Machine Intelligence Research
www.mi-research.net

DOI: 10.1007/s11633-022-1413-5

Optimal Strategy for Aircraft Pursuit-evasion Games via Self-play Iteration

Xin Wang[1,2] Qinglai Wei[1,2,3] Tao Li[1,2] Jie Zhang[1,2]

[1] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2] The School of Artificial Intelligence, The University of Chinese Academy of Sciences, Beijing 100049, China

[3] Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China

Abstract: In this paper, the pursuit-evasion game with state and control constraints is solved to achieve the Nash equilibrium of both the pursuer and the evader with an iterative self-play technique. Under the condition where the Hamiltonian formed by means of Pontryagin's maximum principle has the unique solution, it can be proven that the iterative control law converges to the Nash equilibrium solution. However, the strong nonlinearity of the ordinary differential equations formulated by Pontryagin's maximum principle makes the control policy difficult to figured out. Moreover the system dynamics employed in this manuscript contains a high dimensional state vector with constraints. In practical applications, such as the control of aircraft, the provided overload is limited. Therefore, in this paper, we consider the optimal strategy of pursuit-evasion games with constant constraint on the control, while some state vectors are restricted by the function of the input. To address the challenges, the optimal control problems are transformed into nonlinear programming problems through the direct collocation method. Finally, two numerical cases of the aircraft pursuit-evasion scenario are given to demonstrate the effectiveness of the presented method to obtain the optimal control of both the pursuer and the evader.

Keywords: Differential games, pursuit-evasion games, nonlinear control, optimal control, Nash equilibrium solution.

Citation: X. Wang, Q. Wei, T. Li, J. Zhang. Optimal strategy for aircraft pursuit-evasion games via self-play iteration. *Machine Intelligence Research*. http://doi.org/10.1007/s11633-022-1413-5

1 Introduction

Differential games are originally proposed by Issacs to address the missile interception problem[1]. Due to the sophisticated theoretical problems and wide application in various fields, differential games have drawn widespread attention from researchers in the field of economics[2, 3], control[4, 5], etc. In particular, the study of pursuit-evasion (PE) games that include aircraft dogfight problems[6, 7], orbital PE problems[8, 9], and multi-pursuer multi-evader problems[10–12] are extensively focused in recent years.

PE games are normally formed as zero-sum differential games consisting of two-players, in which one player is named the pursuer and the other is named the evader. The two players in the game have opposite targets. Generally, in the finite-horizon case, where the terminal time of the game is determinate, the objective of the pursuer is to minimize the distance between the pursuer and the evader, while the evader aims at keeping away from the pursuer as far as possible[13]. The desired optimal control strategy in PE games for the pursuer and the evader corresponds to the Nash equilibrium solution of zero-sum differential games. At the Nash equilibrium points, both players achieve the extremum and no one can improve its own expected performance by changing their strategy while the other player remains unchanged. This solution can be achieved by dynamic programming (DP)[14–17] or Pontryagin's maximum principle (PMP)[18, 19]. For two-player zero-sum nonlinear differential games that consist of nonlinear dynamics and nonlinear objective functions, the dynamic programming method is simplified to a Hamilton-Jacobi-Isaacs (HJI) partial differential equation[5], while the PMP method is transformed into a two-point boundary value problem (TPBVP). Generally, figuring out the Nash equilibrium solution of such a game is equivalent to addressing a bilateral optimization problem, which is much more difficult than the well-studied unilateral optimization problem[20], since it usually requires the solution of high-dimensional TPBVP. In particular, PE games in three-dimensional space are more intricate to deal with than those in one or two dimensions.

Optimal control problems of differential games are

Research Article
Manuscript received on August 7, 2022; accepted on December 28, 2022
Recommended by Associate Editor Wei He
Colored figures are available in the online version at https://link.springer.com/journal/11633
This article has been accepted and is being edited now. Its final version will be assigned to and published in a specific journal issue later.

widely studied by many researchers[21–23]. In [5], through an iterative adaptive dynamic programming method, the optimal control strategy for a class of nonlinear zero-sum games is achieved. Kartal[24] proposes a closed-loop optimal control law for PE games by on-policy reinforcement learning. The continuous-time multiple-agent PE games are solved in [25]. For the finite-horizon differential games problem, an approximate optimal strategy for nonlinear zero-sum differential games is presented in [26]. The optimal control problems for nonlinear nonzero-sum differential games in the environment of no initial admissible policies with control constraints are studied by Mu et al.[27] for discrete-time systems. The constraints are incorporated into this optimization by introducing the nonquadratic value function, while policy iterations are utilized to obtain the optimal control law. Cui et al.[28] design an online learning algorithm for finite-horizon non-zero-sum games with constrained inputs. In [29], a comprehensive overview of the researches on PE differential games is presented.

The main methods of solving open-loop optimal control problems include direct method, indirect method and semi-direct method. In [19], a semi-direct collocation method is designed to handle a differential game entailing a missile evasion scenario. In [30], an indirect multiple shooting method is presented to solve the PE game between a missile and an aircraft. While, direct method is considered to be robuster compared to indirect methods[20]. The key factor for this view is that only an initial state needs to be provided to the nonlinear programming problem (NLP) solver. The difficulty of having to guess the initial values of the nonintuitive Lagrange multipliers are thus avoided when TPBVP solvers is utilized. However, the direct method cannot be easily utilized in a differential game or minmax problem since the NLP solver on which the method relies must have just one objective function to optimize.

Motivated by the references mentioned above, the aim of our paper is to address the optimal control problem of pursuit-evasion differential games with constraints. In practical applications, such as the control of aircraft, the overload that can be provided at the moment is limited and some state vectors are restricted by the function of it. To solve this kind of differential game problem, we divide PE games into two optimal problems and design a method called self-play iteration to achieve the optimal control strategy for both the pursuer and the evader. By regarding the control strategy of the rival as a parameter and updating its own optimal control law, the iterative control strategy is able to converge to the Nash equilibrium solution if the game has a unique Nash equilibrium solution. Owing to the strong nonlinearity of the system dynamics and state and control constraints, we transform the optimal problem into nonlinear programming problem through direct collocation. The simulation results indicate that the aircraft PE game investigated in this manuscript can be solved by the designed method.

The main contribution of this article is mainly reflected in the following aspects. Instead of using semi-direct method in [19, 20], we propose a novel self-play iteration method that can transform the two-sided optimization problem into iteratively solving one of the two one-sided optimal control problems for each updating process. After that direct method can be applied and the original problem is transformed into NLP with constraints, which can be solved by various full-fledged optimization algorithms. Under the condition that the differential game contains a unique Nash equilibrium solution, it can be proven that as the number of iterations increases, the iterative optimal control strategies obtained by the present method converge to the equilibrium solution of the game.

The remaining sections of this paper are organized as follows. Section 2 describes PE games and studies the open-loop Nash equilibrium solution by PMP. In Section 3, the self-play iteration algorithm is presented and its convergence is proven under the condition that the PE game contains a unique Nash equilibrium solution. Then, the implementation is given by transforming the optimization problem into nonlinear programming through the direct collocation method. Section 4 constructs an aircraft PE game by introducing the dynamics of three-degree-of-freedom aircraft and demonstrates the simulation results of the game to validate the effectiveness of the proposed method. Finally, Section 5 draws conclusions.

2 Problem formulation

Throughout this paper, we assume that both the pursuer and the evader can not observe the state of each other. Therefore, we employ the Pontryagin Maximum Principle to solve the open-loop optimal control problems for this PE game. To obtain the optimal strategy, we consider the zero-sum differential game as two optimization problems.

2.1 Nash equilibrium

For generality and concision, in this manuscript, we consider a normal continuous time (CT) nonlinear system for PE game written as

$$\begin{aligned}&\dot{x}(t)=f\left(x,u_1,u_2,t\right), \quad t\in[t_0,t_f]\\ &\text{s.t.}\quad x(t_0)=x_0,\ x(t)\in X(u_1(t),\ u_2(t))\\ &\qquad X(u_1(t),\ u_2(t))\subseteq \mathbf{R}^{n_1}\\ &\qquad u_1(t)\in U_1,\ U_1\subseteq\mathbf{R}^{m_1}\\ &\qquad u_2(t)\in U_2,\ U_2\subseteq\mathbf{R}^{m_2}\end{aligned}\qquad(1)$$

where x is the state vector of the system, $u_1(t)$, $u_2(t)$ are the control strategies implemented by the pursuer and the evader, respectively. n_1, m_1 and m_2 represent the dimension of the corresponding space. Then, the cost function of the game is given as

$$J\left(u_1, u_2\right) \doteq \Phi(x(t_f)) + \int_{t_0}^{t_f} l\left(x(t), u_1(t), u_2(t), t\right) \mathrm{d}t \quad (2)$$

where $\Phi(\cdot)$ is the terminal cost, while $l(\cdot)$ represents the running cost. Note that system (1) and cost function (2) form a standard differential game problem, therefore, they can be applied to model and analyze the aircraft dog-fight, economics competition and H_∞ control problems.

Definition 1 (Open-loop Nash equilibrium). Control functions $u_1^*(t)$ and $u_2^*(t)$ form a open-loop Nash equilibrium for the game (1) and (2) if the following holds:

1) The control function $u_1^*(\cdot)$ is the optimal control strategy for the pursuer that

$$\text{Minimize}: J\left(u_1, u_2^*\right) = \Phi(x(t_f)) + \int_{t_0}^{t_f} l\left(x(t), u_1(t), u_2^*(t), t\right) \mathrm{d}t \quad (3)$$

for the system $x(t_0) = x_0$, $\dot{x}(t) = f\left(x, u_1, u_2^*, t\right)$, $x(t) \in X(u_1(t))$, $t \in [t_0, t_f]$, $u_1(t) \in U_1$.

2) The control function $u_2^*(\cdot)$ is the optimal control strategy for the evader that

$$\text{Maximize}: J\left(u_1^*, u_2\right) = \Phi(x(t_f)) + \int_{t_0}^{t_f} l\left(x(t), u_1^*(t), u_2(t), t\right) \mathrm{d}t \quad (4)$$

where $x(t_0) = x_0$, $\dot{x}(t) = f\left(x, u_1^*, u_2, t\right)$, $x(t) \in X(u_2(t))$, $t \in [t_0, t_f]$, $u_2(t) \in U_2$.

Definition 2 (Admissible control). A control policy $[u_1(t), u_2(t)]$ is called an admissible control for system (1) on $[t_0, t_f]$, if $[u_1(t), u_2(t)]$ is continuous on $[t_0, t_f]$, and $[u_1(t), u_2(t)]$ stabilizes system (1), which simultaneously ensures $J\left(u_1, u_2\right)$ to be finite under $[u_1(t), u_2(t)]$.

Employing the definitions above, we can formulate the aircraft PE game as follows:

$$V = \min_{u_1} \max_{u_2} J\left(u_1, u_2\right) \quad (5)$$

where $V(\cdot)$ is the value of the game. The set of admissible controls can be denoted as $\Omega_{[t_0,t_f]} \triangleq U_{1[t_0,t_f]} \times U_{2[t_0,t_f]}$.

According to (5), the PE game problem is defined as to determining a strategy function $[u_1^*(t), u_2^*(t)] \in \Omega_{[t_0,t_f]}$ that minimizes $J(u_1, u_2)$ with respect to $u_1(t)$, meanwhile maximizes $J(u_1, u_2)$ with respect to $u_2(t)$. Then, $[u_1^*(t), u_2^*(t)]$ can be called optimal control strategy for the PE game. Through the analysis above, we have

$$J(u_1^*, u_2) \le J(u_1^*, u_2^*) \le J(u_1, u_2^*). \quad (6)$$

By definition, the strategy $[u_1^*(t), u_2^*(t)]$ forms a Nash equilibrium if the inequality (6) holds.

To achieve the solution of Nash equilibrium, these two optimization problems need to be solved. It is notable that the optimal control $u_1^*(\cdot)$ of the first optimization can be regarded as a parameter in the second one, and vice versa.

Now we need to figure out a pair of open-loop control schemes (u_1^*,u_2^*) that yield Nash equilibrium by the PMP. For this purpose, we make the following assumption.

Assumption 1. For every $(x,t) \in X(u_1(t), u_2(t)) \times [t_0, t_f]$ and Lagrange multiplier $\lambda \in \mathbf{R}^N$, there exists a unique pair of control $(\bar{u}_1^*, \bar{u}_2^*) \in U_1 \times U_2$ that satisfies[31]

$$\begin{aligned} \bar{u}_1^* &= \arg\min_{\mu \in U_1} \{\lambda^{\mathrm{T}} f(x, \mu, \bar{u}_2^*, t) + l(x, \mu, \bar{u}_2^*, t)\} \\ \bar{u}_2^* &= \arg\max_{\mu \in U_2} \{\lambda^{\mathrm{T}} f(x, \bar{u}_1^*, \mu, t) + l(x, \bar{u}_1^*, \mu, t)\}. \end{aligned} \quad (7)$$

Form the Hamilton function as

$$H(x, u_1, u_2, \lambda, t) = l(x, u_1, u_2, t) + \lambda^{\mathrm{T}} f(x, u_1, u_2, t). \quad (8)$$

Suppose that Assumption 1 holds, let $x^*(\cdot)$ denote the trajectory of the open-loop optimal control, while (u_1^*, u_2^*) is the Nash equilibrium solution. Then, applying the PMP to obtain the appropriate Lagrange multipliers that make the following equation hold:

$$\begin{cases} \dot{x}^*(t) = f(x^*, u_1^*, u_2^*, t) \\ \dot{\lambda}(t) = -\dfrac{\partial H(x^*, u_1^*, u_2^*, \lambda, t)}{\partial x^*(t)} \end{cases} \quad (9)$$

with boundary and extremum conditions

$$\begin{cases} x(t_0) = x_0 \\ \lambda(t_f) = \dfrac{\Phi(x^*(t_f), t_f)}{\partial x^*(t_f)} \\ \dfrac{\partial H(x^*, u_1^*, u_2^*, \lambda, t)}{\partial u_1^*(t)} = \dfrac{\partial H(x^*, u_1^*, u_2^*, \lambda, t)}{\partial u_2^*(t)} = 0. \end{cases} \quad (10)$$

By (9) and the corresponding boundary data, one can compute the Nash equilibrium solution for the PE game.

3 Solution of optimal control for the PE game

Note that (9) consists of two ordinary differential equations (ODEs) together with the strong nonlinearity of the maps u_1^*, u_2^* in (7), which make problem (9) and (10) difficult to figure out, in general.

3.1 Self-play iteration

According to definition 1, we divide the PE differential game (19) into two optimal control problems. As the control strategy of the rival can be regarded as a parameter in the process of computing its own optimal control law, we present a self-play iteration method to achieve the two optimal control policies step by step.

Initially, we need to select an initial control strategy for evader denoted as $u_E^0 \in U_{E[t_0,t_f]}$. Then, the iterative control scheme for pursuer is obtained by

$$u_P^1 = \underset{u_P \in U_P}{\arg\min}\, J\left(u_P, u_E^0\right). \tag{11}$$

We can write the process of updating the iterative control strategy of the PE game as

$$\begin{aligned} u_P^i &= \underset{u_P \in U_P}{\arg\min}\, J\left(u_P, u_E^i\right) \\ u_E^{i+1} &= \underset{u_E \in U_E}{\arg\max}\, J\left(u_P^i, u_E\right) \quad \forall i = 0, 1, \dots. \end{aligned} \tag{12}$$

Algorithm 1. The Self-Paly Iteration for Aircraft PE Game with State and Control Constraints

```
Initialization:
1) Set an initial state $s_0$ for the PE game that satisfies the state constraints;
2) Select a calculation precision $\delta$;
3) Choose an initial control strategy $u_E^0$ for evader s.t. $u_E^0 \in U_{E[t_0,t_f]}$;
4) Given the maximum number $i_{\max}$ of the iteration;
5) Let iteration index $i = 0$.
Iteration:
6) Compute the iterative control strategy by
       $u_P^i = \arg\min_{u_P \in U_P} J\left(u_P, u_E^i\right)$;
7) Introduce the solution of Step 6, update the iteration control scheme for evader by
       $u_E^{i+1} = \arg\max_{u_E \in U_E} J\left(u_P^i, u_E\right)$;
8) If $|J(u_P^{i-1}, u_E^i) - J(u_P^i, u_E^i)| < \delta$ and $|J(u_P^i, u_E^{i+1}) - J(u_P^i, u_E^i)| < \delta$ hold simultaneously, goto Step 10. Else goto Step 9;
9) Let $i = i + 1$, if $i < i_{\max}$, then goto step 6. Else goto Step 11;
10) Return $u_P^i$ and $u_E^{i+1}$. The Nash equilibrium solution is obtained;
11) Return the Nash equilibrium solution is not obtained within $i_{\max}$ iterations.
```

Theorem 1. Suppose Assumption 1 holds, choose an arbitrary initial control strategy that satisfies $u_E^0 \in U_{E[t_0,t_f]}$. Update the control policy of pursuer and evader by (11) and (12). Then, the pair of controls (u_P^i, u_E^{i+1}) converge to the Nash equilibrium solution as $i \to \infty$.

Proof. According to the Assumption 1, there exists a unique Nash equilibrium in the PE game. Meanwhile, when the number of iterations is sufficient, (u_P^i, u_E^{i+1}) converges to the optimal control for both the pursuer and evader. That means they achieve the extremum and no player can improve its own expected performance by changing their strategy while the other player keeps unchanging. That is the definition of Nash equilibrium. □

As the number of iterations increases, the control strategies of the pursuer and the evader asymptotically approximate the Nash equilibrium solution, which means the iterative control strategy yields (9) and condition (10). The conditions that stop the update are given as $|J(u_P^{i-1}, u_E^i) - J(u_P^i, u_E^i)| < \delta$ and $|J(u_P^i, u_E^{i+1}) - J(u_P^i, u_E^i)| < \delta$ satisfied simultaneously, otherwise execute the updates to the maximum number of iterations. The detailed implementation of the self-play iteration is displayed in Algorithm 1.

3.2 Implementation details

In this subsection, to solve the optimal problems in Steps 6 and 7 of Algorithm 1, we employ direct collocation method that can transform the original problem into a nonlinear programming problem.

Consider the optimal control problem of the general CT nonlinear system with state and control constraints as follows:

$$\begin{aligned} J(u) = \Phi(x(t_f)) &+ \int_0^{t_f} l(x(\tau), u(\tau), \tau)\mathrm{d}\tau \\ \text{s.t.} \quad & \min_{u \in U} J(u) \\ & \dot{x} = f(x, u, t) \\ & h(u) \le 0,\ F(x, u) \le 0. \end{aligned} \tag{13}$$

For direct collection, the constraints are formed through constructing the CT dynamics in integral and approximating it using trapezoidal quadrature:

$$\begin{aligned} & \dot{x} = f(x, u, t) \\ & \int_{t_k}^{t_{k+1}} \dot{x}\mathrm{d}t = \int_{t_k}^{t_{k+1}} f(x, u, t)\mathrm{d}t \\ & x_{k+1} - x_k \approx \frac{1}{2}\Delta T(f(x_{k+1}, u_{k+1}, k+1) + \\ & \qquad f(x_k, u_k, k)),\ k = 0, \cdots, \mathcal{T}_f. \end{aligned} \tag{14}$$

After that, the cost function is also approximated into the same form:

$$\begin{aligned} & \Phi(x(t_f)) + \int_0^{t_f} l(x(\tau), u(\tau))\mathrm{d}\tau \approx \\ & \quad \Phi(x_{\mathcal{T}_f}) + \sum_{k=0}^{\mathcal{T}_f - 1} \frac{1}{2}\Delta T(l(x_{k+1}, u_{k+1}, k+1) + l(x_k, u_k, k)). \end{aligned} \tag{15}$$

Then, the nonlinear programming problems are given as

$$
\begin{aligned}
&\min\{\Phi(x_{\mathcal{T}_f})+\sum_{k=0}^{\mathcal{T}_f-1}\frac{1}{2}\Delta T(l(x_{k+1},u_{k+1},k+1)+\\
&\quad l(x_k,u_k,k))\}\\
&\text{s.t.}\quad x_{k+1}-x_k=\frac{1}{2}\Delta T(f(x_{k+1},u_{k+1},k+1)+\\
&\qquad\qquad f(x_k,u_k,k)),\ k=0,\cdots,\mathcal{T}_f\\
&\qquad h\left(u_k\right)\leq 0,\ F\left(x_k,u_k\right)\leq 0.
\end{aligned}
\tag{16}
$$

In this way, problem (16) can be solved by various full-fledged optimization algorithms, such as the interior point method[32], the sequential quadratic programming algorithm[33], the trust region reflective algorithm[34], etc.

4 Simulations

In this section, we apply this scheme to an aircraft PE game to demonstrate its effectiveness.

4.1 Dynamics of aircraft

In this subsection, we introduce a classic point-mass model of aircraft system dynamics for the pursuer and the evader, and present the objective of each other in the game to formulate the optimal control problem.

The three-degree-of-freedom aircraft motion equation is given as follows[35]:

$$
\begin{cases}
\dot{v}=g\left(N_{x_a}-\sin\gamma\right)\\
\dot{\psi}=-gN_{z_a}\sin\phi/(v\cos\gamma)\\
\dot{\gamma}=-g\left(\cos\gamma+N_{z_a}\cos\phi\right)/v\\
\dot{x}=v\cos\gamma\cos\psi\\
\dot{y}=v\cos\gamma\sin\psi\\
\dot{z}=-v\sin\gamma
\end{cases}
\tag{17}
$$

where g is the acceleration of gravity. x, y and z are three-dimensional coordinates in the ground coordinate system with the units of meters which refer to north, east and ground respectively. v is the velocity of aircraft with the unit meter per second. γ is the included angle between the aircraft velocity direction and the horizontal plane called the flight path bank angle (up is taken to be the positive direction). ψ is the included angle between the north and the plane velocity vector in the horizontal projection named flight path azimuth angle (the right side of the projection is positive and the left side is negative). ϕ is the flight path bank angle (rolling right is positive, rolling left is negative). N_{x_a} and N_{z_a} are the axial and normal overload factors. The coordinate system and flight path angles are defined as shown in Fig. 1.

Let $u=[N_{x_a},N_{z_a},\phi]^{\mathrm{T}}$, $s=[v,\psi,\gamma]^{\mathrm{T}}$ represent the control vector and state vector respectively. Considering the limitation of the actual capacity of the normal aircraft, the control variables are constrained with $N_{x_a}\in[-2,3]$, $N_{z_a}\in[-9,9]$ and $\phi\in[-45°,45°]$. Meanwhile the state variables are also limited by $\alpha_1 s+\alpha_2 u\leq 0$, where $\alpha_1=[0,0,0.8]^{\mathrm{T}}$, $\alpha_2=[0,0,1]^{\mathrm{T}}$.

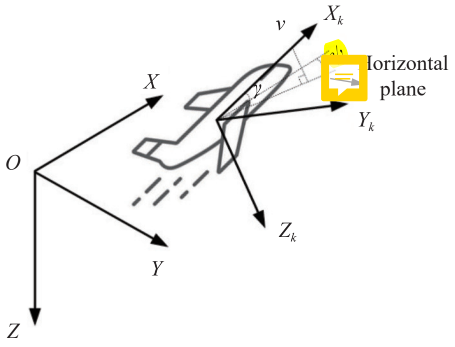


Fig. 1 Diagram of the coordinate system and flight path Angles. $S-Oxyz$ is the earth-fixed coordinate reference systems and $S_k-O_kx_ky_kz_k$ is the flight path axis system.

In terms of PE games, both the pursuer and the evader regard adjustment of the relative distance between them as the control objective. The difference is that the pursuer scheme reduces the distance to complete intercept, while the evader plans to increase it. In this way, we can construct a system that involves both the pursuer and the evader. The dynamics of aircraft PE games can be written as follows:

$$
\begin{cases}
\dot{v}_P=g\left(N_{x_a P}-\sin\gamma_P\right)\\
\dot{\psi}_P=-gN_{z_a P}\sin\phi_P/(v_P\cos\gamma_P)\\
\dot{\gamma}_P=-g\left(\cos\gamma_P+N_{z_a P}\cos\phi_P\right)/v_P\\
\dot{v}_E=g\left(N_{x_a E}-\sin\gamma_E\right)\\
\dot{\psi}_E=-gN_{z_a E}\sin\phi_E/(v_E\cos\gamma_E)\\
\dot{\gamma}_E=-g\left(\cos\gamma_E+N_{z_a}\cos\phi_E\right)/v_E\\
\Delta\dot{x}=v_P\cos\gamma_P\cos\psi_P-v_E\cos\gamma_E\cos\psi_E\\
\Delta\dot{y}=v_P\cos\gamma_P\sin\psi_P-v_E\cos\gamma_E\sin\psi_E\\
\Delta\dot{z}=-v_P\sin\gamma_P+v_E\sin\gamma_E
\end{cases}
\tag{18}
$$

where subscript P and E represent the pursuer and the evader, respectively. Δx, Δy and Δz represent the relative distance in the corresponding direction. For convenience and brevity, let $\bar{s}=[v_P,\psi_P,\gamma_P,v_E,\psi_E,\gamma_E,\Delta x,\Delta y,\Delta z]^{\mathrm{T}}$ denote the state vector.

In this manuscript, we define the optimal control strategy of the aircraft PE game from the following aspects. Let the control objective of the pursuer is minimizing its distance from the evader, while reducing energy consumption as much as possible. Conversely, the evader aims at maximizing the distance with the least control energy. The three-degree-of-freedom aircraft PE game in three-dimensional space is demonstrated in Fig. 2.

Then, we can construct the performance index of the

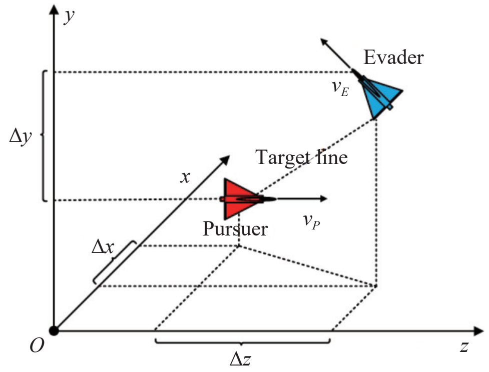


Fig. 2 Diagram of the aircraft PE game in three-dimensional space.

PE game as follows:

$$J\left(u_P, u_E\right) = \bar{s}(t_f)^{\mathrm{T}} H \bar{s}(t_f) + \int_{t_0}^{t_f} \left[\bar{s}^{\mathrm{T}} Q \bar{s} + u_P^{\mathrm{T}} R_P u_P - u_E^{\mathrm{T}} R_E u_E\right] \mathrm{d}t \tag{19}$$

where $H = \begin{pmatrix} 0_{6\times6} & 0_{6\times3} \\ 0_{3\times6} & I_{3\times3} \end{pmatrix}$, $Q = \begin{pmatrix} 0_{6\times6} & 0_{6\times3} \\ 0_{3\times6} & Q_s \end{pmatrix}$. R_P, R_E and subblock Q_s in Q are positive semi-definite symmetric matrices.

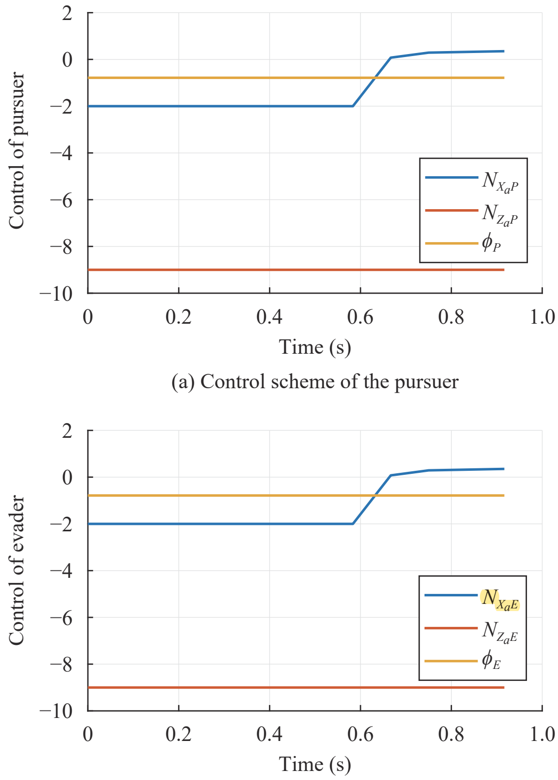
Corollary 1. The pursuit-evasion game described by (18) and (19) in this paper satisfies the Assumption 1 under the condition that a Nash equilibrium solution exists. i.e., if a Nash equilibrium solution exists, it is the unique one.

Proof. The detailed process of the relative proof is given in Appendix. □

4.2 Simulation resuslts

To perform the simulation, we set the time length to one second and select the initial velocity, flight path bank angle, flight path azimuth angle of the pursuer and evader as $[150, 0, 0]^{\mathrm{T}}$. The relative distance in the corresponding direction between them is $[200, 200, 200]^{\mathrm{T}}$. Meanwhile, we assume the initial admissible control law of the evader is $u_E^0 = [3, 5, 0]^{\mathrm{T}}$. Matrices R_P and R_E in (19) is chosen as identity matrix.

First, we perform the simulation under symmetrical experimental conditions, that is, the pursuer and the evader have the same initial state except for the relative distance. As shown in Figs. 3 and 4, in the situation where the pursuer and the evader have the same initial velocity, bank angle and azimuth angle, their optimal control strategy is also the same. Meanwhile, the changes of their state variables display similar trends. Finally, by executing their own optimal control scheme, the distance between pursuer and evader keeps almost unchanged which means that the two players reach the Nash equilibrium. The flight trajectories of the two aircrafts are displayed in Fig. 5.



(a) Control scheme of the pursuer

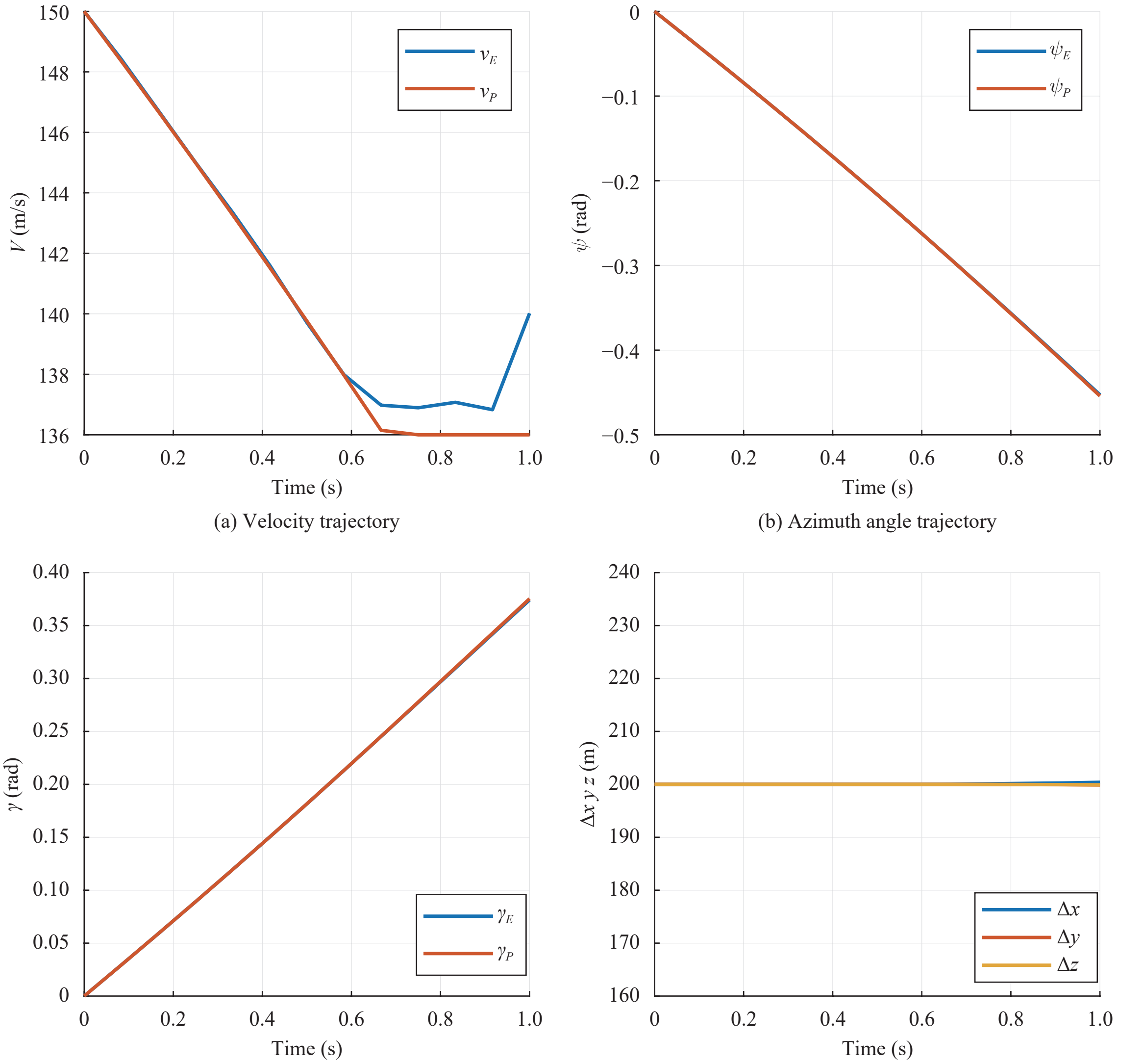
(b) Control scheme of the evader

Fig. 3 Control scheme of the pursuer and evader with the same initial state.

After the symmetric experiment, we conduct an asymmetric experiment, in which two players possess different initial states. For different initial velocities, the results can be seen in Figs. 6 and 7, the control laws of pursuer and evade become dissimilar. The distance between them increases at the beginning, then slows down and finally stays constant.

To demonstrate the effectiveness of the presented method, we introduce a model predictive control (MPC) method as supplemental experiments. MPC is an online optimization technique that allows the controller to update the control strategy according to the current system information at each decision time. MPC algorithm can deal with nonlinear systems with multiple inputs multiple outputs (MIMO), and it is also suitable for solving optimal control problems with state and control constraints. Selecting the same initial condition as the above experiment, then we have the results displayed as follows

As is shown in Figs. 8 and 9, the results have a similar conclusion to the experiment above. For the symmetrical situation, the pursuer and the evader have the same optimal control strategy. For asymmetrical situation, the control laws of pursuer and evade become dissimilar at the beginning. However, after a series of adjustments, their control law stays constant.



(a) Velocity trajectory

(b) Azimuth angle trajectory

(c) Bank angle trajectory

(d) Changes of relative distance in the corresponding direction

Fig. 4 State trajectory of the pursuer and evader with the same velocity

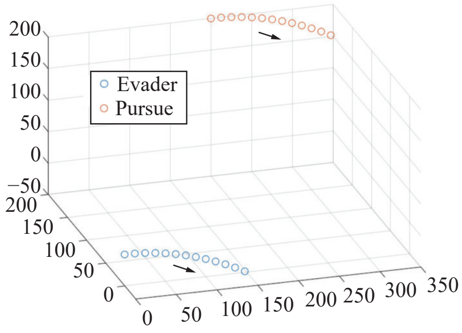
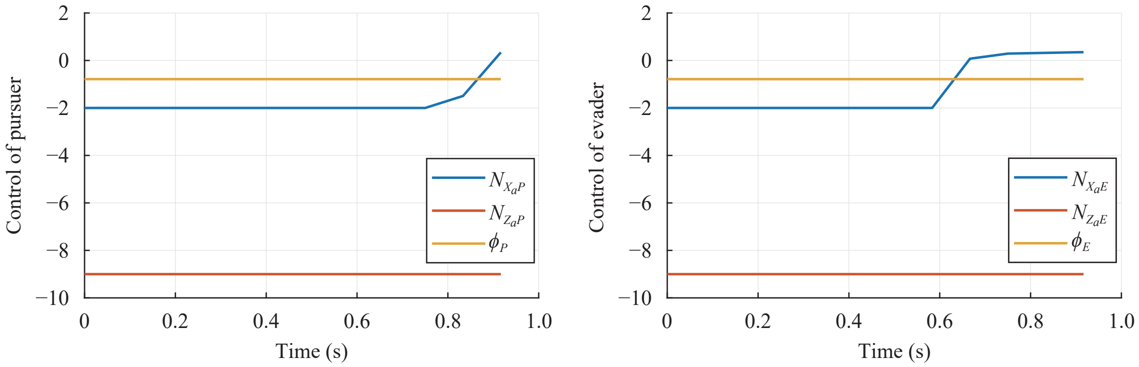


Fig. 5 Aircraft PE game in finite-horizon. The pursuer and evader utilize Algorithm 1 to make strategy.

5 Conclusions

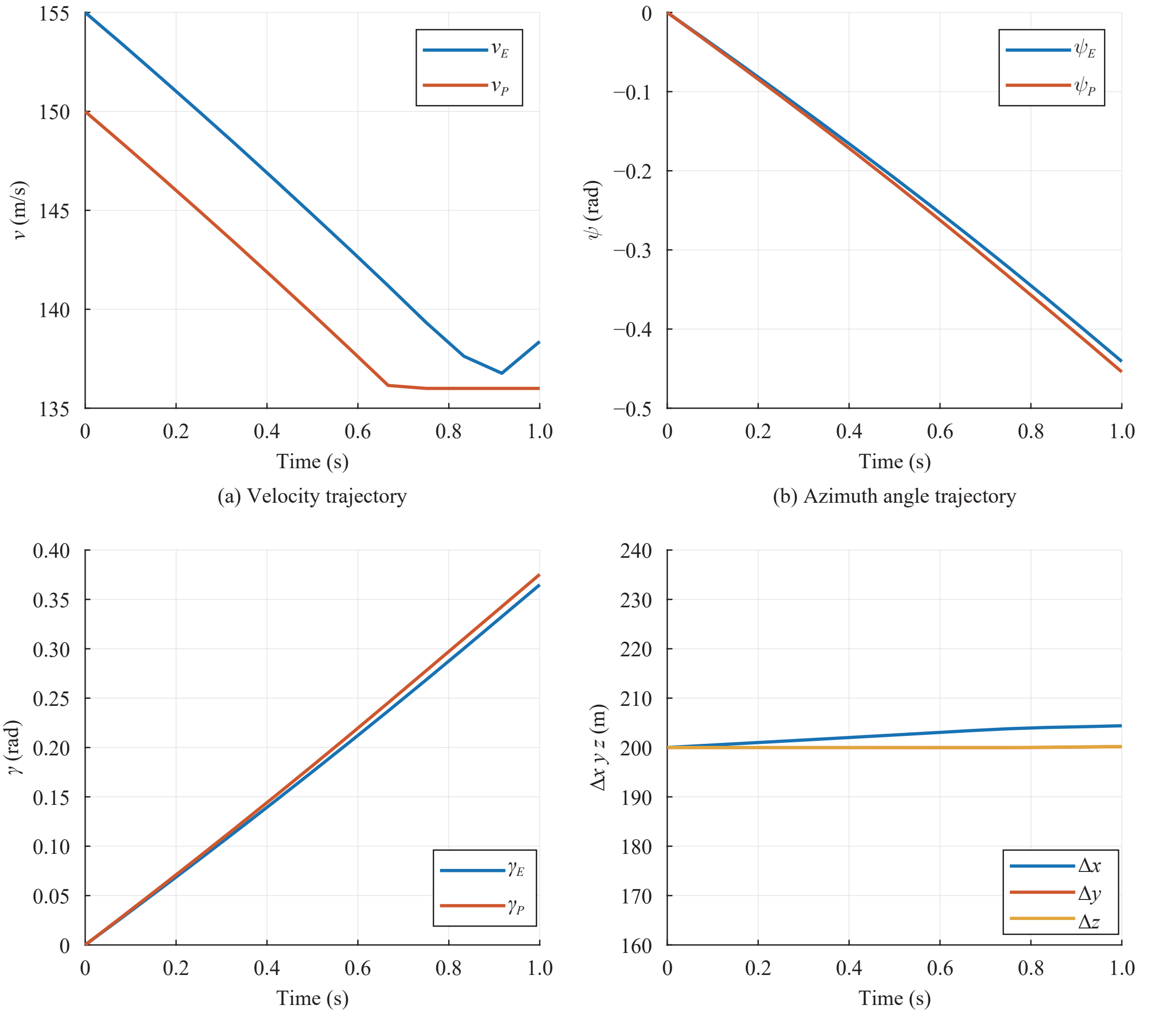
In this note, a self-paly iteration algorithm is presented towards a class of PE games to obtain the Nash equilibrium solution. To transform the two-sided optimization problem into two one-sided ones so that the direct collocation method can be applied, we regard the control strategy of one player as a parameter and iteratively update the optimal control law of the other one. Then, by solving optimization problems (11) and (12), which are approximated as nonlinear programming problems through the direct collocation method, the optimal control strategy of both the pursuer and the evader is achieved. Moreover, we prove the uniqueness of the Nash equilibrium solution in this game through analyzing the condition when Assumption 1 holds. The uniqueness of the Nash equilibrium solution ensures that the optimal control scheme solved by Algorithm 1 converges to the Nash equilibrium point. Finally, simulation results of an aircraft PE game are given to display effectiveness of the presented method. The main limitation of the proposed method is that there should be a unique Nash equilibrium point in the PE game to guarantee the convergence of the iterative control policy, which renders the applica-



(a) Control scheme of the pursuer

(b) Control scheme of the evader

Fig. 6 Control scheme of the pursuer and evader with different initial state



(a) Velocity trajectory

(b) Azimuth angle trajectory

(c) Bank angle trajectory.

(d) Changes of relative distance in the corresponding direction

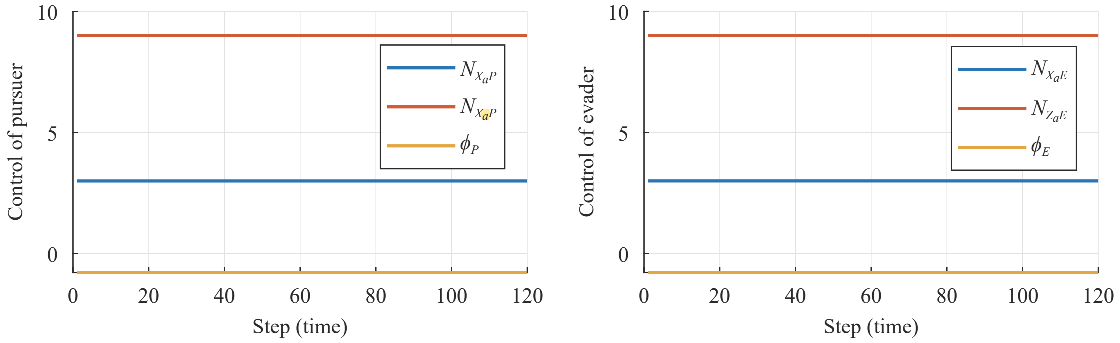
Fig. 7 State trajectory of the pursuer and evader with different velocity

tion of this method limited. The implication for practice is highlighted in the way that the proposed method can apply to PE games with the state and control constraints, which are unavoidable in practical applications. In the future, we will intensively study the multi-agent PE games with nonlinear dynamics and constraints.

Appendix Proof of Corollary 1

Introduce the Lagrange multipliers λ_{vP}, $\lambda_{\gamma P}$, $\lambda_{\varphi P}$, λ_{vE}, $\lambda_{\gamma E}$, $\lambda_{\varphi E}$, $\lambda_{\Delta x}$, $\lambda_{\Delta y}$, $\lambda_{\Delta z}$. For convenience and concision, we regard matrices R_P and R_E as identity matrices.

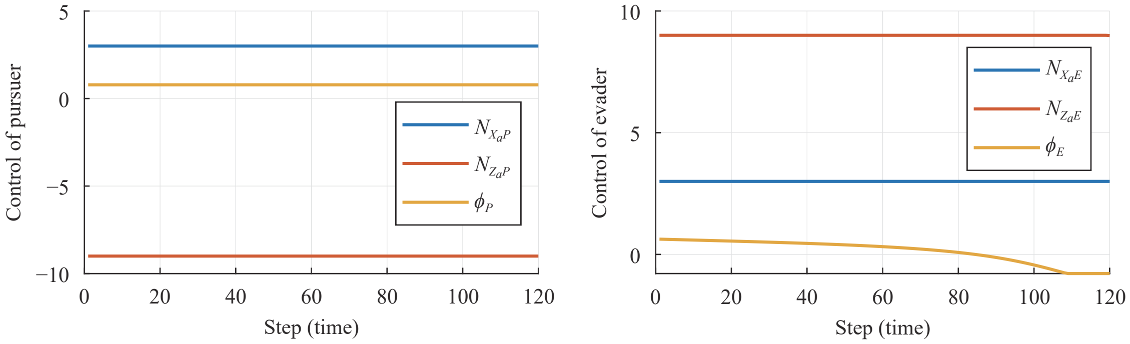
According to the system dynamics (18), the Hamilto-



(a) Control scheme of the pursuer

(b) Control scheme of the evader

Fig. 8 MPC control scheme of the pursuer and evader with the same initial states



(a) Control scheme of the pursuer

(b) Control scheme of the evader

Fig. 9 MPC control scheme of the pursuer and evader with different initial states

nian function can be written as follows:

$$
\begin{aligned}
H &= l\left(s, u_P, u_E, t\right) + \lambda_{v_P} g\left[N_{x_aP} - \sin\gamma_P\right] + \\
&\lambda_{\gamma_P}\frac{g\left[N_{z_aP}\cos\phi_P - \cos\gamma_P\right]}{v_P} + \\
&\lambda_{\varphi_P}\frac{gN_{z_aP}\sin\phi_P}{v_P\cos\gamma_P} + \lambda_{v_E} g\left[N_{x_aE} - \sin\gamma_E\right] + \\
&\lambda_{\gamma_E}\frac{g\left[N_{z_aE}\cos\phi_E - \cos\gamma_E\right]}{v_E} + \lambda_{\varphi_E}\frac{gN_{z_aE}\sin\phi_E}{v_E\cos\gamma_E} + \\
&\lambda_{\Delta x}\left(v_P\cos\gamma_P\cos\varphi_P - v_E\cos\gamma_E\cos\varphi_E\right) + \\
&\lambda_{\Delta y}\left(v_P\cos\gamma_P\sin\varphi_P - v_E\cos\gamma_E\sin\varphi_E\right) + \\
&\lambda_{\Delta z}\left(-v_P\sin\gamma_P + v_E\sin\gamma_E\right).
\end{aligned}
\tag{A1}
$$

Based on PMP, the Nash equilibrium solution satisfies

$$
H^* = \min_{u_P}\max_{u_E} H = \max_{u_E}\min_{u_P} H. \tag{A2}
$$

Calculate the partial derivative of the Hamiltonian function with respect to N_{x_aP}, N_{z_aP}, N_{x_aE}, N_{z_aE} at the Nash equilibrium point:

$$
\frac{\partial H}{\partial N_{x_aP}} = 2N_{z_aP}{+}\lambda_{v_P}g{=}0 \tag{A3}
$$

$$
\frac{\partial H}{\partial N_{x_aE}} = 2N_{x_aE}{+}\lambda_{v_P}g{=}0 \tag{A4}
$$

$$
\frac{\partial H}{\partial N_{z_aP}} = 2N_{z_aP} + \lambda_{\gamma_P}\frac{g\cos\phi_P}{v_P} + \lambda_{\varphi_P}\frac{g\sin\phi_P}{v_P\cos\gamma_P} = 0 \tag{A5}
$$

$$
\frac{\partial H}{\partial N_{z_aE}} = 2N_{z_aE} + \lambda_{\gamma_E}\frac{g\cos\phi_E}{v_E} + \lambda_{\varphi_E}\frac{g\sin\phi_E}{v_E\cos\gamma_E} = 0. \tag{A6}
$$

Then, we have

$$
N^*_{x_aP}(t) = -\frac{1}{2}\lambda_{v_P}(t)g \tag{A7}
$$

$$
N^*_{x_aE}(t) = -\frac{1}{2}\lambda_{v_E}(t)g \tag{A8}
$$

$$
N^*_{z_aP}(t) = -\frac{1}{2}\left(\lambda_{\gamma_P}\frac{g\cos\phi_P}{v_P} + \lambda_{\varphi_P}\frac{g\sin\phi_P}{v_P\cos\gamma_P}\right) \tag{A9}
$$

$$
N^*_{z_aE}(t) = -\frac{1}{2}\left(\lambda_{\gamma_E}\frac{g\cos\phi_E}{v_E} + \lambda_{\varphi_E}\frac{g\sin\phi_E}{v_E\cos\gamma_E}\right). \tag{A10}
$$

Calculate the partial derivative of the Hamiltonian function with respect to ϕ_P, ϕ_E at the Nash equilibrium point

$$\frac{\partial H}{\partial \phi_P} = 2\phi_P - \lambda_{\gamma_P}\frac{gN_{z_a P}\sin\phi_P}{v_P} + \lambda_{\varphi_P}\frac{gN_{z_a P}\cos\phi_P}{v_P\cos\gamma_P} = 0 \tag{A11}$$

$$\frac{\partial H}{\partial \phi_E} = 2\phi_E - \lambda_{\gamma_E}\frac{gN_{z_a E}\sin\phi_E}{v_E} + \lambda_{\varphi_E}\frac{gN_{z_a E}\cos\phi_E}{v_E\cos\gamma_E} = 0. \tag{A12}$$

Let

$$G_P = 2\phi_P - \lambda_{\gamma_P}\frac{gN_{z_a P}\sin\phi_P}{v_P} + \lambda_{\varphi_P}\frac{gN_{z_a P}\cos\phi_P}{v_P\cos\gamma_P} \tag{A13}$$

and calculate the partial derivative with respect to ϕ_P again

$$\frac{\partial G_P}{\partial \phi_P} = 2 - \lambda_{\gamma_P}\frac{gN_{z_a P}\cos\phi_P}{v_P} - \lambda_{\varphi_P}\frac{gN_{z_a P}\sin\phi_P}{v_P\cos\gamma_P}. \tag{A14}$$

For λ_{γ_P} in (A14), according to (9)

$$\begin{aligned} \dot{\lambda}_{\gamma_P} &= \frac{\partial H}{\partial \gamma_P} = -\frac{\lambda_{\gamma_P}\sin\gamma_P}{v_P} \\ \lambda_{\gamma_P} &= \mathrm{e}^{-\frac{\sin\gamma_P}{v_P}t} < 1 \end{aligned} \tag{A15}$$

and for λ_{φ_P}

$$\begin{aligned} \dot{\lambda}_{\varphi_P} &= \frac{\partial H}{\partial \varphi_P} = 0 \\ \lambda_{\varphi_P}(t_f) &= \frac{\partial \Phi(x^*(t_f), t_f)}{\partial \varphi_P} = 0 \\ \lambda_{\varphi_P} &= 0. \end{aligned} \tag{A16}$$

Based on (A15), (A16) and the constraints of the dynamics, we get

$$\frac{\partial G_P}{\partial \phi_P} = 2 - \lambda_{\gamma_P}\frac{gN_{z_a P}\cos\phi_P}{v_P} - \lambda_{\varphi_P}\frac{gN_{z_a P}\sin\phi_P}{v_P\cos\gamma_P} > 0 \tag{A17}$$

for any condition.

From the analysis above, the Nash equilibrium point is uniquely obtained at the boundary of the constrains or the point where the derivative is zero. □

References

[1] R. Isaacs. *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization*, New York, USA: Dover Publications, 1999.

[2] P. K. Chintagunta, V. R. Rao. Pricing strategies in a dynamic duopoly: A differential game model. *Management Science*, vol. 42, no. 11, pp. 1501–1514, 1996. DOI: 10.5555/2777472.2777473.

[3] L. A. Petrosyan, N. A. Zenkevich. *Game Theory*, World Scientific, 1996. (查阅所有网上资料，未找到出版地信息，请联系作者补充)

[4] Y. Mousavi, A. Zarei, A. Mousavi, M. Biari. Robust optimal higher-order-observer-based dynamic sliding mode control for VTOL unmanned aerial vehicles. *International Journal of Automation and Computing*, vol. 18, no. 5, pp. 802–813, 2021. DOI: 10.1007/s11633-021-1282-3.

[5] H. G. Zhang, Q. L. Wei, D. R. Liu. An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games. *Automatica*, vol. 47, no. 1, pp. 207–214, 2011. DOI: 10.1016/j.automatica.2010.10.033.

[6] N. Greenwood. A differential game in three dimensions: The aerial dogfight scenario. *Dynamics and Control* , vol. 2, no. 2, pp. 161–200, 1992. DOI: 10.1007/BF02169496.

[7] K. Horie, B. A. Conway. Optimal fighter pursuit-evasion maneuvers found via two-sided optimization. *Journal of Guidance, Control, and Dynamics*, vol. 29, no. 1, pp. 105–112, 2006. DOI: 10.2514/1.3960.

[8] Z. Y. Li, H. Zhu, Z. Yang, Y. Z. Luo. A dimension-reduction solution of free-time differential games for spacecraft pursuit-evasion. *Acta Astronautica*, vol. 163, pp. 201–210, 2019. DOI: 10.1016/j.actaastro.2019.01.011.

[9] J. F. Zhou, L. Zhao, H. Li, J. H. Cheng, S. Wang. Compensation control strategy for orbital pursuit-evasion problem with imperfect information. *Applied Sciences*, vol. 11, no. 4, Article number 1400, 2021. DOI: 10.3390/app11041400.

[10] M. Salimi, M. Ferrara. Differential game of optimal pursuit of one evader by many pursuers. *International Journal of Game Theory*, vol. 48, no. 2, pp. 481–490, 2019. DOI: 10.1007/s00182-018-0638-6.

[11] V. G. Lopez, F. L. Lewis, Y. Wan, E. N. Sanchez, L. L. Fan. Solutions for multiagent pursuit-evasion games on communication graphs: Finite-time capture and asymptotic behaviors. *IEEE Transactions on Automatic Control*, vol. 65, no. 5, pp. 1911–1923, 2020. DOI: 10.1109/TAC.2019.2926554.

[12] E. Garcia, D. W. Casbeer, A. von Moll, M. Pachter. Multiple pursuer multiple evader differential games. *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2345–2350, 2021. DOI: 10.1109/TAC.2020.3003840.

[13] D. W. Oyler. Contributions to Pursuit-Evasion Game Theory, Ph. D. dissertation, University of Michigan, USA, 2016.

[14] J. F. Zhang, L. Y. Liu, S. Z. Fu, S. Li. Event-triggered control of positive switched systems with actuator saturation and time-delay. *International Journal of Automation and Computing*, vol. 18, no. 1, pp. 141–155, 2021. DOI: 10.1007/s11633-020-1245-0.

[15] D. Wang, M. M. Ha, M. M. Zhao. The intelligent critic framework for advanced optimal control. *Artificial Intelligence Review*, vol. 55, no. 1, pp. 1–22, 2022. DOI: 10.1007/s10462-021-10118-9.

[16] P. Soravia. Pursuit-evasion problems and viscosity solutions of isaacs equations. *SIAM Journal on Control and Optimization*, vol. 31, no. 3, pp. 604–623, 1993. DOI: 10.1137/0331027.

[17] Q. L. Wei, D. R. Liu, Q. Lin, R. Z. Song. Adaptive dynamic programming for discrete-time zero-sum games. *IEEE*

Transactions on Neural Networks and Learning Systems, vol. 29, no. 4, pp. 957–969, 2018. DOI: 10.1109/TNNLS.2016.2638863.

[18] L. S. Pontryagin. *Mathematical Theory of Optimal Processes*, Boca Raton, USA: CRC Press, 1987.

[19] R. W. Carr, R. G. Cobb, M. Pachter, S. Pierce. Solution of a pursuit-evasion game using a near-optimal strategy. *Journal of Guidance*, *Control*, *and Dynamics*, vol. 41, no. 4, pp. 841–850, 2018. DOI: 10.2514/1.G002911.

[20] M. Pontani, B. A. Conway. Numerical solution of the three-dimensional orbital pursuit-evasion game. *Journal of Guidance*, *Control*, *and Dynamics*, vol. 32, no. 2, pp. 474–487, 2009. DOI: 10.2514/1.37962.

[21] Y. L. Yang, K. G. Vamvoudakis, H. Modares. Safe reinforcement learning for dynamical games. *International Journal of Robust and Nonlinear Control*, vol. 30, no. 2, pp. 3706–3726, 2020. DOI: 10.1002/rnc.4962.

[22] M. M. Ha, D. Wang, D. R. Liu. Discounted iterative adaptive critic designs with novel stability analysis for tracking control. *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1262–1272, 2022. DOI: 10.1109/JAS.2022.105692.

[23] Y. Yang, D. Ding, H. Xiong, Y. Yin, D. Wunsch. Online barrier-actor-critic learning for H_∞ control with full-state constraints and input saturation. *Journal of the Franklin Institute*, vol. 357, no. 7, pp. 3316–3344, 2020. DOI: 10.1016/j.jfranklin.2019.12.017.

[24] Y. Kartal, K. Subbarao, A. Dogan, F. Lewis. Optimal game theoretic solution of the pursuit-evasion intercept problem using on-policy reinforcement learning. *International Journal of Robust and Nonlinear Control*, vol. 31, no. 16, pp. 7886–7903, 2021. DOI: 10.1002/rnc.5719.

[25] J. Selvakumar, E. Bakolas. Feedback strategies for a reach-avoid game with a single evader and multiple pursuers. *IEEE Transactions on Cybernetics*, vol. 51, no. 2, pp. 696–707, 2021. DOI: 10.1109/TCYB.2019.2914869.

[26] H. Xu. Finite-horizon near optimal design of nonlinear two-player zero-sum game in presence of completely unknown dynamics. *Journal of Control*, *Automation and Electrical Systems*, vol. 26, no. 4, pp. 361–370, 2015. DOI: 10.1007/s40313-015-0180-8.

[27] C. X. Mu, K. Wang, C. Y. Sun. Policy-iteration-based learning for nonlinear player game systems with constrained inputs. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*: *Systems*, vol. 51, no. 10, pp. 6488–6502, 2021. DOI: 10.1109/TSMC.2019.2962629.

[28] X. H. Cui, H. G. Zhang, Y. H. Luo, P. F. Zu. Online finite-horizon optimal learning algorithm for nonzero-sum games with partially unknown dynamics and constrained inputs. *Neurocomputing*, vol. 185, pp. 37–44, 2016. DOI: 10.1016/j.neucom.2015.12.021.

[29] I. E. Weintraub, M. Pachter, E. Garcia. An introduction to pursuit-evasion differential games. In *Proceedings of American Control Conference*, IEEE, Denver, USA, pp. 1049–1066, 2020. DOI: 10.23919/ACC45564.2020.9147205.

[30] M. H. Breitner, H. J. Pesch, W. Grimm. Complex differential games of pursuit-evasion type with state constraints, Part 1: Necessary conditions for optimal open-loop strategies. *Journal of Optimization Theory and Applications*, vol. 78, no. 3, pp. 419–441, 1993. DOI: 10.1007/BF00939876.

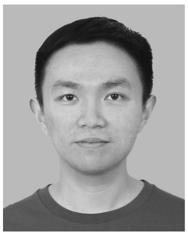
[31] A. Bressan. Noncooperative Differential Games. A Tutorial. Department of Mathematics, Penn State University, USA, 2010. (查阅所有网上资料，未能确认文献类型，请联系作者确认)(查阅所有网上资料，未找到报告编号信息，请联系作者确认)

[32] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, Y. Zhang. On the formulation and theory of the newton interior-point method for nonlinear programming. *Journal of Optimization Theory and Applications*, vol. 89, no. 3, pp. 507–541, 1996. DOI: 10.1007/BF02275347.

[33] P. T. Boggs, J. W. Tolle. Sequential quadratic programming. *Acta Numerica*, vol. 4, pp. 1–51, 1995. DOI: 10.1017/S0962492900002518.

[34] A. R. Conn, N. I. M. Gould, P. L. Toint. *Trust-Region Methods*, Philadelphia, USA: SIAM, 2000.

[35] F. Austin, G. Carbone, M. Falco, H. Hinz, M. Lewis. Automated maneuvering decisions for air-to-air combat. In *Guidance, Navigation and Control Conference*, Monterey, USA: AIAA, pp. 2393, 1987. DOI: 10.2514/6.1987-2393.



Xin Wang received the B.Sc. degree in electronic information engineering from Zhengzhou University, China in 2012, and the M.Sc. degree in control engineering from University of Science and Technology Beijing, China in 2015. He is currently a Ph.D. degree candidate in control theory and control engineering at State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences China, and University of Chinese Academy of Sciences, China.

His research interests include reinforcement learning, adaptive dynamic programming, optimal control and multi-agent system.

E-mail: wangxin2019@ia.ac.cn
ORCID iD: 0000-0002-8035-5586

Qinglai Wei received the B.Sc. degree in automation, and the Ph.D. degree in control theory and control engineering from Northeastern University, China in 2002 and 2009, respectively. From 2009 to 2011, he was a postdoctoral fellow with State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China. He is currently a professor of the institute and the associate director of the laboratory. He has authored four books, and published over 80 international journal papers.

He is the Secretary of IEEE Computational Intelligence Society (CIS) Beijing Chapter since 2015. He was Guest Editors for several international journals. He was a recipient of IEEE/CAA Journal of Automatica Sinica Best Paper Award, IEEE System, Man, and Cybernetics Society Andrew P. Sage Best Transactions Paper Award, IEEE Transactions on Neural Networks and Learning Systems Outstanding Paper Award, the Outstanding Paper Award of Acta Automatica Sinica, IEEE 6th Data Driven Control and Learning Systems Conference (DDCLS2017) Best Paper Award, and Zhang Siying Outstanding Paper Award of Chinese Control and Decision Conference (CCDC). He was a recipient of Shuang-Chuang Talents in Jiangsu Province, China, Young Researcher Award of Asia Pacific Neural Network Soci-

ety (APNNS), Young Scientst Award and Yang Jiachi Tech Award of Chinese Association of Automation (CAA). He is a Board of Governors (BOG) member of the International Neural Network Society (INNS) and a council member of CAA.

His research interests include adaptive dynamic programming, neural-networks-based control, optimal control, nonlinear systems and their industrial applications.

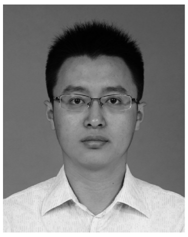
E-mail: qinglai.wei@ia.ac.cn

ORCID iD: 0000-0001-7002-9800

Tao Li received the B. Sc. degree in automation from Northeastern University, China in 2019. He is currently a Ph. D. degree candidate in control theory and control engineering at State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, China.

His research interests include adaptive dynamic programming, reinforcement learning and approximate dynamic programming.

E-mail: litao2019@ia.ac.cn



Jie Zhang received the B. Sc. degree in information and computing science from Tsinghua University, China in 2005, and the Ph. D. degree in technology of computer application from University of Chinese Academy of Sciences, China in 2015. He has been an associate professor with State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China, since 2016.

His research interests include parallel control, mechanism design, optimal control and multiagent reinforcement learning.

E-mail: jie.zhang@ia.ac.cn (Corresponding author)

ORCID iD: 0000-0001-6046-4497