Nonfragile Output Feedback Tracking Control for Markov Jump Fuzzy Systems Based on Integral Reinforcement Learning Scheme

Jing Wang, Jiacheng Wu, Jinde Cao, *Fellow, IEEE*, Mohammed Chadli, *Senior Member, IEEE*, and Hao Shen, *Member, IEEE*

***Abstract*—In this article, a novel integral reinforcement learning (RL)-based nonfragile output feedback tracking control algorithm is proposed for uncertain Markov jump nonlinear systems presented by the Takagi–Sugeno fuzzy model. The problem of nonfragile control is converted into solving the zero-sum games, where the control input and uncertain disturbance input can be regarded as two rival players. Based on the RL architecture, an offline parallel output feedback tracking learning algorithm is first designed to solve fuzzy stochastic coupled algebraic Riccati equations for Markov jump fuzzy systems. Furthermore, to overcome the requirement of a precise system information and transition probability, an online parallel integral RL-based algorithm is designed. Besides, the tracking object is achieved and the stochastically asymptotic stability, and expected $\mathcal{H}_\infty$ performance for considered systems is ensured via the Lyapunov stability theory and stochastic analysis method. Furthermore, the effectiveness of the proposed control algorithm is verified by a robot arm system.**

***Index Terms*—Fuzzy tracking control, integral reinforcement learning (RL), Markov jump fuzzy systems (MJFSs), nonfragile control.**

I. Introduction

MARKOV jump systems (MJSs) are a kind of random jumping system composed of many interconnected subsystems, in which the Markov process characterizes the jumping among the subsystems. In the last few decades, since MJSs could be capable of modeling many practical physical systems, where their structures or parameters change randomly, such systems have attracted lots of attention [1], [2], [3], [4], [5], [6], [7], [8] . On the other front, it should be mentioned that nonlinearities exist in many physical systems. The Takagi–Sugeno (T–S) fuzzy model is an effective method to approximate these nonlinearities [9], [10], [11]. Combining MJSs and T–S fuzzy model, a broader type of system, namely, Markov jump fuzzy systems (MJFSs), have received significantly increasing attention and numerous important results have emerged [12], [13], [14], [15], [16]. To be specific, by virtue of the linear matrix inequality technique, Wu *et al.* [17] investigated the problem of nonfragile guaranteed cost controller design for MJFSs with time-varying delays. The problems of $\mathcal{H}_\infty$ control and filtering for Markov jump nonlinear singularly perturbed systems were studied in [18]. In [19], the finite-time asynchronous filtering issue for MJFSs with incomplete transition rate was addressed.

Besides, in the era of the fourth industrial revolution, artificial intelligent methods have attracted lots of attention and gradually becomes one key focus in the control field [20], [21], [22], [23], [24]. Through interaction with the environment, reinforcement learning (RL) methods can remove the curse of dimensionality effectively in the traditional dynamic programming method [25], [26], [27]. On the other hand, the RL method has been successfully employed to solve the optimal control policy issue [28], in which optimal solution can be converted into algebraic Riccati equation or Hamilton–Jacobi–Bellman equation. Furthermore, the optimal tracking control has become an important problem in robot engineering and aircraft transportation field [29], [30], [31], where adaptive dynamic programming and the RL method were widely introduced to approximate the optimal policy. Mohammadi *et al.* [32] investigated the optimal tracking controller design problem for nonlinear continuous-time systems, where the integral RL (IRL) method was introduced to determine the control input. Optimal tracking control of switched systems was studied in [33]. By virtue of the IRL technology, Zhang *et al.* [34] developed a parallel tracking control algorithm for fuzzy interconnected systems. Furthermore, the above-mentioned works are both based on the state feedback control. However, it is not easy to measure the full states of the systems in some practical situations. The static output feedback control strategy allows flexibility and simplicity of implementation, which is extremely important in

Manuscript received 16 May 2022; revised 30 June 2022; accepted 29 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62273006, Grant 62173001, Grant 61873002; in part by the Natural Science Foundation for Excellent Young Scholars of Anhui Province under Grant 2108085Y21; and in part by the Major Technologies Research and Development Special Program of Anhui Province under Grant 202003a05020001. This article was recommended by Associate Editor W. Yu. *(Corresponding author: Hao Shen.)*

Jing Wang, Jiacheng Wu, and Hao Shen are with the Anhui Province Key Laboratory of Special Heavy Load Robot and School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan 243032, China (e-mail: jingwang08@126.com; jc980507@163.com; haoshen10@gmail.com).

Jinde Cao is with the School of Mathematics, Southeast University, Nanjing 210096, China, and also with the Yonsei Frontier Lab, Yonsei University, Seoul 03722, South Korea (e-mail: jdcao@seu.edu.cn).

Mohammed Chadli is with the University Paris-Saclay, Univ Evry, IBISC, 91020 Evry, France (e-mail: mchadli20@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2022.3203795.

Digital Object Identifier 10.1109/TCYB.2022.3203795

current controller design applications. Hence, it is of great significance to study output feedback tracking control design in the framework of RL.

What is more, inaccuracies or uncertainties may exist in many practical applications because of the effect of the highly complex working environment of control systems [35], [36], [37], [38], [39], [40]. A smaller controller gain fluctuations may cause the control system to be unstable or fail to guarantee the expected performance [41], [42], [43]. Therefore, it is necessary to investigate the nonfragile control problem. In [44], a resilient controller was designed by solving zero-sum games, where cyber defenders and denial-of-service attackers were two rival players. Nevertheless, the research on these points is quietly few, let alone with the RL algorithm, which motivates our current investigation.

In summary, by employing the IRL technology and the strategy of output feedback, we devote ourselves to designing a nonfragile tracking controller for MJFSs based on the T–S fuzzy model. The main contributions are listed as follows.

1) For the first attempt, an offline parallel fuzzy nonfragile output feedback tracking learning control algorithm and an online IRL-based fuzzy nonfragile output feedback tracking control algorithm are designed for MJFSs presented by the T–S fuzzy model. Moreover, the system information can be partially unknown in the proposed online algorithm.
2) A nonfragile fuzzy output feedback tracking controller is developed by solving a class of zero-sum games, where the control policy and disturbance policy that have different impacts on the system performance are regarded as two rival players, which provides new insights to design the nonfragile tracking learning controller.
3) By designing two kinds of learning algorithms, a set of fuzzy stochastic coupled algebraic Riccati equations (FSCAREs) is solved. Moreover, the stochastic stability of the presented online IRL-based algorithm is verified by the Lyapunov stability theory. Furthermore, the applicability to the real engineering systems of the proposed method is proven by a robot arm system model.

The remainder of this article is organized as follows. Based on the T–S fuzzy model, the system and controller model are elaborated in Section II. In Section III, an offline learning algorithm and an IRL tracking algorithm are presented for MJFSs. A robot arm example is provided in Section IV to demonstrate the effectiveness of the proposed method. In Section V, we conclude this article. Moreover, the structure of the nonfragile output-feedback tracking learning control scheme is presented in Fig. 1.

Notations: The notations used in this article are standard. $\mathbb{E}\{\cdot\}$ denotes mathematical expectation. $\mathcal{M}^{\mathrm{T}}$ means the transpose of the matrix $\mathcal{M}$; $\mathcal{M} > 0$ represents that matrix $\mathcal{M}$ is positive definite. $\mathbf{R}$, $\mathbf{R}^n$, and $\mathbf{R}^{n\times m}$ denote the set of real numbers, the n-dimensional Euclidean space, and the set of all real matrices, respectively. Symbol $\exp(\cdot)$ represents a function as $\exp([\varsigma_1, \ldots, \varsigma_d]) = [\exp(\varsigma_1), \ldots, \exp(\varsigma_d)]$.

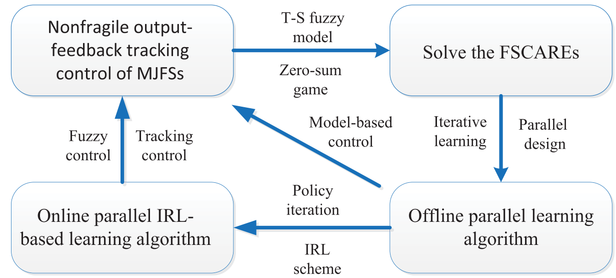

Fig. 1. Structure of the nonfragile output-feedback tracking learning control scheme.

II. Preliminaries

A. System Description

Consider a class of continuous-time MJFSs with uncertainties.

Plant Rule α: **IF** $\theta_1(\ell)$ is $R_{\alpha 1}$, $\theta_2(\ell)$ is $R_{\alpha 2}, \ldots, \theta_g(\ell)$ is $R_{\alpha g}$ **THEN**

$$\begin{aligned}\dot{x}(\ell) &= \big(A_{\alpha\eta(\ell)} + \Delta A_{\alpha\eta(\ell)}(\ell)\big)x(\ell)\\ &\quad + \big(B_{\alpha\eta(\ell)} + \Delta B_{\alpha\eta(\ell)}(\ell)\big)\mu(\ell)\\ y(\ell) &= C_{\alpha\eta(\ell)}x(\ell)\end{aligned} \tag{1}$$

where $\alpha \in \mathbb{M}\{1, 2, \ldots, m\}$, and m is the number of **IF-THEN** rules of system. $R_{\alpha a}$ with $a = 1, \ldots, g$ is the fuzzy set. $\theta(\ell) = [\theta_1(\ell), \ldots, \theta_g(\ell)]^{\mathrm{T}}$ is the premise variable. $x(\ell) \in \mathbf{R}^{n_x}$, $u(\ell) \in \mathbf{R}^{n_u}$, and $y(\ell) \in \mathbf{R}^{n_y}$ represent system state, control input, and system output, respectively. $A_{\alpha\eta(\ell)}$, $B_{\alpha\eta(\ell)}$, and $C_{\alpha\eta(\ell)}$ are system matrices. $\Delta A_{\alpha\eta(\ell)}$ and $\Delta B_{\alpha\eta(\ell)}$ denote time-varying uncertainties. $\{\eta(\ell)\}$ is a continuous-time Markov process, which takes values in a set $\mathbb{S} = \{1, 2, \ldots, S\}$ with the transition probability rate matrix $\Pi \triangleq [\pi_{ij}]_{S\times S}$ is given by

$$\Pr\{\eta(\ell+\Delta) = j|\eta(\ell) = i\} = \begin{cases}\pi_{ij}\Delta + o(\Delta), & i \neq j\\ 1 - \pi_{ii}\Delta + o(\Delta), & i = j\end{cases}$$

where $\Delta > 0$, $\lim_{\Delta\to 0} o(\Delta)\Delta^{-1} = 0$, for $i \neq j$, $\pi_{ij} \geq 0$ is the transition rate from mode i at time t to mode j at time $t + \Delta$ and $\pi_{ii} = -\sum_{j=1, j\neq i}^{S} \pi_{ij}$.

Fuzzy weighting functions are defined as

$$f_\alpha(\theta(\ell)) \triangleq \frac{\Pi_{z=1}^{g} R_{\alpha z}(\theta_z(\ell))}{\sum_{\alpha=1}^{m} \Pi_{z=1}^{g} R_{\alpha z}(\theta_z(\ell))}$$

where $R_{\alpha z}(\theta_z(\ell))$ is the membership degree of $\theta_z(\ell)$ in the fuzzy rule α. It follows that:

$$f_\alpha(\theta(\ell)) \geq 0,\ \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell)) = 1.$$

To simplify the notation, for $\forall\alpha \in \mathbb{M}$, $\eta(\ell) = i \in \mathbb{S}$, we denote $A_{\alpha i} \triangleq A_{\alpha\eta(\ell)}$, $B_{\alpha i} \triangleq B_{\alpha\eta(\ell)}$ and $C_{\alpha i} \triangleq C_{\alpha\eta(\ell)}$. Based on (1), the overall system dynamic can be obtained as follows:

$$\begin{aligned}\dot{x}(\ell) &= \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))[(A_{\alpha i} + \Delta A_{\alpha i}(\ell))x(\ell)\\ &\quad + (B_{\alpha i} + \Delta B_{\alpha i}(\ell))\mu(\ell)]\\ y(\ell) &= \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))C_{\alpha i}x(\ell)\end{aligned} \tag{2}$$

where $\Delta A_{\alpha i}(\ell)$ and $\Delta B_{\alpha i}(\ell)$ are considered as the following form:

$$[\,\Delta A_{\alpha i}(\ell)\quad \Delta B_{\alpha i}(\ell)\,] = H_{\alpha i}T(\ell)[D_{1\alpha i}\quad D_{2\alpha i}]$$

where $H_{\alpha i}$, $D_{1\alpha i}$,and $D_{2\alpha i}$ are constant matrices. The unknown time-varying matrix $T(\ell)$ satisfies $T^{\mathrm{T}}(\ell)T(\ell) \le I$.

Define the desired reference trajectory as the following form:

$$\dot{x}_r(\ell) = Fx_r(\ell),\;\; y_r(\ell) = Gx_r(\ell) \tag{3}$$

where $x_r \in \mathbf{R}^{n_{x_r}}$ and $y_r \in \mathbf{R}^{n_{y_r}}$ are system state and system output of reference trajectory system, respectively. F and G are constant matrices.

Let $X(\ell) \triangleq [\,x^{\mathrm{T}}(\ell)\;\; x_r^{\mathrm{T}}(\ell)\,]^{\mathrm{T}} \in \mathbf{R}^n$, $n = n_x + n_{x_r}$, and $Y(\ell) \triangleq [\,y^{\mathrm{T}}(\ell)\;\; y_r^{\mathrm{T}}(\ell)\,]^{\mathrm{T}} \in \mathbf{R}^{n_y+n_{y_r}}$. From (2) and (3), we can deduce

$$\begin{aligned}\dot{X}(\ell) &= \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))\big\{\big(\mathbb{A}_{\alpha i} + \mathcal{I}\Delta A_{\alpha i}(\ell)\mathcal{I}^{\mathrm{T}}\big)X(\ell)\\ &\qquad + (\mathbb{B}_{\alpha i} + \mathcal{I}\Delta B_{\alpha i}(\ell))\mu(\ell)\big\}\\ Y(\ell) &= \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))\{\mathbb{C}_{\alpha i}X(\ell)\}\end{aligned} \tag{4}$$

where

$$\mathbb{A}_{\alpha i} \triangleq \begin{bmatrix} A_{\alpha i} & 0\\ 0 & F\end{bmatrix},\; \mathbb{B}_{\alpha i} \triangleq \begin{bmatrix} B_{\alpha i}\\ 0\end{bmatrix}$$
$$\mathbb{C}_{\alpha i} \triangleq \begin{bmatrix} C_{\alpha i} & 0\\ 0 & G\end{bmatrix},\; \mathcal{I} \triangleq \begin{bmatrix} I & 0\end{bmatrix}^{\mathrm{T}}.$$

Then, the tracking error is defined as

$$y_e(\ell) \triangleq y(\ell) - y_r(\ell) = \mathbb{C}_{0i}X(\ell)$$

where $\mathbb{C}_{0i} \triangleq \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))\mathbb{C}_{0\alpha i}$ and $\mathbb{C}_{0\alpha i} \triangleq [\,C_{\alpha i}\;\; -G\,]$.

Control Rule α: **IF** $\theta_1(\ell)$ is $R_{\alpha 1}$, $\theta_2(\ell)$ is $R_{\alpha 2}, \ldots, \theta_g(\ell)$ is $R_{\alpha g}$, **THEN**

$$\mu(\ell) = (K_{\alpha i} + \Delta K_{\alpha i})Y(\ell)$$

where $K_{\alpha i}$ is controller gain and $\Delta K_{\alpha i} \triangleq \tilde{N}_{\alpha i}T(\ell)\tilde{M}_{\alpha i}$ denotes controller gain uncertainties, where $\tilde{N}_{\alpha i}$ and $\tilde{M}_{\alpha i}$ are known matrices.

Then, we can deduce the overall controller as

$$\begin{aligned}\mu(\ell) &= \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))(K_{\alpha i} + \Delta K_{\alpha i})Y(\ell)\\ &= \breve{K}_i Y(\ell)\end{aligned}$$

where $\breve{K}_i \triangleq \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))(K_{\alpha i} + \Delta K_{\alpha i})$.

Substituting $\mu(\ell)$ to (4), it yields the following system equation:

$$\dot{X}(\ell) = \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))(\mathbb{A}_{\alpha i}X(\ell) + \mathbb{B}_{\alpha i}u(\ell) + \mathbb{E}_{\alpha i}w(\ell)) \tag{5}$$

where $u(\ell) \triangleq \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))K_{\alpha i}Y(\ell)$, $w(\ell) \triangleq (\mathcal{I}H_{\alpha i}T(\ell)(D_{1\alpha i}\mathcal{I}^{\mathrm{T}} + D_{2\alpha i}\breve{K}_i\mathbb{C}_{\alpha i}) + \mathbb{B}_{\alpha i}\Delta K_i\mathbb{C}_{\alpha i})X(\ell)$ denote the impact of system and controller uncertainties on the system. $\mathbb{E}_{\alpha i}$ is the identity matrix.

Remark 1: It is noticed that the system parameter uncertainties and the gain fluctuation are considered as $w(\ell)$, which may worsen the system performance. Furthermore, when $T(\ell) = 0$, we can deduce $w(\ell) = 0$, which means that uncertainties and the gain fluctuation are not occur. Thus, inspired by [44], system parameter uncertainties and nonfragile control problems can be discussed within the framework of zero-sum games theory.

B. Control Design With Zero-Sum Games

The overall performance index is defined as

$$\begin{aligned}V(X(\ell), i) = \mathbb{E}\{&\int_\ell^\infty e^{-\delta(\tau-\ell)}((\mathbb{C}_{0i}X(\tau))^{\mathrm{T}}Q(\mathbb{C}_{0i}X(\tau))\\ &+ u^{\mathrm{T}}(\tau)u(\tau) - \gamma^2 w^{\mathrm{T}}(\tau)w(\tau))\mathrm{d}\tau|X(\ell), i\}.\end{aligned} \tag{6}$$

The solvability of zero-sum games problem can be described as

$$\begin{aligned}V(X(\ell), i) &= \min_{u(\ell)}\max_{w(\ell)} \mathbb{E}\left\{\int_\ell^\infty \mathcal{U}(X, u, w)\mathrm{d}\tau|X(\ell), i\right\}\\ &= \max_{w(\ell)}\min_{u(\ell)} \mathbb{E}\left\{\int_\ell^\infty \mathcal{U}(X, u, w)\mathrm{d}\tau|X(\ell), i\right\}\end{aligned}$$

where $\mathcal{U}(X, u, w) \triangleq e^{-\delta(\tau-\ell)}[(\mathbb{C}_{0i}X(\tau))^{\mathrm{T}}Q(\mathbb{C}_{0i}X(\tau)) + u^{\mathrm{T}}(\tau)u(\tau) - \gamma^2 w^{\mathrm{T}}(\tau)w(\tau)]$.

By employing the parallel distributed compensation scheme, the mode-dependent feedback control policies and performance index under fuzzy rule can be developed as follows.

Control Rule α: **IF** $\theta_1(\ell)$ is $R_{\alpha 1}$, $\theta_2(\ell)$ is $R_{\alpha 2}, \ldots, \theta_g(\ell)$ is $R_{\alpha g}$, **THEN**

$$\begin{aligned}u_\alpha(\ell) &= K_{\alpha i}Y(\ell)\\ w_\alpha(\ell) &= K_{w\alpha i}X(\ell)\\ V_\alpha(X(\ell), i) &= \mathbb{E}\{\int_\ell^\infty \mathcal{U}(X, u_\alpha, w_\alpha)\mathrm{d}\tau|X(\ell), i\}.\end{aligned} \tag{7}$$

Therefore, we have overall control policies expressed as

$$\begin{aligned}u(\ell) &= \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))K_{\alpha i}Y(\ell) = K_iY(\ell)\\ w(\ell) &= \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))K_{w\alpha i}X(\ell) = K_{wi}X(\ell).\end{aligned} \tag{8}$$

Inserting the control policies (8) into (5), it yields

$$\dot{X}(\ell) = (\bar{\mathbb{A}}_i + \bar{\mathbb{B}}_iK_i\bar{\mathbb{C}}_i + \bar{\mathbb{E}}_iK_{wi})X(\ell) \tag{9}$$

where $\bar{\mathbb{A}}_i \triangleq \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))\mathbb{A}_{\alpha i}$, $\bar{\mathbb{B}}_i \triangleq \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))\mathbb{B}_{\alpha i}$, $\bar{\mathbb{C}}_i \triangleq \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))\mathbb{C}_{\alpha i}$, $\bar{\mathbb{E}}_i \triangleq \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))\mathbb{E}_{\alpha i}$.

Before further presentation, the following definitions and lemmas are introduced.

Definition 1 [3]: The continuous-time MJSs (9) with $w(\ell) \equiv 0$ are stochastically stable, if there exists a control $u(\ell)$ satisfying

$$\lim_{T\to\infty} \mathbb{E}\{\int_0^{\mathrm{T}} X^{\mathrm{T}}(\ell, u)X(\ell, u)\mathrm{d}\ell|X_0, \eta_0\} \le X_0^{\mathrm{T}}\mathbb{M}X_0$$

where $\mathbb{M} = \mathbb{M}^{\mathrm{T}}$ is a symmetric positive-definite matrix.

Definition 2 [45]: The continuous-time MJSs (9) are stochastically stable along with an $\mathcal{H}_\infty$ performance index, if for all X_0, η_0, and nonzero w, the following inequality holds:

$$\mathbb{E}\left\{\int_t^\infty e^{-\delta(\tau-\ell)}\left(y_e^{\mathrm{T}}Qy_e+u^{\mathrm{T}}u\right)\mathrm{d}\tau\right\}\\ \le \gamma^2\int_t^\infty e^{-\delta(\tau-\ell)}w^{\mathrm{T}}w\mathrm{d}\tau.$$

Lemma 1 [34]: Given $\mathbb{S}$, $\mathbb{S}_\alpha > 0$ be symmetrical positive-definite matrices, which meet $\mathbb{S}\mathbb{S}_\alpha = \mathbb{S}_\alpha\mathbb{S}$. Then, $\mathbb{S}\mathbb{S}_\alpha$ and $\mathbb{S}_\alpha\mathbb{S}$ are positive definite.

Lemma 2 [44]: Let $V(\ell) = x^{\mathrm{T}}(\ell)SZSx(\ell)$ and $V_\alpha(\ell) = x^{\mathrm{T}}(\ell)SZS_\alpha x(\ell)$, where S, $S_\alpha > 0$ are symmetrical, $SS_\alpha = S_\alpha S$ and $Z^{\mathrm{T}} = Z$ are the symmetric matrix. One can deduce that $V(\ell) = \sigma V_\alpha(\ell)$, $\sigma > 0$ is a constant.

Lemma 3: Considering the MJFSs (9), both the overall performance index $V(X(\ell), i)$ and the fuzzy-based performance index $V_\alpha(X(\ell), i)$ can be expressed as the quadratic form $V(X(\ell), i) = X^{\mathrm{T}}(\ell)P_iX(\ell)$ and $V_\alpha(X(\ell), i) = X^{\mathrm{T}}(\ell)P_{\alpha i}X(\ell)$.

Proof: For the control policies u_α and w_α, the performance index $V_\alpha(X(\ell), i)$ can be developed as

$$\begin{aligned}V_\alpha(X(\ell), i) &= \mathbb{E}\left\{\int_\ell^\infty \mathcal{U}(X, u_\alpha, w_\alpha)\mathrm{d}\tau|X(\ell), i\right\}\\ &= \mathbb{E}\left\{\int_0^\infty e^{-\delta\tau}\left[X^{\mathrm{T}}(\tau+\ell)\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i}X(\tau+\ell)\right.\right.\\ &\quad + X^{\mathrm{T}}(\tau+\ell)\mathbb{C}_{\alpha i}^{\mathrm{T}}K_\alpha^{\mathrm{T}}K_\alpha\mathbb{C}_{\alpha i}X(\tau+\ell)\\ &\quad \left.\left. - \gamma^2X^{\mathrm{T}}(\tau+\ell)K_{w\alpha}^{\mathrm{T}}K_{w\alpha}X(\tau+\ell)\right]\mathrm{d}\tau|X(\ell), i\right\}\end{aligned}$$

which implies

$$\begin{aligned}V_\alpha(X(\ell), i) &= X^{\mathrm{T}}(\ell)\left(\mathbb{E}\left\{\int_0^\infty e^{-\delta\tau}\left[\exp^{\mathrm{T}}(\Gamma_i\tau)(\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i}\right.\right.\right.\\ &\quad \left. + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}K_{\alpha i}\mathbb{C}_{\alpha i} - \gamma^2K_{w\alpha i}^{\mathrm{T}}K_{w\alpha i})\right]\\ &\quad \left.\left.\times\ \mathrm{d}\tau|X(\ell), i\right\}\right)X(\ell)\\ &= X^{\mathrm{T}}(\ell)P_{\alpha i}X(\ell)\end{aligned}$$

where $\Gamma_i = \mathbb{A}_{\alpha i}+\mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i}+\mathbb{E}_{\alpha i}K_{w\alpha i}$, $P_{\alpha i} = \mathbb{E}\{\int_0^\infty e^{-\delta\tau}[\exp^{\mathrm{T}}(\Gamma_i\tau)(\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i} + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}K_{\alpha i}\mathbb{C}_{\alpha i} - \gamma^2K_{w\alpha i}^{\mathrm{T}}K_{w\alpha i})]\mathrm{d}\tau|X(\ell), i\}$. Similarly, one can deduce that $V(X(\ell), i) = X^{\mathrm{T}}(\ell)P_iX(\ell)$. This end the proof. ■

III. MAIN RESULTS

In this section, the problem of nonfragile output feedback tracking control for MJFSs is converted to solving a set of zero-sum games. Two kinds of learning algorithms are presented to solve FSCAREs, respectively. Furthermore, the equivalence of solutions of offline Algorithm 1 and online Algorithm 2 is proved. Besides, the stability is proved by the Lyapunov stability theory.

A. Problem Transformation

Theorem 1: Considering the MJFSs (9) in the αth rule, the solution optimal matrix $P_{\alpha i}$ of $V_\alpha(X(\ell), i)$ satisfies the following FSCAREs:

$$\begin{aligned}&\mathbb{A}_{\alpha i}^{\mathrm{T}}P_{\alpha i}+P_{\alpha i}\mathbb{A}_{\alpha i}-\delta P_{\alpha i}-P_{\alpha i}\mathbb{B}_{\alpha i}\mathbb{B}_{\alpha i}^{\mathrm{T}}P_{\alpha i}\\ &\quad+\left(1/\gamma^2\right)P_{\alpha i}\mathbb{E}_{\alpha i}\mathbb{E}_{\alpha i}^{\mathrm{T}}P_{\alpha i}+\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i}\\ &\quad+L_{\alpha i}^{\mathrm{T}}L_{\alpha i}+\sum_{j=1}^{S}\pi_{ij}P_{\alpha j}=0\end{aligned}\tag{10}$$

Algorithm 1: Offline Parallel Output Feedback Nonfragile Tracking Control Algorithm for MJFSs

Select a sequence $P_\alpha(0) = (P_{\alpha 1}(0), \ldots, P_{\alpha m}(0))$ for every rule. Start with initial admissible feedback matrices $K_{\alpha i}(0)$, $K_{w\alpha i}(0)$ and $L_{\alpha i}(0)$.
repeat
  Parallel evaluates $P_{\alpha i}(\lambda+1)$ from

$$\begin{aligned}&(\tilde{\mathbb{A}}_{\alpha i}+\mathbb{B}_{\alpha i}K_{\alpha i}(\lambda)\mathbb{C}_{\alpha i}+\mathbb{E}_{\alpha i}K_{w\alpha i}(\lambda)-0.5\delta I)^{\mathrm{T}}\\ &P_{\alpha i}(\lambda+1)+P_{\alpha i}(\lambda+1)(\tilde{\mathbb{A}}_{\alpha i}+\mathbb{B}_{\alpha i}K_{\alpha i}(\lambda)\\ &\times\mathbb{C}_{\alpha i}+\mathbb{E}_{\alpha i}K_{w\alpha i}(\lambda)-0.5\delta I)\\ &=-\mathbb{C}_{0i}^{\mathrm{T}}Q\mathbb{C}_{0i}-\mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}(\lambda)K_{\alpha i}(\lambda)\mathbb{C}_{\alpha i}\\ &\quad+\gamma^2K_{w\alpha i}^{\mathrm{T}}(\lambda)K_{w\alpha i}(\lambda)-\sum_{j\neq i}^{S}\pi_{ij}P_{\alpha j}(\lambda)\end{aligned}\tag{11}$$

  Parallel policy is updated by

$$\begin{aligned}K_{\alpha i}(\lambda+1) &= -(\mathbb{B}_{\alpha i}^{\mathrm{T}}P_{\alpha i}(\lambda+1)+L_{\alpha i}(\lambda))\\ &\quad\times\mathbb{C}_{\alpha i}^{\mathrm{T}}(\mathbb{C}_{\alpha i}\mathbb{C}_{\alpha i}^{\mathrm{T}})^{-1}\\ K_{w\alpha i}(\lambda+1) &= \left(1/\gamma^2\right)\mathbb{E}_{\alpha i}^{\mathrm{T}}P_{\alpha i}(\lambda+1)\\ L_{\alpha i}(\lambda+1) &= K_{\alpha i}(\lambda+1)\mathbb{C}_{\alpha i}-\mathbb{B}_i^{\mathrm{T}}P_{\alpha i}(\lambda+1)\end{aligned}\tag{12}$$

until $\max_{\alpha i}\|P_{\alpha i}(\lambda+1)-P_{\alpha i}(\lambda)\| < \epsilon$, *$\epsilon > 0$ is a small constant*;

Algorithm 2: Online Parallel IRL-Based Output Feedback Nonfragile Tracking Control Algorithm for MJFSs

Initialization
Select a sequence $P_\alpha(0) = (P_{\alpha 1}(0), \ldots, P_{\alpha m}(0))$ for every rule and given inital admissible feedback matrices $K_{\alpha i}(0)$ $K_{w\alpha i}(0)$ and $L_{\alpha i}(0)$. Let $\Delta\ell$ be a small enough positive constant, and $\lambda = 1$.
Procedure
while $\max_{\alpha i}\|P_{\alpha i}(\lambda+1)-P_{\alpha i}(\lambda)\| > \epsilon$, *$\epsilon > 0$ is a small constant* **do**
  for *for all $\eta(\ell) = i \in \mathbb{S}$* **do**
    for *for all $\alpha \in \mathbb{M}$* **do**
      Evaluate matrix $P_{\alpha i}$ from
      $X^{\mathrm{T}}(\ell)P_{\alpha i}(\lambda)X(\ell) - e^{-\delta\Delta\ell}X^{\mathrm{T}}(\ell+\Delta\ell)P_{\alpha i}(\lambda)X(\ell+\Delta\ell)$
      $= \int_\ell^{\ell+\Delta\ell} e^{-\delta(\tau-\ell)}X^{\mathrm{T}}(\tau)(\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i} + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}K_{\alpha i}\mathbb{C}_{\alpha i}-\gamma^2K_{w\alpha i}^{\mathrm{T}}K_{w\alpha i}X(\tau))\mathrm{d}\tau$
      Update control policies by
      $K_{\alpha i}(\lambda+1) = -\left(\mathbb{B}_{\alpha i}^{\mathrm{T}}P_{\alpha i}(\lambda)+L_{\alpha i}(\lambda)\right)\mathbb{C}_{\alpha i}^{\mathrm{T}}\left(\mathbb{C}_{\alpha i}\mathbb{C}_{\alpha i}^{\mathrm{T}}\right)^{-1}$
      $K_{w\alpha i}(\lambda+1) = \left(1/\gamma^2\right)\mathbb{E}_{\alpha i}^{\mathrm{T}}P_{\alpha i}(\lambda)$
      $L_{\alpha i}(\lambda+1) = K_{\alpha i}(\lambda+1)\mathbb{C}_{\alpha i}-\mathbb{B}_i^{\mathrm{T}}P_{\alpha i}(\lambda)$
      Update the feedback control policies as
      $u_{\alpha i}(\ell) = K_{\alpha i}(\lambda+1)Y(\ell)$
      $w_{\alpha i}(\ell) = K_{w\alpha i}(\lambda+1)X(\ell)$
    end
  end
  Set $\lambda = \lambda+1$
end
Use the overall control policies with
$u(\ell) = \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))K_{\alpha i}Y(\ell)$
$w(\ell) = \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))K_{w\alpha i}X(\ell)$
End Procedure

and corresponding optimal control policies are obtained as $u_\alpha(\ell) = K_{\alpha i}Y(\ell)$ and $w_\alpha(\ell) = K_{w\alpha i}X(\ell)$, where $K_{\alpha i} = -(\mathbb{B}_i^{\mathrm{T}}P_{\alpha i} + L_{\alpha i})\mathbb{C}_{\alpha i}^{\mathrm{T}}(\mathbb{C}_{\alpha i}\mathbb{C}_{\alpha i}^{\mathrm{T}})^{-1}$ and $K_{w\alpha i} = (1/\gamma^2)\mathbb{E}_{\alpha i}^{\mathrm{T}}P_{\alpha i}$.

Proof: See Appendix A. ■

Remark 2: According to [23], the output feedback optimal control design problem may not have a global optimal solution. Thus, the matrix $L_{\alpha i}$ is employed in this article to show the difference between the optimal state-feedback control gain and the proposed output-feedback control gain for MJFSs.

To solve FSCAREs (10), a novel offline parallel algorithm and an IRL-based online parallel algorithm are designed in the following part.

B. Offline Parallel Algorithm Design

In this part, an offline parallel output feedback tracking algorithm is proposed for MJFSs. By mathematical transformation, (10) can be rewritten as

$$\begin{aligned}&\Big[\tilde{\mathbb{A}}_{\alpha i} + \mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i} + \mathbb{E}_{\alpha i}K_{w\alpha i} - 0.5\delta I\Big]^{\mathrm{T}}P_{\alpha i}\\ &\quad + P_{\alpha i}\Big[\tilde{\mathbb{A}}_{\alpha i} + \mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i} + \mathbb{E}_{\alpha i}K_{w\alpha i} - 0.5\delta I\Big]\\ &= -\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i} - \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}K_{\alpha i}\mathbb{C}_{\alpha i} + \gamma^2K_{w\alpha i}^{\mathrm{T}}K_{w\alpha i}\\ &\quad - \sum_{j=1, j\neq i}^{S}\pi_{ij}P_{\alpha j}\end{aligned}$$

where $\tilde{\mathbb{A}}_{\alpha i} = \mathbb{A}_{\alpha i} + ((\pi_{ii}/2)I)$, matrices $P_{\alpha i}$ and $P_{\alpha j}$ are symmetric positive definite. Using parallel Lyapunov (11) to evaluate the matrix $P_{\alpha i}$, a parallel two-step learning method is proposed in Algorithm 1.

Remark 3: Due to the existence of the term $\sum_{j=1}^{S}\pi_{ij}P_{\alpha j}$, the existing algorithm cannot be directly used to solve $P_{\alpha i}$ from FSCAREs (10). Thus, we transform $\sum_{j=1}^{S}\pi_{ij}P_{\alpha j}$ into $\sum_{j\neq i}^{S}\pi_{ij}P_{\alpha j}+\pi_{ii}P_{\alpha i}$ and rewrite $\tilde{\mathbb{A}}_{\alpha i}$ as $\mathbb{A}_{\alpha i}+((\pi_{ii}/2)I_n)$. Based on this transformation, $P_{\alpha i}$ can be parallel evaluated from (11).

Note that the system dynamics and the exact information of stationary transition probability are required in Algorithm 1, which may be difficult to meet in practical applications.

C. Online Parallel IRL-Based Algorithm Design

To relax the requirement of exact information of stationary transition probability and system dynamic $\mathbb{A}_{\alpha i}$ of MJFSs, a novel IRL-based output feedback tracking nonfragile control algorithm is designed to solving a class of FSCAREs.

Considering MJFSs (9) and performance index (7), we have the following integral Bellman equation:

$$\begin{aligned}&X^{\mathrm{T}}(\ell)P_{\alpha i}X(\ell) - e^{-\delta\Delta\ell}X^{\mathrm{T}}(\ell+\Delta\ell)P_{\alpha i}X(\ell+\Delta\ell)\\ &= \int_{\ell}^{\ell+\Delta\ell} e^{-\delta(\tau-\ell)}X^{\mathrm{T}}(\tau)(\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i} + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}K_{\alpha i}\mathbb{C}_{\alpha i}\\ &\quad - \gamma^2K_{w\alpha i}^{\mathrm{T}}K_{w\alpha i})X(\tau)\mathrm{d}\tau. \end{aligned}\tag{13}$$

Note that stationary transition probability and system dynamic are avoided in (13). In the following, we will prove that the solution obtained by (13) is equal to the solution of Lyapunov (11) in Algorithm 1.

Theorem 2: For every fuzzy rule α, the positive-define matrix $P_{\alpha i}$ calculated from (13) in Algorithm 2 is equal to the solution by solving the Lyapunov (11) in Algorithm 1.

Proof: See Appendix B. ■

Based on integral Bellman (13), we present a novel IRL-based nonfragile output feedback tracking control method for MJFSs, which is shown in Algorithm 2.

Remark 4: Compared with Algorithm 1, the proposed IRL-based algorithm exhibits the following distinct advantages.

1) The computational burden is reduced. The main computation is to solve (13) by using the least square method. The calculation process is shown in the following.
2) In Algorithm 2, the requirement of accurate dynamic information can be avoided in the process of learning.

In the following, we show the computation method to implement Algorithm 2, where the matrix $P_{\alpha i}(\lambda)$ is obtained from the following equality:

$$\begin{aligned}&X^{\mathrm{T}}(\ell)P_{\alpha i}(\lambda)X(\ell) - e^{-\delta\Delta\ell}X^{\mathrm{T}}(\ell+\Delta\ell)P_{\alpha i}(\lambda+1)X(\ell+\Delta\ell)\\ &= \int_{\ell}^{\ell+\Delta\ell} e^{-\delta(\tau-\ell)}X^{\mathrm{T}}(\tau)(\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i} + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}(\lambda)\\ &\quad \times K_{\alpha i}(\lambda)\mathbb{C}_{\alpha i} - \gamma^2K_{w\alpha i}^{\mathrm{T}}(\lambda)K_{w\alpha i}(\lambda)X(\tau))\mathrm{d}\tau\end{aligned}$$

and introduce batch least squares method to obtain the solution of matrix $P_{\alpha i}^{\lambda}$. Define the following operators:

$$\begin{aligned}\bar{P}_{\alpha i}^{\lambda} &= \mathrm{vecs}(P_{\alpha i}(\lambda)) \in \mathbf{R}^{n(n+1)/2}\\ &\triangleq \begin{bmatrix} P_{\alpha i}^{\lambda 11} & 2P_{\alpha i}^{\lambda 12} & \dots & P_{\alpha i}^{\lambda 22} & \dots & 2P_{\alpha i}^{\lambda 2n} & \dots & P_{\alpha i}^{\lambda nn}\end{bmatrix}^{\mathrm{T}}\\ \bar{X}(\ell) &\triangleq \begin{bmatrix} X_1^2 & X_1X_2 & \dots & X_2^2 & \dots & X_2X_n & \dots & X_n^2\end{bmatrix}^{\mathrm{T}}.\\ &\in \mathbf{R}^{n(n+1)/2}\end{aligned}$$

Then, one can deduce

$$X^{\mathrm{T}}(\ell)P_{\alpha i}^{\lambda}X(\ell) = \big(\bar{P}_{\alpha i}^{\lambda}\big)^{\mathrm{T}}\bar{X}(\ell).$$

Furthermore, the parallel policy evaluation (13) can be rewritten as

$$\big(\bar{P}_{\alpha i}^{\lambda}\big)^{\mathrm{T}}\tilde{X}(\ell) = v\big(X(\ell), K_{\alpha i}^{\lambda}, K_{w\alpha i}^{\lambda}\big)$$

where $\tilde{X}(\ell) = \bar{X}(\ell) - e^{-\delta\Delta\ell}\bar{X}(\ell+\Delta\ell)$ and $v(X(\ell), K_{\alpha i}^{\lambda}, K_{w\alpha i}^{\lambda}) = \int_{\ell}^{\ell+\Delta\ell} e^{-\delta(\tau-\ell)}X^{\mathrm{T}}(\tau)(\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i} + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}(\lambda)K_{\alpha i}(\lambda)\mathbb{C}_{\alpha i} - \gamma^2K_{w\alpha i}^{\mathrm{T}}(\lambda)K_{w\alpha i}(\lambda)X(\tau))\mathrm{d}\tau$.

By using the least squares method, we need to collect $N \geq n(n+1)/2$ data and matrix $\bar{P}_{\alpha i}^{\lambda}$ can be solved by

$$\bar{P}_{\alpha i}^{\lambda} = \Big(\tilde{X}^{\lambda}\big(\tilde{X}^{\lambda}\big)^{\mathrm{T}}\Big)^{-1}\tilde{X}^{\lambda}\big(\Theta_{\alpha i}^{\lambda}\big)$$

where

$$\begin{aligned}\tilde{X}^{\lambda} &= \begin{bmatrix}\tilde{X}_1^{\lambda} & \tilde{X}_2^{\lambda} & \dots & \tilde{X}_N^{\lambda}\end{bmatrix}^{\mathrm{T}} \in \mathbf{R}^{\frac{n(n+1)}{2}\times N}\\ \Theta_{\alpha i}^{\lambda} &= \Big[v\big(X_1^{\lambda}(\ell), K_{\alpha i}^{\lambda}, K_{w\alpha i}^{\lambda}\big) \;\; v\big(X_2^{\lambda}(\ell), K_{\alpha i}^{\lambda}, K_{w\alpha i}^{\lambda}\big)\\ &\qquad \dots \;\; v\big(X_N^{\lambda}(\ell), K_{\alpha i}^{\lambda}, K_{w\alpha i}^{\lambda}\big)\Big]^{\mathrm{T}} \in \mathbf{R}^{N}.\end{aligned}$$

Finally, the obtained matrix $P_{\alpha i}^{\lambda}$ can be employed to compute $K_{\alpha i}^{\lambda+1}$ and $K_{w\alpha i}^{\lambda+1}$.

Remark 5: Note that matrix $P_{\alpha i}$ has $n(n+1)/2$ unknown independent elements, thus, we need at least $N \geq n(n+1)/2$ sampled data to evaluate $P_{\alpha i}$ by least squares method. $\Delta\ell$ can be regarded as control policy update interval, where stationary transition probability π_{ij} and system dynamic $\mathbb{A}_{\alpha i}$ are avoided by collecting system state and control input information. As shown in [45], the value of $\Delta\ell$ may have little influence on the convergence of Algorithm 2. It is necessary to select a big enough $\Delta\ell$ to learn the matrix $P_{\alpha i}$, which can precise reflect the jumping characteristics on system data.

D. Stability Analysis

From Algorithm 2, the overall fuzzy control policies are obtained. The following theorem will prove that overall fuzzy control policies integrated by local fuzzy control policies can make MJFSs (9) stochastically asymptotically stable with an $\mathcal{H}_\infty$ performance.

Theorem 3: Consider the performance index with quadratic from $V(X(\ell), i) = X^{\mathrm{T}}(\ell)P_iX(\ell)$, where $P_i > 0$ is symmetric, then, system (9) is stochastically asymptotically stable with $\mathcal{H}_\infty$ performance if there exist matrices $\bar{K}_i$, $\bar{K}_{wi}$, and L_i such that

$$\bar{K}_i\bar{\mathbb{C}}_i = -\left(\bar{\mathbb{B}}_i^{\mathrm{T}}P_i + L_i\right)$$
$$\bar{K}_{wi} = \left(1/\gamma^2\right)\mathbb{E}_i^{\mathrm{T}}P_i$$

where $P_i = P_i^{\mathrm{T}}$ is the solution of

$$\left(\tilde{\mathbb{A}}_i - 0.5\delta I\right)^{\mathrm{T}}P_i + P_i\left(\tilde{\mathbb{A}}_i - 0.5\delta I\right) + \sum_{j=1, j\neq i}^{S}\pi_{ij}P_j + \mathbb{C}_{0i}^{\mathrm{T}}Q\mathbb{C}_{0i} - P_i\bar{\mathbb{B}}_i\bar{\mathbb{B}}_i^{\mathrm{T}}P_i + \frac{1}{\gamma^2}P_i\mathbb{E}_i\mathbb{E}_i^{\mathrm{T}}P_i + L_i^{\mathrm{T}}L_i = 0$$

with $\bar{K}_i = \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))K_{\alpha i} = -\sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))(\mathbb{B}_{\alpha i}^{\mathrm{T}}P_{\alpha i} + L_{\alpha i})\ \mathbb{C}_{\alpha i}^{\mathrm{T}}(\mathbb{C}_{\alpha i}\mathbb{C}_{\alpha i}^{\mathrm{T}})^{-1}$, $\bar{K}_{wi} = \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))K_{w\alpha i} = \sum_{\alpha=1}^{m} f_\alpha(\theta(\ell))(1/\gamma^2)\mathbb{E}_{\alpha i}^{\mathrm{T}}P_{\alpha i}$.

Proof: See Appendix C. ■

Remark 6: Compared with the existing related control methods, the method proposed in this article has the following features.

1) Compared with [46] and [47] that investigate MJSs by using the RL method, the nonlinearity, system uncertainty, and controller gain fluctuation are employed in this article simultaneously, which are not involved in [46] and [47]. The system model considered in this article is more general. Moreover, we first develop two kinds of nonfragile output feedback tracking learning algorithms for MJFSs.
2) Although the nonfragile control problem is solved in [7] and [12] for MJSs by utilizing the linear matrix inequality and matrix scaling techniques, it may bring some conservatism. In this article, we regard control input and the impact of system parameter uncertainties and the controller gain fluctuation as two hostile players and transform the nonfragile control into solving the zero-sum games for MJFSs.
3) The core characteristic distinguishing this article and [14], [17] lies in that the requirement of system dynamic $\mathbb{A}_{\alpha i}$ and stationary transition probability π_{ij} are avoided in the proposed method by using the IRL scheme. Thus, the presented control method for MJFSs is more practical.

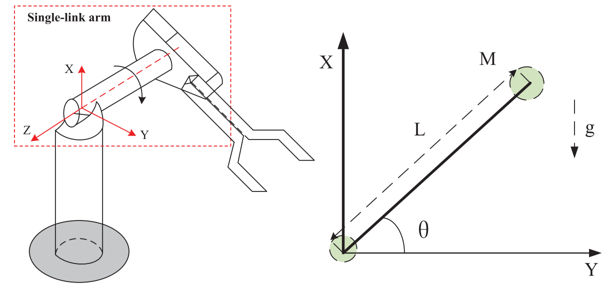

Fig. 2. Single-link robot arm.

TABLE I
PARAMETER VALUES OF SINGLE-LINK ROBOT ARM SYSTEM

Mode	Parameter $\mathbb{M}$	Parameter $\mathbb{J}$	Parameter L	Parameter g
1	1	1	0.5	9.81
2	2	2	0.5	9.81

IV. SIMULATION

In this section, a single-link robot arm model is adopted to verify the availability of the main results. We assume that the reference trajectory system has the following form:

$$\dot{x}_r(\ell) = Fx_r(\ell) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x_r(\ell)$$
$$y_r(\ell) = Gx_r(\ell) = \begin{bmatrix} 1 & 1 \end{bmatrix} x_r(\ell)$$

where $\dot{x}_r(\ell) = [x_{1r}^{\mathrm{T}}(\ell)\ \ x_{2r}^{\mathrm{T}}(\ell)]^{\mathrm{T}}$ and the initial state $\dot{x}_r(0) = [1\ \ 1]^{\mathrm{T}}$.

The single-link robot arm system is presented in Fig. 2, in which the dynamic equation is given by

$$\ddot{\vartheta}(\ell) = -\frac{MgL}{\mathbb{J}}\sin(\vartheta(\ell)) - \frac{\mathbb{D}(\ell)}{\mathbb{J}}\dot{\vartheta}(\ell) + \frac{1}{\mathbb{J}}u(\ell)$$

where $u(\ell)$, $\vartheta(\ell)$, $\mathbb{J}$, g, L, M, and $\mathbb{D}(\ell)$ represent control input, the angle position, the total inertia, acceleration of gravity, the length of the arm, load mass, and the damping, respectively. The values of parameters $\mathbb{D}(\ell)$ are selected with its nominal value $\mathbb{D} = 2$ and the other parameters are presented in Table I with two modes. The transition probability rate matrix is given as follows:

$$\Pi = \begin{bmatrix} -0.3 & 0.3 \\ 0.5 & -0.5 \end{bmatrix}.$$

Let $x_1(\ell) = \vartheta(\ell)$ and $x_2(\ell) = \dot{\vartheta}(\ell)$. Under the condition $-179.4270° < \vartheta(\ell) < 179.4270°$, the nonlinear function $\sin(x_1(\ell))$ can be described as

$$\sin(x_1(\ell)) = f_1(x_1(\ell)) \cdot x_1(\ell) + f_2(x_1(\ell)) \cdot \xi \cdot x_1(\ell)$$

where $\xi = 10^{-2}/\pi$. Then, $f_1(x_1(\ell))$ and $f_2(x_1(\ell))$ can be obtained as

$$f_1(x_1(\ell)) = \begin{cases} \frac{\sin(x_1(\ell)) - \xi x_1(\ell)}{x_1(\ell)(1-\xi)}, & x_1(\ell) \neq 0 \\ 1, & x_1(\ell) = 0 \end{cases}$$
$$f_2(x_1(\ell)) = 1 - f_1(x_1(\ell)).$$

Then, MJFSs can be represented as the following T–S fuzzy model.

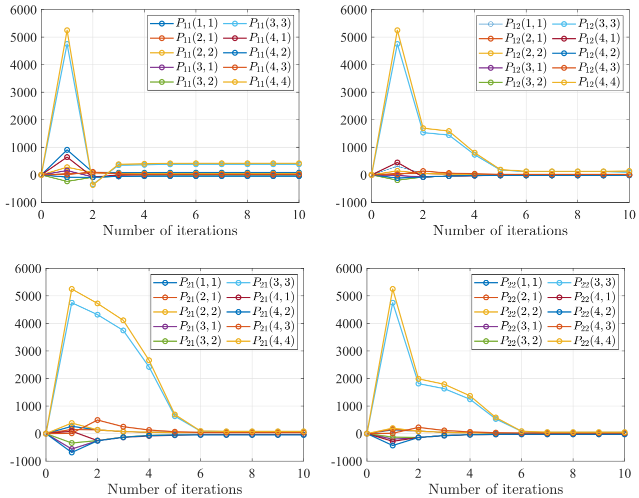

Fig. 3. Convergence of matrices P_{11}, P_{12}, P_{21}, and P_{22}.

Plant Rule 1:
IF $x_1(\ell)$ is "around 0 rad " **THEN**

$$\dot{x}(\ell) = (A_{1i} + \Delta A_{1i}(\ell))x(\ell) + (B_{1i} + \Delta B_{1i}(\ell))u(\ell)$$
$$y(\ell) = C_{1i}x(\ell).$$

Plant Rule 2:
IF $x_1(\ell)$ is "around π rad or $-\pi$ rad " **THEN**

$$\dot{x}(\ell) = (A_{2i} + \Delta A_{2i}(\ell))x(\ell) + (B_{2i} + \Delta B_{2i}(\ell))u(\ell)$$
$$y(\ell) = C_{2i}x(\ell)$$

where $x(\ell) = [\,x_1^{\mathrm{T}}(\ell)\ x_2^{\mathrm{T}}(\ell)\,]^{\mathrm{T}}$

$$A_{11} = \begin{bmatrix} 0 & 1 \\ -gL & -\mathbb{D}_0 \end{bmatrix},\ A_{12} = \begin{bmatrix} 0 & 1 \\ -gL & -0.5\mathbb{D}_0 \end{bmatrix}$$
$$A_{21} = \begin{bmatrix} 0 & 1 \\ -gL/\pi & -\mathbb{D}_0 \end{bmatrix},\ A_{22} = \begin{bmatrix} 0 & 1 \\ -gL/\pi & -0.5\mathbb{D}_0 \end{bmatrix}$$
$$B_{11} = B_{21} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ B_{12} = B_{22} = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$$
$$C_{\alpha i} = \begin{bmatrix} 1 & 1 \end{bmatrix},\ H_{\alpha i} = \tilde{N}_{\alpha i} = 0.01$$
$$T(\ell) = \begin{cases} \sin(\ell), & 0 \le \ell \le 2 \\ 0, & \ell > 2 \end{cases}$$
$$D_{1\alpha i} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix},\ D_{2\alpha i} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$
$$\tilde{M}_{\alpha i} = \begin{bmatrix} 0.1 & 0.2 & 0.1 & 0.2 \end{bmatrix},\ i = 1, 2,\ \alpha = 1, 2.$$

We select $Q = 500$, $\delta = 0.1$, $\gamma = 100$. The initial state is $X(0) = [2.2\ \ -3\ \ 1\ \ 1]^{\mathrm{T}}$. The simulation is employed using the system data at every 0.05 s. The control policies are considered to update at every 2 s. By applying parallel Algorithm 1, the output feedback tracking control gains are obtained as

$$K_{11} = \begin{bmatrix} 17.9173 & -21.4952 \end{bmatrix}$$
$$K_{12} = \begin{bmatrix} 19.8191 & -22.2238 \end{bmatrix}$$
$$K_{21} = \begin{bmatrix} 20.5846 & -21.9922 \end{bmatrix}$$
$$K_{22} = \begin{bmatrix} 21.2753 & -22.4877 \end{bmatrix}.$$

Furthermore, along with the learning process in Algorithm 2, the parameter matrices P_{11}, P_{12}, P_{21}, and P_{22} are presented in Fig. 3, where both stability and convergence of the IRL-based Algorithm 2 are achieved. And the gains of controller are obtained without using the system dynamic $\mathbb{A}_{\alpha i}$ and transition probability π_{ij}, which are presented as follows:

$$K_{11} = \begin{bmatrix} 18.3226 & -21.4508 \end{bmatrix}$$
$$K_{12} = \begin{bmatrix} 19.6216 & -22.0838 \end{bmatrix}$$
$$K_{21} = \begin{bmatrix} 20.7476 & -22.2003 \end{bmatrix}$$
$$K_{22} = \begin{bmatrix} 21.0660 & -22.2516 \end{bmatrix}.$$

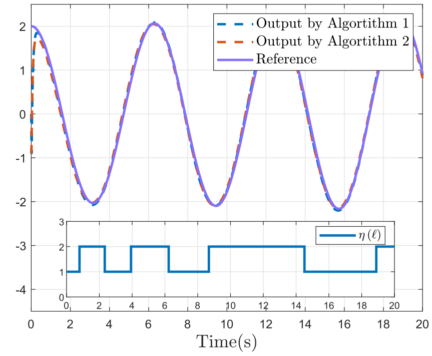

Fig. 4. Tracking trajectories by Algorithms 1 and 2.

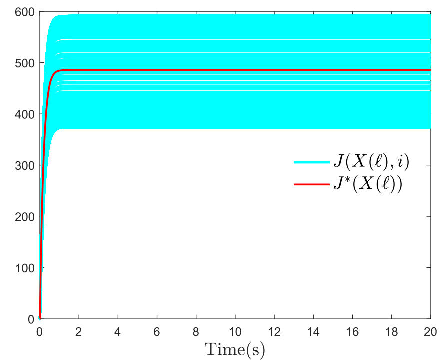

Fig. 5. Monte Carlo simulation.

By using the control gains calculated from Algorithms 1 and 2, the system tracking trajectories are shown in Fig. 4, respectively, where the reference trajectories get well tracked. Moreover, the mode evaluation of the Markov chain during the learning process is also depicted in Fig. 4. It indicates the validity and practicability of the proposed learning method. Moreover, based on the control policies obtained from Algorithm 2, we perform 1000 Monte Carlo experiments in Fig. 5 to show the convergence of mode-based performance index $J(X(\ell), i)$ and overall optimal performance index $J^*(X(\ell), i)$.

V. Conclusion

A novel online IRL-based nonfragile output feedback tracking control algorithm has been first proposed for MJFSs. The proposed online algorithm does not require the information of all the system dynamics. Moreover, by considering the control policy and disturbance as two rival players, the problem of nonfragile control can be converted into the issue of zero-sum games. Two kinds of learning algorithms have been developed for solving the FSCAREs for MJFSs. For the proposed algorithm, stochastic stability has been proven via the Lyapunov stability theory. In addition, we have applied the designed algorithm in an application, and the simulation results have demonstrated the effectiveness. In our future work, we will attempt to apply the obtained results to semi-Markov systems or hidden-Markov systems.

Appendix A
Proof of Theorem 1

Proof: Together with Lemma 3 and weak infinitesimal operator $\mathcal{L}$, we have

$$\begin{aligned}
&\mathcal{L}V_a(X(\ell), i)\\
&= \lim_{\Delta\ell\to 0}\frac{1}{\Delta\ell}[\mathbb{E}\{V_a(X(\ell+\Delta\ell), i+\Delta\ell)|X(\ell), i\}\\
&\quad - V_a(X(\ell), i)]\\
&= X^{\mathrm{T}}(\ell)[(\mathbb{A}_{\alpha i}+\mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i}+\mathbb{E}_{\alpha i}K_{w\alpha i})^{\mathrm{T}}P_{\alpha i}\\
&\quad + P_{\alpha i}(\mathbb{A}_{\alpha i}+\mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i}+\mathbb{E}_{\alpha i}K_{w\alpha i})\\
&\quad + \sum_{j=1}^{S}\pi_{ij}P_{\alpha j}]X(\ell)\\
&= -X^{\mathrm{T}}(\ell)\Big[\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i} + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}K_{\alpha i}\mathbb{C}_{\alpha i}\\
&\qquad - \gamma^2K_{w\alpha i}^{\mathrm{T}}K_{w\alpha i} - \delta P_{\alpha i}\Big]X(\ell)
\end{aligned}$$

which implies the following Bellman equation:

$$\begin{aligned}
&(\mathbb{A}_{\alpha i}+\mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i}+\mathbb{E}_{\alpha i}K_{w\alpha i})^{\mathrm{T}}P_{\alpha i}\\
&\quad + P_{\alpha i}(\mathbb{A}_{\alpha i}+\mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i}+\mathbb{E}_{\alpha i}K_{w\alpha i}) + \sum_{j=1}^{S}\pi_{ij}P_{\alpha j} - \delta P_{\alpha i}\\
&\quad + \mathbb{C}_{0i}^{\mathrm{T}}Q\mathbb{C}_{0i} + \mathbb{C}_{\alpha i}^{\mathrm{T}}K_{\alpha i}^{\mathrm{T}}K_{\alpha i}\mathbb{C}_{\alpha i} - \gamma^2K_{w\alpha i}^{\mathrm{T}}K_{w\alpha i} = 0. \qquad (14)
\end{aligned}$$

Based on the Bellman's optimization principle, we can further obtain the optimal policy as $u_\alpha(\ell) = -\mathbb{B}_{\alpha i}^{\mathrm{T}}P_{\alpha i}\mathbb{C}_{\alpha i}X(\ell)$ and $w_\alpha(\ell) = (1/\gamma^2)\mathbb{E}_{\alpha i}^{\mathrm{T}}P_{\alpha i}X(\ell)$. From (14), we have

$$\begin{aligned}
&\mathbb{A}_{\alpha i}^{\mathrm{T}}P_{\alpha i} + P_{\alpha i}\mathbb{A}_{\alpha i} - \delta P_{\alpha i} - P_{\alpha i}\mathbb{B}_{\alpha i}\mathbb{B}_{\alpha i}^{\mathrm{T}}P_{\alpha i}\\
&\quad + (1/\gamma^2)P_{\alpha i}\mathbb{E}_{\alpha i}\mathbb{E}_{\alpha i}^{\mathrm{T}}P_{\alpha i} + \mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i}\\
&\quad + L_{\alpha i}^{\mathrm{T}}L_{\alpha i} + \sum_{j=1}^{S}\pi_{ij}P_{\alpha j} = 0
\end{aligned}$$

where equation $K_{\alpha i}\mathbb{C}_{\alpha i} = -(\mathbb{B}_{\alpha i}^{\mathrm{T}}P_{\alpha i} + L_{\alpha i})$ is used. This completes the proof. ■

Appendix B
Proof of Theorem 2

Proof: Dividing both sides of (13) with $\Delta\ell$, we have

$$\begin{aligned}
&\lim_{\Delta\ell\to 0}\frac{1}{\Delta\ell}\Big[\int_{\ell}^{\ell+\Delta\ell} e^{-\delta(\tau-\ell)}\Big[X^{\mathrm{T}}(\ell)\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i}X(\ell)\\
&\quad + u_\alpha^{\mathrm{T}}(\ell)u_\alpha(\ell) - \gamma^2w_\alpha^{\mathrm{T}}(\ell)w_\alpha(\ell)\Big]\mathrm{d}\tau\\
&= X^{\mathrm{T}}(\ell)\mathbb{C}_{0\alpha i}^{\mathrm{T}}Q\mathbb{C}_{0\alpha i}X(\ell) + u_\alpha^{\mathrm{T}}(\ell)u_\alpha(\ell)\\
&\quad - \gamma^2w_\alpha^{\mathrm{T}}(\ell)w_\alpha(\ell).
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\lim_{\Delta\ell\to 0}\frac{1}{\Delta\ell}\Big[\mathbb{E}\Big\{e^{-\delta\Delta\ell}X^{\mathrm{T}}(\ell+\Delta\ell)P_{\alpha i}X(\ell+\Delta\ell)\\
&\quad - X^{\mathrm{T}}(\ell)P_{\alpha i}X(\ell)\Big\}\Big]\\
&= -X^{\mathrm{T}}(\ell)\Big\{[\mathbb{A}_{\alpha i} + \mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i} + \mathbb{E}_{\alpha i}K_{w\alpha i}]^{\mathrm{T}}P_{\alpha i}\\
&\qquad + P_{\alpha i}[\mathbb{A}_{\alpha i} + \mathbb{B}_{\alpha i}K_{\alpha i}\mathbb{C}_{\alpha i} + \mathbb{E}_{\alpha i}K_{w\alpha i}]\\
&\qquad + \sum_{j=1}^{S}\pi_{ij}P_{\alpha j} - \delta P_{\alpha i}\Big\}X(\ell)
\end{aligned}$$

Thus, the matrix $P_{\alpha i}$ calculated from (13) is equal to the solution by solving (11). This completes the proof. ■

Appendix C
Proof of Theorem 3

Proof: For the system (9), $V(X(\ell), i)$ is selected as the overall stochastic quadratic Lyapunov function. Consider the weak infinitesimal operator $\mathcal{L}$, we have

$$\begin{aligned}
\mathcal{L}V(X(\ell), i) &= X^{\mathrm{T}}(\ell)\Big[\Big[\tilde{\mathbb{A}}_i + \bar{\mathbb{B}}_iK_i\bar{\mathbb{C}}_i + \bar{\mathbb{E}}_iK_{wi}\Big]^{\mathrm{T}}P_i\\
&\quad + P_i\Big[\tilde{\mathbb{A}}_i + \bar{\mathbb{B}}_iK_i\bar{\mathbb{C}}_i + \bar{\mathbb{E}}_iK_{wi}\Big] + \sum_{j=1, j\neq i}^{S}\pi_{ij}P_j\Big]X(\ell). \qquad (15)
\end{aligned}$$

According to the optimal principle, we can deduce that

$$\begin{aligned}
\mathcal{L}V(X(\ell), i) &= X^{\mathrm{T}}(\ell)\Big[\Big(\tilde{\mathbb{A}}_i - 0.5\delta I\Big)^{\mathrm{T}}P_i + P_i\Big(\tilde{\mathbb{A}}_i - 0.5\delta I\Big)\\
&\quad - \bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\big(\bar{K}_i\bar{\mathbb{C}}_i + L_i\big) - \big(\bar{K}_i\bar{\mathbb{C}}_i + L_i\big)^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i\\
&\quad + 2\gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi} + \sum_{j=1, j\neq i}^{S}\pi_{ij}P_j\Big]X(\ell)\\
&= -X^{\mathrm{T}}(\ell)[\bar{\mathbb{C}}_{0i}^{\mathrm{T}}Q\bar{\mathbb{C}}_{0i} + \bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}]X(\ell). \qquad (16)
\end{aligned}$$

By using Lemmas 1 and 2, there exists $\sigma \in (0, 2)$ satisfying

$$\begin{aligned}
&X^{\mathrm{T}}(\ell)\Big[\bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big]X(\ell)\\
&= \sigma X^{\mathrm{T}}(\ell)\Big[\bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}K_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}K_{wi}\Big]X(\ell). \qquad (17)
\end{aligned}$$

Then, combining (15)–(17) can further become

$$\begin{aligned}
&\mathcal{L}V(X(\ell), i)\\
&= -2X^{\mathrm{T}}(\ell)\Big[\Big(\bar{\mathbb{C}}_i^{\mathrm{T}}K_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2K_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big)\\
&\qquad - \Big(\bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big)\Big]X(\ell)\\
&\quad - 2X^{\mathrm{T}}(\ell)L_i^{\mathrm{T}}L_iX(\ell)\\
&\quad - X^{\mathrm{T}}(\ell)\Big[\bar{\mathbb{C}}_{0i}^{\mathrm{T}}Q\bar{\mathbb{C}}_{0i} + \bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big]X(\ell)\\
&\le -2X^{\mathrm{T}}(\ell)\Big[\Big(\bar{\mathbb{C}}_i^{\mathrm{T}}K_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2K_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big)\\
&\qquad - \Big(\bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big)\Big]X(\ell)\\
&\quad - X^{\mathrm{T}}(\ell)\Big[\bar{\mathbb{C}}_{0i}^{\mathrm{T}}Q\bar{\mathbb{C}}_{0i} + \bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big]X(\ell)\\
&= -\frac{1}{\sigma}\Big\{2X^{\mathrm{T}}(\ell)\Big[\Big(\bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big)
\end{aligned}$$

$$
\begin{aligned}
&- 2\sigma\Big(\bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big)\Big]X(\ell)\\
&+ \sigma X^{\mathrm{T}}(\ell)\Big[\bar{\mathbb{C}}_{0i}^{\mathrm{T}}Q\bar{\mathbb{C}}_{0i} + \bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big]X(\ell)\Big\}\\
&= -\frac{1}{\rho}X^{\mathrm{T}}(\ell)\Big[\bar{\mathbb{C}}_{0i}^{\mathrm{T}}Q_1\bar{\mathbb{C}}_{0i} + \bar{\mathbb{C}}_i^{\mathrm{T}}\bar{K}_i^{\mathrm{T}}\bar{K}_i\bar{\mathbb{C}}_i - \gamma^2\bar{K}_{wi}^{\mathrm{T}}\bar{K}_{wi}\Big]X(\ell)
\end{aligned} \tag{18}
$$

where $\rho = \sigma/(2-\sigma)$ and $Q_1 = \rho Q$.

When $w(\ell) \equiv 0$, there is $\mathcal{L}V(X(\ell), i) \le 0$, which ensures the stochastically asymptotic stability of the system.

Multiplying $e^{-\delta\ell}$ to both sides of (18) and integrating both sides in [0, T], then, one can deduce that

$$
\begin{aligned}
&\mathbb{E}\Big\{\int_0^{\mathrm{T}} e^{-\delta\ell}\big(X^{\mathrm{T}}(\ell)\bar{\mathbb{C}}_{0i}^{\mathrm{T}}Q_1\bar{\mathbb{C}}_{0i}X(\ell) + u^{\mathrm{T}}(\ell)u(\ell)\big)d\ell\Big\}\\
&\le \mathbb{E}\Big\{\gamma^2\int_0^{\mathrm{T}} e^{-\delta\ell}w^{\mathrm{T}}(\ell)w(\ell)d\ell\Big\}
\end{aligned}
$$

taking $T \to \infty$, which matches the Definition 2. This completes the proof. ■

References

[1] P. Shi, Y. Xia, G. Liu, and D. Rees, "On designing of sliding-mode control for stochastic jump systems," *IEEE Trans. Autom. Control*, vol. 51, no. 1, pp. 97–103, Jan. 2006.

[2] P. Shi, E.-K. Boukas, and R. K. Agarwal, "Control of Markovian jump discrete-time systems with norm bounded uncertainty and unknown delay," *IEEE Trans. Autom. Control*, vol. 44, no. 11, pp. 2139–2144, Nov. 1999.

[3] P. Shi, M. Liu, and L. Zhang, "Fault-tolerant sliding-mode-observer synthesis of Markovian jump systems using quantized measurements," *IEEE Trans. Ind. Electron.*, vol. 62, no. 9, pp. 5910–5918, Sep. 2015.

[4] Z.-G. Wu, S. Dong, H. Su, and C. Li, "Asynchronous dissipative control for fuzzy Markov jump systems," *IEEE Trans. Cybern.*, vol. 48, no. 8, pp. 2426–2436, Aug. 2018.

[5] D. Yao, B. Zhang, P. Li, and H. Li, "Event-triggered sliding mode control of discrete-time Markov jump systems," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 49, no. 10, pp. 2016–2025, Oct. 2019.

[6] G. W. Gabriel and J. C. Geromel, "Performance evaluation of sampled-data control of Markov jump linear systems," *Automatica*, vol. 86, pp. 212–215, Dec. 2017.

[7] H. Shen, X. Hu, J. Wang, J. Cao, and W. Qian, "Non-fragile H_∞ Synchronization for Markov jump singularly perturbed coupled neural networks subject to double-layer switching regulation," *IEEE Trans. Netw. Learn. Syst.*, early access, Sep. 6, 2021, doi: 10.1109/TNNLS.2021.3107607.

[8] L. Zhang, B. Cai, and Y. Shi, "Stabilization of hidden semi-Markov jump systems: Emission probability approach," *Automatica*, vol. 101, pp. 87–95, Mar. 2019.

[9] C. Zhang, J. Hu, J. Qiu, and Q. Chen, "Reliable output feedback control for T-S fuzzy systems with decentralized event triggering communication and actuator failures," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2592–2602, Sep. 2017.

[10] Y. Xue, B.-C. Zheng, and X. Yu, "Robust sliding mode control for T-S fuzzy systems via quantized state feedback," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 2261–2272, Aug. 2018.

[11] J. Wang, C. Yang, J. Xia, Z.-G. Wu, and H. Shen, "Observer-based sliding mode control for networked fuzzy singularly perturbed systems under weighted try-once-discard protocol," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 6, pp. 1889–1899, Jun. 2022.

[12] R. Kavikumar, R. Sakthivel, O. Kwon, and B. Kaviarasan, "Reliable non-fragile memory state feedback controller design for fuzzy Markov jump systems," *Nonlinear Anal. Hybrid. Syst.*, vol. 35, Feb. 2020, Art. no. 100828.

[13] S. Dong, C. L. P. Chen, M. Fang, and Z.-G. Wu, "Dissipativity-based asynchronous fuzzy sliding mode control for T–S fuzzy hidden Markov jump systems," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 4020–4030, Sep. 2020.

[14] M. Zhang, P. Shi, L. Ma, J. Cai, and H. Su, "Network-based fuzzy control for nonlinear Markov jump systems subject to quantization and dropout compensation," *Fuzzy Sets Syst.*, vol. 371, pp. 96–109, Sep. 2019.

[15] J. Tao, R. Lu, H. Su, P. Shi, and Z.-G. Wu, "Asynchronous filtering of nonlinear Markov jump systems with randomly occurred quantization via T–S fuzzy models," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 1866–1877, Aug. 2018.

[16] B. Cai, R. Weng, R. Zhang, Y. Liang, and L. Zhang, "Stabilization of a class of fuzzy stochastic jump systems with partial information on jump and sojourn parameters," *Sci. China Technol. Sci.*, vol. 64, no. 2, pp. 353–363, 2021.

[17] Z.-G. Wu, S. Dong, P. Shi, H. Su, T. Huang, and R. Lu, "Fuzzy-model-based nonfragile guaranteed cost control of nonlinear Markov jump systems," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 47, no. 8, pp. 2388–2397, Aug. 2017.

[18] Y. Wang, C. K. Ahn, H. Yan, and S. Xie, "Fuzzy control and filtering for nonlinear singularly perturbed Markov jump systems," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 297–308, Jan. 2021.

[19] H. Ren, G. Zong, and H. R. Karimi, "Asynchronous finite-time filtering of Markov jump nonlinear systems and its applications," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 3, pp. 1725–1734, Mar. 2021.

[20] S.-M. Chen and S.-W. Chen, "Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups and the probabilities of trends of fuzzy logical relationships," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 391–403, Mar. 2015.

[21] C. Mu and Y. Zhang, "Learning-based robust tracking control of quadrotor with time-varying and coupling uncertainties," *IEEE Trans. Netw. Learn. Syst.*, vol. 31, no. 1, pp. 259–273, Jan. 2020.

[22] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.

[23] H. Modares, F. L. Lewis, and Z.-P. Jiang, "Optimal output-feedback control of unknown continuous-time linear systems using off-policy reinforcement learning," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2401–2410, Nov. 2016.

[24] L. Moussaoui, S. Aouaouda, M. Chadli, O. Bouhali, and I. Righi, "State and output feedback control for constrained discrete-time nonlinear systems," *Eur. J. Control*, vol. 50, pp. 79–87, Nov. 2019.

[25] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 14–25, Feb. 2011.

[26] T. Bian, Y. Jiang, and Z.-P. Jiang, "Adaptive dynamic programming for stochastic systems with state and control dependent noise," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 4170–4175, Dec. 2016.

[27] K. Zhang, R. Su, H. Zhang, and Y. Tian, "Adaptive resilient event-triggered control design of autonomous vehicles with an iterative single critic learning framework," *IEEE Trans. Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5502–5511, Dec. 2021.

[28] H. Modares, F. L. Lewis, and M.-B. Naghibi-Sistani, "Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems," *Automatica*, vol. 50, no. 1, pp. 193–202, 2014.

[29] W. Gao and Z.-P. Jiang, "Adaptive dynamic programming and adaptive optimal output regulation of linear systems," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 4164–4169, Dec. 2016.

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

[31] T. Çimen and S. P. Banks, "Nonlinear optimal tracking control with application to super-tankers for autopilot design," *Automatica*, vol. 40, no. 11, pp. 1845–1863, 2004.

[32] M. Mohammadi, M. M. Arefi, P. Setoodeh, and O. Kaynak, "Optimal tracking control based on reinforcement learning value iteration algorithm for time-delayed nonlinear systems with external disturbances and input constraints," *Inf. Sci.*, vol. 554, pp. 84–98, Apr. 2021.

[33] C. Qin, H. Zhang, and Y. Luo, "Optimal tracking control of a class of nonlinear discrete-time switched systems using adaptive dynamic programming," *Neural Comput. Appl.*, vol. 24, no. 3, pp. 531–538, 2014.

[34] K. Zhang, H. Zhang, Y. Mu, and C. Liu, "Decentralized tracking optimization control for partially unknown fuzzy interconnected systems via reinforcement learning method," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 4, pp. 917–926, Apr. 2021.

[35] X. Zhao, X. Wang, G. Zong, and H. Li, "Fuzzy-approximation-based adaptive output-feedback control for uncertain nonsmooth nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3847–3859, Dec. 2018.

[36] M. Chen, S.-Y. Shao, and B. Jiang, "Adaptive neural control of uncertain nonlinear systems using disturbance observer," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3110–3123, Oct. 2017.

[37] H. Shen, M. Xing, H. Yan, and J. Cao, "Observer-based l_2–l_∞ control for singularly perturbed semi-Markov jump systems with improved weighted TOD protocol," *Sci. China Inf. Sci.*, vol. 65, no. 9, Jun. 2022, Art. no. 199204.

[38] L. Zhang, H.-K. Lam, Y. Sun, and H. Liang, "Fault detection for fuzzy semi-Markov jump systems based on interval type-2 fuzzy approach," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 10, pp. 2375–2388, Oct. 2020.

[39] L. Zhang, B. Cai, T. Tan, and Y. Shi, "Stabilization of non-homogeneous hidden semi-Markov jump systems with limited sojourn-time information," *Automatica*, vol. 117, Jul. 2020, Art. no. 108963.

[40] G. Zong, C. Huang, and D. Yang, "Bumpless transfer fault detection for switched systems: A state-dependent switching approach," *Sci. China Inf. Sci.*, vol. 64, no. 7, pp. 1–15, 2021.

[41] G. Zong, D. Yang, J. Lam, and X. Song, "Fault-tolerant control of switched LPV systems: A bumpless transfer approach," *IEEE/ASME Trans. Mech.*, vol. 27, no. 3, pp. 1436–1446, Jun. 2022.

[42] H. Zhang, Y. Liu, and Y. Wang, "Observer-based finite-time adaptive fuzzy control for nontriangular nonlinear systems with full-state constraints," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1110–1120, Mar. 2021.

[43] D. Yang, G. Zong, S. K. Nguang, and X. Zhao, "Bumpless transfer H_∞ anti-disturbance control of switching Markovian LPV systems under the hybrid switching," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 2833–2845, May 2022, doi: 10.1109/TCYB.2020.3024988.

[44] K. Zhang, R. Su, and H. Zhang, "A novel resilient control scheme for a class of Markovian jump systems with partially unknown information," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8191–8200, Aug. 2022.

[45] R. Moghadam and F. L. Lewis, "Output-feedback H_∞ quadratic tracking control of linear systems using reinforcement learning," *Int. J. Adapt. Control Signal Process.*, vol. 33, no. 2, pp. 300–314, 2017.

[46] K. Zhang, H. Zhang, Y. Cai, and R. Su, "Parallel optimal tracking control schemes for mode-dependent control of coupled Markov jump systems via integral RL method," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1332–1342, Jul. 2020.

[47] X. Zhong, H. He, H. Zhang, and Z. Wang, "A neural network based online learning and control approach for Markov jump systems," *Neural Comput.*, vol. 149, pp. 116–123, Feb. 2015.

Jing Wang received the Ph.D. degree in power system and automation from Hohai University, Nanjing, China, in 2019.

She is currently an Associate Professor with the School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan, China. Her current research interests include Markov jump nonlinear systems, singularly perturbed systems, power systems, and nonlinear control.

Jiacheng Wu received the B.Sc. degree in electrical engineering and automation from the Anhui University of Technology, Ma'anshan, China, in 2020, where he is currently pursuing the M.S. degree.

His current research interests include fuzzy systems, complex networks, and robust control.

Jinde Cao (Fellow, IEEE) received the B.S. degree in mathematics/applied mathematics from Anhui Normal University, Wuhu, China, in 1986, the M.S. degree in mathematics/applied mathematics from Yunnan University, Kunming, China, in 1989, and the Ph.D. degree in mathematics/applied mathematics from Sichuan University, Chengdu, China, in 1998.

He is an Endowed Chair Professor, the Dean of the School of Mathematics, the Director of the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence of China, and the Director of the Research Center for Complex Systems and Network Sciences, Southeast University, Nanjing, China, and also with the Yonsei Frontier Lab, Yonsei University, Seoul, South Korea.

Prof. Cao was a recipient of the National Innovation Award of China, the Gold medal of Russian Academy of Natural Sciences, the Obada Prize, and the Highly Cited Researcher Award in Engineering, Computer Science, and Mathematics by Thomson Reuters/Clarivate Analytics. He is elected as a member of the Academy of Europe and the European Academy of Sciences and Arts, a Foreign Member of the Russian Academy of Sciences, the Russian Academy of Engineering, and the Lithuanian Academy of Sciences, a Foreign Fellow of the Pakistan Academy of Sciences, a Fellow of the African Academy of Sciences, and an IASCYS academician.

Mohammed Chadli (Senior Member, IEEE) received the M.Sc.Eng. (DEA) from the Engineering School, INSA-Lyon, Villeurbanne, France, in 1999, the Ph.D. degree in automatic control from the University of Lorraine (UL), CRAN, Nancy, France, in 2002, and the habilitation degree with the University of Picardie Jules Verne (UPJV), Amiens, France, in 2011.

From 2000 to 2004, he was a Lecturer and an Assistant Professor with the UL Institut National Polytechnique de Lorraine, Nancy. Since 2004, he has been an Associate Professor with UPJV. He is currently a Full Professor with the University Paris-Saclay, Univ Evry, IBISC Lab., Evry, France. He was a Visiting Professorship with the Technical University of Ostrava, Ostrava, Czech Republic; Universitetet i Agder, Kristiansand, Norway; Shanghai Maritime University, Shanghai, China, from 2014 to 2017; Nanjing University of Aeronautics and Astronautics, Nanjing, China, from 2018 to 2019; and the University of Naples Federico II, Naples, Italy. He is a coauthor of books and book chapters (Wiley, Springer, and Hermes), and numerous articles published in international journals and conferences. His research interests include fuzzy/LPV and switched systems, singular systems, robust control, fault detection and isolation, fault tolerant control via LMI, SOS, and Lyapunov methods. On the application side, he is mainly interested in automotive control and renewable energy.

Dr. Chadli is an Editor/Associate Editor/Editorial Board Member of several international journals, including the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Automatica*, and was a guest editor for special issues in international journals.

Hao Shen (Member, IEEE) received the Ph.D. degree in control theory and control engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2011.

Since 2011, he has been with Anhui University of Technology, Ma'anshan, China, where he is currently a Professor. His current research interests include stochastic hybrid systems, complex networks, fuzzy systems and control, and nonlinear control.

Dr. Shen was a recipient of the Highly Cited Researcher Award by Clarivate Analytics (formerly, Thomson Reuters) in 2019–2021. He has served on the technical program committee for several international conferences. He is an Associate Editor for several international journals, including *Journal of The Franklin Institute*, *Applied Mathematics and Computation*, and *Neural Processing Letters*.