Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Reinforcement Learning-Based Dynamic Event-Triggered Nonlinear Games Under Incomplete and Asymmetric Information

Tao Ren
Air Force Engineering University
Lei Shao
zj_shaolei_2021@163.com

Air Force Engineering University
Shu Li
Liaoning University of Technology

Research Article

Keywords: Reinforcement learning (RL), dynamic event-triggered (DET), nonlinear game, maneuverability constraints, incomplete and asymmetric information

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11373276/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Reinforcement Learning-Based Dynamic Event-Triggered Nonlinear Games Under Incomplete and Asymmetric Information

Tao Ren[1], Lei Shao[1*†], Shu Li[2†]

[1]Air Defence and Antimissile School, Air Force Engineering University, Xi'an, 710038, China.

[2]Provincial Key Laboratory of Intelligent Control Theory and Application, Liaoning University of Technology, Jinzhou, 121001, China.

*Corresponding author(s). E-mail(s): zj_shaolei_2021@163.com;
Contributing authors: rentaolab2022@outlook.com;
li_shu43@lnut.edu.cn;
†These authors contributed equally to this work.

Abstract

This article proposes a dynamic event-triggered reinforcement-learning-based composite game interception method for unknown maneuvering targets under incomplete and asymmetric information and maneuverability constraints. Unlike many existing game-based interception methods that rely on a worst-case target response assumption, the proposed method evaluates the real-time effect of target maneuvers and adaptively switches between a nongame optimal strategy and a game-based strategy. This reduces unnecessary policy conservatism while preserving robustness against strongly evasive maneuvers. To address unknown target intention preference, a disturbance observer is designed to reconstruct the target maneuvering input, and the intention weighting matrix is inferred through the target-side optimality condition. Critic, actor, and virtual actor neural networks are developed to approximate the value function, the interceptor policy, and the virtual target policy, respectively. Dynamic event-triggered mechanisms are further introduced to reduce redundant policy updates and situation evaluations. The closed-loop system state, neural-network weight estimation errors, and disturbance observation error are proved to be uniformly ultimately bounded, and Zeno behavior is excluded. Simulation results demonstrate the effectiveness of the proposed method.

Keywords: Reinforcement learning (RL), dynamic event-triggered (DET), nonlinear game, maneuverability constraints, incomplete and asymmetric information

1 Introduction

With the rapid development of aerospace technologies and intelligent unmanned systems, adversarial missions such as aerospace defense, target capture, on-orbit pursuit evasion, and unmanned swarm confrontation have received increasing attention [1–4] . Intercepting an unknown maneuvering target is a representative problem in such missions. In this problem, the interceptor is required to generate an effective interception strategy under complex dynamic environments, limited sensing information, and constrained maneuverability, while the target may actively maneuver to degrade interception performance and improve its survivability. Since the objectives of the interceptor and the target are intrinsically conflicting, their interaction can be naturally modeled as a two-player game under the ideal assumption of complete and symmetric information [5, 6]. However, the relative dynamics between the interceptor and the target exhibit pronounced nonlinearities that cannot be accurately represented by conventional linearized models. Consequently, interception against maneuvering targets constitutes a challenging nonlinear game problem, whose solution is substantially more difficult than that of linear-quadratic differential games.

For nonlinear games, the optimal interception and evasion strategies are generally characterized by the Nash equilibrium (NE) associated with the Hamilton–Jacobi–Isaacs (HJI) equation. However, the analytical solution of the HJI equation is difficult to obtain in practical interception scenarios due to nonlinear dynamics, input constraints, and uncertainties in target behavior. RL and adaptive dynamic programming (ADP) provide effective tools for approximating value functions and updating policies online [7–10]. Nevertheless, most existing RL-based game-learning methods are developed under relatively ideal information structures, where the opponent's cost-function parameters or intention-related weights are assumed to be known or fixed. This assumption is restrictive for intercepting unknown noncooperative targets, since the target weighting matrix usually reflects its implicit maneuvering intention, namely the tradeoff between evasion performance and maneuvering energy consumption. If such intention information is unavailable or inaccurately described, the HJI equation constructed by the interceptor may deviate from the actual adversarial interaction, resulting in inaccurate equilibrium policy learning. Although inverse optimal control and inverse RL have been used to infer unknown cost-function parameters from observed behaviors [13–15], integrating online target intention estimation into nonlinear game-based interception under asymmetric and incomplete information remains insufficiently studied.

Another important issue is the conservatism of game-based interception strategies. The NE essentially corresponds to a stable worst-case response: no player can improve its own payoff by unilaterally changing its strategy when the strategy of the other player remains unchanged [17]. From the interceptor's perspective, such a design improves robustness against strongly maneuvering targets. However, in real missions,

the target may not always adopt the most adverse rational evasion strategy. When the target exhibits weak, intermittent, or phase-dependent maneuvers, always treating it as a fully rational adversary may lead to unnecessary energy consumption and excessive policy conservatism. Therefore, it is necessary to evaluate the actual influence of target maneuvers on the current interception situation and adaptively select an appropriate interception mode.

Moreover, the online implementation of RL-based game strategies usually requires iterative policy updates and frequent information exchange. In high-speed interception scenarios, continuous state evolution and policy interaction generate substantial data transmission and computational demands. Conventional time-triggered schemes update neural-network weights and control policies at every sampling instant, which may be inefficient for aerospace interception systems with stringent real-time requirements and limited onboard resources. Event-triggered mechanisms offer a promising solution by updating information only when necessary [18–20]. However, most existing event-triggered methods rely on static triggering rules [21, 22], which are difficult to adapt to rapidly evolving adversarial situations. Although dynamic event-triggered mechanisms improve flexibility by introducing internal dynamic variables or auxiliary dynamics [23–26]. their application to unknown maneuvering target interception remains limited, especially when the target intention is unknown, the interceptor is maneuverability constrained, and the adversarial situation evolves continuously.

Based on the above analysis, this article proposes an RL-based dynamic event-triggered composite game interception scheme for nonlinear games with maneuverability constraints under incomplete and asymmetric information. The proposed scheme simultaneously addresses target-intention uncertainty, strategy conservatism, and online computational burden. The main contributions are summarized as follows.

1) Unlike existing game-based interception methods [4], [8], [16], which treat the target as a worst-case adversarial opponent throughout the entire engagement, this article proposes a situation-aware composite interception architecture. The constructed adaptive switching function quantifies the real-time influence of target maneuvers on the current interception situation, and the game-based mode is activated only when this influence exceeds a designed threshold; otherwise, the nongame optimal mode is adopted. This design preserves the robustness of worst-case game strategies when the target performs adversarial maneuvers, while avoiding the excessive conservatism and redundant control effort induced by persistent game-based treatment when the target maneuver is weak or nongame.

2)Different from game-learning methods that assume the opponent's intention weighting parameters to be known [5], and from intention inference methods developed for human-machine cooperative interaction scenarios [13], [14], [15], this article proposes an online target-intention estimation mechanism embedded in adversarial game learning. First, a disturbance observer is employed to reconstruct the target's actually executed maneuver from engagement measurements; then, by inverting the optimality mapping between the target's optimal strategy and its cost weights, the unknown intention weighting matrix is recovered online, with a provably bounded estimation error. The estimated intention is incorporated into the virtual actor network,

so that the virtual target policy employed in the interceptor's game solver continuously approaches the true target policy, thereby improving the accuracy of equilibrium strategy learning under incomplete information.

3)Compared with time-triggered reinforcement learning implementations and static triggering rules [21], [22] this article proposes a nested dual dynamic event-triggered mechanism, in which the two triggering mechanisms undertake distinct responsibilities: the state-triggered mechanism schedules state transmission and neural-network policy updates, while the situation-triggered mechanism is responsible for judging the switching of interception modes and is evaluated only on the triggering-time sequence of the state-triggered mechanism. Since both mechanisms verify their triggering conditions only at periodic sampling instants, their inter-event times are lower-bounded by the sampling period, thereby excluding Zeno behavior by construction. While reducing the computational and transmission burden of online learning, the nested design suppresses frequent mode switching caused by small-amplitude fluctuations of the situation function.

The remainder of this article is organized as follows. Section 2 formulates the nonlinear game model of the two players and derives the composite game equations. Section 3 develops two dynamic triggering mechanisms. Section 4 presents the dynamic event-triggered game-learning method, including neural networks (NNs) design, stability analysis, and the target intention matrix estimation mechanism. Section 5 provides numerical simulations to validate the proposed method. Finally, Section 6 concludes this article.

2 Problem formulation and preliminaries

2.1 Nonlinear games under incomplete and asymmetric information

To enable the development and validation of the proposed method, a typical interception engagement scenario is considered, as shown in Fig. 1. The following three -dimensional relative motion dynamics of the two players are established [28]:

$$\begin{cases} \ddot{r} = r\dot{q}_\varepsilon^2 + r\dot{q}_\beta^2\cos^2 q_\varepsilon + a_{Tr} - a_{Mr} \\ \ddot{q}_\varepsilon = -2\dot{r}\dot{q}_\varepsilon/r - \dot{q}_\beta^2\cos q_\varepsilon \sin q_\varepsilon - \left(a_{M\varepsilon} - a_{T\varepsilon}\right)/r \\ \ddot{q}_\beta = -2\dot{r}\dot{q}_\beta/r + 2\dot{q}_\varepsilon\dot{q}_\beta\tan q_\varepsilon + \left(a_{M\beta} - a_{\mathrm{T}\beta}\right)/r\cos q_\varepsilon \\ \dot{a}_{M\varepsilon} = -\left(a_{M\varepsilon} - u_\varepsilon\right)/\tau_\varepsilon \\ \dot{a}_{M\beta} = -\left(a_{M\beta} - u_\beta\right)/\tau_\beta \end{cases} \tag{1}$$

where, r is the relative distance between the interceptor and the target, q_ε and q_β are the relative line-of-sight angles. The acceleration components of the interceptor and the target along the line-of-sight coordinate axes are respectively denoted by $a_M = [a_{Mr}, a_{M\varepsilon}, a_{M\beta}]^{\mathrm{T}}$ and $a_T = [a_{Tr}, a_{T\varepsilon}, a_{T\beta}]^{\mathrm{T}}$. τ_ε and τ_β are the time constants of the interceptor acceleration variation. In general, the interception process aims to drive the line-of-sight angular rates to zero. Therefore, a_{Mr} is not treated as the interceptor input. Instead, $a_{M\varepsilon}$ and $a_{M\beta}$ are used as the input variables for designing

the interception strategy, The corresponding target input is also given by $a_{T\varepsilon}$ and $a_{T\beta}$ [28].

According to (1), the continuous-time nonlinear game system can be formulated as follows [16]:

$$\dot{x}(t) = f(x) + g(x)u(t) + h(x)v(t) \tag{2}$$

where, the system states are $x = [q_\varepsilon, \dot{q}_\varepsilon, q_\beta, \dot{q}_\beta]^{\mathrm{T}}$, considering the limited maneuverability of the interceptor,$u(t)$ is the constrained interceptor input, where each component satisfies $|u| < \delta$, and δ is a positive constant, and $v(t)$ is the control input of the target. The function $f(x)$ represents the system dynamics, while $g(x)$ and $h(x)$ are known continuous nonlinear functions satisfying $\|g(x)\| \leqslant \bar{g}$, $\|h(x)\| \leqslant \bar{h}$, where $\bar{g}$ and $\bar{h}$ are positive constants.

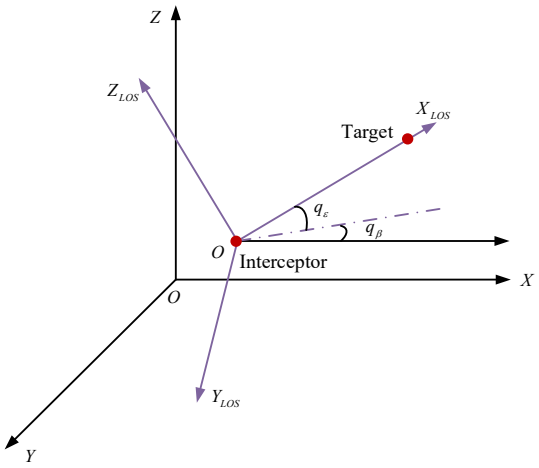


Fig. 1 Geometry of the three-dimensional interception engagement in the inertial and LOS frames.

Define the following finite-horizon cost function:

$$J(x,t) = \Phi(x(t_f), t_f) + \int_t^{t_f} r(x,u,v)dt \tag{3}$$

where, $r(x,u,v) = Q(x) + U(u,R_1) - v^{\mathrm{T}}R_2 v$, Q is a positive definite function, R_1 and R_2 denote the intention-related weighting matrices of the two players. In this article, target intention does not refer of whether the target performs evasive maneuvers. Instead, it refers to the latent strategic preference embedded in the target cost function, where R_2 characterizes the target's tradeoff between evasion performance and maneuvering expenditure. $\Phi = x^{\mathrm{T}}(t_f)x(t_f)$ is the terminal cost function, t_f is the terminal time, and $U(u,R_1)$ is a nonquadratic utility function used to handle input saturation constraints [16], defined as

$$U(u, R_1) = 2\int_0^u \delta \tanh^{-\mathrm{T}}\left(\frac{\tau}{\delta}\right) R_1 d\tau$$
$$= 2\delta\left(\tanh^{-1}\left(\frac{u}{\delta}\right)\right)^{\mathrm{T}} R_1 u + \delta^2 R_1 \ln\left(1 - \left(\frac{u}{\delta}\right)^2\right) \tag{4}$$

where the remaining time is denoted by $\tau = t_f - t$.

The interceptor and the target optimize their respective strategies by minimizing and maximizing the predefined cost function, respectively, to achieve their individual objectives. Such an interaction is typically modeled as a two-player zero-sum (ZS) game, which aims to obtain the saddle-point solution of the associated optimization problem. Assume that the saddle point exists and is unique. Then, the optimal strategy pair $\{u^*, v^*\}$ satisfies the following saddle point equilibrium condition [16]:

$$J^*(x_0, u^*, v) \leqslant J^*(x_0, u^*, v^*) \leqslant J^*(x_0, u, v^*)$$

Based on the continuous differentiability of the cost function, define the Hamiltonian function as [27]:

$$H(x, u, v) = (\nabla_x J)^{\mathrm{T}} [f(x) + g(x) u(t) + h(x) v(t)] + \nabla_t J$$
$$+ Q(x) + U(u, R_1) - v^{\mathrm{T}} R_2 v$$

By setting $\frac{\partial H}{\partial u} = 0$ and $\frac{\partial H}{\partial v} = 0$, the optimal strategies can be obtained as:

$$u^* = -\frac{1}{2}\delta \tanh(\eta^*) \tag{5}$$

$$v^* = \frac{1}{2} {R_2}^{-1} h^{\mathrm{T}}(x) \nabla_x J^* \tag{6}$$

where, $\eta^* = \frac{1}{\delta} R_1^{-1} g^{\mathrm{T}}(x) \nabla_x J^*$.

Based on the (5) and (6), yields the finite-horizon HJI equation:

$$\begin{aligned} H(x, u^*, v^*) &= (\nabla_x J^*)^{\mathrm{T}} [f(x) + g(x) u^*(t) + h(x) v^*(t)] \\ &\quad + Q(x) + U^*(u^*, R_1) - {v^*}^{\mathrm{T}} R_2 v^* + \nabla_t J^* \\ &= (\nabla_x J^*)^{\mathrm{T}} f(x) + \nabla_t J^* + Q(x) \\ &\quad - \frac{1}{2} R_1 \delta^2 \ln\left(1 - \tanh^2(\eta^*)\right) \\ &\quad + \frac{1}{4} \nabla_x {J^*}^{\mathrm{T}} h(x) {R_2}^{-1} h^{\mathrm{T}}(x) \nabla_x J^* \\ &= 0 \end{aligned} \tag{7}$$

Remark 1: In the above game-strategy solution process, both players have their own cognition of the opponent's intention parameters and their own judgment of the opponent's intention. Such judgments are private information rather than public information. Let the intention information possessed by the interceptor be $I_m = (R_1^m, R_2^m)$, and that possessed by the target be $I_m = (R_1^t, R_2^t)$. The above derivations of the optimal strategies are established under an ideal game setting, where both players are assumed to have identical intention preference matrices, i.e.,$R_1^m = R_1^t = R_1$, $R_2^m = R_2^t = R_2$, or the interceptor has complete knowledge of the target's intention.

Definition 1: In practical interception scenarios, the interceptor and the target generally possess different mission objectives and decision preferences. Due to factors such as maneuverability constraints, control costs, and engagement environments, the interceptor tends to prioritize reducing interception risks and control consumption, whereas the target aims to improve its survivability through evasive maneuvers. Therefore, the weighting matrices characterizing the decision preferences of the two players are generally different. Moreover, the target intention preference matrix is typically private decision information that cannot be directly obtained by the interceptor, resulting in an incomplete and asymmetric information structure.

Accordingly, instead of the conventional complete-information assumption, this article considers an interception scenario with unknown target intention under incomplete and asymmetric information. Under such an information structure, the interceptor cannot directly access the actual target strategy required for equilibrium strategy computation due to the lack of target intention information. Consequently, the interceptor cannot accurately update its policy according to the actual adversarial interaction. To address this issue, a virtual target strategy is introduced to approximate the unknown target policy and provide the necessary adversarial information for game-based strategy learning [16].

For the interceptor, the uncertainty of the target intention means that its decision-making can only rely on locally available private information when designing the interception strategy. Therefore, the cost functions of the interceptor are redesigned as:

$$J_m = \Phi\left(x\left(t_f\right), t_f\right) + \int_t^{t_f} Q\left(x\right) + U\left(u, R_1^m\right) - v_v^{\mathrm{T}} R_2^m v_v dt \tag{8}$$

where, v_v denotes the virtual target strategy, namely the interceptor's estimate of the target strategy.

For the target, although the target knows the interceptor's intention, it does not know the actual strategy adopted by the interceptor during the game process.

$$J_t = -\left(\Phi\left(x\left(t_f\right), t_f\right) + \int_t^{t_f} Q\left(x\right) + u_v^{\mathrm{T}} R_1^t u_v - v^{\mathrm{T}} R_2^t v dt\right) \tag{9}$$

where u_v denotes the virtual interceptor strategy, namely the target's estimate of the interceptor strategy. Moreover, $J_m \neq -J_t$.

Remark 2: For interception scenarios with unknown target intentions, the complete-information assumption in Remark 1 is relaxed to an incomplete and asymmetric information setting. The matrix $Q(x)$ and R_1 are assumed to be common knowledge,$R_1^m = R_1^t = R_1$.The target intention weighting matrix R_2 is treated as private information that cannot be directly accessed by the interceptor.

2.2 Composite game equations

The cost function in (8) reflects both instantaneous maneuvering cost and long-term interception performance [16]. However, the target may not remain strongly adversarial during the entire engagement. Its maneuvering effect can vary with the flight phase, relative geometry, and remaining interception time. If the target only performs weak or intermittent maneuvers, always treating it as a rational adversary may lead to an overly conservative interception strategy. Therefore, a situation-dependent switching mechanism is introduced to evaluate the actual influence of the target maneuver and select an appropriate interception mode.

Therefore, define the following adaptive switching function based on the game situation of the two players:

$$c = \begin{cases} 1, \ \varsigma > h_\rho \\ 0, \ \varsigma \leqslant h_\rho \end{cases} \tag{10}$$

where h_ρ is a threshold parameter to be designed. The variable $\varsigma = \dot{\rho}_\pi, \ \pi = \varepsilon, \beta$ is used to characterize the degree to which the target maneuver affects the current interception situation. By analyzing this quantity, one can determine whether the current target maneuver has a sufficiently significant adverse influence on the interception process. To quantitatively describe the influence of the target maneuver on the interception situation, the following situation-evaluation functions are introduced:

$$\rho_\pi = k_z Z_{t\pi}^2 \left(t_{k+1}^s\right) + \int_{t_k^s}^{t_{k+1}^s} v_\pi^2 (t) \, dt$$

where, $k_z > 0$ is a weighting coefficient, and $\{t_k^s\}_{k=0}^{\infty}$ is the dynamic event-triggered sampling-time sequence designed later. This avoids continuous judgment throughout the entire game process, excludes Zeno behavior, and reduces unnecessary computational burden. The terms v_π is the target-strategy estimates obtained by the observer designed later. $Z_{t\pi}$ is the zero-effort-miss, which describe interception accuracy [28] , and are expressed as:

$$Z_{t\pi}(t) = -\dot{r}\dot{q}_\pi t_{go}^2 + Z_{\pi 0}$$

where, $Z_{\pi 0}$ is the initial relative distances.

Based on the above switching mechanism, the interceptor can adaptively select the strategy-solution mode according to the influence of the target maneuver on the current interception situation. If the target maneuver has a significant influence on the current interception situation, the interception problem is formulated as a nonlinear game problem and solved to obtain the game-based interception strategy. If the target

maneuver has a limited influence on the current interception situation, it is treated as an estimable external disturbance, and the problem is transformed into a unilateral optimal-control problem for the interceptor. Accordingly, the interceptor cost function is transformed into:

$$J_m = \Phi\left(x\left(t_f\right), t_f\right) + \int_t^{t_f} Q\left(x\right) + U\left(u\right) - c v_v^{\mathrm{T}} R_2^m v_v dt \tag{11}$$

Based on the nonlinear game system in (2) , the finite-horizon cost functions with maneuverability constraints in (3) and (4) , and the incomplete and asymmetric information structures in (8), (9) and (11), the objective of this article is to develop a dynamic event-triggered composite game-learning framework for unknown maneuvering target interception under incomplete and asymmetric information. The overall architecture of the proposed method is shown in Fig. 2.

Remark 3: It is worth noting that the proposed composite strategy switching mechanism does not aim to identify the existence of target maneuvering behavior itself. Rather, it evaluates the extent to which target maneuvers affect the current interception process by integrating target motion information with the engagement situation, and accordingly determines whether the game-based mode should be activated. This design focuses on the practical influence of target maneuvers during the current interception phase instead of relying on a fixed maneuver classification criterion, thereby providing a more general applicability.

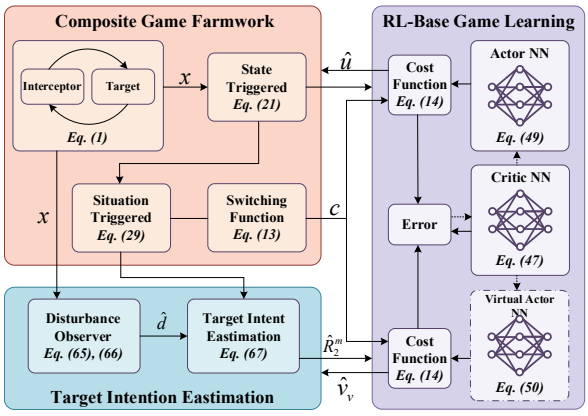


Fig. 2 The overall structural farmwork in this paper.

3 Dynamic event-triggered strategy design

This section proposes two event-triggered mechanisms. The first is a state-triggered mechanism, which is used to reduce the computational and transmission burden of the RL network and thus decrease the computational cost of the system. The second is a

situation triggered mechanism, which is designed to avoid frequent strategy switching caused by small-amplitude fluctuations of the situation function.

Define the sampled state and the error between the actual state and the sampled state as:

$$\hat{x}_k(t) = \begin{cases} x(t_k^x), & \forall t \in \left[t_k^x, t_{k+1}^x\right) \\ x(t), & t = t_{k+1}^x \end{cases} \tag{12}$$

$$e_x(t) = \hat{x}_k(t) - x(t) \tag{13}$$

where, $\{t_k^x\}_{k=0}^{\infty}$ is a monotonically increasing triggering-time sequence.

Based on (5) and (6), the event-triggered optimal strategies are obtained as:

$$u^*(t_k^x) = -\frac{1}{2}\delta \tanh\left(\eta^*(t_k^x)\right) \tag{14}$$

$$v^*(t_k^x) = \frac{1}{2}{R_2}^{-1} h^{\mathrm{T}}(x(t_k^x)) \nabla_x J^*(t_k^x) \tag{15}$$

where $\eta^*(t_k^x) = \frac{1}{\delta} R_1^{-1} g^{\mathrm{T}}(t_k^x) \nabla_x J^*$.

Accordingly, system (2) can be expressed as:

$$\dot{x}(t) = f(x) + g(x) u^*(t_k^x) + h(x) v^*(t_k^x)$$

Assumption 1[17]: The optimal strategies are Lipschitz continuous, there exist positive constants λ_u and λ_v such that:

$$\|u^*(t_k^x) - u^*(t)\| \leqslant \lambda_u \|e_x(t)\|$$

$$\|v^*(t_k^x) - v^*(t)\| \leqslant \lambda_v \|e_x(t)\|$$

Assumption 2[16]:The optimal cost function exists and is continuously differentiable. That is, there exist positive constants λ_x and λ_t such that $\|\nabla_x J_m^*\| \leqslant \lambda_{Jx}$ and $\|\nabla_t J_m^*\| \leqslant \lambda_{Jt}$.

Theorem 1: For the nonlinear zero-sum game system and its finite-horizon cost function, if Assumption 2 and the following event-triggered condition hold, then the system state is uniformly ultimately bounded (UUB)[17].

The dynamic event-triggered condition is defined as:

$$t_{k+1}^x = \inf\left\{ t = t_k^x + \iota\hbar \middle| \|e_x(t)\|^2 \geqslant \frac{\left(1 - {k_D}^2\right) \lambda_{\min}(Q) \|x\|^2 + \gamma}{s \|R_1\| \lambda_u^2 + \|R_2\| \lambda_v^2} \right\} \tag{16}$$

where, the triggering mechanism designed in this section transmits and updates the system state only at periodic sampling instants. $\iota \in \mathbb{N}$, $\hbar > 0$ is the sampling period of the event-triggered mechanism, $\lambda_{\min}(Q)$ is the minimum eigenvalue of Q, $0 < k_D < 1$ is an event-triggered parameter, $s = \delta\lambda_{\tanh}$, $\lambda_{\tanh}$ is the Lipschitz constant on compact sets of $\tanh^{-1}(\cdot)$.

The internal dynamic variable γ satisfies:

$$\dot{\gamma} = -k_\eta \left(\gamma\left(t\right)\right) + \left(1 - {k_D}^2\right) \lambda_{\min}\left(Q\right) \left\|x\right\|^2 - \left(s \left\|R_1\right\| \lambda_u^2 + \left\|R_2^m\right\| \lambda_v^2\right) \left\|e_x\left(t\right)\right\|^2$$

where, k_η is a positive constant and $\gamma\left(0\right) > 0$.
Lemma 1[9]: For any $t \in [0, t_\infty)$, the following inequalities hold:

$$\gamma + k_\eta \left(\left(1 - {k_D}^2\right) \lambda_{\min}\left(Q\right) \left\|x\right\|^2 - \left(s \left\|R_1\right\| \lambda_u^2 + \left\|R_2^m\right\| \lambda_v^2\right) \left\|e_x\left(t\right)\right\|^2\right) \geqslant 0$$

Proof: Choose the following Lyapunov function:

$$V_k = J_m^* + \gamma$$

Taking the first derivative and substituting (2) yield:

$$\begin{aligned}
\dot{V}_k &= \dot{J}_m^* + \dot{\gamma} \\
&= \left(\nabla_x \dot{J}_m^*\right)^{\mathrm{T}} \left(f\left(x\right) + g\left(x\right) u^*\left(t\right) + h\left(x\right) v^*\left(t_k^x\right)\right) \\
&\quad + \left(\nabla_x \dot{J}_m^*\right)^{\mathrm{T}} g\left(x\right) \left(u^*\left(t_k^x\right) - u^*\left(t\right)\right) + \nabla_t J_m^* \\
&\quad + \left(\nabla_x \dot{J}_m^*\right)^{\mathrm{T}} h\left(x\right) \left(v^*\left(t_k^x\right) - v^*\left(t\right)\right) \\
&\quad - k_\eta \left(\gamma\left(t\right)\right) + \left(1 - {k_D}^2\right) \lambda_{\min}\left(Q\right) \left\|x\right\|^2 \\
&\quad - \left(s \left\|R_1\right\| \lambda_u^2 + \left\|R_2^m\right\| \lambda_v^2\right) \left\|e_x\left(t\right)\right\|^2
\end{aligned} \tag{17}$$

Based on (5), (6) and (7) we get:

$$\begin{aligned}
\dot{V}_k &= 2\left(v^*\right)^{\mathrm{T}} R_2^m \left(v^*\left(t_k^x\right) - v^*\left(t\right)\right) + c\left(v^*\right)^{\mathrm{T}} R_2^m v^* \\
&\quad - 2\delta \tanh^{-\mathrm{T}}\left(\frac{u^*\left(t\right)}{\delta}\right) R_1 \left(u^*\left(t_k^x\right) - u^*\left(t\right)\right) \\
&\quad - k_\eta \left(\gamma\left(t\right)\right) + \left(1 - {k_D}^2\right) \lambda_{\min}\left(Q\right) \left\|x\right\|^2 \\
&\quad - \left(s \left\|R_1\right\| \lambda_u^2 + \left\|R_2^m\right\| \lambda_v^2\right) \left\|e_x\left(t\right)\right\|^2 - Q\left(x\right) - U^*\left(u^*\right)
\end{aligned} \tag{18}$$

Further, based on (4), we obtain:

$$U^*\left(u^*\right) = 2 \int_{u^*\left(t_k^x\right)}^{u^*\left(t\right)} \delta \tanh^{-\mathrm{T}}\left(\frac{\tau}{\delta}\right) R_1 d\tau + U\left(u^*\left(t_k^x\right)\right) \tag{19}$$

By using Assumptions 1 and 2 as well as the Lipschitz property of $\tanh^{-1}(\cdot)$, substituting (19) into (18), yield:

$$\begin{aligned}\dot{V}_k \leqslant & 2\int_{u^*\left(t_k^x\right)}^{u^*(t)} \delta \tanh^{-\mathrm{T}}\left(\frac{u^*}{\delta}\right) R_1 d\tau - U\left(u^*\left(t_k^x\right)\right) \\ & -2\int_{u^*\left(t_k^x\right)}^{u^*(t)} \delta \tanh^{-\mathrm{T}}\left(\frac{\tau}{\delta}\right) R_1 d\tau + 2\left(v^*\right)^{\mathrm{T}} R_2^m v^* \\ & +\left(v^*\left(t_k^x\right)-v^*(t)\right)^{\mathrm{T}} R_2^m\left(v^*\left(t_k^x\right)-v^*(t)\right)-Q(x) \\ & -k_\eta(\gamma(t))+\left(1-{k_D}^2\right)\lambda_{\min}(Q)\|x\|^2 \\ & -\left(s\|R_1\|\lambda_u^2+\|R_2^m\|\lambda_v^2\right)\left\|e_x(t)\right\|^2 \end{aligned} \tag{20}$$

Further collated to get:

$$\begin{aligned}\dot{V}_k \leqslant & -U\left(u^*\left(t_k^x\right)\right)-\lambda_{\min}(Q)\|x\|^2+\frac{1}{2}\lambda_{Jx}^2\|R_2^m\|\bar{h}^2 \\ & -k_\eta(\gamma(t))+\left(1-{k_D}^2\right)\lambda_{\min}(Q)\|x\|^2 \\ & +\left(s\|R_1\|\lambda_u^2+\|R_2^m\|\lambda_v^2\right)\left\|e_x(t)\right\|^2 \end{aligned} \tag{21}$$

Therefore, when the triggering condition is satisfied and the state norm is satisfied $\|x\|>\sqrt{\frac{\lambda_{Jx}^2\left\|R_2^m\right\|\bar{h}^2-U\left(u^*\left(t_k^x\right)\right)}{\left(1-{k_D}^2\right)\lambda_{\min}(Q)}}$, $\dot{V}_k \leqslant -k_\eta(\gamma(t))-\lambda_{\min}(Q)\|x\|^2-\gamma \leqslant 0$.

Define the sampled situation state and the error between the situation state and its sampled value as:

$$\hat{\rho}_k(t)=\begin{cases}\rho\left(t_k^s\right), & \forall t \in\left[t_k^s, t_{k+1}^s\right) \\ \rho(t), & t=t_{k+1}^s\end{cases}$$

$$e_s=\hat{\rho}_k(t)-\rho(t)$$

The following situation-triggered condition is designed:

$$t_{k+1}^s=\inf\left\{t_{k+i}^x \mid e_s(t)>\bar{\sigma}_s \text{ or } e_s(t)<\underline{\sigma}_s\right\} \tag{22}$$

The interception-strategy switching mechanism designed above evaluates the target characteristic only at the triggering instants of the situation triggered mechanism.

Moreover, the triggering judgment instants of the situation- triggered mechanism are selected from the state-triggered time sequence $\{t_k^x\}_{k=0}^{\infty}$, $\{t_k^s\} \subseteq \{t_k^x\} \subseteq \mathbb{R}^+$[29].The t_{k+i}^x denotes the $k+i$ th triggering instant of the state-triggered mechanism in (16), which is also the $k+1$ th triggering instant of the situation triggered mechanism in

(22). Let $t_0^s = t_0^x = t_0$ which means that the initial instant is also the first triggering instant.

The triggering sampling structure of the two mechanisms is shown in Fig. 3.

Notably, since the situation-triggered mechanism is judged on the triggering-time sequence of the state-triggered mechanism, while the state-triggered mechanism verifies the triggering condition only at periodic sampling instants, the inter-event time of both triggering mechanisms is greater than or equal to the periodic sampling interval. Therefore, Zeno behavior is naturally excluded [29].

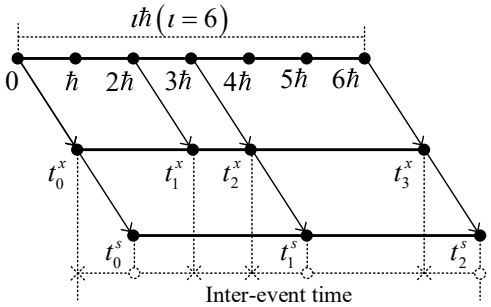


Fig. 3 Event-triggered sampling structure.

4 Event-triggered game learning and target intention estimation based on RL

4.1 Critic NN approximation

The critic NN is designed to approximate the cost function and the terminal cost function:

$$J_m^*(x,t) = (W_1)^{\mathrm{T}}\theta_1(x,\tau) + \varepsilon_1 \tag{23}$$

$$\begin{aligned} J_m^*(x(t_f),t_f) &= \Phi^*(x(t_f),t_f) \\ &= (W_1)^{\mathrm{T}}\theta_1(x(t_f),0) + \varepsilon_{1f} \end{aligned} \tag{24}$$

From (23), the partial derivatives of the cost function with respect to x and t are:

$$\nabla_x J_m(x,t) = (\nabla_x \theta_1(x,\tau))^{\mathrm{T}} W_1 + \nabla_x \varepsilon_1(x,\tau)$$

$$\nabla_t J_m(x,t) = (\nabla_x \theta_1(x,\tau))^{\mathrm{T}} W_1 + \nabla_t \varepsilon_1(x,t)$$

Substituting into (14) and (15), the event-triggered optimal strategies can be expressed as:

$$u^*\left(t_k^x\right) = -\frac{1}{2}\delta \tanh\left(\eta^*\left(t_k^x\right)\right) \tag{25}$$

$$\begin{aligned} v_v^*\left(t_k^x\right) = \frac{1}{2}(R_2^m)^{-1} h^{\mathrm{T}}\left(t_k^x\right) \Big(& \left(\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right)\right)^{\mathrm{T}} W_1 \\ & + \nabla_t \varepsilon_1\left(x\left(t_k^x\right), t\right)\Big) \end{aligned} \tag{26}$$

The ideal weight estimates are defined as:

$$\hat{J}_m^*(x,t) = \hat{W}_1^{\mathrm{T}} \theta_1(x,\tau) \tag{27}$$

$$\hat{\Phi}^*\left(x\left(t_f\right), t_f\right) = \hat{W}_1^{\mathrm{T}} \theta_1\left(\hat{x}\left(t_f\right), 0\right) \tag{28}$$

According to (23) and (24), the Hamiltonian can be represented as:

$$\begin{aligned} H(x,u,v_v) = & \left(W_1^*\right)^{\mathrm{T}} \nabla_x \theta_1(x,\tau)\left(f(x) + g(x) u^*\left(t_k^x\right)\right. \\ & \left. + h(x) {v_v}^*\left(t_k^x\right)\right) + \left(W_1^*\right)^{\mathrm{T}} \nabla_t \theta_1(x,\tau) + x^{\mathrm{T}} Q x \\ & + U^*\left(u^*\left(t_k^x\right), R_1^m\right) - \left({v_v}^*\left(t_k^x\right)\right)^{\mathrm{T}} R_2^m {v_v}^*\left(t_k^x\right) \\ = & \varepsilon_H \end{aligned} \tag{29}$$

where $\varepsilon_H = -\nabla_t \varepsilon_1(x,t) - \nabla_x {\varepsilon_1}^{\mathrm{T}}(x,t)\left(f(x) + g(x) u^*\left(t_k^x\right)\right.$
$\left. + h(x) {v_v}^*\left(t_k^x\right)\right.$ is the approximation error of the critic NN.

4.2 Actor NN and virtual action NN approximation

The actor NN and virtual actor NN are used to approximate the interceptor strategy and the virtual target strategy, respectively:

$$\hat{u}\left(t_k^x\right) = -\frac{1}{2}\delta \tanh\left(\hat{\eta}\left(t_k^x\right)\right) \tag{30}$$

$$\hat{v}_v\left(t_k^x\right) = \frac{1}{2}(R_2^m)^{-1} h^{\mathrm{T}}\left(t_k^x\right)\left(\nabla_x \theta_1\left(t_k^x\right)\right)^{\mathrm{T}} \hat{W}_3 \tag{31}$$

where $\hat{\eta}\left(t_k^x\right) = \frac{1}{\delta}(R_1^m)^{-1} g^{\mathrm{T}}\left(t_k^x\right)\left(\nabla_x \theta_1\left(t_k^x\right)\right)^{\mathrm{T}} \hat{W}_2$.
Based on (30) and (31), the approximate HJI equation is obtained as:

$$\begin{aligned} \hat{H}(x,\hat{u},\hat{v}_v) = & -\hat{v}_v(t_k^x)^{\mathrm{T}} R_2^m \hat{v}_v\left(t_k^x\right) + \hat{W}_1^{\mathrm{T}} \nabla_x \theta_1(x,\tau) * \\ & * \left(f(x) + g(x)\hat{u}\left(t_k^x\right) + h(x)\hat{v}_v\left(t_k^x\right)\right) \\ & + \hat{W}_1^{\mathrm{T}} \nabla_t \theta_1(x,\tau) + Q(x) + \hat{U}\left(\hat{u}\left(t_k^x\right), R_1\right) \\ = & e_H \end{aligned} \tag{32}$$

Combining ε_H, the estimation error of the HJB equation is given by:

$$e_H = \varepsilon_H + \tilde{W}_1^{\mathrm{T}} \nabla_t \theta_1 (x, \tau) + \tilde{W}_1^{\mathrm{T}} \nabla_x \theta_1 (x, \tau) * \\ * (f(x) + g(x) \hat{u}(t_k^x) + h(x) \hat{v}_v(t_k^x)) \tag{33}$$

where, $\tilde{W}_1 = \hat{W}_1 - W_1$ is the NN weight approximation error.
Based on (24) and (28), the terminal constraint estimation error is

$$\begin{aligned} e_{t_f} &= \hat{\Phi}^* (x(t_f), t_f) - \Phi (x(t_f), t_f) \\ &= \hat{W}_1^{\mathrm{T}} \theta_1 (x(t_f), 0) - {W_1}^{\mathrm{T}} \theta_1 (x(t_f), 0) + \varepsilon_1 (x, t_f) \\ &= \tilde{W}_1^{\mathrm{T}} \theta_1 (x(t_f), 0) - \varepsilon_1 (x, t_f) \end{aligned} \tag{34}$$

Construct the following objective function:

$$V_e = \frac{1}{2} e_H^{\mathrm{T}} e_H + \frac{1}{2} e_{t_f}^{\mathrm{T}} e_{t_f} \tag{35}$$

The objective is to minimize (35) and ensure that all signals remain bounded during learning. Based on the gradient descent method, the adaptive update law of the critic network weight is designed as:

$$\dot{\hat{W}}_1 = -k_1 \frac{\partial V_e}{\partial \hat{W}_1} = -k_1 e_H \frac{\partial e_H}{\partial \hat{W}_1} - k_1 e_{t_f} \frac{\partial e_{t_f}}{\partial \hat{W}_1} \tag{36}$$

Using (32) and (34),(36), the critic weight update law can be rewritten as:

$$\begin{aligned} \dot{\hat{W}}_1 &= -k_1 \bar{k}_H \left(r(x, \hat{u}(t_k^x), \hat{v}_v(t_k^x)) + k_H^{\mathrm{T}} \hat{W}_1 \right) - k_1 \bar{k}_{t_f} e_{t_f} \\ &= -k_1 \bar{k}_{t_f} e_{t_f} - k_1 \bar{k}_H \left(\delta^2 R_1 \ln \left(1 - \tanh^2 (\hat{\eta}(t_k^x)) \right) \right. \\ &\quad + Q(x) - \frac{1}{4} \hat{W}_3^{\mathrm{T}} \nabla_x \theta_3 (x(t_k^x), \tau) (R_2^m)^{-1} h^{\mathrm{T}}(x) * \\ &\quad \left. * (\nabla_x \theta_3 (x(t_k^x), \tau))^{\mathrm{T}} \hat{W}_3 \right) \\ &\quad + \hat{W}_2^{\mathrm{T}} \nabla_x \theta_2 (x(t_k^x), \tau) g(x) \tanh (\hat{\eta}(t_k^x)) + k_H^{\mathrm{T}} \hat{W}_1 \end{aligned} \tag{37}$$

where $\bar{k}_H = \frac{k_H}{\left(1 - k_H^{\mathrm{T}} k_H\right)^2}, \bar{k}_{t_f} = \frac{k_{t_f}}{\left(1 - k_{t_f}^{\mathrm{T}} k_{t_f}\right)^2}$, $\frac{1}{\left(1 - k_H^{\mathrm{T}} k_H\right)^2}$ are normalized regressor vectors. The normalization terms are used to avoid divergence of the weight update law caused by excessively large k_H and k_{t_f}. The parameter k_1 is the critic network learning rate. $k_{t_f} = -\theta_1 (x(t_f), 0)$, $k_H = \nabla_x \theta_1 (x, \tau) (f(x) + g(x) \hat{u}(t) + h(x) \hat{v}_v(t)) + \nabla_t \theta_1^{\mathrm{T}} (x, \tau)$.

The actor and virtual actor networks are used to approximate the interceptor strategy and the virtual target strategy, respectively. Their weights are updated by:

$$\dot{\hat{W}}_2 = -k_2 \frac{\partial V_e}{\partial \hat{W}_2} = -k_2 e_H \frac{\partial e_H}{\partial \hat{W}_2} - k_2 e_{t_f} \frac{\partial e_{t_f}}{\partial \hat{W}_2} \tag{38}$$

$$\dot{\hat{W}}_3 = -k_3 \frac{\partial V_e}{\partial \hat{W}_3} = -k_3 e_H \frac{\partial e_H}{\partial \hat{W}_3} - k_3 e_{t_f} \frac{\partial e_{t_f}}{\partial \hat{W}_3} \tag{39}$$

Furthermore, a Lyapunov gradient term is introduced to remove the requirement of an initial admissible policy in traditional gradient descent methods [16] . Combining (7), (30), (31), (32) and (34), the weight update laws are obtained as:

$$\begin{aligned}\dot{\hat{W}}_2 = &-\frac{1}{2} k_2 \Xi\left(x, \hat{u}\left(t_k^x\right), \hat{v}_v\left(t_k^x\right)\right) \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) * \\ &* g(x) R_1^{-1}\left(1-\tanh^2\left(\hat{\eta}\left(t_k^x\right)\right)\right) g^{\mathrm{T}}(x) \nabla_x \bar{J} \\ &+ k_2 \hat{W}_2 - k_2 \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) g(x) \delta \tanh\left(\hat{\eta}\left(t_k^x\right)\right) \bar{k}_H^{\mathrm{T}} \hat{W}_1\end{aligned} \tag{40}$$

$$\begin{aligned}\dot{\hat{W}}_3 = &-\frac{1}{2} k_3 \Xi\left(x, \hat{u}\left(t_k^x\right), \hat{v}_v\left(t_k^x\right)\right) \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) * \\ &* h(x)\left(R_2^m\right)^{-1} h^{\mathrm{T}}(x) \nabla_x \bar{J} + k_3 \hat{W}_3 \\ &- \frac{1}{4} k_3 \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) h(x)\left(R_2^m\right)^{-1} h^{\mathrm{T}}(x) * \\ &* \left(\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right)\right)^{\mathrm{T}} \hat{W}_3 \bar{k}_H^{\mathrm{T}} \hat{W}_1\end{aligned} \tag{41}$$

The sign function $\Xi\left(x, \hat{u}, \hat{v}_v\right)$ is defined as [16]:

$$\Xi\left(x, \hat{u}, \hat{v}_v\right) = \begin{cases} 0, & \left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x) + g(x) \hat{u}\left(t_k^x\right) + h(x) \hat{v}_v\left(t_k^x\right)\right) < 0 \\ 1, & otherwise \end{cases}$$

where, $\bar{J}(x)$ is a continuously differentiable Lyapunov function for system (2). There exists a positive definite matrix Θ such that

$$\begin{aligned}\dot{\bar{J}}(x) &= \left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x) + g(x) u^*\left(t_k^x\right) + h(x) {v_v}^*\left(t_k^x\right)\right) \\ &< -\left(\nabla_x \bar{J}\right)^{\mathrm{T}} \Theta \nabla_x \bar{J}\end{aligned} \tag{42}$$

Remark 4: When the interception system degenerates into a unilateral optimal control problem, only the actor–critic network is used to approximate the optimal value function and the optimal strategy, in order to reduce network complexity and computational burden. In this case, the virtual actor network is no longer used to approximate the target strategy.

4.3 Stability analysis

Theorem 2: Consider the nonlinear game system (2), the event-triggered strategy (16), and the NN weight update laws(37), (40) and (41). Under Assumptions 1-3, the proposed triggering strategy guarantees that the states of the event-triggered closed-loop system and the weight approximation errors are UUB [17].

Assumption 3[11]: For the NNs in (23),(30) and (31), the ideal weights W_i approximation errors ε_i are norm bounded. That is, there exist positive constants $\bar{\varepsilon}_i$, $\bar{W}_i$ such that $\|W_i\| \leqslant \bar{W}_i$, $\|\varepsilon_i\| \leqslant \bar{\varepsilon}_i$, $i = 1, 2, 3$. In addition, for $\nabla\varepsilon_i$ and $\nabla\theta_i$, there exist positive constants such that $\bar{\varepsilon}_{\nabla i}$ and $\bar{\psi}_{\nabla i}$ are bounded for $\|\nabla\varepsilon_i\| \leqslant \bar{\varepsilon}_{\nabla i}$, $\|\nabla\theta_1\| \leqslant \bar{\theta}_{\nabla 1}$, $i = 1, 2, 3$.

Proof: Choose the following Lyapunov function:

$$\begin{aligned} V = V_1 + V_2 = {} & J_m^*(x,t) + J_m^*\left(x\left(t_k^x\right),t\right) + \bar{J}(x) \\ & + \bar{J}\left(x\left(t_k^x\right)\right) + \gamma + \frac{1}{2k_1}\tilde{W}_1^T\tilde{W}_1 + \frac{1}{2k_2}\tilde{W}_2^T\tilde{W}_2 + \frac{1}{2k_3}\tilde{W}_3^T\tilde{W}_3 \end{aligned} \tag{43}$$

where $J_m^*(x,t)$ is the optimal cost function and $J_m^*\left(x\left(t_k^x\right),t\right)$ is its event-triggered representation.

According to whether an event is triggered, the proof considers the following two cases.

Case1: The event is not triggered, $t \in \left[t_k^x, t_{k+1}^x\right)$.

In this case, $\dot{J}_m^*\left(x\left(t_k^x\right),t\right) = 0$, $\dot{\bar{J}}\left(x\left(t_k^x\right)\right) = 0$. By taking the time derivative of (43):

$$\begin{aligned} \dot{V} = \dot{V}_1 + \dot{V}_2 = {} & \dot{\bar{J}}(x) + \frac{1}{k_1}\tilde{W}_1^T\dot{\tilde{W}}_1 + \frac{1}{k_2}\tilde{W}_2^T\dot{\tilde{W}}_2 \\ & + \frac{1}{k_3}\tilde{W}_3^T\dot{\tilde{W}}_3 + \dot{J}^*(x,t) + \dot{\gamma} \end{aligned} \tag{44}$$

where,

$$\dot{V}_1 = \dot{\bar{J}}(x) + \frac{1}{k_1}\tilde{W}_1^T\dot{\tilde{W}}_1 + \frac{1}{k_2}\tilde{W}_2^T\dot{\tilde{W}}_2 + \frac{1}{k_3}\tilde{W}_3^T\dot{\tilde{W}}_3 \tag{45}$$

$$\begin{aligned} \frac{1}{k_1}\tilde{W}_1^T\dot{\tilde{W}}_1 = {} & -\tilde{W}_1^T\bar{k}_H\left(\delta^2 R_1 \ln\left(1-\tanh^2\left(\hat{\eta}\left(t_k^x\right)\right)\right)\right. \\ & + \hat{W}_2^{\mathrm{T}}\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)g(x)\tanh\left(\hat{\eta}\left(t_k^x\right)\right) + k_H^{\mathrm{T}}\hat{W}_1 \\ & + Q(x) - \frac{1}{4}\hat{W}_3^{\mathrm{T}}\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\left(R_2^m\right)^{-1}h^{\mathrm{T}}(x) * \\ & \left.*\left(\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\right)^{\mathrm{T}}\hat{W}_3\right) - \tilde{W}_1^T\bar{k}_{t_f}e_{t_f} \end{aligned}$$

After simplification:

$$\begin{aligned}\frac{1}{k_1}\tilde{W}_1^T\dot{\tilde{W}}_1 = &-\tilde{W}_1^T\bar{k}_H\left(R_1\,\mathrm{sgn}\left(\hat{\eta}\left(t_k^x\right)\right)R_1^{-1}-\tanh^{\mathrm{T}}\left(\hat{\eta}\left(t_k^x\right)\right)\right)*\\ &*\delta g^{\mathrm{T}}\left(x\right)\left(\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\right)^{\mathrm{T}}\hat{W}_2\\ &+\tilde{W}_1^T\bar{k}_H R_1\,\mathrm{sgn}\left(\hat{\eta}\left(t_k^x\right)\right)R_1^{-1}g^{\mathrm{T}}\left(x\right)*\\ &*\left(\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\right)^{\mathrm{T}}W_1+\tilde{W}_1^Tk_{k_H}k_{k_H}^{\mathrm{T}}\tilde{W}_1\\ &+\tilde{W}_1^T\bar{k}_H\left(\varepsilon_H+\delta^2R_1\left(2\,\mathrm{sgn}\left(\eta^*\left(t_k^x\right)\right)\eta^*\left(t_k^x\right)\right.\right.\\ &\left.\left.-\varepsilon_\eta-\hat{\varepsilon}_\eta\right)+W_1^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)f\left(x\right)\right)\\ &-\frac{1}{4}\tilde{W}_1^T\bar{k}_H\hat{W}_3^{\mathrm{T}}\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\left(R_2^m\right)^{-1}h^{\mathrm{T}}\left(x\right)*\\ &*\left(\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\right)^{\mathrm{T}}\tilde{W}_3-\tilde{W}_1^T\bar{k}_{t_f}e_{t_f}\end{aligned}\tag{46}$$

where, $\ln\left(1-\tanh^2\left(\hat{\eta}\left(t_k^x\right)\right)\right)=\ell-2\,\mathrm{sgn}\left(\eta\right)\eta+\varepsilon_\eta$, $\ell=diag\left\{\ln 4,\ln 4\right\}$, ε_η is approximate error,$\mathrm{sgn}\left(\eta\right)=diag\left\{\mathrm{sgn}\left(\eta\right),\mathrm{sgn}\left(\eta\right)\right\}$,$k_{k_H}=\bar{k}_H=\frac{k_H}{1-k_H^{\mathrm{T}}k_H}$[16].

$$\begin{aligned}\frac{1}{k_2}\tilde{W}_2^T\dot{\tilde{W}}_2\leqslant&\,\tilde{W}_2^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)g\left(x\right)\delta\tanh\left(\hat{\eta}\left(t_k^x\right)\right)k_H^{\mathrm{T}}W_1\\ &+\tilde{W}_2^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)g\left(x\right)\delta\tanh\left(\hat{\eta}\left(t_k^x\right)\right)k_H^{\mathrm{T}}\tilde{W}_1\\ &-\frac{1}{2}\tilde{W}_2^T\Xi\left(x,\hat{u}\left(t_k^x\right),\hat{v}_v\left(t_k^x\right)\right)\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)*\\ &*g\left(x\right)R_1^{-1}\left(1-\tanh^2\left(\hat{\eta}\left(t_k^x\right)\right)\right)g^{\mathrm{T}}\left(x\right)\nabla_x\bar{J}+\tilde{W}_2^T\tilde{W}_2\end{aligned}\tag{47}$$

$$\begin{aligned}\frac{1}{k_3}\tilde{W}_3^T\dot{\tilde{W}}_3\leqslant&\,\tilde{W}_3^T\tilde{W}_3+\frac{1}{4}\tilde{W}_3^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)h\left(x\right)\left(R_2^m\right)^{-1}*\\ &*h^{\mathrm{T}}\left(x\right)\left(\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\right)^{\mathrm{T}}\hat{W}_3k_H^{\mathrm{T}}W_1\\ &+\frac{1}{4}\tilde{W}_3^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)h\left(x\right)\left(R_2^m\right)^{-1}h^{\mathrm{T}}\left(x\right)*\\ &*\left(\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)\right)^{\mathrm{T}}\hat{W}_3k_H^{\mathrm{T}}\tilde{W}_1\\ &-\frac{1}{2}k_3\Xi\left(x,\hat{u}\left(t_k^x\right),\hat{v}_v\left(t_k^x\right)\right)\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)*\\ &*h\left(x\right)\left(R_2^m\right)^{-1}h^{\mathrm{T}}\left(x\right)\nabla_x\bar{J}+k_3\hat{W}_3\end{aligned}\tag{48}$$

Based on Assumptions 1-3, substituting (46)-(48) into (45) :

$$
\begin{aligned}
\dot{V}_1 \leqslant & -\lambda_{\min}(\Theta)^{\mathrm{T}}\left\|\nabla_x \bar{J}\right\|^2 - K_1^{\mathrm{T}} K_2 K_1 + K_3 \left\|K_1\right\| \\
& - \frac{1}{2}\tilde{W}_2^T \Xi\left(x, \hat{u}\left(t_k^x\right), \hat{v}_v\left(t_k^x\right)\right) \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) * \\
& * g(x) R_1^{-1}\left(1-\tanh^2\left(\hat{\eta}\left(t_k^x\right)\right)\right) g^{\mathrm{T}}(x) \nabla_x \bar{J} \\
& - \frac{1}{2} k_3 \Xi\left(x, \hat{u}\left(t_k^x\right), \hat{v}_v\left(t_k^x\right)\right) \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) * \\
& * h(x)\left(R_2^m\right)^{-1} h^{\mathrm{T}}(x) \nabla_x \bar{J} + k_3 \hat{W}_3
\end{aligned} \tag{49}
$$

where

$$
K_1 = \left[k_{k_H}^{\mathrm{T}} \tilde{W}_1, \tilde{W}_2^{\mathrm{T}}, \tilde{W}_3^3\right]^{\mathrm{T}},
$$

$$
K_2 = \begin{bmatrix} K_{21} & K_{22} \\ K_{23} & K_{24} \end{bmatrix},
$$

$$
K_{21} = \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{I} \end{bmatrix}, \quad K_{22} = \begin{bmatrix} 0 & 0 \\ k_{21} & 0 \end{bmatrix}, \quad K_{23} = \begin{bmatrix} 0 & k_{21} \\ 0 & k_{22} \end{bmatrix}, \quad K_{24} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & k_{44} \end{bmatrix}.
$$

$$
k_{21} = \left(\frac{\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) g(x) \delta R_1^{-1} \operatorname{sgn}\left(\hat{\eta}\left(t_k^x\right)\right) R_1^{\mathrm{T}}}{2\left(1+k_H^{\mathrm{T}} k_H\right)}\right)^{\mathrm{T}},
$$

$$
k_{44} = \frac{1}{4} \bar{k}_H^{\mathrm{T}} W_1 \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) h(x)\left(R_2^m\right)^{-1}\left(\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right)\right)^{\mathrm{T}} + \frac{1}{2},
$$

$$
K_3 = \begin{bmatrix} K_{31} \\ \left\|\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) g(x) \delta \tanh\left(\hat{\eta}\left(t_k^x\right)\right) \bar{k}_H^{\mathrm{T}} W\right\| \\ K_{33} \end{bmatrix},
$$

$$
\begin{aligned}
K_{31} = & \frac{1}{1+k_H^{\mathrm{T}} k_H}\left(\varepsilon_H + \delta^2 R_1\left(2 \operatorname{sgn}\left(\eta^*\left(t_k^x\right)\right) \eta^*\left(t_k^x\right) - \varepsilon_\eta - \hat{\varepsilon}_\eta\right)\right. \\
& + \tilde{W}_1^T \bar{k}_H R_1 \operatorname{sgn}\left(\hat{\eta}\left(t_k^x\right)\right) R_1^{-1} g^{\mathrm{T}}(x)\left(\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right)\right)^{\mathrm{T}} W_1 \\
& \left. + W_1^T \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) f(x)\right),
\end{aligned}
$$

$$
\begin{aligned}
K_{33} = & \left\| \frac{1}{4} \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) h(x)\left(R_2^m\right)^{-1} h^{\mathrm{T}}(x)\right. \\
& \left(\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right)\right)^{\mathrm{T}} W_1 \bar{k}_H^{\mathrm{T}} W_1 + \frac{1}{2} \nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right) * \\
& \left. * h(x)\left(R_2^m\right)^{-1} h^{\mathrm{T}}(x)\left(\nabla_x \theta_1\left(x\left(t_k^x\right), \tau\right)\right)^{\mathrm{T}} W_1 \right\|.
\end{aligned}
$$

According to the definition of Ξ, two subcases are considered.

(1) When $\left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x)+g(x)\hat{u}\left(t_k^x\right)+h(x)\hat{v}_v\left(t_k^x\right)\right) < 0$, then $\Xi\left(x,\hat{u}\left(t_k^x\right),\hat{v}_v\left(t_k^x\right)\right)=0$, (49) can be rewritten as:

$$\begin{aligned}\dot{V}_1 \leqslant & -\lambda_{\min}(\Theta)\left\|\nabla_x \bar{J}\right\|^2-\lambda_{\min}\left(K_2\right)\left(\|K_1\|-\frac{\lambda_{K_3}}{2\lambda_{\min}\left(K_2\right)}\right)^2 \\ & +\frac{\lambda_{K_3}^2}{4\lambda_{\min}\left(K_2\right)}\end{aligned} \tag{50}$$

where, $\|K_3\| \leqslant \lambda_{K_3}$,when $\left\|\nabla_x \bar{J}\right\| > \sqrt{\frac{\lambda_{K_3}^2}{4\lambda_{\min}(\Theta)\lambda_{\min}(K_2)}}$ or $\|K_1\| > \frac{\lambda_{K_3}}{\lambda_{\min}(K_2)}$, $\dot{V}_1 < 0$.

(2) When $\left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x)+g(x)\hat{u}\left(t_k^x\right)+h(x)\hat{v}_v\left(t_k^x\right)\right) \geqslant 0$, $\Xi\left(x,\hat{u}\left(t_k^x\right),\hat{v}_v\left(t_k^x\right)\right)=1$, (49) can be rewritten as:

$$\begin{aligned}\dot{V}_1 \leqslant & \left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x)+g(x)\hat{u}\left(t_k^x\right)+h(x)\hat{v}_v\left(t_k^x\right)\right) \\ & -K_1^{\mathrm{T}}K_2K_1+K_3\|K_1\|-\frac{1}{2}\tilde{W}_2^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)* \\ \dot{V}_1 \leqslant & \left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x)+g(x)\hat{u}\left(t_k^x\right)+h(x)\hat{v}_v\left(t_k^x\right)\right) \\ & -K_1^{\mathrm{T}}K_2K_1+K_3\|K_1\|-\frac{1}{2}\tilde{W}_2^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)* \\ & *g(x)R_1^{-1}\left(1-\tanh^2\left(\hat{\eta}\left(t_k^x\right)\right)\right)g^{\mathrm{T}}(x)\nabla_x\bar{J} \\ & -\frac{1}{2}k_3\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)h(x)\left(R_2^m\right)^{-1}h^{\mathrm{T}}(x)\nabla_x\bar{J}+k_3\hat{W}_3\end{aligned} \tag{51}$$

Based on (25), (26), (30) and (31):

$$\begin{aligned}\dot{V}_1 \leqslant & \left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x)+g(x)u^*\left(t_k^x\right)+h(x)v^*{}_v\left(t_k^x\right)\right) \\ & -\left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(g(x)\varepsilon_2+h(x)\varepsilon_3\right) \\ & -\frac{1}{2}g(x)\delta\left(\tanh\left(\hat{\eta}\left(t_k^x\right)\right)-\tanh\left(\eta\left(t_k^x\right)\right)\right) \\ & +\frac{1}{2}h(x)\left(R_2^m\right)^{-1}h^{\mathrm{T}}\left(t_k^x\right)\left(\nabla_x\theta_1\left(t_k^x\right)\right)^{\mathrm{T}}\tilde{W}_3 \\ & -K_1^{\mathrm{T}}K_2K_1+K_3\|K_1\|-\frac{1}{2}\tilde{W}_2^T\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)* \\ & *g(x)R_1^{-1}\left(1-\tanh^2\left(\hat{\eta}\left(t_k^x\right)\right)\right)g^{\mathrm{T}}(x)\nabla_x\bar{J} \\ & -\frac{1}{2}k_3\nabla_x\theta_1\left(x\left(t_k^x\right),\tau\right)h(x)\left(R_2^m\right)^{-1}h^{\mathrm{T}}(x)\nabla_x\bar{J}+k_3\hat{W}_3\end{aligned} \tag{52}$$

Based on the $\tanh\left(\hat{\eta}\left(t_k^x\right)\right)-\tanh\left(\eta\left(t_k^x\right)\right)$ Taylor expansion:

$$
\begin{aligned}
\dot{V}_1 \leqslant & \left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(f(x)+g(x) u^*\left(t_k^x\right)+h(x) v^*{}_v\left(t_k^x\right)\right) \\
& -\left(\nabla_x \bar{J}\right)^{\mathrm{T}}\left(g(x) \varepsilon_2+h(x) \varepsilon_3\right)-K_1^{\mathrm{T}} K_2 K_1+K_3\left\|K_1\right\| \\
& +\left(\nabla_x \bar{J}\right)^{\mathrm{T}} g(x)\left(\left(\hat{\eta}\left(t_k^x\right)-\eta\left(t_k^x\right)\right)^2\right) \\
\leqslant & -\lambda_{\min }(\Theta)\left\|\nabla_x \bar{J}\right\|^2+\left(\bar{g} \bar{\varepsilon}_2+\bar{h} \bar{\varepsilon}_3\right)\left\|\nabla_x \bar{J}\right\|^2 \\
& +\lambda_{\min }\left(\frac{1}{\delta}\left(R_1^m\right)^{-1}\right) \bar{g}^2 \bar{\theta}_{\nabla 1}^2\left\|\nabla_x \bar{J}\right\|^2\left\|\tilde{W}_2\right\|-K_1^{\mathrm{T}} K_2 K_1+K_3\left\|K_1\right\|
\end{aligned}
$$

It follows that:

$$
\dot{V}_1 \leqslant -\hbar_1 \lambda_{\min }(\Theta)\left\|\nabla_x \bar{J}\right\|^2+\hbar_2\left\|\nabla_x \bar{J}\right\|^2+\frac{\hbar_3\left\|K_1\right\|^2}{4(1-\hbar) \lambda_{\min }\left(\frac{1}{\delta}\left(R_1^m\right)^{-1}\right)}
$$

where $0<\hbar_1<1$, $\hbar_3=\lambda_{\max }\left(\frac{1}{\delta}\left(R_1^m\right)^{-1}\right) \bar{g}^2 \bar{\theta}_{\nabla 1}^2\left\|\nabla_x \bar{J}\right\|^2$, $\hbar_2=\bar{g} \bar{\varepsilon}_2+\bar{h} \bar{\varepsilon}_3$.
Based on (50) gives:

$$
\begin{aligned}
\dot{V}_1 \leqslant & -\hbar_1 \lambda_{\min }(\Theta)\left\|\nabla_x \bar{J}\right\|^2+\hbar_2\left\|\nabla_x \bar{J}\right\|+\lambda_{K_3}^2\left\|K_1\right\|-\lambda_{\min }\left(K_2\right)\left\|K_1\right\|^2 \\
& +\frac{\lambda_{\min }\left(K_2\right) \hbar_3\left\|K_1\right\|^2}{4(1-\hbar) \lambda_{\min }\left(\frac{1}{\delta}\left(R_1^m\right)^{-1}\right)} \\
= & -\hbar_1 \lambda_{\min }(\Theta)\left(\left\|\nabla_x \bar{J}\right\|-\frac{\hbar_2}{2 \hbar_1 \lambda_{\min }(\Theta)}\right)^2-\hbar_4\left(\left\|K_1\right\|-\frac{\lambda_{K_3}^2}{2 \hbar_4}\right)^2+\hbar_5
\end{aligned}
$$

where, $\hbar_4=\lambda_{\min }\left(K_2\right)-\left(\frac{\hbar_3}{4(1-\hbar_1) \lambda_{\min }(\Theta)}\right)>0$, $\hbar_5=\frac{\lambda_{K_3}^2}{4 \hbar_1}+\frac{\hbar_2^2}{4 \hbar_1 \lambda_{\min }(\Theta)}$.
Thus, when $\left\|\nabla_x \bar{J}\right\|>\frac{\hbar_2}{2 \hbar_1 \lambda_{\min }(\Theta)}+\sqrt{\frac{\hbar_5}{\hbar_1 \lambda_{\min }(\Theta)}}$ or $\left\|K_1\right\|>\frac{\lambda_{K_3}}{2 \hbar_4}+\sqrt{\frac{\hbar_5}{\hbar_4}}$,$\dot{V}_1<0$,the weight approximation error is UUB, and its norm upper bound is defined as $\lambda_{\tilde{W}}$.

$$
\begin{aligned}
\dot{V}_2 & =\dot{J}^*(x, t)+\dot{\gamma} \\
& =\left(\nabla_x J^*\right)^{\mathrm{T}}\left(f(x)+g(x) u^*(t)+h(x) v^*(t)\right) \\
& \quad+\left(\nabla_x J^*\right)^{\mathrm{T}} g(x)\left(\hat{u}\left(t_k^x\right)-u^*(t)\right)+\nabla_t J^* \\
& \quad+\left(\nabla_x J^*\right)^{\mathrm{T}} g(x)\left(\hat{v}\left(t_k^x\right)-v^*(t)\right)-k_\eta(\gamma(t)) \\
& \quad+\left(1-{k_D}^2\right) \lambda_{\min }(Q)\|x\|^2-\left(s\left\|R_1\right\| \lambda_u^2+\left\|R_2\right\| \lambda_v^2\right)\left\|e_x(t)\right\|^2
\end{aligned}
$$

From Theorem 1 gives:

$$\begin{aligned}\dot{V}_2 \leqslant & 2\delta \left\|R_1\right\| \lambda_{\tanh}\left\|\hat{u}\left(t_k^x\right)-u^*(t)\right\|^2 - U\left(\hat{u}\left(t_k^x\right)\right) + \frac{1}{2}\left(\bar{W}_1\bar{\theta}_1+\bar{\varepsilon}_1\right)^2 * \\ & * \left\|R_2\right\| \bar{h}^2 + \left\|R_2\right\| \left\|\hat{v}\left(t_k^x\right)-v^*(t)\right\|^2 - \lambda_{\min}(Q)\|x\|^2 - k_\eta\left(\gamma(t)\right) \\ & - \left(s\left\|R_1\right\|\lambda_u^2\|e_x(t)\|^2 + \left\|R_2\right\|\lambda_v^2\|e_x(t)\|^2\right) + \left(1-{k_D}^2\right)\lambda_{\min}(Q)\|x\|^2\end{aligned}$$

Based on (5), (14) and (30) gives:

$$\begin{aligned}\left\|\hat{u}\left(t_k^x\right)-u^*(t)\right\|^2 &= \left\|\hat{u}\left(t_k^x\right)-u^*\left(t_k^x\right)+u^*\left(t_k^x\right)-u^*(t)\right\|^2 \\ &\leqslant 2\lambda_u^2\|e_x(t)\|^2 + \frac{\lambda_{\tanh}^2}{\delta}\bar{g}^2(R_1^m)^{-1} * \\ &\quad * \left\|(\nabla_x\theta_1)^{\mathrm{T}}W_1+\varepsilon_2-(\nabla_x\theta_1)^{\mathrm{T}}\hat{W}_2\right\|^2 \\ &\leqslant 2\lambda_u^2\|e_x(t)\|^2+\beta_u\end{aligned}$$

where $\beta_u = \frac{\lambda_{\tanh}^2}{\delta}\bar{g}^2(R_1^m)^{-1}\left(\bar{\theta}_{\nabla 1}\lambda_{\tilde{W}_2}+\bar{\varepsilon}_2\right)^2$.
In the same way, gives:

$$\left\|\hat{v}\left(t_k^x\right)-v^*(t)\right\|^2 \leqslant 2\lambda_v^2\|e_x(t)\|^2+\beta_v$$

where $\beta_v = \bar{h}^2(R_2^m)^{-1}\left(\bar{\theta}_{\nabla 1}\lambda_{\tilde{W}_3}+\bar{\varepsilon}_3\right)^2$.
And then get:

$$\begin{aligned}\dot{V}_2 \leqslant & 4\delta\left\|R_1\right\|\lambda_{\tanh}\beta_u + \left\|R_2^m\right\|\beta_v - U\left(\hat{u}\left(t_k^x\right)\right) \\ & + \frac{1}{2}\left(\bar{W}_1\bar{\theta}_1+\bar{\varepsilon}_1\right)^2\left\|R_2^m\right\|\bar{h}^2 - k_\eta\left(\gamma(t)\right) \\ & - \lambda_{\min}(Q)\|x\|^2 + \left(1-{k_D}^2\right)\lambda_{\min}(Q)\|x\|^2 \\ & + \left(s\left\|R_1\right\|\lambda_u^2 + \left\|R_2^m\right\|\lambda_v^2\right)\|e_x(t)\|^2\end{aligned}$$

Therefore, when the event is not triggered, when $\|x\| > \sqrt{\frac{2\delta\|R_1\|\lambda_{\tanh}\beta_u - U\left(\hat{u}\left(t_k^x\right)\right)+\frac{1}{2}\left(\bar{W}_1\bar{\theta}_1+\bar{\varepsilon}_1\right)^2\left\|R_2^m\right\|\bar{h}^2}{\left(1-{k_D}^2\right)\lambda_{\min}(Q)}}$, $\dot{V}_2 \leqslant -k_\eta\left(\gamma(t)\right) - \lambda_{\min}(Q)\|x\|^2 - \gamma \leqslant 0$.

Case2: At the triggering instant $t = t_{k+1}^x$. For $J_m^*(x,t)$, $\bar{J}(x)$ and $\frac{1}{2k_1}\tilde{W}_i^T\tilde{W}_i$, there is no jump at the triggering instant. The difference of V satisfies:

$$\begin{aligned}\Delta V &= \bar{J}\left(x\left(t_{k+1}^x\right)\right) - \bar{J}\left(x\left(t_k^x\right)\right) + J_m^*\left(x\left(t_{k+1}^x\right)\right) - J_m^*\left(x\left(t_k^x\right)\right) \\ &\leqslant -K_{\bar{J}}\left\|x\left(t_k^x\right)\right\| - K_J\left\|x\left(t_k^x\right)\right\|\end{aligned}$$

where, $K_{\bar{J}}$ and K_J represents a class K function. Thus, the UUB property of all system states and weight approximation errors is guaranteed [17]. This completes the proof.

4.4 Target intention estimation

Although the game-based interception mode explicitly considers the adversarial interaction between the interceptor and the target, solving the game strategy still requires prior knowledge of the target's cost function. Under incomplete and asymmetric information, such prior knowledge is generally unavailable to the interceptor. In particular, the weighting matrix R_2^m, which characterizes the target's latent preference between evasion effectiveness and maneuvering expenditure, cannot be directly obtained from onboard measurements. Therefore, even after the activation of the game-based mode, the interceptor still faces uncertainty regarding the target's strategic preference.

To address this issue, a target intention-preference estimation mechanism is developed to estimation the unknown R_2^m from the observed target maneuver. Specifically, the target maneuver induced by the unknown preference is first regarded as a bounded disturbance, and a disturbance observer is constructed to reconstruct the target control input online. Then, by exploiting the structural relationship between the target optimal strategy and its intention weighting matrix, the unknown R_2^m is inferred from the reconstructed target maneuver.

The transformed disturbed system is expressed as:

$$\dot{x}(t) = f(x) + g(x)\,\hat{u}(t) + d(t) \tag{53}$$

The following disturbance observer is designed [30]:

$$\begin{cases} \hat{d}(t) = z(t) + \kappa x(t) \\ \dot{z}(t) = -\kappa\left(z(t) + x(t) + f(x) + g(x)\,\hat{u}(t)\right) \end{cases} \tag{54}$$

where $z(t)$ is the auxiliary variable of the observer, $\hat{d}(t)$ is the disturbance estimate, and $\kappa > 0$ is the observer gain. Which yields $\dot{\hat{d}}(t) = \kappa\left(d(t) - \hat{d}(t)\right)$.

By using the mapping relationship between the maneuvering strategy and the intention matrix in the optimality condition of the target, the target intention parameter can be obtained as:

$$\hat{R}_2^m = \frac{1}{2} h^{-1}\left(t_k^x\right) \hat{d}^{-1} h^{\mathrm{T}}\left(t_k^x\right) \left(\nabla_x \theta_1\left(t_k^x\right)\right)^{\mathrm{T}} \hat{W}_3 \tag{55}$$

Assumption 4[30]: From the perspective of practical mission scenarios, the target maneuvering input is also subject to physical maneuverability constraints, and its variation rate does not exhibit unrealistically abrupt changes. Therefore, for system (53), the disturbance d and its derivative $\dot{d}$ are bounded and satisfy $\|d\| \leqslant \bar{d}, \left\|\dot{d}\right\| \leqslant \bar{\dot{d}}$.

Theorem 3: By using the designed disturbance observer to estimate the system disturbance term, under Assumption 4, the disturbance observation error $e_d(t) = d(t) - \hat{d}(t)$ is UUB.

Proof: To analyze the boundedness of the proposed disturbance observer, choose the following Lyapunov function:

$$V_d = \frac{1}{2} e_d^{\mathrm{T}} e_d \tag{56}$$

Taking the first derivative of (56) gives

$$\dot{V}_d = e_d^{\mathrm{T}} \dot{e}_d = -e_d^{\mathrm{T}} \dot{d}\,(t) - \kappa e_d^{\mathrm{T}} e_d$$

According to assumption 4 and Young's inequality, one obtains:

$$\dot{V}_d \leqslant -\frac{\kappa}{2} e_d^{\mathrm{T}} e_d + \frac{\bar{d}^2}{2\kappa}$$

Therefore, the proposed disturbance observer guarantees bounded disturbance estimation error. Through the above process, the interceptor can estimate the target intention weighting matrix R_2^t under incomplete information, so that the virtual target strategy solved by the interceptor approaches the true target strategy.

Remark 5: Although both the target policy observer and the virtual actor network aim to obtain target policy information, they serve different purposes. The target policy observer reconstructs the target's executed maneuvering policy based on the observed state responses during the engagement and the system dynamics. Therefore, it inevitably suffers from dynamic response delays and estimation errors. It mainly characterizes the actual maneuvering behavior already generated by the target and provides the data basis for target intention identification. Different from the target policy observer, the virtual actor network does not rely on the observed motion response for policy reconstruction. Instead, it learns the target's potential optimal response policy according to the current engagement state and the estimated target intention information, thereby providing the required adversarial strategy approximation for online game-policy computation.

5 Numerical simulation verification

To verify the effectiveness of the proposed RL-based dynamic event-triggered composite game interception method, simulations are conducted in the three-dimensional interception scenario shown in Fig. 1. The terminal time is set as $t_f = 20s$, and the simulation step is 0.001s. The initial position of the interceptor is $(14.5, 7.5, 6.5)$, the initial line-of-sight angles are $q_\varepsilon = -31^\circ$ and $q_\beta = 25^\circ$, and the initial speed is 0.6.,The saturation constraint of the target input is set to 0.3. The saturated input is 0.25. The interceptor weights are selected as $Q = diag\left(10^4, 10^2, 10^4, 10^2\right)$, $R_1 = [0.1, 0.1]$.The initial position of the target is $(10, 15, 10.5)$, and the initial speed is 0.5,The saturation constraint of the target input is set to 0.3. The true target intention weighting matrix is $R_2^t = [0.6, 0.6]$, whereas the initial estimate of the interceptor is $R_2^m = [1.8, 1.5]$. Therefore, the simulation contains incomplete and asymmetric information from the initial instant. The target is considered intercepted when the relative distance between the interceptor and the target is less than 0.03, indicating that the target is within the effective interception range. In addition, to satisfy the persistence of excitation requirement during the network learning process, an exploration signal is added during the first 8s of policy learning, thereby enriching the learning information.

To examine the applicability of the proposed method, two types of target maneuvers are considered: game-based maneuvers and nongame square-wave maneuvers. The square-wave maneuver is defined as:

$$\begin{cases} a_{T\varepsilon} = 0.25\,\mathrm{sgn}\left(\sin\left(1.5t + \pi/4\right)\right) \\ a_{T\beta} = 0.01\,\mathrm{sgn}\left(\sin\left(0.5t + 6\right)\right) \end{cases}$$

The game-based maneuver is obtained online by the target according to its own cost function.

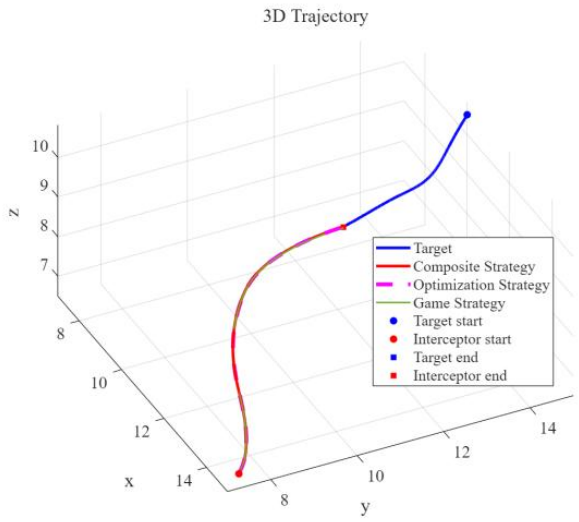


Fig. 4 Comparison of three-dimensional interception trajectories against a game-driven maneuvering target.

The simulation results are shown in Figs. 4-14. Figs. 4-7 show the interception performance under the game-based maneuvering scenario. The nongame optimal strategy fails to intercept the game target, while both the game strategy and the proposed composite strategy achieve successful interception. This is mainly attributed to the fact that the optimal strategy does not account for the adversarial decision-making behavior of the target, when the target performs adversarial maneuvers, the mismatch between the assumed target model and the actual target behavior leads to insufficient robustness and interception failure. The game strategy overcomes this limitation by continuously considering the target as a rational adversary, but this persistent worst-case treatment may result in excessive conservatism and redundant control effort. In contrast, the proposed composite strategy activates the game mode only when the target maneuver has a significant influence on the current interception situation and otherwise adopts the nongame optimal mode. Consequently, the proposed method retains the robustness required for adversarial interception while avoiding unnecessary

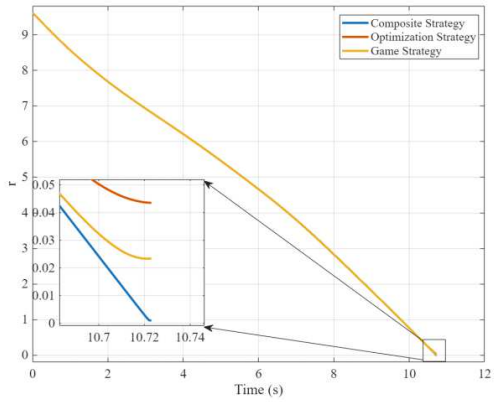


Fig. 5 Relative range histories under different interception strategies.

worst-case compensation, thereby achieving superior interception accuracy. The interceptor input remains within the prescribed saturation bound, and the LOS angular rates converge to a small neighborhood of zero, indicating that the maneuverability constraint and interception requirement are simultaneously satisfied.

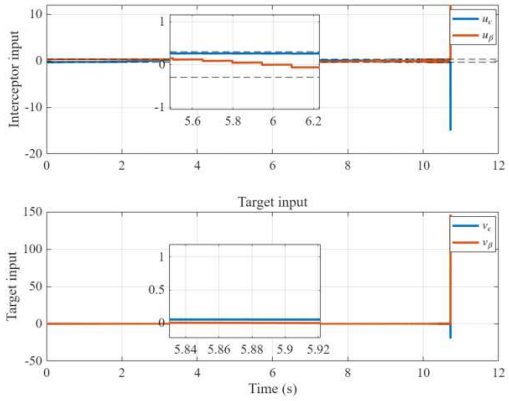


Fig. 6 Maneuvering commands for the interceptor and the target.

Figs. 8 and 9 illustrate the learning and intention estimation processes. The critic, actor, and virtual actor weight norms remain bounded and converge to steady values. Although the initial estimate of $\hat{R}_2$ deviates from the true target intention matrix, the proposed estimator gradually drives the estimated weights toward the true values. The nonmonotonic transient response is mainly caused by the activation of the estimator only in the game-based interception mode.

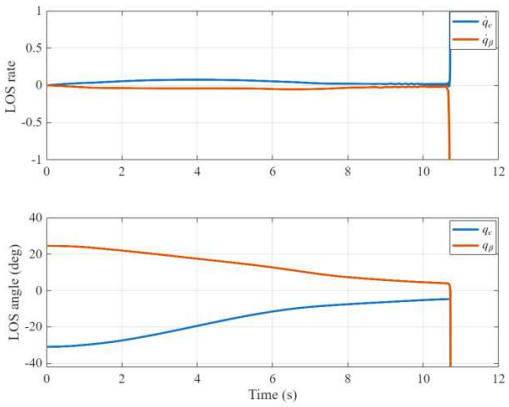


Fig. 7 LOS-angle and LOS-rate under the game-driven maneuvering scenario.

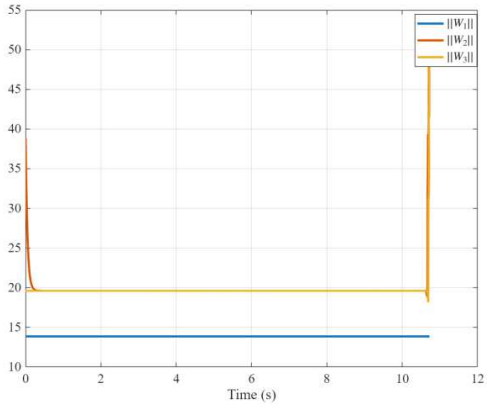


Fig. 8 Norms of the critic, actor, and virtual actor NN weights.

Fig. 10 verify the effectiveness of the triggering mechanism. The interception mode is updated only when the target maneuver significantly affects the engagement situation, and the state information is transmitted only at event-triggered instants. Therefore, unnecessary policy updates and strategy switching are reduced.

Fig. 11 presents the energy consumption under different strategies. Together with Fig. 5, it shows that the proposed method can adaptively switch between game-based interception and nongame optimal interception according to the real-time influence of the target maneuver on the current interception situation, while incorporating the estimated target intention matrix into the game-learning process. Therefore, the proposed method achieves a more reasonable tradeoff among interception performance, control effort, and policy conservatism.

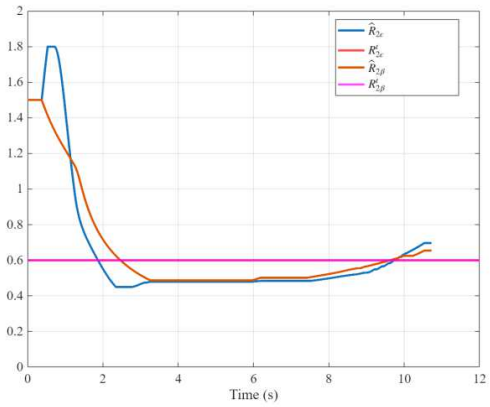


Fig. 9 Online estimation of the target intention weighting matrix.

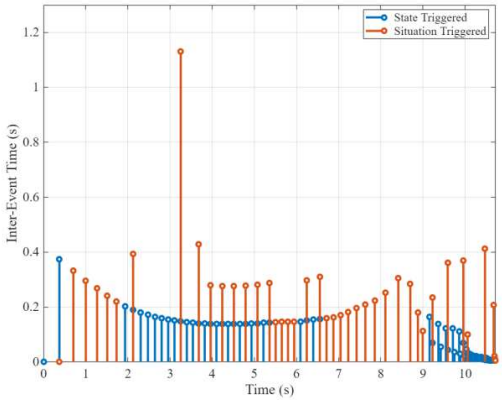


Fig. 10 Event-triggered instants under the game-driven maneuvering scenario.

Figs. 12-14 further examine the proposed method under the square-wave maneuvering scenario. The interceptor successfully approaches the target, and the LOS angular rates converge to a neighborhood of zero, which demonstrates the adaptability of the proposed method to different maneuvering patterns.

It should be noted that in the terminal phase of interception, unavoidable divergence and chattering may appear in the simulation curves. The main reason is that when the interceptor hits the target, the entire interception system is destroyed, which causes abrupt changes in the system parameters [28].

Table 1 compares the interception performance with and without the event-triggered strategy. In the game-based maneuvering scenario, the interception accuracy is improved from 0.70 to 0.44. In the square-wave maneuvering scenario, the interception accuracy is improved from 0.25 to 0.14. These results indicate that, in the

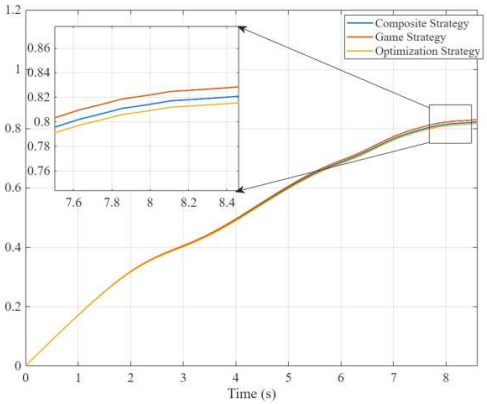


Fig. 11 Energy consumption comparison under different strategies.

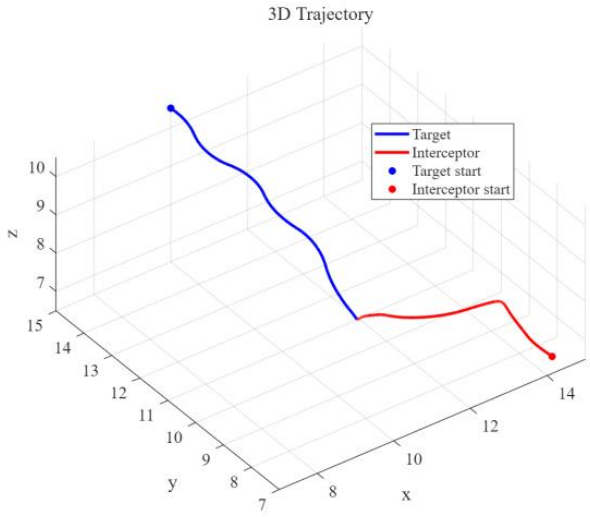


Fig. 12 Three-dimensional interception trajectory against a square-wave maneuvering target.

Table 1 Comparison of interception accuracy of different triggering strategies.

Maneuvering scenario	With the event-triggered	Without the event-triggered
Game maneuvering	0.44	0.7
Square-wave maneuvering	0.14	0.25

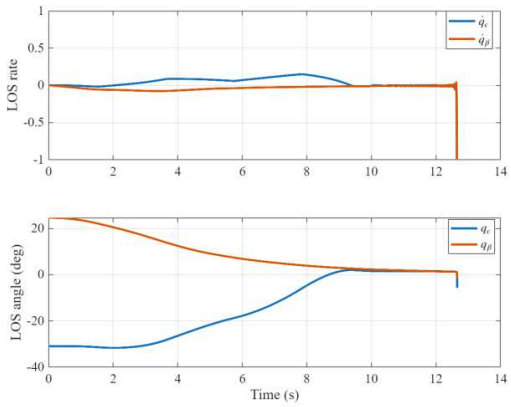


Fig. 13 LOS-angle and LOS-rate under the square-wave maneuvering scenario.

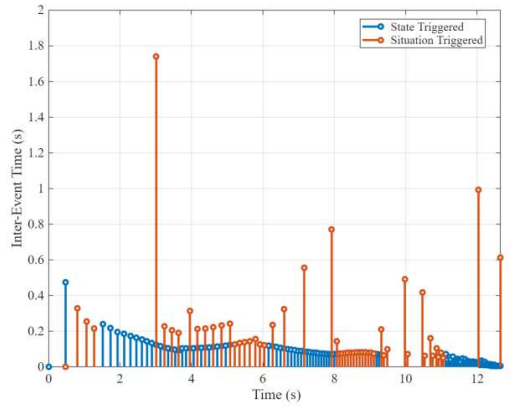


Fig. 14 Event-triggered instants under the square-wave maneuvering scenario.

considered simulation scenarios, the dynamic event-triggered scheme achieves better terminal interception accuracy than the time-triggered scheme. This does not imply that fewer updates necessarily lead to higher interception accuracy. Instead, it shows that the designed triggering rule can suppress redundant updates caused by small-amplitude state variations and transient estimation errors. By avoiding unnecessary policy updates and frequent switching near the triggering threshold, the proposed mechanism makes the input commands smoother and improves the terminal interception performance in these scenarios, thereby verifying the effectiveness of the designed triggering mechanism for online game-learning-based interception tasks.

6 Conclusions

This article investigated the interception problem of an unknown maneuvering target under incomplete and asymmetric information, and proposed a dynamic event-triggered reinforcement-learning-based composite game method. The constructed adaptive switching function balances interception performance and policy conservatism. Two triggering mechanisms were then designed to reduce the computational burden in online RL-based solution and decrease the strategy switching frequency, while excluding Zeno behavior. Sufficient conditions ensuring the UUB property of the closed-loop system were derived. In addition, a target intention estimation mechanism was designed to address the difficulty of solving game strategies caused by the unknown target cost function. Simulation results verified the effectiveness of the proposed method. Future research can be extended to cooperative interception with multiple interceptors, multitarget confrontation, and cooperative game tasks of heterogeneous unmanned systems, so as to further improve the applicability of the proposed method in complex aerospace adversarial environments.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- Funding
  This study is supported in part by the National Natural Science Foundation of China under Grant 62173339/62573215/62333009; in part by the Project of Liaoning Province Science and Technology Department under Grant 2026JH6/101000016; in part by the Liaoning Revitalization Talents Program under Grant XLYC2403070.

- Conflict of interest
  The authors declare no conflicts of interest.
- Data availability
  Data sharing does not apply to this article as no new data were created or analyzed in this study.
- Author Contributions
  All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Tao Ren, Lei Shao and Shu Li. The first draft of the manuscript was written by Tao Ren and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

[1] Wu, X., Wang, W., Tan, S.: Deception in target defense game with unknown attackers speed within a continuous range. IEEE Trans. Aerosp. Electron. Syst., 1–17 (2025) https://doi.org/10.1109/TAES.2025.3620771

[2] Kong, L., He, S., Wu, P., Cui, X., Liu, C.: Distributed state estimation with hybrid game-reinforced consensus for maneuvering target tracking. Aerosp. Sci. Technol. **168**, 111188 (2026) https://doi.org/10.1016/j.ast.2025.111188

[3] Wu, Y., Chen, M., Chadli, M.: Zero-Sum-Game-Based Distributed Fuzzy Adaptive Self-Triggered Control of Swarm UAVs Under Intermittent Communication and DoS Attacks. IEEE Trans. Fuzzy Syst. **32**(9), 5371–5384 (2024) https://doi.org/10.1109/TFUZZ.2024.3423709

[4] Xu, X., Liu, C., Xu, L., Wang, Q., Meng, Y.: Integral reinforcement learning-based optimal pursuit-evasion control for multi-QUAVs: A zero-sum game approach. J. Franklin Inst. **362**(8), 107685 (2025) https://doi.org/10.1016/j.jfranklin.2025.107685

[5] Liu, S., Chen, Z., Gao, D.: Safe reinforcement learning for discrete-time nonlinear zero-sum games with unknown state constraints and asymmetric input constraints. Neurocomputing **651**, 131026 (2025) https://doi.org/10.1016/j.neucom.2025.131026

[6] Leudo, S.J., Sanfelice, R.G.: Two-player zero-sum hybrid games. Automatica **185**, 112688 (2026) https://doi.org/10.1016/j.automatica.2025.112688

[7] Li, Y., Yu, C., Fang, H., Chen, J.: Inverse optimal control for linear quadratic tracking with unknown target states. Automatica **185**, 112819 (2026) https://doi.org/10.1016/j.automatica.2026.112819

[8] Wang, X., Yang, M., Wang, S., Hou, M., Chao, T.: Linear-quadratic and norm-bounded combined differential game guidance scheme with obstacle avoidance for attacking defended aircraft in three-player engagement. Def. Technol. **42**, 136–155 (2024) https://doi.org/10.1016/j.dt.2024.06.018

[9] Ren, T., Li, S., Liu, Y.-J., Wan, F., Liu, L., Sun, C.: Adaptive Event-Triggered Optimal Tracking Control for Wheeled Mobile Robots Considering Force-Velocity Hybrid Constraints. IEEE Trans. Intell. Transp. Syst., 1–12 (2025) https://doi.org/10.1109/TITS.2025.3556530

[10] Xue, S., Luo, B., Liu, D., Gao, Y.: Event-Triggered ADP for Tracking Control of Partially Unknown Constrained Uncertain Systems. IEEE Trans. Cybern. **52**(9), 9001–9012 (2022) https://doi.org/10.1109/TCYB.2021.3054626

[11] Li, S., Liu, Y.-J., Ding, L., Liu, L., Wan, F.: Neural Adaptive Optimal Control of Inequality-Constrained Nonlinear System With Partial Uncertain Time Delay. IEEE Trans. Syst. Man Cybern, Syst., 1–11 (2024) https://doi.org/10.1109/TSMC.2024.3376595

[12] Li, S., Ding, L., Gao, H., Liu, Y.-J., Huang, L., Deng, Z.: ADP-Based Online Tracking Control of Partially Uncertain Time-Delayed Nonlinear System and

Application to Wheeled Mobile Robots. IEEE Trans. Cybern. **50**(7), 3182–3194 (2020) https://doi.org/10.1109/TCYB.2019.2900326

[13] Liu, Y., Wu, H.-N., Zhang, X.-M.: Stability and H∞ performance of human-in-the-loop control systems through hidden semi-Markov human behavior modeling. Appl. Math. Model. **116**, 799–815 (2023) https://doi.org/10.1016/j.apm.2022.12.013

[14] Wu, H.-N.: Online learning human behavior for a class of human-in-the-loop systems via adaptive inverse optimal control. IEEE Trans. Hum.-Mach. Syst. **52**(5), 1004–1014 (2022) https://doi.org/10.1109/THMS.2022.3155369

[15] Wu, M.-E., Wang, M., Wu, H.-N.: Intent inferring based human-machine game control strategy for spacecraft rendezvous. J. Franklin Inst. **361**(9), 106918 (2024) https://doi.org/10.1016/j.jfranklin.2024.106918

[16] Wang, Y., *et al.*: Reinforcement-Learning-Based Counter Deception for Nonlinear Pursuit–Evasion Game With Incomplete and Asymmetric Information. IEEE Trans. Syst. Man Cybern, Syst. **55**(5), 3261–3274 (2025) https://doi.org/10.1109/TSMC.2025.3541105

[17] Xue, S., Zhang, W., Luo, B., Liu, D.: Integral Reinforcement Learning-Based Dynamic Event-Triggered Nonzero-Sum Games of USVs. IEEE Trans. Cybern. **55**(4), 1706–1716 (2025) https://doi.org/10.1109/TCYB.2025.3533139

[18] Li, S., Ren, T., Liu, Y.-J., Liu, L., Wan, F.: Force Feedback Event Triggering-Based Tracking Control for Wheeled Mobile Robots. IEEE Trans. Autom. Sci. Eng., 1–11 (2024) https://doi.org/10.1109/TASE.2024.3461953

[19] Tabuada, P.: Event-triggered real-time scheduling of stabilizing control tasks. IEEE Trans. Automat. Contr. **52**(9), 1680–1685 (2007) https://doi.org/10.1109/TAC.2007.904277

[20] Wang, K., Gu, Q., Huang, B., Wei, Q., Zhou, T.: Adaptive Event-Triggered Near-Optimal Tracking Control for Unknown Continuous-Time Nonlinear Systems. IEEE Access **10**, 9506–9518 (2022) https://doi.org/10.1109/ACCESS.2021.3140076

[21] Liu, Y.-J., Sun, X., Liao, Y., Li, S., Liu, L., Liu, J.J.R.: Adaptive reinforcement learning tracking control of vehicle based on threshold band event-triggered. IEEE Trans. Autom. Sci. Eng. **22**, 13764–13776 (2025)

[22] Sun, X., Liu, L., Liu, Y.-J., Li, S., Wan, F.: Disturbance event triggered-model predictive tracking control for 4WIS–4WID mobile robot. Signal, Image Video Process. (2024) https://doi.org/10.1007/s11760-024-03405-9

[23] Ding, S., Wang, Y., Xie, X.: Periodic Event-Triggered Dynamic Feedback Synchronization Control of Discrete-Time Neural Networks. IEEE Trans. Cybern. **53**(8), 5380–5386 (2023) https://doi.org/10.1109/TCYB.2021.3131475

[24] Ding, S., Wang, Z., Xie, X.: Dynamic Periodic Event-Triggered Synchronization of Complex Networks: The Discrete-Time Scenario. IEEE Trans. Cybern. **53**(10), 6571–6576 (2023) https://doi.org/10.1109/TCYB.2022.3208012

[25] Guo, H., Chen, M., Zhang, P.: Distributed performance constraint control for heterogeneous multiagent systems with dynamic event-triggered mechanism. Chin. J. Aeronaut **38**(3), 103319 (2025) https://doi.org/10.1016/j.cja.2024.11.029

[26] Zhang, H., Wang, J., Luo, Y.: Dynamic event-triggered adaptive critic control for nonlinear systems. IEEE Trans. Neural Netw. Learn. Syst. **33**(4), 1502–1514 (2022)

[27] Xu, H., Zhao, Q., Jagannathan, S.: Finite-Horizon Near-Optimal Output Feedback Neural Network Control of Quantized Nonlinear Discrete-Time Systems With Input Constraint. IEEE Trans. Neural Netw. Learn. Syst. **26**(8), 1776–1788 (2015) https://doi.org/10.1109/TNNLS.2015.2409301

[28] Zhang, B., Zhou, D.: Optimal predictive sliding-mode guidance law for intercepting near-space hypersonic maneuvering target. Chin. J. Aeronaut. **35**(4), 320–331 (2022) https://doi.org/10.1016/j.cja.2021.05.021

[29] Xu, Z., Tang, R., Sun, Y., Li, X., Yang, X.: Secure Synchronization of Coupled Systems via Double Event-Triggering Mechanisms With Actuator Fault. IEEE Trans. Netw. Sci. Eng. **9**(5), 3580–3589 (2022) https://doi.org/10.1109/TNSE.2022.3177718

[30] Song, R., Lewis, F.L.: Robust optimal control for a class of nonlinear systems with unknown disturbances based on disturbance observer and policy iteration. Neurocomputing **390**, 185–195 (2020) https://doi.org/10.1016/j.neucom.2020.01.082