追逃博弈问题研究综述

迟嵩禹 [1] 李 帅 [1] 王 晨 [1, 2] 谢广明 [1, 3]

摘 要 作为多智能体对抗博弈问题的重要分支, 追逃博弈 (Pursuit-evasion, PE) 问题在控制和机器人领域得到广泛应用, 受到众多研究者的密切关注. 追逃博弈问题主要聚焦于追逐者和逃跑者双方为实现各自目标而展开的动态博弈: 追逐者试图在最短时间内抓到逃跑者, 逃跑者的目标则是避免被捕获. 本文概述追逃博弈问题的相关研究进展, 从空间环境、信息获取等五个方面介绍追逃博弈问题的各类设定; 简述理论求解、数值求解等四种当下主流的追逃博弈问题求解方法. 通过对现有研究的总结和分析, 给出几点研究建议, 对未来追逃博弈问题的发展具有一定指导意义.

关键词 追逃博弈问题, 多智能体, 对抗博弈, 微分对策

引用格式 迟嵩禹, 李帅, 王晨, 谢广明. 追逃博弈问题研究综述. 自动化学报, 2025, **51**(4): 705−726

DOI 10.16383/j.aas.c240396 **CSTR** 32138.14.j.aas.c240396

A Review of Research on Pursuit-evasion Games

CHI Song-Yu[1] LI Shuai[1] WANG Chen[1, 2] XIE Guang-Ming[1, 3]

Abstract As an important branch of multi-agent adversarial games, pursuit-evasion (PE) games have found widespread applications in the fields of control and robotics, attracting considerable attention from researchers. PE games primarily focus on the dynamic games between pursuer and evader, each striving to achieve their respective objectives: The pursuer aims to capture the evader as quickly as possible, while the evader′s goal is to avoid capture. This article provides an overview of the research progress in PE games, and introduces various settings of PE games across five key dimensions, including spatial environment, information acquisition, and so on. It briefly describes four mainstream methods for solving PE games, including theoretical approaches, numerical approaches, and so on. By summarizing and analyzing existing researches, this article offers several research suggestions, which are expected to provide significant guidance for future developments in PE games.

Key words Pursuit-evasion (PE) games, multi-agent, adversarial games, differential games

Citation Chi Song-Yu, Li Shuai, Wang Chen, Xie Guang-Ming. A review of research on pursuit-evasion games. *Acta Automatica Sinica*, 2025, **51**(4): 705−726

随着系统与控制科学的飞速发展, 智能群体控制与博弈问题逐渐成为研究热点, 其中追逃博弈 (Pursuit-evasion, PE) 问题 (以下简称为追逃博弈) 受到广泛关注, 积累了丰硕的成果. 追逃博弈中的参与者主要分为追逐者与逃跑者, 二者为实现各自目标而展开博弈. 追逐者的主要目标是尽快捕获逃跑者, 或者将其驱赶至特定区域内; 而逃跑者则致力于避免被追逐者捕获, 或尽可能延长被捕获的时间.

收稿日期 2024-06-28 录用日期 2024-10-08
Manuscript received June 28, 2024; accepted October 8, 2024
国家自然科学基金 (12272008, U22A2062, U23B2037) 资助
Supported by National Natural Science Foundation of China (12272008, U22A2062, U23B2037)
本文责任编委 唐漾
Recommended by Associate Editor TANG Yang
1. 北京大学工学院智能仿生设计实验室 北京 100871 2. 北京大学软件工程国家工程研究中心 北京 100871 3. 北京大学海洋研究院 北京 100871
1. Intelligent Biomimetic Design Laboratory, College of Engineering, Peking University, Beijing 100871 2. National Engineering Research Center of Software Engineering, Peking University, Beijing 100871 3. Institute of Ocean Research, Peking University, Beijing 100871

追逃博弈源于自然界. 捕食者追逐猎物的过程就是最自然的追逃博弈. 在非洲草原上, 猎豹追逐羚羊, 这是典型的一追一逃问题; 美国黄石公园里, 狼群围捕野牛群, 属于多追逐者−多逃跑者问题. 在海洋里, 鲨鱼攻击鱼群, 对应的是一追多逃问题; 虎鲸群猎捕冰上的海豹, 这是多追逐者−单逃跑者问题. 捕猎双方目标相反, 捕食者往往要在最小代价下抓到猎物, 比如消耗最少的体力、从某个安全角度抓捕等, 而猎物要保证自己不被抓到, 或是不被捉到的同时跑到某一安全区域.

有关动物追逃行为的研究发现, 一些捕食者/猎物会采取一些先天的或是后天习得的策略[1−2]: 比如羚羊可以迅速改变其逃跑方向, 而大多数捕食者转向需要一定的转弯半径[3]; 狮群围捕角马时各自分工, 有些负责收缩包围圈, 有些负责从角马群中分割出落单的个体[4]. 追逃博弈的基本情形设定、追逃策略设计、双方胜利目标等均受到动物追逃行为的启发[5]. 例如, Li[6−7] 受动物追逃行为启发, 研究速

度更快的追逐者与转弯更灵活的逃跑者之间的追逃博弈, 并设计了仿生追逃策略.

追逃博弈被广泛应用于人类社会. 在导航上, 追逃策略可用于规划行驶路线以保证安全车距; 在治安管理上, 需要合理的抓捕策略来实现对犯罪嫌疑人的追捕; 在机器人和智能控制领域上, 无人机 (无人车、无人艇) 集群执行任务时, 常常涉及追逃情境下的策略; 在战斗机空战 (Dogfight) 中, 追击敌机、围捕目标等都可以用到追逃博弈中的策略.

目前, 追逃博弈发展出多个分支方向, 也随之产生了各种求解方法. Weintraub 等[8] 从追逃双方数量对比、两追一问题以及目标防御问题入手来分析追逃博弈, Mu 等[9] 则从群体智能的视角来讨论追逃博弈. 但这些工作都侧重于从某个角度切入来介绍追逃博弈, 缺少对追逃博弈系统全面的论述.

本文综合国内外有关追逃博弈的大量文献, 对追逃博弈的发展历史、研究现状进行介绍, 分别从个体动态模型、个体数量变化、空间环境、信息获取、状态空间五个角度介绍追逃博弈的各种设定, 并详细叙述理论求解法、数值求解法、积分强化学习法和几何法这四种求解方法, 分析每种方法的优势与缺点. 进一步地, 本文探讨追逃博弈的应用场景, 并且对代表性的追逃博弈文献做了分类总结 (详见附录 A 中的表 A1). 基于这些讨论, 本文给出几点研究建议, 对于今后追逃博弈的研究具有一定的指导意义.

1 追逃博弈研究简史

追逃博弈历史悠久, 在图 1 中总结了追逃博弈研究的概况. 追逃博弈始于上世纪中叶, Isaacs[10] 在给兰德公司提交的报告里首次提到了“微分对策”这一概念, 并且首次利用微分对策的方法来建模分析杀人司机问题 (Homicidal chauffeur problem)、各向同性火箭问题 (Isotropic rocket problem) 等一系列经典的追逃博弈问题, 给出了理论解. 利用微分对策求解追逃问题, 通常首先根据问题设定建立动力学模型, 然后确定优化目标 (追逃问题中一般以抓捕时间为目标, 也可将能量、距离等实际因素考虑在内, 形成一个综合的目标), 最后进行求解得到追逃双方的最优策略. 进一步地, Isaacs[10] 还提出了 HJI (Hamilton-Jacobi-Isaacs) 方程, 用于确定最优策略和值函数应该满足的条件. 这些研究成果收录在其著作 *Differential Games*[10] 里. 此外, Isaacs[10] 也给出了一些微分对策的基本概念, 如分散曲面 (Dispersal surface)、值函数 (Value function)、界栅 (Barrier)、可用部分 (Useable part) 等. Isaacs 的工作奠定了追逃博弈的理论基础, 影响深远. 基于 Isaacs[10] 的工作, Starr 和 Ho[11−12] 也做了一系列研究.

我国的微分对策和追逃博弈研究始于上世纪八十年代, 自动控制专家张嗣瀛院士最早对此开展了研究. 在他的著作《微分对策》[13] 里, 详细地介绍了 Isaacs[10] 的研究成果, 首次将定量微分对策 (Game of degree)、定性微分对策 (Game of kind) 和奇异曲面 (Singular surface) 等引入国内, 讨论了追逃博弈的具体应用, 对国内追逃博弈的研究具有重要的指导意义. 此后, 我国学者不断发掘追逃博弈的研究价值, 拓宽其应用方向[14−15].

2 追逃博弈中的个体动态模型

建立合适的动态模型是研究追逃博弈的第一

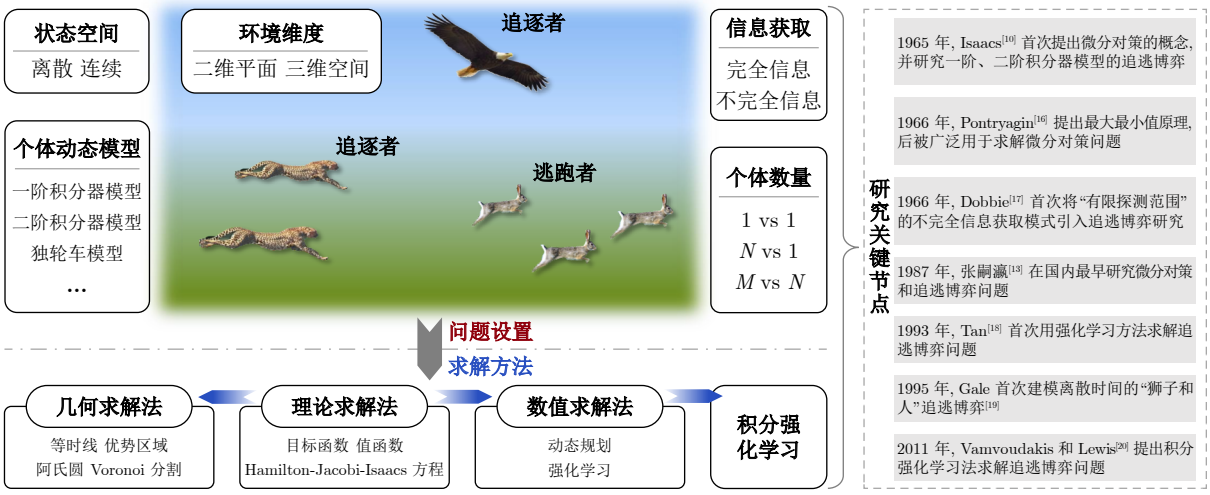

图 1 追逃博弈研究概况

Fig. 1 Research overview of PE games

步, 因此本文首先从动态模型的分类来讨论追逃博弈. 动态模型是用于描述智能体系统的输入、运动状态等之间关系的数学模型. 在追逃博弈中, 动态模型主要有一阶积分器模型、二阶积分器模型和独轮车模型等. 其他条件相同时, 动态模型结构越简洁, 相应的最优策略求解复杂度越低. 此外, 若双方遵循的动态模型不一致, 其策略求解难度往往高于双方遵循相同动态模型的情况. 接下来将分别介绍一阶积分器模型、二阶积分器模型、独轮车模型和航天器模型的研究情况.

2.1 一阶积分器模型

一阶积分器模型是追逃博弈中最简单、最常见的动态模型. 追逃博弈的早期研究中设定的大都是一阶积分器模型[12, 21]. 在此模型下, 多为速度大小恒定、方向任意改变, 也称之为简单动态模型. 目前, 一阶积分器模型仍被广泛应用[22−24].

在讨论一些特殊设定时, 研究者们为了剥离非关键复杂因素、专注探索核心创新点, 往往会选择采用简单动态模型. 例如, Oyler 等[22] 在考虑障碍物对等时线的形状变化影响时, 由于一阶模型的简单运动特性, 其等时线形状更为规则, 便于分析; 再比如 Garcia 等[23] 研究的三维空间中的追逃博弈, 当空间维数增加时, 一阶积分器模型仍然为追逃双方动态模型的首选. 最基本的一阶积分器模型一般定义为

$$\dot{\boldsymbol{x}} = \boldsymbol{v} \tag{1}$$

其中, $\boldsymbol{x}$ 为智能体的位置坐标向量, $\boldsymbol{v}$ 为速度向量. 以二维平面上的追逃博弈为例, 可进一步将一阶积分器模型细化为

$$\begin{cases} \dot{x} = v\cos\theta \\ \dot{y} = v\sin\theta \end{cases} \tag{2}$$

其中, $\boldsymbol{x} = (x,\ y)^{\mathrm{T}}$ 为平面坐标, v 为速度大小, θ 为智能体前进方向. 在最优策略下, 追逃双方均以最大速度移动[10], 所以速度值一般取常数, 因此, 整个动态模型的输入仅为运动方向, 控制简单. 在一阶积分器模型下, 可将整个问题区域分为三部分: 追逐者先到达的区域、逃跑者先到达的区域以及双方同时到达的区域. 其中双方同时到达的区域不仅是前两个区域的分界线, 而且在该区域中会产生一些重要的概念, 如优势区域[25]、阿波罗尼斯圆[26] (Apollonius circle, 以下简称阿氏圆) 及 Voronoi 分割[27] 等. 在第 7.4 节中, 我们将详细解释这些概念在追逃博弈中的应用.

一阶积分器模型被研究者们用来研究各类追逃博弈. 除了传统的追逃博弈[28], Shishika 和 Kumar[29] 研究了多入侵者多防御者的边界防御问题; Liang 等[30] 分阶段考虑了 TAD (Target-attacker-defender) 问题的最佳策略; Yan 等[31] 研究了一个有界凸区域里的 Reach-avoid 问题. 尽管追逃博弈问题的研究领域持续涌现新颖的理论与方法, 一阶积分器模型凭借其无可比拟的简洁性和广泛的适用性, 在这一研究领域中始终占据举足轻重的地位.

2.2 二阶积分器模型

在一阶积分器模型下智能体可以随意改变方向, 这是不符合实际的. 因此, 可以加上对加速度的控制. 沿此思路, 二阶积分器模型应运而生. 基本的二阶积分器模型为

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{v} \\ \dot{\boldsymbol{v}} = \boldsymbol{u} \end{cases} \tag{3}$$

其中, $\boldsymbol{x}$, $\boldsymbol{v}$ 定义同前文, $\boldsymbol{u}$ 为对速度 $\boldsymbol{v}$ 的控制, 模型中第二个方程控制速度是可微的, 因此其运动路线一定是光滑的. 二阶积分器模型最早由 Isaacs[10] 提出, 在研究各向同性火箭问题时, Isaacs[10] 又进一步提出了带有阻尼的二阶积分器模型

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{v} \\ \dot{\boldsymbol{v}} = -\mu\boldsymbol{v} + \boldsymbol{u} \end{cases} \tag{4}$$

其中, 称 $\mu > 0$ 为阻尼系数, 用来限制速度大小. 相比普通二阶模型, 此模型中速度与加速度间的相互作用更为复杂, 可以将二阶模型扩展到更多应用场景中.

二阶积分器模型得到了广泛的使用[32−33]. Isaacs[10] 把追逐者建模为二阶积分器模型, 逃跑者建模为一阶积分器模型, 讨论一追一逃问题. Bakolas[34] 研究了二阶积分器下多智能体的追逃博弈, 提出一种类似于 Voronoi 分割法的数值方法来求解, Selvakumar 和 Bakolas[35] 对此做了跟进研究. Coon 和 Panagou[36] 把所有的智能体都建模为纯二阶积分器模型, 以此来研究 TAD 问题, 但是文中仅给出了仿真结果, 并未给出理论解. 这也是目前二阶积分器模型研究中普遍存在的问题: 多数停留在情境设定的创新、利用仿真结果检验最优性这些层面, 缺乏严谨的理论分析和数学证明. 例如, Chipade 和 Panagou[37−38] 讨论了二阶积分器下的多智能体防御、围捕问题, 然而他们使用动态规划的方法求解, 最后得到的是数值解, 也没有给出严格的理论解. 针对带有阻尼的二阶积分器, Wang 等[5] 提出一套包含对齐、避碰、追逃规则在内的智能体控制系统, 并进行了仿真讨论以验证其有效性, 然而文中同样缺少深入的理论

分析. Li 等[39] 率先给出了二阶积分器模型下一追一逃问题的理论解, 通过严格的微分对策方法推导出了最优策略, 这是具有开拓性的工作, 可见二阶积分器模型蕴含着巨大的研究潜力, 等待我们去发掘.

2.3 独轮车模型

追逃博弈中常常将智能体描述为汽车、无人机等常见的物体, 这些物体都有转弯半径, 而一阶、二阶模型没有考虑到这一点, 因此有人对独轮车模型的追逃问题开展了研究. 在二维平面上, 独轮车模型的微分方程通常表述为

$$\begin{cases} \dot{x} = v\cos\theta \\ \dot{y} = v\sin\theta \\ \dot{\theta} = \omega \end{cases} \tag{5}$$

其中, $(x,\ y)$ 为智能体的平面坐标, v 为线速度, θ 为运动方向, ω 为角速度. 在独轮车模型里, 线速度、运动方向和角速度这三个参数都是可变的, 从而精确模拟汽车等物体在平面上的运动.

当独轮车模型的智能体只能前向运动且速度恒定时, 就得到了 Dubins 模型[40]. 最早将 Dubins 模型引入追逃博弈分析的是 Isaacs, 他在研究杀人司机问题时, 将追逐者建模为 Dubins 模型, 逃跑者建模为一阶积分器模型. 基于此框架, Patsko 和 Turova[41] 回顾了杀人司机问题的研究历程, 总结了各类动态模型; Pachter 和 Coates[42] 给出了经典杀人司机问题的完整解. 将逃跑者建模为 Dubins 模型, 令追逐者为一阶模型, 就得到了杀人司机的对偶问题、自杀行人问题, Exarchos 等[43] 首次提出这一构思. 进一步地, Nath 和 Ghose[24, 44] 将一追一逃问题中的追逃双方都建模为 Dubins 模型, 而且令双方不了解彼此的转弯半径, 在此情况下寻找逃跑者最佳策略. Kokolakis 和 Vamvoudakis[40] 将平面的 Dubins 模型拓展到了三维飞行器上, 研究了自主无人机集群跟踪问题, 文中所有智能体都采用了 Dubins 模型, 难以直接得到理论解, 所以文中利用强化学习的方法拟合值函数来求解.

对于一般的独轮车模型, 控制系统本身的复杂性使得相关研究更为困难. 对此, 一些研究者做出了尝试. Sani 等[45] 运用非线性模型预测控制 (NMPC) 技术研究独轮车模型下存在障碍物的一追一逃问题; 同样利用 NMPC, Manoharan 等[46] 考虑了 TAD 问题中追逃各方均为独轮车模型的情况.

2.4 航天器模型

近年来, 航空航天技术快速发展, 推动了追逃博弈研究向这一领域延伸. 但是传统的一阶、二阶积分器模型无法适配宇宙空间, 因此, 结合航天器所处环境与自身特点的航天器模型顺势而生, 用于解决航天器间的追逃问题.

航天器一般定义在 LVLH 坐标系中[47]. 如图 2 所示, 该坐标系的原点位于参考航天器的质心 o, ox 轴沿参考航天器的矢径方向, oy 轴沿参考航天器运动的迹向, oz 轴为轨道面的法线方向并与 ox, oy 构成右手坐标系.

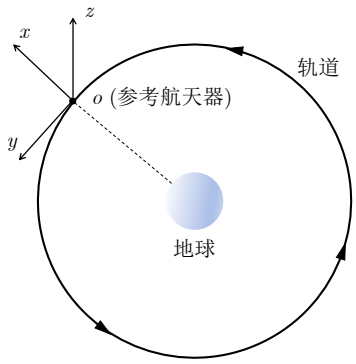

图 2 LVLH 坐标系

Fig. 2 LVLH coordinate system

在此坐标系下, 当参考轨道为圆轨道时, 并且航天器与参考航天器之间的距离相对于参考轨道的半径为小量时, 基本的航天器相对运动模型如下

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \tag{6}$$

其中, $\boldsymbol{x} = (x,\ y,\ z,\ \dot{x},\ \dot{y},\ \dot{z})^{\mathrm{T}}$ 为航天器在 LVLH 坐标系中的位置和速度矢量. $\boldsymbol{u} = (a_x,\ a_y,\ a_z)^{\mathrm{T}}$ 为推力加速度, $\boldsymbol{A}$, $\boldsymbol{B}$ 分别为系数矩阵, 将此式展开, 可得

$$\begin{cases} \ddot{x} - 2\beta\dot{y} - 3\beta^2 x = a_x \\ \ddot{y} + 2\beta\dot{x} = a_y \\ \ddot{z} + \beta^2 z = a_z \end{cases} \tag{7}$$

其中, $\beta = \sqrt{G/r^3}$, G 为引力常数, r 为轨道半径. 式 (7) 即航天领域著名的 C-W 方程[48]. 研究者们探讨了一系列基于航天器模型的追逃博弈. Zhang 等[49] 利用神经网络研究了航天器模型下的自由轨道 (Free-time orbit) 追逃博弈, 给出了一个最佳引导策略. Venigalla 和 Scheeres[50] 采用可达集 (Reachable-sets) 的方法分析了航天飞船的追逃博弈. Li 等[51] 分析了近似圆形轨道上的两航天飞船追逃博弈, 将其中的 24 维两点边界值问题降维成一个 4 维非线性方程组, 并提出了混合数值算法来求其鞍点解.

除了以上几种模型, 还有基于生物启发的动态模型[5]、基于无人艇的动态模型[52] 等. 随着追逃博弈不断向实际应用靠近, 动态模型也在不断改进, 这

些异构的动态模型为追逃博弈的研究提供了丰富的理论框架, 对应更多的现实场景, 以求更好地适应当下研究需要.

3 追逃博弈的个体数量

确定了个体的动态模型以后, 还需要明确追逃博弈中个体的数量. 在追逃博弈的研究中, 参与者个数对于求解方法的选择及问题复杂度具有显著影响, 参与者越多, 问题复杂度越高. 具体而言, 我们将聚焦于一对一、多对一和多对多追逃博弈三种情境.

3.1 一对一追逃博弈

追逃博弈的早期研究大都设定为单追逐者–单逃跑者. Isaacs、Ho 等研究的杀人司机问题、两车问题就是一追一逃问题[53], 经典的“狮子和人”问题也是设定在一追一逃情形下的[21]. 一追一逃的策略可以拓展到多个个体的追逃博弈中, 而多追多逃有时也可化简为若干个一追一逃来考虑. 一追一逃问题常常建模为一个目标函数取最大最小值的问题, 追逐者旨在找到最优策略使目标函数取最大值 (最小值), 以最大程度地接近其目标状态; 而逃跑者则力求通过反制策略使该函数取最小值 (最大值), 以最有效地远离不利局面[54], 这里由于值函数的设置不同, 双方对最值的追求可能正好相反.

在一追一逃这一基本框架下, 研究者们从追逃双方自身特性出发开展了研究. 有的研究者从追逃双方速度大小关系入手, 一些令追逐者与逃跑者速度一致[55], 一些令追逐者更快[6, 56], 还有一些令逃跑者更快[57]. 此外, 一追一逃的简洁性使得智能体能更灵活地选择动态模型. 例如 Isaacs[10] 把车建模为 Dubins 模型, 把人建模为一阶积分器模型. 显然行人灵活度更高, 但车辆速度更快, 这符合实际. 受此启发, Exarchos 和 Tsiotras[57] 考虑了追逐者更灵活、逃跑者更快的情况. 此外, 研究者们也对一追一逃在各种复杂环境约束下的情境展开了探究. 譬如, Oyler 等[56] 研究了存在障碍物的二维平面上的一追一逃博弈; Das 等[58] 把一攻一防一逃的 TAD 问题加以改造, 将防守者限制在一条线段上运动; Liang 等[59] 则考虑了将防守者限制在一定区域内的 TAD 问题.

3.2 多对一追逃博弈

多对一追逃博弈是指多个追逐者与一个逃跑者或多个逃跑者与单一追逐者的追逃博弈. 与一对一追逃相比, 多对一追逃博弈需要考虑多个智能体之间的合作, 分别对每一个智能体的策略进行设计.

当追逐者有多个时, 最简单的情形是两追一逃问题[60]. 而对于 N 追 1 逃问题 ($N \geq 2$), 由于追逐者数量上占据绝对优势, 为平衡局势, 常设定逃跑者速度更快或更加灵活, 或是限定追逐者的运动范围. 例如, Chen 等[61] 设定了一群追逐者抓一个高速逃跑者的问题, 从追逐者的数量和初始分布位置这两个角度探讨抓捕成功的必要条件. 另一方面, 逃跑者在面对多个追逐者的围堵时, 如何找到一套反制策略, 也是目前的一个研究方向. Ramana 和 Kothari[28] 洞察到, 即使在面对看似完全封闭的围捕队形时, 逃跑者仍能通过改进策略诱导围捕者的行动, 在敌方防线中制造出漏洞, 开辟逃生通道.

当存在多个逃跑者时, 一追两逃是最基本的情况. 由于追逐者只有一个, 两逃跑者相互配合可以显著地提高博弈优势; 若两逃跑者各自为政, 则很难实现利益最大化. Yan 等[62] 就一追两逃的情况给出了两逃跑者合作的策略: 一个逃跑者吸引追逐者远离目标区域, 另一个逃跑者向目标区域前进, 从而最大程度接近目标区域甚至是进入目标区域. 在这里, 逃跑者们可以通过牺牲其中某个个体, 来确保另一个逃跑者能成功逃脱或是延长其生存时间, 实现群体目标. 这种合作的策略同样适用于逃跑者数量更多的一追多逃问题. 例如, Liu 等[63] 考虑了一个更快的追逐者抓捕一队逃跑者的问题, 逃跑者互相合作, 向多个方向分散逃跑, 延迟最终抓捕时间, 将整体利益最大化. 此外, 在一追多逃问题中, 抓捕顺序也是值得考虑的因素. 追逐者试图找到一个最优的抓捕序列使总抓捕时间最小[64].

多对一的追逃博弈还可用于研究某些动物集群行为. 例如, Scott 和 Leonard[65] 通过研究单个一阶积分器模型的追逐者与多个独轮车模型的逃跑者间的博弈发现, 生物形成集群的核心动机在于通过集体行动来有效降低个体遭受捕获的风险, 这为理解生物集群行为提供了新的理论依据.

3.3 多对多追逃博弈

多对多追逃博弈是指追逐者和逃跑者的数量均大于等于 2 的追逃博弈. 追逃双方各自形成一支团队, 双方的策略设计更加灵活, 预期目标也更多样, 问题复杂度也随之提升. 通常先分配双方的目标任务, 然后再考虑追逃双方之间的博弈[66–67]. 这样可将问题中具有相同性质的智能体予以整合, 使之化简为少数关键主体之间的博弈, 从而降低问题复杂度.

多对多追逃博弈受到了学术界的关注. Wei 等[68] 提出一个“去中心化”抓捕方案, 该方案以贪心策略为基础, 针对多追逐者追捕多个高速逃跑者的情形, 探讨了追逐者如何有效地协同合作, 确保至少捕获

一名逃跑者. Xu 等[69] 考虑了以下问题: 对于一组已经训练好的追逐者, 当其中某几个出现故障时, 如何调整追捕策略、保证对逃跑者们的抓捕.

在多对多的追逃博弈里, 还需要考虑目标任务的分配. Wei 等[67] 假设追逐者之间信息共享, 采用相同目标分配算法进行计算, 初步实现了分布式的目标分配. Li 等[70−71] 则考虑了用任务完成时间作为追逐者的目标, 并采用集中式规划来求解, 发表了一系列文章. 随着障碍物、噪声、复杂动态模型等引入追逃问题的研究, 分布式目标分配难以实现: 若每个个体都有独立的目标, 则问题复杂度过高难以求解; 而相同目标分配在实际应用中又是不现实的. 因此, 目前大都采用集中式分配方法以降低目标分配的难度[23, 72].

追逐者和逃跑者不同的数量组合提高了追逃博弈的普适性. 不论是一对一追逃博弈展现出的两个个体间策略的对抗, 还是多对一追逃博弈中团队的相互配合, 亦或是多对多追逃问题里两个团体的相互较量, 这些都可以与现实生活中的某些场景一一对应, 进一步提升追逃博弈的应用广度.

4 追逃博弈的空间环境

探讨完智能体本身的动态模型、数量, 还需要从外部空间环境来分类讨论追逃博弈. 我们日常处理的问题, 大多是二维和三维的, 因此可以考虑智能体在二维平面和三维空间中的追逃博弈.

4.1 二维平面的追逃博弈

二维平面的追逃博弈主要关注平面内追逃双方的运动轨迹和策略选择. 二维平面中计算复杂度较低、轨迹较为简单, 平面几何的成熟理论也为追逃博弈提供了理论基础, 因此早期的追逃研究都设定在二维平面上[10, 12].

此外, 地形、障碍物等因素也可能对平面追逃博弈产生影响, 尽管这些因素增加了问题的复杂性, 但也使得追逃问题的建模更贴近现实条件. 目前, 平面追逃博弈中障碍物的研究主要聚焦于线段 (或简单曲线) 和凸多边形. 这些障碍物位于二维平面的特定位置, 不仅能阻挡视线, 还能阻止智能体通过[73−75]. 微分对策解法往往难以判断障碍物对轨迹和策略的影响, 因此有学者探索了其他方法. 例如, Oyler 等[22] 针对线段形障碍物做了研究, 通过对比有障碍物与没有障碍物时等时线簇形状的变化, 确定追逐者和逃跑者各自优势区域的分界线, 进而给出追逐者的最佳方案. 在现实中, 相比于线段形的障碍物, 更常见的是那些形状不规则、边界复杂的障碍物, 通常将其近似为多边形障碍物来分析. 多边形障碍物的边界由若干线段组成, 可以考虑用解决线段形障碍物的方法来处理多边形障碍物. 此外, 一些学者还加入了圆形障碍物, 基于几何方法和强化学习等手段对求解算法进行了探讨[76−78].

也可聚焦于二维平面上某一特定区域, 把追逃双方的活动范围限定在一个圆[79]、一个角落[80] 或一个凸多边形[81] 区域内. 早在上世纪三十年代, Rado 就提出了圆形区域内的“狮子和人”追逃博弈[82]. Yan 等[83] 提出圆形区域防御问题并给出了界栅上的几何解. 在圆形区域基础上, Ruiz 和 Isler[84] 将追逃博弈定义在一个凸多边形区域内, 给追逐者加上转向限制, 在此条件下给出了追逐者策略, 证明了该策略在某时间上界内一定可以抓捕成功. 为研究有界区域内的追逃博弈, Okabe 等[85] 引入了 Voronoi 分块, 这些分块在研究多对多追逃博弈时展现了强大的效用[27].

4.2 三维空间的追逃博弈

随着无人机集群技术的广泛应用, 二维追逃博弈研究成果已无法满足当前的应用场景. 这一局限性促使有必要寻求更加先进的解决方案. 在这样的背景下, 更好地适配无人机运动特性的三维空间追逃博弈应运而生. 但相比于平面上的追逃博弈, 由于维度的提升, 对三维追逃博弈的建模、值函数的推导以及 HJI 方程的计算复杂度也会提高. 直观上讲, 在三维空间中, 追逃双方可以在立体空间内进行自由运动, 这使得运动轨迹和策略选择更加多样. 追逃双方不仅需要关注对方的水平位置, 还需考虑高度、深度等因素. 因此, 在三维空间中, 智能体需要具备更强的空间感知能力和策略制定能力. 此外, 三维空间中的障碍物、地形等也可能使追逃博弈变得更加复杂. 针对三维追逃博弈中的难题, 主要形成了以下几种解决思路.

其一是借助二维追逃的现有成果来研究三维. 在一定条件下, 许多二维平面中的策略和方法都可以类比到三维空间中. 例如, Pierson 等[86] 结合前人在二维有界凸区域上的研究, 先考虑二维平面中的抓捕策略, 提出一个追逐者策略算法, 随后将该算法代入三维空间中, 通过将仿真模拟结果与其他方法对比来检验策略在抓捕过程中的最优性. Li 等[87] 把二维上的阿氏圆扩展到三维, 用于解决三维空间中的一追一逃问题. 进一步地, Zhang 等[88] 也研究了两个三维空间中的追逐者追捕一个二维平面上的逃跑者的博弈问题. 此外, Zhi 等[89] 在二维视野算法的基础上提出了在三维空间中始终保持可以观测到

目标的最大化视野算法. 文中将三维空间下的算法时间复杂度降低了两个数量级, 提高了算法效率. 这对于三维空间的处理是十分重要的.

其二是从实际应用场景出发, 直接聚焦于三维空间的研究. 例如, Garcia 等[23] 针对三维空间中的自主飞行器追逃博弈展开了研究. 他们将智能体都建模为一阶模型. 首先探讨了三维空间中单追逐者与单逃跑者之间的策略互动. 随后, 将这一成果应用于两追逐者−单逃跑者的情况. 这些场景对应着许多现实中的情形, 如两战机合作追逐空中目标物, 防空导弹拦截敌机等. Chen 等[90] 将三维空间中的智能体建模为二阶积分器模型, 将追逐者设置为三自由度, 逃跑者是自由移动的. 利用 HJI 方程探讨了此情况下的均衡策略, 这说明了追逃解法在三维空间中应用的可行性. 随着航空航天技术的发展, 轨道追逃问题近些年也成为追逃博弈领域研究的热门方向. Shen 和 Casalino[91] 研究了三维轨道上的航天器追逃问题. 这里用到的模型也不同于原来的传统一阶、二阶模型, 要加入离地高度、推力和重力加速度等参数. 这一工作可被用于近地轨道航天器的对接、卫星跟踪等方向上. 另外, Manoharan 等[46] 也着眼于现实应用, 研究了三维空间中的无人机集群防御问题. 为了更好地模拟实际情况, 得到更具实用性的追逃策略, 他们按照实际的无人机参数设定和现实环境搭建仿真环境, 对无人机群进行训练.

此外, 也有学者对复杂的三维空间追逃博弈做了拆分, 化简为若干简单问题分别讨论. Yan 等[92] 设定了多追逐者−多逃跑者的三维场景, 并巧妙地将此问题分解为多个多追逐者−单逃跑者的子问题, 针对这些子问题, 他们探讨了追逐者队伍取得胜利的条件, 并提出了一套确保追逐者胜利的策略. 更进一步地, 他们还将研究结果推广至一个有界的凸三维区域内, 使得其更具实际应用价值.

二维和三维上的追逃是追逃博弈最基本的分类之一. 通过以上讨论不难发现, 目前二维平面上的追逃研究较为成熟, 各种场景的设定比较全面; 而三维上的讨论相对较少, 其科研潜力巨大. 所以越来越多的国内外团队正将目光聚焦于三维空间的追逃博弈研究, 这预示着追逃领域未来的重要发展方向.

5 追逃博弈的信息获取

除了智能体所处的空间环境, 信息的获取也是博弈需要考虑的因素. 在实际场景中, 全局信息的获取是一项艰巨的任务. 直观来讲, 人类的视距具有局限性, 而战机的雷达系统同样受限于探测范围. 根据信息的完整性, 追逃博弈可被分为完全信息追逃博弈和不完全信息追逃博弈. 在完全信息情境中, 追逃双方均能够全面掌握对方的位置、速度等关键信息, 这种信息的对称性有利于简化问题. 然而, 在不完全信息追逃中至少有一方对另一方的信息存在不确定性, 这种信息的不对称性对决策者提出了更高的要求. 前文所介绍的有关追逃的研究大都设定为完全信息. 因此, 本节将重点介绍不完全信息追逃博弈.

5.1 有限探测下的不完全信息追逃博弈

有限探测下的不完全信息追逃博弈指的是智能体只能接收以自身为中心的一定半径范围内的信息, 包括其他智能体的位置、速度等. 此类问题始于上世纪六十年代, Dobbie[17] 提出了“监控逃跑问题” (Surveillance-evasion problem), 引入了一个假设: 追逐者装备了类似雷达的探测装置, 仅能获取其探测范围内的信息. 当逃跑者处于此范围内时, 追逐者能实时掌握其信息; 而一旦逃跑者逃至探测范围外, 则追逐者完全无法获取关于逃跑者的任何信息[17, 93]. “监控逃跑问题”首次使得追逐者不再拥有全局信息的掌控权, 这也奠定了不完全信息追逃博弈研究的基础.

许多研究者都考虑了有限探测范围条件下的追逃博弈. Greenfeld[94] 设定两车博弈中每个车辆只能获取有限范围内的信息, 研究了探测范围与转弯半径的比值对追逃博弈最优解的影响. Coon 和 Panagou[36] 将探测范围引入到 TAD 问题中. Bopardikar 等[95] 针对探测范围有限的追逐者, 提出一种遍历游戏区域寻找逃跑者的算法, 在探测到逃跑者后, 可以使逃跑者一直在探测范围内. 这一算法有效地解决了追逐者获取信息范围有限的问题, 从而将问题转化为完全信息的情况. Lopez 等[96] 提出将探测范围加入到通信图上的追逃博弈研究中, 他们基于通信图的空间离散化特性和图论相关知识, 提出有限探测范围下追逃博弈问题的解决方案.

自然界中生物的视力、听觉范围都是有限的, 因而获取的信息也是有限的. 在这些制约下, 捕食者依然能成功地抓到猎物, 其策略对于有限探测下的追逃博弈研究具有一定的启发. Wang 等[5] 从生物学的视角研究了多个追逐者合作围捕一个逃跑者的问题. 他们给追逃双方加入了时间开销、通信限制、能量开销等代价函数, 用来解决多追逐者合作以及追逃之间的博弈问题.

5.2 不确定因素下的不完全信息追逃博弈

除了监控逃跑问题, 还有一些学者构建了其他不完全信息追逃博弈的模型, 令追逃博弈中的某一

方或是双方的位置、速度等信息具有不确定性, 称之为不确定因素下的不完全信息追逃博弈. 例如, Zemskov 和 Pashkow[97] 构造了一个一追两逃问题, 两个逃跑者中的任意一个均有可能在任意时刻“消失”. 一旦逃跑者消失, 追逐者将无法再获取其任何信息. 这个问题无法直接给出一个确定性的结论, 需要借助一些概率论的相关知识来计算一个平均收益最高的解. 这一工作进一步丰富了不完全信息追逃博弈的定义.

沿此思路, Wei 等[68] 在考虑环境中的噪声对追逃双方接收信息的影响时, 通过分析其概率密度函数, 设计了一种“粒子过滤器”来处理非高斯分布的复杂噪声. 紧接着, Li 等[71] 讨论了双方信息不对称的情况: 追逐者获取信息时受到随机噪声干扰; 而逃跑者不受噪声影响. 这会导致逃跑者占优, 因而追逐者需要更为高效的策略. 还可以从逃跑者信息受限的情况入手, 例如, Liu 等[98] 构造了一个具有“隐藏追逐者”的一追多问题, 追逐者的实时位置按照某概率模型分布, 逃跑者通过提高总抓捕时间的期望 (加权平均值) 来寻找最优逃跑策略. 这些工作成功地把概率论融合进了追逃问题的研究. 概率论在分析追逃博弈应用场景中的不确定因素时, 展现出了强大的作用. 利用概率论的知识求解也是追逃博弈的发展方向之一.

5.3 不完全信息追逃博弈的现实意义

随着研究的不断深入, 不完全信息追逃博弈的研究也逐渐向实际应用靠拢. 譬如当下比较热门的无人机集群围捕地面目标问题: 无人机处于三维空间, 地面目标处于二维平面. 追逃双方由于视野、障碍物等限制, 其获取的信息是不对称的, 显然地面目标是处于劣势的, 如何才能使其达到最大效益是一个值得思考的问题. von Moll 等[99] 对此进行了研究. 除了空对地, 不完全信息追逃博弈也可以应用到海上. 由于海上信号覆盖率低、天气多变等因素, 水上舰艇的信息获取常常出现延迟甚至丢失的情况. 所以在不完全信息下, 如何实现对水上的追逃是一个具有现实意义的课题. Lin 等[52] 研究了水面无人艇 (USV) 的 TAD 问题, 令双方都不知道对方策略, 在此条件下利用非线性模型预测控制 (NMPC) 来探讨双方如何以最小代价实现各自目标.

在各个领域, 完整信息的获取都是一个难题, 因而在不完全信息下考虑追逃博弈具有深远的现实意义. 通过以上列举的文献不难发现, 不完全信息问题的处理思路有两种: 一种是直接求解; 另一种是将不完全信息约化为完全信息的情况. 总之, 无论是基于当前的研究成果, 还是针对实际应用的探索, 或是受自然界启发, 不完全信息追逃博弈的研究都在逐步深化.

6 追逃博弈的状态空间

除上述几个方面之外, 智能体所处的状态空间也是追逃博弈中的关键要素. 根据状态空间的设定, 可以把追逃博弈分为离散的和连续的. 离散设定下的追逃博弈通常涉及在离散的时间空间内的追逃行为, 而连续设定则允许追逃双方在连续的时空中运动. Isaacs[10] 在研究追逃问题时主要从 HJI 方程、几何法等数学角度求解, 考虑的大多是连续设定的追逃博弈. 直到今天, 大多数的追逃博弈研究都是设定在连续情况下的[39, 42, 100]. 连续的设定有助于建立数学模型、求解 HJI 方程等. 但随着追逃博弈的设定不断复杂化, 纯数学方法很多时候无法求解, 需要借助计算机仿真模拟或是强化学习等方法. 但根据计算机的设计原理, 即便追逃博弈设定为连续状态, 实际在计算机中的处理都是离散化的. 接下来将着重讨论离散状态空间下的追逃博弈, 按照离散时间、离散空间的顺序分别介绍.

6.1 离散时间–连续空间的追逃博弈

离散时间的追逃是最常见的离散设定. 时间的离散化能使智能体的运动轨迹分段化, 便于计算机仿真模拟的处理; 离散的时间也对应着离散时间马尔科夫过程, 可以与概率论、随机过程等理论结合.

离散时间追逃博弈里有一个经典的命题, “狮子和人”问题. “狮子和人”问题最早由 Rado 提出, 随后 Littlewood[101] 对此进行了整理和补充. Gale 率先提出离散时间设定下的“狮子和人”问题[19]: 狮子和人都处于二维非负半平面上, 狮子和人按回合制移动, 即每回合分别移动一次 (人先移动), 各自到达下一个位置, 待下个回合再重复此过程. 狮子和人每次移动的最大距离相同, 狮子的目标同追逐者, 人被视为逃跑者. Gale 的这一工作开启了“狮子和人”问题在离散时间设定下的新篇章.

针对 Gale 的问题, 许多学者考虑了狮子的抓捕策略[95, 102]. 由于追逃双方只在固定时间点依次移动, 因而抓捕策略的重点在于如何由人的当前位置确定狮子的下一轮最优位置. Sgall[103] 提出“固定圆心”策略, 使得每轮移动过后, 人、狮子、圆心三点在一条直线上, 从而把人限制在一个原点和坐标轴围成的角落里, 逐步压缩人的安全区域. 但是在离散时间下, 人在每轮移动中具有优先权, 因此该解法存在一个较强的初始条件. 所以有研究者继续了 Sgall[103]

的工作. Casini 和 Garulli[80] 优化了 Sgall[103] 的“固定圆心”策略, 给出了“升级圆心”策略, 即在双方每轮移动以后, 重新计算一个圆心. 按此策略, Casini 和 Garulli[104−105] 给出了数值解, 证明了该策略优于“固定圆心”策略. Casini 等[106] 也对区域进行了限制, 在一个有界凸多边形区域里研究追逐者策略, 证明了在此区域内“固定圆心”策略无需满足 Sgall[103] 设定的初始位置条件.

除了狮子和人问题, 还有将连续时间追逃博弈改造为离散时间的. Bopardikar 等[95, 107] 假设追逐者只在固定时间点移动, 每次移动距离不超过其探测半径, 把有限探测范围下的不完全信息模型同离散时间设定相结合, 分析了两种情况: 一是平面有界凸区域内的一追一逃问题[95]; 二是二维平面上的多追一逃问题[107]. 紧接着, Bopardikar 等[95] 提出“sweep-capture-strategy”算法, 该算法的基本思路是对区域进行折返遍历, 直至探测到逃跑者, 进而把问题简化为无探测范围的完全信息情况. 而且, 还证明了在探测范围与移动步长满足一定条件时, 该算法可以保证抓捕成功率.

6.2　离散空间−连续时间的追逃博弈

除了离散时间的设定外, 追逃双方所处的空间环境同样可以是离散的. 常见的离散空间形式主要包括树状图以及有向无环图 (图 3) 等. 图类空间是一种抽象的空间形式, 通过定义图的顶点集和边集来规定追逐者和逃跑者的活动范围. 追逃双方只能沿着图的边在顶点之间进行移动. 图的构造十分灵活. 一方面, 可以根据实际需求组合顶点和边, 应用到更多场景中; 另一方面, 图的理论较为成熟, 可以使用很多现成的定理、结论来处理图上的追逃博弈. 一些科研工作者[96, 108] 基于图理论讨论了多个追逐者间的合作方案, 提出了可行的抓捕策略. 而且, 可以对点与点之间边的权重进行定义. 以城市间道路上的追逃为例, 可以将城市抽象为节点, 道路抽象为边, 而城市间的实际距离则对应为边的权重[109−110].

目前离散空间的追逃博弈研究, 大都建立在各种类型的图上. 例如, Sundaram 等[111] 研究了有向无环图上的部分信息追逃博弈. 他们构造了一个类似于图 3 的有向无环图, 图上有一个起点和若干终点. 追逃双方先后从起点出发 (追逐者速度更快, 所以让逃跑者先出发), 追逐者要在逃跑者到达终点前捕获之. 类似于自然界中捕食者按气味追踪猎物, 他们在每个节点上安装传感器, 追逐者通过传感器分析逃跑者是否经过该节点. 据此信息, 追逐者根据文中的算法作出决策, 移动到下一个位置. 此外,

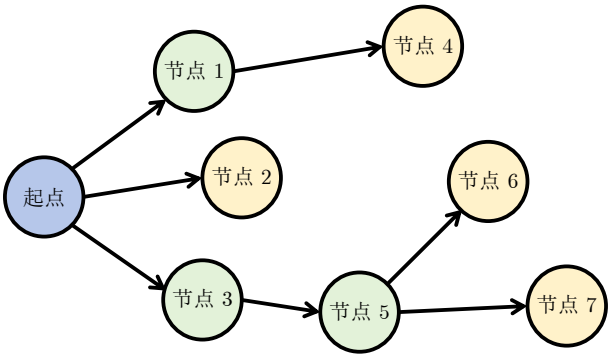

图 3　有向无环图

Fig. 3　Directed acyclic graph

也可以借助强化学习中的 Q-learning 方法来求解图上的追逃博弈. 例如, Dong 等[112] 利用 Q-learning 和图论的方法不断升级迭代策略来求解 Hamilton-Jacobi-Bellman 方程, 从而寻找多追多逃问题里每个智能体的最优策略.

6.3　离散时间−离散空间的追逃博弈

除了单纯离散时间、单纯离散空间的追逃博弈外, 也有时间、空间均为离散的追逃博弈. 这类追逃博弈主要设定在一种类似于棋盘的网格化空间上. 其特点在于追逃双方只能在空间中固定的坐标点上进行活动, 并且位置的移动也只能限制在特定的时间点进行.

此类问题也被称为“警察与强盗”问题[113−114], 以其独特的设定吸引了众多研究者的目光. 譬如, Bhattacharya 等[115] 考虑了二维网格上的追逃博弈, 设定追逃双方只能在网格顶点之间按照离散时间移动, 讨论了此模型在多智能体系统中的应用. 另外, Bhattacharya 等[116] 也构造了一个 N 维网格中的警察与强盗问题. 警察与强盗只能在顶点上移动, 保证空间离散性; 而且双方只在特定时间点按步移动, 确保时间离散性. Bhattacharya 等[116] 分析了此情况下至少需要几个警察才能确保抓到该强盗. 进一步地, 也探讨了二维网格下, 只有一个警察时如何选取策略保证抓捕成功.

同样是在上述设定下, Das 和 Gahlawat[117] 加入了强盗攻击警察、懒惰警察等假设, 更深入地研究了网格图上的路径规划问题.

离散状态空间的追逃博弈是伴随着计算机技术兴起的, 与计算机更好的适配性使得离散状态空间追逃博弈受到了广泛关注, 展现出广阔的应用前景.

7　追逃博弈的求解方法

在追逃博弈几十年的研究历史中, 涌现出了复

杂多变的问题设定, 随之也形成了一系列求解方法, 主要包括理论求解法、数值求解法、积分强化学习法以及几何法等. 接下来, 将逐一探讨这四种求解方法.

7.1 理论求解法

理论求解法是追逃研究中最早出现的方法之一, 由 Isaacs[10] 提出. 理论求解法主要利用微分对策理论求解追逃博弈, 被广泛应用到各类追逃博弈当中, 包括前面提到的杀人司机问题[41]、自杀行人问题[43]、Dubins 模型的追逃问题[40]、各向同性火箭问题[118]、多独轮车追逃博弈[119]、多航天器追逃博弈[120]等. 理论求解逻辑严密、精确度高, 是其他解法的理论依据和参考标准, 下面将详细介绍.

7.1.1 微分对策与追逃博弈

理论求解主要基于微分对策理论. 微分对策通常涉及多个由微分方程控制的系统之间的博弈, 对策双方都欲使目标函数最优, 一方使最大, 一方使最小. 这与追逃双方为达到各自目标而发生的博弈过程紧密相关, 可将追逃博弈建模为微分对策问题. 为了简化问题, 以一追一逃为例, 分别用 $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ 来表示追逃双方的策略. 则状态方程为

$$\dot{\boldsymbol{x}} = f(t,\ \boldsymbol{x}(t),\ \boldsymbol{u}_1(t),\ \boldsymbol{u}_2(t)) \tag{8}$$

目标函数为

$$J = \varphi(t_f,\ \boldsymbol{x}(t_f)) + \int_{t_0}^{t_f} L(t,\ \boldsymbol{x}(t),\ \boldsymbol{u}_1(t),\ \boldsymbol{u}_2(t))\mathrm{d}t \tag{9}$$

其中, φ, L 为给定的函数, 分别表示终端和博弈过程中的收益或损失. 结束时间 t_f 和目标集 $\mathcal{E}$ 为

$$\begin{cases} t_f = \inf\{t \in \mathbf{R}^+ : \boldsymbol{x}(t) \in \mathcal{E}\} \\ \mathcal{E} = \{\boldsymbol{x} | E(t,\ \boldsymbol{x}) = 0\} \end{cases} \tag{10}$$

其中, inf 表示集合的下确界, E 为终端条件. 目前大多研究通常以结束时间作为目标[36, 39, 63], 考虑时间的最优性, 此时目标函数通常表示为 $J = \int_0^{t_f} \mathrm{d}t = t_f$.

在追逃博弈里, 追逐者和逃跑者分别寻求最优策略以期望目标函数 J 取得最大值或最小值. 这涉及到寻求纳什均衡解. 所谓纳什均衡, 是指在包含两个或以上参与者的非合作博弈 (Non-cooperative game) 中, 假设在每个参与者都知道其他参与者的均衡策略的情况下, 没有参与者可以通过改变自身策略使自身受益时的一个概念解[121]. 假设 $\boldsymbol{u}_1$ 欲使目标函数 J 取最大, $\boldsymbol{u}_2$ 意图使 J 取最小, 双方的纳什均衡解 $\boldsymbol{u}^* = (\boldsymbol{u}_1^*,\ \boldsymbol{u}_2^*)$ 应满足

$$J(\boldsymbol{u}_1,\ \boldsymbol{u}_2^*) \le J(\boldsymbol{u}_1^*,\ \boldsymbol{u}_2^*) \le J(\boldsymbol{u}_1^*,\ \boldsymbol{u}_2) \tag{11}$$

此纳什均衡策略涉及到最大最小值, 故也称之为鞍点策略. 在鞍点策略下, 目标函数的值称为追逃博弈的值函数

$$V(t,\ \boldsymbol{x}) = \max_{\boldsymbol{u}_1(t)} \min_{\boldsymbol{u}_2(t)} \{\varphi(t_f,\ \boldsymbol{x}(t_f)) + \int_{t_0}^{t_f} L(t,\ \boldsymbol{x}(t),\ \boldsymbol{u}_1(t),\ \boldsymbol{u}_2(t))\mathrm{d}t\} \tag{12}$$

这里由于目标函数的设置不同, 追逐者和逃跑者对最大值、最小值的追求不是固定的, 追逐者可能对应 $\boldsymbol{u}_1$, 逃跑者对应 $\boldsymbol{u}_2$; 或者正好相反.

若值函数 $V(t,\ \boldsymbol{x})$ 存在且关于 t, $\boldsymbol{x}$ 连续可微, 则其一定满足 HJI 方程

$$-\frac{\partial V}{\partial t} = \max_{\boldsymbol{u}_1} \min_{\boldsymbol{u}_2} \left\{ \frac{\partial V^{\mathrm{T}}}{\partial \boldsymbol{x}} f(t,\ \boldsymbol{x},\ \boldsymbol{u}_1,\ \boldsymbol{u}_2) + L(t,\ \boldsymbol{x},\ \boldsymbol{u}_1,\ \boldsymbol{u}_2) \right\} \tag{13}$$

则可以利用 HJI 方程来检验最优性: 首先构造值函数, 代入 HJI 方程计算, 若方程成立, 则此函数为值函数, 对应策略为鞍点策略. 这一过程被总结为如下定理[122]:

定理 1. 对于一个连续可微函数 $V(t,\ \boldsymbol{x})$, 如果有

1) $V(t,\ \boldsymbol{x})$ 满足 HJI 方程 (13);

2) 在目标集边界上, 即 $E(t,\ \boldsymbol{x}) = 0$ 时, 满足 $V(t_f,\ \boldsymbol{x}) = \varphi(t_f,\ \boldsymbol{x})$;

3) 满足 1) 中 HJI 方程的策略 $\boldsymbol{u}_1^*$, $\boldsymbol{u}_2^*$ 可在有限时间内到达目标集;

则上述函数 $V(t,\ \boldsymbol{x})$ 为值函数, 策略 $\boldsymbol{u}_1^*$, $\boldsymbol{u}_2^*$ 为最优策略, 构成鞍点解.

除此之外, Krasovskii 等[123] 提出了极限瞄准 (Extremal aiming) 方法. 该方法也被应用到多人博弈问题中[124−125]. 针对 (12) 中构造的值函数, Pontryagin[16] 提出了最大最小值原理.

定理 2. 策略 $(\boldsymbol{u}_1^*,\ \boldsymbol{u}_2^*)$ 为最优策略的必要条件是

$$\max_{\boldsymbol{u}_1} \min_{\boldsymbol{u}_2} H(\boldsymbol{x},\ \lambda,\ \boldsymbol{u}_1,\ \boldsymbol{u}_2,\ t) = \min_{\boldsymbol{u}_2} \max_{\boldsymbol{u}_1} H(\boldsymbol{x},\ \lambda,\ \boldsymbol{u}_1,\ \boldsymbol{u}_2,\ t) = H(\boldsymbol{x},\ \lambda,\ \boldsymbol{u}_1^*,\ \boldsymbol{u}_2^*,\ t) \tag{14}$$

其中, H 为 Hamilton 量, λ 为参数. 此定理又称双方极值原理, 常用于探讨追逃博弈中抓捕发生的条件[126−128].

7.1.2 微分对策空间划分

在微分对策理论的基础上, Isaacs[10] 给出了几

个关键的划分空间的概念, 如可用部分 (Usable part)、不可用部分 (Nonusable part)、界栅 (Barrier) 和散射曲面 (Dispersal surface) 等, 这些概念与上述理论一同组成了微分对策下的追逃理论体系. 《微分对策》[13] 中对这些概念做了总结和提炼, 下面将一一介绍.

多数情况下, 微分对策的轨迹终点仅能落在目标集内的某一特定区域, 称为目标集的可用部分. 对于 m 维系统, 其状态方程满足式 (8), 任取一点 $\boldsymbol{y} \in \mathcal{E}$, $\boldsymbol{y}$ 属于目标集的可用部分当且仅当

$$\max_{\boldsymbol{u}_1} \min_{\boldsymbol{u}_2} \sum_{i=1}^{m} n_i f_i(\boldsymbol{y},\ \boldsymbol{u}_1,\ \boldsymbol{u}_2) < 0 \tag{15}$$

其中, $\boldsymbol{n} = (n_1,\ n_2,\ \cdots,\ n_m)$ 为目标集边界的法向量. 这些点的集合构成了目标集的可用部分. 而与之对应的不可用部分, 正是目标集中除去可用部分的点的集合.

一追一逃问题中存在两个区域: 捕捉区和躲避区. 在捕捉区中, 对逃跑者的任意策略, 追逐者总能找到抓捕方案将其捕获; 而躲避区中只要逃跑者策略得当, 可保证成功逃脱. 这两个区域的分界面被称为界栅[13]. 问题的纳什均衡解, 往往出现在界栅上.

另一个重要的概念是散射曲面. 散射曲面一般是指两个正则最优策略区域的分界面[129]. 以图 4 为例能更形象地解释散射曲面的概念. 这里, 圆形障碍物会阻挡智能体通过, 追逐者 P、逃跑者 E 初始位置分别在该障碍物某条直径的两侧延长线上, 双方速度大小恒定, P, E 分别存在两条“同等好”的最优轨线 (图中轨线 1、2, 最优性证明见文献 [13]), 而这两条轨线都是从图中所示虚线上散射出去的, 对于其他初始位置, 也是如此. 则称该虚线为“散射曲面”. 散射曲面上的追逃双方策略选择存在“两难问题”, 这也会引出许多有趣的性质.

理论求解法也存在着不足, 由于现在追逃博弈已经拓展到 N 追 M 逃的情况, 而且还加上了障碍物、信息受限等条件, HJI 方程的求解成为一个难题[130−131]. 当下微分对策理论的局限性限制了理论求解法的发展, 需要数学上的进一步突破.

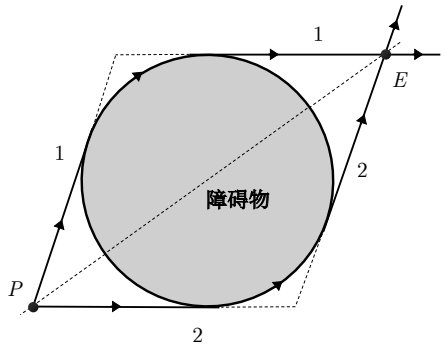

图 4　散射曲面示例

Fig. 4　Dispersal surface example

7.2　数值求解法

当下追逃博弈的研究中, 求解 HJI 方程是具有挑战性的. 理论求解法计算复杂度过高, 寻求一种创新解法尤为迫切. 在此背景下, 数值求解法顺势而生. 作为求解优化问题的常用手段, 动态规划方法在追逃问题上得到了广泛的使用[132−133], 也可以利用深度强化学习方法研究追逃博弈[134]. 下面将对这几种方法做详细介绍.

7.2.1　动态规划方法

动态规划 (Dynamic programming) 方法是一种广泛应用于数学、计算机科学等领域的方法. 它通过将复杂问题分解为一系列相对简单的子问题来求解原问题最优解. 在追逃博弈研究中, 常常利用动态规划法来求解 HJI 方程的近似解, 以及不断更新迭代得到局部最优策略.

若给定状态方程 (8) 和目标函数 (9), 对于 $t \in [t_0,\ t_f]$, 最优控制 $(\boldsymbol{u}_1^*,\ \boldsymbol{u}_2^*)$ 对应 J 的最优值 J^*, 可得

$$\begin{aligned} J^*(\boldsymbol{x}(t),\ t) = &\max_{\boldsymbol{u}_1} \min_{\boldsymbol{u}_2} \{\varphi(\boldsymbol{x}(t_f),\ t_f) + \\ &\int_t^{t_f} L(t,\ \boldsymbol{x}(t),\ \boldsymbol{u}_1(t),\ \boldsymbol{u}_2(t))\mathrm{d}t\} = \\ &\max_{\boldsymbol{u}_1} \min_{\boldsymbol{u}_2} \{\varphi(\boldsymbol{x}(t_f),\ t_f) + \\ &\int_t^{t+\Delta t} L(t,\ \boldsymbol{x}(t),\ \boldsymbol{u}_1(t),\ \boldsymbol{u}_2(t))\mathrm{d}t + \\ &\int_{t+\Delta t}^{t_f} L(t,\ \boldsymbol{x}(t),\ \boldsymbol{u}_1(t),\ \boldsymbol{u}_2(t))\mathrm{d}t\} = \\ &\max_{\boldsymbol{u}_1} \min_{\boldsymbol{u}_2} \left\{\int_t^{t+\Delta t} L(t,\ \boldsymbol{x}(t),\ \boldsymbol{u}_1(t),\ \boldsymbol{u}_2(t))\mathrm{d}t\right\} + \\ &J^*(\boldsymbol{x}(t+\Delta t),\ t+\Delta t) \end{aligned} \tag{16}$$

上述推导证明了 $J^*(\boldsymbol{x}(t),\ t)$ 与 $J^*(\boldsymbol{x}(t+\Delta t),\ t+\Delta t)$ 的关系, 将动态规划方法和微分对策联系到一起, 从而将动态规划与追逃博弈结合起来.

Isaacs[10] 最早提出使用动态规划的方法来解决追逃博弈. 随后, Kachroo 等[135] 给出了“牧羊犬赶羊群”的动态规划解. Hespanha 等[136] 证明了用动态规划方法能够求解随机移动的追逃博弈. 此外, Cristiani 和 Falcone[137] 也探讨了基于动态规划思想来解决一定限制下的追逃博弈的可行性. 动态规划问题在航空追逃领域也大显身手. Anderson[138] 最早通过动态规划法来分析两个航空器之间的追逃博弈. 但在他们的工作里, 动态规划和传统方法的优势对

比不够明显. 后来, Alexopoulos 等[139] 在利用动态规划方法解决无人机集群追逃博弈时, 在计算机中做了大量仿真, 更直观地展示了动态规划法的高效性.

总而言之, 动态规划方法为解决追逃博弈打开了新思路, 具有一定现实意义; 但目前的动态规划相关研究大多停留在控制层面, 应用到追逃博弈中的场景较为单一, 可以继续拓宽动态规划方法的应用范围, 例如, 在二阶积分器、凸多边形内或是存在障碍物的追逃博弈等.

7.2.2 强化学习方法

近些年强化学习发展迅速, 已成为研究追逃博弈的热门方法. 强化学习法的思路通常是训练追逐者形成一套高效的抓捕策略, 或通过神经网络迭代求 HJI 方程的近似解. 下面将介绍强化学习方法与追逃博弈的联系以及研究现状.

在追逃博弈的研究中, 常将追逐者和逃跑者的动力学模型、各自目标、抓捕条件等作为输入, 把追逃过程建模为马尔科夫决策过程, 利用强化学习算法不断更新求解最优策略. 理论求解法需要精确的模型且不能保证计算实时性, 在应用到像无人机集群对抗等实际场景中时效率较低. 而强化学习法实时性较高, 在应对不同场景时游刃有余, 因此研究者们对此开展了探索. 1993 年, Tan[18] 首次尝试了用强化学习方法来解决简单情形下的两追两逃问题. 近些年, 随着深度学习等技术的不断发展, 许多研究者提出了基于深度学习的强化学习方法[69, 140−141]. 相比于传统强化学习方法, 其普适性更强、效率更高, 可应用到各种设定上.

深度确定性策略梯度框架 (DDPG) 是深度强化学习中常用的结构, 一些研究者基于 DDPG 展开了研究. Xu 等[69] 采用 DDPG 和双向循环递归神经网络 (Bi-RNN) 来设计网络, 提出了可拓展的深度强化学习方法来解决多对多追逃问题. 相比于现有方法, 该策略下得到的结果更具稳定性, 而且可用于更大数量规模的情况, 有利于与大型无人机集群相关任务结合. Xing 和 Zeng[140] 利用基于 DDPG 的深度强化学习方法研究了狮子和人问题. 他们利用贪心策略对搜索策略做了提升, 并引入了学习重置机制来提高学习效率, 同时给出了该方法与经典 DDPG 方法的对比, 可见该方法更优. 进一步地, DDPG 算法与多智能体理论结合, 出现了 MADDPG 算法, 这一结构与追逃博弈的适配性更高, 同样有许多研究者对此开展了研究. Wan 等[142] 基于经典的 MADDPG 算法提出新的深度强化学习算法 A2-MADDPG, 用于处理多智能体追逃中的环境感知、决策等问题. 此外也有其他研究团队对 MADDPG 算法做出了改进, 提出了 CEL-MADDPG 算法[143]、FA-MADDPG 算法[144] 等. 这些不同的 MADDPG 算法适配于不同的追逃场景, 也展现出了 MADDPG 的普适性.

此外, 也有研究者使用基于其他框架的强化学习方法来解决追逃博弈问题. 前文提到, 在多对多追逃里, 任务分配是困难的. Asgharnia 等[134] 利用分层强化学习来解决这一困境, 分别生成了底层的追逃策略和上层的任务分配策略, 前者专注于单个个体的追逃行为; 后者负责在多个智能体之间协调分配任务. 这样, 不仅提高了任务分配的效率, 也有助于简化问题. de Souza 等[76] 使用基于课程学习的深度强化学习方法研究了圆形竞技场内的多追一问题, 以竞技场半径为自变量, 给出了该方法与其他三种方法的抓捕成功率对比情况, 证明了该方法效率更高.

强化学习法也并非尽善尽美. 强化学习实际上是一种试错型的方法, 通过大量的模拟试错和奖励诱导, 使智能体学到有效的策略. 但是, 强化学习方法的收敛性和有效性都难以保证, 这无疑会增加时间开销、浪费计算资源.

7.3 积分强化学习法

第 7.2 节已指出强化学习方法存在局限性. 为克服这些局限, 从理论上确保强化学习过程的收敛性, Vamvoudakis 和 Lewis[20] 提出了积分强化学习 (Integral reinforcement learning, IRL) 方法. 该方法用线性神经网络来拟合复杂的值函数和策略输入等关键参数. 这种处理方式使得原本非线性、动态变化的问题转化为网络参数的学习优化过程.

在实际操作中, 积分强化学习主要采取在线学习机制, 实时地根据系统状态与行为调整网络参数. 其核心在于设计一种与 HJI 方程误差关联的奖励信号. 当神经网络输出与方程期望解间的差距减小时, 给予正向的奖励反馈; 反之, 则施以惩罚. 随着学习过程的深入, 神经网络逐步收敛至某一稳定状态. 此时, 其对值函数及策略输入的模拟已经在预定的误差范围内, 保证结果高度逼近真实值. 下面将具体梳理积分强化学习的研究历程.

7.3.1 积分强化学习方法的基本思路

在 IRL 方法中, 首先要由追逃博弈的动力学模型、双方目标等构造问题的值函数 (这里主要参考文献 [100])

$$V^{\pi^p,\,\pi^e}(\delta)=\int_t^{\infty}(\delta^{\mathrm{T}}Q\delta+U(\pi^p(\delta))-U(\pi^e(\delta)))\mathrm{d}\tau \tag{17}$$

这里 Q, U 具体值都由问题设定决定. δ 表示

追逃之间的状态之差, π 表示策略. 然后将值函数 $V^{\pi^p,\,\pi^e}(\delta)$ 写成以下形式

$$V^{\pi^p,\,\pi^e}(\delta(t)) = \int_t^{t+T} (\delta^{\mathrm{T}}Q\delta + U(\pi^p(\delta)) - U(\pi^e(\delta)))\mathrm{d}\tau + V^{\pi^p,\,\pi^e}(\delta(t+T)) \tag{18}$$

其中, T 表示时间间隔. 按上述过程, 追逐者和逃跑者不断地给出反馈, 得到奖惩数据, 进而更新各自策略. 随着值函数迭代次数的增加, 可以证明博弈的目标函数有可能收敛到真实值函数 V^*, 追逃双方也可以达到纳什均衡策略 $(\pi^{p*},\ \pi^{e*})$ 附近[100, 145]. 这里在算法中逼近最优值函数时, 一般采用魏尔斯特拉斯 (Weierstrass) 逼近值来近似值函数

$$\begin{cases} \hat{V}(\delta) = \hat{\mathbf{W}}^{\mathrm{T}}\phi(\delta) \\ \nabla\hat{V}(\delta) = \nabla\phi(\delta)^{\mathrm{T}}\hat{\mathbf{W}} \end{cases} \tag{19}$$

其中, $\hat{\mathbf{W}}$ 为神经网络权重向量, 是策略更新的关键变量, 其他变量释义详见文献 [100]. 所以想求解 HJI 方程, 就要不断更新权重向量 $\hat{\mathbf{W}}$ 使误差尽可能的小, 从而使得值函数近似值 $\hat{V}(\delta)$ 近似为一定误差范围内的纳什均衡解, 达到预期目标. 目前大多数积分强化学习方法的相关研究也都是沿着这一思路展开的[146−147].

7.3.2　积分强化学习方法的发展历程和关键工作

积分强化学习起源于自适应动态规划方法. Vrabie[148] 根据追逃博弈的特点, 对传统的动态规划方法进行了改进, 提出了自适应动态规划 (Adaptive dynamic programming, ADP) 方法. ADP 方法普适性强、效率高, 被广泛用于处理各类追逃博弈的研究中, 如一追一逃[149]、多追一逃[150]、导弹防御问题[151] 以及卫星跟踪问题[152] 等.

受 ADP 方法的启发, Vamvoudakis 和 Lewis[20] 在研究两个智能体的非零和博弈过程中提出了积分强化学习方法, 紧接着又考虑了多个玩家的非零和博弈[153], 在这两篇文献中, 他们主要聚焦于像 Riccati 方程、Hamilton-Jacobi 方程求解等最优控制的内容, 这些基础控制理论是解决追逃博弈的重要手段. 许多研究者试图将积分强化学习方法直接应用到追逃博弈中. 例如, Kokolakis 等[145] 在研究多个无人机围捕一个高速目标时, 将无人机建模为 Dubins 模型, 显然在此情况下利用微分对策方法解决这类问题是不容易的, 因此他们引入了积分强化学习方法来计算双方的最优策略. 另外, 他们也证明了该策略能在有限时间内收敛到纳什均衡. 这一工作成功地将积分强化学习方法融入到了追逃博弈的讨论中, 证明了此方法的可行性.

此后有许多学者对积分强化学习方法做了进一步研究. 例如, Kartal 等[100] 将速度作为一个关键变量, 用积分强化学习方法求解 HJI 方程, 使之收敛到鞍点值的一定范围内, 得到双方的纳什均衡速度策略. Xiong 和 Zhang[146] 在考虑多个四旋翼无人机的追逃博弈时, 根据四旋翼无人机的自身特点、外部干扰构造了值函数, 然后利用强化学习方法来保证跟踪误差的稳定性, 最后利用积分强化学习方法求解 HJI 方程, 从而求解纳什均衡策略. 此外, Zhou 和 Xu[147] 提出一种 ACMO (Actor-Critic-Mass-Opponet) 方法来解决大规模数量的多对多追逃博弈, 其实质也是一种积分强化学习方法. 很多强化学习方法都是以 Actor-Critic 为框架搭建的. 譬如基于 Actor-Critic 模型提出一个可在线自适应同步积分的安全导引方案, 并将其应用到导弹防御的追逃博弈中[154].

从以上分析不难发现, 积分强化学习的特点在于: 一方面不同设定下的追逃博弈可以构造不同的值函数; 另一方面由于在更新策略和值函数时用到的一些细节技术 (如神经网络的选择、参数的调配等) 各不相同, 会导致收敛效果的优劣. 因此, 可以从以上两个方面继续积分强化学习的研究. 总之, 积分强化学习是一个年轻的方法, 有待继续探索.

7.4　几何法

几何法简洁直观, 在近几年的追逃博弈研究中崭露头角. 几何法应用广泛, 无论是不完全信息追逃博弈[155]、多对多追逃博弈[66], 还是三维空间中的追逃博弈[156], 都可以看到几何法的身影. 几何法的核心在于剖析各个区域的几何特征, 推导出抓捕点的可能位置以及最优策略. 下面将详细介绍.

"几何法"始于上世纪对"两车问题"的研究, 一些学者利用平面几何的相关知识, 给出了若干种"几何法"的解决方案[124, 157−158]. 接下来将逐一介绍等时线、优势区域、阿波罗尼斯圆及 Voronoi 多边形分割等几何法里的重要概念.

7.4.1　等时线

首先介绍等时线 (Isochrones). 等时线被广泛用于天文、道路交通规划等领域, 一般指的是连接同一时间发生的事件所在位置的曲线. 在追逃博弈中, 智能体的等时线定义为一组点的集合: 智能体在某一策略集合 π 中的所有策略下, 行进至这些点所需的时间完全一致. 在一阶积分器模型中, 追逃双方的最优策略都是以最大速度朝某一方向做匀速直线运动[10]. 因此对于初始位置分别为 $\boldsymbol{x}_p^0$, $\boldsymbol{x}_e^0$, 最大速度分别为 $\boldsymbol{v}_p$, $\boldsymbol{v}_e$ 的追逐者和逃跑者, 可分别定

义其时刻 t 的等时线为

$$\mathcal{I}_p = \{\boldsymbol{x}_p | \|\boldsymbol{x}_p - \boldsymbol{x}_p^0\| = \|\boldsymbol{v}_p\| t\} \tag{20}$$

$$\mathcal{I}_e = \{\boldsymbol{x}_e | \|\boldsymbol{x}_e - \boldsymbol{x}_e^0\| = \|\boldsymbol{v}_e\| t\} \tag{21}$$

由于速度恒定, $\|\boldsymbol{v}\|$ 为常数, t 给定, 则上述等时线为以智能体当前位置为圆心、$\|\boldsymbol{v}\|t$ 为半径的圆, 随着 t 的变化, 等时线簇为一组同心圆. 若追逐者和逃跑者的速度比 $\|\boldsymbol{v}_p\|/\|\boldsymbol{v}_e\| = \gamma$, 联立上式可定义追逐者和逃跑者在任意 t 时刻的等时线交点集合为

$$\mathcal{A} = \{\boldsymbol{x},\ \boldsymbol{x} \in \mathbf{R}^2 | \|\boldsymbol{x} - \boldsymbol{x}_p^0\| = \gamma \|\boldsymbol{x} - \boldsymbol{x}_e^0\|\} \tag{22}$$

其中, $\boldsymbol{x}$ 为等时线上任一点的坐标, $\|\cdot\|$ 为欧氏距离.

等时线的交点集合可以用于分析追逃博弈的求解方案, 甚至还能用来分析有障碍物存在的情况[56]. 如图 5 所示, 当智能体周围有 2 个障碍物时, 对应的等时线会发生形变, 使得空间分为两类区域: 白色区域为不受影响的等时线区域; 深色区域为绕过障碍物两侧之后的等时线区域. 通过分析多个智能体的等时线簇的交汇情况, 可以进一步研究 TAD 问题. 如 Coon 和 Panagou[36] 基于最优时间策略建立了各个智能体的等时线簇, 利用等时线簇的交汇来确定防御者是否能拦下攻击者.

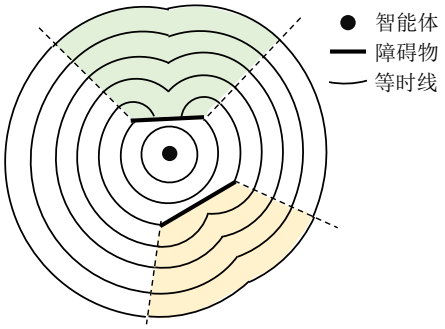

图 5 障碍物对等时线的影响

Fig. 5 Influence of obstacles on isochrones

7.4.2 优势区域和阿波罗尼斯圆

接下来定义智能体的“优势区域”. 给定某个智能体, 对于平面上的任一点, 若该智能体可以先于其他智能体到达该点, 则称该智能体在该点是“占优的”, 所有“占优的”点的集合称为该智能体的“优势区域”. 等时线和优势区域也存在着联系, 在一阶积分器模型下, 追逐者和逃跑者的等时线束交汇即为双方优势区域的分界线. “优势区域”的引入将二维平面有效划分为若干子区域, 有利于提高求解效率. “优势区域”的方法可以用于解决众多追逃博弈. 例如, 文献 [25] 和文献 [155] 分别用“优势区域”的方法解决了经典的“杀人司机问题”和不确定因素下的追逃博弈; Scott 和 Leonard[159] 则通过“优势区域”来解决一追多逃问题.

在一阶积分器模型下, 由式 (22) 可得, 当追逐者和逃跑者速度不等时, 追逃双方优势区域的分界线是一个圆周, 即阿氏圆. 这里我们给出一个详细的推导过程来定义阿氏圆. 当式 (22) 中的 $\gamma \neq 1$ 时, 不妨设 $\boldsymbol{x}_p^0 = (x_p,\ y_p)$, $\boldsymbol{x}_e^0 = (x_e,\ y_e)$, 所以式 (22) 展开为

$$(x-x_p)^2 + (y-y_p)^2 = \gamma^2 (x-x_e)^2 + \gamma^2 (y-y_e)^2 \tag{23}$$

进一步展开, 得到

$$\left(x - \frac{x_p - \gamma^2 x_e}{1-\gamma^2}\right)^2 + \left(y - \frac{y_p - \gamma^2 y_e}{1-\gamma^2}\right)^2 = \left(\frac{\gamma}{1-\gamma^2}\|\boldsymbol{x}_p - \boldsymbol{x}_e\|\right)^2 \tag{24}$$

显然此区域为一个圆周, 即前文提到的阿氏圆. 其中, 阿氏圆的圆心 C_B 和半径 r_B 分别为

$$C_B = \frac{\boldsymbol{x}_p - \gamma^2 \boldsymbol{x}_e}{1-\gamma^2} \tag{25}$$

$$r_B = \frac{\gamma \|\boldsymbol{x}_p - \boldsymbol{x}_e\|}{1-\gamma^2} \tag{26}$$

阿氏圆圆周上的每个点都是双方可以同时到达的位置. 当 $\gamma > 1$ 时, 逃跑者在阿氏圆内部, 其优势区域在阿氏圆内, 阿氏圆外为追逐者的优势区域; 当 $\gamma < 1$ 时, 正好相反. 对于上述简单动态模型下且没有任何限制的二维平面上的追逃博弈, 其最佳抓捕点一定在阿氏圆上[10]. 当追逐者速度大于逃跑者时, 逃跑者一定会被抓到, 因此逃跑者的最优策略是尽量延长生存时间, 因而选择向阿氏圆上的最远点逃跑, 追逐者的策略为直接向该点移动[160]. 基于以上分析, 阿氏圆被广泛用于平面追逃博弈的研究. Pachter 等[26] 利用阿氏圆研究多追一逃博弈, 指出最终的抓捕结果主要是由至多三个关键追逐者决定的; Liang 等[78] 用基于阿氏圆的协同算法来处理存在障碍物的多追一无人机集群对抗问题.

7.4.3 Voronoi 分割

基于优势区域和阿氏圆的概念可以给出 Voronoi 分割的定义. 若给定二维平面上的一个追逐者 p 和一个逃跑者 e, 追逃双方都设定为一阶积分器模型, 其他设定参数同式 (22). 当 $\gamma = 1$ 时, $\mathcal{A}$ 为追逐者和逃跑者实时位置连线的垂直平分线, 该直线 (线段) 把空间分为两块, 显然追逐者所在的一侧为其优势区域, 另一侧为逃跑者优势区域. 当问题限

定在有界区域内时, 对任一智能体, 会和每个与其相邻的智能体形成垂直平分线. 这些平分线相互交错, 并可能结合边界, 共同构成一个封闭区域, 即 Voronoi 小块, 该小块将智能体包含在内. 这样, 便实现了对区域空间的 Voronoi 划分. 在图 6 中给出了一个 Voronoi 分割法的例子.

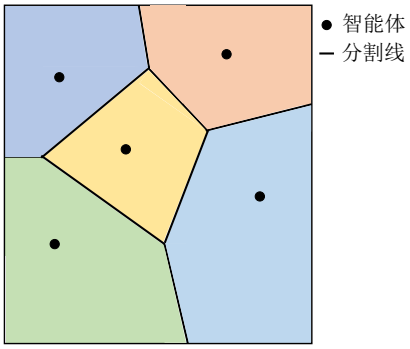

图 6　Voronoi 分割示例

Fig. 6　Voronoi partition example

由此, Pierson 等[86] 提出了利用 Voronoi 图形分割法来设计追逐者的策略, 保证一队追逐者在有限时间内抓到所有逃跑者. 进一步地, Zhou 等[161] 针对平面有界区域内多追逐者–单逃跑者问题开展了研究, 提出了一个创新的围捕算法. 该算法以缩小逃跑者的 Voronoi 小块的面积为基本思路, 有效地实现了对逃跑者的围捕. 这一算法后来也被广泛应用到其他情形的追逃博弈研究中[81, 162−163].

几何法就是一种构造–验证方法, 先利用几何特征推导博弈的策略和值函数, 然后把值函数代入 HJI 方程来验证策略的最优性. 几何法直观高效, 已成为如今解决追逃博弈的重要方法. 但几何法也存在局限性: 该方法主要被用于一阶动态模型玩家的各类博弈中. 然而, 近些年随着追逃博弈不断向无人机、机器人控制领域发展, 一阶积分器模型难以精准模拟机器人系统的真实动态行为和物理属性. Li 等[39] 率先将几何法拓展到二阶积分器模型, 他们将一追一逃两个智能体建模为二阶积分器模型, 分别构建两个智能体的等时线簇, 利用此来求解“抓捕点”, 而且提出一个算法来计算抓捕时间, 根据“抓捕点”和“抓捕时间”来确定策略. 这一工作创造性地将几何法拓展到了二阶积分器模型上, 证明了几何法在复杂模型下的可行性.

因此, 总结几何法的研究方向有两点: 一方面, 受文献 [39] 的启发, 可以进一步探索如何运用几何法解决二阶积分器模型下更为复杂的追逃问题; 另一方面, 如何把几何法拓展到其他更复杂的动态模型中, 如独轮车模型、航天器模型等, 仍是一个亟待探索的问题.

本节对理论求解法、数值求解法、积分强化学习法以及几何法进行了全面探讨, 每种方法均展现了其独特的优势和局限性. 理论求解法为追逃博弈提供了坚实的理论基础, 但往往受限于问题的复杂性和计算难度, 因此适用于问题设定较为简单的情况; 数值求解法凭借数值计算处理更为复杂的情况, 但可能牺牲一定的精度和效率; 积分强化学习法结合了强化学习的自适应性和积分控制的稳定性, 但训练过程可能较为耗时且依赖于大量的数据, 因此数值求解法和积分强化学习法适用于与实际结合比较强的情况; 几何法则因其直观高效性在追逃博弈中得到了广泛应用, 然而对于复杂动态模型的处理能力还有待提高, 因此常用于一些运动轨迹比较规则的情况. 未来, 随着技术的进步和方法的不断完善, 相信这些方法将在追逃博弈中发挥更大的作用.

8　追逃博弈的应用场景

追逃博弈诞生之初就与实际应用紧密结合, Isaacs[10] 研究的两车问题、杀人司机问题都设定在实际场景中. 目前, 追逃博弈问题的相关理论和成果不仅被广泛应用到了智能交通、导航、安全管理等民用领域, 在导弹拦截以及战机空战等军事领域也得到了大量的使用. 特别地, 追逃博弈与无人机相结合, 在无人机集群控制等方面发挥了重要作用.

在智能交通、导航等领域, 追逃博弈扮演着重要角色[164]. 例如, Isler 等[165] 考虑了移动机器人在路径图上的追逃问题, 并研究了追逃策略在避碰方面的应用, 路径图通常由城市道路抽象而成, 因而这一工作可用于智慧交通、路径规划等. 进一步, Sun 等[164] 在结合追逃博弈与避碰问题时, 探讨了如何利用合作策略来实现更高效的避碰.

在安全管理上, Reach-avoid 类型的追逃博弈也发挥着重要作用[166−167]. 此问题中智能体需绕开防守者到达目标区域. 所以从智能体策略的角度, 此问题可用于无人机、无人车躲避守卫, 到达目标点; 从防守者的角度, 则可用于保护目标区域免遭入侵, 用于像城市巡逻、犯罪嫌疑人抓捕等情形中. 对 Reach-avoid 做改进, 把防守者限定在目标区域边界上, 则得到所谓的边界防守追逃博弈[167]. 在边界上, 当有入侵者出现时, 安保人员需要以最低代价在边界附近抓捕入侵者, 这需要用到边界防守追逃博弈.

追逃博弈在军事领域上也被广泛使用. 一些基本追逃博弈设定可以应用到许多场景中. 例如, Coon 和 Panagou[36] 考虑了二阶积分器的 TAD 问题, 这一模型正是对导弹、战斗机等带有转弯半径的飞行物的简化, 可被用于多个场景. Garcia 等[168] 直接设

定情形为: 一架友军战机被导弹追逐, 我方发射防御导弹将攻击导弹击落.

无人机集群的跟踪和围捕等任务与追逃博弈密切相关, 追逃博弈的相关理论在无人机集群上得到了广泛使用[77, 99, 110, 146]. 多旋翼无人机机动性强, 可建模为二阶积分器模型, 可用于对城市中目标的跟踪与围捕[77, 146]. 固定翼无人机需要转弯半径, 可建模为独轮车模型[78], 常用不对称障碍物和不完全信息的追逃博弈来建模固定翼无人机攻击地面坦克、士兵的场景: 障碍物只对逃跑者有效, 追逐者无视障碍物; 逃跑者信息受限, 追逐者掌握全局信息, 这正对应于三维空中的飞机和二维地面的目标[99, 110].

近年来, 海洋强国战略的提出使海洋方面研究吸引了大量关注, 因而追逃博弈也在不断向海洋靠拢. 水面无人艇 (USV) 操控简单、灵活度高, 被广泛用于商业、军事等用途. 因此不少学者以水面无人艇来建模水面上的追逃博弈. Lin 等[52] 研究了水面无人艇的 TAD 问题. Qu 等[141] 就复杂且多障碍物环境中的水上无人艇的追逃策略进行了研究. 有众多海洋勘探、安全救援以及军事任务涉及到追逃博弈, 因此其应用前景极为广阔.

总而言之, 不论是直接应用亦或是间接应用, 追逃博弈都扮演着重要的角色, 蕴含着巨大的应用价值, 还有许多尚未开发的应用方向等待我们去研究.

9 建议与展望

9.1 研究建议

结合上述追逃博弈的发展历史、研究现状, 可总结归纳以下几个方面的建议供研究者参考:

1) 设定上, 三维空间中的追逃博弈对应更多的现实场景, 目前, 二维平面还是占大多数, 三维空间下的追逃博弈研究较少. 三维空间中由于维度升高, 一方面会增大计算量, 求解 HJI 方程等难度也会加大; 另一方面智能体的自由度会增加, 其轨迹更为复杂、策略更灵活. 因此如何简化计算、利用积分强化学习等先进方法求解三维下的方程, 探索复杂轨迹下的新策略, 是三维追逃问题重要的研究方向.

2) 理论上, 将数学理论、新兴技术等引入追逃博弈的研究中也是追逃博弈重要的发展方向. 一方面, 针对不确定因素下的追逃博弈, 可以利用概率论的知识求解, 以期望值的最优性来减弱不确定因素带来的影响; 另一方面, 强化学习的最新成果也能应用于模拟现实情况的追逃博弈中.

3) 方法上, 可以对现有方法做进一步拓展. 对于几何法, 目前主要应用在一阶模型上, 可以进一步拓展到独轮车等其他复杂模型中. 该模型下的等时线形状是否规则、能否根据等时线几何性质找到高效解法, 都是有待研究的. 对于积分强化学习法, 目前许多有关研究都基于 DDPG 框架, 可以进一步探索新的算法框架来得到更稳定的收敛性和更快的收敛速度.

4) 应用上, 随着海洋强国战略推进, 水面和水下追逃博弈研究重要性日益凸显. 作为新兴领域, 追逃问题在各类水面和水下场景应用前景巨大, 但其基础理论研究尚待探索.

9.2 总结展望

本文介绍了追逃博弈的生物学起源, 回顾了追逃博弈的简要历史. 从动态模型、个体数量、空间环境、信息获取以及状态空间五个方面简述了追逃博弈的不同设定, 详细介绍了理论求解、数值求解、积分强化学习和几何法这四种当下主流的追逃博弈求解方法. 并且对当下追逃博弈的应用方向和前景做出了总结和概括. 最后, 针对当前的研究情况和不足, 分别从设定、理论、方法和应用四个方面提出了研究建议.

综上所述, 追逃博弈的研究体系已突破传统生物学范畴, 在智能交通规划、军事防御策略设计及无人机集群控制等领域展现出多维应用价值, 其理论框架亦在持续完善. 值得注意的是, 当前研究仍需在复杂场景构造、数学理论融合及研究方法创新等方向取得突破. 未来, 追逃博弈有望在更多维度和场景中发挥关键作用, 为解决实际问题提供更为有效的策略与方法.

附录 A

表 A1 代表性追逃博弈文献分类

Table A1 Classification of representative literature on PE games

求解方法	文献	动态模型				追逃双方数量			维度		信息完整程度		状态空间	
		一阶积分器	二阶积分器	独轮车	其他	一对一	多对一	多对多	二维平面	三维空间	完全信息	不完全信息	连续	离散
理论求解法	[60]	✓	✓				✓		✓		✓		✓	
	[57]	✓		✓		✓			✓		✓		✓	
	[169]			✓		✓			✓		✓		✓	
	[96]	✓						✓	✓			✓		✓

表 A1 代表性追逃博弈文献分类 (续表)

Table A1 Classification of representative literature on PE games (continued table)

求解方法	文献	动态模型				追逃双方数量			维度		信息完整程度		状态空间	
		一阶积分器	二阶积分器	独轮车	其他	一对一	多对一	多对多	二维平面	三维空间	完全信息	不完全信息	连续	离散
数值求解法	[73]	✓					✓		✓			✓	✓	
	[6]		✓			✓			✓		✓		✓	
	[86]	✓						✓		✓		✓	✓	
	[49]				✓	✓				✓	✓		✓	
	[44]			✓			✓		✓			✓	✓	
积分强化学习法	[147]	✓						✓	✓			✓	✓	
	[141]		✓				✓		✓		✓		✓	
	[46]			✓			✓			✓	✓		✓	
	[15]				✓	✓				✓	✓		✓	
	[40]			✓				✓		✓	✓		✓	
	[146]		✓					✓		✓	✓		✓	
几何法	[157]			✓		✓			✓		✓		✓	
	[95]				✓	✓			✓			✓		✓
	[155]	✓						✓	✓			✓	✓	
	[161]	✓					✓		✓		✓		✓	
	[105]				✓	✓			✓		✓			✓
	[23]	✓					✓			✓	✓		✓	
	[31]	✓						✓	✓			✓	✓	
	[39]		✓			✓			✓		✓		✓	

References

1 Domenici P, Blagburn J M, Bacon J P. Animal escapology Ⅱ: Escape trajectory case studies. *Journal of Experimental Biology*, 2011, **214**(15): 2474−2494

2 Tay N E, Warburton N M, Moseby K E, Fleming P A. Predator escape behaviour in threatened marsupials. *Animal Conservation*, 2023, **26**(4): 587−601

3 FitzGibbon C D. The costs and benefits of predator inspection behaviour in Thomson′s gazelles. *Behavioral Ecology and Sociobiology*, 1994, **34**(2): 139−148

4 Scheel D, Packer C. Group hunting behaviour of lions: A search for cooperation. *Animal Behaviour*, 1991, **41**(4): 697−709

5 Wang J N, Li G L, Liang L, Wang C Y, Deng F. Pursuit-evasion games of multiple cooperative pursuers and an evader: A biological-inspired perspective. *Communications in Nonlinear Science and Numerical Simulation*, 2022, **110**: Article No. 106386

6 Li W. A dynamics perspective of pursuit-evasion: Capturing and escaping when the pursuer runs faster than the agile evader. *IEEE Transactions on Automatic Control*, 2017, **62**(1): 451−457

7 Li W. The confinement-escape problem of a defender against an evader escaping from a circular region. *IEEE Transactions on Cybernetics*, 2016, **46**(4): 1028−1039

8 Weintraub I E, Pachter M, Garcia E. An introduction to pursuit-evasion differential games. In: Proceedings of the American Control Conference (ACC). Denver, USA: IEEE, 2020. 1049−1066

9 Mu Z X, Pan J, Zhou Z Y, Yu J Z, Cao L. A survey of the pursuit-evasion problem in swarm intelligence. *Frontiers of Information Technology & Electronic Engineering*, 2023, **24**(8): 1093−1116

10 Isaacs R. *Differential Games: A Mathematical Theory With Applications to Warfare and Pursuit, Control and Optimization*. New York: John Wiley & Sons, Inc., 1965.

11 Starr A W, Ho Y C. Further properties of nonzero-sum differential games. *Journal of Optimization Theory and Applications*, 1969, **3**(4): 207−219

12 Ho Y C. Differential games, dynamic optimization, and generalized control theory. *Journal of Optimization Theory and Applications*, 1970, **6**(3): 179−209

13 Zhang Si-Ying. *Differential Games*. Beijing: Science Press, 1987.
(张嗣瀛. 微分对策. 北京: 科学出版社, 1987.)

14 Liu Kun, Zheng Xiao-Shuai, Lin Ye-Ming, Han Le, Xia Yuan-Qing. Design of optimal strategies for the pursuit-evasion problem based on differential game. *Acta Automatica Sinica*, 2021, **47**(8): 1840−1854
(刘坤, 郑晓帅, 林业茗, 韩乐, 夏元清. 基于微分博弈的追逃问题最优策略设计. 自动化学报, 2021, **47**(8): 1840−1854)

15 Geng Yuan-Zhuo, Yuan Li, Huang Huang, Tang Liang. Terminal-guidance based reinforcement-learning for orbital pursuit-evasion game of the spacecraft. *Acta Automatica Sinica*, 2023, **49**(5): 974−984
(耿远卓, 袁利, 黄煌, 汤亮. 基于终端诱导强化学习的航天器轨道追逃博弈. 自动化学报, 2023, **49**(5): 974−984)

16 Pontryagin L S. On the theory of differential games. *Russian Mathematical Surveys*, 1966, **21**(4): 193−246

17 Dobbie J M. Solution of some surveillance-evasion problems by the methods of differential games. In: Proceedings of the International Conference on Operational Research. New York, USA: John Wiley & Sons, 1966.

18 Tan M. Multi-agent reinforcement learning: Independent versus cooperative agents. In: Proceedings of the 10th International Conference on Machine Learning. Amherst, USA: Morgan Kaufmann Publishers Inc., 1993. 330−337

19 Guy R K. Unsolved problems in combinatorial games. *Combinatorics Advances*. Boston: Springer, 1995. 161−179

20 Vamvoudakis K G, Lewis F L. Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton-Jacobi equations. *Automatica*, 2011, **47**(8): 1556−1569

21 Flynn J. Lion and man: The general case. *SIAM Journal on Control*, 1974, **12**(4): 581−597

22 Oyler D W, Kabamba P T, Girard A R. Pursuit-evasion games in the presence of a line segment obstacle. In: Proceedings of the 53rd IEEE Conference on Decision and Control (CDC). Los Angeles, USA: IEEE, 2014. 1149−1154

23 Garcia E, Casbeer D W, Pachter M. Optimal strategies for a class of multi-player reach-avoid differential games in 3D space. *IEEE Robotics and Automation Letters*, 2020, **5**(3): 4257−4264

24 Nath S, Ghose D. A two-phase evasive strategy for a pursuit-evasion problem involving two non-holonomic agents with incomplete information. *European Journal of Control*, 2022, **68**: Article No. 100677

25 Oyler D W, Girard A R. Dominance regions in the homicidal chauffeur problem. In: Proceedings of the American Control Conference (ACC). Boston, USA: IEEE, 2016. 2494−2499

26 Pachter M, Moll A V, Garcia E, Casbeer D, Milutinović D. Cooperative pursuit by multiple pursuers of a single evader. *Journal of Aerospace Information Systems*, 2020, **17**(8): 371−389

27 Bakolas E, Tsiotras P. Optimal pursuit of moving targets using dynamic Voronoi diagrams. In: Proceedings of the 49th IEEE Conference on Decision and Control (CDC). Atlanta, USA: IEEE, 2010. 7431−7436

28 Ramana M V, Kothari M. Pursuit-evasion games of high speed evader. *Journal of Intelligent & Robotic Systems*, 2017, **85**: 293−306

29 Shishika D, Kumar V. Local-game decomposition for multiplayer perimeter-defense problem. In: Proceedings of the IEEE Conference on Decision and Control (CDC). Miami, USA: IEEE, 2018. 2093−2100

30 Liang L, Deng F, Lu M B, Chen J. Analysis of role switch for cooperative target defense differential game. *IEEE Transactions on Automatic Control*, 2021, **66**(2): 902−909

31 Yan R, Shi Z Y, Zhong Y S. Task assignment for multiplayer reach-avoid games in convex domains via analytical barriers. *IEEE Transactions on Robotics*, 2020, **36**(1): 107−124

32 Zhao Y, Tao Q L, Xian C X, Li Z K, Duan Z S. Prescribed-time distributed Nash equilibrium seeking for noncooperation games. *Automatica*, 2023, **151**: Article No. 110933

33 Xue L, Ye J F, Wu Y B, Liu J, Wunsch D C. Prescribed-time Nash equilibrium seeking for pursuit-evasion game. *IEEE/CAA Journal of Automatica Sinica*, 2024, **11**(6): 1518−1520

34 Bakolas E. Evasion from a group of pursuers with double integrator kinematics. In: Proceedings of the 52nd IEEE Conference on Decision and Control. Firenze, Italy: IEEE, 2013. 1472−1477

35 Selvakumar J, Bakolas E. Evasion from a group of pursuers with a prescribed target set for the evader. In: Proceedings of the American Control Conference (ACC). Boston, USA: IEEE, 2016. 155−160

36 Coon M, Panagou D. Control strategies for multiplayer target-attacker-defender differential games with double integrator dynamics. In: Proceedings of the 56th IEEE Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE, 2017. 1496−1502

37 Chipade V S, Panagou D. IDCAIS: Inter-defender collision-aware interception strategy against multiple attackers. arXiv preprint arXiv: 2112.12098, 2021.

38 Chipade V S, Panagou D. Multiagent planning and control for swarm herding in 2-D obstacle environments under bounded inputs. *IEEE Transactions on Robotics*, 2021, **37**(6): 1956−1972

39 Li S, Wang C, Xie G M. Optimal strategies for pursuit-evasion differential games of players with damped double integrator dynamics. *IEEE Transactions on Automatic Control*, 2024, **69**(8): 5278−5293

40 Kokolakis N M T, Vamvoudakis K G. Bounded rational Dubins vehicle coordination for target tracking using reinforcement learning. *Automatica*, 2023, **149**: Article No. 110732

41 Patsko V S, Turova V L. Homicidal chauffeur game: History and modern studies. *Advances in Dynamic Games: Theory, Applications, and Numerical Methods for Differential and Stochastic Games*. Boston: Birkhäuser, 2011. 227−251

42 Pachter M, Coates S. The classical homicidal chauffeur game. *Dynamic Games and Applications*, 2019, **9**(3): 800−850

43 Exarchos I, Tsiotras P, Pachter M. On the suicidal pedestrian differential game. *Dynamic Games and Applications*, 2015, **5**(3): 297−317

44 Nath S, Ghose D. Worst-case scenario evasive strategies in a two-on-one engagement between Dubins′ vehicles with partial information. *IEEE Control Systems Letters*, 2023, **7**: 25−30

45 Sani M, Robu B, Hably A. Pursuit-evasion game for nonholonomic mobile robots with obstacle avoidance using NMPC. In: Proceedings of the 28th Mediterranean Conference on Control and Automation (MED). Saint-Rapha, France: IEEE, 2020. 978−983

46 Manoharan A, Thakur P, Singh A K. Multi-agent target defense game with learned defender to attacker assignment. In: Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS). Warsaw, Poland: IEEE, 2023. 297−304

47 Zhu Hai. Optimal Control of Spacecraft Orbital Pursuit-evasion Based on Differential Game [Master thesis], National University of Defense Technology, China, 2017.
(祝海. 基于微分对策的航天器轨道追逃最优控制策略 [硕士学位论文], 国防科技大学, 中国, 2017.)

48 Clohessy W H, Wiltshire R S. Terminal guidance system for satellite rendezvous. *Journal of the Aerospace Sciences*, 1960, **27**(9): 653−658

49 Zhang C M, Zhu Y W, Yang L P, Zeng X. An optimal guidance method for free-time orbital pursuit-evasion game. *Journal of Systems Engineering and Electronics*, 2022, **33**(6): 1294−1308

50 Venigalla C, Scheeres D. Spacecraft rendezvous and pursuit/evasion analysis using reachable sets. In: Proceedings of the Space Flight Mechanics Meeting. Kissimmee, USA: American Institute of Aeronautics and Astronautics, Inc., 2018.

51 Li Z Y, Zhu H, Yang Z, Luo Y Z. A dimension-reduction solution of free-time differential games for spacecraft pursuit-evasion. *Acta Astronautica*, 2019, **163**: 201−210

52 Lin B, Qiao L, Jia Z H, Sun Z J, Sun M, Zhang W D. Control strategies for target-attacker-defender games of USVs. In: Proceedings of the 6th International Conference on Automation, Control and Robotics Engineering (CACRE). Dalian, China: IEEE, 2021. 191−198

53 Ho Y, Bryson A, Baron S. Differential games and optimal pursuit-evasion strategies. *IEEE Transactions on Automatic Control*, 1965, **10**(4): 385−389

54 Kothari M, Manathara J G, Postlethwaite I. Cooperative multiple pursuers against a single evader. *Journal of Intelligent & Robotic Systems*, 2017, **86**(3): 551−567

55 Yufereva O. Lion and man game in compact spaces. *Dynamic Games and Applications*, 2019, **9**(1): 281−292

56 Oyler D W, Kabamba P T, Girard A R. Pursuit-evasion games in the presence of obstacles. *Automatica*, 2016, **65**: 1−11

57 Exarchos I, Tsiotras P. An asymmetric version of the two car pursuit-evasion game. In: Proceedings of the 53rd IEEE Confer-

ence on Decision and Control (CDC). Los Angeles, USA: IEEE, 2014. 4272−4277

58 Das G, Dorothy M, Bell Z I, Shishika D. Guarding a non-maneuverable translating line with an attached defender. arXiv preprint arXiv: 2209.09318, 2022.

59 Liang L, Deng F, Wang J N, Lu M B, Chen J. A reconnaissance penetration game with territorial-constrained defender. *IEEE Transactions on Automatic Control*, 2022, **67**(11): 6295−6302

60 Levchenkov A Y, Pashkov A G. Differential game of optimal approach of two inertial pursuers to a noninertial evader. *Journal of Optimization Theory and Applications*, 1990, **65**(3): 501−518

61 Chen J, Zha W, Peng Z H, Gu D B. Multi-player pursuit-evasion games with one superior evader. *Automatica*, 2016, **71**: 24−32

62 Yan R, Shi Z Y, Zhong Y S. Cooperative strategies for two-evader-one-pursuer reach-avoid differential games. *International Journal of Systems Science*, 2021, **52**(9): 1894−1912

63 Liu S Y, Zhou Z Y, Tomlin C, Hedrick K. Evasion as a team against a faster pursuer. In: Proceedings of the American Control Conference (ACC). Washington, USA: IEEE, 2013. 5368−5373

64 Wang D, Peng Z H. Pursuit-evasion games of multi-players with a single faster player. In: Proceedings of the 35th Chinese Control Conference (CCC). Chengdu, China: IEEE, 2016. 2583−2588

65 Scott W L, Leonard N E. Optimal evasive strategies for multiple interacting agents with motion constraints. *Automatica*, 2018, **94**: 26−34

66 Garcia E, Casbeer D W, von Moll A, Pachter M. Multiple pursuer multiple evader differential games. *IEEE Transactions on Automatic Control*, 2021, **66**(5): 2345−2350

67 Wei M, Chen G S, Cruz J B, Haynes L S, Pham K, Blasch E. Multi-pursuer multi-evader pursuit-evasion games with jamming confrontation. *Journal of Aerospace Computing, Information, and Communication*, 2007, **4**(3): 693−706

68 Wei M, Chen G S, Cruz J B, Haynes L S, Chang M H, Blasch E. A decentralized approach to pursuer-evader games with multiple superior evaders in noisy environments. In: Proceedings of the IEEE Aerospace Conference. Big Sky, USA: IEEE, 2007. 1−10

69 Xu L, Hu B, Guan Z H, Cheng X M, Li T, Xiao J W. Multi-agent deep reinforcement learning for pursuit-evasion game scalability. In: Proceedings of the Chinese Intelligent Systems Conference. Haikou, China: Springer, 2019. 658−669

70 Li D X, Cruz J B, Chen G S, Kwan C, Chang M H. A hierarchical approach to multi-player pursuit-evasion differential games. In: Proceedings of the 44th IEEE Conference on Decision and Control (CDC). Seville, Spain: IEEE, 2005. 5674−5679

71 Li D X, Cruz Jr J B, Schumacher C J. Stochastic multi-player pursuit-evasion differential games. *International Journal of Robust and Nonlinear Control*, 2008, **18**(2): 218−247

72 Yan R, Deng R L, Lai H W, Zhang W X, Shi Z Y, Zhong Y S. Multiplayer homicidal chauffeur reach-avoid games via guaranteed winning strategies. arXiv preprint arXiv: 2107.04709, 2021.

73 LaValle S M, Lin D, Guibas L J, Latombe J C, Motwani R. Finding an unpredictable target in a workspace with obstacles. In: Proceedings of the International Conference on Robotics and Automation. Albuquerque, USA: IEEE, 1997. 737−742

74 Razali S, Meng Q G, Yang S H. A refined immune systems inspired model for multi-robot shepherding. In: Proceedings of the Second World Congress on Nature and Biologically Inspired Computing (NaBIC). Kitakyushu, Japan: IEEE, 2010. 473−478

75 Bhadauria D, Gosse S, Pipp J. Capturing an evader in a polygonal environment with obstacles [Online], available: https://www.conservancy.umn.edu/items/256a22f6-ba8c-4b12-a80d-300b7cba947c, June 15, 2024

76 de Souza C, Newbury R, Cosgun A, Castillo P, Vidolov B, Kulić D. Decentralized multi-agent pursuit using deep reinforcement learning. *IEEE Robotics and Automation Letters*, 2021, **6**(3): 4552−4559

77 Zhang R L, Zong Q, Zhang X Y, Dou L Q, Tian B L. Game of drones: Multi-UAV pursuit-evasion game with online motion planning by deep reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(10): 7900−7909

78 Liang X, Zhou B R, Jiang L P, Meng G L, Xiu Y. Collaborative pursuit-evasion game of multi-UAVs based on Apollonius circle in the environment with obstacle. *Connection Science*, 2023, **35**(1): Article No. 2168253

79 Garcia E, Casbeer D W, Pachter M. Optimal strategies of the differential game in a circular region. *IEEE Control Systems Letters*, 2020, **4**(2): 492−497

80 Casini M, Garulli A. A novel family of pursuit strategies for the lion and man problem. In: Proceedings of the 56th IEEE Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE, 2017. 6436−6441

81 Yan R, Mi S, Duan X M, Chen J T, Ji X Y. Pursuit winning strategies for Reach-Avoid games with polygonal obstacles. *IEEE Transactions on Automatic Control*, DOI: 10.1109/TAC.2024.3438806

82 Flynn J O. Lion and man: The boundary constraint. *SIAM Journal on Control*, 1973, **11**(3): 397−411

83 Yan R, Shi Z Y, Zhong Y S. Defense game in a circular region. In: Proceedings of the 56th IEEE Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE, 2017. 5590−5595

84 Ruiz U, Isler V. Capturing an omnidirectional evader in convex environments using a differential drive robot. *IEEE Robotics and Automation Letters*, 2016, **1**(2): 1007−1013

85 Okabe A, Boots B, Sugihara K. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. New York: John Wiley, 1992.

86 Pierson A, Wang Z J, Schwager M. Intercepting rogue robots: An algorithm for capturing multiple evaders with multiple pursuers. *IEEE Robotics and Automation Letters*, 2017, **2**(2): 530−537

87 Li S, Wang C, Xie G M. Pursuit-evasion differential games of players with different speeds in spaces of different dimensions. In: Proceedings of the American Control Conference (ACC). Atlanta, USA: IEEE, 2022. 1299−1304

88 Zhang R Q, Li S, Wang C, Xie G M. Optimal strategies for the game with two faster 3D pursuers and one slower 2D evader. In: Proceedings of the 41st Chinese Control Conference (CCC). Hefei, China: IEEE, 2022. 1767−1772

89 Zhi J X, Hao Y, Vo C, Morales M, Lien J M. Computing 3-D from-region visibility using visibility integrity. *IEEE Robotics and Automation Letters*, 2019, **4**(4): 4286−4291

90 Chen N, Li L J, Mao W J. Equilibrium strategy of the pursuit-evasion game in three-dimensional space. *IEEE/CAA Journal of Automatica Sinica*, 2024, **11**(2): 446−458

91 Shen H X, Casalino L. Revisit of the three-dimensional orbital pursuit-evasion game. *Journal of Guidance, Control, and Dynamics*, 2018, **41**(8): 1823−1831

92 Yan R, Duan X M, Shi Z Y, Zhong Y S, Bullo F. Matching-based capture strategies for 3D heterogeneous multiplayer reach-avoid differential games. *Automatica*, 2022, **140**: Article

No. 110207

93 Lewin J, Breakwell J V. The surveillance-evasion game of degree. *Journal of Optimization Theory and Applications*, 1975, **16**(3): 339−353

94 Greenfeld I. A differential game of surveillance evasion of two identical cars. *Journal of Optimization Theory and Applications*, 1987, **52**(1): 53−79

95 Bopardikar S D, Bullo F, Hespanha J P. Sensing limitations in the lion and man problem. In: Proceedings of the American Control Conference (ACC). New York, USA: IEEE, 2007. 5958−5963

96 Lopez V G, Lewis F L, Wan Y, Sanchez E N, Fan L L. Solutions for multiagent pursuit-evasion games on communication graphs: Finite-time capture and asymptotic behaviors. *IEEE Transactions on Automatic Control*, 2020, **65**(5): 1911−1923

97 Zemskov K A, Pashkow A G. Construction of optimal position strategies in a differential pursuit-evasion game with one pursuer and two evaders. *Journal of Applied Mathematics and Mechanics*, 1997, **61**(3): 391−399

98 Liu S Y, Zhou Z Y, Tomlin C, Hedrick J K. Evasion of a team of dubins vehicles from a hidden pursuer. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE, 2014. 6771−6776

99 von Moll A, Garcia E, Casbeer D, Suresh M, Swar S C. Multiple-pursuer, single-evader border defense differential game. *Journal of Aerospace Information Systems*, 2020, **17**(8): 407−416

100 Kartal Y, Subbarao K, Dogan A, Lewis F. Optimal game theoretic solution of the pursuit-evasion intercept problem using on-policy reinforcement learning. *International Journal of Robust and Nonlinear Control*, 2021, **31**(16): 7886−7903

101 Littlewood J E. *A Mathematician's Miscellany*. Cambridge: Cambridge University Press, 1953.

102 Kohlenbach U, López-Acedo G, Nicolae A. A uniform betweenness property in metric spaces and its role in the quantitative analysis of the "Lion-Man" game. *Pacific Journal of Mathematics*, 2021, **310**(1): 181−212

103 Sgall J. Solution of David Gale′s lion and man problem. *Theoretical Computer Science*, 2001, **259**(1−2): 663−670

104 Casini M, Garulli A. An improved lion strategy for the lion and man problem. *IEEE Control Systems Letters*, 2017, **1**(1): 38−43

105 Casini M, Garulli A. A new class of pursuer strategies for the discrete-time lion and man problem. *Automatica*, 2019, **100**: 162−170

106 Casini M, Criscuoli M, Garulli A. A discrete-time pursuit-evasion game in convex polygonal environments. *Systems & Control Letters*, 2019, **125**: 22−28

107 Bopardikar S D, Bullo F, Hespanha J P. Cooperative pursuit with sensing limitations. In: Proceedings of the American Control Conference (ACC). New York, USA: IEEE, 2007. 5394−5399

108 Li D X, Cruz J B. Graph-based strategies for multi-player pursuit evasion games. In: Proceedings of the 46th IEEE Conference on Decision and Control. New Orleans, USA: IEEE, 2007. 4063−4068

109 Chen H, Kalyanam K, Zhang W, Casbeer D. Intruder isolation on a general road network under partial information. *IEEE Transactions on Control Systems Technology*, 2017, **25**(1): 222−234

110 Kalyanam K, Casbeer D, Pachter M. Graph search of a moving ground target by a UAV aided by ground sensors with local information. *Autonomous Robots*, 2020, **44**(5): 831−843

111 Sundaram S, Kalyanam K, Casbeer D W. Pursuit on a graph under partial information from sensors. In: Proceedings of the American Control Conference (ACC). Seattle, USA: IEEE, 2017. 4279−4284

112 Dong X, Zhang H G, Ming Z Y. Adaptive optimal control via Q-learning for multi-agent pursuit-evasion games. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2024, **71**(6): 3056−3060

113 Sugihara K, Suzuki I. Optimal algorithms for a pursuit-evasion problem in grids. *SIAM Journal on Discrete Mathematics*, 1989, **2**(1): 126−143

114 Dawes R W. Some pursuit-evasion problems on grids. *Information Processing Letters*, 1992, **43**(5): 241−247

115 Bhattacharya S, Banerjee A, Bandyopadhyay S. CORBA-based analysis of multi agent behavior. *Journal of Computer Science and Technology*, 2005, **20**(1): 118−124

116 Bhattacharya S, Paul G, Sanyal S. A cops and robber game in multidimensional grids. *Discrete Applied Mathematics*, 2010, **158**(16): 1745−1751

117 Das S, Gahlawat H. Variations of cops and robbers game on grids. *Discrete Applied Mathematics*, 2021, **305**: 340−349

118 Lewin J, Olsder G J. The isotropic rocket——A surveillance evasion game. *Computers & Mathematics With Applications*, 1989, **18**(1−3): 15−34

119 Altaher M, Elmougy S, Nomir O. Intercepting a superior missile: A reachability analysis of an Apollonius circle-based multiplayer differential game. *International Journal of Innovative Computing, Information and Control*, 2019, **15**(1): 369−381

120 Jang J S, Tomlin C. Control strategies in multi-player pursuit and evasion game. In: Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit. San Francisco, USA: AIAA, 2005. Article No. 6239

121 Osborne M J, Rubinstein A. *A Course in Game Theory*. Cambridge: MIT Press, 1994.

122 Başar T, Olsder G J. *Dynamic Noncooperative Game Theory* (Second edition). New York: Academic Press Inc., 1998.

123 Krasovskii N N, Subbotin A I, Kotz S. *Game-theoretical Control Problems*. New York: Springer, 1988.

124 Mizukami K, Eguchi K. A geometrical approach to problems of pursuit-evasion games. *Journal of the Franklin Institute*, 1977, **303**(4): 371−384

125 Samatov B T, Soyibboev U B. Differential game with a lifeline for the inertial movements of players. *Ural Mathematical Journal*, 2021, **7**(2): 94−109

126 Petrov N N. "Soft" capture in pontryagin′s example with many participants. *Journal of Applied Mathematics and Mechanics*, 2003, **67**(5): 671−680

127 Blagodatskikh A I. On group pursuit problem in Pontryagin′s nonstationary example. *Vestn. Udmurt. Gos. Univ. Ser. Mat*, 2007, **1**: 17−24

128 Petrov N N, Solov′eva N A. Multiple capture in Pontryagin′s recurrent example. *Automation and Remote Control*, 2016, **77**(5): 855−861

129 Lewin J. *Differential Games: Theory and Methods for Solving Game Problems With Singular Surfaces*. London: Springer, 2012.

130 Wise K A, Sedwick J L. Successive approximation solution of the HJI equation. In: Proceedings of the 33rd IEEE Conference on Decision and Control. Lake Buena Vista, USA: IEEE, 1994. 1387−1391

131 Yang X, Liu D R, Ma H W, Xu Y C. Online approximate solution of HJI equation for unknown constrained-input nonlinear continuous-time systems. *Information Sciences*, 2016, **328**: 435−454

132 Earl M G, D′Andrea R. Modeling and control of a multi-agent system using mixed integer linear programming. In: Proceedings of the 41st IEEE Conference on Decision and Control. Las

Vegas, USA: IEEE, 2002. 107−111

133 Ni Y J, Gao S H, Huang S N, Xiang C, Ren Q Y, Lee T H. Multi-agent cooperative pursuit-evasion control using gene expression programming. In: Proceedings of the 47th Annual Conference of the IEEE Industrial Electronics Society. Toronto, Canada: IEEE, 2021. 1−6

134 Asgharnia A, Schwartz H M, Atia M. Multi-invader multi-defender differential game using reinforcement learning. In: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). Padua, Italy: IEEE, 2022. 1−8

135 Kachroo P, Shedied S A, Bay J S, Vanlandingham H. Dynamic programming solution for a class of pursuit evasion problems: The herding problem. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 2001, **31**(1): 35−41

136 Hespanha J P, Pappas G J, Prandini M. Greedy control for hybrid pursuit-evasion games. In: Proceedings of the European Control Conference (ECC). Los Angeles, USA: 2001. 2621−2626

137 Cristiani E, Falcone M. Fully-discrete schemes for the value function of pursuit-evasion games with state constraints. *Advances in Dynamic Games and Their Applications: Analytical and Numerical Developments*. Boston: Birkhäuser, 2009. 1−30

138 Anderson G. Feedback control for pursuit-evasion problems between two spacecraftbased on differential dynamic programming. In: Proceedings of the 15th Aerospace Sciences Meeting. Los Angeles, USA: AIAA, 1977. Article No. 34

139 Alexopoulos A, Schmidt T, Badreddin E. A pursuit-evasion game between unmanned aerial vehicles. In: Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO). Vienna, Austria: IEEE, 2014. 74−81

140 Xing J B, Zeng X L. A deep reinforcement learning method for lion and man problem. In: Proceedings of the 40th Chinese Control Conference (CCC). Shanghai, China: IEEE, 2021. 8366−8371

141 Qu X Q, Gan W H, Song D L, Zhou L Q. Pursuit-evasion game strategy of USV based on deep reinforcement learning in complex multi-obstacle environment. *Ocean Engineering*, 2023, **273**: Article No. 114016

142 Wan K F, Wu D W, Zhai Y W, Li B, Gao X G, Hu Z J. An improved approach towards multi-agent pursuit-evasion game decision-making using deep reinforcement learning. *Entropy*, 2021, **23**(11): Article No. 1433

143 Li B, Wang J M, Song C, Yang Z P, Wan K F, Zhang Q F. Multi-UAV roundup strategy method based on deep reinforcement learning CEL-MADDPG algorithm. *Expert Systems With Applications*, 2024, **245**: Article No. 123018

144 Li X L, Li Z, Zheng X L, Yang X B, Yu X H. The study of crash-tolerant, multi-agent offensive and defensive games using deep reinforcement learning. *Electronics*, 2023, **12**(2): Article No. 327

145 Kokolakis N M T, Kanellopoulos A, Vamvoudakis K G. Bounded rational unmanned aerial vehicle coordination for adversarial target tracking. In: Proceedings of the American Control Conference (ACC). Denver, USA: IEEE, 2020. 2508−2513

146 Xiong H, Zhang Y. Reinforcement learning-based formation-surrounding control for multiple quadrotor UAVs pursuit-evasion games. *ISA Transactions*, 2024, **145**: 205−224

147 Zhou Z J, Xu H. Decentralized optimal large scale multi-player pursuit-evasion strategies: A mean field game approach with reinforcement learning. *Neurocomputing*, 2022, **484**: 46−58

148 Vrabie D. Online Adaptive Optimal Control for Continuous-time Systems [Ph.D. dissertation], The University of Texas at Arlington, USA, 2010.

149 Gong Z F, He B, Liu G, Zhang X B. Solution for pursuit-evasion game of agents by adaptive dynamic programming. *Electronics*, 2023, **12**(12): Article No. 2595

150 Gong Z F, He B, Hu C, Zhang X B, Kang W J. Online adaptive dynamic programming-based solution of networked multiple-pursuer and single-evader game. *Electronics*, 2022, **11**(21): Article No. 3583

151 Sun J L, Liu C S. Finite-horizon differential games for missile-target interception system using adaptive dynamic programming with input constraints. *International Journal of Systems Science*, 2018, **49**(2): 264−283

152 Zhang Z X, Zhang K, Xie X P, Sun J Y. Fixed-time zero-sum pursuit-evasion game control of multisatellite via adaptive dynamic programming. *IEEE Transactions on Aerospace and Electronic Systems*, 2024, **60**(2): 2224−2235

153 Vamvoudakis K G, Lewis F L, Hudas G R. Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality. *Automatica*, 2012, **48**(8): 1598−1611

154 Peng C, Liu X M, Ma J J. Design of safe optimal guidance with obstacle avoidance using control barrier function-based actor-critic reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, **53**(11): 6861−6873

155 Oyler D W, Kabamba P T, Girard A R. Dominance in pursuit-evasion games with uncertainty. In: Proceedings of the 54th IEEE Conference on Decision and Control (CDC). Osaka, Japan: IEEE, 2015. 5859−5864

156 Agasti A, Reddy P V, Bhikkaji B. Optimal role assignment for multiplayer reach-avoid differential games in 3D space. arXiv preprint arXiv: 2303.07885, 2023.

157 Meier L. A new technique for solving pursuit-evasion differential games. *IEEE Transactions on Automatic Control*, 1969, **14**(4): 352−359

158 Getz W M, Pachter M. Capturability in a two-target "game of two cars". *Journal of Guidance and Control*, 1981, **4**(1): 15−21

159 Scott W, Leonard N E. Dynamics of pursuit and evasion in a heterogeneous herd. In: Proceedings of the 53rd IEEE Conference on Decision and Control (CDC). Los Angeles, USA: IEEE, 2014. 2920−2925

160 Garcia E, Fuchs Z E, Milutinovic D, Casbeer D W, Pachter M. A geometric approach for the cooperative two-pursuer one-evader differential game. *IFAC-PapersOnLine*, 2017, **50**(1): 15209−15214

161 Zhou Z Y, Zhang W, Ding J, Huang H M, Stipanović D M, Tomlin C J. Cooperative pursuit with Voronoi partitions. *Automatica*, 2016, **72**: 64−72

162 Zhou S B, Li H P, Chen Z Y. Optimal containment strategies on high-speed evader using multiple pursuers with point-capture. In: Proceedings of the 42nd Chinese Control Conference (CCC). Tianjin, China: IEEE, 2023. 1−6

163 Zhang Z, Zhang D Y, Zhang Q R, Pan W, Hu T J. DACOOP-A: Decentralized adaptive cooperative pursuit via attention. *IEEE Robotics and Automation Letters*, 2024, **9**(6): 5504−5511

164 Sun Z Y, Sun H B, Li P, Zou J. Cooperative strategy for pursuit-evasion problem with collision avoidance. *Ocean Engineering*, 2022, **266**: Article No. 112742

165 Isler V, Sun D F, Sastry S. Roadmap based pursuit-evasion and collision avoidance. *Robotics: Science and Systems*. Cambridge: MIT Press, 2005. 257−264

166 Selvakumar J, Bakolas E. Feedback strategies for a reach-avoid game with a single evader and multiple pursuers. *IEEE Transactions on Cybernetics*, 2021, **51**(2): 696−707

167 Shishika D, Paulos J, Kumar V. Cooperative team strategies for multi-player perimeter-defense games. *IEEE Robotics and Automation Letters*, 2020, **5**(2): 2738−2745

168 Garcia E, Casbeer D W, Pachter M. Active target defense us-

ing first order missile models. *Automatica*, 2017, **78**: 139−143

169 Bera R, Makkapati V R, Kothari M. A comprehensive differential game theoretic solution to a game of two cars. *Journal of Optimization Theory and Applications*, 2017, **174**(3): 818−836

迟嵩禹 北京大学工学院硕士研究生. 主要研究方向为多智能体博弈. E-mail: 2301213114@stu.pku.edu.cn
(**CHI Song-Yu** Master student at the College of Engineering, Peking University. His main research interest is multi-agent games.)

李 帅 北京大学工学院博士后. 主要研究方向为多智能体系统, 多机器人控制, 人工智能和博弈论. E-mail: shuaier@pku.edu.cn
(**LI Shuai** Postdoctor at the College of Engineering, Peking University. His research interest covers multi-agent systems, multi-robot control, artificial intelligence, and game theory.)

王 晨 北京大学软件工程国家工程研究中心副研究员. 主要研究方向为仿生机器人, 多智能体系统和深度学习. E-mail: wangchen@pku.edu.cn
(**WANG Chen** Associate researcher at the National Engineering Research Center of Software Engineering, Peking University. Her research interest covers biomimetic robotics, multi-agent systems, and deep learning.)

谢广明 北京大学工学院教授. 主要研究方向为智能群体理论, 仿生机器人, 网络控制系统和深度学习. 本文通信作者.
E-mail: xiegming@pku.edu.cn
(**XIE Guang-Ming** Professor at the College of Engineering, Peking University. His research interest covers smart swarm theory, biomimetic robotics, networked control systems, and deep learning. Corresponding author of this paper.)