基于微分对策理论的非线性控制回顾与展望

谭拂晓 [1, 2] 刘德荣 [3] 关新平 [4] 罗 斌 [1]

摘 要 微分对策是使用微分方程处理双方或多方连续动态冲突、竞争或合作问题的一种数学工具. 它已经广泛应用于生物学、经济学、国际关系、计算机科学和军事战略等诸多领域. 微分对策实质上是一种双方或多方的最优控制问题, 它将现代控制理论与对策论相融合, 从而比控制理论具有更强的竞争性、对抗性和适用性. 本文根据非线性微分对策理论的控制、均衡及算法阐述了微分对策的理论发展历史, 综述了已有结论与算法的本质, 总结了现有的研究成果. 最后对基于微分对策理论非线性系统的鲁棒性与最优性进行了展望.

关键词 微分对策, 非线性系统, 均衡, HJI 方程, 代价函数

引用格式 谭拂晓, 刘德荣, 关新平, 罗斌. 基于微分对策理论的非线性控制回顾与展望. 自动化学报, 2014, **40**(1): 1−15

DOI 10.3724/SP.J.1004.2014.00001

Review and Perspective of Nonlinear Systems Control Based on Differential Games

TAN Fu-Xiao[1, 2] LIU De-Rong[3] GUAN Xin-Ping[4] LUO Bin[1]

Abstract Differential game is a mathematical tool for dealing with the problems of continuous dynamic conflict, competition or cooperation with two or more control actions using differential equations. It has been widely employed in biology, economics, international relations, computer science, military strategy and so on. Differential game is essentially an optimal control problem of two or more parties. By integration of modern control theory and game theory, differential game thus has stronger competitiveness, confrontation ability and applicability than control theory. Based on control, equilibrium, and algorithms of nonlinear differential game theory, the paper elaborates on the development history of control, surveys the essence of existing conclusions and algorithms, and summarizes the existing research results. Finally, the perspective of robustness and optimality of nonlinear systems based on differential game are discussed and explored.

Key words Differential games, nonlinear system, equilibrium, HJI equation, cost function

Citation Tan Fu-Xiao, Liu De-Rong, Guan Xin-Ping, Luo Bin. Review and perspective of nonlinear systems control based on differential games. *Acta Automatica Sinica*, 2014, **40**(1): 1−15

1 微分对策简介

对策论是描述现实世界中包含矛盾、冲突、对抗、合作等数学模型的理论与方法. 微分对策是指局中人在进行对策活动时, 用微分方程 (组) 来描述对策现象或规律的一种策略. 它是处理双方或多方连续动态冲突、竞争或合作问题的一种数学工具. 微分对策实质上是一种双 (多) 方的最优控制问题, 它将现代控制理论与对策论相融合, 从而比控制理论具有更强的竞争性和对抗性[1].

1944 年数学家冯 · 诺依曼 (Von Neumann) 和经济学家奥 · 摩根斯特恩 (Osker Mor-Genstern) 合著的《对策论与经济行为》(*Theory of Games and Economic Behavior*) 一书, 被认为是对策论发展的一块里程碑, 它完善了对策论的数学理论, 使之系统化和公理化.

微分对策与经典的静态对策都具有一些共同的基本要素: 局中人、策略集、代价函数 (支付函数) 等, 但又有本质区别. 微分对策不能看成是静态对策论的简单扩展. 自 20 世纪 50 年代以来, 由于制导系统拦截飞行器的引入、人造卫星的发射和航空航天

收稿日期 2013-06-14 录用日期 2013-09-18
Manuscript received June 14, 2013; accepted September 18, 2013
国家自然科学基金 (61073116), 安徽省自然科学基金 (1208085MF111), 中国科学院自动化研究所复杂系统管理与控制国家重点实验室开放基金 (20120102), 安徽省教育厅自然科学研究项目 (KJ2011B123), 安徽省博士后基金, 安徽省工业图像处理与分析重点实验室开放基金资助
Supported by National Natural Science Foundation of China (61073116), Anhui Provincial Natural Science Foundation of China (1208085MF111), the Open Research Project from State Key Laboratory of Management and Control for Complex Systems (20120102), Natural Science Research Project of the Education Department of Anhui Province (KJ2011B123), Anhui Postdoctoral Foundation, and Open Fund of Key Laboratory of Anhui Industrial Image Processing and Analysis
本文责任编委 王占山
Recommended by Associate Editor WANG Zhan-Shan
1. 安徽大学计算机科学与技术学院 合肥 230601 2. 阜阳师范学院计算机与信息学院 阜阳 236037 3. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190 4. 上海交通大学电子信息与电气工程学院 上海 200240
1. School of Computer Science and Technology, Anhui University, Hefei 230601 2. School of Computer and Information, Fuyang Teachers College, Fuyang 236037 3. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 4. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240

中有关机动追击技术的发展, 需要处理双方或多方连续动态对抗冲突、竞争与合作问题的一种数学工具, 从而诞生了微分对策. 美国著名的 Rand 公司在空军资助下, 以美国数学家 Issacs 博士为首的研究小组开展了对抗双方都能自由决策行动的追逃问题研究. 他们把现代控制理论中的一些概念、原理与方法引入对策论中, 取得了突破性的成果, 撰写了多篇研究报告, 形成了微分对策的最初研究成果[2].

进入 20 世纪 60 年代, 微分对策与最优控制理论相结合, 使得微分对策理论得到长足发展. 1965 年, Issacs 整理出版了世界上第一部微分对策专著《微分对策》, 标志着微分对策理论的正式诞生. 在此之后, 由于军事方面的原因, 微分对策的研究引起了世界各国的普遍关注, 特别是美国和前苏联出于军备竞赛的需要, 对空战、核导弹与人造卫星拦截、电子战等方面提出了各种类型的微分对策模型, 使得军事微分对策得以迅速发展[3]. 1971 年, 美国科学家 Friedman 采用了两个近似离散对策序列精确定义了微分对策, 建立了微分对策值与鞍点存在性理论, 从而奠定了微分对策理论的数学基础[4−5].

微分对策理论从诞生发展到今天经历了四十多年的突飞猛进发展. 随着定量微分对策和定性微分对策不断完善外, 随机微分对策、多人合作微分对策、非合作微分对策和主从微分对策等方面的研究也取得了很大进展. 美国著名数学家 Nash 最先将微分对策理论引入经济学研究领域, 研究了二人合作对策的谈判问题, 解决了非合作微分对策的混合平衡解的存在性, 因此获得 1994 年诺贝尔经济学奖[6].

微分对策种类繁多, 根据划分标准不同, 分类也各不相同. 按照局中人的数目可分为多人或者二人微分对策; 按照局中人是否有合作的行为, 可分为合作与非合作微分对策; 按照信息结构的不同可划分为不完全信息、完全信息和无信息微分对策; 根据有无支付函数可分为定量与定性微分对策; 对于定量微分对策, 若其支付函数之和为零, 则称为二人零和微分对策, 若不为零, 则称二人非零和微分对策; 按照对策问题中动态系统分类不同可分为偏微分对策和随机微分对策; 按照局中人间合作程度的不同, 可分为纳什平衡、帕雷托最优、组队最优、协商微分对策和主从微分对策等多种形式[7−12].

在微分对策系统中, 由于控制所测信息被噪声破坏、各种干扰、环境的恶劣等因素, 往往具有不确定性, 微分对策问题本身的数学模型也不能完全要求是精确的, 不确定问题是微分对策中的一个主要研究内容. 针对非线性控制系统的不确定性, 目前已提出了一系列诸如 H_∞ 控制、L_2 增益等新的研究成果. 如果把非线性系统中的不确定性考虑成微分对策中的一局中人, 那么把非线性系统的鲁棒控制纳入到微分对策系统中, 研究基于微分对策理论的非线性系统鲁棒性问题是必然和急迫的. 但由于对策现象中局中人相互作用, 使得微分对策比单方最优控制要繁杂得多, 不能完全看成双边或多边最优控制问题. 因此, 基于微分对策理论的非线性系统鲁棒控制研究非常困难, 目前, 关于微分对策鲁棒控制的研究成果也很少[13−17].

在工程实际中, 微分对策、最优控制、变分法和动态规划都是求解变分问题的理论, 并且已经融合成为现代控制理论的一部分. 本文主要讨论零和与非零和、合作与非合作微分对策及在控制理论中的应用, 研究无记忆完全状态信息闭环微分对策的 Nash 平衡解的充分必要条件, 即决策是基于当前状态信息完全已知的条件下进行. 此外, 局中人完全掌握系统当前和过去的状态信息, 并且基于这些信息进行个人决策, 该微分对策成为闭环完整信息结构或模式. 另一方面, 如果对策仅仅基于系统状态的当前值进行, 那么该决策称为 “无记忆性”. 当对策中只包含一个局中人, 那么该决策问题就变为最优控制, 即可以通过 Pontryagin 的 “最大原理” 或 Bellman 的 “动态规划原理” 来求解[18−20].

2 非线性系统微分对策基本定义

多人微分对策是微分对策研究领域的一个重要方面, 在军事、经济和社会问题中广泛应用. 由于参与对策的人不止两个, 且利益关系不一致, 这些局中人从个人的利益出发会寻求一种合适的参与方式参与对策. 如果各个局中人两两之间不相互合作, 那么每一个局中人必须选择一个策略, 从而共同寻求一个平衡点. 如果是非线性的这无疑会增大问题研究的难度. 对于二人非线性微分对策的方法可采用极大极小值法、HJB 方法等. 但是对于多人微分对策, 由于参与对策的局中人增多, 对策情况复杂的多, 传统的方法已经无法确定其均衡解[21].

2.1 离散时间非线性非零和微分对策

基于二人零和微分对策理论, Issacs 在 1954 年 ∼1956 年间首次提出了非零和微分对策, 随后经过 Starr, Ho, Friedman 等的深入研究, 并最终发展成随机微分对策[4, 13−14, 22].

非线性离散确定性 N 人微分对策系统, 其状态方程可以描述如下:

$$\begin{cases} x(k+1) = f_k(x(k), u_1(k), \cdots, u_N(k)) \\ x(0) = x_0, \ \ k \in \{0, \cdots, n-1\} \end{cases} \tag{1}$$

其中, x_0 为初始状态, $x(k) \in \mathbf{R}^n$, $k = 0, \cdots, K$ 是系统状态向量; $u_i(k)$, $i = 1, \cdots, N$, $N \in \mathbf{N}$ 是属

于可测控制集合 $\{U^i \subset \mathbf{R}^{m_i}, m_i \in \mathbf{N}\}$ 的局中人 $\{1, \cdots, N\}$ 的决策变量或控制输入; $f_k : U_1 \times \cdots \times U_N \to \mathbf{R}$ 满足 C^1 函数. 微分对策有关的 N 个局中人的代价函数 (支付函数)$J_i(u_1, \cdots, u_N) : \Gamma^1 \times \cdots \times \Gamma^N \to \mathbf{R}$ 并且满足

$$J_i = \sum_{k=0}^{K} L_k^i\left(x(k+1), x(k), u_1(k), \cdots, u_N(k)\right) \tag{2}$$

其中, Γ^i, $i = 1, \cdots, N$ 是局中人的容许策略空间; $L_k^i : U^1 \times \cdots \times U^N \to \mathbf{R}$, $i = 1, \cdots, N$, $k = 0, \cdots, K$ 是实 C^1 函数; $u_i(k) \in \Gamma^i$.

如果局中人之间的关系是非合作, 那么控制目标是最小化式 (2) 中每一个代价函数, 这样的微分对策称为非零和微分对策. Nash 均衡即为最优决策 $u_i^*(k)$, $i = 1, \cdots, N$ 在每一阶段使其代价函数最小.

定义 1[22]**.** 假设 N 个局中人最优控制策略为 $u_i^*(k) \in \Gamma^i$, 其中 Γ^i, $i = 1, \cdots, N$. 如果对于 $\forall i = 1, \cdots, N$, 代价函数 J_i^* 满足

$$\begin{aligned} J_i^* = J_i(u_1^*, \cdots, u_N^*) \leq \\ J_i\left(u_1^*, \cdots, u_{i-1}^*, u_i, u_{i+1}^*, \cdots, u_N^*\right) \end{aligned} \tag{3}$$

那么就称基于非合作微分对策的 N 个局中人构建了 Nash 均衡. 其中, $u_i^* = \{u_i^*(k), k = 0, \cdots, K\}$, $u_i = \{u_i(k), k = 0, \cdots, K\}$, $i = 1, \cdots, N$.

文献 [7] 基于动态规划原理, 给出了 Nash 均衡解的策略集合存在的充分必要条件.

定理 1[7]**.** 对于 N 人非线性离散非零和微分对策系统 (1) ∼ (3), N 个控制策略集合 $u_i^* \in \Gamma^i$, $i = 1, \cdots, N$ 提供一个 Nash 均衡解的充分必要条件是当且仅当存在 $N \times K$ 个函数 V_i 满足以下递归方程

$$\begin{aligned} V_i(x,k) = \min_{u_i(k) \in U_i} \{L_k^i(x(k+1), x(k), u_1^*(k), \cdots, \\ u_i(k), \cdots, u_N^*(k)) + V_i(x(k+1), k+1)\} = \\ \min_{u_i(k) \in U_i} \{L_k^i(f_k(x, u_1^*(k), \cdots, u_i(k), \cdots, \\ u_N^*(k)), x, u_1^*(k), \cdots, u_i(k), \cdots, u_N^*(k)) + \\ V_i(f_k(x, u_1^*(k), \cdots, u_i(k), \cdots, u_N^*(k))), k+1\} \end{aligned} \tag{4}$$

其中, $i = 1, \cdots, N$, $k \in \{0, 1, \cdots, K\}$, $V_i(x, K+1) = 0$.

从定理 1 中可以得到在每个阶段决策过程中, 不同的代价函数都需要每一个局中人参与, 因此需要存在 $N \times K$ 个函数. 然而, 每一个局中人的单个方程是通过 N 个满足上述递归条件的函数来获得, 这在线性条件下尤其成立.

在非线性控制系统中, 大多考虑二人非零和微分对策, 其状态方程为

$$\begin{cases} x(k+1) = f_k\left(x(k), \omega(k), u(k)\right) \\ x(0) = x_0 \end{cases} \tag{5}$$

其中, $\omega(k) \in W \subset \mathbf{R}^{m_\omega}$ 和 $u(k) \in U \subset \mathbf{R}^{m_u}$ 分别为干扰向量与控制向量, 并且代表两个局中人的决策, $k \in \{0, \cdots, K\}$.

定义二人代价函数 J_1, J_2 分别为

$$J_1(\omega, u) = \sum_{k=0}^{K} L_k^1\left(x(k+1), x(k), \omega(k), u(k)\right) \tag{6}$$

$$J_2(\omega, u) = \sum_{k=0}^{K} L_k^2\left(x(k+1), x(k), \omega(k), u(k)\right) \tag{7}$$

如果存在一对策略集合 $u^*(k)$, $\omega^*(k)$, $k = 0, \cdots, K$ 满足以下两式

$$J_1(\omega^*, u^*) \leq J_1(\omega^*, u), \ \forall u \in U \tag{8}$$

$$J_2(\omega^*, u^*) \leq J_2(\omega, u^*), \ \forall \omega \in W \tag{9}$$

那么该对策集合构成二人非零和微分对策 Nash 均衡.

此外, 根据定理 1, 存在以下两式

$$\begin{aligned} V_1(x,k) = \min_{u(k) \in U} \{L_k^1\left(f_k(x, \omega^*(k), u(k)), x, \right. \\ \left. \omega^*(k), u(k)\right) + V_1\left(f_k(x, \omega^*(k), u(k)), k+1\right)\} \end{aligned} \tag{10}$$

$$\begin{aligned} V_2(x,k) = \min_{\omega(k) \in W} \{L_k^2\left(f_k(x, \omega(k), u^*(k)), x, \right. \\ \left. \omega(k), u^*(k)\right) + V_1\left(f_k(x, \omega(k), u^*(k)), k+1\right)\} \end{aligned} \tag{11}$$

其中, $V_1(x, K+1) = 0$, $V_2(x, K+1) = 0$, $k = 0, \cdots, K$.

方程 (10) 和 (11) 为一对耦合的 HJ 方程对, 对于该方程组求解非常困难, 现有的文献并没有对其进行深入研究. 然而, 对于基于微分对策理论的离散混合 H_2/H_∞ 问题的求解, 该方程组将起着至关重要的作用.

二人零和微分对策是非零和微分对策的一个特例, 其目标方程满足

$$\begin{aligned} J_1(\omega, u) = -J_2(\omega, u) = J(\omega, u) = \\ \sum_{k=0}^{K} L_k\left(x(k+1), x(k), \omega(k), u(k)\right) \end{aligned} \tag{12}$$

在这种情况下, 式 (8) 和式 (9) 满足 $J_1(\omega^*,u^*)+J_2(\omega^*,u^*)=0$, 其中控制 u 是需要最小化, 干扰 ω 需要最大化. 因此, 在控制系统设计中可以把控制 u 作为最小化局中人, 干扰 ω 作为最大化局中人. 此时, Nash 均衡条件 (8) 和 (9) 简化为鞍点存在的条件

$$J(\omega,u^*)\le J(\omega^*,u^*)\le J(\omega^*,u),\ \forall\omega\in W,u\in U \tag{13}$$

其中, (u^*,ω^*) 是最优控制策略, 称为 "鞍点", 故存在以下定理.

定理 2[3, 7−8]**.** 对于式 (5) 和式 (13) 中定义的二人离散零和微分对策系统, 如果一对控制策略 (u^*,ω^*) 构成了鞍点平衡点的解, 那么当且仅当存在 K 个函数集合 V, 对于每个 $k\in[0,K]$, 满足以下迭代方程

$$\begin{aligned}V(x,k)=&\min_{u(k)\in U}\max_{\omega(k)\in W}\{L_k(f_k(x,\omega(k),u(k)),x,\\&\omega(k),u(k))+V(f_k(x,\omega(k),u(k)),k+1)\}=\\&\max_{\omega(k)\in W}\min_{u(k)\in U}\{L_k(f_k(x,\omega(k),u(k)),x,\omega(k),\\&u(k))+V(f_k(x,\omega(k),u(k)),k+1)\}=\\&L_k(f_k(x,\omega^*(k),u^*(k)),x,\omega^*(k),u^*(k))+\\&V(f_k(x,\omega^*(k),u^*(k)),k+1)\end{aligned} \tag{14}$$

其中, $V(x,K+1)=0$.

方程 (14) 为著名的 Isaacs 方程, 并且定理 2 中 "max" 和 "min" 之间的可交换性被称为 Isaacs 条件. 该方程对于离散时间非线性 H_∞ 控制起了至关重要的作用.

2.2 连续非线性非零和微分对策

考虑 N 个局中人参与的非零和微分对策, 其状态方程定义如下:

$$\dot{x}(t)=f(x,u_1,\cdots,u_N,t),\ \ x(t_0)=x_0 \tag{15}$$

其中, $x(t)\in\mathbf{R}^n$ 是 t 时刻的系统状态, $x(t_0)$ 是局中人都已知的初始状态. $u_i\in U^i$, $i=1,\cdots,N$, $N\in\mathbf{N}$ 是 N 个局中人的决策变量或者控制输入, 属于可测控制集合 $U^i\in\mathbf{R}^{m_i}$, $m_i\in\mathbf{N}$, 并且 $f:U^1\times\cdots\times U^N\to\mathbf{R}$ 是满足 C^1 函数.

每个局中人 $i=1,\cdots,N$ 所相关的代价函数为 J_i: $\Gamma^1\times\cdots\times\Gamma^N\to\mathbf{R}$, 其中 Γ^i, $i=1,\cdots,N$ 是局中人最小化策略空间. 定义代价函数为

$$\begin{aligned}J_i(u_1,\cdots,u_N)=&\varphi_i(x(t_f),t_f)+\\&\int_{t_0}^{t_f}L_i(x,u_1,\cdots,u_N,t)\mathrm{d}t,\ \ i=1,\cdots,N\end{aligned} \tag{16}$$

其中, $L_i:U^1\times\cdots\times U^N\times\mathbf{R}^+\to\mathbf{R}$ 和 φ_i, $i=1,\cdots,N$ 分别满足实 C^1 函数. 为了简单, 假设终端时间 t_f 不变, 函数 $\varphi_i(\cdot)$, $i=1,\cdots,N$ 称为终端代价函数.

在非合作微分对策中, 每个局中人都知道自己当前系统状态、系统参数和代价函数, 但是每个局中人不知道其他竞争对手的策略. 因此非合作微分对策控制的目标是从理论上分析在闭环无记忆完全信息结构下的 Nash 均衡策略, 即基于闭环无记忆完全信息条件下非合作微分对策 Nash 均衡解.

定义 2[8]**.** 考虑 N 人参与的非合连续作微分对策系统, 其控制策略为 $u_i^*\in\Gamma^i$, $i=1,\cdots,N$, 其中 Γ^i, $i=1,\cdots,N$ 是其策略集合, 如果对于 $\forall i=1,\cdots,N$, J_i^* 满足

$$\begin{aligned}J_i^*=&J_i(u_1^*,\cdots,u_N^*)\le\\&J_i\left(u_1^*,\cdots,u_{i-1}^*,u_i,u_{i+1}^*,\cdots,u_N^*\right)\end{aligned} \tag{17}$$

那么 u_i^* 为式 (15) 和式 (16) 所构成非合作连续微分对策系统的 Nash 均衡解, 其中, $u_i^*=\{u_i^*(t),\ t\in[t_0,t_f]\}$, $u_i=\{u_i(t),t\in[t_0,t_f]\}$.

为推导出最优性存在条件, 考虑 N 局中人构成的 Piecewise 连续控制策略 $u=\{u_1,\cdots,u_N\}$, 并且第 i 个局中人的值函数为

$$\begin{aligned}V_i(x,t)=&\inf_{u\in U}\Big\{\varphi_i(x(t_f),t_f)+\\&\int_t^{t_f}L_i(x,u,\tau)\mathrm{d}\tau\Big\}=\\&\varphi_i(x(t_f),t_f)+\int_t^{t_f}L_i(x,u^*,\tau)\mathrm{d}\tau\end{aligned} \tag{18}$$

根据定义 2 和动态规划原理, 值函数 V_i, $i=1,\cdots,N$ 的解满足下面 Hamilton-Jacobi 方程

$$\begin{cases}\dfrac{\partial V_i}{\partial t}=-\inf\limits_{u_i\in U^i}H_i\Big(x,t,u_1^*,\cdots,u_{i-1}^*,u_i,u_{i+1}^*,\\\qquad\qquad\cdots,u_N^*,\dfrac{\partial V_i}{\partial x}\Big)\\V_i(x(t_f),t_f)=\varphi_i(x(t_f),t_f)\end{cases} \tag{19}$$

其中, H_i, $i=1,\cdots,N$ 满足

$$H_i\left(x,t,u,\lambda_i^{\mathrm{T}}\right)=L_i(x,u,t)+\lambda_i^{\mathrm{T}}f(x,u,t) \tag{20}$$

均衡控制策略 $u^*=(u_1^*,\cdots,u_N^*)$ 的作用是最小化方程 (19) 的右半部分.

对于方程 (19) 按照末端 $x(t_f)$ 向后求积分, 并且在每个 (x,t) 处必须求解 Hamiltonians 方程中 H_i, $i=1,\cdots,N$ 的静态对策来寻找 Nash 均衡.

基于以上分析, 对于所有的 $\{x,\lambda,t\}$ 及其向量 $H=[H_1,\cdots,H_N]$, Nash 唯一均衡解 u^* 满足

$$\begin{cases}\dfrac{\partial V_i}{\partial t}(x,t)=-H_i\left(x,t,u^*\left(x,t,\dfrac{\partial V_1}{\partial t},\cdots,\dfrac{\partial V_N}{\partial t}\right),\right.\\ \qquad\qquad\left.\dfrac{\partial V_i}{\partial x}\right)\\ \dot{x}=f(x,u_1,\cdots,u_N,t)\end{cases}\tag{21}$$

其中, $u^*=u^*\left(x,t,\frac{\partial V_1}{\partial t},\cdots,\frac{\partial V_N}{\partial t}\right)$. 对方程 (21) 从终端平面上的所有点向后求积分, 即可获得非线性连续微分对策系统的运动轨迹.

对于一个闭环无记忆完全信息状态下 N 人微分对策系统 (21), 其 Nash 均衡解存在的充分条件是满足以下定理.

定理 3[7]**.** 对于式 (15) 和式 (16) 所构成的一个闭环无记忆完全信息状态下的 N 人微分对策系统, 在固定时间 $[t_0,t_f]$ 下, 如果存在 N 个 C^1 函数 V_i, $i=1,\cdots,N$, 在 N 个控制策略 $u^*=[u_1^*,\cdots,u_N^*]$ 下 Nash 均衡解必须满足式 (19) 和式 (20) 所构成的 HJ 方程.

定理 3 的主要结果可以推广到二人非零和连续时间微分对策系统, 并且对于求解混合 H_2/H_∞ 问题, 定理 3 起着至关重要的作用.

针对二人零和连续时间微分对策系统, 其系统方程为

$$\dot{x}=f(x,u,\omega,t),\quad x(t_0)=x_0\tag{22}$$

并且代价函数为

$$J_1(u,\omega)=-J_2(u,\omega)=J(u,\omega)=\varphi(x(t_f),t_f)+\int_{t_0}^{t_f}L(x,u,\omega,t)\,\mathrm{d}t\tag{23}$$

其中, 控制 $u\in U$, 干扰 $\omega\in W$ 分别是两局中人的策略, 并且属于可测集合 $U\in\mathbf{R}^{m_u}$, $W\in\mathbf{R}^{m_\omega}$, x_0 是两局中人都预先知道的初始状态. 那么有与定理 2 对应的连续时间微分对策鞍点存在性定理.

定理 4[3]**.** 在固定期间 $[t_0,t_f]$ 和闭环无记忆信息状态下二人零和连续时间微分对策系统 (22) 和 (23), 如果存在一对控制策略 (u^*,w^*), 并且存在满足 C^1 函数的值函数 $V(x,t)$, 那么, 该微分对策系统的 Nash 均衡解满足

$$-\frac{\partial V(x,t)}{\partial t}=\inf_{u\in U}\sup_{\omega\in W}\left\{\frac{\partial V(x,t)}{\partial x}f(x,u,\omega,t)+L(x,u,\omega,t)\right\}=\sup_{\omega\in W}\inf_{u\in U}\left\{\frac{\partial V(x,t)}{\partial x}f(x,u,\omega,t)+L(x,u,\omega,t)\right\}=\frac{\partial V(x,t)}{\partial x}f(x,u^*,\omega^*,t)+L(x,u^*,\omega^*,t)\tag{24}$$

其中, $V(x(t_f),t_f)=\varphi(x(t_f),t_f)$.

定理 4 中的方程 (24) 为著名的 Isaacs 方程或者 Hamilton-Jacobi-Isaacs (HJI) 方程. HJI 方程对于连续时间非线性系统 H_∞ 控制问题求解和推导中起到非常重要的作用.

3　基于微分对策的非线性系统鲁棒控制

在非线性鲁棒控制中, 得到广泛关注的是二人零和微分对策的最小最大 H_∞ 优化控制[8]. 其中控制器的设计是最小化一方局中人, 而干扰抑制是满足 Nash 均衡最大化一方局中人. 在线性系统中, 二人零和微分对策具有无限时间二次代价函数, Nash 均衡解的确定等价于求泛化代数 Riccati 方程. 但是, 对于非线性微分对策系统, 需要求解 HJI 偏微分方程, 该方程的解析解通常是棘手和非光滑的.

基于微分对策理论, 文献 [23−32] 研究了非线性系统的鲁棒控制与最优控制.

3.1　非线性微分对策与 L_2 增益

考虑以下非线性连续时间系统

$$\begin{cases}\dot{x}=f(x,u,\omega)\\ y=h_y(x,\omega)\\ z=h_z(x,u)\end{cases}\tag{25}$$

和非线性离散系统

$$\begin{cases}x(k+1)=f(x(k),u(k),\omega(k))\\ y(k)=h_y(x(k),\omega(k))\\ z(k)=h_z(x(k),u(k))\end{cases}\tag{26}$$

其中, $x\subseteq\mathbf{R}^n$ 是系统的状态向量, 初始状态 $x(0)=x_0$, $u\subset\mathbf{R}^p$ 和 $\omega\subset\mathbf{R}^r$ 分别是控制输入和干扰信号; f: $U\times W\to\mathbf{R}^n$ 是光滑的 C^∞ 函数, $h_y:\omega\subset\mathbf{R}^m$ 和 $h_z:u\to\mathbf{R}^s$ 是光滑函数; y 是系统的可测量输出, $z\in\mathbf{R}^s$ 是表示系统跟踪误差或固定参考位置的控制输出; 假设系统是零状态可检测, 初始状态 $x=0$ 是系统的平衡点, 即 $f(0,0,0)=0$, $h_2(0,0)=0$.

定义 3[23−24]**.** 对于任意初始状态 x_0, 存在一反馈控制 u, 在时间 $[t_0,T)$ 上, 如果系统 (25) 存在从 ω 到 z 的 L_2 增益, 对于任意扰动 $\omega\in L_2$ 所对应的输出响应 z 满足

$$\int_{t_0}^{T}\|z(t)\|^2\mathrm{d}t\le\gamma^2\int_{t_0}^{T}\|\omega(t)\|^2\mathrm{d}t+d(x_0),\quad\forall T>t_0\tag{27}$$

其中, 有界函数 d 满足 $d(0)=0$, 那么该非线性系统具有小于或者等于 γ 的 L_2 增益. 对于非线性离散系统 (26), 在时间 $[0,K)$ 上, 如果对于任意扰动 $\omega(k)\in L_2$ 所对应的输出响应 $z(k)$ 满足

$$\sum_{k=0}^{K}\|z(k)\|^2\le\gamma^2\sum_{k=0}^{K}\|\omega(k)\|^2+d(x_0)\tag{28}$$

就称为该非线性离散系统具有小于或者等于 γ 的 L_2 增益. 其中 $\|\cdot\|$ 表示在 $\mathbf{R}^n$ 上的 Euclidean 范数.

如果把控制 u 和干扰 ω 看作是二人零和微分对策的局中人, 其中控制 u 需要最小化, 而扰动 ω 需要最大化, 那么该二人零和微分对策鞍点平衡点存在的充分必要条件是存在以下值函数.

连续状态

$$V^c(x,t)=\inf_{u\in U}\sup_{\omega\in L_2}\int_t^T\left[\|z(\tau)\|^2-\gamma^2\|\omega(\tau)\|^2\right]\mathrm{d}\tau\tag{29}$$

离散状态

$$V^d(x,k)=\inf_{u\in U}\sup_{\omega\in L_2}\sum_{k=0}^{K}\left[\|z(k)\|^2-\gamma^2\|\omega(k)\|^2\right]\tag{30}$$

满足以下 HJI 方程 (Isaacs 方程)[8].

连续状态

$$\begin{cases}-\dfrac{\partial V^c(x,t)}{\partial t}=\inf\limits_u\sup\limits_\omega\left\{\dfrac{\partial V^c(t,x)}{\partial x}f(x,u,\omega)+\right.\\ \qquad\left.\left[\|z(t)\|^2-\gamma^2\|\omega(t)\|^2\right]\right\}\\ V^c(T,x)=0\end{cases}\tag{31}$$

离散状态

$$\begin{cases}V^d(x,k)=\inf\limits_{u(k)}\sup\limits_{\omega(k)}\left\{V^d(f(x,u(k),\omega(k)),\right.\\ \qquad k+1)+\left.\left[\|z(k)\|^2-\gamma^2\|\omega(k)\|^2\right]\right\}\\ V^d(K+1,x)=0\end{cases}\tag{32}$$

在反馈完全信息条件下, 上述连续状态与离散状态下二人零和非线性微分对策鞍点存在的充分必要条件是存在一对最优策略 (u^*,ω^*), 满足

$$J(u^*,\omega)\le J(u^*,\omega^*)\le J(u,\omega^*)\tag{33}$$

为了进一步分析基于微分对策的非线性系统的最优控制和最坏干扰情况下的 H_∞ 控制问题, 连续状态和离散状态下仿射非线性系统分别为

$$\begin{cases}\dot{x}=f(x)+g_1(x)\omega+g_2(x)u\\ z=h_1(x)+d_{12}(x)u\\ y=h_2(x)+d_{21}(x)\omega\end{cases}\tag{34}$$

$$\begin{cases}x(k+1)=f(x(k))+g_1(x(k))\omega(k)+\\ \qquad g_2(x(k))u(k)\\ z(k)=h_1(x(k))+d_{12}(x(k))u(k)\\ y(k)=h_2(x(k))+d_{21}(x(k))\omega(k)\end{cases}\tag{35}$$

其中, $x(0)=0$, u 为控制输入, ω 为干扰信号; d_{12}, $d_{21}\in C^r$, $r\ge2$ 为适当维数的函数向量或者函数矩阵.

一般意义下, 对于无限时间最优控制问题, 连续系统的控制策略应满足 $\lim_{T\to\infty}J^c(u,\omega)$, 离散系统为 $\lim_{K\to\infty}J^d(u(k),\omega(k))$ 存在和有界.

如果非线性系统存在有限的 L_2 增益, 那么需要寻找一个不受时间约束的半正定存储函数 $V(x)$ 满足以下 HJI 方程.

连续系统 HJI 方程

$$\min_u\sup_\omega\left\{\frac{\partial V^c(x)}{\partial x}[f(x)+g_1(x)\omega(t)+\right.$$
$$\left.g_2(x)u(t)]+\frac{1}{2}\left(\|z(t)\|^2-\gamma^2\|\omega(t)\|^2\right)\right\}=0\tag{36}$$

离散系统 HJI 方程

$$V^d(x)=\min_u\sup_\omega\left\{V^d(f(x)+g_1(x)\omega+\right.$$
$$\left.g_2(x)u)+\frac{1}{2}\left(\|z(k)\|^2-\gamma^2\|\omega(k)\|^2\right)\right\}=0\tag{37}$$

其中, $V(0)=0$.

如果非线性系统是耗散系统, 那么最坏干扰情况下基于二人零和微分对策的非线性 H_∞ 控制应该满足以下耗散 HJI 不等式.

连续状态

$$\inf_{u\in U}\sup_{\omega\in W}\left\{\frac{\partial V^c(x)}{\partial x}[f(x)+g_1(x)\omega+g_2(x)u]+\right.$$

$$\frac{1}{2}\left(\|z\|^2-\gamma^2\|\omega\|^2\right)\Bigg\}\le 0 \tag{38}$$

离散状态

$$\inf_{u\in U}\sup_{\omega\in W}\Bigg\{V^d(f(x)+g_1(x)\omega(k)+g_2(x)u(k))+\frac{1}{2}\left(\|z(k)\|^2-\gamma^2\|\omega(k)\|^2\right)\Bigg\}\le V(x) \tag{39}$$

其中, $V(x)$ 为半正定存储函数.

3.2 非线性混合 H_2/H_∞ 控制

对于非线性系统 (34) 和 (35), 基于二人非零和微分对策理论, 其混合 H_2/H_∞ 控制系统具有如下代价函数.

连续状态

$$\begin{cases}J_1(u,\omega)=\displaystyle\int_{t_0}^{T}\left(\gamma^2\|\omega(\tau)\|^2-\|z(\tau)\|^2\right)\mathrm{d}\tau\\ J_2(u,\omega)=\displaystyle\int_{t_0}^{T}\|z(\tau)\|^2\mathrm{d}\tau\end{cases} \tag{40}$$

离散状态

$$\begin{cases}J_1(u,\omega)=\displaystyle\sum_{k=0}^{K}\left(\gamma^2\|\omega(k)\|^2-\|z(k)\|^2\right)\\ J_2(u,\omega)=\displaystyle\sum_{k=0}^{K}\|z(k)\|^2\end{cases} \tag{41}$$

式 (40) 与式 (41) 都是由两个方程构成, 第一个方程是 H_∞ 约束准则, 第二个方程是与系统输出有关的 H_2 性能准则. 如果使 $J_1\ge 0$, 那么满足 H_∞ 约束的条件是非线性系统具有小于或者等于 γ 的增益; 此时, 如果最小化 J_2, 那么可以实现 H_2/H_∞ 控制[19, 27−28].

假设控制 $U\subset L_2([0,\infty),\mathbf{R}^r)$, 那么在闭环完全信息状态下, 二人非零和微分对策的 Nash 均衡解存在的充分必要条件是存在一对策略 (u^*,ω^*), 满足

$$J_1(u^*,\omega^*)\le J_1(u^*,\omega),\ \forall\omega\in W \tag{42}$$

$$J_2(u^*,\omega^*)\le J_1(u,\omega^*),\ \forall u\in U \tag{43}$$

式 (42) 与式 (43) 所构成微分对策有解的充分必要条件是满足以下耦合 HJI 方程.

连续状态

$$\begin{cases}-\dfrac{\partial Y(x,t)}{\partial t}=\inf\limits_{\omega\in W}\left\{\dfrac{\partial Y(x,t)}{\partial x}f(x,u^*(x),\omega(x))+\gamma^2\|\omega(x)\|^2-\|z^*(x)\|^2\right\};\ Y(x,T)=0\\ -\dfrac{\partial V(x,t)}{\partial t}=\inf\limits_{u\in U}\left\{\dfrac{\partial V(x,t)}{\partial x}f(x,u(x),\omega^*(x))+\|z^*(x)\|^2\right\}=0;\ V(x,T)=0\end{cases} \tag{44}$$

离散状态

$$\begin{cases}Y(x,k)=\min\limits_{\omega_k\in W}\left\{Y(f_k(x,u_k^*,\omega_k,k+1))+\gamma^2\|\omega_k\|^2-\|z_k^*\|^2\right\};\ Y(x,K+1)=0\\ V(x,k)=\min\limits_{u_k\in U}\left\{V(f_k(x,u_k,\omega_k^*,k+1))+\|z_k^*\|^2\right\};\ V(x,K+1)=0;\ k=0,\cdots,K\end{cases} \tag{45}$$

其中, Y 为负定函数, V 为正定函数. 对于耦合函数 Y 和 V 分析和计算十分复杂, 可参阅文献 [17, 19, 27−28].

3.3 非线性连续系统 H_∞ 控制

对于仿射非线性系统

$$\begin{cases}\dot{x}=f(x)+g_1(x)\omega+g_2(x)u\\ y=x\\ z=h_1(x)+d_{12}(x)u\end{cases} \tag{46}$$

其中, x 是系统的状态向量, $x(0)=x_0$, $u\in U\subseteq\mathbf{R}^p$ 是 p 维控制输入, 干扰信号满足 $\omega\in W\subset L_2$, $y\in\mathbf{R}^n$ 是可以直接测量的状态向量, $z\in\mathbf{R}^s$ 是控制输出, 假设系统零状态可检测.

根据定义 (3) 和微分对策理论, 对于非线性系统 (46), 选择合适的控制 $u^*(\cdot)$ 使系统从 ω 到 z 具有小于 γ 的 L_2 增益问题可以转换为二人零和微分对策: 控制 u 为最小化局中人; 干扰 ω 为最大化局中人[29−31].

对于 $\forall T>t_0$, 其代价函数为

$$\min_{u\in U}\max_{\omega\in W}J(u,\omega)=\frac{1}{2}\int_{t_0}^{T}\left[\|z(t)\|^2-\gamma^2\|\omega(t)\|^2\right]\mathrm{d}t \tag{47}$$

基于微分对策理论, 非线性状态反馈鲁棒控制可以转换为以下两个问题: 1) 实现干扰抑制; 2) 实现渐近稳定.

对于第 1 个问题, 假设存在最坏条件下干扰 ω, 那么根据当前状态信息构建反馈控制器 u, 使得性能函数 $J(u,\omega)$ 最小化.

根据微分对策连续系统 (46), 定义值函数为

$$V(x,t)=\inf_{u}\sup_{\omega}\frac{1}{2}\int_{t}^{T}\left[\|z(\tau)\|^{2}-\gamma\|\omega(\tau)\|^{2}\right]\mathrm{d}\tau \tag{48}$$

定理 5[27−28]**.** 考虑二人零和非线性微分对策系统 (46) 具有代价函数 (47), 在反馈信息状态下, 如果存在一对最优策略 $[u^{*}(x,t),\omega^{*}(x,t)]$, 系统鞍点满足

$$J(u^{*},\omega)\leq J(u^{*},\omega^{*})\leq J(u,\omega^{*}) \tag{49}$$

的充分必要条件是存在 C^{1} 函数 V 满足以下 HJI 偏微分方程

$$\begin{aligned}&-\frac{\partial V^{c}(x,t)}{\partial t}=\min_{u}\sup_{\omega}\left\{\frac{\partial V^{c}(x,t)}{\partial x}[f(x)+\right.\\&\left.g_{1}(x)\omega+g_{2}(x)u]+\frac{1}{2}\left(\|z\|^{2}-\gamma^{2}\|w\|^{2}\right)\right\}=\\&\sup_{\omega}\min_{u}\left\{\frac{\partial V^{c}(x,t)}{\partial x}[f(x)+g_{1}(x)\omega+\right.\\&\left.g_{2}(x)u]+\frac{1}{2}\left(\|z\|^{2}-\gamma^{2}\|\omega\|^{2}\right)\right\}=\\&\frac{\partial V^{c}(x,t)}{\partial x}[f(x)+g_{1}(x)\omega^{*}(x,t)+\\&g_{2}(x)u^{*}(x,t)]+\frac{1}{2}\|h_{1}(x)+d_{12}(x)u^{*}(x,t)\|^{2}-\\&\frac{1}{2}\gamma^{2}\|w^{*}(x,t)\|^{2}\end{aligned} \tag{50}$$

其中, $V(x,T)=0$.

其次, 为了求得合适的反馈策略 (u^{*},ω^{*}) 满足 Isaccs 方程, 那么需要构建一个 Hamiltonian 函数

$$\begin{aligned}H(x,\lambda,u,\omega)=&\lambda^{\mathrm{T}}(f(x)+g_{1}(x)\omega+g_{2}(x)u)+\\&\frac{1}{2}\|h_{1}(x)+d_{12}(x)u\|^{2}-\frac{1}{2}\gamma^{2}\|\omega\|^{2}\end{aligned} \tag{51}$$

求得唯一鞍点 (u^{*},ω^{*}) 满足

$$H(x,\lambda,u^{*},\omega)\leq H(x,\lambda,u^{*},\omega^{*})\leq H(x,\lambda,u,\omega^{*}) \tag{52}$$

其中, λ 是伴随向量.

对于非线性系统 (46), 令 $d_{12}^{\mathrm{T}}(x)d_{12}(x)=I$, $h_{1}^{\mathrm{T}}(x)d_{12}(x)=0$, 那么, 二人零和非线性微分对策系统的 HJI 方程为

$$\begin{aligned}&\frac{\partial V(x)}{\partial x}f(x)+\frac{1}{2}\frac{\partial V(x)}{\partial x}\Big[\frac{1}{\gamma^{2}}g_{1}(x)g_{1}^{\mathrm{T}}(x)-\\&g_{2}(x)g_{2}^{\mathrm{T}}(x)\Big]\frac{\partial V^{\mathrm{T}}(x)}{\partial x}+\frac{1}{2}h_{1}^{\mathrm{T}}(x)h_{1}(x)=0\end{aligned} \tag{53}$$

其中, $V(\cdot)$ 是满足 HJI 方程的半正定解, $V(0)=0$. 那么此时最优反馈策略为

$$\begin{cases}u^{*}(x)=-g_{2}^{\mathrm{T}}(x)\dfrac{\partial V^{\mathrm{T}}(x)}{\partial x}\\[2mm]\omega^{*}(x)=\dfrac{1}{\gamma^{2}}g_{1}^{\mathrm{T}}(x)\dfrac{\partial V^{\mathrm{T}}(x)}{\partial x}\end{cases} \tag{54}$$

对于第 2 个问题, 如果闭环系统渐近稳定, 那么令

$$\alpha(x)=u^{*}(x)=-g_{2}^{\mathrm{T}}(x)V_{x}^{\mathrm{T}}(x) \tag{55}$$

当在 $\omega=0$ 时, 对 $V(\cdot)$ 沿着闭环系统状态求导, 可以得到

$$\begin{aligned}\dot{V}(x)=&\frac{\partial V(x)}{\partial x}\left(f(x)-g_{2}(x)g_{2}^{\mathrm{T}}(x)\frac{\partial V(x)}{\partial x}\right)=\\&-\frac{1}{2}\|u^{*}\|^{2}-\frac{1}{2}\gamma^{2}\|\omega^{*}\|^{2}-\frac{1}{2}h_{1}^{\mathrm{T}}(x)h_{1}(x)\leq 0\end{aligned} \tag{56}$$

因此, 该二人零和微分对策在 Lyapunov 意义下稳定.

3.4 非线性离散系统 H_{∞} 控制

仿射非线性离散系统为

$$\begin{cases}x(k+1)=f(x(k))+g_{1}(x(k))\omega(k)+\\\qquad g_{2}(x(k))u(x(k))\\z(k)=h_{1}(x(k))+d_{11}(x(k))\omega(k)+\\\qquad d_{12}(x(k))u(k)\\y(k)=h_{2}(x(k))+d_{21}(x(k))\omega(x(k))\end{cases} \tag{57}$$

其中, x 是系统的状态向量, 初始状态 $x(0)=x_{0}$, $u\in\mathbf{R}^{p}$ 是 p 维控制输入, 干扰信号 $\omega\in W\subset L_{2}$, $y\in\mathbf{R}^{n}$ 是可以直接测量的状态向量, $z\in\mathbf{R}^{s}$ 是控制输出, 假设系统零状态可检测.

仿射非线性离散系统 (57) 的 H_{∞} 控制问题也可以转换为二人零和微分对策问题[32−36]. 其有限时间代价函数为

$$J(u,\omega)=\frac{1}{2}\sum_{k=0}^{K}\left(\|z(k)\|-\lambda^{2}\|\omega(k)\|^{2}\right) \tag{58}$$

其中, 二局中人的策略分别是最小化控制函数 $u(k)=\alpha_{2}(x(k))$ 和最大化扰动函数 $\omega(k)=\alpha_{1}(x(k))$. 因此, 存在以下定理.

定理 6[26, 32−33]**.** 考虑二人零和离散微分对策系统 (57), 在完全信息状态下, 该系统的最优控制策略 $u^*(x(k))$ 和 $\omega^*(x(k))$ 所构成的反馈鞍点满足

$$J(u^*,\omega)\le J(u^*,\omega^*)\le J(u,\omega^*),\quad \forall u\in U,\ \forall \omega\in W \tag{59}$$

的充分必要条件是: 当且仅当存在 K 个函数 $V_k(\cdot):[0,K]\to\mathbf{R}$ 满足离散时间 HJI 方程

$$\begin{aligned}V_k(x)=&\min_{u_k\in U}\max_{\omega_k\in W}\left\{\frac{1}{2}\left(\|z_k\|^2-\gamma^2\|\omega_k\|^2\right)+\right.\\&\left.V_{k+1}(f(x,u_k,\omega_k))\right\}=\\&\max_{\omega_k\in W}\min_{u_k\in U}\left\{\frac{1}{2}\left(\|z_k\|^2-\gamma^2\|\omega_k\|^2\right)+\right.\\&\left.V_{k+1}(f(x,u_k,\omega_k))\right\}=\\&\frac{1}{2}\left(\|z_k(x,u_k^*,\omega_k^*)\|^2-\gamma^2\|\omega_k^*\|^2\right)+\\&V_{k+1}(f(x(k))+g_1(x(k))\omega_k^*+g_2(x(k))u_k^*)\end{aligned} \tag{60}$$

其中, $V_{K+1}(x(k))=0$.

由于本文主要研究时不变控制系统, 那么当 $k\to\infty$ 时, 时不变离散 Isaac 方程转换为

$$\begin{aligned}&V(f(x(k))+g_1(x(k))\omega_k^*+g_2(x(k))u_k^*)-\\&V(x(k))+\frac{1}{2}\left(\|z(x,u_k^*,\omega_k^*)\|^2-\gamma^2\|\omega_k^*\|^2\right)=0\end{aligned} \tag{61}$$

其中, $V(0)=0$. 控制器的设计可参考文献 [8, 11, 17].

4　非线性微分对策的均衡理论及主要算法

在非线性系统中, 微分对策理论为鲁棒控制和最优控制的发展提供一个自然扩展[1−8]. 对于多局中人参与的非线性控制系统, 不同的局中人都在计算和优化各自的性能函数. 此时, 控制器设计是确定一个容许控制策略, 来保证系统的稳定和最小化各自性能函数, 并最终产生均衡解[7]. 在微分对策中, 最优性是以具体的均衡表现出来: Pareto 均衡、Nash 均衡和 Stackelberg 均衡等.

4.1　Pareto 均衡

如果局中人策略集合中的子集能使决策一致, 并且具有共同的利益, 那么可以得到一个相互有利的结果, 因此便可获得合作型微分对策. 在这种情况下, 局中人通过合作进行行动, 就可以获得最优策略. 合作微分对策即局中人形成具有约束力的协议, 从而形成微分对策整体最优, 即达到 Pareto 均衡[37−39].

定义代价函数 $J_i(t,x,u_1,\cdots,u_N)$ 为

$$J_i=\int_0^T L(\tau,x(\tau),u_1(\tau),\cdots,u_N(\tau))\mathrm{d}\tau \tag{62}$$

其中, $i=1,\cdots,N$. 系统的状态方程为

$$\dot{x}=f(t,x(t),u_1(t),\cdots,u_N(t)),\ x(0)=x_0 \tag{63}$$

定义 4[39]**.** 令 $\alpha_i\in(0,1)$, 如果存在一个参数集合

$$\Phi=\left\{\alpha=(\alpha_1,\cdots,\alpha_N)\ \middle|\ \alpha_i\ge 0;\sum_{i=1}^N\alpha_i=1\right\} \tag{64}$$

使得 $u^*\in U$ 满足

$$u^*\in\arg\min_{u\in U}\left\{\sum_{i=1}^N\alpha_iJ_i(u)\right\} \tag{65}$$

那么就称 u^* 为 Pareto 有效解.

对于 Pareto 有效解, 存在最优控制策略 u^* 满足不等式

$$J_i(u)\le J_i(u^*),\ i=1,\cdots,N \tag{66}$$

相应的 $(J_1(u^*),\cdots,J_N(u^*))$ 即为 Pareto 优化解, 所有 Pareto 解的集合称为 Pareto 边界. Pareto 最优策略为 $u^*=\arg\min_{u\in U}\{\sum_{i=1}^N\alpha_iJ_i(u)\}$.

定理 7[38]**.** 对于式 (62) 和式 (63) 所构成的非线性 N 人微分对策系统, 其 Pareto 优化解为 $(J_1(u^*),\cdots,J_N(u^*))$, 那么存在一连续可微共态函数 $\lambda^{\mathrm{T}}(t)$

$$H(t,x,u,\lambda)=\sum_{i=1}^N\alpha_iL_i(t,x,u)+\lambda f(t,x,u) \tag{67}$$

使得最优策略 u^* 满足

$$\begin{cases}\dot{x}^*(t)=f(t,x^*(t),u_1^*(t),\cdots,u_N^*(t))\\H(t,x^*,u^*,\lambda)\le H(t,x^*,u,\lambda)\\\dot{\lambda}(t)=-\left(\sum\limits_{i=1}^N\alpha_i\dfrac{\partial L_i}{\partial x}+\lambda(t)\dfrac{\partial f}{\partial x}\right),\ x^*(0)=x_0\end{cases} \tag{68}$$

4.2　Nash 均衡

一个 Nash 微分对策由多局中人同时做出决策, 并且整个最优策略结果不能因每一个局中人的策略的改变而改变[14].

基于系统的初始状态、模型和最小化代价函数, 局中人需要执行预先给定的策略. N 人微分对策系统 (63) 和代价函数 (62), 如果存在容许控制策略集合 $u_i^* \in U_i$, $i \in N$ 满足以下 N 个不等式

$$\begin{cases} J_1^* = J_1\left(x\left(t\right), u_1^*, u_2^*, \cdots, u_N^*\right) \leq \\ \qquad J_1\left(x\left(t\right), u_1, u_2^*, \cdots, u_N^*\right) \\ J_2^* = J_2\left(x\left(t\right), u_1^*, u_2^*, \cdots, u_N^*\right) \leq \\ \qquad J_2\left(x\left(t\right), u_1^*, u_2, \cdots, u_N^*\right) \\ \qquad \vdots \\ J_N^* = J_N\left(x\left(t\right), u_1^*, u_2^*, \cdots, u_N^*\right) \leq \\ \qquad J_N\left(x\left(t\right), u_1^*, u_2^*, \cdots, u_N\right) \end{cases} \tag{69}$$

那么该控制集合 $\left(u_1^*, \cdots, u_N^*\right)$ 为 N 人非零和微分对策系统的 Nash 均衡[4, 7].

文献 [14, 40−41] 研究了非零和微分对策中 Nash 均衡非唯一性问题. 在开环非零和微分对策系统中, 如果每个局中人在时间 $t \in [0, T]$ 上都知道自己的初始状态 x_0, 并给出了唯一 Nash 平衡解存在的条件. 在闭环完全信息状态下, 每个局中人在时间 $t \in [0, T]$ 上都知道系统的完整历史状态, 那么系统将存在无限多 Nash 均衡[41]. 在这种情况下, 把 Nash 平衡解限定到反馈解的子集中, 因此称为反馈 Nash 平衡. 文献 [14] 和文献 [41] 同时通过求解 Hamilton-Jacobi 不等式获得值函数, 最终根据反馈策略给出了微分对策完美 Nash 均衡, 该算法已经成功地运用于线性二次微分对策系统中[16]. 在控制理论中, Nash 均衡的一个特例是在未知动态最坏情况下最小最大 (min-max) 鞍点使系统达到平衡[8].

鞍点平衡已经广泛运用于 H_∞ 控制理论中. 如果存在最小增益 $\gamma \geq 0$, 其代价函数

$$J_\gamma\left(u, \omega\right) = \int_0^\infty \left(Q\left(x_t\right) + u\left(x_t\right)^2 - \gamma^2 \left\|\omega\left(x_t\right)\right\|^2\right) \mathrm{d}t \tag{70}$$

有界, 那么该控制问题转化为设计相应控制器以确定该界值. 与二次微分对策相关的 H_∞ 控制问题在某种意义上等价于最差情况对目标函数 (70) 确定上确界和下确界, 并最终通过二次微分对策系统中鞍点的求解来获得该有界值. 基于二人零和微分对策的 H_∞ 控制的目标是把控制器作为局中人最小化, 把干扰作为局中人的另一方最大化[18, 21].

4.3 Stackelberg 均衡

分层非零和微分对策起源于 Von Stackelberg, 即当一个局中人的策略影响到其他局中人的策略时怎样确定均衡解. Stackelberg 方法现在已经广泛应用于解决一大类分层决策控制问题[42]. Stackelberg 策略是一种求解二人非零和对策的方法.

对于式 (62) 并有式 (63) 代价函数约束的二人零和微分对策系统, 令 $N = 2$. 二局中人控制策略分别为 u_1 和 u_2, 局中人 1 的代价函数为 $J_1\left(u_1, u_2\right)$, 局中人 2 的代价函数为 $J_2\left(u_1, u_2\right)$. 局中人 1 为领导者 (Leader), 局中人 2 为跟随者 (Follower). 局中人 1 知道选择控制策略 u_1 后, 局中人 2 的执行策略 $u_2 = T_2\left(u_1\right)$, 其中 T_2 为从 u_1 到 u_2 的一一映射.

假设二人零和微分对策中局中人 2 知道自身控制策略 u_2, 并知道局中人 1 的控制策略 u_1. 二局中人试图选择自己的策略使得各自的代价函数最小, 那么局中人 1 选择控制策略 u_1 后, 局中人 2 选择策略 u_2 将使

$$J_2\left(u_1, T_2\left(u_1\right)\right) \leq J_2\left(u_1, u_2\right) \tag{71}$$

对于局中人 2 的策略 u_2, 局中人 1 选择最优控制策略 u_1^* 后, 使得:

$$J_1\left(u_1^*, T_2\left(u_1^*\right)\right) \leq J_1\left(u_1, T_2\left(u_1\right)\right) \tag{72}$$

那么, 最优策略 u_1^* 称为局中人 1 的 Stackelberg 策略, 而策略 $u_2^* = T_2\left(u_1^*\right)$ 称为当局中人 1 为领导者时局中人 2 的 Stackelberg 策略[43−44].

与 Nash 均衡不同, 领导者使用 Stackelberg 策略的一个重要动机是产生与 Nash 策略不同的代价函数对比值. 因此对于领导者, 其 Stackelberg 策略至少与任意 Nash 策略一样最优[44−45]. 对于动态 Stackelberg 对策理论的求解与算法可参阅文献 [7].

4.4 主要求解算法

基于微分对策理论的非线性系统最优控制与鲁棒控制问题分别等价于非线性系统 HJB 方程与 HJI 方程的求解[8−9, 13]. 分析非线性 H_∞ 控制问题需要大量的计算, 并且需要求解 HJB 方程或者 HJI 方程, 因此在实时系统中的应用通常是不可能的. 由于 HJB 方程和 HJI 方程求解困难, 部分学者试图通过增强学习与动态规划方法研究最优控制与鲁棒控制, 并确定最优解[46−50].

增强学习 (Reinforcement learning, RL) 是通过对环境的评价反馈学习产生控制的学习算法[51]. 一个广泛使用的增强学习算法是基于控制–评价结构 (Action-critic, AC) 的, 其中控制器通过和环境交互产生动作 (控制), 评价器评价动作, 向控制器提供反馈, 引起下一步动作性能的改善. AC 算法在机器学习中很普遍, 在有限空间离散 Markov 决策问题中, AC 算法用于在线学习以获得最优策略[52−53]. 近似动态规划 (Approximate dynamic programming, ADP) 通过离散/迭代特性可以使

ADP 算法很容易设计离散最优控制器来确定最优控制的次优解[54−55].

根据系统的部分特性, 文献 [56] 基于策略迭代 (Policy interactive, PI) 设计了仅连续时间与离散时间混杂系统的采样数据控制器, 该反馈控制器的执行时间比自适应评价学习的时间更快. Vamvoudakis 和 Lewis 扩展了混杂系统模型在线学习方法, 设计了同步策略迭代算法, 该算法实现了控制器与评价器两个神经网络的连续时间上的同步[57]. Bhasin 等设计了连续控制–评价–辨识器 (Actor-Critic-Identifier, ACI) 对无线时间范围内一人微分对策进行控制, 并通过构建鲁棒动态神经网络 (Dynamic neural network, DNN) 辨识系统参数, 评价神经网络来获得近似值函数[58]. 基于文献 [59] 的近似最优控制方法, 文献 [60] 通过构造 Lyapunov 函数, 证明了非零和非线性连续闭环系统一致有界, 并获得次优解.

针对二人零和非线性系统的微分对策及与之相关的 HJB 方程与 HJI 方程的求解, 文献 [61] 采用 ADP 算法设计两个迭代代价函数, 以保证鞍点的收敛序列的性能指标具有上界和下界. 在假设鞍点不存在的情况下, 文献 [62] 设计一个迭代 ADP 算法, 实现了零和微分对策的求解, 并在混合最优控制策略下获得最优性能指标函数. 文献 [63] 提出了一种新的 ADP 迭代算法, 把非仿射非线性二次零和微分对策转换为对等序列的线性二次零和微分对策以获得最优鞍点的近似解. 文献 [64] 运用增强学习理论设计没有完全信息的二人非零和线性微分对策系统的在线学习算法, 以获得最优解. 基于文献 [57] 的同步策略迭代算法, 文献 [18] 深入研究了二人零和微分对策问题. 基于贪婪迭代 HDP (Heuristic dynamic programming) 算法, 文献 [65] 研究了仿射非线性离散系统的零和微分对策问题. 文献 [66] 研究了未知模型的非线性连续时间多人非零和微分对策系统的求解算法问题.

马尔科夫对策 (Markov games, MG) 也称为随机对策, 是将对策论应用到类 Markov 决策过程 (Markov decision processes, MDP) 环境中, 是 MDP 推广到多 Agent 环境下的泛化[67−68]. 马尔科夫对策也可以看作是矩阵对策的概念在多状态下的延伸. MG 理论是研究具有离散时间特性多 Agent 协作的重要理论框架[69]. 基于 Markov 对策理论, Michael 于 1994 年提出了多 Agent 强化学习的理论框架. 针对二人零和对策问题, Littman 首次提出 Minmax-Q 增强学习算法来寻找最优策略[70]. 在此基础上, 文献 [71] 基于增强学习理论研究了 Markov 对策的值函数问题, 并给出了 Nash 平衡解的收敛条件. 针对一般基于数据的控制系统, Markov 对策可用于设计结构简单的直接鲁棒控制器. 该鲁棒控制器在处理噪音和干扰方面优于基于 MDP 的 Q-learning 和传统 H_∞ 的控制方法[71]. 进一步, Hu 与 Wellman 针对 Markov 零和对策研究了 Nash-Q 学习算法, 在该算法中, 每一个阶段的 Nash 平衡点作为每个 Agent 的行动策略, 在满足一定约束条件下保证 Nash-Q 学习的收敛性. 但是, 该算法强加了一个严格限制的假设: 每个 Agent 的决策必须拥有一个独一无二的平衡点, 这对于一般意义下的 MG 零和对策问题未必成立[72].

5 非线性微分对策主要应用与展望

5.1 非线性微分对策主要应用

5.1.1 在经济学领域[73−74]

当今经济社会高速发展, 现实中的经济问题已经并不能由静态博弈所刻画和描述, 静态博弈框架下的最优均衡无法保证经济的顺利进行. 近几十年来, 世界经济运行出现了许多超预期的新变化、新趋势和新规律, 这不仅让传统经济学理论和模型失去了应有的解释力, 也让决策者在应对一系列复杂问题以及危机治理方面表现得十分乏力, 而微分对策理论为经济学的进一步发展提供了更广阔的视角和更加科学的方法. 当前, 基于对策论和演化经济学的理论更加关注微观, 关注宏观决策中常常被忽视的“个体”, 更加关注经济系统变量之间的作用机制, 关注经济演变的过程而不是结果. 自 1994 年普林斯顿大学 Nash 等 3 位博弈论专家被授予诺贝尔经济学奖以来, 至今共有 6 届诺贝尔经济学奖与博弈论的研究有关: 1996 年、2001 年、2005 年、2007 年和 2012 年. 作为一门工具学科能够在经济学中如此广泛运用并得到科学界垂青实为罕见.

5.1.2 在计算机科学与信息领域[75−79]

2006 年 10 月 27 日剑桥大学计算机系 Ross Anderson 和 TylerMoore 在《科学》杂志上刊发题为《信息安全经济学》(*The Economics of Information Security*) 一文, 标志着一个新兴的学科领域: 信息安全经济学的产生, 它属于普适计算安全领域与计算机理论应用的交叉方向. 传统信息安全通常将用户分成两类: 诚实用户和恶意用户, 但在垃圾邮件、自私实体的 TCP 效应、建立路由、网络创建等领域的失效往往是由理性且自私、没有恶意的“策略用户”引起. 与以往信息安全技术限制非法用户的使用不同, 基于微分对策理论的信息安全的使用安全机制通过影响系统参与用户的策略选择, 使用户之间相互影响和制约以完成系统的预定目标.

计算机经济学不仅包括将商业活动信息化的传统电子商务领域, 还包括利用微分对策、微观经

济学等理论解决计算机科学中所遇到的问题, 计算机经济学也被称作算法博弈论 (Algorithmic game theory). 算法博弈论作为计算机理论科学的一个新领域, 重点关注并解决有关拍卖、网络和人类行为的根本问题. 算法博弈论的研究大致包括以下几个方面: 1) 研究各种均衡 (如 Nash 均衡、子博弈 Nash 均衡、Pareto 均衡等) 的计算复杂性问题; 2) 从博弈论的观点研究计算机学科中的许多问题; 3) 算法机制设计领域; 4) 计算性社会选择问题. 基于微分对策理论的计算机经济学将成为今后计算机科学与技术领域非常热的一个研究方向.

5.1.3 在生物学领域[80−83]

生物群体现象是自然界中常见的, 例如编队迁徙的鸟群、结对巡游的鱼群、协同工作的蚂蚁等. 这些现象的共同特征是拥有一定数量的自主个体通过相互联系、相互合作和自组织组成的网络系统, 在集体层面上呈现出有序的协同运动的动力学行为. 基于微分对策的多智能体合作控制问题可以描述为: 根据某些特定的任务、种类或性能要求, 设计单个智能体的控制策略, 使其通过与其他智能体的信息交互和共同约定的简单相互作用规则, 达到某些关键量的一致或共享. 多智能体合作控制具有分布式特点, 在个体层面上要求每个智能体具有有限信息采集、计算和通讯功能, 而在集体层面上则表现出复杂的协同配合智能行为, 并实现单个智能体不能完成的工作. 多智能体的合作目标是指它们所呈现的群体现象或特征, 这些群体特征是根据智能体所处的局部环境而设计的算法或控制律. 多智能体合作的群体目标主要有: 群集 (Swarming)、蜂涌 (Flocking)、集结 (Aggregation)、聚集 (Rendezvous)、编队 (Formation) 等.

5.2 非线性微分对策研究展望

1) 对于非合作微分对策, 由于每个局中人只追求自己利益的最大化, 从而不可避免地与其他局中人发生利益冲突. 利益冲突的最极端形式是零和微分对策, 即局中人之间的利益完全对立. 非合作微分对策均衡解即设计控制策略, 为利益相互冲突的局中人寻找了一种平衡点. 现有的研究成果大都集中在一个有限时间范线性二次微分对策, 但是对于 Nash 均衡的主要问题即一般意义下 Nash 平衡点的确定研究的比较少.

2) 对于特定类型的多人微分对策系统 (主要是非零和微分对策), 最小化代价函数的结果是对一对耦合 HJB 方程求解. 大量的文献致力于确定线性动态微分对策解唯一性的条件, 尤其是在线性二次代价函数的微分对策领域. 对于非线性系统, 耦合 HJB 方程解的存在性和唯一性的推导是非常困难的 (稀疏性).

3) Pareto 最优性原理是针对合作型微分对策, Pareto 优化解是基于一个前提: 任何一个特定局中人的代价不是单独唯一确定, 其最优解是当所有局中人的代价同时得不到提升才能确定. 在控制理论中, Pareto 优化解的方法是求解参数化的最优控制, 然而, 该方法是否能产生所有的 Pareto 解有待进一步深入研究.

4) 控制理论中开环 Stackelberg 对策的早期研究主要致力于推导一个解析解并提供增益约束, 但是这些结果受限于已知被控对象的线性系统, 并且没有证明控制律的稳定性.

5) 马尔科夫对策模型是描述多 Agent 系统的一种常用数学模型, 可以清楚地描述多 Agent 学习中交互的本质. 在基于微分对策的非线性鲁棒控制领域, 马尔科夫对策模型分为零和对策与非零和对策. 两人零和对策可以使用极大极小 Q 学习算法求取最优解, 但是极大极小 Q 学习算法只能解决对策双方具有对抗性质的问题, 同时状态量不能过于复杂, 非零和对策的适用范围更具有普遍性, 因此解决该类问题的模型都具有局限性.

6) 多智能体合作控制与动态优化是一类典型的矛盾求解问题, 因此, 微分对策理论为多智能体动态优化问题的求解提供了极其合适的工具. 基于非合作微分对策理论的多智能体动态优化, 就是把每个智能体 (博弈的参与人) 根据其可觉察的环境和自身的利益进行决策 (控制), 智能体的最优代价不仅取决于自己的行为选择, 而且受到其他智能体的行为的影响. 基于合作微分对策理论的多智能体动态优化, 就是假定所有智能体有一个可实施的共同行动协议, 在冲突和利益一致的群体决策与管理过程中, 达成各方共同认可的有约束力的合作协议, 而有约束力的协议的达成是通过智能体之间的有效的合作决策来实现的.

7) 1994 年诺贝尔经济学奖得主 Nash 证明每个博弈必定存在一个纳什均衡点, 其他经济学家则推测纳什均衡点极难找到, 但如果找到的话, 它将能精确描述市场行为. 中国学者 Chen 与 Deng 对 Nash 均衡进行了深入研究, 证明了二人 Nash 均衡是 PPAD (Polynomial parity arguments on directed graphs) 完全问题, 并因此获得 2006 年度 IEEE FOCS 最佳论文[84]. 目前工作在麻省理工学院电机工程和计算机科学系的助理教授 Daskalakis 在他的博士论文中证明了纳什均衡属于 NP (Non-deterministic polynomial) 问题的一个子集, 不是通常认为的 NP 完全 (NP-complete) 问题, 而是 PPAD 完全问题[85]. 该博士论文因此获得 2008 年度美国计算机协会最佳学位论文奖. Daskalakis 并

认为对于某些特殊的博弈, 计算机找不到 Nash 均衡点, 人类也不可能最终找到.

6 结论

本文基于微分对策理论研究了非线性控制问题, 介绍了非线性微分对策的基本定理、定义以及连续与离散状态下的 HJI 方程; 总结了基于微分对策理论的非线性鲁棒控制与最优控制研究成果; 介绍了微分对策理论的 Pareto 均衡、Nash 均衡和 Stackelberg 均衡理论; 分析了微分对策理论在经济领域、计算机科学领域和生物学领域的相关应用及当前研究进展. 经历了 50 年左右的发展, 微分对策理论已经广泛地应用于各学科, 并且已经取得重要研究成果. 但是至今, 基于微分对策理论的非线性的鲁棒控制与最优控制的研究仍然具有挑战性, 求解算法需要进一步创新, 已有均衡理论有待进一步完善, 理论上的突破需要进一步加强.

References

1 Nian Xiao-Hong, Huang Lin. New development on differential game theory and its application. *Control and Decision*, 2004, **19**(2): 128−133
(年晓红, 黄琳. 微分对策理论及其应用研究的新进展. 控制与决策, 2004, **19**(2): 128−133)

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

3 Issacs R. *Differential Games: SIAM Series in Applied Mathematics*. New York: John Wiley and Sons, 1965

4 Friedman A. *Differential Games: Pure and Applied Mathematics Series*. New York: Wiley Interscience, 1971

5 Friedman A. *Differential Games*. Rhode Island: American Mathematical Society, 1974

6 Nash J. Non-cooperative games. *Annals of Mathematics*, 1951, **54**(3): 286−295

7 Basar T, Olsder G J. *Dynamic Noncooperative Game Theory (2nd Edition)*. New York: SIAM, Society for Industrial and Applied Mathematics, 1999

8 Basar T, Bernhard P. *H_∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach (2nd Edition)*. Boston: Birkhäuser Boston Inc., 2008

9 Song Chong-Hui, Bian Chun-Yuan, Zhang Xie, Shi Cheng-Long. Numerical optimization method for HJI equations derived from robust receding horizon control schemes and controller design. *Scientia Sinica Informationis*, 2011, **41**(9): 1156−1170
(宋崇辉, 边春元, 张勰, 史成龙. 鲁棒后退时域控制中 HJI 方程的数值解法及控制器设计. 中国科学: 信息科学, 2011, **41**(9): 1156−1170)

10 Isaacs R. Differential games: their scope, nature, and future. *Journal of Optimization Theory and Applications*, 1969, **3**(5): 283−292

11 Bardi M, Capuzzo-Dolcetta I. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Boston: Birkhäuser Boston Inc., 1997

12 Beard R, Saridis G, Wen J. Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation. *Automatica*, 1997, **33**(12): 2159−2177

13 Sassano M, Astolfi A. Dynamic approximate solutions of the HJ inequality and of the HJB equation for input-affine nonlinear systems. *IEEE Transactions on Automatic Control*, 2012, **57**(10): 2490−2503

14 Frihauf P, Krstic M, Basar T. Nash equilibrium seeking in noncooperative games. *IEEE Transactions on Automatic Control*, 2012, **57**(5): 1192−1207

15 Shamma J S, Arslan G. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 2005, **50**(3): 312−327

16 Engwerda J C. *LQ Dynamic Optimization and Differential Games*. New York: John Wiley and Sons Ltd, 2005

17 Aliyu M D S. *Nonlinear H_∞ Control, Hamiltonian Systems and Hamilton-Jacobi Equations*. New York: CRC Press, 2011

18 Vamvoudakis K G, Lewis F L. Online solution of nonlinear two-player zero-sum games using synchronous policy iteration. *International Journal of Robust and Nonlinear Control*, 2012, **22**(13): 1460−1483

19 Limebeer D J N, Anderson B D O, Hendel B. A Nash game approach to mixed H_2/H_∞ control. *IEEE Transactions on Automatic Control*, 1994, **39**(1): 69−82

20 Liu D R, Wei Q L. Finite-approximation-error based optimal control approach for discrete-time nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2013, **43**(2): 779−789

21 Abu-Khalaf M, Lewis F L, Huang J. Neuro-dynamic programming and zero-sum games for constrained control systems. *IEEE Transactions on Neural Networks*, 2008, **19**(7): 1243−1252

22 Starr A W, Ho Y C. Nonzero-sum differential games. *Journal of Optimization Theory and Applications*, 1969, **3**(3): 184−206

23 Pavel L, Fairman F W. Robust stabilization of nonlinear plants: an L_2-approach. *International Journal of Robust and Nonlinear Control*, 1996, **6**(7): 691−726

24 van der Schaft A J. L_2-gain analysis of nonlinear systems and nonlinear state feedback H_∞-control. *IEEE Transactions on Automatic Control*, 1992, **37**(6): 770−784

25 Lu W M, Doyle J C. H_∞ control of nonlinear systems: a convex characterization. *IEEE Transactions on Automatic Control*, 1995, **40**(9): 1668−1675

26 Lin W, Byrnes C I. H_∞-control of discrete-time nonlinear systems. *IEEE Transactions on Automatic Control*, 1996, **41**(4): 494−509

27 Lin W. Mixed H_2/H_∞-control for nonlinear systems. *International Journal of Control*, 1996, **64**(5): 899−922

28 Chen B S, Chang Y C. Nonlinear mixed H_2/H_∞-control for robust tracking of robotic systems. *International Journal of Control*, 1998, **67**(6): 837−857

29 Isidori A. Feedback control of nonlinear systems. *International Journal of Robust and Nonlinear Control*, 1992, **2**(4): 291−311

30 Isidori A. H_∞ control via measurement feedback for affine nonlinear systems. *International Journal of Robust and Nonlinear Control*, 1994, **4**(4): 553−574

31 Isidori A, Kang W. H_∞ control via measurement feedback for general class of nonlinear systems. *IEEE Transactions on Automatic Control*, 1995, **40**(3): 466−472

32 Lin W, Byrnes C I. Dissipativity, L_2-gain and H_∞-control for discrete-time nonlinear systems. In: Proceedings of the 1994 American Control Conference. Baltimore, Maryland, 1994. 2257−2260

33 Lin W, Byrnes C I. Discrete-time nonlinear H_∞ control with measurement feedback. *Automatica*, 1996, **31**(3): 419−434

34 Guillard H, Monaco S, Normand-Cyrot D. Approximate solutions to nonlinear discrete-time H_∞-control. *IEEE Transactions on Automatic Control*, 1995, **40**(12): 2143−2148

35 Guillard H, Monaco S, Normand-Cyrot D. On H_∞-control of discrete-time nonlinear systems. *International Journal of Robust and Nonlinear Control*, 1996, **6**(7): 633−643

36 James M R, Baras J S. Robust H_∞ output-feedback control for nonlinear systems. *IEEE Transactions on Automatic Control*, 1995, **40**(6): 1007−1017

37 Engwerda J C. The regular convex cooperative linear quadratic control problem. *Automatica*, 2008, **44**(9): 2453−2457

38 Engwerda J C, Salmah S. Necessary and sufficient conditions for Pareto optimal solutions of cooperative differential games. *SIAM Journal on Control and Optimization*, 2010, **48**(6): 3859−3881

39 Reddy P V, Engwerda J C. Pareto optimality in infinite horizon linear quadratic differential games. *Automatica*, 2013, **49**(6): 1705−1714

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

41 Engwerda J C, Salmah S. Feedback nash equilibria for linear quadratic descriptor differential games. *Automatica*, 2012, **48**(4): 625−631

42 von Stackelbe H. *The Theory of the Market Economy*. Oxford: Oxford University Press, 1952

43 Cruz J B. Leader-follower strategies for multilevel systems. *IEEE Transactions on Automatic Control*, 1978, **23**(2): 244−255

44 Cruz J B. Survey of Nash and Stackelberg equilibrium strategies in dynamic games. *Annals Economic and Social Measurement*, 1975, **4**(2): 339−344

45 Papavassilopoulos G P, Cruz J B. Nonclassical control problems and Stackelberg games. *IEEE Transactions on Automatic Control*, 1979, **24**(2): 155−166

46 Al-Tamimi A, Lewis F L, Abu-Khalaf M. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, **38**(4): 943−949

47 Barto A G, Sutton R S, Anderson C W. Neuron-like adaptive elements that can solve difficult learning control problems. *IEEE Transactions on System, Man, and Cybernetic, Part B*, 1983, **13**(5): 834−846

48 Abu-Khalaf M, Lewis F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 2005, **41**(5): 779−791

49 Zhang Ping, Fang Yang-Wang, Hui Xiao-Bin, Liu Xin-Ai, Li Liang. Near optimal strategy for nonlinear stochastic differential games based on the technique of statistical linearization. *Acta Automatica Sinica*, 2013, **39**(4): 390−399
(张平, 方洋旺, 惠晓滨, 刘新爱, 李亮. 基于统计线性化的随机非线性微分对策逼近最优策略. 自动化学报, 2013, **39**(4): 390−399)

50 Zhao Dong-Bin, Liu De-Rong, Yi Jian-Qiang. An overview on the adaptive dynamic programming based urban city traffic signal optimal control. *Acta Automatica Sinica*, 2009, **35**(6): 677−681
(赵冬斌, 刘德荣, 易建强. 基于自适应动态规划的城市交通信号优化控制方法综述. 自动化学报, 2009, **35**(6): 677−681)

51 Sutton R S, Barto A G. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press, 1998

52 Zhang Hua-Guang, Zhang Xin, Luo Yan-Hong, Yang Jun. An overview of research on adaptive dynamic programming. *Acta Automatica Sinica*, 2013, **39**(4): 303−311
(张化光, 张欣, 罗艳红, 杨珺. 自适应动态规划综述. 自动化学报, 2013, **39**(4): 303−311)

53 Lewis F L, Vrabie D, Vamvoudakis K G. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems Magazine*, 2012, **32**(6): 76−105

54 Wang F Y, Jin N, Liu D R, Wei Q L. Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ϵ-error bound. *IEEE Transactions on Neural Networks*, 2011, **22**(1): 24−36

55 Wang F Y, Zhang H G, Liu D R. Adaptive dynamic programming: an introduction. *IEEE Computational Intelligence Magazine*, 2009, **4**(2): 39−47

56 Vrabie D, Lewis F L. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Networks*, 2009, **22**(3): 237−246

57 Vamvoudakis K G, Lewis F L. Online synchronous policy iteration method for optimal control. *Recent Advances in Intelligent Control Systems*. Berlin: Springer-Verlag, 2009. 357−374

58 Bhasin S, Kamalapurkar R, Johnson M, Vamvoudakis K G, Lewis F L, Dixon W E. A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. *Automatica*, 2013, **49**(1): 82−92

59 Zhang H G, Luo Y H, Liu D R. Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. *IEEE Transactions on Neural Networks*, 2009, **20**(9): 1490−1503

60 Zhang H G, Cui L L, Luo Y H. Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP. *IEEE Transactions on Cybernetics*, 2013, **43**(1): 206−216

61 Wei Q L, Zhang H G. A new approach to solve a class of continuous-time nonlinear quadratic zero-sum game using ADP. In: Proceedings of the 2008 IEEE International Conference on Networking, Sensing and Control. Sanya, China: IEEE, 2008. 507−512

62 Zhang H G, Wei Q L, Liu D R. An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games. *Automatica*, 2010, **47**(1): 207−214

63 Vamvoudakis K G, Lewis F L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 2010, **46**(5): 878−888

64 Zhang X, Zhang H G, Luo Y H, Dong M. Iteration algorithm for solving the optimal strategies of a class of nonaffine nonlinear quadratic zero-sum games. In: Proceedings of the 2010 Chinese Control and Decision Conference (CDC). Xuzhou, China: IEEE, 2010. 1359−1364

65 Liu D R, Li H L, Wang D. Neural-network-based zero-sum game for discrete-time nonlinear systems via iterative adaptive dynamic programming algorithm. *Neurocomputing*, 2013, **110**(13): 92−100

66 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

67 Xu Xin, Shen Dong, Gao Yan-Qing, Wang Kai. Learning control of dynamical systems based on Markov decision processes: research frontiers and outlooks. *Acta Automatica Sinica*, 2012, **38**(5): 673−687
(徐昕, 沈栋, 高岩青, 王凯. 基于马氏决策过程模型的动态系统学习控制: 研究前沿与展望. 自动化学报, 2012, **38**(5): 673−687)

68 Sharma R, Gopal M. Synergizing reinforcement learning and game theory — a new direction for control. *Applied Soft Computing*, 2010, **10**(3): 675−688

69 Littman M L. Value-function reinforcement learning in markov games. *Journal of Cognitive Systems Research*, 2001, **2**(1): 55−56

70 Littman M L. Markov games as a framework for multi-agent reinforcement learning. In: Proceedings of the 11th International Conference on Machine Learning. New Brunswick, NJ: Morgan Kaufmann Publishers, 1994. 157−163

71 Frénay B, Saerens M. QL_2, A simple reinforcement learning scheme for two-player zero-sum Markov games. *Neurocomputing*, 2009, **72**(7−9): 1494−1507

72 Hu J L, Wellman M P. Multiagent reinforcement learning: theoretical framework and an algorithm. In: Proceedings of the 15th International Conference on Machine Learning. New Brunswick, NJ: Morgan Kaufmann Publishers, 1998. 242−250

73 Dockner E J, Steffen J, van Ngo L, Sorger G. *Differential Games in Economics and Management Science*. Cambridge: Cambridge University Press, 2001

74 Weber T A, Kryazhimskiy A V. *Optimal Control Theory with Applications in Economics*. Cambridge: The MIT Press, 2011

75 Wang Fei-Yue. Parallel control: a method for data-driven and computational control. *Acta Automatica Sinica*, 2013, **39**(4): 293−302
(王飞跃. 平行控制: 数据驱动的计算控制方法. 自动化学报, 2013, **39**(4): 293−302)

76 Anderson R, Moore T. The economics of information security. *Science*, 2006, **314**(5799): 610−613

77 Nisan N, Roughgarden T, Tardos E, Vazirani V V. *Algorithmic Game Theory*. Cambridge: Cambridge University Press, 2007

78 Roughgarden T. Algorithmic game theory. *Communications of the ACM*, 2010, **53**(7): 78−86

79 Wei Zhi-Qiang, Zhou Wei, Ren Xiang-Jun, Wei Qing, Jia Dong-Ning, Kang Mi-Jun, Yin Bo, Cong Yan-Ping. A strategy-proof trust based decision mechanism for pervasive computing environments. *Chinese Journal of Computer*, 2012, **35**(5): 871−882
(魏志强, 周炜, 任相军, 魏青, 贾东宁, 康密军, 殷波, 丛艳平. 普适计算环境中防护策略的信任决策机制研究. 计算机学报, 2004, **35**(5): 871−882)

80 Semsar-Kazerooni E, Khorasani K. Multi-agent team cooperation: a game theory approach. *Automatica*, 2009, **45**(10): 2205−2213

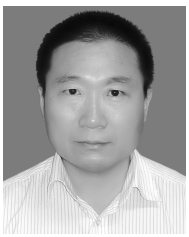
81 Fax J A, Murray R M. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 2004, **49**(9): 1465−1476

82 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

83 Jadbabaie A, Lin J, Morse A S. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 2003, **48**(6): 988−1001

84 Chen X, Deng X T. Settling the complexity of two-player Nash equilibrium. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS′06). Berkeley, USA: IEEE, 2006. 261−272

85 Daskalakis C. The Complexity of Nash Equilibria. Electrical Engineering and Computer Sciences [Ph. D. dissertation], University of California at Berkeley, USA, 2008



谭拂晓 安徽大学计算机科学与技术学院博士后, 阜阳师范学院计算机与信息学院副教授. 主要研究方向为多智能体网络系统的协调控制, 非线性系统的鲁棒控制, 基于增强学习的非线性系统动态优化. 本文通信作者.
E-mail: fuxiaotan@gmail.com
(**TAN Fu-Xiao** Postdoctor at the School of Computer Science and Technology, Anhui University, and associate professor at the School of Computer and Information, Fuyang Teachers College. His research interest covers coordinated control of networked multi-agent systems, nonlinear robust control, and dynamic optimization of nonlinear system based on reinforcement learning. Corresponding author of this paper.)

刘德荣 中国科学院自动化研究所复杂系统管理与控制国家重点实验室研究员. 主要研究方向为智能系统和复杂系统的建模, 分析与控制.
E-mail: derong.liu@ia.ac.cn
(**LIU De-Rong** Professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interest covers modeling, analysis, and control of intelligent systems and complex systems.)

关新平 上海交通大学电子信息与电气工程学院系教授. 主要研究方向为多智能体系统, 非线性系统的鲁棒控制.
E-mail: xpguan@sjtu.edu.cn
(**GUAN Xin-Ping** Professor at the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University. His research interest covers multi-agent systems and nonlinear robust control.)

罗 斌 安徽大学计算机科学与技术学院教授. 主要研究方向为模式识别与数字图像处理.
E-mail: luobin@ahu.edu.cn
(**LUO Bin** Professor at the School of Computer Science and Technology, Anhui University. His research interest covers image and graph matching, statistical pattern recognition, and image feature extraction.)