基于广义模糊双曲模型的自适应动态规划最优控制设计

张吉烈 [1] 张化光 [1, 2] 罗艳红 [1] 梁洪晶 [1]

摘 要 为连续非线性系统提出了一种有效的最优控制设计方法. 广义模糊双曲模型 (Generalized fuzzy hyperbolic model, GFHM) 首次作为逼近器用来估计 HJB (Hamilton-Jacobi-Bellman) 方程的解 (值函数, 即它是状态与代价函数之间的映射), 然后, 利用该近似解获得最优控制. 本文方法只需要一个 GFHM 估计值函数. 首先, 阐述了对于连线非线性系统最优控制的设计过程; 然后, 证明了逼近误差是一致最终有界的 (Uniformly ultimately bounded, UUB); 最后, 一个数值例子验证了本文方法的有效性. 另一个例子通过与神经网络自适应动态规划的方法作比较, 演示了本文方法的优点.

关键词 广义模糊双曲模型, 最优控制, 自适应动态规划, 近似最优, 自适应控制

引用格式 张吉烈, 张化光, 罗艳红, 梁洪晶. 基于广义模糊双曲模型的自适应动态规划最优控制设计. 自动化学报, 2013, **39**(2): 142−149

DOI 10.3724/SP.J.1004.2013.00142

Nearly Optimal Control Scheme Using Adaptive Dynamic Programming Based on Generalized Fuzzy Hyperbolic Model

ZHANG Ji-Lie[1] ZHANG Hua-Guang[1, 2] LUO Yan-Hong[1] LIANG Hong-Jing[1]

Abstract An effective scheme is presented to design the nearly optimal control for continuous-time (C-T) nonlinear systems. The generalized fuzzy hyperbolic model (GFHM) is used to approximate the solution of the Hamilton-Jacobi-Bellman (HJB) equation (i.e., the value function) for the first time. Further, the approximate solution is utilized to obtain the nearly optimal control. The value function is estimated by only using single GFHM, which captures the mapping between the state and value function. First, we illustrate the design process for the nearly optimal control involving nonlinear systems. Then stability conditions and conservatism analysis are given, and the approximate errors are proven to be uniformly ultimately bounded (UUB). Finally, a numerical example illustrates the effectiveness of our method and an example compared with the adaptive method based on dual neural-network models is used to demonstrate the advantages of our method.

Key words Generalized fuzzy hyperbolic model (GFHM), optimal control, adaptive dynamic programming (ADP), approximation optimization, adaptive control

Citation Ji-Lie Zhang, Hua-Guang Zhang, Yan-Hong Luo, Hong-Jing Liang. Nearly optimal control scheme using adaptive dynamic programming based on generalized fuzzy hyperbolic model. *Acta Automatica Sinica*, 2013, **39**(2): 142−149

近十年来, 自适应动态规划 (Adaptive dynamic programming, ADP) 在求解 HJB (Hamilton-Jacobi-Bellman) 方程方面起着重要的作用[1−9]. 文献 [10−12] 概述了自适应动态规划技术的发展历史. 最近, 连续时间系统的近似最优控制设计问题吸引了很多科研工作者的关注[13−17]. 设计中的主要问题是如何求解 HJB 方程[11, 18]. 如今, 针对这个问题有三种主流的解决方法: Galerkin 逐次逼近法、基于增强学习[12, 19−20] 的策略迭代 (Policy iteration, PI)[19] 法和基于神经网络的自适应方法.

如何求解 HJB 方程是一直困扰着科研工作者的问题. Beard 在文献 [21−22] 中提出了一种获得近似解的方法. 该方法中, HJB 方程分两步逼近求解, 首先, 将 HJB 方程的非线性偏微分形式化简成线性偏微分方程, 然后, 利用 Galerkin 谱方法逼近 GHJB 方程. 这种逐次逼近法不断地改进控制率,

收稿日期 2012-04-10 录用日期 2012-08-20
Manuscript received April 10, 2012; accepted August 20, 2012
国家重点基础研究发展计划 (973 计划) (2009CB320601), 国家高技术研究发展计划 (863 计划) (2012AA040104), 国家自然科学基金 (50977008, 61034005), 辽宁省教育厅科技研究项目 (LT2010040) 资助
Supported by National Basic Research Program of China (973 Program) (2009CB320601), National High Technology Research and Development Program of China (863 Program) (2012AA040104), National Natural Science Foundation of China (50977008, 61034005), and Science and Technology Research Program of the Education Department of Liaoning Province (LT2010040)
本文责任编委 刘德荣
Recommended by Associate Editor LIU De-Rong
1. 东北大学信息科学与工程学院 沈阳 110819 2. 流程工业综合自动化国家重点实验室 (东北大学) 沈阳 110819
1. School of Information Science and Engineering, Northeastern University, Shenyang 110819 2. State Key Laboratory of Synthetical Automation for Process Industries (Northeastern University), Shenyang 110819
该文的英文版同时发表在 *Acta Automatica Sinica*, vol. 39, no. 2, pp. 142−149, 2013.

使其逼近理想的最优控制. 但是, 这种方法的计算是离线形式的.

为此, Vrabie 等结合了最小二乘和增强学习方法提出了一种对于系统部分未知的 PI 迭代在线算法[13−14]. 该方法以一个容许控制策略作为初始值估计值函数, 然后更新控制策略使其比之前策略得到一个更小的值函数. 控制策略反复地被更新直到不再变化为止. 这意味着策略迭代收敛到了最优控制. 通过加权残量法可以得到参数权值的最小二乘解, 即 HJB 方程的近似解. Abu-Khalaf 等也利用这种方法对约束非线性系统设计了近似最优控制器[23]. 但是, 这种方法需要离散化 HJB 方程, 并且利用采样数据计算权值.

最近发表的文章中利用两个神经网络模型 (执行网和评价网, 这里称之为双网) 设计近似最优控制器比较流行. 文献 [15−16, 24] 将自适应控制技术用在求解双网的参数上. 2008 年, Al-Tamimi 等利用这种方法求解了关于离散非线性系统的 HJB 方程[25]. 该方法引入了 Weierstrass 逼近定理[26], 即存在一组完备相互独立的基可以逼近 HJB 方程解.

理论上, 对于 N 个完备基 (N 趋近于无穷大), Weierstrass 逼近定理是收敛的. 但是对于有限的 N 来说, 这个定理对选择的完备基是敏感的. 如果一个光滑的函数不能被有限个完备基函数张成, 那么神经网络模型就不能严格地逼近它. 我们的目标是找到这样一组基, 尽可能地使 N 小, 并且能够捕获函数的主要特征. 但是, 选取一组合适的基函数并不是一件容易的事. 选择不好会导致精度的下降, 甚至不能逼近该函数. 这激发我们寻找一个新的估计器克服这些缺点. 我们使用了具有这样功能的广义模糊双曲模型 (Generalized fuzzy hyperbolic model, GFHM) 作为估计器解决这些问题. 因为广义模糊双曲模型是强非线性模型, 基函数是相互独立的, 又有万能逼近性[27] (在完备集中, 可以逼近任意非线性函数). 同时, 因为广义模糊双曲模型 GFHM 可以视为神经网络模型, 所以模型权值可以通过有效的学习方法进行优化. 另一方面, GFHM 仅仅使用了与输入变量相同数量的权值参数. 与传统的双神经网络 (双网) 相比, 很大程度上减少了计算量. 当处理高维系统时, 本文方法仅需要一个 GFHM (权值数量与输入变量数相同). 最重要的一点是因为双曲正切函数导数的范数小于 1, 所以本文方法对于稳定条件有更小的保守性. 证明将在文中给出.

本文提出了一个利用 GFHM 设计近似最优控制的新方法. 只使用一个 GFHM 估计值函数, 所以与双网[16] 相比至少减少了一半的存储空间. 之后, 得到连续时间系统 HJB 方程的近似误差. 为了使误差达到最小, 利用梯度下降法搜索最优值. 本文也给出了我们方法的稳定条件, 并且证明了参数权值误差、最优控制输入误差和状态误差是一致最终有界的 (Uniformly ultimately bounded, UUB).

本文方法的主要贡献包括:

1) 使用 GFHM 逼近值函数, 只需要更新 n 个权值;

2) 这种简单的方法仅仅需要一个 GFHM 足以设计近似最优控制, 不需要双网 (包括执行网和评价网), 从而至少减少了一半的存储空间;

3) 由于双曲正切函数导数的范数小于 1, 所以该方法的稳定条件有更小的保守性.

本文安排如下: 第 1 节介绍了一些基本定义和概念; 第 2 节和第 3 节给出了近似最优控制的设计过程; 在第 4 节中, 对该方法进行了稳定性和保守性分析, 并且证明了逼近误差是最终一致有界的 (UUB); 最后, 一个数值例子证明了本文方法的有效性, 也通过与双神经网络自适应方法比较的例子给出该方法的优点.

1　预备知识

定义 1. 已知一个 MISO 系统的 n 个输入变量组成的向量为 $\boldsymbol{x} = [x_1(t), \cdots, x_n(t)]^{\mathrm{T}}$, 输出变量为 y. 如果用来描述此系统的模糊规则基满足以下条件, 则称这组模糊规则基为双曲正切模糊规则基.

1) 输出变量 y 与以下模糊规则一一对应:

规则: IF x_1 is $F_{x_1}$and x_2 is $F_{x_2}, \cdots,$ and x_n is F_{x_n}

$$\text{THEN}\ \ y = \theta^{\pm}_{F_{x_1}} + \theta^{\pm}_{F_{x_2}} + \cdots + \theta^{\pm}_{F_{x_n}}$$

其中, F_{x_i} $(i = 1, \cdots, n)$ 是 x_i 的模糊子集, 其中包括 P_{x_i}(正) 和 N_{x_i}(负), 并且 $\theta^{\pm}_{F_{x_i}} (i = 1, \cdots, n)$ 有 $2n$ 个实常数与 F_{x_i} 相对应.

2) THEN 中常数项 $\theta^{\pm}_{F_{x_i}}$ 与 IF 中 F_{x_i} 一一对应, 即如果 IF 中 F_{x_i} 的语言值是正值 P_{x_i}, 那么 THEN 中一定是 $\theta^{+}_{F_{x_i}}$; 如果 IF 中 F_{x_i} 的语言值是负值 N_{x_i}, 那么 THEN 中一定是 $\theta^{-}_{F_{x_i}}$; 但是, 如果 IF 中没有 F_{x_i}, 那么 THEN 中与之对应的 $\theta^{\pm}_{F_{x_i}}$ 是 0.

3) 规则基中存在 2^n 个模糊规则; 即在 IF 中对于所有的 P_{x_i} 和 N_{x_i} 存在 2^n 个组合方式.

引理 1. 已知一组双曲正切型模糊规则基, 若定义 P_{x_i} 和 N_{x_i} (x_i 为输入变量) 的隶属度函数为

$$\mu_{P_{x_i}}(x_i) = \mathrm{e}^{-\frac{1}{2}(x_i - \phi_i)^2}$$

$$\mu_{N_{x_i}}(x_i) = \mathrm{e}^{-\frac{1}{2}(x_i + \phi_i)^2}$$

其中, $i = 1, \cdots, n$ 且 ϕ_i 是正常数. $\theta_{P_{x_i}}$ 和 $\theta_{N_{x_i}}$ 分别表示 $\theta^{+}_{F_{x_i}}$ 和 $\theta^{-}_{F_{x_i}}$, 利用单点模糊化, 乘积推理和重心去模糊化方法可以推出:

$$y = \boldsymbol{\theta}^{\mathrm{T}} \tanh(\Phi \boldsymbol{x}) + \zeta$$

其中, $\boldsymbol{\theta} = [\theta_1 \quad \cdots \quad \theta_n]^{\mathrm{T}}$ 是最优的权值向量; $\tanh(\Phi\boldsymbol{x}) = [\tanh(\phi_1 x_1) \quad \cdots \quad \tanh(\phi_n x_n)]^{\mathrm{T}}$, 并且 $\Phi = \mathrm{diag}\{\phi_i\}$ $(i = 1 \cdots n)$; ζ 是一个常数标量. 它被称作广义模糊双曲模型 (GFHM).

引理 2[28]**.** 紧集 $U \subset \mathbf{R}^n$ 上已知实连续函数 $f(\boldsymbol{x})$, 有任意的 $\delta > 0$, 存在一个 $h(\boldsymbol{x}) \in F$ (F 是所有模糊基函数的集合) 满足:

$$\sup_{\boldsymbol{x}\in U} |f(\boldsymbol{x}) - h(\boldsymbol{x})| < \delta$$

注 1. 引理 2 意味着 GFHM 是个万能逼近器.

定义 2[29]**.** 矩阵 $A \in \mathbf{R}^{n\times n}$ 的 Frobenius 范数定义为 $\|A\|_F^2 = \mathrm{tr}(A^{\mathrm{T}}A) = \sum a_{ij}^2$, 其中, $\mathrm{tr}(A^{\mathrm{T}}A)$ 表示矩阵 $A^{\mathrm{T}}A$ 的迹. 则有以下不等式成立:

$$\|A\boldsymbol{x}\| \le \|A\|_F \|\boldsymbol{x}\|$$

2 最优控制设计

考虑下面连续非线性系统:

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

$\boldsymbol{x} \in \mathbf{R}^n$ 是状态向量, $\boldsymbol{u} \in \mathbf{R}^p$ 是控制输入向量, 并且 $f(\boldsymbol{x}) \in \mathbf{R}^n$, $g(\boldsymbol{x}) \in \mathbf{R}^{n\times p}$, $f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}$ 在 $\boldsymbol{\Omega} \subseteq \mathbf{R}^n$ 上是 Lipschitz 连续非线性函数向量 ($f(\boldsymbol{0}) = \boldsymbol{0}$, $\boldsymbol{\Omega}$ 包含原点).

假设 1. 系统 (1) 在 $\boldsymbol{\Omega}$ 上是可镇定的 (即存在一个连续控制函数 $\boldsymbol{u}$ 使该系统在 $\boldsymbol{\Omega}$ 上是渐近稳定的).

定义性能指标为

$$J(\boldsymbol{x_0}) = \int_0^\infty r(\boldsymbol{x},\boldsymbol{u})\mathrm{d}t \tag{2}$$

其中, $r(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{x}^{\mathrm{T}}Q\boldsymbol{x} + \boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u}$, $Q = Q^{\mathrm{T}} > 0$ 并且 $R = R^{\mathrm{T}} > 0$.

定义 3 (容许控制)[15]**.** 如果一个控制策略 $\boldsymbol{\mu}(\boldsymbol{x})$ 不仅能够在 $\boldsymbol{\Omega}$ 上使系统稳定, 而且也能使其性能指标积分有界的, 那么我们称之为容许控制.

定义 4 (一致最终有界(UUB))[30]**.** 已知系统 (1) 的平衡点是 $\boldsymbol{x}_e = \boldsymbol{0}$, 若紧集 $S \subset \mathbf{R}^n$, $\boldsymbol{x}_0 \in S$, 则存在有界的 B 和时间 $T(B,\boldsymbol{x}_0)$ 使得对于所有 $t \ge t_0 + T$ 满足 $\|\boldsymbol{x}(t) - \boldsymbol{x}_e\| \le B$, 那么称该平衡点为一致最终有界 (UUB).

问题 1. 如何找到一个容许控制 $\boldsymbol{u}$ 使系统 (1) 稳定, 并且最小化性能指标 (2) 是我们要解决的问题.

对于任意容许控制 $\boldsymbol{\mu}$, 如果相应的性能指标:

$$V^{\boldsymbol{\mu}}(\boldsymbol{x_0}) = \int_{t_0}^\infty r(\boldsymbol{x},\boldsymbol{\mu})\mathrm{d}t \tag{3}$$

属于 C^1, 那么它的极小形式称作非线性 Lyapunov 方程:

$$0 = r(\boldsymbol{x},\boldsymbol{\mu}) + (V_{\boldsymbol{x}}^{\boldsymbol{\mu}})^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{\mu}) \tag{4}$$

其中, $V_{\boldsymbol{x}}^{\boldsymbol{\mu}}$ 是 $V^{\boldsymbol{\mu}}(\boldsymbol{x})$ 关于 $\boldsymbol{x}$ 的偏导数 ($V^{\boldsymbol{\mu}}(\boldsymbol{x})$ 不显含 t). 式 (4) 是非线性系统在容许控制 $\boldsymbol{\mu}(\boldsymbol{x})$ 下的一个 Lyapunov 方程. 控制可以通过先获得与非线性系统相关的值函数 $V^{\boldsymbol{\mu}}(\boldsymbol{x})$ 得到. 已知容许控制 $\boldsymbol{\mu}(\boldsymbol{x})$, 如果 $V^{\boldsymbol{\mu}}(\boldsymbol{x})$ 满足式 (4), 并且 $r(\boldsymbol{x},\boldsymbol{\mu}) > 0$, 那么 $V^{\boldsymbol{\mu}}(\boldsymbol{x})$ 是系统 (1) 在控制 $\boldsymbol{\mu}(\boldsymbol{x})$ 下的一个 Lyapunov 函数.

定义问题的 Hamiltonian 函数是:

$$H(\boldsymbol{x},\boldsymbol{\mu},V_{\boldsymbol{x}}) = r(\boldsymbol{x},\boldsymbol{\mu}) + (V_{\boldsymbol{x}})^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{\mu})$$

那么最优值函数 $V^*(\boldsymbol{x})$ 满足以下的 HJB 方程:

$$0 = \min_{\boldsymbol{u}\in\Psi(\boldsymbol{\Omega})} H(\boldsymbol{x},\boldsymbol{u},V_{\boldsymbol{x}}^*) = r(\boldsymbol{x},\boldsymbol{u}^*) + (V_{\boldsymbol{x}}^*)^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}^*) \tag{5}$$

假设方程 (5) 的右边的最小值存在且唯一, 那么最优控制为

$$\boldsymbol{u}^* = -\frac{1}{2}R^{-1}g(\boldsymbol{x})^{\mathrm{T}}V_{\boldsymbol{x}}^* \tag{6}$$

为找到问题的最优解, 依据假设 1 只要求解 HJB 方程 (5) 获得 $V^*(\boldsymbol{x})$, 然后将其代入式 (6) 中, 就可以得到最优控制 $\boldsymbol{u}^*$. 但在实际中, 求解 HJB 方程是相当困难的, 甚至是不可能的. 神经网络现在广泛地被用作估计器求解近似解[15−16, 23]. 而我们利用广义模糊双曲模型作为估计器, 提出了一种求解 HJB 方程 (5) 的新方法. 下节中, 将详细给出本文的方法.

3 广义模糊双曲模型近似逼近 HJB 方程的解

根据引理 1 和引理 2, 我们首次利用 GFHM 估计值函数 $V(\boldsymbol{x})$, 并称之为广义模糊双曲评价估计器 (Generalized fuzzy hyperbolic critic estimator, GFHCE), 如下:

$$V(\boldsymbol{x}) = \boldsymbol{\theta}^{\mathrm{T}}\tanh(\Phi\boldsymbol{x}) + \zeta + \varepsilon$$

其中, $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_n]^{\mathrm{T}} \in \mathbf{R}^n$ 是未知的理想权值向量, $\Phi = \mathrm{diag}\{\phi_i\} \in \mathbf{R}^{n\times n}$ $(i = 1, \cdots, n)$ 是未知的理想权值矩阵, ζ 一个常标量, 并且 ε 是 GFHCE 逼近误差.

$V(\boldsymbol{x})$ 关于 $\boldsymbol{x}$ 的导数是

$$V_{\boldsymbol{x}} = \Lambda(\Phi\boldsymbol{x})\Phi^{\mathrm{T}}\boldsymbol{\theta} + \Delta\boldsymbol{\varepsilon} \tag{7}$$

其中, $\Lambda(\Phi\boldsymbol{x}) = [\frac{\partial\tanh(\Phi\boldsymbol{x})}{\partial\Phi\boldsymbol{x}}]^{\mathrm{T}}$, $\Delta\boldsymbol{\varepsilon} = \frac{\partial\varepsilon}{\partial\boldsymbol{x}}$.

令 $\hat{\boldsymbol{\theta}}$, $\hat{\Phi}$ 和 $\hat{\zeta}$ 分别为 $\boldsymbol{\theta}$, Φ 和 ζ 的估计量, 那么可以得到 $V(\boldsymbol{x})$ 和 $V_{\boldsymbol{x}}$ 的估计如下:

$$\hat{V}(\boldsymbol{x}) = \hat{\boldsymbol{\theta}}^{\mathrm{T}} \tanh(\hat{\Phi}\boldsymbol{x}) + \hat{\zeta} \tag{8}$$

$$\hat{V}_{\boldsymbol{x}} = \hat{\Lambda}(\Phi\boldsymbol{x})\hat{\Phi}^{\mathrm{T}}\hat{\boldsymbol{\theta}} \tag{9}$$

其中, $\hat{\Lambda}(\Phi\boldsymbol{x}) = [\frac{\partial \tanh(\hat{\Phi}\boldsymbol{x})}{\partial \hat{\Phi}\boldsymbol{x}}]^{\mathrm{T}}$. 进而得到近似的 Hamiltonian 函数如下:

$$\begin{aligned} e = H(\boldsymbol{x}, \boldsymbol{u}, \hat{\boldsymbol{\theta}}, \hat{\Phi}) = \\ \hat{V}_{\boldsymbol{x}}^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}) + \boldsymbol{x}^{\mathrm{T}}Q\boldsymbol{x} + \boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u} = \\ \hat{\varphi}(\hat{\boldsymbol{\theta}}, \hat{\Phi}) + r(\boldsymbol{x}, \boldsymbol{u}) \end{aligned}$$

其中, $\hat{\varphi}(\hat{\boldsymbol{\theta}}, \hat{\Phi}) = (\Lambda(\hat{\Phi}\boldsymbol{x})\hat{\Phi}^{\mathrm{T}}\hat{\theta})^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u})$.

已知任意容许控制 $\boldsymbol{u}$, 选择权值 $\hat{\boldsymbol{\theta}}$ 和 $\hat{\Phi}$ 最小化下式:

$$E(\hat{\boldsymbol{\theta}}, \hat{\Phi}) = \frac{1}{2}e^{\mathrm{T}}e$$

$\hat{\boldsymbol{\theta}}$ 的权值更新率可以通过梯度下降法[31] 得到, 如下:

$$\dot{\hat{\boldsymbol{\theta}}} = -a_1\boldsymbol{\sigma}(\boldsymbol{\sigma}^{\mathrm{T}}\hat{\boldsymbol{\theta}} + r(\boldsymbol{x}, \boldsymbol{u})) \tag{10}$$

其中, $a_1 > 0$ 是 $\hat{\boldsymbol{\theta}}$ 的自适应增益. $\boldsymbol{\sigma} = h(\hat{\Phi})(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u})$, 其中, $h(\hat{\Phi}) = \hat{\Phi}[\Lambda(\hat{\Phi}\boldsymbol{x})]^{\mathrm{T}}$.

$\hat{\Phi}$ 的权值更新率也可由梯度下降法得到, 如下:

$$\dot{\hat{\Phi}} = -a_2(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u})\hat{\boldsymbol{\theta}}^{\mathrm{T}}\left[\frac{\partial(h(\hat{\Phi}))}{\partial\hat{\Phi}}\right]e \tag{11}$$

其中, $a_2 > 0$ 是 $\hat{\Phi}$ 的自适应增益.

由于 e 不显含 $\hat{\zeta}$, 即 $\hat{\zeta}$ 不影响 e, $\hat{\zeta}$ 的权值更新率是:

$$\dot{\hat{\zeta}} = 0 \tag{12}$$

其中, $a_3 > 0$ 是 $\hat{\zeta}$ 的自适应增益.

注 2. 因为 $\hat{\zeta}$ 不影响误差 e, 那么式 (12) 成立, 所以 $\hat{\zeta}$ 可以任意选择一个常数. 当 $\hat{\zeta}$ 选为 0, 则我们完全可以称之为模糊双曲模型 (Fuzzy hyperbolic model, FHM), 进而取代 GFHM.

注 3. 为了使 $\hat{\boldsymbol{\theta}}$ 和 $\hat{\Phi}$ 收敛, 必须加入持续激励, 所以控制输入混合了噪声和控制输入信号.

将式 (9) 插入式 (6) 中, 得到最优控制:

$$\boldsymbol{u} = -\frac{1}{2}R^{-1}g(\boldsymbol{x})^{\mathrm{T}}\Lambda(\hat{\Phi}\boldsymbol{x})\hat{\Phi}^{\mathrm{T}}\hat{\boldsymbol{\theta}} \tag{13}$$

该控制能够最小化性能指标 (2) ($\hat{\boldsymbol{\theta}}$ 和 $\hat{\Phi}$ 的自适应控制率分别是式 (10) 和式 (11)).

注 4. 因为 $\hat{\Phi}$ 是一个非线性参数 (Nonlinear in the parameters, NLIP), 所以在实际应用中参数的调整和 UUB 分析是相当困难的. 庆幸的是 GFHM 可以视为两层神经网络模型[27, 31−32], 而激励函数是 $\tanh(\cdot)$. 如果权值 $\hat{\Phi}$ 固定, 那么 GFHM 就是一个关于 $\boldsymbol{\theta}$ 线性参数模型 (Linear in the parameters, LIP).

根据注 4, 令 $\hat{\Phi} = I$, 可以得到简单的最优控制:

$$\boldsymbol{u} = -\frac{1}{2}R^{-1}g(\boldsymbol{x})^{\mathrm{T}}\Lambda(\boldsymbol{x})\hat{\boldsymbol{\theta}} \tag{14}$$

该控制使性能指标 (2) 最小化, 而 $\hat{\boldsymbol{\theta}}$ 的自适应率是

$$\dot{\hat{\boldsymbol{\theta}}} = -a\boldsymbol{\sigma}(\boldsymbol{\sigma}^{\mathrm{T}}\hat{\boldsymbol{\theta}} + \boldsymbol{x}^{\mathrm{T}}Q\boldsymbol{x} + \boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u}) \tag{15}$$

其中, $\boldsymbol{\sigma} = [\frac{\partial \tanh(\boldsymbol{x})}{\partial \boldsymbol{x}}](f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u})$, 并且 a 是 $\hat{\boldsymbol{\theta}}$ 的自适应增益.

4 性能分析

本节对本文方法给出了稳定性和保守性分析.

4.1 稳定性分析

本节对本文提出的方法进行了稳定性分析. 以下假设贯穿于本节:

假设 2.

1) $\|f(\boldsymbol{x})\| \le \kappa$ 和 $\|g(\boldsymbol{x})\|_F \le \bar{\beta}$, κ 和 $\bar{\beta}$ 都是正常数;

2) 持续激励条件能够使得 $\boldsymbol{\sigma}$ 和 $\boldsymbol{\theta}$ 满足 $\sigma_m < \|\boldsymbol{\sigma}\| < \sigma_M$ 和 $\|\boldsymbol{\theta}\| < \theta_M$, σ_m, σ_M 和 θ_M 也都是正常数;

3) $\|\Delta\boldsymbol{\varepsilon}\| < \varepsilon_{\Delta M}$, $\varepsilon_{\Delta M}$ 是一个正常数;

4) HJB 方程的误差 $\varepsilon_{\mathrm{HJB}}$ 有上界, 满足 $\|\varepsilon_{\mathrm{HJB}}\| < \bar{\varepsilon}$, $\bar{\varepsilon}$ 是一个正常数.

将式 (7) 插入式 (4), Hamiltonian 函数是

$$\begin{aligned} H(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}) = r(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\theta}^{\mathrm{T}}\Lambda^{\mathrm{T}}(\boldsymbol{x})(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}) + \\ \Delta\boldsymbol{\varepsilon}^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}) \end{aligned}$$

此时, 式 (5) 可以写为

$$r(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\sigma} = \varepsilon_{\mathrm{HJB}} \tag{16}$$

其中, $\varepsilon_{\mathrm{HJB}} = -\Delta\boldsymbol{\varepsilon}^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u})$, $\varepsilon_{\mathrm{HJB}}$ 是由 GFHM 逼近值函数产生的剩余误差.

定义值函数的权值估计误差为 $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. 由式 (10) 和式 (16) 得到:

$$\dot{\tilde{\boldsymbol{\theta}}} = -a\boldsymbol{\sigma}(\boldsymbol{\sigma}^{\mathrm{T}}\tilde{\boldsymbol{\theta}} + \varepsilon_{\mathrm{HJB}}) \tag{17}$$

定理 1. 在控制 (14) 和 $\hat{\boldsymbol{\theta}}$ 的更新率 (15) 作用下, 考虑系统 (1). 若式 (19) 和式 (20) 成立, 则状态

$\boldsymbol{x}$ 和权值估计误差 $\tilde{\theta}$ 是一致最终有界. 并且, 控制输入 $\boldsymbol{u}$ 近似于最优控制 $\boldsymbol{u}^*$; 即对于任意 $\varepsilon_u > 0$, 有 $\|\boldsymbol{u} - \boldsymbol{u}^*\| \leq \varepsilon_u$ $(t \to \infty)$.

证明. 选择 Lyapunov 函数如下:

$$L = L_1 + L_2 + L_3 \tag{18}$$

其中, $L_1 = \mathrm{tr}(\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\tilde{\boldsymbol{\theta}})/2a$, $L_2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + 2\Gamma V(\boldsymbol{x})$ $(\Gamma > 0)$, 并且 $L_3 = \mathrm{tr}(\tilde{\zeta}^{\mathrm{T}}\tilde{\zeta})/2$.

根据假设 2 和式 (14), Lyapunov 函数 (18) 沿着系统 (1) 轨迹的时间导数存在下面不等式:

$$\begin{aligned}
\dot{L}_1 =& \frac{1}{a}\mathrm{tr}(\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\dot{\tilde{\boldsymbol{\theta}}}) = \frac{1}{a}(\tilde{\boldsymbol{\theta}}^{\mathrm{T}}(-a\boldsymbol{\sigma}(\boldsymbol{\sigma}^{\mathrm{T}}\tilde{\boldsymbol{\theta}} + \varepsilon_{\mathrm{HJB}}))) \leq \\
& \left(\frac{1}{a}\sigma_M^2 - \sigma_m^2\right)\|\tilde{\boldsymbol{\theta}}\|^2 + \frac{a}{4}\bar{\varepsilon}^2 \\
\dot{L}_2 =& 2\boldsymbol{x}^{\mathrm{T}}\dot{\boldsymbol{x}} + 2\Gamma\dot{V}(\boldsymbol{x}) = \\
& 2\boldsymbol{x}^{\mathrm{T}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}) + 2\Gamma(-\boldsymbol{x}^{\mathrm{T}}Q\boldsymbol{x} - \boldsymbol{u}^{\mathrm{T}}R\boldsymbol{u}) \leq \\
& (1 + \bar{\beta}^2 - 2\Gamma\lambda_{\min}(Q))\|\boldsymbol{x}\|^2 + \\
& \|\boldsymbol{u}\|^2 - 2\Gamma\lambda_{\min}(R)\|\boldsymbol{u}\|^2 + \kappa^2
\end{aligned}$$

由式 (14), 得到以下不等式:

$$\begin{aligned}
0 \leq& \|\boldsymbol{u}\|^2 = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\frac{1}{4}\|R^{-1}g(\boldsymbol{x})^{\mathrm{T}}V_{\boldsymbol{x}}\|^2 \leq \\
& \frac{1}{4}\|R^{-1}\|_F^2\|g(\boldsymbol{x})\|_F^2\|\Lambda(\boldsymbol{x})\|_F^2(\|\tilde{\boldsymbol{\theta}}\| + \|\boldsymbol{\theta}\|)^2 \leq U_M^2
\end{aligned}$$

其中, $U_M^2 = \frac{1}{4}D(\|\tilde{\boldsymbol{\theta}}\|^2 + 2\theta_M\|\tilde{\boldsymbol{\theta}}\| + \theta_M^2)$, 而 $D = \|R^{-1}\|_F^2\bar{\beta}^2$. 那么有:

$$\begin{aligned}
\dot{L} =& \dot{L}_1 + \dot{L}_2 + \dot{L}_3 \leq \\
& \left(\left(\frac{1}{a}\sigma_M^2 - \sigma_m^2 + \frac{1}{4}D\right)\|\tilde{\boldsymbol{\theta}}\| + \frac{1}{2}\theta_M D\right)\|\tilde{\boldsymbol{\theta}}\| + \\
& (1 + \bar{\beta}^2 - 2\Gamma\lambda_{\min}(Q))\|\boldsymbol{x}\|^2 + D_M
\end{aligned}$$

其中, $D_M = \frac{a}{4}\bar{\varepsilon}^2 + \kappa^2 + \frac{1}{4}D\theta_M^2$.

若 a 和 Γ 选择满足下面不等式:

$$a > \frac{\sigma_M^2}{\sigma_m^2 - \frac{1}{4}D}, \quad \Gamma > \frac{1 + \bar{\beta}^2}{2\lambda_{\min}(Q)}$$

并且式 (19) 和式 (20) 两个不等式也成立:

$$\|\tilde{\boldsymbol{\theta}}\| > \frac{\frac{1}{2}\theta_M D}{\sigma_m^2 - \frac{1}{a}\sigma_M^2 - \frac{1}{4}D} := b_{\tilde{\theta}} \tag{19}$$

$$\|\boldsymbol{x}\| > \sqrt{\frac{D_M}{2\Gamma\lambda_{\min}(Q) - 1 - \bar{\beta}^2}} := b_x \tag{20}$$

那么 $\dot{L} < 0$. 因此, 通过 Lyapunov 理论[33] 可以得出系统状态和权值估计误差是一致最终有界的 (UUB).

下面证明当 $t \to \infty$ 时, $\|\hat{\boldsymbol{u}} - \boldsymbol{u}^*\| \leq \varepsilon_u$. 回顾一下 $\boldsymbol{u}^*$ 的表达式 (6), 得到:

$$\hat{\boldsymbol{u}} - \boldsymbol{u}^* = -\frac{1}{2}R^{-1}g(\boldsymbol{x})^{\mathrm{T}}\Lambda(\tilde{\boldsymbol{\theta}} - \Delta\boldsymbol{\varepsilon}) \tag{21}$$

当 $t \to \infty$, 式 (21) 的上界是

$$\begin{aligned}
&\|\hat{\boldsymbol{u}} - \boldsymbol{u}^*\| \leq \\
&\quad \frac{1}{2}\|R^{-1}\|_F\|g(\boldsymbol{x})\|_F\|\Lambda\|_F\sqrt{(\tilde{\boldsymbol{\theta}}^{\mathrm{T}} - \Delta\boldsymbol{\varepsilon}^{\mathrm{T}})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\Delta\varepsilon})} \leq \\
&\quad \frac{1}{2}\|R^{-1}\|_F\bar{\beta}\sqrt{2\|\tilde{\boldsymbol{\theta}}\|^2 + 2\|\Delta\boldsymbol{\varepsilon}\|^2} \leq \varepsilon_u
\end{aligned}$$

其中, $\varepsilon_u = \frac{1}{2}\|R^{-1}\|_F\bar{\beta}\sqrt{2(b_{\tilde{\theta}}^2 + \varepsilon_{\Delta M}^2)}$. □

4.2 保守性分析

这里我们通过与神经网络模型的稳定性条件作比较, 对保守性进行了分析. 如果激励函数选作 $\phi(\boldsymbol{x})$, 且 $\|\Delta\phi(\boldsymbol{x})\|_F < \phi_M$, 例如在文献 [15−16] 中选为 $\phi(\boldsymbol{x}) = [x_1^2 \;\; x_1x_2 \;\; x_2^2 \;\; \cdots \;\;]^{\mathrm{T}}$, 那么有:

$$\|\tilde{\boldsymbol{\theta}}\| > \frac{\frac{1}{2}\theta_M D'}{\sigma_m^2 - \frac{1}{a}\sigma_M^2 - \frac{1}{4}D'} := b'_{\tilde{\theta}} \tag{22}$$

$$\|\boldsymbol{x}\| > \sqrt{\frac{D'_M}{2\Gamma\lambda_{\min}(Q) - 1 - \bar{\beta}^2}} := b'_x \tag{23}$$

其中, $D'_M = \frac{a}{4}\bar{\varepsilon}^2 + \kappa^2 + \frac{1}{4}D'\theta_M^2$, 而 $D' = \|R^{-1}\|_F^2\bar{\beta}^2\phi_M^2$.

对于系统, 若

$$1 \leq \phi_M \tag{24}$$

那么 $D_M < D'_M$, $b_{\tilde{\theta}} < b'_{\tilde{\theta}}$ 并且 $b_x < b'_x$.

下面可以得到 $\|\hat{\boldsymbol{u}} - \boldsymbol{u}^*\| < \varepsilon'_u$, 其中

$$\varepsilon'_u = \frac{1}{2}\|R^{-1}\|_F\bar{\beta}\sqrt{2(b_{\tilde{\theta}}^2\phi_M^2 + \varepsilon_{\Delta M}^2)}$$

如果式 (24) 成立, 则 $\varepsilon_u < \varepsilon'_u$. 这意味着本文方法使得 $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{x}}$ 和 $\hat{\boldsymbol{u}}$ 更能接近 $\boldsymbol{\theta}^*, \boldsymbol{x}^*$ 和 $\boldsymbol{u}^*$.

由于 $\|\Lambda(\boldsymbol{x})\|_F < 1 < \phi_M$, 所以在上一节的推导中去掉了 $\|\Lambda(\boldsymbol{x})\|_F$ 的上界表达 (因为上界是 1). 这样对于稳定性条件就会存在一个小的下界, 则系统就更容易被镇定, 权值估计 $\hat{\boldsymbol{\theta}}$ 也更接近最优值 $\boldsymbol{\theta}^*$. 因此如果式 (24) 成立, 我们提出的方法比文献 [15−16] 中方法有更小的保守性.

注 5. 神经网络中的激励函数导数的上界 ϕ_M 由系统的性质决定的. 但无论对于什么系统, 双曲正切函数导数的上界都是 1.

5　仿真与对比

考虑以下非线性系统 (1):

$$f(\boldsymbol{x})=\begin{bmatrix} -x_1^5+\tanh^3(x_2) \\ -0.5x_1-0.5x_2(1-(\cos(2x_1)+2)^2) \end{bmatrix}$$

$$g(\boldsymbol{x})=\begin{bmatrix} -2\sin(x_1) \\ 0 \end{bmatrix}$$, 并且在式 (2) 中 $Q=R=I$.

我们要达到的控制目的是使 $\boldsymbol{x}$ 收敛到 $\mathbf{0}$, 并且最小化性能指标 (2).

首先, 通过这个数值例子阐述该方法的有效性, 且为系统 (1) 设计了近似最优控制.

由式 (14), 利用 LIP-$\hat{\boldsymbol{\theta}}$ GFHM 设计了以下最优控制:

$$\boldsymbol{u}^*=-\frac{1}{2}\begin{bmatrix} -2\sin(x_1) & 0 \end{bmatrix}\times \tag{25}$$

$$\begin{bmatrix} \operatorname{sech}^2(x_1) & 0 \\ 0 & \operatorname{sech}^2(x_2) \end{bmatrix}\hat{\boldsymbol{\theta}} \tag{26}$$

从图 1 中可以看到 $\hat{\boldsymbol{\theta}}$ 的变化曲线. 125 s 后, $\hat{\boldsymbol{\theta}}$ 收敛到理想 (最优) 值. 图 2 描述了系统在控制 (25) 下的状态曲线. 125 s 后, 状态趋近平衡点 $\boldsymbol{x}=\mathbf{0}$.

下面用双网方法[15−16] 对同样系统进行仿真.

图 3 描述了状态变化曲线. 从图 4 和图 5 中, 我们看到 1 850 s 后, 可调权值收敛到了理想 (最优) 值.

通过仿真对比, 得到以下结论:

1) 显而易见, 本文方法的状态和可调权值收敛速度比双神经网络方法快很多;

2) 本例是针对 2 阶系统的仿真, 仅仅需要 2 个可调权值, 而双网的方法需要 6 个. 若对于高阶系统来说, 本文方法不但减少了存储空间, 而且很大程度上降低了计算负担.

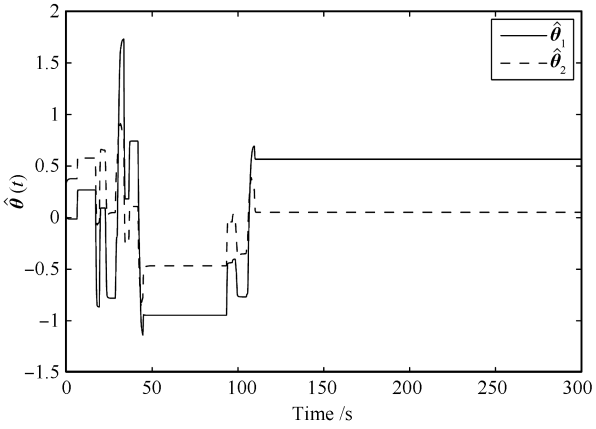


图 1　参数 $\hat{\boldsymbol{\theta}}$ 的收敛轨迹

Fig. 1　Convergence trajectory of parameter $\hat{\boldsymbol{\theta}}$

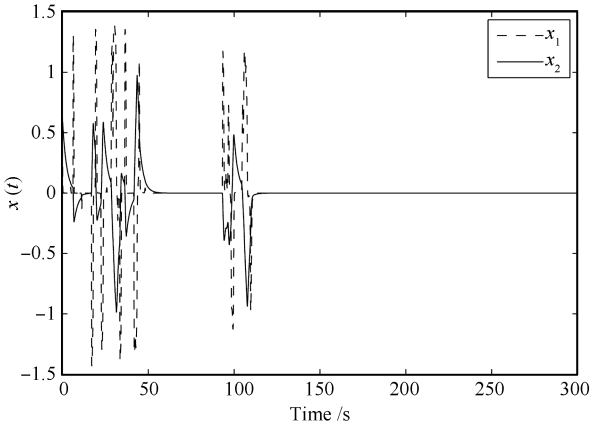


图 2　系统状态轨迹

Fig. 2　State trajectories of the system

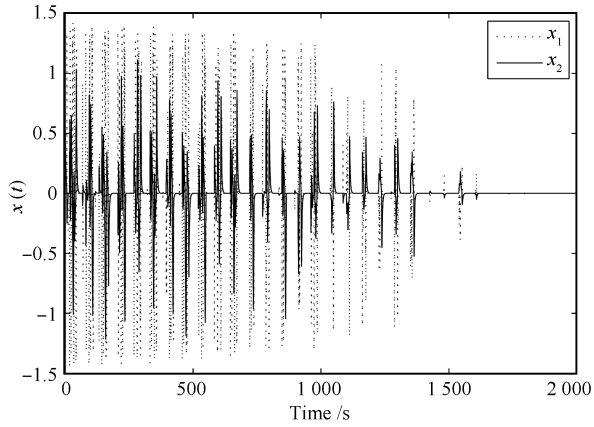


图 3　系统状态轨迹 (双网方法)

Fig. 3　State trajectories of the system (dual-neural-network method)

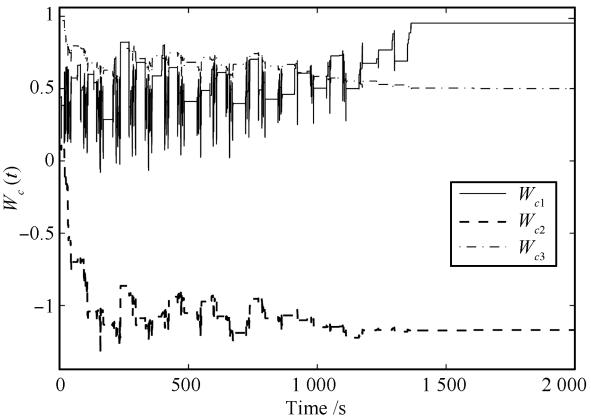


图 4　评价网参数收敛轨迹

Fig. 4　Convergence trajectories of the critic neural network parameter W_c

6 结论

本文提出了一种连续系统近似最优控制器的设计方法. 本文方法的主要思路是利用广义模糊双曲模型逼近 HJB 方程的解. 最后通过该逼近解获得最优控制. 一个数值例子证明了本文方法的有效性.

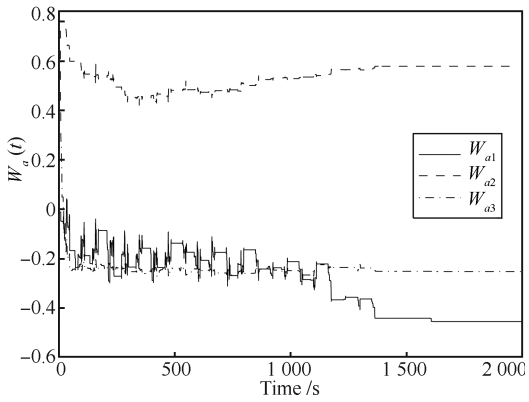


图 5 执行网参数收敛轨迹

Fig. 5 Convergence trajectories of the action neural network parameter W_a

致谢

作者衷心感谢东北大学信息科学与工程学院王占山教授和冯涛博士研究生的帮助.

References

1 Prokhorov D V, Wunsch D C. Adaptive critic designs. *IEEE Transactions on Neural Networks*, 1997, **8**(5): 997−1007

2 Murray J J, Cox C J, Lendaris G G, Saeks R. Adaptive dynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2002, **32**(2): 140−153

3 Wang F Y, Jin N, Liu D R, Wei Q L. Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ε-error bound. *IEEE Transactions on Neural Networks*, 2011, **22**(1): 24−36

4 Dierks T, Thumati B T, Jagannathan S. Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence. *Neural Network*, 2009, **22**(5−6): 851−860

5 Zhang H G, Luo Y H, Liu D R. Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. *IEEE Transactions on Neural Networks*, 2009, **20**(9): 1490−1503

6 Zhang H G, Cui L L, Luo Y H. Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2012, doi: 10.1109/TSMCB.2012.2203336

7 Wei Q L, Zhang H G, Cui L L. Data-based optimal control for discrete-time zero-sum games of 2-D systems using adaptive critic designs. *Acta Automatica Sinica*, 2009, **35**(6): 682−692

8 Wei Qing-Lai, Zhang Hua-Guang, Liu De-Rong, Zhao Yan. An optimal control scheme for a class of discrete-time nonlinear systems with time delays using adaptive dynamic programming. *Acta Automatica Sinica*, 2010, **36**(1): 121−129 (魏庆来, 张化光, 刘德荣, 赵琰. 基于自适应动态规划的一类带有时滞的离散时间非线性系统的最优控制策略. 自动化学报, 2010, **36**(1): 121−129)

9 Wei Q L, Liu D R. An iterative 2-optimal control scheme for a class of discrete-time nonlinear systems with unfixed initial state. *Neural Networks*, 2012, **32**: 236−244

10 Si J, Barto A G, Powell W B, Wunsch D. *Handbook of Learning and Approximate Dynamic Programming*. New York: Wiley, 2004

11 Kirk D E. *Optimal Control Theory: An Introduction*. New York: Dover, Inc., 2004

12 Lewis F L, Vrabie D. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits and Systems Magazine*, 2009, **9**(3): 32−50

13 Vrabie D, Pastravanu O, Abu-Khalaf M, Lewis F L. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 2009, **45**(2): 477−484

14 Vrabie D, Lewis F. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Networks*, 2009, **22**(3): 237−246

15 Vamvoudakis K G, Lewis F L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 2010, **46**(5): 878−888

16 Zhang H G, Cui L L, Zhang X, Luo Y H. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2226−2236

17 Haddad W M, Chellaboina V. *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. United Kingdom: Princeton University Press, 2008

18 Lewis F L. *Optimal Control*. New York: John Wiley and Sons, 1986

19 Sutton R S, Barto A G. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press, 1998

20 Al-Tamimi A, Lewis F L, Abu-Khalaf M. Model-free Q-learning designs for linear discrete-time zero-sum games with application to H_∞ control. *Automatica*, 2007, **43**(3): 473−481

21 Beard R W. Improving the Closed-Loop Performance of Nonlinear Systems [Ph. D. dissertation], Rensselaer Polytechnic Institute, Troy, 1995

22 Beard R W, Saridis G N, Wen J T. Approximate solutions to the time-invariant Hamilton-Jacobi-Bellman equation. *Journal of Optimization Theory and Applications*, 1998, **96**(3): 589−626

23 Abu-Khalaf M, Lewis F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 2005, **41**(5): 779−791

24 Wang D, Liu D R, Wei Q L. Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach. *Neurocomputing*, 2012, **78**(1): 14−22

25 Al-Tamimi A, Lewis F L, Abu-Khalaf M. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, **38**(4): 943−949

26 Finlayson B A. *The Method of Weighted Residuals and Variational Principles*. New York: Academic Press, 1972

27 Zhang H G, Quan Y B. Modeling identification and control of a class of nonlinear system. *IEEE Transactions on Fuzzy Systems*, 2001, **9**(2): 349−354

28 Kim Y H, Lewis F L, Dawson D M. Hamilton-Jacobi-Bellman optimal design of functional link neural network controller for robot manipulators. In: Proceedings of the 36th IEEE Conference on Decision and Control. San Diego, California USA, 1997, 2: 1038−1043

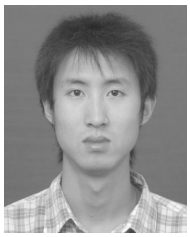
29 Wang L X, Mendel J M. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, 1992, **3**(5): 807−814

30 Lewis F W, Jagannathan S, Yesildirek A. *Neural Network Control of Robot Manipulators and Nonlinear Systems*. USA: Taylor and Francis, Inc., 1998

31 Hagan M T, Demuth H B, Beale M H. *Neural Network Design*. Boston, MA: PWS Publishing, 1996

32 Abdollahi F, Talebi H A, Patel R V. A stable neural network observer with application to flexible-joint manipulators. In: Proceedings of the 9th International Conference on Neural Information Processing (ICONIP′02). Singapore: IEEE, 2002, 4: 1910−1914

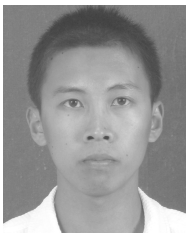
33 Khalil H K. *Nonlinear Systems*, Third edition. New Jersey: Prentice Hall, 2001



张吉烈 东北大学信息科学与工程学院博士研究生. 主要研究方向为模糊自适应动态规划和故障诊断.
E-mail: jilie0226@163.com
(**ZHANG Ji-Lie** Ph. D. candidate at the School of Information Science and Engineering, Northeastern University. His research interest covers fuzzy adaptive dynamic programming and fault diagnosis.)

张化光 东北大学信息科学与工程学院教授. 主要研究方向为神经网络控制和模糊控制. 本文通信作者.
E-mail: jilie0226@163.com
(**ZHANG Hua-Guang** Professor at the School of Information Science and Engineering, Northeastern University. His research interest covers neural-network-based control and fuzzy control. Corresponding author of this paper.)

罗艳红 东北大学信息科学与工程学院副教授. 主要研究方向为近似最优控制和神经网络控制.
E-mail: neuluo@gmail.com
(**LUO Yan-Hong** Associate professor at the School of Information Science and Engineering, Northeastern University. Her research interest covers approximate optimal control and neural network control.)



梁洪晶 东北大学信息科学与工程学院博士研究生. 主要研究方向为多智能体和复杂网络.
E-mail: lianghongjing99@163.com
(**LIANG Hong-Jing** Ph. D. candidate at the School of Information Science and Engineering, Northeastern University. His research interest covers multi-agent systems and complex networks.)