, modification, and improvement. IEEE Trans. Industr. Inf. **14**(7), 3087–3096 (2017)
4. Jiang, B., Karimi, H.R., Kao, Y., Gao, C.: A novel robust fuzzy integral sliding mode control for nonlinear semi-Markovian jump T-S fuzzy systems. IEEE Trans. Fuzzy Syst. **26**(6), 3594–3604 (2018)
5. Van, M., Ge, S.S.: Adaptive fuzzy integral sliding-mode control for robust fault-tolerant control of robot manipulators with disturbance observer. IEEE Trans. Fuzzy Syst. **29**(5), 1284–1296 (2020)
6. Qin, C.B., Zhang, H.G., Luo, Y.H.: Online optimal tracking control of continuous-time linear systems with unknown dynamics by using adaptive dynamic programming. Int. J. Control **87**(5), 1000–1009 (2014)
7. Jiang, H.Y., Zhou, B., Liu, G.P.: H_∞ optimal control of unknown linear systems by adaptive dynamic programming with applications to time-delay systems. Int. J. Robust Nonlinear Control **31**(12), 5602–5617 (2021)
8. Wen, Y., Si, J., Brandt, A., Gao, X., Huang, H.: Online reinforcement learning control for the personalization of a robotic knee prosthesis. IEEE Trans. Cybern. **50**(6), 2346–2356 (2019)
9. Bian, T., Jiang, Z.: Reinforcement learning for linear continuous-time systems: an incremental learning approach. IEEE/CAA J. Automatica Sinica **6**(2), 433–440 (2019)
10. Moghadam, R., Lewis, F.L.: Output-feedback H_∞ quadratic tracking control of linear systems using reinforcement learning. Int. J. Adapt. Control Signal Process. **33**(2), 300–314 (2017)
11. Wang, G., Luo, B., Xue, S.: Integral reinforcement learning based optimal feedback control for nonlinear continuous time systems with input delay. Neurocomputing **460**, 31–38 (2021)
12. Yang, Q.: Robust control for near space vehicle with input saturation. Nanjing University of Aeronautics and Astronautics (2017)