控制与决策
Control and Decision

基于强化学习的多智能体系统容错编队最优控制

申思凯, 江南, 刘小洋

引用本文:
申思凯, 江南, 刘小洋. 基于强化学习的多智能体系统容错编队最优控制[J]. 控制与决策, 2025, 40(12): 3565-3575.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2025.0218

基于强化学习的多智能体系统容错编队最优控制

申思凯[1]，江 南[2†]，刘小洋[1]

(1. 江苏师范大学 数学与统计学院，江苏 徐州 221116；2. 江苏师范大学 商学院，江苏 徐州 221116)

摘 要: 研究执行器故障下多智能体系统的最优编队控制问题, 提出一种基于强化学习的预设时间最优容错编队控制方法. 首先, 借助预设时间调谐函数, 结合评论家-演员算法, 可显著降低算法复杂度; 然后, 在此基础上, 设计一种高效的容错控制机制, 以确保系统在执行器发生故障时仍然能够稳定地实现预期编队控制目标; 接着, 基于强化学习的自适应算法能力, 进一步增强算法对于复杂动态环境的适应性和鲁棒性; 最后, 通过仿真结果验证了所提出方法在故障场景下的有效性和优越性.

关键词: 强化学习；预设时间；多智能体系统；编队控制；执行器故障

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.0218

引用格式: 申思凯, 江南, 刘小洋. 基于强化学习的多智能体系统容错编队最优控制 [J]. 控制与决策, 2025, 40(12): 3565-3575.

Fault-tolerant formation optimized control for multi-agent systems via reinforcement learning

SHEN Si-kai[1]，*JIANG Nan*[2†]，*LIU Xiao-yang*[1]

(1. School of Mathematics and Statistics， Jiangsu Normal University， Xuzhou 221116， China； 2. Business School，Jiangsu Normal University，Xuzhou 221116，China)

Abstract: This paper investigates the optimal formation control problem for multi-agent systems under actuator faults and proposes a predefined-time optimized fault-tolerant formation control method based on reinforcement learning. The method employs a predefined-time tuning function and combines with an actor-critic algorithm and significantly reduces algorithmic complexity. Based on this, an efficient fault-tolerant control mechanism is designed to ensure the system achieve the desired formation control objectives, even in the presence of actuator faults. Furthermore, the reinforcement learning-based adaptability further enhances the algorithm's robustness and adaptability in complex dynamic environments. Finally, simulation results validate the effectiveness and superiority of the proposed method across fault scenarios.

Keywords: reinforcement learning；predefined-time；multi-agent systems；formation control；actuator faults

0 引 言

近 10 年来, 多智能体系统因其在固定翼无人机系统[1]、智能电网[2]、机器人[3] 等领域的广泛应用, 逐渐成为研究的重点, 其中编队控制因其极具潜力的应用前景, 已然发展为备受关注的研究热点[4]. 然而, 当前的研究多集中于理想条件下的控制策略, 对于系统在实际运行中可能面临的故障问题关注得较少. 事实上, 工程实践中由于组件损坏或外部干扰引发的故障不可避免, 若不能及时处理, 则可能会导致财产损失、人员伤亡、环境污染甚至生态灾难. 因此, 提升多智能体系统的容错能力以确保其可靠性和安全性具有重要意义[5]. 针对这一问题, 文献 [6] 在具有随机噪声和执行器故障的非线性系统下通过引入自适应故障补偿项, 提出了一种新的容错跟踪控制策略; 文献 [7] 提出了一种自适应神经网络动态事件触发一致性方法, 该方法进一步拓展了多智能体系统在复杂环境下的容错控制能力.

强化学习作为一种高效的优化方法, 已广泛应用于复杂控制问题[8]. 其中评论家-演员架构是一种被广泛采用的强化学习策略结构: 演员通过与环境交互执行控制动作; 评论家则对动作效果进行评估,

收稿日期: 2025-03-03；**录用日期:** 2025-09-15.

基金项目: 国家自然科学基金项目 (62276119)；江苏省基础研究计划项目 (BK20241764)；江苏省研究生实践创新计划项目 (SJCX25_1523)；江苏师范大学研究生科研与实践创新计划项目 (2024XKT2630).

[†]通信作者. E-mail: lxyjn@163.com.

并将结果反馈给演员以优化其控制策略[9]. 在多智能体系统中, 最优控制问题旨在在完成控制任务的同时最小化预定义的性能指标, 以提升系统整体效率[10]. 然而, 由于多智能体间的邻域耦合关系, 贝尔曼方程中通常包含高维非线性项, 使得最优控制器的求解变得复杂. 为克服这一挑战, 研究者们引入了强化学习方法, 如: 文献 [11] 针对非线性严格反馈动态多智能体系统, 提出了基于强化学习的领导-跟随一致性控制方案; 文献 [12] 使用强化学习方法和基于事件触发的完全分布式控制策略研究了信息未知的多智能体系统的二分一致性问题; 文献 [13] 针对一类状态不可测的非线性系统, 采用强化学习和模糊逻辑系统, 设计了自适应优化反步控制方案. 尽管已有研究取得初步进展, 但是, 关于具有执行器故障的多智能体系统最优编队控制的强化学习方法仍然较少, 需要进一步探索.

传统多智能体系统控制策略通常能够确保系统在时间趋于无穷大时达到一致. 然而, 在某些实际应用中, 系统的收敛速度成为关键指标. 为此, 文献 [14] 提出了一种多智能体模糊自适应快速有限时间控制策略, 用于实现领导-跟随编队控制; 文献 [15] 提出了多智能体固定时间控制方法, 通过缩短收敛时间, 显著提升了控制效率. 然而, 这些方法的一个共同局限性在于均无法预先设定系统的具体收敛时间. 针对这一不足, 文献 [16] 首次提出了能够在预设时间内实现收敛的控制方法. 该控制方法的最大优势在于系统的收敛时间可通过参数调节提前设定, 从而更好地满足工程实践中对于时间精确度的要求.

综上, 本文围绕基于强化学习的多智能体系统预设时间容错编队控制展开研究, 主要内容如下:

1) 与现有最优控制方法[10-13] 相比, 本文同时考虑锁定故障 (LIP) 和部分效能丧失故障 (LOE) 模型, 通过在线补偿故障效应, 提出新的容错控制策略, 可解决多智能体系统的编队控制问题.

2) 相较于传统的预设时间控制方法[17], 本文基于预设时间调谐函数[18-19], 构建一种新型预设时间容错编队控制方法. 该方法能够确保编队误差在预设时间内收敛至指定界限, 显著提升控制速度.

3) 与传统的最优控制[11] 不同, 本文结合强化学习, 通过设计简化的评论家-演员神经网络更新策略, 降低传统最优控制中因梯度下降算法带来的计算复杂度.

1 问题描述与预备知识

考虑由一个领导者和N个跟随者组成的多智能体系统, 跟随者的动态模型可描述为

$$\begin{cases}\dot{x}_{i,j}(t)=x_{i,j+1}(t)+f_{i,j}(\bar{x}_{i,j}(t)),\\ \quad j=1,2,\ldots,n-1;\\ \dot{x}_{i,n}(t)=\psi_i^{\mathrm{T}}u_i(t)+f_{i,n}(\bar{x}_{i,n}(t));\\ y_i(t)=x_{i,1}(t).\end{cases} \tag{1}$$

其中: $\bar{x}_{ij}(t)=[x_{i,1}(t),x_{i,2}(t),\ldots,x_{i,j}(t)]^{\mathrm{T}}\in\mathbb{R}^j$为系统的状态向量, $f_{i,j}(\bar{x}_{i,j}(t))\in\mathbb{R}$为未知的连续非线性函数, $u_i(t)=[u_{i,1}(t),u_{i,2}(t),\ldots,u_{i,l}(t)]^{\mathrm{T}}\in\mathbb{R}^l$为其组件存在故障的执行器输入, $y_i(t)\in\mathbb{R}$为系统的输出, $\psi_i=[\psi_{i,1},\psi_{i,2},\ldots,\psi_{i,l}]^{\mathrm{T}}\in\mathbb{R}^l$为输入增益的正常向量.

定义 1 考虑多智能体系统 (1), 当满足$\lim\limits_{t\to T_f}\|y_i(t)-y_l(t)-h_i\|=0$时, 称多智能体系统实现预设时间编队控制. 其中: T_f为预设时间, $h_i\in\mathbb{R}$为第i个智能体与领导者$y_l(t)$间的相对距离.

假设 1 信息交互图G是无向的, 且存在一棵生成树, 领导者节点为树根.

假设 2 领导者的输出信号y_l对时间的一阶导数连续且有界.

本文考虑执行器故障包括锁定故障 (LIP) 和部分效能丧失 (LOE) 故障[20], 这些故障模型可用参数化形式[21] 表示如下.

1) LIP 如下所示:

$$u_{i,m}(t)=\underline{u}_{i,m},\\ m\in\{m_1,m_2,\ldots,m_p\}\subset\{1,2,\ldots,l\}. \tag{2}$$

其中: $\underline{u}_{i,m}$为 LIP 故障的常数, p为发生 LIP 故障的执行器数量.

2) LOE 如下所示:

$$u_{i,m}(t)=\rho_{i,m}v_{i,m}(t),\ m\notin\{m_1,m_2,\ldots,m_p\}. \tag{3}$$

其中: $v_{i,m}$为实际执行器信号, $\rho_{i,m}$为故障系数. 由于执行器不能完全失效, $\rho_{i,m}\in[\underline{\rho_{i,m}},1]$, $0<\rho_{i,m}\leqslant 1$, $\underline{\rho_{i,m}}$为故障系数$\rho_{i,m}$的未知下界. 当$\rho_{i,m}=1$时, 第m个执行器没有故障.

记 $v_i(t)=[v_{i,1}(t),v_{i,2}(t),\ldots,v_{i,l}(t)]^{\mathrm{T}}$, $\underline{u}_i(t)=[\underline{u}_{i,1}(t),\underline{u}_{i,2}(t),\ldots,\underline{u}_{i,l}(t)]^{\mathrm{T}}$, $\rho_i=\mathrm{diag}\{\rho_{i,1},\rho_{i,2},\ldots,\rho_{i,l}\}$, $\sigma_i=\mathrm{diag}\{\sigma_{i,1},\sigma_{i,2},\ldots,\sigma_{i,l}\}$, 其中

$$\sigma_{i,n}=\begin{cases}1,\ u_{i,n}(t)=\underline{u}_{i,n};\\ 0,\ \text{otherwise}.\end{cases} \tag{4}$$

基于式 (2) ∼ (4), 可得到

$$u_i(t)=\rho_iv_i(t)+\sigma_i(\underline{u}_i-\rho_iv_i(t)). \tag{5}$$

定义执行器的结构为

$$v_{i,m}(t)=\omega_{i,m}(x)u_{i,0}(t). \tag{6}$$

其中: $u_{i,0}(t)$为后续要设计的输入信号, 增益$\omega_{i,m}>0$满足$\underline{\omega}_{i,m}\leqslant\omega_{i,m}(x)\leqslant\bar{\omega}_{i,m}(\forall x\in\mathbb{R}^n, m=1,2,\ldots,l)$.

假设 3　对于系统 (1), 至少有一个执行器不受式 (2) 所述的 LIP 故障影响.

假设 4　对于非线性函数$f_{i,j}(\cdot)(i=1,2,\ldots,N, j=1,2,\ldots,n)$有界.

1.1　神经网络逼近

对于未知连续函数$g(x)$, 其可由紧集Ω_x上的径向基函数 (RBF) 神经网络逼近[22], 其形式如下所示:

$$g(x)=W^{*\mathrm{T}}S(x)+\varepsilon(x).$$

其中: $W^*=[W_1^*,W_2^*,\ldots,W_p^*]^{\mathrm{T}}\in R^p$为理想权重向量, 满足$W^*:=\arg\min\limits_{W\in R^p}\{\sup\limits_{x\in\Omega_x}|\psi(x)-W^{\mathrm{T}}S(x)|\}$; 基函数向量$S(x)=[\eta_1(x),\eta_2(x),\ldots,\eta_p(x)]^{\mathrm{T}}$, 这里$\eta_i(x)=\exp[-(x-\nu_i)^{\mathrm{T}}(x-\nu_i)/\zeta_i^2]$, ν_i和ζ_i分别为接受域的中心和高斯函数的宽度; $\varepsilon(x)$为近似误差.

1.2　最优控制

定义 2　对于非线性系统

$$\dot{x}(t)=f(x)+u(t), \tag{7}$$

若在控制器$u(t)$下系统 (7) 稳定, 且性能指标函数$J=\int_t^{\infty}(Q(x)+R(u))\mathrm{d}s$有限, 则称$u$为容许控制器, 其中$Q(x)$和$R(u)$为正函数. 进一步地, 若存在容许控制$u$, 使得$J=\int_t^{\infty}(Q(x)+R(u))\mathrm{d}s$的值最小, 则该控制称为最优控制.

2　最优容错控制器设计

在本节中, 将最优控制算法与反步控制法相结合, 所有虚拟控制和实际控制统一设计为各子系统的最优解, 从而实现对整个系统性能的优化. 智能体i的最优控制通过反步法分n步进行设计, 其设计思路如图 1 所示.

引入坐标变换, 如下所示:

$$\begin{aligned}&z_{i,1}(t)=y_i(t)-y_l(t)-h_i-\mu(t),\\&z_{i,j}(t)=x_{i,j}(t)-q_{i,j}(t),\ j=2,3,\ldots,n.\end{aligned} \tag{8}$$

利用动态面控制技术来消除“复杂性爆炸”, 设计如下一阶滤波器:

$$\begin{cases}\tau_{i,j}(t)\dot{q}_{i,j}(t)+q_{i,j}(t)=\hat{\alpha}_{i,j-1}^*(t),\\q_{i,j}(0)=\hat{\alpha}_{i,j-1}^*(0),\\\varpi_{i,j-1}(t)=q_{i,j}(t)-\hat{\alpha}_{i,j-1}^*(t).\end{cases} \tag{9}$$

其中: $q_{i,j}(t)$为滤波器输出, $\varpi_{i,j}(t)$为输出误差.

定义同步误差为

$$\begin{aligned}e_i(t)=&\sum_{j=1}^n a_{i,j}((y_i(t)-h_i)-(y_j(t)-h_j))+\\&b_{i0}(y_i(t)-y_l(t)-h_i-\mu(t)).\end{aligned} \tag{10}$$

其中: $a_{i,j}$为邻接矩阵的元素, 若智能体i与智能体j存在通信, 则$a_{i,j}>0$; 否则, $a_{i,j}=0$. b_{i0}为智能体i与领导者间的通信. $\mu(t)$为时变调谐函数, 其形式如下所示:

$$\mu(t)=\begin{cases}q_1t^4+q_2t^3+q_3t^2+o_1t+\iota_1,\ t<T_f;\\0,\ t\geqslant T_f.\end{cases} \tag{11}$$

这里: $q_1=-\left(\dfrac{6.5l_1}{T_f^4}+\dfrac{o_1}{T_f^3}\right)$, $q_2=\left(\dfrac{9.5l_1}{T_f^3}+\dfrac{3o_1}{T_f^2}\right)$, $q_3=-\left(\dfrac{6.2l_1}{T_f^2}+\dfrac{3o_1}{T_f}\right)$, ι_1和o_1是由系统状态的初始条件和参考信号决定的常数.

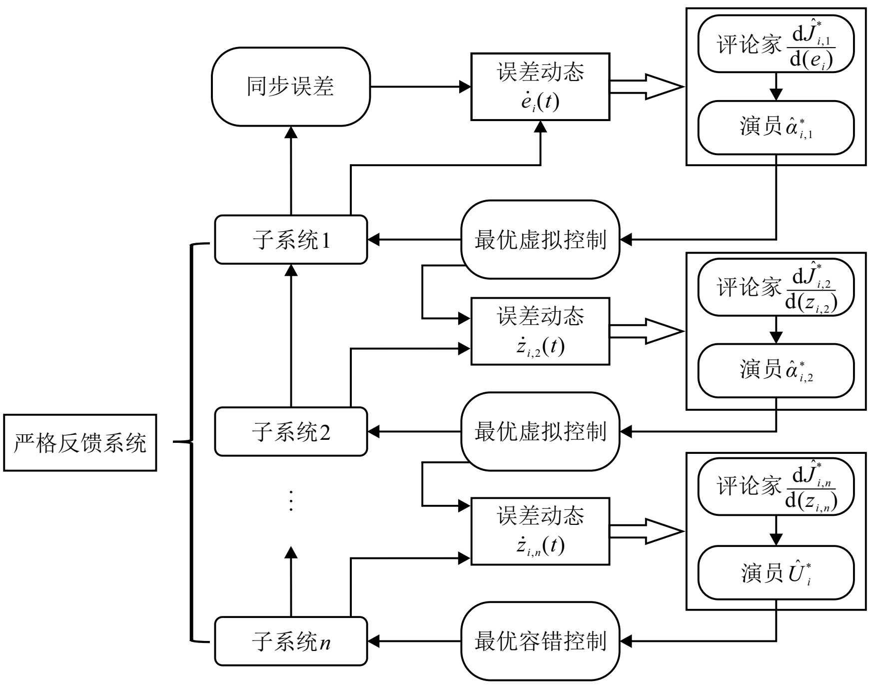


图1　智能体i的最优反步设计

step 1: 基于坐标变换, 同步误差 (10) 更新为

$$e_i(t)=\sum_{j=1}^{n}a_{i,j}(z_{i,1}(t)-z_{j,1}(t))+b_{i0}z_{i,1}(t),\quad(12)$$

其满足

$$\dot{e}_i(t)=p_i(x_{i,2}+f_{i,1}(\bar{x}_{i,1}(t))-\dot{y}_l(t)-\dot{\mu}(t))-\sum_{j=1}^{n}a_{i,j}\dot{z}_{j,1}(t),\quad(13)$$

其中$p_i=\sum_{j=1}^{n}a_{i,j}+b_{i0}$.

定义性能指标函数为

$$J_{i,1}(e_i)=\int_t^{\infty}c_{i,1}(e_i(s),\alpha_{i,1}(s))\mathrm{d}s.\quad(14)$$

其中: $c_{i,1}(e_i,\alpha_{i,1})=e_i^2(t)+\alpha_{i,1}^2(t)$为成本函数, $\alpha_{i,1}$为虚拟控制器. 令$\alpha_{i,1}^*$为最优虚拟控制, 将式 (14) 中的$\alpha_{i,1}$替换为$\alpha_{i,1}^*$, 得到最优性能指标函数为

$$J_{i,1}^*(e_i)=\int_t^{\infty}c_{i,1}(e_i(s),\alpha_{i,1}^*(s))\mathrm{d}s=\min_{\alpha_{i,1}\in\Psi(\Omega)}\Big\{\int_t^{\infty}c_{i,1}(e_i(s),\alpha_{i,1}(s))\mathrm{d}s\Big\}.\quad(15)$$

将$x_{i,2}$看作最优虚拟控制器$\alpha_{i,1}^*$, 则与式 (13) 和 (15) 相关的 HJB 方程为

$$H_{i,1}(e_i,\alpha_{i,1}^*,J_{i,1}^*)=e_i^2(t)+\alpha_{i,1}^{*2}(t)+\frac{\mathrm{d}J_{i,1}^*(e_i)}{\mathrm{d}e_i}\times\Big(p_i(\alpha_{i,1}^*(t)+f_{i,1}(\bar{x}_{i,1}(t))-\dot{y}_l(t)-\dot{\mu}_1(t))-\sum_{j=1}^{n}a_{i,j}\dot{z}_{j,1}(t)\Big)=0.\quad(16)$$

通过求解$\partial H_{i,1}/\partial\alpha_{i,1}^*=0$, 得到最优虚拟控制

$$\alpha_{i,1}^*(e_i)=-\frac{p_i}{2}\frac{\mathrm{d}J_{i,1}^*(e_i)}{\mathrm{d}(e_i)}.\quad(17)$$

将$\dfrac{\mathrm{d}J_{i,1}^*(e_i)}{\mathrm{d}(e_i)}$分解为

$$\frac{\mathrm{d}J_{i,1}^*(e_i)}{\mathrm{d}(e_i)}=\frac{2\beta_{i,1}}{p_i}e_i(t)-\frac{2(\dot{y}_l(t)+\dot{\mu}(t))}{p_i}+\frac{1}{p_i}J_{i,1}^o(e_i).\quad(18)$$

其中: $\beta_{i,1}$为正常数, $J_{i,1}^o(e_i)=-2\beta_{i,1}e_i+2(\dot{y}_l+\dot{\mu})+p_i\dfrac{\mathrm{d}J_{i,1}^*(e_i)}{\mathrm{d}(e_i)}$. 将式 (18) 代入 (17), 有

$$\alpha_{i,1}^*(e_i)=-\beta_{i,1}e_i(t)+\dot{y}_l(t)+\dot{\mu}(t)-\frac{1}{2}J_{i,1}^o(e_i).\quad(19)$$

由于未知函数$J_{i,1}^o(e_i)$是连续的, 可通过紧集Ω_x上的神经网络来逼近, 其形式如下所示:

$$J_{i,1}^o(e_i)=W_{i,1}^{*\mathrm{T}}S_{i,1}(e_i)+\varepsilon_{i,1}.\quad(20)$$

这里: $W_{i,1}^*\in\mathbb{R}^{q_{i1}}$为理想的神经网络权重, $S_{i,1}(e_i)\in\mathbb{R}^{q_{i1}}$为基函数向量, $\varepsilon_{i,1}\in\mathbb{R}$为有界近似误差. 将式 (20) 代入 (18) 和 (19), 可得到

$$\frac{\mathrm{d}J_{i,1}^*(e_i)}{\mathrm{d}(e_i)}=\frac{2\beta_{i,1}}{p_i}e_i(t)-\frac{2(\dot{y}_l(t)+\dot{\mu}(t))}{p_i}+\frac{1}{p_i}(W_{i,1}^{*\mathrm{T}}S_{i,1}(e_i)+\varepsilon_{i,1}),\quad(21)$$

$$\alpha_{i,1}^*(e_i)=-\beta_{i,1}e_i(t)+\dot{y}_l(t)+\dot{\mu}(t)-\frac{1}{2}W_{i,1}^{*\mathrm{T}}(t)S_{i,1}(e_i)-\frac{1}{2}\varepsilon_{i,1}.\quad(22)$$

由于理想参数向量$W_{i,1}^*(t)$未知, 导致最优虚拟控制 (22) 不可用. 为得到可用的最优虚拟控制, 通过以下评论家和演员来实现强化学习算法.

由式 (21), 设计用于评估控制性能的评论家, 如下所示:

$$\frac{\mathrm{d}\hat{J}_{i,1}^*(e_i)}{\mathrm{d}(e_i)}=\frac{2\beta_{i,1}}{p_i}e_i(t)-\frac{2(\dot{y_l(t)}+\dot{\mu}(t))}{p_i}+\frac{1}{p_i}\hat{W}_{c,i1}^{\mathrm{T}}(t)S_{i,1}(e_i).\quad(23)$$

其中: $\dfrac{\mathrm{d}\hat{J}_{i,1}^*(e_i)}{\mathrm{d}(e_i)}\in\mathbb{R}$为对$\dfrac{\mathrm{d}J_{i,1}^*(e_i)}{\mathrm{d}(e_i)}$的估计; $\hat{W}_{c,i1}(t)\in\mathbb{R}^{q_{i1}}$为评论家神经网络权重, 其更新率如下所示:

$$\dot{\hat{W}}_{c,i1}(t)=-\gamma_{c,i1}S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)\hat{W}_{c,i1}(t)-\frac{1}{2}S_{i,1}(e_i)e_i,\quad(24)$$

这里$\gamma_{c,i1}>0$为评论家学习率.

由式 (22), 设计用于实现控制动作的演员, 如下所示:

$$\hat{\alpha}_{i,1}^*(t)=-\beta_{i,1}e_i(t)+\dot{y}_l(t)+\dot{\mu}(t)-\frac{1}{2}\hat{W}_{a,i1}^{\mathrm{T}}(t)S_{i,1}(e_i),\quad(25)$$

其中$\hat{W}_{a,i1}(t)\in\mathbb{R}^{q_{i1}}$为演员神经网络权重, 其更新率如下所示:

$$\dot{\hat{W}}_{a,i1}(t)=-\gamma_{a,i1}S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)\times(\hat{W}_{a,i1}(t)-\hat{W}_{c,i1}(t)),\quad(26)$$

这里$\gamma_{a,i1}>0$为演员学习率. 常数$\beta_{i,1}$、$\gamma_{c,i1}$和$\gamma_{a,i1}$的选择需要满足以下条件:

$$\gamma_{c,i1}>\gamma_{a,i1},\ \gamma_{a,i1}>\frac{\epsilon_{i,1}}{2},\quad \gamma_{c,i1}>\frac{\varsigma_{i,1}}{2},\ \beta_{i,1}>\frac{3}{2}+\frac{1}{4\epsilon_{i,1}}+\frac{1}{4\varsigma_{i,1}}.\quad(27)$$

选取 Lyapunov 函数为$V_i^1(t)=V_0(t)+V_{i,1}(t)$. 其中

$$V_0(t)=(1/2)z_1^{\mathrm{T}},\ \tilde{L}=L+B,$$
$$z_1=[z_{1,1},z_{2,1},\ldots,z_{N,1}]^{\mathrm{T}},$$

$$V_{i,1}(t)=(1/2)\tilde{W}_{c,i1}^{\mathrm{T}}\tilde{W}_{c,i1}+(1/2)\tilde{W}_{a,i1}^{\mathrm{T}}\tilde{W}_{a,i1},$$
$$\tilde{W}_{c,i1}(t)=\hat{W}_{c,i1}(t)-W_{i,1}^{*}(t),$$
$$\tilde{W}_{a,i1}(t)=\hat{W}_{a,i1}(t)-W_{i,1}^{*}(t).$$

令$e(t)=[e_1(t),e_2(t),\dots,e_N(t)]^{\mathrm{T}}$, 由式 (10), 可得到$e(t)=\tilde{L}z_1(t)$. 此外, 有如下等式成立:

$$z_1^{\mathrm{T}}(t)\tilde{L}z_1=z_1^{\mathrm{T}}(t)\tilde{L}\tilde{L}^{-1}\tilde{L}z_1(t)=\mathrm{e}^{\mathrm{T}}(t)\tilde{L}^{-1}e(t). \tag{28}$$

其中: $L=D-A$为拉普拉斯矩阵, $D=\mathrm{diag}(d_1,d_2,\dots,d_N)$, $d_i=\sum_{j\in\mathcal{N}_i}a_{ij}$为节点$i$的入度, $A=[a_{ij}]\in\mathbb{R}^{N\times N}$为邻接矩阵. $B=\mathrm{diag}(b_{10},b_{20},\dots,b_{N0})$为智能体与领导者间的通信权重, 若第$i$个智能体与领导者通信, 则$b_{i0}>0$; 否则, $b_{i0}=0$. 由式 (28), 可得到以下结果:

$$\frac{1}{\lambda_{\tilde{L}}^{\max}}\|e(t)\|^2\leqslant z_1^{\mathrm{T}}(t)\tilde{L}z_1(t)\leqslant\frac{1}{\lambda_{\tilde{L}}^{\min}}\|e(t)\|^2, \tag{29}$$

这里$\lambda_{\tilde{L}}^{\max}$和$\lambda_{\tilde{L}}^{\min}$分别为矩阵$\tilde{L}$的最大特征值和最小特征值.

沿着式 (24) ~ (26), 计算$V_i^1(t)$对时间的导数, 如下所示:

$$\begin{aligned}&\dot{V}_i^1(t)=\\&\sum_{l=1,l\neq i}^{N}e_l^{\mathrm{T}}(t)\dot{z}_{l,1}(t)+e_i(t)(z_{i,2}(t)+f_{i,1}+\varpi_{i,1}(t)-\\&\beta_{i,1}e_i(t))-\frac{1}{2}e_i(t)\hat{W}_{a,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)-\\&\frac{1}{2}\tilde{W}_{c,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)e_i(t)-\gamma_{c,i1}\tilde{W}_{c,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)\times\\&S_{i,1}^{\mathrm{T}}(e_i)\hat{W}_{c,i1}(t)-\gamma_{a,i1}\tilde{W}_{a,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)\times\\&(\hat{W}_{a,i1}(t)-\hat{W}_{c,i1}(t)).\end{aligned} \tag{30}$$

基于$\tilde{W}_{c,i1}(t)=\hat{W}_{c,i1}(t)-W_{i,1}^{*}(t)$和$\tilde{W}_{a,i1}(t)=\hat{W}_{a,i1}(t)-W_{i,1}^{*}(t)$, 可得到如下等式:

$$\begin{aligned}&-\frac{1}{2}e_i(t)\hat{W}_{a,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)-\frac{1}{2}\tilde{W}_{c,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)e_i(t)=\\&-\frac{1}{2}e_i(t)\tilde{W}_{a,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)-\frac{1}{2}e_i(t)\hat{W}_{c,i1}^{\mathrm{T}}(t)S_{i,1}(e_i),\end{aligned} \tag{31}$$

$$\hat{W}_{a,i1}(t)-\hat{W}_{c,i1}(t)=\tilde{W}_{a,i1}(t)-\tilde{W}_{c,i1}(t). \tag{32}$$

将式 (31) 和 (32) 代入 (30), 并根据 Young 不等式, 可得到

$$\begin{aligned}&\dot{V}_i^1(t)\leqslant\\&\sum_{l=1,l\neq i}^{N}e_l^{\mathrm{T}}(t)\dot{z}_{l,1}(t)-\Big(\beta_{i,1}-\frac{1}{4\epsilon_{i,1}}-\frac{1}{4\varsigma_{i,1}}-\frac{3}{2}\Big)e_i^2(t)-\\&\frac{\gamma_{c,i1}}{2}\tilde{W}_{c,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)\tilde{W}_{c,i1}(t)-\\&\Big(\gamma_{a,i1}-\frac{\epsilon_{i,1}}{4}\Big)\tilde{W}_{a,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)\tilde{W}_{a,i1}(t)+\\&\gamma_{a,i1}\tilde{W}_{a,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)\tilde{W}_{c,i1}(t)-\\&\Big(\frac{\gamma_{c,i1}}{2}-\frac{\varsigma_{i,1}}{4}\Big)\hat{W}_{c,i1}^{\mathrm{T}}(t)S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)\hat{W}_{c,i1}(t)+\\&\frac{1}{2}z_{i,2}^2(t)+\frac{1}{2}h_{i,1}^2(t)+B_{i,1}.\end{aligned} \tag{33}$$

其中

$$\epsilon_{i,1}>0,\ \varsigma_{i,1}>0,$$
$$B_{i,1}=\frac{\gamma_{c,i1}}{2}W_{i,1}^{*\mathrm{T}}S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)W_{i,1}^{*}+\frac{1}{2}f_{i,1}^2.$$

这里$B_{i,1}$的所有项均是有界的, 即存在常数$b_{i,1}$, 使得$|B_{i,1}|\leqslant b_{i,1}$.

通过选择合适的参数$\beta_{i,1}$、$\gamma_{c,i1}$和$\gamma_{a,i1}$满足条件 (27), 式 (33) 可进一步写为

$$\begin{aligned}&\dot{V}_i^1\leqslant\\&\sum_{l=1,l\neq i}^{N}e_l^{\mathrm{T}}(t)\dot{z}_{l,1}(t)-\Big(\beta_{i,1}-\frac{1}{4\epsilon_{i,1}}-\frac{1}{4\varsigma_{i,1}}-\frac{3}{2}\Big)e_i^2(t)-\\&\Big(\frac{\gamma_{c,i1}}{2}-\frac{\gamma_{a,i1}}{2}\Big)\lambda_{S_{i,1}}^{\min}\tilde{W}_{c,i1}^{\mathrm{T}}(t)\tilde{W}_{c,i1}(t)-\\&\Big(\frac{\gamma_{a,i1}}{2}-\frac{\epsilon_{i,1}}{4}\Big)\lambda_{S_{i,1}}^{\min}\tilde{W}_{a,i1}^{\mathrm{T}}(t)\tilde{W}_{a,i1}(t)+\\&\frac{1}{2}z_{i,2}^2(t)+\frac{1}{2}h_{i,1}^2(t)+b_{i,1},\end{aligned} \tag{34}$$

其中$\lambda_{S_{i,1}}^{\min}$为$S_{i,1}(e_i)S_{i,1}^{\mathrm{T}}(e_i)$的最小特征值.

令

$$a_{i,1}=\min\{(\gamma_{c,i1}-\gamma_{a,i1})\lambda_{S_{i,1}}^{\min},\gamma_{a,i1}-(\epsilon_{i,1}/2)\lambda_{S_{i,1}}^{\min}\},$$

结合式 (34), 可得到

$$\begin{aligned}\dot{V}_i^1\leqslant&\sum_{l=1,l\neq i}^{N}e_l^{\mathrm{T}}(t)\dot{z}_{l,1}(t)-\Big(\beta_{i,1}-\frac{1}{4\epsilon_{i,1}}-\\&\frac{1}{4\varsigma_{i,1}}-\frac{3}{2}\Big)e_i^2(t)+\frac{1}{2}z_{i,2}^2(t)-\\&a_{i,1}V_{i,1}(t)+\frac{1}{2}h_{i,1}^2(t)+b_{i,1}.\end{aligned}$$

step $j\,(j=2,3,\dots,n-1)$: 由式 (9), 可得到

$$\begin{aligned}\dot{h}_{i,j-1}(t)=&-\frac{h_{i,j-1}(t)}{\tau_{i,j}(t)}-\dot{\hat{\alpha}}_{i,j-1}^{*}(t)=\\&-\frac{\varpi_{i,j-1}(t)}{\tau_{i,j}(t)}-\dot{\hat{\alpha}}_{i,j-1}^{*}(t).\end{aligned}$$

定义性能指标函数为

$$J_{i,j}(z_{i,j})=\int_t^{\infty}c_{i,j}(z_{i,j}(s),\alpha_{i,j}(z_{i,j}(s)))\mathrm{d}s, \tag{35}$$

其中$c_{i,1}(z_{i,j},\alpha_{i,1})=z_{i,j}^2(t)+\alpha_{i,j}^2(t)$. 令$\alpha_{i,j}^{*}(t)$为最优虚拟控制, 将式 (35) 中的$\alpha_{i,j}(t)$替换为$\alpha_{i,j}^{*}(t)$, 可得到最优性能指标函数

$$J_{i,j}^{*}(z_{i,j})=\int_t^{\infty}c_{i,j}(z_{i,j}(s),\alpha_{i,j}^{*}(s))\mathrm{d}s=$$

$$\min_{\alpha_{i,j}\in\Psi(\Omega)}\left\{\int_t^{\infty}c_{i,j}(z_{i,j}(s),\alpha_{i,j}(s))\mathrm{d}s\right\}. \tag{36}$$

由式 (36), 可得到 HJB 方程

$$H_{i,j}(z_{i,j},\alpha^*_{i,j},J^*_{i,j})=z^2_{i,j}(t)+\alpha^{*2}_{i,j}(t)+\frac{\mathrm{d}J^*_{i,j}(z_{i,j})}{\mathrm{d}z_{i,j}}\times(\alpha^*_{i,j}(t)+f_{i,j}-\dot{q}_{i,j}(t))=0.$$

通过求解$\partial H_{i,j}/\partial\alpha^*_{i,j}=0$, 可得到最优虚拟控制

$$\alpha^*_{i,1}(z_{i,j})=-\frac{1}{2}\frac{\mathrm{d}J^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}. \tag{37}$$

进一步地, 将$\dfrac{\mathrm{d}J^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}$分解为

$$\frac{\mathrm{d}J^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}=2\beta_{i,j}z_{i,j}(t)-2\dot{q}_{i,j}(t)+J^o_{i,j}(z_{i,j}). \tag{38}$$

其中: $\beta_{i,j}$为正常数, $J^o_{i,j}(z_{i,j})=-2\beta_{i,j}z_{i,j}(t)+2\dot{q}_{i,j}(t)+\dfrac{\mathrm{d}J^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}$. 将式 (38) 代入 (37), 将最优虚拟控制器改写为

$$\alpha^*_{i,j}(z_{i,j})=-\beta_{i,j}z_{i,j}(t)+\dot{q}_{i,j}(t)-\frac{1}{2}J^o_{i,j}(z_{i,j}). \tag{39}$$

对$J^o_{i,j}(z_{i,j})$使用神经网络逼近, 形式如下所示:

$$J^o_{i,j}(z_{i,j})=W^{*\mathrm{T}}_{i,j}S_{i,j}(z_{i,j})+\varepsilon_{i,j}. \tag{40}$$

将式 (40) 代入 (38) 和 (39), 可得到

$$\frac{\mathrm{d}J^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}=2\beta_{i,j}z_{i,j}(t)-2\dot{q}_{i,j}(t)+W^{*\mathrm{T}}_{i,j}(t)S_{i,j}(z_{i,j})+\varepsilon_{i,j}, \tag{41}$$

$$\alpha^*_{i,j}(z_{i,j})=-\beta_{i,j}z_{i,j}(t)+\dot{q}_{i,j}(t)-\frac{1}{2}W^{*\mathrm{T}}_{i,j}(t)S_{i,j}(z_{i,j})-\frac{1}{2}\varepsilon_{i,j}. \tag{42}$$

由式 (41), 设计用于评估控制性能的评论家, 如下所示:

$$\frac{\mathrm{d}\hat{J}^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}=2\beta_{i,j}z_{i,j}(t)-2\dot{q}_{i,j}(t)+\hat{W}^{\mathrm{T}}_{c,ij}(t)S_{i,j}(z_{i,j}). \tag{43}$$

其中: $\dfrac{\mathrm{d}\hat{J}^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}\in\mathbb{R}$为对$\dfrac{\mathrm{d}J^*_{i,j}(z_{i,j})}{\mathrm{d}(z_{i,j})}$的估计; $\hat{W}_{c,ij}(t)\in\mathbb{R}^{q_{ij}}$为评论家神经网络, 其更新率设计为

$$\dot{\hat{W}}_{c,ij}(t)=-\gamma_{c,ij}S_{i,j}(z_{i,j})S^{\mathrm{T}}_{i,j}(z_{i,j})\hat{W}_{c,ij}(t)-\frac{1}{2}S_{i,j}(z_{i,j})z_{i,j}(t), \tag{44}$$

这里$\gamma_{c,ij}>0$为评论家学习率.

由式 (42), 设计用于实现控制动作的演员, 如下所示:

$$\hat{\alpha}^*_{i,j}=-\beta_{i,j}z_{i,j}(t)+\dot{q}_{i,j}(t)-\frac{1}{2}\hat{W}^{\mathrm{T}}_{a,ij}S_{i,j}(z_{i,j}), \tag{45}$$

其中$\hat{W}_{a,ij}\in\mathbb{R}^{q_{ij}}$为演员神经网络, 其更新率为

$$\dot{\hat{W}}_{a,ij}(t)=-\gamma_{a,ij}S_{i,j}(z_{i,j})S^{\mathrm{T}}_{i,j}(z_{i,j})\times(\hat{W}_{a,ij}(t)-\hat{W}_{c,ij}(t)), \tag{46}$$

这里$\gamma_{a,ij}>0$为演员学习率.

选取 Lyapunov 函数为

$$V^j_i(t)=V_0(t)+\sum_{k=1}^{j}V_{i,k}(t).$$

其中

$$\begin{aligned}&V_{i,k}(t)=\\&(1/2)z^2_{i,k}(t)+(1/2)h^2_{i,k-1}(t)+\\&(1/2)\tilde{W}^{\mathrm{T}}_{c,ik}(t)\tilde{W}_{c,ik}(t)+\\&(1/2)\tilde{W}^{\mathrm{T}}_{a,ik}(t)\tilde{W}_{a,ik}(t),\ 2\leqslant k\leqslant j;\\&\tilde{W}_{c,ik}(t)=\hat{W}_{c,ik}(t)-W^*_{i,k}(t);\\&\tilde{W}_{a,ik}(t)=\hat{W}_{a,ik}(t)-W^*_{i,k}(t).\end{aligned}$$

沿着式 (44) $\sim$ (46), 计算$V^j_i(t)$对时间的导数, 如下所示:

$$\begin{aligned}&\dot{V}^j_i(t)\leqslant\\&\dot{V}_0(t)+\sum_{k=1}^{j-1}\dot{V}_{i,k}(t)-\Big(\beta_{i,j}-\frac{1}{4\varsigma_{i,j}}-\frac{1}{4\epsilon_{i,j}}-\frac{3}{2}\Big)\times\\&z^2_{i,j}(t)-\Big(\frac{\gamma_{c,ij}}{2}-\frac{\gamma_{a,ij}}{2}\Big)\tilde{W}^{\mathrm{T}}_{c,ij}(t)S_{i,j}(z_{i,j})\times\\&S^{\mathrm{T}}_{i,j}(z_{i,j})\tilde{W}_{c,ij}(t)-\Big(\frac{\gamma_{a,ij}}{2}-\frac{\epsilon_{i,j}}{4}\Big)\tilde{W}^{\mathrm{T}}_{a,ij}(t)S_{i,j}(z_{i,j})\times\\&S^{\mathrm{T}}_{i,j}(z_{i,j})\tilde{W}_{a,ij}(t)-\Big(\frac{\gamma_{c,ij}}{2}-\frac{\varsigma_{i,j}}{4}\Big)\hat{W}^{\mathrm{T}}_{c,ij}(t)S_{i,j}(z_{i,j})\times\\&S^{\mathrm{T}}_{i,j}(z_{i,j})\hat{W}_{c,ij}(t)+\frac{1}{2}h^2_{i,j}(t)+\frac{1}{2}z^2_{i,j+1}(t)-\\&\Big(\frac{1}{\tau_{i,j}}-\frac{1}{2}\Big)h^2_{i,j-1}(t)+B_{i,j}\leqslant\\&\sum_{l=1,l\neq i}^{N}e^{\mathrm{T}}_l(t)\dot{z}_{l,1}(t)-\Big(\beta_{i,1}-\frac{1}{4\epsilon_{i,1}}-\frac{1}{4\varsigma_{i,1}}-\frac{3}{2}\Big)e^2_i(t)+\\&\sum_{k=1}^{j}(-a_{i,k}V_{i,k}(t)+b_{i,k})+h^2_{i,j}(t)+\frac{1}{2}z^2_{i,j+1}(t).\end{aligned} \tag{47}$$

其中

$$\begin{aligned}&B_{i,j}=\frac{\gamma_{c,ij}}{2}W^{*\mathrm{T}}_{i,j}(t)S_{i,j}(z_{i,j})S^{\mathrm{T}}_{i,j}(z_{i,j})W^*_{i,j}(t)+\\&\qquad\frac{1}{2}f^2_{i,j}+\frac{1}{2}(\dot{\alpha}^*)^2_{i,j-1},\\&|B_{i,j}|\leqslant b_{i,j};\\&\beta_{i,j}>2+\frac{1}{4\epsilon_{i,j}}+\frac{1}{4\varsigma_{i,j}};\\&\gamma_{c,ij}>\gamma_{a,ij},\ \gamma_{c,ij}>\frac{\varsigma_{i,j}}{2},\ \gamma_{a,ij}>\frac{\epsilon_{i,j}}{2};\end{aligned}$$

$$a_{i,k}=\min\Big\{(\gamma_{c,ik}-\gamma_{a,ik})\lambda_{S_{i,k}}^{\min},\Big(2\beta_{i,k}-4-k_{i,k}-\frac{1}{2\epsilon_{i,k}}-\frac{1}{2\varsigma_{i,k}}\Big),\frac{2}{\tau_{i,k}}-2,\gamma_{a,ik}-(\epsilon_{i,k}/2)\lambda_{S_{i,k}}^{\min}\Big\},\ 2\leqslant k\leqslant j.$$

这里$\lambda_{S_{i,j}}^{\min}$和$\lambda_{S_{i,k}}^{\min}$分别为$S_{i,j}(z_{i,j})S_{i,j}^{\mathrm{T}}(z_{i,j})$和$S_{i,k}(z_{i,k})S_{i,k}^{\mathrm{T}}(z_{i,k})$的最小特征值.

step n: 在反步法的最后一步中将得到实际的最优容错控制.

基于式 (5) 和 (6), 含有执行器故障的输入为

$$\begin{aligned}\psi_i^{\mathrm{T}}u_i(t)=\ &\psi_i^{\mathrm{T}}(\rho_iv_i(t)+\sigma_i(\underline{u}_i-\rho_iv_i(t)))=\\&\psi_i^{\mathrm{T}}((I-\sigma_i)\rho_iv_i(t)+\sigma_i\underline{u}_i)=\\&\psi_i^f(t)u_{i,0}(t)+u_i^f(t).\end{aligned}$$

其中

$$\psi_i^f(t)=\sum_{m\neq m_1,m_2,\ldots,m_p}\rho_{i,m}\psi_{i,m}\omega_{i,m}>0,$$

$$u_i^f(t)=\sum_{m=m_1,m_2,\ldots,m_p}\psi_{i,m}\underline{u}_{i,m}.$$

令$\theta_i(t)=1/\psi_i^f(t)$, 此时$u_{i,0}(t)$可被设计为

$$u_{i,0}(t)=\hat{\theta}_i(t)(U_i(t)-\hat{u}_i^f(t)).\tag{48}$$

其中: $\hat{\theta}_i(t)$和$\hat{u}_i^f(t)$分别为对$\theta_i(t)$和$u_i^f(t)$的估计, $U_i(t)$将会在下文进行设计, 自适应更新率设计如下所示:

$$\dot{\hat{\theta}}_i(t)=-z_{i,n}(t)(U_i(t)-\hat{u}_i^f(t))-2\hat{\theta}_i(t),\tag{49}$$

$$\dot{\hat{u}}_i^f(t)=z_{i,n}(t)-2\hat{u}_i^f(t).\tag{50}$$

定义性能指标函数为

$$J_{i,n}(z_{i,n})=\int_t^{\infty}c_{i,n}(z_{i,n}(s),U_i(z_{i,n}(s)))\mathrm{d}s,\tag{51}$$

其中$c_{i,n}(z_{i,n},U_i)=z_{i,n}^2(t)+U_i^2(t)$为成本函数. 令$U_i^*(t)$为最优容错控制, 将式 (51) 中的$U_i(t)$替换为$U_i^*(t)$, 可得到最优性能指标函数

$$\begin{aligned}J_{i,n}^*(z_{i,n})=&\int_t^{\infty}c_{i,n}(z_{i,n}(s),U_i^*(s))\mathrm{d}s=\\&\min_{U_i\in\Psi(\Omega)}\Big\{\int_t^{\infty}c_{i,n}(z_{i,n}(s),U_i(s))\mathrm{d}s\Big\}.\end{aligned}\tag{52}$$

由式 (52), 可得到 HJB 方程

$$\begin{aligned}&H_{i,n}(z_{i,n},U_i^*,J_{i,n}^*)=\\&z_{i,n}^2(t)+U_i^{*2}(t)+\frac{\mathrm{d}J_{i,n}^*(z_{i,n})}{\mathrm{d}z_{i,n}}\times\\&(\psi_i^f(t)(\hat{\theta}_i(t)(U_i^*(t)-\hat{u}_i^f(t)))+u_i^f(t)+\\&f_{i,n}-\dot{q}_{i,n}(t))=0.\end{aligned}$$

通过求解$(\partial H_{i,n}/\partial U_i^*)=0$, 可得到最优控制

$$U_i^*(z_{i,n})=-\frac{1}{2}\psi_i^f(t)\hat{\theta}_i(t)\frac{\mathrm{d}J_{i,n}^*(z_{i,n})}{\mathrm{d}(z_{i,n})}.\tag{53}$$

为了实现最优容错控制, 将$\dfrac{\mathrm{d}J_{i,n}^*(z_{i,n})}{\mathrm{d}(z_{i,n})}$分解为

$$\frac{\mathrm{d}J_{i,n}^*(z_{i,n})}{\mathrm{d}(z_{i,n})}=\frac{2\beta_{i,n}z_{i,n}(t)-2\dot{q}_{i,n}(t)}{\psi_i^f(t)\hat{\theta}_i(t)}+\frac{J_{i,n}^o(z_{i,n})}{\psi_i^f(t)\hat{\theta}_i(t)}.\tag{54}$$

其中: $\beta_{i,n}$为正常数, $J_{i,n}^o(z_{i,n})=-2\beta_{i,n}z_{i,n}(t)+2\dot{q}_{i,n}(t)+\dfrac{\mathrm{d}J_{i,n}^*(\hat{z}_{i,n})}{\mathrm{d}(z_{i,n})}\psi_i^f(t)\hat{\theta}_i(t)$. 将式 (54) 代入 (53), 有

$$U_i^*(z_{i,n})=-\beta_{i,n}z_{i,n}(t)+\dot{q}_{i,n}(t)-\frac{1}{2}J_{i,n}^o(z_{i,n}).\tag{55}$$

对$J_{i,n}^o(z_{i,n})$使用神经网络逼近, 可得到

$$J_{i,n}^o(z_{i,n})=W_{i,n}^{*\mathrm{T}}(t)S_{i,n}(z_{i,n})+\varepsilon_{i,n}.\tag{56}$$

将式 (56) 代入 (54) 和 (55), 可得到

$$\begin{aligned}\frac{\mathrm{d}J_{i,n}^*(z_{i,n})}{\mathrm{d}(z_{i,n})}=&\frac{1}{\psi_i^f(t)\hat{\theta}_i(t)}(2\beta_{i,n}z_{i,n}(t)-2\dot{q}_{i,n}(t)+\\&W_{i,n}^{*\mathrm{T}}S_{i,n}(z_{i,n})+\varepsilon_{i,n}),\end{aligned}\tag{57}$$

$$\begin{aligned}U_i^*(z_{i,n})=&-\beta_{i,n}z_{i,n}(t)+\dot{q}_{i,n}(t)-\frac{1}{2}\varepsilon_{i,n}-\\&\frac{1}{2}W_{i,n}^{*\mathrm{T}}(t)S_{i,n}(z_{i,n}).\end{aligned}\tag{58}$$

由式 (57), 设计用于评估控制性能的评论家, 有

$$\begin{aligned}\frac{\mathrm{d}\hat{J}_{i,n}^*(z_{i,n})}{\mathrm{d}(z_{i,n})}=&\frac{1}{\psi_i^f(t)\hat{\theta}_i(t)}(2\beta_{i,n}z_{i,n}(t)-2\dot{q}_{i,n}(t)+\\&\hat{W}_{c,in}^{\mathrm{T}}(t)S_{i,n}(z_{i,n})).\end{aligned}\tag{59}$$

其中: $\dfrac{\mathrm{d}\hat{J}_{i,n}^*(z_{i,n})}{\mathrm{d}(z_{i,n})}\in\mathbb{R}$为对$\dfrac{\mathrm{d}J_{i,n}^*(z_{i,n})}{\mathrm{d}(z_{i,n})}$的估计; $\hat{W}_{c,in}(t)\in\mathbb{R}^{q_{in}}$为评论家神经网络权重, 其更新率如下所示:

$$\begin{aligned}\dot{\hat{W}}_{c,in}(t)=&-\gamma_{c,in}S_{i,n}(z_{i,n})S_{i,n}^{\mathrm{T}}(z_{i,n})\hat{W}_{c,in}(t)-\\&\frac{1}{2}S_{i,n}(z_{i,n})z_{i,n}(t),\end{aligned}\tag{60}$$

这里$\gamma_{c,in}>0$为评论家学习率.

由式 (58), 设计用于实现控制动作的演员, 即

$$\begin{aligned}\hat{U}_i^*(t)=&-\beta_{i,n}z_{i,n}(t)+\dot{q}_{i,n}(t)-\\&\frac{1}{2}\hat{W}_{a,in}^{\mathrm{T}}(t)S_{i,n}(z_{i,n}),\end{aligned}\tag{61}$$

其中$\hat{W}_{a,in}(t)\in\mathbb{R}^{q_{in}}$为演员神经网络权重, 其更新率如下所示:

$$\begin{aligned}\dot{\hat{W}}_{a,in}(t)=&-\gamma_{a,in}S_{i,n}(z_{i,n})S_{i,n}^{\mathrm{T}}(z_{i,n})\times\\&(\hat{W}_{a,in}(t)-\hat{W}_{c,in}(t)),\end{aligned}\tag{62}$$

这里$\gamma_{a,in}>0$为演员学习率.

对于智能体i, 选取整个反步控制系统的Lyapunov函数为

$$V_i^n(t)=V_0(t)+\sum_{k=1}^{n}V_{i,k}(t).$$

其中

$$\begin{aligned}V_{i,n}(t)=&(1/2)z_{i,n}^2(t)+(1/2)h_{i,n-1}^2(t)+\\&(1/2)\tilde{W}_{c,in}^{\mathrm{T}}(t)\tilde{W}_{c,in}(t)+\\&(1/2)\tilde{W}_{a,in}^{\mathrm{T}}(t)\tilde{W}_{a,in}(t)+\\&\frac{\psi_i^f(t)}{2}\tilde{\theta}_i^2(t)+\frac{1}{2}(\tilde{u}_i^f(t))^2,\end{aligned}$$

$$\tilde{W}_{c,in}=\hat{W}_{c,in}-W_{i,n}^*,$$

$$\tilde{W}_{a,in}(t)=\hat{W}_{a,in}(t)-W_{i,n}^*(t),$$

$$\tilde{\theta}_i(t)=\theta_i(t)-\hat{\theta}_i(t),$$

$$\tilde{u}_i^f(t)=u_i^f(t)-\hat{u}_i^f(t).$$

沿着式(60) ∼ (62), 计算$V_i^n(t)$对时间的导数, 有

$$\begin{aligned}&\dot{V}_i^n(t)\leqslant\\&\dot{V}_0(t)+\sum_{k=1}^{n-1}\dot{V}_{i,k}(t)-\Big(\beta_{i,n}-\frac{1}{2}-\frac{1}{4\epsilon_{i,n}}-\\&\frac{1}{4\varsigma_{i,n}}\Big)z_{i,n}^2(t)-\Big(\frac{\gamma_{c,in}}{2}-\frac{\gamma_{a,in}}{2}\Big)\tilde{W}_{c,in}^{\mathrm{T}}(t)\times\\&S_{i,n}(z_{i,n})S_{i,n}^{\mathrm{T}}(z_{i,n})\tilde{W}_{c,in}(t)-\Big(\frac{\gamma_{a,in}}{2}-\frac{\epsilon_{i,n}}{4}\Big)\times\\&\tilde{W}_{a,in}^{\mathrm{T}}(t)S_{i,n}(z_{i,n})S_{i,n}^{\mathrm{T}}(z_{i,n})\tilde{W}_{a,in}(t)-\\&\Big(\frac{\gamma_{c,in}}{2}-\frac{\varsigma_{i,n}}{4}\Big)\hat{W}_{c,in}^{\mathrm{T}}(t)S_{i,n}(z_{i,n})S_{i,n}^{\mathrm{T}}(z_{i,n})\times\\&\hat{W}_{c,in}(t)-\Big(\frac{1}{\tau_{i,n}}-\frac{1}{2}\Big)h_{i,n-1}^2(t)-\psi_i^f(t)\tilde{\theta}_i^2(t)-\\&(\tilde{u}_i^f(t))^2+B_{i,n}\leqslant\\&\sum_{l=1,l\neq i}^{N}e_l^{\mathrm{T}}(t)\dot{z}_{l,1}(t)-\Big(\beta_{i,1}-\frac{3}{2}-\frac{1}{4\epsilon_{i,1}}-\\&\frac{1}{4\varsigma_{i,1}}\Big)e_i^2(t)+\sum_{k=1}^{n}(-a_{i,k}V_{i,k}(t)+b_{i,k}).\end{aligned}\tag{63}$$

其中

$$\begin{aligned}&B_{i,n}=\\&\frac{\gamma_{c,in}}{2}W_{i,n}^{*\mathrm{T}}(t)S_{i,n}(z_{i,n})S_{i,n}^{\mathrm{T}}(z_{i,n})W_{i,n}^*(t)+\frac{1}{2}f_{i,n}^2+\\&\frac{1}{2}(\dot{\alpha}^*(t))_{i,n-1}^2+\psi_i^f(t)\theta_i^2(t)+(u_i^f(t))^2,\end{aligned}$$

$$|B_{i,n}|\leqslant b_{i,n};$$

$$\beta_{i,n}>1+\frac{1}{4\epsilon_{i,j}}+\frac{1}{4\varsigma_{i,j}};$$

$$\gamma_{c,in}>\gamma_{a,ij},\ \gamma_{c,in}>\frac{\varsigma_{i,j}}{2},\ \gamma_{a,in}>\frac{\epsilon_{i,j}}{2};$$

$$\begin{aligned}a_{i,n}=\min\Big\{&(\gamma_{c,in}-\gamma_{a,in})\lambda_{S_{i,n}}^{\min},\\&\Big(2\beta_{i,n}-2-\frac{1}{2\epsilon_{i,n}}-\frac{1}{2\varsigma_{i,n}}\Big),\\&\gamma_{a,in}-(\epsilon_{i,n}/2)\lambda_{S_{i,n}}^{\min},\\&\frac{2}{\tau_{i,n}}-2,2\psi_i^f,2\Big\}.\end{aligned}$$

这里$\lambda_{S_{i,n}}^{\min}$为$S_{i,n}(z_{i,n})S_{i,n}^{\mathrm{T}}(z_{i,n})$的最小特征值.

综上, 在上述最优反步法的每一步中, 通过构建评论家神经网络和演员神经网络, 引入强化学习算法以提升控制效果. 其中: 评论家神经网络用于评估系统的优化性能, 而演员神经网络则负责执行具体的控制策略. 在此基础上, 通过自适应容错控制机制, 进一步实现容错最优控制, 确保系统的鲁棒性和可靠性. 评论家-演员神经网络最优容错控制算法如算法1所示.

算法1 评论家-演员神经网络最优容错控制算法.

---

1. 初始化系统状态$x_{i,j}(0)$, 参考信号y_l, 神经网络权重$\hat{W}_{a,ij}(0)\in R^6$, $\hat{W}_{c,ij}(0)\in R^6$和各项参数$\gamma_{a,ij}$, $\gamma_{c,ij}$, $\beta_{i,j}$, $\omega_{i,j}$.

for 智能体$i=1,2,\ldots,N$ do

  对于$j=1,2,\ldots,n$

  计算$z_{i,j}(t)$, 同步误差$e_i(t)$

end

step $1,2,\ldots,n-1$

1. 通过求解$\partial H_{i,j}/\partial\alpha_{i,j}^*=0$, 得到最优虚拟控制$\alpha_{i,j}^*$.

2. 设计评估控制性能的评论家$\dfrac{\mathrm{d}\hat{J}_{i,1}^*(e_i)}{\mathrm{d}(e_i)}$, $\dfrac{\mathrm{d}\hat{J}_{i,j}^*(z_{i,j})}{\mathrm{d}(z_{i,j})}$, 评论家更新率为$\dot{\hat{W}}_{c,ij}(t)$.

3. 设计实现控制动作的演员$\hat{\alpha}_{i,j}^*$, 演员更新率为$\dot{\hat{W}}_{a,ij}(t)$.

step n

1. 设计故障估计的更新率$\dot{\hat{\theta}}_i(t)$, $\dot{\hat{u}}_i^f(t)$.

2. 通过求解$(\partial H_{i,n}/\partial U_i^*)=0$, 得到最优控制$U_i^*$.

3. 设计评估控制性能的评论家$\dfrac{\mathrm{d}\hat{J}_{i,n}^*(z_{i,j})}{\mathrm{d}(z_{i,n})}$, 评论家更新率为$\dot{\hat{W}}_{c,in}(t)$.

4. 设计实现容错控制动作的演员$\hat{U}_i^*$, 演员更新率为$\dot{\hat{W}}_{a,in}(t)$.

直至$\dot{V}(t)\leqslant-\xi V(t)+b$.

---

3 稳定性分析

定理1 在评论家(23)、(43)和(59)的更新率分别为式(24)、(44)和(60), 演员(25)、(45)和(61)的更新率分别为式(26)、(46)和(62)下, 多智能体系统(1)能够在预设时间内实现编队控制.

证明　选取整个多智能体系统 (1) 的总 Lyapunov 函数为

$$V(t)=V_0(t)+\sum_{i=1}^{N}\sum_{k=1}^{n}V_{i,k}.$$

由式 (63) 可知, V 对时间的导数为

$$\begin{aligned}\dot V(t)\leqslant&-\sum_{i=1}^{N}\Big(\beta_{i,1}-\frac{3}{2}-\frac{1}{4\epsilon_{i,1}}-\frac{1}{4\varsigma_{i,1}}\Big)e_i^2(t)+\\&\sum_{i=1}^{N}\sum_{k=1}^{n}(-a_{i,k}V_{i,k}(t)+b_{i,k})\leqslant\\&-2\beta\lambda_{\bar L}^{\min}V_0(t)-a\sum_{i=1}^{N}\sum_{k=1}^{n}V_{i,k}(t)+b\leqslant\\&-\xi V(t)+b.\end{aligned}\tag{64}$$

其中

$$\beta=\min_{i=1,2,\dots,N}\Big\{\beta_{i,1}-\frac{3}{2}-\frac{1}{4\epsilon_{i,1}}-\frac{1}{4\varsigma_{i,1}}\Big\},$$

$$a=\min_{\substack{i=1,2,\dots,N,\\k=1,2,\dots,n}}\{a_{i,k}\},$$

$$b=\sum_{i=1}^{N}\sum_{k=1}^{n}b_{i,k},$$

$$\xi=\min\{2\beta\lambda_{\bar L}^{\min},a\}.$$

由文献 [18, 23] 可知, 系统 (1) 可在实际预设时间内实现编队控制. □

注 1　由于本文同时考虑了 LIP 和 LOE 故障, 且关于故障的先验信息是未知的, 控制器的设计变得更加复杂. 为此, 所设计容错控制器 (48) 包含了自适应故障更新率 (49) 和 (50).

4　仿真例子

考虑如下由 4 个智能体构成的非线性系统:

$$\begin{aligned}&\dot x_{i,1}(t)=x_{i,2}-a_{i,1}\cos^2(x_{i,1})+a_{i,2}\sin(x_{i,1}),\\&\dot x_{i,2}(t)=\psi_i^{\mathrm T}u_i-b_{i,1}x_{i,2}\sin(x_{i,1})+b_{i,2}\cos(x_{i,2}),\\&\qquad i=1,2,3,4.\end{aligned}\tag{65}$$

其中: $a_{1,1}=1.5$, $a_{2,1}=-0.8$, $a_{3,1}=0.6$, $a_{4,1}=-1.3$, $a_{1,2}=-0.8$, $a_{2,2}=0.4$, $a_{3,2}=-0.7$, $a_{4,2}=0.8$; $b_{1,1}=0.7$, $b_{2,1}=1.4$, $b_{3,1}=-1.5$, $b_{4,1}=-1.2$, $b_{1,2}=0.5$, $b_{2,2}=-0.6$, $b_{3,2}=1.1$, $b_{4,2}=-1.9$. 输入信号 $u_i=[u_{i,1},u_{i,2}]^{\mathrm T}$, $\psi_i=[\psi_{i,1},\ \psi_{i,2}]^{\mathrm T}=[1,1]^{\mathrm T}$. 领导者信号为 $y_l=2\cos(0.6t)$.

智能体与领导者间的拓扑结构如图 2 所示.

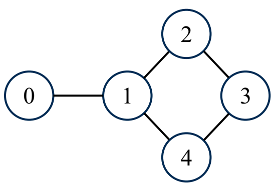


图2　通信拓扑结构

选取合适的初值如下: $x_{1,1}(0)=0$, $x_{2,1}(0)=1$, $x_{3,1}(0)=3$, $x_{4,1}(0)=4$, $x_{1,2}(0)=0.5$, $x_{2,2}(0)=1.5$, $x_{3,2}(0)=3.5$, $x_{4,2}(0)=4.5$; $\hat W_{a,i1}(0)=\hat W_{c,i1}(0)=[2.3,2.3,\dots,2.3]\in R^6$, $\hat W_{a,i2}(0)=\hat W_{c,i2}(0)=[3.4,3.4,\dots,3.4]\in R^6$; $\gamma_{a,11}=\gamma_{a,31}=19.5$, $\gamma_{a,21}=\gamma_{a,41}=59.5$, $\gamma_{c,11}=\gamma_{c,31}=20$, $\gamma_{c,21}=\gamma_{c,41}=60$, $\gamma_{a,12}=24$, $\gamma_{a,22}=\gamma_{a,23}=\gamma_{a,24}=44.5$, $\gamma_{c,12}=25$, $\gamma_{c,22}=\gamma_{c,32}=\gamma_{c,42}=45$; $\beta_{i,1}=6.48$, $\beta_{i,2}=2.78$. 执行器参数 $\omega_{i,1}=\omega_{i,2}=1$. $\hat u_i^f(0)=0$, $\hat\theta_i(0)=1$. 当 $t\geqslant 1$ s 时, 发生 LOE 故障, 即 $u_{i,1}=0.7u_{i,0}$; 当 $t\geqslant 1.5$ s 时, 发生 LIP 故障, 即 $u_{i,2}=2$.

当选取预设时间 $T_f=3$ s, 编队参数 $h_1=-1$, $h_2=1$, $h_3=2$, $h_4=3$ 时, 图 3 ∼ 图 8 为仿真结果. 由图 3 可见, 所设计预设时间容错编队控制方案能够实现指定的队形控制. 图 4 和图 5 分别为反步法中第 1 步和第 2 步的成本函数 $c_{i,1}(e_i,\hat\alpha_{i,1}^*)$ 和 $c_{i,2}(z_{i,2},\hat U_i^*)$. 图 6 和图 7 显示了评论家和演员的神经网络权重的有界性.

为了验证所提出方法在实现编队控制的同时具有良好的资源节约性能, 在相同参数条件下, 分别采用所提出方法和文献 [18] 设计控制器, 定义总性能指标为 $J_i=\int_0^{20}(z_{i,1}^2+\alpha_{i,1}^2+z_{i,2}^2+U_i^2)\mathrm{d}t$. 图 8 为

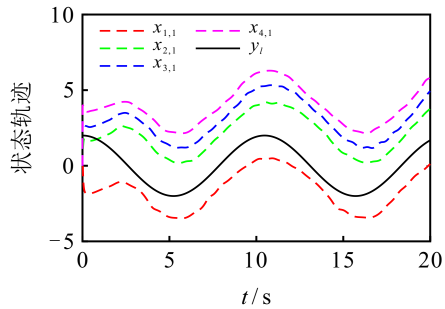
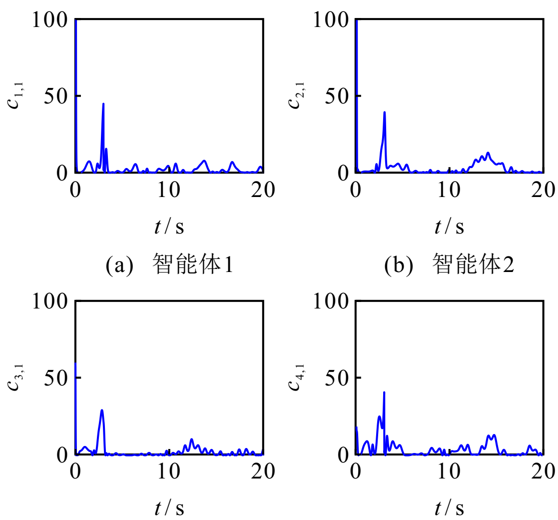


图3　预设时间编队性能



(a)　智能体1　　(b)　智能体2

(c)　智能体3　　(d)　智能体4

图4　成本函数的第一维图像

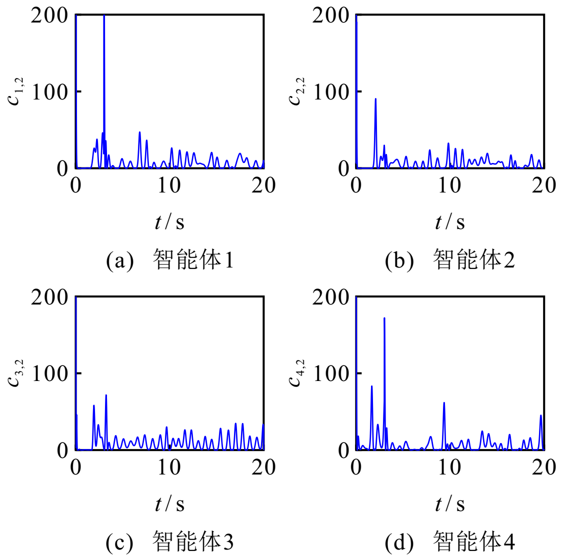


图5 成本函数的第二维图像

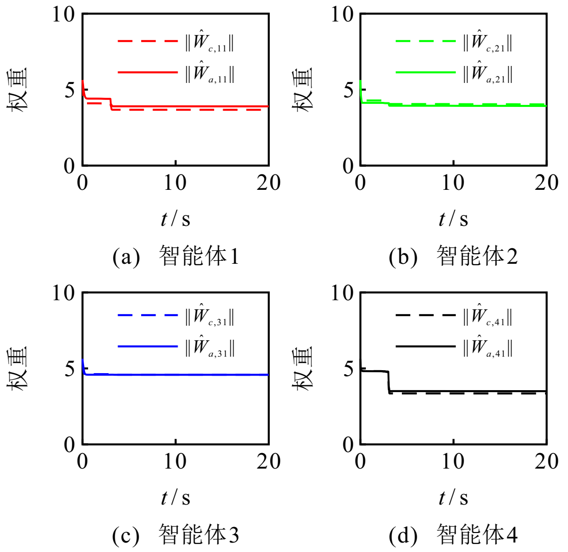


图6 评论家和演员神经网络权重的第一维图像

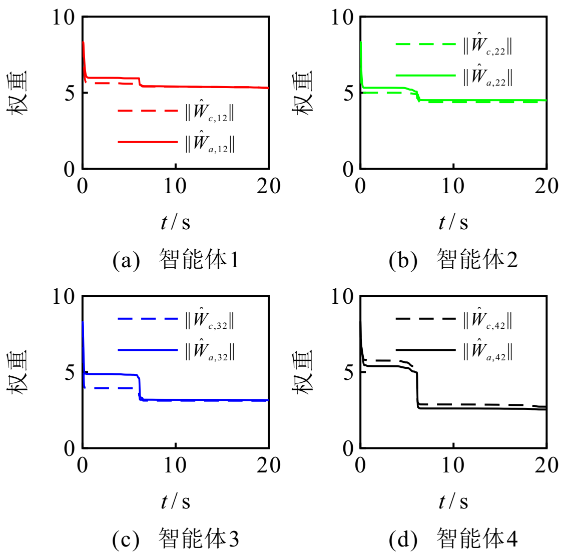


图7 评论家和演员神经网络权重的第二维图像

在未考虑优化的情况下, 文献 [18] 的方法与所提出方法的对比结果. 可以看出, 本文的性能指标 J_i 明显

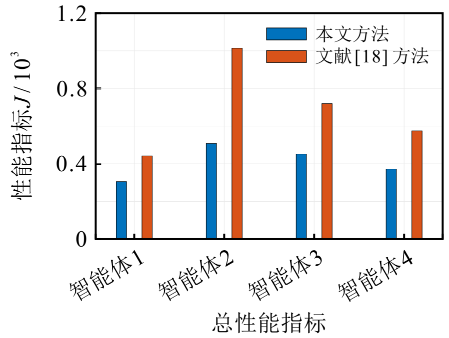


图8 本文方法和文献 [18] 方法的总性能指标

低于文献 [18] 的方法.

5 结 论

本文研究了基于强化学习的多智能体系统最优容错编队控制问题. 设计了基于反步法的容错控制器, 保证了智能体在预设时间内实现编队. 通过设计评论家和演员的神经网络逼近方法, 并在反步法的每一步中执行强化学习算法, 实现了多智能体系统的最优控制. 仿真验证了所提出方案的有效性.

参考文献 (References)

[1] Kim A R, Keshmiri S, Blevins A, et al. Control of multi-agent collaborative fixed-wing UASs in unstructured environment[J]. Journal of Intelligent & Robotic Systems, 2020, 97(1): 205-225.

[2] Gherairi S. Design and implementation of an intelligent energy management system for smart home utilizing a multi-agent system[J]. Ain Shams Engineering Journal, 2023, 14(3): 101897.

[3] Yu D X, Philip C C L, Xu H. Fuzzy swarm control based on sliding-mode strategy with self-organized omnidirectional mobile robots system[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(4): 2262-2274.

[4] 向雨竹, 邹文成, 郭健, 等. 无人机领导的多无人艇系统固定时间优化编队控制[J]. 控制与决策, 2025, 40(1): 223-230.
(Xiang Y Z, Zou W C, Guo J, et al. Fixed-time optimal formation control for multi-unmanned surface vessels under the leadership of unmanned aerial vehicle[J]. Control and Decision, 2025, 40(1): 223-230.)

[5] 高升, 张伟, 郭延宁. 输入饱和的机器人固定时间预设性能容错控制[J]. 控制与决策, 2025, 40(9): 2639-2646.
(Gao S, Zhang W, Guo Y N, et al. Prescribed performance-based fixed-time fault-tolerant control for robotic systems with input saturation[J]. Control and Decision, 2025, 40(9): 2639-2646.)

[6] Jia F L, He X. Adaptive fault-tolerant tracking control for discrete-time nonstrict-feedback nonlinear systems with stochastic noises[J]. IEEE Transactions on Automation Science and Engineering, 2024, 21(3): 3344-3356.

[7] Xiao J W, Liu Y C. Adaptive neural network dynamic

event-triggered consensus control for nonlinear multi-agent systems subject to sensor deception attacks and actuator faults[J]. Nonlinear Dynamics, 2024, 112(22): 20019-20034.

[8] Wang D, Li C, Liu D R, et al. Data-based robust optimal control of continuous-time affine nonlinear systems with matched uncertainties[J]. Information Sciences, 2016, 366: 121-133.

[9] Bhasin S, Kamalapurkar R, Johnson M, et al. A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems[J]. Automatica, 2013, 49(1): 82-92.

[10] 高伟男, 杨涛, 柴天佑. 基于自适应动态规划和梯度下降法的自适应最优输出调节[J]. 控制与决策, 2023, 38(8): 2425-2432.
(Gao W N, Yang T, Chai T Y. Adaptive optimal output regulation based on adaptive dynamic programming and gradient descent method[J]. Control and Decision, 2023, 38(8): 2425-2432.)

[11] Wen G X, Philip C C L. Optimized backstepping consensus control using reinforcement learning for a class of nonlinear strict-feedback-dynamic multi-agent systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(3): 1524-1536.

[12] 蔡玉良, 吕春慧, 何强, 等. 基于强化学习的完全分布式事件驱动二分一致性算法[J]. 计算机科学, 2025, 52(2): 279-290.
(Cai Y L, Lv C H, He Q, et al. Fully distributed event driven bipartite consensus algorithm based on reinforcement learning[J]. Computer Science, 2025, 52(2): 279-290.)

[13] Wen G X, Li B, Niu B. Optimized backstepping control using reinforcement learning of observer-critic-actor architecture based on fuzzy system for a class of nonlinear strict-feedback systems[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(10): 4322-4335.

[14] Lan J, Liu Y J, Xu T Y, et al. Adaptive fuzzy fast finite-time formation control for second-order MASs based on capability boundaries of agents[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(9): 3905-3917.

[15] 王加朋, 杨家宁, 罗家祥, 等. 切换拓扑下多智能体的动态事件触发固定时间一致性控制[J]. 控制理论与应用, 2025, 42(1): 87-95.
(Wang J P, Yang J N, Luo J X, et al. Dynamic event-triggered fixed-time consensus control of multi-agent systems under switching topologies[J]. Control Theory & Applications, 2025, 42(1): 87-95.)

[16] Jiménez-Rodríguez E, Muñoz-Vázquez A J, Sánchez-Torres J D, et al. A Lyapunov-like characterization of predefined-time stability[J]. IEEE Transactions on Automatic Control, 2020, 65(11): 4922-4927.

[17] Xu H, Yu D X, Liu Y J. Observer-based fuzzy adaptive predefined time control for uncertain nonlinear systems with full-state error constraints[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(3): 1370-1382.

[18] Ren D J, Liu Y J, Lan J. Event-based predefined-time fuzzy formation control for nonlinear multiagent systems with unknown disturbances[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(6): 3497-3507.

[19] Liu B J, Hou M S, Wu C H, et al. Predefined-time backstepping control for a nonlinear strict-feedback system[J]. International Journal of Robust and Nonlinear Control, 2021, 31(8): 3354-3372.

[20] Li P, Yang G H. Backstepping adaptive fuzzy control of uncertain nonlinear systems against actuator faults[J]. Journal of Control Theory and Applications, 2009, 7(3): 248-256.

[21] Tang X D, Tao G, Joshi S M. Adaptive actuator failure compensation for nonlinear MIMO systems with an aircraft control application[J]. Automatica, 2007, 43(11): 1869-1883.

[22] Dai S L, He S D, Ma Y F, et al. Distributed cooperative learning control of uncertain multiagent systems with prescribed performance and preserved connectivity[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(7): 3217-3229.

[23] Liu B J, Wang W C, Li Y K, et al. Adaptive quantized predefined-time backstepping control for nonlinear strict-feedback systems[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2022, 69(9): 3859-3863.

作者简介

申思凯 (2001−), 男, 博士生, 主要研究方向为复杂网络、神经网络、多智能体协同控制, E-mail: 2020230565@jsnu.edu.cn;

江南 (1979−), 女, 高级实验师, 主要研究方向为多智能体协同控制, E-mail: lxyjn@163.com;

刘小洋 (1979−), 男, 教授, 博士, 博士生导师, 主要研究方向为复杂网络、神经网络、多智能体协同控制, E-mail: xyliu@jsnu.edu.cn.