Parallel Control for Optimal Tracking via Adaptive Dynamic Programming

Jingwei Lu, Qinglai Wei, Senior Member, IEEE, and Fei-Yue Wang, Fellow, IEEE

Abstract—This paper studies the problem of optimal parallel tracking control for continuous-time general nonlinear systems. Unlike existing optimal state feedback control, the control input of the optimal parallel control is introduced into the feedback system. However, due to the introduction of control input into the feedback system, the optimal state feedback control methods can not be applied directly. To address this problem, an augmented system and an augmented performance index function are proposed firstly. Thus, the general nonlinear system is transformed into an affine nonlinear system. The difference between the optimal parallel control and the optimal state feedback control is analyzed theoretically. It is proven that the optimal parallel control with the augmented performance index function can be seen as the suboptimal state feedback control with the traditional performance index function. Moreover, an adaptive dynamic programming (ADP) technique is utilized to implement the optimal parallel tracking control using a critic neural network (NN) to approximate the value function online. The stability analysis of the closed-loop system is performed using the Lyapunov theory, and the tracking error and NN weights errors are uniformly ultimately bounded (UUB). Also, the optimal parallel controller guarantees the continuity of the control input under the circumstance that there are finite jump discontinuities in the reference signals. Finally, the effectiveness of the developed optimal parallel control method is verified in two cases.

Index Terms—Adaptive dynamic programming (ADP), nonlinear optimal control, parallel controller, parallel control theory, parallel system, tracking control, neural network (NN).

I. INTRODUCTION

Remarkable strides have been made in the past decades in the area of control theory and technology due to the development of science and computational capacity of computers, and the intelligent control is one of the most rapidly developing technologies recently [1]—[9]. In the area of intelligent control, adaptive dynamic programming (ADP), proposed by Werbos [10], [11], is an effective technique to solve optimal control problems of nonlinear systems. Such optimal control problems are often required to solve a nonlinear partial differential equation called the Hamilton-Jacobi-Bellman (HJB) equation [12]—[15], and the analytic solutions of the HJB equation can not be obtained directly in most cases. Thus, the ADP technique emerges to obtain the solution of the HJB equation forward-in-time and has attracted much attention from researchers [16]—[26]. In many cases, value or policy iteration is used in the ADP technique to solve the HJB equation. The difference between value or policy iteration lies in an initial admissible control, policy iteration method requires an initial admissible controller while it is not necessary for value iteration method [22], [27], [28]. Recently, several online methods based on the Lyapunov theory are proposed to solve the HJB equation without sequential updates of neural networks (NNs) of the critic and actor [29]—[32].

In engineering applications, the tracking control problem is more common than the regulation problem [31], [33]—[35]. Many intelligent methods including the ADP have been applied by many researchers to solve the optimal tracking control problem [36]—[39]. In [40], a greedy heuristic dynamic programming (HDP) method with a new performance index is proposed for the discrete affine nonlinear systems. In [41], a data-driven robust approximate optimal tracking control scheme is proposed for the unknown continuous-time nonlinear systems based on the ADP technique for the first time. In [31], an ADP-based optimal tracking control method without using value and policy iterations is for the helicopter unmanned aerial vehicles. In [42], a data-driven adaptive tracking control approach is proposed for a class of continuous-time nonlinear systems using the goal representation HDP (GrHDP) with the filter-based action network. Notice that most previous researches about optimal tracking control focus on affine systems and the reference signals are usually assumed to be differentiable. In [43], an ADP-based optimal control method with experience replay is proposed for underactuated snake robots. It is worth pointing out that one of the difficulties in implementing ADP online for nonaffine nonlinear systems directly lies in the desired control input that can not be obtained directly using the first-order necessary condition. Besides, a problem in tracking control with a non-differentiable or discontinuous reference signal is that the error dynamics are hard to establish. Also, discontinuous reference
signals (i.e., step signal) could cause the sudden change of the control input, which is hard to implement on the mechanical actuators. According to the problems mentioned above, it is urged to design a new type of controller.

The parallel control, proposed by Wang [1], [2], [44]−[46], is a powerful method to solve many control problems based on the parallel system theory. The structure of the parallel system is shown in Fig. 1. The main idea of parallel control is to expand practical problems into virtual space, and then to solve the control problems by virtual-reality interaction [1]. Parallel control can also be referred to as the ACP methodology [1], [44], which is a trilogy: artificial systems (A), computational experiments (C), and parallel execution (P).

![Fig. 1. Structure of parallel system.](image1)

Fig. 2 shows the parallel execution that is established based on the parallel system theory. Many processes, such as the experiment and evaluation, learning and training, management and control can be executed based on the interaction between the artificial systems and physical systems. Researchers have made considerable fruits on the parallel control and the parallel systems, such as intelligent transportation systems, intelligent vehicle systems, and computer vision, see e.g., [2], [47]−[51] and references therein. However, the study of optimal parallel control for nonlinear systems is insufficient, which motivates our research.

This article studies the optimal tracking control for general nonlinear systems under the parallel control theory framework of the literatures [1], [2], [44]−[46]. The parallel control theory is introduced into the nonlinear optimal control. The main contributions of this article are summarized as follows.

1) An augmented system and an augmented performance index function are proposed, and the general nonlinear system is transformed into an affine system. It is proven that the optimal parallel control with the augmented performance index function can be seen as the suboptimal state feedback control with the traditional performance index function.

2) Based on the augmented system and the augmented performance index function, the ADP technique is employed to implement the optimal parallel control by using a critic NN to estimate the solution of the HJB equation. It is proven that the tracking error states and the NN weight errors are uniformly ultimately bounded (UUB). The continuity of the control signal can be guaranteed by the parallel controller with certain discontinuous reference signals, which is a good property in industrial applications.

The remainder of this paper is organized as follows. In Section II, a brief description of the parallel control is given and the problem of optimal parallel tracking control is formulated. In Section III, an augmented system is proposed with an augmented performance index function, and the difference between the optimal parallel control and the optimal state feedback control is analyzed theoretically. In Section IV, an ADP-based optimal parallel control is developed with the theoretical analysis. Simulation results are provided and discussed in Section V. Section VI gives some conclusions of this study.

II. PROBLEM FORMULATION

In this section, a brief description of the parallel control is given and the comparisons between the parallel control and state feedback control are shown. Then, the problem of the optimal parallel tracking control is presented.

A. Parallel Control

In this subsection, the parallel control is introduced briefly. More detailed descriptions for parallel control can be found in [1], [2], [44]−[46], [52]. Consider the following system

\[ \dot{x} = h(x, u) \]  

with states \( x \in \mathbb{R}^n \), control input \( u \in \mathbb{R}^m \), and \( h(x, u) \in \mathbb{R}^n \). The parallel control, which is shown in Fig. 3, can be established. The parallel control is given by

\[ \dot{u} = g(x, u) \]  

![Fig. 2. Structure of parallel execution.](image2)

Notice that the control input is introduced into the feedback system. Since control input \( u \) is no longer generated passively by the state \( x \), \( \dot{u} = g(x, u) \) can be seen as an artificial system. The dynamic system (1) and the dynamic parallel controller (2) are executed in parallel with information interaction. It
is obvious that there exist the differences between traditional feedback control and parallel control. The state feedback control input is generated based on the state.

According to (1) and (2), we have
\[ \dot{z} = \begin{bmatrix} h(z) \\ g(z) \end{bmatrix} \triangleq \mathcal{F}(z). \] (3)

where \( z = \begin{bmatrix} x^T \\ u^T \end{bmatrix}^T \). Distinguished merit of parallel control is that two dynamic systems \( \dot{x} = h(x, u) \) and \( \dot{u} = g(x, u) \) are constructed based on parallel control theory. It has great potential to improve the performance of controllers by gaming amongst the two dynamic systems.

**B. Optimal Parallel Tracking Control**

Consider the following general systems:
\[ \dot{x} = f(x, u) \] (4)

where \( x = \begin{bmatrix} x_1 \ x_2 \ldots \ x_n \end{bmatrix}^T \in \mathbb{R}^n \), \( u \in \mathbb{R}^m \) is the control input.

The objective of tracking control based on the state feedback is to design a controller \( u = \mathcal{K}(x) \) for system (4) which ensures that the \( x \) tracks the reference signal \( x_d(t) \).

Assume that the reference signal \( x_d(t) \) is given by
\[ \dot{x}_d = f(x_d, u_d) \] (5)

where \( x_d = \begin{bmatrix} x_{d,1} \ x_{d,2} \ldots \ x_{d,n} \end{bmatrix}^T \in \mathbb{R}^n \) and \( u_d \in \mathbb{R}^m \) are state and control input of the reference system respectively.

Let \( e = x - x_d \) and \( u_e = u - u_d \). Then, the error dynamics system is defined as
\[ \dot{e} = \dot{x} - \dot{x}_d = f(x, u) - f(x_d, u_d) = f(e, u_e). \] (6)

The objective of the optimal state feedback tracking control is to design a controller \( u_e = \mathcal{K}(e) \) for system (6) which ensures that the \( x \) tracks the reference signal \( x_d(t) \) and minimizes the following infinite horizon performance index function:
\[ J(e, u_e) = \int_t^\infty U(e(\tau), u_e(\tau))d\tau \] (7)

where \( U(e, u_e) = e^T Q e + u_e^T R u_e \) is the utility function and \( Q \in \mathbb{R}^{n \times n} \) and \( R \in \mathbb{R}^{m \times m} \) are symmetric positive definite matrices.

The objective of the optimal parallel tracking control is to design a controller \( u_e = g_e(e, u_e) \) for system (6) which ensures that the \( x \) tracks the reference signal \( x_d(t) \) and minimizes the following infinite horizon performance index function:
\[ J_p(e, u_e, \dot{u}_e) = \int_t^\infty U_p(e(\tau), u_e(\tau), \dot{u}_e(\tau))d\tau \] (8)

where \( U_p(e, u_e, \dot{u}_e) = [e \ u_e]^T Q_p [e \ u_e]^T + \dot{u}_e^T R_p \dot{u}_e \) is the utility function and \( Q_p \in \mathbb{R}^{(n+m) \times (n+m)} \) and \( R_p \in \mathbb{R}^{m \times m} \) are symmetric positive definite matrices.

According to (3), a closed-loop system can be obtained:
\[ \begin{align*}
\dot{e} &= f_e(e, u_e) \\
\dot{u}_e &= g_e(e, u_e).
\end{align*} \] (9)

According to the system (9), parallel control of the system (4) is different since control input \( u \) is introduced into the feedback system, the energy function of \( u \) is considered as a part of the Lyapunov function. So different from the Lyapunov function \( V(e) \) for state feedback control, an augmented Lyapunov function \( V_p(e, u_e) \) should be considered. The optimal state feedback control methods can not be used directly, i.e., ADP and linear quadratic regulator (LQR).

The following assumptions are made for the system (4) for convenience of analysis.

**Assumption 1**: Assume that:
1) \( f(x, u) \) is Lipschitz continuous on a compact set \( \Omega \) including the origin, system is stabilizable on the \( \Omega \), and \( f(0,0) = 0 \).
2) State \( x \) is measurable;
3) Reference signal \( x_d(t) \) is a bounded piecewise smooth function with bounded derivatives and finite jump discontinuities.

**Remark 1**: The reference signals mentioned in Assumption 1 contain the most common signals, i.e., step signal, square wave signal, sine signal.

**III. OPTIMAL PARALLEL TRACKING CONTROL DESIGN**

In this section, the optimal parallel tracking control is developed. First, an augmented system and an augmented performance index function are proposed. Then, the difference between the optimal parallel control and optimal state feedback control is analyzed theoretically.

**A. Augmented Error System**

As mentioned above, the traditional optimal control methods can not be used directly due to the introduction of parallel controller (2). Inspired by [53], [54], a variable is introduced as follows:
\[ e_u = u_e. \] (10)

Differentiating the equation (10) and letting \( s = [e \ e_u]^T \), we obtain the following affine system
\[ \dot{s} = f_s(s) + g_s v_s \] (11)

where
\[ f_s(s) = \begin{bmatrix} f_e(s) \\ 0_m \end{bmatrix}, \quad g_s = \begin{bmatrix} 0_{n \times n} \\ I_{m \times m} \end{bmatrix}. \]

\( v_s = v - v_d, \quad v = \dot{u} \) and \( v_d = \dot{u}_d \).

**Theorem 1**: Consider the system (6) and the system (11). If there exists at least a differentiable \( u \) such that the system (6) and the system (11) are globally asymptotic stable (GAS) and let Assumption 1 be met. Then, the system (6) has a GAS equilibrium at the origin if and only if the system (11) has a GAS equilibrium at the origin with the same control input.

**Proof**: \((\Rightarrow)\) If the system (6) has a GAS equilibrium at the origin, then
\[ \lim_{t \to \infty} ||e(t, e(t_0))|| = 0 \] (12)

for all \( e(t_0) \in \mathbb{R}^n \).
The construction of such a differentiable $\Omega$ guaranteed based on Assumption 1, which means the system for all equilibrium at the origin.

$u$ following.

The difference between the optimal parallel control with different symmetric positive definite matrix $R$. Comparison With The Optimal State Feedback Control

Since the new virtual variable is introduced, the existence of such a differentiable $u$ can be guaranteed based on Assumption 1, which means the system (11) is stabilizable on a compact set $\Omega_p$ including the origin. The construction of such a differentiable $u$ will be shown in the next section.

Remark 2: The existence of such a differentiable $u$ can be guaranteed based on Assumption 1, which means the system (11) is stabilizable on a compact set $\Omega_p$ including the origin. The construction of such a differentiable $u$ will be shown in the next section.

Remark 3: Since the new virtual variable is introduced, the general system (6) is transformed into the affine system (11) with the invariable control weighting $g_s$.

B. Comparison With The Optimal State Feedback Control

Compared with the optimal state feedback control, a new symmetric positive definite matrix $R_p$ is introduced in performance index function to restrict the energy of $\dot{u}_e$. The difference between the optimal parallel control with different $R_p$ and optimal state feedback control will be analyzed in the following.

Consider the system (6) and the system (11), the value function of an admissible control $u_e(t)$ is defined as

$$V_{u_e}(e) = \int_t^\infty U(e(\tau), u_e(\tau))d\tau.$$  (18)

The optimal control $u^*_e(e)$ is given by

$$u^*_e(e) = \arg\min_{\mu \in \Psi(t)} \int_t^\infty U(e(\tau), \mu(\tau))d\tau.$$  (19)

Based on the principle of optimality, the optimal performance index function is given by

$$J^*(e) \triangleq \min_{u_e} J(e, u_e) = \int_t^\infty U(e(\tau), u^*_e(\tau))d\tau.$$  (20)

For the optimal control $u^*_e(e)$ in (19), denoting its optimal value function as $V^*(e) \triangleq V_{u^*_e}(e)$, we have

$$V^*(e) = J^*(e).$$  (21)

Similarly, consider the system (11) and the performance index (8). The value function of an admissible control $v_s(s)$ is defined as

$$V_{v_s}^*(s) = \int_t^\infty U_p(s(\tau), v_s(\tau))d\tau.$$  (22)

The optimal parallel control $v^*_s(s)$ is given by

$$v^*_s(s) = \arg\min_{v \in \Psi(t)} \int_t^\infty U_p(s(\tau), v(\tau))d\tau.$$  (23)

The optimal performance index function is considered as

$$J^*_p(s) \triangleq \min_{v_s} J_p(s, v_s) = \int_t^\infty U_p(s(\tau), v^*_s(\tau))d\tau.$$  (24)

Denote the optimal value function of the optimal control $v^*_s(s)$ by $V_{v_s}^*(s) \triangleq V_{v^*_s}^*(s)$. Then,

$$V_{v_s}^*(s) = J^*_p(s).$$  (25)

In the following, the effect on the value function with different $R_p$ is analyzed theoretically. Before starting, two value functions $V_{v^*_s}^*(e)$ and $V_{v^*_s}^*(s)$ with certain control policies are defined as follows:

$$V_{v^*_s}^*(e) = \int_t^\infty U(e(\tau), u_{e,v^*_s}(\tau))d\tau$$  (26)

and

$$V_{v^*_s}^*(s) = \int_t^\infty U_p(s(\tau), v_{s,v^*_s}(\tau))d\tau.$$  (27)

where $u_{e,v^*_s}(\tau) = \int v^*_s(t)dt$, and $v_{s,v^*_s}(\tau) = \dot{u}^*_s(\tau)$.

Corollary 1: Consider the system (6) and the system (11), and let

$$Q_p = \begin{bmatrix} Q & 0_{m \times n} \\ 0_{n \times m} & R \end{bmatrix}$$

and initial control input $e_u(t) = u_e(t)$. Let $\|v\|_{R_p}^2 \triangleq v^TR_p v$ for any vector $v \in \mathbb{R}^m$. Then:

1) $V^*(e) \leq V_{v^*_s}^*(e) \leq V_{v^*_s}^*(s) \leq V_{v^*_s}^*(s)$;

2) $V^*(s)$ and $V_{v^*_s}^*(e)$ are close to the optimal value function of the state feedback control $V^*(e)$ within a bound $\varepsilon_V$. the bound $\varepsilon_V$ decreases with the decrease of $\|v\|_{R_p}^2$, i.e., $V_{v^*_s}^*(s) - V^*(e) \leq \varepsilon_V$ and $V_{v^*_s}^*(e) - V^*(e) \leq \varepsilon_V$, $\varepsilon_V \rightarrow 0^+$ as $\|v\|_{R_p}^2 \rightarrow 0^+$.

3) $u_{e,v^*_s}(\tau) \leq 0^+$ is close to the optimal state feedback control $u^*_e$ within a small bound $\varepsilon_u$ as $\|v\|_{R_p}^2 \rightarrow 0^+$, i.e., $\|u_{e,v^*_s} - u^*_e\| \leq \varepsilon_u$ as $\|v\|_{R_p}^2 \rightarrow 0^+$.

Proof:

1) Based on the principle of optimality, we have $V^*(e) \leq V_{v^*_s}^*(e)$ and $V_{v^*_s}^*(s) \leq V_{v^*_s}^*(s)$. 
According to Theorem 1, \( u, v^* \) and \( v_s, u_s^* \) are admissible control inputs for system (6) and system (11) respectively and guarantee \( V^v(e) \) and \( V_p^{u}(s) \) to be finite. Then,

\[
V_p^{u*}(s) - V^*(e) = \int_t^\infty \left( U_p(s(\tau), v_s^*(\tau)) - U(e(\tau), u_e, v^*_e(\tau)) \right) d\tau
\]

\[
= \int_t^\infty \left( s(\tau)^T Q_p s(\tau) + v_s^*(\tau)^T R_p v_s^*(\tau) - e(\tau)^T Q e(\tau) - u_e^*(\tau)^T R u_e^*(\tau) \right) d\tau
\]

\[
= \int_t^\infty v_s^*(\tau)^T R_p v_s^*(\tau) d\tau
\]

\[
\geq 0
\]

(28)

for all \( e(t) \in \mathbb{R}^n \). Since \( R_p \) is a symmetric positive definite matrix, equality holds if and only if \( s = 0 \). Thus \( V^*(e) \leq V_p^{u*}(s) \leq V_p^{u}(s) \leq V^v(e) \).

2) According to (21) and (27), we have

\[
V_p^{u*}(s) - V^*(e) = \int_t^\infty \left( U_p(s(\tau), v_s^*(\tau)) - U(e(\tau), u_e, v^*_e(\tau)) \right) d\tau
\]

\[
= \int_t^\infty \left( s(\tau)^T Q_p s(\tau) + v_s^*(\tau)^T R_p v_s^*(\tau) - e(\tau)^T Q e(\tau) - u_e^*(\tau)^T R u_e^*(\tau) \right) d\tau
\]

\[
= \int_t^\infty v_s^*(\tau)^T R_p v_s^*(\tau) d\tau
\]

\[
\geq 0
\]

(29)

for all \( e(t) \in \mathbb{R}^n \).

According to the 1) of Corollary 1, we have

\[
V_p^v*(s) - V^*(e) \leq V_p^{u*}(s) - V^*(e) = \varepsilon_V
\]

(30)

and

\[
V^v(e) - V^*(e) \leq V_p^{u*}(s) - V^*(e) = \varepsilon_V.
\]

(31)

Note that the \( v_s, u_s^* (t) \) is constructed based on the \( u_s^*(t) \), which means \( v_s, u_s^* (t) \) is independent of \( R_p \). The \( \varepsilon_V \) decreases with the decrease of \( \|v\|_R^2 \), which means \( \varepsilon_V \rightarrow 0^+ \) as \( \|v\|_R^2 \rightarrow 0^+ \).

3) According to (21), (26), and the 2) of Corollary 1, we have

\[
V^v(e) - V^*(e) = \int_t^\infty \left( U(e(\tau), u_e, v^*_e(\tau)) - U(e(\tau), u_e, v^*_e(\tau)) \right) d\tau
\]

\[
\leq \varepsilon_V.
\]

(32)

for all \( e(t) \in \mathbb{R}^n \). Thus \( \|U(e(\tau), u_e, v^*_e) - U(e, u_e, v^*_e)\| \rightarrow 0^+ \) as \( \|v\|_R^2 \rightarrow 0^+ \), which means \( \|u_e, v^*_e - u_e^*\| \leq \varepsilon_u \) as \( \|v\|_R^2 \rightarrow 0^+ \).

Remark 4: According to the properties of Rayleigh quotient, \( \|v\|_R^2 \) decreases with the decrease of eigenvalue of the \( R_p \). Since \( \varepsilon_V \rightarrow 0^+ \) as \( \|v\|_R^2 \rightarrow 0^+ \), the optimal parallel control with the performance index function (8) can be seen an suboptimal control with performance index function (7).

Remark 5: In the optimal state feedback control, researchers pay more attention to the quality and constraints of control input \( u \) and show less attention to the quality of \( \dot{u} \). But it is not proper for many control problems, e.g., 160 MW Boiler-Turbine-Alternator Units [33] proposed by Aström. The parallel control theory combines with the existing control methods can achieve better performance, e.g., backstepping, LQR, and ADP.

IV. IMPLEMENTATION WITH ADAPTIVE DYNAMIC PROGRAMMING

In this section, the implementation of the optimal parallel tracking control based on the ADP is presented and the theoretical analysis is provided.

According to system (11), define the Hamiltonian function as

\[
H(s, v_s(s), \nabla V_p(s)) = U_p(s, v_s(s)) + \nabla^T V_p(s)
\]

\[
\times (f_s(s) + g_s v_s(s))
\]

(33)

where \( \nabla V_p \) denotes gradient of the \( V_p(s) \) with respect to \( s \).

The optimal value function \( V_p^*(s) \) is given by (25) and satisfies the following HJB equation:

\[
0 = \min_{\nu \in \Psi (s)} H(s, v_s(s), \nabla V_p^*(s)).
\]

(34)

Since the general system (5) is transformed into the affine system (11), the optimal parallel control input \( v_s^* \) can be obtained by solving \( \partial H(s, v_s(s), \nabla V_p^*(s)) \)/\( \partial v_s = 0 \) as

\[
v_s^*(s) = -\frac{1}{2} R_p^{-1} g_s^T \nabla V_p^*(s).
\]

(35)

Then, the overall control is \( v^* = v_d + v_s^* \).

Substituting (35) into (34) yields

\[
H(s, v_s^*(s), \nabla V_p^*(s)) = 0
\]

(36)

which can be written as

\[
\nabla^T V_p^*(s) f_s(s) + s^T Q_p s - \frac{1}{4} \nabla^T V_p^*(s) g_s R_p^{-1} \nabla g_s \nabla V_p^*(s) = 0.
\]

(37)

As mentioned above, the analytic solutions of HJB equation (37) can not be obtained directly in most cases, thus the ADP method is utilized to the implementation of the optimal parallel tracking control in the following.

A. Offline Implementation

HJB equation (37) can be solved offline based on the HDP with a classical three NNs framework that is suggested by Werbos [10], [11]. The detailed training strategies of the iterative ADP can be found in [3], [22].

The difference with the existing work is that the developed method does not try to obtain the optimal control \( u^*_e \), which is acting on the real physical system, but to obtain the optimal derivative of the control \( u^*_e \).
Furthermore, it should be noted that HJB equation (37) can be solved without concrete equation of system (4) by using HDP, because \( g_s \) is an artificial one and is given by

\[
g_s = \begin{bmatrix} 0_{n \times m} & I_{m \times m} \end{bmatrix}^T.
\]

Once iterative \( V_p(s) \) is obtained, the iterative \( v_s(s) \) can be obtained using the first-order necessary condition.

The difficulty in solving HJB equation (37) without concrete equation of system (4) is to design feedforward control \( u_d \) and \( v_d \). When it comes to optimal regulator problem, \( u_d = v_d = 0 \). Then, the problem mentioned above would not exist.

B. Online Implementation

To achieve online implementation, a critic NN is utilized to approximate \( V_p(s) \) as follows:

\[
V_p(s) = W^T \phi(s) + \varepsilon(s)
\]

where \( W^* \) is the ideal weight vector, \( \phi(s) : \mathbb{R}^{n+m} \to \mathbb{R}^N \) is the vector of critic NN activation functions, \( N \) is the number of activation functions, and \( \varepsilon(s) \) is the critic NN approximation error.

The derivative of the \( V_p(s) \) with respect to \( s \) is:

\[
\nabla V_p(s) = \nabla T \phi(s) W^* + \nabla \varepsilon(s)
\]

where \( \nabla \phi(s) \triangleq \partial \phi(s) / \partial s \) and \( \nabla \varepsilon(s) \triangleq \partial \varepsilon(s) / \partial s \).

The optimal control (35) and HJB equation (37) can be rewritten as

\[
v_s(s) = -\frac{1}{2} R_p^{-1} g_s^T \nabla \phi(s) W^* - \frac{1}{2} R_p^{-1} g_s^T \nabla \varepsilon(s)
\]

and

\[
s^T Q_p s + W^T \nabla \phi(s) f_s(s) - \frac{1}{4} W^T \nabla \phi(s) D \nabla T \phi(s) - W^* + \varepsilon_H = 0
\]

where \( D = g_s R_p^{-1} g_s^T > 0 \) is a constant matrix.

In implementation, the estimation of \( V_p(s) \) is

\[
\hat{V}_p(s) = \hat{W}^T \phi(s)
\]

where \( \hat{W} \) is the estimation of the \( W^* \).

It follows from (40) and (42) that the control input is given by

\[
v_s(s) = -\frac{1}{2} R_p^{-1} g_s^T \nabla \phi(s) \hat{W}.
\]

The approximate Hamiltonian function with the estimation error of the critic NN is given by:

\[
H(s, v_s(s), \hat{W}) = U(s, v_s(s)) + \hat{W}^T \nabla \phi(s) (f_s(s) + g_s v_s(s)) = \xi(\hat{W}).
\]

The objective is to minimize the squared residual error given below by tuning the \( \hat{W} \):

\[
E(\hat{W}) = \xi^T(\hat{W}) \xi(\hat{W}).
\]

Inspired by works \([29],[30],[31]\), an online ADP method is employed to obtain the \( W^* \). Before starting, the following assumption is made.

**Assumption 2:** Let \( P(s) \in C^1(\Omega_p) \) be a radially unbounded Lyapunov function candidate that satisfies

\[
\nabla^T P(s)(f_s(s) + g_s v_s(s)) \leq 0
\]

Moreover, there exists a positive definite matrix \( \bar{Q}(s) \in \mathbb{R}^{(n+m) \times (n+m)} \) that satisfies

\[
\nabla^T P(s)(f_s(s) + g_s v_s(s)) = -\nabla^T P(s) \bar{Q}(s) \nabla P(s)
\]

with \( 0 < \bar{Q}_{\text{min}} \leq \bar{Q}(s) \leq \bar{Q}_{\text{max}} \).

It is desired to guarantee the system stability while tuning \( \hat{W} \). Therefore, the following tuning rule is developed:

\[
\dot{\hat{W}} = -\frac{\alpha_1 \sigma(s)}{(\sigma(s) \sigma(s) + 1)^2} (\hat{W}^T \sigma(s) + s^T Q_p s + v_s^T R_p v_s(s) + \kappa \alpha_2 \nabla \phi(s) g_s R_p^{-1} g_s^T \nabla P(s)
\]

where \( \alpha_1 > 0 \) and \( \alpha_2 > 0 \) are tuning parameters and

\[
\sigma(s) = \nabla \phi(s) (f_s(s) + g_s v_s(s))
\]

The key advantage of the tuning rule (48) is that it does not require an initial admissible control input for (41) by introducing (50). Note that the tuning rule is different than that in \([30]\) as it is designed to obtain the optimal derivative of control input.

The stability of system (4) with the parallel control (43) and tuning rule (48) is provided in the following. Before starting, the definition of UUB and Assumption 3 are given as follows.

**Definition 1** \([55]\): An equilibrium point \( s_0 = 0 \) of (11) is said to be UUB, if there exists a compact set \( \Omega_p \subset \mathbb{R}^{n+m} \) so that for all \( s_0 \in \Omega_p \) there exists a bound \( B \) and a time \( T(B, s_0) \) such that \( \| s(t) - s_0 \| \leq B \) for all \( t \geq t_0 + T \).

**Assumption 3:** Assume that:

1. The ideal critic NN weights vector \( W^* \) is bounded, i.e., \( \| W^* \| \leq W_M \) with \( W_M > 0 \);
2. The estimation error of critic NN \( \varepsilon(s) \) and its gradient \( \nabla \varepsilon(s) \) are bounded on a compact set containing \( \Omega_p \), i.e., \( \varepsilon(s) \leq \varepsilon_M \) and \( \nabla \varepsilon(s) \leq d_{\varepsilon_M} \) with \( \varepsilon_M, d_{\varepsilon_M} > 0 \);
3. The optimal closed loop dynamics are bounded by a function of the system states, i.e., \( \| f_s(s) + g_s v_s(s) \| \leq \delta(s) \leq \sqrt{K^* \| \nabla P(s) \|} \).

**Theorem 2:** Consider the system (5) with the parallel control input (40) and tuning rule (48). Let Assumptions 1–3 be met. Then, the tracking error \( e \) and the estimation error of the critic NN weights vector \( \hat{W} = W^* - \hat{W} \) are UUB. The control input \( u \) is continuous.

**Proof:** As motivated in \([30],[31]\), the detailed proof of Theorem 2 is given in the Appendix.

**Remark 6:** The virtual variable \( e_v \) is introduced in constructing the augmented error system (11). It is worth pointing
out that \( e_u = u - u_d \), so no actual measurement is added when implementing the optimal parallel control.

Remark 7: Generally speaking, the stability of the system (1) can be guaranteed theoretically in optimal state feedback control with jump discontinuities, but it requires control input \( u \) is able to make the step-changes at the jump discontinuities, which is difficult to execute on the mechanical actuators.

V. Numerical Analysis

In this section, two simulations are employed to evaluate the effectiveness of the developed optimal parallel control.

Example 1: In the first example, we consider the following linear system [56]

\[
\dot{x} = Ax + Bu
\]

where the system matrices are given by

\[
A = \begin{bmatrix}
0 & 1 \\
-2 & 3
\end{bmatrix},
B = \begin{bmatrix}
0 \\
1
\end{bmatrix}.
\]

Reference signal \( x_d(t) \) is given by

\[
x_d(t) = \begin{cases}
1 & 0 \leq t < 10 \\
2 & 10 \leq t \leq 20.
\end{cases}
\]

The augmented error system can be obtained according to (11):

\[
\dot{s} = Ap s + Bp v_s
\]

where the matrices are given by

\[
A_p = \begin{bmatrix}
0 & 1 & 0 \\
-2 & 3 & 1 \\
0 & 0 & 0
\end{bmatrix},
B_p = \begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}.
\]

Let \( Q_p = I_{3\times3} \) and \( R_p = 0.1 \). Then, the optimal value function can be obtained and is given by \( V_p(s) = s^T P_p s \) with

\[
P_p = \begin{bmatrix}
15.6447 & -0.30786 & -1.1539 \\
-3.0786 & 25.2959 & 3.8291 \\
-1.1539 & 3.8291 & 0.9305
\end{bmatrix}.
\]

Then, the optimal parallel controller is given by

\[
v_s = -K_p s
\]

where optimal gain matrix

\[
K_p = \begin{bmatrix}
-11.9193 & 24.4568 & 7.0777
\end{bmatrix}.
\]

To validate the effectiveness of the developed optimal parallel controller, an optimal state feedback controller with \( Q = I_{2\times2} \) and \( R = 1 \) is design as

\[
u_e = -K e
\]

where optimal gain matrix \( K = \begin{bmatrix}
0.2361 & 6.2361
\end{bmatrix} \).

The initial state \( x(0) = [0, 0]^T \). The initial control input \( e_u(0) = u_e(0) = 0.2361 \). The trajectories of the system state are shown in Fig. 4, and the trajectories of the control input are shown in Fig. 5. It can be seen from the Figs. 4 and 5 that the system is stable with the parallel controller, and state \( x \) is able to track \( x_d \) steadily.

Furthermore, it can be seen from the Figs. 4 and 5 that within the first four seconds, the optimal parallel controller is capable of achieving almost the same performance with the optimal state feedback controller. Performance index (7) \( V^{v_s}(e(0)) = 13.2367 \) and \( V^{u_e}(e(0)) = 13.2361 \).

With the sudden change of the reference signal \( x_d \) in the fourth second, the overall control input \( u \) of the optimal state feedback controller abruptly changes at this point as the steady-state controller \( u_d \) steps from 2 to 4 and error state \( e_1 \) steps from 0 to 1, which is difficult to achieve on actuators. Compared with the optimal state feedback controller, the optimal parallel controller provides a more reasonable way by introducing the control input into feedback. The control input of the optimal parallel controller is continuous even with the sudden change of the reference signal. In engineering applications, it is important to choose a reasonable \( R_p \) according to the practical restrictions of mechanical actuators.

To compare the difference between the optimal state feedback controller and parallel controller with different \( R_p \), the relative difference (RD) is defined as:
The initial states are $u$ and $u_v = e_u + u_d$. The RD with different $R_p$ is shown in Fig. 6. It can be seen from Fig. 6 that the difference between $u_e$ and $e_u$ is decreasing as the $R_p$ decreases. The optimal parallel controller with $R_p \leq 10^{-5}$ behaves pretty much the same as the state feedback controller as Corollary 1 predicted. So different performance can be achieved by choosing $R_p$ according to the control objectives and the practical restrictions.

$$\text{RD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (u_v(i) - u(i))^2 / \sum_{i=1}^{N} u(i)^2}$$  \hspace{1cm} (57)$$

where $N$ is the length of $u_v$ and $u$, and $u_v = e_u + u_d$. The RD with different $R_p$ is shown in Fig. 6. It can be seen from Figs. 7 and 8 that the trajectory of the control input is shown in Fig. 7, and the trajectory of the system state is shown in Fig. 9, the trajectory of the system state is shown in Fig. 9, the trajectory of the control input is shown in Fig. 10, and the trajectories of the system state are shown in Fig. 11. It can be seen from Figs. 9–11 that the developed optimal parallel tracking controller is effective with the time-varying reference signal.

The reference signal is given by

$$x_d(t) = \begin{cases} [0 \ 0]^T & t \in [0, 20), [40, 60), [80, 100) \\ [1 \ 1]^T & t \in [20, 40), [60, 80). \end{cases}$$  \hspace{1cm} (59)$$

The augmented error system can be obtained according to (11):

$$\begin{aligned} \dot{s}_1 &= - (s_1 + x_{d,1})^3 + s_2 + x_{d,2} \\ \dot{s}_2 &= (s_1 + x_{d,1})^2 - s_1 - x_{d,1} - s_2 - x_{d,2} \\ \dot{s}_3 &= -0.1(s_3 + u_d)^3 + \sin(0.1(s_3 + u_d)) \end{aligned}$$  \hspace{1cm} (60)$$

To approximate the value function $V_p(s)$, the activation functions of the critic NN are selected as

$$\phi(s) = \begin{bmatrix} s_1^2 & s_1s_2 & s_1s_3 & s_2s_3 & s_2^2 & s_3^2 & s_4 & s_1s_2 & s_1s_3 & s_1^2s_2 & s_2^2s_3 & s_3^2s_2 & s_2^2s_3 & s_3^2s_3 & s_4^2 \end{bmatrix}^T.$$  \hspace{1cm} (61)$$

$Q_p$ is chosen to be $I_{3 \times 3}$ and $R_p$ is chosen to be 1. The initial state is $x(0) = [1 \ -1]^T$, initial control input $u(0) = 1.7647$, and the initial critic NN weights are chosen to be 0. The parameter of tuning rule $\alpha_1 = 10$ and $\alpha_2 = 1$. The trajectory of the system state is shown in Fig. 9, the trajectory of the control input is shown in Fig. 10, and the trajectories of the critic NN weights are shown in Fig. 11. It can be seen from Figs. 9–11 that the developed optimal parallel tracking controller is able to stabilize nonlinear system effectively, the tracking error converges to zero or a small neighborhood of zero.

**Example 2:** In the second example, we examine the performance of the developed optimal parallel controller in a nonaffine system:

$$\begin{aligned} \dot{x}_1 &= -x_1^3 + x_2 \\ \dot{x}_2 &= x_1^2 - x_1 - x_2 + 0.15u^3 + \sin(0.1u) \end{aligned}$$  \hspace{1cm} (58)$$

Fig. 7. State trajectories of Example 1 with time-varying reference signal.

Fig. 8. Control input of Example 1 with time-varying reference signal.

The reference input of Example 1 with time-varying reference signal.

$$x_d(t) = \begin{cases} [0 \ 0]^T & t \in [0, 20), [40, 60), [80, 100) \\ [1 \ 1]^T & t \in [20, 40), [60, 80). \end{cases}$$  \hspace{1cm} (59)$$

Fig. 6. RD with different $R_p$. To further verify the effectiveness of the developed optimal parallel controller, a time-varying reference signal $x_d = \begin{bmatrix} \sin(t) \\ \cos(t) \end{bmatrix}^T$, $t \geq 0$ is used. Let $Q_p = I_{3 \times 3}$ and $R_p = 0.1$. The initial states are $x(0) = [1 \ 1]^T$. The initial control input $u_e(0) = e_u(0) + u_d(0) = 0$. Then, the trajectory of the system state is shown in Fig. 7, and the trajectory of the control input is shown in Fig. 8. It can be seen from Figs. 7 and 8 that the developed optimal parallel controller is effective with the time-varying reference signal.
Furthermore, it can be seen from Figs. 9 and 11 that the critic NN weighs converge as the tracking error converges to zero. The performance with the final critic NN weighs is better than with the initial critic NN weighs. In parallel control, control input is seen as a state, so the control input of the optimal parallel controller is continuous with the sudden change of the reference signal as predicted. This property is very useful in industrial applications, especially for the control of the switched systems, as long as the candidate Lyapunov function for system is radially unbounded, the system will be stable with any initial control input $u_0$.

**VI. Conclusion**

Optimal parallel tracking control for general nonlinear systems via ADP is studied in this research. Different from the optimal state feedback controller, the variation of the control of the optimal parallel controller is constructed based on both system state and control input. First, an augmented system and an augmented performance index function are proposed. Then, it is proven that the optimal parallel control with the augmented performance index function can be seen as the suboptimal state feedback control with the traditional performance index function. Moreover, the ADP method is utilized to implement the optimal parallel control without using value iteration or policy iteration. We prove that the tracking error state and the NN weights error are UUB and control input is continuous with the optimal parallel controller. Finally, the effectiveness of the developed optimal parallel control method is verified in two cases.

**APPENDIX A**

**Proof of Theorem 2**

*Proof:* Consider the interval between any two adjacent discontinuities of the $x_d(t)$, we have

$$\dot{\hat{W}} = -\hat{W}. \quad (62)$$

It follows from (41) that

$$s^T Q_{ps} s = -W^{s^T} \nabla \phi(s) f_s(s) + \frac{1}{4} W^{s^T} \nabla \phi(s) D \nabla^T \phi(s) - \varepsilon_H. \quad (63)$$

The error dynamics of (48) are

$$\dot{\hat{W}} = -\frac{\alpha_1}{\rho^2(s)} \left( \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) + \nabla \phi(s) D \nabla^T \phi(s) \hat{W} + \varepsilon_H \right) - \kappa \alpha_2 \nabla \phi(s) D \nabla P(s) \quad (64)$$

where $\rho(s) = \sigma^T(s) \sigma(s) + 1$ and $\dot{s}^* = f_s(s) + g_s v_s^*$. Consider the following Lyapunov function candidate:

$$L = \alpha_2 P(s) + \hat{W}^T \hat{W} \quad (65)$$

If $\|s\| = 0$, $\dot{L} = 0$, and $\hat{W}$ remains a bounded constant. Consider the case that $\|s\| > 0$, by taking the derivative along with respect to time, we can get (66)
\[ \dot{L} = \alpha_2 \nabla P(s) \dot{s} + \dot{W}^T \dot{W} \]
\[ = \alpha_2 \nabla P(s) \left( f_s(s) - \frac{D \nabla^T \phi(s) \dot{W}}{2} \right) - \frac{\alpha_1}{\rho^2} \left( \dot{W}^T \nabla \phi(s) \left( \dot{s}^* - \frac{D \nabla \varepsilon(s)}{2} \right) \right)^2 - \frac{\alpha_1}{8 \rho^2} \left( \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} \right)^2 \]
\[ - \frac{3 \alpha_1}{4 \rho^2} \dot{W}^T \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} - \frac{\alpha_1}{\rho^2} \dot{W}^T \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) \varepsilon_H \]
\[ - \frac{\alpha_1}{2 \rho^2} \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} \varepsilon_H - \kappa_2 \dot{W}^T \nabla \phi(s) D \nabla P(s) \]

\[ = - \frac{\alpha_1}{2 \rho^2} \left( \dot{W}^T \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) \right)^2 - \frac{\alpha_1}{16 \rho^2} \left( \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} \right)^2 + \frac{3 \alpha_1}{4 \rho^2} \varepsilon_H^2 \]
\[ - \frac{\alpha_1}{2 \rho^2} \left( \dot{W}^T \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) + \varepsilon_H \right)^2 - \frac{\alpha_1}{16 \rho^2} \left( \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} + 4 \varepsilon_H \right)^2 - \frac{3 \alpha_1}{4 \rho^2} \dot{W}^T \nabla \phi(s) \frac{1}{2} \nabla \phi(s) \nabla \phi(s) \dot{W} + \alpha_2 \nabla P(s) \left( f_s(s) - \frac{D \nabla^T \phi(s) \dot{W}}{2} \right) - \kappa_2 \dot{W}^T \nabla \phi(s) D \nabla P(s) \]

\[ \leq - \frac{\alpha_1}{2 \rho^2} \left( \dot{W}^T \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) \right)^2 - \frac{\alpha_1}{16 \rho^2} \left( \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} \right)^2 + \frac{3 \alpha_1}{4 \rho^2} \dot{W}^T \nabla \phi(s) \frac{1}{2} \nabla \phi(s) \nabla \phi(s) \dot{W} + \alpha_2 \nabla P(s) \left( f_s(s) - \frac{D \nabla^T \phi(s) \dot{W}}{2} \right) - \kappa_2 \dot{W}^T \nabla \phi(s) D \nabla P(s) \]

\[ = - \frac{\alpha_1}{8 \rho^2} \left( \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} \right)^2 + 6 \dot{W}^T \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) - \frac{\alpha_1}{32 \rho^2} \left( \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} \right)^2 \]
\[ + \frac{4 \alpha_1}{\rho^2} \left( \dot{W}^T \nabla \phi(s) \frac{1}{2} \nabla \phi(s) \nabla \phi(s) \dot{W} \right)^2 + \frac{3 \alpha_1}{2 \rho^2} \varepsilon_H^2 + \alpha_2 \nabla P(s) \left( f_s(s) - \frac{D \nabla^T \phi(s) \dot{W}}{2} \right) - \kappa_2 \dot{W}^T \nabla \phi(s) D \nabla P(s) \]

\[ \leq - \frac{\alpha_1}{32 \rho^2} \left( \dot{W}^T \nabla \phi(s) D \nabla^T \phi(s) \dot{W} \right)^2 + \frac{4 \alpha_1}{\rho^2} \left( \dot{W}^T \nabla \phi(s) \left( \dot{s}^* + \frac{D \nabla \varepsilon(s)}{2} \right) \right)^2 + \frac{3 \alpha_1}{4 \rho^2} \dot{W}^T \nabla \phi(s) \frac{1}{2} \nabla \phi(s) \nabla \phi(s) \dot{W} \]
\[ + \alpha_2 \nabla P(s) \left( f_s(s) - \frac{D \nabla^T \phi(s) \dot{W}}{2} \right) - \kappa_2 \dot{W}^T \nabla \phi(s) D \nabla P(s). \] (66)

Taking bounds on (66) and completing the square with respect to \( \| \dot{W}^T \nabla \phi(s) \|^2 \), we have

\[ \dot{L} \leq \frac{3 \alpha_1}{2 \rho^2} \varepsilon_H^2 - \frac{\alpha_1}{64 \rho^2} \| \dot{W}^T \nabla \phi(s) \|^4 D_c^2 - \frac{\alpha_1 D_c}{\rho^2} \]
\[ \times \left( \| \dot{W}^T \nabla \phi(s) \|^2 - \frac{16 \| \dot{s}^* + D \nabla \varepsilon(s) / 2 \|^2}{D_c^2} \right)^2 \]
\[ + \frac{256 \alpha_1}{\rho^2} \| \dot{s}^* + D \nabla \varepsilon(s) / 2 \|^4 + \varphi(s, \dot{W}) \]
\[ \leq \frac{3 \alpha_1}{2 \rho^2} \varepsilon_H^2 - \frac{\alpha_1}{64 \rho^2} \| \dot{W}^T \nabla \phi(s) \|^4 D_c^2 \]
\[ + \frac{256 \alpha_1}{\rho^2} \| \dot{s}^* + D \nabla \varepsilon(s) / 2 \|^4 + \varphi(s, \dot{W}) \] (67)

where \( D_c = \| D \| \) and

\[ \varphi(s, \dot{W}) = \alpha_2 \nabla P(s) \left( f_s(s) - \frac{D \nabla^T \phi(s) \dot{W}}{2} \right) - \kappa_2 \dot{W}^T \nabla \phi(s) D \nabla P(s) \]

Assumption 3, it follows from (67) that

\[ \dot{L} \leq \frac{3 \alpha_1}{2 \rho^2} \left( \varphi(s)^4 + d_{\varepsilon_m}^2 + 2 d_{\varepsilon_m}^2 \right) - \frac{\alpha_1}{64 \rho^2} \| \dot{W}^T \nabla \phi(s) \|^4 D_c^2 \]
\[ + \frac{256 \alpha_1}{\rho^2} \left( \| \dot{s}^* \|^2 + 2 \| \frac{D \nabla \varepsilon(s)}{2} \|^2 \right)^2 + \varphi(s, \dot{W}) \]
\[ \leq \frac{3 \alpha_1}{2 \rho^2} \left( \varphi(s)^4 + d_{\varepsilon_m}^2 + 2 d_{\varepsilon_m}^2 \right) - \frac{\alpha_1}{64 \rho^2} \| \dot{W}^T \nabla \phi(s) \|^4 D_c^2 \]
\[ + \frac{2048 \alpha_1}{\rho^2} \left( \| \dot{s}^* \|^4 + \| \frac{D \nabla \varepsilon(s)}{2} \|^4 \right) + \varphi(s, \dot{W}) \] (68)

where \( \beta_1 = \nabla \phi_{m}^4 / 64, \beta_2 = 2048 / D^2 + 3 / 2, \eta(\varepsilon) = 128 d_{\varepsilon_m}^2 D_c^2 + 3 \left( d_{\varepsilon_m}^2 + 2 d_{\varepsilon_m}^2 \right) / 2, \) and \( 0 < \nabla \phi_{m} \leq \| \nabla \phi(s) \| . \)

According to the definition of \( \kappa \) in (50), it has two cases.

**Case 1:** if \( \nabla P(s) \left( f_s(s) + g_s v_s \right) \leq 0, \kappa = 0. \) Then, it follows from (68) that
\[
\dot{\hat{L}} \leq -\alpha_2 \|\nabla P(s)\| \hat{s} - \frac{\alpha_1 \beta_1}{\rho^2} \|\hat{W}\|^4 + \frac{\alpha_1 \eta(z)}{\rho^2} + \frac{\alpha_2 \beta_2 \delta(s)}{\rho^2} \\
\leq -\alpha_2 \|\nabla P(s)\| \hat{s}_{\text{min}} - \frac{\alpha_1 \beta_1}{\rho^2} \|\hat{W}\|^4 \\
+ \frac{\alpha_1 \eta(z)}{\rho^2} + \frac{\alpha_2 \beta_2 K^*}{\rho^2} \|\nabla P(s)\| \\
= -\|\nabla P(s)\| \left( \frac{\alpha_2 \hat{s}_{\text{min}} - \alpha_2 \beta_2 K^*}{\rho^2} \right) \\
- \frac{\alpha_1 \beta_1}{\rho^2} \|\hat{W}\|^4 + \frac{\alpha_1 \eta(z)}{\rho^2} \\
\tag{69}
\]

where \( \hat{s}_{\text{min}} \) is a constant and satisfies \( 0 < \hat{s}_{\text{min}} \leq \|s\| \). \( \hat{L} \) keeps negative definite provided that \( \hat{s}_{\text{min}} > \alpha_2 \beta_2 K^*/\alpha_2 \rho^2 \), and \( \|\hat{W}\| > \sqrt[4]{\eta(z)/\beta_1} \).

Case 2: if \( \nabla^T P(s) (f_s(s) + g_s v_s) > 0, \kappa = 1/2 \). Then, it follows from (68) that

\[
\dot{\hat{L}} \leq \alpha_2 \nabla^T P(s) \left( f_s(s) - \frac{D \nabla^T \phi(s) \hat{W}}{2} - \frac{\alpha_1 \beta_1}{\rho^2} \|\hat{W}\|^4 \\
+ \frac{\alpha_1 \eta(z)}{\rho^2} + \frac{\alpha_2 \hat{W} \nabla \phi(s) D \nabla P(s)}{2} \\
+ \frac{1}{2} \alpha_2 \nabla^T P(s) D \left( \nabla^T \phi(s) W^* + \nabla^2 \phi(s) \right) \\
- \frac{1}{2} \alpha_2 \nabla^T P(s) D \left( \nabla^T \phi(s) W^* + \nabla^2 \phi(s) \right) \\
\leq -\alpha_2 \hat{Q}_{\text{min}} \|\nabla^T P(s)\|^2 - \frac{\alpha_1 \beta_1}{\rho^2} \|\hat{W}\|^4 + \frac{\alpha_1 \eta(z)}{\rho^2} \\
+ \frac{\alpha_2 \beta_2 K^*}{\rho^2} \|\nabla P(s)\| + \frac{1}{2} \alpha_2 \|\nabla^T P(s)\| D_e d_{x_M} \\
= -\frac{\alpha_1 \eta(z)}{\rho^2} - \frac{\alpha_2 \hat{Q}_{\text{min}}}{2} \|\nabla^T P(s)\|^2 - \frac{\alpha_1 \beta_1}{\rho^2} \|\hat{W}\|^4 \\
- \frac{\alpha_2 \hat{Q}_{\text{min}}}{2} \left( \frac{\alpha_2 \beta_2 K^*}{\rho^2} \|\nabla P(s)\| - \frac{2 \alpha_2 \beta_2 K^*}{\rho^2} \|D_e d_{x_M} \|^2 \right) \\
+ \frac{\alpha_2 \hat{Q}_{\text{min}}}{2} \left( \frac{2 \alpha_2 \beta_2 K^*}{\rho^2} \|D_e d_{x_M} \|^2 + \frac{\alpha_1 \beta_1}{\rho^2} \hat{W}^2 \right) \\
\tag{70}
\]

Since the fourth term in (70) is negative semidefinite, we have

\[
\dot{\hat{L}} \leq -\frac{\alpha_2 \hat{Q}_{\text{min}}}{2} \|\nabla^T P(s)\|^2 - \frac{\alpha_1 \beta_1}{\rho^2} \|\hat{W}\|^4 \\
+ \frac{\alpha_1 \eta(z)}{\rho^2} + \frac{\alpha_2 D_e d_{x_M}^2}{4 \hat{Q}_{\text{min}}} + \frac{\alpha_2 \beta_2 K^*}{\rho^2} \|\nabla P(s)\| \\
\tag{71}
\]

Based on (71), \( \hat{L} \) keeps negative definite provided that

\[
\|\nabla P(s)\| > \sqrt{\frac{2 \alpha_1 \eta(z)}{\alpha_2 \rho^2 \hat{Q}_{\text{min}}} + \frac{\alpha_2 D_e d_{x_M}^2}{4 \hat{Q}_{\text{min}}} + \frac{2 \alpha_2 \beta_2 K^*}{\alpha_2 \rho^2 \hat{Q}_{\text{min}}} \|\hat{W}\|^4 \\
\text{or} \\
\|\hat{W}\| > \sqrt{\frac{\eta(z)}{\beta_1} + \frac{\alpha_2 D_e d_{x_M}^2}{4 \alpha_2 \beta_2 \hat{Q}_{\text{min}}} + \frac{\alpha_2 \beta_2 K^*}{\alpha_2 \beta_1 \rho^2 \hat{Q}_{\text{min}}} \|\hat{W}\|^4} \\
\]

Therefore, state \( s \) and the critic NN weights error vector \( \hat{W} \) are UUB.

Note that \( u_d \), which is constructed based on the \( x_d \), is a feedback control input and \( x_d \) is a bounded function with bounded derivatives. Therefore, \( u_d \) is UUB. Moreover, \( e^T e \leq s^T s \), which means \( e \) is UUB.

Jump discontinuities of the \( x_d \) lead to the sudden change of state \( s \), which means tuning rule (48) could be activated. As the tuning rule (48) guarantees the stability of (11) with any initial critic NN weights vector \( \hat{W} \) and jump discontinuities are finite, the errors \( s \) and \( \hat{W} \) are UUB.

Finite jump discontinuities of the \( x_d \) lead to the finite jump discontinuities or removing discontinuities of \( v_s \) and \( \hat{u}_d \). According to (5), (10), and (11), we have

\[
u = u_d + e_u = \int (\hat{u}_d(\tau) + v_s(\tau)) \, d\tau. \tag{72}
\]

Thus, the global continuity of \( u \) can be guaranteed.

REFERENCES


Jingwei Lu received the bachelor degree in automation from University of Science and Technology Liaoning, Anshan, China, in 2014, and the master degree in control science and engineering from Beijing University, Beijing, China, in 2017. He was an assistant engineer with the AVIC Manufacturing Technology Institute, Beijing, China, from 2017 to 2019. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence, University of Chinese Academy of Sciences, and the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interests include optimal control, neural networks-based control, reinforcement learning, adaptive dynamic programming, and their industrial applications.

Fei-Yue Wang (S’87–M’89–SM’94–F’03) received his Ph.D. degree in computer and systems engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990. He joined The University of Arizona in 1990 and became a Professor and the Director of the Robotics and Automation Laboratory and the Program in Advanced Research for Complex Systems. In 1999, he founded the Intelligent Control and Systems Engineering Center at the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, and, in 2002, was appointed as the Director of the Key Laboratory of Complex Systems and Intelligence Science, CAS. In 2011, he became the State Specially Appointed Expert and the Director of the State Key Laboratory for Management and Control of Complex Systems.

His current research focuses on methods and applications for parallel intelligence, social computing, and knowledge automation. He is a Fellow of INCOSE, IFAC, ASME, and AAAS. In 2007, he received the National Prize in Natural Sciences of China and became an Outstanding Scientist of ACM for his work in intelligent control and social computing. He received the IEEE ITS Outstanding Application and Research Awards in 2009 and 2011, respectively. In 2014, he received the IEEE Norbert Wiener Award. Since 1997, he has been serving as the General or Program Chair of over 30 IEEE, INFORMS, IFAC, ACM, and ASME conferences. He was the President of the IEEE ITS Society from 2005 to 2007, the Chinese Association for Science and Technology, USA, in 2005, the American Zhu Kezhen Education Foundation from 2007 to 2008, the Vice President of the ACM China Council from 2010 to 2011, the Vice President and the Secretary General of the Chinese Association of Automation from 2008–2018. He was the Founding Editor-in-Chief (EiC) of the International Journal of Intelligent Control and Systems from 1995 to 2000, the IEEE ITS Magazine from 2006 to 2007, the IEEE/CAA Journal of Automatica Sinica from 2014–2017, and the China’s Journal of Command and Control from 2015–2020. He was the EiC of the IEEE Intelligent Systems from 2009 to 2012, the IEEE Transactions on Intelligent Transportation Systems from 2009 to 2016, and the EiC of the IEEE Transactions on Computational Social Systems since 2017, and the Founding EiC of China’s Journal of Intelligent Science and Technology since 2019. Currently, he is the President of CAA’s Supervision Council, IEEE Council on RFID, and Vice President of IEEE Systems, Man, and Cybernetics Society.