

## RESEARCH ARTICLE

# Disturbance Compensation-Based Optimal Tracking Control for Perturbed Nonlinear Systems

YANPING GAO<sup>1</sup> AND ZUOJUN LIU<sup>1</sup>

School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China

Corresponding author: Yanping Gao (zbkz\_gao@163.com)

**ABSTRACT** This paper investigates the disturbance compensation-based optimal tracking control for nonlinear systems in the presence of uncertain dynamic drift and extraneous disturbance by using the adaptive dynamic programming (ADP). First, an extended state observer (ESO)-based disturbance rejection controller is designed to estimate the comprehensive disturbances of system. Then, a novel composite controller capable of online learning is developed based on disturbance rejection controller and optimal regulation law, where the optimal regulation law is conducted by ADP framework to stabilize the dynamics of tracking error and minimize predefined value function. Particularly, an improved critic-only weight updating algorithm is inserted in ADP for ensuring the finite time convergence of critic weight without resorting to traditional actor-critic structures enduring remarkable computational burden. Based on Lyapunov analysis, it is proved that the tracking errors and weight estimation errors of critic network are uniformly ultimately bounded and the pursued controller approximates to the optimal policy. Finally, simulation results are shown to check the superiority of involved strategy, and the value function can be decreased by 25% with consistent tracking performance.

**INDEX TERMS** Disturbance compensation, extended state observer, perturbed nonlinear system, critic-only, adaptive dynamic programming.

## I. INTRODUCTION

In recent years, tracking control for nonlinear systems has always been investigated for many real-world plants in practices, such as unmanned aerial vehicles [1], [2], robots [3], and quadrotors [4], [5]. To realize the performance of tracking, massive various control algorithms are employed for tracking control problems, including backstepping control [6], sliding model control [7], [8] and model predictive control [9]. In fact, it is inevitable that the modeling error and external disturbance exist in practical applications. Since classical control is sensitive to the model variation, the tracking performance will be weakened and the system stability cannot be guaranteed. Thus, achieving disturbance accommodation and nominal tracking performance are the chief concerns in current controller designs.

To ensure robust tracking and performance recovery, a large number of control methods or tools have emerged to

deal with uncertainties, such as sliding mode control [10], [11],  $H_\infty$  controller [12], [13], neural network (NN) [14], [15], fuzzy logic system (FLS) [16], [17], and so on. In [11], a finite time sliding mode controller is proposed for robotic manipulators, where a better performance and robust tracking control can be derived and the chattering phenomenon is eliminated, but it requires the specific bound of uncertainties. In [12], a robust  $H_\infty$  tracking controller is considered for nonlinear multi-UAV networked system to fulfill the stability via transforming network systems into leader-follower structure. The controller typically consumes unnecessary control actions due to the worst-case consideration. In [15], the NN-based backstepping controller is established based on distance for nonlinear multiagent systems to ensure formation and tracking control, where the unknown nonlinearity of the system dynamics is recovered by NN. In addition, FLS is employed in the backstepping-based control for stochastic nonlinear systems to estimate the unknown dynamic drift while reduced-order state observer is applied for reconstructing immeasurable states, such that

The associate editor coordinating the review of this manuscript and approving it for publication was Emre Koyuncu<sup>1</sup>.

the tracking errors belong to a compact set in the mean square sense [17]. However, a heavy computation burden is usually accompanied with FLS due to the fact that heuristic parameter adjustments are closely depending on trial and error, which is hard to apply in applications. To achieve efficient closed-loop tracking, active disturbance rejection control (ADRC) proposed by Han [18] is a widespread and easy implementation solution to handle the uncertainties and disturbances. The key module of ADRC is the so-called extended state observer (ESO), which could approximate and compensate for the negative influences of total uncertainty in a timely fashion, and has been successfully validated both from theoretical and industrial advancements. In general, the compensated system is then converted into a chain of integrators with nominal controller and it is a common selection to devise a proportional derivative controller to achieve reference tracking in the existing paradigms [19], [20], [21], [22], [23], [24], [25]. Although the reported ADRC-based outcomes can achieve nominal tracking versus a wide class of uncertainties, it lacks the online optimization ability in terms of tracking performance and control consumption due to the time-invariant control structure and few works explore optimizing the overall control performance to retain the optimality, which results in investigating the controllers with improved performance and low control consumption.

Optimal tracking control can be obtained via the dynamic programming by minimizing a predefined quadratic function consisting of control input and tracking errors subject to nonlinear systems [26]. Noting that corresponding Hamilton-Jacobi-Bellman (HJB) equation for nonlinear systems is in form of nonlinear partial derivative equations, thus it is not easy to directly obtain analytical solutions and may cause curse of dimensionality compared to linear systems. Recently, reinforcement learning (RL) provides an efficient way for approximation of optimal value function and optimal control by interacting with the environment via pursuing the maximum reward function. Adaptive dynamic programming (ADP), which is a representative branch of RL in control community, is demonstrated that the optimal control can be derived by replacing the analytic solution of HJB equation with NN approximators. In particular, actor-critic networks are prevailing structures utilized in ADP, wherein the critic NN is constructed for approximation of value function, while the actor NN is used for approximating the optimal policy. For instance, under the condition that system model is known, the policy iteration in [27] updates critic and actor NN synchronously, and the system state and NN errors are proved to be uniformly ultimately bounded (UUB). However, the iteration algorithms are only effective under the assumption that model is precisely known. Thus, many scholars have shed lights on ADP methods for nonlinear uncertain systems. In [28], the identifier-critic-actor NN is proposed using identifier NN to reconstruct the uncertainty in model. To avoid overhigh learning time for suppressing modeling errors, the identifier in [29] is designed by functional-link NN

for permanent-magnet synchronous motor. However, the use of triple neural structures may lead to heavy computation and slow convergence rate. To solve this difficulty, a critic NN is used to approximate the optimal value function and optimal control input whilst the unknown dynamics are reconstructed by identifier, avoiding complexity of introducing the actor NN [30]. Besides, the sliding mode control is employed in the ADP framework for uncertain nonlinear systems, which is insensitive to external disturbance. In [31], the ADP with actor-critic structure is employed to online learning the optimal control of the sliding-mode dynamics, where lumped disturbances consisting of unknown term and input disturbance are estimated by NN and disturbance observer. In [32], a modified value function with control inputs and disturbance from neighbor nodes is considered for multiagent systems to derive optimal control in critic-actor framework such that the consensus errors are asymptotically stable, where a distributed event-triggered integral sliding mode controller is constructed to eliminate the matched disturbance. Furtherly, actor-critic ADP is conducted by incorporating a novel fast nonsingular sliding mode into controller design and input constraints are handled via a nonlinear anti-windup compensator in [33], leading to convergence of tracking error in fixed-time. Besides, a novel ADP in [34] is capable of deriving optimal strategy under the framework of zero-sum game. However, the upper bound is required to be known before the controller design, which is hard to guarantee in real scenarios. Nevertheless, the above mentioned controllers need to introduce actor NN for weight convergence to induce optimal control. Note that ESO structure does not depend on the upper bound of lumped disturbances. As far as we know, there is no work considering ESO-based optimal tracking control within critic-only ADP scheme and it is also challenging originating from single NN to address the high coupling of disturbance observation errors by ESO, NN weight updating deviation as well as reconstruction errors.

Inspired by foregoing observations, there exist following shortcomings in above developed controller schemes: The traditional control design is carried without consideration of control consumption, while the optimal controllers are devised based on actor-critic framework for weight convergence. This article focuses on disturbance compensation-based optimal tracking control for nonlinear systems in the presence of uncertain dynamic drift and extraneous disturbance with critic-only ADP. The main contributions of this work are summarized as follows:

- 1) Unlike [10], [11], [12], [13], [14], [15], [16], [17] that only focus on disturbance rejection without caring about control consumptions, herein an efficient compromise between tracking performance and control costs is handled under the devised ADP framework, achieving performance optimality and robustness simultaneously. In addition, the disturbance rejection controller is imported to deal with nonzero reference trajectory such that optimal tracking problem is converted to optimal stabilization problem.

2) Contrasting to [28], [29], and [31] using multiple NNs to design optimal policy for perturbed nonlinear systems, the critic-only ADP with a lower computational burden is established. Additionally, unlike nonlinear disturbance observer [31] and sliding mode controller [32] with entailed disturbance upper bound information, the proposed optimal anti-disturbance control based on uncertainty estimates offered by ESO can be derived without resorting to the prior information, which is not conservative and computationally convenient compared with game scheme in [34].

3) Different from [28], [29], [31], [32], [33], [35], and [36], where Bellman error is minimized via prevailing gradient descent or least squares methods, by extracting weight errors from designated intermittent arguments, an improved neural weight updating rule by inclusion of weight errors is designed ingeniously to achieve finite time convergence. On account of weight rule, the tracking errors can converge to the neighborhood of origin under critic-only ADP framework with synthesized consideration of disturbance observation error by ESO, NN evaluated error and tracking error. In addition, the persistent excitation (PE) condition can be easily checked by judging the minimum eigenvalue of constructed matrix, which provides an online feasible examining avenue compared to [28] and [29].

The rest of paper is organized as follows. The problem formulation is shown in Sect. II. The disturbance rejection controller is described in Sect. III, while composite controller is designed based on estimation of comprehensive disturbances and optimal regulation law. The simulations are carried out for quadrotor tracking to verify the effective of investigated method in Sect. IV. Finally, we give the conclusion in Sect. V.

Notation: The bold characters are used to express vectors or matrices.  $\mathbf{0}$  and  $\mathbf{I}$  denote zero and the identity matrices with dimension determined by the context. The  $\text{diag}(\cdot)$  indicates a diagonal matrix.  $\lambda_{\max}(\mathbf{A})$  and  $\lambda_{\min}(\mathbf{A})$  represent the maximum and minimize eigenvalues of matrix  $\mathbf{A}$ , respectively.

## II. PROBLEM FORMULATION

Consider the nonlinear affine multi-input multi-output (MIMO) system,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u} + \mathbf{d}(t) \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{u} \in \mathbb{R}^m$  are the system state and control input,  $\mathbf{d}(t) \in \mathbb{R}^n$  denotes unknown time-varying disturbance.  $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^{n \times n}$  is unknown continuously differentiable function and bounded on a compact set  $\Omega$ , and control gain matrix  $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^{n \times m}$  is known, bounded and reversible. The tracking error  $\mathbf{e}$  can be expressed as

$$\mathbf{e} = \mathbf{x} - \mathbf{x}_d \quad (2)$$

where  $\mathbf{x}_d$  is bounded and differential trajectory to be specified. From [27], the system can be stabilizable by designing a continuous control input  $\mathbf{u}$  to track desired

trajectory  $\mathbf{x}_d$ . Following (1), the error dynamics can be computed as

$$\dot{\mathbf{e}} = \dot{\mathbf{x}} - \dot{\mathbf{x}}_d = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u} + \mathbf{d}(t) - \dot{\mathbf{x}}_d \quad (3)$$

The control objective is to design the controller  $\mathbf{u}$  such that  $\mathbf{x}$  can track the desired trajectory  $\mathbf{x}_d$  in an optimal way by minimizing following function consisting of tracking errors and control inputs:

$$V(\mathbf{e}) = \int_t^\infty \mathbf{e}^T \mathbf{Q} \mathbf{e} + \mathbf{u}^T \mathbf{R} \mathbf{u} d\tau \quad (4)$$

where  $\mathbf{Q} \in \mathbb{R}^{n \times n}$  and  $\mathbf{R} \in \mathbb{R}^{m \times m}$  are positive definite matrices. Noting that the unknown term and disturbance exist in system (3), it can be regarded as comprehensive disturbances in disturbance rejection controller.

*Assumption 1:* The comprehensive disturbances  $\delta = \mathbf{f}(\mathbf{x}) + \mathbf{d}(t)$  and its time derivative  $\dot{\delta}$  can be bounded with an unknown positive constant.

*Remark 1:* The model (1) considered in this paper can cover a wide range of plants in applications, such as unmanned aerial vehicles [2], robotic manipulators [11], MEMS gyroscopes [22] and quadrotors [37]. For high-order systems, one can use iterative backstepping design to convert high-order dynamics into first-order systems in form of (1). Thus, the proposed controller can be feasible to stabilize high-order tracking error system. In addition, similar to numerous preceding ESO works, the assumption is necessary in the disturbance compensation framework, which can be found in [22], [24], and [37]. And the bound of  $\dot{\delta}$  is only utilized for proving the convergence of disturbance approximation rather than control design in [11] and [31].

Furtherly, following lemma and definition are prepared for weight convergence.

*Lemma 1 [38]:* Assume that there exists a positive function  $V$  satisfying

$$\dot{V}(t) \leq -\ell V^\beta(t) + \eta \quad (5)$$

with  $\ell > 0$ ,  $0 < \beta < 1$  and positive bounded  $\eta$ , then  $V(t)$  can converge to neighborhood of origin before a settling time

$$T \leq \frac{V^{1-\beta}(0)}{\ell(1-\beta)} \quad (6)$$

*Definition 1 [31]:* If there exist positive  $\varepsilon$  and  $\tau$  such that  $\int_t^{t+\tau} \Phi \Phi^T d\tau > \varepsilon \mathbf{I}$  holds, variable  $\Phi$  satisfies PE condition.

## III. MAIN RESULTS

In this section, a novel composite controller capable of online learning is developed based on disturbance rejection controller and optimal regulation law. First, in light of the uncertainties estimation result of ESO, a disturbance rejection controller is constructed. Then, to simultaneously accomplish a minimum tracking error and a lowest control cost, an optimal regulation law is established by devising a critic-only ADP technique with novel adaptive law such that

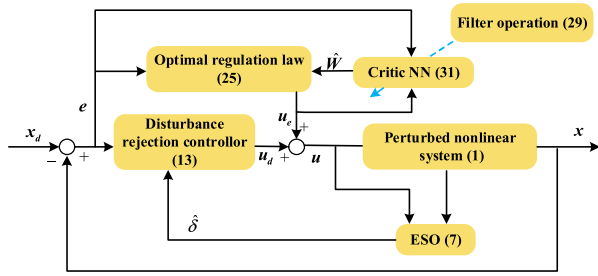


FIGURE 1. ESO-based optimal tracking controller architecture.

the preset quadratic value function can be minimized. Note that the whole control architecture of the proposed control is summarized in Figure 1.

**A. DISTURBANCE REJECTION CONTROLLER DESIGN**

According to kernel of disturbance compensation and with reference to [24] and [37], by treating comprehensive disturbances  $\delta$  as the extended state, an ESO is designed to estimate disturbance of system (1) as follows:

$$\begin{cases} \dot{\hat{x}} = \hat{\delta} + \mathbf{g}(\mathbf{x})\mathbf{u} + \varpi_1 (\mathbf{x} - \hat{\mathbf{x}}) \\ \dot{\hat{\delta}} = \varpi_2 (\mathbf{x} - \hat{\mathbf{x}}) \end{cases} \quad (7)$$

where  $\hat{\mathbf{x}}$  and  $\hat{\delta}$  indicate the estimation of state  $\mathbf{x}$  and disturbance  $\delta$  respectively,  $\varpi_1 = \text{diag}(2w_0, 2w_0, \dots, 2w_0) \in \mathbb{R}^{n \times n}$  and  $\varpi_2 = \text{diag}(w_0^2, w_0^2, \dots, w_0^2) \in \mathbb{R}^{n \times n}$  are positive definite matrices with  $w_0 > 0$ . Compared to NN-based observers [14], [15], [28], [29], it is easy to tune the bandwidth  $w_0$  for accurate estimation of  $\delta$  without adjusting the weights and choosing suitable regressors.

From (1) and (7), the dynamics of estimation errors can be expressed as

$$\begin{cases} \dot{\tilde{\mathbf{x}}} = \tilde{\delta} - \varpi_1 \tilde{\mathbf{x}} \\ \dot{\tilde{\delta}} = -\varpi_2 \tilde{\mathbf{x}} + \mathbf{h} \end{cases} \quad (8)$$

where  $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$  and  $\tilde{\delta} = \delta - \hat{\delta}$  denote the estimation errors of state and disturbance,  $\mathbf{h} = \dot{\delta}$ . Let  $\mathbf{y} = [\tilde{\mathbf{x}}^T, \tilde{\delta}^T]^T$ , it can be rewritten into a compact form:

$$\dot{\mathbf{y}} = \mathbf{H}\mathbf{y} + \Theta\mathbf{h} \quad (9)$$

where

$$\mathbf{H} = \begin{bmatrix} -\varpi_1 & \mathbf{I}_{n \times n} \\ -\varpi_2 & \mathbf{0}_{n \times n} \end{bmatrix} \in \mathbb{R}^{2n \times 2n}, \quad \Theta = \begin{bmatrix} \mathbf{0}_{n \times n} \\ \mathbf{I}_{n \times n} \end{bmatrix} \in \mathbb{R}^{2n \times n} \quad (10)$$

Observing that  $\mathbf{H}$  is Hurwitz, there exists a positive definite matrix  $\mathbf{P}_x \in \mathbb{R}^{2n \times 2n}$  satisfying  $\mathbf{H}^T \mathbf{P}_x + \mathbf{P}_x \mathbf{H} = -\mathbf{I}$ . According to Assumption 1, the convergence of  $\tilde{\delta}$  can be easily inferred from [22] and [24].

Thus, the dynamics of tracking errors (3) can be rewritten into

$$\dot{\mathbf{e}} = \dot{\mathbf{x}} - \dot{\mathbf{x}}_d = \hat{\delta} + \mathbf{g}(\mathbf{x})\mathbf{u} - \dot{\mathbf{x}}_d + \tilde{\delta} \quad (11)$$

In order to derive  $e \rightarrow 0$ , the composite control is divided into two parts,

$$\mathbf{u} = \mathbf{u}_d + \mathbf{u}_e \quad (12)$$

where disturbance rejection controller  $\mathbf{u}_d$  is designed to compensate for the effect of disturbance:

$$\mathbf{u}_d = (\mathbf{g}^T \mathbf{g})^{-1} \mathbf{g}^T [\dot{\mathbf{x}}_d - \hat{\delta} - \mathbf{K}\mathbf{e}] \quad (13)$$

where  $(\mathbf{g}^T \mathbf{g})^{-1} \mathbf{g}^T$  represents the generalized inverse of matrix  $\mathbf{g}$  and positive diagonal gain matrix  $\mathbf{K} \in \mathbb{R}^{n \times n}$  is selected to maintain the tracking error close to zero at the steady-state stage. Furthermore, by substitute (13) into (11), the reconstituted error dynamics is converted to

$$\dot{\mathbf{e}} = -\mathbf{K}\mathbf{e} + \mathbf{g}(\mathbf{x})\mathbf{u}_e + \tilde{\delta} \quad (14)$$

Thus, the optimal regulation law  $\mathbf{u}_e$  will be considered within ADP framework for balancing tracking performance and control expenses.

*Remark 2:* By employing the disturbance rejection controller  $\mathbf{u}_d$  to compensate the disturbance and nonlinear time-varying desired trajectory, the optimal control considered in subsection B is designed for stabilization system, avoiding conducting augmented systems for bounded value functions.

**B. OPTIMAL REGULATION LAW DESIGN**

1) OPTIMAL CONTROL

To stabilize system (14), the value function is predefined for balancing control cost and tracking performance. Thus, the value function (4) can be reconducted related to  $\mathbf{u}_e$  as

$$V(\mathbf{e}) = \int_t^\infty \mathbf{e}^T \mathbf{Q}\mathbf{e} + \mathbf{u}_e^T \mathbf{R}\mathbf{u}_e d\tau \quad (15)$$

For deriving the optimal control policy, the Hamiltonian is denoted as

$$H(\mathbf{e}, \mathbf{u}_e, V) = \mathbf{V}_e^T [-\mathbf{K}\mathbf{e} + \mathbf{g}(\mathbf{x})\mathbf{u}_e + \tilde{\delta}] + \mathbf{e}^T \mathbf{Q}\mathbf{e} + \mathbf{u}_e^T \mathbf{R}\mathbf{u}_e \quad (16)$$

where  $\mathbf{V}_e = \partial V(\mathbf{e})/\partial \mathbf{e}$ . By taking derivative of (15) along (14) and following the optimality principle [39], the HJB equation can be derived as

$$\begin{aligned} H(\mathbf{e}, \mathbf{u}_e^*, \mathbf{V}_e^*) \\ = \mathbf{V}_e^{*T} [-\mathbf{K}\mathbf{e} + \mathbf{g}(\mathbf{x})\mathbf{u}_e^* + \tilde{\delta}] + \mathbf{e}^T \mathbf{Q}\mathbf{e} + \mathbf{u}_e^{*T} \mathbf{R}\mathbf{u}_e^* = 0 \end{aligned} \quad (17)$$

with  $\mathbf{V}_e^* = \partial V^*(\mathbf{e})/\partial \mathbf{e}$  and the optimal value function  $V^*(\mathbf{e})$  is given as

$$V^*(\mathbf{e}) = \min_{\mathbf{u}_e \in \Psi(\Omega)} \int_t^\infty \mathbf{e}^T \mathbf{Q}\mathbf{e} + \mathbf{u}_e^T \mathbf{R}\mathbf{u}_e d\tau \quad (18)$$

The optimal control  $\mathbf{u}_e^*$  can be obtained by utilizing the stationarity condition  $\partial H(\mathbf{e}, \mathbf{u}_e^*, \mathbf{V}_e^*)/\partial \mathbf{u}_e^* = 0$  as

$$\mathbf{u}_e^* = -\frac{1}{2} \mathbf{R}^{-1} \mathbf{g}^T \mathbf{V}_e^* \quad (19)$$



By virtue of (19), HJB equation (17) can be represented as

$$V_e^{*T}[-\mathbf{K}e + \tilde{\delta}] + e^T \mathbf{Q}e - \frac{1}{4} V_e^{*T} \mathbf{g} \mathbf{R} \mathbf{g}^T V_e^* = 0 \quad (20)$$

Note that (20) is a nonlinear partial differential equation relevant to optimal value function, it is difficult to obtain its analytical solution due to its complex dynamics and nonlinearity. Here, the RL strategy is considered for approximating  $V^*(e)$  by introducing NN. With the aid of critic NN, the optimal value function  $V^*(e)$  and its gradient can be obtained as

$$V^*(e) = \mathbf{W}^T \phi(e) + \varepsilon \quad (21)$$

And

$$\frac{\partial V^*(e)}{\partial e} = \nabla \phi(e)^T \mathbf{W} + \nabla \varepsilon \quad (22)$$

where  $l$  represents the number of neurons in hidden layer,  $\mathbf{W} \in \mathbb{R}^l$  is the ideal weight of critic NN,  $\phi \in \mathbb{R}^l$  is the regressor vector, and  $\varepsilon$  is the error caused by critic NN, which can involve to zero with sufficiently large  $l$  [27]. And  $\nabla \phi = \frac{\partial \phi}{\partial e} \in \mathbb{R}^{l \times n}$  and  $\nabla \varepsilon = \frac{\partial \varepsilon}{\partial e} \in \mathbb{R}^n$  denote the partial derivatives of  $\phi$  and  $\varepsilon$ , satisfying  $\|\nabla \phi\| \leq \bar{\phi}$  and  $\|\nabla \varepsilon\| \leq \bar{\varepsilon}$  with positive constants  $\bar{\phi}$  and  $\bar{\varepsilon}$ . Thus, the optimal control  $u_e^*$  can be given as

$$u_e^* = -\frac{1}{2} \mathbf{R}^{-1} \mathbf{g}^T (\nabla \phi(e)^T \mathbf{W} + \nabla \varepsilon) \quad (23)$$

Taking the unavailability of  $\mathbf{W}$ , the optimal value  $V^*$  can be estimated by following form

$$V(e) = \hat{\mathbf{W}}^T \phi(e) \quad (24)$$

where  $\hat{\mathbf{W}} \in \mathbb{R}^l$  is estimation of  $\mathbf{W}$  and its online update will be given later to realize optimal control and optimal value function simultaneously. According to (23) and (24), the related optimal control can be approximated by

$$u_e = -\frac{1}{2} \mathbf{R}^{-1} \mathbf{g}^T \nabla \phi(e)^T \hat{\mathbf{W}} \quad (25)$$

*Remark 3:* In the most ADP schemes, the Hamiltonian function is considered for the nominal system in [27], [28] and [29]. In particular, the HJB equation is structured related to nominal system to solve optimal control. In contrary to these methods, the HJB equation (17) is conducted related to (11) consisting of the estimation error  $\tilde{\delta}$ , which leads to challenges for adjusting weight and deducing optimal regulation law. Differing from the existing ADP control design for nominal system with actor-critic NN framework, where utilizes critic NN and actor NN for approximation of optimal value function and optimal policy [28], [29], optimal value function and optimal control policy can be updated simultaneously by the critic-only NN with adaptive weight update, reducing computational consumption in actor NN significantly.

## 2) IMPROVED UPDATING RULE FOR CRITIC NN

In this subsection, the update of critic NN weight  $\hat{\mathbf{W}}$  is considered to make sure that optimal regulation law  $u_e$  not only can stabilize the system (14) but also enable an optimal property. Based on NN tools, the HJB equation (17) can be restated from (22) as

$$\mathbf{W}^T \nabla \phi(-\mathbf{K}e + \mathbf{g}u_e) + e^T \mathbf{Q}e + u_e^T \mathbf{R}u_e + \varepsilon_{HJB} = 0 \quad (26)$$

where  $\varepsilon_{HJB} = \nabla \varepsilon^T(-\mathbf{K}e + \mathbf{g}u_e + \tilde{\delta}) + (\mathbf{W}^T \nabla \phi) \tilde{\delta}$  is residual error of HJB equation.

For the sake of description, by denoting  $\psi = e^T \mathbf{Q}e + u_e^T \mathbf{R}u_e$  and  $\Phi = \nabla \phi(-\mathbf{K}e + \mathbf{g}u_e)$ , HJB equation (26) can be expressed as

$$\psi = -\Phi^T \mathbf{W} - \varepsilon_{HJB} \quad (27)$$

Instead of minimizing quadratic function defined by the Bellman error, we design the improved method to update the weight  $\hat{\mathbf{W}}$  with finite time stability. To describe the update rule, we multiply both sides of (27) by  $\Phi$ , resulting in

$$\Phi \psi = -\Phi \Phi^T \mathbf{W} - \Phi \varepsilon_{HJB} \quad (28)$$

Thus, for a constant  $c > 0$ , the auxiliary matrix  $\mathbf{N} \in \mathbb{R}^{l \times l}$  and vector  $\mathbf{S} \in \mathbb{R}^l$  are introduced to perform filter operation, which are defined as

$$\begin{cases} \dot{\mathbf{N}} = -c\mathbf{N} + \Phi \Phi^T, \mathbf{N}(0) = 0 \\ \dot{\mathbf{S}} = -c\mathbf{S} + \Phi \psi, \mathbf{S}(0) = 0 \end{cases} \quad (29)$$

where positive constant  $c$  is chosen to ensure the boundedness of  $\mathbf{S}$  and  $\mathbf{N}$ . From (27), there exists the auxiliary variable  $\mathbf{P} \in \mathbb{R}^l$ , which satisfies

$$\mathbf{P} = \mathbf{N} \hat{\mathbf{W}} + \mathbf{S} \quad (30)$$

Thus, the improved update rule for weight is

$$\dot{\hat{\mathbf{W}}} = -\Gamma \frac{\mathbf{N}^T \mathbf{P}}{\|\mathbf{P}\|} \quad (31)$$

where  $\Gamma \in \mathbb{R}^{l \times l}$  is a positive definite diagonal matrix. There may emerge singularity of norm-normalized term, which has been well investigated by a saturation or sign function among sliding mode controllers. Thus, one can employ the similar technique to make sure weight boundedness, where discussion about sign function and norm-normalized term can be found in [40]. With the aid of improved update rule (31), the finite time weight convergence of critic NN can be derived in following theorem.

*Theorem 1:* Consider the value function (15) for nonlinear system (14), if the estimation of value function is updated with approximate weight  $\hat{\mathbf{W}}$  given by (31). Then, the approximation error of weight, i.e.,  $\tilde{\mathbf{W}} = \mathbf{W} - \hat{\mathbf{W}}$ , can converge to the neighborhood of origin in a finite time.

*Proof:* By imposing integral operation on (29), one has

$$\begin{cases} N(t) = \int_0^t e^{-c(t-r)} \Phi \Phi^T dr, & N(0) = 0 \\ S(t) = \int_0^t e^{-c(t-r)} \Phi \psi dr, & S(0) = 0 \end{cases} \quad (32)$$

The auxiliary variable  $\varsigma$  related to HJB error is introduced as  $\varsigma = -\int_0^t e^{-c(t-r)} \Phi \varepsilon_{HJB} dr$ , resulting in  $S = -N\tilde{W} + \varsigma$  from (28). Further, due to  $\tilde{W} = W - \hat{W}$ , we can infer that estimation error of  $W$  is implicitly contained in matrix  $P$  in following form

$$P = N\hat{W} + S = -N\tilde{W} + \varsigma \quad (33)$$

which is crucial to guarantee weight convergence. Suppose that the PE condition of  $\Phi$  holds, one can infer from [30] that  $\sigma = \lambda_{\min}(N) > 0$  with  $\lambda_{\min}(N)$  representing the minimize eigenvalue of matrix  $N$ . And the candidate Lyapunov function is selected as

$$V_W = \frac{1}{2} \tilde{W}^T \Gamma^{-1} \tilde{W} \quad (34)$$

By substituing (31) and (33) into the time derivative of  $V_W$ , one can yield that

$$\begin{aligned} \dot{V}_W &= \tilde{W}^T \Gamma^{-1} \dot{\tilde{W}} = \tilde{W}^T \frac{N^T P}{\|P\|} \\ &= \tilde{W}^T N^T \frac{N\hat{W} + S}{\|N\hat{W} + S\|} = \tilde{W}^T N^T \frac{-N\tilde{W} + \varsigma}{\|-N\tilde{W} + \varsigma\|} \\ &= \frac{(\tilde{W}^T N^T - \varsigma^T) (-N\tilde{W} + \varsigma) + \varsigma^T (-N\tilde{W} + \varsigma)}{\|-N\tilde{W} + \varsigma\|} \\ &= -\|-N\tilde{W} + \varsigma\| + \frac{\varsigma^T (-N\tilde{W} + \varsigma)}{\|-N\tilde{W} + \varsigma\|} \\ &\leq -\|N\tilde{W}\| + 2\|\varsigma\| \\ &\leq -\sigma \|\tilde{W}\| + 2\|\varsigma\| \leq -\eta V_W^{\frac{1}{2}} + 2\varepsilon_\varsigma \end{aligned} \quad (35)$$

with a postive constant  $\varepsilon_\varsigma$  being assured to satisfy  $\|\varsigma\| \leq \varepsilon_\varsigma$  and  $\eta = \sigma \sqrt{2/\lambda_{\max}(\Gamma^{-1})}$  with  $\lambda_{\max}(\Gamma^{-1})$  representing the maximum eigenvalue of matrix  $\Gamma^{-1}$ . Thus, one can conclude that the weight estimation errors  $\tilde{W}$  converge to the neighborhood of origin after the finite time  $T \leq 2V_W^{\frac{1}{2}}(0)/\eta$  by recalling Lemma 1.

*Remark 4:* In the most ADP framework, the Bellman error is minimized by gradient descent or least squares methods [28], [29]. Unlike these methods, weight errors can be extracted from designated intermittent arguments via filter operation such that proposed adaptive rule (31), which is driven by weight errors, can achieve finite-time weight convergence, avoiding the hysteresis phenomenon of weight convergence by minimizing Bellman error. Furthermore, approximate optimal regulation law can be derived with estimation of weight, resulting in zero coupling errors

with disturbance observation error NN evaluated error and tracking error. In addition, the PE condition is the basic assumption to guarantee the convergence of the NN weight among ADP frameworks [28], [29], which is not easy to verify in the simulation. By introducing the auxiliary matrix  $N$ , PE condition can be judged by computing whether  $\sigma > 0$ .

The proposed controller with critic-only ADP can be operated in Algorithm 1 with specific steps, where consists of the key equations.

**Algorithm 1** The ESO-Based Optimal Tracking Controller

- 1 Initialize  $e(0), u(0), N = S = 0$ ;
- 2 Apply (7) to compute  $u_d$  (13);
- 3 Apply (29) to update the weight with (31);
- 4 Compute the optimal controller (13) and (25);
- 5  $t = t + \Delta t$ , step to 2.

3) STABILITY ANALYSIS

*Theorem 2:* For the nonlinear system (1), consider composite control (12) consisting of disturbance rejection controller (13) and optimal regulation law (25), where ESO (7) and improved adaptive rule (31) are provided for updating  $\hat{\delta}$  and  $\tilde{W}$ . If  $\Phi$  satisfies PE condition, then the tracking errors can be proved within the neighborhood of origin. And  $u_e$  will converge to a neighborhood of optimal regulation law  $u_e^*$ .

*Proof:* It follows from (14) and (25) that

$$\begin{aligned} \dot{e} &= -Ke + gu_e + \tilde{\delta} \\ &= -Ke - \frac{1}{2}gR^{-1}g^T \nabla \phi^T \hat{W} + \frac{1}{2}gR^{-1}g^T (\nabla \phi^T W + \nabla \varepsilon) \\ &\quad + gu_e^* + \tilde{\delta} \\ &= -Ke + \frac{1}{2}gR^{-1}g^T \nabla \phi^T \tilde{W} + \frac{1}{2}gR^{-1}g^T \nabla \varepsilon + gu_e^* + \tilde{\delta} \end{aligned} \quad (36)$$

Considering the following Lyapunov function:

$$V = \Gamma' e^T e + K' V^* \quad (37)$$

where  $V^*$  is optimal value function given by (18), constants  $K' > 0$  and  $\Gamma' > 0$ . According to (18) and (36), it follows from Young's inequality  $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$  that

$$\begin{aligned} \dot{V} &= 2\Gamma' e^T \dot{e} + K' \dot{V}^* \\ &= 2\Gamma' e^T \dot{e} + K' (-e^T Qe - u_e^{*T} R u_e^*) \\ &= 2\Gamma' e^T [-Ke + \frac{1}{2}gR^{-1}g^T \nabla \phi^T \tilde{W} + \frac{1}{2}gR^{-1}g^T \nabla \varepsilon \\ &\quad + gu_e^* + \tilde{\delta}] + K' (-e^T Qe - u_e^{*T} R u_e^*) \\ &\leq -[2\lambda_{\min}(K)\Gamma' + K' \lambda_{\min}(Q) - (\|gR^{-1}g^T \nabla \phi^T\| \\ &\quad + \|gR^{-1}g^T\| + 2)\Gamma'] \|e\|^2 - (K' \lambda_{\min}(R) \\ &\quad - \Gamma' \|g\|^2) \|u_e^*\|^2 + \frac{1}{4}\Gamma' \|gR^{-1}g^T\| \|\nabla \varepsilon\|^2 + \Gamma' \|\tilde{\delta}\|^2 \\ &\quad + \frac{1}{4}\Gamma' \|gR^{-1}g^T \nabla \phi^T\| \|\tilde{W}\|^2 \end{aligned} \quad (38)$$

Under the condition that  $K'\lambda_{\min}(\mathbf{R}) - \Gamma'\|\mathbf{g}\|^2 \geq 0$ , (38) can be rewritten as

$$\dot{V} \leq -v\|\mathbf{e}\|^2 + \gamma' \quad (39)$$

with positive constants  $v$  and  $\gamma'$

$$\begin{aligned} v &= 2\lambda_{\min}(\mathbf{K})\Gamma' + \mathbf{K}'\lambda_{\min}(\mathbf{Q}) - (\|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\nabla\phi^T\| \\ &\quad + \|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\| + 2)\Gamma' \\ \gamma' &= \frac{1}{4}\Gamma'\|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\nabla\phi^T\|\|\tilde{\mathbf{W}}\|^2 + \frac{1}{4}\Gamma'\|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\|\|\nabla\varepsilon\|^2 \\ &\quad + \Gamma'\|\tilde{\delta}\|^2 \end{aligned}$$

Noting that  $\gamma'$  contains weight estimation errors  $\tilde{\mathbf{W}}$ , ESO estimation errors  $\tilde{\delta}$  and NN approximation errors  $\nabla\varepsilon$  from critic NN, which implies that constant  $\gamma'$  is bounded from Theorem 1 and the boundedness of  $\nabla\varepsilon, \tilde{\delta}$ . In addition, the parameters  $\Gamma'$  and  $\mathbf{K}'$  are chosen as follows, shown in the equation at the bottom of the page, which guarantees the positiveness of  $v$  to induce the stability of (38).

Due to the fact that the disturbance estimation errors and critic NN approximation errors are upper limited by finite values, i.e.,  $\gamma' > 0$ . For any

$$\|\mathbf{e}\| > \sqrt{\frac{\gamma'}{v}} \quad (40)$$

We can derive  $\dot{V} < 0$  from (39). From Lyapunov Theorem, tracking errors are UUB. Thus, one can conclude that the tracking errors converge to region  $\|\mathbf{e}\| \leq \sqrt{\gamma'/v}$ . In addition, according to the definition of  $\mathbf{u}_e^*$  in (23) and  $\mathbf{u}_e$  in (25), the boundness of  $\|\mathbf{u}_e - \mathbf{u}_e^*\|$  can be conducted by

$$\begin{aligned} \lim_{t \rightarrow \infty} \|\mathbf{u}_e - \mathbf{u}_e^*\| &= \lim_{t \rightarrow \infty} \left\| \frac{1}{2}\mathbf{R}^{-1}\mathbf{g}^T\nabla\phi^T\tilde{\mathbf{W}} + \frac{1}{2}\mathbf{R}^{-1}\mathbf{g}^T\nabla\varepsilon \right\| \\ &\leq \frac{1}{2}\|\mathbf{R}^{-1}\mathbf{g}^T\|(\|\tilde{\phi}\|\|\tilde{\mathbf{W}}\| + \bar{\varepsilon}) \\ &\triangleq \varepsilon_u \end{aligned} \quad (41)$$

The proof is completed.

#### IV. SIMULATION RESULTS

To verify the effectiveness and advantages of proposed disturbance compensation-based optimal tracking control, simulations are conducted on a class of typical perturbed nonlinear system. The quadrotors, which is one of indispensable aircrafts widely applied in many fields [24], [37], has been investigated by many scholars, due to its nonlinearity,

TABLE 1. Model parameters of the quadrotor and disturbance.

Variable	Value
$m$	2 kg
$g$	9.8 m/s <sup>2</sup>
$\Pi$	diag(0.01, 0.01, 0.01) Nms <sup>2</sup>
$\mathbf{d}$	$[2(\sin(t) + \sin(0.5t) - \cos(0.8t)), 2(\cos(t) + \sin(0.5t) - \cos(0.8t)), 2\sin(0.5t)]^T$

unknown dynamic model, and wind disturbance. As modeled in [37], the position loop of quadrotors is described as

$$m\ddot{\mathbf{p}} = -\Pi\dot{\mathbf{p}} + \mathbf{H}\mathbf{u}_p - \mathbf{G} + \mathbf{d} \quad (42)$$

where  $\mathbf{p} = [p_x, p_y, p_z]^T$  denotes the position of quadrotor,  $\mathbf{v} = \dot{\mathbf{p}} = [v_x, v_y, v_z]^T$  represents the velocity,  $m$  is mass of quadrotor and  $\mathbf{G} = [0, 0, mg]^T$  with  $g$  standing for the gravity acceleration. The coupling matrix  $\mathbf{H} = [\cos\varphi\cos\psi\sin\theta + \sin\varphi\sin\psi, \cos\varphi\sin\psi\sin\theta - \sin\varphi\cos\psi, \cos\varphi\cos\theta]^T$  reveals the relationship of translation and rotational motion with  $\varphi, \theta, \psi$  being roll, pitch and yaw angles.  $\Pi$  and  $\mathbf{d}$  stand for unknown positive definite aerodynamic matrix and bounded disturbance. With an effort to boost controller design, we reformulate (42) into a compact form to adopt proposed composited controller

$$\dot{\mathbf{X}} = \mathbf{f}(\mathbf{X}) + \mathbf{g}(\mathbf{X})\mathbf{u} + \mathbf{D}(t) \quad (43)$$

where

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{p} \\ \mathbf{v} \end{bmatrix} \in \mathbb{R}^{6 \times 1}, \quad \mathbf{f}(\mathbf{X}) = \begin{bmatrix} \mathbf{v} \\ -\Pi\mathbf{v}/m \end{bmatrix} \in \mathbb{R}^{6 \times 1}, \\ \mathbf{g}(\mathbf{X}) &= \begin{bmatrix} 0 \\ \mathbf{I}_3 \end{bmatrix} \in \mathbb{R}^{6 \times 3}, \quad \mathbf{u} = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \in \mathbb{R}^{3 \times 1}, \\ \mathbf{D}(t) &= \begin{bmatrix} 0 \\ \mathbf{d}/m \end{bmatrix} \in \mathbb{R}^{6 \times 1} \end{aligned}$$

with  $u_x = [u_p(\cos\varphi\cos\psi\sin\theta + \sin\varphi\sin\psi)]/m$ ,  $u_y = [u_p(\cos\varphi\sin\psi\sin\theta - \sin\varphi\cos\psi)]/m$  and  $u_z = [u_p(\cos\varphi\cos\theta - mg)]/m$ . Since the  $\mathbf{f}(\mathbf{X})$  and  $\mathbf{D}(t)$  are unknown, we consider them as comprehensive disturbances  $\delta = \mathbf{f}(\mathbf{X}) + \mathbf{D}(t)$ . The model parameter of the quadrotor and disturbance are shown in Table 1.

The trajectory reference command is

$$\mathbf{p}_d = [10(1 - \cos(\pi t/10)), 5\sin(\pi t/5), 9]^T$$

$$\begin{aligned} K' &> \max\left(\frac{(\|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\nabla\phi^T\| + \|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\| + 2)\Gamma' - 2\lambda_{\min}(\mathbf{K})\Gamma'}{\lambda_{\min}(\mathbf{Q})}, \frac{\|\mathbf{g}\|^2}{\lambda_{\min}(\mathbf{R})}\right) \\ \lambda_{\min}(\mathbf{K}) &> \frac{\|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\nabla\phi^T\| + \|\mathbf{g}\mathbf{R}^{-1}\mathbf{g}^T\| + 2}{2} \end{aligned}$$

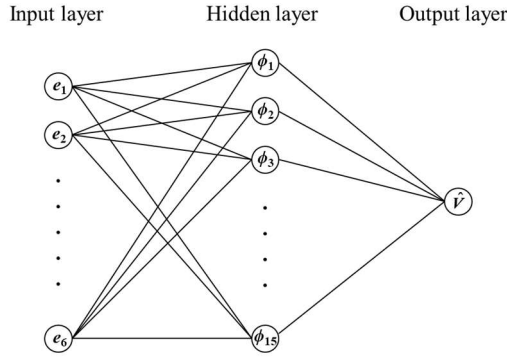


FIGURE 2. The structure of critic NN.

The reference velocity is

$$v_d = [\pi \sin(\pi t/10), \pi \cos(\pi t/5), 0]^T$$

with the initial position and velocity of the quadrotor being  $p(0) = [-5, 3, 2]^T$  and  $v(0) = [0, 0, 0]^T$ .

Based on the aforementioned control design, proposed control is exploited based on disturbance rejection controller and optimal regulation law, where a disturbance rejection controller is designed based on ESO to compensate the impact of comprehensive disturbances. The optimal regulation law is derived by minimizing a value function in ADP framework equipped with adaptive law (31) for weight update. To approximate the optimal value function by critic NN, the 6-15-1 structure is established as following: six input neurons, fifteen hidden neurons, and one output neuron, as shown in Figure 2.

In specific, we select Gaussian basis functions  $\phi = [\phi_1, \phi_2, \dots, \phi_{15}]^T$  as regressors in following form

$$\phi_i = \exp\left[-\frac{\|e - c_i\|^2}{b_i^2}\right], \quad i = 1, 2, \dots, 15 \quad (44)$$

where  $e = X - X_d = [p^T - p_d^T, v^T - v_d^T]^T$ , the centers  $c_i = [c_{i1}, c_{i2}, \dots, c_{i6}]^T$  are evenly distributed over the state space  $[-5, 5]$  and the width of the Gaussian function is set as  $b_i = 5$ . Correspondingly, the critic NN weights are  $W = [W_1, W_2, \dots, W_{15}]^T \in \mathbb{R}^{15}$  with initial weights  $W(0) = [0, 0, \dots, 0]^T \in \mathbb{R}^{15}$ . For fair comparison, we choose control parameters via a trial and error manner such that the disturbance is estimated precisely and steady-state accuracy is consistent with following contrastive controller schemes.

1) Disturbance rejection controller (abbreviated as DR) [41]: It involves the ESO and proportion feedback to achieve tracking performance, which is expressed as

$$u = u_d + (g^T g)^{-1} g^T [\dot{X}_d - \hat{\delta} - K e] \quad (45)$$

where a gain matrix  $K = \text{diag}(1, 1, 1, 4, 4, 4)$  and  $\hat{\delta}$  is derived from (7) with  $w_0 = 10$ .

2) Proposed controller (abbreviated as ADP1): Based on the disturbance rejection control, a value function (15) is minimized to obtain optimal regulation law by updating weight.

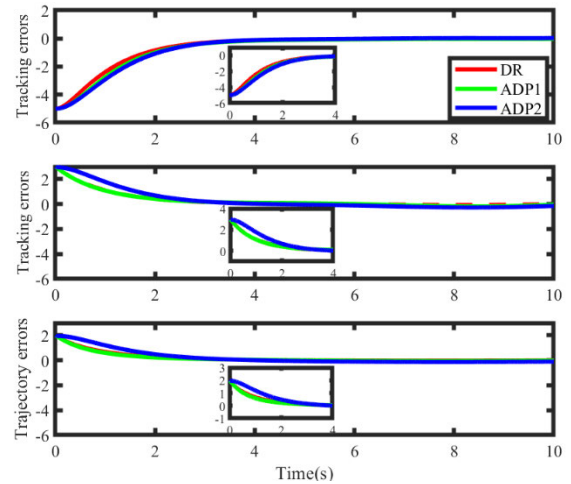


FIGURE 3. The evolution of trajectory errors under different controls.

The relevant parameters are set as  $K = \text{diag}(1, 1, 1, 2, 2, 2)$ ,  $c = 1$ ,  $\Gamma = I$ .

3) Critic-actor ADP [29] (abbreviated as ADP2): The controller in [29] is achieved by identifier NN and critic-actor NN within ADP scheme, which is the state-of-art ADP frameworks and can be found in [28]. For fair comparison, we augment critic-actor ADP with disturbance rejection controller. The learning gains in critic NN and actor NN are chosen as  $\Gamma_c = I$  and  $\Gamma_a = I$ .

### A. COMPARISON WITH DIFFERENT CONTROLLERS

The simulation results are listed in Figures 3-8. The evolution of position tracking errors among above-mentioned controllers are depicted in the Figure 3, implying that all controllers can make sure that trajectory tracking errors of position can convergence to the neighborhood of the origin in a short time despite of comprehensive disturbance consisting of unknown dynamic drift and extraneous disturbance, where comprehensive disturbances can be well estimated by ESO, as shown in Figure 4. The estimation of comprehensive disturbances can track the real lumped disturbances in a short time. The control inputs are shown in Figure 5.

In virtue of ADP framework, the optimal controllers of ADP1 and ADP2 are derived by different NN structure with update rules. The transient tracking performance by proposed ADP1 controller is faster than that of ADP2, which can be originated from the implicit representation and employment of weight errors, such that the proposed adaptive law can drive the weight errors to converge to zero in a prescribed time. Related weight convergence profiles for both methods are revealed in Figures 6-8. The proposed update rule drives the weight approximation to convergence real weight in 2s, furtherly inducing the convergence of optimal value function and optimal policy with weight estimation, which also indicate that the involved ADP1 is capable of learning the optimal control policy as weights tend to constant values. Oppositely, the traditional weight update is designed by minimizing Bellman error via gradient descent or least



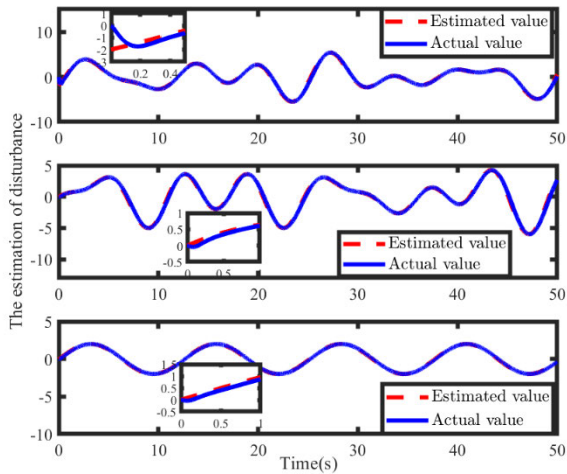


FIGURE 4. The approximation of the comprehensive disturbances by ESO.

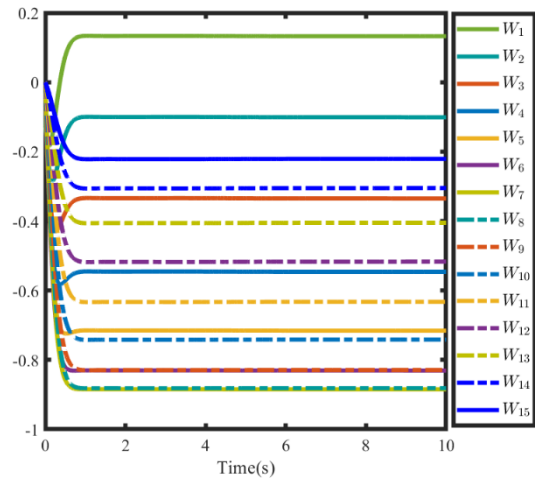


FIGURE 7. The evolution of actor weight with ADP2.

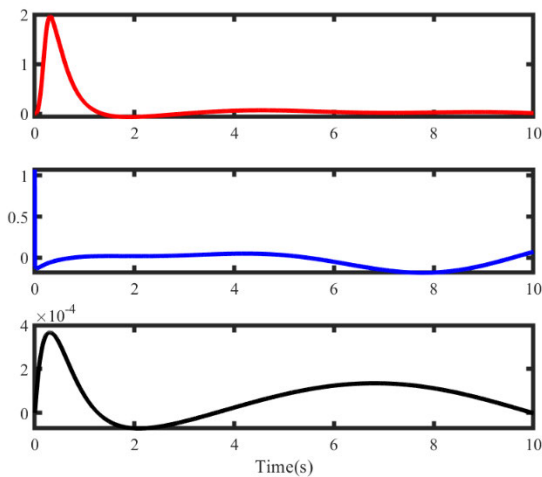


FIGURE 5. The evolution of control inputs.

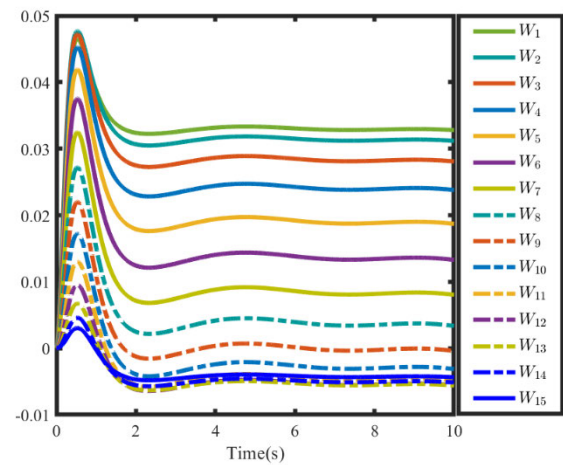


FIGURE 8. The evolution of critic weight with ADP2.

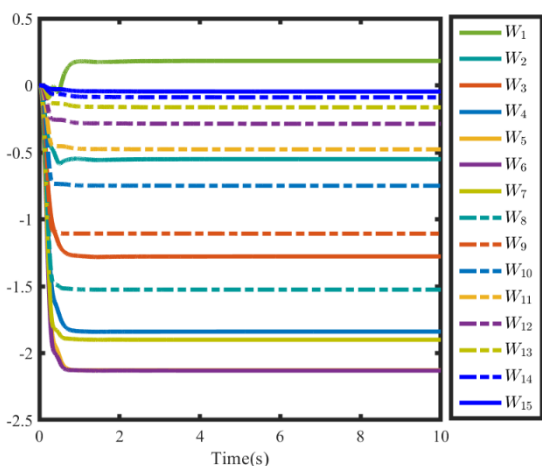


FIGURE 6. The evolution of critic weight with ADP1.

squares method in ADP2, resulting in the weight convergence with obvious lag phenomenon in the presence of uncertainties after 2s.

Further, taking tracking errors and control consumptions into consideration, the value function is given to evaluate the performance of different controllers:

$$V(e) = \int_t^\infty e^T Q e + u^T R u d\tau \quad (46)$$

Herein, the value function is denoted with positive matrices  $Q = I$  and  $R = I$ , where  $I$  represents the identity matrix. The evolution of value function under different controllers are shown in Figure 9. It interprets intuitively that proposed controller is equipped with online optimization ability in terms of tracking performance and control consumptions compared with DR controller, wherein the comparable optimization capability with critic-only ADP make the value function decrease by 25% and its performance is slightly better than that of critic-actor architecture.

### B. EFFECT OF DIFFERENT OBSERVER BANDWIDTH ON PERFORMANCE AND CONSUMPTION

To explore the optimality ability of proposed optimal tracking controller, we make simulation comparisons with DR controller, where the different observer bandwidths are

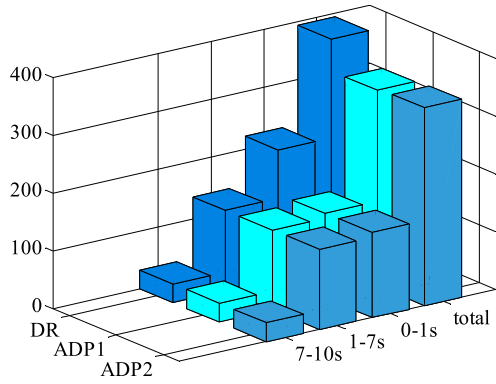


FIGURE 9. The value functions under different controllers.

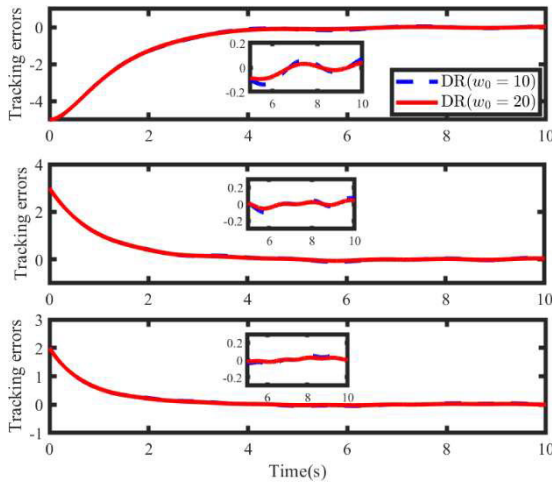


FIGURE 10. The evolution of trajectory errors among DR controllers with different observer bandwidths.

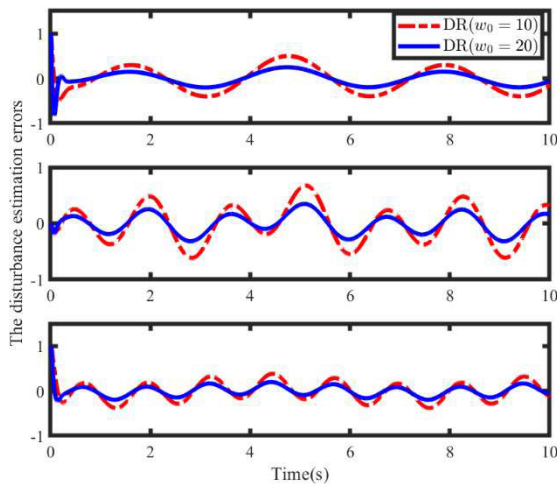


FIGURE 11. The evolution of disturbance estimation errors with different observer bandwidths.

considered. To highlight the immunity ability of proposed controller, we impose following disturbance on systems:

$$d(t) = [\sin(4t) - \cos(t), 2(\cos(4t) + \sin(2t) - \cos(t)), \sin(3t) \times \cos(2t) - \cos(t)]^T$$

TABLE 2. The performance and consumption with different observer bandwidth.

	$E$			$U$		
	$E_x$	$E_y$	$E_z$	$U_x$	$U_y$	$U_z$
DR ( $w_0 = 10$ )	0.0261	0.0151	0.0077	192.45	97.98	29.12
DR ( $w_0 = 20$ )	0.0144	0.0088	0.0044	193.77	98.99	29.42
ADP	0.0113	0.0075	0.0038	119.99	59.61	15.80

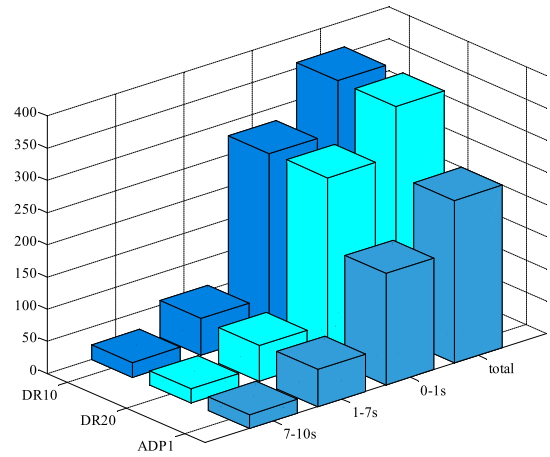


FIGURE 12. Comparison of the value functions among ADP1 and DR with different observer bandwidths.

Although the disturbance estimation errors can be decreased with increasing bandwidth, the tracking errors can be minimized with high control consumptions among DR controllers with different observer bandwidths, as revealed in Figures 10-11. Thanks to the fact that the disturbance compensation of ESO and online learning ability for updating the weight to reduce coupling errors, tracking errors can be minimized with the whole effective integration. To prove the necessity of the optimal regulation law  $u_e$  to improve tracking performance and minimize cost consumption, we employ following indices to measure steady state tracking performance and control costs:

$$E = \int_{T_i}^T (e - e_0)^T Q (e - e_0) dt$$

$$U = \int_0^T u^T Q u dt$$

with  $T_i$  being the time of entering steady state and  $e_0$  representing the mean error. Herein, we choose  $T_i = 4s$ . As is exhibited in Table 2, the ADP1 controller can achieve the better tracking errors, due to the fact that optimal regulation law acts a part of controller for minimizing tracking errors, whilst the control consumptions are not significant increasing, which further demonstrates the superiority of investigated ADP framework. With help of optimal regulation law  $u_e$ , proposed controller can achieve comparative tracking

errors, even there exists disturbance estimation errors. To further compare the tracking performance and control costs, the overall value functions are depicted during the whole optimization procedure in Figure 12. The optimizing ability can be observed significantly that inserted ADP1 scheme has less value functions than DR controllers in spite of ESO designed with different bandwidths.

## V. CONCLUSION

In this paper, the disturbance rejection-based optimal tracking control including a disturbance rejection controller and an optimal regulation law is proposed for perturbed nonlinear systems by utilizing the constructed ESO and proposed ADP techniques. Thereinto, the constructed ESO can well estimate the extraneous disturbance in system, such that the established disturbance rejection controller can ensure satisfactory disturbance compensation. Then, an optimal regulation law is designed to tradeoff the control input and tracking performance by minimizing the preassigned efficiency function. It should be pointed out that critic-only NN is used to approximate the value function in suggested framework with novel adaptive rule for updating weight, which has less computation and better control performance than existing ADP schemes with critic-actor NN. Additionally, convergence of weights can be assured in a finite time with proposed improved update rule, and tracking errors can be guaranteed to be UUB. The investigated optimal regulation law goes close to optimal controller. The simulations show the effectiveness and advantages on a practical example. Noting that proposed scheme is designed based on time-trigger, we will employ event-triggered mechanism of [42] and [43] to further taking event-triggered optimal tracking controller into account.

## REFERENCES

- [1] Y. Yang, L. Liao, H. Yang, and S. Li, "An optimal control strategy for multi-UAVs target tracking and cooperative competition," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 12, pp. 1931–1947, Dec. 2021.
- [2] A. Biswas, M. A. O. Reon, P. Das, Z. Tasneem, S. M. Mueyeen, S. K. Das, F. R. Badal, S. K. Sarker, M. M. Hassan, S. H. Abhi, M. R. Islam, M. F. Ali, M. H. Ahamed, and M. M. Islam, "State-of-the-art review on recent advancements on lateral control of autonomous vehicles," *IEEE Access*, vol. 10, pp. 114759–114786, 2022, doi: [10.1109/ACCESS.2022.3217213](https://doi.org/10.1109/ACCESS.2022.3217213).
- [3] L. Song, J. Huang, Q. Liang, L. Nie, X. Liang, and J. Zhu, "Trajectory tracking strategy for sliding mode control with double closed-loop for lawn mowing robot based on ESO," *IEEE Access*, vol. 11, pp. 1867–1882, 2023, doi: [10.1109/ACCESS.2022.3166816](https://doi.org/10.1109/ACCESS.2022.3166816).
- [4] M. Aliyari, W.-K. Wong, Y. Bouteraa, S. Najafinia, A. Fekih, and S. Mobayen, "Design and implementation of a constrained model predictive control approach for unmanned aerial vehicles," *IEEE Access*, vol. 10, pp. 91750–91762, 2022, doi: [10.1109/ACCESS.2022.3202020](https://doi.org/10.1109/ACCESS.2022.3202020).
- [5] W. Zhang, X. Shao, W. Zhang, J. Qi, and H. Li, "Unknown input observer-based appointed-time funnel control for quadrotors," *Aerosp. Sci. Technol.*, vol. 126, Jul. 2022, Art. no. 107351.
- [6] W. Guo and D. Liu, "Adaptive neural network command filtered backstepping control for the underactuated TORA system," *IEEE Access*, vol. 11, pp. 13679–13689, 2023, doi: [10.1109/ACCESS.2023.3243497](https://doi.org/10.1109/ACCESS.2023.3243497).
- [7] H.-H. An and T.-H. Cheng, "Finite-time geometric tracking controller for multirotor systems," *IEEE Access*, vol. 10, pp. 110905–110915, 2022, doi: [10.1109/ACCESS.2022.3215689](https://doi.org/10.1109/ACCESS.2022.3215689).
- [8] A. J. Muñoz-Vázquez, G. Fernández-Anaya, and J. D. Sánchez-Torres, "Fractional integro-differential sliding mode control of a class of distributed-order nonlinear systems," *J. Appl. Math. Comput.*, vol. 68, no. 4, pp. 2743–2760, Oct. 2021.
- [9] Z. Pan, Z. Sun, H. Deng, and D. Li, "A multilayer graph for multiagent formation and trajectory tracking control based on MPC algorithm," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13586–13597, Dec. 2022.
- [10] V. K. Tripathi, A. K. Kamath, L. Behera, N. K. Verma, and S. Nahavandi, "An adaptive fast terminal sliding-mode controller with power rate proportional reaching law for quadrotor position and altitude tracking," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 6, pp. 3612–3625, Jun. 2022.
- [11] S. Mobayen, K. A. Alattas, and W. Assawinchaichote, "Adaptive continuous barrier function terminal sliding mode control technique for disturbed robotic manipulator," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 10, pp. 4403–4412, Oct. 2021.
- [12] B.-S. Chen, C.-P. Wang, and M.-Y. Lee, "Stochastic robust team tracking control of multi-UAV networked system under Wiener and Poisson random fluctuations," *IEEE Trans. Cybern.*, vol. 51, no. 12, pp. 5786–5799, Dec. 2021.
- [13] Z. Gu, C. K. Ahn, D. Yue, and X. Xie, "Event-triggered  $H_\infty$  filtering for T-S fuzzy-model-based nonlinear networked systems with multisensors against DoS attacks," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 5311–5321, Jun. 2022.
- [14] R. Rout, R. Cui, and W. Yan, "Sideslip-compensated guidance-based adaptive neural control of marine surface vessels," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 2860–2871, May 2022.
- [15] K. Aryankia and R. R. Selmic, "Neural network-based formation control with target tracking for second-order nonlinear multiagent systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 1, pp. 328–341, Feb. 2022.
- [16] M. Chen, H. Wang, and X. Liu, "Adaptive practical fixed-time tracking control with prescribed boundary constraints," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 4, pp. 1716–1726, Apr. 2021.
- [17] B. Niu, J. Liu, P. Duan, J. Li, and D. Yang, "Reduced-order observer-based adaptive fuzzy tracking control scheme of stochastic switched nonlinear systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 7, pp. 4566–4578, Jul. 2021.
- [18] J. Han, "From PID to active disturbance rejection control," *IEEE Trans. Ind. Electron.*, vol. 56, no. 3, pp. 900–906, Mar. 2009.
- [19] P. Grelewicz, P. Nowak, J. Czeczot, and J. Musial, "Increment count method and its PLC-based implementation for autotuning of reduced-order ADRC with Smith predictor," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12554–12564, Dec. 2021.
- [20] A. Basit, M. Tufail, and M. Rehan, "An adaptive gain based approach for event-triggered state estimation with unknown parameters and sensor nonlinearities over wireless sensor networks," *ISA Trans.*, vol. 129, pp. 41–54, Oct. 2022.
- [21] Y. Yang, X. Si, D. Yue, and Y.-C. Tian, "Time-varying formation tracking with prescribed performance for uncertain nonaffine nonlinear multiagent systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 4, pp. 1778–1789, Oct. 2021.
- [22] X. Shao, Y. Shi, and W. Zhang, "Input-and-measurement event-triggered output-feedback chattering reduction control for MEMS gyroscopes," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 9, pp. 5579–5590, Sep. 2022.
- [23] V.-T. Nguyen, C.-Y. Lin, S.-F. Su, W. Sun, and M. J. Er, "Global finite time active disturbance rejection control for parallel manipulators with unknown bounded uncertainties," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 12, pp. 7838–7849, Dec. 2021.
- [24] X. Shao, J. Zhang, and W. Zhang, "Distributed cooperative surrounding control for mobile robots with uncertainties and aperiodic sampling," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18951–18961, Oct. 2022.
- [25] A. Basit, M. Tufail, and M. Rehan, "Event-triggered distributed state estimation under unknown parameters and sensor saturations over wireless sensor networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1772–1776, Mar. 2022.
- [26] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.
- [27] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, May 2010.

- [28] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2226–2236, Dec. 2011.
- [29] F. F. M. El-Sousy, M. M. Amin, and A. Al-Durra, "Adaptive optimal tracking control via actor-critic-identifier based adaptive dynamic programming for permanent-magnet synchronous motor drive system," *IEEE Trans. Ind. Appl.*, vol. 57, no. 6, pp. 6577–6591, Nov./Dec. 2021.
- [30] J. Na, Y. Lv, K. Zhang, and J. Zhao, "Adaptive identifier-critic-based optimal tracking control for nonlinear systems with experimental validation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 459–472, Jan. 2022.
- [31] Q.-Y. Fan and G.-H. Yang, "Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 165–177, Jan. 2016.
- [32] J. Wang, Z. Zhang, B. Tian, and Q. Zong, "Event-based robust optimal consensus control for nonlinear multiagent system with local adaptive dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 27, 2022, doi: [10.1109/TNNLS.2022.3180054](https://doi.org/10.1109/TNNLS.2022.3180054).
- [33] S. Cao, L. Sun, J. Jiang, and Z. Zuo, "Reinforcement learning-based fixed-time trajectory tracking control for uncertain robotic manipulators with input saturation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 15, 2021, doi: [10.1109/TNNLS.2021.3116713](https://doi.org/10.1109/TNNLS.2021.3116713).
- [34] G. Hu, J. Guo, Z. Guo, J. Cieslak, and D. Henry, "ADP-based intelligent tracking algorithm for reentry vehicles subjected to model and state uncertainties," *IEEE Trans. Ind. Informat.*, vol. 19, no. 4, pp. 6047–6055, Apr. 2023, doi: [10.1109/TII.2022.3171327](https://doi.org/10.1109/TII.2022.3171327).
- [35] R. Moghadam and S. Jagannathan, "Optimal adaptive control of uncertain nonlinear continuous-time systems with input and state delays," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 29, 2021, doi: [10.1109/TNNLS.2021.3112566](https://doi.org/10.1109/TNNLS.2021.3112566).
- [36] H. Dong and X. Yang, "Learning-based online optimal sliding-mode control for space circumnavigation missions with input constraints and mismatched uncertainties," *Neurocomputing*, vol. 484, pp. 13–25, May 2022.
- [37] X. Shao, X. Yue, and J. Li, "Event-triggered robust control for quadrotors with preassigned time performance constraints," *Appl. Math. Comput.*, vol. 392, Mar. 2021, Art. no. 125667.
- [38] Y. Liu, H. Li, R. Lu, Z. Zuo, and X. Li, "An overview of finite/fixed-time control and its application in engineering systems," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 12, pp. 2106–2120, Dec. 2022.
- [39] F. L. Lewis, D. L. Vrabie, and V. L. Syrmos, *Optimal Control*. New York, NY, USA: Wiley, 2012.
- [40] D. Li, S. S. Ge, and T. H. Lee, "Simultaneous arrival to origin convergence: Sliding-mode control through the norm-normalized sign function," *IEEE Trans. Autom. Control*, vol. 67, no. 4, pp. 1966–1972, Apr. 2022.
- [41] Y. Wang, M. Piao, Z. Chen, M. Sun, and Q. Sun, "Capability exploration of extended-state observer-based control under the uncertain case of disturbance and actuator saturation," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 1841–1851, Feb. 2023.
- [42] J. Liu, Y. Dong, L. Zha, E. Tian, and X. Xie, "Event-based security tracking control for networked control systems against stochastic cyber-attacks," *Inf. Sci.*, vol. 612, pp. 306–321, Oct. 2022.
- [43] J. Liu, N. Zhang, Y. Li, X. Xie, E. Tian, and J. Cao, "Learning-based event-triggered tracking control for nonlinear networked control systems with unmatched disturbance," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 5, pp. 3230–3240, May 2023, doi: [10.1109/TSMC.2022.3224432](https://doi.org/10.1109/TSMC.2022.3224432).



**YANPING GAO** received the B.S. degree from the Hebei University of Technology, Tianjin, China, in 2015, where she is currently pursuing the Ph.D. degree in control theory and control engineering with the School of Artificial Intelligence. Her research interests include optimal control and quadrotors.



**ZUOJUN LIU** received the B.S. degree in automation and the M.S. degree in control theory and control engineering from the Hebei University of Technology, Tianjin, China, in 1994 and 2000, respectively, and the Ph.D. degree in control theory and control engineering from Nankai University, Tianjin, in 2004. He completed his visiting scholar research with Nanyang Technological University, Singapore, in 2004, and Case Western Reserve University, USA, in 2013, respectively. He is currently a Professor with the School of Artificial Intelligence, Hebei University of Technology. His research interests include intelligent control theory and robotics.

• • •