Proceedings of the 14th Asian Control Conference
July 5-8, 2024, Dalian, China

Safe Stabilization Control for Interconnected Virtual-Real Systems via Model-based Reinforcement Learning

Junkai Tan
School of Electrical Engineering, Xi'an Jiaotong University,
Xi'an, China
15958024@stu.xjtu.edu.cn

Shuangsi Xue
School of Electrical Engineering, Xi'an Jiaotong University,
Xi'an, China
xssxjtu@stu.xjtu.edu.cn

Huan Li
School of Electrical Engineering, Xi'an Jiaotong University,
Xi'an, China
lh2000dami@stu.xjtu.edu.cn

Hui Cao
School of Electrical Engineering, Xi'an Jiaotong University,
Xi'an, China
huicao@mail.xjtu.edu.cn

Dongyu Li
School of Cyber Science and Technology, Beihang University,
Beijing, China
dongyuli@buaa.edu.cn

***Abstract*—In this paper, a safe-guarding controller is designed for the interconnected virtual-real system based on a reinforcement learning framework to achieve stabilization control. We established the mathematical formulation of the interconnected virtual-real system and the safety-guaranteed stabilization optimization problem. Online reinforcement learning methods are utilized to solve the Hamilton-Jacobi-Bellman(HJB) equation on the established optimal control problem. The safe-guarding term is introduced to achieve safe-guarding control for the real part. Single network is used to approximate the value function. Concurrent Learning methods are introduced to train the network without excitation risks. We prove that the dynamics of the estimation error of the designed critic network are uniform and ultimately bounded. Finally, a numerical simulation example is provided to illustrate the effectiveness of the proposed control method.**

***Index Terms*—Interconnected virtual-real system, safety-guaranteed, stabilization control, reinforcement learning.**

I. Introduction

Interconnected systems have attracted extensive attention in various fields due to the diverse application scenarios [1]–[3]. With the integration of virtual and real worlds in many fields, interconnected virtual-reality systems have gradually become an important research focus. This system, which combines virtual elements and real-world components, builds a unique and powerful form of combination. Nevertheless, due to its complex composition and dynamic characteristics, the stabilization and control problem of the interconnected virtual-reality system has become particularly complicated. The stable operation of the system is directly related to safety issues. Especially in the real world, any failure or instability may lead to serious losses. Therefore, it is an essential task to explore the safety stabilization and control of interconnected virtual-reality systems.

There are many researches on stabilization control for interconnected systems that have been published. In [4], stabilization control of a type of nonlinear interconnected systems is implemented based on an optimal control method. In [5], the tracking control of nonlinear interconnected systems is modified into the regulation problem of the augmented subsystem and solved. Combining the continuous and event sampling feedback information, an approximate optimal distributed control strategy is proposed for a class of nonlinear interconnected systems with strong interconnectivity in [6]. A behavioral macro modeling approach for electronic interconnect systems in [7], where simulated behavioral modeling is used to capture the frequency-dependent properties and nonlinear termination of multiconductor interconnects. In [8], a new decentralized adaptive strategy for fractional-order interconnect systems with unknown relation between subsystems is proposed. The state estimation and unconstrained reconstruction of unknown inputs for a class of interconnected systems with uncertain outputs and unknown input distribution matrices in [9]. However, a clear gap exists in the research on stabilization control for interconnected virtual-real systems.

Reinforcement learning-based methods can learn optimal policies through interactions with the environment, which makes them well-suited for addressing the challenges posed by interconnected virtual-real systems. A reinforcement learning-based method for the control of nonlinear systems is introduced in [10], which provides a framework with an adjustable policy learning rate. Based on a novel hybrid reinforcement learning scheme, [11] proposed a distributed control scheme consisting of uncertain input nonlinear subsystems. By using a reinforcement learning method with both slow and fast sampled data, [12] constructed a data-driven sliding mode tracking control strategy that avoids numerical problems

caused by the coupling of different dynamics. Nevertheless, traditional reinforcement learning-based methods focus on performance optimization without explicit consideration of safety constraints [13], [14], which can lead to unsafe behavior during the learning process.

Safe-Guarding Control is a critical aspect of interconnected virtual-reality systems, as failures or accidents can lead to severe consequences. Ensuring the safe operation of these systems is challenging due to the complex interactions between virtual and real subsystems and the need to satisfy various security constraints [15]. Several studies have proposed solutions, such as the enhancement of learning-based control policies with a novel control barrier function to ensure safety [16], or the development of an intermittent framework for safe reinforcement learning [1]. Learning barrier-certified security controllers were introduced in [17]. A secure pursuit-avoidance game for adaptation to unknown clutter environments is developed in [18]. To conclude, the control barrier function (CBF) is vital in safe-guarding control.

Inspired by the above discussion, this paper proposed a reinforcement learning-based stabilization control strategy for a novel type of interconnected virtual-real systems. Our approach builds upon recent advances in safe reinforcement learning and concurrent learning methods. First, the online reinforcement learning approach is used to solve the HJB equations on the optimal control problem, providing an innovative solution that enhances system stability and responsiveness. To achieve safety control for the real part, a safe-guarding term generated by the control barrier function is introduced, which mitigates risk and enhances overall system reliability. Finally, using the single neural network technique with concurrent learning methods to approximate the value functions, optimized system performance by facilitating more accurate and efficient function approximations.

The remainder of this paper is organized as follows: Section II presents the system description and basic concepts are presented. Section III presents the main results of this paper, including the safe-guarding controller design, the RL-based framework and the stability analysis. In Section IV we present the numerical simulation. Finally, the conclusions of this paper are presented in Section V.

II. Problem Statement

This paper mainly studies a novel type of interconnected systems, defined as virtual-real systems. The systems are composed of $2N$ subsystems and the dynamics are described as:

$$\dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))\left(u_i(x_i(t)) - \mathbf{I}(x(t))\right), \quad (1)$$

where $i = 1, ..., 2N$, the $x_i(t) \in \mathbb{R}^n$ are the state of each subsystem, $u_i(x_i(t)) \in \mathbb{R}^n$ are the control input vectors of the i th subsystem, and $x(t) = [x_1(t), \ldots, x_{2N}(t)]$ is a state vector composed of all the subsystem states.

We define the $[x_1, ...x_N]$ as the virtual part and the $[x_{N+1}(t), ...x_{2N}(t)]$ as the real part. Furthermore, $f_i(x_i)$ represent the nonlinear internal functions, $g_i(x_i)$ represent the input gain matrix, and $\mathbf{I}(x(t))$ is the interconnected term upper bounded by $\hat{I}$.

Assumption 1: The funtion $f_i(\cdot)$ and $g_i(\cdot)$ are locally Lipschitz.

In addition, we set $x_i = 0$ to be the equilibrium state of the ith subsystem with $i = 1, \ldots, 2N$. We assume the control input vector $u_i(x_i) = 0$ when $x_i = 0$, where $i = 1, 2, \ldots, 2N$, which means that the control input stops once the system reaches the equilibrium point. The system structure is demonstrated in Fig. 1.

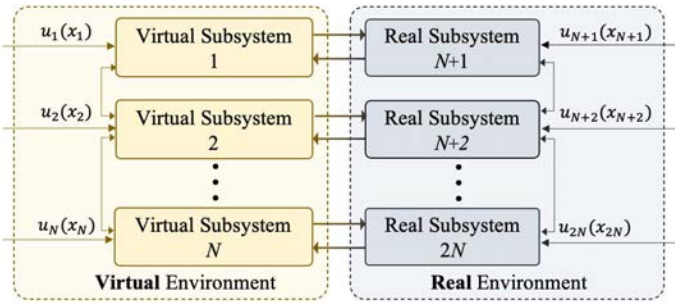

Fig. 1. Structure of the interconnected virtual-real systems.

For efficient subsequent controller design, we give the following definitions:

Definition 1: (Forward invariant) To present the safe-guarding control, we introduce the forward invariant property of a set $\mathbf{s} \subset \mathbb{R}^n$. For a pre-defined period $\mathcal{I}(x_0)$ and any $x_0 \in \mathbf{s}$, if system dynamic satisfies $x(t) \in \mathbf{s}$ for any $t \in \mathcal{I}(x_0)$. We define set s as a forward invariant set and can be separated as interior part and boundary part as

$$\begin{aligned} \mathbf{s} &= \{x \in \mathbb{R}^n \mid h(x) \geq 0\}, \\ \partial\mathbf{s} &= \{x \in \mathbb{R}^n \mid h(x) = 0\}, \\ \mathrm{Int}(\mathbf{s}) &= \{x \in \mathbb{R}^n \mid h(x) > 0\}, \end{aligned} \quad (2)$$

where $h \in \mathbb{R}^n$ is the boundary function, which vanishes in the boundary of s. If the system states in the period $t \in \mathcal{I}(x_0)$ satisfies $x(t) \in \partial\mathbf{s}$, we mark the system as operating in a safe region.

Definition 2: A continuous function $b(x)$ is called a barrier function with the following properties:

1) The function $b(x)$ does not go to infinity when $x(t) \in \mathrm{Int}(s)$, that is, $|b(x)| < \infty$.
2) As the state x approaches the boundary of the forward invariant set, the function $b(x)$ goes to infinity, expressed as $\lim_{z \to \partial\mathbf{s}} b(x) = \infty$.
3) The equilibrium value of the barrier function vanishes, that is, $b(0) = 0$.

Control objective: Consider the system (1) and the set $\mathbf{s} \subset \mathbb{R}^n$. Given the control barrier function b as defined in Definition 2. The control objective of this paper is to design a control strategy u_i such that $\mathrm{Int}(\mathbf{s})$ is forward-invariant for the system (1), and all the subsystems asymptotically converges to the origin.

III. Main Results

A. Optimal Controller Design

To achieve the control objective of the interconnected system, we first design the optimal control strategy that steers the system to the equilibrium point. Since the two subsystems eventually reach the same target point in the control process, we design the optimal controllers of the two subsystems together. Thus, considering $2N$ isolated subsystems corresponding to system (1), denoted as:

$$\dot{x}_i(t) = f_i\left(x_i(t)\right) + g_i\left(x_i(t)\right) u_i\left(x_i(t)\right), \tag{3}$$

where $i = 1, 2, \ldots, 2N$. Then we make the assumption that every individual subsystem i is controllable and that there is a continuous control strategy on $\Omega \in \mathbb{R}^n$ that is capable of asymptotically stabilizing the individual subsystem i. To deal with the optimal control problem on the infinite horizon, the design objective is to find the control policy $u_i\left(x_i\right)$ with $i = 1, \ldots, 2N$ that minimizes the local cost function as:

$$J_i(x_i, u(\cdot)) = \int_0^\infty Q_i^2(x_i(\tau)) + u_i^T(x_i(\tau)) R_i u_i(x_i(\tau)) d\tau \tag{4}$$

where $i = 1, 2, \ldots, 2N$ and $Q_i\left(x_i\right)$ is a positive definite function with $q_i\left(x_i\right) \leq Q_i\left(x_i\right)$.

According to the optimal control theory, the planned feedback control should stabilize the subsystem on Ω_i, $i = 1, \ldots, 2N$ and ensure that the cost function is finite. For any set of control policies $\mu_i \in \Phi_i(\Omega_i)$, $i = 1, \ldots, 2N$ that can achieve the aforementioned objective, if the associated cost function $\hat{J}_i(x_i, \mu(\cdot))$ is continuously differentiable, which given as:

$$\hat{J}_i(x_i, \mu(\cdot)) = \int_0^\infty Q_i^2(x_i(\tau)) + \mu_i^T(x_i(\tau)) R_i \mu_i(x_i(\tau)) d\tau. \tag{5}$$

Then the Hamiltonian function for each isolated subsystem is obtained by taking derivatives on both sides of (5) as:

$$\begin{aligned} H_i(x_i, \mu_i, \nabla \hat{J}_i(x_i, \mu_i)) =& Q_i^2(x_i) + \mu_i^T(x_i) R_i \mu_i(x_i) \\ &+ (\nabla \hat{J}_i(x_i))^T (f_i(x_i) + g_i(x_i)\mu_i(x_i)) \end{aligned} \tag{6}$$

where $i = 1, \ldots, 2N$. The optimal cost function for each subsystem can be expressed as:

$$\begin{aligned} V_i = \min_{\mu_i \in \Phi_i(\Omega_i)} \int_0^\infty \{ & Q_i^2(x_i(\tau)) \\ & + \mu_i^T(x_i(\tau)) R_i \times \mu_i(x_i(\tau))\} d\tau, \end{aligned} \tag{7}$$

where $i = 1, \ldots, 2N$ and $J_i^*\left(x_i\right)$ satisfies the HJB equation $0 = \min_{\mu_i \in \Psi_i(\Omega_i)} H_i(x_i, \mu_i, \nabla V_i)$, with $V_i = \partial V_i / \partial x_i$. The optimal control policy for the $2N$ subsystems can then be derived as follows:

$$\begin{aligned} u_i^*\left(x_i\right) &= \arg \min_{\mu_i \in \Psi_i(\Omega_i)} H_i\left(x_i, \mu_i, \nabla V_i\right) \\ &= -\frac{1}{2} R_i^{-1} g_i^T\left(x_i\right) \nabla V_i, i = 1, \ldots, 2N. \end{aligned} \tag{8}$$

To establish a stable control scheme for the interconnected system (1), we now show that the asymptotic stability of the isolated subsystem can be guaranteed by proportionally increasing the feedback gains. Here we present the following lemma:

Lemma 1: Considering the isolated subsystems in (3), the following feedback control law could ensure that the $2N$ subsystems are asymptotically stable:

$$\begin{aligned} \hat{u}_i =& \beta_i u_i^*\left(x_i\right) \\ =& -\frac{1}{2} \beta_i R_i^{-1} g_i^T\left(x_i\right) \nabla \hat{J}_i(x_i), \end{aligned} \tag{9}$$

where $i = 1, \ldots, 2N$ and β_i are positive constants satisfies $\beta_i \geq \frac{1}{2}$.

proof: We prove the asymptotic stability of the controller by showing the value function $\hat{J}_i(x_i)$, $i = 1, 2, \ldots, 2N$ are Lyapunov candidates. First, according to (5), we can find that $\hat{J}_i(x_i) > 0$ when any $x_i \neq 0$ and $\hat{J}_i = 0$ when $x_i = 0$, which implies that $\hat{J}_i(x_i), i = 1, 2, \ldots, 2N$ are positive definite.

Then we substituting the optimal control policies (24) into the infinitesimal version of (7), and the HJB equations in terms of $\nabla \hat{J}_i(x_i), i = 1, 2, \ldots, 2N$ can be obtained as:

$$\begin{aligned} 0 =& Q_i^2\left(x_i\right) + \left(\nabla \hat{J}_i(x_i)\right)^T f_i\left(x_i\right) \\ & - \frac{1}{4} \left(\nabla \hat{J}_i(x_i)\right)^T g_i(x) R_i^{-1} g_i^T\left(x_i\right) \nabla \hat{J}_i(x_i) \\ =& Q_i^2\left(x_i\right) + \left(\nabla \hat{J}_i(x_i)\right)^T f_i\left(x_i\right) \\ & - \frac{1}{4} \left\| R_i^{-1/2} g_i^T\left(x_i\right) \nabla \hat{J}_i(x_i)) \right\|^2 \end{aligned} \tag{10}$$

Denoting $\mathbb{J} = \left\| R_i^{-1/2} g_i^T\left(x_i\right) \nabla \hat{J}_i(x_i)) \right\|^2$, we have $\mathbb{J} \geq 0$. Then we calculate the time derivatives of $\hat{J}_i(x_i), i = 1, 2, \ldots, 2N$ along the isolated subsystem, adding and subtracting $(1/2)\left(\nabla J_i^*\left(x_i\right)\right)^T g_i\left(x_i\right) u_i^*\left(x_i\right)$ and consideing (8), (9) and (10) we have:

$$\begin{aligned} \dot{\hat{J}}_i(x_i) =& (\nabla \hat{J}_i(x_i))^T \dot{x}_i \\ =& (\nabla \hat{J}_i(x_i))^T (f_i(x_i) + g_i(x_i)\hat{u}_i(x_i)) \\ =& \left[(\nabla \hat{J}_i(x_i))^T f_i\left(x_i\right) - \frac{1}{4}\mathbb{J} \right] - \frac{1}{2}\left(\beta_i - \frac{1}{2}\right)\mathbb{J} \\ =& -Q_i^2\left(x_i\right) - \frac{1}{2}\left(\beta_i - \frac{1}{2}\right)\mathbb{J} \end{aligned} \tag{11}$$

With the condition $\beta_i \geq \frac{1}{2}$, we have $\dot{\hat{J}}_i(x_i) < 0$. Thus, the conditions for the Lyapunov theory of local stability are obtained and the proof is completed.

B. Safe-Guarding Control Design

As the focus of this paper, the safety of the real part of the interconnected system during the stabilization process is of primary importance. In this section, we adopt the control barrier function to achieve safe control of the system.

To facilitate the subsequent design of the safe-guarding controller. We select the barrier function $b(x_i)$ in the form of

$$b(x_i) = \left(\frac{1}{h(x)} - \frac{1}{h(0)}\right)^2, \ i = N+1, \ldots, 2N, \tag{12}$$

where $h(x_i)$ is the continuous boundary function that ensures $b(x_i)$ meet all three properties in Definition 2. In order to ensure the forward invariance of $\mathbf{s}$, we establish the condition on $b(x_i)$ by the following Lemma.

Lemma 2: [16] We consider the system (1) and assume that the origin is contained in the set $\mathrm{Int}(\mathbf{s})$. As long as $b(x) < \infty$ in all $t \in \mathcal{I}(x_0)$, then $\mathrm{Int}(\mathbf{s})$ is forward invariant for the system (1).

Next, we introduce the safety-guarding term of the real part, inspired by the [16], we directly give the safety controller as follows:

$$u_b^*(x_i) = -\alpha_i g_i^T(x_i)\nabla b^T(x_i),\ i = N+1,\ldots,2N, \quad (13)$$

where α_i is the selected control gain and $b(x)$ is the barrier function as we defined in (12).

To place a limit on the growth of drift in the system and to ensure that the direction of control is not lost where control is needed to keep the system safe, we make the following assumptions:

Assumption 2: (1) There exists a positive, non-decreasing function l such that $\|f_i(x_i)\| \le l(\|x_i\|)\|x_i\|$ for all $x_i \in \mathbf{s}$ and $\lim_{x\to\partial\mathbf{s}} l(\|x_i\|)\|x_i\| < \infty$. (2) There exists a positive constant $\underline{g}$ such that $\underline{g} \le \|g_i(x_i)\|$ for all $x_i \in \mathbf{s}$. (3) There exists a neighborhood of $\partial\mathbf{s}$, denoted by $\mathbf{N}(\partial\mathbf{s})$, such that $0 \notin \mathbf{N}(\partial\mathbf{s})$ and $\|\nabla b(x)g_i(x)\| \neq 0$ for all $x_i \in \mathbf{N}(\partial\mathbf{s})$.

To show that under Assumption 2, the control policy makes $\mathrm{Int}(\mathbf{s})$ forward invariant for system (1), we present the following theorem.

Theorem 1: Consider system (1), a set $\mathbf{s}$ that satisfies the previously stated conditions. Also, set b be a candidate control barrier function of system (1) on $\mathbf{s}$. Suppose that the Assumption 2 holds, the controller $u_b^*(x_i)$ makes the closed-loop system (1) forward invariant on $\mathrm{Int}(\mathbf{s})$.

proof: We begin the proof by taking the derivative of $b(x_i)$ along the system dynamic as:

$$\begin{aligned}\dot{b}(x_i) =&\nabla b(x_i)(f_i(x_i(t)) + g_i(x_i)\,(u_b^*(x_i) + \mathbf{I}(x)))\\ =&\nabla b(x_i)f_i(x_i - \alpha_i\|\nabla b(x_i)g_i(x_i)\|^2\\ &- \nabla b(x_i)g_i(x_i)\mathbf{I}(x(t)).\end{aligned} \quad (14)$$

We suppose that Assumption 2 holds, then we have:

$$\begin{aligned}\dot{b}(x_) \le&\|\nabla b(x_i)\|l(\|x_i\|)\|x\| - \alpha_i\|\nabla b(x_i)\|^2\underline{g}^2\\ &- \|\nabla b(x_i)\|\underline{g}\hat{I}\\ =&\|\nabla b(x_i)\|^2\cdot\left(\frac{l(\|x_i\|)\|x_i\| - \underline{g}\hat{I}}{\|b(x_i)\|} - \alpha_i\underline{g}\right).\end{aligned} \quad (15)$$

With $x \to \partial\mathbf{s}$, we take the limits in (15) and have:

$$\lim_{x\to\partial\mathbf{s}} \dot{b}(x_i) = -\infty < 0, \quad (16)$$

This result rules out the existence of trajectories of the system into $\mathbf{s}$, thus the proof are completed.

The feedback control for the real part are rewritten into:

$$u_i = \beta_i u_i^*(x_i) + \gamma_i u_b^*(x_i),\ i = N+1,\ldots,2N, \quad (17)$$

where γ_i are positive constants set by users. According to Theorem 1 in [4], by choosing the appropriate γ_i and β_i, the designed controller is sufficient to make the interconnected system achieve asymptotic stability. Due to our safety-guaranteed term, the real part of the system can operate within set $\mathbf{s}$, that is, always in a safe region.

C. Concurrent Learning

To obtain the analytical solution of control policies u_i and value functions V_i, we utilize a single critic network for the approximation of value function V_i, which in the form of

$$V_i = W_i^T\phi(x) + \epsilon_i(x), \quad (18)$$

where $W_i \in \mathbb{R}^{p_i}$ is the ideal weight for the single network and $\phi(x) \in \mathbb{R}^{n\times p_i}$ is the vector of the activation function, p_i is the hidden layer neuron number and $\epsilon_i(x)$ is the critic network's approximation error. The estimated approximation of the ideal value function V_i is defined as

$$\hat{V}_i = \hat{W}_i^T\phi(x), \quad (19)$$

where $\hat{W}_i \in \mathbb{R}^{p_i}$ is the estimated weight of the single network, which is implemented in the single network to estimate the actual value of V_i. To reduce the computational load, the approximation of control u_i is implemented through the single network method as:

$$u_i = -\frac{1}{2}R_i^{-1}g_i^T(\nabla\phi_i^T(x)W_i + \nabla\epsilon_i^T(x)). \quad (20)$$

With the estimated value's gradient using the weights W_i in (19), the actual controller can be expressed as:

$$\hat{u}_i = -\frac{1}{2}R_i^{-1}g_i^T\nabla\phi_i^T(x)\hat{W}_i. \quad (21)$$

Based on (20)), and (21), we can define the error of approximating the Hamilton-Jacobi equation as

$$\begin{aligned}H_i(x,u_i,W_i) =&u_i^TR_iu_i + Q_i^2\\ &+ [W_i^T\nabla\phi_i + (\nabla\epsilon_i)^T](f_i + g_iu_i),\\ =&-\nabla\epsilon_i^T(f_i + g_iu_i),\end{aligned} \quad (22)$$

$$\begin{aligned}H_i(x,\hat{u}_i,\hat{W}_i) =&\hat{u}_i^TR_i\hat{u}_i + Q_i^2 + (\hat{W}_i^T\nabla\phi_i)(f_i + g_i\hat{u}_i)\\ =&e_i.\end{aligned} \quad (23)$$

To simplify the notation, we denote $\nabla\epsilon_i^T(f_i+g_iu_i) = -e_{H,i}$ and $\omega_i = \nabla\phi_i(f_i + g_i\hat{u}_i)$. To obtain an admissible control policy u and facilitate the following optimization, we first combine the historical and instantaneous data in the form of the total energy-like objective E_i expressed as

$$E_i = \frac{1}{2}\left[\frac{e_i^2}{\left(1+\omega_i^T\omega_i\right)^2} + \sum_{k=1}^{M}\frac{(e_i^k)^2}{\left(1+(\omega_i^k)^T\omega_i^k\right)^2}\right], \quad (24)$$

where ω_i^k is the k-th historical data of ω_i. M is the total number of historical data. Let $\overline{\omega}_i = \left[\omega_i^1 ... \omega_i^M\right]$ be the historical data stack.

According to the property of the above objective function, we can obtain the adaptation law based on least squares for the estimated critic network weight $\hat{W}_i$ as follows.

$$\begin{aligned}\dot{\hat{W}}_i =& -a_i \frac{\partial E_i}{\partial \hat{W}_i} \\ =& -a_i \frac{\omega_i e_i}{\left(1+\omega_i^T \omega_i\right)^2} - a_i \sum_{k=1}^{M} \frac{\omega_i^k e_i^k}{\left(1+(\omega_i^k)^T \omega_i^k\right)^2},\end{aligned} \tag{25}$$

where a_i is the learning gain for each subsystem, determine the convergence speed of each single network weight W_i of the subsystem.

D. Stability Analysis

This section presents the Lyapunov stability analysis to investigate the stability of the proposed concurrent learning law. The error dynamic of $\tilde{W}_i$ is given as:

$$\begin{aligned}\dot{\tilde{W}}_i(t) =& -a_i \frac{\omega_i}{\omega_i^T \omega_i + 1}\left[\frac{\omega_i^T \tilde{W}(t) + e_{H,i}}{\omega_i^T \omega_i + 1}\right] \\ & - a_i \sum_{k=1}^{M} \frac{\omega_i^k}{(\omega_i^k)^T \omega_i^k + 1}\left[\frac{\omega^T(t_i)\tilde{W}(t) + e_{H,i}^k}{(\omega_i^k)^T \omega_i^k + 1}\right].\end{aligned} \tag{26}$$

Theorem 1: Critic Weights is uniformly ultimately bounded (UUB) under the following conditions:

1.$rank(\overline{\omega}_i) = p_i$;

2.$e_{H,i}$ is upper bounded by $e_{Hmax,i}$.

Proof: We define the following Lyapunov function for stability analysis:

$$V_i(t) = \frac{1}{2a_i}\tilde{W}_i^T \tilde{W}_i. \tag{27}$$

For each subsystem, we have

$$\dot{V}_i = -\tilde{W}_i^T \left(\zeta_a(t) + \zeta_i(t)\right)\tilde{W}_i + \tilde{W}_i^T \eta_i, \tag{28}$$

with

$$\zeta_a(t) = \frac{\omega_i \left(\omega_i\right)^T}{\left[1+\left(\omega_i\right)^T \omega_i\right]^2}, \tag{29}$$

$$\zeta_i = \sum_{k=1}^{p} \frac{\omega_i \left(\omega_i^k\right)^{\mathrm{T}}}{\left[1+\left(\omega_i^k\right)^{\mathrm{T}} \omega_i\right]^2}, \tag{30}$$

$$\eta_i = \frac{\omega_i e_{H,i}}{\left[1+\left(\omega_i\right)^T \omega_i\right]^2} + \sum_{k=1}^{p} \frac{\omega_i^k e_{H,i}^k}{\left[1+\left(\omega_i^k\right)^T \omega_i^k\right]^2}. \tag{31}$$

With $\zeta_a > 0$, we have:

$$\dot{V}_i \leq -\tilde{W}_i^T \zeta_i(t) \tilde{W}_i + \tilde{W}_i^T \eta_i. \tag{32}$$

With the assumption that $rank(\overline{\omega}_i) = k$, we get

$$\dot{V}_i \leq -\lambda_{min}(\zeta_i)\|\tilde{W}_i\|^2 + \|\tilde{W}_i\|\left(\frac{M+1}{2}\right)e_{Hmax,i}. \tag{33}$$

$\dot{V}_i$ can be guarranteed negative if $\|\tilde{W}_i\| \geq \frac{(M+1)e_{Hmax,i}}{2\lambda_{min}(\zeta_i)}$. As a result, the critics weights' error dynamic is UUB, and the proof is completed.

IV. Simulation

In this section, we visually demonstrate the effectiveness of the proposed control method through numerical experiments. In this paper, we consider the following consisting of two interconnected subsystems, where the system uses the same dynamic parameters as in [4], in the following form:

$$\begin{aligned}\dot{x}_1 =& \begin{bmatrix} -x_{11} + x_{12} \\ -0.5(x_{11}+x_{12}) - 0.5x_{12}\left(\cos\left(2x_{11}\right)+2\right)^2 \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ \cos\left(2x_{11}\right)+2 \end{bmatrix}\left(u_1 + (x_{11}+x_{22})\sin x_{12}^2 \cos\left(0.5x_{21}\right)\right), \\ \dot{x}_2 =& \begin{bmatrix} x_{22} \\ -x_{21} - 0.5x_{22} + 0.5x_{21}^2 x_{22} \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ x_{21} \end{bmatrix}\left(u_2 + 0.5\left(x_{12}+x_{22}\right)\cos\left(e^{x_{21}^2}\right)\right),\end{aligned} \tag{34}$$

where $x_1 = [x_{11}\ x_{12}]^T \in \mathbb{R}^2$ denote the state of the virtual subsystem and u_1 denote the virtual part control input. $x_2 = [x_{21}\ x_{22}]^T \in \mathbb{R}^2$ denote the real subsystem state and u_1 denote the real part control input.

To stabilize the decentralized virtual-real system, the objective of our proposed controller is to guarantee that the state $x_i(t)$ converges to zero while making sure that the state of the real subsystem does not move out of the safe boundary set ∂s. In this numerical simulation, the boundary set ∂s have a specified boundary function $h(x_i) = x_{i,2}^2 - x_{i,1} + 1$. The initial state is selected as $x_{i,0} = [1.2, 0.47]$. By choosing the barrier function as $b(x_i) = (1/h(x_i) - 1/h(0))^2$ with the gain $\alpha_i = 1$, a safety-guarding controller is obtained. For each subsystem, the learning rates are selected as $a_1 = 1$, $a_2 = 1$, and the weights are initialized as $\hat{\omega}_i\left(t_0\right) = [1,1,1]^T$.

To demonstrate the generality of the safety-guaranteed controller, we set a non-convex boundary restriction. The learning process is shown in Fig. 2 and Fig. 3, the approximated value functions of both virtual and real subsystems are convergent respectively.

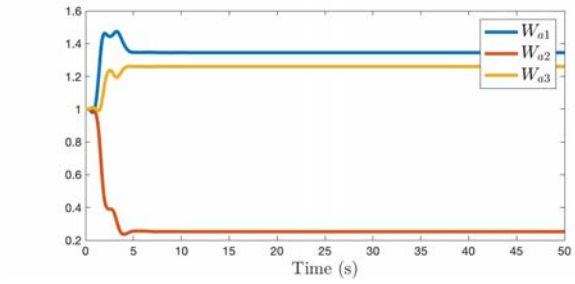

Fig. 2. Real subsystem critic NN weights.

The main result is presented in Fig. 4 and Fig. 5, where the black dotted line in Fig. 5 represents the security boundary. As illustrated in Fig. 5, the state trajectories of the two virtual subsystems have similar convergence trends. However, for the real subsystems, when the safety-guarding controller approaches the boundary, the safety-guarding term drags the state trajectory away from the direction of the reverse gradient

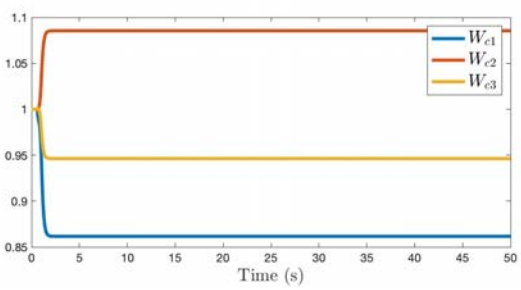

Fig. 3. Virtual subsystem critic NN weights.

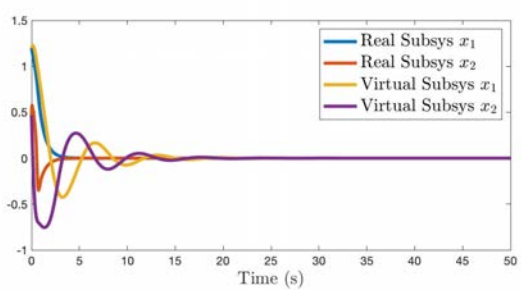

Fig. 4. States of each sbsystems.

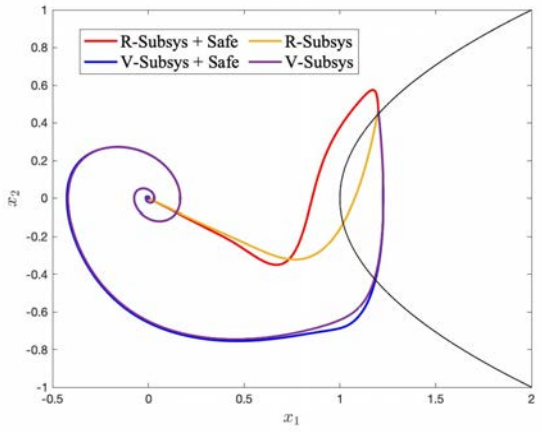

Fig. 5. State trajectories of each subsystem.

of the barrier function. While the distance from the boundary increases gradually, the influence of the safety-guarding term disappears. The states eventually converge to zero. In contrast, the state trajectory of the original real subsystem crosses the state limitation directly.

V. Conclusions

This paper designs a reinforcement learning-based safety-guaranteed stabilizing controller for the interconnected virtual-real system. We formulated the interconnected virtual-real system and safety-guaranteed stabilizing optimization problem. Online reinforcement learning solves the HJB equation for optimal stabilization control. A safe-guarding term ensures safe-guranteed control for the real part. Single network is used to approximate the value function.The estimation error dynamics of the critic network is verified to be uniform ultimately bounded. Lastly, the simulation example illustrates the effectiveness of the present stabilization control scheme.

References

[1] Y. Yang, K. G. Vamvoudakis, H. Modares, Y. Yin, and D. C. Wunsch, "Safe Intermittent Reinforcement Learning With Static and Dynamic Event Generators," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 31, no. 12, pp. 5441–5455, 2020.

[2] F. Tatari, H. Modares, C. Panayiotou, and M. Polycarpou, "Finite-Time Distributed Identification for Nonlinear Interconnected Systems," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 7, pp. 1188–1199, 2022.

[3] X. Yang and H. He, "Decentralized Event-Triggered Control for a Class of Nonlinear-Interconnected Systems Using Reinforcement Learning," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 635–648, 2021.

[4] D. Liu, D. Wang, and H. Li, "Decentralized Stabilization for a Class of Continuous-Time Nonlinear Interconnected Systems Using Online Learning Optimal Control Approach," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 25, no. 2, pp. 418–428, Feb. 2014.

[5] Q. Qu, H. Zhang, T. Feng, and H. Jiang, "Decentralized adaptive tracking control scheme for nonlinear large-scale interconnected systems via adaptive dynamic programming," *Neurocomputing*, vol. 225, pp. 1–10, 2017.

[6] V. Narayanan, A. Sahoo, S. Jagannathan, and K. George, "Approximate Optimal Distributed Control of Nonlinear Interconnected Systems Using Event-Triggered Nonzero-Sum Games," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 30, no. 5, pp. 1512–1522, 2019.

[7] Z. Du, N. Dong, and Y.-z. Xie, "Behavioral Modeling Method of Macro-models for Interconnected Systems with Frequency Characteristics and Nonlinear Termination Networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, pp. 1–1, 2023.

[8] B. Liang, S. Zheng, C. K. Ahn, and F. Liu, "Adaptive Fuzzy Control for Fractional-Order Interconnected Systems With Unknown Control Directions," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 1, pp. 75–87, 2022.

[9] G. Su and Z. Wang, "Joint Estimation of State and Unknown Input for Nonlinear Interconnected Systems Based on Multiple Intermediate Estimator," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, pp. 1–1, 2023.

[10] M. Chen, H. K. Lam, Q. Shi, and B. Xiao, "Reinforcement Learning-Based Control of Nonlinear Systems Using Lyapunov Stability Concept and Fuzzy Reward Scheme," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 10, pp. 2059–2063, 2020.

[11] V. Narayanan and S. Jagannathan, "Event-Triggered Distributed Control of Nonlinear Interconnected Systems Using Online Reinforcement Learning With Exploration," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2510–2519, 2018.

[12] Y. Zhang, L. Ma, C. Yang, and W. Dai, "Reinforcement Learning-Based Sliding Mode Tracking Control for the Two-Time-Scale Systems: Dealing With Actuator Attacks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 9, pp. 3819–3823, Sep. 2022.

[13] C. Liu, H. Zhang, G. Xiao, and S. Sun, "Integral reinforcement learning based decentralized optimal tracking control of unknown nonlinear large-scale interconnected systems with constrained-input," *Neurocomputing*, vol. 323, pp. 1–11, 2019.

[14] W. Gao, C. Deng, Y. Jiang, and Z.-P. Jiang, "Resilient reinforcement learning and robust output regulation under denial-of-service attacks," *Automatica*, vol. 142, p. 110366, 2022.

[15] B. Li, S. Wen, Z. Yan, G. Wen, and T. Huang, "A Survey on the Control Lyapunov Function and Control Barrier Function for Nonlinear-Affine Control Systems," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 3, pp. 584–602, 2023.

[16] M. H. Cohen and C. Belta, "Safe exploration in model-based reinforcement learning using control barrier functions," *Automatica*, vol. 147, p. 110684, Jan. 2023.

[17] Z. Marvi and B. Kiumarsi, "Barrier-Certified Learning-Enabled Safe Control Design for Systems Operating in Uncertain Environments," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 437–449, 2022.

[18] N.-M. T. Kokolakis and K. G. Vamvoudakis, "Safety-Aware Pursuit-Evasion Games in Unknown Environments Using Gaussian Processes and Finite-Time Convergent Reinforcement Learning," *IEEE Trans. Neural Netw. Learning Syst.*, pp. 1–14, 2022.