



Letter

Secure Tracking Control via Fixed-Time Convergent Reinforcement Learning for a UAV CPS

Zhenyu Gong  and Feisheng Yang 

Dear Editor,

This letter is concerned with the secure tracking control problem in the unmanned aerial vehicle (UAV) system by fixed-time convergent reinforcement learning (RL). By virtue of the zero-sum game, the false data injection (FDI) attacker and secure controller are viewed as game players. Then, the attack-defense process is recast as a min-max problem. For solving the problem and acquiring the optimal secure control policy, a single-critic RL algorithm with fixed-time convergence is presented. Meanwhile, the associated convergence and stability proofs are given. A simulation is provided to show the effectiveness of the raised method.

The UAV system integrates data sensing, information interaction and decision making by the network, which can be considered as a cyber physical system (CPS) [1]. To complete the specified mission, the UAV is required to receive the control command from the ground control station by network transmission. Nevertheless, the open network environment is vulnerable to cyber attacks. In the transmission process, the FDI attacker can inject misleading data into control commands, resulting in the UAV actuators receiving the compromised control signals. Hence, designing a secure control scheme for the UAV system is of significance [2].

Optimal control can maintain control performance while minimizing the specified cost. RL provides an effective scheme for solving optimal control problems, which can obviate the curse of dimensionality. By the neural network (NN) architecture, [3]–[5] addressed the optimal tracking control problem. Although these works designed effective RL-based optimal tracking controllers, they did not consider the impact of cyber attacks. It inspires us to develop a secure control scheme for the UAV system.

For the attacker and secure controller, the impact of the opponent's policies needs to be assessed while executing their policies. Game theory provides a unified framework to describe this interaction. Researchers devised the defense mechanism using the Stackelberg game [6] and hybrid game [7]. RL can be utilized to solve the secure game. References [8], [9] employed the off-policy and Q-learning algorithms in designing secure controllers. These works have developed a series of schemes by RL and game theory. Besides, they did not consider the convergence rate of RL. Recently, it is worth noting that [10] introduced a finite-time convergent RL algorithm. For improving the flexibility of settling time, the fixed-time convergence technique can be combined with the RL algorithm, which conducts this work.

Motivated by addressing the secure tracking control problem via fixed-time convergent RL for the UAV system, this letter adopts the zero-sum game to describe the attack-defense process. The main contributions of this letter are summarized as follows.

1) In the zero-sum game framework, the secure controller and the FDI attacker are regarded as game players. The secure tracking control problem in the UAV system is recast as a min-max problem.

2) By the single-critic NN structure and experience replay (ER) mechanism, a fixed-time convergent RL algorithm is proposed, which can remove the dependence of the settling time on the initial NN weights compared to [10].

Notations: The $m \times n$ dimensional zero matrix and $m \times m$ dimensional identity matrix are denoted by $0_{m \times n}$ and I_m , respectively. The

Corresponding author: Feisheng Yang.

Citation: Z. Gong and F. Yang, "Secure tracking control via fixed-time convergent reinforcement learning for a UAV CPS," *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 7, pp. 1699–1701, Jul. 2024.

The authors are with the School of Automation, Northwestern Polytechnical University, Xi'an 710129, and also with the Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518063, China (e-mail: gongzhenyu@mail.nwpu.edu.cn; yangfeisheng@nwpu.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JAS.2023.124149

minimum singular value and the minimum eigenvalue are denoted by $\sigma_{\min}(\cdot)$ and $\lambda_{\min}(\cdot)$, respectively. The operator $\lceil \cdot \rceil$ represents $\text{sgn}(\cdot) \cdot |\cdot|$, where $\text{sgn}(\cdot)$ is the sign function and $|\cdot|$ represents the calculation of the absolute value. $\nabla_e(\cdot)$ represents $\partial(\cdot)/\partial e$. The set $\mathcal{B}_\zeta[x_0]$ is the closed-ball with radius ζ and center x_0 . The Cartesian product of two sets \mathcal{S}_1 and \mathcal{S}_2 is given by $\mathcal{S}_1 \times \mathcal{S}_2$.

Problem statement: Consider the following UAV model:

$$\begin{aligned} \dot{p} &= v \\ \dot{v} &= -cg + \bar{u} \end{aligned} \quad (1)$$

where $p = [p_x \ p_y \ p_z]^T$ and $v = [v_x \ v_y \ v_z]^T$ represent the position vector and velocity vector for the center of mass of the UAV in the inertial coordinate system, respectively. The x -axis, y -axis, and z -axis point to the east, north, and center of the earth, respectively. g is the gravity acceleration, $c = [0 \ 0 \ 1]^T$, and \bar{u} is the controller. Define the system state as $x = [p^T \ v^T]^T$. The UAV dynamics becomes

$$\dot{x} = Ax + B\bar{u} + Cg \quad (2)$$

$$\text{where } A = \begin{bmatrix} 0_{3 \times 3} & I_3 \\ 0_{3 \times 3} & 0_{3 \times 3} \end{bmatrix}, B = \begin{bmatrix} 0_{3 \times 3} \\ I_3 \end{bmatrix}, C = [0_{1 \times 5} \quad -1]^T.$$

The aim of tracking control is designing a controller to guide the UAV to reach the expected trajectory and velocity. For the expected position $p_r = [p_{xr} \ p_{yr} \ p_{zr}]^T$ and velocity $\dot{p}_r = v_r$ with $v_r = [v_{xr} \ v_{yr} \ v_{zr}]^T$, define the position and velocity tracking error as $e_p = p - p_r$, $e_v = v - v_r$, respectively. Let $e = [e_p^T \ e_v^T]^T$. The error dynamics is

$$\dot{e} = Ae + B\bar{u} + Cg + f \quad (3)$$

where $f = [0_{1 \times 3} \quad -\dot{p}_r^T]^T$. The tracking task is completed if $e \rightarrow 0$.

Consider the cyber attacker will inject the FDI attack w into the control signal. Then, the actuators will receive false control signals. Hence, one has

$$\bar{u} = \check{u} + w \quad (4)$$

where \check{u} is the secure tracking controller to be designed.

Note that there exist the constant term g and the time-dependent term f in the error system. We can eliminate them by introducing their opposite terms. Hence, design the following secure tracking controller:

$$\check{u} = u + cg + \dot{p}_r \quad (5)$$

where u is the controller to be further designed. Substituting (4) and (5) into (3), one has

$$\dot{e} = Ae + Bu + Bw. \quad (6)$$

Main results: Consider the following performance function:

$$\begin{aligned} J(e(0), u, w) &= \int_0^\infty U(e, u, w) dt \\ &= \int_0^\infty (e^T Q e + u^T R u - \gamma^2 w^T T w) dt \end{aligned} \quad (7)$$

where symmetric weight matrices $Q, R, T > 0$ and the attack attenuation level $\gamma > 0$. In the attack-defense process, the FDI attacker will deteriorate system performance while the secure controller aims to improve it. It indicates that one side's gain leads to the other side's loss, which can be viewed as a zero-sum game. Correspondingly, the secure controller and FDI attacker are players, and the secure controller aims to minimize the performance function while the FDI attacker intends to maximize it. The zero-sum game can be recast as the following min-max problem:

$$V(e(0)) = \min_u \max_w J(e(0), u, w) \quad (8)$$

where $V(e(0))$ is the game value. Moreover, the saddle point (u^*, w^*) is the Nash equilibrium if the following condition holds:

$$J(e(0), u, w^*) \geq J(e(0), u^*, w^*) \geq J(e(0), u^*, w). \quad (9)$$

For the value function $V(e)$, define the Hamiltonian function as

$$H(e, u, w, \nabla_e V(e)) = U(e, u, w) + \nabla_e V^T(e) \dot{e}. \quad (10)$$

By the stationary condition, we can obtain optimal policies below:

$$u^* = -\frac{1}{2} R^{-1} B^T \nabla_e V(e) \quad (11)$$

$$w^* = \frac{1}{2\gamma^2} T^{-1} B^T \nabla_e V(e). \quad (12)$$

Then, the following theorem is given for analyzing the fixed-time stability with the existence of the Nash equilibrium.

Theorem 1: Consider the error system (6) and the given attack attenuation level γ . Set $l_1 > 0$, $l_2 > 0$, $0 < d_1 < 1$, $d_2 > 1$. Suppose that there exists a radially unbounded, continuously differentiable, positive function $V(e)$ satisfying

$$\nabla_e V^T(e) \dot{e} \leq -l_1(V(e))^{d_1} - l_2(V(e))^{d_2}, \quad V(0) = 0. \quad (13)$$

Then, the system (6) with optimal policies (u^*, w^*) is globally fixed-time stable, and the settling time fulfills

$$\tilde{\mathcal{T}} \leq \frac{1}{l_1(1-d_1)} + \frac{1}{l_2(d_2-1)}. \quad (14)$$

Furthermore, (u^*, w^*) is the Nash equilibrium and the game value is $V(e(0))$.

Proof: According to [11], the settling-time function can be derived directly. Moreover, note that $\lim_{t \rightarrow \tilde{\mathcal{T}}} e(t) = \lim_{t \rightarrow \infty} e(t) = 0$, $V(e(\infty)) = V(0) = 0$. By completing the squares for (7), one yields

$$\begin{aligned} J(e(0), u, w) &= V(e(0)) + \int_0^\infty (u - u^*)^T R(u - u^*) dt \\ &\quad - \int_0^\infty \gamma^2 (w - w^*)^T T(w - w^*) dt. \end{aligned} \quad (15)$$

Set policies as (u, w^*) , (u^*, w^*) , and (u^*, w) , respectively. One has

$$J(e(0), u, w^*) \geq J(e(0), u^*, w^*) \geq J(e(0), u^*, w).$$

Then, the Nash equilibrium condition (9) is met. It leads to

$$J(e(0), u^*, w^*) = V(e(0)) \quad (16)$$

which gives the zero-sum game value. \blacksquare

Next, the fixed-time convergent RL algorithm is designed to obtain optimal policies.

The value function can be approximated by the NN. Consider the following critic NN:

$$V(e) = W^T \phi(e) + \varepsilon(e), \quad e \in \mathcal{E} \quad (17)$$

where W , $\phi(\cdot)$, $\varepsilon(\cdot)$, $\mathcal{E} \subset \mathbb{R}^n$ represent expected NN weights, the basis function, the approximation error of critic NN and the error state set, respectively. Moreover, the differentiable basis function should be selected. Since the expected NN weights are unknown, estimation NN weights are utilized to estimate $V(e)$. Then, one has

$$\hat{V}(e) = \hat{W}^T \phi(e), \quad e \in \mathcal{E}. \quad (18)$$

Correspondingly, the optimal policies can be estimated by

$$\hat{u} = -\frac{1}{2} R^{-1} B^T \nabla_e \phi^T(e) \hat{W} \quad (19)$$

$$\hat{w} = \frac{1}{2\gamma^2} T^{-1} B^T \nabla_e \phi^T(e) \hat{W}. \quad (20)$$

Substituting (18)–(20) into the Hamiltonian function (10), it yields the following residual error function:

$$\xi(t) = \hat{W}^T(t) \nabla_e \phi(e(t)) \dot{e}(t) + U(e(t), \hat{u}(t), \hat{w}(t)). \quad (21)$$

According to the ER mechanism, the historical residual error in the time series $0 < t_0, \dots, t_f < t$ is defined as

$$\xi^{[l]}(s) = \hat{W}^T(s) \nabla_e \phi(e(s)) \dot{e}(s) + U(e(s), \hat{u}(s), \hat{w}(s)). \quad (22)$$

Construct the following loss function:

$$\begin{aligned} E(t) &= \frac{1}{q+1} \left(\left| \frac{\xi(t)}{1 + \psi^T(t)\psi(t)} \right|^{q+1} + \sum_{s=t_0}^{t_f} \left| \frac{\xi^{[l]}(s)}{1 + \psi^T(s)\psi(s)} \right|^{q+1} \right) \\ &\quad + \frac{1}{r+1} \left(\left| \frac{\xi(t)}{1 + \psi^T(t)\psi(t)} \right|^{r+1} + \sum_{s=t_0}^{t_f} \left| \frac{\xi^{[l]}(s)}{1 + \psi^T(s)\psi(s)} \right|^{r+1} \right) \end{aligned} \quad (23)$$

where $\psi(\cdot) = \nabla_e \phi(e(\cdot)) \dot{e}(\cdot)$, $0 < q < 1$ and $r > 1$. Let $\tilde{\psi}(\cdot) = \frac{\psi(\cdot)}{1 + \psi^T(\cdot)\psi(\cdot)}$. By the gradient descent principle, we can derive the following NN weight updating law:

$$\begin{aligned} \dot{W}(t) &= -\alpha \left(\tilde{\psi}(t) \left[\frac{\xi(t)}{1 + \psi^T(t)\psi(t)} \right]^q + \sum_{s=t_0}^{t_f} \tilde{\psi}(s) \left[\frac{\xi^{[l]}(s)}{1 + \psi^T(s)\psi(s)} \right]^q \right. \\ &\quad \left. + \tilde{\psi}(t) \left[\frac{\xi(t)}{1 + \psi^T(t)\psi(t)} \right]^r + \sum_{s=t_0}^{t_f} \tilde{\psi}(s) \left[\frac{\xi^{[l]}(s)}{1 + \psi^T(s)\psi(s)} \right]^r \right) \end{aligned} \quad (24)$$

where $\alpha > 0$ is the learning rate of the critic NN. Furthermore, the following assumption is needed to relax the persistence of excitation (PE) condition.

Assumption 1: The matrix $\Psi = [\psi(t_0), \dots, \psi(t_f)]$ is comprised by k historical data and full row rank.

Remark 1: The PE condition is required in the weight training process. A common practice is to add probing noise to control policies. However, selecting proper probing noise signals is a tricky problem and they may deteriorate system stability. The ER mechanism needs to fulfill Assumption 1, which can be achieved by introducing sufficient data. Moreover, the historical data is updated as time, resulting in time complexity $O(n)$ and space complexity $O(1)$, where n represents the scale of the problem.

Define the weight error as $\tilde{W} = \hat{W} - W$. The weight error dynamics is

$$\begin{aligned} \dot{\tilde{W}}(t) &= -\alpha \left(\tilde{\psi}(t) [\tilde{\psi}^T(t) \tilde{W}(t) + \tilde{\varepsilon}(t)]^q + \sum_{s=t_0}^{t_f} \tilde{\psi}(s) [\tilde{\psi}^T(s) \tilde{W}(t) \right. \\ &\quad \left. + \tilde{\varepsilon}(s)]^q + \tilde{\psi}(t) [\tilde{\psi}^T(t) \tilde{W}(t) + \tilde{\varepsilon}(t)]^r \right. \\ &\quad \left. + \sum_{s=t_0}^{t_f} \tilde{\psi}(s) [\tilde{\psi}^T(s) \tilde{W}(t) + \tilde{\varepsilon}(s)]^r \right) \end{aligned} \quad (25)$$

where $\tilde{\varepsilon}(\cdot) = \frac{\varepsilon(\cdot)}{1 + \psi^T(\cdot)\psi(\cdot)}$ with $\varepsilon(\cdot) = W^T \nabla_e \phi(e(\cdot)) \dot{e}(\cdot) + U(e(\cdot), \hat{u}(\cdot), \hat{w}(\cdot))$.

Next, the fixed-time convergence will be analyzed by the following theorem.

Theorem 2: Suppose that Assumption 1 is met. Let $\tilde{\Psi} = [\tilde{\psi}(t_0), \dots, \tilde{\psi}(t_f)]$, $0 < \beta < 1$, $\bar{k} = k^{\frac{1-r}{2}}$, $\iota = (1 - \beta) \sigma_{\min}^{q+1}(\tilde{\Psi})$ and $\bar{\iota} = 2^{1-r} \bar{k} \sigma_{\min}^{r+1}(\tilde{\Psi})$.

Define $\varsigma = \left(\frac{(k+1)(\bar{\varepsilon}_m^q + \bar{\varepsilon}_m^r) \bar{\psi}_m}{\beta \sigma_{\min}^{q+1}(\tilde{\Psi})} \right)^{\frac{1}{q}}$ with $\bar{\varepsilon}_m > 0$, $\bar{\psi}_m > 0$. Then, 1) If $\varepsilon = 0$,

the solution to (25) is globally fixed-time stable with the settling-time function

$$\mathcal{T} \leq \frac{1}{\sigma_{\min}^{q+1}(\tilde{\Psi})(2\alpha)^{\frac{q+1}{2}}(1 - \frac{q+1}{2})} + \frac{1}{\bar{k} \sigma_{\min}^{r+1}(\tilde{\Psi})(2\alpha)^{\frac{r+1}{2}}(\frac{r+1}{2} - 1)}.$$

2) If $\varepsilon \neq 0$, the solution to (25) is globally fixed-time uniformly ultimately bounded (UUB) with the settling-time function

$$\mathcal{T} \leq \frac{2^{\frac{1-q}{2}} \alpha^{-\frac{q+1}{2}} - \alpha^{-1} \varsigma^{1-q}}{\iota(1-q)} + \frac{2^{\frac{1-r}{2}} \alpha^{-\frac{r+1}{2}}}{\bar{\iota}(r-1)}.$$

Proof: Choose the following Lyapunov function:

$$\tilde{V}(\tilde{W}) = \frac{1}{2\alpha} \tilde{W}^T \tilde{W}. \quad (26)$$

Taking the time derivative of (26) along (25), it yields

$$\begin{aligned} \dot{\tilde{V}}(\tilde{W}) &= -\alpha \left(\tilde{W}^T(t) \tilde{\psi}(t) [\tilde{\psi}^T(t) \tilde{W}(t) + \tilde{\varepsilon}(t)]^q \right. \\ &\quad \left. + \sum_{s=t_0}^{t_f} \tilde{W}^T(t) \tilde{\psi}(s) [\tilde{\psi}^T(s) \tilde{W}(t) + \tilde{\varepsilon}(s)]^q \right. \\ &\quad \left. + \tilde{W}^T(t) \tilde{\psi}(t) [\tilde{\psi}^T(t) \tilde{W}(t) + \tilde{\varepsilon}(t)]^r \right. \\ &\quad \left. + \sum_{s=t_0}^{t_f} \tilde{W}^T(t) \tilde{\psi}(s) [\tilde{\psi}^T(s) \tilde{W}(t) + \tilde{\varepsilon}(s)]^r \right). \end{aligned} \quad (27)$$

1) If $\varepsilon = 0$, by p -norm monotonicity property and the Hölder inequality, (27) can be scaled as

$$\begin{aligned} \dot{\tilde{V}}(\tilde{W}) &\leq -\sigma_{\min}^{q+1}(\tilde{\Psi})(2\alpha)^{\frac{q+1}{2}} \tilde{V}^{\frac{q+1}{2}}(\tilde{W}) \\ &\quad - \bar{k} \sigma_{\min}^{r+1}(\tilde{\Psi})(2\alpha)^{\frac{r+1}{2}} \tilde{V}^{\frac{r+1}{2}}(\tilde{W}). \end{aligned}$$

It shows that the solution $\tilde{W} = 0$ to (25) is globally fixed-time stable. By the comparison lemma, the corresponding settling time \mathcal{T} satisfies

$$\mathcal{T} \leq \frac{1}{\sigma_{\min}^{q+1}(\tilde{\Psi})(2\alpha)^{\frac{q+1}{2}}(1 - \frac{q+1}{2})} + \frac{1}{\bar{k} \sigma_{\min}^{r+1}(\tilde{\Psi})(2\alpha)^{\frac{r+1}{2}}(\frac{r+1}{2} - 1)}. \quad (28)$$

2) According to Proposition 1 in [10], one has $\text{sgn}(\tilde{\psi}^T \tilde{W} + \tilde{\varepsilon}) = \text{sgn}(\tilde{\psi}^T \tilde{W})$. By Lemmas 3.3 and 3.4 in [12], one can derive that

$$|\bar{\psi}^T \bar{W}^q - |\bar{e}|^q \leq |\bar{\psi}^T \bar{W} + \bar{e}|^q. \quad (29)$$

$$2^{1-r} |\bar{\psi}^T \bar{W}^r - |\bar{e}|^r \leq |\bar{\psi}^T \bar{W} + \bar{e}|^r. \quad (30)$$

By Proposition 1 in [10], one has $|\bar{e}| \leq \bar{\epsilon}_m$. There exists $\bar{\psi}_m > 0$ that satisfies $\|\bar{\psi}\|_2 \leq \bar{\psi}_m$. Based on (29) and (30), one has

$$\begin{aligned} \dot{\bar{V}}(\bar{W}) &\leq -(1-\beta)\sigma_{\min}^{q+1}(\bar{\Psi})\|\bar{W}\|_2^{q+1} - \beta\sigma_{\min}^{q+1}(\bar{\Psi})\|\bar{W}\|_2^{q+1} \\ &\quad - 2^{1-r}\bar{k}\sigma_{\min}^{r+1}(\bar{\Psi})\|\bar{W}\|_2^{r+1} + (k+1)(\bar{\epsilon}_m^q + \bar{\epsilon}_m^r)\bar{\psi}_m\|\bar{W}\|_2. \end{aligned}$$

Then, the solution to (25) is globally fixed-time UUB with the bound ζ . By the comparison lemma, the settling time fulfills

$$\bar{\mathcal{T}} \leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\zeta^{1-q}}{\iota(1-q)} + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{\iota}(r-1)}. \quad (31)$$

Next, the following theorem is provided to discuss the closed-loop stability.

Theorem 3: Consider the error closed-loop system (6) using policies (19) and (20). Let $\epsilon \neq 0$. Given $\bar{\zeta} = \left(\frac{h+(\bar{\epsilon}_m^q + \bar{\epsilon}_m^r)\bar{\psi}_m}{\beta\sigma_{\min}^{q+1}(\bar{\Psi})}\right)^{\frac{1}{q}}$ with proper $0 < \beta < 1$ and $h > 0$, the augmented state $\Pi = [\bar{W}^T, e^T]^T \in \mathbb{R}^N \times \mathcal{E}$ is fixed-time UUB with the settling time

$$\bar{\mathcal{T}} \leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\bar{\zeta}^{1-q}}{\iota(1-q)} + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{\iota}(r-1)}.$$

Proof: Select the Lyapunov function

$$\bar{V}(\Pi) = V(e) + \bar{V}(\bar{W}). \quad (32)$$

Assume that $\|W\|_2 \leq W_m$, $\|\nabla_e \phi(e)\|_2 \leq \phi_m$ and $\|\nabla_e \varepsilon(e)\|_2 \leq \varepsilon_m$. Combining with Theorem 1, the time derivative of $V(e)$ satisfies

$$\dot{V}(e) \leq -l_1(V(e))^{d_1} - l_2(V(e))^{d_2} + \bar{m}(\phi_m\|\bar{W}\|_2 + \varepsilon_m)$$

where $\bar{m} = (W_m\phi_m + \varepsilon_m)(\frac{1}{2}\|B\|_2^2\lambda_{\min}(R) + \frac{1}{2\gamma^2}\|B\|_2^2\lambda_{\min}(T))$. Choose $h > 0$ such that $\bar{m}(\phi_m\|\bar{W}\|_2 + \varepsilon_m) \leq h\|\bar{W}\|_2$. Let $\bar{\zeta} = \frac{\varepsilon_m}{\frac{h}{m} - \phi_m}$. One yields

$$\dot{V}(e) \leq -l_1(V(e))^{d_1} - l_2(V(e))^{d_2} + h\|\bar{W}\|_2, \quad \bar{W} \in \mathcal{B}_{\bar{\zeta}}[0], \quad e \in \mathcal{E}. \quad (33)$$

By (33) and Theorem 2, choosing the parameter β allows:

$$\begin{aligned} \dot{\bar{V}}(\Pi) &\leq -l_1(V(e))^{d_1} - l_2(V(e))^{d_2} - \iota\|\bar{W}\|_2^{q+1} - \bar{\iota}\|\bar{W}\|_2^{r+1} \\ \bar{W} &\in \mathcal{B}_{\bar{\zeta}}[0], \quad e \in \mathcal{E} \end{aligned} \quad (34)$$

where $\mathcal{B}_{\bar{\zeta}}[0] \subset \mathcal{B}_{\zeta}[0]$ holds for sufficiently small β . Likewise, the settling time fulfills

$$\bar{\mathcal{T}} \leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\bar{\zeta}^{1-q}}{\iota(1-q)} + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{\iota}(r-1)}. \quad (35)$$

Therefore, Π is fixed-time UUB with the bound $\mathcal{B}_{\bar{\zeta}}[0] \times \mathcal{E}$. ■

Simulation example: The reference trajectory is set as $p_r = [10\cos(\frac{\pi}{5}t) \ 12\sin(\frac{\pi}{5}t) \ 1.5 + 0.4t]^T$. The optimal secure controller and optimal attack are solved by the raised scheme. Fig. 1 depicts that the UAV can reach the desired trajectory and velocity under optimal FDI attack by the raised secure controller. Fig. 2(a) shows that the critic NN weights will converge in 11.63 s. As a comparison, the finite-time convergent RL [10] is introduced. In Fig. 2(b), one can observe that the critic NN weights converge in 13.81 s. It indicates that the

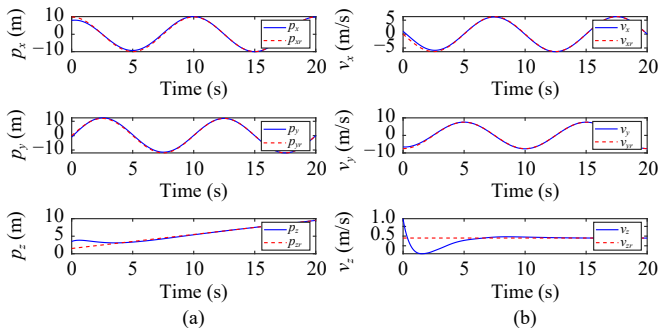


Fig. 1. Evolution of UAV states and reference states. (a) The trajectories of the UAV and reference; (b) The velocities of the UAV and reference.

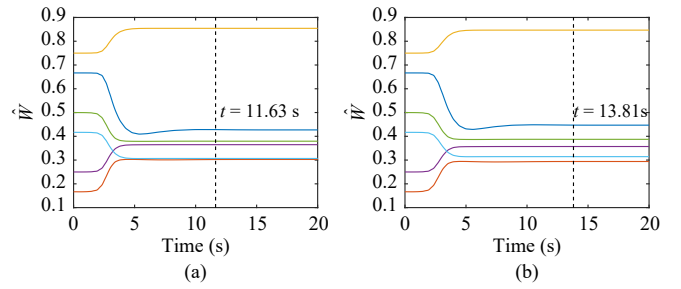


Fig. 2. Evolution of critic NN weights by fixed-time convergent RL and finite-time convergent RL. (a) The fixed-time convergent RL case; (b) The finite-time convergent RL case.

presented algorithm has a faster convergence rate.

Conclusion: This letter studies the secure tracking control problem under the FDI attack for the UAV system. The zero-sum game framework is built to analyze the attack-defense process. Combined with the fixed-time convergence technique, the online single-critic NN is utilized and a novel weight updating law is presented to obtain optimal policies. In the future, secure control schemes for multi-UAV systems will be designed.

Acknowledgments: This work was supported partially by Guangdong Basic and Applied Basic Research Foundation (2023A1515 011220), National Natural Science Foundation of China (62073269), Key Research and Development Program of Shaanxi (2022GY-244), Aeronautical Science Foundation of China (2020Z034053002), and Natural Science Foundation of Chongqing, China (CSTB2022NSCQ-MSX0963).

References

- [1] H. Wang, H. Zhao, J. Zhang, D. Ma, J. Li, and J. Wei, "Survey on unmanned aerial vehicle networks: A cyber physical system perspective," *IEEE Communi. Surveys & Tutorials*, vol. 22, no. 2, pp. 1027–1070, 2020.
- [2] H. Lin, P. Sun, C. Cai, S. Lu, and H. Liu, "Secure LQG control for a quadrotor under false data injection attacks," *IET Control Theory & Applications*, vol. 16, no. 9, pp. 925–934, Jun. 2022.
- [3] L. Dou, X. Su, X. Zhao, Q. Zong, and L. He, "Robust tracking control of quadrotor via on-policy adaptive dynamic programming," *Int. J. Robust and Nonlinear Control*, vol. 31, no. 7, pp. 2509–2525, May 2021.
- [4] M. Ha, D. Wang, and D. Liu, "Discounted iterative adaptive critic designs with novel stability analysis for tracking control," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 7, pp. 1262–1272, Jul. 2022.
- [5] W. Yang, G. Cui, Q. Ma, J. Ma, and S. Guo, "Finite-time adaptive optimal tracking control for a QUAV," *Nonlinear Dynamics*, vol. 111, no. 11, pp. 10063–10076, Mar. 2023.
- [6] Y. Huang and J. Zhao, "Switching defence for switched systems under malicious attacks: A Stackelberg game approach," *Nonlinear Analysis: Hybrid Systems*, vol. 42, p. 101092, Nov. 2021.
- [7] J. Shen, X. Ye, and D. Feng, "A game-theoretic method for resilient control design in industrial multi-agent CPSs with Markovian and coupled dynamics," *Int. J. Control*, vol. 94, no. 11, pp. 3079–3090, Nov. 2021.
- [8] C. Wu, X. Li, W. Pan, J. Liu, and L. Wu, "Zero-sum game based optimal secure control under actuator attacks," *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 3773–3780, Aug. 2021.
- [9] Y. Zhou, K. G. Vamvoudakis, W. M. Haddad, and Z. P. Jiang, "A secure control learning framework for cyber-physical systems under sensor and actuator attacks," *IEEE Trans. Cyber.*, vol. 51, no. 9, pp. 4648–4660, Sept. 2021.
- [10] 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 Trans. Neural Networks and Learning Systems*, vol. 35, no. 3, pp. 3130–3143, Mar. 2024.
- [11] A. Polyakov, "Nonlinear feedback design for fixed-time stabilization of linear control systems," *IEEE Trans. Autom. Control*, vol. 57, no. 8, pp. 2106–2110, Aug. 2012.
- [12] Z. Zuo and L. Tie, "A new class of finite-time nonlinear consensus protocols for multi-agent systems," *Int. J. Control*, vol. 87, no. 2, pp. 363–370, Feb. 2014.