

异构非线性多智能体系统无模型输出一致性控制

孙一仆^{1,2,3} 陈鑫^{1,2,3} 贺文朋^{1,2,3} 余锦华⁴ 吴敏^{1,2,3}

摘要 针对异构非线性多智能体系统 (Multi-agent system, MAS) 的输出一致性控制难题, 设计了一种基于同胚分布式控制协议的无模型方法. 通过将输出反馈线性化理论与自适应动态规划相结合, 可以在不需要精确系统模型的情况下实现非线性智能体的线性化, 简化分布式控制器的设计复杂性. 具体而言, 设计一种双层分布式控制结构, 在物理空间层通过无模型反馈线性化方法实现未知系统线性化, 在微分同构空间层利用线性控制技术进行分布式共识控制. 通过两个实验验证了所提方法在处理未知异构非线性多智能体系统中的有效性, 将传统的线性分布式控制方法扩展到未知非线性多智能体系统的控制器设计.

关键词 非线性多智能体系统, 无模型输出共识控制, 微分同胚, 输入输出反馈线性化, 自适应动态规划

引用格式 孙一仆, 陈鑫, 贺文朋, 余锦华, 吴敏. 异构非线性多智能体系统无模型输出一致性控制. 自动化学报, 2025, 51(3): 604–616

DOI 10.16383/j.aas.c240459 **CSTR** 32138.14.j.aas.c240459

Model-free Output Consensus Control for Heterogeneous Nonlinear Multi-agent Systems

SUN Yi-Pu^{1,2,3} CHEN Xin^{1,2,3} HE Wen-Peng^{1,2,3} SHE Jin-Hua⁴ WU Min^{1,2,3}

Abstract A model-free method based on homeomorphic distributed control protocol is proposed to address the output consensus control problem of heterogeneous nonlinear multi-agent systems (MASs). By integrating output feedback linearization theory with adaptive dynamic programming, this approach linearizes nonlinear agents without requiring precise system models, simplifying the design of distributed controllers. Specifically, a two-layer distributed control structure is designed: In the physical space layer, model-free feedback linearization is applied to linearize unknown systems, while in the diffeomorphic space layer, linear control techniques are used for distributed consensus control. The effectiveness of the proposed method in handling unknown heterogeneous nonlinear multi-agent systems is validated through two experiments, extending traditional linear distributed control methods to the design of controllers for unknown nonlinear multi-agent systems.

Key words Nonlinear multi-agent system, model-free output consensus control, diffeomorphic, input-output feedback linearization, adaptive dynamic programming

Citation Sun Yi-Pu, Chen Xin, He Wen-Peng, She Jin-Hua, Wu Min. Model-free output consensus control for heterogeneous nonlinear multi-agent systems. *Acta Automatica Sinica*, 2025, 51(3): 604–616

在刚性航天器一致性^[1]和欧拉-拉格朗日系统

的编队控制^[2]等应用场景中, 直接测量和反馈系统的输出变量更为方便和可靠. 例如, 在多无人车编队中, 通过全球定位系统 (Global positioning system, GPS) 等技术直接测量每辆车的位置和速度, 比估计和控制内部状态更简单易行^[3]. 因此, 输出一致性跟踪控制在多智能体系统 (Multi-agent system, MAS) 的工程应用中更具实用性.

线性控制方法在传统多智能体控制理论中占据重要地位^[4–5], 其通过将复杂的非线性系统线性化为多个局部线性系统来简化控制问题^[6–8]. 然而, 异构非线性多智能体系统的高度非线性和动态特性使得这些方法难以有效应用. 具体来说, 线性控制方法在处理大范围动态变化和强耦合非线性特性时表现出较大局限性, 例如在多机器人协同任务中, 简化模型无法准确地反映各机器人不同的动力学特性, 导致控制精度和鲁棒性下降.

收稿日期 2024-07-01 录用日期 2024-11-11

Manuscript received July 1, 2024; accepted November 11, 2024
高等学校学科创新引智计划 (B17040), 湖北省科技创新重大专项 (2020AEA010), 国家自然科学基金 (61873248), 湖北省自然科学基金 (2020CFA031), 国家电网公司科技专项 (52153216000R) 资助
Supported by the 111 Project (B17040), Technical Innovation Major Project of Hubei Province (2020AEA010), National Natural Science Foundation of China (61873248), Natural Science Foundation of Hubei Province (2020CFA031), and Science and Technology Project of State Grid Corporation of China (52153216000R)

本文责任编辑 孙健

Recommended by Associate Editor SUN Jian

1. 中国地质大学 (武汉) 自动化学院 武汉 430074 中国 2. 复杂系统先进控制与智能化湖北省重点实验室 武汉 430074 中国 3. 地球探测智能化技术教育部工程研究中心 武汉 430074 中国 4. 东京工业大学 东京 192-0982 日本

1. School of Automation, China University of Geosciences, Wuhan 430074, China 2. Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China 3. Engineering Research Center of Intelligent Technology for Geo-exploration, Ministry of Education, Wuhan 430074, China 4. Tokyo University of Technology, Tokyo 192-0982, Japan

非线性控制方法直接处理系统的非线性特性, 通过 Lyapunov 方法^[9-10]、反馈线性化^[11-12]等理论设计控制策略. 尽管理论上能够解决线性方法的不足, 但其应用面临诸多困难: 需要精确的系统模型、设计和实现复杂, 特别是在异构多智能体系统中, 要求各智能体之间的协调和实时响应, 增加了计算量和实现难度^[13]. 此外, 非线性控制方法在处理高维度和外界扰动时, 稳定性和鲁棒性也受到挑战.

无模型自适应动态规划方法作为一种数据驱动的控制策略^[14]逐渐受到关注, 通过与环境交互, 基于奖励机制自主学习最优策略, 无需系统模型即可实现复杂任务的控制. Jiang 等^[15]提出一种数据驱动的自适应动态规划方法, 使用输入和输出序列作为基础状态的等效表示, 解决了部分可观测系统状态的离散线性多智能体系统的最优输出一致性控制问题. 对于部分未知动力学的严格反馈非线性多智能体系统, 文献^[16]在输出调节理论下, 提出基于实测数据结合神经网络和自适应动态规划求解最优输出反馈控制的方法. 然而, 对于异构非线性系统的无模型输出一致性控制研究仍处于起步阶段.

无模型学习控制方法也存在明显不足: 自适应动态规划方法的训练过程对参数选择和奖励设计高度敏感, 可能导致策略的鲁棒性和稳定性不佳; 可解释性差, 使得控制策略的进一步调整变得困难; 在系统跟踪时变信号时, 自适应动态规划方法本身不具备预测未来状态的能力, 这使其更适合镇定控制而非跟踪控制.

混合控制策略利用不同方法的互补特性解决异构非线性多智能体系统的一致性控制问题^[17]. 结合自适应动态规划与经典控制理论, 可以在数据驱动的基础上引入稳定性分析, 提升控制策略的可靠性^[18]. 然而, 混合控制策略设计和实现难度大, 需在不同方法之间找到平衡点, 确保整体系统的稳定性和性能.

上述背景下, 本工作结合输入输出反馈线性化理论和自适应动态规划, 从简化分布式控制器设计、增加控制器可解释性、降低学习对奖励设计的敏感度的角度出发, 开发了异构非线性多智能体系统的无模型输出一致性控制方法. 具体来说, 通过构建一个同胚分布式两层控制结构, 将异构非线性多智能体系统的无模型输出一致性控制问题转化为两个问题进行求解: 在物理空间层中利用观测数据, 提出能够动态调整奖励信号的两阶段双启发式自适应动态规划方法实现非线性系统的无模型输入输出反馈线性化; 在同胚线性化空间层中, 基于线性化系统设计一致性分布式控制器, 实现被控多智能体系统的输出一致性控制. 本文的主要创新点和贡献如下:

1) 现有分布式控制方法在处理异构多智能体输出一致性控制时^[15-16], 因模型未知和非线性动态的影响, 会造成黎卡提方程或贝尔曼方程求解困难的问题. 为此, 本文提出一种基于无模型反馈线性化的同胚分布式控制协议, 不依赖精确模型的情况下实现输出一致性控制. 不同于传统无模型分布式控制方法, 分层分布式控制协议包含两层控制策略, 在物理空间层通过构建自适应动态规划算法求解无模型反馈线性化控制器, 将未知非线性多智能体系统转化为已知的线性系统. 结合同胚空间层的一致性控制协议, 该线性化系统可以根据协同任务的性能需求进行预设计或二次设计, 当控制任务发生改变时无需重新学习, 从而降低一致性策略设计难度.

2) 解决物理空间层中反馈线性化控制器对精确模型的依赖问题是分层分布式方法实施的关键, 本文设计一种基于两阶段迭代学习的无模型自适应动态规划算法. 算法在值函数学习过程中引入目标依赖, 可以动态调整奖励信号以适应异构的智能体, 无需设计不同奖励信号, 同时通过一个双启发式评价网络实现线性化控制策略快速更新.

1 图论和问题描述

本节首先详细描述图论的相关概念, 然后针对异构非线性多智能体输出一致性问题, 分析其求解难度和存在问题.

1.1 图论

存在一个有向图 $\mathcal{G}(\mathcal{K}, \Gamma, \mathcal{A})$ 包含领导者和 N 个跟随者节点, 其中 $\mathcal{K} = \{\kappa_1, \kappa_2, \dots, \kappa_N\}$ 是一个非空有限节点集, 表示有向边集; $\mathcal{A} = [a_{ij}] \in \mathbf{R}^{N \times N}$ 是一个相关的邻接矩阵, $a_{ij} = 1$ 表示节点 j 到 i 之间存在一个有向边, 满足 $(\kappa_j, \kappa_i) \in \Gamma$, $\Gamma \subseteq \mathcal{K} \times \mathcal{K}$, 否则, $a_{ij} = 0$. 设增益 $b_i \geq 0$, 只有与领导节点直接相连的节点才不为零, $\mathcal{B} = \text{diag}\{\sum b_i\}$. 令与节点 κ_i 存在有向图相连的邻居集合为 $\mathfrak{N}_i = \{\kappa_j : (\kappa_j, \kappa_i) \in \Gamma\}$, 进一步定义一个入度矩阵为 $\mathcal{D} = \text{diag}\{\sum_{j \in \mathfrak{N}_i} a_{ij}\}$, $i = 1, 2, \dots, N$, 则有向图 \mathcal{G} 的 Laplacian 矩阵表示为 $\mathcal{L} = \mathcal{D} - \mathcal{A}$.

1.2 问题描述

考虑 N 个异构仿射非线性多智能体系统, 智能体分布在有向图 \mathcal{G} 上, 系统动力学模型可描述为

$$\begin{cases} x_{i, k+1} = f_i(x_{i, k}) + g_i(x_{i, k})u_{i, k} \\ y_{i, k} = h_i(x_{i, k}) \end{cases} \quad (1)$$

其中, $i \in \mathcal{N}$, $\mathcal{N} = 1, 2, \dots, N$, $x_{i, k} \in \mathbf{R}^n$ 为状态向量, $u_i \in \mathbf{R}^m$ 表示控制策略. 光滑向量场 $f_i(x_{i, k}) \in$

\mathbf{R}^n 和 $g_i(x_{i,k}) \in \mathbf{R}^{n \times m}$ 表示未知的系统动力学漂移阵和输入阵, $h_i(x_{i,k}) \in \mathbf{R}$ 为输出矩阵, 均满足在 \mathbf{R}^n 上 Lipschitz 连续且有界, $f_i(0) = 0$.

假设 1. 智能体的相对阶 $\rho_i = n$.

假设 2. 对于 $\forall i \in \mathcal{N}$, 总存在一个 $j \in \mathcal{N}$ 且 $j \neq i$, 使得 $f_i(x_{i,k}) \neq f_j(x_{i,k})$; 总存在一个 $k \in \mathcal{N}$ 且 $k \neq i$, 使得 $g_i(x_{i,k}) \neq g_k(x_{i,k})$.

在跟踪同步问题中, 需要设计分布式控制输入 $u_{i,k}$, 使所有节点的输出与领导节点 y_r 的输出同步. 领导节点可以是一个期望轨迹生成器, 也可以是智能决策的结果, 或者人工示教的轨迹, 它代表所需的期望轨迹. 领导者的动力学模型为

$$\begin{cases} x_{r,k+1} = f_r(x_{r,k}) \\ y_{r,k} = h_r(x_{r,k}) \end{cases} \quad (2)$$

其中, $x_{r,k} \in \mathbf{R}^n$. 函数 $f_r(\cdot)$ 和 $h_r(\cdot)$ 假设为 C_∞ 类. 输出 $y_{r,k}$ 是跟踪领导者输出所需的期望性能输出. 假设所有的智能体状态都是可测量的, 或者在系统对于输出满足能观性时, 也可以添加观察器.

为了解决智能体 (1) 和期望轨迹 (2) 的输出一致性跟踪问题, 智能体与期望轨迹的跟踪误差为 $e_{p,i,k} = y_{i,k} - y_{r,k}$, 多智能体协同局部邻域跟踪误差可表示为

$$\mathcal{E}_{i,k} = \sum_{j \in \mathcal{N}_i} a_{ij}(y_{i,k} - y_{j,k}) + b_i e_{p,i,k} \quad (3)$$

假设 3. 有向图 \mathcal{G} 存在一个生成树结构, 且至少有一个根节点的增益 b_i 是非零的, 意味着至少有一个智能体直接与领导者通讯.

由式 (3) 可知, 有向图 \mathcal{G} 的全局邻域误差向量为

$$E = [(\mathcal{L} + \mathcal{B}) \otimes I_\rho](Y - Y_r) \equiv [(\mathcal{L} + \mathcal{B}) \otimes I_\rho] \delta \quad (4)$$

其中, $Y = [y_{1,k}, y_{2,k}, \dots, y_{N,k}]^T \in \mathbf{R}^N$ 表示系统全局输出向量, $Y_r = 1_N \otimes y_r$, 1_N 表示元素全为 1 的 N 维向量, \otimes 表示 Kronecker 积, $E = [\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_N]^T \in \mathbf{R}^N$. δ 为全局跟踪误差向量, 由于其是一个全局向量, 无法在每个节点局部计算.

为了实现完全分布式的控制结构, 本文利用式 (3) 中的局部邻域跟踪误差来解决输出同步问题. 由式 (1) 和式 (3) 联例可得智能体 i 的局部跟踪误差动力学:

$$\begin{aligned} \mathcal{E}_{i,k+1} = & \sum_{j \in \mathcal{N}_i} a_{ij} h_i [f_i(x_{i,k}) + g_i(x_{i,k}) u_{i,k}] - \\ & \sum_{j \in \mathcal{N}_i} a_{ij} h_j [f_j(x_{j,k}) + g_j(x_{j,k}) u_{j,k}] + \\ & b_i \{ h_i [f_i(x_{i,k}) + g_i(x_{i,k}) u_{i,k}] - \\ & h_r [f_r(x_{r,k})] \} \end{aligned} \quad (5)$$

对于包含复杂非线性部分的误差动力学 (5), 传统控制理论在解决输出一致性控制问题时, 常受到黎卡提方程难以求解的困扰, 尤其是在系统的非线性动态未知且异构的情况下, 输出一致性控制器求解极其复杂.

输入输出反馈线性化技术能够通过微分同胚映射将非线性系统的输出 $y_{i,k}$ 与输入 $u_{i,k}$ 之间的动态关系转化为线性关系, 从而实现非线性系统的严格线性化. 基于模型的反馈线性化控制器求解形式如下所示:

$$u_{i,k} = \frac{-L_{f_i}^\rho h_i(x_{i,k})}{L_{g_i} L_{f_i}^{\rho-1} h_i(x_{i,k})} + \frac{v_{i,k}}{L_{g_i} L_{f_i}^{\rho-1} h_i(x_{i,k})} = \beta_i(x_{i,k}) + \alpha_i(x_{i,k}) v_{i,k} \quad (6)$$

其中, L 为李导数运算符, $u_{i,k}$ 为实际控制输入, $v_{i,k}$ 是一个虚拟输入, 在本文中作为分布式控制的输入端. 经过严格反馈线性化, 可消除系统非线性项并得到:

$$y_{i,k}^{(\rho)} = v_{i,k} \quad (7)$$

此时, 非线性多智能体通过微分同胚映射 $\Phi(x_{i,k})$ 投影到同胚线性空间中的动力学方程为

$$\begin{cases} \xi_{i,k+1} = A \xi_{i,k} + B v_{i,k} \\ y_{i,k} = C \xi_{i,k} \end{cases} \quad (8)$$

其中, $A = \begin{bmatrix} 0^{(n-1) \times 1} & I_{n-1} \\ 0 & 0_{1 \times (n-1)} \end{bmatrix}$, $B = \begin{bmatrix} 0^{(n-1) \times 1} \\ I \end{bmatrix}$, $C = [I \quad 0_{1 \times (n-1)}]$. 由此, 每个智能体均被映射为系统结构已知的线性化系统.

然而, 在原系统模型未知的情况下, $\alpha_i(x_{i,k})$ 和 $\beta_i(x_{i,k})$ 的精确求解变得极为困难, 不严格的反馈线性化将影响分布式控制器的执行效果. 本文提出的控制策略核心在于无模型自适应动态规划方法, 在不依赖精确模型的前提下, 实现非线性多智能体系统的精确线性化, 使每个智能体的动力学行为近似为同一期望的线性系统动力学, 进而能够利用传统的线性控制理论设计分布式控制器, 实现全局系统的输出一致性.

2 同胚分布式控制协议

为解决模型未知的输出一致性控制问题, 本文提出一种同胚分布式控制协议 (如图 1). 通过无模型自适应动态规划实现输入输出反馈线性化, 将异构非线性多智能体系统转化为同构线性系统, 从而简化分布式控制器的设计. 在物理空间中, 利用自适应动态规划方法设计输入输出反馈线性化控制器, 将智能体的闭环动态通过微分同胚映射为期望的线性系统, 实现与之一致的输出响应; 在同胚空

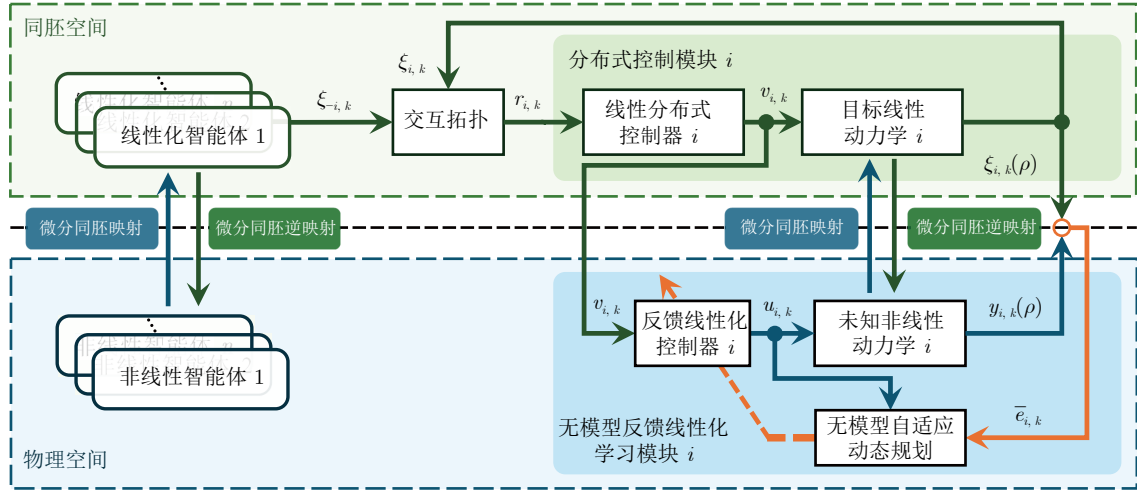


图 1 同胚分布式控制协议结构图

Fig. 1 Structure diagram of homeomorphic distributed control protocol

间中, 以期望线性系统为基础设计分布式一致性控制器. 通过物理空间的线性化处理和同胚空间的协同作用, 将控制性能优化与分布式决策设计相结合, 以实现异构非线性智能体的输出一致性控制.

2.1 无模型输入输出反馈线性化

为近似求解未知的反馈线性化控制器 (6), 首先需设计评价指标引导反馈线性化控制器学习. 考虑到系统输入输出未完成线性化前, 结合式 (7), 存在如下微分状态误差:

$$\bar{e}_{i,k} = v_{i,k} - y_{i,k}^{(\rho)} \quad (9)$$

自适应动态规划的目标是调整控制器使得 \$\bar{e}_{i,k}\$ 最小, 此时系统将被反馈线性化. 为得到 \$y_{i,k}^{(\rho)}\$, 采用式 (8) 作为期望转化的目标线性系统, 构造龙伯格状态观测器用以重构被控对象线性化状态:

$$\begin{cases} \hat{\xi}_{i,k+1} = A\hat{\xi}_{i,k} + Bv_{i,k} + H(y_{i,k} - \hat{y}_{i,k}) \\ \hat{y}_{i,k} = C\hat{\xi}_{i,k} \end{cases} \quad (10)$$

其中, \$v_{i,k}\$ 为分布式控制输入, \$H\$ 为滤波增益. 观测误差动力学可以表示为

$$e_{m,i,k+1} = \frac{\partial \Phi(x_{i,k})}{\partial x_{i,k}} \{f_i(x_{i,k}) + g_i(x_{i,k})[\beta_i(x_{i,k}) + \alpha_i(x_{i,k})v_{i,k}]\} - A\hat{\xi}_{i,k} - Bv_{i,k} - H(y_{i,k} - \hat{y}_{i,k}) \quad (11)$$

注 1. 在智能体完成线性化之前, 由于被控智能体与目标线性系统异构, 状态误差 \$\bar{e}_{i,k}\$ 无法渐近收敛. 仅当满足 \$\alpha_i(x_{i,k}) = \frac{1}{L_{g_i} L_{f_i}^{\rho-1} h_i(x_{i,k})}\$ 和 \$\beta_i(x_{i,k}) = \frac{-L_{f_i}^{\rho} h_i(x_{i,k})}{L_{g_i} L_{f_i}^{\rho-1} h_i(x_{i,k})}\$ 时, \$\lim_{t \rightarrow \infty} \bar{e}_{i,k} = 0\$, 被控系统线性

化为目标线性系统 (8).

考虑 \$\alpha_i(\cdot)\$ 和 \$\beta_i(\cdot)\$ 的两组李导数是关于 \$x_{i,k}\$ 的多项式, 因此利用 \$x_{i,k}\$ 各个元素及相关表达式作为基向量, 设计两组多项式近似未知的反馈线性化控制器 \$u_i = \beta_i(x_{i,k}) + \alpha_i(x_{i,k})v_i\$, 有

$$\hat{\alpha}_i(x_{i,k}) = W_{\alpha_i,k}^T \omega(x_{i,k}) \quad (12)$$

$$\hat{\beta}_i(x_{i,k}) = W_{\beta_i,k}^T \omega(x_{i,k}) \quad (13)$$

其中, \$W_{\alpha_i,k}^T, W_{\beta_i,k}^T\$ 为多项式权值, \$\omega(\cdot)\$ 是由 \$x_{i,k}\$ 及其多项式组合构成的基向量. 接下来, 通过数据驱动自适应动态规划算法, 学习得到 \$\alpha_i(\cdot)\$ 和 \$\beta_i(\cdot)\$ 的最优近似.

由于 \$\alpha_i(x_{i,k})\$ 和 \$\beta_i(x_{i,k})\$ 作用于同一控制通道, 一个网络的变化会影响另一个网络的学习空间. 这使得 \$\alpha_i(x_{i,k})\$ 和 \$\beta_i(x_{i,k})\$ 的学习均处于非平稳空间, 贝尔曼方程求解将是一个非凸优化问题, 容易使学习陷入局部最优.

为避免非线性项耦合, 利用历史采样输入输出数据, 结合极限差分方法重构 \$\alpha_i(\cdot)\$ 观测值的倒数:

$$L_{g_i} L_{f_i}^{\rho-1} h_i(x_{i,k}) = \frac{1}{\alpha_i(x_{i,k})} = \frac{\partial y_{i,k}(\rho)}{\partial u_{i,k}} \quad (14)$$

采用监督学习训练网络 (12) 得到 \$\hat{\alpha}_i(x_{i,k}) = \alpha_i(x_{i,k}) + d_{i,\alpha}\$, 可将式 (11) 表示为

$$\begin{cases} e_{m,i,k+1}(1) = e_{m,i,k}(2) - H_1(y_{i,k} - \hat{y}_{i,k}) \\ e_{m,i,k+1}(2) = e_{m,i,k}(3) - H_2(y_{i,k} - \hat{y}_{i,k}) \\ \vdots \\ e_{m,i,k+1}(\rho) = \beta_i(x_{i,k}) + \hat{\beta}_i(x_{i,k}) - H_{\rho}(y_{i,k} - \hat{y}_{i,k}) + \sigma_{i,k} \end{cases} \quad (15)$$

其中, $\sigma_{i, k} = d_{i, \beta} + d_i, \alpha v_{i, k}, d_{i, \beta} = \hat{\beta}_i(x_{i, k}) - \beta_i(x_{i, k})$. 理论上多项式可以无限逼近一条光滑曲线, 因此 $\sigma_{i, k}$ 满足 $\|\sigma_{i, k}\| \leq d^m < \varepsilon_d$ 和 $\|\sigma_{i, k} - \sigma_{i, k-1}\| \leq \Delta\sigma^m, \sigma^m, \Delta\sigma^m \in \mathbf{R}^+$ 是未知的, ε_d 为极小值. 基于此, 分布式反馈线性化控制器学习问题转为一个模型参考跟踪控制问题, 通过状态误差 $\bar{e}_{i, k}$ 作为强化信号优化网络 $\hat{\beta}_i(\cdot)$ 的输出以消除非线性动态, 使得观测误差动力学 (15) 能够快速收敛, 同时完成系统线性化.

值得注意的是, 传统的启发式动态规划在求解最优跟踪策略时通常需要考虑误差-动作对信息. 反馈线性化控制器通过消除系统的非线性特征, 使得线性控制器能够得到更好的控制效果, 间接影响跟踪误差, 而非直接通过误差反馈减小跟踪误差. 因此, 在反馈线性化控制器的无模型学习中, 执行网络和值函数不应与误差相关. 为了有效引导优化方向, 避免陷入局部最优, 需将反馈线性化的程度指标作为系统长期目标融入值函数的优化过程. 但是由于模型信息缺失, 难以预先设计一个奖励信号来正确引导反馈线性化的学习.

为此, 本文定义反馈线性化奖励作为各智能体线性化程度的指标:

$$C_{i, k} = \begin{cases} 0, & \|\bar{e}_{i, k}\|_1 + \|\bar{e}_{i, k} - \bar{e}_{i, k-1}\|_1 < \varepsilon_i \\ 1, & \|\bar{e}_{i, k}\|_1 + \|\bar{e}_{i, k} - \bar{e}_{i, k-1}\|_1 \geq \varepsilon_i \end{cases} \quad (16)$$

同时为正确引导学习方向, 设计奖励网络 $R_{i, k}$

$$\hat{R}_{i, k}^l = W_{r_i}^{lT} \omega(X_{i, k}) \quad (17)$$

该网络用于在学习过程中动态调整奖励值, 无需针对不同异构智能体分别设计奖励信号.

为了同时调整奖励信号和求解反馈线性化控制器, 设计了双启发式评价网络同时逼近最优值函数和一个启发函数. 其中, 启发式函数用于快速估计值函数梯度方向和大小, 优化控制策略. 本文在奖

励网络、评价网络与执行网络之间构建两阶段双启发式自适应动态规划问题, 通过两阶段循环迭代, 实现对高维奖励信息、值函数、启发函数和最优策略的同步逼近. 如图 2 所示, 两阶段双启发式自适应动态规划方法的每轮迭代包括两个阶段: 在奖励评估阶段, 根据反馈线性化奖励, 迭代优化奖励网络和双评价网络; 在动作评估阶段, 通过上一阶段得到的启发网络直接估计值函数梯度. 进而快速更新动作网络, 实现控制器的性能提升. 具体实现如下所述.

首先, 给出累计折扣奖励值函数的表达式:

$$J_{i, k} = \sum_{\delta=0}^{\infty} \gamma_{J_i}^{\delta} C_{i, k+\delta} \quad (18)$$

其中, $\gamma_{J_i} \in (0, 1)$ 是一个折扣因子. 定义一个双启发式评价网络结构同时近似最优值函数 $J_i^*(\cdot)$ 和一个最优启发函数 $\lambda_i^*(\cdot)$:

$$\begin{bmatrix} \hat{J}_{i, k}^l \\ \hat{\lambda}_{i, k}^l \end{bmatrix} = \begin{bmatrix} W_{J_i}^{lT} \\ W_{\lambda_i}^{lT} \end{bmatrix} \omega(X_{i, k}, R_{i, k}^l) \quad (19)$$

其中, $\hat{J}_{i, k}^l$ 和 $\hat{\lambda}_{i, k}^l$ 分别表示在 l 次迭代后对 $J_{i, k}$ 和 $\lambda_{i, k}$ 的估计值. $\lambda_{i, k}$ 是值函数 $J_{i, k}$ 关于 $X_{i, k}$ 的各元素偏导组成的向量.

学习过程中, 采用异策略学习方式, 利用 k 和 $k-1$ 的数据更新网络权值. 根据贝尔曼原理, 定义 $e_{c, i, k}$ 为双评价网络的估计误差:

$$e_{c, i, k} = \mu_j \frac{e_{J, i, k}^2}{2} + \mu_{\lambda} \frac{e_{\lambda, i, k}^2}{2} \quad (20)$$

其中, $e_{J, i, k} = \hat{R}_{i, k} - 1 + \gamma_{J_i} \hat{J}_{i, k} - \hat{J}_{i, k-1}$; $e_{\lambda, i, k} = \frac{\hat{R}_{i, k-1}}{X_{i, k-1}} + \gamma_{J_i} \hat{\lambda}_{i, k} \Xi_{i, k} - \hat{\lambda}_{i, k-1}$, 其中 $\mu_j \in (0, 1]$ 和 $\mu_{\lambda} \in (0, 1]$ 为学习步长; $\Xi_{i, k} = \frac{\partial X_{i, k}}{\partial X_{i, k-1}}$ 为增广状态的雅克比矩阵. 根据梯度下降原则, 双评价网络

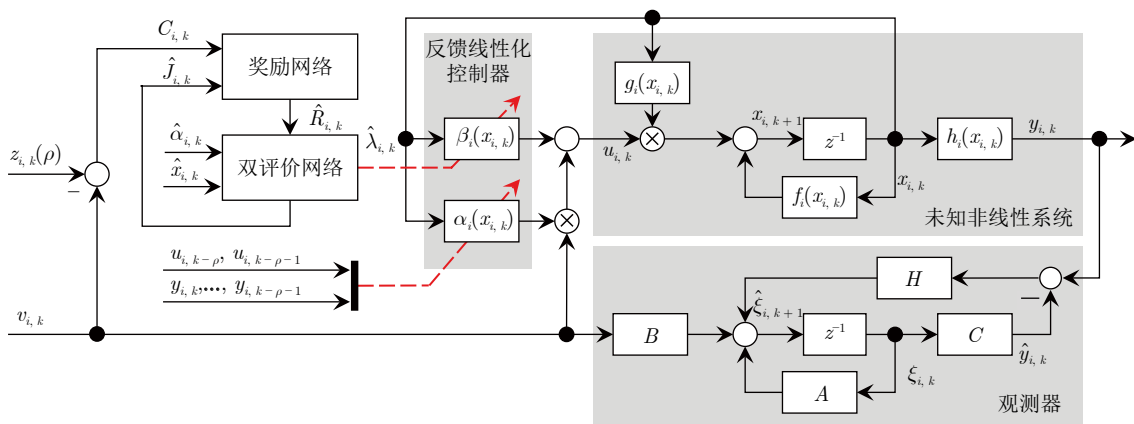


图 2 无模型反馈线性化学习模块

Fig. 2 Model-free feedback linearized learning modules

通过如下更新规则进行更新:

$$\begin{aligned} \begin{bmatrix} W_{J_i}^{l+1} \\ W_{\lambda_i}^{l+1} \end{bmatrix}^T &= \begin{bmatrix} W_{J_i}^l \\ W_{\lambda_i}^l \end{bmatrix}^T - \\ \eta_c \begin{bmatrix} \mu_j \frac{\partial e_{J_i, i, k}}{\partial \hat{J}_{i, k}^l} \frac{\partial \hat{J}_{i, k}^l}{\partial W_{J_i}^l} e_{J_i, i, k} \\ \mu_\lambda \frac{\partial e_{\lambda_i, i, k}}{\partial \hat{\lambda}_{i, k}^l} \frac{\partial \hat{\lambda}_{i, k}^l}{\partial W_{\lambda_i}^l} e_{\lambda_i, i, k} \end{bmatrix}^T &= \begin{bmatrix} W_{J_i}^l \\ W_{\lambda_i}^l \end{bmatrix}^T - \\ \eta_c \begin{bmatrix} \mu_j \gamma_{J_i} \omega \left(X_{i, k}, R_{i, k}^l \right) e_{J_i, i, k} \\ \mu_\lambda \gamma_{\lambda_i} \omega \left(X_{i, k}, R_{i, k}^l \right) \left(\Xi_{i, k} e_{\lambda_i, i, k} \right)^T \end{bmatrix}^T & \quad (21) \end{aligned}$$

其中, η_c 是评价网络的权值更新步长.

定义 $e_{R_i, i, k}$ 为奖励网络估计误差:

$$e_{R_i, i, k} = C_{i, k-1} - \hat{R}_{i, k}^l = C_{i, k-1} - \left(\hat{J}_{i, k-1}^l - \gamma_{J_i} \hat{J}_{i, k}^l \right) \quad (22)$$

奖励网络通过如下更新规则进行更新,

$$\begin{aligned} W_{r_i}^{l+1} &= W_{r_i}^l - \eta_r \frac{\partial e_{R_i, i, k}}{\partial \hat{J}_{i, k}^l} \frac{\partial \hat{J}_{i, k}^l}{\partial \hat{R}_{i, k}^l} \frac{\partial \hat{R}_{i, k}^l}{\partial W_{r_i}^l} e_{R_i, i, k} = \\ W_{r_i}^l - \eta_r e_{R_i, i, k} \gamma_{J_i} W_{J_i}^{lT} \omega' \left(X_{i, k}, R_{i, k}^l \right) \omega \left(X_{i, k} \right) & \quad (23) \end{aligned}$$

其中, η_r 是奖励网络的权值更新步长.

基于启发网络, 动作网络的误差函数可定义为

$$e_{\beta_i, i, k} = \hat{\lambda}_{i, k}^l \quad (24)$$

动作网络通过最小化误差函数 $e_{\beta_i, i, k}$ 求解最优动作, 更新规则如下:

$$\begin{aligned} W_{\beta_i}^{(l+1)T} &= W_{\beta_i}^{lT} - \eta_a \frac{\partial e_{\beta_i, i, k}}{\partial X_{i, k}} \frac{\partial X_{i, k}}{\partial \hat{\beta}_{i, k}^l} \frac{\partial \hat{\beta}_{i, k}^l}{\partial W_{\beta_i}^{lT}} \hat{\lambda}_{i, k}^l = \\ W_{\beta_i}^{lT} - \eta_a \hat{\lambda}_{i, k}^l \left(\xi_k \right) W_{\lambda_i}^{lT} \omega' \left(X_{i, k}, R_{i, k}^l \right) \omega \left(x_{i, k} \right) & \quad (25) \end{aligned}$$

其中, η_a 是执行网络的权值更新步长.

2.2 线性化系统分布式控制

在同胚空间中, 通过无模型反馈线性化, 非线性多智能体输入输出关系在控制器 (6) 的作用下由非线性动力学 (1) 映射为同胚空间中的能控标准型系统, 由此可将虚拟领导者设计为更简单的线性形式:

$$\begin{cases} \xi_{r, k} = A \xi_{r, k} + BK \xi_{r, k} \\ y_{r, k} = C \xi_{r, k} \end{cases} \quad (26)$$

其中, K 为反馈控制增益, 局部邻域输出跟踪误差

可由一个虚拟局部邻域状态跟踪误差等效:

$$\mathcal{E}_{i, k} = \sum_{j \in N_i} a_{ij} (\xi_{i, k} - \xi_{j, k}) + b_i e_{p, i, k} \quad (27)$$

其中, $e_{p, i, k} = \xi_{i, k} - \xi_{r, k}$.

令 $\xi = [\xi_1, \xi_2, \dots, \xi_N]$, 则全局动力学方程为

$$\begin{cases} \xi_k = (I_N \otimes A) \xi + (I_N \otimes B) v \\ y = (I_N \otimes C) \xi_r \end{cases} \quad (28)$$

定义 $Q = Q^T$ 和 $R = R^T$ 为正定矩阵. 令反馈控制增益为

$$K = R^{-1} B^T \mathcal{P} \quad (29)$$

其中, \mathcal{P} 是代数黎卡提方程的唯一正定解:

$$A^T \mathcal{P} + \mathcal{P} A + Q - \mathcal{P} B R^{-1} B^T \mathcal{P} = 0 \quad (30)$$

令 ζ_i ($i \in \mathcal{N}$) 为 $\mathcal{L} + \mathcal{B}$ 的特征根, 当满足 $C \geq \frac{1}{2 \min_{i \in \mathcal{N}} \text{Re}(\zeta_i)}$ 时, $\forall i \in \mathcal{N}$, 所有 $A - C \zeta_i B K$ 满足 Hurwitz 条件, $C \in \mathbf{R}$ 为耦合增益.

引理 1^[41]. 选择

$$v_{i, k} = -CK \mathcal{E}_{i, k} \quad (31)$$

为分布式线性控制输入, 其中 $C \geq \frac{1}{2 \min_{i \in \mathcal{N}} \text{Re}(\zeta_i)}$, $K = R^{-1} B^T \mathcal{P}$, 则 $\forall i \in \mathcal{N}$, 有 ξ_i 关于 ξ_r 协同一致渐近有界, 且所有节点与 ξ_r 同步.

注 2. 由于输入输出反馈线性化特性, 可将期望线性系统动力学设计为统一形式. 根据假设, 当所有智能体相对阶一致时, 采用同样的反馈控制增益 K 即可实现所有智能体动态品质趋同, 显著减小分布式控制器设计复杂度.

3 学习收敛性证明

本节讨论分布式无模型反馈线性化算法的收敛性. 考虑跟踪误差的收敛性以及双评价网络、奖励网络、动作网络的学习收敛问题. 定义分布式无模型反馈线性化算法中三种网络的最优权值表达式为

$$\begin{cases} W_{J_i}^* = \arg \min_{W_{J_i}} \left\| \hat{J}_i^l(X_{i, k}, \hat{R}_{i, k}^l) - J_{i, k} \right\| \\ W_{\lambda_i}^* = \arg \min_{W_{\lambda_i}} \left\| \hat{\lambda}_i^l(X_{i, k}, \hat{R}_{i, k}^l) - \frac{\partial J_{i, k}}{\partial X_{i, k}} \right\| \\ W_{r_i}^* = \arg \min_{W_{r_i}} \left\| \hat{R}_i^l(X_{i, k}) - C_{i, k} \right\| \\ W_{a_i}^* = \arg \min_{W_{a_i}} \left\| \hat{\beta}_i^l(x_{i, k}) - L_{f_i}^p h_i(x_{i, k}) \right\| \end{cases} \quad (32)$$

其中, $J_i(X_{i, k})$ 为理想值函数. 可得权值的估计误差为

$$\begin{cases} \tilde{W}_{J_i}^l = W_{J_i}^l - W_{J_i}^* \\ \tilde{W}_{\lambda_i}^l = W_{\lambda_i}^l - W_{\lambda_i}^* \\ \tilde{W}_{r_i}^l = W_{r_i}^l - W_{r_i}^* \\ \tilde{W}_{a_i}^l = W_{a_i}^l - W_{a_i}^* \end{cases} \quad (33)$$

为了简化表示, 令 $\omega_{a,i,k} = \omega(x_{i,k})$, $\omega_{c,i,k} = \omega(X_{i,k}, \tilde{R}_{i,k}^l)$, $\omega_{r,i,k} = \omega(X_{i,k})$, $\tilde{u}_{i,k}^l = \tilde{W}_{a_i}^{l,T} \omega_{a,i,k}$, $\tilde{J}_{i,k}^l = \tilde{W}_{J_i}^{l,T} \omega_{c,i,k}$, $\tilde{\lambda}_{i,k}^l = \tilde{W}_{\lambda_i}^{l,T} \omega_{c,i,k}$, $\tilde{R}_{i,k}^l = \tilde{W}_{r_i}^{l,T} \omega_{r,i,k}$.

假设 4. 网络的权值 W_{J_i} , W_{λ_i} , W_{a_i} , W_{r_i} 和基向量输出 $\omega(\cdot)$ 均有界, 且上界分别表示为 $W_{J_i}^m$, $W_{\lambda_i}^m$, $W_{a_i}^m$, $W_{r_i}^m$, ω^m .

首先讨论系统跟踪误差的收敛性, 若期望模型的状态 $z_{i,k}$ 和输入 $r_{i,k}$ 有界, 且假设 4 成立, 令 $e_{m,i,k}$ 的 Lyapunov 函数候选为 $L_{e_i} = \frac{1}{3} e_{m,i,k}^T e_{m,i,k}$, 则 L_{e_i} 的一阶差分满足:

$$\begin{aligned} \Delta L_{e_i} &= e_{m,i,k+1}^T e_{m,i,k+1} - e_{m,i,k}^T e_{m,i,k} \leq \\ &\left(\lambda_{\max} - \frac{1}{3} \right) \|e_{m,i,k}\|^2 + \|\hat{\beta}_{i,k}^l\|^2 + \|d_{i,k}\|^2 \end{aligned} \quad (34)$$

其中, λ_{\max} 表示 $H^T H$ 最大特征根.

接下来讨论学习过程的收敛性. 为了分析双重评价函数权值更新的稳定性, 考虑四个部分的收敛性: 值函数权值的估计误差、值函数的估计误差、启发式函数权值的估计误差和启发式函数的估计误差. 根据式 (21), 双评价网络权值估计误差如下:

$$\begin{aligned} \begin{bmatrix} \tilde{W}_{J_i}^{l+1} \\ \tilde{W}_{\lambda_i}^{l+1} \end{bmatrix}^T &= \begin{bmatrix} \tilde{W}_{J_i}^l \\ \tilde{W}_{\lambda_i}^l \end{bmatrix}^T - \\ &\eta_c \begin{bmatrix} \mu_j \gamma_j \omega_{c,i,k} e_{J,i,k}^T \\ \mu_\lambda \gamma_\lambda \omega_{c,i,k} (\Xi_{i,k} e_{\lambda,i,k})^T \end{bmatrix}^T \end{aligned} \quad (35)$$

引理 2. 令双评价网络的 Lyapunov 函数候选为

$$\begin{aligned} L_{c_i} &= L_{W_{J_i}} + L_{J_i} + L_{W_{\lambda_i}} + L_{\lambda_i} = \\ &\frac{1}{\eta_c} \text{tr} \left(\tilde{W}_{J_i}^{l,T} \tilde{W}_{J_i}^l \right) + \frac{1}{2} \mu_j \left\| \tilde{J}_i^l(X_{i,k}) \right\|^2 + \\ &\frac{1}{\eta_c} \text{tr} \left(\tilde{W}_{\lambda_i}^{l,T} \tilde{W}_{\lambda_i}^l \right) + \frac{1}{2} \mu_\lambda \left\| \tilde{\lambda}_i^l(X_{i,k}) \right\|^2 \end{aligned}$$

则有 L_{c_i} 的一阶差分满足以下不等式:

$$\begin{aligned} \Delta L_{c_i} &\leq -\mu_j \gamma_{J_i}^2 \left\| \tilde{J}_{i,k}^l \right\|^2 + \frac{\mu_j}{2} \left\| \tilde{J}_{i,k-1}^l \right\|^2 + \\ &\frac{\mu_\lambda}{2} \left\| \tilde{\lambda}_{i,k-1}^l \right\|^2 - \mu_j \gamma_{J_i}^2 (I - \chi_{J_k}) \times \\ &\left\| \tilde{J}_{i,k}^l + \gamma_{J_i}^{-1} \varepsilon_{J_k}^* \right\|^2 - \end{aligned}$$

$$\begin{aligned} &\mu_\lambda \gamma_{J_i}^2 \left\| \Xi_{i,k} \right\|^2 \left\| \tilde{\lambda}_{i,k}^l \right\|^2 - \\ &\mu_\lambda \gamma_{J_i}^2 \left(I - \eta_c \mu_\lambda \gamma_{J_i}^2 \left\| \Xi_{i,k} \right\|^2 \left\| \omega_{c,i,k} \right\|^2 \right) \times \\ &\left\| \Xi_{i,k}^T \tilde{\lambda}_{i,k}^l + \gamma_{J_i}^{-1} \varepsilon_{\lambda_k}^* \right\|^2 + 2\mu_j \left\| \hat{R}_{i,k-1}^l + \right. \\ &\left. \gamma_{J_i} W_{J_i}^* \omega_{c,i,k} - \frac{1}{2} (W_{J_i}^l + W_{J_i}^*) \omega_{c,i,k-1} \right\|^2 + \\ &2\mu_\lambda \left\| \frac{\partial \hat{R}_{i,k-1}^l}{\partial X_{i,k-1}} + \gamma_{J_i} \Xi_{i,k} W_{\lambda_i}^* \omega_{c,i,k} - \right. \\ &\left. \frac{1}{2} (W_{\lambda_i}^l - W_{\lambda_i}^*) \omega_{c,i,k-1} \right\|^2 + \\ &\frac{1}{2} \mu_j \left(\left\| \tilde{J}_{i,k}^l \right\|^2 - \left\| \tilde{J}_{i,k-1}^l \right\|^2 \right) + \\ &\frac{1}{2} \mu_\lambda \left(\left\| \tilde{\lambda}_{i,k}^l \right\|^2 - \left\| \tilde{\lambda}_{i,k-1}^l \right\|^2 \right) \end{aligned} \quad (36)$$

其中, $\frac{\partial \hat{R}_{i,k-1}^l}{\partial X_{i,k-1}}$, $\Xi_{i,k}$ 的上界分别为 R^m 和 Ξ^m .

证明. $L_{W_{J_i}}$ 一阶差分为

$$\begin{aligned} \Delta L_{W_{J_i}} &= \frac{1}{\eta_c} \text{tr} \left[\tilde{W}_{J_i}^{l+1,T} \tilde{W}_{J_i}^{l+1} - \tilde{W}_{J_i}^{l,T} \tilde{W}_{J_i}^l \right] = \\ &\frac{1}{\eta_c} \text{tr} \left[\tilde{W}_{J_i}^{l,T} (I - \chi_c)^T (I - \chi_c) \tilde{W}_{J_i}^l - \right. \\ &\varepsilon_{J_k}^* \omega_{c,i,k}^T \eta_c \mu_j \gamma_j (I - \chi_c) \tilde{W}_{J_i}^l + \\ &\varepsilon_{J_k}^* \omega_{c,i,k}^T \eta_c^2 \mu_j^2 \gamma_j^2 \omega_{c,i,k} \varepsilon_{J_k}^{*,T} - \\ &\tilde{W}_{J_i}^{l,T} (I - \chi_c)^T \eta_c \mu_j \gamma_j \omega_{c,i,k} \varepsilon_{J_k}^{*,T} - \\ &\left. \tilde{W}_{J_i}^{l,T} \tilde{W}_{J_i}^l \right] \end{aligned} \quad (37)$$

其中, $\varepsilon_{J_k}^* = \hat{R}_{i,k}^l - W_{J_i}^{l,T} \omega_{c,i,k-1} + \gamma_{J_i} W_{J_i}^{*,T} \omega_{c,i,k}$, $\chi_c = \eta_c \mu_j \gamma_j^2 \omega_{c,i,k} \omega_{c,i,k}^T$.

对上式进行如下变换:

$$\begin{aligned} &\tilde{W}_{J_i}^{l,T} (I - \chi_c)^T (I - \chi_c) \tilde{W}_{J_i}^l - \tilde{W}_{J_i}^{l,T} \tilde{W}_{J_i}^l = \\ &\tilde{W}_{J_i}^{l,T} (I - \chi_c) \tilde{W}_{J_i}^l - \tilde{W}_{J_i}^{l,T} \tilde{W}_{J_i}^l - \\ &\tilde{W}_{J_i}^{l,T} \chi_c (I - \chi_c) \tilde{W}_{J_i}^l = -\eta_c \mu_j \gamma_j^2 \left\| \tilde{J}_{i,k}^l \right\|^2 - \\ &\eta_c \mu_j \gamma_j^2 (I - \chi_c) \left\| \tilde{J}_{i,k}^l \right\|^2 \end{aligned} \quad (38)$$

则 $\Delta L_{W_{J_i}}$ 可重写为

$$\begin{aligned} \Delta L_{W_{J_i}} &= \frac{1}{\eta_c} \text{tr} \left[-\eta_c \mu_j \gamma_j^2 (I - \chi_c) \left\| \tilde{J}_{i,k}^l \right\|^2 - \right. \\ &\eta_c \mu_j \gamma_j^2 \left\| \tilde{J}_{i,k}^l \right\|^2 + \varepsilon_{J_k}^* \omega_{c,i,k}^T \eta_c^2 \mu_j^2 \gamma_j^2 \omega_{c,i,k} \varepsilon_{J_k}^{*,T} - \\ &\varepsilon_{J_k}^* \omega_{c,i,k}^T \eta_c \mu_j \gamma_j (I - \chi_c) \tilde{W}_{J_i}^l - \\ &\left. \tilde{W}_{J_i}^{l,T} (I - \chi_c)^T \eta_c \mu_j \gamma_j \omega_{c,i,k} \varepsilon_{J_k}^{*,T} \right] = \end{aligned}$$

$$\begin{aligned} & \mu_j \|\varepsilon_{jk}^*\|^2 - \mu_j \gamma_j^2 \left\| \tilde{J}_{i,k}^l \right\|^2 - \\ & \mu_j \gamma_j^2 (I - \chi_c) \left\| \tilde{J}_{i,k}^l + \gamma_j^{-1} \varepsilon_{jk}^* \right\|^2 \end{aligned} \quad (39)$$

根据 Cauchy-Schwarz 不等式^[19], $\Delta L_{W_{J_i}}$ 满足:

$$\begin{aligned} \Delta L_{W_{J_i}} & \leq -\mu_j \gamma_j^2 \left\| \tilde{J}_{i,k}^l \right\|^2 + \frac{\mu_j}{2} \left\| \tilde{J}_{i,k-1}^l \right\|^2 - \\ & \mu_j \gamma_j^2 (I - \chi_c) \left\| \tilde{J}_{i,k}^l + \gamma_j^{-1} \varepsilon_{c,k}^* \right\|^2 + \\ & 2\mu_j \left\| \hat{R}_{i,k}^l + \gamma_J W_{J_i}^* \omega_{c,i,k} - \frac{1}{2} (W_{J_i}^l + W_{J_i}^*) \omega_{c,i,k-1} \right\|^2 \end{aligned} \quad (40)$$

同理, 可得 $\Delta L_{W_{\lambda_i}}$ 满足

$$\begin{aligned} \Delta L_{W_{\lambda_i}} & \leq \frac{\mu_\lambda}{2} \left\| \tilde{\lambda}_{i,k-1}^l \right\|^2 - \mu_\lambda \gamma_\lambda^2 \left\| \Xi_{i,k} \right\|^2 \left\| \tilde{\lambda}_{i,k}^l \right\|^2 - \\ & \mu_\lambda \gamma_\lambda^2 (I - \eta_c \mu_\lambda \gamma_\lambda^2 \left\| \Xi_{i,k} \right\|^2 \left\| \omega_{c,i,k} \right\|^2) \times \\ & \left\| \Xi_{i,k}^T \tilde{\lambda}_{i,k}^l + \gamma_\lambda^{-1} \varepsilon_\lambda^* \right\|^2 + \\ & 2\mu_\lambda \left\| \frac{\partial \hat{R}_{i,k-1}^l}{\partial X_{i,k-1}} + \gamma_J \Xi_{i,k} W_{\lambda_i}^* \omega_{c,i,k} - \right. \\ & \left. \frac{1}{2} (W_{\lambda_i}^l + W_{\lambda_i}^*) \omega_{c,i,k-1} \right\|^2 \end{aligned} \quad (41)$$

其中, $\varepsilon_\lambda^* = \frac{\partial \hat{R}_{i,k-1}^l}{\partial X_{i,k-1}} + \gamma_J \Xi_{i,k} W_{\lambda_i}^* \omega_{c,i,k} - W_{\lambda_i}^l \omega_{c,i,k-1}$, $\chi_\lambda = \eta_c \mu_\lambda \gamma_\lambda^2 \left\| \Xi_{i,k} \right\|^2 \left\| \omega_{c,i,k} \right\|^2 \omega_{c,i,k}^T$.

对于 L_{J_i} 和 L_{λ_i} , 可直接表示为

$$\Delta L_{J_i} = \frac{1}{2} \mu_j \left(\left\| \tilde{J}_{i,k}^l \right\|^2 - \left\| \tilde{J}_{i,k-1}^l \right\|^2 \right) \quad (42)$$

$$\Delta L_{\lambda_i} = \frac{1}{2} \mu_\lambda \left(\left\| \tilde{\lambda}_{i,k}^l \right\|^2 - \left\| \tilde{\lambda}_{i,k-1}^l \right\|^2 \right) \quad (43)$$

结合上述计算式, 可得 $\Delta L_{c,i}$ 满足式 (36). \square

根据式 (23), 奖励网络权值误差方程如下:

$$\tilde{W}_{r_i}^{l+1} = \tilde{W}_{r_i}^l - \eta_r e_{R,i,k} \gamma_{J_i} W_{J_i}^l \omega'_{c,i,k} \omega_{r,i,k} \quad (44)$$

引理 3. 奖励网络的 Lyapunov 函数候选为

$$L_{r_i} = \frac{1}{2\eta_r} \text{tr} \left(\tilde{W}_{r_i}^l \tilde{W}_{r_i}^l \right)$$

Lyapunov 函数 L_{r_i} 的一阶差分满足以下不等式:

$$\begin{aligned} \Delta L_{r_i} & \leq \left\| \tilde{W}_{r_i}^l \omega_{r,i,k} \right\|^2 + \left\| W_{J_i}^l \omega'_{c,i,k} \right\|^2 + \left\| J_{i,k} \gamma_{J_i} \right\|^2 - \\ & \left(1 - \eta_r \left\| \omega_{r,i,k} \right\|^2 \right) \left\| W_{J_i}^l \omega'_{c,i,k} \right\|^2 \left\| J_{i,k} \gamma_{J_i} \right\|^2 \end{aligned} \quad (45)$$

证明. 根据式 (44), L_{r_i} 的一阶差分为

$$\begin{aligned} \Delta L_{r_i} & = \frac{1}{\eta_r} \text{tr} \left(\tilde{W}_{r_i}^{l+1} \tilde{W}_{r_i}^{l+1} - \tilde{W}_{r_i}^l \tilde{W}_{r_i}^l \right) = \\ & \text{tr} \left(-2J_{i,k} \gamma_{J_i} W_{J_i}^l \omega'_{c,i,k} \omega_{r,i,k} \tilde{W}_{r_i}^l + \right. \\ & \left. \eta_r \left\| \omega_{r,i,k} \right\|^2 \left\| W_{J_i}^l \omega'_{c,i,k} \right\|^2 \left\| \gamma_{J_i} J_{i,k} \right\|^2 \right) \end{aligned} \quad (46)$$

对式 (46) 第 1 项进行变换可得:

$$\begin{aligned} \Delta L_{r_i} & = \eta_r \left\| \omega_{r,i,k} \right\|^2 \left\| W_{J_i}^l \omega'_{c,i,k} \right\|^2 \left\| \gamma_{J_i} J_{i,k} \right\|^2 - \\ & \left\| J_{i,k} \gamma_{J_i} W_{J_i}^l \omega'_{c,i,k} \right\|^2 - \left\| \tilde{W}_{r_i}^l \omega_{r,i,k} \right\|^2 + \\ & \left\| \tilde{W}_{r_i}^l \omega_{r,i,k} - J_{i,k} \gamma_{J_i} W_{J_i}^l \omega'_{c,i,k} \right\|^2 = \\ & \left\| \tilde{W}_{r_i}^l \omega_{r,i,k} - J_{i,k} \gamma_{J_i} W_{J_i}^l \omega'_{c,i,k} \right\|^2 - \\ & \left\| \tilde{W}_{r_i}^l \omega_{r,i,k} \right\|^2 - \\ & \left(1 - \eta_r \left\| \omega_{r,i,k} \right\|^2 \right) \left\| J_{i,k} \gamma_{J_i} W_{J_i}^l \omega'_{c,i,k} \right\|^2 \end{aligned} \quad (47)$$

同样, 根据 Cauchy-Schwarz 不等式^[19] 进行缩放, 可得 ΔL_{r_i} 满足式 (45). \square

根据式 (25), 执行网络权值估计误差方程如下:

$$\tilde{W}_{a_i}^{l+1} = \tilde{W}_{a_i}^l - \eta_a \hat{\lambda}_{i,k}^l W_{\lambda_i}^l \omega'_{c,i,k} \omega_{a,i,k} \quad (48)$$

引理 4. 执行网络的 Lyapunov 函数候选为

$$L_{a_i} = \frac{1}{\eta_a} \text{tr} \left(\tilde{W}_{a_i}^l \tilde{W}_{a_i}^l \right)$$

Lyapunov 函数 L_{a_i} 的一阶差分满足以下不等式:

$$\begin{aligned} \Delta L_{a_i} & \leq \left\| \tilde{\beta}_{i,k}^l \right\|^2 + \left\| W_{\lambda_i}^l \omega'_{c,i,k} \right\|^2 + \left\| \hat{\lambda}_{i,k}^l \right\|^2 - \\ & \left(1 - \eta_a \left\| \omega_{a,i,k} \right\|^2 \right) \left\| W_{\lambda_i}^l \omega'_{c,i,k} \right\|^2 \left\| \hat{\lambda}_{i,k}^l \right\|^2 \end{aligned} \quad (49)$$

证明. L_{a_i} 的一阶差分为

$$\begin{aligned} \Delta L_{a_i} & = \frac{1}{\eta_a} \text{tr} \left(\tilde{W}_{a_i}^{l+1} \tilde{W}_{a_i}^{l+1} - \tilde{W}_{a_i}^l \tilde{W}_{a_i}^l \right) = \\ & \text{tr} \left\{ -2\tilde{\beta}_{i,k}^l (W_{\lambda_i}^l \omega'_{c,i,k})^T \hat{\lambda}_{i,k}^l + \right. \\ & \left. \eta_a \left\| \omega_{a,i,k} \right\|^2 \left\| \hat{\lambda}_{i,k}^l \right\|^2 \left\| W_{\lambda_i}^l \omega'_{c,i,k} \right\|^2 \right\} \end{aligned} \quad (50)$$

与引理 3 证明类似, 易得 ΔL_{a_i} 满足式 (49). \square

通过上述分析, 可以给出算法收敛性定理.

定理 1. 考虑非线性智能体 i 的输入输出反馈线性化控制器学习过程, 动作网络、奖励网络和双评价网络分别如式 (13)、(17) 和 (19) 所定义. 各网络权值根据式 (25)、(23) 和 (21) 给出的更新规律进行更新. 如果学习参数满足以下不等式:

$$\begin{cases} 3\lambda_{\max} < 1, \frac{\sqrt{2}}{2} < \gamma_{J_i} < 1 \\ \eta_c < \frac{1}{\mu_{J_i} \gamma_{J_i}^2 \|\omega^m\|^2}, \eta_r < \frac{1}{\|\omega^m\|^2}, \eta_a < \frac{1}{\|\omega^m\|^2} \end{cases} \quad (51)$$

则有基于输入输出数据的两阶段自适应双评价设计算法的跟踪性能误差 $e_{m, i, k} \in \mathcal{P}_{e_{m, i}}$ 和学习误差 $\tilde{J}_{i, k}^l \in \mathcal{P}_{J_i}$ 最终一致有界. 其中

$$\begin{aligned} \mathcal{P}_{e_{m, i}} &= \left\{ e_{m, i, k} \in \mathbf{R}^n : \|e_{m, i, k}\| \leq \sqrt{\frac{\Gamma_{\max}}{1 - 3\lambda_{\max}}} \right\} \\ \mathcal{P}_{J_i} &= \left\{ J_{i, k}^l \in \mathbf{R} : \|\tilde{J}_{i, k}^l\| \leq \sqrt{\frac{\Gamma_{\max}}{\mu_j (2\gamma_{J_i}^2 - 1)}} \right\} \end{aligned} \quad (52)$$

证明. 基于引理 2~4 以及不等式 (34), 无模型反馈线性化算法的 Lyapunov 候选函数满足如下不等式:

$$\begin{aligned} \Delta L_i &= \Delta L_{e_i} + \Delta L_{c_i} + \Delta L_{r_i} + \Delta L_{a_i} \leq \\ &\quad - \left(\frac{1}{3} - \lambda_{\max} \right) \|e_{m, i, k}\|^2 - \\ &\quad \mu_j \gamma_{J_i}^2 (I - \chi_{J_k}) \left\| \tilde{J}_{i, k}^l + \gamma_{J_i}^{-1} \varepsilon_{J_k}^* \right\|^2 - \\ &\quad \mu_J \gamma_{J_i}^2 \left\| \tilde{J}_{i, k}^l \right\|^2 - \\ &\quad \mu \lambda \gamma_{J_i}^2 \left(I - \eta_c \mu \lambda \gamma_{J_i}^2 \|\Xi_{i, k}\|^2 \|\omega_{c, i, k}\|^2 \right) \times \\ &\quad \left\| \Xi_{i, k}^T \tilde{\lambda}_{i, k}^l + \gamma_{J_i}^{-1} \varepsilon_{\lambda_k}^* \right\|^2 - \\ &\quad \mu \lambda \gamma_{J_i}^2 \|\Xi_{i, k}\|^2 \left\| \tilde{\lambda}_{i, k}^l \right\|^2 - \\ &\quad \left(1 - \eta_r \|\omega_{r, i, k}\|^2 \right) \left\| W_{J_i}^{l, T} \omega_{c, i, k}' \right\|^2 \|J_{i, k} \gamma_{J_i}\|^2 - \\ &\quad \left(1 - \eta_a \|\omega_{a, i, k}\|^2 \right) \left\| W_{\lambda_i}^{l, T} \omega_{c, i, k}' \right\|^2 \left\| \tilde{\lambda}_{i, k}^l \right\|^2 + \Gamma_i \end{aligned} \quad (53)$$

其中, 对 Γ_i 进行缩放可得:

$$\begin{aligned} \Gamma_i &= 2 \left\| \tilde{\beta}_{i, k}^l \right\|^2 + \|d_{i, k}\|^2 + \left\| W_{\lambda_i}^{l, T} \omega_{c, i, k}' \right\|^2 + \\ &\quad \left\| \tilde{\lambda}_{i, k}^l \right\|^2 + \left\| \tilde{W}_{r_i}^{l, T} \omega_{r, i, k} \right\|^2 + \left\| W_{J_i}^{l, T} \omega_{c, i, k}' \right\|^2 + \\ &\quad \|J_{i, k} \gamma_{J_i}\|^2 + \frac{1}{2} \mu_j \left\| \tilde{J}_{i, k}^l \right\|^2 + 2\mu_j \left\| \hat{R}_{i, k-1}^l \right\|^2 + \\ &\quad \gamma_{J_i} W_{J_i}^* \omega_{c, i, k} - \frac{1}{2} (W_{J_i}^l + W_{J_i}^*) \omega_{c, i, k-1} \left\|^2 + \right. \\ &\quad \left. \frac{1}{2} \mu \lambda \left\| \tilde{\lambda}_{i, k}^l \right\|^2 + 2\mu \lambda \left\| \frac{\partial \hat{R}_{i, k-1}^l}{\partial X_{i, k-1}} \right\|^2 + \right. \\ &\quad \left. \gamma_{J_i} \Xi_{i, k} W_{\lambda_i}^* \omega_{c, i, k} - \frac{1}{2} (W_{\lambda_i}^l - W_{\lambda_i}^*) \omega_{c, i, k-1} \right\|^2 \end{aligned}$$

$$\begin{aligned} &\left\| \tilde{W}_{r_i}^{l, T} \omega_{r, k} \right\|^2 + \left\| W_{J_i}^{l, T} \omega_{c, k}' \right\|^2 + \\ &\|d_{i, k}\|^2 + \|J_{i, k} \gamma_{J_i}\|^2 \end{aligned} \quad (54)$$

则 Γ_i 的上界为

$$\begin{aligned} \Gamma_{\max} &= 2\|W_a^m \omega^m\|^2 + 2\|W_\lambda^m \omega^m\|^2 + \|W_r^m \omega^m\|^2 + \\ &\quad \|W_J^m \omega^m\|^2 + \|d^m\|^2 + \|\gamma_{J_i} W_J^m \omega^m\|^2 + \\ &\quad 8\mu \lambda \|W_r^m \omega^m\|^2 + \\ &\quad \frac{1}{2} \mu \lambda \left(5 + 8\|\gamma_J \Xi^m\|^2 \right) \|W_\lambda^m \omega^m\|^2 + \\ &\quad 8\mu_j \|W_r^m \omega^m\|^2 + \frac{1}{2} \mu_j \left(5 + 8\|\gamma_{J_i}\|^2 \right) \|W_J^m \omega^m\|^2 \end{aligned} \quad (55)$$

当学习参数满足式 (51), 且对于任意的跟踪误差和值函数估计误差

$$\begin{cases} \|e_{m, i, k}\| > \sqrt{\frac{\Gamma_{\max}}{1 - 3\lambda_{\max}}} \\ \|\tilde{J}_{i, k}^l\| > \sqrt{\frac{\Gamma_{\max}}{\mu_j (2\gamma_{J_i}^2 - 1)}} \end{cases} \quad (56)$$

有 $\Delta L_i \leq 0$. 因此, 根据 Lyapunov 扩展定理, 可得跟踪误差和学习误差最终一致有界收敛. \square

定理 1 及相关证明通过数学推导给出学习收敛的条件, 这些条件的满足确保了系统的收敛性. 接下来将展示两个案例实验验证所提方法在模型未知的异构非线性多智能体系统中的应用效果.

4 实验验证

在本节中, 通过对异构未知非线性多智能体系统的仿真算例说明同胚分布式控制协议的可开发性和有效性. 系统的网络拓扑如图 3 所示. 考虑由 6 个两轮小车横向动力学构成的多智能体系统, 智能体的动力学如下所示:

$$\begin{aligned} f_i(\xi_i) &= \begin{bmatrix} \bar{v} \cos(\psi_i) \\ \frac{h_i}{2m_i} \dot{\psi} \sin(\psi_i) \\ \bar{v} \sin(\psi_i) \\ -\frac{h_i}{2m_i} \dot{\psi} \cos(\psi_i) \\ \dot{\psi}_i \end{bmatrix} \\ g_i(\xi_i) &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{m_i}{h_i} \end{bmatrix}, \quad h_i(\xi_i) = \begin{bmatrix} x_{i, k} \\ y_{i, k} \end{bmatrix} \end{aligned}$$

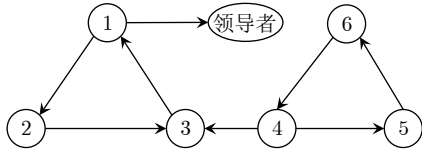


图 3 通讯拓扑

Fig.3 Communication topology

其中, $\xi_i = [x \ \dot{x} \ y \ \dot{y} \ \psi]^T$, x, y, \dot{x}, \dot{y} 分别为小车中心沿 x 轴和 y 轴方向的位移和速度, ψ 和 $\dot{\psi}$ 为航向角和角速度, m_i 为车轮到小车中心距离, h_i 为万向轮到小车中心的距离, 模型参数 (表 1) 和模型结构 $f_i(\xi_i), g_i(\xi_i)$ 在学习过程中被设定为未知. \bar{v} 为小车前进速度.

表 1 异构多智能体系统参数

变量	值 (m)	变量	值 (m)	变量	值 (m)
m_1	0.04	m_2	0.04	m_3	0.06
h_1	0.06	h_2	0.04	h_3	0.06
m_4	0.06	m_5	0.08	m_6	0.08
h_4	0.04	h_5	0.06	h_6	0.04

为了降低分布式控制难度, 将各智能体目标线性系统设定为如下同构系统:

$$\begin{cases} \dot{\xi}_i = A\xi_i + Bv_{i,k} \\ y_{i,k} = C\xi_i \end{cases}, i = 1, \dots, 6 \quad (57)$$

$$\text{其中, } A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^T.$$

4.1 案例 1: 学习有效性验证

在本案例中, 采用预设计的式 (57) 和以其作为对象设计的线性分布式控制器, 基于两阶段双启发式自适应动态规划算法优化反馈线性化控制器, 进行学习前后控制效果对比实验. 学习参数如表 2 所示.

奖励网络以扩展状态-动作对 $X_{i,k}$ 作为输入, 输出奖励值 $R_{i,k}$. 双评价网络以 $X_{i,k}$ 和 $R_{i,k}$ 为输入, 输出值函数 $J_{i,k}^l$ 和启发式函数值 $\hat{\lambda}_{i,k}^l$. 动作网络的输入为状态 $x_{i,k}$, 输出未知非线性项 $\hat{\beta}_i(x_{i,k})$ 的估计值. 网络的初始权值服从均值为 0、方差为 0.1 的分布.

在实验的初始阶段, 采用未训练的同胚分布式控制器对系统进行控制. 图 4(a) 和图 4(c) 显示了学习前的系统状态演化曲线. 结果表明未训练的控

表 2 学习参数

Table 2 Learning parameters

参数	值	参数	值	参数	值
η_r	0.05	η_c	0.02	η_a	0.01
γ	0.9	μ_j	0.01	μ_λ	0.01
ε_i	0.08	H	[1, 0.2]		

制器在应对异构非线性智能体系统时, 表现出较大的误差和不稳定性, 系统输出无法与期望轨迹一致. 原因在于系统的非线性动态和显著的异构性, 使得线性化控制策略无法适应所有智能体, 导致一致性控制效果不理想. 通过引入无模型反馈线性化算法, 并结合经验池和梯度下降对每个智能体的反馈线性化控制器进行训练, 系统控制性能显著提升. 学习收敛后, 系统收敛性和稳定性明显提高 (图 4(b) 和图 4(d)), 智能体输出与期望轨迹趋于一致, 跟踪误差显著减少, 验证了同胚分布式控制协议在模型未知的异构智能体系统中的有效性.

与现有动态规划方法不同, 本文无需预设计奖励信号的超参数. 但所提双启发式自适应动态规划算法仍然能够快速使各智能体的值函数网络和奖励函数网络的权值收敛 (图 5 和图 6), 体现出算法在应对非线性系统时具备较高的效率. 具体来说, 值函数网络通过奖励函数学习智能体线性化特征的长期动态行为, 逐步优化系统性能. 而奖励函数网络则动态调整奖励信号, 引导系统线性化效果.

值得注意的是, 图 7 中奖励函数的损失高于值函数损失, 说明直接使用原始奖励信号来驱动值函数学习可能会导致较大的波动性, 增加学习收敛的难度. 因此实验中引入的奖励值动态调整机制能够通过平滑奖励信号减少值函数网络的学习波动, 增强学习的稳定性.

4.2 案例 2: 方法优越性验证

在本案例中, 为验证所提方法的可扩展性和优越性, 在反馈线性化控制器学习收敛后, 将其与预设的分布式控制器共同作用于系统. 系统在稳定运行 30 s 后, 仅通过调整分布式控制器 $v_{i,k}$, 实现编队构型的快速调整 (图 8). 实验结果表明, 同胚分布式控制协议能通过调整虚拟输入端的线性控制器输入适应不同的动态性能要求, 无需重新学习.

所提无模型分布式控制方法与现有方法的显著区别在于, 本方法在学习收敛后, 得到的反馈线性化控制器与被控系统共同组成已知的线性化系统, 可利用线性系统理论进行控制与综合. 如果系统性能需求或环境发生改变, 也可以方便地调整线性控制输入, 而完全依赖学习的无模型分布式控制器设计方法由于状态空间发生改变, 则需要重新学习.

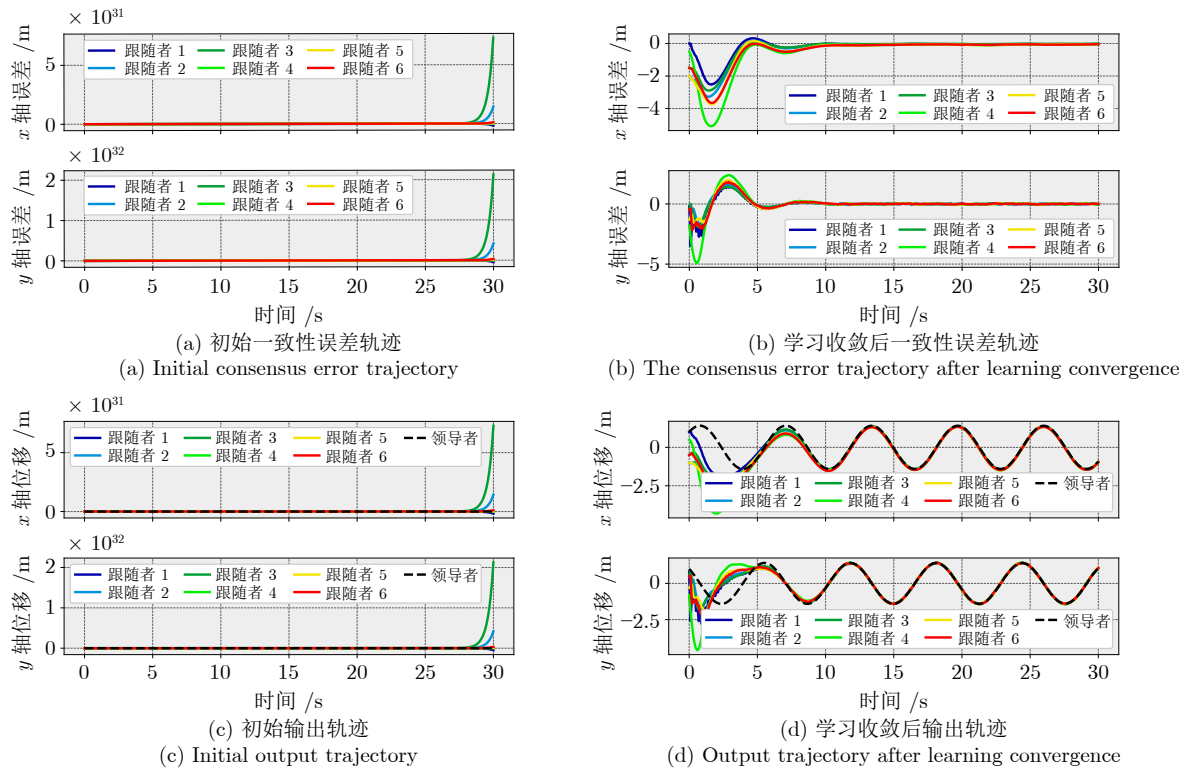


图 4 学习前后输出和一致性误差轨迹对比

Fig. 4 The output and consensus error trajectory comparison before and after learning

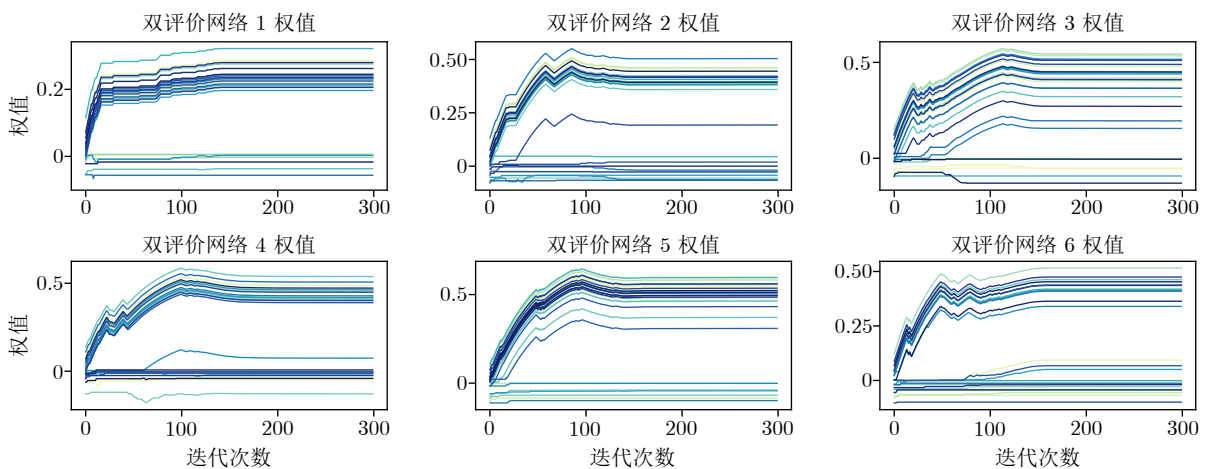


图 5 智能体双评价网络权值更新轨迹

Fig. 5 Agent dual-critic network weight update trajectory

5 结束语

本文提出一种同胚分布式控制协议, 解决了异构非线性多智能体系统的无模型输出一致性控制问题. 结合输入输出反馈线性化理论和自适应动态规划技术, 实现了无需系统模型的非线性系统线性化. 通过将异构非线性多智能体系统转为预设的同构线性系统, 简化了分布式控制器的设计, 使得线性控

制理论得以应用. 动态调整的奖励值和双阶段学习机制在训练过程中不断优化控制器, 增强了学习的稳定性和收敛速度. 实验结果表明, 各智能体的轨迹在所提方法下能够快速收敛到期望输出, 验证了控制策略的适应性和二次设计能力. 未来的研究将进一步讨论方法的泛化性, 考虑存在输入时滞、饱和、受限等情况, 扩展同胚分布式控制协议的适用范围, 以应对更复杂的实际应用场景.

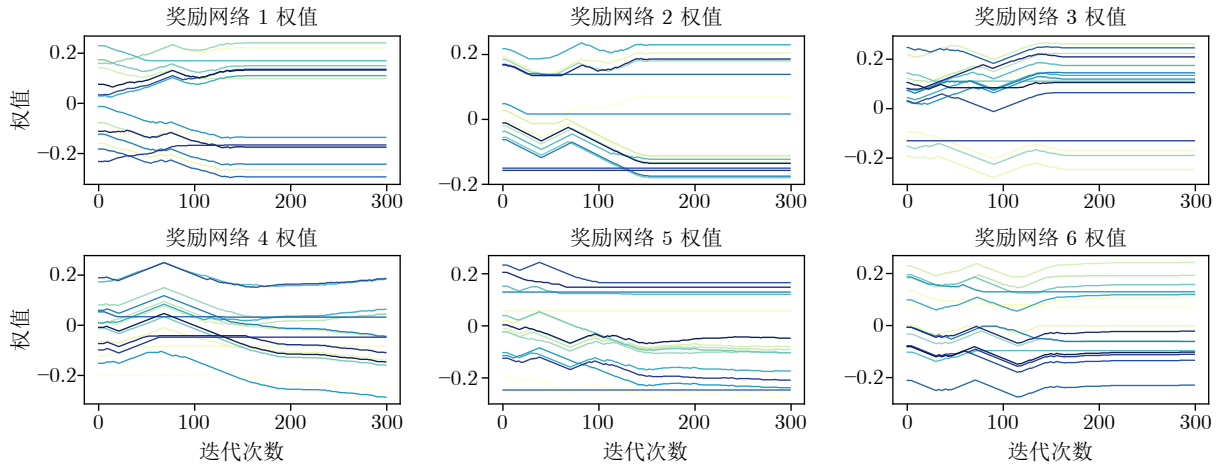


图 6 智能体奖励网络权值更新轨迹

Fig.6 Agent reward network weight update trajectory

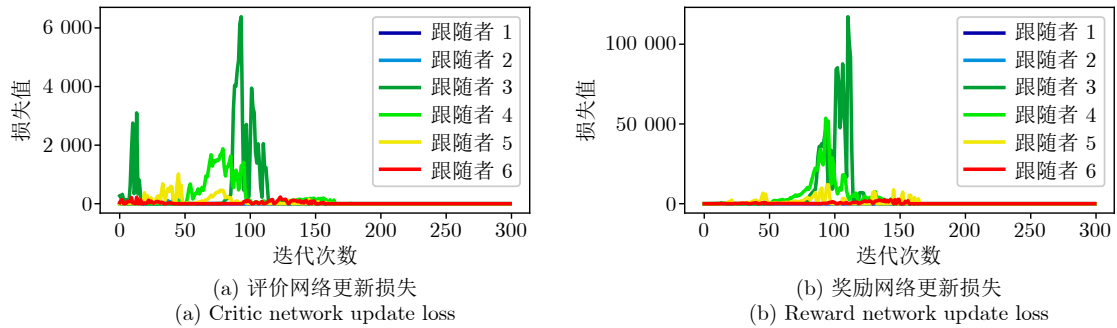


图 7 网络更新损失演化轨迹

Fig.7 Evolution trajectory of network update loss

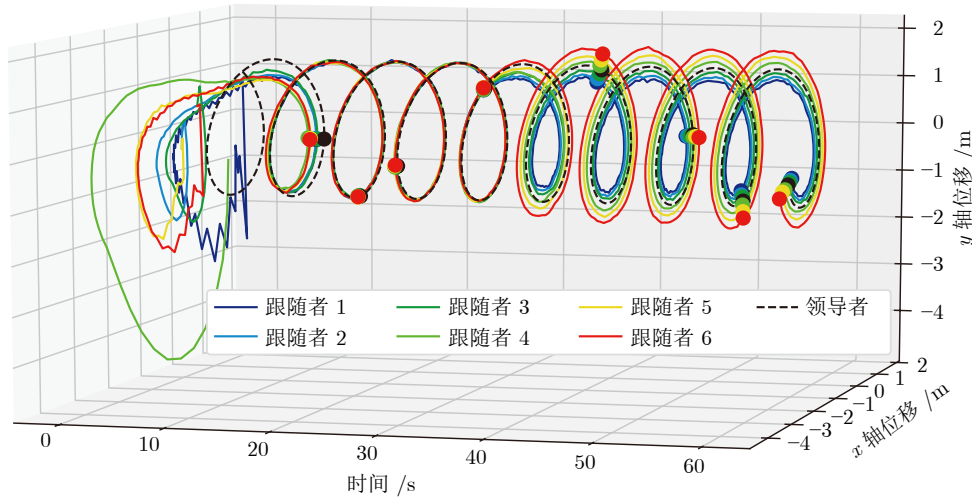


图 8 学习收敛后输出一致性轨迹切换实验

Fig.8 Output consensus trajectory switching experiment after learning convergence

References

- 1 Nair R R, Behera L. Robust adaptive gain higher order sliding mode observer based control-constrained nonlinear model predictive control for spacecraft formation flying. *IEEE/CAA Journal of Automatica Sinica*, 2016, 5(1): 367-381
- 2 Guo X C, Wei G L, Yao M, Zhang P J. Consensus control for multiple Euler-Lagrange systems based on high-order disturbance observer: An event-triggered approach. *IEEE/CAA Journal of Automatica Sinica*, 2022, 9(5): 945-948
- 3 Peng Z H, Wang D, Li T S, Han M. Output-feedback cooperative formation maneuvering of autonomous surface vehicles with

- connectivity preservation and collision avoidance. *IEEE Transactions on Cybernetics*, 2019, **50**(6): 2527–2535
- 4 Simões D, Lau N, Reis L P. Multi-agent actor centralized-critic with communication. *Neurocomputing*, 2020, **390**: 40–56
 - 5 Wu J, Lou Y C. Efficient centralized traffic grid signal control based on meta-reinforcement learning. *IEEE/CAA Journal of Automatica Sinica*, DOI: 10.1109/JAS.2023.123270
 - 6 Yan B, Shi P, Lim C C. Robust formation control for nonlinear heterogeneous multiagent systems based on adaptive event-triggered strategy. *IEEE Transactions on Automation Science and Engineering*, 2021, **19**(4): 2788–2800
 - 7 Bai C C, Yan P, Pan W, Guo J F. Learning-based multi-robot formation control with obstacle avoidance. *IEEE Transactions on Intelligent Transportation Systems*, 2021, **23**(8): 11811–11822
 - 8 Huang J Y, Zhou S Y, Tu H, Yao Y H, Liu Q S. Distributed optimization algorithm for multi-robot formation with virtual reference center. *IEEE/CAA Journal of Automatica Sinica*, 2022, **9**(4): 732–734
 - 9 Ju Y M, Ding D R, He X, Han Q L, Wei G L. Consensus control of multi-agent systems using fault-estimation-in-the-loop: Dynamic event-triggered case. *IEEE/CAA Journal of Automatica Sinica*, 2021, **9**(8): 1440–1451
 - 10 Yu X Y, Yang F, Zou C, Ou L L. Stabilization parametric region of distributed PID controllers for general first-order multi-agent systems with time delay. *IEEE/CAA Journal of Automatica Sinica*, 2019, **7**(6): 1555–1564
 - 11 Bidram A, Lewis F L, Davoudi A. Synchronization of nonlinear heterogeneous cooperative systems using input-output feedback linearization. *Automatica*, 2014, **50**(10): 2578–2585
 - 12 Sun Y P, Chen X, He W P, Zhang Z Y, Fukushima E F, She J. Q-learning based model-free input-output feedback linearization control method. *IFAC-PapersOnLine*, 2023, **56**(2): 9534–9539
 - 13 Li K, Hua C C, You X, Guan X P. Output feedback-based consensus control for nonlinear time delay multiagent systems. *Automatica*, 2020, **111**: Article No. 108669
 - 14 Wang D, Gao N, Liu D R, Li J N, Lewis F L. Recent progress in reinforcement learning and adaptive dynamic programming for advanced control applications. *IEEE/CAA Journal of Automatica Sinica*, 2024, **11**(1): 18–36
 - 15 Jiang H, He H B. Data-driven distributed output consensus control for partially observable multiagent systems. *IEEE Transactions on Cybernetics*, 2018, **49**(3): 848–858
 - 16 Jiang Y, Fan J L, Gao W N, Chai T Y, Lewis F L. Cooperative adaptive optimal output regulation of nonlinear discrete-time multi-agent systems. *Automatica*, 2020, **121**: Article No. 109149
 - 17 Lu X D, Li H T. Consensus of singular linear multiagent systems via hybrid control. *IEEE Transactions on Control of Network Systems*, 2022, **9**(2): 647–656
 - 18 Wen G X, Chen C L P, Feng J, Zhou N. Optimized multi-agent formation control based on an identifier-actor-critic reinforcement learning algorithm. *IEEE Transactions on Fuzzy Systems*, 2018, **26**(5): 2719–2731
 - 19 Bayili G, Nicaise S, Silga R. Rational energy decay rate for the wave equation with delay term on the dynamical control. *Journal of Mathematical Analysis and Applications*, 2021, **495**(1): Article No. 124693



孙一仆 中国地质大学(武汉)自动化学院博士研究生. 主要研究方向为多智能体系统, 强化学习.
E-mail: 20141000976@cug.edu.cn
(**SUN Yi-Pu** Ph.D. candidate at the School of Automation, China University of Geosciences. His re-

search interest covers multi-agent system and reinforcement learning.)



陈鑫 中国地质大学(武汉)自动化学院教授. 主要研究方向为智能控制, 过程控制, 机器人运动控制. 本文通信作者.

E-mail: chenxin@cug.edu.cn

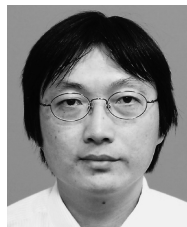
(**CHEN Xin** Professor at the School of Automation, China University of Geosciences. His research interest covers intelligent control, process control, and robot motion control. Corresponding author of this paper.)



贺文朋 中国地质大学(武汉)自动化学院博士研究生. 主要研究方向为多智能体系统分布式控制.

E-mail: wenpenghe@cug.edu.cn

(**HE Wen-Peng** Ph.D. candidate at the School of Automation, China University of Geosciences. His main research interest is multi-agent system distributed control.)



余锦华 日本东京工科大学教授. 主要研究方向为重复控制, 机电系统的高精度控制, 康复机器人, 计算智能的工业应用.

E-mail: she@stf.teu.ac.jp

(**SHE Jin-Hua** Professor at the Tokyo University of Technology, Japan. His research interest covers repetitive control, high precision control of mechatronic systems, rehabilitation robots, and industrial applications of computational intelligence.)



吴敏 中国地质大学(武汉)自动化学院教授. 主要研究方向为过程控制, 鲁棒控制和智能系统.

E-mail: wumin@cug.edu.cn

(**WU Min** Professor at the School of Automation, China University of Geosciences. His research interest covers process control, robust control, and intelligent systems.)