Decentralized Stabilization for a Class of Continuous-Time Nonlinear Interconnected Systems Using Online Learning Optimal Control Approach

Derong Liu, Fellow, IEEE, Ding Wang, and Hongliang Li

Abstract—In this paper, using a neural-network-based online learning optimal control approach, a novel decentralized control strategy is developed to stabilize a class of continuous-time nonlinear interconnected large-scale systems. First, optimal controllers of the isolated subsystems are designed with cost functions reflecting the bounds of interconnections. Then, it is proven that the decentralized control strategy of the overall system can be established by adding appropriate feedback gains to the optimal control policies of the isolated subsystems. Next, an online policy iteration algorithm is presented to solve the Hamilton–Jacobi–Bellman equations related to the optimal control problem. Through constructing a set of critic neural networks, the cost functions can be obtained approximately, followed by the control policies. Furthermore, the dynamics of the estimation errors of the critic networks are verified to be uniformly and ultimately bounded. Finally, a simulation example is provided to illustrate the effectiveness of the present decentralized control scheme.

Index Terms—Adaptive dynamic programming, decentralized control, large-scale systems, neural networks, nonlinear interconnected systems, optimal control, policy iteration, reinforcement learning.

I. INTRODUCTION

VARIOUS complex systems in social and engineering areas, such as ecosystems, transportation systems, and power systems, are considered as large-scale systems. Generally speaking, a large-scale system is comprised of several subsystems with obvious interconnections, which leads to the increasing difficulty of analysis and synthesis when using classical centralized control techniques. Bakule [1] pointed out with similar results that it is, therefore, necessary to partition the design issue of the overall system into manageable subproblems. Then, the overall plant is no longer controlled by a single controller but by an array of independent controllers that all together represent a decentralized controller. Therefore, the decentralized control has been a control of choice for large-scale systems because it is computationally efficient to formulate control law that use only locally available subsystem states or outputs [2]. Actually, considerable attention has been paid to the decentralized stabilization of large-scale systems during the last several decades [3]–[7].

As previously mentioned, a decentralized strategy consists of some noninteracting local controllers corresponding to the isolated subsystems, not the overall system. Thus, in many situations, the design of the isolated subsystems is a matter of great significance. In [8], it was shown that the decentralized control of the interconnected system was related to the optimal control of the isolated subsystems. Therefore, the optimal control method can be employed to facilitate the design process of the decentralized control strategy. However, in [8], the cost functions of the isolated subsystems were not chosen as the general forms, not to mention that the detailed procedure was not given. For this reason, in this paper, by employing the online policy iteration algorithm, we will investigate the decentralized stabilization problem using neural-network-based learning optimal control approach.

The optimal control of nonlinear system often leads to solving the Hamilton–Jacobi–Bellman (HJB) equation instead of the Riccati equation of the linear case. Though dynamic programming is a useful technique to solve the optimization and optimal control problems, in many cases, it is computationally difficult to apply it because of the curse of dimensionality. Fortunately, based on function approximators, such as neural networks, adaptive (or approximate) dynamic programming (ADP) was proposed by Werbos [9], [10] as an alternative method to solve the optimal control problems forward-in-time. There are several synonyms used for ADP, including adaptive dynamic programming [11]–[15], approximate dynamic programming [16]–[18], neuro-dynamic programming [19], and reinforcement learning [20]. In the recent years, great efforts have been made to ADP and related research in theory and applications. Numerous excellent results have been obtained that greatly promotes the development of relevant disciplines [23]–[40].

In light of [13], the ADP technique is closely related to reinforcement learning when engaging in the research of feedback control. In general, value and policy iterations are fundamental algorithms for the reinforcement learning-based ADP in optimal control. Policy iteration starts with a stabilizing control, whereas value iteration cannot always guarantee the stability of control during the implementation process. Al-Tamimi et al. [18], Zhang et al. [23], and Liu et al. [26]...
studied the optimal control problem of discrete-time nonlinear systems using value iteration algorithm. Specifically, policy iteration represents a class of algorithms containing two basic iterations, i.e., policy evaluation and policy improvement [41]-[46]. Abu-Khalaf and Lewis [41] derived an offline optimal control scheme for nonlinear systems with saturating actuators. Then, Vrabie and Lewis [42] and Yamvoudakis and Lewis [43] used online policy iteration algorithm to study the infinite horizon optimal control of continuous-time nonlinear systems, respectively. The former was performed based on the sequential updates of two neural networks, namely, critic network and action network, whereas in the latter, the two networks were trained simultaneously. Recently, Liu et al. [44] extended the policy iteration algorithm to nonlinear optimal control problem with unknown internal dynamics and discounted cost function. Besides, Bhasin et al. [46] constructed an actor-critic-identifier architecture to deal with the infinite horizon optimal control of uncertain nonlinear systems, characterized by the introduction of a robust dynamic neural network.

In this paper, we employ the online policy iteration algorithm to tackle the decentralized control of a class of nonlinear interconnected systems. To design the decentralized control scheme of the overall system, the optimal controllers of the isolated subsystems are designed at first with the cost functions modified to account for the interconnections. Then, the decentralized control strategy can be established by adding appropriate feedback gains to the local optimal control policies. Next, the online policy iteration algorithm is developed to solve the HJB equations related to the optimal control by constructing and training some critic networks. It is shown that the approximate closed-form expressions of the optimal control policies are available. Hence, there is no need to build action networks. Additionally, the uniform ultimate boundedness (UUB) of the dynamics of the weight estimation errors is analyzed using the Lyapunov approach. Remarkably, considering the effectiveness of ADP and reinforcement learning techniques in solving the nonlinear optimal control problem, the decentralized control approach established here is natural and convenient. More importantly, it can be employed to stabilize a broad class of nonlinear large-scale systems.

This paper is organized as follows. In Section II, the decentralized control problem of the large-scale system is described. In Section III, the optimal control of isolated subsystems is presented in the framework of HJB equations, based on which, the decentralized control strategy can be developed. In Section IV, the online policy iteration algorithm is introduced to solve the HJB equations with convergence analysis. In addition, critic networks are constructed for facilitating the implementation of online algorithm. The UUB of the dynamics of the weight estimation errors is proved as well. In Section V, an example is given to demonstrate the effectiveness of the established approach. In Section VI, concluding remarks are provided.

II. PROBLEM STATEMENT

In this paper, we study a class of continuous-time nonlinear large-scale systems composed of $N$ interconnected subsystems described by

\[
\dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t)) (\bar{u}_i(x_i(t)) + Z_i(x(t))) \quad i = 1, 2, \ldots, N
\]

where $x_i(t) \in \mathbb{R}^n$ and $\bar{u}_i(x_i(t)) \in \mathbb{R}^m_i$ are the state and control vectors of the $i$th subsystem, respectively. In large-scale system (1), $x \in [x_1^T, x_2^T, \ldots, x_N^T]^T \in \mathbb{R}^n$ denotes the overall state, where $n = \sum_{i=1}^{N} n_i$. Correspondingly, $x_1, x_2, \ldots, x_N$ are called local states, whereas $\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N)$ are local controls. Note that for subsystem $i$, $f_i(x_i), g_i(x_i)$, and $g_i(x_i)Z_i(x)$ represent the nonlinear internal dynamics, input gain matrix, and interconnected term, respectively.

Let $x_i(0) = x_{i0}$ be the initial state of the $i$th subsystem, $i = 1, 2, \ldots, N$. Additionally, we let the following assumptions hold throughout this paper.

Assumption 1: The state vector $x_i = 0$ is the equilibrium of the $i$th subsystem, $i = 1, 2, \ldots, N$.

Assumption 2: The functions $f_i(\cdot)$ and $g_i(\cdot)$ are differentiable in their arguments with $f_j(0) = 0$, where $i = 1, 2, \ldots, N$.

Assumption 3: The feedback control vector $\bar{u}_i(x_i) = 0$ when $x_i = 0$, where $i = 1, 2, \ldots, N$.

Let $R_i \in \mathbb{R}^{m_i \times m_i}, i = 1, 2, \ldots, N$, be symmetric positive definite matrices. Then, we denote $Z_i(x) = R_i^{1/2}Z_i(x)$, where $Z_i(x) \in \mathbb{R}^{m_i}, i = 1, 2, \ldots, N$, are bounded as follows:

\[
\|Z_i(x)\| \leq \sum_{j=1}^{N} \rho_{ij} h_{ij}(x_j), \quad i = 1, 2, \ldots, N.
\]

In (2), $\rho_{ij}$ are nonnegative constants and $h_{ij}(x_j)$ are positive semidefinite functions with $i, j = 1, 2, \ldots, N$.

If we define $h_i(x_i) = \max(h_{i1}(x_i), h_{i2}(x_i), \ldots, h_{IN}(x_i))$, $i = 1, 2, \ldots, N$, then (2) can be formulated as

\[
\|Z_i(x)\| \leq \sum_{j=1}^{N} \lambda_{ij} h_j(x_j), \quad i = 1, 2, \ldots, N
\]

where $\lambda_{ij} \geq \rho_{ij} h_{ij}(x_j)/h_j(x_j), i, j = 1, 2, \ldots, N$, are also nonnegative constants.

When dealing with the decentralized control problem, we aim at finding $N$ control policies $\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N)$ to stabilize the large-scale system (1). It is important to note that in the control pair $(\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N))$, $\bar{u}_i(x_i)$ is only a function of the corresponding local state, namely $x_i$, where $i = 1, 2, \ldots, N$. The schematic diagram of the decentralized control problem is shown in Fig. 1.
III. DECENTRALIZED CONTROLLER DESIGN VIA OPTIMAL CONTROL SCHEME

In this section, we investigate the methodology for decentralized controller design. Two sections are included in this part. In the first section, the optimal control of the isolated subsystems is described under the framework of HJB equations, whereas in the second section, the decentralized control strategy can be constructed based on the optimal control policies.

A. Optimal Control and the HJB Equations

Now, we consider the N isolated subsystems corresponding to (1) that are given by
\[ \dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))u_i(x_i(t)), \quad i = 1, 2, \ldots, N. \] (4)

For the ith isolated subsystem, we further assume that \( f_i + g_iu_i \) is Lipschitz continuous on a set \( \Omega_i \) in \( \mathbb{R}^{n_i} \) containing the origin, and the subsystem is controllable in the sense that there exists a continuous control policy on \( \Omega_i \) that asymptotically stabilizes the subsystem.

In this paper, to deal with the infinite horizon optimal control problem, we have to find the control policies \( u_i(x_i) \), \( i = 1, 2, \ldots, N \), which minimize the local cost functions
\[ J_i(x_i(0)) = \int_0^{\infty} \left( Q_i(x_i(\tau)) + u_i^T(\dot{x}_i(\tau))R_iu_i(x_i(\tau)) \right) d\tau, \quad i = 1, 2, \ldots, N \] (5)

where \( Q_i(x_i) \), \( i = 1, 2, \ldots, N \), are positive definite functions satisfying
\[ h_i(x_i) \leq Q_i(x_i), \quad i = 1, 2, \ldots, N. \] (6)

Based on optimal control theory, here, the designed feedback controls must not stabilize the subsystems on \( \Omega_i \), \( i = 1, 2, \ldots, N \), but also guarantee that the cost functions (5) are finite. In other words, the control policies must be admissible. Below is the definition of admissible control.

Definition 1: Consider the isolated subsystem \( i \), a control policy \( \mu_i \) is defined as admissible with respect to (5) on \( \Omega_i \), denoted by \( \mu_i \in \Psi_i(\Omega_i) \), if \( \mu_i(x_i) \) is continuous on \( \Omega_i \), \( \mu_i(0) = 0 \), and \( u_i(x_i) = \mu_i(x_i) \) stabilizes (4) on \( \Omega_i \), and \( J_i(x_i(0)) \) is finite for all \( x_i(0) \in \Omega_i \).

For any set of admissible control policies \( \mu_i \in \Psi_i(\Omega_i) \), \( i = 1, 2, \ldots, N \), if the associated cost functions
\[ V_i(x_i(0)) = \int_0^{\infty} \left( Q_i(x_i(\tau)) + \mu_i^T(\dot{x}_i(\tau))R_i\mu_i(x_i(\tau)) \right) d\tau, \quad i = 1, 2, \ldots, N \] (7)

are continuously differentiable, then the infinitesimal versions of (7) are the so-called nonlinear Lyapunov equations
\[ 0 = Q_i(x_i) + \mu_i^T(\dot{x}_i(x_i)) + (\nabla V_i(x_i))^T f_i(x_i) + g_i(x_i)\mu_i(x_i), \quad i = 1, 2, \ldots, N \] (8)

with \( V_i(0) = 0, i = 1, 2, \ldots, N \). In (8), the terms \( \nabla V_i(x_i) \), \( i = 1, 2, \ldots, N \), denote the partial derivatives of the local cost functions \( V_i(x_i) \) with respect to local states \( x_i \), i.e., \( \nabla V_i(x_i) = \partial V_i(x_i)/\partial x_i \), where \( i = 1, 2, \ldots, N \).

Define the Hamiltonian functions of the N isolated subsystems as follows:
\[ H_i(x_i, \mu_i, \nabla V_i(x_i)) = Q_i^T(x_i) + \mu_i^T(\dot{x}_i(x_i))R_i\mu_i(x_i) + (\nabla V_i(x_i))^T f_i(x_i) + g_i(x_i)\mu_i(x_i) \] (9)

where \( i = 1, 2, \ldots, N \).

The optimal cost functions of the N isolated subsystems can be formulated as
\[ J_i^*(x_i(0)) = \min_{\mu_i \in \Psi_i(\Omega_i)} \int_0^{\infty} \left( Q_i^T(x_i(\tau)) + \mu_i^T(\dot{x}_i(\tau))R_i \right. \
\times \mu_i(x_i(\tau)) \left. \right) d\tau, \quad i = 1, 2, \ldots, N. \] (10)

In view of optimal control theory, the optimal control policies \( J_i^*(x_i) \), \( i = 1, 2, \ldots, N \), satisfy the HJB equations
\[ 0 = \min_{\mu_i \in \Psi_i(\Omega_i)} H_i(x_i, \mu_i, \nabla J_i^*(x_i)) = 0, \quad i = 1, 2, \ldots, N \] (11)

where \( \nabla J_i^*(x_i) = \partial J_i^*(x_i)/\partial x_i \), \( i = 1, 2, \ldots, N \). Assume that the minima on the right hand side of (11) exist and are unique. Then, the optimal control policies for the N isolated subsystems are
\[ u_i^*(x_i) = \arg \min_{\mu_i \in \Psi_i(\Omega_i)} H_i(x_i, \mu_i, \nabla J_i^*(x_i)) \]
\[ = -1/2R_i^{-1}g_i^T(x_i)\nabla J_i^*(x_i) \] (12)

Substituting the optimal control policies (12) into the nonlinear Lyapunov equations (8), we can obtain the formulation of the HJB equations in terms of \( \nabla J_i^*(x_i) \), \( i = 1, 2, \ldots, N \), as follows:
\[ 0 = Q_i^T(x_i) + (\nabla J_i^*(x_i))^T f_i(x_i) + \frac{1}{4}(\nabla J_i^*(x_i))^T g_i(x_i)R_i^{-1}g_i^T(x_i)\nabla J_i^*(x_i) \] (13)

with \( J_i^*(0) = 0 \) and \( i = 1, 2, \ldots, N \).

Remark 1: The formulas developed in (12) display an array of closed-form expressions of the optimal control policies, which obviates the need to search for the optimal control policies via optimization process. However, the knowledge of \( J_i^*(x_i) \), \( i = 1, 2, \ldots, N \), is required, which implies the importance of the solutions of HJB equations.

B. Decentralized Control Strategy

According to (12), we have expressed the optimal control policies, i.e., \( u_1^*(x_1), u_2^*(x_2), \ldots, u_N^*(x_N) \), for the N isolated subsystems (4). In the following, we will show that by proportionally increasing some local feedback gains, a stabilizing decentralized control scheme can be established for the interconnected system (1). Now, we give the following lemma, indicating how the feedback gains can be added, to guarantee the asymptotic stability of the isolated subsystems.

Lemma 1: Consider the isolated subsystems (4), the feedback controls
\[ \tilde{u}_i(x_i) = \pi_iu_i^*(x_i) \]
\[ = -1/2\pi_iR_i^{-1}g_i^T(x_i)\nabla J_i^*(x_i), \quad i = 1, 2, \ldots, N \] (14)
ensure that the $N$ closed-loop isolated subsystems are asymptotically stable for all $\pi_i \geq 1/2$, where $i = 1, 2, \ldots, N$.

Proof: The lemma can be proved by showing $J_i^*(x_i)$, $i = 1, 2, \ldots, N$, are Lyapunov functions. First of all, in light of (10), we can find that $J_i^*(x_i) > 0$ for any $x_i \neq 0$ and $J_i^*(x_i) = 0$ when $x_i = 0$, which implies that $J_i^*(x_i)$, $i = 1, 2, \ldots, N$, are positive definite functions. Next, the derivatives of $J_i^*(x_i)$, $i = 1, 2, \ldots, N$, along the corresponding trajectories of the closed-loop isolated subsystems are given by

$$
\dot{J}_i^*(x_i) = (\nabla J_i^*(x_i))^T \dot{x}_i = (\nabla J_i^*(x_i))^T (f_i(x_i) + g_i(x_i)\dot{u}_i(x_i))
$$

where $i = 1, 2, \ldots, N$. Then, by adding and subtracting $1/(2\pi_i)\nabla J_i^*(x_i))^T g_i(x_i)\dot{u}_i(x_i)$ to (15) and considering (12)–(14), we have

$$
J_i^*(x_i) = (\nabla J_i^*(x_i))^T f_i(x_i)
- \frac{1}{4}(\nabla J_i^*(x_i))^T g_i(x_i)R_i^{-1}g_i^T(x_i)\nabla J_i^*(x_i)
- \frac{1}{2}(\pi_i - \frac{1}{2})(\nabla J_i^*(x_i))^T \times g_i(x_i)R_i^{-1}g_i^T(x_i)\nabla J_i^*(x_i)
= -Q_i^2(x_i) - \frac{1}{2}(\pi_i - \frac{1}{2})\|R_i^{-1/2}g_i^T(x_i)\nabla J_i^*(x_i)\|^2
$$

(16)

where $i = 1, 2, \ldots, N$. Observing (16), we can obtain that $J_i^*(x_i) < 0$ for all $\pi_i \geq 1/2$ and $x_i \neq 0$, where $i = 1, 2, \ldots, N$. Therefore, the conditions for Lyapunov local stability theory are satisfied and the proof is completed.

Remark 2: Lemma 1 reveals that any feedback controls $\dot{u}_i(x_i)$, $i = 1, 2, \ldots, N$, can ensure the asymptotic stability of the closed-loop isolated subsystems as long as $\pi_i \geq 1/2$, $i = 1, 2, \ldots, N$. However, only when $\pi_i = 1$, $i = 1, 2, \ldots, N$, the feedback controls are optimal. In fact, similar results have been given in [47]–[49], showing that the optimal controls $u_i^*(x_i)$, $i = 1, 2, \ldots, N$, are robust in the sense that they have infinite gain margins.

Now, we present the main theorem of this paper based on that the acquired decentralized control strategy can be established.

**Theorem 1:** For interconnected system (1), there exist $N$ positive numbers $\pi_i^* > 0$, $i = 1, 2, \ldots, N$, such that for any $\pi_i \geq \pi_i^*$, $i = 1, 2, \ldots, N$, the feedback controls developed by (14) ensure that the closed-loop interconnected system is asymptotically stable. In other words, the control pair $(\dot{u}_1(x_1), \dot{u}_2(x_2), \ldots, \dot{u}_N(x_N))$ is the decentralized control strategy of large-scale system (1).

**Proof:** In accordance with Lemma 1, we observe that $J_i^*(x_i)$, $i = 1, 2, \ldots, N$, are all Lyapunov functions. Here, we select a composite Lyapunov function given by

$$
\mathcal{L}(x) = \sum_{i=1}^{N} \theta_i J_i^*(x_i)
$$

(17)

where $\theta_i$, $i = 1, 2, \ldots, N$, are arbitrary positive constants.

Taking the time derivative of $\mathcal{L}(x)$ along the trajectories of the closed-loop interconnected system, we can obtain

$$
\dot{\mathcal{L}}(x) = \sum_{i=1}^{N} \theta_i \dot{J}_i^*(x_i)
= \sum_{i=1}^{N} \theta_i \{ (\nabla J_i^*(x_i))^T (f_i(x_i) + g_i(x_i)\dot{u}_i(x_i))
+ (\nabla J_i^*(x_i))^T g_i(x_i)\dot{Z}_i(x_i) \}.
$$

(18)

Then, taking (3), (6), and (16) into consideration, (18) can be turned into the following form:

$$
\dot{\mathcal{L}}(x) \leq -\sum_{i=1}^{N} \theta_i \left[ Q_i^2(x_i)
+ \frac{1}{2}(\pi_i - \frac{1}{2})\left\| R_i^{-1/2}g_i^T(x_i)\nabla J_i^*(x_i) \right\|^2
- \left\| (\nabla J_i^*(x_i))^T g_i(x_i)R_i^{-1/2} \right\| Z_i(x) \right]
\leq -\sum_{i=1}^{N} \theta_i \left[ Q_i^2(x_i)
+ \frac{1}{2}(\pi_i - \frac{1}{2})\left\| (\nabla J_i^*(x_i))^T g_i(x_i)R_i^{-1/2} \right\| Z_i(x) \right]
- \left\| (\nabla J_i^*(x_i))^T g_i(x_i)R_i^{-1/2} \right\| \sum_{j=1}^{N} \lambda_{ij} Q_j(x_j) \right]
$$

(19)

Here, we denote

$$
\Theta = \text{diag}[\theta_1, \theta_2, \ldots, \theta_N]
$$

$$
\Lambda = \begin{bmatrix}
\lambda_{11} & \lambda_{12} & \cdots & \lambda_{1N} \\
\lambda_{21} & \lambda_{22} & \cdots & \lambda_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{NN}
\end{bmatrix}
$$

(20)

(21)

and

$$
\Pi = \text{diag}\left[ \frac{1}{2}(\pi_1 - \frac{1}{2}), \frac{1}{2}(\pi_2 - \frac{1}{2}), \ldots, \frac{1}{2}(\pi_N - \frac{1}{2}) \right].
$$

(22)

Therefore, through introducing a $2N$-dimensional vector

$$
\xi = \begin{bmatrix}
Q_1(x_1) \\
Q_2(x_2) \\
\vdots \\
Q_N(x_N) \\
\left\| (\nabla J_1^*(x_1))^T g_1(x_1)R_1^{-1/2} \right\| \\
\left\| (\nabla J_2^*(x_2))^T g_2(x_2)R_2^{-1/2} \right\| \\
\vdots \\
\left\| (\nabla J_N^*(x_N))^T g_N(x_N)R_N^{-1/2} \right\|
\end{bmatrix}
$$

(23)
we can transform (19) to the following compact form:
\[
\dot{L}(x) \leq -\varepsilon^T \begin{bmatrix}
\Theta & -\frac{1}{2}\Lambda^T \Theta \\
-\frac{1}{2} \Theta \Lambda & \Theta \Pi
\end{bmatrix} \xi
\]
\[
\triangleq -\varepsilon^T \mathcal{A} \xi.
\] (24)

According to (24), sufficiently large \( \pi_i \), \( i = 1, 2, \ldots, N \), can be chosen such that the matrix \( \mathcal{A} \) is positive definite. That is to say, there exist \( \pi_i^* \), \( i = 1, 2, \ldots, N \), so that any \( \pi_i \geq \pi_i^* \), \( i = 1, 2, \ldots, N \), are large enough to guarantee the positive definiteness of \( \mathcal{A} \). Then, we have \( \dot{L}(x) < 0 \). Therefore, the closed-loop interconnected system is asymptotically stable under the action of the control pair \((\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N))\). The proof is completed. \(\blacksquare\)

Clearly, the focal point of designing the decentralized control strategy becomes to derive the optimal controllers for the \( N \) isolated subsystems on the basis of Theorem 1. Then, we should put our emphasis on solving the HJB equations, which yet, is regarded as a difficult task \[12\], \[13\]. Hence, in the following, we will employ a more pragmatic approach to obtain the approximate solutions based on online policy iteration algorithm and neural network techniques.

IV. NEURAL-NETWORK-BASED LEARNING OPTIMAL CONTROL OF THE ISOLATED SUBSYSTEMS USING ONLINE POLICY ITERATION ALGORITHM

Three sections are included here. In the first section, the online policy algorithm is introduced to tackle the optimal control problem of the isolated subsystems, whereas the neural network implementation process is given in the second section. The stability proof of the dynamics of the estimation errors is developed in the last section.

A. Online Policy Iteration Algorithm and Its Convergence

Here, the online policy iteration algorithm is introduced to solve the HJB equations. The policy iteration algorithm consists of policy evaluation based on (8) and policy improvement based on (12), as shown in \[22\]. Specifically, its iteration procedure can be described as follows.

Step 1: Choose a small positive number \( \epsilon \). Let \( p = 0 \) and \( V_i^{(0)}(x_i) = 0 \), where \( i = 1, 2, \ldots, N \). Then, start with \( N \) initial admissible control policies \( \mu_1^{(0)}(x_1) \), \( \mu_2^{(0)}(x_2) \), \ldots, \( \mu_N^{(0)}(x_N) \).

Step 2: Based on the control policies \( \mu_i^{(p)}(x_i), i = 1, 2, \ldots, N \), solve the following nonlinear Lyapunov equations
\[
0 = Q_i^2(x_i) + \left( \mu_i^{(p)}(x_i) \right)^T R_i \mu_i^{(p)}(x_i) + \left( \nabla V_i^{(p+1)}(x_i) \right) \left( f_i(x_i) + g_i(x_i) \mu_i^{(p)}(x_i) \right)
\] (25)
with \( V_i^{(p+1)}(0) = 0 \) and \( i = 1, 2, \ldots, N \).

Step 3: Update the control policies via
\[
\mu_i^{(p+1)}(x_i) = -\frac{1}{2} R_i^{-1} s_i^T (x_i) \nabla V_i^{(p+1)}(x_i)
\] (26)
where \( i = 1, 2, \ldots, N \).

Step 4: If \( ||V_i^{(p+1)}(x_i) - V_i^{(p)}(x_i)|| \leq \epsilon, i = 1, 2, \ldots, N \), stop and obtain the approximate optimal controls of the \( N \) isolated subsystems; else, let \( p = p + 1 \) and go back to Step 2.

Note that \( N \) initial admissible control policies are required in the above algorithm. In the following, we present the convergence analysis of the online policy iteration algorithm for the isolated subsystems.

Theorem 2: Consider the \( N \) isolated subsystems (4), given \( N \) initial admissible control policies \( \mu_1^{(0)}(x_1), \mu_2^{(0)}(x_2), \ldots, \mu_N^{(0)}(x_N) \). Then, using the policy iteration algorithm established in (25) and (26), the cost functions and control policies converge to the optimal ones as \( p \to \infty \), i.e., \( V_i^{(p)}(x_i) \to J_i^*(x_i) \) and \( \mu_i^{(p)}(x_i) \to u_i^*(x_i) \) as \( p \to \infty \), where \( i = 1, 2, \ldots, N \).

Proof: First, we consider the subsystem \( i \). According to \[41\] and \[44\], when given an initial admissible control policy \( \mu_i^{(0)}(x_i) \), we have \( \mu_i^{(p)}(x_i) \in \Psi_i(\Omega_i) \) for any \( p \geq 0 \). Additionally, for any \( \zeta > 0 \), there exists an integer \( p_0 \), such that for any \( p \geq p_0 \), the formulas
\[
\sup_{x_i \in \Omega_i} |V_i^{(p)}(x_i) - J_i^*(x_i)| < \zeta
\] (27)
and
\[
\sup_{x_i \in \Omega_i} |\mu_i^{(p)}(x_i) - u_i^*(x_i)| < \zeta
\] (28)
hold simultaneously.

Next, we consider the \( N \) isolated subsystems. When given \( \mu_1^{(0)}(x_1), \mu_2^{(0)}(x_2), \ldots, \mu_N^{(0)}(x_N) \), where \( \mu_i^{(0)}(x_i) \) is the initial admissible control policy corresponding to the \( i \)th subsystem, we can acquire that \( \mu_i^{(p)}(x_i) \in \Psi_i(\Omega_i) \) for any \( p \geq 0 \), where \( i = 1, 2, \ldots, N \). In addition, we denote \( p_0 = \max\{p_0, p_0, \ldots, p_0\} \). Thus, we can conclude that for any \( \zeta > 0 \), there exists an integer \( p_0 \), such that for any \( p \geq p_0 \), (27) and (28) are true with \( i = 1, 2, \ldots, N \).

In other words, the algorithm will converge to the optimal cost functions and optimal control policies of the \( N \) isolated subsystems. The proof is completed. \(\blacksquare\)

B. Implementation Procedure via Neural Networks

For the \( N \) isolated subsystems, assume that the cost functions \( V_i(x_i), i = 1, 2, \ldots, N \), are continuously differentiable. Then, according to the universal approximation property of neural networks, \( V_i(x_i) \) can be reconstructed by a single-layer neural network on a compact set \( \Omega_i \) as
\[
V_i(x_i) = \omega_i^T \sigma_{ci}(x_i) + e_{ci}(x_i), i = 1, 2, \ldots, N
\] (29)
where \( \omega_{ci} \in \mathbb{R}^l \) is the ideal weight, \( \sigma_{ci}(x_i) \in \mathbb{R}^l \) is the activation function, \( l_i \) is the number of neurons in the hidden layer, and \( e_{ci}(x_i) \) is the approximation error of the \( i \)th neural network, \( i = 1, 2, \ldots, N \).

The derivatives of the cost functions with respect to their state vectors are formulated as
\[
\nabla V_i(x_i) = (\nabla \sigma_{ci}(x_i))^T \omega_{ci} + \nabla e_{ci}(x_i),
\] (30)
\( i = 1, 2, \ldots, N \).
where \( \nabla \sigma_{ci}(x_i) = \partial \sigma_{ci}(x_i)/\partial x_i \in \mathbb{R}^{l_i \times n_i} \) and \( \nabla e_{ci}(x_i) = \partial e_{ci}(x_i)/\partial x_i \in \mathbb{R}^{n_i} \) are the gradient of the activation function and approximate error of the \( i \)th neural network, respectively, \( i = 1, 2, \ldots, N \). Based on (30), the Lyapunov equations (8) becomes

\[
0 = Q_i^T(x_i) + \mu_i^T R_i \mu_i + (\omega_{ci}^T \nabla \sigma_{ci}(x_i) + (\nabla e_{ci}(x_i))^T) \dot{x}_i \tag{31}
\]

where \( i = 1, 2, \ldots, N \).

For the \( i \)th neural network, \( i = 1, 2, \ldots, N \), assume that the neural network weight vector \( \omega_{ci} \), the gradient \( \nabla \sigma_{ci}(x_i) \), and the approximation error \( e_{ci}(x_i) \) and its derivative \( \nabla e_{ci}(x_i) \) are all bounded on the compact set \( \Omega_i \). In addition, according to [43], we have \( e_{ci}(x_i) \to 0 \) and \( \nabla e_{ci}(x_i) \to 0 \) as \( l_i \to \infty \), where \( i = 1, 2, \ldots, N \).

Because the ideal weights are unknown, \( N \) critic neural networks can be built to approximate the cost functions as follows:

\[
\hat{V}_i(x_i) = \hat{\omega}_{ci}^T \sigma_{ci}(x_i), \quad i = 1, 2, \ldots, N \tag{32}
\]

where \( \hat{\omega}_{ci}, i = 1, 2, \ldots, N \), is the estimated weights. Here, \( \sigma_{ci}(x_i), i = 1, 2, \ldots, N \), is selected such that \( \hat{V}_i(x_i) > 0 \) for any \( x_i \neq 0 \) and \( \hat{V}_i(x_i) = 0 \) when \( x_i = 0 \).

Similarly, the derivatives of the approximate cost functions with respect to the state vectors can be expressed by

\[
\nabla \hat{V}_i(x_i) = (\nabla \sigma_{ci}(x_i))^T \hat{\omega}_{ci}, \quad i = 1, 2, \ldots, N \tag{33}
\]

where \( \nabla \hat{V}_i(x_i) = \partial \hat{V}_i(x_i)/\partial x_i \), \( i = 1, 2, \ldots, N \). Then, the approximate Hamiltonian functions can be expressed as

\[
H_i(x_i, \mu_i, \hat{\omega}_{ci}) = Q_i^T(x_i) + \mu_i^T R_i \mu_i + \hat{\omega}_{ci}^T \nabla \sigma_{ci}(x_i) \dot{x}_i = e_{ci}, \quad i = 1, 2, \ldots, N. \tag{34}
\]

For the purpose of training the critic networks of the isolated subsystems, it is desired to design \( \hat{\omega}_{ci}, i = 1, 2, \ldots, N \), to minimize the following objective functions:

\[
E_{ci} = \frac{1}{2} e_{ci}^T e_{ci}, \quad i = 1, 2, \ldots, N. \tag{35}
\]

The standard steepest descent algorithm is introduced to tune the critic networks, then their weights are updated through

\[
\dot{\hat{\omega}}_{ci} = -a_{ci} \left[ \frac{\partial E_{ci}}{\partial \hat{\omega}_{ci}} \right], \quad i = 1, 2, \ldots, N \tag{36}
\]

where \( a_{ci} > 0, i = 1, 2, \ldots, N \), is the learning rates of the critic networks.

On the other hand, based on (30), the Hamiltonian functions take the following forms:

\[
H_i(x_i, \mu_i, \hat{\omega}_{ci}) = Q_i^T(x_i) + \mu_i^T R_i \mu_i + \hat{\omega}_{ci}^T \nabla \sigma_{ci}(x_i) \dot{x}_i = e_{ci}, \quad i = 1, 2, \ldots, N \tag{37}
\]

where \( e_{ci} = -(\nabla e_{ci}(x_i))^T \dot{x}_i \), \( i = 1, 2, \ldots, N \), is the residual errors because of the neural network approximation.

Denote \( \delta_i = \nabla \sigma_{ci}(x_i) \dot{x}_i \), \( i = 1, 2, \ldots, N \). We assume that there exist \( N \) positive constants \( \delta_{Mi}, i = 1, 2, \ldots, N \), such that

\[
\| \delta_i \| \leq \delta_{Mi}, \quad i = 1, 2, \ldots, N. \tag{38}
\]

In addition, we define the weight estimation errors of the critic networks as \( \hat{\omega}_{ci} = \omega_{ci} - \hat{\omega}_{ci} \), where \( i = 1, 2, \ldots, N \). Then, combining (34) with (37) yields

\[
e_{ci} H_i - e_{ci} = \hat{\omega}_{ci}^T \delta_i, \quad i = 1, 2, \ldots, N. \tag{39}
\]

Therefore, the dynamics of the weight estimation errors can be given as follows:

\[
\dot{\hat{\omega}}_{ci} = a_{ci} (e_{ci} H_i - \hat{\omega}_{ci}^T \delta_i) \delta_i, \quad i = 1, 2, \ldots, N. \tag{40}
\]

Incidentally, the persistency of excitation condition is required to tune the \( i \)th critic network to guarantee that \( \| \delta_i \| \geq \delta_{mi}, \) where \( \delta_{mi}, i = 1, 2, \ldots, N \), are positive constants. Thus, a set of probing noises will be added to the isolated subsystems to satisfy the condition in practice.

When implementing the online policy iteration algorithm, to accomplish the policy improvement, we should obtain the control polices that minimize the current cost functions. Hence, according to (12) and (30), we have

\[
\mu_i(x_i) = -\frac{1}{2} R_i^{-1} g_i^T(x_i) \nabla V_i(x_i) = -\frac{1}{2} R_i^{-1} g_i^T(x_i) ((\nabla \sigma_{ci}(x_i))^T \omega_{ci} + \nabla e_{ci}(x_i)) \tag{41}
\]

where \( i = 1, 2, \ldots, N \). Correspondingly, the approximate control policies can be obtained by

\[
\hat{\mu}_i(x_i) = -\frac{1}{2} R_i^{-1} g_i^T(x_i) \nabla \hat{V}_i(x_i) = -\frac{1}{2} R_i^{-1} g_i^T(x_i) (\nabla \sigma_{ci}(x_i))^T \hat{\omega}_{ci} \tag{42}
\]

where \( i = 1, 2, \ldots, N \).

Remark 3: According to (42), it is obvious to observe that the approximate control polices of the \( N \) isolated subsystems can be derived directly based on the trained critic networks. Therefore, unlike the traditional actor-critic architecture, the action neural networks are not required any more.

C. Stability Analysis

When considering the critic networks, the weight estimation dynamics are UUB as described in the following theorem.

*Theorem 3:* For the \( N \) isolated subsystems (4), the weight update laws for tuning the critic networks are given by (36). Then, the dynamics of the weight estimation errors of the critic networks are UUB.

*Proof:* Choose \( N \) Lyapunov function candidates described as follows:

\[
\mathcal{L}_i(t) = \frac{1}{a_{ci}} tr(\hat{\omega}_{ci}^T \hat{\omega}_{ci}), \quad i = 1, 2, \ldots, N. \tag{43}
\]

The time derivatives of the Lyapunov functions \( \dot{\mathcal{L}}_i(t), i = 1, 2, \ldots, N \), along the trajectories of the error dynamics (40) are

\[
\dot{\mathcal{L}}_i(t) = \frac{2}{a_{ci}} tr(\hat{\omega}_{ci}^T \dot{\hat{\omega}}_{ci}) = \frac{2}{a_{ci}} tr(\hat{\omega}_{ci}^T \omega_{ci} (e_{ci} H_i - \hat{\omega}_{ci}^T \delta_i) \delta_i) \tag{44}
\]
Fig. 2. Convergence of the weight vector of the critic network 1 ($\omega_{ac11}$, $\omega_{ac12}$, and $\omega_{ac13}$ represent $\hat{\omega}_{c11}$, $\hat{\omega}_{c12}$, and $\hat{\omega}_{c13}$, respectively).

Fig. 3. Convergence of the weight vector of the critic network 2 ($\omega_{ac21}$, $\omega_{ac22}$, and $\omega_{ac23}$ represent $\hat{\omega}_{c21}$, $\hat{\omega}_{c22}$, and $\hat{\omega}_{c23}$, respectively).

Fig. 4. 3-D plot of the approximation error of the cost function of isolated subsystem 1, i.e., $J_1^*(x_1) - \hat{V}_1(x_1)$.

Fig. 5. 3-D plot of the approximation error of the control policy of isolated subsystem 1, i.e., $u_1^*(x_1) - \hat{\mu}_1(x_1)$.

where $i = 1, 2, \ldots, N$. After some basic manipulations, it yields

$$\dot{L}_i(t) \leq -(2 - a_{ci}) \| \tilde{\omega}_{ci} \|_2^2 + \frac{1}{a_{ci}} e_i^2 c_{Hi}$$

where $i = 1, 2, \ldots, N$. In view of the Cauchy–Schwarz inequality and (38), we can conclude that $\dot{L}_i(t) < 0$ as long as

$$0 < a_{ci} < 2 \quad \| \tilde{\omega}_{ci} \| > \frac{e_i^2 c_{Hi}}{a_{ci}(2 - a_{ci}) \delta_{Mi}}$$

where $i = 1, 2, \ldots, N$. In accordance with the Lyapunov stability theory, we obtain that the dynamics of the weight estimation errors of the critic networks are all UUB. Meanwhile, the norms of the weight estimation errors are bounded as well. The proof is completed.

Remark 4: Let $\hat{u}_i(x_i) = \pi_i \hat{\mu}_i(x_i)$, where $\hat{\mu}_i(x_i)$, $i = 1, 2, \ldots, N$, are obtained by (42). According to the selections of the activation functions of the critic networks, we can easily find that the approximate optimal cost functions $\hat{V}_i(x_i)$, $i = 1, 2, \ldots, N$, are also Lyapunov functions. Furthermore, similar to the proof of Theorem 1, we have $\dot{L}(x) \leq -c^T \alpha(x) + \Sigma_{e}$, where $\Sigma_{e}$ is the sum of the approximation errors. Hence, we can conclude that based on the approximate optimal control policies $\hat{\mu}_i(x_i)$, $i = 1, 2, \ldots, N$, the developed control pair $(\hat{u}_1(x_1), \hat{u}_2(x_2), \ldots, \hat{u}_N(x_N))$ can ensure the UUB of the state trajectories of the closed-loop interconnected system. It is in this sense that we accomplish the design of the decentralized control scheme by adopting the learning optimal control approach based on online policy iteration algorithm.

Remark 5: Note that the controller presented here is a decentralized stabilization one. Though the optimal decentralized controller of interconnected systems has been studied before [50], in this paper, we aim at developing a novel decentralized control strategy based on ADP. How to extend the present results to the design of optimal decentralized control for nonlinear interconnected systems is part of our future research.
V. SIMULATION STUDY

A simulation example is provided in this section to show the applicability of the decentralized control strategy established in this paper.

Consider the following continuous-time nonlinear large-scale system consisting of two interconnected subsystems:

\[ \dot{x}_1 = \begin{bmatrix} -x_{11} + x_{12} \\ -0.5x_{11} - 0.5x_{12} - 0.5x_{12} \cos(2x_{11}) + 2 \end{bmatrix} + \begin{bmatrix} 0 \\ \cos(2x_{11}) + 2 \end{bmatrix} \left( \bar{u}_1(x_1) + (x_{11} + x_{12}) \sin x_{12} \cos(0.5x_{21}) \right) \]

\[ \dot{x}_2 = \begin{bmatrix} x_{22} \\ -x_{21} - 0.5x_{22} + 0.5x_{21}^2x_{22} \end{bmatrix} + \begin{bmatrix} 0 \\ x_{21} \end{bmatrix} \left( \bar{u}_2(x_2) + 0.5(x_{12} + x_{22}) \cos (e^{x_{22}}) \right) \]  

(47)

where \( x_1 = [x_{11}, x_{12}]^T \in \mathbb{R}^2 \) and \( \bar{u}_1(x_1) \in \mathbb{R} \) are the state and control variables of subsystem 1, and \( x_2 = [x_{21}, x_{22}]^T \in \mathbb{R}^2 \) and \( \bar{u}_2(x_2) \in \mathbb{R} \) are the state and control variables of subsystem 2. Let \( R_1 = R_2 = I \), where \( I \) denotes the identity matrix with suitable dimension. Additionally, let \( h_1(x_1) = \|x_1\| \) and \( h_2(x_2) = \|x_{22}\| \). Then, we find that \( Z_1(x) \) and \( Z_2(x) \) with \( x = [x_{11}^T \ x_{12}^T]^T \) are upper bounded as in (3). For example, we can select \( \lambda_{11} = \lambda_{12} = 1 \) and \( \lambda_{21} = \lambda_{22} = 1/2 \).

To design the decentralized controller of interconnected system (47), we first deal with the optimal control problem of two isolated subsystems. Here, we choose \( Q_1(x_1) = \|x_1\| \) and \( Q_2(x_2) = \|x_{22}\| \). Hence, the cost functions of the optimal control problem are

\[ J_1(x_{10}) = \int_0^\infty \{x_{11}^2 + x_{12}^2 + u_1^Tu_1\}dT \]  

(48)

and

\[ J_2(x_{20}) = \int_0^\infty \{x_{21}^2 + u_2^Tu_2\}dT. \]  

(49)

We adopt the online policy iteration algorithm to tackle the optimal control problem, where two critic networks are constructed to approximate the cost functions. We denote the weight vectors of the two critic networks as

\[ c_1 = [\hat{\omega}_{c11} \hat{\omega}_{c12} \hat{\omega}_{c13}]^T \] and \( \hat{\omega}_{c2} = [\hat{\omega}_{c21} \hat{\omega}_{c22} \hat{\omega}_{c23}]^T \). During the simulation process, the initial weights of the critic networks are chosen randomly in \([0, 2]\). In addition, the activation functions of the two critic networks are chosen as \( \sigma_{c1}(x_1) = [x_{11}^2 \ x_{11}x_{12} \ x_{12}^2]^T \) and \( \sigma_{c2}(x_2) = [x_{21}^2 \ x_{21}x_{22} \ x_{22}^2]^T \). Besides, let the learning rates of the critic networks be \( \alpha_{c1} = \alpha_{c2} = 0.1 \) and the initial states of the two isolated subsystems be \( x_{10} = x_{20} = [1 - 1]^T \).

During the implementation process of the online policy iteration algorithm, for each isolated subsystem, we add a probing noise to satisfy the persistency of excitation condition. We can observe that the convergence results of the weights have occurred after 750 and 180 s, respectively. Then, the probing signals are turned off. Actually, the weights of the critic networks converge to

\[ \hat{\omega}_{c1} = [0.498969 0.000381 0.999843]^T \]  

(50)

and

\[ \hat{\omega}_{c2} = [1.000002 - 0.000021 0.999992]^T \]  

(51)

that are shown in Figs. 2 and 3.
The cost functions can be approximated by constructing several critic networks and the expressions of the control policies can be obtained directly. In addition, the dynamics of the estimation errors are proved to be UUB. Simulation study is presented to demonstrate the validity of the decentralized control strategy in the end.

VI. CONCLUSION

In this paper, a novel decentralized control strategy is developed to deal with the stabilization problem of a class of continuous-time nonlinear large-scale systems using online policy iteration algorithm. Initially, the optimal controllers of the isolated subsystems are designed. Then, it is shown that the decentralized control strategy of the overall system can be established by adding feedback gains to the obtained optimal control policies. In addition, the online policy iteration algorithm is introduced to solve the HJB equations iteratively.

REFERENCES


Ding Wang received the B.S. degree in mathematics from the Zhengzhou University of Light Industry, Zhengzhou, China, the M.S. degree in operations research and cybernetics from Northeastern University, Shenyang, China, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007, 2009, and 2012, respectively.

He is currently an Assistant Professor with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His current research interests include adaptive dynamic programming, neural networks and learning systems, and complex systems and intelligent control.

Hongliang Li received the B.S. degree in mechanical engineering and automation from the Beijing University of Posts and Telecommunications, Beijing, China, in 2010. He is currently pursuing the Ph.D. degree with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing.

He is with the University of Chinese Academy of Sciences, Beijing. His current research interests include machine learning, neural networks, reinforcement learning, adaptive dynamic programming, and game theory.