控制与决策
Control and Decision

基于单网络评判学习的非线性系统鲁棒跟踪控制

霍煜, 王鼎, 乔俊飞

引用本文:
霍煜,王鼎,乔俊飞. 基于单网络评判学习的非线性系统鲁棒跟踪控制[J]. 控制与决策, 2023, 38(11): 3066−3074.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.0124

基于单网络评判学习的非线性系统鲁棒跟踪控制

霍 煜[1,2,3,4], 王 鼎[1,2,3,4], 乔俊飞[1,2,3,4†]

(1. 北京工业大学 信息学部, 北京 100124;
2. 北京工业大学 计算智能与智能系统北京市重点实验室，北京 100124;
3. 北京工业大学 北京人工智能研究院，北京 100124; 4. 北京工业大学 智慧环保北京实验室，北京 100124)

摘 要: 针对一类具有不确定性的连续时间非线性系统,提出一种基于单网络评判学习的鲁棒跟踪控制方法. 首先建立由跟踪误差与参考轨迹构成的增广系统,将鲁棒跟踪控制问题转换为镇定设计问题. 通过采用带有折扣因子和特殊效用项的代价函数,将鲁棒镇定问题转换为最优控制问题. 然后,通过构建评判神经网络对最优代价函数进行估计,进而得到最优跟踪控制算法. 为了放松该算法的初始容许控制条件,在评判神经网络权值更新律中增加一个额外项. 利用Lyapunov方法证明闭环系统的稳定性及鲁棒跟踪性能. 最后,通过仿真结果验证该方法的有效性和适用性.

关键词: 单网络评判学习；非线性系统；不确定性；神经网络；最优控制；鲁棒跟踪控制

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.0124

引用格式: 霍煜, 王鼎, 乔俊飞. 基于单网络评判学习的非线性系统鲁棒跟踪控制 [J]. 控制与决策, 2023, 38(11): 3066-3074.

Robust tracking control for nonlinear systems based on critic learning formulation with single network

HUO Yu[1,2,3,4], *WANG Ding*[1,2,3,4], *QIAO Jun-fei*[1,2,3,4†]

(1. Faculty of Information Technology，Beijing University of Technology，Beijing 100124，China；2. Beijing Key Laboratory of Computational Intelligence and Intelligent System，Beijing University of Technology，Beijing 100124，China；3. Beijing Institute of Artificial Intelligence，Beijing University of Technology，Beijing 100124，China；4. Beijing Laboratory of Smart Environmental Protection，Beijing University of Technology，Beijing 100124，China)

Abstract: For a kind of continuous-time nonlinear systems with uncertainties, a robust tracking control method is established based on critic learning formulation with single network. Firstly, an augmented system consisting of the tracking error and the reference trajectory is established, and the robust tracking control problem is transformed into a stabilization design problem. By adopting a cost function with a discount factor and a special utility term, the robust stabilization problem is transformed into an optimal control problem. Then, the optimal cost function is estimated by building a critic neural network, and consequently the optimal tracking control algorithm can be derived. In order to relax the initial admissible control conditions in the proposed algorithm, an extra term is added to the weight updating law of the critic neural network. Furthermore, the stability of the closed-loop system and the robust tracking performance are proved using the Lyapunov approach. Finally, the effectiveness and applicability of the developed approach are demonstrated via simulation results.

Keywords: critic learning formulation with single network；nonlinear system；uncertainties；neural networks；optimal control；robust tracking control

0 引 言

非线性系统的最优控制依赖于求解Hamilton-Jacobi-Bellman(HJB)方程,这是一个具有挑战性的问题. 近年来,基于神经网络的函数逼近特性[1],自适应动态规划(adaptive dynamic programming, ADP)作为一种有效的近似最优控制设计方法出现. 这一方法

收稿日期: 2022-01-17；**录用日期:** 2022-05-24.
基金项目: 北京市自然科学基金项目(JQ19013)；国家自然科学基金项目(61773373, 61890930-5, 62021003)；科技创新2030-“新一代人工智能”重大项目(2021ZD0112302)；国家重点研发计划项目(2018YFC1900800-5).
责任编委: 陈家伟.
†通讯作者. E-mail: adqiao@bjut.edu.cn.

的成功应用主要归功于其独特的执行-评判双网络框架[2]. 具体而言,一个执行者向系统或环境释放一个动作,一个评判者对这个动作进行评估,并向执行者反馈一个积极或消极的信号. 在此设计框架下,得到HJB方程的近似解,从而避免所谓的维数灾难(即计算复杂度随系统维数增加而急剧增加的现象)[3]. 自20世纪70年代,ADP受到了控制界中众多学者的广泛关注,并被用于解决各种最优控制问题[4-7]. 在之后的研究中,为了简化设计过程,采用单评判网络设计了一些关于非线性系统的最优控制方案[8-9]. 在计算智能领域中,自适应动态规划、强化学习[10]与自适应评判设计[11]都采用相似的实现架构.

在复杂系统的实际应用中,控制系统通常会受到外部干扰、建模误差和系统老化等不确定性的影响. 这些无法避免的不确定性会严重降低系统性能,甚至会导致系统不稳定,因此有必要针对不确定非线性系统设计鲁棒控制器. 根据外部扰动是否在输入矩阵的范围内,通常将不确定性分为匹配不确定性和非匹配不确定性,前者是后者的特殊情况. 针对具有匹配或不匹配的不确定性非线性系统,一些先进的技术被引入鲁棒控制领域. 其中,鲁棒控制与最优控制的结合受到了特别关注. Lin等[12]通过设计相应的最优控制器,将具有匹配不确定性的非线性系统鲁棒控制问题与最优控制问题联系起来,该思想是先构造以最优代价函数为约束条件的标称系统,通过对标称系统进行等价最优控制,为求解不确定系统的鲁棒控制问题提供了新途径. 此后,在考虑ADP思想的基础上,逐步提出了相关的研究方法. Huang[13]针对一类非线性不确定系统的最优保成本控制问题,提出了一种基于ADP的并行学习方法; Fan等[14]将自适应执行—评判方法与滑模控制设计相结合,处理存在输入扰动的部分未知动态非线性系统; Jiang等[15]利用基于数据驱动的强化学习方法,研究了一类完全未知动态的不确定非线性系统鲁棒控制问题; Wang等[16]提出了一种基于学习的鲁棒镇定方法,通过系统变换和自适应评判设计,并利用对标称系统的近似最优控制,实现了对原系统中匹配不确定部分的鲁棒镇定;此后, Wang[17]进一步采用辅助系统方法和积分策略迭代算法实现了对不匹配非线性系统的鲁棒镇定; Zhao等[18]针对不匹配不确定性非线性连续系统,提出了一种鲁棒控制设计方法,其核心思想是通过构造含有代价函数的增广辅助系统,将鲁棒控制问题转换为最优控制问题.

非线性系统的跟踪控制是一个经典问题,在系统和控制领域中具有重要意义[19-20]. 因此,除了基于ADP的鲁棒镇定问题,一些学者也针对基于自适应评判的最优跟踪控制设计进行了研究[21-23]. 最优跟踪问题的目的是找到一种控制策略,使被考虑的系统能够最优地跟踪设定的期望参考轨迹. 目前,基于ADP的最优跟踪问题的研究结果大多集中在确定性系统上,例如, Modares等[24]针对带有输入约束的部分未知非线性系统,提出了一种基于积分强化学习的最优跟踪控制方法; Hou等[25]利用迭代自学习算法,提出了一种受约束非线性系统的数据驱动最优跟踪控制方法. 当考虑被控对象受到不确定性的影响时,鲁棒跟踪控制将成为被关注的问题. Wang等[26]研究了一类不考虑折扣因子的匹配不确定系统的鲁棒跟踪控制问题; Mu等[27]针对一类具有不匹配不确定性的连续非线性系统,提出了一种近似最优控制策略; Cui等[28]对于一类带有扰动和输入约束的不确定非线性系统,提出了一种H_∞跟踪控制方案.

自适应评判设计通常需要选择初始稳定控制器. 需要注意的是,在实际应用中总是很难获得这种控制器. 非线性系统经常涉及到不确定性,如电力系统、倒立摆系统和质量-弹簧-阻尼系统等. 因此,对这类系统研究鲁棒跟踪控制具有重要意义. 针对上述问题,本文研究一般不确定非线性系统的鲁棒跟踪问题. 本文提出的控制算法将为解决上述问题提供一种有效的方法. 利用问题转换和自适应评判学习的思想,提出一种鲁棒跟踪控制方法. 通过采用一种增加额外项的权值更新律,消除传统方法中需要选择初始稳定控制器的条件,为单评判网络设计带来独特的优势. 同时,考虑带有折扣因子的代价函数,通过折扣因子中的可调参数可以进一步保证代价函数的有界性. 此外,与采用执行-评判双网络的结构相比,采用基于单评判网络结构的在线ADP方法,可以减少计算成本,并且利用Lyapunov理论证明闭环系统的稳定性. 最后通过仿真实例,验证该控制方法的有效性.

1　问题描述

考虑一类具有如下形式的不确定非线性系统:

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) + \Delta f(x(t)). \quad (1)$$

其中: $x(t) \in \boldsymbol{R}^n$和$u(t) \in \boldsymbol{R}^m$分别为系统状态和控制输入; $f(x) \in \boldsymbol{R}^n$和$g(x) \in \boldsymbol{R}^{n\times m}$为已知函数,且$f(0) = 0$; $\Delta f(x) \in \boldsymbol{R}^n$为未知扰动,且$\Delta f(0) = 0$. 通常,令$x(0) = x_0$为系统初始状态. 为了后续分析的简便性,给出如下假设.

假设1　$f(x) + g(x)u$在包含原点的紧集$\Omega \in$

$\boldsymbol{R}^n$上满足局部Lipschitz连续条件,且系统(1)是可控的. 不确定项$\Delta f(x)$以一个已知函数$k_M(x)$为界,即$\|\Delta f(x)\| \leqslant k_M(x)$且$k_M(0)=0$.

为了研究跟踪控制问题,给定如下参考系统:

$$\dot{x}_d(t)=\delta(x_d(t)). \tag{2}$$

其中:$x_d(t)\in\boldsymbol{R}^n$为期望轨迹,$x_d(0)=x_{d0}$;$\delta(x_d(t))$满足局部Lipschitz连续条件,且$\delta(0)=0$. 定义跟踪误差为$\eta(t)=x(t)-x_d(t)$,且初始值为$\eta(0)=\eta_0=x_0-x_{d0}$. 那么,结合式(1)与(2)可以得到跟踪误差动态方程

$$\begin{aligned}\dot{\eta}(t)=&f(\eta(t)+x_d(t))+g(\eta(t)+x_d(t))u(t)-\\&\delta(x_d(t))+\Delta f(\eta(t)+x_d(t)).\end{aligned} \tag{3}$$

定义增广状态为$S(t)=[\eta^{\mathrm{T}}(t),x_d^{\mathrm{T}}(t)]^{\mathrm{T}}\in\boldsymbol{R}^{2n}$,且相应的初始值为$S(0)=S_0=[\eta_0^{\mathrm{T}},x_{d0}^{\mathrm{T}}]^{\mathrm{T}}$. 根据式(2)和(3),增广系统的动态方程可表示为

$$\dot{S}(t)=F(S(t))+G(S(t))u(t)+\Delta F(S(t)). \tag{4}$$

其中:$F:\boldsymbol{R}^{2n}\to\boldsymbol{R}^{2n}$, $G:\boldsymbol{R}^{2n}\to\boldsymbol{R}^{2n\times m}$, $\Delta F(S(t))\in\boldsymbol{R}^{2n}$,同时有

$$\begin{aligned}F(S(t))=&\begin{bmatrix}f(\eta(t)+x_d(t))-\delta(x_d(t))\\ \delta(x_d(t))\end{bmatrix},\\ G(S(t))=&\begin{bmatrix}g(\eta(t)+x_d(t))\\ 0_{n\times m}\end{bmatrix},\\ \Delta F(S(t))=&\begin{bmatrix}\Delta f(\eta(t)+x_d(t))\\ 0_{n\times 1}\end{bmatrix}.\end{aligned} \tag{5}$$

值得注意的是,上述新构造的不确定项$\Delta F(S(t))$有界条件仍然成立,可表示为$\|\Delta F(S)\|=\|\Delta f(x)\|\leqslant k_M(x)\triangleq k_M(S)$.

为了实现系统(1)对参考轨迹(2)的鲁棒跟踪目的,构造具有不确定性$\Delta F(S(t))$的增广系统(4),目标是找到一个反馈控制律$u(S)$确保闭环系统是稳定的. 接下来,证明鲁棒跟踪控制问题可以转换为对相应标称系统的最优控制问题. 那么,标称系统的动态方程可写为

$$\dot{S}(t)=F(S(t))+G(S(t))u(t). \tag{6}$$

对于式(6)这个非线性系统,希望找到一个反馈控制律$u(S)$来最小化如下代价函数:

$$J(S(t))=\int_t^{\infty}\mathrm{e}^{-\xi(\tau-t)}\{\Psi(S(\tau))+U(S(\tau),u(\tau))\}\mathrm{d}\tau. \tag{7}$$

其中:$\xi>0$为折扣因子;$\Psi(S)\geqslant 0$为仅与不确定性有关的额外项;$U(S,u)$为基本效用函数,$U(0,0)=0$,且对于任意S和u,$U(S,u)\geqslant 0$均成立. 令$U(S,u)=S^{\mathrm{T}}Q_SS+u^{\mathrm{T}}Ru$,则$Q_S=\mathrm{diag}\{Q,0_{n\times n}\}$.

在这一部分中,考虑具有代价函数(7)的标称系统(6)的最优控制问题,可以得出非线性Lyapunov方程

$$\begin{aligned}0=&\Psi(S)+U(S,u(S))+(\nabla J(S))^{\mathrm{T}}\cdot\\&[F(S)+G(S)u(S)]-\xi J(S),\end{aligned} \tag{8}$$

其中$J(0)=0$. 定义Hamiltonian为

$$\begin{aligned}&H(S,u(S),\nabla J(S))=\\&\Psi(S)+U(S,u(S))+\\&(\nabla J(S))^{\mathrm{T}}[F(S)+G(S)u(S)]-\xi J(S),\end{aligned} \tag{9}$$

其中$\nabla(\cdot)\triangleq\partial(\cdot)/\partial S$.

定义最优代价函数

$$J^*(S)=\min_{u\in\Gamma(\Omega)}J(S), \tag{10}$$

其中$\Gamma(\Omega)$为容许控制集合. $J^*(S)$满足如下HJB方程:

$$0=\min_{u\in\Gamma(\Omega)}H(S,u(S),\nabla J^*(S)), \tag{11}$$

相应的最优反馈控制律为

$$u^*(S)=-\frac{1}{2}R^{-1}G^{\mathrm{T}}(S)\nabla J^*(S). \tag{12}$$

结合式(8)与(12),可以将HJB方程重写为

$$\begin{aligned}0=&\Psi(S)+U(S,u^*(S))+(\nabla J^*(S))^{\mathrm{T}}\cdot\\&[F(S)+G(S)u^*(S)]-\xi J^*(S),\end{aligned} \tag{13}$$

其中$J^*(0)=0$. 随后,令额外效用项$\Psi(S)$为

$$\Psi(S)=\frac{1}{4}(\nabla J^*(S))^{\mathrm{T}}\nabla J^*(S)+k_M^2(S). \tag{14}$$

由于直接求解HJB方程(13)十分困难,接下来将通过设计评判神经网络得到该方程的近似解,从而进一步实现鲁棒跟踪控制.

2 基于ADP的鲁棒跟踪控制设计

2.1 评判网络设计

采用如下神经网络将最优代价函数表示为

$$J^*(S)=\omega_c^{\mathrm{T}}\varphi_c(S)+\varsigma_c(S). \tag{15}$$

其中:$\omega_c\in\boldsymbol{R}^N$为未知的理想权值向量;$\varphi_c(S)\in\boldsymbol{R}^N$为激活函数,$N$为隐藏神经元数量;$\varsigma_c(S)\in\boldsymbol{R}$为重构误差. 相应的梯度向量可写为

$$\nabla J^*(S)=(\nabla\varphi_c(S))^{\mathrm{T}}\omega_c+\nabla\varsigma_c(S). \tag{16}$$

在自适应评判网络设计中,考虑到理想权值未知的情况,通常以权值向量的估计值$\hat{\omega}_c$构建评判网络来逼近最优代价函数,即近似最优代价函数可写为

$$\hat{J}^*(S)=\hat{\omega}_c^{\mathrm{T}}\varphi_c(S). \tag{17}$$

与式(16)相似,可以得到

$$\nabla \hat{J}^*(S) = (\nabla \varphi_c(S))^{\mathrm{T}} \hat{\omega}_c. \tag{18}$$

利用最优代价函数的梯度(16),将最优控制律(12)改写为

$$u^*(S) = -\frac{1}{2} R^{-1} G^{\mathrm{T}}(S)[(\nabla \varphi_c(S))^{\mathrm{T}} \omega_c + \nabla \varsigma_c(S)]. \tag{19}$$

同理,基于近似最优代价函数的梯度(18)得到近似最优反馈控制律

$$\hat{u}^*(S) = -\frac{1}{2} R^{-1} G^{\mathrm{T}}(S)(\nabla \varphi_c(S))^{\mathrm{T}} \hat{\omega}_c. \tag{20}$$

将近似最优控制律(20)代入标称增广系统(6),闭环系统动态方程(即$\dot{S} = F(S) + G(S)\hat{u}^*(S)$)可表示为

$$\dot{S} = F(S) - \frac{1}{2} G(S) R^{-1} G^{\mathrm{T}}(S)(\nabla \varphi_c(S))^{\mathrm{T}} \hat{\omega}_c. \tag{21}$$

为了后续分析的简便性,令

$$\mathcal{V}(S) = (\nabla \varphi_c(S)) G(S) R^{-1} G^{\mathrm{T}}(S)(\nabla \varphi_c(S))^{\mathrm{T}},$$

$$\mathcal{W}(S) = (\nabla \varphi_c(S))(\nabla \varphi_c(S))^{\mathrm{T}}.$$

在评判网络框架下,代价函数和控制律都可以表示为权值向量的函数. 因此,相应的Hamiltonian可以写成一个包含S和ω_c的新方程,即

$$\begin{aligned} &H(S, \omega_c) = \\ &S^{\mathrm{T}} Q_S S + \omega_c^{\mathrm{T}} \nabla \varphi_c(S) F(S) - \xi \omega_c^{\mathrm{T}} \varphi_c(S) - \\ &\frac{1}{4} \omega_c^{\mathrm{T}} \mathcal{V}(S) \omega_c + \frac{1}{4} \omega_c^{\mathrm{T}} \mathcal{W}(S) \omega_c + k_M^2(S) + \varsigma_{\mathrm{HJB}} = 0. \end{aligned} \tag{22}$$

其中:ς_{HJB}为残差,可写为

$$\begin{aligned} &\varsigma_{\mathrm{HJB}} = \\ &(\nabla \varsigma_c(S))^{\mathrm{T}} F(S) - \xi \varsigma_c(S) + \frac{1}{4} (\nabla \varsigma_c(S))^{\mathrm{T}} \nabla \varsigma_c(S) - \\ &\frac{1}{2} (\nabla \varsigma_c(S))^{\mathrm{T}} G(S) R^{-1} G^{\mathrm{T}}(S)(\nabla \varphi_c(S))^{\mathrm{T}} \omega_c - \\ &\frac{1}{4} (\nabla \varsigma_c(S))^{\mathrm{T}} G(S) R^{-1} G^{\mathrm{T}}(S) \nabla \varsigma_c(S) + \\ &\frac{1}{2} (\nabla \varsigma_c(S))^{\mathrm{T}} (\nabla \varphi_c(S))^{\mathrm{T}} \omega_c. \end{aligned} \tag{23}$$

通过估计的权值向量,近似Hamiltonian可以表示为

$$\begin{aligned} \hat{H}(S, \hat{\omega}_c) = {} & S^{\mathrm{T}} Q_S S + \hat{\omega}_c^{\mathrm{T}} \nabla \varphi_c(S) F(S) - \xi \hat{\omega}_c^{\mathrm{T}} \varphi_c(S) - \\ & \frac{1}{4} \hat{\omega}_c^{\mathrm{T}} \mathcal{V}(S) \hat{\omega}_c + \frac{1}{4} \hat{\omega}_c^{\mathrm{T}} \mathcal{W}(S) \hat{\omega}_c + k_M^2(S). \end{aligned} \tag{24}$$

定义误差$e_c = \hat{H}(S, \hat{\omega}_c) - H(S, \omega_c)$,并且考虑式(22)可以得出$e_c = \hat{H}(S, \hat{\omega}_c)$. 定义评判网络的权值估计误差为$\tilde{\omega}_c = \omega_c - \hat{\omega}_c$,结合式(22)与(24),$e_c$的表达式可以写为

$$\begin{aligned} e_c = \hat{H}(S, \hat{\omega}_c) - H(S, \omega_c) = \\ - \tilde{\omega}_c^{\mathrm{T}} \nabla \varphi_c(S) F(S) + \xi \tilde{\omega}_c^{\mathrm{T}} \varphi_c(S) - \\ \frac{1}{4} \tilde{\omega}_c^{\mathrm{T}} \mathcal{V}(S) \tilde{\omega}_c + \frac{1}{2} \tilde{\omega}_c^{\mathrm{T}} \mathcal{V}(S) \omega_c + \\ \frac{1}{4} \tilde{\omega}_c^{\mathrm{T}} \mathcal{W}(S) \tilde{\omega}_c - \frac{1}{2} \tilde{\omega}_c^{\mathrm{T}} \mathcal{W}(S) \omega_c - \varsigma_{\mathrm{HJB}}. \end{aligned} \tag{25}$$

将目标函数定义为$E_c = (1/2) e_c^{\mathrm{T}} e_c$,通过训练评判网络使$E_c$的值最小.

为了放松对初始控制条件的约束,构造一种评判网络权值更新律,同时给出以下假设,也用于之后的稳定性分析.

假设2　考虑具有代价函数(7)的标称增广系统(6)和最优反馈控制律(12),令$L_2(S)$为Lyapunov函数,并满足

$$\dot{L}_2(S) = (\nabla L_2(S))^{\mathrm{T}}(F(S) + G(S)u^*(S)) < 0, \tag{26}$$

其中$\nabla L_2(S)$为$L_2(S)$对于状态变量S的偏导数. 另外,存在一个正定矩阵$\mathcal{M}$使得下式成立:

$$\begin{aligned} &(\nabla L_2(S))^{\mathrm{T}}(F(S) + G(S)u^*(S)) \leqslant \\ &- \lambda_{\min}(\mathcal{M}) \|\nabla L_2(S)\|^2. \end{aligned} \tag{27}$$

注1　式(27)是已有文献中常用的假设(如文献[23, 26, 29]),便于讨论闭环系统的稳定性. 根据文献[29]的结果,利用最优控制律的闭环系统是有界的,而且这个上界可以表示为一个状态向量的函数. 在这种情况下,假设$\|F(S) + G(S)u^*(S)\| \leqslant \beta \|\nabla L_2(S)\|$,其中$\beta > 0$,可以得到$\|(\nabla L_2(S))^{\mathrm{T}}(F(S) + G(S)u^*(S))\| \leqslant \beta \|\nabla L_2(S)\|^2$. 结合式(26)与$\lambda_{\min}(\mathcal{M}) \|\nabla L_2(S)\|^2 \leqslant (\nabla L_2(S))^{\mathrm{T}} \mathcal{M} \nabla L_2(S) \leqslant \lambda_{\max}(\mathcal{M}) \|\nabla L_2(S)\|^2$,可以得出假设2是合理的.

为了实现该算法,可以通过合理地选择状态变量的多项式得到$L_2(S)$,例如,令$L_2(S) = 0.5 S^{\mathrm{T}} S$.

然后,通过如下改进的更新律调整评判网络的权值:

$$\begin{aligned} &\dot{\hat{\omega}}_c = \\ &- \alpha_c \Big(\frac{\partial E_c}{\partial \hat{\omega}_c} \Big) + \frac{\alpha_s}{2} \nabla \varphi_c(S) G(S) R^{-1} G^{\mathrm{T}}(S) \nabla L_2(S). \end{aligned} \tag{28}$$

其中:$\alpha_c > 0$为评判网络的学习率,$\alpha_s > 0$为附加稳定项的学习率,$\nabla L_2(S)$由假设2给出.

2.2　稳定性分析

下面推导权值估计误差的动力学方程. 通过式(24)可得

$$\begin{aligned} &\frac{\partial e_c}{\partial \hat{\omega}_c} = \\ &\nabla \varphi_c(S) F(S) - \frac{1}{2} \mathcal{V}(S) \hat{\omega}_c + \frac{1}{2} \mathcal{W}(S) \hat{\omega}_c - \xi \varphi_c(S). \end{aligned} \tag{29}$$

根据式(28)以及$\dot{\tilde{\omega}}_c=-\dot{\hat{\omega}}_c$得到权值估计误差的动力学方程

$$\dot{\tilde{\omega}}_c=\alpha_c\Big(\frac{\partial E_c}{\partial\tilde{\omega}_c}\Big)-\frac{\alpha_s}{2}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla L_2(S). \tag{30}$$

结合式(25)与(29),将式(30)展开重写为

$$\begin{aligned}\dot{\tilde{\omega}}_c=&\alpha_c\Big(-\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)F(S)-\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{V}(S)\tilde{\omega}_c+\\&\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{V}(S)\omega_c+\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\tilde{\omega}_c-\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\omega_c-\\&\varsigma_{\mathrm{HJB}}+\xi\tilde{\omega}_c^{\mathrm{T}}\varphi_c(S)\Big)\cdot\\&\Big(\nabla\varphi_c(S)F(S)-\frac{1}{2}\mathcal{V}(S)\omega_c+\frac{1}{2}\mathcal{V}(S)\tilde{\omega}_c+\\&\frac{1}{2}\mathcal{W}(S)\omega_c-\frac{1}{2}\mathcal{W}(S)\tilde{\omega}_c-\xi\varphi_c(S)\Big)-\\&\frac{\alpha_s}{2}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla L_2(S).\end{aligned} \tag{31}$$

接下来将证明在近似最优控制器作用下,评判网络权值估计误差和闭环系统状态为一致最终有界(uniformly ultimately bounded, UUB). 证明前,作以下常见假设[2,16,26].

假设3 $\|\nabla\varphi_c(S)\|\leqslant k_{\varphi_c},\|\nabla\varsigma_c(S)\|\leqslant k_{\varsigma_c},\|\varsigma_{\mathrm{HJB}}\|\leqslant k_{\varsigma_{\mathrm{HJB}}},\|G(S)R^{-1}G^{\mathrm{T}}(S)\|\leqslant k_1,G(S)\leqslant k_G$,其中$k_{\varphi_c}$、$k_{\varsigma_c}$、$k_{\varsigma_{\mathrm{HJB}}}$、$k_1$、$k_G$均为正常数.

定理1 对于系统(6),反馈控制律由式(20)给出,评判网络的权值更新律由式(28)给出. 在该反馈控制器的作用下,闭环系统状态和评判网络权值估计误差均为UUB.

证明 选取合适的Lyapunov函数

$$L(t)=\frac{1}{2\alpha_c}\tilde{\omega}_c^{\mathrm{T}}\tilde{\omega}_c+\frac{\alpha_s}{\alpha_c}L_2(S), \tag{32}$$

其中$L_2(S)$的定义见假设2.

对Lyapunov函数(32)求关于时间的导数,可得

$$\dot{L}(t)=\frac{1}{\alpha_c}\tilde{\omega}_c^{\mathrm{T}}\dot{\tilde{\omega}}_c+\frac{\alpha_s}{\alpha_c}(\nabla L_2(S))^{\mathrm{T}}\dot{S}. \tag{33}$$

将式(21)和(31)代入(33),可以将式(33)重写为

$$\begin{aligned}\dot{L}(t)=&\tilde{\omega}_c^{\mathrm{T}}\Big(-\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)F(S)-\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{V}(S)\tilde{\omega}_c+\\&\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{V}(S)\omega_c+\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\tilde{\omega}_c-\\&\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\omega_c-\varsigma_{\mathrm{HJB}}+\xi\tilde{\omega}_c^{\mathrm{T}}\varphi_c(S)\Big)\cdot\\&\Big(\nabla\varphi_c(S)F(S)-\frac{1}{2}\mathcal{V}(S)\omega_c+\frac{1}{2}\mathcal{V}(S)\tilde{\omega}_c+\\&\frac{1}{2}\mathcal{W}(S)\omega_c-\frac{1}{2}\mathcal{W}(S)\tilde{\omega}_c-\xi\varphi_c(S)\Big)-\\&\frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla L_2(S)+\\&\frac{\alpha_s}{\alpha_c}(\nabla L_2(S))^{\mathrm{T}}\dot{S}.\end{aligned} \tag{34}$$

结合式(21)与$\mathcal{V}(S)$的表达式,可得

$$\begin{aligned}&\tilde{\omega}_c^{\mathrm{T}}\Big(\nabla\varphi_c(S)F(S)-\frac{1}{2}\mathcal{V}(S)\omega_c+\frac{1}{2}\mathcal{V}(S)\tilde{\omega}_c\Big)=\\&\tilde{\omega}_c^{\mathrm{T}}(\nabla L_2(S))^{\mathrm{T}}\dot{S}.\end{aligned} \tag{35}$$

进一步可得

$$\begin{aligned}\dot{L}(t)=&-\Big(\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)\dot{S}-\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{V}(S)\tilde{\omega}_c-\xi\tilde{\omega}_c^{\mathrm{T}}\varphi_c(S)-\\&\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\tilde{\omega}_c+\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\omega_c+\varsigma_{\mathrm{HJB}}\Big)\cdot\\&\Big(\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)\dot{S}-\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\tilde{\omega}_c+\\&\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\omega_c-\xi\varphi_c(S)\Big)-\\&\frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla L_2(S)+\\&\frac{\alpha_s}{\alpha_c}(\nabla L_2(S))^{\mathrm{T}}\dot{S}.\end{aligned} \tag{36}$$

采用式(19)中的控制律$u^*(S)$,并考虑最优闭环系统$\dot{S}^*=F(S)+G(S)u^*(S)$,可推导

$$\begin{aligned}\dot{L}(t)=&-\Big(\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)\dot{S}^*+\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{V}(S)\tilde{\omega}_c-\xi\tilde{\omega}_c^{\mathrm{T}}\varphi_c(S)+\\&\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla\varsigma_c(S)-\\&\frac{1}{4}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\tilde{\omega}_c+\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\omega_c+\varsigma_{\mathrm{HJB}}\Big)\cdot\\&\Big(\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)\dot{S}^*+\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{V}(S)\tilde{\omega}_c+\\&\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla\varsigma_c(S)-\\&\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\tilde{\omega}_c+\frac{1}{2}\tilde{\omega}_c^{\mathrm{T}}\mathcal{W}(S)\omega_c-\xi\varphi_c(S)\Big)-\\&\frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla L_2(S)+\\&\frac{\alpha_s}{\alpha_c}(\nabla L_2(S))^{\mathrm{T}}\dot{S}.\end{aligned} \tag{37}$$

接着,将式(37)的所有项展开,进行相应的数学运算,并通过假设3中给定的有界条件,得到如下不等式:

$$\begin{aligned}\dot{L}(t)\leqslant&-k_2\|\tilde{\omega}_c\|^4+k_3\|\tilde{\omega}_c\|^2+k_4^2-\\&\frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathrm{T}}\nabla\varphi_c(S)G(S)R^{-1}G^{\mathrm{T}}(S)\nabla L_2(S)+\\&\frac{\alpha_s}{\alpha_c}(\nabla L_2(S))^{\mathrm{T}}\dot{S},\end{aligned} \tag{38}$$

其中k_2、k_3和k_4均为正常数. 考虑式(38)、假设2和假设3,进一步得到

$$\begin{aligned}\dot{L}(t)\leqslant&-k_2\|\tilde{\omega}_c\|^4+k_3\|\tilde{\omega}_c\|^2+k_4^2+\\&\frac{\alpha_s}{\alpha_c}(\nabla L_2(S))^{\mathrm{T}}(F(S)+G(S)u^*(S))+\\&\frac{\alpha_s}{2\alpha_c}(\nabla L_2(S))^{\mathrm{T}}G(S)R^{-1}G^{\mathrm{T}}(S)\nabla\varsigma_c(S)\leqslant\\&-k_2\|\tilde{\omega}_c\|^4+k_3\|\tilde{\omega}_c\|^2+k_4^2-\\&\frac{\alpha_s}{\alpha_c}\lambda_{\min}(\mathcal{M})\|\nabla L_2(S)\|^2+\end{aligned}$$

$$\frac{\alpha_s}{2\alpha_c}k_1k_{\varsigma_c}\|\nabla L_2(S)\| \leqslant -k_2\Big(\|\tilde{\omega}_c\|^2-\frac{k_3}{2k_2}\Big)^2+k_5-\frac{\alpha_s}{\alpha_c}\lambda_{\min}(\mathcal{M})\Big(\|\nabla L_2(S)\|-\frac{k_1k_{\varsigma_c}}{4\lambda_{\min}(\mathcal{M})}\Big)^2, \tag{39}$$

其中常数项为

$$k_5=\frac{k_3^2+4k_2k_4^2}{4k_2}+\frac{\alpha_sk_1^2k_{\varsigma_c}^2}{16\alpha_c\lambda_{\min}(\mathcal{M})}. \tag{40}$$

由此可知,当

$$\|\tilde{\omega}_c\|\geqslant\sqrt{\frac{k_3}{2k_2}+\sqrt{\frac{k_5}{k_2}}}\triangleq k_{\tilde{\omega}_c}, \tag{41}$$

或者

$$\|\nabla L_2(S)\|\geqslant\frac{k_1k_{\varsigma_c}}{4\lambda_{\min}(\mathcal{M})}+\sqrt{\frac{\alpha_ck_5}{\alpha_s\lambda_{\min}(\mathcal{M})}}\triangleq k_{L_2} \tag{42}$$

成立时, $\dot{L}(t)<0$. 根据Lyapunov理论可以得出闭环系统状态和权值估计误差均为UUB. □

根据定理1的结论可以得出以下推论,并证明了近似控制律函数的收敛性.

推论1　由式(20)推导出的近似控制律$\hat{u}^*(S)$收敛到具有有限界的最优反馈控制律$u^*(S)$的邻域内.

证明　考虑式(19)和(20)表示的控制律,可得出

$$\|u^*(S)-\hat{u}^*(S)\|=-\frac{1}{2}R^{-1}G^{\mathrm{T}}(S)[(\nabla\varphi_c(S))^{\mathrm{T}}\tilde{\omega}_c+\nabla\varsigma_c(S)]. \tag{43}$$

根据定理1得出结论$\|\tilde{\omega}_c\|<k_{\tilde{\omega}_c}$. 然后,结合假设3,可以确定存在一个有限界$k_u$使下式成立:

$$\|u^*(S)-\hat{u}^*(S)\|\leqslant\frac{1}{2}\|R^{-1}\|k_G(k_{\varphi_c}k_{\tilde{\omega}_c}+k_{\varsigma_c})\triangleq k_u. \tag{44}$$

由此推论1得证. □

2.3　鲁棒跟踪性能分析

接下来将证明不确定增广系统的跟踪误差是稳定的.

定理2　考虑标称系统(6)、代价函数(7),以及附加稳定项(14),反馈控制器(20)可以保证系统(4)在闭环形式下的跟踪误差为UUB.

证明　选取最优代价函数$J^*(S)$为Lyapunov函数,并根据式(12)可得

$$(\nabla J^*(S))^{\mathrm{T}}G(S)=-2u^{*\mathrm{T}}(S)R. \tag{45}$$

结合式(13),可推导

$$(\nabla J^*(S))^{\mathrm{T}}F(S)=-\Psi(S)-U(S,u^*(S))+\xi J^*(S)-(\nabla J^*(S))^{\mathrm{T}}G(S)u^*(S)=-\Psi(S)+u^{*\mathrm{T}}(S)Ru^*(S)-S^{\mathrm{T}}Q_SS+\xi J^*(S). \tag{46}$$

基于式(20)和(46),将近似最优控制律应用于不确定增广系统(4),得到最优代价函数的时间导数为

$$\begin{aligned}\dot{J}^*(S)=&(\nabla J^*(S))^{\mathrm{T}}[F(S)+G(S)\hat{u}^*(S)+\Delta F(S)]=\\&-\Psi(S)-S^{\mathrm{T}}Q_SS+u^{*\mathrm{T}}(S)Ru^*(S)+\xi J^*(S)-\\&2u^{*\mathrm{T}}(S)R\hat{u}^*(S)+(\nabla J^*(S))^{\mathrm{T}}\Delta F(S)=\\&-S^{\mathrm{T}}Q_SS+u^{*\mathrm{T}}(S)Ru^*(S)-2u^{*\mathrm{T}}(S)R\hat{u}^*(S)-\\&\Big[\frac{1}{2}\nabla J^*(S)-\Delta F(S)\Big]^{\mathrm{T}}\Big[\frac{1}{2}\nabla J^*(S)-\Delta F(S)\Big]-\\&k_M^2(S)+(\Delta F(S))^{\mathrm{T}}\Delta F(S)+\xi J^*(S)\leqslant\\&-S^{\mathrm{T}}Q_SS+u^{*\mathrm{T}}(S)Ru^*(S)-2u^{*\mathrm{T}}(S)R\hat{u}^*(S)-\\&[k_M^2(S)-(\Delta F(S))^{\mathrm{T}}\Delta F(S)]+\xi J^*(S).\end{aligned} \tag{47}$$

根据矩阵Q_S的结构可以得到$S^{\mathrm{T}}Q_SS=\eta^{\mathrm{T}}Q\eta$. 那么,式(47)中与控制相关的项可被写为

$$u^{*\mathrm{T}}(S)Ru^*(S)-2u^{*\mathrm{T}}(S)R\hat{u}^*(S)\leqslant[u^*(S)-\hat{u}^*(S)]^{\mathrm{T}}R[u^*(S)-\hat{u}^*(S)]. \tag{48}$$

由于$\|\Delta F(S)\|\leqslant k_M(S)$,而最优代价函数的上界是一个大于零的常数$k_{J^*}$,再结合式(48)可以将式(47)写为

$$\begin{aligned}\dot{J}^*(S)\leqslant&[u^*(S)-\hat{u}^*(S)]^{\mathrm{T}}R[u^*(S)-\hat{u}^*(S)]-\\&\eta^{\mathrm{T}}Q\eta+\xi J^*(S)\leqslant\\&\lambda_{\max}(R)k_u^2-\lambda_{\min}(Q)\|\eta\|^2+\xi k_{J^*}.\end{aligned} \tag{49}$$

其中: $\lambda_{\min}(Q)$为矩阵Q的最小特征值, $\lambda_{\max}(R)$为矩阵R的最大特征值.

由式(49)可以得出,当$\eta(t)$的取值在集合

$$\Omega_\eta=\Big\{\eta:\|\eta\|\leqslant\sqrt{\frac{\lambda_{\max}(R)k_u^2+\xi k_{J^*}}{\lambda_{\min}(Q)}}\triangleq k_\eta\Big\} \tag{50}$$

之外时,存在$\dot{J}^*(S)<0$. 其中k_η为一个正常数. 通过上述推导得到在近似最优控制器(20)的作用下,闭环不确定增广系统的跟踪误差为UUB. □

由此可以得出结论:对于给定的参考系统(2),原不确定非线性系统(1)实现了鲁棒轨迹跟踪. 为了清晰起见,给出鲁棒轨迹跟踪控制算法的整个设计过程,主要步骤如下.

step 1: 考虑非线性系统(1)和参考系统(2),通过跟踪误差和期望信号构造一个增广系统(4),将鲁棒轨迹跟踪问题转化为镇定设计;

step 2: 对于标称系统(6),采用特殊效用项(14)并定义带有折扣因子的代价函数(7),将鲁棒镇定问题转化为非线性最优控制问题;

step 3: 通过构建单评判网络(17)逼近最优代价函数,实现了最优控制算法;

step 4: 选择合适的学习率α_c和α_s,采用权值更新律(28)训练评判网络的权值,经过充足的学习阶段后,收敛权值保持不变;

step 5: 将近似最优控制律(20)应用于系统(4),可获得良好的鲁棒跟踪效果;

step 6: 结束设计步骤.

注 2 由于不能直接求解非线性最优控制律(12),本文设计了基于单评判网络的近似最优控制律(20). 将式(12)和(20)用于讨论轨迹跟踪的鲁棒镇定问题,所提算法的优势在于利用式(20)得到轨迹跟踪问题的近似解,克服了难以通过式(12)求其最优值的缺点.

3 仿 真

考虑如下形式的连续时间不确定非线性系统:

$$\dot{x}=\begin{bmatrix}-x_1+0.5x_2\\ -0.5(x_1+x_2)-0.5x_2\cos^2(x_1)\end{bmatrix}+\begin{bmatrix}0\\1\end{bmatrix}u+\begin{bmatrix}\rho_1 x_1\sin(x_2)\\ \rho_2 x_2\cos(x_1^2)\end{bmatrix}. \tag{51}$$

其中: $x=[x_1,x_2]^{\mathrm{T}}\in\boldsymbol{R}^2$为状态变量, $u\in\boldsymbol{R}$为控制变量,参数ρ_1、$\rho_2\in[-1,1]$. 系统(51)的最后一项为不确定项,其上界为$k_M(x)=\sqrt{x_1^2+x_2^2}$,即$\|\Delta f(x)\|\leqslant\|x\|$. 效用函数中的基本项选择为$U(x,u)=x^{\mathrm{T}}Qx+u^{\mathrm{T}}Ru$,令$Q=I_2,R=I$. 令系统的初始状态向量为$x_0=[1.5,-2]^{\mathrm{T}}$. 参考系统可设定为

$$\dot{x}_d=\begin{bmatrix}x_{d2}\\ -x_{d1}\end{bmatrix}. \tag{52}$$

其中: $x_d=[x_{d1},x_{d2}]^{\mathrm{T}}\in\boldsymbol{R}^2$为参考状态,令初始参考状态为$x_{d0}=[0.5,-0.5]^{\mathrm{T}}$.

定义增广状态为$S=[\eta^{\mathrm{T}},x_d^{\mathrm{T}}]^{\mathrm{T}}$,结合非线性系统(51)与参考系统(52),可以将增广系统方程写为

$$\dot{S}=\begin{bmatrix}-(S_1+S_3)+0.5(S_2-S_4)\\ -0.5(S_1+S_2-S_3+S_4)-0.5(S_2+S_4)\cos^2(S_1+S_3)\\ S_4\\ -S_3\end{bmatrix}+\begin{bmatrix}0\\1\\0\\0\end{bmatrix}u+\begin{bmatrix}\rho_1(S_1+S_3)\sin(S_2+S_4)\\ \rho_2(S_2+S_4)\cos(S_1+S_3)^2\\ 0\\ 0\end{bmatrix}. \tag{53}$$

其中: $S=[S_1,S_2,S_3,S_4]^{\mathrm{T}}\in\boldsymbol{R}^4$, $S_1=\eta_1$, $S_2=\eta_2$, $S_3=x_{d1}$, $S_4=x_{d2}$. 不确定项的上界为$k_M(S)=\sqrt{(S_1+S_3)^2+(S_2+S_4)^2}$,即$\|\Delta F(S)\|\leqslant\|S\|$. 初始跟踪误差向量为$\eta_0=x_0-x_{d0}=[1,-1.5]^{\mathrm{T}}$. 那么,增广系统的初始状态向量为$S_0=[1,-1.5,0.5,-0.5]^{\mathrm{T}}$. 令$Q_S=\mathrm{diag}\{I_2,0_{2\times2}\}$, $R=I$, $\xi=0.2$. 在本例中,选取$[S_1^2,S_2^2,S_1^2S_2^2,S_1^2S_3^2,S_1^2S_4^2,S_2^2S_3^2,S_2^2S_4^2]^{\mathrm{T}}$作为评判网络的激活函数,接着构造一个评判神经网络来近似最优代价函数,即

$$\hat{J}^*(s)=\hat{\omega}_{c1}S_1^2+\hat{\omega}_{c2}S_2^2+\hat{\omega}_{c3}S_1^2S_2^2+\hat{\omega}_{c4}S_1^2S_3^2+\hat{\omega}_{c5}S_1^2S_4^2+\hat{\omega}_{c6}S_2^2S_3^2+\hat{\omega}_{c7}S_2^2S_4^2. \tag{54}$$

在仿真过程中,选择权值更新的学习率$\alpha_c=1.7$和$\alpha_s=0.03$,并引入探测噪声来满足持续激励条件. 由于不需要初始允许控制条件,将评判网络权值向量的初始值均设置为零. 经过$t=300\,\mathrm{s}$的学习训练过程,最终评判网络权值向量收敛为

$$\hat{\omega}_c=[0.752\,1,-0.008\,9,-0.131\,2,1.132\,9,-0.826\,2,0.442\,7,-0.595\,1]^{\mathrm{T}}. \tag{55}$$

评判网络权值收敛曲线如图1所示.

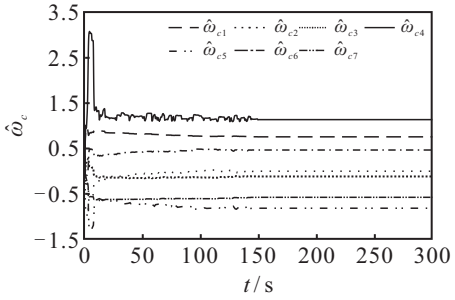

图 1 评判网络权值收敛曲线

根据图1评判网络权值收敛曲线可以看出,在该学习阶段获得了预期的收敛效果,进而验证了上述提出的鲁棒跟踪控制策略的有效性.

下面利用近似最优控制器(20)验证鲁棒跟踪性能. 对于增广系统(53)的不确定性部分,选择参数$\rho_1=-0.5$和$\rho_2=0.5$,将控制器应用于增广不确定系统(53),得到跟踪误差和相应的控制输入轨迹分别如图2和图3所示.

为了进行比较,另外选择参数$\rho_1=1$和$\rho_2=-1$,然后再次进行仿真验证,得到跟踪误差和相应的控制

输入轨迹分别如图4和图5所示.

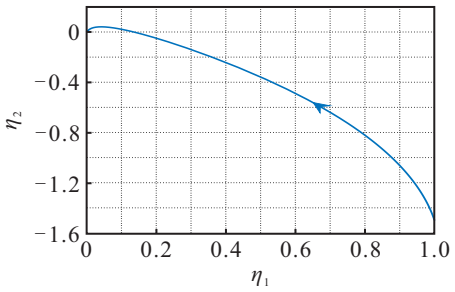

图 2　跟踪误差轨迹($\rho_1 = -0.5$且$\rho_2 = 0.5$)

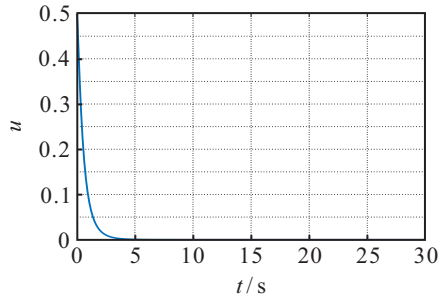

图 3　跟踪控制轨迹($\rho_1 = -0.5$且$\rho_2 = 0.5$)

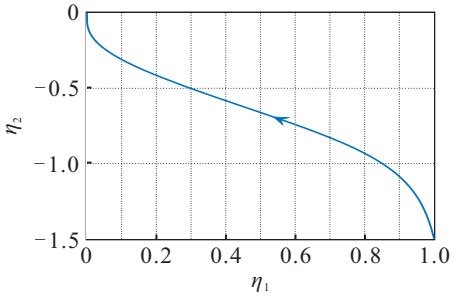

图 4　跟踪误差轨迹($\rho_1 = 1$且$\rho_2 = -1$)

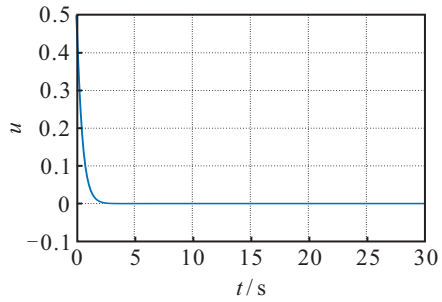

图 5　跟踪控制轨迹($\rho_1 = 1$且$\rho_2 = -1$)

由图4可以看出,跟踪误差同样有明显的稳定趋势,进一步验证了鲁棒跟踪控制器的良好性能.

注3　从上述仿真结果研究中发现,本文所提方法与文献[25-26]相比,更适用于包含一般不确定性的复杂非线性系统. 同时,与文献[30]相比,降低了初始容许控制律的要求,在本文中,初始权值只需设置为零,使算法的实现过程变得简单方便. 此外,与文献[16,23]相比,定义了具有折扣因子的代价函数并改进了评判网络权值更新律,保证了代价函数的有界性,同时,使权值收敛效果较好. 因此,本文所提方法具有一定的优越性.

4　结　论

本文主要研究了一种基于神经网络的自适应评判学习策略,并将其应用于一类不确定非线性系统的鲁棒跟踪控制问题. 首先,采用构造增广系统和改进传统代价函数的方法进行问题转换;其次,通过构造单评判网络近似求解与增广系统相关的HJB方程,并推导出系统的最优控制律;再次,在评判网络的权值更新律中引入额外项,消除了所提算法中的初始容许控制条件;进一步,利用Lyapunov理论分析了闭环系统的稳定性;最后,通过典型非线性系统的仿真实例,验证了该算法具有良好的跟踪控制性能. 在未来工作中,将研究系统模型未知情况下,通过放松持续性激励条件解决最优跟踪问题.

参考文献(References)

[1] Liu D, Wei Q, Wang D, et al. Adaptive dynamic pogramming with applications in optimal control[M]. Advances in Industrical Control. Switzerland: Springer, 2017: 223-264.

[2] 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.

[3] Powell W B. Approximate dynamic programming[M]. Approximaee Dynamic Programming: Solving the Curses of Dimensionality. Hoboken: John Wiley & Sons, Inc., 2011: 253-270.

[4] Wang D, Ha M M, Zhao M M. The intelligent critic framework for advanced optimal control[J]. Artificial Intelligence Review, 2022, 55(1): 1-22.

[5] 王鼎, 赵明明, 哈明鸣, 等. 基于折扣广义值迭代的智能最优跟踪及应用验证[J]. 自动化学报, 2022, 48(1): 182-193.
(Wang D, Zhao M M, Ha M M, et al. Intelligent optimal tracking with application verifications via discounted generalized value iteration[J]. Acta Automatica Sinica, 2022, 48(1): 182-193.)

[6] 王鼎. 一类离散动态系统基于事件的迭代神经控制[J]. 工程科学学报, 2022, 44(3): 411-419.
(Wang D. Event-based iterative neural control for a type of discrete dynamic plant[J]. Chinese Journal of Engineering, 2022, 44(3): 411-419.)

[7] 林小峰, 丁强. 基于评价网络近似误差的自适应动态规划优化控制[J]. 控制与决策, 2015, 30(3): 495-499.
(Lin X F, Ding Q. Adaptive dynamic programming optimal control based on approximation error of critic network[J]. Control and Decision, 2015, 30(3): 495-499.)

[8] Xue S, Luo B, Liu D R, et al. Event-triggered ADP for tracking control of partially unknown constrained

uncertain systems[J]. IEEE Transactions on Cybernetics, 2022, 52(9): 9001-9012.

[9] Wang D, Qiao J F, Cheng L. An approximate neuro-optimal solution of discounted guaranteed cost control design[J]. IEEE Transactions on Cybernetics, 2022, 52(1): 77-86.

[10] Lewis F L, Vrabie D. Reinforcement learning and adaptive dynamic programming for feedback control[J]. IEEE Circuits and Systems Magazine, 2009, 9(3): 32-50.

[11] Wang D, Mu C. Adaptive critic control with robust stabilization for uncertain nonlinear systems[M]. Studies in Systems, Decision and Control. Singapore: Springer, 2019: 1-43.

[12] Lin F, Brandt R D, Sun J. Robust control of nonlinear systems: Compensating for uncertainty[J]. International Journal of Control, 1992, 56(6): 1453-1459.

[13] Huang Y Z. Optimal guaranteed cost control of uncertain non-linear systems using adaptive dynamic programming with concurrent learning[J]. IET Control Theory & Applications, 2018, 12(8): 1025-1035.

[14] Fan Q Y, Yang G H. Adaptive actor–critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(1): 165-177.

[15] Jiang H, Zhang H G, Cui Y, et al. Robust control scheme for a class of uncertain nonlinear systems with completely unknown dynamics using data-driven reinforcement learning method[J]. Neurocomputing, 2018, 273: 68-77.

[16] Wang D, Liu D R, Mu C X, et al. Neural network learning and robust stabilization of nonlinear systems with dynamic uncertainties[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(4): 1342-1351.

[17] Wang D. Robust policy learning control of nonlinear plants with case studies for a power system application[J]. IEEE Transactions on Industrial Informatics, 2020, 16(3): 1733-1741.

[18] Zhao J, Na J, Gao G B. Adaptive dynamic programming based robust control of nonlinear systems with unmatched uncertainties[J]. Neurocomputing, 2020, 395: 56-65.

[19] 季政, 楼旭阳, 吴炜. 基于神经动态优化的非线性系统近似最优跟踪控制[J]. 控制与决策, 2021, 36(1): 97-104.
(Ji Z, Lou X Y, Wu W. Approximate optimal tracking control for nonlinear systems based on neurodynamic optimization[J]. Control and Decision, 2021, 36(1): 97-104.)

[20] Zou W C, Shi P, Xiang Z R, et al. Consensus tracking control of switched stochastic nonlinear multiagent systems via event-triggered strategy[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(3): 1036-1045.

[21] Xu D G, Wang Q L, Li Y. Optimal guaranteed cost tracking of uncertain nonlinear systems using adaptive dynamic programming with concurrent learning[J]. International Journal of Control, Automation and Systems, 2020, 18(5): 1116-1127.

[22] 王鼎. 基于学习的鲁棒自适应评判控制研究进展[J]. 自动化学报, 2019, 45(6): 1031-1043.
(Wang D. Research progress on learning-based robust adaptive critic control[J]. Acta Automatica Sinica, 2019, 45(6): 1031-1043.)

[23] Wang D, Cheng L, Yan J. Self-learning robust control synthesis and trajectory tracking of uncertain dynamics[J]. IEEE Transactions on Cybernetics, 2022, 52(1): 278-286.

[24] Modares H, Lewis F L. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning[J]. Automatica, 2014, 50(7): 1780-1792.

[25] Hou J X, Wang D, Liu D R, et al. Model-free H_∞ optimal tracking control of constrained nonlinear systems via an iterative adaptive learning algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(11): 4097-4108.

[26] Wang D, Liu D R, Zhang Y, et al. Neural network robust tracking control with adaptive critic framework for uncertain nonlinear systems[J]. Neural Networks, 2018, 97: 11-18.

[27] Mu C X, Zhang Y, Gao Z K, et al. ADP-based robust tracking control for a class of nonlinear systems with unmatched uncertainties[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(11): 4056-4067.

[28] Cui X, Zhang H, Luo Y, et al. Adaptive dynamic programming for tracking design of uncertain nonlinear systems with disturbances and input constraints[J]. International Journal of Adaptive Control and Signal Processing, 2017, 31(11): 1567–1583.

[29] Dierks T, Jagannathan S. Optimal control of affine nonlinear continuous-time systems[C]. Proceedings of the 2010 American Control Conference. Baltimore, 2010: 1568-1573.

[30] 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.

作者简介

霍煜 (1993–), 女, 博士生, 从事自适应动态规划、智能控制等研究, E-mail: HuoYu@emails.bjut.edu.cn;

王鼎 (1984–), 男, 教授, 博士生导师, 从事自适应评判控制、强化学习等研究, E-mail: dingwang@bjut.edu.cn;

乔俊飞 (1968–), 男, 教授, 博士生导师, 从事智能控制与智能信息处理、复杂过程建模与优化控制等研究, E-mail: adqiao@bjut.edu.cn.