This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Learning-Based Optimal Cooperative Formation Tracking Control for Multiple UAVs: A Feedforward-Feedback Design Framework

Boyang Zhang, Maolong Lv, Shaohua Cui, Xiangwei Bu, and Ju H. Park, *Senior Member, IEEE*

***Abstract*— Notwithstanding the successful design of state-of-the-art cooperative control protocols to accomplish formation tracking for multiple unmanned aerial vehicles (UAVs), the assurance of performance optimality cannot be guaranteed in the face of complex disturbances affecting these multi-UAV systems. In order to surmount this challenge, this research endeavor aims to establish a feedforward-feedback learning-based optimal control methodology to facilitate cooperative UAV formation tracking in the presence of intricate disturbances. To be more precise, by leveraging backstepping-based feedback control, the problem of UAV formation tracking is transformed into an equivalent optimal regulation problem. Consequently, a learning-based feedforward control scheme is devised, wherein the cooperative policy iteration algorithm is formulated based on a two-player zero-sum game. The critic-only echo state network (ESN) is employed to approximate the optimal feedforward control policies, with the inclusion of an online adaptive tuning law and compensation terms to alleviate the persistence of excitation condition and eliminate the need for an initial admissible control. As a result, the closed-loop stability is guaranteed in terms of uniformly ultimately boundedness for tracking errors and ESN weights.**

***Note to Practitioners*—In real-world scenarios, the flight of multiple UAVs is invariably affected by intricate disturbances, resulting in compromised tracking precision. There is an urgent need to enhance resistance to disturbances and ensure optimal performance for cooperative formation tracking of multiple UAVs. Beyond the capabilities of model-based controllers, the integration of reinforcement learning has shown promise in achieving robust control actions. By introducing the cooperative policy iteration algorithm based on a two-player zero-sum game, the tracking performances of UAV formation can be further optimized. In order to facilitate the practical application of reinforcement learning in UAV systems, our proposed algorithm addresses the persistency of excitation condition by incorporating innovative compensation terms into the ESN tuning law. Furthermore, we resolve the requirement for initial admissible control by introducing a novel piecewise compensation term into the ESN tuning law, which is based on a newly proposed Lyapunov function.**

***Index Terms*— Feedforward-feedback learning-based control, two-player zero-sum game, unmanned aerial vehicle formation tracking.**

I. Introduction

In RECENT years, the utilization of unmanned aerial vehicle (UAV) formations has proven successful in various applications such as load transport [1], surveillance [2], and target enclosing [3]. Numerous formation control methodologies have been proposed to address these applications. For instance, in [4], a decentralized sliding mode controller was introduced to achieve consensus in altitude and heading angle for UAV formations. Reference [5] developed a time-varying formation controller that leveraged local status information sharing to solve the consensus control problem in UAV formations. References [6], [7], [8], and [9] directed their attention towards coordinated formation stabilization and formation tracking of UAVs, considering fixed topologies and switching topologies, respectively. However, it is important to note that most existing results, including [4], [5], [6], [7], [8], and [9], fail to guarantee performance optimality. Complicating matters further, UAV formations are often exposed to intricate disturbances [10], [11], [12], [13] when operating in complex environments. This reality poses challenges to ensuring optimality, thus greatly limiting the performance of UAV formations. Consequently, there is a critical need to pursue both optimality and robustness to achieve stable formation control in the presence of complex disturbances. This objective falls under the domain of mixed robust control and optimal control [14], [15], [16], [17].

Primarily, the current focus in the realm of mixed robust control and optimal control design lies in the framework

Manuscript received 12 June 2023; accepted 21 September 2023. This article was recommended for publication by Associate Editor W. He and Editor Z. Li upon evaluation of the reviewers' comments. This work was supported in part by the Young Talent Fund of Association for Science and Technology in Shanxi under Grant 20220101, in part by the Young Talent Support Project for Military Science and Technology under Grant 2022-JCJQ-QT-018, in part by the Post-Doctoral International Exchange Project under Grant YJ20220347, in part by the National Natural Science Foundation under Grant 62303489, in part by the Post-Doctoral Science Foundation General Program under Grant 2022M723877, in part by the Post-Doctoral Science Foundation Special Funding under Grant 2023T160790, and in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government under Grant RS-2023-00208078. *(Corresponding author: Maolong Lv.)*

Boyang Zhang is with the Beijing Blue Sky Science and Technology Innovation Center, Beijing 100010, China (e-mail: boyang_530@163.com).

Maolong Lv is with Air Traffic Navigation College, Air Force Engineering University, and also with the National Key Lab of Aerospace Power System and Plasma Technology, Xi'an 710051, China (e-mail: maolonglv@163.com).

Shaohua Cui is with the School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: shaoh_cui@buaa.edu.cn).

Xiangwei Bu is with Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China (e-mail: buxiangwei1987@126.com).

Ju H. Park is with the Department of Electrical Engineering, Yeungnam University, Kyongsan 38541, Republic of Korea (e-mail: jessie@ynu.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2023.3322028.

Digital Object Identifier 10.1109/TASE.2023.3322028

1545-5955 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

of the two-player zero-sum game. This framework offers a viable solution to the robust control problem, where the controller acts as the minimizing player while the disturbance represents the maximizing counterpart [18], [19]. Recent advancements have presented various approaches for addressing two-player zero-sum games through online learning of control and disturbance policies [20], [21]. These methods involve the adaptation of control and disturbance policies using reinforcement learning, with neural networks employed to identify the corresponding value function.

Reinforcement learning provides an optimal design technique for control systems [22]. Different from the traditional optimal control designs, controllers incorporating reinforcement learning technique are capable of learning the approximate solution to optimal control from the feedback of surroundings [23], [24]. With respect to learning-based optimal control designs [25], [26], [27], [28], [29], online approximate solutions based on policy iteration are developed, whereby the neural networks are deployed to approximate the value function and control policy. This advantage has motivated researchers to develop several learning-based control approaches for the path planning [30], obstacle avoidance [31] and resource allocation [32] of UAV formation.

Taking inspiration from foregoing literature, a feedforward-feedback learning-based optimal control is developed for multiple UAVs. The objective is to seek both the optimality and robustness of cooperative formation tracking. To this end, a backstepping-based feedback control technique is employed to transform the UAV formation tracking problem into an equivalent optimal regulation problem. Subsequently, a learning-based optimal control is derived by utilizing a two-player zero-sum game framework based on Echo State Network (ESN) approximation. The main contributions of this paper can be summarized as follows:

- In response to the challenge posed by the inability of most UAV formation control methods [4], [5], [6], [7], [8], [9] to ensure performance optimality, a feedforward-feedback learning-based optimal control scheme is devised. The proposed control scheme addresses the complex disturbances encountered in cooperative UAV formation tracking while also guaranteeing performance optimality.
- In addition to the model-based backstepping controllers [5], [6], [7], [8], [9], the integration of reinforcement learning offers the potential to achieve robust control actions. Through the introduction of the cooperative policy iteration algorithm based on two-player zero-sum game, the tracking performances of UAV formation are further optimized.
- In contrast to most learning-based control methods, such as [25], [26], [27], [28], [29], [30], [31], and [32], which satisfy the persistency of excitation condition by introducing probing noise to the system dynamics or through a replay strategy involving the collection and training of large amounts of recorded data, our proposed algorithm takes a different way. The proposed algorithm attempts to eliminate the persistency of excitation condition by introducing innovative compensation terms into the ESN tuning law. Furthermore, our proposed algorithm addresses the requirement for initial admissible control by incorporating a novel piecewise compensation term into the ESN tuning law, which is based on a newly proposed Lyapunov function.

The rest of paper is organized as follows: Section II presents the problem formulation and preliminaries. This is followed by the presentation of feedforward-feedback learning-based optimal control design in Section III. In Section IV, an ESN-based approximation for the value function is provided. The stability analysis is covered in Section V. Experiment and simulation validations are discussed in Section VI. Section VII draws the conclusion.

Notations: $\Re$, $\Re^m$ and $\Re^{m\times n}$ denote the real number, the real m-vector and the real $m\times n$ matric, respectively. $|\cdot|$ represents the absolute value, $\|\cdot\|$ is the Euclidean norm of a vector or the Frobenius norm of a matrix, $\mathrm{tr}(\cdot)$, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the trace, minimum and maximum eigenvalues of the matrix respectively, $\boldsymbol{I}_n$ is a $n\times n$ identity matrix, operator $\otimes$ denotes the kronecker product, and $[\cdot\,;\,\cdot]$ is a two-vector concatenation operation.

II. Problem Formulation and Preliminaries

A. Algebraic Graph Theory Basics

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ to describe the connections among the multiple UAVs, it consists of nodes $\mathcal{V} = \{v_1, \cdots, v_n\}$ and the sets of edges $\mathcal{E} = \{(i,j),\ \ i,j \in \mathcal{V}, \text{and } i \neq j\}$, and $\mathcal{A} = [a_{ij}] \in \Re^{n\times n}$ denotes the weighted adjacency matrix of $\mathcal{G}$. If there exists an edge between the ith UAV and jth UAV, then $a_{ij} = a_{ji} \neq 0$ and otherwise $a_{ij} = a_{ji} = 0$. Moreover, the neighbors of ith UAV is denoted by the set $\mathcal{N}_i = \{v_j :\ (v_j,\ v_i) \in \mathcal{E}\}$. The Laplacian matrix is defined by $\mathcal{L} = \mathcal{D} - \mathcal{A}$, where $\mathcal{D} = \mathrm{diag}\{d_1, \cdots, d_n\}$ with $d_i = \sum_{j=1}^{n} a_{ij}$. $\mathcal{G}$ is connected if there is a path from each UAV to others.

For the sake of multi-task need, a switching graph is considered for UAV formation. Specifically, define an infinite sequence of time intervals $[t_k, t_{k+1})$, where $t_0 = 0$ and $0 < t_\tau \leq t_{k+1} - t_k$ with t_τ denoting the dwell time. The graph keeps fixed during $[t_k, t_{k+1})$ and switches at time t_{k+1}. Let $\sigma(t) : [0, +\infty) \to \{1, 2, \ldots, n\}$ stand for a switching signal, where n represents the number of all probable graphs. The value of $\sigma(t)$ is the index of switching topology. Define $a_{ij}^{\sigma(t)}$ as the weight of $\mathcal{A}$ for $\sigma(t)$. Let $\mathcal{N}_i^{\sigma(t)}$, $\mathcal{G}^{\sigma(t)}$ and $\mathcal{L}^{\sigma(t)}$ be the neighbor set, the topology, and the Laplacian matrix for $\sigma(t)$, respectively.

B. Problem Description

Consider a formation of N identical fixed wing UAVs. By introducing the auxiliary dynamics with a path parameter θ_i, the kinematic model for each UAV is expressed as [33], [34]

$$\begin{cases} \dot{x}_i = V_i \cos\psi_i \cos\gamma_i + w_{xi} \\ \dot{y}_i = V_i \sin\psi_i \cos\gamma_i + w_{yi} \\ \dot{z}_i = V_i \sin\gamma_i + w_{zi} \\ \dot{\psi}_i = g\tan\phi_{ci} \big/ V_i \\ \dot{\gamma}_i = \kappa(\gamma_{ci} - \gamma_i) \\ \dot{\theta}_i = \varphi_i \\ \dot{\varphi}_i = \mu_i \end{cases} \tag{1}$$

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where x_i, y_i and z_i denote the position in the inertial frame, V_i is the airspeed, ψ_i is the heading angle, γ_i is the air-relative flight path angle. The control inputs are selected as ϕ_{ci} and γ_{ci}, which represent the commands of roll angle and flight path angle, respectively. w_{xi}, w_{yi} and w_{zi} are complicated unknowns along $x-$, $y-$ and $z-$ axes, respectively, μ_i is a virtual control law to generate the path. Note that we follow the coordinated turn assumption to achieve the motion of ψ_i. And the dynamics of γ_i is described using a first-order system with a time constant κ to accommodate the relatively slow response set up by the autopilot. θ_i is a path parameter, which can be any physical quantity, and we take $\dot{\theta}_i = \varphi_i$ and $\dot{\varphi}_i = \mu_i$.

Definition 1: (*Cooperative UAV Formation Tracking*) Multiple UAVs are said to achieve cooperative formation tracking if

$$\lim_{t\to\infty}\left(\boldsymbol{p}_i(t) - \Delta\boldsymbol{p}_i(t) - \boldsymbol{p}_0(t)\right) = \boldsymbol{0} \tag{2}$$

where $\boldsymbol{p}_i(t) = \left[x_i(t), y_i(t), z_i(t)\right]^{\mathrm{T}}$ denotes the position of the ith UAV, $\Delta\boldsymbol{p}_i(t) = \left[\Delta x_i(t), \Delta y_i(t), \Delta z_i(t)\right]^{\mathrm{T}}$ represents the relative position with respect to the virtual leader which specifies the expected time-varying formation, and $\boldsymbol{p}_0(t)$ is the position of virtual leader, expressed as

$$\begin{aligned}\boldsymbol{p}_0(t) = \{&x_0, y_0, z_0 \in \Re | \theta_0 \in [\theta_{\min}, \theta_{\max}] \\ &\mapsto x_0 = r_x(\theta_0), y_0 = r_y(\theta_0), z_0 = r_z(\theta_0)\}\end{aligned} \tag{3}$$

where $\dot{\theta}_0 = f(\theta_0, t)$ with function $f : \Re^n \to \Re^n$, $\theta_{\min}$ and $\theta_{\max}$ are the minimum and maximum values of θ_0, respectively.

From (2), define $e_{xi} = x_i - \Delta x_i - x_0$, $e_{yi} = y_i - \Delta y_i - y_0$ and $e_{zi} = z_i - \Delta z_i - z_0$. Taking the time derivative of e_{xi}, e_{yi} and e_{zi} along (1) yields

$$\begin{cases}\dot{e}_{xi} = f_{xi}(\boldsymbol{X}_i) + w_{xi} - \Delta\dot{x}_i \\ \dot{e}_{yi} = f_{yi}(\boldsymbol{X}_i) + w_{yi} - \Delta\dot{y}_i \\ \dot{e}_{zi} = f_{zi}(\boldsymbol{X}_i) + w_{zi} - \Delta\dot{z}_i\end{cases} \tag{4}$$

where $f_{xi}(\boldsymbol{X}_i) = V_i\cos\psi_i\cos\gamma_i - \frac{\partial x_0}{\partial\theta_0}\varphi_0$, $f_{yi}(\boldsymbol{X}_i) = V_i\sin\psi_i\cos\gamma_i - \frac{\partial y_0}{\partial\theta_0}\varphi_0$ and $f_{zi}(\boldsymbol{X}_i) = V_i\sin\gamma_i - \frac{\partial z_0}{\partial\theta_0}\varphi_0$, $\boldsymbol{X}_i = \left[x_i, y_i, z_i, \psi_i, \gamma_i, \theta_0, \varphi_0\right]^{\mathrm{T}}$.

For the sake of simplification, let $u_{1i} = g\tan\phi_{ci}/V_i$ and $u_{2i} = \kappa(\gamma_{ci} - \gamma_i)$. Then the implementable control laws can be calculated as $\phi_{ci} = \arctan\left(u_{1i}V_i/g\right)$ and $\gamma_{ci} = \gamma_i + u_{2i}/\kappa$. Define $\boldsymbol{e}_i = \left[e_{xi}, e_{yi}, e_{zi}\right]^{\mathrm{T}}$, $\boldsymbol{d}_i = \left[w_{xi}, w_{yi}, w_{zi}\right]^{\mathrm{T}}$, $\boldsymbol{u}_i = [u_{i1}, u_{i2}, \mu_0]^{\mathrm{T}}$ and $\boldsymbol{f}_i = \left[f_{xi}, f_{yi}, f_{zi}\right]^{\mathrm{T}}$. It follows from (4) that

$$\begin{cases}\dot{\boldsymbol{e}}_i = \boldsymbol{f}_i + \boldsymbol{d}_i - \Delta\dot{\boldsymbol{p}}_i \\ \dot{\boldsymbol{f}}_i = \boldsymbol{F}_i + \boldsymbol{G}_i\boldsymbol{u}_i\end{cases} \tag{5}$$

where $\boldsymbol{F}_i = -\left[\frac{\partial\left(\frac{\partial x_0}{\partial\theta_0}\right)}{\partial\theta_0}\varphi_0^2, \frac{\partial\left(\frac{\partial y_0}{\partial\theta_0}\right)}{\partial\theta_0}\varphi_0^2, \frac{\partial\left(\frac{\partial z_0}{\partial\theta_0}\right)}{\partial\theta_0}\varphi_0^2\right]^{\mathrm{T}}$,

$$\boldsymbol{G}_i = \begin{bmatrix} -V_i\sin\psi_i\cos\gamma_i & -V_i\cos\psi_i\sin\gamma_i & -\frac{\partial x_0}{\partial\theta_0} \\ V_i\cos\psi_i\cos\gamma_i & -V_i\sin\psi_i\sin\gamma_i & -\frac{\partial y_0}{\partial\theta_0} \\ 0 & V_i\cos\gamma_i & -\frac{\partial z_0}{\partial\theta_0}\end{bmatrix}.$$

The objective of this paper is to design a feedforward-feedback learning-based optimal control scheme such that the multiple UAVs follow the virtual leader and maintain a prescribed formation configuration in the presence of complicated disturbances, while respecting the switching topologies $\mathcal{G}^{\sigma(t)}$.

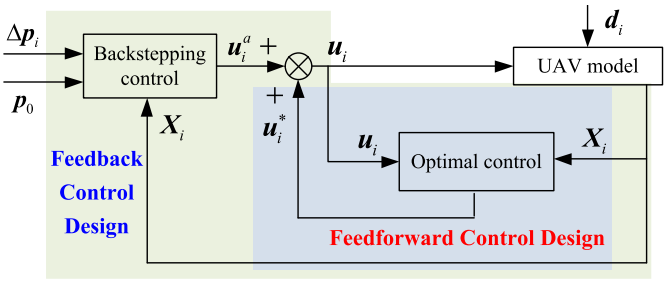


Fig. 1. Configuration of feedforward-feedback control scheme.

III. Feedforward-Feedback Learning-Based Optimal Control Design

This section presents a mixed feedforward-feedback based optimal control scheme to achieve the cooperative UAV formation tracking (see Definition 1) with complicated disturbances. Firstly, the feedback control law, denoted by $\boldsymbol{u}_i^a$ is designed using the backstepping control method to obtain a new transformed error dynamics. Subsequently, an optimal feedforward control action, denoted by $\boldsymbol{u}_i^*$ is derived. As a result, the mixed feedforward-feedback based control signal is given as $\boldsymbol{u}_i = \boldsymbol{u}_i^a + \boldsymbol{u}_i^*$. The overall control configuration is shown in Fig. 1.

A. Backstepping-Based Feedback Control Design

A backstepping-based feedback control design is proposed as follows:

Step 1: To carry out the controller design, some notations are defined as follows: $\widetilde{\boldsymbol{f}}_i = \boldsymbol{f}_i - \boldsymbol{f}_{id}$, $\boldsymbol{f}_{id}$ is the virtual control such that $\boldsymbol{f}_{id} = \boldsymbol{f}_{id}^a + \boldsymbol{f}_{id}^*$, where $\boldsymbol{f}_{id}^a$ is the feedback virtual control and $\boldsymbol{f}_{id}^*$ is the feedforward optimal term. Then it follows from (5) that

$$\dot{\boldsymbol{e}}_i = \widetilde{\boldsymbol{f}}_i + \boldsymbol{f}_{id}^a + \boldsymbol{f}_{id}^* + \boldsymbol{d}_i - \Delta\dot{\boldsymbol{p}}_i \tag{6}$$

Consider the following Lyapunov function candidate

$$L_{i1} = \frac{1}{2}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{e}_i \tag{7}$$

Take the time derivative of L_{i1} along (6) yields

$$\dot{L}_{i1} = \boldsymbol{e}_i^{\mathrm{T}}\left(\widetilde{\boldsymbol{f}}_i + \boldsymbol{f}_{id}^a + \boldsymbol{f}_{id}^* + \boldsymbol{d}_i - \Delta\dot{\boldsymbol{p}}_i\right) \tag{8}$$

The feedback virtual control $\boldsymbol{f}_{id}^a$ is chosen as

$$\boldsymbol{f}_{id}^a = -k_{2i}\sum_{j\in\mathcal{N}_i^{\sigma(t)}} a_{ij}^{\sigma(t)}\left(\boldsymbol{e}_i - \boldsymbol{e}_j\right) - k_{1i}\boldsymbol{e}_i + \Delta\dot{\boldsymbol{p}}_i \tag{9}$$

where k_{1i} and k_{2i} are positive design parameters.

Substituting (9) into (8) gives

$$\begin{aligned}\dot{L}_{i1} = &-k_{2i}\boldsymbol{e}_i^{\mathrm{T}}\sum_{j\in\mathcal{N}_i^{\sigma(t)}} a_{ij}^{\sigma(t)}\left(\boldsymbol{e}_i - \boldsymbol{e}_j\right) + \boldsymbol{e}_i^{\mathrm{T}}\widetilde{\boldsymbol{f}}_i \\ &- k_{1i}\|\boldsymbol{e}_i\|^2 + \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{d}_i + \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{f}_{id}^*\end{aligned} \tag{10}$$

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Step 2: Choose the following Lyapunov function candidate

$$L_{i2} = \frac{1}{2}\widetilde{\boldsymbol{f}}_i^{\mathrm{T}}\widetilde{\boldsymbol{f}}_i \tag{11}$$

As shown in Fig. 1, the control input $\boldsymbol{u}_i$ is taken as $\boldsymbol{u}_i = \boldsymbol{u}_i^a + \boldsymbol{u}_i^*$. The time derivative of L_{i2} along (5) is

$$\dot{L}_{i2} = \widetilde{\boldsymbol{f}}_i^{\mathrm{T}}\left\{\boldsymbol{F}_i + \boldsymbol{G}_i\left(\boldsymbol{u}_i^a + \boldsymbol{u}_i^*\right) - \dot{\boldsymbol{f}}_{id}\right\} \tag{12}$$

The feedback control $\boldsymbol{u}_i^a$ is designed as

$$\begin{aligned}\boldsymbol{u}_i^a =& -k_{2i}\boldsymbol{G}_i^{-1}\sum_{j\in\mathcal{N}_i^{\sigma(t)}}\left\{a_{ij}^{\sigma(t)}\left(\boldsymbol{e}_i-\boldsymbol{e}_j\right)+a_{ij}^{\sigma(t)}\left(\widetilde{\boldsymbol{f}}_i-\widetilde{\boldsymbol{f}}_j\right)\right\}\\ &-k_{2i}\boldsymbol{G}_i^{-1}\frac{\tilde{\boldsymbol{f}}_i}{\|\tilde{\boldsymbol{f}}_i\|^2}\boldsymbol{e}_i^{\mathrm{T}}\sum_{j\in\mathcal{N}_i^{\sigma(t)}}a_{ij}^{\sigma(t)}\left(\tilde{\boldsymbol{f}}_i-\tilde{\boldsymbol{f}}_j\right)\\ &-\boldsymbol{G}_i^{-1}\left(\boldsymbol{F}_i+k_{3i}\widetilde{\boldsymbol{f}}_i+\boldsymbol{e}_i-\dot{\boldsymbol{f}}_{id}\right)\end{aligned} \tag{13}$$

where k_{3i} is a positive design parameter.

Substituting (13) into (12) yields

$$\begin{aligned}\dot{L}_{i2} =& -k_{2i}\widetilde{\boldsymbol{f}}_i^{\mathrm{T}}\sum_{j\in\mathcal{N}_i^{\sigma(t)}}\left\{a_{ij}^{\sigma(t)}\left(\boldsymbol{e}_i-\boldsymbol{e}_j\right)+a_{ij}^{\sigma(t)}\left(\widetilde{\boldsymbol{f}}_i-\widetilde{\boldsymbol{f}}_j\right)\right\}\\ &-k_{2i}\boldsymbol{e}_i^{\mathrm{T}}\sum_{j\in\mathcal{N}_i^{\sigma(t)}}\left\{a_{ij}^{\sigma(t)}\left(\tilde{\boldsymbol{f}}_i-\tilde{\boldsymbol{f}}_j\right)\right\}\\ &-k_{3i}\left\|\widetilde{\boldsymbol{f}}_i\right\|^2-\tilde{\boldsymbol{f}}_i^{\mathrm{T}}\boldsymbol{e}_i+\widetilde{\boldsymbol{f}}_i^{\mathrm{T}}\boldsymbol{G}_i\boldsymbol{u}_i^*\end{aligned} \tag{14}$$

Theorem 1: Given the error dynamics (5), define the augmented control input vector $\boldsymbol{\tau}_i = \left[\boldsymbol{f}_{id}^{\mathrm{T}}, \boldsymbol{u}_i^{\mathrm{T}}\right]^{\mathrm{T}}$ such that $\boldsymbol{\tau}_i = \boldsymbol{\tau}_i^a + \boldsymbol{\tau}_i^*$, where $\boldsymbol{\tau}_i^a = \left[\boldsymbol{f}_{id}^{\mathrm{aT}}, \boldsymbol{u}_i^{\mathrm{aT}}\right]^{\mathrm{T}}$ is the augmented feedback control input in (9) and (13), and $\boldsymbol{\tau}_i^* = \left[\boldsymbol{f}_{id}^{*\mathrm{T}}, \boldsymbol{u}_i^{*\mathrm{T}}\right]^{\mathrm{T}}$ is the augmented feedforward control action which optimally stabilizes the transformed error dynamics

$$\dot{\mathfrak{F}}_i = \boldsymbol{C}_i\boldsymbol{\tau}_i^* + \boldsymbol{K}_i\boldsymbol{d}_i \tag{15}$$

where $\boldsymbol{C}_i = \begin{bmatrix}\boldsymbol{1}_{3\times3} & \boldsymbol{0}_{3\times3}\\ \boldsymbol{0}_{3\times3} & \boldsymbol{G}_i\end{bmatrix}$, $\boldsymbol{K}_i = \begin{bmatrix}\boldsymbol{1}_{3\times3}\\ \boldsymbol{0}_{3\times3}\end{bmatrix}$ and $\mathfrak{F}_i = \left[\boldsymbol{e}_i^{\mathrm{T}}, \widetilde{\boldsymbol{f}}_i^{\mathrm{T}}\right]^{\mathrm{T}}$.

Then the optimal formation tracking of UAVs is accomplished by forcing $\mathfrak{F}_i$ to converge to an arbitrarily small region around origin in an optimal manner.

Proof: Based on (7) and (11), let us consider the entire Lyapunov function

$$L = \sum_{i=1}^{N}\left(L_{i1}+L_{i2}\right) \tag{16}$$

Taking the time derivative of L along (10) and (14) gives

$$\begin{aligned}\dot{L} =& -k_{2i}\sum_{i=1}^{N}\left(\boldsymbol{e}_i+\widetilde{\boldsymbol{f}}_i\right)^{\mathrm{T}}\sum_{j\in\mathcal{N}_i^{\sigma(t)}}\left\{a_{ij}^{\sigma(t)}\left(\boldsymbol{e}_i-\boldsymbol{e}_j\right)+a_{ij}^{\sigma(t)}\left(\widetilde{\boldsymbol{f}}_i-\widetilde{\boldsymbol{f}}_j\right)\right\}\\ &+\sum_{i=1}^{N}\left(\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{f}_{id}^*+\widetilde{\boldsymbol{f}}_i^{\mathrm{T}}\boldsymbol{G}_i\boldsymbol{u}_i^*+\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{d}_i\right)\\ &-\sum_{i=1}^{N}\left(k_{1i}\|\boldsymbol{e}_i\|^2+k_{3i}\left\|\widetilde{\boldsymbol{f}}_i\right\|^2\right)\end{aligned} \tag{17}$$

Let $\boldsymbol{e} = \left[\boldsymbol{e}_1^{\mathrm{T}}, \boldsymbol{e}_2^{\mathrm{T}}, \cdots, \boldsymbol{e}_N^{\mathrm{T}}\right]^{\mathrm{T}}$, $\widetilde{\boldsymbol{f}} = [\widetilde{\boldsymbol{f}}_1^{\mathrm{T}}, \widetilde{\boldsymbol{f}}_2^{\mathrm{T}}, \cdots, \widetilde{\boldsymbol{f}}_N^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{\Lambda}^{\sigma(t)} = \begin{bmatrix}\mathcal{L}^{\sigma(t)} & \mathcal{L}^{\sigma(t)}\\ \mathcal{L}^{\sigma(t)} & \mathcal{L}^{\sigma(t)}\end{bmatrix}$. Then, the following equality holds

$$\begin{aligned}&\sum_{i=1}^{N}\left(\boldsymbol{e}_i+\widetilde{\boldsymbol{f}}_i\right)^{\mathrm{T}}\sum_{j\in\mathcal{N}_i^{\sigma(t)}}\left\{a_{ij}^{\sigma(t)}\left(\boldsymbol{e}_i-\boldsymbol{e}_j\right)+a_{ij}^{\sigma(t)}\left(\widetilde{\boldsymbol{f}}_i-\widetilde{\boldsymbol{f}}_j\right)\right\}\\ &=\begin{bmatrix}\boldsymbol{e}\\ \widetilde{\boldsymbol{f}}\end{bmatrix}^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{\sigma(t)}\otimes\boldsymbol{I}_3\right)\begin{bmatrix}\boldsymbol{e}\\ \widetilde{\boldsymbol{f}}\end{bmatrix}\end{aligned} \tag{18}$$

Assuming that graph $\mathcal{G}^{\sigma(t)}$ is uniformly connected, $\mathcal{L}^{\sigma(t)}$ is symmetric and positive semi-definite, and further $\boldsymbol{\Lambda}^{\sigma(t)}$ is positive semi-definite. Then, it holds that

$$-k_{2i}\begin{bmatrix}\boldsymbol{e}\\ \widetilde{\boldsymbol{f}}\end{bmatrix}^{\mathrm{T}}\left(\boldsymbol{\Lambda}^{\sigma(t)}\otimes\boldsymbol{I}_3\right)\begin{bmatrix}\boldsymbol{e}\\ \widetilde{\boldsymbol{f}}\end{bmatrix}\le 0 \tag{19}$$

Substituting (19) into (17) gives rise to

$$\begin{aligned}\dot{L} \le& \sum_{i=1}^{N}\left(\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{f}_{id}^*+\widetilde{\boldsymbol{f}}_i^{\mathrm{T}}\boldsymbol{G}_i\boldsymbol{u}_i^*+\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{d}_i\right)\\ &-\sum_{i=1}^{N}\left(k_{1i}\|\boldsymbol{e}_i\|^2+k_{3i}\left\|\widetilde{\boldsymbol{f}}_i\right\|^2\right)\\ \le& \sum_{i=1}^{N}\left\{\mathfrak{F}_i^{\mathrm{T}}\left(\begin{bmatrix}\boldsymbol{1}_{3\times3} & \boldsymbol{0}_{3\times3}\\ \boldsymbol{0}_{3\times3} & \boldsymbol{G}_i\end{bmatrix}\boldsymbol{\tau}_i^*+\begin{bmatrix}\boldsymbol{1}_{3\times3}\\ \boldsymbol{0}_{3\times3}\end{bmatrix}\boldsymbol{d}_i\right)\right\}\\ &-c_i\sum_{i=1}^{N}\|\mathfrak{F}_i\|^2\end{aligned} \tag{20}$$

It follows from (20) that when $\boldsymbol{\tau}_i^*$ stabilizes the system (15), the first term in right-hand side of (20) turns negative.

This completes the proof of Theorem 1. ■

B. Learning-Based Optimal Feedforward Control Design

In this section, a zero-sum game based optimal controller is designed to stabilize the transformed error dynamics (15), which forces $\mathfrak{F}_i$ to converge to an arbitrarily small region around origin in an optimal manner.

Given the feedforward control policies $\boldsymbol{\tau}_i^*$ and $\boldsymbol{d}_i$, the infinite horizon integral cost is defined as

$$V_i = \int_t^{\infty}\left(Q(\mathfrak{F}_i)+\boldsymbol{\tau}_i^{*\mathrm{T}}\boldsymbol{R}_\tau\boldsymbol{\tau}_i^*-\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{R}_d\boldsymbol{d}_i\right)dt \tag{21}$$

where $Q(\mathfrak{F}_i) = \mathfrak{F}_i^{\mathrm{T}}\boldsymbol{Q}\mathfrak{F}_i$ is a penalty on the error $\mathfrak{F}_i$, $\boldsymbol{R}_\tau \in \Re^{6\times6}$ and $\boldsymbol{R}_d \in \Re^{6\times6}$ are positive design matrixes.

Then, the optimal cost for two-player zero-sum game can be obtained

$$V_i^* = \min_{\boldsymbol{\tau}_i^*}\max_{\boldsymbol{d}_i}\int_t^{\infty}\left(Q(\mathfrak{F}_i)+\boldsymbol{\tau}_i^{*\mathrm{T}}\boldsymbol{R}_\tau\boldsymbol{\tau}_i^*-\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{R}_d\boldsymbol{d}_i\right)dt \tag{22}$$

In view of the value function (21), the Hamiltonian function associated with feedforward control policies $\boldsymbol{\tau}_i^*$ and $\boldsymbol{d}_i$ is defined as

$$H_i = Q(\mathfrak{F}_i)+\boldsymbol{\tau}_i^{*\mathrm{T}}\boldsymbol{R}_\tau\boldsymbol{\tau}_i^*-\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{R}_d\boldsymbol{d}_i+\nabla V_i^{\mathrm{T}}\left(\boldsymbol{C}_i\boldsymbol{\tau}_i^*+\boldsymbol{K}_i\boldsymbol{d}_i\right) \tag{23}$$

where $\nabla V_i = \partial V_i/\partial\mathfrak{F}_i \in \Re^6$.

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

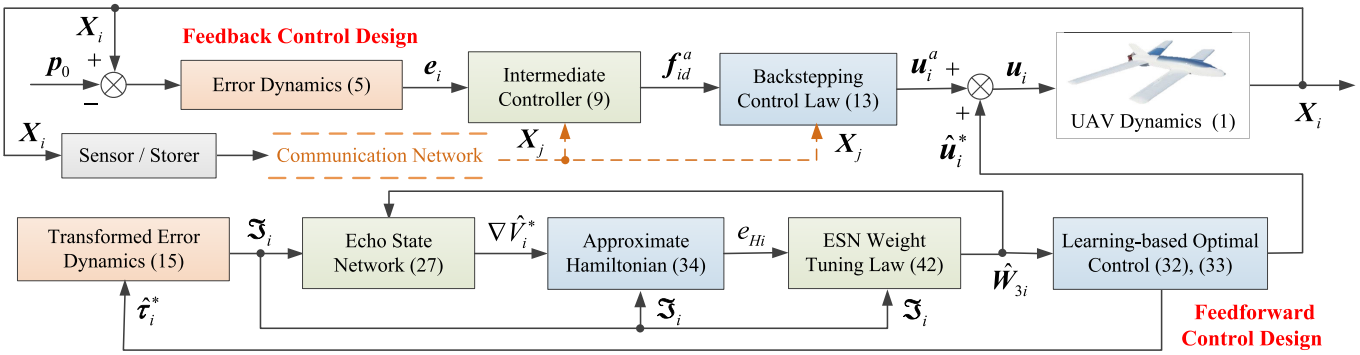


Fig. 2. Overall design framework of proposed control methodology.

Based on the stationarity condition of Nash equilibrium for (23), the optimal feedforward control policies are obtained by

$$\bar{\boldsymbol{\tau}}_i^* = -\frac{1}{2}\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}\nabla V_i^* \tag{24}$$

$$\boldsymbol{d}_i^* = \frac{1}{2}\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}\nabla V_i^* \tag{25}$$

where the optimal value function V_i^* is the solution of following Hamilton-Jacobi-Isaacs (HJI) equation

$$0 = -\frac{1}{4}\nabla V_i^{*\mathrm{T}}\boldsymbol{C}_i\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}\nabla V_i^* + \frac{1}{4}\nabla V_i^{*\mathrm{T}}\boldsymbol{K}_i\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}\nabla V_i^* + Q(\mathfrak{I}_i) \tag{26}$$

where $\nabla V_i^*(0) = 0$.

If the HJI equation can be solved by V_i^*, the optimal feedforward control policies $\bar{\boldsymbol{\tau}}_i^*$ and $\boldsymbol{d}_i^*$ can be implemented by (24) and (25). However, the analytic solution to the HJI equation is generally hard to obtain due to its inherently nonlinear feature. To overcome this difficulty, a learning-based control algorithm is proposed to learn the solution to the HJI equation online using a critic ESN in order to get the optimal feedforward control policies $\bar{\boldsymbol{\tau}}_i^*$ and $\boldsymbol{d}_i^*$.

IV. ESN-Based Approximation for Value Function

To implement the feedforward control policies, an online critic ESN is used to approximate the optimal value function as [35]

$$\begin{aligned}\dot{\boldsymbol{b}}_i(t) &= \frac{1}{\beta}(-\alpha\boldsymbol{b}_i(t) + \phi(\boldsymbol{W}_{1i}\mathfrak{I}_i(t) + \boldsymbol{W}_{2i}\boldsymbol{b}_i(t)))\\ \nabla V_i^*(t) &= \boldsymbol{W}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)] + \boldsymbol{\varepsilon}_i(\mathfrak{I}_i(t),\boldsymbol{b}_i(t))\end{aligned} \tag{27}$$

where α is a positive leaky rate, β is a positive time constant, $\boldsymbol{W}_{1i}$, $\boldsymbol{W}_{2i}$ and $\boldsymbol{W}_{3i}$ are weighted matrices for input, reservoir states, and output of ESN, respectively. Thereinto, $\boldsymbol{W}_{1i}$ and $\boldsymbol{W}_{2i}$ are always sparse matrices generated stochastically and do not need to be tuned. Only $\boldsymbol{W}_{3i}$ is needed to trained. The training of $\boldsymbol{W}_{3i}$ uses the gradient descent method in the following. $\phi(\cdot)$ is the reservoir unit function, $\mathfrak{I}_i(t)$ is the input of ESN, $\boldsymbol{\varepsilon}_i(\mathfrak{I}_i(t),\boldsymbol{b}_i(t))$ is the approximation error. The output activation function is taken as an identity function.

Remark 1: The main reasons of using ESN to approximate the optimal value functions is that compared with the existing NNs [36], the ESN uses a dynamical reservoir to replace the hidden layer of the recurrent NNs. Hence, we only need to train the output weight $\boldsymbol{W}_{3i}$, which reduces the online computation of weight tuning for UAV system.

Using (27), the optimal feedforward control policies are given by

$$\bar{\boldsymbol{\tau}}_i^* = -\frac{1}{2}\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}\boldsymbol{W}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)] + \boldsymbol{\varepsilon}_{\tau i} \tag{28}$$

$$\boldsymbol{d}_i^* = \frac{1}{2}\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}\boldsymbol{W}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)] + \boldsymbol{\varepsilon}_{di} \tag{29}$$

where $\boldsymbol{\varepsilon}_{\tau i} = -\frac{1}{2}\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_{di} = \frac{1}{2}\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i$.

Assumption 1: [37] The approximation error $\boldsymbol{\varepsilon}_i$ is bounded by $\|\boldsymbol{\varepsilon}_i\| \le \kappa_\varepsilon$ with κ_ε a positive constant. Then it follows that $\|\boldsymbol{\varepsilon}_{\tau i}\| \le \kappa_{\varepsilon\tau}$ and $\|\boldsymbol{\varepsilon}_{di}\| \le \kappa_{\varepsilon d}$ with $\kappa_{\varepsilon\tau}$ and $\kappa_{\varepsilon d}$ positive constants.

Substituting (27) into the HJI equation (26) gives

$$\begin{aligned}\varepsilon_i^{\mathrm{HJI}} = &-\frac{1}{4}(\boldsymbol{W}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}\boldsymbol{P}_{\tau i}\boldsymbol{W}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)]\\ &+\frac{1}{4}(\boldsymbol{W}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}\boldsymbol{P}_{di}\boldsymbol{W}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)]\\ &+Q(\mathfrak{I}_i)\end{aligned} \tag{30}$$

where $\boldsymbol{P}_{\tau i} = \boldsymbol{C}_i\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}$, $\boldsymbol{P}_{di} = \boldsymbol{K}_i\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}$, and $\varepsilon_i^{\mathrm{HJI}}$ is the residual error arising from ESN approximation.

Since the ideal $\boldsymbol{W}_{3i}$ is unknown, we let $\widehat{\boldsymbol{W}}_{3i}$ be the estimate of $\boldsymbol{W}_{3i}$, and we define $\widetilde{\boldsymbol{W}}_{3i} = \boldsymbol{W}_{3i} - \widehat{\boldsymbol{W}}_{3i}$. Then, $\nabla V_i^*(\mathfrak{I}_i(t))$ can be estimated as

$$\nabla\widehat{V}_i^*(t) = \widehat{\boldsymbol{W}}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)] \tag{31}$$

Substituting (31) into (24) and (25), the estimated optimal feedforward control policies are given by

$$\widehat{\bar{\boldsymbol{\tau}}}_i^* = -\frac{1}{2}\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}\widehat{\boldsymbol{W}}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)] \tag{32}$$

$$\widehat{\boldsymbol{d}}_i = \frac{1}{2}\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}\widehat{\boldsymbol{W}}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)] \tag{33}$$

The approximate Hamiltonian function is obtained as

$$\begin{aligned}\widehat{H}_i = &\left(\widehat{\boldsymbol{W}}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)]\right)^{\mathrm{T}}\left(\boldsymbol{C}_i\widehat{\bar{\boldsymbol{\tau}}}_i^* + \boldsymbol{K}_i\widehat{\boldsymbol{d}}_i\right) + Q(\mathfrak{I}_i)\\ &+\widehat{\bar{\boldsymbol{\tau}}}_i^{*\mathrm{T}}\boldsymbol{R}_\tau\widehat{\bar{\boldsymbol{\tau}}}_i^* - \widehat{\boldsymbol{d}}_i^{\mathrm{T}}\boldsymbol{R}_d\widehat{\boldsymbol{d}}_i \triangleq e_{Hi}\end{aligned} \tag{34}$$

Then, it follows along (30)-(33) that

$$e_{Hi} = \left(\widetilde{\boldsymbol{W}}_{3i}[\mathfrak{I}_i(t);\boldsymbol{b}_i(t)]\right)^{\mathrm{T}}\left(\boldsymbol{C}_i\widehat{\bar{\boldsymbol{\tau}}}_i^* + \boldsymbol{K}_i\widehat{\boldsymbol{d}}_i\right) - \varepsilon_i^{\mathrm{HJI}}$$

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

$$- \frac{1}{4}\big(\widetilde{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\big)^{\mathrm{T}}\boldsymbol{P}_{\tau i}\widetilde{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)] + \frac{1}{4}\big(\widetilde{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\big)^{\mathrm{T}}\boldsymbol{P}_{di}\widetilde{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)] \quad (35)$$

It is desired to select $\widetilde{\boldsymbol{W}}_{3i}$ to minimize the squared residual error $E_{Hi} = \frac{1}{2}e_{Hi}^2$ by the gradient descent algorithm. Then, the tuning law of $\widehat{\boldsymbol{W}}_{3i}$ is designed as

$$\dot{\widehat{\boldsymbol{W}}}_{3i} = -\lambda_i \frac{\partial E_{Hi}}{\partial \widehat{\boldsymbol{W}}_{3i}} = -\lambda_i\big(\boldsymbol{C}_i\widehat{\boldsymbol{\tau}}_i^* + \boldsymbol{K}_i\widehat{\boldsymbol{d}}_i\big)[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}e_{Hi} \quad (36)$$

where λ_i is a positive tuning parameter.

Remark 2: Before starting the training process, an initial admissible control signal is necessary. In [38, Definition 1], [39, Definition 1], and [40, Assumption 1], the initial admissible control solution is assumed to be existent *a priori* in the stability analysis and is chosen manually in simulation and experimental examples. This priori assumption is practically difficult or even impossible to be satisfied since the output weights of the ESNs generally have dozens of dimensions [35]. We remove the initial admissible control condition by proposing new piecewise adaptation laws (38) for the critic ESN with the help of a newly proposed operator Γ_i (39) which is selected based on Lyapunov's sufficient condition (37). This operation is desired to pull the controller back to the admissible range when the control is not admissible.

Remark 3: In terms of tuning law (36), the persistency of excitation (PE) condition is needed to guarantee the convergence of $\widehat{\boldsymbol{W}}_{3i}$. Generally, the PE condition is directly assumed to be satisfied [38, eq. (27)], and is always satisfied by exerting probing noise on the system dynamics [39, eq. (28)]. While in [40, Section III.C], the PE condition is solved by so-called experience replay strategy in the sense of collecting and training massive amounts of recorded data. In contrast with aforementioned methods, we attempt to remove the PE assumption via an adaptive manner by introducing appropriate compensation terms (40), (41) in the adaptation laws (36) of the critic ESN. On the one hand, to stabilize the instability terms of (36) caused by the absence of the PE assumption, we introduce the compensation terms (40) in the adaptation law (36). On the other hand, to offset the superfluous terms of $\widehat{\boldsymbol{W}}_{3i}$ in (46), the compensation terms (41) are designed combined with the adaptation laws (36).

To remove the requirement of initial admissible control pair, we introduce a Lyapunov function candidate L_{i3} satisfying

$$\dot{L}_{i3} = -\frac{1}{2}\nabla L_{i3}^{\Im_i\mathrm{T}}\boldsymbol{P}_{\tau i}\widehat{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)] + \frac{1}{2}\nabla L_{i3}^{\Im_i\mathrm{T}}\boldsymbol{P}_{di}\widehat{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)] \quad (37)$$

where $\nabla L_{i3}^{\Im_i}$ denotes the partial derivative of L_{i3} with respect to $\Im_i$.

Lemma 1: [41] Given the transformed error dynamics (15) with associated value function (21) and optimal control policies (24) and (25), it is supposed that there exists a continuously differentiable Lyapunov function L_{i3} such that $\dot{L}_{i3} = \nabla L_{i3}^{\Im_i\mathrm{T}}\big(\boldsymbol{C}_i\bar{\boldsymbol{\tau}}_i^* + \boldsymbol{K}_i\boldsymbol{d}_i^*\big) < 0$. Let $\boldsymbol{R}_{\Im i}$ be an appropriate positive definite matrix. Then it holds that $\nabla L_{i3}^{\Im_i\mathrm{T}}\big(\boldsymbol{C}_i\bar{\boldsymbol{\tau}}_i^* + \boldsymbol{K}_i\boldsymbol{d}_i^*\big) = -\nabla L_{i3}^{\Im_i\mathrm{T}}\boldsymbol{R}_{\Im i}\nabla L_{i3}^{\Im_i}$.

Based on (37), a compensation term is designed as

$$\boldsymbol{T}_{1i} = -\Gamma_i\lambda_i\frac{\partial \dot{L}_{i3}}{\partial \widehat{\boldsymbol{W}}_{3i}} = \frac{1}{2}\Gamma_i\lambda_i\boldsymbol{P}_{\tau i}\nabla L_{i3}^{\Im_i}[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} - \frac{1}{2}\Gamma_i\lambda_i\boldsymbol{P}_{di}\nabla L_{i3}^{\Im_i}[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \quad (38)$$

where the operator Γ_i is defined as

$$\Gamma_i = \begin{cases} 0, & \text{if } \nabla L_{i3}^{\Im_i\mathrm{T}}\big(\boldsymbol{C}_i\widehat{\boldsymbol{\tau}}_i^* + \boldsymbol{K}_i\widehat{\boldsymbol{d}}_i\big) < 0 \\ 1, & \text{else} \end{cases} \quad (39)$$

Remark 4: The operator Γ_i is chosen based on Lyapunov's sufficient condition for stability. This operation is desired to pull the controller back to the admissible range when the control is not admissible. If the closed-loop system is unstable, the operator $\Gamma_i = 1$, (38) will be activated, which turns $\dot{L}_{i3} < 0$ hold true. Otherwise, the operator $\Gamma_i = 0$, and (38) do not take effect.

To relax the PE condition, the following compensation terms are introduced in the tuning law for $\widehat{\boldsymbol{W}}_{3i}$ as follows.

$$\boldsymbol{T}_{2i} = \lambda_i\sigma_i\boldsymbol{M}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \quad (40)$$

$$\boldsymbol{T}_{3i} = -\frac{1}{4}\lambda_i\boldsymbol{M}_i{}^{\mathrm{T}}\boldsymbol{P}_{\tau i}\boldsymbol{M}_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} + \frac{1}{4}\lambda_i\boldsymbol{M}_i{}^{\mathrm{T}}\boldsymbol{P}_{di}\boldsymbol{M}_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \quad (41)$$

where $\boldsymbol{M}_i = \widehat{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]$, $\boldsymbol{H}_i = \boldsymbol{C}_i\widehat{\boldsymbol{\tau}}_i^* + \boldsymbol{K}_i\widehat{\boldsymbol{d}}_i$, and σ_i is a learning rate.

Based on (38)-(41), the new tuning law of $\widehat{\boldsymbol{W}}_{3i}$ denotes

$$\begin{aligned}\dot{\widehat{\boldsymbol{W}}}_{3i} =& -\lambda_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}e_{Hi} + \lambda_i\sigma_i\boldsymbol{M}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \\ &- \frac{1}{4}\lambda_i\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{P}_{\tau i}\boldsymbol{M}_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \\ &+ \frac{1}{4}\lambda_i\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{P}_{di}\boldsymbol{M}_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \\ &+ \frac{1}{2}\Gamma_i\lambda_i\boldsymbol{P}_{\tau i}\nabla L_{i3}^{\Im_i}[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \\ &- \frac{1}{2}\Gamma_i\lambda_i\boldsymbol{P}_{di}\nabla L_{i3}^{\Im_i}[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}\end{aligned} \quad (42)$$

V. Stability Analysis

Theorem 2: Consider the transformed error dynamics (15). Let the ESN weight tuning law be provided by (42). The feedforward control policies are given by (32) and (33). With the Assumption 1, the tracking error $\Im_i$ and the weight estimation error $\widetilde{\boldsymbol{W}}_{3i}$ are uniformly ultimately bounded (UUB).

Proof: Consider the following Lyapunov function candidate

$$L_{i4} = L_{i3} + \frac{1}{2\lambda_i}\mathrm{tr}\Big(\widetilde{\boldsymbol{W}}_{3i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big) \quad (43)$$

Taking the time derivative of (43) gives

$$\dot{L}_{i4} = \nabla L_{i3}^{\Im_i}\boldsymbol{H}_i + \frac{1}{\lambda_i}\mathrm{tr}\big(\dot{\widetilde{\boldsymbol{W}}}_{3i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\big) \quad (44)$$

Using (35) and (42), it follows that

$$\dot{\widetilde{\boldsymbol{W}}}_{3i} = -\lambda_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}\Big(\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\boldsymbol{H}_i - \frac{1}{4}\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\boldsymbol{P}_{\tau i}\widetilde{\boldsymbol{M}}_i$$

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

$$+\frac{1}{4}\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\boldsymbol{P}_{di}\widetilde{\boldsymbol{M}}_i-\varepsilon_i^{\mathrm{HJI}}\Big)-\lambda_i\sigma_i\boldsymbol{M}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}$$
$$+\frac{1}{4}\lambda_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{P}_{\tau i}\boldsymbol{M}_i$$
$$-\frac{1}{4}\lambda_i\boldsymbol{H}_i[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{P}_{di}\boldsymbol{M}_i$$
$$-\frac{1}{2}\Gamma_i\lambda_i\boldsymbol{P}_{\tau i}\nabla L_{i3}^{\Im_i}[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}}$$
$$+\frac{1}{2}\Gamma_i\lambda_i\boldsymbol{P}_{di}\nabla L_{i3}^{\Im_i}[\Im_i(t);\boldsymbol{b}_i(t)]^{\mathrm{T}} \tag{45}$$

where $\widetilde{\boldsymbol{M}}_i=\widetilde{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]$.

Substituting (45) into (44), the term $\mathrm{tr}\big(\dot{\widetilde{\boldsymbol{W}}}_{3i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\big)$ becomes

$$\frac{1}{\lambda_i}\mathrm{tr}\big(\dot{\widetilde{\boldsymbol{W}}}_{3i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\big)=-\mathrm{tr}\Big(\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\boldsymbol{H}_i\boldsymbol{H}_i^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\Big)+\mathrm{tr}\Big(\sigma_i\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\Big)$$
$$+\frac{1}{4}\mathrm{tr}\big((\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}\boldsymbol{P}_{\tau i}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\big)$$
$$-\frac{1}{4}\mathrm{tr}\big((\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}\boldsymbol{P}_{di}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\big)$$
$$-\mathrm{tr}\big(\sigma_i(\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\big)+\mathrm{tr}\big(\varepsilon_i^{\mathrm{HJI}}\boldsymbol{H}_i^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\big)$$
$$-\frac{1}{2}\mathrm{tr}\Big(\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\boldsymbol{P}_{\tau i}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\Big)$$
$$+\frac{1}{2}\mathrm{tr}\Big(\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\boldsymbol{P}_{di}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}\widetilde{\boldsymbol{M}}_i\Big)$$
$$-\frac{1}{2}\mathrm{tr}\Big(\Gamma_i[\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{\tau i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big)$$
$$+\frac{1}{2}\mathrm{tr}\Big(\Gamma_i[\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{di}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big) \tag{46}$$

Define

$$\boldsymbol{N}_i=\boldsymbol{H}_i\boldsymbol{H}_i^{\mathrm{T}}+\frac{1}{2}\boldsymbol{P}_{\tau i}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}$$
$$-\frac{1}{2}\boldsymbol{P}_{di}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}-\sigma_i\boldsymbol{I}_6 \tag{47}$$
$$\boldsymbol{S}_i=\frac{1}{4}(\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}\boldsymbol{P}_{\tau i}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}$$
$$-\frac{1}{4}(\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}\boldsymbol{P}_{di}\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\boldsymbol{H}_i^{\mathrm{T}}$$
$$-\sigma_i(\boldsymbol{W}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)])^{\mathrm{T}}+\varepsilon_i^{\mathrm{HJI}}\boldsymbol{H}_i^{T} \tag{48}$$

Combining (46)-(48), (44) can be rewritten as

$$\dot{L}_{i4}=\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{H}_i-\mathrm{tr}\Big(\widetilde{\boldsymbol{M}}_i^{\mathrm{T}}\boldsymbol{N}_i\widetilde{\boldsymbol{M}}_i\Big)+\mathrm{tr}\big(\boldsymbol{S}_i\widetilde{\boldsymbol{M}}_i\big)$$
$$-\frac{1}{2}\mathrm{tr}\Big(\Gamma_i[\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{\tau i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big)$$
$$+\frac{1}{2}\mathrm{tr}\Big(\Gamma_i[\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{di}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big) \tag{49}$$

It can be concluded from [41] that $\boldsymbol{S}_i$ is bounded by $\|\boldsymbol{S}_i\|\le\kappa_s$. Let the parameters $\boldsymbol{P}_{\tau i}$, $\boldsymbol{P}_{\tau i}$ and σ_i be chosen such that $\lambda_{\min}(\boldsymbol{N}_i)>0$. Then it follows that

$$\dot{L}_{i4}\le\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{H}_i-\lambda_{\min}(\boldsymbol{N}_i)\big\|\widetilde{\boldsymbol{M}}_i\big\|^2+\kappa_s\big\|\widetilde{\boldsymbol{M}}_i\big\|$$
$$-\frac{1}{2}\mathrm{tr}\Big(\Gamma_i[\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{\tau i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big)$$
$$+\frac{1}{2}\mathrm{tr}\Big(\Gamma_i[\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{di}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big) \tag{50}$$

According to (39), two cases of $\Gamma_i=0$ and $\Gamma_i=1$ are analyzed as follows.

Case 1: $\Gamma_i=0$, which indicates that $\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{H}_i<0$. It follows from (50) that

$$\dot{L}_{i4}\le-\kappa_{\Im}\Big\|\nabla L_{i3}^{\Im_i}\Big\|-\lambda_{\min}(\boldsymbol{N}_i)\big\|\widetilde{\boldsymbol{M}}_i\big\|^2+\kappa_s\big\|\widetilde{\boldsymbol{M}}_i\big\|$$
$$=-\kappa_{\Im}\Big\|\nabla L_{i3}^{\Im_i}\Big\|-\lambda_{\min}(\boldsymbol{N}_i)\Big(\big\|\widetilde{\boldsymbol{M}}_i\big\|-\frac{\kappa_s}{2\lambda_{\min}(\boldsymbol{N}_i)}\Big)^2$$
$$+\frac{\kappa_s^2}{4\lambda_{\min}(\boldsymbol{N}_i)} \tag{51}$$

Give the following inequalities

$$\Big\|\nabla L_{i3}^{\Im_i}\Big\|>\frac{\kappa_s^2}{4\kappa_{\Im}\lambda_{\min}(\boldsymbol{N}_i)} \tag{52}$$

or

$$\big\|\widetilde{\boldsymbol{M}}_i\big\|>\frac{\kappa_s}{\lambda_{\min}(\boldsymbol{N}_i)} \tag{53}$$

hold true, then $\dot{L}_{i4}<0$ is satisfied.

Case 2: $\Gamma_i=1$, which means that $\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{H}_i\ge0$. Then (50) further satisfies

$$\dot{L}_{i4}\le\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{H}_i-\lambda_{\min}(\boldsymbol{N}_i)\big\|\widetilde{\boldsymbol{M}}_i\big\|^2+\kappa_s\big\|\widetilde{\boldsymbol{M}}_i\big\|$$
$$-\frac{1}{2}\mathrm{tr}\Big([\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{\tau i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big)$$
$$+\frac{1}{2}\mathrm{tr}\Big([\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{di}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big) \tag{54}$$

By adding and subtracting $\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{C}_i\bar{\boldsymbol{\tau}}_i^*$ and $\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{K}_i\boldsymbol{d}_i^*$ to the right-hand side of (54), one arrives at

$$\dot{L}_{i4}\le\nabla L_{i3}^{\Im_i{\mathrm{T}}}\Big(\boldsymbol{C}_i\bar{\boldsymbol{\tau}}_i^*+\boldsymbol{K}_i\boldsymbol{d}_i^*\Big)+\kappa_s\big\|\widetilde{\boldsymbol{M}}_i\big\|-\lambda_{\min}(\boldsymbol{N}_i)\big\|\widetilde{\boldsymbol{M}}_i\big\|^2$$
$$+\nabla L_{i3}^{\Im_i{\mathrm{T}}}\Big\{\boldsymbol{C}_i\Big(\widehat{\boldsymbol{\tau}}_i^*-\bar{\boldsymbol{\tau}}_i^*\Big)+\boldsymbol{K}_i\big(\widehat{\boldsymbol{d}}_i-\boldsymbol{d}_i^*\big)\Big\}$$
$$-\frac{1}{2}\mathrm{tr}\Big([\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{\tau i}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big)$$
$$+\frac{1}{2}\mathrm{tr}\Big([\Im_i(t);\boldsymbol{b}_i(t)]\nabla L_{i3}^{\Im_i{\mathrm{T}}}\boldsymbol{P}_{di}^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}\Big) \tag{55}$$

Substituting (28), (29), (32) and (33) into (55), it follows

$$\dot{L}_{i4}\le\nabla L_{i3}^{\Im_i{\mathrm{T}}}\Big(\boldsymbol{C}_i\bar{\boldsymbol{\tau}}_i^*+\boldsymbol{K}_i\boldsymbol{d}_i^*\Big)-\nabla L_{i3}^{\Im_i{\mathrm{T}}}(\boldsymbol{C}_i\boldsymbol{\varepsilon}_{\tau i}+\boldsymbol{K}_i\boldsymbol{\varepsilon}_{\tau i})$$
$$-\lambda_{\min}(\boldsymbol{N}_i)\big\|\widetilde{\boldsymbol{M}}_i\big\|^2+\kappa_s\big\|\widetilde{\boldsymbol{M}}_i\big\| \tag{56}$$

Recalling Assumption 1, it can be concluded that $\|\boldsymbol{C}_i\boldsymbol{\varepsilon}_{\tau i}+\boldsymbol{K}_i\boldsymbol{\varepsilon}_{\tau i}\|\le\kappa_c$ with κ_c a positive parameter.

$$\dot{L}_{i4}\le-\lambda_{\min}(\boldsymbol{R}_{\Im i})\Big\|\nabla L_{i3}^{\Im_i}\Big\|^2+\kappa_c\Big\|\nabla L_{i3}^{\Im_i}\Big\|$$
$$-\lambda_{\min}(\boldsymbol{N}_i)\big\|\widetilde{\boldsymbol{M}}_i\big\|^2+\kappa_s\big\|\widetilde{\boldsymbol{M}}_i\big\|$$
$$\le-\lambda_{\min}(\boldsymbol{R}_{\Im i})\Big(\Big\|\nabla L_{i3}^{\Im_i}\Big\|-\frac{\kappa_c}{2\lambda_{\min}(\boldsymbol{R}_{\Im i})}\Big)^2$$
$$-\lambda_{\min}(\boldsymbol{N}_i)\Big(\big\|\widetilde{\boldsymbol{M}}_i\big\|-\frac{\kappa_s}{2\lambda_{\min}(\boldsymbol{N}_i)}\Big)^2+\zeta_i \tag{57}$$

where $\zeta_i=\frac{\kappa_s^2}{4\lambda_{\min}(\boldsymbol{N}_i)}+\frac{\kappa_c^2}{4\lambda_{\min}(\boldsymbol{R}_{\Im i})}$.

If the following inequalities

$$\Big\|\nabla L_{i3}^{\Im_i}\Big\|>\frac{\kappa_c}{2\lambda_{\min}(\boldsymbol{R}_{\Im i})}+\sqrt{\frac{\zeta_i}{\lambda_{\min}(\boldsymbol{R}_{\Im i})}} \tag{58}$$

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

or

$$\left\|\widetilde{\boldsymbol{M}}_i\right\| > \frac{\kappa_s}{2\lambda_{\min}(\boldsymbol{N}_i)} + \sqrt{\frac{\zeta_i}{\lambda_{\min}(\boldsymbol{N}_i)}} \tag{59}$$

hold true, there exists $\dot{L}_{i4} < 0$.

This completes the proof of Theorem 2. ■

Theorem 3: With the Assumption 1, the feedback policy pair $(\widehat{\boldsymbol{\tau}}_i^*, \widehat{\boldsymbol{d}}_i)$ converges to the approximate Nash equilibrium solution of the zero-sum game, i.e., $\left\|\bar{\boldsymbol{\tau}}_i^* - \widehat{\boldsymbol{\tau}}_i^*\right\|$ and $\|\boldsymbol{d}_i^* - \widehat{\boldsymbol{d}}_i\|$ are UUB.

Proof: From (28), (29), (32) and (33), it follows that

$$\left\|\bar{\boldsymbol{\tau}}_i^* - \widehat{\boldsymbol{\tau}}_i^*\right\| \le \frac{1}{2}\|-\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\| \tag{60}$$

$$\|\boldsymbol{d}_i^* - \widehat{\boldsymbol{d}}_i\| \le \frac{1}{2}\|-\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}\widetilde{\boldsymbol{W}}_{3i}[\Im_i(t);\boldsymbol{b}_i(t)]\| \tag{61}$$

Invoking the Theorem 2, $[\Im_i(t);\boldsymbol{b}_i(t)]$ is bounded by $\|[\Im_i(t);\boldsymbol{b}_i(t)]\| \le b_z$, then it follows that

$$\left\|\bar{\boldsymbol{\tau}}_i^* - \widehat{\boldsymbol{\tau}}_i^*\right\| \le \frac{1}{2}\lambda_{\max}\left(\boldsymbol{R}_\tau^{-1}\boldsymbol{C}_i^{\mathrm{T}}\right)b_z\widetilde{\boldsymbol{W}}_{3i} \tag{62}$$

$$\|\boldsymbol{d}_i^* - \widehat{\boldsymbol{d}}_i\| \le \frac{1}{2}\lambda_{\max}\left(\boldsymbol{R}_d^{-1}\boldsymbol{K}_i^{\mathrm{T}}\right)b_z\widetilde{\boldsymbol{W}}_{3i} \tag{63}$$

Recalling the boundedness of $\widetilde{\boldsymbol{W}}_{3i}$, it is obvious that $\left\|\bar{\boldsymbol{\tau}}_i^* - \widehat{\boldsymbol{\tau}}_i^*\right\|$ and $\|\boldsymbol{d}_i^* - \widehat{\boldsymbol{d}}_i\|$ are UUB. This completes the proof of Theorem 3. ■

VI. Experiment and Simulation Validation

To showcase the effectiveness of our proposed control algorithm, comprehensive flight tests encompassing both experimental and simulated validations are conducted in the presence of complex disturbances. In Section A, an experimental test utilizing the Links-RT UAV Platform is constructed to validate the effectiveness of proposed controller. Additionally, in Section B, the comparisons with state-of-the-art approaches are presented through simulation validation, thereby demonstrating the superiority of our proposed controller. These two sections encompass two representative flight tasks: low altitude penetration with formation reconfiguration and coordinated turn with formation keeping.

A. Support of Proposed Algorithm Validation by Experiment Test

To demonstrate the feasibility of proposed controller, the experimental validation is carried out based on the Links-RT UAV Platform (supported by Beijing Links Co., Ltd.). The overall experiment setup and detailed hardcore parts of Links-RT UAV Platform are presented in Figs. 3-4. The interactive relationship between functional units is depicted in Fig. 5. The proposed controller is firstly implemented in MATLAB. Then the code generation by Links-Auto Coder is used to convert the MATLAB language into C code, which is downloaded by Pixhawk. The Pixhawk is responsible for executing the proposed control algorithm with a sampling time of 2ms and generating pulse-width modulation (PWM) signals, which are sent to the real-time simulator, where the multi-UAV dynamics model is compiled and loaded into the real-time simulator, such that the velocity and attitude motion are calculated and transmitted to Pixhawk and monitoring software, while presenting the 3D visual scenes in Tacview and plotting the velocity and attitude tracking curves by MATLAB.

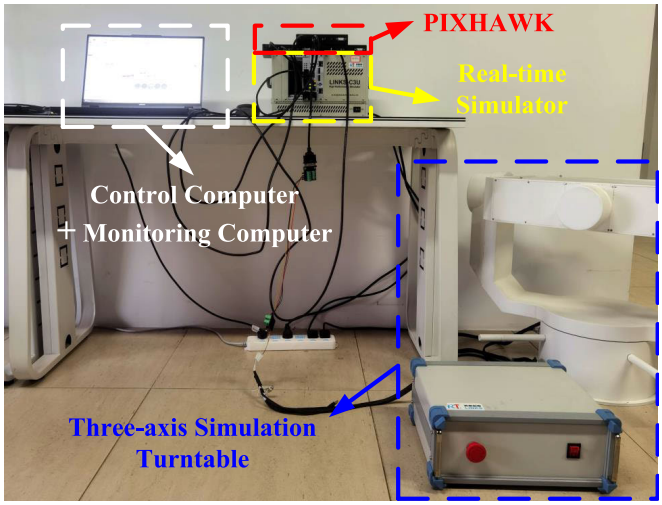


Fig. 3. Experiment prototype of the Links-RT UAV Platform.

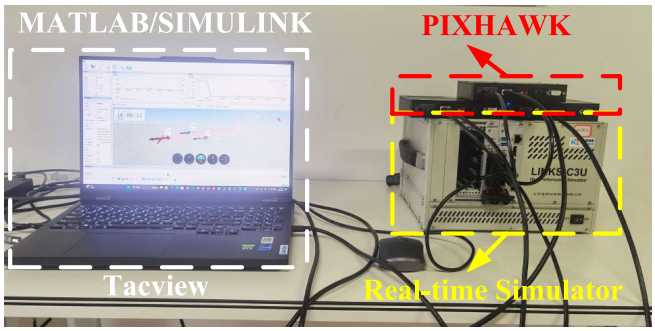


Fig. 4. Hardcore parts of the Links-RT UAV Platform.

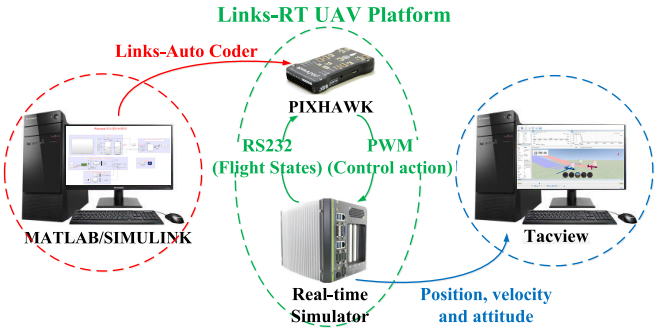
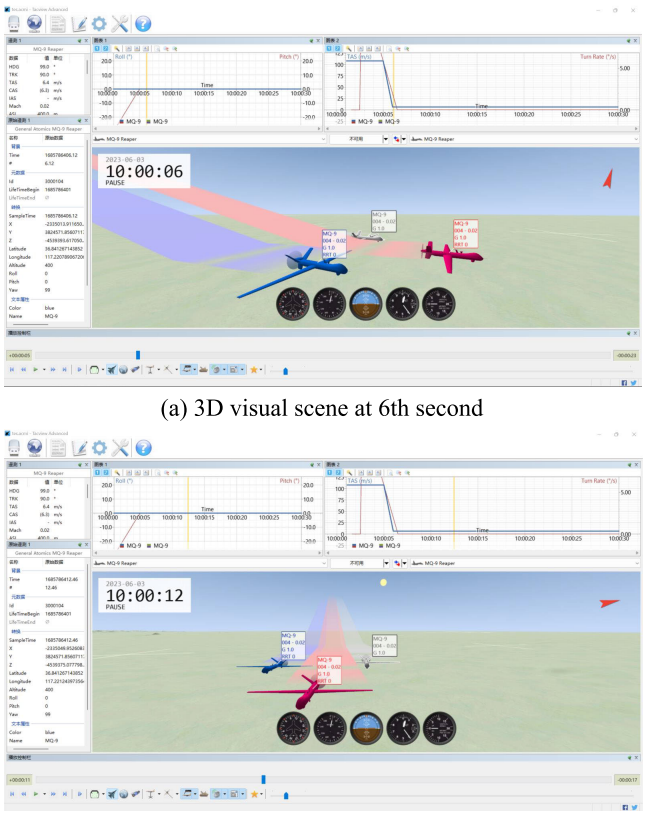


Fig. 5. Interactive relationship between functional units.

In the experiment test, the performance of our proposed control scheme is assessed through a low altitude penetration task for UAV formation reconfiguration. The planner path profile of references is expressed as $x_0(\theta_0) = 0$, $y_0(\theta_0) = 32\theta_0$, $z_0(\theta_0) = 500 - 20\theta_0 (t \le 5)$ and $z_0(\theta_0) = 400 (5 < t \le 14)$. The initial values of the path parameters are taken as $\theta_0 = 0$rad and $\dot{\theta}_0 = 1\text{rad}/\text{s}$. The initial airspeed, heading angle and pitch angle of each UAV are taken as $V_0 = 30\text{m/s}$, $\psi_0 = \pi/2\text{rad}$ and $\gamma_0 = -\arctan(2/3)\text{rad}$ respectively. The time constant of pitch

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) 3D visual scene at 6th second

(b) 3D visual scene at 12th second

Fig. 6. 3D visual scenes displayed by Tacview.

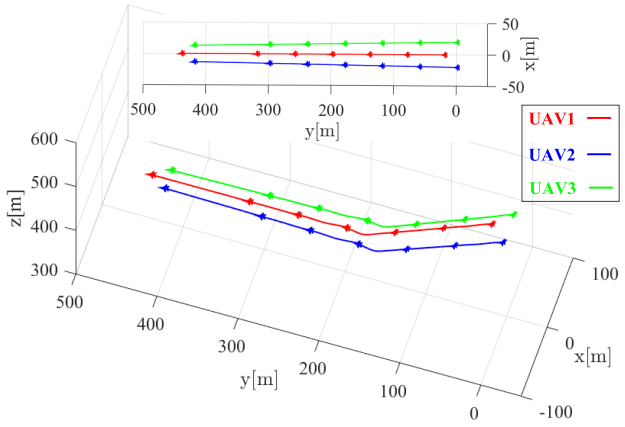


Fig. 7. Position trajectories of UAV formation.

dynamics is chosen by $\kappa = 1$. The desired relative positions of UAVs with respect to virtual leader and the preset topologies are listed in Table I. Moreover, the flight test is performed for the cases of complicated disturbances as follows.

$$\begin{aligned} \omega_x &= 5\sin(0.1t) + \xi_x \\ \omega_y &= 3\cos(0.1t) + \xi_y \\ \omega_z &= 3\cos(0.5t) + \xi_z \end{aligned} \tag{64}$$

where ξ_x, ξ_y and ξ_z represent random noise signal with a normal distribution.

TABLE I

FORMATION RECONFIGURATION AND SWITCHING TOPOLOGIES

	Δx(m)	Δy(m)	Δz(m)	$a_{ij}^{\sigma(t)}$
UAV1	0	$-2t$	0	$a_{12}^{\sigma(t)} = 1\ (t \in [0, 14]\,\text{s})$ $a_{13}^{\sigma(t)} = 1\ (t \in [0, 14]\,\text{s})$
UAV2	$-20 + 0.5t$	0	0	$a_{21}^{\sigma(t)} = 1\ (t \in [0, 14]\,\text{s})$ $a_{23}^{\sigma(t)} = \begin{cases} 1\ (t \in [0, 5]\,\text{s}) \\ 0\ (t \in (5, 14]\,\text{s}) \end{cases}$
UAV3	$20 - 0.5t$	0	0	$a_{31}^{\sigma(t)} = 1\ (t \in [0, 14]\,\text{s})$ $a_{32}^{\sigma(t)} = \begin{cases} 0\ (t \in [0, 5]\,\text{s}) \\ 1\ (t \in (5, 14]\,\text{s}) \end{cases}$

TABLE II

FORMATION KEEPING AND FIXED TOPOLOGIES

	Δx(m)	Δy(m)	Δz(m)	$a_{ij}^{\sigma(t)}$
UAV1	0	0	0	$a_{12}^{\sigma(t)} = 1\ (t \in [0, 12]\,\text{s})$ $a_{13}^{\sigma(t)} = 1\ (t \in [0, 12]\,\text{s})$
UAV2	-20	0	0	$a_{21}^{\sigma(t)} = 1\ (t \in [0, 12]\,\text{s})$ $a_{23}^{\sigma(t)} = 1\ (t \in [0, 12]\,\text{s})$
UAV3	20	0	0	$a_{31}^{\sigma(t)} = 1\ (t \in [0, 12]\,\text{s})$ $a_{32}^{\sigma(t)} = 1\ (t \in [0, 12]\,\text{s})$

Remark 5: Different from the constant disturbances in [42], [43] and low-frequency disturbance with small-value amplitude in [44] and [45], the assumed disturbances (71) takes into account the factors of electromagnetic interferences and gusts among disturbances, which integrates the characteristics of high-frequency vibration and large-value amplitude. This disturbance set fits in the complex flight environment.

The control parameters of proposed feedback controller are chosen as $k_{1i} = k_{2i} = k_{3i} = 90$, $i = 1, 2, 3$. For learning-based feedforward controller, we take $\boldsymbol{Q} = \boldsymbol{I}_6$, $\boldsymbol{R}_\tau = 10^8\boldsymbol{I}_6$ and $\boldsymbol{R}_d = 10^8\boldsymbol{I}_3$. For ESN, the reservoir is selected as $\boldsymbol{W}_2 \in \Re^{100\times 100}$, the internal unit function ϕ is defined as an identity function, and we select $\alpha = 10$ and $\beta = 10$. To relax the initial admissible control, the Lyapunov function L_{i3} is chosen as $L_{i3} = \mathfrak{F}_i^{\mathrm{T}}\boldsymbol{Q}_{\mathfrak{F}i}\mathfrak{F}_i$ with $\boldsymbol{Q}_{\mathfrak{F}i} = \boldsymbol{I}_6$, then it follows $\dot{L}_{i3} = 2\mathfrak{F}_i^{\mathrm{T}}\boldsymbol{Q}_{\mathfrak{F}i}(\boldsymbol{C}_i\hat{\boldsymbol{\tau}}_i^* + \boldsymbol{K}_i\hat{\boldsymbol{d}}_i)$. The initial output weights are set to be zero. The learning rates are selected by $\lambda_i = \sigma_i = 10$, $i = 1, 2, 3$.

The 3D visual scenes displayed by Tacview are presented in Fig. 6 (a)-(b). And the corresponding experimental results drawn by MATLAB are exhibited in Figs. 7-11. Fig. 7 shows the position trajectories of UAV formation, where the initial configuration is described by a loose V shape, and the final positions of UAV formation are realized by a close V one. Fig. 8 presents the state responses for each UAV during flight. And Fig. 9 describes the tracking errors for each UAV. It can be observed from Fig. 9 that cooperative formation tracking is well achieved by our proposed control algorithm. Taking UAV1 for example, Fig. 10 gives the control signals and the evolution of operator Γ_i is shown in Fig.11.

B. Comparison With State-of-the-Art by Simulation Validation

To highlight the disturbance-rejection ability of our proposed method, our proposed control algorithm is compared

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

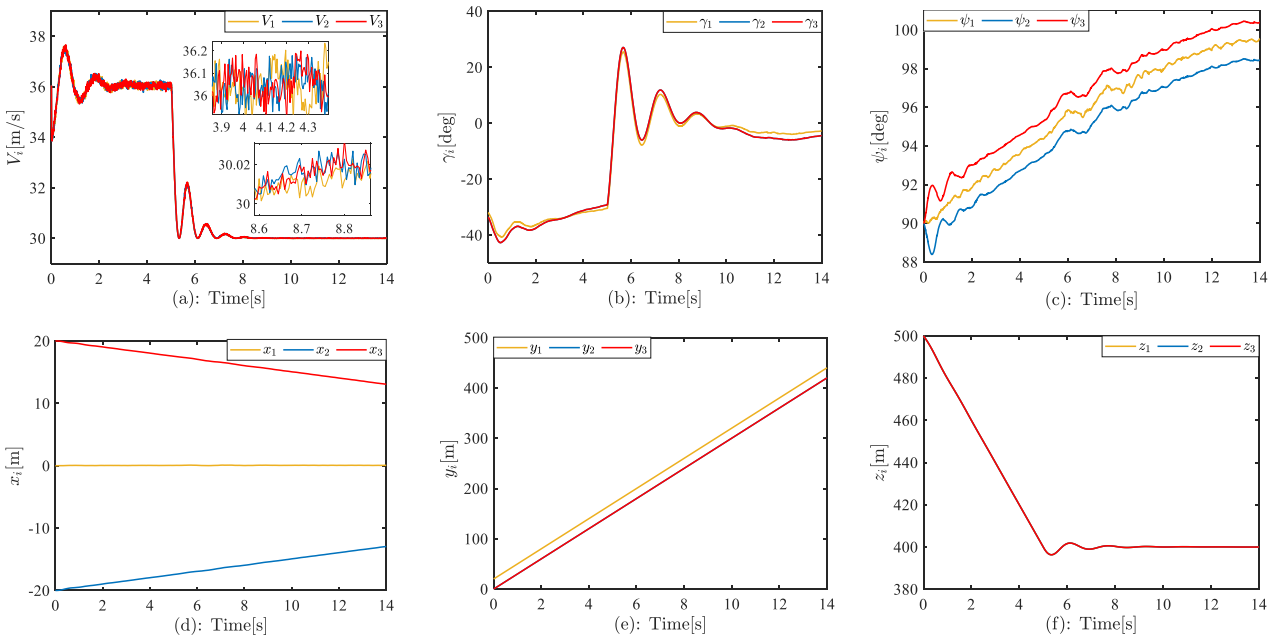


Fig. 8. Steady-state responses under proposed controller for low altitude penetration task: (a) V_i response (b) γ_i response (c) ψ_i response (d) x_i response (e) y_i response (f) z_i response.

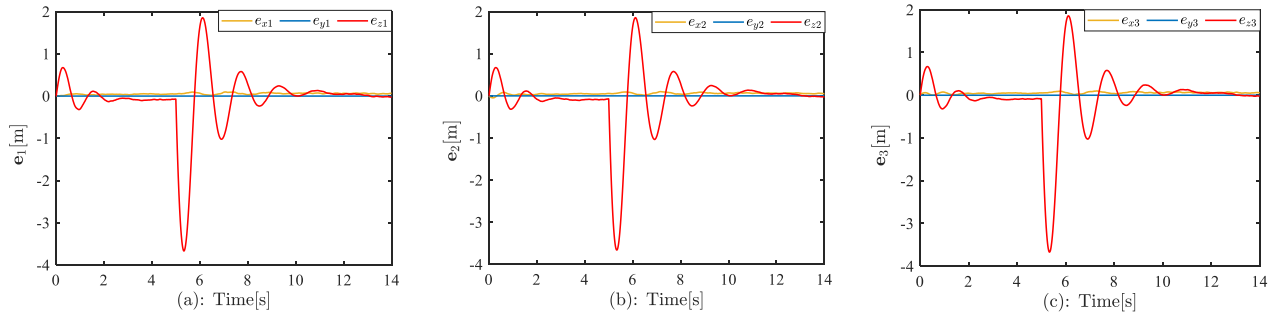


Fig. 9. Tracking errors under proposed controller for low altitude penetration task: (a) Tracking errors for UAV1 (b) Tracking errors for UAV2 (c) Tracking errors for UAV3.

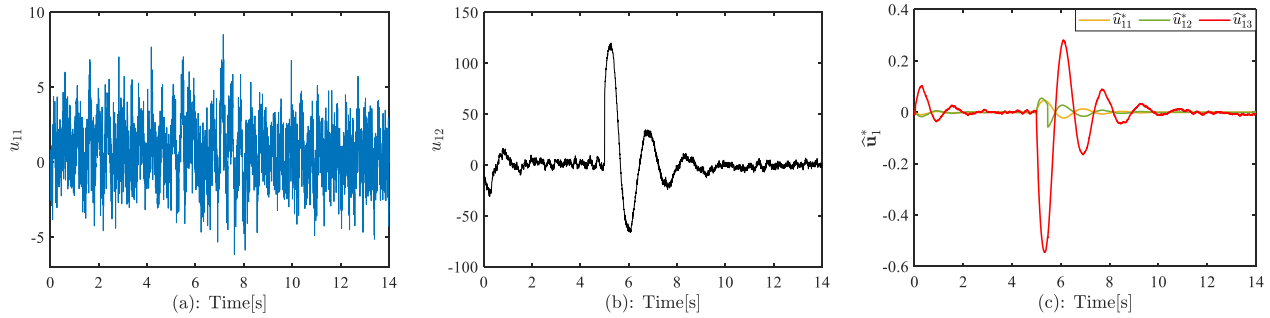


Fig. 10. Control actions for UAV1: (a) Control input u_{11} (b) Control input u_{12} (c) Estimated optimal control policy $\widehat{\boldsymbol{u}}_1^*$.

with state-of-the-art approaches [3], [34] by simulation validation. The first comparative method is taken from [3], where a conventional backstepping controller is considered. The second comparative approach is a disturbance observer based backstepping controller in [34].

In this section, the superiority of our proposed control scheme is emphasized through a coordinated turn with formation keeping. The path profile of virtual leader is expressed as $x_0(\theta_0) = 1000\cos\theta_0$, $y_0(\theta_0) = 1500\sin(2\theta_0)$ and $z_0(\theta_0) = 1000$. The initial values of path parameters are taken as $\theta_0 = 0$rad and $\dot{\theta}_0 = 0.01$rad$/$s. The initial states for each UAV are taken as $V_0 = 30$m/s, $\psi_0 = \pi/2$rad and $\gamma_0 = 0$rad. The time constant of pitch dynamics is chosen as $\kappa = 1$. The preset desired relative positions of UAVs 1-3 with respect to virtual leader and the fixed topologies are shown in Table II. The complicated disturbances ω_x, ω_y and ω_z are same as (64).

The control parameters of proposed feedback controller are chosen as $k_{1i} = 100$, $k_{2i} = 0.1$, $k_{3i} = 100$, $i = 1, 2, 3$.

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

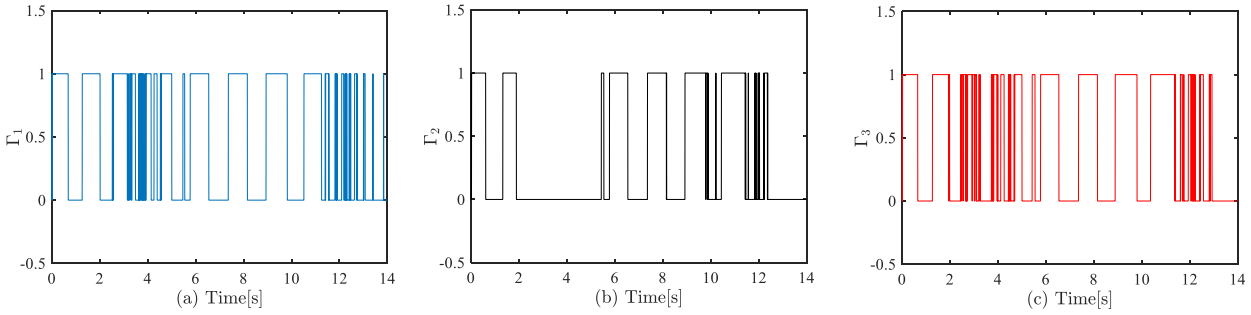


Fig. 11. Evolution of the operators Γ_i: (a) Γ_1 (b) Γ_2 (c) Γ_3.

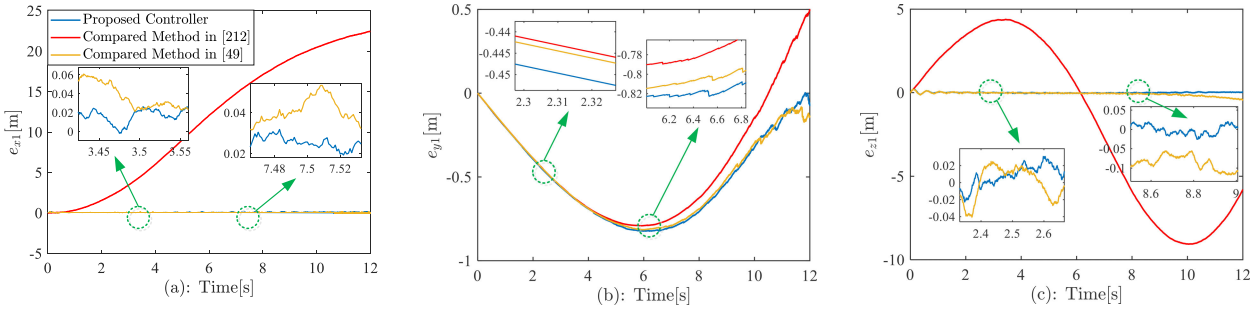


Fig. 12. Tracking errors under the proposed controller and comparative controllers. (a) e_{x1} response (b) e_{y1} response (c) e_{z1} response.

TABLE III
TRANSIENT AND STEADY-STATE PERFORMANCES UNDER DIFFERENT CONTROLLERS

	Controller	Maximum error fluctuations(m)	Offset error(m)	Integral of absolute error(m.s)
e_{x1}	**Proposed Controller**	**0.0836**	**0.0496**	**0.3773**
	Compared method in [3]	22.4448	24.4430	136.4567
	Compared method in [34]	0.7999	0.277014	0.4535
e_{y1}	**Proposed Controller**	**0.8233**	**-0.0724**	**5.9935**
	Compared method in [3]	0.7916	0.4951	5.7112
	Compared method in [34]	0.8148	-0.1372	5.9513
e_{z1}	**Proposed Controller**	**0.2035**	**0.0294**	**0.2813**
	Compared method in [3]	9.0657	-5.7929	54.5148
	Compared method in [34]	0.4535	-0.4531	0.9986

For learning-based feedforward controller, we take $\boldsymbol{Q} = \boldsymbol{I}_6$, $\boldsymbol{R}_\tau = \boldsymbol{I}_6$ and $\boldsymbol{R}_d = \boldsymbol{I}_3$. For ESN, the reservoir is selected as $\boldsymbol{W}_2 \in \Re^{1000\times 1000}$, the internal unit function ϕ is defined as an identity function, and we select $\alpha = 10$ and $\beta = 10$. To relax the initial admissible control, the Lyapunov function L_{i3} is chosen as $L_{i3} = \mathfrak{F}_i^{\mathrm{T}} \boldsymbol{Q}_{\mathfrak{F}i} \mathfrak{F}_i$ with $\boldsymbol{Q}_{\mathfrak{F}i} = \boldsymbol{I}_6$, then one has $\dot{L}_{i3} = 2\mathfrak{F}_i^{\mathrm{T}} \boldsymbol{Q}_{\mathfrak{F}i}(\boldsymbol{C}_i \hat{\boldsymbol{\tau}}_i^* + \boldsymbol{K}_i \dot{\boldsymbol{d}}_i)$. The initial output weights are set to be zero. The learning rates are selected by $\lambda_i = \sigma_i = 10$, $i = 1, 2, 3$.

Taking the UAV1 for example, Fig. 8 shows the curves of formation tracking errors under three different controllers. It is observed from Fig. 8 that with the help of feedback compensation of the learning-based algorithm, our proposed method exhibits a stronger disturbance-rejection ability in the presence of complicated disturbances. To visually evaluate the tracking performances of different controllers, the indices of the maximum error fluctuations, offset errors and integral of absolute errors are listed Table III, which validates the superiority of our proposed control scheme.

VII. CONCLUSION

This paper has presented a feedforward-feedback learning-based optimal control scheme to provide the cooperative UAV formation tracking in the presence of complicated disturbances. A two-player zero-sum game framework is designed, where the critic ESN is derived to approximate the optimal feedforward policies. To remove the PE condition and the requirement of initial admissible control in weight tuning law, appropriate compensation terms and a new Lyapunov function have been introduced into adaptive tuning laws. Simulation results have validated the effectiveness and superiority of our proposed control algorithm over the state-of-the-art approaches. One of future researches would be the extension of our control design to complex networked systems.

REFERENCES

[1] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Wireless communication using unmanned aerial vehicles (UAVs): Optimal transport theory for hover time optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8052–8066, Dec. 2017.

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[2] A. Brown and D. Anderson, "Trajectory optimization for high-altitude long-endurance UAV maritime radar surveillance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 2406–2421, Jun. 2020.

[3] X. Dong, Y. Li, C. Lu, G. Hu, Q. Li, and Z. Ren, "Time-varying formation tracking for UAV swarm systems with switching directed topologies," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3674–3685, Dec. 2019.

[4] S. Rao and D. Ghose, "Sliding mode control-based autopilots for leaderless consensus of unmanned aerial vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 5, pp. 1964–1972, Sep. 2014.

[5] Y. Kang et al., "Robust leaderless time-varying formation control for unmanned aerial vehicle swarm system with Lipschitz nonlinear dynamics and directed switching topologies," *Chin. J. Aeronaut.*, vol. 35, no. 1, pp. 124–136, Jan. 2022.

[6] M. Lv, C. K. Ahn, B. Zhang, and A. Fu, "Fixed-time anti-saturation cooperative control for networked fixed-wing unmanned aerial vehicles considering actuator failures," *IEEE Trans. Aerosp. Electron. Syst.*, early access, Sep. 24, 2023, doi: 10.1109/TAES.2023.3311420.

[7] Y. Zou, Z. Zhou, X. Dong, and Z. Meng, "Distributed formation control for multiple vertical takeoff and landing UAVs with switching topologies," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 4, pp. 1750–1761, Aug. 2018.

[8] X. Dong, B. Yu, Z. Shi, and Y. Zhong, "Time-varying formation control for unmanned aerial vehicles: Theories and applications," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 1, pp. 340–348, Jan. 2015.

[9] X. Dong, Y. Zhou, Z. Ren, and Y. Zhong, "Time-varying formation tracking for second-order multi-agent systems subjected to switching topologies with application to quadrotor formation flying," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 5014–5024, Jun. 2017.

[10] M. Lv, B. De Schutter, and S. Baldi, "Nonrecursive control for formation-containment of HFV swarms with dynamic event-triggered communication," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 3188–3197, Mar. 2023.

[11] M. Lv, Z. Chen, B. De Schutter, and S. Baldi, "Prescribed-performance tracking for high-power nonlinear dynamics with time-varying unknown control coefficients," *Automatica*, vol. 146, Dec. 2022, Art. no. 110584.

[12] Y. Ke, K. Wang, and B. M. Chen, "Design and implementation of a hybrid UAV with model-based flight capabilities," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 3, pp. 1114–1125, Jun. 2018.

[13] P. Hemakumara and S. Sukkarieh, "Learning UAV stability and control derivatives using Gaussian processes," *IEEE Trans. Robot.*, vol. 29, no. 4, pp. 813–824, Aug. 2013.

[14] F. Santoso, M. A. Garratt, and S. G. Anavatti, "State-of-the-art intelligent flight control systems in unmanned aerial vehicles," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 2, pp. 613–627, Apr. 2018.

[15] Z. Zhen, Y. Chen, L. Wen, and B. Han, "An intelligent cooperative mission planning scheme of UAV swarm in uncertain dynamic environment," *Aerosp. Sci. Technol.*, vol. 100, May 2020, Art. no. 105826.

[16] L. He, N. Aouf, and B. Song, "Explainable deep reinforcement learning for UAV autonomous path planning," *Aerosp. Sci. Technol.*, vol. 118, Nov. 2021, Art. no. 107052.

[17] C. Wang, J. Wang, Y. Shen, and X. Zhang, "Autonomous navigation of UAVs in large-scale complex environments: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2124–2136, Mar. 2019.

[18] Y. Fu and T. Chai, "Online solution of two-player zero-sum games for continuous-time nonlinear systems with completely unknown dynamics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2577–2587, Dec. 2016.

[19] J. Zhao, Y. Lv, and Z. Zhao, "Adaptive learning based output-feedback optimal control of CT two-player zero-sum games," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1437–1441, Mar. 2022.

[20] H. Li, D. Liu, and D. Wang, "Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 3, pp. 706–714, Jul. 2014.

[21] S. Mehraeen, T. Dierks, S. Jagannathan, and M. L. Crow, "Zero-sum two-player game theoretic formulation of affine nonlinear discrete-time systems using neural networks," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1641–1655, Dec. 2013.

[22] V. Saxena, J. Jaldén, and H. Klessig, "Optimal UAV base station trajectories using flow-level models for reinforcement learning," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 4, pp. 1101–1112, Dec. 2019.

[23] X. Liu, Y. Liu, and Y. Chen, "Reinforcement learning in multiple-UAV networks: Deployment and movement design," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8036–8049, Aug. 2019.

[24] T. Guo, N. Jiang, B. Li, X. Zhu, Y. Wang, and W. Du, "UAV navigation in high dynamic environments: A deep reinforcement learning approach," *Chin. J. Aeronaut.*, vol. 34, no. 2, pp. 479–489, Feb. 2021.

[25] H. Zargarzadeh, T. Dierks, and S. Jagannathan, "Optimal control of nonlinear continuous-time systems in strict-feedback form," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2535–2549, Oct. 2015.

[26] S. Rad-Moghadam and M. Farrokhi, "Optimal output feedback control of a class of uncertain systems with input constraints using parallel feedforward compensator," *J. Franklin Inst.*, vol. 357, no. 18, pp. 13449–13476, Dec. 2020.

[27] B. Zhao, D. Liu, and C. Luo, "Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4330–4340, Oct. 2020.

[28] N. Wang, Y. Gao, H. Zhao, and C. K. Ahn, "Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3034–3045, Jul. 2021.

[29] X. Yang and H. He, "Decentralized event-triggered control for a class of nonlinear-interconnected systems using reinforcement learning," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 635–648, Feb. 2021.

[30] L. Wang, K. Wang, C. Pan, W. Xu, N. Aslam, and L. Hanzo, "Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 1, pp. 73–84, Mar. 2021.

[31] A. Singla, S. Padakandla, and S. Bhatnagar, "Memory-based deep reinforcement learning for obstacle avoidance in UAV with limited environment knowledge," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 107–118, Jan. 2021.

[32] J. Cui, Y. Liu, and A. Nallanathan, "Multi-agent reinforcement learning-based resource allocation for UAV networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 729–743, Feb. 2020.

[33] J. Yang, C. Liu, M. Coombes, Y. Yan, and W.-H. Chen, "Optimal path following for small fixed-wing UAVs under wind disturbances," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 3, pp. 996–1008, May 2021.

[34] X. Yu, J. Yang, and S. Li, "Finite-time path following control for small-scale fixed-wing UAVs under wind disturbances," *J. Franklin Inst.*, vol. 357, no. 12, pp. 7879–7903, Aug. 2020.

[35] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, "Optimization and applications of echo state networks with leaky- integrator neurons," *Neural Netw.*, vol. 20, no. 3, pp. 335–352, Apr. 2007.

[36] J. M. Keller, D. Liu, and D. B. Fogel, *Recurrent Neural Networks*. Hoboken, NJ, USA: Wiley, 2016.

[37] Z. Shi and M. Han, "Support vector echo-state machine for chaotic time-series prediction," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 359–372, Mar. 2007.

[38] 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.

[39] J. Qin, M. Li, Y. Shi, Q. Ma, and W. X. Zheng, "Optimal synchronization control of multiagent systems with input saturation via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 85–96, Jan. 2019.

[40] X. Yang and H. He, "Adaptive critic learning and experience replay for decentralized event-triggered control of nonlinear interconnected systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4043–4055, Nov. 2020.

[41] H. Zhang, L. Cui, and Y. Luo, "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 206–216, Feb. 2013.

[42] Z. Yu, Y. Qu, and Y. Zhang, "Distributed fault-tolerant cooperative control for multi-UAVs under actuator fault and input saturation," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2417–2429, Nov. 2019.

[43] K. Klausen, C. Meissen, T. I. Fossen, M. Arcak, and T. A. Johansen, "Cooperative control for multirotors transporting an unknown suspended load under environmental disturbances," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 2, pp. 653–660, Mar. 2020.

[44] S. Shao, M. Chen, and Y. Zhang, "Adaptive discrete-time flight control using disturbance observer and neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3708–3721, Dec. 2019.

[45] Y. Chen, R. Yu, Y. Zhang, and C. Liu, "Circular formation flight control for unmanned aerial vehicles with directed network and external disturbance," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 2, pp. 505–516, Mar. 2020.

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Boyang Zhang received the Ph.D. degree from the Department of Equipment Management and Unmanned Aerial Vehicle Engineering, Air Force Engineering University, Xi'an, China, in 2022. He is currently with the Beijing Blue Sky Science and Technology Innovation Center. His research interests include adaptive learning control, distributed control, reinforcement learning, and intelligent decision-making, with applications in multi-agent systems, hypersonic vehicles, and unmanned autonomous systems.

Maolong Lv received the Ph.D. degree from the Delft Center for Systems and Control, Delft University of Technology, The Netherlands, in 2021. He is currently with Air Force Engineering University. His research interests include adaptive learning control, distributed control, reinforcement learning, and intelligent decision-making, with applications in multi-agent systems, hypersonic vehicles, unmanned autonomous systems.

He received the Descartes Excellence Fellowship from the French Government in 2018, which allowed him a research visit with the University of Grenoble, from 2018 to 2019, working on adaptive networked systems, with an emphasis on traffic with human driven and autonomous vehicles. He also received the Young Talent Support Project for Military Science and Technology, the Young Talent Fund of Association for Science and Technology in Shaaxi, and the Post-Doctoral International Exchange Program in 2022. He is currently an Editor of *Aerospace* and *Measurement and Control*.

Shaohua Cui received the M.S. degree from the School of Traffic and Transportation, Systems Science Institute, Beijing Jiaotong University. He is currently pursuing the Ph.D. degree with the School of Transportation Science and Engineering, Beihang University. He has focused on traffic flow analysis, vehicle control, adaptive control, robust control, and non-smooth nonlinearities.

Xiangwei Bu received the B.S., M.S., and Ph.D. degrees from Air Force Engineering University, Xi'an, China, in 2010, 2012, and 2016, respectively.

He is currently an Associate Professor with the Air and Missile Defense College, Air Force Engineering University, and also with the School of Astronautics, Northwestern Polytechnical University. His research interests include advanced control theory and its applications. He is an Editorial Board Member of *Advances in Mechanical Engineering* and *Measurement and Control*.

Ju H. Park (Senior Member, IEEE) received the Ph.D. degree in electronics and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea, in 1997. He joined Yeungnam University, Kyongsan, Republic of Korea, in March 2000, where he is currently the Chuma Chair Professor. His research interests include control engineering, neural/complex networks, and fuzzy systems.

He is a fellow of the Korean Academy of Science and Technology (KAST). Since 2015, he has been a recipient of the Highly Cited Researchers Award by Clarivate Analytics and listed in three fields, engineering, computer sciences, and mathematics, from 2019 to 2022. He is an Associate Editor of some journals, including IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and *Nonlinear Dynamics*.

Authorized licensed use limited to: China University of Petroleum. Downloaded on August 02,2024 at 02:29:55 UTC from IEEE Xplore. Restrictions apply.