控制与决策
Control and Decision

具有误差约束的机械臂系统自适应强化学习控制

苏航, 张滋林

引用本文:
苏航, 张滋林. 具有误差约束的机械臂系统自适应强化学习控制[J]. 控制与决策, 2026, 41(5): 1331-1337.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2025.1010

具有误差约束的机械臂系统自适应强化学习控制

苏 航†，张滋林

(山东科技大学 电气与自动化工程学院， 山东 青岛 266590)

摘 要: 为了提高机械臂控制过程中的安全性, 针对机械臂系统的误差约束问题提出一种基于强化学习的自适应控制方法. 将机械臂的动力学系统转化为关于跟踪误差的动态方程, 然后利用一类误差转换函数, 将受约束的误差系统转换为新的不受约束系统, 并基于此系统设计最优控制器. 为了解决最优控制问题, 利用强化学习的方法求解系统的 HJB 方程, 其中评价网络用于逼近系统最优值函数, 执行网络用于逼近最优控制器的输出, 并利用一类正定函数来大幅简化评价-执行网络的自适应率. 基于李雅普诺夫稳定性理论, 证明系统所有误差信号半全局一致最终有界. 最后通过一个 2 自由度机械臂的仿真案例验证所提出方法的有效性.

关键词: 机械臂；误差约束；强化学习；神经网络；自适应控制；最优控制

中图分类号: TP13 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.1010

引用格式: 苏航, 张滋林. 具有误差约束的机械臂系统自适应强化学习控制 [J]. 控制与决策, 2026, 41(5): 1331-1337.

Adaptive reinforcement learning control for robotic manipulator systems with error constraints

SU Hang†，*ZHANG Zi-lin*

(College of Electrical Engineering and Automation， Shandong University of Science and Technology， Qingdao 266590， China)

Abstract: To enhance safety in the control process of robotic manipulators, an adaptive reinforcement learning-based control method is proposed for addressing error constraints in manipulator systems. For this purpose, the dynamic system of the manipulator is transformed into a dynamic equation concerning tracking errors. Then, an error transformation function is employed to convert the constrained error system into a new unconstrained system, based on which an optimal controller is designed. To solve the optimal control problem, a reinforcement learning approach is used to approximate the solution of the Hamilton-Jacobi-Bellman (HJB) equation, where a critic network approximates the optimal value function and an actor network approximates the output of the optimal controller. A positive definite function is introduced to significantly simplify the adaptation laws of the critic and actor networks. Based on Lyapunov stability theory, all error signals of the system are proven to be semi-globally uniformly ultimately bounded. Finally, the effectiveness of the proposed method is verified through a simulation example of a two-degree-of-freedom robotic manipulator.

Keywords: manipulator； error constraints； reinforcement learning； neural network； adaptive control； optimal control

0 引 言

在过去十几年中, 机器人技术得到了快速发展, 机械臂已广泛应用于工业制造、航空航天和医疗等多种行业[1-3], 并且取得了显著效果. 机械臂系统是一个复杂的非线性系统, 针对机械臂系统的控制问题, 研究人员提出了许多方法, 如滑模控制、自适应控制、神经网络控制等[4-10]. 其中: 文献 [11] 中利用神经网络逼近系统的不确定性和输入死区的未知参数, 并结合误差转换函数和障碍函数设计滑模面进行控制器设计; 针对柔性机械臂, 文献 [12] 利用自适应神经网络逼近未知的非线性函数, 并借助三角函数和障碍李雅普诺夫函数来处理系统的输入饱和约束和

收稿日期: 2025-09-24；**录用日期:** 2025-12-24.
基金项目: 国家自然科学基金项目 (62103243)；山东省自然科学基金项目 (ZR2025MS994).
责任编委: 訾斌.
†通信作者. E-mail: suhang0102@163.com.

输出约束问题.

自适应动态规划 (ADP) 方法是解决最优控制问题的一种重要方法. 通过逼近最优值函数来求解系统的哈密尔顿-雅可比-贝尔曼 (HJB) 方程. 基于强化学习的智能控制方法有着优越的学习能力和优化能力, 是现在解决最优控制问题应用最广泛的方法, 并且取得了许多成果[13-20]. 在文献 [21-22] 中将反步法与最优控制相结合, 解决了一类高阶系统的最优跟踪控制问题, 并成功应用至船舶系统[23]; 文献 [24] 和文献 [25], 通过设计非二次型代价函数得到满足输入饱和约束的最优控制器, 分别解决了系统模型未知的最优控制问题以及系统部分动态未知的跟踪控制问题; 文献 [26] 中为了保证机械臂系统的跟踪性能, 将控制器分解为最优控制器与稳态控制器两部分, 但无法保证整个控制器的最优性.

值得注意的是, 在现实条件下, 由于实际应用条件和环境的限制, 系统往往会受到一定的约束, 对系统施加规定的约束成为一个值得研究的问题[27-34]. 针对系统的约束问题, 文献 [35] 通过在反步法中的每一步引入障碍李雅普诺夫函数, 解决了一类高阶系统的全状态约束问题; 文献 [36] 通过设计辅助系统解决系统输入饱和约束问题; 文献 [37] 通过一类转换函数解决了机械臂系统的误差约束问题, 但其设计的自适应律较为复杂且收敛速度相对较慢.

综上所述, 从提升机械臂控制安全性和智能化的角度出发, 本文提出一种基于强化学习的机械臂优化控制方案, 用来解决具有误差约束的机械臂系统的跟踪控制问题. 所提出的方案具有如下优势:

1) 与文献 [9] 和文献 [26] 相比, 本文考虑了误差约束, 利用一类正定函数来大幅简化自适应律的形式, 减少控制过程中所需的计算量, 同时能够保证所设计的控制器整体为系统的最优解.

2) 与文献 [37] 相比, 本文所使用的方法有着更快的收敛速度以及更高的跟踪精度, 并且系统能够跟踪所设计的动态信号, 更具有一般性, 且不需要持续性激励假设.

1 问题描述

根据欧拉-拉格朗日公式, 对于n个自由度的机械臂, 其动力学方程可表示为

$$M(q)\ddot{q}+C(q,\dot{q})\dot{q}+G(q)+F(q)=\tau(t). \tag{1}$$

其中: $q=[q_1,\dots,q_n]^{\mathrm{T}}\in\mathbb{R}^n$为关节位置, $\dot{q}=[\dot{q}_1,\dots,\dot{q}_n]^{\mathrm{T}}\in\mathbb{R}^n$为关节速度, $\ddot{q}=[\ddot{q}_1,\dots,\ddot{q}_n]^{\mathrm{T}}\in\mathbb{R}^n$为关节加速度, $M(q)\in\mathbb{R}^{n\times n}$为正定对称惯性矩阵, $C(q,\dot{q})\in\mathbb{R}^n$为科里奥利和离心力项, $G(q)\in\mathbb{R}^n$为重力项, $F(\dot{q})\in\mathbb{R}^n$为摩擦力项, $\tau(t)\in\mathbb{R}^n$为输入力矩.

定义跟踪误差为

$$e(t)=q_d(t)-q(t), \tag{2}$$

其中$q_d(t)$是关节的期望轨迹. 根据文献 [9] 定义瞬时性能函数为

$$r(t)=\dot{e}(t)+\Lambda e(t), \tag{3}$$

其中$\Lambda\in\mathbb{R}^{n\times n}$是常数增益矩阵. 根据式 (2) 和 (3), 动力学方程 (1) 可以写为

$$M(q)\dot{r}(t)=-C(q,\dot{q})r(t)-\tau(t)+p(x), \tag{4}$$

其中$p(x)=M(q)(\ddot{q}_d+\Lambda\dot{e})+C(q,\dot{q})(\dot{q}_d+\Lambda\dot{e})+G(q)+F(\dot{q})$. 定义辅助控制器$u(t)=p(x)-\tau(t)$, 则式 (4) 可以写为

$$\dot{r}(t)=-M^{-1}(q)C(q,\dot{q})r(t)+M^{-1}(q)u(t). \tag{5}$$

利用式 (2) 和 (5) 能够得到以下系统:

$$\dot{x}=\begin{bmatrix}-\Lambda & I\\ 0_n & -M^{-1}C\end{bmatrix}x(t)+\begin{bmatrix}0_n\\ M^{-1}\end{bmatrix}u(t)=c(x)+h(x)u(t). \tag{6}$$

其中: $x(t)=[e(t),r(t)]^{\mathrm{T}}\in\mathbb{R}^{2n}$, $h(x)\in\mathbb{R}^{2n\times n}$, $c(x)\in\mathbb{R}^{2n\times 2n}$. 系统约束条件为: 跟踪误差$e(t)$要求保持在区间$\Omega_i:=\{\underline{k}_i<e_i<\overline{k}_i\}$, $\forall t>0$内.

控制目标是为系统 (6) 设计最优控制器并满足: 1) 系统所有误差信号半全局一致最终有界 (SGUUB); 2) 机械臂关节能够跟踪给定的参考信号, 且跟踪误差能够收敛并保持在原点附近的一个小邻域内.

2 预备知识

2.1 非线性转换函数

定义一个非线性转换函数$F:x\to\zeta$如下:

$$\zeta_i=F_i(\overline{k}_i,\underline{k}_i,x_i)=\log_a\left(\frac{\overline{k}_i(\underline{k}_i-x_i)}{\underline{k}_i(\overline{k}_i-x_i)}\right). \tag{7}$$

其中: $a>0$为常数, $\overline{k}_i>0$和$\underline{k}_i<0$为常数; ζ_i代表转换后不受约束的误差状态, x和ζ具有一一对应的映射关系; F为连续的初等函数, 具有以下性质:

$$\begin{cases}\lim_{x_i\to\underline{k}_i^+}\zeta_i\to-\infty,\\ \lim_{x_i\to\overline{k}_i^-}\zeta_i\to+\infty,\\ F_i(\overline{k}_i,\underline{k}_i,0)=0.\end{cases} \tag{8}$$

由式 (7) 可得

$$x_i=F_i^{-1}=\frac{\overline{k}_i\underline{k}_i(a^{\frac{1}{2}\zeta_i}-a^{-\frac{1}{2}\zeta_i})}{\underline{k}_ia^{\frac{1}{2}\zeta_i}-\overline{k}_ia^{-\frac{1}{2}\zeta_i}}. \tag{9}$$

根据式 (7), 定义$\zeta_1=F_1$, $\zeta_2=F_2$, 则对于式 (6) 中的$\dot{x}$, 有

$$\dot{\zeta} = f(\zeta) + g(\zeta)u(t). \tag{10}$$

其中: $D = \mathrm{diag}\left\{\dfrac{\partial F_1}{\partial x_1}, \dfrac{\partial F_2}{\partial x_2}\right\}$, $f(\zeta) = Dc(x)$, $g(\zeta) = Dh(x)$.

注 1　式 (10) 中的控制输入$u(t)$与式 (6) 中的$u(t)$为完全相同的控制输入. 本节的非线性转换仅针对受约束状态x进行可逆映射, 即$x \leftrightarrow \zeta$, 以将约束系统转换为无约束系统, 未对$u(t)$的表达式、物理意义及设计目标做任何修改, 仅系统动态项$c(x)$、$h(x)$因状态转换变为ζ的函数.

2.2　最优控制背景

考虑如下非线性连续时间动态系统:

$$\dot{x} = f(x) + g(x)u(t). \tag{11}$$

其中: $x(t)$表示系统状态, $u(t)$表示系统的控制输入. 定义系统的性能指标为

$$J(x) = \int_t^{\infty} r(x(s), u(s))\mathrm{d}s. \tag{12}$$

其中: $r(x(s), u(s)) = x^{\mathrm{T}}Nx + u^{\mathrm{T}}Ru$为系统的成本函数, $N = N^{\mathrm{T}}$, R为正定矩阵. 系统 (11) 的 HJB 方程为

$$H(x, u^*, \nabla J_x) = x^{\mathrm{T}}Nx + u^{\mathrm{T}}Ru + \nabla J_x\dot{x}(t). \tag{13}$$

其中: $\nabla J_x(x) = \partial J(x)/\partial x(t)$, 即$J(x)$对$x(t)$的梯度; u^*为系统的最优控制器. 若最优控制器存在, 此时的性能指标函数是$J^*(x)$, 则 HJB 方程可以写为

$$H(x, u^*, \nabla J_x^*) = x^{\mathrm{T}}Nx + u^{*\mathrm{T}}Ru^* + \nabla J_x^*(f(x) + g(x)u^*(t)) = 0, \tag{14}$$

其中$\nabla J_x^* = \partial J^*/\partial x$. 通过求解$\partial H(x, u^*, \nabla J_x^*)/\partial u^*$即可得到最优控制器为

$$u^* = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\nabla J_x^*(x), \tag{15}$$

将最优控制器 (15) 代入 HJB 方程 (14) 中即可求得$\nabla J_x^*(x)$, 再将求得的$\nabla J_x^*(x)$代入最优控制器 (15) 中即可得到系统 (11) 的最优控制器. 但由于其存在强非线性和耦合性, 使得对$\nabla J_x^*(x)$的求解非常困难. 因此, 本文采用一类简化的执行-评价网络强化学习方法来自适应逼近求解.

定义 1[20]　系统 (11) 中的控制器u是定义在集合Ω上的容许控制器, 即u在集合Ω上连续, $u(0) = 0$, 且u可以使得系统 (11) 在Ω上稳定. 容许控制器集记为$u \in \Psi(\Omega)$, 本文所研究的最优问题是设计最优控制器$u \in \Psi(\Omega)$, 使所设计的代价函数最小.

2.3　神经网络逼近

神经网络有出色的函数逼近和自适应学习能力. 任意定义在紧集Ω_z上的连续非线性函数$\varphi(z)$可以用神经网络逼近, 具体形式如下:

$$\psi(z) = \theta^{\mathrm{T}}S(z). \tag{16}$$

其中: $\theta = [\theta_1, \theta_2, \ldots, \theta_p]^{\mathrm{T}} \in \mathbb{R}^p$为权值向量, p表示神经元的个数; $z \in \Omega_z \subset \mathbb{R}^p$是输入向量; $S(z) = [s_1(z), s_2(z), \ldots, s_n(z)]^{\mathrm{T}} \in \mathbb{R}^p$为基函数向量, 基函数选择为

$$s_i(z) = \exp[-(z_i - \mu_i)^{\mathrm{T}}(z_i - \mu_i)/\phi_i^2], \tag{17}$$

ϕ_i是高斯函数的宽度, $\mu_i = [\mu_{i,1}, \mu_{i,2}, \ldots, \mu_{in}]^{\mathrm{T}}$是高斯函数的中心. 基于神经网络逼近特性, 存在理想权值θ^*, 函数$\psi(z)$可以重新描述为以下形式:

$$\psi(z) = \theta^{*\mathrm{T}}S(z) + \varepsilon(z),\ \forall z \in \Omega_z \subset \mathbb{R}^p; \tag{18}$$

$$\theta^* \triangleq \arg\min_{\theta \in \mathbb{R}^p}\{\sup_{z \in \Omega_z}|\psi(z) - \theta^{\mathrm{T}}S(z)|\}. \tag{19}$$

其中: $\theta^* = [\theta_1^*, \theta_2^*, \ldots, \theta_p^*]^{\mathrm{T}} \in \mathbb{R}^p$; $\varepsilon(z)$是逼近误差, 且满足$|\varepsilon(z)| \leqslant \delta^*$, δ^*是正常数.

3　控制器设计及稳定性证明

3.1　控制器设计

针对机械臂动力学模型 (1) 转换的误差系统 (10), 定义系统的性能指标为

$$J(\zeta) = \int_t^{\infty} r(\zeta(s), u(\zeta))\mathrm{d}s, \tag{20}$$

其中$r(\zeta(s), u(\zeta)) = \zeta^{\mathrm{T}}\zeta + u^{\mathrm{T}}u$为成本函数. 定义$u^*$为系统 (10) 的最优控制器, 则系统的最优性能指标为

$$J^*(\zeta) = \min_{u \in \Psi(\Omega)}\left(\int_t^{\infty} r(\zeta(s), u(\zeta))\mathrm{d}s\right) = \int_t^{\infty} r(\zeta(s), u^*(\zeta))\mathrm{d}s, \tag{21}$$

其中$u \in \Psi(\Omega)$是控制器u在集合Ω上的可容许控制器.

注 2　对于机械臂原误差系统 (6) 中的性能指标$V(x) = \int_t^{\infty}(x^{\mathrm{T}}x + u^{\mathrm{T}}u)\mathrm{d}s$, 转换后的性能指标$J(\zeta)$可以看作是$V(x)$的上界. 通过选择合适的转换函数$F$使$\partial F_i/\partial x_i \geqslant 1$, 基于其性质 (7) 和 (8), 有$|\zeta_i| \geqslant |x_i|$. 因此, 能够得到

$$\int_t^{\infty}(x^{\mathrm{T}}x + u^{\mathrm{T}}u)\mathrm{d}s \leqslant \int_t^{\infty}(\zeta^{\mathrm{T}}\zeta + u^{\mathrm{T}}u)\mathrm{d}s. \tag{22}$$

上式说明, 若能够保证转换后系统的最优性, 则能够保证原来系统的最优性.

系统的 HJB 方程为

$$H(\zeta, u(\zeta), \nabla J_\zeta) = r(\zeta(s), u(\zeta)) + \nabla J_\zeta\dot{\zeta}, \tag{23}$$

其中$\nabla J_\zeta = \partial J/\partial\zeta$, 表示$J(\zeta)$沿$\zeta$方向的梯度. 将式 (21) 代入系统 HJB 方程 (23) 中得到

$$H(\zeta,u^*,\nabla J_\zeta^*)=$$
$$\zeta^{\mathrm{T}}\zeta+u^{*\mathrm{T}}u^*+\nabla J_\zeta^*(f(\zeta)+g(\zeta)u^*(t))=0. \quad (24)$$

假设式 (24) 存在唯一解, 则可以通过求解$\partial H(\zeta,u^*,\nabla J_\zeta^*)/\partial u^*$得到最优控制器为

$$u^*=-\frac{1}{2}g^{\mathrm{T}}(\zeta)\nabla J_\zeta^*(\zeta). \quad (25)$$

将式 (25) 代入 (24) 可以求得$\nabla J_\zeta^*(\zeta)$, 从而得到实际的最优控制器, 但求解过程非常困难. 因此采用强化学习的方法来代替直接求解 HJB 方程, 利用神经网络逼近系统的最优性能指标

$$J_\zeta^*(\zeta)=\theta^{*\mathrm{T}}S(\zeta)+\varepsilon(\zeta). \quad (26)$$

其中: θ^*为理想神经网络权重向量; $S(\zeta)$为基函数向量; $\varepsilon(\zeta)$为神经网络的逼近误差, 并且$|\varepsilon(\zeta)|\leqslant\delta(\zeta)$, $\delta(\zeta)$是正常数.

基于式 (26), 最优控制器 (25) 为

$$u^*=-\frac{1}{2}g^{\mathrm{T}}(\zeta)(\nabla S^{\mathrm{T}}(\zeta)\theta^*+\nabla\varepsilon(\zeta)). \quad (27)$$

其中: $\nabla S(\zeta)=\partial S(\zeta)/\partial\zeta$, $\nabla\varepsilon(\zeta)=\partial\varepsilon(\zeta)/\partial\zeta$. 由于理想权重向量$\theta^*$是未知的, 最优控制器 (27) 不能直接使用. 为此, 构建如下评价网络和执行网络, 分别用于评估控制性能和调节最优控制器:

$$\hat{J}^*=\hat{\theta}_c^{\mathrm{T}}S(\zeta), \quad (28)$$

$$\hat{u}^*=-\frac{1}{2}g^{\mathrm{T}}(\zeta)\nabla S^{\mathrm{T}}(\zeta)\hat{\theta}_a. \quad (29)$$

其中: $\hat{J}^*(\zeta)$为$J^*(\zeta)$的估计, $\hat{u}^*$为u^*的估计, $\hat{\theta}_c$和$\hat{\theta}_a$分别是评价和执行网络的权重向量.

将式 (28) 和 (29) 代入 (23) 得到估计后的 HJB 方程

$$H(\zeta,\hat{u}^*,\nabla\hat{J}_\zeta^*)=\zeta^{\mathrm{T}}\zeta+\Big\|-\frac{1}{2}g^{\mathrm{T}}(\zeta)S^{\mathrm{T}}(\zeta)\hat{\theta}_a\Big\|^2+(\nabla S(\zeta)\hat{\theta}_c^{\mathrm{T}})(f(\zeta)+g(\zeta)\hat{u}^*(t)), \quad (30)$$

其中$\nabla\hat{J}_\zeta^*=\partial\hat{J}^*/\partial\zeta$.

根据式 (24) 和 (30) 可以得到系统的贝尔曼误差为

$$e(t)=H(\zeta,\hat{u}^*,\nabla\hat{J}_\zeta^*)-H(\zeta,u^*,\nabla J_\zeta^*)=H(\zeta,\hat{u}^*,\nabla\hat{J}_\zeta^*). \quad (31)$$

若最优控制器$\hat{u}^*$能够满足$e(t)\to0$, 即若$e(t)=0$成立并且存在唯一解, 则以下等式成立:

$$\frac{H(\zeta,\hat{u}^*,\nabla\hat{J}_\zeta^*)}{\hat{\theta}_a}=\frac{1}{2}\|g(\zeta)\|^2\nabla S\nabla S^{\mathrm{T}}(\hat{\theta}_a-\hat{\theta}_c)=0. \quad (32)$$

为了使所设计的自适应率能够保证式 (32) 成立, 定义以下正定函数:

$$P(t)=(\hat{\theta}_a-\hat{\theta}_c)^{\mathrm{T}}(\hat{\theta}_a-\hat{\theta}_c). \quad (33)$$

显然, 当$P(t)=0$成立时, 式 (32) 成立.

设计评价网络和执行网络权重的自适应律如下:

$$\dot{\hat{\theta}}_c=-\gamma_c\nabla S\nabla S^{\mathrm{T}}\hat{\theta}_c, \quad (34)$$

$$\dot{\hat{\theta}}_a=-\nabla S\nabla S^{\mathrm{T}}(\gamma_a(\hat{\theta}_a-\hat{\theta}_c)+\gamma_c\hat{\theta}_c). \quad (35)$$

其中: $\gamma_c>0$和$\gamma_a>0$分别为评价和执行网络的学习率, 并满足以下条件:

$$\gamma_a>\frac{1}{2},\ \gamma_c>\frac{\gamma_a}{2}. \quad (36)$$

根据式 (34) 和 (35) 能够得到

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t}=\frac{\partial P(t)}{\partial\hat{\theta}_c^{\mathrm{T}}}\dot{\hat{\theta}}_c+\frac{\partial P(t)}{\partial\hat{\theta}_a^{\mathrm{T}}}\dot{\hat{\theta}}_a=-\frac{\gamma_a}{2}\frac{\partial P(t)}{\partial\hat{\theta}_a^{\mathrm{T}}}\nabla S\nabla S^{\mathrm{T}}\frac{\partial P(t)}{\partial\hat{\theta}_a^{\mathrm{T}}}\leqslant0. \quad (37)$$

式 (37) 意味着当采用自适应律 (34) 和 (35) 时, 式 (32) 成立, 能够最小化贝尔曼误差.

3.2 稳定性分析

选择如下候选 Lyapunov 函数:

$$L=\frac{1}{2}\zeta^{\mathrm{T}}\zeta+\frac{1}{2}\tilde{\theta}_a^{\mathrm{T}}\tilde{\theta}_a+\frac{1}{2}\tilde{\theta}_c^{\mathrm{T}}\tilde{\theta}_c, \quad (38)$$

其中$\tilde{\theta}_a=\hat{\theta}_a-\theta^*$和$\tilde{\theta}_c=\hat{\theta}_c-\theta^*$分别是执行和评价网络权重的估计误差.

根据式 (10)、(29)、(34)、(35), 可得L对时间的导数为

$$\dot{L}=\zeta^{\mathrm{T}}f(\zeta)-\frac{1}{2}\zeta^{\mathrm{T}}Q(\zeta)\nabla S^{\mathrm{T}}\hat{\theta}_a-\tilde{\theta}_a^{\mathrm{T}}\nabla S\nabla S^{\mathrm{T}}(\gamma_a(\hat{\theta}_a-\hat{\theta}_c)+\gamma_c\hat{\theta}_c)-\gamma_c\tilde{\theta}_c^{\mathrm{T}}\nabla S\nabla S^{\mathrm{T}}\hat{\theta}_c, \quad (39)$$

其中$Q(\zeta)=g^{\mathrm{T}}(\zeta)g(\zeta)$. 应用杨氏不等式和式 (36) 的条件, 并结合$\tilde{\theta}_a$、$\tilde{\theta}_c$的定义, 式 (39) 可写为

$$\dot{L}\leqslant\zeta^{\mathrm{T}}\Big(\frac{1}{4}Q^{\mathrm{T}}Q+\frac{1}{2}I\Big)\zeta-\frac{\gamma_a}{2}\tilde{\theta}_a^{\mathrm{T}}\nabla S\nabla S^{\mathrm{T}}\tilde{\theta}_a-\frac{\gamma_c}{2}\tilde{\theta}_c^{\mathrm{T}}\nabla S\nabla S^{\mathrm{T}}\tilde{\theta}_c+B, \quad (40)$$

其中$B=\dfrac{\gamma_a+\gamma_c}{2}(\theta^{*\mathrm{T}}\nabla S)^2+\dfrac{1}{2}\|f(\zeta)\|^2$. 由于$B$中所有项都是有界的, 有$||B||\leqslant b$. 令$\lambda_{\min}^Q$和$\lambda_{\min}^S$分别为$\dfrac{1}{4}Q^{\mathrm{T}}Q+\dfrac{1}{2}I$和$\nabla S\nabla S^{\mathrm{T}}$的最小特征值, 则式 (40) 可以写为

$$\dot{L}\leqslant-\alpha L+b, \quad (41)$$

其中$\alpha=\min\{-\lambda_{\min}^Q,\gamma_a\lambda_{\min}^S,\gamma_c\lambda_{\min}^S\}$. 式 (41) 能够证明系统信号$\zeta$、$\tilde{\theta}_a$、$\tilde{\theta}_c$为 SGUUB.

注 3 基于转换函数 (7) 及其性质, 由于F是可逆的严格单调递增光滑函数, ζ的 SGUUB 能够保证

x的 SGUUB, 进而保证了规定的约束条件.

4　数值仿真

本节通过一个 2 自由度机械臂模型进行数值仿真, 以验证所提出的控制方法的有效性. 机械臂的每个关节都配有单独的电机来提供输入扭矩. 机械臂的物理参数如表 1 所示.

表1　机械臂物理参数

i	m_i/kg	l_i/m	l_{ci}/m	I_i/(kg · m²)
1	4	0.5	0.25	1
2	2	0.25	0.15	0.8

利用拉格朗日方程, 可以得到机械臂的动力学方程为

$$M(q)\ddot{q}+C(q,\dot{q})\dot{q}+G(q)+F(q)=\tau(t). \quad (42)$$

其中

$$\begin{aligned}
&m_{11}=\\
&m_1l_{c1}^2+m_2(l_1^2+l_{c2}^2+2l_1l_2\cos(q_2))+I_1+I_2,\\
&m_{12}=m_{21}=m_2l_{c2}^2+I_2+l_1m_2l_{c2}\cos(q_2),\\
&m_{22}=m_2l_{c2}^2+I_2,\\
&c_{11}=-2l_1m_2l_{c2}\sin(q_2)\dot{q}_2,\\
&c_{12}=-l_1m_2l_{c2}\sin(q_2)(\dot{q}_1+\dot{q}_2),\\
&c_{21}=-l_1m_2l_{c2}\sin(q_2)\dot{q}_2\\
&c_{22}=0,\\
&g_1=g(m_1l_{c1}+m_2l_1)\sin(q_1)+gm_2l_{c2}\sin(q_1+q_2),\\
&g_2=gm_2l_{c2}\sin(q_1+q_2),\\
&f_1=5\dot{q}_1+3\mathrm{sgn}(\dot{q}_1),\\
&f_2=4\dot{q}_2+2\mathrm{sgn}(\dot{q}_2),\\
&g=9.8.
\end{aligned}$$

设计参数选择为 $a=\exp(1)$, $\gamma_a=30$, $\gamma_c=0.02$. 机械臂关节初始位置和速度为$x_1(0)=q(0)=[0.5,0.5]^{\mathrm{T}}$, $x_2(0)=\dot{q}(0)=[0.3,-0.4]^{\mathrm{T}}$. 关节位置的期望信号为$q_d=[\sin t,\cos t]^{\mathrm{T}}$. 将上述机械臂系统转换为式 (6) 所示的误差系统; 约束范围设置为 $\overline{k}_1=-\underline{k}_1=1$, $\overline{k}_2=-\underline{k}_2=1$; 采用式 (29) 所示的实际控制器; 采用高斯函数作为基函数, 其中μ_i在$[-3,3]$内均匀分布, $\phi_i=1\,(i=1,2,\ldots,7)$为宽度; 评价-执行神经网络更新率采用式 (34) 和 (35) 的形式; 神经网络权值的初始值选择为

$$\hat{\theta}_a(0)=[0.2,\ldots,0.2]^{\mathrm{T}}\in\mathbb{R}^{7\times1},$$
$$\hat{\theta}_c(0)=[0.5,\ldots,0.5]^{\mathrm{T}}\in\mathbb{R}^{7\times1}.$$

评价网络的输入为式 (7) 中转换后的无约束状态$\zeta=[\zeta_1,\zeta_2]^{\mathrm{T}}$, 输出为式 (24) 最优函数的估计值, 执行网络的输入与评价网络一致, 输出为式 (25) 最优控制器的估计值. 机械臂两个关节位置和所设计的参考信号轨迹如图 1 所示. 可以看出, 所设计的控制器可以实现良好的跟踪效果.

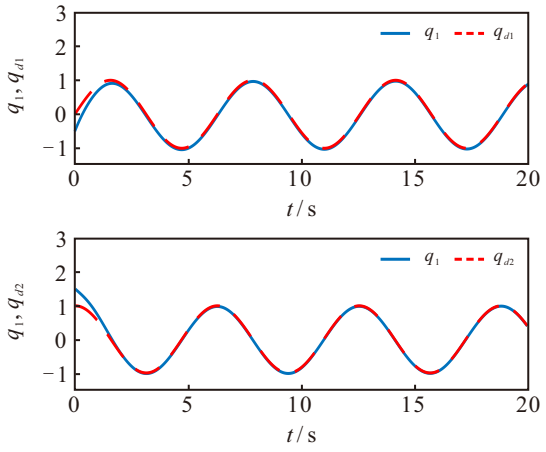

图1　关节位置跟踪效果

两个关节位置的跟踪误差以及转换后的系统误差状态如图 2 所示. 可以看出: 系统的跟踪误差能够保持在约束范围内, 且跟踪误差能够在 3 s 内收敛至零附近; 本文控制方法相较于文献 [21] 和文献 [37], 在收敛速度和跟踪误差方面具有优势. 系统输入力矩的变化曲线如图 3 所示. 需要说明的是, 图 3 中所示的输入力矩是施加于机械臂的实际力矩τ. 评价网络和执行网络权重的收敛情况如图 4 所示, 这些权重都能够在短暂的瞬态变化后收敛至常值. 为验证所提方法在误差约束方面的鲁棒性, 在$10\,\mathrm{s}<t<13\,\mathrm{s}$时添加$d(t)=[4\sin t,4\sin t]^{\mathrm{T}}$的突发扰动, 此时的机械臂动力学方程为 $M(q)\ddot{q}+C(q,\dot{q})\dot{q}+G(q)+F(q)=\tau(t)+d(t)$. 如图 5 所示, 此时的跟踪误差依旧保持在约束范围之内.

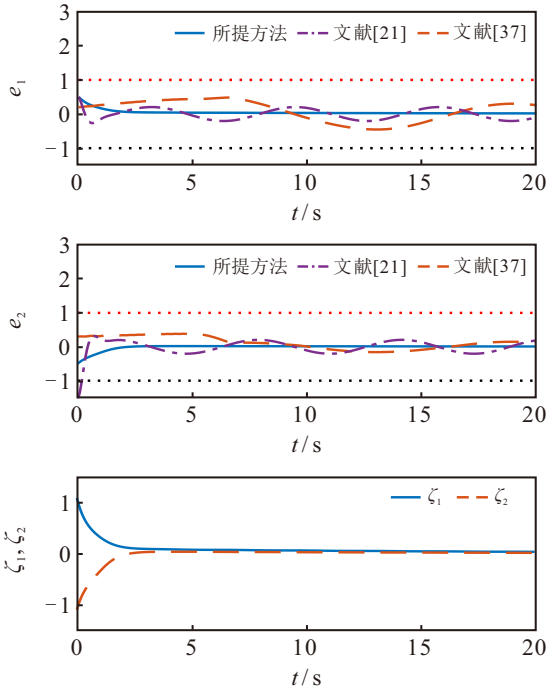

图2　3 种方法跟踪误差及转换后跟踪误差的轨迹

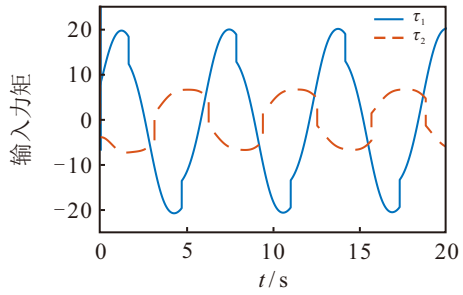

图3 关节输入力矩

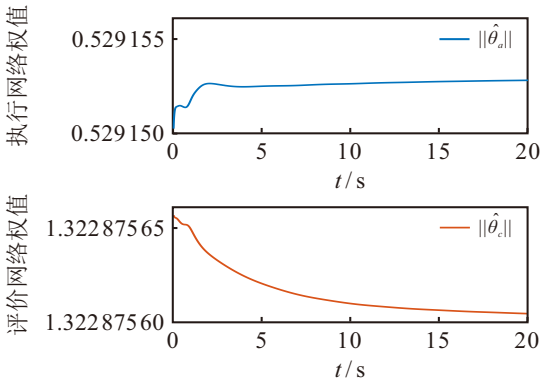

图4 执行与评价网络权值

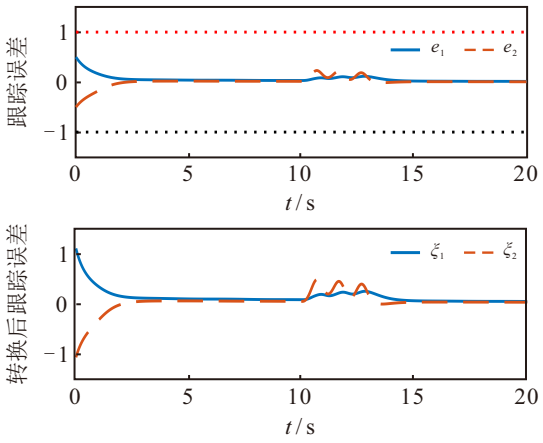

图5 突发扰动下跟踪误差

5 结 论

考虑到机械臂控制的安全性需求, 需保证其在运动过程中跟踪误差约束在预定范围内, 否则可能因环境干扰等因素引发安全问题. 因此, 考虑到机械臂的跟踪误差约束控制, 本文采用了一类转换函数将机械臂的误差系统转换为一类不受约束的系统, 并利用强化学习的方法来解决其最优控制问题. 在保证误差约束的前提下, 利用一类等效的正定函数替代传统的基于贝尔曼误差的梯度下降法来推导自适应律, 显著简化了其设计. 基于 Lyapunov 稳定性理论, 证明了系统所有信号均为 SGUUB. 通过 Matlab 仿真实验, 验证了所设计控制器在保证误差约束的同时, 能够精准地跟踪期望轨迹.

参考文献 (References)

[1] Lu Z L, Zhou Y J, Hu L, et al. A wearable human-machine interactive instrument for controlling a wheelchair robotic arm system[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 4005315.

[2] Zhou Y J, Yu T Y, Gao W, et al. Shared three-dimensional robotic arm control based on asynchronous BCI and computer vision[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 3163-3175.

[3] Wang M, Chen Z S, Guo K X, et al. Millimeter-level pick and peg-in-hole task achieved by aerial manipulator[J]. IEEE Transactions on Robotics, 2024, 40: 1242-1260.

[4] Chávez-Vázquez S, Lavín-Delgado J E, Gómez-Aguilar J F, et al. Trajectory tracking of Stanford robot manipulator by fractional-order sliding mode control[J]. Applied Mathematical Modelling, 2023, 120: 436-462.

[5] Zhang D, Hu J B, Cheng J, et al. A novel disturbance observer based fixed-time sliding mode control for robotic manipulators with global fast convergence[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(3): 661-672.

[6] Xu K, Wang Z L. The design of a neural network-based adaptive control method for robotic arm trajectory tracking[J]. Neural Computing and Applications, 2023, 35(12): 8785-8795.

[7] Liu Z T, Gao H J, Yu X H, et al. B-spline wavelet neural-network-based adaptive control for linear-motor-driven systems via a novel gradient descent algorithm[J]. IEEE Transactions on Industrial Electronics, 2024, 71(2): 1896-1905.

[8] Yan J K, Jin L, Hu B. Data-driven model predictive control for redundant manipulators with unknown model[J]. IEEE Transactions on Cybernetics, 2024, 54(10): 5901-5911.

[9] Kim Y H, Lewis F L, Dawson D M. Intelligent optimal control of robotic manipulators using neural networks[J]. Automatica, 2000, 36(9): 1355-1364.

[10] 范亚洲, 孙林祥, 白雪剑, 等. 基于自适应反正切非奇异终端滑模的水下机械臂轨迹跟踪控制[J]. 控制与决策, 2025, 40(1): 205-213.
(Fan Y Z, Sun L X, Bai X J, et al. Trajectory tracking control of underwater manipulator based on adaptive arctangent non-singular terminal sliding mode control[J]. Control and Decision, 2025, 40(1): 205-213.)

[11] Zhang Y, Kong L H, Zhang S, et al. Improved sliding mode control for a robotic manipulator with input deadzone and deferred constraint[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(12): 7814-7826.

[12] Shi M, Yu J J, Zhang T P. Command filter-based adaptive control of flexible-joint manipulator with input saturation and output constraints[J]. Asian Journal of Control, 2024, 26(1): 42-55.

[13] Shen H, Wang Y, Wang J, et al. A fuzzy-model-based approach to optimal control for nonlinear Markov jump singularly perturbed systems: A novel integral reinforcement learning scheme[J]. IEEE Transactions on Fuzzy Systems, 2023, 31(10): 3734-3740.

[14] Sun Y, Chen M, Peng K X, et al. Finite-time adaptive

optimal control of uncertain strict-feedback nonlinear systems based on fuzzy observer and reinforcement learning[J]. International Journal of Systems Science, 2024, 55(8): 1553-1570.

[15] Li D D, Dong J X. Fuzzy weight-based reinforcement learning for event-triggered optimal backstepping control of fractional-order nonlinear systems[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(1): 214-225.

[16] Xiao B, Zhang H C, Chen Z Y, et al. Fixed-time fault-tolerant optimal attitude control of spacecraft with performance constraint via reinforcement learning[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(6): 7715-7724.

[17] Zhao Y W, Wang H Q, Xu N, et al. Reinforcement learning-based decentralized fault tolerant control for constrained interconnected nonlinear systems[J]. Chaos, Solitons & Fractals, 2023, 167: 113034.

[18] Zhang Y, Chadli M, Xiang Z R. Prescribed-time formation control for a class of multiagent systems via fuzzy reinforcement learning[J]. IEEE Transactions on Fuzzy Systems, 2023, 31(12): 4195-4204.

[19] 李小华, 刘莹, 邹嵩楠, 等. 基于可变障碍函数和强化学习的预设性能最优安全跟踪控制[J]. 控制与决策, 2025, 40(3): 803-812.
(Li X H, Liu Y, Zou S N, et al. Optimal safety tracking control with prescribed performance based onvariable barrier function and reinforcement learning[J]. Control and Decision, 2025, 40(3): 803-812.)

[20] 申思凯, 江南, 刘小洋. 基于强化学习的多智能体系统容错编队最优控制[J]. 控制与决策, 2025, 40(12): 3565-3575.
(Shen S K, Jiang N, Liu X Y. Fault-tolerant formation optimized control for multi-agent systems via reinforcement learning[J]. Control and Decision, 2025, 40(12): 3565-3575.)

[21] 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.

[22] Wen G X, Chen C L P, Ge S S. Simplified optimized backstepping control for a class of nonlinear strict-feedback systems with unknown dynamic functions[J]. IEEE Transactions on Cybernetics, 2021, 51(9): 4567-4580.

[23] Wen G X, Ge S S, Chen C L P, et al. Adaptive tracking control of surface vessel using optimized backstepping technique[J]. IEEE Transactions on Cybernetics, 2019, 49(9): 3420-3431.

[24] Yang X, Liu D R, Wang D. Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints[J]. International Journal of Control, 2014, 87(3): 553-566.

[25] Zhu Y H, Zhao D B, He H B, et al. Event-triggered optimal control for partially unknown constrained-input systems via adaptive dynamic programming[J]. IEEE Transactions on Industrial Electronics, 2017, 64(5): 4101-4109.

[26] Long X Y, He Z, Wang Z Y. Online optimal control of robotic systems with single critic NN-based reinforcement learning[J]. Complexity, 2021, 2021: 8839391.

[27] Mishra P K, Jagtap P. Approximation-free prescribed performance control with prescribed input constraints[J]. IEEE Control Systems Letters, 2023, 7: 1261-1266.

[28] Luo A, Zhou Q, Ma H, et al. Observer-based consensus control for MASs with prescribed constraints via reinforcement learning algorithm[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(12): 17281-17291.

[29] Vázquez C R. Prescribed-time control of disturbed systems with state and input constraints[J]. IEEE Transactions on Automatic Control, 2025, 70(1): 619-626.

[30] Han S I, Lee J M. Improved prescribed performance constraint control for a strict feedback non-linear dynamic system[J]. IET Control Theory & Applications, 2013, 7(14): 1818-1827.

[31] Zhang J X, Xu K D, Wang Q G. Prescribed performance tracking control of time-delay nonlinear systems with output constraints[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(7): 1557-1565.

[32] Yao Y G, Kang Y, Zhao Y B, et al. A novel prescribed-time control approach of state-constrained high-order nonlinear systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(5): 2941-2951.

[33] Yang X R, Lin X X, Yang Y J, et al. Finite-time attitude tracking control of rigid spacecraft with multiple constraints[J]. IEEE Transactions on Aerospace and Electronic Systems, 2024, 60(3): 3688-3697.

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

[35] Liu Y J, Li J, Tong S C, et al. Neural network control-based adaptive learning design for nonlinear systems with full-state constraints[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(7): 1562-1571.

[36] Li L N, Liu Z X, Guo S F, et al. Adaptive practical predefined-time control for uncertain teleoperation systems with input saturation and output error constraints[J]. IEEE Transactions on Industrial Electronics, 2024, 71(2): 1842-1852.

[37] Ouyang Y C, Dong L, Sun C Y. Critic learning-based control for robotic manipulators with prescribed constraints[J]. IEEE Transactions on Cybernetics, 2022, 52(4): 2274-2283.

作者简介

苏航 (1989–), 女, 副教授, 博士, 主要研究方向为非线性系统控制, E-mail: suhang0102@163.com;

张滋林 (1999–), 男, 硕士, 主要研究方向为自适应神经网络控制, E-mail: zlzhang0123@163.com.