

引用:李熙,孙伟,殷县龙.单连杆柔性关节机械臂系统的优化反步跟踪控制[J].聊城大学学报(自然科学版),2026,39(2): 183-191

Citation: LI Xi, SUN Wei, YIN Xianlong. Optimized backstepping tracking control for a single-link flexible joint robotic arm system[J]. Journal of Liaocheng University (Natural Science Edition), 2026, 39(2): 183-191.

文章编号 1672-6634(2026)02-0183-09

DOI 10.19728/j.issn1672-6634.2025070001

单连杆柔性关节机械臂系统的优化反步跟踪控制

李熙,孙伟,殷县龙

(聊城大学 数学科学学院,山东 聊城 252059)

摘要 针对单连杆柔性关节机械臂系统,提出了一种基于强化学习策略的优化反步跟踪控制方法。首先,在每个反步步骤中利用 Actor-critic 结构来执行强化学习策略,进而得到每个子系统的最优虚拟控制器,最终得到系统的最优控制器。然后基于设定的李雅普诺夫函数,推导出单连杆柔性关节机械臂系统的所有信号是有界的,证明了所设计控制方法的稳定性。最后,仿真实验验证了基于强化学习策略的优化反步跟踪控制在单连杆柔性关节机械臂系统中的有效性。

关键词 强化学习;单连杆柔性关节机械臂;稳定性分析

中图分类号 TP241

文献标志码 A

开放科学(资源服务)标识码(OSID)



Optimized backstepping tracking control for a single-link flexible joint robotic arm system

LI Xi, SUN Wei, YIN Xianlong

(School of Mathematical Sciences, Liaocheng University, Liaocheng 252059, China)

Abstract In this paper, an optimized backstepping tracking control method based on a reinforcement learning strategy is proposed for a single-link flexible-joint robotic arm system. Firstly, the actor-critic architecture is employed at each backstepping step to implement the reinforcement learning strategy, thereby obtaining the optimal virtual controller for each subsystem and ultimately deriving the system's optimal controller. Then, based on the defined Lyapunov function, it is demonstrated that all signals of the single-link flexible-joint robotic arm system are bounded, proving the stability of the designed control method. Finally, simulation experiments verify the effectiveness of the optimized backstepping tracking control based on reinforcement learning strategy in a single-link flexible-joint robotic arm system.

Keywords reinforcement learning; single-link flexible-joint robotic arm; stability analysis

0 引言

强化学习算法最初由 Werbos^[1]针对离散系统提出,随后被扩展到连续系统中,如文献[2-6]所述。在强化学习框架下,智能体能够通过与环境的动态交互学习最优策略,从而达成长期收益最大化的目标^[7]。具体的学习机制包含:初始阶段智能体通过环境交互获取知识;之后对学习成效进行评分并分配奖励;继而基于

收稿日期:2025-07-01

基金项目:国家自然科学基金重点项目(62473185),山东省自然科学基金优秀青年项目(ZR2024YQ033)资助

通信作者:孙伟,男,汉族,博士,教授,泰山学者青年专家,研究方向:非线性系统控制、智能算法,E-mail:sunw8617@163.com。

奖励优劣制定后续决策;经过充分训练后,目标对象将建立优化的行为策略模型,实现目标最终想要的预期收益^[8]。其基本思想是根据来自环境的评价生成所需的控制动作。通常,强化学习通过 Actor-critic 结构的在线相互迭代来实现,其中 Critic 函数用于评估控制性能并将评估结果反馈给 Actor 函数,而 Actor 函数则用于执行控制动作^[9]。强化学习无需依赖监督信号即可实现收敛,这一特性使其特别适用于非结构化环境下的机械臂控制。目前,强化学习在机器人控制领域的主要应用集中于避障与路径规划,同时也有少量研究探索其在振动抑制方面的潜力^[10]。随着强化学习策略的深入发展,研究者发现其在非线性系统的优化控制方面表现优异,因此越来越多的学者采用强化学习策略来提升柔性关节机械臂系统的控制性能。

柔性关节机械臂是工业领域的研究热点之一,其核心构成部分包括臂端、关节系统、传感器、控制系统及末端执行器。关节系统用于实现柔性臂段的运动和控制,它由电机、液压缸或气动驱动系统构成;传感器包括位置传感器和力传感器,主要用于监测机械臂的环境条件和状态,控制系统主要负责处理传感器反馈并及时调整关节系统,通过这些来实现机械臂所需的运动和任务;末端执行器位于机械臂的末端,主要用于执行抓取、操纵或操作物体等具体的任务。柔性关节机械臂的各组成部分协同工作,使其能够适应复杂多变的任务和工作环境^[11]。相较于传统刚性机械臂,柔性关节机械臂具有更高的灵活性、安全性和能效,使其在工业制造、医疗康复和航空航天等领域得到广泛应用。然而,该系统不仅保留了刚性机械臂的非线性、不确定性和强耦合特性,还因柔性结构的引入带来了额外变量,导致系统复杂度显著增加。因此,如何为其设计高效的优化控制策略至关重要^[12]。

在许多实际系统中,控制器设计存在误差是无法避免的,柔性关节机械臂系统同样如此^[13]。基于强化学习策略设计控制柔性关节机械臂系统可以通过 Actor-critic 结构的迭代来让每个子系统的控制达到最优,从而使整个系统的控制达到最优。因此可以有效优化控制器,减小控制器误差,实现跟踪控制。本文聚焦于单连杆柔性关节机械臂系统的控制优化问题,重点探讨强化学习策略在反步控制方法中的应用价值。本文的理论创新和实践意义主要体现在以下三个方面:(1) 与传统反步法相比,本文利用基于强化学习策略的优化反步跟踪控制方法来设计控制单连杆柔性关节机械臂系统,可以得到系统的最优控制器,实现跟踪期望目标,控制效果更好。(2) 利用构造的李雅普诺夫函数,使单连杆柔性关节机械臂系统的所有信号均为半全局最终一致有界,实现预期的控制性能。(3) 通过仿真实验,在单连杆柔性关节机械臂系统上验证了优化反步跟踪控制方法的有效性,对单连杆柔性关节机械臂的控制设计研究具有重要意义。

符号表示 用 $\mathbf{R}$ 表示实数集, $\mathbf{R}^n$ 表示 n 维欧几里德空间。

1 系统建模、引理与定义

本文基于欧拉-拉格朗日方程,采用如下动力学模型对单连杆柔性关节机械臂系统进行建模描述^[14-15]

$$M(q_1)\ddot{q}_1 + C(q_1, \dot{q}_1) + G(q_1) + F(\dot{q}_1) + Kq_1 = Kq_2, J\ddot{q}_2 + B\dot{q}_2 + K(q_2 - q_1) = u, \quad (1)$$

式中 $M(q_1) \in \mathbf{R}$ 表示连杆惯性; $C(q_1, \dot{q}_1) \in \mathbf{R}$ 表示向心力; $G(q_1) \in \mathbf{R}$ 表示重力; $F(\dot{q}_1) \in \mathbf{R}$ 表示摩擦力; $K \in \mathbf{R}$ 表示关节弹性; $J \in \mathbf{R}$ 表示执行器惯性; $B \in \mathbf{R}$ 表示自然阻尼项; $u \in \mathbf{R}$ 表示控制输入; $q_1 \in \mathbf{R}$ 表示连杆位置, $\dot{q}_1 \in \mathbf{R}$ 表示速度, $\ddot{q}_1 \in \mathbf{R}$ 表示加速度; $q_2 \in \mathbf{R}$ 表示转子角位置, $\dot{q}_2 \in \mathbf{R}$ 表示速度, $\ddot{q}_2 \in \mathbf{R}$ 表示加速度。

定义 $x_1 = q_1, x_2 = \dot{q}_1, x_3 = q_2$ 和 $x_4 = \dot{q}_2$, 然后将动态模型(1)转换为以下单输入单输出的非线性系统

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = M^{-1}(x_1)Kx_3 + M^{-1}(x_1)(-C(x_1, x_2) - G(x_1) - F(x_2) - Kx_1), \\ \dot{x}_3 = x_4, \\ \dot{x}_4 = J^{-1}u + J^{-1}(-Bx_4 - K(x_3 - x_1)), \end{cases} \quad (2)$$

式中 $x_i \in \mathbf{R}, i=1, 2, 3, 4$ 为系统状态。

本文的控制目标是基于强化学习策略针对单连杆柔性关节机械臂系统设计优化跟踪控制方案,使得机械臂在运动过程中跟踪上期望轨迹 y_d , 并使跟踪误差收敛至原点附近的邻域内,同时确保单连杆柔性关节机械臂系统中的所有信号均有界。为了达到控制目标,引入以下引理及定义。

神经网络 对于连续函数 $f(v): \mathbf{R} \rightarrow^n \mathbf{R}$, 可以在紧集 Ω_v 选择神经网络作为其逼近器, 具体模型为^[16] $f(v) = W^T S(v)$ 式中 $W \in \mathbf{R}^q$ 为神经网络权重 (q 为神经元数量)。 $S(v) = [s_1(v), \dots, s_q(v)] \in \mathbf{R}^q$ 为基函数向量, 式中 $s_1(v) = \exp[-(v - o_i)^T (v - o_i) / 2]$, $o_i = [o_{i1}, o_{i2}, \dots, o_{in}]^T, i = 1, 2, \dots, q$ 。

对于神经网络, 一定会存在一个理想权重 $W^* \in \mathbf{R}^p$, 其定义 $W^* = \operatorname{argmin}_{W \in \Omega_v} \|f(v) - W^T S(v)\|$, 使得函数重新描述为 $f(v) = W^{*T} S(v) + \varepsilon(v)$, 式中 $\varepsilon(v) \in \mathbf{R}$ 表示有界近似误差, 若神经元数量设置的足够大, 该误差可控制在期望范围内。

引理 1 若正函数 $\varphi(t) \in \mathbf{R}$ 满足 $\dot{\varphi}(t) \leq -a\varphi(t) + b$ (式中 $a > 0, b > 0$), 则以下不等式成立

$$\varphi(t) \leq e^{-at} \varphi(0) + \frac{b}{a} (1 - e^{-at}).$$

定义 1^[17] 对于位于紧集 Ω 内的初始值 $x(0)$, 若存在 $Z > 0$ 和 $T(x(0), Z) > 0$ 使得

$$x(t) \in \{x(t) \mid \|x(t)\| \leq Z\}, \forall t > T(x(0), Z),$$

则称微分方程 $\dot{x}(t) = f(t, x)$ 的解 $x(t)$ 具有半全局一致最终有界性。

2 主要结果

第一步: 将跟踪误差定义为 $z_1 = x_1 - y_d$ 。根据式(2), 误差动态方程为

$$\dot{z}_1 = x_2 - \dot{y}_d. \quad (3)$$

结合误差动态方程(3), 定义最优性能指标函数为

$$J_1^* = \min_{\alpha_1 \in \psi(\Omega)} \left\{ \int_t^\infty c_1(z_1, \alpha_1) ds \right\} = \int_t^\infty c_1(z_1, \alpha_1^*) ds, \quad (4)$$

式中 $c_1(z_1, \alpha_1^*) = z_1^2 + \alpha_1^{*2}$ 为成本函数, α_1^* 为该反步步骤的最优虚拟控制器。

通过将 x_2 视为最优控制 α_1^* , 与(3)和(4)相关联的 HJB 方程为

$$H_1 = z_1^2 + \alpha_1^{*2} + \frac{dJ_1^*}{dz_1} (\alpha_1^* - \dot{y}_d) = 0. \quad (5)$$

通过求解 $\frac{\partial H_1}{\partial \alpha_1^*} = 0$ 可推导出最优虚拟控制器 α_1^* 为

$$\alpha_1^* = -\frac{1}{2} \frac{dJ_1^*}{dz_1}. \quad (6)$$

为了实现跟踪控制, 将 $\frac{dJ_1^*}{dz_1}$ 分解为

$$\frac{dJ_1^*}{dz_1} = 2r_1 z_1 + J_1^0 - 2\dot{y}_d, \quad (7)$$

式中 r_1 是常数, $J_1^0 = -2r_1 z_1 + \frac{dJ_1^*}{dz_1} + 2\dot{y}_d$ 为连续函数。因为未知函数 J_1^0 是连续的, 因此 J_1^0 可以在紧集 Ω 上通过神经网络被近似为

$$J_1^0 = \theta_{j_1}^T \phi_{j_1} + \varepsilon_{j_1}, \quad (8)$$

式中 θ_{j_1} 为神经网络权重, ϕ_{j_1} 为基函数向量, ε_{j_1} 为有界的近似误差。

将公式(8)代入公式(7)可得

$$\frac{dJ_1^*}{dz_1} = 2r_1 z_1 + \theta_{j_1}^T \phi_{j_1} + \varepsilon_{j_1} - 2\dot{y}_d. \quad (9)$$

将公式(9)代入公式(6)可得

$$\alpha_1^* = -r_1 z_1 - \frac{1}{2} \theta_{j_1}^T \phi_{j_1} - \frac{1}{2} \varepsilon_{j_1} + \dot{y}_d. \quad (10)$$

因为神经网络权重 θ_{j_1} 是未知的, 因此 α_1^* 是未知的。为了得到有效的最优虚拟控制器, 设计 Actor-critic 结构来执行强化学习策略。根据公式(9), 设计评估控制性能的 Critic 为

$$\frac{d\hat{J}_1^*}{dz_1} = 2r_1 z_1 + \hat{\theta}_{c_1}^T \phi_{j_1} - 2\dot{y}_d, \quad (11)$$

式中 $\frac{d\hat{J}_1^*}{dz_1}$ 为 $\frac{dJ_1^*}{dz_1}$ 的估计值, $\hat{\theta}_{c_1}$ 为 Critic 理想权重。根据公式(10),设计可用于执行控制动作的 Actor 为

$$\hat{\alpha}_1^* = -r_1 z_1 - \frac{1}{2} \hat{\theta}_{a_1}^T \phi_{j_1} + \dot{y}_d, \quad (12)$$

式中 $\hat{\alpha}_1^*$ 是期望最优虚拟控制器, $\hat{\theta}_{a_1}$ 是 Actor 理想权重。设计理想权重 $\hat{\theta}_{c_1}$ 和 $\hat{\theta}_{a_1}$ 的权重更新率为 $\dot{\hat{\theta}}_{c_1} = -r_{c_1} \phi_{j_1} \phi_{j_1}^T \hat{\theta}_{c_1}$, $\dot{\hat{\theta}}_{a_1} = -\phi_{j_1} \phi_{j_1}^T (r_{a_1} (\hat{\theta}_{a_1} - \hat{\theta}_{c_1}) + r_{c_1} \hat{\theta}_{c_1})$, 式中 $r_{c_1} > 0$ 和 $r_{a_1} > 0$ 是常数。

将公式(11)和公式(12)代入公式(5),得到近似 HJB 方程

$$\begin{aligned} \hat{H}_1 &= z_1^2 + \hat{\alpha}_1^{*2} + \frac{d\hat{J}_1^*}{dz_1} (\hat{\alpha}_1^* - \dot{y}_d) \\ &= z_1^2 + (-r_1 z_1 - \frac{1}{2} \hat{\theta}_{a_1}^T \phi_{j_1} + \dot{y}_d)^2 + (2r_1 z_1 + \hat{\theta}_{c_1}^T \phi_{j_1} - 2\dot{y}_d) (-r_1 z_1 - \frac{1}{2} \hat{\theta}_{a_1}^T \phi_{j_1}). \end{aligned}$$

根据前面的分析,期望最优虚拟控制器 $\hat{\alpha}_1^*$ 能使近似函数 $\hat{H}_1 \rightarrow 0$ 。如果能满足 $\hat{H}_1 = 0$, 且因为 $\hat{\alpha}_1^*$ 是 HJB 方程的唯一解,则下式成立

$$\begin{aligned} \frac{\partial \hat{H}_1}{\partial \hat{\theta}_{a_1}} &= 2(-r_1 z_1 - \frac{1}{2} \hat{\theta}_{a_1}^T \phi_{j_1} + \dot{y}_d) (-\frac{1}{2} \phi_{j_1}) + (2r_1 z_1 + \hat{\theta}_{c_1}^T \phi_{j_1} - 2\dot{y}_d) (-\frac{1}{2} \phi_{j_1}) \\ &= \frac{1}{2} \phi_{j_1} \phi_{j_1}^T (\hat{\theta}_{a_1} - \hat{\theta}_{c_1}) = 0. \end{aligned} \quad (13)$$

假设 $e_1 = \hat{\theta}_{a_1} - \hat{\theta}_{c_1}$ 是误差,我们构建李雅普诺夫函数如

$$P_1(t) = \frac{1}{2} e_1^T e_1, \quad (14)$$

公式(14)关于时间的导数为

$$\frac{dP_1(t)}{dt} = e_1^T (\dot{\hat{\theta}}_{a_1} - \dot{\hat{\theta}}_{c_1}) = -r_{a_1} e_1^T \phi_{j_1} \phi_{j_1}^T e_1 \leq 0.$$

因此误差 $e_1 = \hat{\theta}_{a_1} - \hat{\theta}_{c_1}$ 可以渐近收敛到 0,这样可证得式(13)成立。考虑如下李雅普诺夫函数

$$V_1 = \frac{1}{2} z_1^2 + \frac{1}{2} \bar{\theta}_{a_1}^T \bar{\theta}_{a_1} + \frac{1}{2} \bar{\theta}_{c_1}^T \bar{\theta}_{c_1},$$

式中 $\bar{\theta}_{a_1} = \hat{\theta}_{a_1} - \theta_{j_1}$, $\bar{\theta}_{c_1} = \hat{\theta}_{c_1} - \theta_{j_1}$ 。则

$$\begin{aligned} \dot{V}_1 &= z_1 \dot{z}_1 + \bar{\theta}_{a_1}^T \dot{\hat{\theta}}_{a_1} + \bar{\theta}_{c_1}^T \dot{\hat{\theta}}_{c_1} \\ &= z_1 (x_2 - \dot{y}_d) + (r_{a_1} - r_{c_1}) \bar{\theta}_{a_1}^T \phi_{j_1} \phi_{j_1}^T \hat{\theta}_{c_1} - r_{a_1} \bar{\theta}_{a_1}^T \phi_{j_1} \phi_{j_1}^T \hat{\theta}_{a_1} - r_{c_1} \bar{\theta}_{c_1}^T \phi_{j_1} \phi_{j_1}^T \hat{\theta}_{c_1}. \end{aligned} \quad (15)$$

根据杨氏不等式以及 $\bar{\theta}_{a_i} = \hat{\theta}_{a_i} - \theta_{j_i}$, $\bar{\theta}_{c_i} = \hat{\theta}_{c_i} - \theta_{j_i}$ 可得

$$\begin{aligned} z_i z_{i+1} &\leq \frac{1}{2} z_i^2 + \frac{1}{2} z_{i+1}^2, i=1,2,3, \\ \bar{\theta}_{a_i}^T \phi_{j_i} \phi_{j_i}^T \hat{\theta}_{c_i} &\leq \frac{1}{2} (\bar{\theta}_{a_i}^T \phi_{j_i})^2 + \frac{1}{2} (\bar{\theta}_{c_i}^T \phi_{j_i})^2, i=1,2,3,4, \\ \bar{\theta}_{a_i}^T \phi_{j_i} \phi_{j_i}^T \hat{\theta}_{a_i} &\leq \frac{1}{2} (\bar{\theta}_{a_i}^T \phi_{j_i})^2 + \frac{1}{2} (\hat{\theta}_{a_i}^T \phi_{j_i})^2 - \frac{1}{2} (\theta_{j_i}^T \phi_{j_i})^2, i=1,2,3,4, \\ \bar{\theta}_{c_i}^T \phi_{j_i} \phi_{j_i}^T \hat{\theta}_{c_i} &\leq \frac{1}{2} (\bar{\theta}_{c_i}^T \phi_{j_i})^2 + \frac{1}{2} (\hat{\theta}_{c_i}^T \phi_{j_i})^2 - \frac{1}{2} (\theta_{j_i}^T \phi_{j_i})^2, i=1,2,3,4. \end{aligned} \quad (16)$$

将公式(16)代入公式(15)可得

$$\begin{aligned} \dot{V}_1 &\leq z_1 (x_2 - \dot{y}_d) + (r_{a_1} - r_{c_1}) \bar{\theta}_{a_1}^T \phi_{j_1} \phi_{j_1}^T \hat{\theta}_{c_1} - r_{a_1} \bar{\theta}_{a_1}^T \phi_{j_1} \phi_{j_1}^T \hat{\theta}_{a_1} - r_{c_1} \bar{\theta}_{c_1}^T \phi_{j_1} \phi_{j_1}^T \hat{\theta}_{c_1} \\ &\leq - (r_1 - \frac{3}{4}) z_1^2 + \frac{1}{2} z_2^2 - (\frac{1}{2} r_{a_1} - \frac{1}{4}) (\hat{\theta}_{a_1}^T \phi_{j_1})^2 - (\frac{1}{2} r_{c_1}) (\bar{\theta}_{a_1}^T \phi_{j_1})^2 - \\ &\quad (r_{c_1} - \frac{1}{2} r_{a_1}) (\hat{\theta}_{c_1}^T \phi_{j_1})^2 - (\frac{1}{2} r_{c_1}) (\bar{\theta}_{c_1}^T \phi_{j_1})^2 + (\frac{1}{2} r_{a_1} + \frac{1}{2} r_{c_1}) (\theta_{j_1}^T \phi_{j_1})^2. \end{aligned}$$

第二步:将跟踪误差定义为 $z_2 = x_2 - \hat{\alpha}_1^*$, 根据式(2),误差动态方程为

$$\dot{z}_2 = \frac{K}{M}x_3 + \frac{1}{M}(-C - G - F - Kx_1) - \dot{\hat{\alpha}}_1^*。 \quad (17)$$

结合误差动态方程(17),定义最优性能指标函数为

$$J_2^* = \min_{\alpha_2^* \in \psi(\Omega)} \left\{ \int_t^\infty c_2(z_2, \alpha_2) ds \right\} = \int_t^\infty c_2(z_2, \alpha_2^*) ds, \quad (18)$$

式中 $c_2(z_2, \alpha_2^*) = z_2^2 + \alpha_2^{*2}$ 为成本函数, α_2^* 为该反步步骤的最优虚拟控制器。

通过将 x_3 视为最优控制 α_2^* , 与(17)和(18)相关联的 HJB 方程为

$$H_2 = z_2^2 + \alpha_2^{*2} + \frac{dJ_2^*}{dz_2} \left(\frac{K}{M}\alpha_2^* + \frac{1}{M}(-C - G - F - Kx_1) - \dot{\hat{\alpha}}_1^* \right) = 0。$$

通过求解 $\frac{\partial H_2}{\partial \alpha_2^*} = 0$ 可推导出最优虚拟控制器 α_2^* 为

$$\alpha_2^* = \frac{M}{K}(-r_2 z_2 - \frac{1}{2}\theta_{j_2}^T \phi_{j_2} - \frac{1}{2}\epsilon_{j_2} + \dot{\hat{\alpha}}_1^* - \frac{1}{M}(-C - G - F - Kx_1))。 \quad (19)$$

为了得到有效的最优虚拟控制器,设计 Actor-critic 结构来执行强化学习策略。

设计评估控制性能的 Critic 为 $\frac{dJ_2^*}{dz_2} = \frac{M^2}{K^2}(2r_2 z_2 + \hat{\theta}_{c_2}^T \phi_{j_2} - 2\dot{\hat{\alpha}}_1^* + 2\frac{1}{M}(-C - G - F - Kx_1))$, 式中 $\hat{\theta}_{c_2}$

为 Critic 理想权重。根据公式(19),设计可用于执行控制动作的 Actor 为

$$\hat{\alpha}_2^* = \frac{M}{K}(-r_2 z_2 - \frac{1}{2}\hat{\theta}_{a_2}^T \phi_{j_2} + \dot{\hat{\alpha}}_1^* - \frac{1}{M}(-C - G - F - Kx_1))，$$

式中 $\hat{\alpha}_2^*$ 是期望最优虚拟控制器, $\hat{\theta}_{a_2}$ 是 Actor 理想权重。设计理想权重 $\hat{\theta}_{c_2}$ 和 $\hat{\theta}_{a_2}$ 的权重更新率为 $\dot{\hat{\theta}}_{c_2} =$

$-r_{c_2} \phi_{j_2} \phi_{j_2}^T \hat{\theta}_{c_2}, \dot{\hat{\theta}}_{a_2} = -\phi_{j_2} \phi_{j_2}^T (r_{a_2}(\hat{\theta}_{a_2} - \hat{\theta}_{c_2}) + r_{c_2} \hat{\theta}_{c_2})$, 式中 $r_{c_2} > 0, r_{a_2} > 0$ 均是常数。构建如下李雅普诺夫函数 $V_2 = V_1 + \frac{1}{2}z_2^2 + \frac{1}{2}|\hat{\theta}_{a_2}^T \tilde{\theta}_{a_2} + \frac{1}{2}\tilde{\theta}_{c_2}^T \tilde{\theta}_{c_2}|$, 式中 $\tilde{\theta}_{a_2} = \hat{\theta}_{a_2} - \theta_{j_2}, \tilde{\theta}_{c_2} = \hat{\theta}_{c_2} - \theta_{j_2}$ 。则

$$\begin{aligned} \dot{V}_2 \leq & \dot{V}_1 - (r_2 - \frac{K^2}{2M^2} - \frac{1}{4})z_2^2 + \frac{1}{2}z_3^2 - (\frac{1}{2}r_{a_2} - \frac{1}{4})(\hat{\theta}_{a_2}^T \phi_{j_2})^2 - (\frac{1}{2}r_{c_2})(\hat{\theta}_{c_2}^T \phi_{j_2})^2 - \\ & (r_{c_2} - \frac{1}{2}r_{a_2})(\hat{\theta}_{c_2}^T \phi_{j_2})^2 - (\frac{1}{2}r_{c_2})(\tilde{\theta}_{c_2}^T \phi_{j_2})^2 + (\frac{1}{2}r_{a_2} + \frac{1}{2}r_{c_2})(\theta_{j_2}^T \phi_{j_2})^2。 \end{aligned}$$

第三步:将跟踪误差定义为 $z_2 = x_3 - \hat{\alpha}_2^*$, 根据式(2),误差动态方程为

$$\dot{z}_3 = x_4 - \dot{\hat{\alpha}}_2^*。 \quad (20)$$

结合误差动态方程(20),定义最优性能指标函数为

$$J_3^* = \min_{\alpha_3^* \in \psi(\Omega)} \left\{ \int_t^\infty c_3(z_3, \alpha_3) ds \right\} = \int_t^\infty c_3(z_3, \alpha_3^*) ds, \quad (21)$$

式中 $c_3(z_3, \alpha_3^*) = z_3^2 + \alpha_3^{*2}$ 为成本函数, α_3^* 为该反步步骤的最优虚拟控制器。

通过将 x_4 视为最优控制 α_3^* , 与(20)和(21)相关联的 HJB 方程为 $H_3 = z_3^2 + \alpha_3^{*2} + \frac{dJ_3^*}{dz_3}(\alpha_3^* - \dot{\hat{\alpha}}_2^*) = 0。$

通过求解 $\frac{\partial H_3}{\partial \alpha_3^*} = 0$ 可推导出最优虚拟控制器 α_3^* 为

$$\alpha_3^* = -r_3 z_3 - \frac{1}{2}\theta_{j_3}^T \phi_{j_3} - \frac{1}{2}\epsilon_{j_3} + 2\dot{\hat{\alpha}}_2^*。 \quad (22)$$

为了得到有效的最优虚拟控制器,设计 Actor-critic 结构来执行强化学习策略。设计评估控制性能的

Critic 为 $\frac{dJ_3^*}{dz_3} = 2r_3 z_3 + \hat{\theta}_{c_3}^T \phi_{j_3} - 2\dot{\hat{\alpha}}_2^*$, 式中 $\hat{\theta}_{c_3}$ 为 Critic 理想权重。根据公式(22),设计可用于执行控制动

作的 Actor 为 $\hat{\alpha}_3^* = -r_3 z_3 - \frac{1}{2}\hat{\theta}_{a_3}^T \phi_{j_3} + \dot{\hat{\alpha}}_2^*$, 式中 $\hat{\alpha}_3^*$ 是期望最优虚拟控制器, $\hat{\theta}_{a_3}$ 是 Actor 理想权重。设计

理想权重 $\hat{\theta}_{c_3}$ 和 $\hat{\theta}_{a_3}$ 的权重更新率为 $\dot{\hat{\theta}}_{c_3} = -r_{c_3} \phi_{j_3} \phi_{j_3}^T \hat{\theta}_{c_3}$, $\dot{\hat{\theta}}_{a_3} = -\phi_{j_3} \phi_{j_3}^T (r_{a_3} (\hat{\theta}_{a_3} - \hat{\theta}_{c_3}) + r_{c_3} \hat{\theta}_{c_3})$, 式中 $r_{c_3} > 0$ 和 $r_{a_3} > 0$ 均是常数。构建如下李雅普诺夫函数 $V_3 = V_2 + \frac{1}{2} z_3^2 + \frac{1}{2} \tilde{\theta}_{a_3}^T \tilde{\theta}_{a_3} + \frac{1}{2} \tilde{\theta}_{c_3}^T \tilde{\theta}_{c_3}$, 式中 $\tilde{\theta}_{a_3} = \hat{\theta}_{a_3} - \theta_{j_3}$, $\tilde{\theta}_{c_3} = \hat{\theta}_{c_3} - \theta_{j_3}$ 。则

$$\begin{aligned} V_3 \leq & \dot{V}_2 - (r_3 - \frac{3}{4}) z_3^2 + \frac{1}{2} z_4^2 - (\frac{1}{2} r_{a_3} - \frac{1}{4}) (\hat{\theta}_{a_3}^T \phi_{j_3})^2 - (\frac{1}{2} r_{c_3}) (\tilde{\theta}_{a_3}^T \phi_{j_3})^2 - \\ & (r_{c_3} - \frac{1}{2} r_{a_3}) (\hat{\theta}_{c_3}^T \phi_{j_3})^2 - (\frac{1}{2} r_{c_3}) (\tilde{\theta}_{c_3}^T \phi_{j_3})^2 + (\frac{1}{2} r_{a_3} + \frac{1}{2} r_{c_3}) (\theta_{j_3}^T \phi_{j_3})^2. \end{aligned}$$

第四步:将跟踪误差定义为 $z_4 = x_4 - \hat{\alpha}_3$, 根据式(2), 误差动态方程为

$$\dot{z}_4 = \frac{1}{J} u + \frac{1}{J} (-Bx_4 - k(x_3 - x_1)) - \dot{\hat{\alpha}}_3^*. \quad (23)$$

结合误差动态方程(23), 定义最优性能指标函数为

$$J_4^* = \min_{a_4 \in \psi(\Omega)} \left\{ \int_t^\infty c_4(z_4, u^*) ds \right\} = \int_t^\infty c_4(z_4, u^*) ds, \quad (24)$$

式中 $c_4(z_4, u^*) = z_4^2 + u^{*2}$ 为成本函数, u^* 为该反步步骤的最优控制器。与(23)和(24)相关联的 HJB 方程为 $H_4 = z_4^2 + u^{*2} + \frac{dJ_4}{dz_4} (\frac{1}{J} u^* + \frac{1}{J} (-Bx_4 - k(x_3 - x_1)) - \dot{\hat{\alpha}}_3^*) = 0$ 。

通过求解 $\frac{\partial H_4}{\partial u^*} = 0$ 可推导出最优控制器 u^* 为

$$u^* = J [-r_4 z_4 - \frac{1}{2} \theta_{j_4}^T \phi_{j_4} - \frac{1}{2} \epsilon_{j_4} + \dot{\hat{\alpha}}_3^* - \frac{1}{J} (-Bx_4 - k(x_3 - x_1))]. \quad (25)$$

为了得到有效的最优控制器, 设计 Actor-critic 结构来执行强化学习策略。设计评估控制性能的 Critic 为 $\frac{dJ_4^*}{dz_4} = J^2 [2r_4 z_4 + \hat{\theta}_{c_4}^T \phi_{j_4} - 2\dot{\hat{\alpha}}_3^* + 2 \frac{1}{J} (-Bx_4 - k(x_3 - x_1))]$, 式中 $\hat{\theta}_{c_4}$ 为 Critic 理想权重。根据公式

(25), 设计可用于执行控制动作的 Actor 为 $\hat{u}^* = J [-r_4 z_4 - \frac{1}{2} \hat{\theta}_{a_4}^T \phi_{j_4} + \dot{\hat{\alpha}}_3^* - \frac{1}{J} (-Bx_4 - K(x_3 - x_1))]$,

式中 $\hat{u}^*$ 是期望最优虚拟控制器, $\hat{\theta}_{a_4}$ 是 Actor 理想权重。设计理想权重 $\hat{\theta}_{c_4}$ 和 $\hat{\theta}_{a_4}$ 的权重更新率为 $\dot{\hat{\theta}}_{c_4} = -r_{c_4} \phi_{j_4} \phi_{j_4}^T \hat{\theta}_{c_4}$, $\dot{\hat{\theta}}_{a_4} = -\phi_{j_4} \phi_{j_4}^T (r_{a_4} (\hat{\theta}_{a_4} - \hat{\theta}_{c_4}) + r_{c_4} \hat{\theta}_{c_4})$, 式中 $r_{c_4} > 0$ 和 $r_{a_4} > 0$ 均是常数。构建如下李雅普诺夫函数

$$V_4 = V_3 + \frac{1}{2} z_4^2 + \frac{1}{2} \tilde{\theta}_{a_4}^T \tilde{\theta}_{a_4} + \frac{1}{2} \tilde{\theta}_{c_4}^T \tilde{\theta}_{c_4},$$

式中 $\tilde{\theta}_{a_4} = \hat{\theta}_{a_4} - \theta_{j_4}$, $\tilde{\theta}_{c_4} = \hat{\theta}_{c_4} - \theta_{j_4}$ 。则

$$\begin{aligned} \dot{V}_4 \leq & - (r_1 - \frac{3}{4}) z_1^2 - (r_2 - \frac{K^2}{2M^2} - \frac{3}{4}) z_2^2 - (r_3 - \frac{5}{4}) z_3^2 - (r_4 - \frac{3}{4}) z_4^2 - \sum_{i=1}^4 (\frac{1}{2} r_{a_i} - \frac{1}{4}) (\hat{\theta}_{a_i}^T \phi_{j_i})^2 - \\ & \sum_{i=1}^4 (\frac{1}{2} r_{c_i}) (\tilde{\theta}_{a_i}^T \phi_{j_i})^2 - \sum_{i=1}^4 (r_{c_i} - \frac{1}{2} r_{a_i}) (\hat{\theta}_{c_i}^T \phi_{j_i})^2 - \sum_{i=1}^4 (\frac{1}{2} r_{c_i}) (\tilde{\theta}_{c_i}^T \phi_{j_i})^2 + \sum_{i=1}^4 (\frac{1}{2} r_{a_i} + \frac{1}{2} r_{c_i}) (\theta_{j_i}^T \phi_{j_i})^2. \end{aligned}$$

如果 $r_i > \max \left\{ \frac{5}{4}, \frac{K^2}{2M^2} + \frac{3}{4} \right\}$, $r_{c_i} > \frac{1}{2} r_{a_i} > \frac{1}{4}$, 易得

$$\dot{V}_4 \leq -AV_4 + B, \quad (26)$$

式中 $A = \min \left\{ 2r_i - \frac{5}{4}, 2r_i - \frac{K^2}{M^2} - \frac{3}{2}, r_{c_i} \lambda_{\min}(\phi_{j_i} \phi_{j_i}^T) \right\}$, $B = \sum_{i=1}^4 (\frac{1}{2} r_{a_i} + \frac{1}{2} r_{c_i}) (\theta_{j_i}^T \phi_{j_i})^2$ 。

根据引理 1, 可由公式(26)推出以下式子

$$V_4(t) \leq e^{-At} V_4(0) + \frac{A}{B} (1 - e^{-At}). \quad (27)$$

根据不等式(27)可以得出结论,单连杆柔性关节机械臂系统的所有信号均为半全局最终一致有界。由上述推导可知当 B 越小且 A 越大时,跟踪误差就会越小。

3 仿真例子

本节针对单连杆柔性关节机械臂系统^[18],通过仿真实验来验证上述方法的可行性。

$$\begin{cases} ml^2\ddot{q}_1 + mgl\sin q_1 + F(\dot{q}_1) + Kq_1 = Kq_2, \\ J\ddot{q}_2 + B\dot{q}_2 + K(q_2 - q_1) = u, \end{cases}$$

式中 l 表示链路长度, m 表示连杆质量, J 表示关节柔韧度, g 表示重力加速度, K 表示刚度, B 表示灵活度, q_1 表示连杆位置, $\dot{q}_1$ 表示速度, $\ddot{q}_1$ 表示加速度; q_2 表示转子角位置, $\dot{q}_2$ 表示速度, $\ddot{q}_2$ 表示加速度。 u 是控制输入; $F(\dot{q}_1) = \cos\dot{q}_1$ 。

令 $x_1 = q_1, x_2 = \dot{q}_1, x_3 = q_2$ 和 $x_4 = \dot{q}_2$ 。系统被重新写成

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = \frac{K}{ml^2}x_3 - \frac{K}{ml^2}x_1 - \frac{K}{ml^2}\cos x_2 - \frac{g}{l}\sin x_1, \\ \dot{x}_3 = x_4, \\ \dot{x}_4 = \frac{1}{J}u - \frac{B}{J}x_4 - \frac{K}{J}(x_3 - x_1). \end{cases}$$

在本次仿真中,系统初值为 $x_1(0) = 1, x_2(0) = 0, x_3(0) = 0, x_4(0) = 0$, 其他相关控制设计数为 $r_1 = 0.1, r_2 = 0.001, r_3 = 0.000\ 01, r_4 = 800$ 且 $r_{a_1} = 200, r_{a_2} = 10, r_{a_3} = 200, r_{a_4} = 200, r_{c_1} = 300, r_{c_2} = 10, r_{c_3} = 200, r_{c_4} = 200$ 。图1~9是相应的仿真结果。图1是跟踪轨迹。图2是控制器的曲线。图3是跟踪误差。图4至图7是成本函数,图像清晰地表明,所研究的算法能使成本函数最小化。图8是 Actor 权重范数,图9是 critic 权重范数,图像表明权重范数是有界的,并且能收敛到常数值。从仿真图像可以得出结论,所设计的最优控制器实现了预期的跟踪性能。图10的误差对比与图11的控制器对比结果清晰地表明相较于传统方法,强化学习算法在减小跟踪误差与优化控制性能方面展现出显著优势,并且控制过程的稳定性更好、精准度更高。

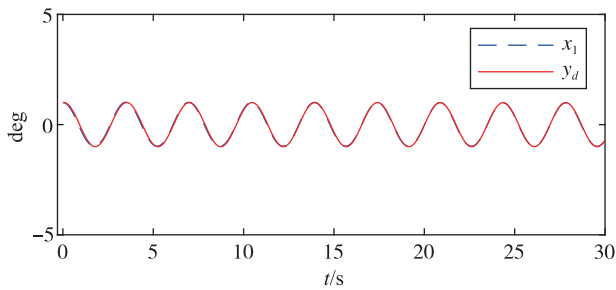


图1 跟踪轨迹

Fig. 1 Tracking trajectory

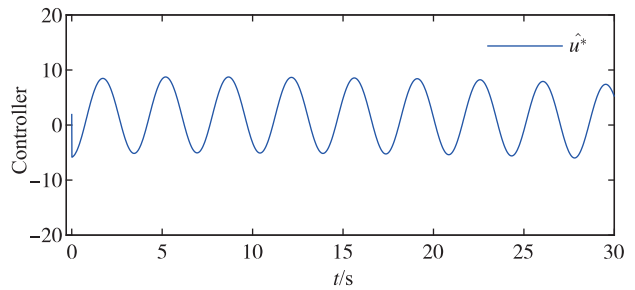


图2 控制器

Fig. 2 Controller

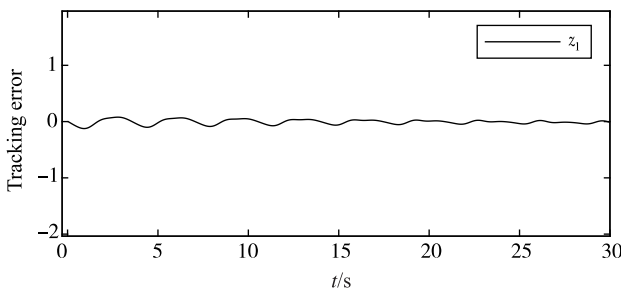


图3 跟踪误差

Fig. 3 Tracking error

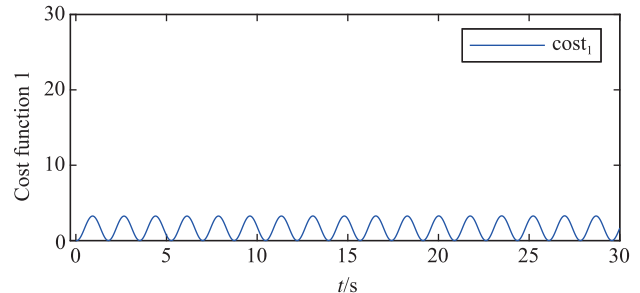


图4 成本函数1

Fig. 4 Cost function 1

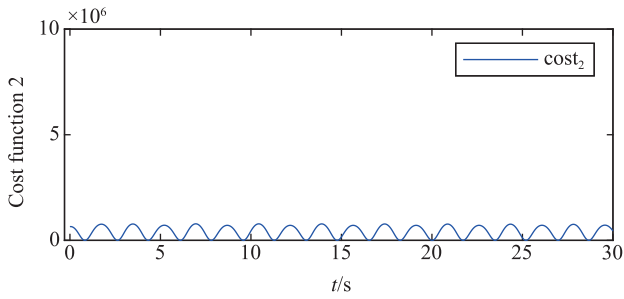


图5 成本函数 2

Fig. 5 Cost function 2

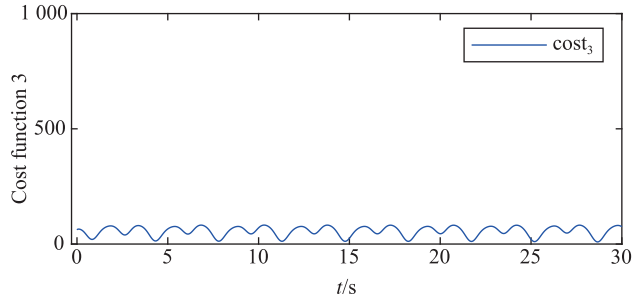


图6 成本函数 3

Fig. 6 Cost function 3

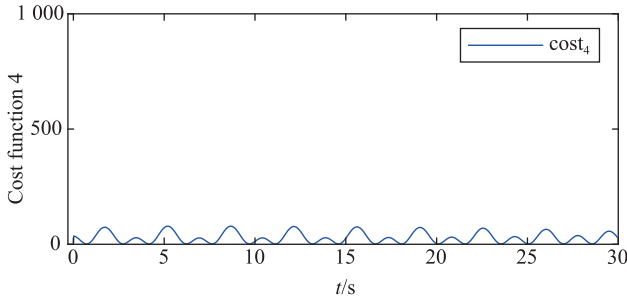


图7 成本函数 4

Fig. 7 Cost function 4

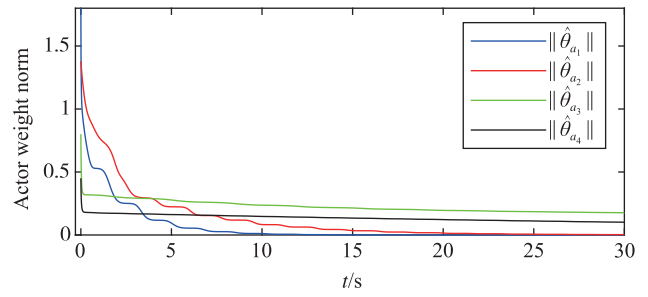


图8 Actor 权重范数

Fig. 8 Actor weight norm

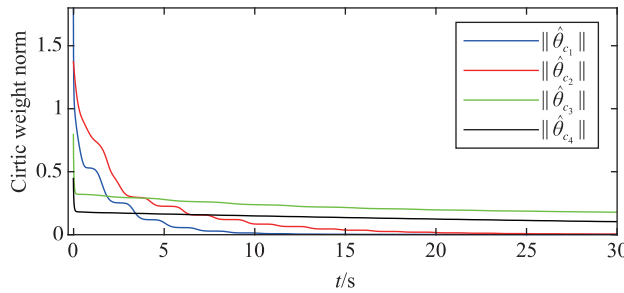


图9 critic 权重范数

Fig. 9 Critic weight norm

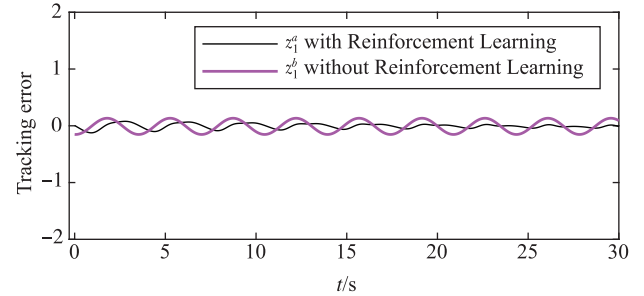


图10 误差对比

Fig. 10 Error comparison

4 结论

在本文中,针对单连杆柔性关节机械臂系统,通过强化学习策略下的优化反步追踪控制方法得到最优控制器,成功实现了跟踪期望目标。所设计的最优控制器简化了控制的设计过程,并使成本函数最小化。然后根据设计的李雅普诺夫函数可以推导出单连杆柔性关节机械臂系统的所有信号均为半全局最终一致有界。最后通过仿真结果验证了基于强化学习策略的优化反步跟踪控制方法对单连杆柔性关节机械臂系统控制的可行性。

在未来的研究中,我们将进一步应用强化学习算法来探究多连杆柔性关节机械臂系统的控制设计,其具有重要的学术与工程价值。在理论上,强化学习算法可克服传统控制方法对精确建模的依赖,解决多连杆柔性关节机械臂系统的高维非线性、强耦合及参数不确定性问题,推动智能自适应控制的发展。在应用上,该研究能提升多连杆柔性关节机械臂在精密装配、医疗手术

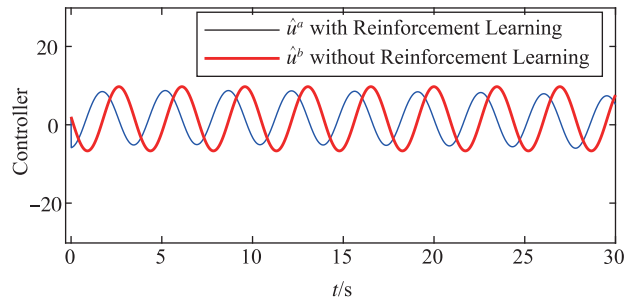


图11 控制器对比

Fig. 11 Controller comparison

等场景的动态响应与抗干扰能力,同时降低硬件振动与能耗,具有广阔的社会经济效益。

参 考 文 献

- [1] WERBOS P J. Approximate dynamic programming for real-time control and neural modeling[C]//Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches, 1992: 493-525.
- [2] ABU-KHALAF M, LEWIS F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach[J]. Automatica, 2005, 41(5): 779-791.
- [3] CHENG T, LEWIS F L, ABU-KHALAF M. A neural network solution for fixed-final time optimal control of nonlinear systems[C]//Proceedings of 14th Mediterranean Conference on Control and Automation IEEE, 2007: 482-490.
- [4] WEN G X, PHILIP CHEN C L, FENG J, et al. Optimized multi-agent formation control based on an identifier-actor-critic reinforcement learning algorithm[J]. IEEE Transactions on Fuzzy Systems, 2018, 26(5): 2719-2731.
- [5] VAMVOUDAKIS K G, LEWIS F L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem [J]. Automatica, 2010, 46(5): 878-888.
- [6] WEN G X, PHILIP CHEN C L, GE S S, et al. Optimized adaptive nonlinear tracking control using actor-critic reinforcement learning strategy[J]. IEEE Transactions on Industrial Informatics, 2019, 15(9): 4969-4977.
- [7] 黄诗毅. 不确定非线性系统的强化学习控制技术研究[D]. 南昌:华东交通大学, 2022.
- [8] 李亚锋. 线驱动柔性机械臂运动控制系统的设计研究[D]. 合肥:合肥工业大学, 2017.
- [9] LEWIS F L, VRABIE D, VAMVOUDAKIS K G. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers[J]. IEEE Control Systems Magazine, 2012, 32(6): 76-105.
- [10] 郑铮. 基于深度强化学习和零化动力学的机械臂轨迹跟踪控制算法的研究[D]. 广州:华南理工大学, 2023.
- [11] 周文烨. 受限柔性机械臂系统抗干扰约束控制算法研究[D]. 扬州:扬州大学, 2024.
- [12] 刁淑贞. 柔性关节机械臂的自适应智能控制[D]. 聊城:聊城大学, 2022.
- [13] 谢仕增, 孙伟. 单连杆柔性关节机械臂的高阶滑模控制器设计[J]. 聊城大学学报(自然科学版), 2024, 37(2): 9-18.
- [14] LING S, WANG H Q, LIU P X. Adaptive fuzzy dynamic surface control of flexible-joint robot systems with input saturation[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(1): 97-107.
- [15] DIAO S Z, SUN W, SU S F, et al. Novel adaptive control for flexible-joint robots with unknown measurement sensitivity[J]. IEEE Transactions on Automation Science and Engineering, 2023, 21(2): 1445-1456.
- [16] WANG X J, CAO Y, NIU B, et al. A novel bipartite consensus tracking control for multiagent systems under sensor deception attacks [J]. IEEE Transactions on Cybernetics, 2023, 53(9): 5984-5993.
- [17] WEN G X, GE S S, TU F W. Optimized backstepping for tracking control of strict-feedback systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(8): 3850-3862.
- [18] DIAO S Z, SUN W, SU S F, et al. Adaptive fuzzy event-triggered control for single-link flexible-joint robots with actuator failures[J]. IEEE Transactions on Cybernetics, 2021, 52(8): 7231-7241.

(责任编辑:刘庆松)