第 33 卷第 1 期
2016 年 1 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 33 No. 1
Jan. 2016

DOI: 10.7641/CTA.2016.40963

针对时变轨迹的非线性仿射系统的鲁棒近似最优跟踪控制

屈秋霞, 罗艳红†, 张化光

(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

摘要: 针对非线性连续系统难以跟踪时变轨迹的问题, 本文首先通过系统变换引入新的状态变量从而将非线性系统的最优跟踪问题转化为一般非线性时不变系统的最优控制问题, 并基于近似动态规划算法(ADP)获得近似最优值函数与最优控制策略. 为有效地实现该算法, 本文利用评价网与执行网来估计值函数及相应的控制策略, 并且在线更新二者. 为了消除神经网络近似过程中产生的误差, 本文在设计控制器时增加一个鲁棒项; 并且通过Lyapunov稳定性定理来证明本文提出的控制策略可保证系统跟踪误差渐近收敛到零, 同时也验证在较小的误差范围内, 该控制策略能够接近于最优控制策略. 最后给出两个时变跟踪轨迹实例来证明该方法的可行性与有效性.

关键词: 非线性仿射系统; 时变轨迹; 最优控制; 跟踪问题; 渐近稳定

中图分类号: TP273　　　**文献标识码**: A

Robust approximate optimal tracking control of time-varying trajectory for nonlinear affine systems

QU Qiu-xia, LUO Yan-hong†, ZHANG Hua-guang

(School of Information Science and Engineering, Northeastern University, Shenyang Liaoning 110819, China)

Abstract: For continuous time nonlinear systems, it is difficult to track their time-varying trajectory. To deal with this problem, we use a system transformation to introduce a new state variable for converting the optimal tracking problem of nonlinear systems into optimal control problem of general nonlinear time-invariant systems. For this system, we obtain the approximate optimal value function and the approximate optimal control policy based on approximate dynamic programming (ADP). Then, we use the critic network and the actor network to estimate the value function and the corresponding control strategy, and update both of them online. Besides, a robust control term is added to the controller to eliminate the residual errors generated in the process of neural network approximation. By using the Lyapunov stability theorem, we prove that the proposed control strategy can guarantee the tracking error to converge asymptotically to zero, and the control strategy is close to the optimal control strategy when the error is in a small bound. Finally, simulations of two time-varying trajectory tracking examples show the feasibility and effectiveness of the proposed method.

Key words: nonlinear affine systems; time-varying trajectory; optimal control; tracking problem; asymptotic stability

1　引言(Introduction)

许多实际控制系统, 比如航空、轮船、机械传动及工业生产过程等, 都对最优跟踪控制器的设计有很大需求. 专家学者近几十年来提出的方案大都是将求解最优跟踪问题转化为求解最优控制问题[1–7]. 多年来的研究成果表明, 近似动态规划方法(approximate dynamic programming, ADP)是有效解决最优控制问题的方法之一[8–10], 其基本思想就是在ADP算法中使用神经网络作为逼近器来近似动态规划方程中的性能指标函数和控制策略, 从而获得最优性能指标函数和最优控制[11–12].

针对无限时域内连续系统的跟踪问题, 跟踪误差以及性能指标函数一般都是关于状态和时间的函数, 文献[3, 13]都是通过构造近似值函数解决连续系统的跟踪问题, 但文中值函数以及控制器都是关于跟踪误差在无限时域内的时变函数, 而用神经网络作为逼近器只能近似紧集上的函数[14]. 到目前为止, 还没有足够的条件说明如何仅将跟踪误差作为输入就能近似

收稿日期: 2014−10−15; 录用日期: 2015−06−08.
†通信作者. E-mail:neuluo@gmail.com; Tel.: +86 13998827616.
本文责任编委: 吴敏.
国家自然科学基金项目(61273029, 61273027), 辽宁省自然科学基金(2013020037), 高等学校博士学科点专项科研基金(20110042120032), 中央高校基本科研基金项目(N130504004, N140404004)资助.
Supported by National Natural Science Foundation of China (61273029, 61273027), Natural Science Foundation of Liaoning (2013020037), Special Research Found for the Doctoral Program of Higher Education (20110042120032) and Fundamental Research Funds for the Central Universities (N130504004, N140404004).

无限时域内的时变值函数和控制器.

对于离散系统的跟踪问题, Dierks基于ADP方法直接求解无限时间最优跟踪问题[6], 但文中参考输入依赖于系统状态, 在实际应用中有很大局限性. Wei[7]等将原系统的跟踪问题转化为最优调节问题, 利用贪婪迭代算法 (HDP) 得到了跟踪问题的最优解. 然而, 这些成果或者缺少相应的稳定性分析, 或者只能保证跟踪误差和相关参数的一致最终有界.

与其他关于非线性连续系统跟踪控制器的设计相比, 本文将跟踪误差以及参考输入同时作为状态变量, 提出了一个新的非线性连续系统的跟踪控制器设计方法, 并需要求解相应的HJB方程. 之前该方程的解, 即值函数是关于跟踪误差的时变函数(从本文的仿真部分也可得到此结论), 是不能用神经网络作为逼近器来近似求解的. 本文的创新就是在原HJB方程中加入值函数对参考输入的偏微分, 使得值函数转变为关于新构造系统状态变量的函数, 而不显含时间t, 故可用神经网络对其近似, 从而可将时变最优跟踪问题转变为时不变最优调节问题, 然后运用策略迭代算法求解最优调节问题. 此外, 本文给出了该算法的稳定性分析证明. 为了保证跟踪误差渐近收敛到零, 本文在控制器上添加鲁棒项, 通过Lyapunov稳定性分析可证明该鲁棒近似最优控制器可使跟踪误差渐近收敛到零, 且使该鲁棒近似控制器与最优控制器存在非常小的误差范围[15]. 特别在稳定性分析中, 本文通过值函数两种等价形式的相互替换, 既证明了新的值函数是Lyapunov备选函数, 同时又证明了算法的稳定性. 在算法的实现上, 通过选用相互独立的两组基函数来设计评价网与执行网, 降低了两个逼近器相互耦合可能产生的训练复杂度, 然后在线同步更新评价网与执行网来得到近似最优值函数与近似最优控制策略.

2 问题描述(Problem formulation)

考虑如下形式的非线性连续仿射系统:

$$\dot{x} = f(x) + g(x)u, \tag{1}$$

其中: $x \in \chi \subset \mathbb{R}^n$为系统的状态变量, $u \subset \mathbb{R}^m$, $f: \chi \to \mathbb{R}^n$, $g: \chi \to \mathbb{R}^{n\times m}$均满足局部李普希茨(Lipschitz)连续条件, 且$f(0) = 0$. 本文目的是跟踪有界连续可微信号$x_{\mathrm{d}}: \mathbb{R}_+ \to \chi$. 定义跟踪误差为$e \triangleq x - x_{\mathrm{d}}$, 则系统的跟踪误差动态方程为

$$\dot{e} = f(x) + g(x)u - \dot{x}_{\mathrm{d}}. \tag{2}$$

在以往研究中, 如文献[3], 针对跟踪系统(2)的性能指标函数定义为

$$V(e) \triangleq \int_t^{\infty} r(e(s), \mu(e(s)))\mathrm{d}s, \tag{3}$$

其中: $r((e),\ \mu(e)) = \mathrm{e}^{\mathrm{T}}\, Qe + \mu^{\mathrm{T}}(e)R\mu(e)$作为效用函数, Q 与R 都是对称正定常数矩阵. 由于跟踪信号x_{d}是一个时变轨迹, 从跟踪误差方程(2)可知误差e是包含x_{d}的, 故误差e与值函数$V(e)$也是时变的, 而在无穷时域内的时变值函数是无法由神经网络逼近器来近似的, 因此本文的主要研究内容是如何引入新的状态变量来将时变值函数转变为时不变值函数, 并用近似动态规划算法来得到非线性系统的最优控制器. 为便于设计最优跟踪控制器, 现做以下假设:

假设 1 存在局部李普希茨函数$h_{\mathrm{d}}: \chi \to \mathbb{R}^n$使得$\dot{x}_{\mathrm{d}} = h_{\mathrm{d}}(x_{\mathrm{d}})$, 并且$h_{\mathrm{d}}(0) = 0$.

假设 2 函数f, g, h_{d}有界, 且g列满秩, 即存在正常数$b_{\mathrm{f}}, b_{\mathrm{g}}$及$b_{\mathrm{h}}$使得$\|f\| \leqslant b_{\mathrm{f}}$, $\|g\| \leqslant b_{\mathrm{g}}$, $\|h_d\| \leqslant b_{\mathrm{h}}$.

以上假设并不是特殊的苛刻要求, 而是便于控制器设计的合理性假设, 在关于最优跟踪控制器设计的参考文献中均有提到[13–14, 16].

令$g^+ \triangleq (g^{\mathrm{T}}g)^{-1}g^{\mathrm{T}}$, g^+有界并且局部李普希茨; $g^+g = I_{m\times m}$, 其中$I_{m\times m} \in \mathbb{R}_{m\times m}$是单位方阵. 关于参考输入的稳态控制策略$u_{\mathrm{d}}: \chi \to \mathbb{R}^m$可表示为

$$u_{\mathrm{d}}(x_{\mathrm{d}}) = g^+(x_{\mathrm{d}})(h_{\mathrm{d}}(x_{\mathrm{d}}) - f(x_{\mathrm{d}})). \tag{4}$$

本文引入新的状态变量$\eta \in \chi \times \chi \subset \mathbb{R}^{2n}$, $\eta \triangleq [\mathrm{e}^{\mathrm{T}}\ \ x_{\mathrm{d}}{}^{\mathrm{T}}]^{\mathrm{T}}$, 根据式(2)和假设1可得新变量的状态微分方程为

$$\dot{\eta} = F(\eta) + G(\eta)\mu, \tag{5}$$

式中: $F: \chi \times \chi \to \mathbb{R}^{2n}$, $G: \chi \times \chi \to \mathbb{R}^{2n\times m}$, 以及控制$\mu \in \mathbb{R}^m$满足:

$$F \triangleq \begin{bmatrix} f(e + x_{\mathrm{d}}) - h_{\mathrm{d}}(x_{\mathrm{d}}) + g(e + x_{\mathrm{d}})u_{\mathrm{d}}(x_{\mathrm{d}}) \\ h_{\mathrm{d}}(x_{\mathrm{d}}) \end{bmatrix},$$

$$G \triangleq \begin{bmatrix} g(e + x_{\mathrm{d}}) \\ 0_{n\times m} \end{bmatrix},$$

$0_{n\times m} \in \mathbb{R}^{n\times m}$是零矩阵. $\mu = u - u_{\mathrm{d}}$. 显然, $F(\eta)$也是局部李普希茨的, 且$F(0) = 0$.

本文的最优控制问题是找到一个容许控制策略[10]使下面值函数V最小:

$$V(\eta, \mu) \triangleq \int_t^{\infty} r(\eta(s), \mu(s))\mathrm{d}s, \tag{6}$$

其中$r: \chi \times \chi \to \mathbb{R}_+$作为效用函数, 定义为

$$r(\eta) = \eta^{\mathrm{T}}\bar{Q}\eta + \mu^{\mathrm{T}}(\eta)R\mu(\eta), \tag{7}$$

$R \in \mathbb{R}^{m\times m}$是正定对称常数矩阵. $\bar{Q} \in \mathbb{R}^{2n\times 2n}$,

$$\bar{Q} \triangleq \begin{bmatrix} Q & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{bmatrix},$$

其中: $Q \in \mathbb{R}^{n\times n}$是对称正定常数矩阵, 最小特征值为$\lambda_{\min}\{Q\}$, $0_{n\times n} \in \mathbb{R}^{n\times n}$是零矩阵.

系统(5)与系统(2)相比, 显然系统(5)不再显含时间t, 从而将求解时变值函数(3)的跟踪问题转变为求

解时不变值函数(6)的最优调节问题. 接下来本文采用策略迭代算法来求解最优值函数.

3 近似最优跟踪控制器设计(Approximate optimal tracking controller design)

定义哈密尔顿函数如下:

$$H(\eta,\mu,V_\eta)=r(\eta,\mu)+{V_\eta}^{\mathrm{T}}(F(\eta)+G(\eta)\mu),\tag{8}$$

式中$V_\eta=\dfrac{\partial V(\eta)}{\partial\eta}$.

最小值函数$V^*(\eta)$定义如下:

$$V^*(\eta)=\min_{\mu\in\psi}\int_t^{\infty}r(\eta(s),\mu(\eta(s)))\mathrm{d}s,\tag{9}$$

式中ψ为容许控制集合. 那么, 若存在最优控制策略μ^*, 则$V^*(\eta)$满足如下HJB方程:

$$H(\eta,\mu^*,V_\eta^*)=r(\eta,\mu^*)+{V_\eta^*}^{\mathrm{T}}(F+G\mu^*)=0,\tag{10}$$

式中$V_\eta^*=\dfrac{\partial V^*(\eta)}{\partial\eta}$. 针对动态系统(5)和效用函数(7), 求解$\dfrac{\partial H(\eta,\mu^*,V_\eta^*)}{\partial\eta}=0$可得最优控制策略

$$\mu^*=-\frac{1}{2}R^{-1}G^{\mathrm{T}}V_\eta^*.\tag{11}$$

但由于HJB方程(10)关于V_η^*的求解是十分困难的, 本文将通过引入评价网与执行网的设计去得到该方程的近似解, 即本文需要的近似最优控制器.

3.1 评价网设计(Critic network design)

在紧集χ上, 值函数V^*可以用如下神经网络表示:

$$V^*(\eta)=W_1^{\mathrm{T}}\phi_1(\eta)+\varepsilon_1(\eta),\tag{12}$$

其中: 理想权值矩阵$W_1\in\mathbb{R}^N$, 连续可微激活函数$\phi_1:\chi\times\chi\to\mathbb{R}^N$, 函数重构误差$\varepsilon_1:\chi\times\chi\to\mathbb{R}$.

值函数关于系统(5)的微分方程为

$$V_\eta^*=\nabla\phi_1^{\mathrm{T}}W_1+\nabla\varepsilon_1,\tag{13}$$

式中: $\nabla\phi_1\triangleq\dfrac{\partial\phi_1}{\partial\eta}$, $\nabla\varepsilon_1\triangleq\dfrac{\partial\varepsilon_1}{\partial\eta}$.

$\hat W_1$作为W_1的神经网络逼近值, 那么值函数的逼近值为

$$\hat V(\eta)=\hat W_1^{\mathrm{T}}\phi_1(\eta).\tag{14}$$

哈密尔顿函数的逼近值可表示为

$$\hat H(\eta,\mu,\hat W_1)=\hat W_1^{\mathrm{T}}\nabla\phi_1(F+G\mu)+\eta^{\mathrm{T}}\bar Q\eta+\mu^{\mathrm{T}}R\mu=e_{\mathrm{H}}.\tag{15}$$

对任意给定容许控制μ, 选择$\hat W_1$使得残差的平方$E_1(\hat W_1)$ 最小, 其中$E_1(\hat W_1)=\frac{1}{2}e_{\mathrm{H}}^{\mathrm{T}}e_{\mathrm{H}}$.

对评价网的权值更新采用如下梯度下降算法:

$$\dot{\hat W}_1=-a_1\sigma_1(\sigma_1^{\mathrm{T}}\hat W_1+\eta^{\mathrm{T}}\bar Q\eta+\mu^{\mathrm{T}}R\mu),\tag{16}$$

式中: $a_1>0$为评价网的自适应增益, $\sigma=\nabla\phi_1(F+G\mu)$, $\sigma_1=\sigma/(\sigma^{\mathrm{T}}\sigma+1)$. 由$\sigma_1$定义可知$\|\sigma_1\|\leqslant\sigma_{1\mathrm{M}}$, $\sigma_{1\mathrm{M}}$是正常数.

定义权值估计误差为$\tilde W_1=W_1-\hat W_1$. 对于固定的控制策略μ, 哈密尔顿函数选为

$$H(\eta,\mu,W_1)=W_1^{\mathrm{T}}\nabla\phi_1(F+G\mu)+\eta^{\mathrm{T}}\bar Q\eta+\mu^{\mathrm{T}}R\mu=\varepsilon_{\mathrm{H}},\tag{17}$$

式中: ε_{H}为神经网络逼近过程产生的残差, $\varepsilon_{\mathrm{H}}=-\nabla\varepsilon_1(F+G\mu)$, 由此值函数的权值逼近误差:

$$\dot{\tilde W}_1=-a_1\sigma_1(\sigma_1^{\mathrm{T}}\tilde W_1+\varepsilon_{\mathrm{H}}).\tag{18}$$

3.2 执行网设计(Actor network design)

控制策略μ由神经网络逼近为

$$\mu=W_2^{\mathrm{T}}\phi_2(\eta)+\varepsilon_2(\eta),\tag{19}$$

其中: 理想权值矩阵$W_2\in\mathbb{R}^N$, 执行网激活函数$\phi_2:\chi\times\chi\to\mathbb{R}^N$, 神经网络逼近误差$\varepsilon_2:\chi\times\chi\to\mathbb{R}$.

$\hat W_2$作为W_2的神经网络逼近值, 则控制策略的逼近值

$$\hat\mu=\hat W_2^{\mathrm{T}}\phi_2(\eta).\tag{20}$$

反馈误差信号为

$$e_\mu=\hat W_2^{\mathrm{T}}\phi_2(\eta)+\frac{1}{2}R^{-1}G^{\mathrm{T}}\nabla\phi_1^{\mathrm{T}}\hat W_1.\tag{21}$$

接下来的设计是要使得此误差的平方最小, 即

$$E_2(\hat W_2)=\frac{1}{2}e_\mu^{\mathrm{T}}e_\mu.\tag{22}$$

执行网权值的更新仍采用梯度下降算法:

$$\dot{\hat W}_2=-a_2\phi_2(\hat W_2^{\mathrm{T}}\phi_2+\frac{1}{2}R^{-1}G^{\mathrm{T}}\nabla\phi_1^{\mathrm{T}}\hat W_1),\tag{23}$$

式中执行网的自适应增益$a_2>0$. 定义执行网的权值逼近误差为$\tilde W_2=W_2-\hat W_2$, 由于控制策略(19)使无限时间的性能指标函数(12)最小, 由式(11)可得

$$\hat W_2^{\mathrm{T}}\phi_2+\varepsilon_2+\frac{1}{2}R^{-1}G^{\mathrm{T}}(\nabla\phi_1^{\mathrm{T}}\hat W_1+\nabla\varepsilon_1)=0,\tag{24}$$

那么, 由式(23)–(24)可得控制策略的权值逼近误差:

$$\dot{\tilde W}_2=-a_2\phi_2(\tilde W_2^{\mathrm{T}}\phi_2+\frac{1}{2}R^{-1}G^{\mathrm{T}}\nabla\phi_1^{\mathrm{T}}\tilde W_1+\varepsilon')^{\mathrm{T}},\tag{25}$$

其中$\varepsilon'=\varepsilon_2+\dfrac{1}{2}R^{-1}G^{\mathrm{T}}\nabla\varepsilon_1$.

为了保证提供足够的状态值来训练评价网和执行网, 需要在训练过程中提供持续性激励, 例如本文在控制输入中加入一系列不同频率的正弦波信号.

3.3 稳定性分析(Stability analysis)

前面部分已经通过引入新的状态变量η将最优跟踪控制系统(2)转变为时不变最优控制系统(5), 从而能够通过评价网与执行网的设计求得最优值函数与最优控制策略. 在稳定性分析中, 我们需要用到值函数用于证明接下来的定理1. 由式(7)可知, $\bar Q$是半正定

函数, 因此不能直接说该值函数可作为Lyapunov候选函数. 但是对于值函数来说, 由于有两种相互等价的表现形式, 即$V^*(\eta,\mu)$ 与$V^*(e,t)$. 前者本文用于最优控制策略的求解，而后者本文用来证明此值函数是关于跟踪误差的正定且下降的函数, 即Lyapunov候选函数. 下面利用两个引理可证明之.

引理 1[16] 设$D\subseteq\mathbb{R}^n$是包含原点的定义域, $\xi: D\times\mathbb{R}_+\subseteq\mathbb{R}_+$是正定函数, 若任给有界变量$x$, $\xi(x,t)$都关于时间t一致有界, 那么$x\to\xi(x,t)$在时间t上一致连续, 且函数ξ在集合D上是下降的.

引理 2[16] $B_{\rm a}$为包含原点, 半径为$a\in\mathbb{R}_+$的封闭球域, 最优值函数$V^*:\chi\times\mathbb{R}_+\to\mathbb{R}$满足下列性质:

$$V^*(e,t)\geqslant\nu_1(\|e\|),\tag{26a}$$

$$V^*(0,t)=0,\ \forall t\in\mathbb{R}_+,\tag{26b}$$

$$V^*(e,t)\leqslant\nu_2(\|e\|).\tag{26c}$$

其中: $\forall t\in\mathbb{R}_+,\forall e\in B_{\rm a}$, $\upsilon_1:\ [0,\ a]\to\mathbb{R}_+$和$\upsilon_2:[0,\ a]\to\mathbb{R}_+$, 都属于$\kappa$类函数, 且$B_{\rm a}\in\chi$.

引理1与引理2的证明在文[16]可看到, 此处不再详细介绍.

由神经网络逼近器的一般特性，本文得到以下假设条件:

假设 3 1) 评价网与执行网的权值矩阵分别满足$\|W_1\|\leqslant W_{1\rm M}$, $\|W_2\|\leqslant W_{2\rm M}$, 激活函数分别满足$\|\phi_1\|\leqslant\phi_{1\rm M}$, $\|\phi_2\|\leqslant\phi_{2\rm M}$, 且$W_{1\rm M}$, $W_{2\rm M}$, $\phi_{1\rm M}$, $\phi_{2\rm M}$是正常数.

2) 权值逼近误差及误差的导数满足$\|\varepsilon_1\|\leqslant\varepsilon_{1\rm M}$, $\|\nabla\varepsilon_1\|\leqslant\varepsilon_{1\rm dM}$, $\|\varepsilon_2\|\leqslant\varepsilon_{2\rm M}$, 此外, HJB方程的残差满足$\|\varepsilon_{\rm H}\|\leqslant\varepsilon_{\rm HM}$, $\varepsilon_{1\rm M}$, $\varepsilon_{1\rm dM}$, $\varepsilon_{2\rm M}$, $\varepsilon_{\rm HM}$都是正常数.

注 1 由式(12)–(13)可知, 通过选择合适的基函数与神经元个数, 值函数及其导数可被一致逼近. 同时当神经元个数$N\to0$时, 逼近器误差$\varepsilon_1\to0$, 及其导数$\nabla\varepsilon_1\to0$[17]. 此外, 对于给定的神经元个数N, 逼近器误差ε_1, 及其导数$\nabla\varepsilon_1$在紧集上均有界[18]. 例如sigmoids函数, tanh函数以及其他常规神经网络激活函数[14].

本文利用以上引理与假设条件来证明以下定理并分析算法的稳定性.

定理 1 对于系统(5), 给定控制策略(20), 评价网和执行网的权值更新律分别为式(16)和(23). 若由执行网的初始权值能得到初始容许控制, 那么在如下不等式条件下, 跟踪误差e以及评价网与执行网的权值估计误差$\tilde W_1$和$\tilde W_2$能达到一致最终有界(UUB). 此外, 性能指标函数$\hat V$收敛到最优性能指标函数V^*的$\varepsilon_{\rm V}$邻域内, 控制输入$\hat\mu$收敛到最优控制输入μ^*的ε_μ邻域内, 即当$t\to\infty$时,

$$\|\hat V-V^*\|\leqslant\varepsilon_{\rm V},\|\hat\mu-\mu^*\|\leqslant\varepsilon_\mu.$$

其中a_1,a_2,Q,R满足

$$a_1<\frac{2\sigma_{1\rm m}^2}{\sigma_{1\rm M}^2}-\frac{1}{4a_2\sigma_{1\rm M}^2}b_{\rm g}^2\phi_{\rm dM}\|R^{-1}\|^2,\tag{27a}$$

$$a_2<1+\frac{\phi_{2\rm m}^2}{\phi_{2\rm M}^2},\tag{27b}$$

$$\lambda_{\min}(Q)>2+4b_{\rm g}^2,\tag{27c}$$

$$\lambda_{\min}(R)>1.\tag{27d}$$

证 选取如下Lyapunov函数:

$$L(t)=\frac12{\rm tr}\{\tilde W_1^{\rm T}{a_1}^{-1}\tilde W_1\}+\frac12{\rm tr}\{\tilde W_2^{\rm T}{a_2}^{-1}\tilde W_2\}+ e^{\rm T}e+V^*(e,\ t).\tag{28}$$

沿着系统(5)对Lyapunov函数(28)进行微分得

$$\dot L(t)={\rm tr}\{\tilde W_1^{\rm T}{a_1}^{-1}\dot{\tilde W}_1\}+{\rm tr}\{\tilde W_2^{\rm T}{a_2}^{-1}\dot{\tilde W}_2\}+ 2e^{\rm T}\dot e+V_\eta^{*\rm T}(F+G\mu^*).\tag{29}$$

通过反复运用$2ab\leqslant a^2+b^2$, 本文可得到如下不等式成立. 其中:

$$\begin{aligned}&{\rm tr}\{\tilde W_1^{\rm T}{a_1}^{-1}\dot{\tilde W}_1\}+{\rm tr}\{\tilde W_2^{\rm T}{a_2}^{-1}\dot{\tilde W}_2\}=\\&-\frac{1}{a_1}{\rm tr}\{a_1\tilde W_1^{\rm T}\sigma_1({\sigma_1}^{\rm T}\tilde W_1+\varepsilon_{\rm H})\}-\\&\frac{1}{a_2}{\rm tr}\{a_2\tilde W_2^{\rm T}\phi_2(\tilde W_2^{\rm T}\phi_2+\\&\frac12R^{-1}G^{\rm T}\nabla\phi_1^{\rm T}\tilde W_1+\varepsilon')^{\rm T}\}\leqslant\\&-(\sigma_{1\rm m}^2-\frac{a_1}{2}\sigma_{1\rm M}^2-\frac{1}{8a_2}b_{\rm g}^2\phi_{\rm dM}\|R^{-1}\|^2)\|\tilde W_1\|^2-\\&(\phi_{2\rm m}^2-a_2\phi_{2\rm M}^2)\|\tilde W_2\|^2+\frac{1}{2a_1}\varepsilon_{\rm HM}^2+\frac{1}{a_2}\varepsilon_{\rm M}'^2,\end{aligned}\tag{30}$$

$$\begin{aligned}&2e^{\rm T}\dot e+V_\eta^{*\rm T}(F+G\mu^*)=\\&2e^{\rm T}[f+gu_{\rm d}-h_{\rm d}+g(\mu^*-\varepsilon_2-\tilde W_2^{\rm T}\phi_2)]+\\&(-e^{\rm T}Qe-{\mu^*}^{\rm T}R\mu^*)\leqslant\\&-(-2-4b_{\rm g}^2+\lambda_{\min}(Q))\|e\|^2+b_{\rm f}^2+\|u_{\rm d}\|^2+b_{\rm h}^2+\\&\varepsilon_{2\rm M}^2+\phi_{2\rm M}^2\tilde W_2^2-(-1+\lambda_{\min}(R))\|\mu^*\|^2.\end{aligned}\tag{31}$$

由式(28)–(31)可得

$$\begin{aligned}\dot L(t)\leqslant&-(\sigma_{1\rm m}^2-\frac{1}{8a_2}b_{\rm g}^2\phi_{\rm dM}\|R^{-1}\|^2-\\&\frac{a_1}{2}\sigma_{1\rm M}^2)\|\tilde W_1\|^2-(\phi_{2\rm m}^2-a_2\phi_{2\rm M}^2+\\&\phi_{2\rm M}^2)\|\tilde W_2\|^2-(-1+\lambda_{\min}(R))\|\mu^*\|^2-\\&(-2-4b_{\rm g}^2+\lambda_{\min}(Q))\|e\|^2+b_{\rm M},\end{aligned}\tag{32}$$

其中

$$b_{\rm M}=\frac{1}{2a_1}\varepsilon_{\rm HM}^2+\frac{1}{a_2}\varepsilon_{\rm M}'^2+b_{\rm f}^2+\|u_{\rm d}\|^2+b_{\rm h}^2+\varepsilon_{2\rm M}^2.$$

选择满足不等式(27a)–(27d)的a_1, a_2, Q, R, 那么, 当

$$\|e\| > \sqrt{\frac{b_{\mathrm{M}}}{-2-4b_{\mathrm{g}}^2+\lambda_{\min}(Q)}} \triangleq l_{\mathrm{e}}, \tag{33}$$

或者

$$\|\tilde{W}_1\| > \sqrt{\frac{b_{\mathrm{M}}}{\sigma_{1\mathrm{m}}^2 - \frac{a_1}{2}\sigma_{1\mathrm{M}}^2 - \frac{1}{8a_2}b_{\mathrm{g}}^2\phi_{\mathrm{dM}}\|R^{-1}\|^2}} \triangleq l_{\tilde{W}_1}, \tag{34}$$

或者

$$\|\tilde{W}_2\| > \sqrt{\frac{b_{\mathrm{M}}}{\phi_{2\mathrm{m}}^2 - a_2\phi_{2\mathrm{M}}^2 + \phi_{2\mathrm{M}}^2}} \triangleq l_{\tilde{W}_2} \tag{35}$$

成立, 则$\dot{L}(t) < 0$. 根据Lyapunov理论可知跟踪误差e以及权值估计误差$\tilde{W}_1$和$\tilde{W}_2$是一致最终有界的(UUB).

接下来本文证明当$t \to \infty$时, $\|V^* - \hat{V}\| \leqslant \varepsilon_{\mathrm{V}}$, $\|\mu^* - \hat{\mu}\| \leqslant \varepsilon_\mu$.

由于

$$V^* - \hat{V} = \tilde{W}_1^{\mathrm{T}}\phi_1(\eta) + \varepsilon_1(\eta), \tag{36}$$

$$\mu^* - \hat{\mu} = \tilde{W}_2^{\mathrm{T}}\phi_2(\eta) + \varepsilon_2(\eta). \tag{37}$$

由式(34)–(35)可得, 当$t \to \infty$时, $\|\tilde{W}_1\| \leqslant l_{\tilde{W}_1}$, $\|\tilde{W}_2\| \leqslant l_{\tilde{W}_2}$. 那么, 结合假设3可得

$$\|V^* - \hat{V}\| \leqslant \|\tilde{W}_1^{\mathrm{T}}\phi_1(\eta)\| + \|\varepsilon_1(\eta)\| \leqslant \varepsilon_{\mathrm{V}}, \tag{38}$$

$$\|\mu^* - \hat{\mu}\| \leqslant \|\tilde{W}_2^{\mathrm{T}}\phi_2(\eta)\| + \|\varepsilon_2(\eta)\| \leqslant \varepsilon_\mu, \tag{39}$$

其中: $\varepsilon_{\mathrm{V}} = l_{\tilde{W}_1}\phi_{1\mathrm{M}} + \varepsilon_{1\mathrm{M}}$, $\varepsilon_\mu = l_{\tilde{W}_2}\phi_{2\mathrm{M}} + \varepsilon_{2\mathrm{M}}$.

证毕.

4　鲁棒最优跟踪控制器设计(Robust approximate optimal tracking controller design)

为了消除神经网络逼近过程中产生的误差ε_{V}和ε_μ(以及他们的微分)对系统的影响, 本文构造了此鲁棒项:

$$\mu_{\mathrm{r}} = \frac{K_{\mathrm{r}}e}{e^{\mathrm{T}}e+\beta}, \tag{40}$$

式中$\beta > 0$, $K_{\mathrm{r}} \in \mathbb{R}_+$满足

$$K_{\mathrm{r}} \geqslant \frac{b_{\mathrm{M}}(e^{\mathrm{T}}e+\beta)}{2\|g\|e^{\mathrm{T}}e}. \tag{41}$$

系统(5)的控制输入变为

$$\mu_{\mathrm{ad}} = \hat{\mu} - \mu_{\mathrm{r}}. \tag{42}$$

定理 2　对于系统(5), 给定控制(40), 评价网和执行网的权值更新律分别为式(16)和式(23). 假设执行网的初始权值能得到初始容许控制, 那么跟踪误差e以及权值估计误差$\tilde{W}_1$和$\tilde{W}_2$最终渐近收敛. 此外, 控制输入μ_{ad}收敛到最优控制输入μ^*的δ_μ邻域内, 即当$t \to \infty$时, $\|\mu_{\mathrm{ad}} - \mu^*\| \leqslant \delta_\mu$.

证　选择Lyapunov函数(28), 利用条件(40)–(41), 沿着系统(5)对其进行微分可得

$$\begin{aligned}\dot{L}(t) = &-\frac{1}{a_1}\mathrm{tr}\{a_1\tilde{W}_1^{\mathrm{T}}\sigma_1({\sigma_1}^{\mathrm{T}}\tilde{W}_1+\varepsilon_{\mathrm{H}})\} - \\ &\frac{1}{a_2}\mathrm{tr}\{a_2\tilde{W}_2^{\mathrm{T}}\phi_2(\tilde{W}_2^{\mathrm{T}}\phi_2 + \\ &\frac{1}{2}R^{-1}G^{\mathrm{T}}\nabla\phi_1^{\mathrm{T}}\tilde{W}_1+\varepsilon')^{\mathrm{T}}\} + \\ &2e^{\mathrm{T}}[f+gu_{\mathrm{d}}-h_{\mathrm{d}}+g(\mu^*-\varepsilon_2-\tilde{W}_2^{\mathrm{T}}\phi_2 - \\ &\mu_{\mathrm{r}})] + (-e^{\mathrm{T}}Qe-\mu^{*\mathrm{T}}R\mu^*) \leqslant \\ &-(\sigma_{1\mathrm{m}}^2 - \frac{1}{8a_2}b_{\mathrm{g}}^2\phi_{\mathrm{dM}}\|R^{-1}\|^2 - \\ &\frac{a_1}{2}\sigma_{1\mathrm{M}}^2)\|\tilde{W}_1\|^2 - (\phi_{2\mathrm{m}}^2 - a_2\phi_{2\mathrm{M}}^2 + \\ &\phi_{2\mathrm{M}}^2)\|\tilde{W}_2\|^2 - (-1+\lambda_{\min}(R))\|\mu^*\|^2 - \\ &(-2-4b_{\mathrm{g}}^2+\lambda_{\min}(Q))\|e\|^2.\end{aligned} \tag{43}$$

选择与定理1相同条件的a_1, a_2, Q, R, 显然$\dot{L}(t) < 0$. 由Barbalat引理, 当$t \to \infty$ 时, 式(28)与(43)可保证跟踪误差e以及权值估计误差$\tilde{W}_1$和$\tilde{W}_2$渐近收敛到零. 下面证明当$t \to \infty$时, $\|\mu^* - \mu_{\mathrm{ad}}\| \leqslant \delta_\mu$. 从式(19)与式(42)可得

$$\mu^* - \mu_{\mathrm{ad}} = \mu_{\mathrm{r}} + \tilde{W}_2^{\mathrm{T}}\phi_2(\eta) + \varepsilon_2(\eta). \tag{44}$$

由式(40)知, 当$t \to \infty$时, $e \to 0$, $\mu_{\mathrm{r}} \to 0$. 式(44)的上界为

$$\|\mu^* - \mu_{\mathrm{ad}}\| \leqslant \delta_\mu, \tag{45}$$

式中$\delta_\mu = \varepsilon_{2\mathrm{M}}$.　　证毕.

式(39)与式(45)相比, 显然加上鲁棒项将使得控制输入的逼近性能更佳.

5　仿真(Simulation)

为了证明本算法的有效性, 以及鲁棒控制的优越性, 本部分给出了非线性系统跟踪两种不同时变轨迹的仿真结果.

例 1　首先考虑如下非线性仿射系统跟踪正余弦时变轨迹的情况.

$$\begin{aligned}&\dot{x}_1 = -x_1 + x_2, \\ &\dot{x}_2 = -0.5x_1 - 0.5x_2(1-(\cos(2x_1)+2)^2) + \\ &\qquad (\cos(2x_1)+2)u.\end{aligned} \tag{46}$$

参考输入为$x_{\mathrm{d}}(t) = [\sin\, t \quad \sin t + \cos\, t]^{\mathrm{T}}$.

通过

$$\begin{aligned}&h_{\mathrm{d}} = [-x_{\mathrm{d1}}+x_{\mathrm{d2}} \quad -2x_{\mathrm{d1}}+x_{\mathrm{d2}}]^{\mathrm{T}}, \\ &f = [-x_1+x_2 \quad -0.5(x_1+x_2) + \\ &\qquad 0.5x_2(\cos(2x_1+2))^2]^{\mathrm{T}}, \\ &g = [0 \ \ \cos(2x_1+2)]^{\mathrm{T}},\ g_{\mathrm{d}}^{+} = [0 \ \ (\cos(2x_{\mathrm{d1}}+2))^{-1}]^{\mathrm{T}},\end{aligned}$$

可将该最优跟踪问题转化为时不变系统(4)的最优调

节问题. 根据本文的算法选取性能指标函数(8)中 $Q=\operatorname{diag}\{15,15\}$, $R=\operatorname{diag}\{3,3\}$.

选择评价网的基函数为

$$\phi_1=[\eta_1^2\ \ \eta_2^2\ \ \eta_1^2\eta_2^2\ \ \eta_1^2\eta_3^2\ \ \eta_1^2\eta_4^2\ \ \eta_2^2\eta_3^2\ \ \eta_2^2\eta_4^2]^{\mathrm{T}}.$$

选择执行网基函数为

$$\phi_2=$$

$$[2\eta_2\ \ 2\eta_1^2\eta_2\ \ 2\eta_2\eta_3^2\ \ 2\eta_2\eta_4^2]^{\mathrm{T}}\cos(2(\eta_1+\eta_2)+2).$$

选择控制增益 $a_1=0.8$, $a_2=0.5$, 鲁棒项参数 $K_{\mathrm{r}}=[30,30]$, $a=1.5$. 此外, 评价网与执行网的初始权值分别选择:

$$W_1=[-5.85\ \ -3.98\ \ -0.58\ \ -10.39\ \ 1.89\ \ -6.10\ \ -5.48]^{\mathrm{T}},$$

$$W_2=[-5.45\ \ -3.29\ \ -1.40\ \ -6.30]^{\mathrm{T}}.$$

为了保证持续性激励条件, 同样在前1100 s添加噪声信号. 执行程序后, 从图1可看出当噪声信号去掉以后, 跟踪误差将收敛到0. 评价网权值与执行网权值的收敛曲线分别见图2和图3. 评价网和执行网权值最终分别收敛到

$$\hat{W}_1=[1.83\ \ 5.03\ \ 11.05\ \ -2.06\ \ 3.02\ \ 1.49\ \ 0.45]^{\mathrm{T}},$$

$$\hat{W}_2=[-0.77\ \ -2.59\ \ -0.22\ \ -0.04]^{\mathrm{T}}.$$

本文注意到与执行网、评价网的基函数中带有期望轨迹η_3, η_4元素一一对应的权值元素都不是0, 那么这就说明最终值函数与最优控制策略都是与期望轨迹相关的, 即都是关于跟踪误差的时变函数.

由于理想权值都是未知的, 本文无法将仿真结果同其进行比较从而得出算法的精确性. 但是, 为比较最终的跟踪效果, 本文选择在同样初始状态下, 分别用设计的鲁棒最优控制器和初始容许控制器来得到跟踪误差曲线. 从图4可以看出, 与初始容许控制器相比, 鲁棒最优控制器实现的跟踪效果更好一些, 它使跟踪误差以更快的收敛速度在更短的时间内收敛到0, 从而体现了本方法的有效性与优越性.

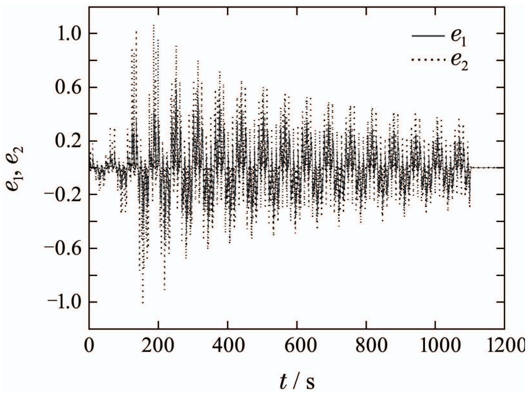

图 1 跟踪误差曲线图

Fig. 1 Error trajectories for the nonlinear system

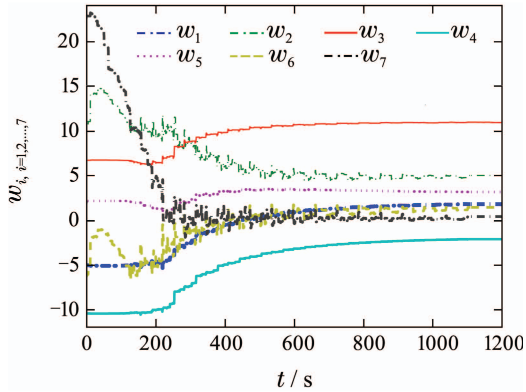

图 2 评价网权值收敛曲线图

Fig. 2 Convergence of critic NN weights

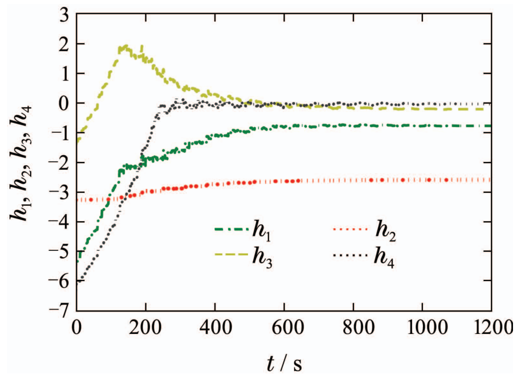

图 3 执行网权值收敛曲线图

Fig. 3 Convergence of action NN weights

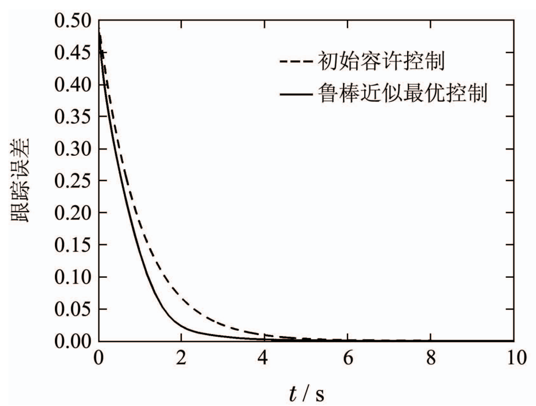

图 4 初始容许控制与鲁棒近似最优控制的比较图

Fig. 4 Comparison result between initial admissible controller and robust approximate optimal controller

例 2 本文选用上述非线性仿射系统来跟踪一组渐进收敛的指数时变轨迹.

参考输入为$x_{\mathrm{d}}(t)=[2^{-t}t+0.5\ \ 2^{-t}+0.5]^{\mathrm{T}}$. 则有$f$, g与例1相同, $h_{\mathrm{d}}=[-\ln 2(x_{\mathrm{d}1}-0.5)+x_{\mathrm{d}2}-0.5\ \ -\ln 2(x_{\mathrm{d}2}-0.5)]^{\mathrm{T}}$, $g_{\mathrm{d}}^{+}=[0\ \ (\cos(2x_{\mathrm{d}1}+2))^{-1}]^{\mathrm{T}}$.

可将该最优跟踪问题转化为时不变系统(4)的最优调节问题. 根据本文的算法, 本文选取同例1相同的性能指标函数Q, R, 评价网的基函数ϕ_1, 执行网基函数ϕ_2, 以及控制增益a_1, a_2, 鲁棒项参数K_{r}, a. 此外, 评价网与执行网的初始权值分别选择:

$$W_1 = [5.31\ \ 5.90\ \ -6.27\ \ -0.20\ \ -1.01\ \ 2.92\ \ 4.18]^{\mathrm{T}},$$
$$W_2 = [-4.48\ \ 3.59\ \ -7.62\ \ -0.03]^{\mathrm{T}}.$$

为了保证持续性激励条件, 在前1100 s添加噪声信号. 执行程序后, 从图5可看出当噪声信号去掉以后，跟踪误差将收敛到0. 评价网权值与执行网权值的收敛曲线分别见图6和图7. 其权值最终分别收敛到:

$$\hat{W}_1 = [6.46\ \ 3.13\ \ 3.98\ \ -0.06\ \ -0.52\ \ 1.38\ \ 5.25]^{\mathrm{T}},$$
$$\hat{W}_2 = [0.24\ \ 0.30\ \ -5.28\ \ 2.56]^{\mathrm{T}}.$$

同例1一样, 与执行网、评价网的基函数中含有期望轨迹η_3, η_4元素相对应的权值都不是0, 也就再次验证了值函数与最优控制策略都是关于跟踪误差的时变函数这一结论.

图8再次说明鲁棒最优控制器与初始容许控制器相比所具有的有效性. 此外, 由于目前在神经网络控制中基函数的选择是没有固定的规律可遵循的, 所以文中基函数的选择也是经过了大量的实验并参考了许多文献而选择的. 这种逼近器的精确性可通过两种方式来得到改善, 其一是选择更合适的基函数, 其二就是以加大计算成本为代价, 增加基函数的元素个数.

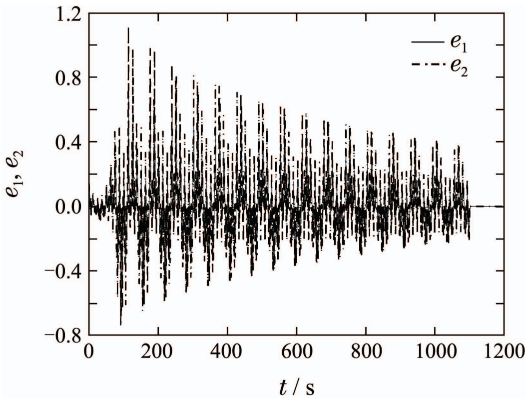

图 5 跟踪误差曲线图

Fig. 5 Error trajectories for the nonlinear system

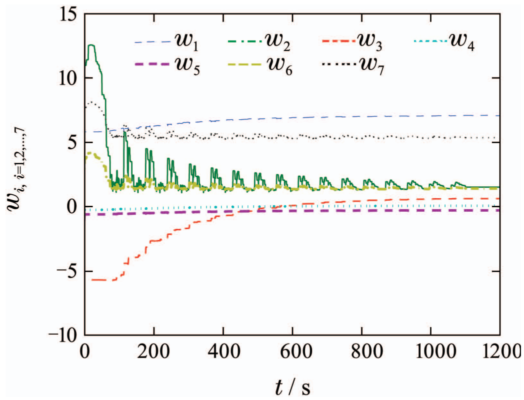

图 6 评价网权值收敛曲线图

Fig. 6 Convergence of critic NN weights

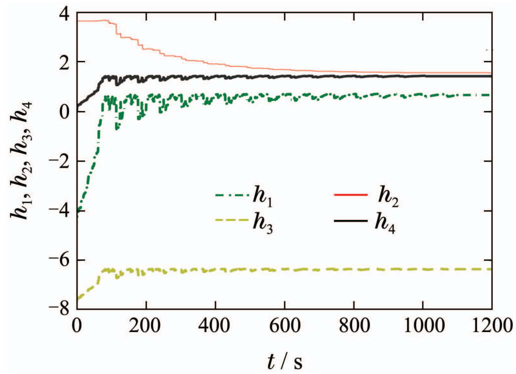

图 7 执行网权值收敛曲线图

Fig. 7 Convergence of action NN weights

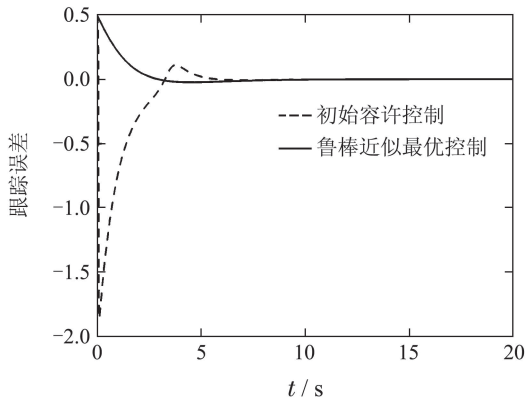

图 8 初始容许控制与鲁棒近似最优控制的比较图

Fig. 8 Comparison result between initial admissible controller and robust approximate optimal controller

6 结论(Conclusions)

本文通过系统变换将原系统的跟踪问题转变为最优调节问题, 且将原系统的时变值函数变换为时不变值函数, 并利用近似动态规划算法中的策略迭代方法来近似求解新的HJB方程, 从而得到非线性连续系统最优控制策略, 同时引入鲁棒控制来消除近似过程中产生的残差, 又通过李雅普诺夫方法证明在该近似最优控制器下跟踪误差渐近收敛到零, 最后给出仿真实例说明该方法的有效性与优越性. 接下来, 本文希望在参考控制未知即系统模型未知情况下, 放松持续性激励条件, 来解决最优跟踪问题.

参考文献(References):

[1] HUANG Y, LIU D. Neural-network-based optimal tracking control scheme for a class of unknown discrete-time nonlinear systems using iterative ADP algorithm [J]. *Neurocomputing*, 2012, 10(3): 325 – 331.

[2] SONG R, XIAO W, SUN C. Optimal tracking control for a class of unknown discrete-time systems with actuator saturation via data-based ADP algorithm [J]. *Neurocomputing*, 2010, 39(9): 1293 – 1302.

[3] ZHANG H, CUI L, ZHANG X, et al. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method [J]. *IEEE Transactions on Neural Networks*, 2011, 22(12): 2226 – 2236.

[4] WANG D, LIU D, WEI Q. Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach [J]. *Neurocomputing*, 2012, 78(1): 14 – 22.

[5] LUO Y, LIANG M. Approximate optimal tracking control for a class of discrete-time non-affine systems based on GDHP algorithm [C] //*The 4th International Workshop on Advanced Computational Intelligence*. Wuhan: International Workshop on Advanced Computational Intelligence, 2011: 143 – 149.

[6] DIERKS T, JAGANNATHAN S. Optimal tracking control of affine nonlinear discrete-time systems with unknown internal dynamics [C] //*Proceedings of the 48th IEEE Conference on Decision and Control*. Shanghai: IEEE, 2009: 6750 – 6755.

[7] ZHANG H, WEI Q, LUO Y. A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy hdp iteration algorithm [J]. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 2008, 38(4): 937 – 942.

[8] WERBOS P J. Approximate dynamic programming for real time control and neural modeling [M] //*Handbook of Intelligent Control*. New York: Multiscience Press, 1992.

[9] LI Xiaoli, LIU Dexin, JIA Chao, et al. Multiple set-points tracking control method based on adaptive dynamic programming [J]. *Control Theory & Applications*, 2013, 30(6): 709 – 716.
(李晓理, 刘德馨, 贾超, 等. 基于自适应动态规划的多设定值跟踪控制方法 [J]. 控制理论与应用, 2013, 30(6): 709 – 716.)

[10] LIN Xiaofeng, HUANG Yuanjun, SONG Chunning. Approximate optimal control with ε-error bound [J]. *Control Theory & Applications*, 2012, 29(1): 104 – 108.
(林小峰, 黄元君, 宋春宁. 带ε误差限的近似最优控制 [J]. 控制理论与应用, 2012, 29(1): 104 – 108.)

[11] ABU-KHALAF M, LEWIS F L. *Nearly optimal HJB solution for constrained input systems using a neural network least-squares approach* [C] //*Proceedings of the 48th IEEE Conference on Decision and Control*. Las Vegas: IEEE, 2002: 943 – 948.

[12] BHASIN S, KAMALAPURKAR R, JOHNSON M, et al. A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems [J]. *Automatica*, 2013, 49(1): 89 – 92.

[13] DIERKS T, JAGANNATHAN S. *Optimal control of affine nonlinear continuous-time systems* [C] //*Proceedings of American Control Conference*. Baltimore: IEEE, 2010: 1568 – 1573.

[14] VAMVOUDARIS K, LEWIS F. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem [J]. *Automatica*, 2010, 46(5): 878 – 888.

[15] YAN Qiuzhen, SUN Mingxuan. Error trajectory tracking by robust learning control for nonlinear systems [J]. *Control Theory & Applications*, 2013, 30(1): 23 – 30.
(严求真, 孙明轩. 一类非线系统的误差轨迹跟踪鲁棒学习控制算法 [J]. 控制理论与应用, 2013, 30(1): 23 – 30.)

[16] KAMALAPURKAR R, DINH H B, HASIN S, et al. Approximate optimal trajectory tracking for continuous-time nonlinear systems [J]. *Automatica*, 2015, 51(1): 40 – 48.

[17] FINLAYSON B. *The Method of Weighted Residuals and Variational Principles* [M]. New York: Academic Press, 1990.

[18] HORNIK K, STINCHCOMBE M, WHITE H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks [J]. *Neural Networks*, 1990, 3(5): 551 – 560.

作者简介:

屈秋霞 (1988–), 女, 博士研究生, 目前研究方向为近似动态规划, E-mail: quqiuxia2010@163.com;

罗艳红 (1981–), 女, 副教授, 目前研究方向为近似动态规划, E-mail: neuluo@gmail.com;

张化光 (1959–), 男, 教授, 博士生导师, 目前研究方向为智能控制、近似动态规划、复杂网络建模与控制、分布式控制等, E-mail: zhanghuaguang@mail.neu.edu.cn.