Reach-Avoid Differential Graphical Games for Single Evader and Multiple Pursuers with Nonlinear Dynamics

Rugang Tang, Chengfeng Luo, Tianqi Wang, Xin Ning, Chih-Yung Wen

***Abstract*—This paper investigates the single-evader and multi-pursuer (SEMP) reach-avoid differential graphical (RADG) games of nonlinear heterogeneous players subject to saturated input and limited communication channels. The evader's objective is to reach a designated target area while avoiding the pursuers, whose aim is to intercept the evader. First, we reformulate the SEMP RADG game as an optimal control problem within a weighted communication topology graph by incorporating the interception and control errors. Then, optimal control strategies that account for input saturation are derived by solving the coupled Hamilton-Jacobi (HJ) equations. These strategies are shown to constitute the Nash Equilibrium (NE) of the game. In addition, four types of pursuers, namely, isolated, passive, invisible, and regular pursuers, are defined based on the communication topology, and the conditions for achieving an Interactive Nash Equilibrium (INE), which is proposed for SEMP RADG games, are analyzed. Moreover, a single-network approximate dynamic programming (ADP) algorithm using concurrent learning (CL) is proposed to provide the near-optimal solutions to the coupled HJ equations. Asymptotic capture conditions are established through an examination of equilibrium points, and extensions to general pursuit-evasion (PE) games and half-space targets are further discussed. Our results are validated through numerical simulations.**

***Note to Practitioners*—This paper addresses the practical challenges encountered in coordinating multiple autonomous agents, such as unmanned aerial vehicles or ground robots, in pursuit-evasion scenarios characterized by nonlinear dynamics and limited communication. Unlike traditional approaches that often rely on ideal communication assumptions or consider linear agent models, the proposed framework enables each agent to make distributed decisions based on local information and restricted communication channels. Practitioners can utilize the developed real-time single-network algorithm to design near-optimal interception and evasion strategies, even under stringent communication limitations and actuator constraints. Future research will extend this framework to address scenarios involving multiple evaders operating in more complex and dynamic environments.**

***Index Terms*—Multi-agent systems (MAS), Nonlinear systems, Approximate dynamic programming (ADP), Reach-Avoid Differential game (RADG), Interactive Nash Equilibrium (INE)**

I. Introduction

Pursuit-evasion (PE) differential games have attracted considerable attention in the field of multi-agent systems (MAS) due to their extensive applications in security, collision avoidance, defense, transportation, and strategic scenarios [1]–[5]. In PE differential games, evaders strive to avoid capture by a group of pursuers, whose primary objective is to capture the evaders. The strategies adopted by both parties are determined by their dynamic behaviors, communication constraints, and performance capabilities. As a special branch of PE games, the Reach-avoid (RA) differential game further complicates the problem by requiring the evader to reach a designated target area while simultaneously avoiding interception [6]–[10].

The prevalent approach addressing PE games and RA games involves formulating the game scenarios as optimal control problems, including identifying the optimal strategies that minimize the cost function while adhering to the system dynamics and constraints [11]. Such optimal control problems can be solved through the application of Hamilton-Jacobi (HJ) equations [12]–[14]. In [15], a distributed control strategy is developed to tackle multi-agent pursuit-evasion (MPE) games, and the conditions for finite-time capture are provided. In addition, the presence of malicious pursuers is modeled using subjectively malicious factors, and the impact of these malicious pursuers on capture time is explicitly quantified [16]. Nevertheless, these studies primarily focus on PE games with linear dynamics, and the extension to RA games with nonlinear dynamics remains an open challenge. On the other hand, geometric tools are quite effective in the analysis of RA games, as demonstrated in several studies [17]–[20]. Reference [21] examines a multiplayer reach-avoid game in which a team of pursuers endeavors to safeguard a target region from an evader within a constant flow field. Even though this analytical method offers the advantage of circumventing the substantial computational demands associated with high-dimensional spaces, it encounters challenges when considering complex system dynamics.

The complexity of RA games is further increased by the presence of multiple pursuers and/or evaders. In this scenario, the interactions among different agents are influenced by the communication topology, which dictates the information exchange between agents. Graph theory has been extensively employed to model these interactions, with agents represented as nodes and their interactions captured by edges [22]. Nevertheless, the application of graph theory has predominantly focused on the cooperative control problems of multi-agent systems (MAS) [23]–[26] and general PE games [15], [16], [27], [28], with limited attention paid to RA games. The major difference between the cooperative control problems and PE or RA games is that the cooperative control problems aim to drive all agents to achieve some common objectives or behaviors, whereas in RA games, pursuers and evaders have distinct objectives and strategies.

In summary, the existing literature on RA games has predominantly focused on simplified dynamics, often neglecting the complexities introduced by nonlinear dynamics. Additionally, the influence of communication constraints on the strategic interactions among agents has not been comprehensively examined. To address these gaps, this paper proposes a unified framework for single-evader and multi-pursuer reach-avoid

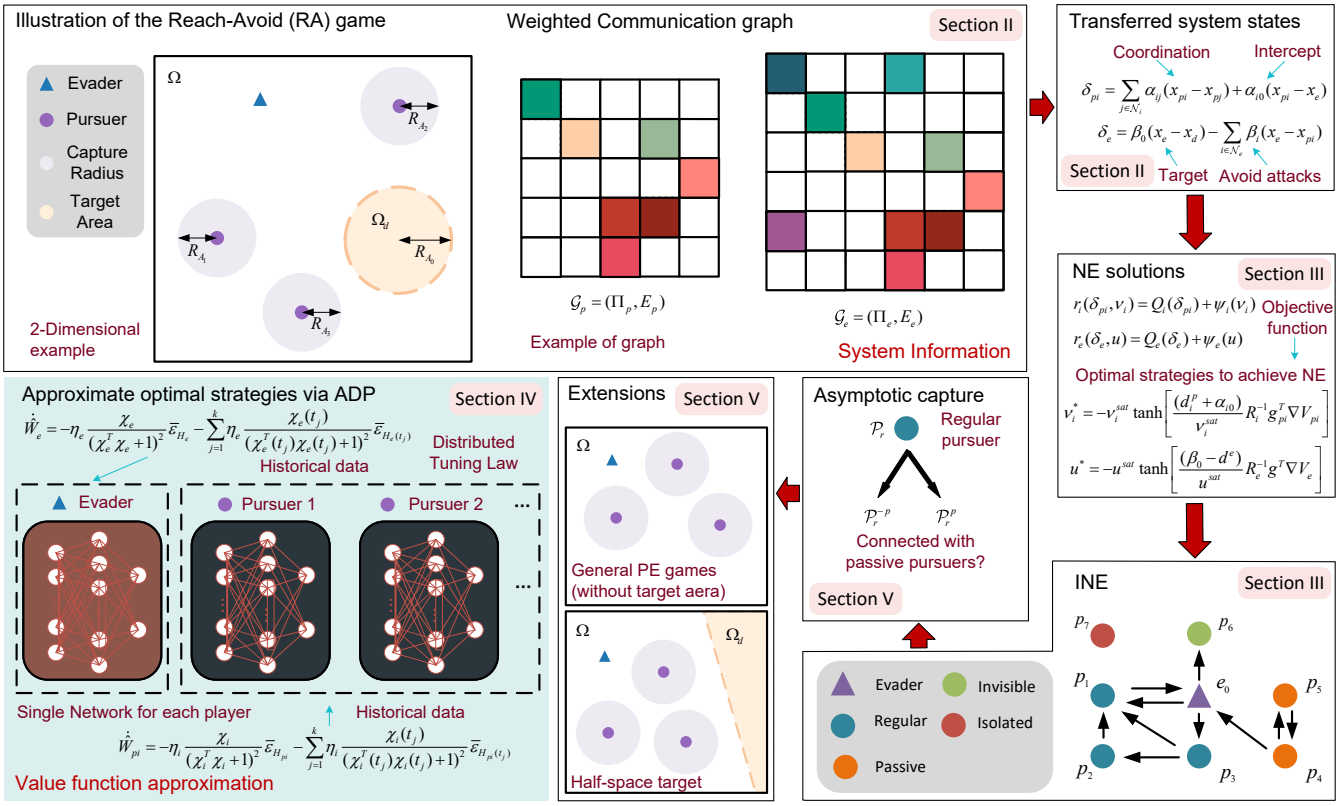

Fig. 1: Overall framework of SEMP RADG with NN approximation.

differential graphical games characterized by nonlinear dynamics and limited communication. The primary contribution of this paper is summarized as follows:

- Single-evader and multi-pursuer reach-avoid differential games are reformulated as an optimal control problem that incorporates both interception and control errors within two weighted communication graphs. In addition, in contrast to the previous studies in [7], [8], [21], our work takes into account the communication topology and more complex dynamics of the agents. Furthermore, unlike in [10], [12], [15], [16], this study considers heterogeneous nonlinear systems with input saturation.
- Optimal strategies considering input saturation are derived from coupled Hamilton-Jacobi (HJ) equations and proven to constitute a Nash Equilibrium (NE). A stricter Interactive Nash Equilibrium (INE) concept is introduced to diagnose the NE's quality by identifying whether the equilibrium is fully interactive or exhibits information asymmetries due to communication constraints, a departure from standard analyses [12], [15], [16]. A computationally efficient single-network ADP algorithm with concurrent learning (CL) is proposed to solve the HJ equations, and its convergence to near-optimal solutions for nonlinear RADG games is rigorously established.
- Conditions for asymptotic capture are derived through a detailed analysis of the equilibrium points, which offers deeper insights into the strategic interactions between the evader and the pursuers. Our proposed method is subsequently extended to general pursuit-evasion games, as discussed in [15], [16], [29], through the design of a virtual target point. Additionally, scenarios involving a half-space target area [10], [30], [31], as opposed to a spherical target area, are investigated, thereby broadening the framework's applicability to a diverse array of situations.

The remainder of this paper is organized as follows, with the overall framework illustrated in Fig. 1. Section II formulates the SEMP RADG game of nonlinear heterogeneous systems over communication networks. Section III derives the optimal strategies for the evader and pursuers and establishes the conditions for Nash Equilibrium (NE) and Interactive Nash Equilibrium (INE). In Section IV, a single-network ADP algorithm is presented to obtain approximate solutions to the coupled HJ equations. Section V analyzes the asymptotic capture conditions and discusses extensions to general PE games and half-space targets. Numerical simulations validating the proposed approach are provided in Section VI, followed by concluding remarks in Section VII.

Notations: $\mathbb{R}^n$ denotes the n-dimensional real Euclidean space. $\mathbb{R}^{m\times n}$ represents the set of all $m\times n$ real matrices. The Euclidean norm of a vector is denoted by $\|\cdot\|$. The symbol $\mathbf{1}_n$ represents an $n\times 1$ column vector with all elements equal to 1. The notation $[a_{ij}]^{m\times n}$ refers to an $m\times n$ matrix where a_{ij} is the element in the i-th row and j-th column. The operator $\mathrm{diag}\{\cdots\}$ is used to construct a diagonal or block-diagonal matrix. For a vector $\mathbf{v}=[v_1,v_2,\cdots,v_n]^T$, the element-wise square is denoted as $\mathbf{v}^2=[v_1^2,v_2^2,\cdots,v_n^2]^T$, and for an operator ϕ, such as tanh or ln, the element-wise

application is expressed as $\phi(\mathbf{v}) = [\phi(v_1), \phi(v_2), \cdots, \phi(v_n)]^T$. The symbols $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of a matrix, respectively. For a matrix $A \in \mathbb{R}^{m\times n}$, the submatrix $A_{P,Q} \in \mathbb{R}^{(m-p)\times(n-q)}$ is obtained by selecting rows indexed by $P = \{a_1, a_2, \cdots, a_p\}$ and columns indexed by $Q = \{b_1, b_2, \cdots, b_q\}$.

II. Problem Formulation and Preliminaries

A. RA games with nonlinear Dynamics

Consider an RA game with a single evader and multiple pursuers taking place in $\Omega \subseteq \mathbb{R}^n$. The target area $\Omega_d(x_d, R_{A_0}) = \{x \in \mathbb{R} \mid \|x - x_d\| \leq R_{A_0}\}$ is a subset of Ω with x_d the center and R_{A_0} as the radius. The evader e_0 is governed by the following nonlinear differential equation:

$$\dot{x}_e(t) = f(x_e(t)) + g(x_e(t)) \cdot u(t),\ x_e(0) = x_e^0, \tag{1}$$

where $x_e \in \mathbb{R}^n$ is the state vector of the evading agent, $u = [u^1, u^2, \cdots, u^m]^T : \mathbb{R}_{t\geq t_0} \to \mathbb{R}^m$ is the input vector satisfying $-u^{sat} \leq u^k \leq u^{sat}$ with $u^{sat} > 0$ for all $k \in \{1, ..., m\}$, $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ with $g \neq 0$ are nonlinear smooth functions.

Meanwhile, there are q malicious attackers in Ω, labeled as p_1 to p_q. The dynamics of the i-th pursuer p_i, $i \in \mathscr{P} \triangleq \{1, 2, \cdots, q\}$, is given by

$$\dot{x}_{pi}(t) = f_{pi}(x_{pi}(t)) + g_{pi}(x_{pi}(t)) \cdot v_i(t),\ x_{pi}(0) = x_{pi}^0, \tag{2}$$

where $x_{pi} \in \mathbb{R}^n$ is the state vector of agent p_i, $v_i = [v_i^1, v_i^2, \cdots, v_i^m]^T : \mathbb{R}_{t\geq t_0} \to \mathbb{R}^m$ is the input vector satisfying $-v_i^{sat} \leq v_i^k \leq v_i^{sat}$ with v_i^{sat} for all $k \in \{1, ..., m\}$, $f_{pi} : \mathbb{R}^n \to \mathbb{R}^n$ and $g_{pi} : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ with $g_{pi} \neq 0$ are nonlinear smooth functions. The capture radius of p_i is denoted by R_{A_i}. For simplicity of expression, t will be omitted in the following derivation if not necessary.

An example of the RA game in a 2-dimensional space is shown in Fig. 1. There are three pursuers with capture radii R_{A_1}, R_{A_2}, and R_{A_3}, and one evader in the space. The radius of the target area is R_{A_0}. The evader aims to reach the target area while avoiding being captured by the pursuer, while the pursuer's objective is to intercept and capture the evader.

The PE game terminates under one of the following three conditions: (i) Evader is captured by one or more pursuers; (ii) Evader runs out of the game space Ω; (iii) Evader reaches the target area Ω_d. The pursuers win in conditions (i) and (iii), and the evader wins in condition (iii). If conditions (i) and (iii) are satisfied simultaneously, i.e., the evader is captured by the pursuer at the boundary of Ω_d, then the evader also wins. Define $X = [x_e^T, x_{p1}^T, \cdots, x_{pq}^T]^T$ as the global states of the game, the termination set is

$$\mathscr{T} := \mathscr{T}_e \bigcup \mathscr{T}_p \tag{3}$$

where

$$\mathscr{T}_e := \{X \mid \|x_e - x_d\|_2 \leq R_{A_0}\} \tag{4}$$

represents the success set of evader where it reaches the target area, and

$$\mathscr{T}_p := \mathscr{T}_p^o \bigcup \mathscr{T}_p^c \tag{5}$$

is the success set of pursuers, which contains the subset

$$\mathscr{T}_p^o := \{X \mid x_e \notin \Omega\}, \tag{6}$$

i.e., the evader is out of Ω, and the subset

$$\mathscr{T}_p^c := \{X \mid \|x_{pi} - x_e\|_2 \leq R_{A_i},\ for\ any\ i = 1, 2, \cdots, q\}, \tag{7}$$

which indicates that the evader is captured by the pursuers.

B. Communication Graphs

A weighted directed graph $\mathscr{G}_p = (\Pi_p, E_p)$ is defined to describe the communication between the pursuers, where $\Pi_p = \{p_1, p_2, \cdots, p_q\}$ is the set of nodes and $E_p \subseteq \Pi_p \times \Pi_p$ is the set of weighted edges. The edge from node p_j to node p_i is denoted as $(p_j, p_i), \forall p_i, p_j \in \Pi_p$, representing that the pursuer i can receive the information from pursuer j. The weight $\alpha_{ij} > 0$ if $(p_j, p_i) \in E_p$; otherwise, $\alpha_{ij} = 0$. The in-degree matrix is defined as a diagonal matrix $\mathscr{D}_p = \text{diag}\{d_1^p, d_2^p, \cdots, d_q^p\}$ where $d_i^p = \sum_{j=1}^q \alpha_{ij}$, and the weighted adjacency matrix is denoted as $\mathscr{A}_p = [\alpha_{ij}]^{q\times q}$. The neighbor set of p_i is represented as $\mathscr{N}_i = \{j \mid (p_j, p_i) \in E_p\}$.

In addition, an extended directed graph $\mathscr{G}_e = (\Pi_e, E_e)$ is established by incorporating the node e_0 with $\Pi_e = \Pi_p \cup \{e_0\}$ and $E_e \subseteq \Pi_e \times \Pi_e$. The edge from node p_i to node e_0 is denoted as $(p_i, e_0), \forall p_i \in \Pi_p$, representing that the evader can access pursuer i's information. The weight $\beta_i > 0$ if $(p_i, e_0) \in E_e$; otherwise, $\beta_i = 0$. Similarly, the edge from node e_0 to node p_i is denoted as (e_0, p_i). $\alpha_{i0} > 0$ if $(e_0, p_i) \in E_e$; otherwise, $\alpha_{i0} = 0$. Define the in-degree of the evader as $d^e = \sum_{i=1}^q \beta_i$, and the adjacency vector of the evader to pursuers as $\mathscr{A}_{pe} = [\alpha_{10}, \alpha_{20}, \cdots, \alpha_{q0}]^T$. The neighbor set of e_0 is denoted as $\mathscr{N}_e = \{i \mid (p_i, e_0) \in E_e\}$. It is assumed that there are no repeated edges or self-loops, i.e., $(p_i, p_i) \notin E_e, \forall p_i \in \Pi_p$ and $(e_0, e_0) \notin E_e$. A sequence of edges $\mathscr{S} = \{(p_{s0}, p_{s1}), (p_{s1}, p_{s_2}), ..., (p_{s(r-1)}, p_{sr})\}$ is called a directed path from p_{s0} to p_{sr} if there exists a directed edge $(p_{si}, p_{s(i+1)}) \in E_e$ for each $i = 0, 1, \cdots, r-1$ (here, p_{si} can represent the evader e_0). If such a path exists, we assert that node p_{s0} is connected to node p_{sr}, or node p_{sr} is reachable from node p_{s0}. A directed graph is said to be strongly connected if there exists a directed path from every node to every other node in Π_e.

C. RADG games with graphs

In RA games, the objective of the pursuers is to intercept the evader while simultaneously moving closer to the other pursuers to achieve coordinated interception. Hence, the interception error of p_i, considering coordination, is defined as

$$\delta_{pi} = \sum_{j\in\mathscr{N}_i} \alpha_{ij}(x_{pi} - x_{pj}) + \alpha_{i0}(x_{pi} - x_e). \tag{8}$$

It can be easily derived using (1)-(2) that δ_{pi} is governed by

$$\begin{aligned}\dot{\delta}_{pi} =& (d_i^p + \alpha_{i0})(f_{pi}(x_{pi}) + g_{pi}(x_{pi}) \cdot v_i) - \alpha_{i0}(f(x_e) + g(x_e) \cdot u) \\ &- \sum_{j\in\mathscr{N}_i} \alpha_{ij}\left(f_{pj}(x_{pj}) + g_{pj}(x_{pj}) \cdot v_j\right).\end{aligned} \tag{9}$$

On the other hand, the evading agent aims to get closer to the target area while avoiding the attacks from the pursuers.

Hence, we define the control error of e_0 considering avoid distance as

$$\delta_e = \beta_0(x_e - x_d) - \sum_{i\in\mathcal{N}_e} \beta_i(x_e - x_{pi}), \tag{10}$$

where β_0 is a user-defined weight representing the evader's tendency to move toward the target area. A detailed analysis of β_0 is provided in Section V-B. The dynamics of e_0 is

$$\begin{aligned}\dot{\delta}_e =&(\beta_0 - d^e)(f(x_e) + g(x_e)\cdot u)\\ &+ \sum_{i\in\mathcal{N}_e} \beta_i(f_{pi}(x_{pi}) + g_{pi}(x_{pi})\cdot v_i).\end{aligned} \tag{11}$$

Moreover, define

$$J_{pi}(\delta_{pi}, v_i) = \int_0^\infty r_i(\delta_{p_i}, v_i)dt \tag{12}$$

as the cost function of p_i, and

$$J_e(\delta_e, u) = \int_0^\infty r_e(\delta_e, u)dt \tag{13}$$

as the cost function of e_0, where $r_i(\delta_{p_i}, v_i)$ and $r_e(\delta_e, u)$ are the costs for every time instant, whose specific form will be given in Section III-A.

Then, the SEMP RADG game investigated in this paper is formulated as the following optimal control problem: Given the constraints of the error system dynamics (9) and (11), which are determined by the initial state $X^0 = (x_e^0, x_{p1}^0, x_{p2}^0, \cdots, x_{pq}^0)$, the dynamics (1) and (2), and the communication graphs $\mathcal{G}_p$ and $\mathcal{G}_e$, the objective is to find the distributed optimal controllers for u and $v_i, i = 1, 2, \cdots, q$, that minimize the cost functions (12) and (13).

Finally, we use the following technical lemma to conclude this section:

Lemma 1. *[32] Let $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ be a diagonally dominant matrix that $|a_{ii}| \geq \sum_{j\neq i} |a_{ij}|$, $\forall\ i = 1, 2, \cdots, n$ and*

$$S = \left\{ i \in N \mid |a_{ii}| > \sum_{j=1; j\neq i}^{n} |a_{ij}| \right\} \neq \emptyset \tag{14}$$

where $N = \{1, 2, \cdots, n\}$. In addition, for each $i \notin S$ there is a sequence of nonzero elements of A of the form $a_{ii_1}, a_{i_1 i_2}, \cdots, a_{i_r s}$ with $s \in S$, then A is nonsingular.

III. Optimal Solutions for RADG Games

A. Optimal Strategy with input saturation

The cost functions, which incorporate input saturation effects, are defined as follows:

$$r_i(\delta_{pi}, v_i) = Q_i(\delta_{pi}) + \psi_i(v_i), \tag{15}$$

$$r_e(\delta_e, u) = Q_e(\delta_e) + \psi_e(u), \tag{16}$$

where $Q_i(\delta_{pi})$, $Q_e(\delta_e) : \mathbb{R}^n \to \mathbb{R}$ are user-defined positive definite functions that penalize the interception errors and the control error, and

$$\begin{aligned}\psi_i(v_i) \triangleq& \int_0^{v_i} v_i^{sat} R_i \tanh^{-1}\left(\frac{\xi}{v_i^{sat}}\right) d\xi\\ =&v_i^{sat} v_i^T R_i \tanh^{-1}\left(\frac{v_i}{v_i^{sat}}\right) + \frac{1}{2}\left(v_i^{sat}\right)^2 \overline{R}_i ln\left(\mathbf{1}_m - \frac{v_i^2}{(v_i^{sat})^2}\right)\end{aligned} \tag{17}$$

and

$$\begin{aligned}\psi_e(u) \triangleq& \int_0^u u^{sat} R_e \tanh^{-1}\left(\frac{\xi}{u^{sat}}\right) d\xi\\ =&u^{sat} u^T R_e \tanh^{-1}\left(\frac{u}{u^{sat}}\right) + \frac{1}{2}\left(u^{sat}\right)^2 \overline{R}_e ln\left(\mathbf{1}_m - \frac{u^2}{(u^{sat})^2}\right)\end{aligned} \tag{18}$$

are positive definite functions penalizing the input of p_i and e_0, respectively. ξ is the variable of integration. $R_i = \text{diag}\{r_i^1, r_i^2, \cdots, r_i^m\}$, $R_e = \text{diag}\{r_e^1, r_e^2, \cdots, r_e^m\}$ are all diagonal matrices with positive diagonal elements. $\overline{R}_i = [r_i^1, r_i^2, \cdots, r_i^m]$, $\overline{R}_e = [r_e^1, r_e^2, \cdots, r_e^m]$ are row vectors.

Assumption 1. *There exist constants $\underline{q}_i, \overline{q}_i, \underline{q}_e, \overline{q}_e \in \mathbb{R}_{>0}$ such that $\underline{q}_i||\delta_{pi}||^2 \leq Q_i(\delta_{pi}) \leq \overline{q}_i||\delta_{pi}||^2$ for all $p_i \in \mathscr{P}$, and $\underline{q}_e||\delta_e||^2 \leq Q_e(\delta_e) \leq \overline{q}_e||\delta_e||^2$ for e_0.*

The value function of p_i and e_0 at time t can be defined according to the cost function

$$V_{pi}(\delta_{pi}, t) = \min_{v_i} \int_t^\infty r_i(\delta_{pi}, v_i)d\tau \tag{19}$$

and

$$V_e(\delta_e, t) = \min_{u} \int_t^\infty r_e(\delta_e, u)d\tau. \tag{20}$$

Define the Hamilton functions as

$$H_{pi}(\delta_{pi}, \nabla V_{pi}) = \nabla V_{pi}^T \dot{\delta}_{pi} + r_i \tag{21}$$

and

$$H_e(\delta_e, \nabla V_e) = \nabla V_e^T \dot{\delta}_e + r_e \tag{22}$$

for p_i and e_0, respectively, where $\nabla V_{pi} \triangleq \frac{\partial V_{pi}}{\partial \delta_{pi}}$, $\nabla V_e \triangleq \frac{\partial V_e}{\partial \delta_e}$.

The optimal control strategies can then be obtained by solving $\frac{\partial H_{pi}}{\partial v_i} = 0$ and $\frac{\partial H_e}{\partial u} = 0$ as

$$v_i^* = -v_i^{sat} \tanh\left[\frac{(d_i^p + \alpha_{i0})}{v_i^{sat}} R_i^{-1} g_{pi}^T \nabla V_{pi}\right] \tag{23}$$

and

$$u^* = -u^{sat} \tanh\left[\frac{(\beta_0 - d^e)}{u^{sat}} R_e^{-1} g^T \nabla V_e\right]. \tag{24}$$

The Hamilton-Jacobi (HJ) equation simplifies to $\min_{v_i} H_{pi}(\delta_{pi}, \nabla V_{pi}) = 0$ and $\min_u H_e(\delta_e, \nabla V_e) = 0$ in infinite-horizon problem. Letting $B_i = (d_i^p + \alpha_{i0}) g_{pi}^T$ and $B_e = (\beta_0 - d^e) g^T$ gives

$$\begin{aligned}&\nabla V_{pi}^T \dot{\delta}_{pi}^* + Q_i(\delta_{pi}) + v_i^{sat} \tanh\left(\frac{1}{v_i^{sat}} R_i^{-1} B_i \nabla V_{pi}\right)^T B_i \nabla V_{pi}\\ &+ v_i^{sat\,2} \overline{R}_i ln\left\{I_{m\times 1} - \left[\tanh\left(\frac{1}{v_i^{sat}} R_i^{-1} B_i \nabla V_{pi}\right)\right]^2\right\} = 0\end{aligned} \tag{25}$$

and

$$\begin{aligned}&\nabla V_e^T \dot{\delta}_e^* + Q_e(\delta_e) + u^{sat} \tanh\left(\frac{1}{u^{sat}} R_e^{-1} B_e \nabla V_e\right)^T B_e \nabla V_e\\ &+ u^{sat\,2} \overline{R}_e ln\left\{I_{m\times 1} - \left[\tanh\left(\frac{1}{u^{sat}} R_e^{-1} B_e \nabla V_e\right)\right]^2\right\} = 0\end{aligned} \tag{26}$$

where

$$\begin{aligned}\dot{\delta}^*_{pi} =&(d^p_i+\alpha_{i0})\left\{f_{pi}(x_{pi})-g_{pi}(x_{pi})\cdot v^{sat}_i\tanh\left[\frac{1}{v^{sat}_i}R^{-1}_i B_i\nabla V_{pi}\right]\right\}\\ &-\sum_{j\in\mathcal{N}_i}\alpha_{ij}\left\{f_{pj}(x_{pj})-g_{pj}(x_{pj})\cdot v^{sat}_j\tanh\left[\frac{1}{v^{sat}_j}R^{-1}_j B_j\nabla V_{pj}\right]\right\}\\ &-\alpha_{i0}\left\{f(x_e)-g(x_e)\cdot u^{sat}\tanh\left[\frac{1}{u^{sat}}R^{-1}_e B_e\nabla V_e\right]\right\}\end{aligned} \tag{27}$$

and

$$\begin{aligned}\dot{\delta}^*_e =&(\beta_0-d^e)\left\{f(x_e)-g(x_e)\cdot u^{sat}\tanh\left[\frac{1}{u^{sat}}R^{-1}_e B_e\nabla V_e\right]\right\}\\ &+\sum_{i\in\mathcal{N}_e}\beta_i\left\{f_{pi}(x_{pi})-g_{pi}(x_{pi})\cdot v^{sat}_i\tanh\left[\frac{1}{v^{sat}_i}R^{-1}_i B_i\nabla V_{pi}\right]\right\}.\end{aligned} \tag{28}$$

B. Nash Equilibrium for RADG Games

The concept of Nash equilibrium (NE) has attracted significant attention in PE games. Although the control strategies of the other players are not explicitly incorporated in (12) and (13), they critically influence the dynamics described by (9) and (11). Therefore, it is realistic to consider the Nash equilibrium strategy for the RADG games. First, we present the definition of the RADG game. Let v_{-i} denote the set of strategies for all pursuers except p_i, v_{-e} denote the set of strategies for all players other than e_0. We have the following definition:

Definition 1. *(Global Nash equilibrium for RADG games) The set of policies* $\mathrm{v}^*=\{u^*,v^*_1,v^*_2,\cdots,v^*_q\}$ *is a* global Nash equilibrium solution *for single evader multiple pursuers reach-avoid differential games if the inequalities*

$$J^*_{pi}\equiv J_{pi}\left(\delta_{pi},v^*_i,v^*_{-i},u^*\right)\le J_{pi}\left(\delta_{pi},v_i,v^*_{-i},u^*\right), \tag{29}$$

$$J^*_e\equiv J_e\left(\delta_e,u^*,v^*_{-e}\right)\le J_e\left(\delta_e,u,v^*_{-e}\right) \tag{30}$$

hold for all agents in the game, then the $(q+1)$*-tuple of the distributed local cost functions* $\{J^*_e,J^*_{p1},J^*_{p2},\ldots,J^*_{pq}\}$ *is the Nash equilibrium of the game, where* $v^*_{-i}=\mathrm{v}^*\backslash\{v^*_i\cup u^*\}$, $v^*_{-e}=\mathrm{v}^*\backslash\{u^*\}$.

The following theorem illustrates the conditions required to achieve Nash equilibrium in RADG games.

Theorem 1. *(Nash equilibrium in RADG games) Let* (1), (2) *be the dynamics of the evader and the pursuers, respectively. Let* (12), (13) *be the cost functions with the local errors* (8) *and* (10)*. Then a RADG game is in Nash equilibrium with the optimal strategies* (23) *and* (24)*. Moreover, the Nash equilibrium of the game is* $\{V_e(\delta_e(0)),V_{p1}(\delta_{p1}(0)),V_{p2}(\delta_{p2}(0)),\cdots,V_{pq}(\delta_{pq}(0))\}$.

Proof. Notice that $V_{pi}(\delta_{pi}(\infty))=V_{pi}(0)=0$ [15], the cost function (12) can be written as

$$J_{pi}=\int_0^\infty[Q_i(\delta_{pi})+\psi_i(v_i)]dt+V_{pi}(\delta_{pi}(0))+\int_0^\infty\nabla V^T_{pi}\dot{\delta}_{pi}(t)dt. \tag{31}$$

Define

$$J'_{pi}=\int_0^\infty[Q_i(\delta_{pi})+\psi_i(v^*_i)]dt+V_{pi}(\delta_{pi}(0))+\int_0^\infty\nabla V^T_{pi}\dot{\delta}^*_{pi}(t)dt. \tag{32}$$

Noting (9), the cost function can be rewritten as

$$\begin{aligned}J_{pi}=&J'_{pi}-\int_0^\infty[\nabla V^T_{pi}(d^p_i+\alpha_{i0})g_{p_i}(x_{p_i})\cdot v^*_i+\psi_i(v^*_i)]dt\\ &+\int_0^\infty[\nabla V^T_{pi}(d^p_i+\alpha_{i0})g_{p_i}(x_{p_i})\cdot v_i+\psi_i(v_i)]dt\end{aligned} \tag{33}$$

if $v_{-i}=v^*_{-i}$ and $u=u^*$, which indicates that all other agents in the RADG game use the optimal strategy. Then, as (25) holds, $J'_{pi}=V_{pi}(\delta_{pi}(0))$. Hence,

$$\begin{aligned}J_{pi}=&V_{pi}(\delta_{pi}(0))-\int_0^\infty[\nabla V^T_{pi}(d^p_i+\alpha_{i0})g_{p_i}(x_{p_i})\cdot v^*_i+\psi_i(v^*_i)]dt\\ &+\int_0^\infty[\nabla V^T_{pi}(d^p_i+\alpha_{i0})g_{p_i}(x_{p_i})\cdot v_i+\psi_i(v_i)]dt.\end{aligned} \tag{34}$$

By (17), we have

$$\begin{aligned}\frac{\partial J_{pi}}{\partial v_i}&=\int_0^\infty[(d^p_i+\alpha_{i0})g^T_{p_i}(x_{p_i})\nabla V_{pi}+\frac{\partial\psi_i(v_i)}{\partial v_i}]dt\\ &=\int_0^\infty[(d^p_i+\alpha_{i0})g^T_{p_i}(x_{p_i})\nabla V_{pi}+v^{sat}_i R_i\tanh^{-1}(\frac{v_i}{v^{sat}_i})]dt.\end{aligned} \tag{35}$$

Letting $\frac{\partial J_{pi}}{\partial v_i}=0$ yields

$$v_i=-v^{sat}_i\tanh\left[\frac{(d^p_i+\alpha_{i0})}{v^{sat}_i}R^{-1}_i g^T_{pi}\nabla V_{pi}\right]=v^*_i, \tag{36}$$

which implies that $v_i=v^*_i$ is the stationary point. Furthermore, it is easy to get $\frac{\partial^2 J_{pi}}{\partial v_i}=\int_0^\infty R_i[\frac{1}{1-(v_i/v^{sat}_i)^2}]dt>0$ at $v_i=v^*_i$, which proofs (29) with the cost of $J_{pi}=V_{pi}(\delta_{pi}(0))$.

The proof of (30) follows a similar argument to the above and is thus omitted. □

C. Interactive Nash Equilibrium under Communication Graphs

In [12], [33], [34], the underlying communication graph in graphical differential game problems is strongly connected, thereby rendering the analysis of Nash equilibrium both effective and comprehensive. Nevertheless, in adversarial reach-avoid games, it becomes essential to examine the diverse communication topologies among agents, particularly with regard to the interactions between the pursuit and evasion groups. Therefore, in this paper, we do not make any assumptions to the communication topology; instead, we classify the pursuers into four categories based on their communication nature with the evaders.

1) *Isolated pursuers*. If there is neither a directed path from e_0 to p_i nor a directed path from p_i to e_0, then p_i is called an isolated pursuer. The set of the index of all isolated pursuers is defined as $\mathscr{P}_{iso}$.
2) *Passive pursuers*. If p_i is connected to e_0, while there is no directed path from e_0 to p_i, then p_i is called a passive pursuer. The set of the index of all passive pursuers is defined as $\mathscr{P}_p$.
3) *Invisible pursuers*. If e_0 is connected to p_i, while there is no directed path from p_i to e_0, then p_i is called an

invisible pursuer. The set of the index of all invisible pursuers is defined as $\mathscr{P}_{inv}$.

4) *Regular pursuers*. If p_i and e_0 are connected to each other, then p_i is called a regular pursuer. The set of the index of all regular pursuers is defined as $\mathscr{P}_r$.

The above four cases cover all communication situations between the pursuer and the evader which leads to $\mathscr{P} = \mathscr{P}_{iso} \cup \mathscr{P}_p \cup \mathscr{P}_{inv} \cup \mathscr{P}_r$. It is worth pointing out that although the different types of pursuers may exert distinct influences on the game, they do not alter the Nash equilibrium as established in Theorem 1. Inspired by Definition 5 in [12], we propose a more stringent equilibrium definition for SEMP RADG games.

Definition 2. *(Interactive Nash equilibrium for SEMP RADG games) The set of policies* $\mathrm{v}^* = \{u^*, v_1^*, v_2^*, \cdots, v_q^*\}$ *is an* interactive Nash equilibrium (INE) solution *for single evader multiple pursuers reach-avoid differential games if the Nash equilibrium condition holds and, in addition, the following conditions hold:*

1) *For all* $p_i \in \mathscr{P}$*, there exists some strategy* v_i' *such that* $J_e\left(\delta_e, u^*, v_{-i}^* \cup v_i'\right) \neq J_e\left(\delta_e, u^*, v_{-e}^*\right)$.
2) *There exists a strategy-cost function pair* (u', J_{pi}) *such that* $J_{pi}\left(\delta_{pi}, v_i^*, v_{-i}^*, u'\right) \neq J_{pi}\left(\delta_{pi}, v_i^*, v_{-i}^*, u^*\right)$.

Remark 1. *Definition 2 imposes stricter requirements than Definition 1 by mandating mutual influence between the pursuers and the evader. Specifically, the set of pursuers must be able to affect the evader's cost function, and conversely, the evader must be able to influence at least one of the pursuers' cost functions. This bidirectional dependency ensures that, in an interactive Nash equilibrium, the strategies of both parties are interdependent, resulting in a more interactive and dynamically coupled equilibrium compared to the standard Nash equilibrium.*

The INE provides a formal criterion for characterizing the NE. Satisfying its condition of bidirectional influence confirms a dynamically coupled equilibrium, whereas a failure to do so reveals structural properties of the communication graph, such as information asymmetries or strategic decoupling.

We then have the following result:

Theorem 2. *(INE in SEMP RADG games) Let an SEMP RADG game be in Nash equilibrium. Then the SEMP RADG game is in interactive Nash equilibrium if and only if* $\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_r$ *and* $\mathscr{P}_r \neq \emptyset$.

Proof. Necessity: If $\mathscr{P}_{iso} \neq \emptyset$ or $\mathscr{P}_{inv} \neq \emptyset$, then there exists some p_i that is not connected to e_0, which contradicts the first condition of Definition 2. Therefore, $\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_r$. Subsequently, if $\mathscr{P}_r = \emptyset$, then $\mathscr{P} = \mathscr{P}_p$ which means that all the pursuers are passive pursuers. Consider that the value of the passive pursuer will not be influenced by the input of the evader, the second condition does not hold. In conclusion, if the two conditions hold, then $\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_r$ and $\mathscr{P}_r \neq \emptyset$.

Sufficiency: If $\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_r$ holds in an SEMP RADG game, then the pursuer is either a passive pursuer or a regular pursuer, which means that all the pursuers are connected to the evader. Then, the control input of p_i can influence the value of e_0 for all $p_i \in \mathscr{P}$. On the other hand, if $\mathscr{P}_r \neq \emptyset$, then there exists p_i such that there is a directed path from e_0 to p_j which makes J_{pi} influenced by the input of e_0. □

Since the pursuers in a strongly connected graph are all regular pursuers, we have $\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_r$ and $\mathscr{P}_r \neq \emptyset$. Thus, we can easily establish the following corollary:

Corollary 1. *Let* $\mathscr{G}_e$ *in an SEMP RADG game be strongly connected, and the game be in Nash equilibrium, then the game is in interactive Nash equilibrium.*

IV. Approximate Optimal Strategy via Approximate Dynamic Programming

A. Value Function Approximation

The optimal strategies (23) and (24) for the RADG game depend on solving the coupled Hamilton-Jacobi (HJ) equations (25) and (26). As these nonlinear partial differential equations are generally intractable, this section introduces an Approximate Dynamic Programming (ADP) algorithm to find numerical solutions. Unlike conventional Actor-Critic (AC) architectures [2], [12], [35]–[37], which use separate networks for the policy (actor) and value function (critic), the proposed method employs a single-network ADP to approximate the value function.

Assumption 2. *[38] The solutions to* (25) *and* (26) *are non-negative and smooth.*

Assumption 2 is a standard condition in ADP that guarantees the applicability of the Weierstrass high-order approximation theorem, thereby enabling the value functions to be approximated with arbitrary precision. We assume that V_{pi}, V_e, and their derivatives respectively with respect to δ_{pi} and δ_e can be expressed as

$$V_{pi} = W_{pi}^T \phi_{pi} + \varepsilon_{pi}, \nabla V_{pi} = \nabla \phi_{pi}^T W_{pi} + \nabla \varepsilon_{pi} \tag{37}$$

and

$$V_e = W_e^T \phi_e + \varepsilon_e, \nabla V_e = \nabla \phi_e^T W_e + \nabla \varepsilon_e \tag{38}$$

where $W_{pi}, W_e \in \mathbb{R}^N$ are the ideal weights of the NN with independent basis $\phi_{pi}, \phi_e : \mathbb{R}^n \to \mathbb{R}^N$, respectively. ε_{pi} and ε_e are the NN approximation errors. $\nabla \phi_{pi} = \frac{\partial \phi_{pi}}{\partial \delta_{pi}}$, $\nabla \phi_e = \frac{\partial \phi_e}{\partial \delta_e}$, $\nabla \varepsilon_{pi} = \frac{\partial \varepsilon_{pi}}{\partial \delta_{pi}}$, $\nabla \varepsilon_e = \frac{\partial \varepsilon_e}{\partial \delta_e}$ are the corresponding gradient operators. According to the Weierstrass high-order approximation theorem, as $N \to \infty$, the corresponding NN estimation errors ε_{pi} and ε_{pi} will both converge to zero. Substituting (37) and (38) into (25) and (26), one can get

$$\bar{H}_{pi}(\delta_{pi}, \nabla \phi_{pi}^T W_{pi}) = W_{pi}^T \nabla \phi_{pi} \dot{\delta}_{pi} + r_i = \varepsilon_{H_{pi}} \tag{39}$$

and

$$\bar{H}_e(\delta_e, \nabla \phi_e^T W_e) = W_e^T \nabla \phi_e \dot{\delta}_e + r_e = \varepsilon_{H_e}, \tag{40}$$

where $\bar{H}_{pi}$ and $\bar{H}_e$ are the Hamilton functions with approximated value functions, $\varepsilon_{H_{pi}} = -\nabla \varepsilon_{pi}^T \dot{\delta}_{pi}$, $\varepsilon_{H_e} = -\nabla \varepsilon_e^T \dot{\delta}_e$ are the residual errors caused by uncompleted basis.

The exact W_{pi}, W_{pi} are unknown, therefore, the estimated weights $\hat{W}_{pi}$ and $\hat{W}_e$ are introduced to approximate V_{pi} and V_e as $\hat{V}_{pi} = \hat{W}_{pi}^T \phi_{pi}$ and $\hat{V}_e = \hat{W}_e^T \phi_e$. Then, the approximated

optimal strategies for the pursuers and evader are respectively given by

$$\hat{v}_i = -v_i^{sat} \tanh\left[\frac{(d_i^p + \alpha_{i0})}{v_i^{sat}} R_i^{-1} g_{pi}^T \nabla\phi_{pi}^T \hat{W}_{pi}\right] \tag{41}$$

and

$$\hat{u} = -u^{sat} \tanh\left[\frac{(\beta_0 - d^e)}{u^{sat}} R_e^{-1} g^T \nabla\phi_e^T \hat{W}_e\right]. \tag{42}$$

B. Distributed Tuning Law

In this subsection, we propose a distributed network update law. Compared to [11], [12], [38], [39], our update law no longer requires persistent excitation conditions.

Define the estimate errors of W_{pi} and W_e as $\tilde{W}_{pi} = W_{pi} - \hat{W}_{pi}$ and $\tilde{W}_e = W_e - \hat{W}_e$, respectively. Then, the Hamilton functions with the estimated weights are given by

$$\hat{H}_{pi}(\delta_{pi}, \nabla\hat{V}_{pi}) = \hat{W}_{pi}^T \nabla\phi_{pi}\dot{\delta}_{pi} + r_i = \bar{\varepsilon}_{H_{pi}} \tag{43}$$

and

$$\hat{H}_e(\delta_e, \nabla\hat{V}_e) = \hat{W}_e^T \nabla\phi_e\dot{\delta}_e + r_e = \bar{\varepsilon}_{H_e}. \tag{44}$$

Recall (39) and (40), it can be seen that the residual errors caused by the uncompleted basis and the estimated weights are $\bar{\varepsilon}_{H_{pi}} = \varepsilon_{H_{pi}} - \tilde{W}_{pi}^T \nabla\phi_{pi}\dot{\delta}_{pi}$ and $\bar{\varepsilon}_{H_e} = \varepsilon_{H_e} - \tilde{W}_e^T \nabla\phi_e\dot{\delta}_e$. Define $E = \frac{1}{2}\sum_{i=1}^{q} \bar{\varepsilon}_{H_{pi}}^T \bar{\varepsilon}_{H_{pi}} + \frac{1}{2}\bar{\varepsilon}_{H_e}^T \bar{\varepsilon}_{H_e}$. To make E sufficiently small, the update law of the NN weights is designed following the modified Levenberg-Marquardt algorithm. In addition, concurrent learning technology is applied, which adds historical data to relax the restriction of the persistent excitation condition.

$$\dot{\hat{W}}_{pi} = -\eta_i \frac{\chi_i}{(\chi_i^T\chi_i + 1)^2}\bar{\varepsilon}_{H_{pi}} - \sum_{j=1}^{k} \eta_i \frac{\chi_i(t_j)}{(\chi_i^T(t_j)\chi_i(t_j) + 1)^2}\bar{\varepsilon}_{H_{pi}(t_j)} \tag{45}$$

and

$$\dot{\hat{W}}_e = -\eta_e \frac{\chi_e}{(\chi_e^T\chi_e + 1)^2}\bar{\varepsilon}_{H_e} - \sum_{j=1}^{k} \eta_e \frac{\chi_e(t_j)}{(\chi_e^T(t_j)\chi_e(t_j) + 1)^2}\bar{\varepsilon}_{H_e(t_j)} \tag{46}$$

where $\chi_i = \nabla\phi_{pi}\dot{\delta}_{pi}$, $\chi_e = \nabla\phi_e\dot{\delta}_e$, $\chi_i(t_j)$ and $\chi_e(t_j)$ are the historical data at each sampling time t_j, $j = 1, 2, \cdots, k$ with $0 = t_1 < t_2 < \cdots < t_k < t$ the sampling instants, η_i and η_e are the user-defined constant gains that dictate the learning rates. It can be easily seen that the update law for the strategy of p_i and e_0 in (45) and (46) depends only on the information of itself and its neighbors, respectively. This ensures distributed decision-making by players.

C. Stability Analysis

First, we recall the concept of the uniformly ultimately boundedness as follows:

Definition 3. *(Uniformly ultimately boundedness) A system* $\dot{\mathrm{x}} = \mathrm{f}(\mathrm{x}, t)$ *is uniformly ultimately bounded (UUB) if there exists a compact set* $S \in \mathbb{R}^n$ *so that for all* $\mathrm{x} \in S$ *there exists a bound* B *and a time* $\mathrm{T}(\mathrm{B}, \mathrm{x})$ *such that* $\|\mathrm{x}(t) - \mathrm{x}_e\| \leq B$ *for all* $t \geq \mathrm{T}$.

Like in [12], [35], [38], we make the following standard assumptions.

Assumption 3.

(i) NN-related bound: *The ideal NN weighs* W_e *and* W_{pi}, *the NN basis* ϕ_e *and* ϕ_{pi} *and their derivative, and the approximation errors* ε_e *and* ε_{pi} *and their derivative for* e_0 *and all* $p_i \in \mathscr{P}$ *are assumed to be bounded with positive constants* b_e^W, b_i^W, b_e^ϕ, b_i^ϕ, $b_e^{\nabla\phi}$, $b_i^{\nabla\phi}$, b_e^ε, b_i^ε, $b_e^{\nabla\varepsilon}$, $b_i^{\nabla\varepsilon}$ *respectively.*

(ii) System bound: *The functions* g *and* g_{pi} *for* e_0 *and all* $p_i \in \mathscr{P}$ *are assumed to be bounded in domain* Ω *with positive constants* b_e^g *and* b_i^g.

Remark 2. *These are standard assumptions in the neural adaptive control literature. Moreover, many practical systems satisfy Assumption 3.ii, such as the satellite pursuit-evasion game [1], single integrator dynamics for quadcopters [40], and Dubins vehicle models [41], [42]. It should be noted that the boundedness conditions specified above are only required for the stability analysis of the closed-loop system, ensuring the boundedness of the overall system. These conditions are not directly utilized in the design of the control strategies. In practice, the system states and network weights will converge to an equilibrium, and the stability of the system typically does not depend on the explicit values of these bounds. Furthermore, under Assumption 3, it follows that* $\varepsilon_{H_{pi}}$ *and* ε_{H_e} *are bounded by* $b_i^{\varepsilon_H}$ *and* $b_e^{\varepsilon_H}$, *respectively.*

We have the following result:

Theorem 3. *Under Assumptions 1 to 3, consider the dynamics of the interception errors (9) and the control error (11), the approximate optimal strategies (41) and (42) with the NN update laws (45) and (46), then the interception errors* δ_{pi} *for* $i = 1, 2, \cdots, q$, *the control error* δ_e *and the NN weights estimation errors* $\tilde{W}_{pi}$, $\tilde{W}_e$ *are UUB.*

Proof. Choose the Lyapunov function candidate as

$$J = \sum_{i=1}^{q}\left(J_i^1 + J_i^2\right) + J_e^1 + J_e^2 \tag{47}$$

where $J_i^1 = V_{pi}$, $J_i^2 = \frac{1}{2}\tilde{W}_{pi}^T\tilde{W}_{pi}$, $J_e^1 = V_e$, $J_e^2 = \frac{1}{2}\tilde{W}_e^T\tilde{W}_e$. By direct calculation, we have

$$\dot{J}_i^1 = -Q_i(\delta_{pi}) - \psi_i(v_i) \leq -\underline{q}_i\|\delta_{pi}\|^2 - \psi_i(v_i) \tag{48}$$

and

$$\dot{J}_e^1 \leq -\underline{q}_0||\delta_e||^2 - \psi_e(u). \tag{49}$$

Substituting $\tilde{W}_{pi} = W_{pi} - \hat{W}_{pi}$ into (45) and using the fact that $\bar{\varepsilon}_{H_{pi}} = \varepsilon_{Hpi} - \tilde{W}_{pi}^T\nabla\phi_{pi}\dot{\delta}_{pi}$ yield

$$\begin{aligned}\dot{\tilde{W}}_{pi} =& -\eta_i[\sigma_i\sigma_i^T + \sum_{j=1}^{k}\sigma_i^T(t_j)\sigma_i(t_j)]\tilde{W}_{pi} \\ &+ \eta_i[\frac{\sigma_i^T\varepsilon_{H_{pi}}}{\chi_i^T\chi_i + 1} + \sum_{j=1}^{k}\frac{\sigma_i(t_j)\varepsilon_{H_{pi}}(t_j)}{\chi_i^T(t_j)\chi_i(t_j) + 1}],\end{aligned} \tag{50}$$

where $\sigma_i = \chi_i/(\chi_i^T\chi_i + 1)$ and $\sigma_i(t_j) = \chi_i(t_j)/[\chi_i^T(t_j)\chi_i(t_j) + 1]$.

Using Young's inequality, we have

$$\frac{\eta_i \tilde{W}_{p_i}^T \sigma_i \varepsilon_{H_{pi}}}{1+\chi_i^T \chi_i} \leq \frac{1}{2}\eta_i \tilde{W}_{p_i}^T \sigma_i \sigma_i^T \tilde{W}_{pi} + \frac{1}{2}\eta_i \varepsilon_{Hp_i}^T \varepsilon_{Hp_i} \tag{51}$$

and

$$\sum_{j=1}^{k} \frac{\eta_i \tilde{W}_{p_i}^T \sigma_i(t_j) \varepsilon_{H_{pi}}(t_j)}{1+\chi_i^T(t_j)\chi_i(t_j)} \leq \sum_{j=1}^{k} \frac{1}{2}\eta_i \tilde{W}_{p_i}^T \sigma_i(t_j)\sigma_i^T(t_j)\tilde{W}_{pi} + \sum_{j=1}^{k} \frac{1}{2}\eta_i \varepsilon_{Hp_i}^T(t_j)\varepsilon_{Hp_i}(t_j), \tag{52}$$

which lead to

$$\dot{J}_i^2 = \tilde{W}_{p_i}^T \dot{\hat{W}}_{p_i} \leq -\frac{1}{2}\eta_i \lambda_{\min}(\Xi_i)\|\tilde{W}_{p_i}\|^2 + \frac{1}{2}\eta_i(k+1)b_i^{\varepsilon_H}, \tag{53}$$

where $\Xi_i = \sigma_i\sigma_i^T + \sum_{j=1}^{k}\sigma_i(t_j)\sigma_i^T(t_j)$.

Similarly, one may obtain

$$\dot{J}_e^2 \leq -\frac{1}{2}\eta_e \lambda_{\min}(\Xi_e)\|\tilde{W}_e\|^2 + \frac{1}{2}\eta_e(k+1)b_e^{\varepsilon_H}, \tag{54}$$

where $\Xi_e = \sigma_e\sigma_e^T + \sum_{j=1}^{k}\sigma_e(t_j)\sigma_e^T(t_j)$.

As a result

$$\begin{aligned}\dot{J} \leq & -\sum_{i=1}^{q}\underline{q}_i||\delta_{pi}||^2 - \underline{q}_0||\delta_e||^2 \\ & -\frac{1}{2}\sum_{i=1}^{q}\left[\eta_i\lambda_{\min}(\Xi_i)||\tilde{W}_{pi}||^2\right] - \frac{1}{2}\eta_e\lambda_{\min}(\Xi_e)||\tilde{W}_e||^2 \\ & +\frac{1}{2}\sum_{i=1}^{q}\eta_i(k+1)b_i^{\varepsilon_H} + \frac{1}{2}\eta_e(k+1)b_e^{\varepsilon_H}.\end{aligned} \tag{55}$$

Note that from the definition of $\Xi_i, i = 1,...,q$ and Ξ_e, there must exist some $T_0 > 0$ such that $\zeta_1 \triangleq \min\{\lambda_{\min}(\Xi_1),...,(\Xi_q),\lambda_{\min}(\Xi_e)\} > 0$. First consider for $0 \leq t < T_0$, (55) can be further written as:

$$\dot{J} \leq \frac{1}{2}\sum_{i=1}^{q}\eta_i(k+1)b_i^{\varepsilon_H} + \frac{1}{2}\eta_e(k+1)b_e^{\varepsilon_H}. \tag{56}$$

Since the right-hand side of (56) is finite and T_0 is finite, we conclude that $V_{pi}, i = 1,...,q$, $\hat{W}_{pi}, i = 1,...,q$, V_e, $\hat{W}_e$ are all finite over $0 \leq t < T_0$.

Now consider for $t \geq T_0$, it can be easily verified that $\dot{J} \leq 0$ if one of the following conditions holds.

1) Error state condition.

$$\sum_{i=1}^{q}||\delta_{pi}||^2 + ||\delta_e||^2 \geq \frac{(k+1)(\sum_{i=1}^{q}\eta_i b_i^{\varepsilon_H} + \eta_e b_e^{\varepsilon_H})}{2min(\underline{q}_i,\underline{q}_0)} = b^{\delta} \tag{57}$$

where $min(\underline{q}_i,\underline{q}_0)$ is the minimum value of $\underline{q}_0$ and $\underline{q}_i$ for $i = 1,2,\cdots,q$.

2) NN estimate error condition.

$$\sum_{i=1}^{q}||\tilde{W}_{pi}||^2 + ||\tilde{W}_e||^2 \geq \frac{(k+1)(\sum_{i=1}^{q}\eta_i b_i^{\varepsilon_H} + \eta_e b_e^{\varepsilon_H})}{2\zeta_1} = b^{\hat{W}}. \tag{58}$$

Therefore, we deduce that the interception errors δ_{pi}, the control error δ_e, and the estimation errors $\tilde{W}_{pi}$ and $\tilde{W}_e$ are uniformly ultimately bounded (UUB). This concludes the proof. □

Remark 3. *The number of historical data points k should be chosen sufficiently large to ensure that Ξ_i and Ξ_e are full rank, thus guaranteeing $\zeta_1 \neq 0$. Unlike previous methods [11], [12], [38], [39], the concurrent learning (CL) technique employed here utilizes both historical and current data, thereby relaxing the persistent excitation requirement and enhancing practical applicability.*

In addition, we have the following corollary:

Corollary 2. *Under Assumptions 1-3, let (42) be the control policy of e_0, (41) be the control policy of $p_i \in \mathscr{P}$, and (46) and (45) be the updating law. Then, the set of control input $\hat{\mathrm{v}} = \{\hat{u}, \hat{v}_1, \hat{v}_2, \cdots, \hat{v}_q\}$ will be attracted to a cube neighborhood $\mathscr{C}$ of $\mathrm{v}^* = \{u^*, v_1^*, v_2^*, \cdots, v_q^*\}$ and the RADG game is in the approximate Nash equilibrium, where*

$$\begin{aligned}\mathscr{C}(\mathrm{v}^*, \zeta_2^e, \zeta_2^i) = \{\mathrm{v} = (\mathrm{u}, \mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_q) \in \mathbb{R}^{m\times(q+1)} \mid \\ \|\mathrm{u} - u^*\| < \zeta_2^e, \|\mathrm{v}_i - v_i^*\| < \zeta_2^i\}\end{aligned} \tag{59}$$

with

$$\zeta_2^e = \frac{\|\beta_0 - d^e\|}{u^{sat}\lambda_{\min}(R_e)} b_e^g b_e^{\nabla\phi}(b_e^{\nabla\varepsilon} + \sqrt{b^{\hat{W}}}),$$

$$\zeta_2^i = \frac{\|d_i^p + \alpha_{i0}\|}{v_i^{sat}\lambda_{\min}(R_i)} b_i^g b_i^{\nabla\phi}(b_i^{\nabla\varepsilon} + \sqrt{b^{\hat{W}}})$$

for $i = 1,2,\cdots,q$.

Proof. From (58), one can get $\|\tilde{W}_e\| < \sqrt{b^{\hat{W}}}$, using the Lagrange mean value theorem, we have $\tanh(a) - \tanh(b) = (1 - \tanh(c))(a-b)$, where $b \leq c \leq a$. Then, by using (24), (42) and (38), if Assumption 3 holds, one can obtain

$$\begin{aligned}\|\hat{u} - u^*\| = & \|u^{sat}\tanh\left[\frac{(\beta_0 - d^e)}{u^{sat}}R_e^{-1}g^T\nabla V_e\right] \\ & - u^{sat}\tanh\left[\frac{(\beta_0 - d^e)}{u^{sat}}R_e^{-1}g^T\nabla\phi_e^T\hat{W}_e\right]\| \\ \leq & \|\left[1 - \tanh^2(c')\right](\beta_0 - d^e)R_e^{-1}g^T(\nabla V_e - \nabla\phi_e^T\hat{W}_e)\| \\ \leq & \frac{\|\beta_0 - d^e\|}{\lambda_{\min}(R_e)} b_e^g b_e^{\nabla\phi}(b_e^{\nabla\varepsilon} + \sqrt{b^{\hat{W}}})\end{aligned} \tag{60}$$

since $0 < 1 - \tanh^2(c') \leq 1$ with c' a mid-value between $\frac{(\beta_0 - d^e)}{u^{sat}}R_e^{-1}g^T\nabla V_e$ and $\frac{(\beta_0 - d^e)}{u^{sat}}R_e^{-1}g^T\nabla\phi_e^T\hat{W}_e$. The same procedure is omitted for

$$\|\hat{v}_i - v_i^*\| \leq \frac{\|d_i^p + \alpha_{i0}\|}{v_i^{sat}\lambda_{\min}(R_i)} b_i^g b_i^{\nabla\phi}(b_i^{\nabla\varepsilon} + \sqrt{b^{\hat{W}}}). \tag{61}$$

□

V. Equilibrium Points and Extensions

The Nash equilibrium solution for the proposed RADG game has been established above, with an ADP approach adopted to approximate the solution. In this section, we extend our analysis by examining the equilibrium points and exploring several extensions of the RADG framework. First, we discuss the characteristics of the equilibrium points and demonstrate the conditions for asymptotic capture. Then, we analyze the configuration of these equilibrium points, revealing that, under certain conditions, RADG games may transition into general

PE games. Finally, we present an extension of the RADG formulation to accommodate a half-space target.

Theorem 2 establishes that an INE solution requires the absence of isolated and invisible pursuers, alongside the presence of at least one regular pursuer. Therefore, this section reasonably assumes that $\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_r$ and $\mathscr{P}_r \neq \emptyset$.

A. Equilibrium Points and Asymptotic Capture

Regular pursuers are divided into two categories $\mathscr{P}_r^{-p}$ and $\mathscr{P}_r^{p}$ based on whether passive pursuers are connected to them. $\mathscr{P}_r^{-p}$ is the set of all the indices of pursuers that satisfy $i \in \mathscr{P}_r$ and there is no directed path from p_j to p_i for all $j \in \mathscr{P}_p$ and $\mathscr{P}_r^{p} = \mathscr{P}_r \backslash \mathscr{P}_r^{-p}$.

Define $\overline{X} = \{\bar{x}_e, \bar{x}_{p1}, \bar{x}_{p2}, \cdots, \bar{x}_{pq}\}$ as the equilibrium set of the SEMP RADG game that satisfies $\delta_{pi} = \sum_{j\in\mathscr{N}_i} \alpha_{ij}(\bar{x}_{pi} - \bar{x}_{pj}) + \alpha_{i0}(\bar{x}_{pi} - \bar{x}_e) = 0$ for all $i \in \mathscr{P}$ and $\delta_e = \beta_0(\bar{x}_e - \bar{x}_d) - \sum_{i\in\mathscr{N}_e} \beta_i(\bar{x}_e - \bar{x}_{pi}) = 0$. The following Theorem demonstrates the asymptotic capture of the pursuers in $\mathscr{P}_r^{-p}$.

Theorem 4. *(Equilibrium and asymptotic capture) Under Assumptions 1 and 3, let (1), (2) be the dynamics of the evader and the pursuers, respectively; and let (12), (13) be the cost functions with the local errors (8) and (10). If Assumption 2 is enhanced to the solutions to (25) and (26) are positive definite, then $\overline{X}$ is asymptotically reached under the optimal strategies (23) and (24).*

Moreover, if $\mathscr{P} = \mathscr{P}_p \cup \mathscr{P}_r$ and $\mathscr{P}_r \neq \emptyset$, then the pursuers in $\mathscr{P}_r^{-p}$ will capture the evader in an asymptotic behavior; that is, if the termination conditions are ignored, then $\lim_{t\to\infty}(\bar{x}_{pi}(t) - \bar{x}_e(t)) = 0$, $\forall\ i \in \mathscr{P}_r^{-p}$.

Proof. Choose the Lyapunov function candidate as $V_{pi}(\delta_{pi})$, with (21) and (22), one can get

$$\dot{V}_{pi}(\delta_{pi}) = \nabla V_{pi}\dot{\delta}_{pi}^{*} = -Q_i(\delta_{pi}) - \psi_i(v_i^{*}) \tag{62}$$

and

$$\dot{V}_{e}(\delta_{e}) = \nabla V_{e}\dot{\delta}_{e}^{*} = -Q_e(\delta_{e}) - \psi_e(u^{*}) \tag{63}$$

as (25) and (26) hold. Since $Q_i(\delta_{pi})$, $\psi_i(v_i)$, $Q_e(\delta_e)$, and $\psi_i(u)$ are all positive definite equations, $\dot{V}_{pi}(\delta_{pi}) < 0$ and $\dot{V}_e(\delta_e) < 0$. Therefore, the equilibrium set is reached and $\delta_{p1} = \delta_{p2} = \cdots = \delta_{pq} = \delta_e = 0$.

To prove $\bar{x}_{pi} = \bar{x}_e$ for all $i \in \mathscr{P}_r^{-p}$, define $\mathscr{D}_{pe} = \mathrm{diag}\{\alpha_{10}, \alpha_{20}, \cdots, \alpha_{q0}\}$ and substituting $\delta_{p1} = \delta_{p2} = \cdots = \delta_{pq} = 0$ into (8), one can get

$$[\mathscr{D}_p - \mathscr{A}_p + \mathscr{D}_{pe}][\bar{x}_{p1}, \bar{x}_{p2}, \cdots, \bar{x}_{pq}]^T = \mathscr{A}_{pe}\bar{x}_e^T. \tag{64}$$

Consider the group of pursuers p_i, $i \in \mathscr{P}_r^{-p}$. Notice that $\alpha_{ij} = 0$ for all $i \in \mathscr{P}_r^{-p}$ and $j \in \mathscr{P}_r^{p} \cup \mathscr{P}_p$, otherwise there exists a path from at least one passive pursuer to p_i which contradicts the definition. Therefore, the following equation holds.

$$[\mathscr{D}_p - \mathscr{A}_p + \mathscr{D}_{pe}]_{\mathscr{P}_r^{-p},\mathscr{P}_r^{-p}} [\bar{x}_{p1}, \bar{x}_{p2}, \cdots, \bar{x}_{pq}]^T_{\mathscr{P}_r^{-p},\{1,2,\cdots,n\}} = \mathscr{A}_{pe\,\mathscr{P}_r^{-p},\{1\}}\bar{x}_e^T. \tag{65}$$

From (8), it is obvious that $\bar{x}_{pi} = \bar{x}_e$, $\forall\ i \in \mathscr{P}_r^{-p}$ is a solution to (65). Next, we show that $\bar{x}_{pi} = \bar{x}_e$, $\forall\ i \in \mathscr{P}_r^{-p}$ is the only solution. Since $\alpha_{i0} + d_i^p \geq \sum_{j=1}^{q} |\alpha_{ij}|$ for each i, $[\mathscr{D}_p - \mathscr{A}_p + \mathscr{D}_{pe}]_{\mathscr{P}_r^{-p},\mathscr{P}_r^{-p}}$ is a diagonally dominant matrix. Notice that there exists i such that $\alpha_{i0} > 0$, otherwise $\mathscr{P}_r = \emptyset$, which leads to $\left\{ i \in \mathscr{P}_r^{-p} \mid \alpha_{i0} + d_i^p > \sum_{j=1}^{q} |\alpha_{ij}| \right\} \neq \emptyset$. In addition, for all $i \in \mathscr{P}_r^{-p}$, p_i are all connected to the evader. According to Lemma 1, $[\mathscr{D}_p - \mathscr{A}_p + \mathscr{D}_{pe}]_{\mathscr{P}_r^{-p},\mathscr{P}_r^{-p}}$ is nonsingular, which completes the proof. □

B. Equilibrium Points Configuration: Choice of β_0

Theorem 4 presents the equilibrium conditions; however, the ultimate positions of the equilibrium points are not addressed. In fact, under the conditions of Theorem 4, we can further establish the following corollary, which discusses the configuration of the equilibrium points and proposes a criterion for choosing β_0.

Corollary 3. *(Equilibrium points configuration) Let the conditions of Theorem 4 hold and assume that $X \notin \mathscr{T}_p$ during the interval under consideration. If $x_d = 0$ and the optimal strategies (23) and (24) are implemented with*

$$\beta_0 \geq \frac{\|\bar{X} - e\|}{R_{A_0}} + d^e \quad or \quad \beta_0 \leq -\frac{\|\bar{X} - e\|}{R_{A_0}} + d^e \tag{66}$$

where $\overline{X}_{-e} = \sum_{i\in\mathscr{N}_e} \beta_i \bar{x}_{pi}$ denotes the center of gravity of the pursuers in $\mathscr{N}_e$, then the equilibrium position of the evader $\bar{x}_e$, will lie within the target area $\Omega_d(x_d, R_{A_0}) = \{x \in \mathbb{R}^n \mid \|x - x_d\| \leq R_{A_0}\}$.

Proof. Substituting $\delta_{pe} = 0$ into (10), one can get $(\beta_0 - d^e)\bar{x}_e = -\sum_{i\in\mathscr{N}_e} \beta_i \bar{x}_{pi} + \beta_0 x_d$. Letting $x_d = 0$ and $\beta_0 \neq d^e$ yields

$$\bar{x}_e = \frac{\overline{X}_{-e}}{\beta_0 - d^e}. \tag{67}$$

Substituting condition (66) into (67) yields $\|\bar{x}_e - x_d\| \leq R_{A_0}$, which ensures that the evader's equilibrium position lies within the target area. Note that, without loss of generality, the coordinate origin can be chosen such that $x_d = 0$. □

Remark 4. *If $\beta_0 = d^e$, then (10) reduces to*

$$\delta_e = -\beta_0 x_d + \sum_{i\in\mathscr{N}_e} \beta_i x_{pi} \tag{68}$$

which implies that the evader's state has no effect on the control error. Consequently, the optimal control becomes $u^{*} = 0$, rendering the evader passive.

Remark 5. *When $\beta_0 - d^e$ is sufficiently small, equation (67) implies that $\|\bar{x}_e\|$ becomes appreciably larger than $\|\overline{X}_{-e}\|$ such that $\|\bar{x}_e\|$ is beyond the target area ($x_d = 0$). Moreover, the pursuers in $\mathscr{P}_r$ continue to converge toward the evader. Consequently, by judiciously positioning the virtual target area, the proposed algorithm can be effectively extended to general pursuit-evasion games [15], [16], [29]. Notably, Theorems 1–4 remain valid under this setting. A numerical example in Section VI-A illustrating this behavior is provided.*

C. Extension to Half-space Targets

In the preceding sections, the target area was modeled as a sphere. However, in certain scenarios [10], [30], [31], the target region may instead be represented as a half-space. In this subsection, we extend the proposed algorithm to accommodate this alternative target geometry.

Definition 4. *(Half-space target) In $\mathbb{R}^n$, a half-space target is defined by the inequality $\mathrm{L}^T\mathbf{x}+b\geq 0$, where $\mathbf{x}\in\mathbb{R}^n$, $\mathrm{L}\in\mathbb{R}^n$ is a nonzero normal vector, and $b\in\mathbb{R}$ denotes the translational offset of the corresponding hyperplane from the origin.*

Considering the target region defined by a half-space $\mathrm{L}^T\mathrm{x}+b\geq 0$, the control error (10) is modified to

$$\delta_e=\frac{\beta_0(\mathrm{L}^T x_e+b)\mathrm{L}}{\|\mathrm{L}\|^2}-\sum_{i\in\mathcal{N}_e}\beta_i(x_e-x_{pi}). \tag{69}$$

Moreover, the HJ equation (26) is modified to

$$\begin{aligned}&\nabla V_e^T\dot{\delta}_e^*+Q_e(\delta_e)+u^{sat}\tanh\left(\frac{1}{u^{sat}}R_e^{-1}B_e\nabla V_e\right)^T B_e\nabla V_e\\&+u^{sat\,2}\overline{R}_e ln\left\{I_{m\times 1}-\left[\tanh\left(\frac{1}{u^{sat}}R_e^{-1}B_e\nabla V_e\right)\right]^2\right\}=0\end{aligned} \tag{70}$$

where

$$\begin{aligned}\dot{\delta}_e^*=&\frac{\beta_0\mathrm{L}^T}{\|\mathrm{L}\|^2}\left\{f(x_e)-g(x_e)\cdot u^{sat}\tanh\left[\frac{1}{u^{sat}}R_e^{-1}B_e\nabla V_e\right]\right\}\mathrm{L}\\&-d^e\left\{f(x_e)-g(x_e)\cdot u^{sat}\tanh\left[\frac{1}{u^{sat}}R_e^{-1}B_e\nabla V_e\right]\right\}\\&+\sum_{i\in\mathcal{N}_e}\beta_i\left\{f_{pi}(x_{pi})-g_{pi}(x_{pi})\cdot v_i^{sat}\tanh\left[\frac{1}{v_i^{sat}}R_i^{-1}B_i\nabla V_{pi}\right]\right\}.\end{aligned} \tag{71}$$

The following theorem demonstrates the conditions required to achieve NE in RADG games with half-space targets.

Theorem 5. *(Nash equilibrium in RADG games with half-space targets) Let (1), (2) be the dynamics of the evader and the pursuers, respectively. Let (12), (13) be the cost functions with the local errors (8) and (69). Then a RADG game is in Nash equilibrium with the optimal strategies (23) and (24). Moreover, the Nash equilibrium of the game is $\{V_e(\delta_e(0)),V_{p1}(\delta_{p1}(0)),V_{p2}(\delta_{p2}(0)),\cdots,V_{pq}(\delta_{pq}(0))\}$.*

Proof. The proof is omitted due to the similarity with the proof of Theorem 1. □

VI. Numerical Simulations

This section presents numerical simulations to validate the efficacy of the proposed algorithm. Two distinct scenarios are examined: the first investigates RADG games within various flow fields, encompassing three cases with different flow configurations, while the second focuses on heterogeneous multi-agent systems and includes a comparative analysis against existing methods. All simulations were implemented in Python 3.9 on a computer equipped with an AMD Ryzen 7840H CPU and 32 GB of RAM, utilizing the fourth-order Runge-Kutta method with a time step of 0.01 seconds.

A. RADG games in Flow Fields

Case 1: RADG game in Steady Uniform Flow. We begin with a simplified scenario involving a reach-avoid game within a steady, uniform flow field. Here, agents are modeled using single integrator dynamics, and the flow velocity remains constant. As depicted in Fig. 2 (a), the trajectories obtained through the proposed algorithm closely approximate the theoretically optimal trajectory, as delineated in [21].

Case 2: RADG game in Non-Uniform Flow. Building on the results from the simplified uniform flow scenario, this case assesses the algorithm's robustness in a more complex environment characterized by a non-uniform flow field and communication constraints. These factors, combined with nonlinear dynamics, introduce challenges that exceed the analytical scope of prior works such as [21]. The RADG game is conducted within a flow field defined by $f(x)=[-0.01x^1+0.01x^2,-0.01x^1-0.01x^2(1-(\cos(2x^1)+2)^2)]^T$, with each agent governed by the following nonlinear dynamics:

$$\begin{bmatrix}\dot{x}^1\\\dot{x}^2\end{bmatrix}=\begin{bmatrix}\sin\left(2x^1\right)+2, & 0\\0, & \cos\left(2x^1\right)+2\end{bmatrix}\begin{bmatrix}u^1\\u^2\end{bmatrix} \tag{72}$$

where x^1 and x^2 represent the agent's states, and u^1,u^2 are the control inputs. The initial positions are set to $x_e^0=[-10,10]^T$ for the evader and $x_{p1}^0=[-10,1]^T$, $x_{p2}^0=[-9,7]^T$, $x_{p3}^0=[-10,-1]^T$, $x_{p4}^0=[-6,7]^T$, and $x_{p5}^0=[1,2.5]^T$ for the pursuers. Each pursuer has a capture radius of 1.0, and the control inputs are saturated at 1.5 for the evader and 0.8 for the pursuers. The communication topology is depicted in Fig. 3. For the critic NN, the basis functions are chosen as $\phi(\mathbf{x})=[(x^1)^2,x^1x^2,(x^2)^2]^T$ for all agents. The initial weights are set to $\hat{W}_e(0)=\hat{W}_{pi}(0)=[1,1,1]^T$, and the learning rates are uniformly set to $\eta_e=\eta_i=1.0$.

In this case, the target area is a sphere centered at $x_d=[0,0]^T$ with radius $R_{A_0}=1.5$, and $\beta_0=2.5$ is set slightly above $d^e=2$. Fig. 2 (b) shows that the evader successfully reaches the target area Ω_d. The passive pursuers p_4 and p_5 act as dynamic obstacles and tend to converge to each other, while p_2 approaches p_3 due to its limited information. Simultaneously, p_1 moves toward p_3 and pursues the evader.

The RADG game terminates when the evader reaches the target area. However, in order to verify the convergence performance of the algorithm, we let the game last up to 20 seconds. Fig. 4 illustrates the state trajectories of the evader and pursuers p_1, p_2, and p_3. The evader's state $[x_e^1,x_e^2]^T$ converges to the estimated equilibrium $[\hat{x}_e^1,\hat{x}_e^2]^T=\frac{\sum_{i\in\mathcal{N}_e}\beta_i x_{pi}}{\beta_0-d^e}$, while the states of pursuers in $\mathscr{P}_r^{-p}$ asymptotically approach the evader. The instantaneous costs for all agents, shown in Fig. 5, converge to zero, consistent with infinite-horizon optimality. The evolution of the transferred states is depicted in Fig. 6. Figs. 7 and 8 present the control inputs and neural network weight updates, respectively.

Case 3: General PE Game in Non-Uniform Flow. This scenario explores a general pursuit-evasion game within a Rankine vortex model, characterized by a core radius of 30.0 and an angular velocity of 0.03 rad/s, ensuring velocity continuity at the vortex core boundary. Employing the same agent settings and parameters as Case 2, except for β_0,

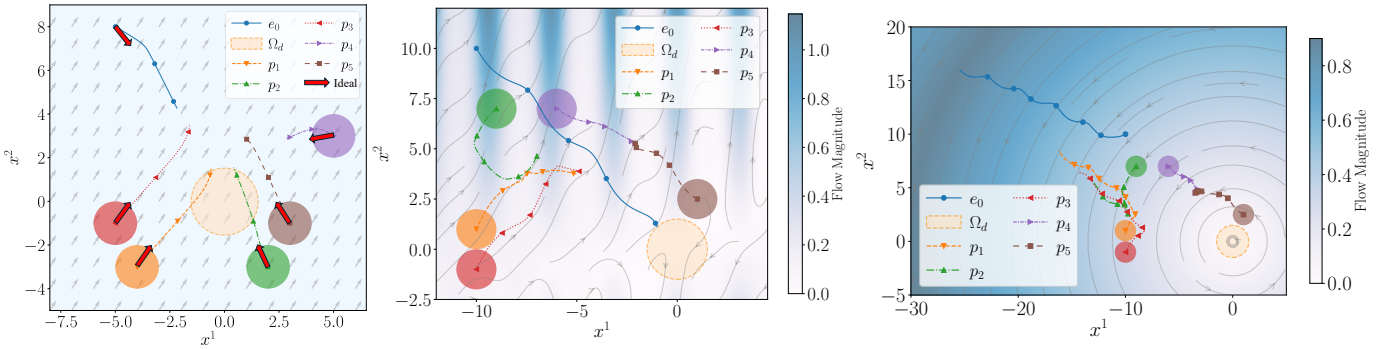

Fig. 2: Agent trajectories across scenarios: (a) RADG in steady, uniform flow, benchmarked against optimal trajectories from [21], (b) RADG in non-uniform flow with communication constraints, and (c) General PE game employing a virtual target.

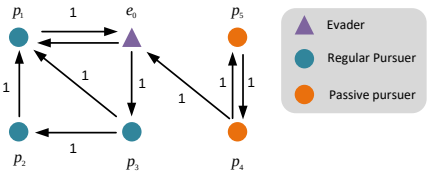

Fig. 3: Communication topology in Case 2.

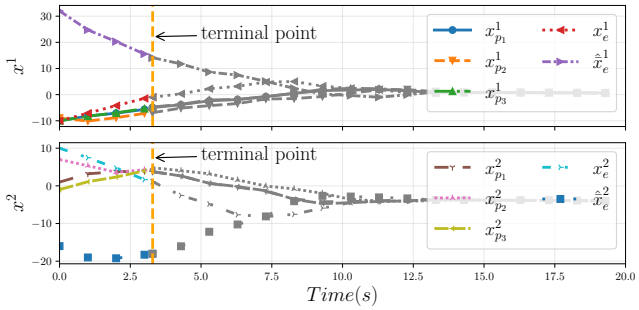

Fig. 4: States of the agents in case 2 with $\beta_0 = 2.5$.

we leverage the principle outlined in Remark 5. By setting β_0 to 1.9, such that $\beta_0 - d^e = -0.1$, the reach-avoid game effectively transitions into a general pursuit-evasion scenario. As illustrated in Fig. 2 (c), the evader diverges from both the pursuers and the virtual target area, while pursuers p_1, p_2, and p_3, being connected to the evader, converge towards it.

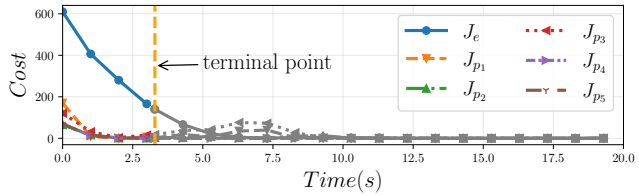

Fig. 5: Costs of agents in case 2 with $\beta_0 = 2.5$.

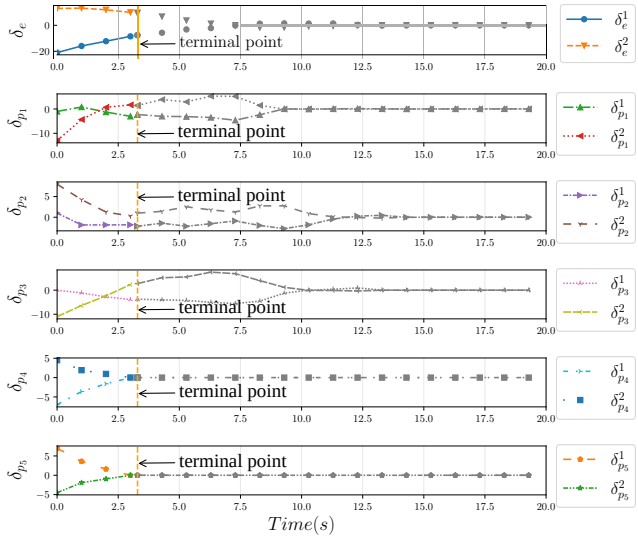

Fig. 6: Transferred states of agents in case 2 with $\beta_0 = 2.5$.

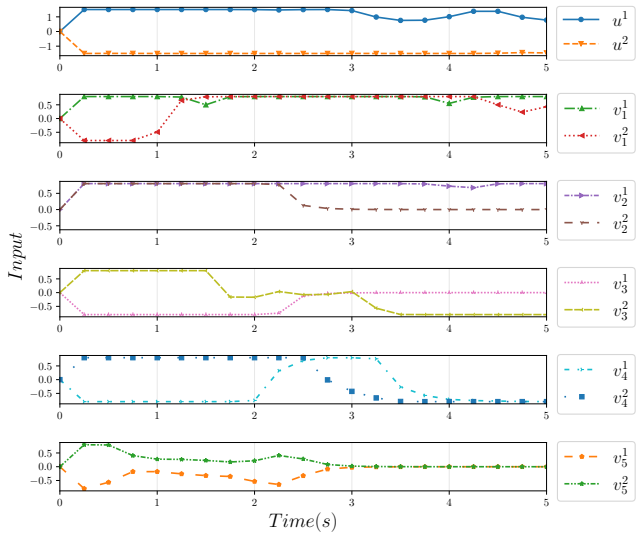

Fig. 7: Inputs of agents in case 2 with $\beta_0 = 2.5$.

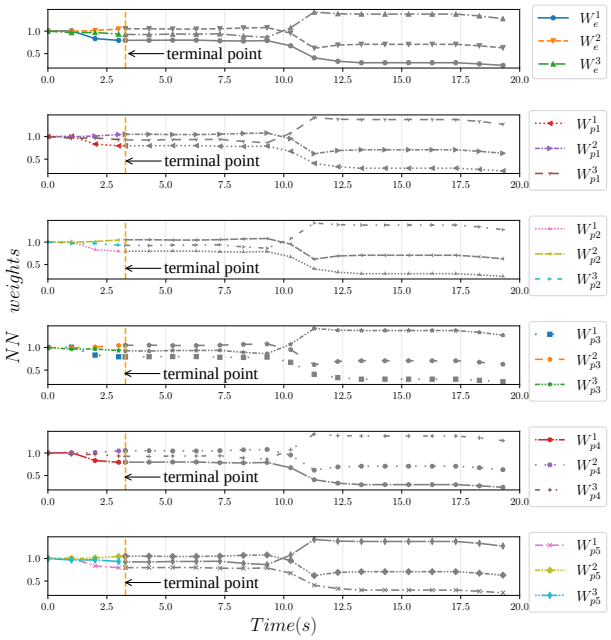

Fig. 8: NN weights in case 2 with $\beta_0 = 2.5$.

B. RADG games with Heterogeneous Agents

This subsection investigates a RADG game with heterogeneous agents governed by distinct dynamics, as summarized in Table I. The models include single/double integrators, Dubins vehicles, planar drones, and the nonlinear system from (72).

Remark 6. *For agents with higher-order dynamics, such as double integrators, Dubins vehicles, and drones, we define their positions as $[x^1, x^2]^T$ and treat the remaining states as auxiliary variables. The interception error (8) and control error (10) are then defined based on these position states. This approach allows the proposed RADG framework to accommodate a wide range of agent dynamics while maintaining the core principles of the reach-avoid game.*

TABLE I: Models of Heterogeneous Agents.

Model	Dynamics	Model	Dynamics
Single Integrator [20], [21], [43]	$\begin{cases} \dot{x}^1 = u^1 \\ \dot{x}^2 = u^2 \end{cases}$	Dubins Vehicle [41], [42]	$\begin{cases} \dot{x}^1 = v\cos(x^3) \\ \dot{x}^2 = v\sin(x^3) \\ \dot{x}^3 = u^1 \end{cases}$
Double Integrator [43]	$\begin{cases} \dot{x}^1 = x^3 \\ \dot{x}^2 = x^4 \\ \dot{x}^3 = u^1 \\ \dot{x}^4 = u^2 \end{cases}$	Drone [44]–[47]	$\begin{cases} \dot{x}^1 = x^3\cos(x^4) \\ \dot{x}^2 = x^3\sin(x^4) \\ \dot{x}^3 = u^1 \\ \dot{x}^4 = u^2 \end{cases}$

Simulations are conducted in a 20×20 workspace, with a spherical target centered at the origin with a radius of $R_{A_0} = 2.0$. Each scenario includes one evader and up to 10 pursuers, with their initial positions and dynamic models randomly assigned. In accordance with Theorem 2, the communication topology is generated to exclude isolated and invisible pursuers. The critic NN employs a 68-dimensional polynomial basis, with weights initialized from a distribution with a mean of 1 and a variance of 0.5. A representative simulation is illustrated in Fig. 9.

The proposed graph-based single-network ADP (GSADP) is benchmarked against three baselines over 50 generated scenarios: Multi-Agent Deep Deterministic Policy Gradient (MADDPG) [2], [37], State-following (StaF) ADP [40], and a Graph-based StaF ADP variant that integrates the graph structure from Section II-B. As detailed in Table II, the proposed online algorithm surpasses the baselines in performance metrics, achieving a decision speed comparable to the offline-trained MADDPG. While integrating the StaF ADP from [40] into our graph-based framework yields a similar win rate, it incurs a substantially higher computational cost, which highlights the computational efficiency of the proposed single-network approach.

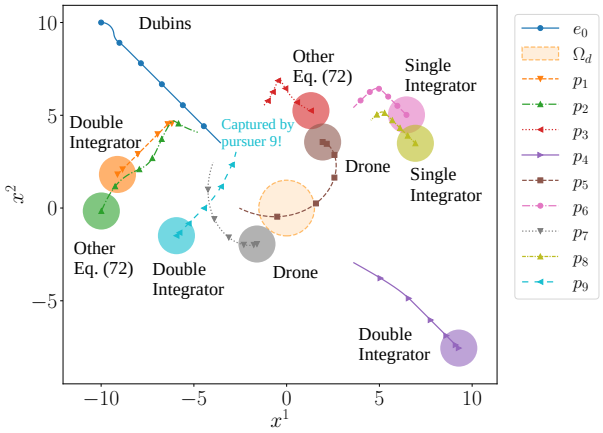

Fig. 9: A representative simulation with proposed method.

TABLE II: Performance Metrics.

Method	Win [a] Rate	Capture [b] Time	Decision [c] Time	Learning Paradigm
GSADP	**42%**	**5.94s**	0.00094s	online
MADDPG [d]	34%	6.74s	**0.00023s**	offline
StaF ADP [e]	10%	/	0.13527s	online
Graph-StaF ADP	40%	6.03s	0.14410s	online

[a] Evader's win rate, all pursuers utilize the proposed algorithm (Eq. (41)).
[b] Average capture time, the evader utilizes the proposed algorithm (Eq. (42)), given as 10s if the evader wins.
[c] Average computational time per decision-making step, including state transfer, control computation, and neural network updates.
[d] The training process for the pursuers was configured in accordance with the specifications provided in [2] while the evader will get a reward when it reaches the target.
[e] StaF ADP in [40] was originally designed for the evader, thus there is no comparison for capture time.

VII. Conclusion and Future Works

This paper addressed SEMP RADG games for nonlinear heterogeneous agents under limited communication and saturated input. First, we reformulated the pursuit-evasion (PE) problem as a distributed optimal control problem to derive Nash equilibrium strategies. Then, by introducing a four-type pursuer categorization (i.e., isolated, passive, invisible, and regular), we defined the Interactive Nash Equilibrium (INE)

for graphical games. A single-network, Concurrent Learning (CL)-based ADP algorithm was developed to compute near-optimal strategies. Furthermore, we established asymptotic capture conditions and extended our result to general PE games and the case with half-space targets. Simulations successfully validated the proposed GSADP framework. This work lays a foundation for future research in the domain of differential games, particularly in enhancing the robustness and adaptability of strategies in complex dynamic environments.

REFERENCES

[1] H. Gong, S. Gong, and J. Li, "Pursuit–evasion game for satellites based on continuous thrust reachable domain," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 6, pp. 4626–4637, 2020.

[2] X. Wang, P. Yi, and Y. Hong, "A hierarchical deep reinforcement learning strategy for collective pursuit-evasion game with partial observations," *IEEE Transactions on Artificial Intelligence*, 2025.

[3] A. Bajcsy, A. Loquercio, A. Kumar, and J. Malik, "Learning vision-based pursuit-evasion robot policies," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 9197–9204.

[4] N. Chen, L. Li, and W. Mao, "Equilibrium strategy of the pursuit-evasion game in three-dimensional space," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 2, pp. 446–458, 2024.

[5] Y. Liu, C. Liu, Y. Meng, B. Jiang, and X. Wang, "Observer-based multi-agent reinforcement learning for pursuit-evasion game with multiple unknown uncertainties," *IEEE Transactions on Automation Science and Engineering*, 2025.

[6] K. Margellos and J. Lygeros, "Hamilton–jacobi formulation for reach–avoid differential games," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1849–1861, 2011.

[7] M. Chen, Z. Zhou, and C. J. Tomlin, "Multiplayer reach-avoid games via pairwise outcomes," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1451–1457, 2016.

[8] Z. Zhou, J. Ding, H. Huang, R. Takei, and C. Tomlin, "Efficient path planning algorithms in reach-avoid problems," *Automatica*, vol. 89, pp. 28–36, 2018.

[9] B. Xue, N. Zhan, M. Fränzle, J. Wang, and W. Liu, "Reach-avoid verification based on convex optimization," *IEEE Transactions on Automatic Control*, vol. 69, no. 1, pp. 598–605, 2023.

[10] H. Fang, P. Yi, D. Deng, and B. He, "Multiple pursuers versus one evader reach-avoid differential games with asymmetric observations," *IEEE Transactions on Automation Science and Engineering*, 2025.

[11] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

[12] K. G. Vamvoudakis, F. L. Lewis, and G. R. Hudas, "Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality," *Automatica*, vol. 48, no. 8, pp. 1598–1611, 2012.

[13] N.-M. T. Kokolakis and K. G. Vamvoudakis, "Safety-aware pursuit-evasion games in unknown environments using gaussian processes and finite-time convergent reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 3, pp. 3130–3143, 2022.

[14] Z. Zheng, P. Zhang, and J. Yuan, "Nonzero-sum pursuit-evasion game control for spacecraft systems: A q-learning method," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 4, pp. 3971–3981, 2023.

[15] V. G. Lopez, F. L. Lewis, Y. Wan, E. N. Sanchez, and L. Fan, "Solutions for multiagent pursuit-evasion games on communication graphs: Finite-time capture and asymptotic behaviors," *IEEE Transactions on Automatic Control*, vol. 65, no. 5, pp. 1911–1923, 2019.

[16] Y. Xu, H. Yang, B. Jiang, and M. M. Polycarpou, "Multiplayer pursuit-evasion differential games with malicious pursuers," *IEEE Transactions on Automatic Control*, vol. 67, no. 9, pp. 4939–4946, 2022.

[17] R. Yan, Z. Shi, and Y. Zhong, "Reach-avoid games with two defenders and one attacker: An analytical approach," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1035–1046, 2018.

[18] L. Liang, F. Deng, Z. Peng, X. Li, and W. Zha, "A differential game for cooperative target defense," *Automatica*, vol. 102, pp. 58–71, 2019.

[19] A. Von Moll, E. Garcia, D. Casbeer, M. Suresh, and S. C. Swar, "Multiple-pursuer, single-evader border defense differential game," *Journal of Aerospace Information Systems*, vol. 17, no. 8, pp. 407–416, 2020.

[20] R. Deng, Y. Sun, Z. Shi, and Y. Zhong, "A geometric tool for two-phase multiplayer reach-avoid games: Ellipses," in *2024 European Control Conference (ECC)*. IEEE, 2024, pp. 46–51.

[21] R. Deng, W. Zhang, R. Yan, Z. Shi, and Y. Zhong, "Multiple-pursuer single-evader reach-avoid games in constant flow fields," *IEEE Transactions on Automatic Control*, vol. 69, no. 3, pp. 1789–1795, 2023.

[22] A. Amirkhani and A. H. Barshooi, "Consensus in multi-agent systems: a review," *Artificial Intelligence Review*, vol. 55, no. 5, pp. 3897–3935, 2022.

[23] Y. Hong, J. Hu, and L. Gao, "Tracking control for multi-agent consensus with an active leader and variable topology," *Automatica*, vol. 42, no. 7, pp. 1177–1182, 2006.

[24] T. Wang and J. Huang, "Leader-following event-triggered adaptive practical consensus of multiple rigid spacecraft systems over jointly connected networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 12, pp. 5623–5632, 2021.

[25] S. Zhang, J. Liu, H. Zhang, W. Wang, and Z. Zhang, "Tracking control for multi-agent systems with model switching and topological switching: A novel dual-switch-based dynamic event-triggered approach," *IEEE Transactions on Automation Science and Engineering*, 2024.

[26] T. Wang and J. Huang, "Leader-following consensus of multiple spacecraft systems with disturbance rejection over switching networks by adaptive learning control," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 5, pp. 3001–3020, 2022.

[27] Y.-Y. Qian, M. Liu, Y. Wan, F. L. Lewis, and A. Davoudi, "Distributed adaptive nash equilibrium solution for differential graphical games," *IEEE Transactions on Cybernetics*, vol. 53, no. 4, pp. 2275–2287, 2021.

[28] X. Ma, M. Lv, K. Duan, W. Xie, J. Yang, and W. Zhang, "Multi-player pursuit-evasion game with interaction constraints: A cooperative game theoretic approach based on coalition structure," *IEEE Transactions on Automation Science and Engineering*, 2025.

[29] X. Dong, H. Zhang, and Z. Ming, "Adaptive optimal control via q-learning for multi-agent pursuit-evasion games," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 71, no. 6, pp. 3056–3060, 2024.

[30] R. Yan, X. Duan, Z. Shi, Y. Zhong, and F. Bullo, "Matching-based capture strategies for 3d heterogeneous multiplayer reach-avoid differential games," *Automatica*, vol. 140, p. 110207, 2022.

[31] R. Yan, R. Deng, H. Lai, W. Zhang, Z. Shi, and Y. Zhong, "Homicidal chauffeur reach-avoid games via guaranteed winning strategies," *IEEE Transactions on Automatic Control*, vol. 69, no. 4, pp. 2367–2382, 2023.

[32] P. Shivakumar and K. H. Chew, "A sufficient condition for nonvanishing of determinants," *Proceedings of the American mathematical society*, pp. 63–66, 1974.

[33] X. He and J. Huang, "Distributed nash equilibrium seeking with dynamics subject to disturbance of unknown frequencies over jointly strongly connected switching networks," *IEEE Transactions on Automatic Control*, vol. 69, no. 1, pp. 606–613, 2023.

[34] Z. Liu and J. Huang, "Distributed nash equilibrium seeking for uncertain euler–lagrange systems over jointly strongly connected networks," *IEEE Transactions on Automatic Control*, vol. 69, no. 12, pp. 8293–8307, 2024.

[35] Vamvoudakis, Kyriakos G and Lewis, Frank L, "Multi-player non-zero-sum games: Online adaptive learning solution of coupled hamilton–jacobi equations," *Automatica*, vol. 47, no. 8, pp. 1556–1569, 2011.

[36] A. Das, D. Wu, B. A. Bhatti, and M. Kamaludeen, "Approximate dynamic programming with enhanced off-policy learning for coordinating distributed energy resources," *IEEE Transactions on Sustainable Energy*, vol. 15, no. 3, pp. 1614–1626, 2024.

[37] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in neural information processing systems*, vol. 30, 2017.

[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, 2010.

[39] X. Yang and H. He, "Self-learning robust optimal control for continuous-time nonlinear systems with mismatched disturbances," *Neural Networks*, vol. 99, pp. 19–30, 2018.

[40] P. Deptula, H.-Y. Chen, R. A. Licitra, J. A. Rosenfeld, and W. E. Dixon, "Approximate optimal motion planning to avoid unknown moving avoidance regions," *IEEE Transactions on Robotics*, vol. 36, no. 2, pp. 414–430, 2019.

[41] N.-M. T. Kokolakis and K. G. Vamvoudakis, "Bounded rational dubins vehicle coordination for target tracking using reinforcement learning," *Automatica*, vol. 149, p. 110732, 2023.

[42] G. Xu, X. Kang, H. Yang, Y. Wu, W. Liu, J. Cao, and Y. Liu, "Distributed multi-vehicle task assignment and motion planning in dense environments," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 4, pp. 7027–7039, 2023.
[43] D. Yao, H. Li, and Y. Shi, "Event-based average consensus of disturbed mass via fully distributed sliding mode control," *IEEE Transactions on Automatic Control*, vol. 69, no. 3, pp. 2015–2022, 2023.
[44] S. Baldi, S. Roy, K. Yang, and D. Liu, "An underactuated control system design for adaptive autopilot of fixed-wing drones," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 4045–4056, 2022.
[45] E. A. Gedefaw, N. B. Abera, and C. M. Abdissa, "A review of modeling and control techniques for unmanned aerial vehicles," *Engineering Reports*, vol. 7, no. 6, p. e70215, 2025.
[46] E. W. Metekia, W. A. Asfaw, C. M. Abdissa, and L. N. Lemma, "Control of a fixed wing unmanned aerial vehicle using a robust fractional order controller," *Scientific Reports*, vol. 15, no. 1, p. 19954, 2025.
[47] T. K. Mohammed, W. A. Asfaw, C. M. Abdissa, and L. N. Lemma, "Constrained robust adaptive control design for fixed wing uav under parameter uncertainties and external disturbances," *Engineering Research Express*, vol. 7, no. 2, p. 025254, 2025.