中国科学 : 信息科学 2025 年 第 55 卷 第 6 期: 1471–1485
SCIENTIA SINICA Informationis
论文

基于强化学习的柔性关节机器人系统模糊优化控制

王锐[1*], 文国兴[2], 刘艳军[3], 于福生[4], 武建[1]

1. 山西财经大学应用数学学院, 太原 030006
2. 山东航空学院理学院, 滨州 256600
3. 辽宁工业大学电气工程学院, 锦州 121001
4. 北京师范大学数学与科学学院, 北京 100875
* 通信作者. E-mail: rui-wang@live.com

收稿日期: 2025–01–21; 修回日期: 2025–04–30; 接受日期: 2025–05–12; 网络出版日期: 2025–06–09

国家自然科学基金重点项目 (批准号: 2023YFB4704403)、国家自然科学基金 (批准号: 12371453, 12102236, 12462005) 和山西自然科学基金 (批准号: 202403021211004, 202203021211334) 资助项目

摘要 针对非严格反馈结构下的柔性关节机器人系统, 本文提出了一种基于强化学习的简化模糊优化跟踪控制算法, 柔性关节机器人通过非严格反馈形式的四阶动态系统来描述, 采用模糊逻辑系统逼近未知函数, 并建立一个辅助自适应系统处理输入饱和问题. 基于 actor-critic 的简化强化学习算法设计模糊最优跟踪控制器. 此外, 采用非负函数的负梯度下降法, 而非贝尔曼 (Bellman) 残差误差平方法实现优化. 通过 Lyapunov 稳定性分析, 确保整个系统的半全局一致最终有界性. 仿真算法说明了基于强化学习的模糊跟踪控制策略的有效性.

关键词 最优控制, 模糊逻辑系统, 强化学习, 非严格反馈结构系统, actor-critic 执行器, 柔性关节机器人系统

1 引言

近年来, 实际工业过程中的机器人跟踪轨迹控制问题得到了广泛的关注, 尤其是对于柔性关节机器人系统 (flexible joint robot systems, FJRS) [1~6], 与刚性关节机器人系统相比, 它由于轴承变形、谐波驱动、轴卷绕等因素而具有关节柔性的高性能. 因此, FJRS 的建模和自适应鲁棒稳定性跟踪控制设计更为困难且极其复杂 [7~13]. 目前基于这类系统的控制问题主要集中于指令滤波反步顺应执行控制 [3,4]、电机驱动的模糊或神经网络全状态约束控制 [4~9] 以及运动规划协同控制问题. 基于事件触发机制的 PID 模糊控制在有限时间内的控制问题也推广到 FJRS 中 [10~12].

同时, 输入饱和 (input saturation, IS) [13] 作为实际控制系统中的一种重要的不确定性输入, 会影响系统性能 [13~17], 当 FJRS 系统受到 IS 的影响时, 如何补偿未知 IS 带来的影响意义重大. 针对带有扰动的自由飞行 FJRS 系统 [14~17] 分别实现了自适应神经网络混合阻抗控制、神经网络动态面控制 [16], 以及基于事件触发机制的网络化神经网络控制. 上述针对具有输入饱和 FJRS 的研究大都基于

引用格式: 王锐, 文国兴, 刘艳军, 等. 基于强化学习的柔性关节机器人系统模糊优化控制. 中国科学: 信息科学, 2025, 55: 1471–1485, doi: 10.1360/SSI-2025-0035
Wang R, Wen G X, Liu Y J, et al. Reinforcement learning-based optimized backstepping fuzzy tracking control for flexible-joint robot systems. Sci Sin Inform, 2025, 55: 1471–1485, doi: 10.1360/SSI-2025-0035

模糊逻辑系统或神经网络逼近的反推 (backstepping) 技术取得了很多成果, 而反推技术在迭代过程中需要对虚拟控制器进行反复求导数而引起计算膨胀问题, 且大多没有考虑优化控制技术. 这是我们的研究动机之一.

最优控制理论由 Bellman 和 Pontryagin 在文献 [18,19] 中首次提出, 不仅可以降低能源消耗还可以最小化成本函数 [20~23], 其主要通过求解 HJB (Hamilton-Jacobi-Bellman) 方程 [20] 来实现目标最优, 由于 HJB 方程具有高度非线性, 这使得问题变得极具挑战性. 文献 [21] 提出了一种新型的基于强化学习 (reinforcement learning, RL) 的多智能体系统控制方法以实现最优控制, 文献 [22,23] 基于识别器的 actor-critic 优化控制策略应用于非线性仿射系统. Wen 等 [24] 将基于识别的 actor-critic 执行器的 RL 优化反推 (optimized backstepping, OB) 技术应用于严格反馈的非仿射系统中. 随后这种优化控制策略被推广到更实际的水面舰艇系统 [25]、无人航空器系统 [26], 并对未知动态系统的简化 OB 跟踪控制技术进行了改进 [27~32]. 文献 [33,34] 实现了非严格反馈结构离散系统的 RL-OB 神经网络控制以及深度神经网络 OB 控制. 现有成果大多数未考虑柔性关节机器人系统, 且大多数基于 FJRSs 将其建模为严格反馈系统 [5~17], 同时也未考虑基于 actor-critic 的强化学习控制机制. 这是我们研究动机之二.

本文重点研究将一类简化的基于模糊强化学习的 actor-critic 控制方法推广到一类具有不确定饱和输入的 FJRSs 非严格反馈结构系统, 其中模糊逻辑系统 (fuzzy logic systems, FLSs) 用于逼近不确定性未知函数, 通过自适应辅助系统处理不确定饱和输入, 基于模糊识别器的 actor-critic 结构实现优化控制. 与相关文献 [24~29] 相比, 本文所考虑的非仿射形式下的非严格反馈结构系统更具有一般性, 也更符合实际系统, 现有的严格反馈系统和纯反馈系统的自适应反推控制方案不能直接应用于非严格反馈系统, 同时采用基于强化学习策略的负的正定函数的梯度下降法更新自适应律到最优控制器.

2　系统描述

2.1　问题引入

考虑一个具有 n 个连杆的 FJRS 系统, 其执行器的动态模型由欧拉 – 拉格朗日 (Euler-Lagrange) 方程 [14~17] 描述, 如下:

$$\begin{cases} N(p_1)\ddot{p_1} + D(p_1,\dot{p}_1)\dot{p}_1 + G(p_1,p_2) + F(p_1,\dot{p}_1) \\ = K(p_2-p_1) + \sigma_1(p_1,p_2,\dot{p}_1,\dot{p}_2), \\ J\ddot{p}_2 + B\dot{p}_2 + K(p_2-p_1) = u(\xi(t)) + \sigma_2(p_1,p_2,\dot{p}_1,\dot{p}_2), \end{cases} \tag{1}$$

其中 $p_1,\dot{p}_1,\ddot{p}_1 \in R^p$ 分别表示连杆位置、速度和加速度, $N(p_1) \in R^{n\times n}$ 是一个非对称的正定矩阵, 表示连杆惯性矩, $F(p_1,\dot{p}_1) \in R^n$, $G(p_1,p_2) \in R^n$, $D(p_1,\dot{p}_1) \in R^{n\times n}$ 分别为摩擦项、重力向量和科里奥利 (Coriolis) – 向心力. $p_2,\dot{p}_2,\ddot{p}_2 \in R^n$ 分别为转子的角位置、角速度和角加速度向量, $K \in R^{n\times n}$ 表示恒定的柔性关节矩阵, $J \in R^{n\times n}$ 是执行器的惯性矩阵, $B \in R^{n\times n}$ 是执行器阻尼项的自然矩阵. $\sigma_1,\sigma_2 \in R^n \times R^n$ 表示电机运行中的外部干扰. $u(\xi(t)) \in R^n$ 表示具有输入饱和的扭矩输入. $\xi(t) \in R^n$ 代表系统输入, $u(\cdot)$ 表示输入饱和, 表示为

$$u(\xi(t)) = \mathrm{sat}(\xi(t)) = \begin{cases} \mathrm{sign}(\xi)\mu_m, |\xi| \geqslant \mu_m, \\ \xi, |\xi| < \mu_m, \end{cases} \tag{2}$$

其中 u_m 表示 $u(\xi(t))$ 的界. 当 $|\xi(t)| = u_m$ 时表示一个奇点且满足 $\xi_{\bar{u}_\mu} = \bar{u}_\mu\xi + (1-\bar{u}_\mu)\eta_0$, 选取 $\eta_0 = 0$, 则上述公式可以重新描述为 $\bar{u}(\xi) = \bar{u}_\mu\xi$. 因此, 输入饱和函数可以表示为 $u(\xi(t)) = \bar{u}_\mu\xi + \eta(\xi)$. 由于 $\bar{u}(\xi)$ 是非增函数, 则存在两个正的常数 $\bar{u}_0$ 和 $\bar{u}_1$ 满足 $0 < \bar{u}_0 \leqslant \bar{u}_\mu \leqslant \bar{u}_1$. 因此, $\bar{u}_\mu$ 是有界的, 用于调

节输入饱和. 假定函数 $D(p_1,\dot{p}_1), G(p_1,p_2), F(p_1,\dot{p}_1), \sigma_1(p_1,p_2,\dot{p}_1,\dot{p}_2), \sigma_2(p_1,p_2,\dot{p}_1,\dot{p}_2)$ 是未知函数且至少满足局部 Lipschitz 条件.

注释 1. 在文献 [4∼9] 中, FJRS 系统考虑了具有外部干扰的控制器设计问题, 其外部干扰函数 σ_1, σ_2 仅是未知有界常数, 此时 FJRS 系统可以被视为严格反馈结构系统. 本文考虑的外部干扰函数是 $p_1,p_2,\dot{p}_1,\dot{p}_2$ 未知函数且该系统可以被建模为更一般的非严格反馈非线性系统. 因为非严格反馈系统更具有一般性也更加符合实际系统, 非严格反馈系统结构可用于描述许多实际系统, 如调谐振荡器、生化过程、机械系统、飞机视觉控制系统、质量和弹簧抑制系统、缓冲器、电气机械系统等, 因此控制非严格反馈非线性系统是一个非常具有挑战性的问题.

为了方便起见, 规定 $\chi_1=p_1,\chi_2=\dot{p}_1,\chi_3=p_2,\chi_4=\dot{p}_2$. 系统 (1) 重新表述为

$$\begin{cases}\dot{\chi}_1=\chi_2,\\ \dot{\chi}_2=\phi(\chi_1,\chi_2)\chi_3+\psi(\chi_1,\chi_2,\chi_3,\chi_4),\\ \dot{\chi}_3=\chi_4,\\ \dot{\chi}_4=J^{-1}u(\xi(t))+\psi_u(\chi_1,\chi_2,\chi_3,\chi_4),\end{cases}\tag{3}$$

其中 $\phi(\chi_1,\chi_2)=-N^{-1}(\chi_1)D(\chi_1,\chi_2)$, $\psi(\chi_1,\chi_2,\chi_3,\chi_4)=N^{-1}(\chi_1)[K(\chi_3-\chi_1)-G(\chi_1,\chi_3)-F(\chi_1,\chi_2)+\sigma_1(\chi_1,\chi_2,\chi_3,\chi_4)]$, $\psi_u(\chi_1,\chi_2,\chi_3,\chi_4)=J^{-1}[\sigma_2(\chi_1,\chi_2,\chi_3,\chi_4)-B\chi_4-K(\chi_3-\chi_1)]$.

假设1　假定连续控制增益函数 $\phi_2(\chi_1,\chi_2)$ 是有界的, 且存在常数 $\bar{\phi},\underline{\phi}$, 使其满足 $\underline{\phi}\leqslant\phi(\chi_1,\chi_2)\leqslant\bar{\phi}$.

控制目标: 针对 n 连杆柔性关节机器人系统 (1) 设计最优控制器, 使得目标成本函数最优, 系统输出可以很好地跟踪给定的参考信号 y_d, 同时确保系统所有信号满足半全局一致最终有界稳定性 (semi-globally uniformly ultimately bounded, SGUUB).

2.2　模糊逻辑系统

假定未知光滑函数 $f(\chi)$ 定义在紧集 Ω 上, 可以采用 FLSs $\beta^{\mathrm{T}}\omega(\chi)$ 对其进行逼近. 模糊规则为 R^l, $l=1,2,\ldots,q$: 如果 $\chi_1(t)$ 是 A_1^l 且如果 $\chi_2(t)$ 是 A_2^l 且 $\cdots$ 且如果 $\chi_m(t)$ 是 A_n^l, 则 y 是 B^l, 其中 A_i^l 和 B^l 为模糊隶属函数. FLSs 表述为

$$y(\chi)=\frac{\sum_{l=1}^{q}\bar{y}^l(\prod_{i=1}^{n}\mu_{A_i^l}(\chi_i))}{\sum_{l=1}^{q}\prod_{i=1}^{n}\mu_{A_i^l}(\chi_i)}.$$

规定 $\Theta^{\mathrm{T}}=[\bar{y}_1,\bar{y}_2,\ldots,\bar{y}_p]^{\mathrm{T}}$, $\omega(\chi)=[\omega^1,\omega^2,\ldots,\omega^p]^{\mathrm{T}}$, $\omega^l(\chi)=\frac{\prod_{i=1}^{n}\mu_{F_i^l}}{\sum_{l=1}^{p}\prod_{i=1}^{n}\mu_{F_i^l(\chi_i)}}$ 是模糊基函数, 则 FLSs 可以描述为 $y(\chi)=\Theta^{\mathrm{T}}\omega(\chi)$.

引理1 ([35])　对于任意常数 $\varepsilon>0$, 以及任意定义在紧集 Ω_f 上的连续函数 $f(\chi)$, 存在 FLSs 使得

$$\sup_{\chi\in\Omega}|(f(\chi)-\Theta^{\mathrm{T}}\omega(\chi))|\leqslant\varepsilon.$$

2.3　最优化控制理论

考虑不确定非线性系统: $\dot{q}(\tau)=f(q(\tau))+g(q(\tau))u(q(\tau))$, 这里 $q(\tau)\in R^n$ 为系统状态, $u(q(\tau))\in R^m$ 为系统输入, $f(q(\tau))\in R^n$, $g(q(\tau))\in R^{n\times m}$ 满足 $f(0)=0,g(0)=0$, 为非线性函数 [27]. 性能指标函数 [24,25] 定义如下:

$$J(q(0))=\int_0^{\infty}r(q(\tau),u(q(\tau)))\mathrm{d}\tau,$$

其中 $r(q(\tau),u(q(\tau)))=q(\tau)^{\mathrm{T}}Qq(\tau)+u^{\mathrm{T}}Pu\in R$ 为成本函数, $Q=Q^{\mathrm{T}}\in R^n\times R^n$ 和 $P\in R^m\times R^m$ 为正定矩阵. 则称定义在紧集 Ω_q 上的控制函数 $u(q(\tau))$ 是可达的, 满足 $u(0)=0$ 且 $u_q\in\Omega_q$. 则规定 u^*

为式 (6) 最优控制器. HJB 函数为

$$H(q,u^*,J^*)=q(t)^{\mathrm{T}}Qq(t)+u^{*\mathrm{T}}Pu^*+\frac{\mathrm{d}J^*(q)}{\mathrm{d}q^{\mathrm{T}}}(f(q)+g(q)u^*(q))=0.$$

注释 2. 如果 u^* 作为最优控制器, 求解 HJB 方程 $\partial H(q,u^*,J^*)/\partial u^*=0$, 解得 $u^*=-\frac{1}{2}P^{-1}g^{\mathrm{T}}(q)\times\frac{\mathrm{d}J^*(q)}{\mathrm{d}q}$. 由于 $\mathrm{d}J^*(q)/\mathrm{d}q$ 是未知且不可用的, 最优控制器 u^* 是不可行的. 因此, 若将 u^* 带入 HJB 方程求解, 可以得到 $H(q,u^*,J^*)=q(t)^{\mathrm{T}}Qq(t)+\frac{\mathrm{d}J^*(q)}{\mathrm{d}q^{\mathrm{T}}}(f(q)+g(q)u^*(q))=0$. 然而由于 HJB 方程具有高度复杂性和非线性, 求解 HJB 方程不切实际, 可以采用基于神经网络的强化学习实现智能识别控制, 本文采用基于模糊强化学习的智能控制识别 actor-critic 策略来解决这个问题, 以克服求解 HJB 方程的困难.

3　模糊自适应 OB 控制设计过程与分析

本节基于强化学习的 actor-critic 策略实现 n 连杆 FJRS 系统的模糊自适应优化控制方案, 首先定义系统的坐标变换为

$$\begin{cases}z_1(t)=\chi_1(t)-y_d(t),\\ z_i(t)=\chi_i(t)-\alpha_{i-1}(t),i=2,3,4,\end{cases}\tag{4}$$

其中 $z_i(t)$ 表示跟踪误差向量.

步骤 1. 针对第一个跟踪误差 $z_1(t)=\chi_1(t)-y_d(t)$, 并对其进行求导, 得到 $\dot{z}_1(t)=\dot{\chi}_1(t)-\dot{y}_d(t)=\chi_2(t)-\dot{y}_d(t)$, 其中 α_1 为虚拟控制器, α_1^* 是最优虚拟控制器, 性能指标函数被重新定义为

$$\begin{aligned}J_1^*(z_1)&=\min_{\alpha_1\in\psi(\Omega)}\left(\int_0^\infty r_1(z_1(\tau),\alpha_1(\chi(\tau)))\right)\mathrm{d}\tau\\&=\int_t^\infty r_1(z_1(\tau),\alpha_1^*(\chi(\tau)))\mathrm{d}\tau,\end{aligned}\tag{5}$$

其中 Ω 表示紧集, 定义局部成本函数如 $r_1(z_1(\tau),\alpha_1(\chi(\tau)))=z_1(\tau)^2+\alpha_1(\chi(\tau))^{*2}$, 同时把 χ_2 看作最优控制器 α_1^*. 结合式 (5), HJB 方程可以被描述为

$$H_1\left(z_1(t),\alpha_1^*,\frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}\right)=z_1^2+\alpha_1^{*2}+\frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}\times(\alpha_1^*-\dot{y}_r)=0.$$

则最优控制器 α_1^* 通过求解方程 $\partial H_1/\partial\alpha_1^*=0$ 可以得到, $\alpha_1^*(z_1)=-\beta_1z_1(t)-1/2J_1^0(\chi_1,z_1)$. 为了得到最优控制器, 规定 $\mathrm{d}J_1^*(z_1)/\mathrm{d}z_1=2\beta_1z_1(t)+J_1^0(\chi_1,z_1)$, 其中 β_1 为设计参数, $J_1^0(\chi_1,z_1)=-2\beta_1\times z_1(t)+\frac{\mathrm{d}J_1^*}{\mathrm{d}(z_1)}$ 为光滑连续函数. 由于 $J_1^0(\chi_1,z_1)$ 是未知的, 采用 FLSs 对其在紧集 Ω 上进行建模,

$$J_1^0(\chi_1,z_1)=\Theta_{J1}^{*\mathrm{T}}\omega_{J1}(\chi_1,z_1)+\varepsilon_{J1}(\chi_1,z_1),$$

其中 $\Theta_{J1}^{*\mathrm{T}}\in R^{q_1}$ 表示模糊逻辑系统的最优权重向量, $\omega_{J1}(\chi_1,z_1)$ 表示模糊基向量函数, ε_{J1} (χ_1,z_1) 为有界逼近误差且满足 $\varepsilon_{J1}(\chi_1,z_1)\leqslant\bar{\varepsilon}_{J1}$. 由于模糊逻辑系统的最优权重向量是未知的, 最优虚拟控制器 α_1^* 不能直接获得, 需采用 FLSs 的万能逼近性, 通过构建 actor-critic 模糊执行强化学习识别策略实现最优控制,

$$\mathrm{d}\hat{J}_1^*(z_1)/\mathrm{d}z_1=2\beta_1z_1(t)+2\hat{\Theta}_{c1}^{\mathrm{T}}(t)\omega_{J1}(\chi_1,z_1),\tag{6}$$

$$\hat{\alpha}_1^*(z_1)=-\beta_1z_1(t)+1/2\hat{\Theta}_{a_1}^{\mathrm{T}}(t)\omega_{J1}(\chi_1,z_1),\tag{7}$$

其中 $\mathrm{d}\hat{J}_1^*(z_1)/\mathrm{d}z_1\in R$ 用来估计 $\mathrm{d}J_1^*(z_1)/\mathrm{d}z_1$, $\hat{\Theta}_{c1}^{\mathrm{T}}\in R^{n_1}$ 且 $\hat{\Theta}_{a1}^{\mathrm{T}}\in R^{n_2}$ 表示系统的模糊 actor-critic 参数向量. 基于模糊逻辑系统 actor-critic 的训练自适应律设计如下:

$$\dot{\hat{\Theta}}_{c1}=-\gamma_{c1}\omega_{J1}(\chi_1(\tau),z_1(t))\omega_{J1}^{\mathrm{T}}(\chi_1(\tau),z_1(t))\hat{\Theta}_{c1}(t),\tag{8}$$

$$\dot{\hat{\Theta}}_{a1} = -\omega_{J1}(\chi_1(\tau), z_1(t))\omega_{J1}^{\mathrm{T}}(\chi_1(\tau), z_1(t)) \times (\gamma_{a1}(\hat{\Theta}_{a1}(t) - \hat{\Theta}_{c1}(t)) + \gamma_{c1}\hat{\Theta}_{c1}(t)), \tag{9}$$

其中 $\gamma_{a1} > 0, \gamma_{c1} > 0$ 为 actor-critic 设计常数, 且 $\beta_3 > 3, \gamma_{a1} > \frac{1}{2}, \gamma_{a1} > \gamma_{c1} > \frac{\gamma_{a1}}{2}$.

定义 Bellman 残差 [27] $e_1(t)$, 形如 $e_1(t) = H_1(z_1, \hat{\alpha}_1^*, \mathrm{d}\hat{J}_1^*/\mathrm{d}z_1) - H_1(z_1, \alpha_1^*, \mathrm{d}J_1^*/\mathrm{d}z_1) = H_1(z_1, \hat{\alpha}_1^*, d\hat{J}_1^*/dz_1)$, 则最优解 $\alpha_1^*(z_1)$ 满足 $e_1(t) = H_1(z_1, \hat{\alpha}_1^*, \mathrm{d}\hat{J}_1^*/\mathrm{d}z_1) \to 0$ 且 $\frac{\partial H_1(z_1, \hat{\alpha}_1^*, \mathrm{d}\hat{J}_1^*/\mathrm{d}z_1)}{\partial \hat{\Theta}_{a1}} = \frac{1}{2}\omega_{J1}^{\mathrm{T}}\omega_{J1}^{\mathrm{T}} \times (\hat{\Theta}_{a1} - \hat{\Theta}_{c1}(t))$. 同时定义如下的非负函数:

$$\rho_1(t) = (\hat{\Theta}_{a1}(t) - \hat{\Theta}_{c1}(t))^{\mathrm{T}}(\hat{\Theta}_{a1}(t) - \hat{\Theta}_{c1}(t)),$$

显然有 $\rho_1(t) = 0$ 成立, 通过适当调整自适应参数, 可以确保 $\rho_1(t) \to 0$ 成立.

注释 3. 上述方程意味着在满足自适应律 (8) 和 (9) 时, $\rho_1(t) \to 0$ 在最优性能下是允许的.

定义第一个 z_1- 子系统的 Lyapunov 能量函数如下:

$$V_1(t) = \frac{1}{2}z_1^2(t) + \frac{1}{2}\tilde{\Theta}_{a1}^{\mathrm{T}}(t)\tilde{\Theta}_{a1}(t) + \frac{1}{2}\tilde{\Theta}_{c1}^{\mathrm{T}}(t)\tilde{\Theta}_{c1}(t), \tag{10}$$

其中 $\tilde{\Theta}_{c1}(t) = \hat{\Theta}_{c1}(t) - \Theta_J^*$ 为基于模糊逻辑系统的自适应评判器, $\tilde{\Theta}_{a1}(t) = \hat{\Theta}_{a1}(t) - \Theta_J^*$ 表示基于 FLSs 的自适应执行器. 则对 Lyapunov 函数 $V_1(t)$ 进行求导, 可以得到

$$\dot{V}_1 = z_1(z_2 + \hat{\alpha}_1^* - \dot{y}_r) + \tilde{\Theta}_{a1}^{\mathrm{T}}\omega_{J1} \times \omega_{J1}^{\mathrm{T}}[\gamma_{a1}(\hat{\Theta}_{a1} - \hat{\Theta}_{c1}) + \gamma_{c1}\hat{\Theta}_{c1}] - \gamma_{c1}\tilde{\Theta}_{c1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{c1}. \tag{11}$$

结合 Young's 不等式, 可以得到 $z_2 z_1 \leqslant \frac{1}{2}z_1^2 + \frac{1}{2}z_2^2, z_1\varepsilon_{J1} \leqslant \frac{1}{2}z_1^2 + \frac{1}{2}\varepsilon_{J1}^2, -z_1\dot{y}_r \leqslant \frac{1}{2}z_1^2 + \frac{1}{2}\dot{y}_r^2$,

$$-\frac{1}{2}z_1\hat{\Theta}_{a1}^{\mathrm{T}}\omega_{J1} \leqslant \frac{1}{4}z_1^2 + \frac{1}{4}\hat{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{a1}, \tag{12}$$

$$\tilde{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{a1} \leqslant \frac{1}{2}\tilde{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\tilde{\Theta}_{a1} + \frac{1}{2}\hat{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{a1} - \frac{1}{2}(\Theta_{J1}^{*\mathrm{T}}\omega_{J1})^2, \tag{13}$$

$$\tilde{\Theta}_{c1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{c1} \leqslant \frac{1}{2}\tilde{\Theta}_{c1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\tilde{\Theta}_{c1} + \frac{1}{2}\hat{\Theta}_{c1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{c1} - \frac{1}{2}(\Theta_{J1}^{*\mathrm{T}}\omega_{J1})^2, \tag{14}$$

$$(\gamma_{c1} - \gamma_{a1})\tilde{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{c1} \leqslant \frac{\gamma_{c1} - \gamma_{a1}}{2}(\tilde{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\tilde{\Theta}_{a1} + \hat{\Theta}_{c1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\hat{\Theta}_{c1}). \tag{15}$$

通过选取恰当的设计参数可以提高系统性能, 其中 $\beta > \frac{7}{4}, \gamma_{a1} > \frac{1}{2}$, $\gamma_{a1} > \gamma_{c1} > \frac{\gamma_{a1}}{2}$, 确保自适应律为正的. 同时结合式 (13)∼(15), $\gamma_{a1} > \gamma_{c1}$ 是满足 Lyapunov 不等式 (16) 成立的基本条件,

$$\dot{V}_1 \leqslant -\left(\beta_1 - \frac{7}{4}\right)z_1^2 + \frac{1}{2}z_2^2 + D_1 - \frac{\gamma_{c1}}{2}\tilde{\Theta}_{c1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\tilde{\Theta}_{c1} - \frac{\gamma_{c1}}{2}\tilde{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\tilde{\Theta}_{a1}, \tag{16}$$

其中 $D_1(t) = \frac{1}{2}\varepsilon_{J1}^2 + \frac{1}{2}\dot{y}_r^2 + \frac{1}{2}z_2^2$, 由于 $\frac{1}{2}\varepsilon_{J1}^2, \frac{1}{2}\dot{y}_r^2, \frac{1}{2}z_2^2$ 的有界性, 存在一个常数 d_1 满足 $D_1(t) \leqslant d_1$. 规定 $\lambda_{\omega_{J1}}^{\min}$ 为 $\omega_{J1}\omega_{J1}^{\mathrm{T}}$ 的最小特征值, 得到下列不等式成立:

$$-\tilde{\Theta}_{c1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\tilde{\Theta}_{c1} \leqslant -\lambda_{\omega_{J1}}^{\min}\tilde{\Theta}_{c1}^{\mathrm{T}}\tilde{\Theta}_{c1}, \tag{17}$$

$$-\tilde{\Theta}_{a1}^{\mathrm{T}}\omega_{J1}\omega_{J1}^{\mathrm{T}}\tilde{\Theta}_{a1} \leqslant -\lambda_{\omega_{J1}}^{\min}\tilde{\Theta}_{a1}^{\mathrm{T}}\tilde{\Theta}_{a1}; \tag{18}$$

规定 $c_1 = \min\{\beta_1 - \frac{7}{4}, \gamma_{c1}\lambda_{\omega_{J1}}^{\min}\}$, 可以推出下列不等式成立:

$$\dot{V}_1 \leqslant -c_1 V_1 + d_1 + \frac{1}{2}z_2^2. \tag{19}$$

步骤 2. 定义第二个跟踪误差如下 $z_2 = \chi_2(t) - \hat{\alpha}_1^*(t)$, α_1, α_1^* 分别为虚拟控制器和最优控制器. 对其求导得 $\dot{z}_2(t) = \dot{\chi}_2 - \dot{\hat{\alpha}}_1^* = \phi(\chi_1, \chi_2)\chi_2 + \psi(\chi_1, \chi_2, \chi_3, \chi_4) - \dot{\hat{\alpha}}_1^*$, 这个子系统的最优性能指标函数可以设计为

$$\begin{aligned} J_2^*(z_2) &= \min_{\alpha_2 \in \psi(\Omega)}\left(\int_0^\infty r_2(z_2(s), \alpha_2(z_1, z_2))\right)\mathrm{d}s \\ &= \int_t^\infty r_2(z_2(s), \alpha_2^*(z_1, z_2))\mathrm{d}s, \end{aligned} \tag{20}$$

其中 $r_2(z_2(s),\alpha_2(z_1,z_2))=z_2^2+\alpha_2(z_1,z_2)^2$ 表示成本函数. 规定 $\chi_3(t)$ 为最优虚拟控制器 α_2^*, HJB 方程可以重新描述为

$$H_2\left(z_2,\alpha_2^*,\frac{\mathrm{d}J_2^*}{\mathrm{d}z_2}\right)=z_2^2+\alpha_2^{*\mathrm{T}}+\mathrm{d}J_2^*/\mathrm{d}z_2\times(\phi(\bar{\chi}_2)\alpha_2^*+\psi(\bar{\chi}_4)-\dot{\hat{\alpha}}_1^*)=0.$$

求解方程 $\partial H_2(z_2,\alpha_2^*,\frac{\mathrm{d}J_2^*}{\mathrm{d}z_2})/\partial\alpha_2^*=0$, 得到最优控制器为 $\alpha_2^*=\bar{\phi}(-\beta_2z_2(t)-\psi(\bar{\chi}_4)-(1/2)J_2^0(\bar{\chi}_4,z_2))$. 分解梯度项 $\mathrm{d}J_2^*(z_2)/\mathrm{d}z_2$, 得到 $\mathrm{d}J_2^*(z_2)/\mathrm{d}z_2=2\beta_2z_2(t)+2\psi(\bar{\chi}_4)+J_2^0(\bar{\chi}_4,z_2)$, 这里 $\beta_2>0$ 为设计参数, 且梯度项表示为 $J_2^0(\bar{\chi}_4,z_2)=-2\beta_2z_2(t)-2\psi(\bar{\chi}_4)+\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}$. 由于未知函数 $\psi(\bar{\chi}_4)$ 和梯度项 $J_2^0(\bar{\chi}_4,z_2)$ 是不可用的, 基于 FLSs 的逼近特性进行补偿,

$$\psi(\bar{\chi}_4)=\Theta_\psi^{*\mathrm{T}}\omega_\psi(\bar{\chi}_4)+\varepsilon_\psi(\bar{\chi}_4),\tag{21}$$

$$J_2^0(\bar{\chi}_4,z_2)=\Theta_{J2}^{*\mathrm{T}}S_{J2}(\bar{\chi}_4,z_2)+\varepsilon_{J2}(\bar{\chi}_4,z_2),\tag{22}$$

其中 $\Theta_\psi^{*\mathrm{T}}\in R^{p_2}$, $\Theta_{J2}^{*\mathrm{T}}\in R^{q_2}$ 表示模糊最优逼近向量, $\omega_\psi(\bar{\chi}_4)\in R^{p_i}$, $\omega_{J2}(\bar{\chi}_4,z_2)\in R^{q_i}$ 为模糊基函数且 $\varepsilon_\psi(\bar{\chi}_4)\in R,\varepsilon_{J2}(\bar{\chi}_4,z_2)\in R$ 表示模糊逼近误差. 由于梯度项是未知的, 采用 FLSs 对其进行建模,

$$\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}=2\beta_2z_2+2\Theta_\psi^{*\mathrm{T}}\omega_\psi(\bar{\chi}_4)+\Theta_{J2}^{*\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)+\varepsilon_2,$$

$$\alpha_2^*=\bar{\phi}\left(-\beta_2z_2-\Theta_\psi^{*\mathrm{T}}S_\psi(\bar{\chi}_4)-\frac{1}{2}\Theta_{J2}^{*\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)-\frac{1}{2}\varepsilon_2\right),$$

其中逼近误差为 $\varepsilon_2=2\varepsilon_\psi(\bar{\chi}_4)+2\varepsilon_{J2}$. 由于最优权重 $\Theta_\psi^{*\mathrm{T}}$ 和 $\Theta_{J2}^{*\mathrm{T}}$ 都是逼近常数向量, 不可以直接用于逼近函数, 也就是说最优控制器是不可用的. 因此规定 $\hat{\Theta}_\psi^{*\mathrm{T}}$ 为 $\Theta_\psi^{*\mathrm{T}}$ 的逼近器且 $\frac{\mathrm{d}\hat{J}_2^*(z_2)}{\mathrm{d}z_2}$, $\hat{\alpha}_2^*$ 分别表示 $\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}$, α_2^* 的逼近器, 基于强化学习 OB 算法构造自适应 actor-critic 更新器如下:

$$\hat{\psi}(\bar{\chi}_2)=\hat{\Theta}_\psi\omega_\psi(\bar{\chi}_2,z_2),\tag{23}$$

$$\frac{\mathrm{d}\hat{J}_2^*(z_2)}{\mathrm{d}z_2}=2\beta_2z_2+2\hat{\Theta}_\psi^{*\mathrm{T}}\omega_\psi(\bar{\chi}_4)+\hat{\Theta}_{c2}^{\mathrm{T}}(t)\omega_{J2}(\bar{\chi}_4,z_2),\tag{24}$$

$$\hat{\alpha}_2^*=\bar{\phi}\left(-\beta_2z_2-\hat{\Theta}_\psi^{\mathrm{T}}\omega_\psi(\bar{\chi}_4)-\frac{1}{2}\hat{\Theta}_{a2}^{\mathrm{T}}(t)\omega_{J2}(\bar{\chi}_4,z_2)\right),\tag{25}$$

其中 $\hat{\Theta}_\psi^{*\mathrm{T}}\in R^{p_2}$, $\hat{\Theta}_{c2}^{\mathrm{T}}(t)\in R^{q_2}$, $\hat{\Theta}_{a2}^{\mathrm{T}}(t)\in R^{q_2}$ 分别表示强化学习的模糊识别器权重、评判更新器权重以及执行更新器权重. 模糊自适应律设计为

$$\dot{\hat{\Theta}}_\psi(t)=\Gamma_2(\omega_\psi(\bar{\chi}_4,z_2)-\gamma_\psi\hat{\Theta}_\psi(t)),\tag{26}$$

$$\dot{\hat{\Theta}}_{c2}(t)=-\gamma_{c2}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{c2}(t),\tag{27}$$

$$\dot{\hat{\Theta}}_{a2}(t)=-\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\times(\gamma_{a2}(\hat{\Theta}_{a2}(t)-\hat{\Theta}_{c2}(t))+\gamma_{c2}\hat{\Theta}_{c2}(t)),\tag{28}$$

其中, Γ_2 表示正定矩阵, $\gamma_\psi>0$ 表示设计常数, 且 $\gamma_{a2}>0,\gamma_{c2}>0$ 分别表示执行评判权重, 选取恰当的自适应更新参数为 $\beta_2>3,\gamma_{a2}>\frac{1}{2},\gamma_{a2}>\gamma_{c2}>\frac{\gamma_{a2}}{2}$. 规定 $z_3=\chi_3-\hat{\alpha}_2^*$, 可以得到 $\dot{z}_2=\phi(\bar{\chi}_2)(z_3+\hat{\alpha}_2^*)+\psi(\bar{\chi}_4)-\dot{\hat{\alpha}}_1^*(t)$. 对于 z_2- 子系统, 定义 Lyapunov 能量函数为

$$V_2=V_1+\frac{1}{2}z_2^2+\frac{1}{2}\tilde{\Theta}_\psi^{\mathrm{T}}\Gamma_2^{-1}\tilde{\Theta}_\psi+\frac{1}{2}\tilde{\Theta}_{a2}^{\mathrm{T}}\tilde{\Theta}_{a2}+\frac{1}{2}\tilde{\Theta}_{c2}^{\mathrm{T}}\tilde{\Theta}_{c2},\tag{29}$$

其中 $\tilde{\Theta}_{\psi}^{\mathrm{T}}=\hat{\Theta}_{\psi}-\Theta_{\psi}^{*},\tilde{\Theta}_{c2}=\hat{\Theta}_{c2}-\Theta_{J2}^{*},\tilde{\Theta}_{a2}=\hat{\Theta}_{a2}-\Theta_{J2}^{*}$ 分别表示模糊逻辑系统的识别误差、执行和评判误差. 对 Lyapunov 函数 $V_2(t)$ 进行求导得到

$$\begin{aligned}\dot{V}_2=\dot{V}_1+z_2\left(-\beta_2 z_2-\frac{1}{2}\hat{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)+\varepsilon_{\psi}(\bar{\chi}_4)-\hat{\alpha}_1^{*}+z_3\phi(\bar{\chi}_2)\right)-\gamma_{\psi}\tilde{\Theta}_{\psi}^{\mathrm{T}}S_{\psi}(\bar{\chi}_4)\hat{\Theta}_{\psi}\\-\gamma_{c2}\tilde{\Theta}_{c2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)S_{J2}(\bar{\chi}_4,z_2)^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{c2}+\tilde{\Theta}_{a2}^{\mathrm{T}}[-\omega_{J2}(\bar{\chi}_4,z_2)S_{J2}(\bar{\chi}_4,z_2)^{\mathrm{T}}\\\times(\gamma_{a2}(\hat{\Theta}_{a2}-\hat{\Theta}_{c2}))+\gamma_{c2}\hat{\Theta}_{c2}].\end{aligned}\tag{30}$$

结合 Young's 不等式可以得到 $z_2\varepsilon_{\psi}\leqslant\frac{1}{2}z_2^2+\frac{1}{2}\varepsilon_{\psi}^2,-z_2\dot{\hat{\alpha}}_1^{*}\leqslant\frac{1}{2}z_2^2+\frac{1}{2}\dot{\hat{\alpha}}_1^{*2},-z_2z_3\phi(\bar{\chi}_2)\leqslant\frac{1}{2}z_2^2+\frac{1}{2}\bar{\phi}^2z_3^2$,

$$z_2\hat{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\leqslant z_2^2+\hat{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\times\omega_{J2}(\bar{\chi}_4,z_2)^{\mathrm{T}}\hat{\Theta}_{a2},\tag{31}$$

$$-\gamma_{\psi_2}\tilde{\Theta}_{\psi_2}^{\mathrm{T}}\hat{\Theta}_{\psi_2}\leqslant\frac{\gamma_{\psi_2}}{2}\tilde{\Theta}_{\psi_2}^{\mathrm{T}}\tilde{\Theta}_{\psi_2}+\frac{\gamma_{\psi_2}}{2}\hat{\Theta}_{\psi_2}^{\mathrm{T}}\hat{\Theta}_{\psi_2}-\frac{\gamma_{\psi_2}}{2}(\Theta_{\psi_2}^{*\mathrm{T}}\Theta_{\psi_2}^{*}),\tag{32}$$

$$\begin{aligned}\gamma_{c2}\tilde{\Theta}_{c2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{c2}\leqslant\frac{\gamma_{c2}}{2}\tilde{\Theta}_{c2}^{\mathrm{T}}(t)\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\tilde{\Theta}_{c2}(t)\\+\frac{\gamma_{c2}}{2}\hat{\Theta}_{c2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{c2}-\frac{\gamma_{c2}}{2}(\Theta_{J2}^{*\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2))^2,\end{aligned}\tag{33}$$

$$\begin{aligned}\gamma_{a2}\tilde{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{a2}\leqslant\frac{\gamma_{a2}}{2}\tilde{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\tilde{\Theta}_{a2}\\+\frac{\gamma_{a2}}{2}\hat{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{a2}-\frac{\gamma_{a2}}{2}(\Theta_{J2}^{*\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2))^2,\end{aligned}\tag{34}$$

$$\begin{aligned}(\gamma_{a2}-\gamma_{c2})\tilde{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{c2}\leqslant\frac{\gamma_{a2}-\gamma_{c2}}{2}(\tilde{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\tilde{\Theta}_{a2}\\+\frac{\gamma_{a2}-\gamma_{c2}}{2}\hat{\Theta}_{c2}^{\mathrm{T}}\omega_{J1}(\bar{\chi}_4,z_2)\omega_{J1}^{\mathrm{T}}(\bar{\chi}_4,z_2)\hat{\Theta}_{c2}),\end{aligned}\tag{35}$$

其中 $D_2=\frac{1}{2}\varepsilon_{\psi}^2+\frac{1}{2}\dot{\hat{\alpha}}_1^{*2}+\frac{\gamma_2}{2}\Theta_{\psi}^{*\mathrm{T}}\Theta_{\psi}^{*}+\frac{(\gamma_{a2}+\gamma_{c2})}{2}(\Theta_{J2}^{*\mathrm{T}}S_{J2}(\bar{\chi}_4,z_2))^2$ 是有界的, 存在 d_2 且满足 $D_2\leqslant d_2$. 规定 $\lambda_{\Gamma_2^{-1}}^{\max}$ 和 $\lambda_{S_{J2}}^{\min}$ 分别为矩阵 $\Gamma_2^{-1},S_{J2}(\bar{\chi}_4,z_2)S_{J2}(\bar{\chi}_4,z_2)^{\mathrm{T}}$ 的最大和最小特征值. 将其带入 Lyapunov 函数 (30), 得到下式成立:

$$\begin{aligned}\dot{V}_2\leqslant-c_1V_1+D_1+\left(-\beta_2+\frac{9}{4}\right)z_2^2+D_2+\frac{1}{2}\bar{\phi}^2z_3^2-\frac{\gamma_2}{2\lambda_{\Gamma_2^{-1}}^{\max}}\tilde{\Theta}_{\psi_2}^{\mathrm{T}}\tilde{\Theta}_{\psi_2}\\-\frac{\gamma_{c2}}{2}\lambda_{\omega_{J2}}^{\min}\tilde{\Theta}_{c2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\tilde{\Theta}_{c2}-\frac{\gamma_{c2}}{2}\lambda_{\omega_{J2}}^{\min}\tilde{\Theta}_{a2}^{\mathrm{T}}\omega_{J2}(\bar{\chi}_4,z_2)\omega_{J2}^{\mathrm{T}}(\bar{\chi}_4,z_2)\tilde{\Theta}_{a2}.\end{aligned}\tag{36}$$

基于设计参数 $\beta_2>\frac{9}{4},\gamma_{a1}>\frac{1}{2},\gamma_{a1}>\gamma_{c1}>\frac{\gamma_{a1}}{2}$, 定义 $c_2=\min\{2(\beta_2-\frac{9}{4}),\frac{\gamma_2}{\lambda_{\Gamma_2^{-1}}^{\max}},\gamma_{c2}\lambda_{\omega_{J2}}^{\min}\}$, 则下列公式成立:

$$\dot{V}_2\leqslant\sum_{j=1}^{2}(-c_jV_j+d_j)+\frac{1}{2}z_3^2\bar{\phi}^2.\tag{37}$$

步骤 3. 对于动态子系统 $z_3=\chi_3-\hat{\alpha}_2^{*}$, 规定 α_3, α_3^{*} 分别为虚拟控制器和最优虚拟控制器. 针对动态误差模型 $\dot{z}_3=\dot{\chi}_3-\dot{\hat{\alpha}}_2^{*}=\chi_4-\dot{\hat{\alpha}}_3^{*}$. 定义 z_3- 子系统的最优性能指标函数,

$$\begin{aligned}J_3^{*}(z_3)&=\min_{\alpha_3\in\psi(\Omega)}\left(\int_t^{\infty}r_3(z_3(s),\alpha_3(z_1,z_2,z_3))\right)\mathrm{d}s\\&=\int_t^{\infty}r_3(z_3(s),\alpha_3^{*}(z_1,z_2,z_3))\mathrm{d}s,\end{aligned}\tag{38}$$

其中 $r_3(z_3(s),\alpha_3(z_1,z_2,z_3))=z_3^2+\alpha_3^2(z_1,z_2,z_3)$ 为成本函数. 规定 $\chi_4(t)$ 为最优虚拟控制器 α_3^{*}, 则 HJB 方程重新表述为

$$H_3\left(z_3,\alpha_3^{*},\frac{\mathrm{d}J_3^{*}}{\mathrm{d}z_3}\right)=z_3^2+\alpha_3^{*2}+\frac{\mathrm{d}J_3^{*}}{\mathrm{d}z_3}\times(\alpha_3^{*}(t)-\dot{\hat{\alpha}}_2^{*}(t))=0.$$

解方程 $\partial H_3(z_3,\alpha_3^*,\frac{\mathrm{d}J_3^*}{\mathrm{d}z_3})/\partial\alpha_3^*=0$, 可以得到最优控制器 α_3^* 为 $\alpha_3^*=-\beta_3 z_3-(1/2)J_3^0(\bar{\chi}_4,z_3)$, 定义梯度 $\mathrm{d}J_3^*(z_3)/\mathrm{d}z_3$ 为 $\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}=2\beta_3 z_3+J_3^0(\bar{\chi}_4,z_3)$, 其中 $\beta_3>0$ 为设计参数且 $J_3^0(\bar{\chi}_4,z_3)=-2\beta_3 z_3+\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}$. 由于 $J_3^0(\bar{\chi}_4,z_3)$ 是不可用的, 采用 FLSs 对其进行逼近,

$$J_3^0(\bar{\chi}_4,z_3)=\Theta_{J3}^{*\mathrm{T}}S_{J3}(\bar{\chi}_4,z_3)+\varepsilon_{J3}(\bar{\chi}_4,z_3),$$

其中 $\Theta_{J3}^{*\mathrm{T}}\in R^{q_1}$ 为模糊逻辑系统的最优权重向量, $S_{J3}(\bar{\chi}_4,z_3)$ 是模糊基函数且 $\varepsilon_{J3}(\bar{\chi}_4,z_3)$ 是逼近误差 $\varepsilon_{J3}(\bar{\chi}_4,z_3)\leqslant\bar{\varepsilon}_{J3}$. 因此, 结合最优虚拟控制器和梯度得到

$$\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}=2\beta_3 z_3(t)+\Theta_{J3}^{*\mathrm{T}}S_{J3}(\bar{\chi}_4,z_3)+\varepsilon_{J3}(\bar{\chi}_4,z_3),$$

$$\alpha_3^*(z_3)=-\beta_3 z_3(t)+\frac{1}{2}\Theta_{J3}^{*\mathrm{T}}S_{J3}(\bar{\chi}_4,z_3)+\varepsilon_{J3}(\bar{\chi}_4,z_3).$$

由于 FLSs 的最优权重向量 $\Theta_{J3}^{*\mathrm{T}}$ 是逼近向量, 基于 RL 识别器, actor-critic 算法设计为

$$\mathrm{d}\hat{J}_3^*(z_3)/\mathrm{d}z_3=2\beta_3 z_3(t)+\hat{\Theta}_{c3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_4,z_3),\tag{39}$$

$$\hat{\alpha}_3^*(z_3)=-\beta_3 z_3(t)+\frac{1}{2}\hat{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_4,z_3),\tag{40}$$

其中 $\frac{\mathrm{d}\hat{J}_3^*(z_3)}{\mathrm{d}z_3}\in R$ 表示 $\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}$ 的估计量, $\hat{\Theta}_{c3}^{\mathrm{T}}\in R^{n_1}$ 和 $\hat{\Theta}_{a3}^{\mathrm{T}}(t)\in R^{n_2}$ 表示模糊 actor-critic 权重, 设计如下:

$$\dot{\hat{\Theta}}_{c3}=-\gamma_{c3}\omega_{J3}(\bar{\chi}_4,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_4,z_3)\hat{\Theta}_{c3}(t),\tag{41}$$

$$\dot{\hat{\Theta}}_{a3}=-\omega_{J3}(\bar{\chi}_4,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_4,z_3)\times(\gamma_{a3}(\hat{\Theta}_{a3}(t)-\hat{\Theta}_{c3}(t))+\gamma_{c3}\hat{\Theta}_{c3}(t)),\tag{42}$$

$\gamma_{a3}>0,\gamma_{c3}>0$ 为执行评判器设计常数.

定义 z_3- 子系统的动态误差 Lyapunov 函数 $\dot{z}_4=\dot{\chi}_4-\dot{\hat{\alpha}}_3^*$ 如下:

$$V_3=V_1+V_2+\frac{1}{2}z_3^2+\frac{1}{2}\tilde{\Theta}_{a3}^{\mathrm{T}}\tilde{\Theta}_{a3}+\frac{1}{2}\tilde{\Theta}_{c3}^{\mathrm{T}}\tilde{\Theta}_{c3},\tag{43}$$

其中 $\tilde{\Theta}_{c3}=\hat{\Theta}_{c3}-\Theta_J^*,\tilde{\Theta}_{a3}=\hat{\Theta}_{a3}-\Theta_J^*$, 表示基于 FLSs-RL 的执行评判器向量. 对其 V_3 进行求导,

$$\begin{aligned}\dot{V}_3=&\dot{V}_1+\dot{V}_2+z_3(z_4+\hat{\alpha}_3^*-\dot{\hat{\alpha}}_2^*)+\tilde{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\\&\times[\gamma_{a3}(\hat{\Theta}_{a3}-\hat{\Theta}_{c3})+\gamma_{c3}\hat{\Theta}_{c3}]-\gamma_{c3}\tilde{\Theta}_{c3}^{\mathrm{T}}S_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{c3}.\end{aligned}\tag{44}$$

基于 Young's 不等式得到下列不等式:

$$z_3z_4\leqslant\frac{1}{2}z_3^2+\frac{1}{2}z_4^2,-z_3\dot{\hat{\alpha}}_2^*\leqslant\frac{1}{2}z_3^2+\frac{1}{2}\dot{\hat{\alpha}}_2^*,z_3\varepsilon_{J3}\leqslant\frac{1}{2}z_3^2+\frac{1}{2}\varepsilon_{J3}^2,$$

$$-z_3\hat{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\leqslant z_3^2+\hat{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J1}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{a3},\tag{45}$$

$$\begin{aligned}&\tilde{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{a3}\\&\quad\leqslant\tilde{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\tilde{\Theta}_{a3}+\hat{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{a3}-(\Theta_{J3}^{*\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3))^2,\end{aligned}\tag{46}$$

$$\begin{aligned}&\tilde{\Theta}_{c3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{c3}\\&\quad\leqslant\tilde{\Theta}_{c3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\tilde{\Theta}_{c3}+\frac{1}{2}\hat{\Theta}_{c3}^{\mathrm{T}}(t)\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{c3}-(\Theta_{J3}^{*\mathrm{T}}\omega_{J3}(\bar{\chi}_3,z_3))^2,\end{aligned}\tag{47}$$

$$\begin{aligned}&(\gamma_{c1}-\gamma_{a3})\tilde{\Theta}_{a3}^{\mathrm{T}}(t)\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{c3}(t)\\&\quad\leqslant\frac{\gamma_{c3}(t)-\gamma_{a1}}{2}(\tilde{\Theta}_{a3}^{\mathrm{T}}(t)\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\tilde{\Theta}_{a3}(t)+\hat{\Theta}_{c3}^{\mathrm{T}}(t)\omega_{J3}(\bar{\chi}_3,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_3,z_3)\hat{\Theta}_{c3}(t)).\end{aligned}\tag{48}$$

将这些不等式带入 Lyapunov 函数 (44), 得到

$$\begin{aligned}\dot{V}_3 \leqslant & \sum_{j=1}^{2}(-c_jV_j+d_j)+\left(-\beta_3+\frac{7}{4}\right)z_3^2+D_3+\frac{z_4^2}{2}\\ & -\frac{\gamma_{c3}}{2}\lambda_{\omega_{J3}}^{\min}\tilde{\Theta}_{c3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_4,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_4,z_3)\tilde{\Theta}_{c3}-\frac{\gamma_{c3}}{2}\lambda_{\omega_{J3}}^{\min}\tilde{\Theta}_{a3}^{\mathrm{T}}\omega_{J3}(\bar{\chi}_4,z_3)\omega_{J3}^{\mathrm{T}}(\bar{\chi}_4,z_3)\tilde{\Theta}_{a3},\end{aligned}\tag{49}$$

其中 $D_3=\frac{1}{2}\varepsilon_{J3}^2+\frac{1}{2}\dot{\hat{\alpha}}_2^{*2}+\frac{(\gamma_{a3}+\gamma_{c3})}{2}(W_{J3}^{*\mathrm{T}}\omega_{J3}(\bar{\chi}_4,z_3))^2$ 是有界的且存在常数 d_3 满足 $D_3\leqslant d_3$. 基于这些不等式 $\beta_3>\frac{7}{4},\gamma_{a3}>\frac{1}{4},\gamma_{a3}>\gamma_{c3}>\frac{\gamma_{a3}}{2}$, 定义 $\lambda_{\omega_{J3}}^{\min}$ 为 $\omega_{J3}(\bar{\chi}_4,z_3)$, $\omega_{J3}(\bar{\chi}_4,z_3)^{\mathrm{T}}$ 的特征值, 规定 $c_3=\min\{2(\beta_3-\frac{7}{4}),\gamma_{c3}\lambda_{\omega_{J3}(\bar{\chi}_4,z_3)}^{\min}\}$, 则下列不等式成立:

$$\dot{V}_3\leqslant\sum_{j=1}^{2}(-c_3V_j+d_j)+\frac{1}{2}z_4^2.\tag{50}$$

步骤 4. 在这一步设计实际的控制器 $u(\xi)$, $z_4=\chi_4-\hat{\alpha}_3^*$, 动态误差 z_4 设计为 $\dot{z}_4=\dot{\chi}_4-\dot{\hat{\alpha}}_3^*=J^{-1}u(\xi)+\psi_4(\chi_1,\chi_2,\chi_3,\chi_4)-\dot{\hat{\alpha}}_3^*$, 规定 $u^*(\xi(t))$ 为最优控制器, 设计积分型最优成本函数为

$$\begin{aligned}J_4^*(z_4)&=\min_{u\in\psi(\Omega)}\left(\int_t^{\infty}r_4(z_4(\tau),u(\xi(t),z_4))\right)\mathrm{d}\tau\\&=\int_t^{\infty}r_4(z_4(\tau),u^*(\xi(t),z_4))\mathrm{d}\tau,\end{aligned}\tag{51}$$

其中 $r_4(z_4(\tau),u^*(\xi,z_4))=z_4^2+u^*(\xi(\tau),z_4)^2$ 表示成本函数. HJB 方程可以描述为

$$H_4\left(z_4,u^*(\xi,z_4),\frac{\mathrm{d}J_4^*}{\mathrm{d}z_4}\right)=z_4^2+u^*(\xi,z_4)+\frac{\mathrm{d}J_4^*}{\mathrm{d}z_4}\times(J^{-1}u(\xi)+\psi_4(\bar{\chi}_4)-\dot{\hat{\alpha}}_3^*)=0.\tag{52}$$

求解方程 $\partial H_4(z_4,u^*(\xi),\frac{\mathrm{d}J_4^*}{\mathrm{d}z_4})/\partial u^*(\xi)=0$, 得到 $u^*(\xi)=-J^{-1}\beta_4z_4-J^{-1}\psi_u(\bar{\chi}_4)-\frac{J^{-1}}{2}J_4^0(\bar{\chi}_4,z_4)$. 规定梯度函数 $\mathrm{d}J_4^*(z_4)/\mathrm{d}z_4$ 为 $\frac{\mathrm{d}J_4^*(z_4)}{\mathrm{d}z_4}=2\beta_4z_4+2\psi_4(\bar{\chi}_4)+J_4^0(\bar{\chi}_4,z_4)$, 其中 $\beta_4>0$ 表示设计参数且 $J_4^0(\bar{\chi}_4,z_4)=-2\beta_4z_4(t)-2\psi_4(\bar{\chi}_4)+\frac{\mathrm{d}J_4^*(z_4)}{\mathrm{d}z_4}$. 由于 $J^{-1}\psi_u(\bar{\chi}_4)$ 和 $J^{-1}J_4^0(\bar{\chi}_4,z_4)$ 不可用, 采用 FLSs 逼近这些不可用函数如下:

$$\psi_u(\bar{\chi}_4)=\Theta_{\psi_u}^{*\mathrm{T}}\omega_{\psi_u}(\bar{\chi}_4)+\varepsilon_{\psi_u}(\bar{\chi}_4),\tag{53}$$

$$J_4^0(\bar{\chi}_4,z_4)=\Theta_{J4}^{*\mathrm{T}}S_{J4}(\bar{\chi}_4,z_4)+\varepsilon_{J4}(\bar{\chi}_4,z_4),\tag{54}$$

其中 $\Theta_{\psi_u}^{*\mathrm{T}}\in R^{p_2}$, $\Theta_{J4}^{*\mathrm{T}}\in R^{q_2}$ 表示模糊最优权重, $\omega_{\psi_u}(\bar{\chi}_4)\in R^{p_i}$, $\omega_{J4}(\bar{\chi}_4,z_4)\in R^{q_i}$ 是模糊基向量函数, $\varepsilon_{\psi_u}(\bar{\chi}_4)\in R,\varepsilon_{J4}(\bar{\chi}_4,z_4)\in R$ 表示逼近误差. 由于梯度项未知, 采用 FLSs 对其进行建模,

$$\frac{\mathrm{d}J_4^*(z_4)}{\mathrm{d}z_4}=2\beta_4z_4+2\Theta_{\psi_u}^{*\mathrm{T}}S_{\psi_u}(\bar{\chi}_4)+\Theta_{J4}^{*\mathrm{T}}S_{J4}(\bar{\chi}_4,z_4)+\varepsilon_4,$$

$$u^*(\xi)=J\beta_4z_4-\frac{J}{2}\Theta_{\psi_u}^{*\mathrm{T}}\omega_{\psi_u}(\bar{\chi}_4)-\frac{J}{2}\Theta_{J4}^{*\mathrm{T}}\omega_{J4}(\bar{\chi}_4,z_4)-\frac{\varepsilon_4}{2},$$

其中 $\varepsilon_4=2\varepsilon_{\psi_u}(\bar{\chi}_4)+2\varepsilon_{J4}$. 由于最优权重 $\Theta_{\psi_u}^{*\mathrm{T}}$ 和 Θ_{J4}^* 作为逼近常数是不可用的, 此时最优控制器是不可行的. 因此, 基于 actor-critic-RL 算法设计有效的最优控制器,

$$\hat{\psi}_u(\bar{\chi}_4)=\hat{\Theta}_{\psi_u}\omega_{\psi_u}(\bar{\chi}_4,z_4),\tag{55}$$

$$\frac{\mathrm{d}\hat{J}_4^*(z_4)}{\mathrm{d}z_4}=2\beta_4z_4+2\hat{\Theta}_{\psi_u}\omega_{\psi_4}(\bar{\chi}_4,z_4)+\hat{\Theta}_{c_4}^{\mathrm{T}}(t)\omega_{J4}(\bar{\chi}_4,z_4),\tag{56}$$

$$\hat{u}^*(\xi)=-J\beta_4z_4-J\hat{\Theta}_{\psi_u}^{\mathrm{T}}\omega_{\psi_u}(\bar{\chi}_4)-\frac{1}{2}J\hat{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4,z_4),\tag{57}$$

其中 $\hat{\psi}_u(\bar{\chi}_4)$, $\frac{\mathrm{d}\hat{J}_4^*(z_4)}{\mathrm{d}z_4}$, $\hat{\Theta}_{\psi_u} \in R^{q_n}$, $\hat{\Theta}_{c2}^{\mathrm{T}}(t) \in R^{q_2}$, $\hat{\Theta}_{a2}^{\mathrm{T}}(t) \in R^{q_2}$ 分别表示基于 FLSs-RL 的识别权重、梯度、识别器、执行器与评判器. 其中模糊自适应识别器、执行器和评判器设计如下:

$$\dot{\hat{\Theta}}_{\psi_u} = \Gamma_{\psi_u}(\omega_{\psi_u}(\bar{\chi}_4, z_4) - \gamma_{\psi_u}\hat{\Theta}_{\psi_u}), \tag{58}$$

$$\dot{\hat{\Theta}}_{c4} = -\gamma_{c4}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\hat{\Theta}_{c4}, \tag{59}$$

$$\dot{\hat{\Theta}}_{a4} = -\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4) \times (\gamma_{a4}(\hat{\Theta}_{a4} - \hat{\Theta}_{c4}) + \gamma_{c4}\hat{\Theta}_{c4}), \tag{60}$$

参数 Γ_4 为正定矩阵和 $\gamma_{\psi_4} > 0, \gamma_{a4} > 0, \gamma_{c4} > 0$ 分别为识别器参数、评判参数和执行参数且满足 $\beta_4 > \frac{7}{4}, \gamma_{a4} > \frac{1}{2}, \gamma_{a4} > \gamma_{c4} > \frac{\gamma_{a4}}{2}$.

构造 Lyapunov 函数如下:

$$V_4 = \sum_{j=1}^{3} V_j + \frac{1}{2}z_4^2 + \frac{1}{2}\tilde{\Theta}_{\psi_4}^{\mathrm{T}}\Gamma_4^{-1}\tilde{\Theta}_{\psi_4} + \frac{1}{2}\tilde{\Theta}_{a4}^{\mathrm{T}}\tilde{\Theta}_{a4} + \frac{1}{2}\tilde{\Theta}_{c4}^{\mathrm{T}}\tilde{\Theta}_{c4}, \tag{61}$$

其中 $\tilde{\Theta}_{\psi_4}^{\mathrm{T}} = \hat{\Theta}_{\psi_4} - \Theta_{\psi_4}^*, \tilde{\Theta}_{c4} = \hat{\Theta}_{c4} - \Theta_{J4}^*, \tilde{\Theta}_{a4} = \hat{\Theta}_{a4} - \Theta_{J4}^*$, 分别表示模糊权重误差, 执行器和评判器误差. 对 Lyapunov 函数 $V_4(t)$ 求导, 结合式 (58)∼(60), 得到 Lyapunov 函数如下:

$$\begin{aligned}\dot{V}_4 = &\sum_{j=1}^{3}\dot{V}_j + z_4\left(-\beta_4 z_4 - \frac{1}{2}\hat{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4) + \varepsilon_{\psi_4}(\bar{\chi}_4) - \dot{\hat{\alpha}}_3^*\right)\\ &+ \tilde{\Theta}_{\psi_4}^{\mathrm{T}}[\omega_{\psi_4}(\bar{\chi}_4, z_4) - \gamma_{\psi_4}\hat{\Theta}_{\psi_4}] - \tilde{\Theta}_{c4}^{\mathrm{T}}[-\gamma_{c4}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)(\bar{\chi}_4, z_4)\hat{\Theta}_{c4}]\\ &+ \tilde{\Theta}_{a4}^{\mathrm{T}}[-\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4) \times (\gamma_{a4}(\hat{\Theta}_{a4} - \hat{\Theta}_{c4})) + \gamma_{c4}\hat{\Theta}_{c4}].\end{aligned} \tag{62}$$

结合 Young's 不等式, 有 $z_4\varepsilon_{\psi_4} \leqslant \frac{1}{2}z_4^2 + \frac{1}{2}\varepsilon_{\psi_4}^2, -z_4\dot{\hat{\alpha}}_3^* \leqslant \frac{1}{2}z_4^2 + \frac{1}{2}\dot{\hat{\alpha}}_3^{*2}$,

$$-\frac{1}{2}z_4\hat{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4) \leqslant \frac{1}{4}z_4^2 + \frac{1}{4}\hat{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}(\bar{\chi}_4, z_4)^{\mathrm{T}}\hat{\Theta}_{a4}, \tag{63}$$

$$-\gamma_{\psi_4}\tilde{\Theta}_{\psi_4}^{\mathrm{T}}\hat{\Theta}_{\psi_4} \leqslant \frac{\gamma_{\psi_4}}{2}\tilde{\Theta}_{\psi_4}^{\mathrm{T}}\tilde{\Theta}_{\psi_4} + \frac{\gamma_{\psi_4}}{2}\hat{\Theta}_{\psi_4}^{\mathrm{T}}\hat{\Theta}_{\psi_4} - \frac{\gamma_{\psi_4}}{2}(\Theta_{\psi_4}^{*\mathrm{T}}\Theta_{\psi_4}^*), \tag{64}$$

$$\begin{aligned}\gamma_{c4}\tilde{\Theta}_{c4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\hat{\Theta}_{c4} = &\frac{\gamma_{c4}}{2}\tilde{\Theta}_{c4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\tilde{\Theta}_{c4}\\ &+ \frac{\gamma_{c4}}{2}\hat{\Theta}_{c4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\hat{\Theta}_{c4} - \frac{\gamma_{c4}}{2}(\Theta_{J4}^{*\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4))^2,\end{aligned} \tag{65}$$

$$\begin{aligned}\gamma_{a4}\tilde{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\hat{\Theta}_{a4} = &\frac{\gamma_{a4}}{2}\tilde{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\tilde{\Theta}_{a4}\\ &+ \frac{\gamma_{a4}}{2}\hat{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\hat{\Theta}_{a4} - \frac{\gamma_{a4}}{2}(\Theta_{J4}^{*\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4))^2,\end{aligned} \tag{66}$$

$$\begin{aligned}&(\gamma_{a4} - \gamma_{c4})\tilde{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\hat{\Theta}_{c4}\\ &\leqslant \frac{\gamma_{a4} - \gamma_{c4}}{2}(\tilde{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\tilde{\Theta}_{a4} + \frac{\gamma_{a4} - \gamma_{c4}}{2}\hat{\Theta}_{c4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\hat{\Theta}_{c4}).\end{aligned} \tag{67}$$

将这些不等式 (63)∼(67) 带入 Lyapunov 函数 (62), 得到

$$\begin{aligned}\dot{V}_4 \leqslant &\sum_{j=1}^{3}(\dot{V}_j + D_j) + \left(-\beta_4 + \frac{9}{4}\right)z_4^2 - \frac{\gamma_4}{2\lambda_{\Gamma_4^{-1}}^{\max}}\tilde{\Theta}_{\psi_u}^{\mathrm{T}}\tilde{\Theta}_{\psi_u} + D_4\\ &- \frac{\gamma_{c4}}{2}\lambda_{\omega_{J4}}^{\min}\tilde{\Theta}_{c4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\tilde{\Theta}_{c4}\\ &- \frac{\gamma_{c4}}{2}\lambda_{\omega_{J4}}^{\min}\tilde{\Theta}_{a4}^{\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}^{\mathrm{T}}(\bar{\chi}_4, z_4)\tilde{\Theta}_{a4},\end{aligned} \tag{68}$$

其中 $D_4 = \frac{1}{2}\varepsilon_{\psi_u}^2 + \frac{1}{2}\dot{\hat{\alpha}}_3^{*2} + \frac{\gamma_4}{2}\Theta_{\psi_u}^{*\mathrm{T}}\Theta_{\psi_u}^* + \frac{(\gamma_{a4}+\gamma_{c4})}{2}(\Theta_{J4}^{*\mathrm{T}}\omega_{J4}(\bar{\chi}_4, z_4))^2$ 是有界的且存在常数 d_4 满足 $D_4 \leqslant d_4$.

规定 $\lambda_{\Gamma_4^{-1}}^{\max}$ 和 $\lambda_{\omega_{J4}}^{\min}$ 分别为 $\Gamma_4^{-1}, \omega_{J4}(\bar{\chi}_4, z_4)\omega_{J4}(\bar{\chi}_4, z_4)^{\mathrm{T}}$ 的最大特征值和最小特征值, 基于设计参数 $\beta_4 > \frac{9}{4}, \gamma_{a4} > \frac{1}{2}$, $\gamma_{a4} > \gamma_{c4} > \frac{\gamma_{a4}}{2}$. 定义 $c_4 = \min\{2(\beta_4 - \frac{9}{4}), \frac{\gamma_4}{\lambda_{\Gamma_4^{-1}}^{\max}}, \gamma_{c4}\lambda_{\omega_{J4}}^{\min}\}$, 得到

$$\dot{V}_4(t) \leqslant \sum_{j=1}^{4}(-c_j(t)V_j(t) + d_j(t)). \tag{69}$$

注释 4. 通常情况下, 最优控制的 RL 通过同时训练评价器和执行器进行更新迭代, 控制器设计较为复杂, 在已有最优控制 [21~23] 方案中, 主要基于贝尔曼 (Bellman) 残差误差的平方来设计, 但由于该方程的复杂非线性特性, 不可避免地增加了控制器设计的复杂性. 本文结合文献 [24] 中基于 RL 的 actor-critic 学习策略, 采用简单的非负函数设计, 等同于 HJB 方程来减少控制器设计的复杂性.

4　仿真实验

本节通过仿真来说明基于 RL 优化模糊 actor-critic 跟踪控制技术对柔性关节机器人进行跟踪控制分析的有效性, 其动态模型具体描述为

$$\begin{cases} \dot{\chi}_1 = \chi_2, \\ \dot{\chi}_2 = \dfrac{K}{ML^2}(\chi_3 - \chi_1) + \dfrac{g}{L}\sin\chi_1 - \dfrac{1}{ML^2}(\cos\chi_4 - \sin\chi_2), \\ \dot{\chi}_3 = \chi_4, \\ \dot{\chi}_4 = \dfrac{1}{J}u(\xi(t)) - \dfrac{K}{J}(\chi_3 - \chi_1) + B\chi_4 + \chi_1\sin\chi_4, \end{cases} \tag{70}$$

系统参数为 $K = 0.5$ N/A, $L = 5$ m, $g = 9.8$ m/s^2, $J = 1$ kg·m^2, $B = 1$ N·m·s/rad, 模型的跟踪信号选取为 $y_r = 4\sin(0.5t)$, 初始参数选取为 $\chi_1(0) = \chi_2(0) = \chi_3(0) = \chi_4(0) = 0.5$.

首先选取设计参数: $\beta_1 = 3.4$, $\gamma_{a_1} = 3.4, \gamma_{c_1} = 3.5$, $\Gamma_2 = 4.5$, $\gamma_{a2} = 2.3, \gamma_{c2} = 2.4$, $\gamma_{a3} = 3, \gamma_{c3} = 2.4$, $J = 1, \beta_4 = 4, \gamma_{\psi_u} = 4.3, \gamma_{a_4} = 4.5, \gamma_{c_4} = 5, \Gamma_4 = 6, \gamma_{a4} = 8.3, \gamma_{c4} = 4.4$. 最优虚拟控制器和最优实际控制器, 基于强化学习的模糊识别器、评判器、执行训练器的设计和选取如第 3 节所示.

其次选取相关模糊向量参数及其初值 $\Theta_{a1} = \Theta_{c1} = [0.1, \ldots, 0.1]^{\mathrm{T}} \in R^{6\times 1}$, 模糊规则数量为 6, 模糊基函数选取为 $\omega_{J1}(z_1) = [\frac{\mu_{F^1}(z_1)}{\sum_{l=1}^{6}\mu_{F^l}(z_1)}, \ldots, \frac{\mu_{F^6}(z_1)}{\sum_{l=1}^{6}\mu_{F^l}(z_1)}]$, 模糊隶属度函数选取为 $\mu_{F^l}(z_1) = \exp[-(z_1 + \mu_l)^2/\eta_l^2]$, 其中心 $\mu_l = 2l - 9$ 在区间 $[-6, 6]$ 上均匀分布且模糊隶属度函数宽度为 $\eta_l = 4, l = 1, 2, \ldots, 6$. $\Theta_\psi = [0.5, \ldots, 0.5]^{\mathrm{T}} \in R^{8\times 1}$, FLSs 参数向量初值选取为 $\Theta_{a_2} = \Theta_{c_2} = [0.5, \ldots, 0.5]^{\mathrm{T}} \in R^{8\times 1}$, 模糊规则数量为 8, 选取模糊基函数为 $\omega_\psi(\bar{\chi}_4) = [\frac{\mu_{F^1}(\chi_2)\mu_{F^1}(\chi_4)}{\sum_{l=1}^{8}\mu_{F^l}(\chi_2)\mu_{F^l}(\chi_4)}, \ldots, \frac{\mu_{F^6}(\chi_2)\mu_{A^6}(\chi_4)}{\sum_{l=1}^{8}\mu_{F^l}(\chi_2)\mu_{F^l}(\chi_4)}]$, $\omega_{J2}(\bar{\chi}_4) = [\frac{\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}{\sum_{l=1}^{8}\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}, \ldots, \frac{\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}{\sum_{l=1}^{8}\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}]$. 对于步骤 3, 模糊隶属度函数选取为 $\mu_{F_j^l}(\chi_j) = \exp[-(\chi_j + \mu_l)^2/\eta_l^2], l = 1, 2, \ldots, 8, j = 1, 2, 3, 4$, 中心 $\mu_l = 2l - 11$ 在区间 $[-8, 8]$ 上均匀分布且模糊隶属度函数宽度选取 $\eta_l = 4$. FLSs 参数向量初值选取为 $\Theta_{a3} = \Theta_{c3} = [0.6, \ldots, 0.6]^{\mathrm{T}}$, 模糊规则数量为 6, 且模糊基函数选取为 $\omega_{J3}(z_3) = [\frac{\mu_{F^1}(z_3)}{\sum_{l=1}^{8}\mu_{F^l}(z_3)}, \ldots, \frac{\mu_{F^6}(z_3)}{\sum_{l=1}^{8}\mu_{F^l}(z_3)}]$. 模糊隶属度函数选取为 $\mu_{A^l}(z_3) = \exp[-(z_3 + \mu_l)^2/\eta_l^2], l = 1, 2, \ldots, 10$, 其中心 $\mu_l = 2l - 13$ 在区间上 $[-10, 10]$ 均匀分布且模糊隶属度函数的宽度为 $\eta_l = 4$.

对于步骤 4, 模糊参数向量初值 $\Theta_{\psi_u} = [0.1, \ldots, 0.1]^{\mathrm{T}} \in R^{6\times 1}$, $\Theta_{a4} = \Theta_{c4} = [0.1, \ldots, 0.1]^{\mathrm{T}} \in R^{6\times 1}$, 模糊规则数量为 6, 模糊基函数选取为 $\omega_{\psi_u}(\bar{\chi}_4) = [\frac{\mu_{F^1}(\chi_1)\mu_{F^1}(\chi_4)}{\sum_{i=1}^{6}\mu_{F^l}(\chi_1)\mu_{F^l}(\chi_4)}, \ldots, \frac{\mu_{F^6}(\chi_1)\mu_{F^6}(\chi_4)}{\sum_{i=1}^{6}\mu_{F^l}(\chi_1)\mu_{F^l}(\chi_4)}]$, $\omega_{J4}(\bar{\chi}_4) = [\frac{\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}{\sum_{i=1}^{6}\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}, \ldots, \frac{\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}{\sum_{i=1}^{6}\prod_{j=1}^{4}\mu_{F_j^l}(\chi_j)}]$. 模糊隶属度函数为 $\mu_{F_j^l}(\chi_j) = \exp[-(\chi_j + \mu_l)^2/\eta_l^2]$, $l = 1, 2, \ldots, 6, j = 1, 2, 3, 4$, 其中心 $\mu_l = 2l - 13$ 在区间上 $[-8, 8]$ 均匀分布且模糊隶属度函数的宽度为 $\eta_l = 4$.

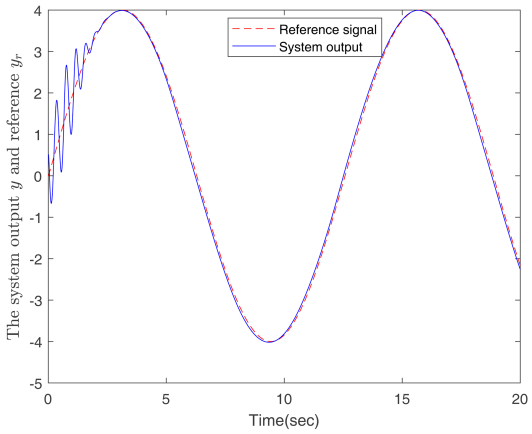

图 1　(网络版彩图) 系统输出 y 和参考信号 y_r 的轨迹.

Figure 1　(Color online) Trajectories of output y and reference signal y_r.

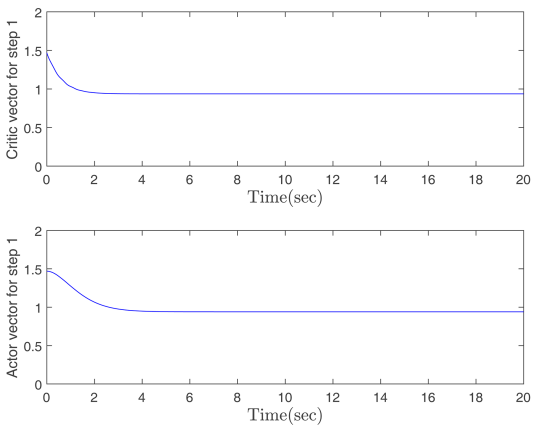

图 2　(网络版彩图) 步骤 1 中的 actor-critic 模糊自适应调节参数.

Figure 2　(Color online) Actor-critic FLS parameters for the 1st step.

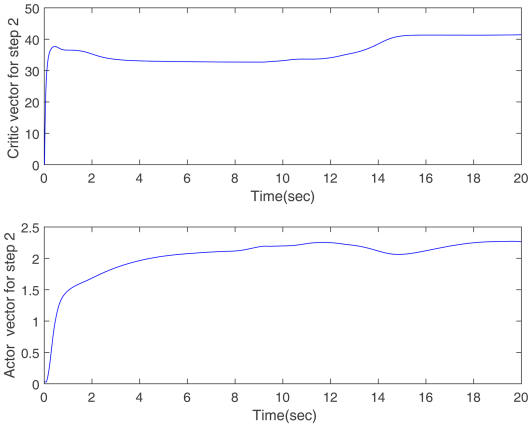

图 3　(网络版彩图) 步骤 2 中的 actor-critic 模糊自适应调节参数.

Figure 3　(Color online) Actor-critic FLS parameters for the 2nd step.

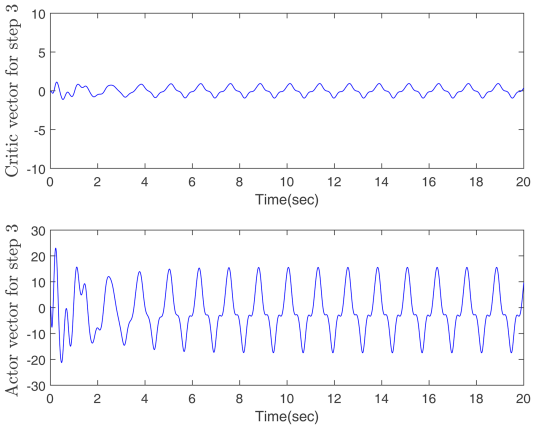

图 4　(网络版彩图) 步骤 3 中的 actor-critic 模糊自适应调节参数.

Figure 4　(Color online) Actor-critic FLS parameters for the 3rd step.

图 1 显示了系统输出和系统参考信号的轨迹的稳定性, 图 2∼5 为系统在每一步迭代更新中模糊执行评判参数图示. 图 6 显示了模糊向量识别器参数的有界性. 成本函数 $r_1(z_1(\tau),\alpha_1), r_2(z_2(\tau),\alpha_2)$, $r_3(z_3(\tau),\alpha_3), r_4(z_4(\tau),u)$ 的轨迹在图 7 和 8 中给出. 基于图 1∼8 可以得到基于强化学习的 OB 最优控制策略用于柔性关节机器人系统的稳定性控制问题的有效性. 与文献 [13∼17, 21∼23] 相比, 基于模糊识别器的 actor-critic 结构实现了柔性关节机器人系统的模糊优化控制问题, 设计过程更简化.

5　结论

本文研究了具有饱和输入的柔性关节机器人非严格反馈系统的模糊跟踪强化学习优化跟踪控制问题, 基于 actor-critic 设计策略实现优化控制, 采用模糊逻辑系统为非线性系统进行建模, 辅助自适应系统用于处理输入饱和问题, 基于 RL 的 OB 技术通过采用执行评判器迭代更新结构设计最优虚拟和实际控制器, 采用非负函数的负梯度下降法, 而非 BRS 平方法, 其中评判器用于评估控制行为并将

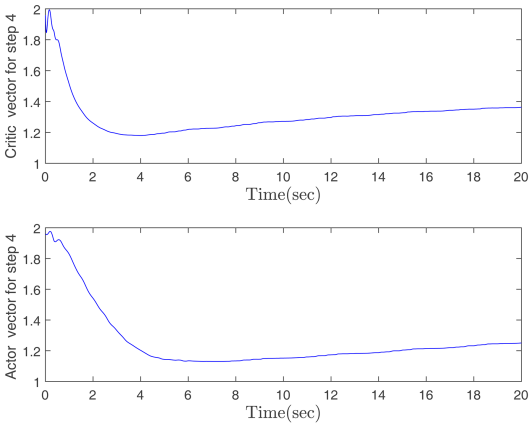

图 5　(网络版彩图) 步骤 4 中的 actor-critic 模糊自适应调节参数.

Figure 5　(Color online) Actor-critic FLS parameters for the 4th step.

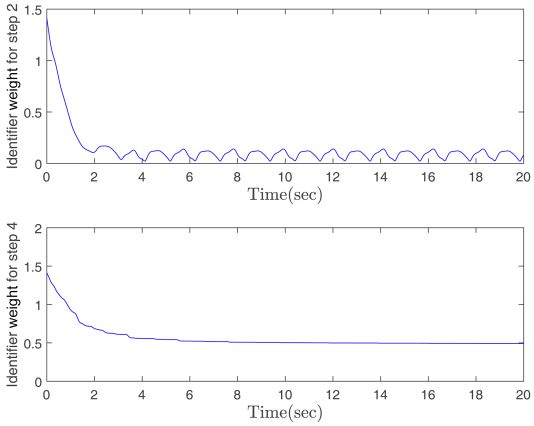

图 6　(网络版彩图) 步骤 2 和 4 中的模糊自适应识别器调节参数.

Figure 6　(Color online) Identifier of FLS parameters for the 2nd, 4th steps.

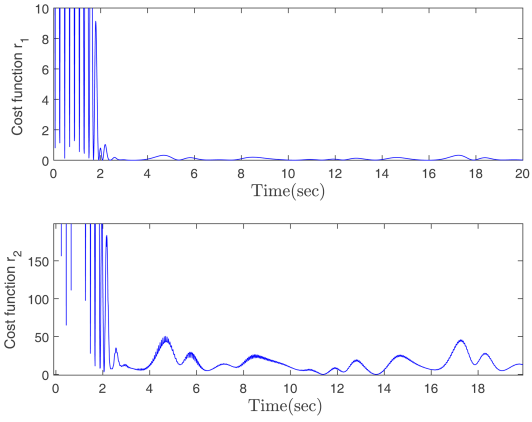

图 7　(网络版彩图) 成本函数 r_1, r_2 的轨迹.

Figure 7　(Color online) Trajectories of cost functions r_1, r_2.

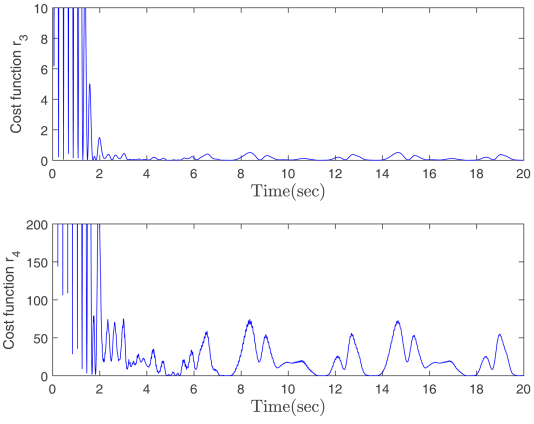

图 8　(网络版彩图) 成本函数 r_3, r_4 的轨迹.

Figure 8　(Color online) Trajectories of cost functions r_3, r_4.

评估反馈给执行器, 最后基于 Lyapunov 理论证明了整个系统中的信号是 SGUUB, 仿真示例说明了所提出的基于 RL 的 OB 模糊跟踪控制策略的有效性和可行性. 下一步将继续探讨基于强化学习的非线性切换系统的状态反馈和输出反馈跟踪控制问题.

参考文献

1　Liu Z G, Wu Y Q. Modelling and adaptive tracking control for flexible joint robots with random noises. Int J Control, 2014, 87: 2499–2510

2　Cheng C, Shen H. Fractional-order dynamics and adaptive dynamic surface control of flexible-joint robots. Asian J Control, 2023, 25: 3029–3044

3　Pan Y, Wang H, Li X, et al. Adaptive command-filtered backstepping control of robot arms with compliant actuators. IEEE Trans Contr Syst Technol, 2018, 26: 1149–1156

4　Zhu Y, Liu J, Yu J, et al. Command filtering-based adaptive fuzzy control of flexible-joint robots with time-varying full-state constraints. IEEE Trans Circuits Syst II, 2024, 71: 682–686

5　Ling S, Wang H, Liu P X. Adaptive fuzzy tracking control of flexible-joint robots based on command filtering. IEEE Trans Ind Electron, 2020, 67: 4046–4055

6 Yu X, Liu S, Zhang S, et al. Adaptive neural network force tracking control of flexible joint robot with an uncertain environment. IEEE Trans Ind Electron, 2024, 71: 5941–5949
7 Guan H, Sui S, Sui Y, et al. NN-based adaptive event-triggered predefined time control of flexible joint robot with full-state error constraints. Neurocomputing, 2025, 631: 129658
8 Zhang Z, Hu X, Huang P. Disturbance observer-based tracking controller for n-link flexible-joint robots subject to time-varying state constraints. Electronics, 2024, 13: 1773–1779
9 Ren X, Li Z, Zhou M, et al. Human intention-aware motion planning and adaptive fuzzy control for a collaborative robot with flexible joints. IEEE Trans Fuzzy Syst, 2023, 31: 2375–2388
10 Diao S, Sun W, Su S, et al. Adaptive fuzzy event-triggered control for single-link flexible-joint robots with actuator failures. IEEE Trans Cybern, 2022, 52: 7231–7241
11 Xie Y, Ma Q, Gu J, et al. Event-triggered fixed-time practical tracking control for flexible-joint robot. IEEE Trans Fuzzy Syst, 2023, 31: 67–76
12 Moyrón J, Moreno-Valenzuela J, Sandoval J. Nonlinear PI"D"-type control of flexible joint robots by using motor position measurements is globally asymptotically stable. IEEE Trans Automat Contr, 2023, 68: 3648–3655
13 Li Z, Zhao J. Adaptive consensus of non-strict feedback switched multi-agent systems with input saturations. IEEE CAA J Autom Sin, 2021, 8: 1752–1761
14 Liu L, Yao W, Guo Y. Prescribed performance tracking control of a free-flying flexible-joint space robot with disturbances under input saturation. J Franklin Institute, 2021, 358: 4571–4601
15 Ding S, Peng J, Zhang H, et al. Neural network-based adaptive hybrid impedance control for electrically driven flexible-joint robotic manipulators with input saturation. Neurocomputing, 2021, 458: 99–111
16 Huang P, Zhang F, Lu Y, et al. Adaptive neural network dynamic surface control of the post-capture tethered system with full state constraints. Atti Take Control Fail Spacec, 2025, 12: 219–247
17 Xu X, Xu S. Event-triggered adaptive neural tracking control of flexible-joint robot systems with input saturation. IEEE Access, 2022, 10: 43367–43375
18 Bellman R. E. Dynamic Programming. Princeton: Princeton University Press, 1957
19 Pontryagin L S, Boltyanskii V G, Gamkrelidze R V, et al. The Mathematical Theory of Optimal Processes. Geneva: Interscience Publishers, 1962
20 Lewis F L, Vrabie D, Syrmos V L. Optimal Control. 3rd ed. Hoboken: Wiley, 2012
21 Yang X, Liu D, Wang D. Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints. Int J Control, 2013, 87: 553–566
22 Bhasin S, Kamalapurkar R, Johnson M, et al. A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. Automatica, 2013, 49: 82–92
23 Modares H, Lewis F L, Naghibi-Sistani M B. Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems. Automatica, 2014, 50: 193–202
24 Wen G, Ge S, Tu F. Optimized backstepping for tracking control of strict-feedback systems. IEEE Trans Neural Netw Learn Syst, 2018, 29: 3850–3862
25 Wen G, Ge S, Chen C L P, et al. Adaptive tracking control of surface vessel using optimized backstepping technique. IEEE Trans Cybern, 2019, 49: 3420–3431
26 Wen G, Chen C L P, Ge S. Simplified optimized backstepping control for a class of nonlinear strict-feedback systems with unknown dynamic functions. IEEE Trans Cybern, 2021, 51: 4567–4580
27 Wen G, Hao W, Feng W, et al. Optimized backstepping tracking control using reinforcement learning for quadrotor unmanned aerial vehicle system. IEEE Trans Syst Man Cybern Syst, 2022, 52: 5004–5015
28 Qin C, Wu Y, Zhu T, et al. Reinforcement-learning-based decentralized event-triggered control of partially unknown nonlinear interconnected systems with state constraints. Appl Intell, 2025, 55: 164
29 Li D, Dong J. Fuzzy control based on reinforcement learning and subsystem error derivatives for strict-feedback systems with an observer. IEEE Trans Fuzzy Syst, 2023, 31: 2509–2521
30 Wen G, Xu L, Li B. Optimized backstepping tracking control using reinforcement learning for a class of stochastic nonlinear strict-feedback systems. IEEE Trans Neural Netw Learn Syst, 2023, 34: 1291–1303
31 Wen G, Chen C L P. Optimized backstepping consensus control using reinforcement learning for a class of nonlinear strict-feedback-dynamic multi-agent systems. IEEE Trans Neural Netw Learn Syst, 2023, 34: 1524–1536
32 Guo Y, Sun Q, Pan Q, et al. Pareto-optimal synchronization control of nonlinear multi-agent systems via integral reinforcement learning. Nonlinear Dyn, 2025, 113: 5339–5357
33 Bai W, Li T, Tong S. NN reinforcement learning adaptive control for a class of nonstrict-feedback discrete-time systems. IEEE Trans Cybern, 2020, 50: 4573–4584

34 Wu J, Lu H, Wang W. Adaptive deep neural network optimized backstepping control for a class of nonlinear strict-feedback systems. In: Proceedings of the 12th International Conference on Intelligent Control and Information Processing (ICICIP), Nanjing, 2024. 172–181
35 Wang L X. Adaptive Fuzzy Systems and Control: Design and Stability Analysis. Englewood Cliffs: Prentice-Hall, Inc., 1994. 89–98

Reinforcement learning-based optimized backstepping fuzzy tracking control for flexible-joint robot systems

Rui WANG[1*], Guoxing WEN[2], Yanjun LIU[3], Fusheng YU[4] & Jian WU[1]

1. *School of Applied Mathematics, Shanxi University of Finance and Economics, Taiyuan 030006, China*
2. *School of Mathematical Science, Shandong University of Aeronautics, Binzhou 256600, China*
3. *School of Electrical Engineering, Liaoning University of Technology, Jinzhou 121001, China*
4. *School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China*
* Corresponding author. E-mail: rui-wang@live.com

Abstract This article focuses on a fuzzy optimized backstepping (OB) practical tracking control algorithm for flexible-joint robot systems (FJRS) based on reinforcement learning (RL) scheme. Since FJRS can be described by fourth-order dynamic systems in non-strict feedback form, uncertainties of this form are approximated by fuzzy logic systems, an auxiliary system is established to handle the input saturation, and we apply the identifier-actor-critic technique based on a simplified RL algorithm to design optimal controllers, in addition, designing the negative gradient descent (GD) for positive function, rather than adopting the GD of the Bellman residual error' square. Semi-global uniformly ultimately boundedness (SGUUB) of the whole system is guaranteed through Lyapunov stability analysis. A simulation example demonstrated the effectiveness of the presentation RL-based OB fuzzy tracking control strategy.

Keywords optimized backstepping (OB), fuzzy logic systems (FLSs), reinforcement learning (RL), non-strict-feedback form systems (NSFFS), actor-critic actuator, flexible-joint robot sysems (FJRS)