

Received June 23, 2020, accepted July 2, 2020, date of publication July 10, 2020, date of current version July 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3008399

# A New Reinforcement Learning Based Adaptive Sliding Mode Control Scheme for Free-Floating Space Robotic Manipulator

ZHICHENG XIE<sup>1</sup>, TAO SUN<sup>1</sup>, TREVOR HOCKSUN KWAN<sup>2</sup>, ZHONGCHENG MU<sup>3</sup>,  
AND XIAOFENG WU<sup>1</sup>, (Senior Member, IEEE)

<sup>1</sup>School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Sydney, NSW 2006, Australia

<sup>2</sup>Department of Thermal Science and Energy Engineering, University of Science and Technology of China, Hefei 230052, China

<sup>3</sup>School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: Xiaofeng Wu (xiaofeng.wu@sydney.edu.au)

**ABSTRACT** This paper presents a new adaptive small chattering sliding mode control (SCSMC) scheme that uses reinforcement learning (RL) and time-delay estimation (TDE) for the motion control of free-floating space robotic manipulators (FSRM) subject to model uncertainty and external disturbance. The proposed sliding mode control scheme can achieve small chattering effects and improve the tracking accuracy by using a new adaptive law for the switching gain and a RL-based robust term to handle the control inputs. In SCSMC, the complicated multiple-input-multiple-output (MIMO) uncertain system of FSRM is transformed into multiple single-input-single-output (SISO) known subsystems with bounded estimation errors by the TDE technique and state feedback compensation. Subsequently, once the sliding variable is inside the designed manifold, the derivative of the switching gain for each subsystem becomes a negative hyperbolic tangent function of the associated sliding variable, which offers the ability to reduce chattering by decreasing the switching gain. Moreover, the RL based robust term for each subsystem is designed to avoid the loss of tracking accuracy caused by the aforementioned switching gain drop. The tracking errors are proven to be uniformly-ultimately-bounded (UUB) with an arbitrarily small bound by using the Lyapunov theory. The effectiveness of the proposed control scheme is verified by numerical simulations.


**INDEX TERMS** Reinforcement learning, sliding mode control, space robotic manipulator, time delay estimation.

## I. INTRODUCTION

Free-floating space robotic manipulators (FSRM) can assist or even substitute astronauts to perform various extravehicular activities (EVAs) such as capturing space debris and maintenance of space structures [1]. To satisfyingly perform such space missions, FSRMs should be controlled to accurately track the desired trajectories. However, FSRMs involve complicated dynamics because the satellite posture (orientation and position) can be easily disturbed by the motion of mounted robotic manipulators, which brings up the challenges of designing the required controller [2], [3]. Moreover, high tracking accuracy ( $<10^{-3}$  radians) under the presence of system uncertainties and unknown external disturbances is also critically important in FSRMs to

successfully guarantee the docking of space objects or establish communication links. Hence, precise motion control of FSRMs remains a hotspot of research.

Various controllers designed for ground manipulators can be directly applied to FSRMs if the system parameters are exactly known [4]. Moreover, the known parameters allow the use of kinematic control methods such as the generalized-Jacobian-matrix based motion rate control scheme [5], [6]. Nevertheless, these parameters are difficult to measure or are even unavailable in practice, which may be because of dynamic mass reductions from fuel consumption [7], unknown payload size during target capturing [8], and the challenges in precisely modeling the physical nonlinearities [9]. Furthermore, the dynamic model of FSRMs cannot be linearly parameterized because of the free-floating base [10], [11], which means the adaptive controllers for linearly parameterized models such as [12] are inapplicable.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhuang Xu .

Hence, many research articles regarding motion control of FSRMs focus on handling nonlinear parameter uncertainties such as [13]–[15].

Many algorithms have been proposed to cope with external disturbances and unmodelled nonlinearities, including neural networks (NN) [16]–[23], fuzzy logic approximators (FLA) [7], [24]–[27] and adaptive disturbance observers (ADO) [28]–[32]. Such algorithms have a stronger capability for motion control of FSRMs over control schemes [13]–[15] that can only handle parameter uncertainties. More precisely, Jia and Shan [16] proposed a finite-time terminal sliding mode controller for space manipulators in which a radius basis function (RBF) neural network is used to compensate model uncertainties. Wang *et al.* [25] proposed the FLA algorithm for uncertain surface vehicles, which self-constructs highly interpretable T-S fuzzy rules by using a decoupled distance measure to manage (add or delete) the fuzzy sets in each dimension of the fuzzy inputs. Chu *et al.* [7] designed a backstepping-formed ADO based controller for FSRMs driven by DC motors, which applied the FLA to compensate model uncertainties and external disturbances.

The time delay estimation (TDE) technique [33]–[35] can estimate system uncertainties and unknown disturbances by using the system states and control inputs in the last sampling instant. Compared to neural networks, FLAs, and adaptive observers, TDE can be easily applied to the control of FSRMs because it does not require any tedious preparation, such as designing the structure of neural networks, setting the fuzzy sets of FLAs or initializing the states of disturbance observers. Although TDE invariably causes estimation errors due to the delay of one sampling step, these errors as applied on Euler-Lagrange systems (e.g. ground manipulators and FSRMs) can be bounded by using a small sampling time and accurate measurements on the inertia matrix [36].

Sliding mode control (SMC), which has many variants including the fast terminal sliding mode (FTSM) [37]–[39], the super-twisting sliding mode [40], [41] and integral sliding mode (ISM) [42], is robust to system uncertainties and unknown disturbances. Therefore, SMC can deal with the bounded estimation errors of various estimation techniques including TDE, neural networks, and the FLA. To achieve the asymptotic stability under the presence of bounded estimation errors, many SMC schemes require either a monotonically increasing switching gain [43]–[45] or a conservative constant switching gain that is greater than the known upper bound of estimation errors [46]–[48]. However, over-estimating the switching gains will invariably lead to chattering effects, which will decrease the tracking accuracy and require the use of large actuator inputs (thus wasting fuel). Hence, many efforts have been made to handle the issues caused by over-estimated switching gains [49]–[54]. For example, Roy and Kar [49] designed an adaptive time-delayed SMC scheme to alleviate the over- and under-estimation problems of SMC. The TDE based SMC schemes for industrial ground manipulators, which mitigate over-estimated switching gains by considering a small vicinity

of the sliding manifold, can be found in [52] and [53] and had significantly small chattering effects. Notably, the SMC schemes in [49]–[54] can reduce the chattering effects by decreasing the switching gains when the sliding variables are inside a designed vicinity, which obtains the uniformly ultimately bounded (UUB) tracking errors rather than asymptotical stable tracking errors. Although the chattering effects are significantly mitigated, such controllers have the risks of compromising the tracking accuracy due to the excessively small switching gains. More precisely, the sliding variables will leave the sliding manifold if the switching gains are too small to handle TDE errors, so tracking accuracy is, once again, badly affected. Thus, to achieve the chattering reduction while retaining a good tracking accuracy, the SMC scheme needs to maintain the advantage of existing controllers [43]–[48] (reduction of chattering effects by decreasing switching gain without loss of stability) while preventing the sliding variables from leaving the manifold when the switching gains are small.

Reinforcement learning (RL) is an artificial intelligence technique that gradually explores the optimal policy by interacting with the environment [55]. This technique imitates the learning process of human brains and has been widely applied in the field of nonlinear control [56]–[60]. The nature of RL is to determine the optimal policy (actions) that either maximizes the cumulative reward or minimizes the cost during the entire learning process. The process of RL starts with applying an action for the initial state according to the initial policy, and the current state is transformed into the next state by the chosen action. Subsequently, the instant reward (or cost) of the action from the current state will be given in the next state to evaluate the new appropriate actions and corresponding policy. This process is repeated until the optimal state is obtained (determined by some tolerance criteria).

Inspired by the RL algorithm's high performance, this paper proposes to use the RL and TDE techniques to resolve the technical limitations of the SMC. Thus, a new adaptive small chattering sliding mode control (SCSMC) scheme is formed to achieve motion control of free-floating space robotic manipulators with model uncertainties and external disturbances. In the proposed SCSMC, the complex uncertain multiple-input-multiple-output (MIMO) system of FSRMs is transformed into multiple single-input-single-output (SISO) known subsystems with bounded estimation errors by the TDE technique and state feedback compensation. In each subsystem, the adaptive law of switching gain is positively proportional to the absolute values of the sliding variable at the beginning of control, which allows the sliding variable to enter the designed vicinity of sliding manifold within a finite time. After that, the adaptive law is switched to a negative hyperbolic tangent function of the associated sliding variable once it is inside the designed manifold, which offers higher chattering reductions by decreasing the switching gains. Subsequently, when the sliding variable leaves the vicinity or the switching gain is no greater than 0, the adaptive law is switched to become positive to steer the sliding variable

back into the vicinity. To mitigate the loss of tracking accuracy resulting from decreased switching gains, the RL based robust term within the control inputs is designed. More precisely, on the one hand, when the sliding variable drifts away from the sliding manifold and moves to the boundary of the designed vicinity (the sliding variable is still inside the vicinity) due to the decreased switching gain insufficient to overcome TDE errors, the RL based robust term can offer the great control input to prevent sliding variables from moving out the given vicinity. On the other hand, when the sliding variable is close to the manifold, the values of the robust terms are mainly determined by RL that learns to achieve the desired dynamics of sliding variables (chattering-free and drifting-free). As a result, the proposed SCSMC can actively prevent the sliding variables moving out the vicinities and attenuate chattering effects of sliding variables near the manifold, which leads to reducing the chattering and improving the tracking accuracy. It is proven through the Lyapunov theory that the tracking errors are uniformly ultimately bounded (UUB), and the effectiveness is verified by numerical simulations including the comparison to the existing approaches. The contributions of this paper are listed as follows: (1) The novel adaptive law of switching gain mitigates chattering effects by decreasing switching gains without loss of stability. (2) The side effects of decreasing the switching gains have been mitigated by applying the novel RL based robust term, so a high tracking accuracy is obtained even under extreme disturbance conditions. In other words, the proposed control scheme has the advantage of existing control methods [49]–[54] (attenuation of the chattering effects by decreasing the switching gain without loss of stability) while also having a better tracking accuracy.

The rest of this paper is organized as follows. In section 2, the preliminaries including the dynamic model of FSRMs are introduced. In section 3, the proposed control scheme is detailed including the stability analysis. Simulation results are given in section 4. In section 5, the conclusion is drawn.

## II. PRELIMINARIES

### A. FSRM DYNAMIC MODEL

A  $n$ -links rigid robotic manipulator mounted on a free-floating base is considered in this paper, as shown in Figure 1. The rigid links are connected by revolute joints, and the base (rigid body 0) can be regarded as being connected to the inertia frame by a free joint with 6 degrees of freedom (DOF).  $\sum E$  is the coordinate of the end-effector (EE).  $\sum B_i$  and  $C_i$  ( $i = 0, 1, \dots, n$ ) respectively denote the local co-ordinate and the centre of mass (CM) of the  $i^{th}$  rigid body.  $J_i$  ( $i = 1, \dots, n$ ) is the joint of the  $i^{th}$  rigid body. It is assumed that the CM of the  $i^{th}$  rigid body, which is located by  $a_i$  and  $b_i$  ( $i = 0, 1, 2, \dots, n$ ), remains fixed in the local frame.

The dynamic model of FSRMs can be presented in the following typical form [10], [17], [61]–[63]:

$$M(\theta)\ddot{\theta} + C(\theta, \dot{\theta})\dot{\theta} + d = \tau \quad (1)$$

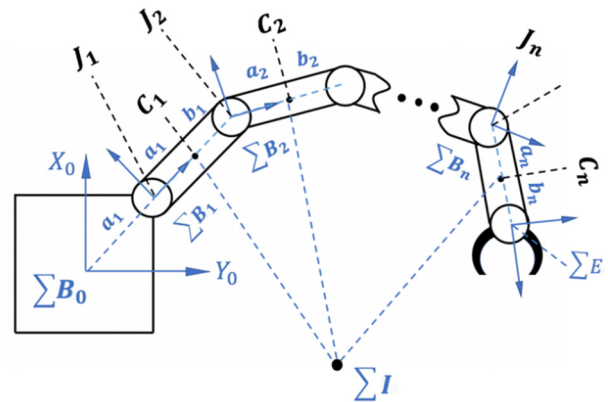


FIGURE 1. The dynamic model of the free-floating space robotic manipulator (FSRM).

where  $\theta = [\theta_1, \theta_2, \dots, \theta_n]^T \in R^{n \times 1}$  is the vector of joints of the manipulator;  $M(\theta) \in R^{n \times n}$  is the uniformly positive definite and symmetric inertia matrix;  $C(\theta, \dot{\theta}) \in R^{n \times n}$  is the skew symmetric matrix;  $d \in R^{n \times 1}$  is the vector of generalized external disturbance;  $\tau \in R^{n \times 1}$  is the vector of control input torques to drive the joints of the manipulator, and  $n$  is the number of joints of the manipulator. To facilitate the design of the controller, the dynamics model of (1) can be re-written as follows.

$$\begin{aligned} \ddot{\theta} &= \widehat{M}^{-1}(\theta) \{ \tau + [\widehat{M}(\theta) - M(\theta)]\ddot{\theta} - C(\theta, \dot{\theta})\dot{\theta} - d \} \\ &= \widehat{M}^{-1}(\theta) \tau + H \end{aligned} \quad (2)$$

where the matrix  $\widehat{M}$  is the nominal part of  $M$ . The vector  $H = -\widehat{M}^{-1}(\theta) [C(\theta, \dot{\theta})\dot{\theta} + d] + [M^{-1}(\theta) - \widehat{M}^{-1}(\theta)] \tau$  is the lumped uncertainty consisting of model uncertainty and external disturbance.

*Remark 1:* Dynamic model (2) is a Euler-Lagrange system as defined in [36] in which the TDE technique can be applied, and TDE errors can be bounded by appropriate  $\widehat{M}$  and sampling times.

### B. FUNDAMENTAL FACT

*Definition 1* [64]: Consider the nonlinear dynamic system  $\dot{x}(t) = f(x(t))$ ,  $x(0) = x_0$ . Uniformly bounded with ultimate bound  $B$  if there exist positive constants  $B$  and  $C$ , as well as  $T = T(A, B)$  independent of  $t_0 \geq 0$ , for every  $A \in (0, C)$ , such that  $x(t_0) \leq A \implies x(t) \leq B, \forall t \geq t_0 + T$ .

## III. SMALL CHATTERING SLIDING MODE CONTROL SCHEME

### A. FORMULATION

The sliding variable is defined:

$$s = \dot{e} + K_s e \quad (3)$$

where the tracking error  $e = \theta - \theta_r \in R^{n \times 1}$ .  $\theta_r \in R^{n \times 1}$  is the vector of references of the manipulator joints.  $K_s = \text{diag}(k_{s1}, k_{s2}, \dots, k_{sn}) \in R^{n \times n}$  is the diagonal matrix where  $k_{si} > 0$  for all  $i = 1, 2, \dots, n$ .

The TDE technique is used to estimate the lumped uncertainty in (2).

$$\widehat{H}(t) = \ddot{\theta}(t - T_s) - \widehat{M}^{-1}(\theta(t))\tau(t - T_s) \quad (4)$$

where  $T_s$  is the sampling time, and  $\widehat{H}(t) \in R^{n \times 1}$  is the estimated vector of  $H(t) = \ddot{\theta}(t) - \widehat{M}^{-1}(\theta(t))\tau(t) \in R^{n \times 1}$ . The estimation errors  $\varepsilon$  therefore exist:

$$\varepsilon(t) = H(t) - \widehat{H}(t) \quad (5)$$

However, the estimation errors  $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T \in R^{n \times 1}$  can be bounded by selecting an appropriately small size for  $T_s$  and sufficiently accurate  $\widehat{M}$ , shown in lemma 1.

*Lemma 1 [36]:* The estimation errors  $\varepsilon$  in (5) will remain bounded, which is shown in (7) for system (2) where the uncertainty is estimated by the TDE technique (4) if the following condition (6) holds:

$$\|M^{-1}(\theta)\widehat{M}(\theta) - I\| < 1 \quad (6)$$

$$\|\varepsilon\| \leq \bar{\varepsilon} \quad (7)$$

where  $I$  is the identity matrix with the appropriate size.  $\bar{\varepsilon}$  is an unknown positive constant. It is easy to conclude from (7) that  $|\varepsilon_i| \leq \bar{\varepsilon}_i$  and  $\bar{\varepsilon}_i \geq 0$  holds for all  $i = 1, 2, \dots, n$ .

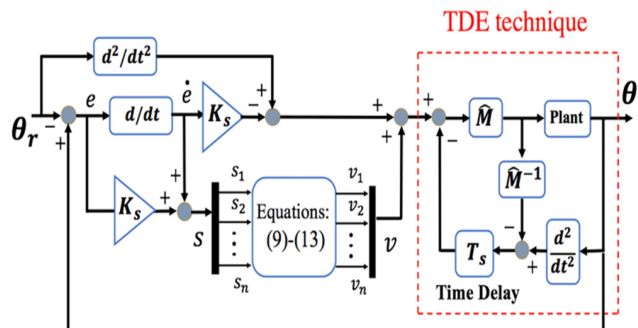


FIGURE 2. Block diagram of the proposed SCSMC.

Therefore, as illustrated in figure 2, the TDE equipped SCSMC scheme is proposed as (8)-(13):

$$\tau = \widehat{M}(\theta)(v - \widehat{H} + \ddot{\theta}_r - K_s \dot{e}) \quad (8)$$

where  $v = [v_1, v_2, \dots, v_n]^T \in R^{n \times 1}$  is the virtual control input, shown explicitly in (9), and  $\beta = \text{diag}(\beta_1, \beta_2, \dots, \beta_n) \in R^{n \times 1}$  is the diagonal matrix where  $\beta_i > 0$  for all  $i = 1, 2, \dots, n$ .

$$v_i = \begin{cases} -\beta_i s_i - \frac{e^{\widehat{k}_i}}{s_i} \mathcal{F}(\widehat{k}_i) - \widehat{k}_i \text{sgn}(s_i), & \text{if } (|s_i| > D_i) \text{ or } (\widehat{k}_i \leq 0) \\ -\beta_i s_i - \frac{e^{\widehat{k}_i}}{s_i} y_i \ln\left(1 + \frac{|s_i|}{e^{\widehat{k}_i}}\right) - \widehat{k}_i \text{sgn}(s_i), & \text{else} \end{cases} \quad (9)$$

where  $D_i > 0$  is the designed vicinity of the sliding manifold for each subsystem ( $i = 1, 2, \dots, n$ ).  $\mathcal{F}(\widehat{k}_i)$  is the switching function that is defined in (11).  $y_i \geq 0$  is a positive function

shown in (12).  $\widehat{k}_i$  is the switching gain, and the adaptive law is presented in (10):

$$\dot{\widehat{k}}_i = \begin{cases} \gamma_i |s_i|, & \text{if } (|s_i| \geq D_i) \text{ or } (\widehat{k}_i \leq 0) \\ -\gamma_i [\mu_i^2 + \frac{\alpha_i}{\kappa_V} \tanh\left(\frac{\alpha_i}{|s_i| + \mu_i^2}\right)], & \text{else} \end{cases} \quad (10)$$

where  $\gamma_i > 0$ ,  $\alpha_i > 0$  are positive constants.  $\mu_i$  is a small constant close to zero.  $\kappa_V$  is a positive constant that satisfies  $\kappa_V = e^{-(\kappa_V + 1)}$ .

*Remark 2:* It can be concluded that  $\widehat{k}_i \geq 0$  holds because the non-positive value of  $\widehat{k}_i$  will result in the positive derivative  $\dot{\widehat{k}}_i = \gamma_i |s_i|$ .

$$F(\widehat{k}_i) = \begin{cases} y_i \ln\left(1 + \frac{|s_i|}{e^{\widehat{k}_i}}\right), & \text{if } \widehat{k}_i < \bar{k}_i \\ \Omega_i, & \text{if } \widehat{k}_i \geq \bar{k}_i. \end{cases} \quad (11)$$

where  $\Omega_i > 0$  is a small positive constant and  $\bar{k}_i > 0$  is a great positive constant.

$$y_i = [(1 - \omega_i) + \omega_i \cdot \Psi(s_i)] \frac{\Phi_i}{\frac{e^{\widehat{k}_i}}{D_i} \ln\left(1 + \frac{D_i}{e^{\widehat{k}_i}}\right)} \quad (12a)$$

$$\Psi(s_i) = \begin{cases} \sin\left(\frac{\pi |s_i|}{2D_i}\right), & \text{if } |s_i| \leq D_i \\ 1, & \text{if } |s_i| > D_i \end{cases} \quad (12b)$$

where  $0 \leq \omega_i \leq 1$  is a positive parameter that is determined by the RL, which will be detailed later. The parameter  $\Phi_i \geq 0$  that starts at zero ( $\Phi_i(t=0) = 0$ ) is discontinuously and monotonically increasing, and it is responsible for preventing the sliding variable from departing the vicinity. The adaptive law of  $\Phi_i$  is shown as follows:

$$\dot{\Phi}_i = \begin{cases} \gamma_i \frac{\alpha_i}{2\kappa_V} \tanh\left(\frac{\alpha_i}{|s_i|}\right) \Pi_i, & \text{if } (|s_i| \leq D_i) \text{ and } (\widehat{k}_i > 0) \\ 0, & \text{else} \end{cases} \quad (13)$$

where  $\Pi_i = (1 - \text{sgn}(\Phi_i - \Phi_i^*))$ .  $\Phi_i^* > 0$  is a positive parameter defining the upper bound of  $\Phi_i$ . In this paper, we design  $\Phi_i^*(t)$  as the value of  $\widehat{k}_i(t_i^*) + \Phi_i(t_i^*)$  in the last time step  $t_i^* (t_i^* < t)$  when the sliding variable  $s_i$  enters the vicinity ( $|s(t_i^*)| = D_i$ ).

The MIMO system of FSRMs (2) can be transformed into  $n$  SISO systems by applying control law (8). The derivative of the sliding variable for each subsystem can then be written as:

$$\dot{s}_i = v_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (14)$$

*Lemma 2:* For subsystems  $i = 1, 2, \dots, n$ , if a period  $t \in [t_{in}, t_{out}]$  exists such that  $|s_i(t = t_{in})| = |s_i(t = t_{out})| = D_i$  and  $|s_i(t_{in} < t < t_{out})| < D_i$ , the inequality of virtual control input,  $|v_i(t = t_{out})| \geq |v_i(t = t_{in})|$ , will hold.

*Proof:* According to (10) and (13), the adaptive law of  $\Phi_i$  is opposite to that of  $\widehat{k}_i$  as long as both conditions  $|s_i| < D_i$



and  $\widehat{k}_i > 0$  hold. It means that the decreased value  $\Delta\widehat{k}_i$  is taken from  $\widehat{k}_i$  and allocated to  $\Phi_i$  within the vicinity:

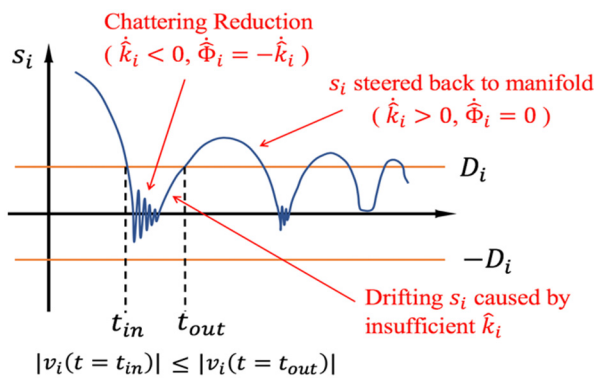
$$\Phi_i(t_{out}) = \begin{cases} \Phi_i(t_{in}) + \widehat{k}_i(t_{in}) - \widehat{k}_i(t_{out}), & \text{if } \widehat{k}_i(t_{in} < t < t_{out}) > 0 \\ \Phi_i(t_{in}) + \widehat{k}_i(t_{in}), & \text{else} \end{cases} \quad (15)$$

According to (9), (12) and (15), the difference of virtual control input between the time instants of  $t_{in}$  and  $t_{out}$  is:

$$\begin{aligned} & |v_i(t_{out})| - |v_i(t_{in})| \\ &= \frac{e^{\widehat{k}_i(t_{out})}}{D_i} y_i(t_{out}) \ln\left(1 + \frac{|D_i|}{e^{\widehat{k}_i(t_{out})}}\right) + \widehat{k}_i(t_{out}) \\ &\quad - \frac{e^{\widehat{k}_i(t_{in})}}{D_i} y_i(t_{in}) \ln\left(1 + \frac{|D_i|}{e^{\widehat{k}_i(t_{in})}}\right) - \widehat{k}_i(t_{in}) \\ &= \Phi_i(t_{out}) - \Phi_i(t_{in}) + \widehat{k}_i(t_{out}) - \widehat{k}_i(t_{in}) \\ &= \begin{cases} 0, & \text{if } \widehat{k}_i(t_{in} < t < t_{out}) > 0 \\ \widehat{k}_i(t_{out}) \geq 0, & \text{else} \end{cases} \quad (16) \end{aligned}$$

*Proof complete.*

*Remark 3:* According to lemma2, when considering the vicinity of the sliding manifold ( $|s_i| \leq D_i$ ), the switching gain  $\widehat{k}_i$  decreases to reduce the chattering effects, while the decreased value of  $\widehat{k}_i$  is transferred to  $\Phi_i$ . If the sliding variable drifts away from the manifold, the greatest value of  $y_i$  will be obtained when the sliding variable arrives at the boundary of the vicinity ( $|s_i| = D_i$ ) to steer the sliding variable back into the manifold. In other words,  $\Phi_i$  preserves the decreased value  $\Delta\widehat{k}_i$  and returns it to the control input  $v_i$  when the sliding variable attempts to leave the designed vicinity. Therefore, the proposed SCSMC can prevent the sliding variable from departing the vicinity and alleviate the problem caused by an under-estimated switching gain. This advantage is illustrated in figure 3:



**FIGURE 3.** Possible dynamics of the sliding variables under the proposed SCSMC.

*Lemma 3 [65]:* According to L’Ho^pital’s Rule, the function  $f(x) = x \cdot \ln(1 + \frac{1}{x})$  is monotonically increasing as long

as  $x \in (0, +\infty)$ , and the following equation holds:

$$\lim_{x \rightarrow \infty} x \cdot \ln(1 + \frac{1}{x}) = 1 \quad (17)$$

(18) can be obtained by replacing  $x$  in (17) with:  $\frac{\widehat{k}_i}{s_i}$

$$\lim_{|s_i| \rightarrow 0} \frac{e^{\widehat{k}_i}}{|s_i|} \cdot \ln(1 + \frac{|s_i|}{e^{\widehat{k}_i}}) = 1 \quad (18)$$

*Remark 4:* According to (18), It is easy to conclude that  $\lim_{|s_i| \rightarrow 0} [-\frac{e^{\widehat{k}_i}}{s_i} y_i \ln(1 + \frac{|s_i|}{e^{\widehat{k}_i}})] = \pm y_i$ . According to (9) and (11), It is easy to conclude that  $|s_i| \geq D_i$  will hold if  $-\frac{e^{\widehat{k}_i}}{s_i} \mathcal{F}(\widehat{k}_i) = \frac{e^{\widehat{k}_i}}{s_i} \Omega_i$ . This is because  $\mathcal{F}(\widehat{k}_i) = \Omega_i$  when  $\widehat{k}_i \geq \bar{k}_i > 0$  that means  $\widehat{k}_i \leq 0$  does not hold in (9), which leads to  $|s_i| \geq D_i$ . Therefore, the robust terms  $-\frac{e^{\widehat{k}_i}}{s_i} \mathcal{F}(\widehat{k}_i)$  and  $-\frac{e^{\widehat{k}_i}}{s_i} y_i \ln(1 + \frac{|s_i|}{e^{\widehat{k}_i}})$  in virtual control input (9) will not cause the problem of singularity when  $|s_i| \rightarrow 0$ .

*Lemma 4 [66]:* the following inequality holds for any  $A > 0$  and for any  $B \in R$ :

$$0 \leq |B| - B \cdot \tanh(\frac{B}{A}) \leq \kappa_V A \quad (19)$$

where  $\kappa_V$  is a positive constant that satisfies  $\kappa_V = e^{-(\kappa_V+1)}$ . The inequality (20) can be derived from (19) by substituting  $A$  with  $|s_i| + \mu_i^2$ , and  $B$  with  $\alpha_i$ .

$$|s_i| + \mu_i^2 + \frac{\alpha_i}{\kappa_V} \tanh(\frac{\alpha_i}{|s_i| + \mu_i^2}) \geq \frac{\alpha_i}{\kappa_V} \quad (20)$$

where  $\alpha_i > 0$  and  $\mu_i \rightarrow 0$  are the constants mentioned in (10).

## B. STABILITY ANALYSIS

*Theorem 1:* For a system of FSRMs (2) controlled by (8)-(13), the sliding variable of each subsystem will enter into the designed vicinity of the sliding manifold within a finite time,  $|s_i(t)| \leq D_i, t \geq \mathcal{T}_i$ . After the sliding variable of any subsystem enters its vicinity, the sliding variables and tracking errors of the FSRM are guaranteed to be UUB, proven as follows:

$$\|s(t)\| \leq 2\sqrt{R_1^* + R_2^*}, \quad t \geq \min_{1 \leq i \leq n} \{\mathcal{T}_i\} \quad (21)$$

where  $R_1^*$  and  $R_2^*$  are the 2 positive constants defined as follows.

$$R_1^* = \max\{\frac{n}{2}(\delta_1 + \delta_2)^2, \frac{n}{2} \cdot \frac{\sum_{i=1}^n \bar{\varepsilon}_i(D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})}{\beta_{min}}\} \quad (22a)$$

$$\begin{aligned} R_2^* = \max\{ & \frac{n\|\bar{\varepsilon}\|^2}{2\gamma_{min}} + \delta_3, \frac{n}{2\gamma_{min}}(\bar{k}_{max} - \bar{\varepsilon}_{min})^2, \\ & \frac{n}{2\gamma_{min}}(\ln(\bar{\varepsilon}_{max} + \frac{\kappa_V}{\alpha_{min}} \sum_{i=1}^n \bar{\varepsilon}_i(D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})) - \bar{\varepsilon}_{i^*})^2, \\ & \frac{n}{2\gamma_{min}}(\ln(\frac{\sum_{i=1}^n \bar{\varepsilon}_i(D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})}{\Omega_{min}}) - \bar{\varepsilon}_{i^*})^2\} \quad (22b) \end{aligned}$$

where  $\delta_1 > 0$  is a small positive constant and  $\delta_2 = \max_{1 \leq i \leq n} \{D_i\}$ .  $\beta_{min} = \min_{1 \leq i \leq n} \{\beta_i\}$ ,  $\gamma_{min} = \min_{1 \leq i \leq n} \{\gamma_i\}$ ,  $\bar{k}_{max} = \max_{1 \leq i \leq n} \{\bar{k}_i\}$ ,  $\bar{\varepsilon}_{min} = \min_{1 \leq i \leq n} \{\bar{\varepsilon}_i\}$ , and  $\delta_3 > 0$  is a positive constant.  $\Omega_{min} = \min_{1 \leq i \leq n} \{\Omega_i\}$ ,  $\alpha_{min} = \min_{1 \leq i \leq n} \{\alpha_i\}$ , and  $\bar{\varepsilon}_{max} = \max_{1 \leq i \leq n} \{\bar{\varepsilon}_i\}$ .  $i^*$  is an unknown integral between 1 and  $n$ , which will be explicitly explained in the following proof.

*Proof:* Consider the first Lyapunov function for each subsystem:

$$V_{1,i} = \frac{1}{2} s_i^2 + \frac{1}{2} \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i}, \quad i = 1, 2, \dots, n \quad (23)$$

Taking the derivative for (23) with respect to time, the follows can be achieved:

$$\begin{aligned} \dot{V}_{1,i} &= s_i (\varepsilon_i + v_i) - \frac{1}{\gamma_i} (\bar{\varepsilon}_i - \hat{k}_i) \dot{\hat{k}}_i \\ &= s_i \varepsilon_i + s_i \left[ -\beta_i s_i - \frac{e^{\hat{k}_i}}{s_i} \mathcal{F}(\hat{k}_i) - \hat{k}_i \text{sgn}(s_i) \right] \\ &\quad - (\bar{\varepsilon}_i - \hat{k}_i) |s_i| \quad (\text{when } |s_i| > D_i) \\ &\leq -\beta_i s_i^2 - \frac{e^{\hat{k}_i}}{s_i} \mathcal{F}(\hat{k}_i) \quad (\text{when } |s_i| > D_i) \\ &< -\beta_i D_i^2 < 0 \quad (\text{when } |s_i| > D_i) \end{aligned} \quad (24)$$

*Remark 5:* In the light of (24), It is clear that the sliding variable of each subsystem  $s_i$  will enter into the associated vicinity  $D_i$  within a finite time  $\mathcal{T}_i$ . After that, the sliding variable could leave the vicinity because of the negative adaptive law of switching gain. Therefore, the next step is to prove that the sliding variables are bounded after any subsystem enters the vicinity ( $t \geq \min_{1 \leq i \leq n} \{\mathcal{T}_i\}$ ).

We consider the second Lyapunov function for the entire system:

$$V_2 = \frac{1}{2} \sum_{i=1}^n V_{1,i} = \frac{1}{2} s^T s + \frac{1}{2} \sum_{i=1}^n \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i} \quad (25)$$

Taking the derivative for (25) with respect to time, the follows can be obtained by using (14):

$$\begin{aligned} \dot{V}_2 &= S^T (\hat{M}^{-1} \tau + H + K_s \dot{e} - \ddot{q}_r) - \sum_{i=1}^n \frac{1}{\gamma_i} (\bar{\varepsilon}_i - \hat{k}_i) \dot{\hat{k}}_i \\ &\leq \sum_{i=1}^n |s_i| \bar{\varepsilon}_i + \sum_{i=1}^n s_i v_i - \sum_{i=1}^n \frac{1}{\gamma_i} (\bar{\varepsilon}_i - \hat{k}_i) \dot{\hat{k}}_i \end{aligned} \quad (26)$$

Next, the subsystems ( $i = 1, 2, \dots, n$ ) are separated into 2 sets of integers in which the elements are arranged in numerical order:

$$E_p = \{p \mid |s_p| > D_p \text{ or } \hat{k}_p \leq 0\} \quad (27a)$$

$$E_q = \{q \mid |s_q| \leq D_q \text{ and } \hat{k}_q > 0\} \quad (27b)$$

The number of elements in  $E_p$  is defined as  $n_1$ , and the number of elements in  $E_q$  is defined as  $n_2$ .  $p_0$  and  $q_0$  are defined as the first element in  $E_p$  and  $E_q$  respectively. It is clear that  $n = n_1 + n_2$  holds because the set  $\{1, 2, \dots, n\}$  is

fully composed of  $E_p$  and  $E_q$ . As a result, combining with (9) and (10), (26) can be further written as:

$$\begin{aligned} \dot{V}_2 &\leq \sum_{p=p_0}^{n_1} [|s_p| \bar{\varepsilon}_p + s_p v_p - \frac{1}{\gamma_p} (\bar{\varepsilon}_p - \hat{k}_p) \dot{\hat{k}}_p] \\ &\quad + \sum_{q=q_0}^{n_2} [|s_q| \bar{\varepsilon}_q + s_q v_q - \frac{1}{\gamma_q} (\bar{\varepsilon}_q - \hat{k}_q) \dot{\hat{k}}_q] \\ &== \sum_{p=p_0}^{n_1} [-\beta_p s_p^2 - e^{\hat{k}_p} \cdot \mathcal{F}(\hat{k}_p) + (\bar{\varepsilon}_p - \hat{k}_p) (|s_p| \\ &\quad - \frac{\dot{\hat{k}}_p}{\gamma_p})] + \sum_{q=q_0}^{n_2} [-\beta_q s_q^2 - e^{\hat{k}_q} \cdot y_q \ln \left( 1 + \frac{|s_q|}{e^{\hat{k}_q}} \right) \\ &\quad + (\bar{\varepsilon}_q - \hat{k}_q) (|s_q| - \frac{\dot{\hat{k}}_q}{\gamma_q})] \\ &= \sum_{p=p_0}^{n_1} [-\beta_p s_p^2 - e^{\hat{k}_p} \cdot \mathcal{F}(\hat{k}_p)] + \sum_{q=q_0}^{n_2} [-\beta_q s_q^2 \\ &\quad - e^{\hat{k}_q} \cdot y_q \ln \left( 1 + \frac{|s_q|}{e^{\hat{k}_q}} \right)] + \sum_{q=q_0}^{n_2} (\bar{\varepsilon}_q - \hat{k}_q) [|s_q| + \mu_q^2 \\ &\quad + \frac{\alpha_q}{\kappa_V} \tanh(\frac{\alpha_q}{|s_q| + \mu_q^2})] \\ &\leq \sum_{p=p_0}^{n_1} [-\beta_p s_p^2 - e^{\hat{k}_p} \cdot \mathcal{F}(\hat{k}_p)] + \sum_{q=q_0}^{n_2} (\bar{\varepsilon}_q - \hat{k}_q) [|s_q| \\ &\quad + \mu_q^2 + \frac{\alpha_q}{\kappa_V} \tanh(\frac{\alpha_q}{|s_q| + \mu_q^2})] \end{aligned} \quad (28)$$

At the beginning of control ( $t = 0$ ), the sliding variables of all subsystems are outside the designed vicinity ( $|s_i| > D_i$ , for all  $i = 1, 2, \dots, n$ ), which means the adaptive laws of switching gains for all subsystems are positive as according to (10). Therefore, the set  $E_q$  is empty with  $n_2 = 0$  while set  $E_p$  is  $\{1, 2, \dots, n\}$  with  $n_1 = n$ , (28) can be written as:

$$\dot{V}_2(t) \leq \sum_{i=1}^n [-\beta_i s_i^2(t) - e^{\hat{k}_i(t)} \cdot \mathcal{F}(\hat{k}_i(t))] < 0 \quad (\text{when } 0 \leq t < \min_{1 \leq i \leq n} \{\mathcal{T}_i\}) \quad (29)$$

Clearly, (29) holds until any subsystem enters the vicinity. After that,  $\dot{V}_2 < 0$  cannot hold because of the negative adaptive law of switching gain according to (10). Therefore, the Lyapunov function  $V_2(t)$  shall be bounded when  $t \geq \min_{1 \leq i \leq n} \{\mathcal{T}_i\}$ . To achieve this goal, we assume a sufficiently large number of the second Lyapunov function (25), namely,  $V_2 = V_2^*$ . Clearly, a sufficiently large number  $V_2^*$  requires at least one of the terms in  $\frac{1}{2} s^T s$  and  $\frac{1}{2} \sum_{i=1}^n \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i}$  to be sufficiently large. Hence, we consider 2 cases.

*Case 1:* The term  $\frac{1}{2} s^T s$  has a sufficiently large number  $R_1^*$ , namely,  $\frac{1}{2} s^T s = R_1^*$ .

*Lemma 5:* If the term  $\frac{1}{2} s^T s$  has a sufficiently large number  $R_1^*$ , the maximum element  $|s_{p^*}|$  in the vector  $[|s_1|, \dots, |s_n|]^T \in R^{n \times 1}$  shall be no less than the positive number:

$$|s_{p^*}| = \max_{1 \leq i \leq n} \{|s_i|\} \geq \sqrt{\frac{2R_1^*}{n}} \quad (30)$$

The proof of Lemma 5 is given in the appendix.

Now, let  $R_1^*$  be no less than a positive number:

$$R_1^* \geq \frac{n}{2}(\delta_1 + \delta_2)^2 \quad (31)$$

The definitions of  $\delta_1$  and  $\delta_2$  are given in the sentence below (22). According to (30) and (31), It is clear that  $|s_{p^*}| \geq \delta_1 + \delta_2 > \delta_2$ . Therefore, the sliding variable  $s_{p^*}$  is outside the vicinity  $D_{p^*}$ , which means the set  $E_p$  is not empty with  $n_1 \geq 1$  because  $p^* \in E_p$ . Combining (27b) and the fact that  $\tanh(\cdot) < 1$ , (28) can be further written as:

$$\begin{aligned} \dot{V}_2 &\leq -\beta_{p^*} s_{p^*}^2 - e^{\hat{k}_{p^*}} \cdot \mathcal{F}(\hat{k}_{p^*}) + \sum_{q=q_0}^{n_2} (\bar{\varepsilon}_q - \hat{k}_q) [|s_q| \\ &\quad + \mu_q^2 + \frac{\alpha_q}{\kappa_V} \tanh(\frac{\alpha_q}{|s_q| + \mu_q^2})] \\ &< -\beta_{p^*} s_{p^*}^2 + \sum_{q=q_0}^{n_2} \bar{\varepsilon}_q (D_q + \mu_q^2 + \frac{\alpha_q}{\kappa_V}) \\ &< -\beta_{min} s_{p^*}^2 + \sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V}) \end{aligned} \quad (32)$$

According to (32),  $\dot{V}_2 < 0$  holds as long as  $|s_{p^*}|$  is no less than the positive number:

$$|s_{p^*}| \geq \sqrt{\frac{\sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})}{\beta_{min}}} \quad (33)$$

Combining (28), (31) and (33), it is concluded that  $\dot{V}_2 < 0$  can hold by a sufficiently large number  $R_1^*$  of the term  $\frac{1}{2} s^T s$  to satisfy the following condition:

$$R_1^* \geq \left\{ \max \frac{n}{2}(\delta_1 + \delta_2)^2, \frac{n}{2} \cdot \frac{\sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})}{\beta_{min}} \right\} \quad (34)$$

*Case 2:* The term  $\frac{1}{2} \sum_{i=1}^n \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i}$  has a sufficiently large number  $R_2^*$ . Namely,  $\frac{1}{2} \sum_{i=1}^n \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i} = R_2^*$ .

*Lemma 6:* If the term  $\frac{1}{2} \sum_{i=1}^n \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i}$  has a sufficiently large number  $R_2^*$ , the maximum element  $\frac{(\hat{k}_{i^*} - \bar{\varepsilon}_{i^*})^2}{2\gamma_{i^*}}$  in the vector  $[\frac{(\hat{k}_1 - \bar{\varepsilon}_1)^2}{2\gamma_1}, \dots, \frac{(\hat{k}_n - \bar{\varepsilon}_n)^2}{2\gamma_n}]^T \in R^{n \times 1}$  will be no less than the positive number:

$$\frac{(\hat{k}_{i^*} - \bar{\varepsilon}_{i^*})^2}{2\gamma_{i^*}} = \max_{1 \leq i \leq n} \left\{ \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{2\gamma_i} \right\} \geq \frac{R_2^*}{n} \quad (35)$$

The proof of *Lemma 6* is the same as that of *Lemma 5*.

It is clear that (35) leads to 2 possible solutions:

$$\hat{k}_{i^*} \geq \bar{\varepsilon}_{i^*} + \sqrt{\frac{2\gamma_{i^*} \cdot R_2^*}{n}} \geq 0 \quad (36a)$$

$$\bar{\varepsilon}_{i^*} \geq \hat{k}_{i^*} + \sqrt{\frac{2\gamma_{i^*} \cdot R_2^*}{n}} \geq 0 \quad (36b)$$

Now, let  $R_2^*$  be no less than such a positive number:

$$R_2^* \geq \max \left\{ \frac{n \|\bar{\varepsilon}\|^2}{2\gamma_{min}} + \delta_3, \frac{n}{2\gamma_{min}} (\bar{k}_{max} - \bar{\varepsilon}_{min})^2 \right\} \quad (37)$$

The definitions of  $\gamma_{min}$ ,  $\bar{k}_{max}$ ,  $\bar{\varepsilon}_{min}$ , and  $\delta_3$  are given in the sentence below (22).

*Remark 6:* It is clear that  $R_2^* > \frac{n \bar{\varepsilon}_{i^*}^2}{2\gamma_{i^*}}$  holds when (37) is satisfied because  $\frac{n \|\bar{\varepsilon}\|^2}{2\gamma_{min}} + \delta_3 > \frac{n \bar{\varepsilon}_{i^*}^2}{2\gamma_{i^*}}$ , which contradicts (36b). As a result, (36a) is the only solution of (35) when (37) is satisfied. Subsequently, combining (37) with (36a) leads to  $\hat{k}_{i^*} \geq \bar{\varepsilon}_{i^*}$ . Therefore, according to (11),  $\mathcal{F}(\hat{k}_{i^*}) = \Omega_{i^*} > 0$  holds when (37) is satisfied.

Now, we consider 2 situations of *case 2* when (37) is satisfied:  $i^* \in E_p$  or  $i^* \in E_q$ .

When  $i^* \in E_p$ , the term  $E_p$  is not empty with  $n_1 \geq 1$ . Combining (27b) and the fact that  $\tanh(\cdot) < 1$ , (28) can be further written as:

$$\begin{aligned} \dot{V}_2 &\leq -\beta_{i^*} s_{i^*}^2 - e^{\hat{k}_{i^*}} \cdot \mathcal{F}(\hat{k}_{i^*}) + \sum_{q=q_0}^{n_2} (\bar{\varepsilon}_q - \hat{k}_q) [|s_q| \\ &\quad + \mu_q^2 + \frac{\alpha_q}{\kappa_V} \tanh(\frac{\alpha_q}{|s_q| + \mu_q^2})] \\ &\leq -e^{\hat{k}_{i^*}} \cdot \Omega_{i^*} + \sum_{q=q_0}^{n_2} (\bar{\varepsilon}_q - \hat{k}_q) [|s_q| + \mu_q^2 \\ &\quad + \frac{\alpha_q}{\kappa_V} \tanh(\frac{\alpha_q}{|s_q| + \mu_q^2})] \\ &< -e^{\hat{k}_{i^*}} \cdot \Omega_{i^*} + \sum_{q=q_0}^{n_2} \bar{\varepsilon}_q (D_q + \mu_q^2 + \frac{\alpha_q}{\kappa_V}) \\ &< -e^{\hat{k}_{i^*}} \cdot \Omega_{min} + \sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V}) \end{aligned} \quad (38)$$

According to (38),  $\dot{V}_2 < 0$  holds as long as  $\hat{k}_{i^*}$  is not less than the positive number:

$$\hat{k}_{i^*} \geq \ln \left( \frac{\sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})}{\Omega_{min}} \right) \quad (39)$$

Combining (36a), (37) and (39), It is concluded that  $\dot{V}_2 < 0$  can hold by a sufficiently large number  $R_2^*$  of the term  $\frac{1}{2} \sum_{i=1}^n \frac{(\hat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i}$  satisfying (40) when  $i^* \in E_p$ .

$$\begin{aligned} R_2^* &\geq \max \left\{ \frac{n \|\bar{\varepsilon}\|^2}{2\gamma_{min}} + \delta_3, \frac{n}{2\gamma_{min}} (\bar{k}_{max} - \bar{\varepsilon}_{min})^2, \right. \\ &\quad \left. \frac{n}{2\gamma_{min}} \left( \ln \left( \frac{\sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})}{\Omega_{min}} \right) - \bar{\varepsilon}_{i^*} \right)^2 \right\} \end{aligned} \quad (40)$$

When  $i^* \in E_q$ , the term  $E_q$  is not empty with  $n_2 \geq 1$ . Notably, (36a) implies that  $\bar{\varepsilon}_{i^*} - \hat{k}_{i^*} \leq 0$ . Combining with (20), (27b) and that  $\tanh(\cdot) < 1$ , (28) can be further written as:

$$\begin{aligned} \dot{V}_2 &\leq \sum_{q=q_0}^{n_2} (\bar{\varepsilon}_q - \hat{k}_q) [|s_q| + \mu_q^2 + \frac{\alpha_q}{\kappa_V} \tanh(\frac{\alpha_q}{|s_q| + \mu_q^2})] \\ &= (\bar{\varepsilon}_{i^*} - \hat{k}_{i^*}) \left[ |s_{i^*}| + \mu_{i^*}^2 + \frac{\alpha_{i^*}}{\kappa_V} \tanh \left( \frac{\alpha_{i^*}}{|s_{i^*}| + \mu_{i^*}^2} \right) \right] \\ &\quad + \sum_{q=q_0}^{n_2-1} (\bar{\varepsilon}_q - \hat{k}_q) [|s_q| + \mu_q^2 + \frac{\alpha_q}{\kappa_V} \tanh \left( \frac{\alpha_q}{|s_q| + \mu_q^2} \right)] \\ &\leq \frac{\alpha_{i^*}}{\kappa_V} (\bar{\varepsilon}_{i^*} - \hat{k}_{i^*}) + \sum_{q=q_0}^{n_2-1} \bar{\varepsilon}_q (D_q + \mu_q^2 + \frac{\alpha_q}{\kappa_V}) \\ &\leq \frac{\alpha_{min}}{\kappa_V} (\bar{\varepsilon}_{max} - \hat{k}_{i^*}) + \sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V}) \end{aligned} \quad (41)$$

The definitions of  $\alpha_{min}$  and  $\bar{\varepsilon}_{max}$  are given in the sentence below (22). According to (41),  $\dot{V}_2 < 0$  holds as long as  $\hat{k}_{i^*}$  is

not less than the positive number:

$$\widehat{k}_i^* \geq \bar{\varepsilon}_{max} + \frac{\kappa_V}{a_{min}} \sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V}) \quad (42)$$

Combining with (36a), (37) and (42), It is concluded that  $\dot{V}_2 < 0$  can hold by a sufficiently large number  $R_2^*$  of the term  $\frac{1}{2} \sum_{i=1}^n \frac{(\widehat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i}$  satisfying (43) when  $i^* \in E_q$ :

$$R_2^* \geq \max \left\{ \frac{n \|\bar{\varepsilon}\|^2}{2\gamma_{min}} + \delta_3, \frac{n}{2\gamma_{min}} [\ln(\bar{\varepsilon}_{max} + \frac{\kappa_V}{a_{min}} \sum_{i=1}^n \bar{\varepsilon}_i (D_i + \mu_i^2 + \frac{\alpha_i}{\kappa_V})) - \bar{\varepsilon}_{i^*}^2] \right\} \quad (43)$$

In the light of (34), (40) and (43), it is shown that  $\dot{V}_2 < 0$  can hold by a sufficiently large number  $V_2^* = \frac{1}{2} s^T s + \frac{1}{2} \sum_{i=1}^n \frac{(\widehat{k}_i - \bar{\varepsilon}_i)^2}{\gamma_i} = R_1^* + R_2^*$ .  $R_1^*$  and  $R_2^*$  are defined in (22). As a result,  $V_2$  will not exceed  $V_2^*$  because  $\dot{V}_2 < 0$  holds as long as  $V_2 \geq V_2^*$ , and all terms including  $s^T s$  in  $V_2$  will not be greater than  $V_2^*$ , namely:

$$\|s\| \leq 2\sqrt{V_2} \leq 2\sqrt{R_1^* + R_2^*} \quad (44)$$

*Proof of theorem 1 complete*

*Remark 7:* According to (12), on one hand, the sliding variable will either arrive at the boundary of the vicinity ( $|s_i| = D_i$ ) or leave the vicinity ( $|s_i| > D_i$ ), and the latter will result in the greatest value of  $y_i$  regardless of the value of  $w_i$  (*remark 3*). On the other hand, when the sliding variable is inside the vicinity and close to the manifold,  $w_i$  plays an important role to determine the value of  $y_i$ . More precisely, the condition  $w_i \rightarrow 0$  leads to an increase in  $y_i$ , which offers the ability to converge the sliding variable but with the risk of increasing the chattering effects, while  $w_i \rightarrow 1$  brings up the decrease of  $y_i$  that offers the ability to reduce chattering effects with the risk of sliding variable drifting away from the manifold.

As a result, an optimal  $w_i$  exists to achieve the drifting-free and chattering-free dynamics of sliding variables between 2 consecutive sampling times of TDE (eg.  $s_i(t) - s_i(t - T_s) = -0.1s_i(t - T_s)$ ). To achieve this goal, a reinforcement learning (RL) based adaptive law for  $w_i$  is designed as:

$$w_i(t) = \begin{cases} [1 - \sigma_i(t)] w_i(t - T_s), \\ \text{if } s_i(t - T_s) [\Delta s_i(t) - \Delta s_i^*(t)] > 0 \\ [1 - \sigma_i(t)] w_i(t - T_s) + \sigma_i(t), \quad \text{else.} \end{cases} \quad (45)$$

where  $0 \leq \sigma_i(t) \leq 1$ ,  $\Delta s_i(t) = s_i(t) - s_i(t - T_s)$  and  $\Delta s_i^*$  is the desired variance. In this paper, we set:  $\Delta s_i^* = -0.1s_i(t - T_s)$ .  $w_i(t)$  starts at zero during the control, namely,  $w_i(t = 0) = 0$ .

*Remark 8:* (45) implies that  $w_i$  increases towards 0 to speed up the convergence of the sliding variables if  $\Delta s_i(t)$  does not meet the desired variance  $\Delta s_i^*$ , namely,  $s_i(t - T_s) [\Delta s_i(t) - \Delta s_i^*(t)] > 0$ . Otherwise,  $w_i$  decreases towards 1 to slow down the convergence of the sliding variable to attenuate the chattering effects. A greater value of

$|\Delta s_i(t) - \Delta s_i^*(t)|$  indicates a greater  $\sigma_i(t)$  is required to tune the value of  $w_i(t)$ .

It is hard to determine the appropriate values of  $\sigma_i(t)$  because of the unknown disturbances and model uncertainties. Hence, instead of manually tuning  $\sigma_i(t)$ , we apply the reinforcement learning-based fuzzy logic inference that can learn the optimal  $\sigma_i(t)$  during the control.

$$\sigma_i(t) = FI(\Gamma_{1,i}(t), \Gamma_{2,i}(t)) \quad (46)$$

where  $FI(\cdot, \cdot)$  refers to the fuzzy logic inference (FLI), and  $\Gamma_{1,i}$  and  $\Gamma_{2,i}$  are 2 fuzzy inputs for the  $i^{th}$  subsystem. In this paper,  $\Gamma_{1,i}(t) = |s_i(t)|$  and  $\Gamma_{2,i}(t) = \text{sgn}(s_i(t - T_s)) (\Delta s_i(t) - \Delta s_i^*(t))$ .

*Remark 9:* The risk of losing stability caused by the bad policies determined from the RL during the initial stage of learning as discovered by [60] is not of significant concern. It is because the proposed control scheme (8)-(13) guarantees the UUB errors by the bounded sliding variables regardless of the value of  $w_i$ , shown in (44). As a result, the proposed SCSMC scheme provides RL with a safe environment to learn optimal policies.

### C. FUZZY Q REINFORCEMENT LEARNING TO DETERMINE $\sigma_i$

Q learning can explore the optimal policies by learning the relationship between interval-values states (discrete states) and applied actions. Fuzzy Q learning is the extension of Q learning, which allows the application of Q learning on continuous systems such as FSRMs.

Intuitively, the following possible linguistic rules to determine  $\sigma_i(t)$  according to *remark 7* and *remark 8* can be given as examples:

IF  $\Gamma_{1,i}(t)$  is small AND  $\Gamma_{2,i}(t)$  is negatively big, THEN  $\sigma_i(t)$  is small.

IF  $\Gamma_{1,i}(t)$  is big AND  $\Gamma_{2,i}(t)$  is closely zero, THEN  $\sigma_i(t)$  is medium.

IF  $\Gamma_{1,i}(t)$  is big AND  $\Gamma_{2,i}(t)$  is negatively big, THEN  $\sigma_i(t)$  is big.

The above terms of small, medium and big are linguistical descriptions for variables  $\Gamma_{1,i}$ ,  $\Gamma_{2,i}(t)$  and  $\sigma_i(t)$ . Fuzzification is required to apply such linguistic descriptions on numerical variables. More precisely, numerical variables  $\Gamma_{1,i}$  and  $\Gamma_{2,i}$  are fuzzified to a series of firing rates of fuzzy rules by the triangular membership function shown in figure 4. Subsequently, the actual numerical values of  $\sigma_i(t)$  is calculated by the fuzzy reasoning based on the firing rates of the rules and the numerical values of the applied actions.

The fuzzy sets for fuzzy inputs, which present linguistic variables in a numerical form (firing rate), are shown in Figure 4 and are detailed as follows:

$$\text{Lin}(\Gamma_{1,i}(t)) = \left\{ \zeta_{1,1}^{(i)}, \dots, \zeta_{1,\mathfrak{a}}^{(i)}, \dots, \zeta_{1,\mathfrak{A}}^{(i)} \right\}, \quad \mathfrak{a} = 1, 2, \dots, \mathfrak{A} \quad (47a)$$

$$\text{Lin}(\Gamma_{2,i}(t)) = \left\{ \zeta_{2,1}^{(i)}, \dots, \zeta_{2,\mathfrak{b}}^{(i)}, \dots, \zeta_{2,\mathfrak{B}}^{(i)} \right\}, \quad \mathfrak{b} = 1, 2, \dots, \mathfrak{B} \quad (47b)$$



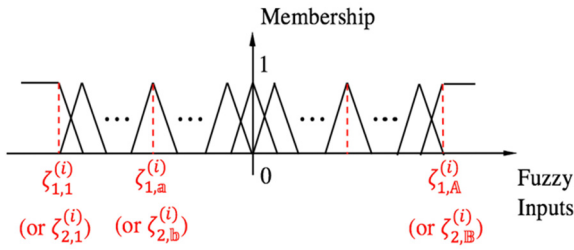


FIGURE 4. Membership functions of fuzzy inputs of the  $i^{th}$  subsystem.

where  $\mathbb{A}$  is the number of fuzzy sets ( $\zeta_{1,a}$ ) for the fuzzy input  $\Gamma_{1,i}$ , and  $\mathbb{B}$  is the number of fuzzy sets ( $\zeta_{1,b}$ ) for the fuzzy input  $\Gamma_{2,i}$ .

The relationship between the  $m^{th}$  fuzzy rule and the candidates of action as well as the q-values evaluating those candidates for the  $i^{th}$  subsystem is defined:

$$R_{m,i}: \text{IF } S_{1,i}^{lk} \text{ is } L_{1,i}^m \text{ and } S_{2,i}^{lk} \text{ is } L_{2,i}^m \text{ and } \dots \text{ and } S_{mm,i}^{lk} \text{ is } L_{mm,i}^m, \\ \text{THEN } u_i^m \in U_{i,m} \text{ that } u_i^m = u_{i,1}^m \text{ with } q_i^{lk}(m, 1) \text{ or } u_i^m = u_{i,2}^m \\ \text{with } q_i^{lk}(m, 2) \text{ or } u_i^m = u_{i,3}^m \text{ with } q_i^{lk}(m, 3), \dots, u_i^m = u_{i,p}^m \\ \text{with } q_i^{lk}(m, p), \dots, u_i^m = u_{i,p}^m \text{ with } q_i^{lk}(m, P). \quad (48)$$

where  $U_{i,m} = \{u_{i,1}^m, \dots, u_{i,p}^m\}$  is the set of action candidates for the parameter  $\sigma_i$  in the rule  $R_{m,i}$ .  $S_i^{lk} = \{S_{1,i}^{lk}, \dots, S_{mm,i}^{lk}\}$  is the set of fuzzy inputs for the  $i^{th}$  subsystem at the moment  $lk$ .  $L_i^m = \{L_{1,i}^m, \dots, L_{mm,i}^m\}$  is the set of linguistic variables of fuzzy inputs. Each action  $u_{i,p}^m$  is evaluated by the associated q value.

Fuzzy inputs are initially fuzzified by the triangular membership function shown in Figure 4 with the fuzzy sets (47) and then matched with the rule antecedents (48), which gives the set of firing rates for fuzzy rules  $\varphi(S_i^{lk}) = \{\varphi_1(S_i^{lk}), \dots, \varphi_N(S_i^{lk})\}$  where  $N$  is the number of fuzzy rules. After that, the numerical values of  $\sigma_i$  can be calculated by the set of firing rates and the actions selected in the light of q-values.

*Remark 10:* Notably, the current time instant  $t$  is equal to the current instant  $lk$  in RL because of the discrete nature of TDE. Similarly,  $t - T_s$  is equal to  $lk - 1$ , and  $t + T_s$  is equal to  $lk + 1$  where  $T_s$  is the sampling time of the TDE. For example,  $\sigma_i(lk)$  is equal to  $\sigma_i(t)$  and  $\sigma_i(lk + 1)$  is equal to  $\sigma_i(t + T_s)$ .

The action with the greatest q-value is considered as the most optimal choice among all candidates of action:

$$u_i^{*m} = \arg \max_{u_i^m \in U_{i,m}} q_i^{lk}(m, p) \quad (49)$$

To alleviate the problem of the local optimum in the learning process, a greed mechanism is used:

$$\hat{u}_i^m = \begin{cases} u_i^{\perp m}, & \text{with probability } \Lambda \\ u_i^{*m}, & \text{with probability } 1 - \Lambda \end{cases} \quad (50)$$

where  $u_i^{\perp m}$  is a random action among  $U_{i,m}$ .  $\hat{u}_i^m$ .  $0 < \Lambda < 1$  is the probability to explore potentially better actions.

The numerical value of  $\sigma_i$  is calculated by firing rates and the selected actions:

$$\sigma_i(t) = \frac{\sum_{m=1}^N \varphi_m(S_i^{lk}) \hat{u}_i^m}{\sum_{m=1}^N \varphi_m(S_i^{lk})} \quad (51)$$

The q-values are updated according to the obtained rewards, which therefore can learn the optimal actions that can achieve higher rewards. To avoid the fuzzy rules from being influenced by the incoherent reinforcements, which is caused by their interactions with other rules, the modified fuzzy Q learning [61] is used.

The Q value of the  $i^{th}$  subsystem at the instant  $lk$  can be calculated as follows:

$$Q(S_i^{lk}) = \frac{\sum_{m=1}^N \varphi_m(S_i^{lk}) [\sum_{p=1}^P \phi_{p,m}(\sigma_i(lk)) q_i^{lk}(m, p)]}{\sum_{m=1}^N \varphi_m(S_i^{lk})} \quad (52)$$

where  $[\phi_1(\sigma_i(lk)), \phi_1(\sigma_i(lk)), \dots, \phi_1(\sigma_i(lk))]$  is the set of firing rates of the aggregated action  $\sigma_i(lk)$  in the rule  $R_{m,i}$ , which is calculated by the triangular membership function shown in figure 5 with fuzzy sets (53).

$$Lin(\sigma_i) = \{u_{i,1}^m, u_{i,2}^m, \dots, u_{i,p}^m\} \quad (53)$$

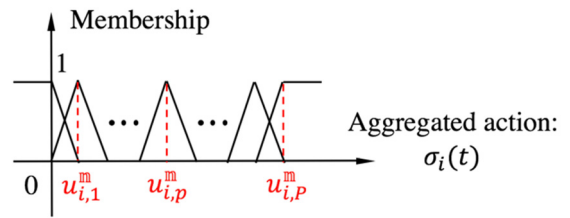


FIGURE 5. Membership function of action  $\sigma_i$  among the action candidates.

It is noticed that the fuzzy sets in (53) fully consist of the action candidates in the rule  $R_{m,i}$ .

The target value is calculated as:

$$\mathcal{V}(S_i^{lk}) = \frac{\sum_{m=1}^N [\varphi_m(S_i^{lk}) \cdot \max_{1 \leq p \leq P} \{q_i^{lk}(m, p)\}]}{\sum_{m=1}^N \varphi_m(S_i^{lk})} \quad (54)$$

When the system transitions from the state  $S_i^{lk}$  to  $S_i^{lk+1}$ , the temporal difference (TD) error is calculated in the light of reward obtained:

$$\Delta Q_i^{lk} = r_i^{lk} + \eta_i \mathcal{V}(S_i^{lk+1}) - Q(S_i^{lk}) \quad (55)$$

where  $r_i^{lk}$  is the reward defined in (56), and  $\eta_i \in [0, 1]$  is the discount factor reflecting the contribution of the current action to obtain the future reward.

$$r_i^{lk} = r_i(t) = e^{-\frac{[\Delta s_i(t) - \Delta s_i^*(t)]^2}{s_i(t)}} \quad (56a)$$

$$s_i(t) = \begin{cases} \frac{-[0.9s_i(t - T_s)]^2}{\ln(0.1)}, & \text{if } |\Delta s_i(t)| > |0.1s_i(t - T_s)| \\ \frac{-[0.1s_i(t - T_s)]^2}{\ln(0.1)}, & \text{else.} \end{cases} \quad (56b)$$

Clearly, the action obtaining the more desired variance of the sliding variable will be given a higher reward and vice versa. Equation (56a) implies that the action to achieve  $\Delta s_i(t) = \Delta s_i^*(t)$  will be given the highest reward with the value of 1. Equation (56b) then indicates that the variance of the sliding variable having either a quicker convergence than  $-s_i(t - T_s)$  or a slower convergence (including divergence) than 0 will result in the reward becoming less than 0.1.

Finally, the adaptive law of q-values is:

$$q_i^{k+1}(m, p) = q_i^k(m, p) + \lambda_i \cdot \Delta Q_i^k \cdot \frac{\varphi_m(S_i^k)}{\sum_{m=1}^N \varphi_m(S_i^k)} \cdot \frac{\phi_{p,m}(\sigma_i(\mathbb{k}_k))}{\sum_{p=1}^P \phi_{p,m}(\sigma_i(\mathbb{k}_k))} \quad (57)$$

where  $\lambda_i \in [0, 1]$  is the learning rate.

The steps of selecting parameters of the proposed SCSMC scheme are given as follows. The detailed method of determining the values of each parameter are provided in *remark 11~remark 19*.

Step (1): Selecting the appropriate values of  $\bar{k}_i$ ,  $\Omega_i$  and  $\mu_i$  in the light of *remark 14*.

Step (2): Selecting proper values of  $\beta_i$  and  $\gamma_i$  in the light of *remark 11* to achieve a satisfyingly fast convergence of sliding variables toward the sliding manifold.

Step (3): Selecting appropriate values of  $D_i$  in the light of *remark 13* to ensure the sliding variables can enter the vicinity before the decline of switching gains.

Step (4): Selecting appropriate values of  $\alpha_i$  in the light of *remark 12* to achieve the satisfying attenuation on chattering effects.

Step (5): If an unacceptable discontinuity of virtual control signal occurs due to the switch function  $\mathcal{F}(\widehat{k}_i)$  during step (2)~step (4), the procedure should go back to step (1) to increase the values of  $\bar{k}_i$  and  $\Omega_i$ .

Step (6): Selecting the appropriate values of fuzzy sets (47) in the light of *remark 15*.

Step (7): Selecting the appropriate values of action group  $U_{i,m} = \{u_{i,1}^m, \dots, u_{i,p}^m\}$  in the light of *remark 16*.

Step (8): Selecting the proper values of probability of mutation  $\Lambda$ , discount factor  $\eta_i$ , learning rate  $\lambda_i$  in the light of *remark 17*, *remark 18* and *remark 19* respectively.

Step (9): If the tracking performance is unsatisfied, the procedure should go back to step (6).

*Remark 11:* The large values of  $\beta_i$  can result in the fast convergence of sliding variables towards zero, but at the expense of large control inputs that increase the chattering effects and fuel consumption. Similarly, large values of  $\gamma_i$  can also lead to a fast convergence in the sliding variables by quickly increasing the value of the switching gain. Although large values of the switching gain can increase the risk of chattering effects, the switching gain can decrease inside the vicinity so that the problem of over-estimated switching gain is alleviated. Therefore,  $\beta_i$  and  $\gamma_i$  should be set to start from a small value, and it should be allowed to gradually increase

until the satisfyingly fast convergence of the sliding variables is achieved.

*Remark 12:* The large values of  $\alpha_i$  can result in a fast decline of the switching gain to zero, which means a significant chattering effect reduction at the cost of the under-estimated switching gain. In the proposed SCSMC, the under-estimated switching gain is not of concern because of the usage of the RL-based robust term that actively prevents the sliding variables from leaving the manifold due to the insufficient switching gain. Therefore, the value of  $\alpha_i$  should initially be small and then gradually increase until the satisfying attenuation of the chattering effect is obtained.

*Remark 13:*  $D_i$  is the vicinity of the sliding manifold. A large value of  $D_i$  could compromise the tracking accuracy because of the insufficient increase of the switching gain. In contrast, a small value of  $D_i$  could lead to the failure of the proposed SCSMC if the sliding variable cannot enter the small  $D_i$  because of the limited computation ability of the hardware. As a result, it is suggested that the trials of selecting  $D_i$  should start at a big value (but no more than the initial value of the sliding variable  $|s_i(t=0)|$ ) and then allow it to gradually decrease until a good tracking accuracy is achieved without the sliding variables failing to enter into the vicinity.

*Remark 14:*  $\bar{k}_i$  and  $\Omega_i$  are used to avoid great values of the virtual control signal by offering a small value to the robust term when the value of switching gain is great. As a result,  $\Omega_i$  should be selected as a small value. To avoid the discontinuity of the virtual control signal caused by the switch function  $\mathcal{F}(\widehat{k}_i)$ , the threshold  $\bar{k}_i$  should be a big value. The term  $\mu_i$  is used to derive the inequality (20) and should be a small number close to zero.

*Remark 15:* Fuzzy sets (47) are imperative because they transform the numerical values of the fuzzy inputs into the group of firing rates corresponding to linguistic variables, which enables the fuzzy reasoning. Hence, we suggest  $\zeta_{1,\mathbb{A}}^{(i)}$  and  $\zeta_{2,\mathbb{B}}^{(i)}$  to be given large values while  $\zeta_{1,1}^{(i)}$  and  $\zeta_{2,1}^{(i)}$  to be given small values for covering the range of all the possible values of  $\Gamma_{1,i}$  and  $\Gamma_{2,i}$  during the control process. Moreover, to well map the relationship between fuzzy outputs and fuzzy inputs, the  $\zeta_{1,\mathbb{A}}^{(i)}$  and  $\zeta_{2,\mathbb{B}}^{(i)}$  are suggested to be evenly distributed among the ranges  $[\zeta_{1,1}^{(i)}, \zeta_{1,\mathbb{A}}^{(i)}]$  and  $[\zeta_{2,1}^{(i)}, \zeta_{2,\mathbb{B}}^{(i)}]$  respectively with the sufficiently large integral numbers  $\mathbb{A}$  and  $\mathbb{B}$ . However, large integral numbers  $\mathbb{A}$  and  $\mathbb{B}$  will increase the computation load, so we suggest that  $\mathbb{A}$  and  $\mathbb{B}$  start at small values, and it should increase until the satisfying performance of fuzzy reasoning is obtained with an acceptable computation load.

*Remark 16:* Action group  $U_{i,m} = \{u_{i,1}^m, \dots, u_{i,p}^m\}$  determines the dynamics of sliding variables inside the vicinity by adopting the values of the robust term.  $u_{i,1}^m$  should be of small values and  $u_{i,p}^m$  should be big values to ensure the optimal action can be calculated by the given candidate actions. However, the significant difference between candidate actions could result in the chattering calculated actions. As a result, the trials of selecting  $u_{i,p}^m$  ( $u_{i,1}^m$ ) is suggested to start at a big (small) value and then be gradually decreased (increased)

until the satisfyingly less-chattering calculated actions are achieved. The remaining members  $[u_{i,2}^m, \dots, u_{i,P-1}^m]$  are suggested to be evenly distributed among between  $u_{i,1}^m$  and  $u_{i,P}^m$ . Although the great values of the integral number  $P$  could increase the possibility of successfully calculating optimal actions by the sufficient action candidates, it will bring up the computation load and the difficulty on the convergence of  $q$  values in RL, and therefore  $P$  should be given a small value between 1~10.

*Remark 17:* The probability of mutation  $\Lambda$  means the trade-off between the exploration for potentially better policies and the exploitation of learned policies. It is widely suggested in RL applications that  $\Lambda$  should be large (e.g. 0.4) at the beginning of learning and be small (e.g. 0) during the later stage of learning.

*Remark 18:* Discount factor  $\eta_i$  reflects the attention paid on the effects of current action on the future performance. In this paper, the designed reward is given according to the variance of sliding variables. The current variance of sliding variables mainly depends on the current control signal, thus  $\eta_i$  should be a small value (e.g. 0.1).

*Remark 19:* The learning rate  $\lambda_i$  means the efficiency of memorizing the new knowledge and forgetting the old knowledge. A large value of  $\lambda_i$  could achieve a fast convergence of the  $q$  values, which means a high learning efficiency. In contrast, a small value of  $\lambda_i$  could achieve the robustness of learning by keeping the old knowledge. Therefore, we suggest medium values (eg 0.4~0.6) for  $\lambda_i$ .

*Remark 20:* The sufficiency of both the fuzzy rules (equally, a large integral numbers  $\mathbb{A}$  and  $\mathbb{B}$  in (47)) and action candidates (equally, a large integral number  $P$  in  $U_{i,m} = \{u_{i,1}^m, \dots, u_{i,P}^m\}$ ) is important to the performance of reinforcement learning because of the reasons detailed in *remark 15* and *remark 16*. However, the complexity increases when the numbers of fuzzy rules and the candidates in the action group increase because of the increased number of  $q$ -values that are being updated during the learning process, which brings up the great computation load. Therefore, there is a trade-off between the complexity and the learning performance of RL. As a result, it is suggested that the fuzzy rules and action candidates should be sufficient (by selecting the large numbers  $\mathbb{A}$ ,  $\mathbb{B}$ , and  $P$ ) if the computation capability is sufficient, while the number of fuzzy rules and action candidates should decrease (by selecting small numbers  $\mathbb{A}$ ,  $\mathbb{B}$ , and  $P$ ) if the computation capability is insufficient.

#### IV. SIMULATION RESULTS

In this section, the numerical simulation for a 2-rigid-links FSRM is executed to verify the effectiveness of the proposed control scheme. The sampling time of the FSRM simulation is set at  $1 \times 10^{-3}s$ , and the sampling time of the controller and TDE is set as  $T_s = 0.02s$  to show the discrete nature of the controller in practices. The details of the used dynamic equations of a 2-rigid-links FSRM shown in figure 6 can be found in [63]. The parameters of the dynamic model are detailed in table 1.

TABLE 1. Parameters of dynamics model of FSRM.

| Link $i$ | $a_i(m)$ | $b_i(m)$ | $m_i(kg)$ | $I_i(kg \cdot m^2)$ |
|----------|----------|----------|-----------|---------------------|
| 0        | -        | 1        | 280       | 23.32               |
| 1        | 0.5      | 0.5      | 8         | 0.67                |
| 2        | 0.5      | 0.5      | 4         | 0.33                |

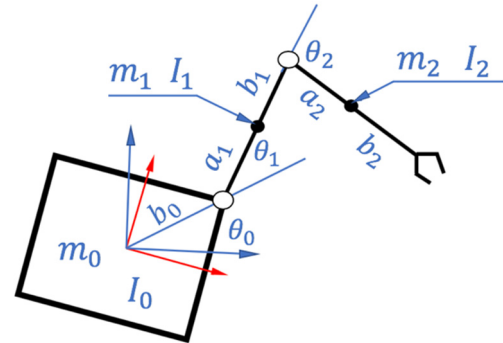


FIGURE 6. 2-rigid-links free-floating space robotic manipulator.

The initial angles and angular velocities of the 2 manipulator joints are  $[\theta_1, \theta_2]^T = [\dot{\theta}_1, \dot{\theta}_2]^T = [0, 0]^T$ . The initial position and attitude of the base and their derivatives are  $[x_0, y_0, \theta_0]^T = [\dot{x}_0, \dot{y}_0, \dot{\theta}_0]^T = [0, 0, 0]^T$ . Similar to [7], [10], the desired trajectories of the joints are selected as the sine and cosine functions, as given below:

$$\begin{cases} \theta_{1,r} = -0.6 \sin\left(\frac{\pi}{5}t\right) + \frac{\pi}{8} \\ \theta_{2,r} = 0.6 \cos\left(\frac{\pi}{5}t\right) - \frac{\pi}{8} \end{cases} \quad (58)$$

Similar to [7], [10], [16], the external disturbances  $d = [d_1, d_2]^T$  are selected as the combination of the sine and cosine functions, which are used to verify the robustness of proposed SCSMC.

$$\begin{cases} d_1 = 7.8 \sin\left(\frac{\pi}{3}t + \frac{\pi}{4}\right) + 0.65 \sin\left(\frac{\pi}{10}t + \frac{\pi}{4}\right) \\ d_2 = 5.85 \cos\left(\frac{\pi}{3}t + \frac{\pi}{4}\right) + 0.91 \sin\left(\frac{\pi}{10}t + \frac{\pi}{4}\right) \end{cases} \quad (59)$$

Apart from the disturbance, model uncertainty is also introduced into the simulation. The proposed control scheme does not require the estimation of  $C(\theta, \dot{\theta})$ , the nominal  $\widehat{M}(\theta) = \begin{bmatrix} 0.8347 & 1.1478 \\ 0.8079 & 1.0124 \end{bmatrix} \cdot M(\theta)$ . The difference between  $M(\theta)$  and  $\widehat{M}(\theta)$  demonstrates the system uncertainty used to verify the robustness of the proposed SCSMC.

The parameters of the proposed SCSMC are carefully tuned in the light of *remarks 11~19* and detailed as follows: In the part of SMC,  $\beta_1 = \beta_2 = 0.65$ ,  $k_{s1} = k_{s2} = 1$ ,  $\gamma_1 = \gamma_2 = 2.5$ ,  $\alpha_1 = 2$ ,  $\alpha_2 = 3$ ,  $D_1 = D_2 = 0.0005$ ,  $\bar{k}_1 = \bar{k}_2 = 100$ ,  $\Omega_1 = \Omega_2 = 0.01$ ,  $\mu_1 = \mu_2 = 0.0001$ ,  $\kappa_V = 0.2758$ . In the part of RL,  $P = 5$ ,  $\mathbb{A} = \mathbb{B} = 11$ ,  $\eta_1 = \eta_2 = 0.1$ ,  $\lambda_1 = \lambda_2 = 0.6$ ,  $\Lambda(t < 20s) = 0.3$ ,  $\Lambda(t \geq 20s) = 0.1$ . The sets of the fuzzy input 1 of the 2 subsystems ( $\Gamma_{1,1}$  and  $\Gamma_{1,2}$ )

are:  $[\zeta_{1,1}^{(1)}, \zeta_{1,2}^{(1)}, \dots, \zeta_{1,\mathbb{A}}^{(1)}] = [\zeta_{1,1}^{(2)}, \zeta_{1,2}^{(2)}, \dots, \zeta_{1,\mathbb{A}}^{(2)}] = [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1] \times 0.0005$ . The sets of fuzzy input 2 of the 2 subsystems ( $\Gamma_{2,1}$  and  $\Gamma_{2,2}$ ) are:  $[\zeta_{2,1}^{(1)}, \zeta_{2,2}^{(1)}, \dots, \zeta_{2,\mathbb{B}}^{(1)}] = [\zeta_{2,1}^{(2)}, \zeta_{2,2}^{(2)}, \dots, \zeta_{2,\mathbb{B}}^{(2)}] = [-1, -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1] \times 0.00002$ . The action group for each rule is given as:  $U_{i,m} = \{u_{i,1}^m, \dots, u_{i,p}^m\} = \{0.01, 0.02, 0.03, 0.04, 0.05\}$ .

To verify the effectiveness of the proposed SCSMC, the ASMC [54] and conventional SMC with monotonically increasing switching gain are used as comparisons. The details of the ASMC [54] and SMC are shown in Table 2:

TABLE 2. Parameters of ASMC [54] and conventional SMC.

| Controller       | Formulation  | Parameters  |
|------------------|--|---|
| ASMC [54]        | $\tau = \widehat{M}(\theta)(v - \widehat{H} + \widehat{\theta}_r - K_s \dot{e})$ $v_i = -\beta_i s_i - \widehat{k}_i \text{sgn}(s_i)$ $\dot{\widehat{k}}_i$ $= \begin{cases} \varphi_i (\alpha_i^{-1})^\ell ( s_i  + \mathcal{K}_i e^{- s_i }) \mathcal{L}, & \text{if } \widehat{k}_i > 0 \\ \varphi_i \alpha_i^{-1} ( s_i  + k_B e^{- s_i }), & \text{else} \end{cases}$ $\mathcal{L} = \text{sgn}(\ s\ _\infty - \epsilon)$ | $K_s = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\beta_1 = \beta_2 = 100$ $\alpha_1 = \alpha_2 = 2$ $\mathcal{K}_1 = \mathcal{K}_2 = 1$ $\varphi_1 = 2, \varphi_2 = 3$ $\epsilon = 0.0005$ |
| Conventional SMC | $\tau = \widehat{M}(\theta)(v - \widehat{H} + \widehat{\theta}_r - K_s \dot{e})$ $v_i = -\beta_i s_i - \widehat{k}_i \text{sgn}(s_i)$ $\dot{\widehat{k}}_i = \gamma_i  s_i $   | $K_s = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\beta_1 = \beta_2 = 0.65$ $\gamma_1 = \gamma_2 = 2.5$   |

We consider 2 situations for the control of FSRMs, which are the uncertain system with the external disturbance of (59) and the uncertain system without any external disturbance ( $d_1 = 0, d_2 = 0$ ).

The sliding variables without any external disturbance are shown in Figure 7. The proposed SCSMC achieved the smallest chattering among 3 control schemes, which is followed by the ASMC [54] and the conventional SMC. Notably, ASMC obtains the fastest convergence of sliding variables to the manifold because of the big value selected for the proportional gain  $\beta_i$  in ASMC.

The tracking errors for 2 joints of FSRM without any external disturbance are presented in Figure 8. It is observed that both the ASMC and SCSMC have good tracking accuracy over the conventional SMC, and there is a negligible difference between the ASMC and SCSMC. Figure 9 shows

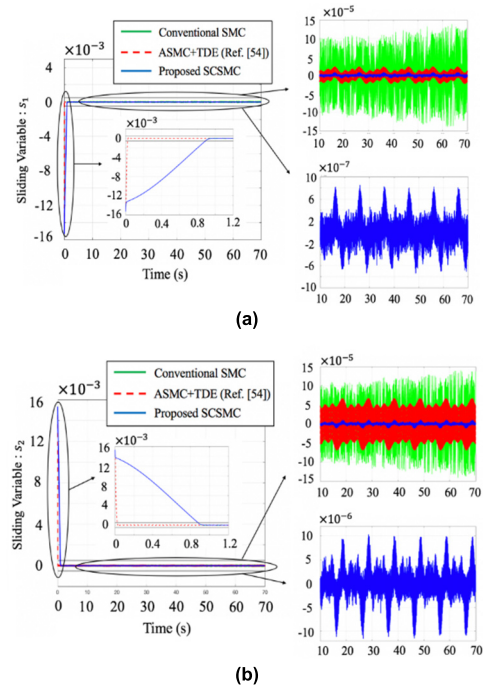


FIGURE 7. Dynamics of sliding variables without the disturbance ( $d_1 = 0, d_2 = 0$ ): (a).  $s_1$ . (b).  $s_2$ .

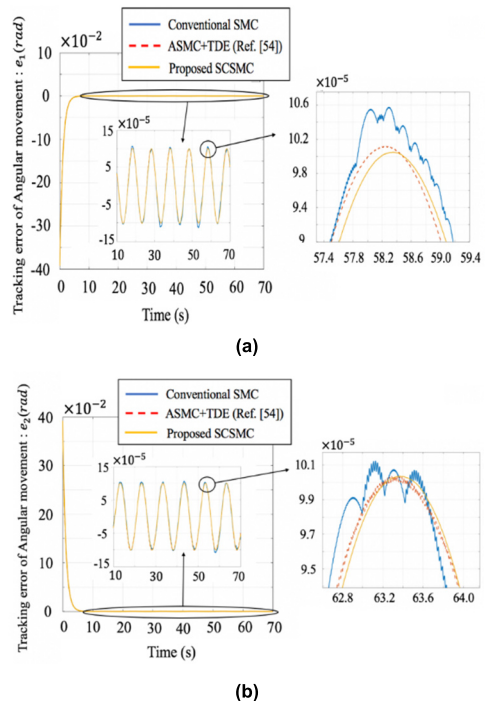
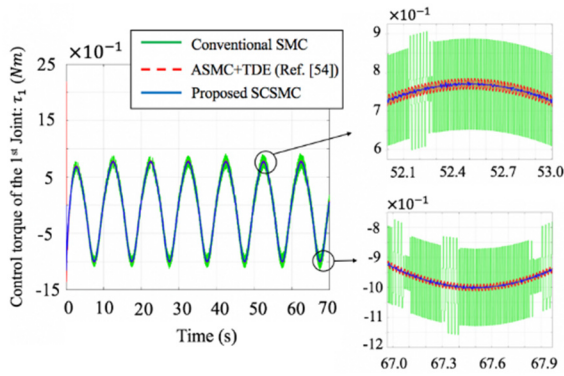


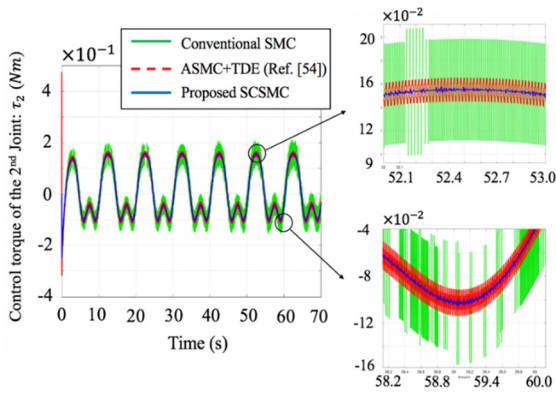
FIGURE 8. Tracking errors without any external disturbance ( $d_1 = 0, d_2 = 0$ ): (a). the 1<sup>st</sup> joint  $\theta_1$ . (b) the 2<sup>nd</sup> joint  $\theta_2$ .

the control inputs generated by conventional SMC, ASMC and SCSMC for the system of FSRM without any external disturbance. The smoothest control signals with the smallest chattering are achieved by SCSMC though the difference on control signals between SCSMC and ASMC is small.





(a)



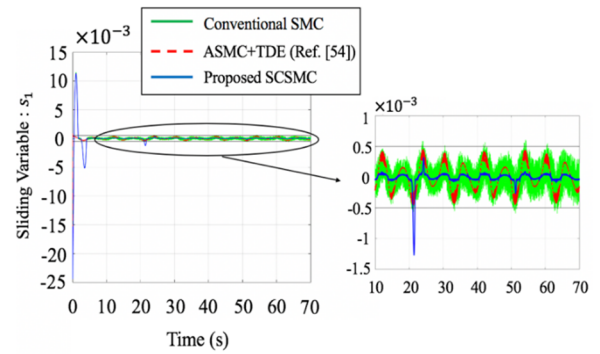
(b)

**FIGURE 9.** Control torques driving the manipulator without the external disturbance ( $d_1 = 0, d_2 = 0$ ): (a). the 1<sup>st</sup> joint. (b). the 2<sup>nd</sup> joint.

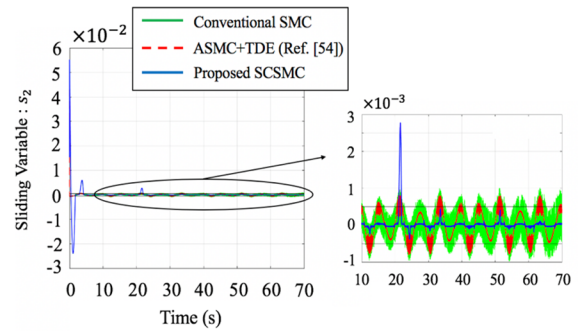
Both of SCSMC and ASMC are less chattering than conventional SMC. The performance of control signals by the 3 control schemes corresponds to their performances with the sliding variables.

Although the proposed SCSMC is superior to the conventional SMC in terms of tracking accuracy and smoothness of control signals, the SCSMC algorithm shows a slight advantage compared to ASMC in the absence of external disturbances. However, the story is different when the system of FSRMs is subject to external disturbances that bring up greater TDE errors.

Figure 10 presents the dynamics of the sliding variables with external disturbances. Clearly, the sliding variables in the proposed SCSMC achieve lower chattering effects over the conventional SMC. Moreover, less drifting is observed by these sliding variables when compared to the ASMC [54]. Although the significant chattering reduction is achieved by the ASMC [54], the sliding variables under ASMC [54] repetitively travel inside and outside the vicinity because of the excessive decrease in the switching gain to handle the TDE error, which results in the declined tracking accuracy. In the proposed SCSMC, the sliding variables travel outside the vicinity during the initial stage of control (e.g.  $t < 5s$  and  $t = 21s$ ), which could be caused by either the bad action



(a)



(b)

**FIGURE 10.** Sliding variables with the disturbance of (59): (a).  $s_1$  (b).  $s_2$ .

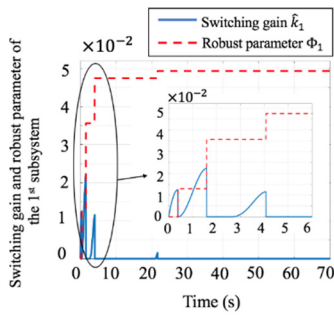
candidates tried by Q learning or the  $\Phi_i$  that have not been accumulated to a large enough value to handle TDE errors. However, the sliding variables stay inside the vicinity with less drifting effects and chattering effects after  $t > 21s$ , which means a high tracking accuracy.

Figure 11 shows the switching gain  $\hat{k}_i$  and the parameter of the robust term  $\Phi_i$  with the presence of external disturbance. It is clear that the values of switching gain are all transferred into the robust parameter  $\Phi_i$ . More precisely, the switching gain  $\hat{k}_i$  increases to overcome the TDE error and  $\Phi_i$  remains unchanged when the sliding variables are outside the vicinity. The switching gain  $\hat{k}_i$  decreases to reduce chattering effects and  $\Phi_i$  preserves the values forsaken by  $\hat{k}_i$  when the sliding variables are inside the vicinity. As a result,  $\hat{k}_i$  eventually declines to zero ( $t > 23s$ ) and  $\Phi_i$  inherits the ability to overcome the TDE errors.

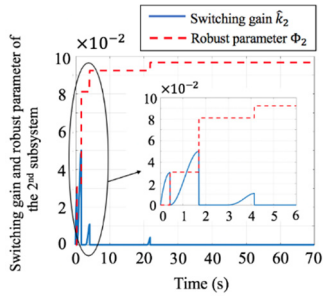
Figure 12 presents the tracking errors. Although the proposed SCSMC show greater errors in  $t = 22s$  than ASMC [54] and conventional SMC because of the sliding variables moving out the vicinities, the smallest tracking error after  $t = 22s$  is achieved by SCSMC. Therefore, the proposed SCSMC is superior to ASMC [54] and the conventional SMC in terms of steady state error (SSE).

The smallest tracking errors mean the best performance on tracking trajectories among 3 control schemes, shown in Figure 13.

Figure 14 shows the control torque that is applied to the joints of the space robotic manipulator, subject to external

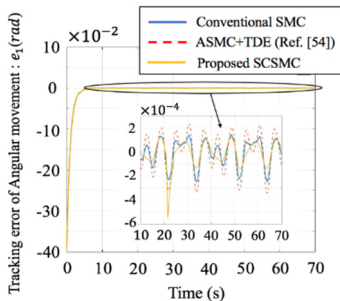


(a)

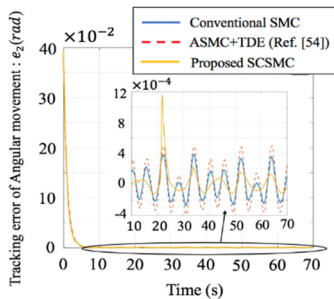


(b)

FIGURE 11. Switching gain  $\hat{k}_i$  and robust parameter  $\Phi_i$  with external disturbance: (a).  $i = 1$ . (b).  $i = 2$ .



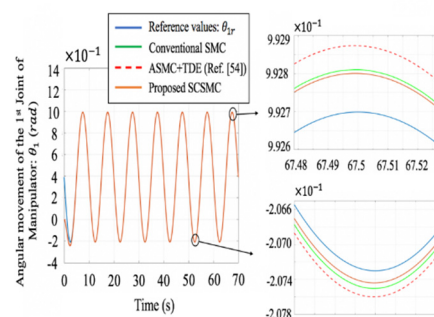
(a)



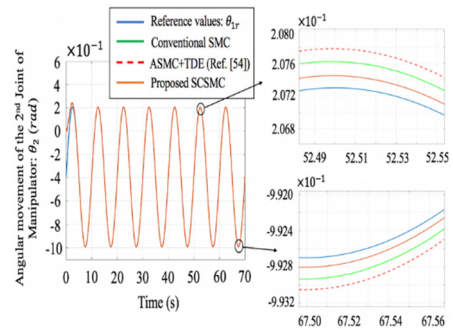
(b)

FIGURE 12. Tracking errors with the external disturbance of (59): (a). the 1<sup>st</sup> joint  $\theta_1$ . (b) the 2<sup>nd</sup> joint  $\theta_2$ .

disturbance. Clearly, both the proposed SCSMC and ASMC [54] can achieve smoother control torques over the conventional SMC, which corresponds to the mitigated chattering effects on the sliding variables. The difference in the smoothness of control torques between the proposed SMC and ASMC is small. In other words, the proposed SCSMC

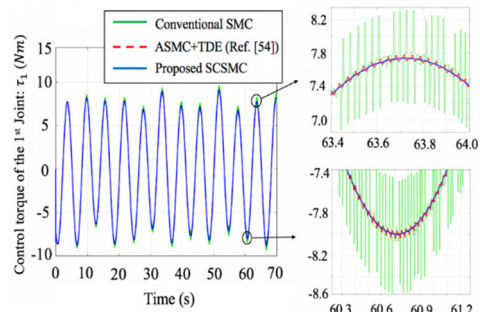


(a)

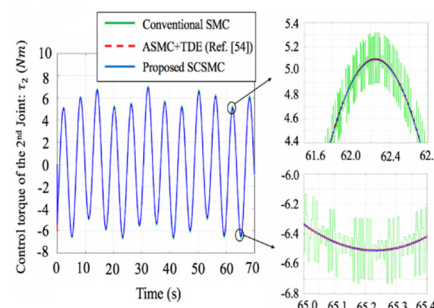


(b)

FIGURE 13. Tracking performance with external disturbance of (59): (a). the 1<sup>st</sup> joint  $\theta_1$ . (b) the 2<sup>nd</sup> joint  $\theta_2$ .



(a)



(b)

FIGURE 14. Control torques driving the manipulator with the external disturbance of (59): (a). the 1<sup>st</sup> joint. (b) the 2<sup>nd</sup> joint.

has a similar high tracking accuracy of the conventional SMC while also having a high control torque smoothness characteristic of the ASMC.

V. CONCLUSION

This paper proposed a new adaptive sliding mode control scheme with the aid of TDE and RL techniques for motion control of free-floating space robotic manipulators subject to model uncertainty and external disturbances. The novel adaptive law of switching gain can reduce chattering effects and the RL based robust term can prevent the sliding variables from drifting away from the manifold, which minimizes the chattering effect and improves the tracking accuracy. The tracking errors are proven to be UUB with an arbitrarily small bound and regardless of the optimal fuzzy rules, which not only ensures a high system stability, the RL could also operate in a safe environment to learn optimal policies. Simulation results have shown that the proposed SCSMC algorithm could obtain a smaller steady-state tracking error ( $< 2 \times 10^{-4}$  radians) than the other algorithms while retaining a smooth control torque characteristic of the ASMC algorithm.

VI. FUTURE WORK

Future work includes the development of the proposed SCSMC scheme where the FSRM system actuators have saturation limits, as demonstrated in [16] and [42]. Also, the strategy to enable the proposed SCSMC to handle measurement noise will be studied. Moreover, the research on designing observers of system states, such as [68] and [69], is included because it can allow the proposed control method to work even when the measurement devices fail.

APPENDIX

*Proof of Lemma 5:* The term of sliding variables in Lyapunov function (24) has a sufficient large number, namely:

$$\frac{1}{2}s^T s = R_1^* \tag{A1}$$

We assume the maximum element  $|s_{p^*}|$  in the vector  $s = [v_1, v_2, \dots, v_n]^T$  is less than such a positive number:

$$|s_{p^*}| = \max_{1 \leq i \leq n} \{|s_i|\} < \sqrt{\frac{2R_1^*}{n}} \tag{A2}$$

According to (A2), the following inequality can be derived:

$$\frac{1}{2}s^T s = \sum_{i=1}^n s_i^2 \leq \frac{n}{2}|s_{p^*}|^2 < R_1^* \tag{A3}$$

(A3) is contradict to (A1), which implies (A2) does not hold. Therefore, the maximum element  $|s_{p^*}|$  is not less than  $\sqrt{\frac{2R_1^*}{n}}$ , namely:

$$|s_{p^*}| \geq \sqrt{\frac{2R_1^*}{n}}. \tag{A4}$$

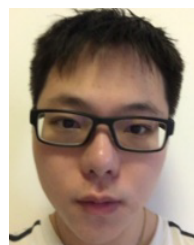
REFERENCES

[1] A. Flores-Abad, O. Ma, K. Pham, and S. Ulrich, "A review of space robotics technologies for on-orbit servicing," *Prog. Aerosp. Sci.*, vol. 68, pp. 1–26, Jul. 2014.  
 [2] T. Rybus, K. Seweryn, and J. Z. Sasiadek, "Control system for free-floating space manipulator based on nonlinear model predictive control (NMPC)," *J. Intell. Robotic Syst.*, vol. 85, nos. 3–4, pp. 491–509, 2017.

[3] X. Lu and Y. Jia, "Adaptive coordinated control of uncertain free-floating space manipulators with prescribed control performance," *Nonlinear Dyn.*, vol. 97, no. 2, pp. 1541–1566, Jul. 2019.  
 [4] E. Papadopoulos and S. Dubowsky, "On the nature of control algorithms for space manipulators," in *Proc. IEEE Proceedings. Int. Conf. Robot. Autom.*, May 1990, pp. 1102–1108.  
 [5] K. Yoshida and Y. Umetani, "Control of space manipulators with generalized Jacobian matrix," in *Space Robotics: Dynamics and Control*. Boston, MA, USA: Cham, Switzerland: Springer, 1993, pp. 165–204.  
 [6] S. Dubowsky and E. Papadopoulos, "The kinematics, dynamics, and control of free-flying and free-floating space robotic systems," *IEEE Trans. Robot. Autom.*, vol. 9, no. 5, pp. 531–543, Oct. 1993.  
 [7] Z. Chu, J. Cui, and F. Sun, "Fuzzy adaptive disturbance-observer-based robust tracking control of electrically driven free-floating space manipulator," *IEEE Syst. J.*, vol. 8, no. 2, pp. 343–352, Jun. 2014.  
 [8] M. Chu and X. Wu, "Modeling and self-learning soft-grasp control for free-floating space manipulator during target capturing using variable stiffness method," *IEEE Access*, vol. 6, pp. 7044–7054, 2018.  
 [9] L. Mostefai, M. Denai, and Y. Hori, "Robust tracking controller design with uncertain friction compensation based on a local modeling approach," *IEEE/ASME Trans. Mechatronics*, vol. 15, no. 5, pp. 746–756, Oct. 2010.  
 [10] Z.-G. Zhou, Y.-A. Zhang, and D. Zhou, "Robust prescribed performance tracking control for free-floating space manipulators with kinematic and dynamic uncertainty," *Aerosp. Sci. Technol.*, vol. 71, pp. 568–579, Dec. 2017.  
 [11] Y. Xu and H. Y. Shum, "Dynamic control of a space robot system with no thrust jets controlled base," Carnegie-Mellon Univ. Pittsburgh Pa Robot. Inst., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-91-33, 1991.  
 [12] C. C. Cheah, C. Liu, and J. J. E. Slotine, "Adaptive tracking control for robots with unknown kinematic and dynamic properties," *Int. J. Robot. Res.*, vol. 25, no. 3, pp. 283–296, Mar. 2006.  
 [13] O. Parlaktuna and M. Ozkan, "Adaptive control of free-floating space manipulators using dynamically equivalent manipulator model," *Robot. Auto. Syst.*, vol. 46, no. 3, pp. 185–193, Mar. 2004.  
 [14] S. Abiko and G. Hirzinger, "An adaptive control for a free-floating space robot by using inverted chain approach," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 2236–2241.  
 [15] Y.-L. Gu and Y. Xu, "A normal form augmentation approach to adaptive control of space robot systems," *Dyn. Control*, vol. 5, no. 3, pp. 275–294, Jul. 1995.  
 [16] S. Jia and J. Shan, "Finite-time trajectory tracking control of space manipulator under actuator saturation," *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 2086–2096, Mar. 2020.  
 [17] W. Zhang, X. Ye, L. Jiang, Y. Zhu, X. Ji, and X. Hu, "Output feedback control for free-floating space robotic manipulators base on adaptive fuzzy neural network," *Aerosp. Sci. Technol.*, vol. 29, no. 1, pp. 135–143, Aug. 2013.  
 [18] W. Zhang, N. Qi, J. Ma, and A. Xiao, "Neural integrated control for a free-floating space robot with suddenly changing parameters," *Sci. China Inf. Sci.*, vol. 54, no. 10, pp. 2091–2099, Oct. 2011.  
 [19] S. Guo, D. Li, Y. Meng, and C. Fan, "Task space control of free-floating space robots using constrained adaptive RBF-NTSM," *Sci. China Technol. Sci.*, vol. 57, no. 4, pp. 828–837, Apr. 2014.  
 [20] S. J. Yoo, J. B. Park, and Y. H. Choi, "Adaptive dynamic surface control of flexible-joint robots using self-recurrent wavelet neural networks," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 36, no. 6, pp. 1342–1355, Dec. 2006.  
 [21] Z.-L. Tang, S. S. Ge, K. P. Tee, and W. He, "Adaptive neural control for an uncertain robotic manipulator with joint space constraints," *Int. J. Control*, vol. 89, no. 7, pp. 1428–1446, Jul. 2016.  
 [22] P. Van Cuong and W. Y. Nan, "Adaptive trajectory tracking neural network control with robust compensator for robot manipulators," *Neural Comput. Appl.*, vol. 27, no. 2, pp. 525–536, Feb. 2016.  
 [23] F. Luan, J. Na, Y. Huang, and G. Gao, "Adaptive neural network control for robotic manipulators with guaranteed finite-time convergence," *Neurocomputing*, vol. 337, pp. 153–164, Apr. 2019.  
 [24] N. Wang and M. Joo Er, "Self-constructing adaptive robust fuzzy neural tracking control of surface vehicles with uncertainties and unknown disturbances," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 3, pp. 991–1002, May 2015.  
 [25] N. Wang, M. J. Er, J.-C. Sun, and Y.-C. Liu, "Adaptive robust online constructive fuzzy control of a complex surface vehicle system," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1511–1523, Jul. 2016.



- [26] C. Sun, H. Gao, W. He, and Y. Yu, "Fuzzy neural network control of a flexible robotic manipulator using assumed mode method," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5214–5227, Nov. 2018.
- [27] Y. Zhu, J. Qiao, and L. Guo, "Adaptive sliding mode disturbance observer-based composite control with prescribed performance of space manipulators for target capturing," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 1973–1983, Mar. 2019.
- [28] T. X. Dinh, T. D. Thien, T. H. V. Anh, and K. K. Ahn, "Disturbance observer based finite time trajectory tracking control for a 3 DOF hydraulic manipulator including actuator dynamics," *IEEE Access*, vol. 6, pp. 36798–36809, 2018.
- [29] W. Zheng and M. Chen, "Tracking control of manipulator based on high-order disturbance observer," *IEEE Access*, vol. 6, pp. 26753–26764, 2018.
- [30] Z. Chen, X. Yang, X. Zhang, and P. X. Liu, "Finite-time trajectory tracking control for rigid 3-DOF manipulators with disturbances," *IEEE Access*, vol. 6, pp. 45974–45982, 2018.
- [31] H. Jun-Pei, H. Qi, L. Yan-Hui, W. Kai, Z. Ming-Chao, and X. Zhen-Bang, "Neural network control of space manipulator based on dynamic model and disturbance observer," *IEEE Access*, vol. 7, pp. 130101–130112, 2019.
- [32] C. Zhongyi, S. Fuchun, and C. Jing, "Disturbance observer-based robust control of free-floating space manipulators," *IEEE Syst. J.*, vol. 2, no. 1, pp. 114–119, Mar. 2008.
- [33] K. Youcef-Toumi and O. Ito, "A time delay controller for systems with unknown dynamics," *J. Dyn. Syst., Meas., Control*, vol. 112, no. 1, pp. 133–142, Mar. 1990.
- [34] T. C. Hsia and L. S. Gao, "Robot manipulator control using decentralized linear time-invariant time-delayed joint controllers," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1990, pp. 2070–2075.
- [35] T. C. S. Hsia, T. A. Lasky, and Z. Guo, "Robust independent joint controller design for industrial robot manipulators," *IEEE Trans. Ind. Electron.*, vol. 38, no. 1, pp. 21–25, 1991.
- [36] S. Roy, I. N. Kar, J. Lee, and M. Jin, "Adaptive-robust time-delay control for a class of uncertain Euler–Lagrange systems," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 7109–7119, Sep. 2017.
- [37] S. Amirkhani, S. Mobayen, M. Ilae, O. Boubaker, and S. H. Hosseinnia, "Fast terminal sliding mode tracking control of nonlinear uncertain mass-spring system with experimental verifications," *Int. J. Adv. Robotic Syst.*, vol. 16, no. 1, Jan. 2019, Art. no. 172988141982817.
- [38] S. Mobayen, M. J. Yazdanpanah, and V. J. Majid, "A finite-time tracker for nonholonomic systems using recursive singularity-free FTSM," in *Proc. Amer. Control Conf.*, Jun. 2011, pp. 1720–1725.
- [39] D. Nojavanzadeh and M. Badamchizadeh, "Adaptive fractional-order nonsingular fast terminal sliding mode control for robot manipulators," *IET Control Theory Appl.*, vol. 10, no. 13, pp. 1565–1572, Aug. 2016.
- [40] Y. Zhao, P. Huang, and F. Zhang, "Dynamic modeling and super-twisting sliding mode control for tethered space robot," *Acta Astronautica*, vol. 143, pp. 310–321, Feb. 2018.
- [41] Y. Shtessel, M. Taleb, and F. Plestan, "A novel adaptive-gain supertwisting sliding mode controller: Methodology and application," *Automatica*, vol. 48, no. 5, pp. 759–769, May 2012.
- [42] S. Mobayen and F. Tchier, "Composite nonlinear feedback integral sliding mode tracker design for uncertain switched systems with input saturation," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 65, pp. 173–184, Dec. 2018.
- [43] M. B. R. Neila and D. Tarak, "Adaptive terminal sliding mode control for rigid robotic manipulators," *Int. J. Autom. Comput.*, vol. 8, no. 2, pp. 215–220, May 2011.
- [44] L. Hung and H. Chung, "Decoupled control using neural network-based sliding-mode controller for nonlinear systems," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1168–1182, May 2007.
- [45] M.-C. Pai, "Dynamic output feedback RBF neural network sliding mode control for robust tracking and model following," *Nonlinear Dyn.*, vol. 79, no. 2, pp. 1023–1033, Jan. 2015.
- [46] L. Shi, S. Kayastha, and J. Katupitiya, "Robust coordinated control of a dual-arm space robot," *Acta Astronautica*, vol. 138, pp. 475–489, Sep. 2017.
- [47] L. Zhang, Y. Wang, Y. Hou, and H. Li, "Fixed-time sliding mode control for uncertain robot manipulators," *IEEE Access*, vol. 7, pp. 149750–149763, 2019.
- [48] Q. Wu, D. Xu, B. Chen, and H. Wu, "Integral fuzzy sliding mode impedance control of an upper extremity rehabilitation robot using time delay estimation," *IEEE Access*, vol. 7, pp. 156513–156525, 2019.
- [49] S. Roy and I. N. Kar, "Adaptive sliding mode control of a class of nonlinear systems with artificial delay," *J. Franklin Inst.*, vol. 354, no. 18, pp. 8156–8179, Dec. 2017.
- [50] X. Luo, S. Ge, J. Wang, and G. Xinping, "Time delay estimation-based adaptive sliding-mode control for nonholonomic mobile robots," *Int. J. Appl. Math. Control Eng.*, vol. 1, no. 1, pp. 1–8, 2018.
- [51] M. Jin, Y. Jin, P. H. Chang, and C. Choi, "High-accuracy tracking control of robot manipulators using time delay estimation and terminal sliding mode," *Int. J. Adv. Robotic Syst.*, vol. 8, no. 4, p. 33, Sep. 2011.
- [52] J. Lee, P. H. Chang, and M. Jin, "Adaptive integral sliding mode control with time-delay estimation for robot manipulators," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6796–6804, Aug. 2017.
- [53] J. Baek, M. Jin, and S. Han, "A new adaptive sliding-mode control scheme for application to robot manipulators," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3628–3637, Jun. 2016.
- [54] S. Baek, J. Baek, and S. Han, "An adaptive sliding mode control with effective switching gain tuning near the sliding surface," *IEEE Access*, vol. 7, pp. 15563–15572, 2019.
- [55] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
- [56] C.-T. Lin and C. S. G. Lee, "Reinforcement structure/parameter learning for neural-network-based fuzzy logic control systems," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 1, pp. 46–63, 1st Quart., 1994.
- [57] K. L. Moore, "A reinforcement-learning neural network for the control of nonlinear systems," in *Proc. Amer. Control Conf.*, Jun. 1991, pp. 21–22.
- [58] C.-F. Juang and C.-M. Lu, "Ant colony optimization incorporated with fuzzy Q-Learning for reinforcement fuzzy control," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 39, no. 3, pp. 597–608, May 2009.
- [59] S. F. Desouky and H. M. Schwartz, " $Q(\lambda)$  learning adaptive fuzzy logic controllers for pursuit-evasion differential games," *Int. J. Adapt. Control Signal Process.*, vol. 25, no. 10, pp. 910–927, 2011.
- [60] A. Kumar and R. Sharma, "Linguistic Lyapunov reinforcement learning control for robotic manipulators," *Neurocomputing*, vol. 272, pp. 84–95, Jan. 2018.
- [61] W. Xu, B. Liang, and Y. Xu, "Survey of modeling, planning, and ground verification of space robotic systems," *Acta Astronautica*, vol. 68, nos. 11–12, pp. 1629–1649, Jun. 2011.
- [62] F.-C. Liu, L.-H. Liang, and J.-J. Gao, "Fuzzy PID control of space manipulator for both ground alignment and space applications," *Int. J. Autom. Comput.*, vol. 11, no. 4, pp. 353–360, Aug. 2014.
- [63] J. He, F. Xu, X. Wang, J. Yang, and B. Liang, "Modeling and control of free-floating space manipulator using the T-S fuzzy descriptor system approach," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 1049–1055.
- [64] A. Thowsen, "Uniform ultimate boundedness of the solutions of uncertain dynamic delay systems with state-dependent and memoryless feedback control," *Int. J. Control*, vol. 37, no. 5, pp. 1135–1143, May 1983.
- [65] S. Lin and W. Zhang, "Chattering reduced sliding mode control for a class of chaotic systems," *Nonlinear Dyn.*, vol. 93, no. 4, pp. 2273–2282, Sep. 2018.
- [66] M. Chen, S. S. Ge, and B. Ren, "Adaptive tracking control of uncertain MIMO nonlinear systems with input constraints," *Automatica*, vol. 47, no. 3, pp. 452–465, Mar. 2011.
- [67] A. Bonarini, A. Lazaric, F. Montrone, and M. Restelli, "Reinforcement distribution in fuzzy Q-learning," *Fuzzy Sets Syst.*, vol. 160, no. 10, pp. 1420–1443, May 2009.
- [68] J. Huang, X. Ma, X. Zhao, H. Che, and L. Chen, "Interval observer design method for asynchronous switched systems," *IET Control Theory Appl.*, vol. 14, no. 8, pp. 1082–1090, May 2020.
- [69] J. Huang, X. Ma, H. Che, and Z. Han, "Further result on interval observer design for discrete-time switched systems and application to circuit systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, early access, Dec. 5, 2019, doi: 10.1109/TCSII.2019.2957945.



**ZHICHENG XIE** was born in 1993. He received the M.S degree from the School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Sydney, Australia, in 2019, where he is currently pursuing the Ph.D. degree. His research interest includes the intelligent control for uncertain nonlinear systems





**TAO SUN** was born in 1991. He received the M.S. degree from the School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Sydney, Australia, in 2019, where he is currently pursuing the Ph.D. degree. His areas of research include stereo-vision based navigation and model estimation.



**ZHONGCHENG MU** graduated from the Harbin Institute of Technology in China. He is currently an Associate Professor, a Senior Satellite System Engineer, a Deputy Director of the Smart Satellite Technology Center (SSTC) of SJTU. He has published more 30 articles and five patents. research topics are intelligent satellite design, satellite mission schedule, and satellite formation.



**TREVOR HOCKSUN KWAN** received the Ph.D. degree with AMME, The University of Sydney, in 2017. He completed a postdoctoral research position at the School of Aerospace and Astronautics, Sun Yat-sen University, in 2019, and is currently an Associate Researcher with the Department of Thermal Science and Energy Engineering, University of Science and Technology of China, Hefei, China. His research interests include the analysis, optimization, and systems integration of renewable energy systems including the fuel cell, thermoelectric cooler, and solar energy systems.



**XIAOFENG WU** (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from Loughborough University, U.K. He is currently a Senior Lecturer in space engineering with the School of Aerospace, Mechanical and Mechatronic Engineering. Since 2019, he has been an Visiting Professor with the School of Aeronautics and Astronautics, Shanghai Jiao Tong University, China. His current researches focus on embedded systems for aerospace applications, and the navigation and control of spacecraft.

...