

含未知动态与扰动的非线性系统神经网络嵌入学习控制

马乐¹ 闫一鸣¹ 徐东甫¹ 李志伟^{1,2} 孙灵芳¹

摘要 针对带有不确定性与扰动的非线性系统的性能优化问题, 提出一种基于神经网络嵌入的学习控制方法. 对一类常见的 Lyapunov 函数导数形式, 将神经网络控制器集成到某种对系统稳定的基准控制器中, 其意义在于将原控制器改进为满足 Lyapunov 稳定的神经网络参数可调控制器, 从而能够利用先进的神经网络学习技术实现控制器的在线优化. 建立了跟踪误差的等效目标函数, 避免了对系统输入-输出的辨识问题. 建立了一种未知非线性与扰动等效值自适应方法, 并依此方法设计基准控制器. 以 RBF (Radial basis function) 反步自适应控制、基于卷积神经网络的滑模控制和深度强化学习控制为对比方法, 对带有死区、饱和、三角函数等数值与物理非线性模型进行仿真分析以测试方法有效性, 并针对上肢康复机器人控制问题进行虚拟实验以验证该方法的实用性. 仿真与实验结果表明, 该方法能在 Lyapunov 稳定条件下有效优化基础控制器性能, 对比结果证实了该方法的实用性与先进性.

关键词 神经网络嵌入, 优化控制, 深度学习技术, 未知非线性动态, 不确定与扰动

引用格式 马乐, 闫一鸣, 徐东甫, 李志伟, 孙灵芳. 含未知动态与扰动的非线性系统神经网络嵌入学习控制. 自动化学报, 2021, 47(8): 2016-2028

DOI 10.16383/j.aas.c200186

Neural Network Embedded Learning Control for Nonlinear System With Unknown Dynamics and Disturbance

MA Le¹ YAN Yi-Ming¹ XU Dong-Fu¹ LI Zhi-Wei^{1,2} SUN Ling-Fang¹

Abstract To address the problem of controlling performance optimization for the nonlinear uncertain system with disturbance, a neural network embedded learning control scheme is proposed in this paper. This method works on a common formal derivative of Lyapunov function, in which a neural network controller is integrated with a benchmark controller that is stable for the system. The main contribution of our work lies in that the benchmark controller is improved to a new one with tunable parameters under Lyapunov stability condition, and the new controller can be online optimized by using frontier technology of neural network. Hence an equivalent objective function based on tracking errors is characterized in this paper, avoiding identification to the relations between inputs and outputs of system. We use a value adaptive method for estimating equivalent term composed of unknown nonlinear function and disturbance, and the benchmark controller is designed based on this method. Some baseline methods are employed for comparison with the proposed method, which contain adaptive control based on RBF (Radial basis function)-backstepping, sliding mode control based on convolutional neural network and deep reinforcement learning control. And for verifying the effectiveness of our method we test some numerical and physical nonlinear model simulations, which contain trigonometric function saturation and dead zone nonlinearities. And virtual experiments of robot arm controlling of upper limb rehabilitation to be tested to verify the practicability of our method. These results show that the method proposed is able to optimize control performance of benchmark controller with Lyapunov stability. And the results of comparisons of tests show our method is efficient and advanced.

Key words Neural network embedded, optimized control, deep learning technology, unknown nonlinear dynamics, uncertainty and disturbance

Citation Ma Le, Yan Yi-Ming, Xu Dong-Fu, Li Zhi-Wei, Sun Ling-Fang. Neural network embedded learning control for nonlinear system with unknown dynamics and disturbance. *Acta Automatica Sinica*, 2021, 47(8): 2016-2028

收稿日期 2020-04-06 录用日期 2020-07-21

Manuscript received April 6, 2020; accepted July 21, 2020

国家自然科学基金 (61673101), 吉林重点行业与产业科技创新计划人工智能专项 (2019001090) 资助

Supported by National Natural Science Foundation of China (61673101), Special Foundation for Artificial Intelligence in Innovative Project of Science and Technology Key Industries of Jilin (2019001090)

本文责任编辑 王占山

非线性系统的高性能控制是控制科学与应用中

Recommended by Associate Editor WANG Zhan-Shan

1. 东北电力大学自动化工程学院 吉林 132012 2. 吉林省精密驱动智能控制国际联合研究中心 吉林 132012

1. School of Automation and Engineering, Northeast Electric Power University, Jilin 132012 2. Jilin Province International Research Center of Precision Drive and Intelligent Control, Jilin 132012

的重要研究问题^[1]. 非线性系统普遍存在于如电机力矩控制、康复系统人机交互控制等应用领域^[2-3]. 由于非线性系统的复杂性, 目前尚无如线性系统那样较为完善的分析与设计体系^[4]. 非线性系统的稳定性分析、不确定性补偿、扰动抑制与性能优化是其中难点问题. 特别是含有非线性动态与扰动因素的非线性系统控制性能优化问题亟待解决.

系统稳定性是非线性控制的首要问题. 基于 Lyapunov 控制 (Lyapunov-based control, LBC) 是非线性系统控制器分析与设计的重要方法^[5]. 其优势在于稳定性分析证明能伴随控制器设计同时生成, Backstepping 方法是其中代表方法^[6]. 文献 [7] 建立控制 Lyapunov 函数 (Control Lyapunov function, CLF), 并利用 Sontag 公式直接设计控制律, 但通常寻找 CLF 较为困难^[8]. 虽然基于 Lyapunov 的分析与设计已被广泛采用, 但其分析相对困难. 同时对于上述方法当考虑不确定性与扰动等因素后, 其稳定性分析的复杂度将陡增.

系统模型中的不确定性给分析与设计带来较大困难. 对于模型形式已知但参数未知的不确定问题, 参数自适应方法将参数误差视作时序量加入 Lyapunov 函数, 并建立参数自适应律, 实现对参数不确定性的自适应. 针对参数数量繁多或者模型形式未知问题, 文献 [9] 采用 RBF (Radial basis function) 神经网络作为系统模型中未知非线性项的逼近器, 将未知非线性模型的逼近问题转化为对 RBF 网络权值自适应逼近问题. 文献 [10] 采用自适应神经网络方法解决了切换互联系统的非线性不确定问题, 但未考虑控制增益的不确定性. 文献 [11] 将逼近器进一步改进为一种双环递归神经网络, 该模型考虑了逼近模型的内部时序状态, 试图增强网络逼近能力. 文献 [12] 采用模糊逻辑方法作为未知非线性项的逼近器. 虽然上述方法能有效逼近系统未知非线性项^[13], 但不确定性可能产生较大的初始逼近误差, 影响系统动态性能.

扰动可视为控制过程中产生的独立于系统模型的不确定因素. 虽然滑模与鲁棒等控制方法对一定范围内扰动具有抑制能力^[14], 但缺乏对扰动必要的量化与补偿机制^[15]. 基于扰动观测器 (Disturbance-observer-based control, DOBC) 控制方法则采用估计扰动并加以补偿的直接技术路线^[16]. 其中文献 [17] 针对线性系统采用不确定扰动估计器 (Uncertainty and disturbance estimator, UDE) 同时估计不确定性与扰动并加以补偿. 文献 [18] 在相关假设条件下, 设计了针对死区非平滑问题的扰动观测器. 扩张状态观测器 (Extended state observer, ESO) 提出了一种“总扰动”思想, 即模型未知动态、控制增益不

确定性与扰动的整体作用视作等效扰动加以估计^[19-20].

通过分析看出, 围绕上述问题展开的工作根本目的是保持控制系统稳定. 然而控制系统的暂态性能与稳态误差等因素也是控制品质的重要指标. 文献 [21] 将反步控制与最优控制结合, 针对模型确定且无扰动的高阶非线性系统逐步建立可学习的最优控制器. 文献 [22] 结合辨识方法提出了自适应迭代学习控制 (Iterative learning control, ILC) 方法, 实现了位值时变线性系统控制. 然而 ILC 需要重复任务条件并且对扰动较为敏感^[23]. 文献 [24] 采用 Hammerstein 神经网络作为辨识模型, 建立跟踪误差目标函数在线调节 PID (Proportion integral differential) 神经网络控制器, 然而辨识初始值和扰动等因素导致的辨识误差会影响性能, 甚至导致系统不稳定.

以上分析得出, 非线性系统中的动态不确定性与扰动问题对于控制性能影响较大, 目前研究仅集中在对上述因素的补偿与抑制方面, 对于带有不确定性与扰动的控制问题缺乏有效的性能优化手段. 另一方面, 近年来深度学习技术成绩斐然^[25], 利用深度学习提升控制性能将是积极有效的途径. 文献 [26] 采用卷积神经网络 (Convolutional neural network, CNN) 作为辨识器结合自适应控制实现了直升机控制, 在基于深度学习的性能优化方面做出了尝试. 文献 [27] 则采用卷积神经网络作为系统不确定项的逼近器并结合滑模控制实现自适应控制. 基于确定性策略梯度的深度强化学习 (Deep deterministic policy gradient, DDPG) 对于深度 Q 神经网络 (Deep Q network, DQN) 作出改进, 使其能适用于状态与控制输出均为连续值的控制问题^[28]. 文献 [29] 对两种深度强化学习控制的优化能力作出了量化比较. 然而大部分深度强化学习控制方法缺乏必要的稳定性分析, 在实际应用中存在隐患, 这也是该技术未能广泛用于实际控制的原因之一^[30].

从上述分析看出, 现有研究未能将深度学习充分应用于控制问题中, 主要原因如下: 1) 上述研究中深度神经网络的作用是作为未知动态的逼近器或辨识器, 仅将问题转为确定性系统控制, 因此控制性能的上限仅为基础控制器对于该确定性系统的性能; 2) 作为逼近器的深度神经网络被简化成了如 RBF 函数的线性基函数形式^[31], 其内部结构与参数未能对系统性能优化作出贡献; 3) 多数深度强化学习控制缺乏必要的稳定性分析, 难以保证实际应用稳定. 因此在稳定条件下, 如何充分灵活地利用神经网络的结构与优化优势提升非线性系统控制性能是值得深入研究的问题.

综上, 本文写作动机可描述为: 针对具有未知动态与扰动的非线性系统, 在少量假设条件下建立既确保 Lyapunov 稳定, 又简洁灵活的学习控制器, 同时该控制器能在无需辨识条件下利用深度学习技术进行在线优化以实现性能提升. 为此本文提出一种 Lyapunov 稳定的神经网络嵌入学习控制方法.

本文主要工作内容如下:

1) 提出了基于神经网络嵌入学习控制器设计方法, 在已知 Lyapunov 稳定控制器中嵌入神经网络控制器构成可优化的学习控制器, 并通过定理 1 证明新的控制器仍保证 Lyapunov 稳定. (详见第 1 节)

2) 利用导数等价条件, 建立改进的性能优化目标函数, 以用于学习控制器优化. 改进后的目标函数可规避辨识环节, 直接利用输出反馈优化控制器. (详见第 2 节)

3) 受文献 [19] 启发, 基于 Lyapunov 方法直接建立未知非线性动态、时变扰动与控制增益不确定的等效值自适应方法. (详见第 3 节)

本文主要创新贡献如下:

1) 不同于现有方法, 本文神经网络并非用于逼近系统非线性函数, 而是作为用于性能优化的控制分量. 在满足 Lyapunov 稳定条件下, 神经网络结构可任意构造与调整, 且无需额外的理论分析过程, 从而解放了神经网络控制对模型形式的束缚. 由于神经网络形式任意, 因此大量的深度神经网络模型可简单直接地嵌入控制器中, 并仍保持 Lyapunov 稳定, 所以本文方法粘合了深度神经网络技术和控制理论与应用的研究缝隙.

2) 改进的性能优化目标函数规避了辨识过程, 从而避免了辨识误差对控制的影响. 同时因上述的稳定性保证, 可在线实现目标函数优化.

3) 建立的值自适应方法相比传统方法, 无需 RBF 网络等模型作为逼近器, 同时具有更快的估计速度与精度.

1 神经网络 Lyapunov 稳定嵌入

设一类带有模型不确定性与扰动的 n 阶 m 维状态反馈系统为 $S: \mathbf{x} \times \mathbf{u} \rightarrow \mathbf{x}$, 其中 $\mathbf{x} \in \mathbf{R}^{n \times m}$ 为系统状态向量, $\mathbf{u} \in \mathbf{R}^m$ 为系统控制输入向量, 第 i ($i = 1, 2, \dots, n-1$) 阶系统为 $\dot{\mathbf{x}}^i = \mathbf{x}^{i+1}$, 第 n 阶为:

$$\dot{\mathbf{x}}^n = \mathbf{F}(\mathbf{x}) + \mathbf{b}\mathbf{u} + \mathbf{d} \quad (1)$$

式中, $\mathbf{F} \in \mathbf{R}^m$ 为未知非线性函数向量, $\mathbf{d} \in \mathbf{R}^m$ 为有界未知扰动向量, $\mathbf{b} > \mathbf{0}$ 为 m 阶未知可逆对角常数矩阵.

定理 1. 设 Lyapunov 函数 V 导数满足:

$$\dot{V} = \mathcal{B}^T (\mathcal{M} - \mathbf{F} - \mathbf{d} - \mathbf{b}\mathbf{u}) + \varphi \quad (2)$$

其中, \mathcal{B} 与 \mathcal{M} 为以状态 \mathbf{x} 为变量的 m 维已知函数向量, φ 为不含 \mathbf{u} 的其余项之和. 若满足: 1) 存在基础控制器 \mathbf{u}_b 使系统 Lyapunov 稳定, 即 $\dot{V} < 0$; 2) 存在向量函数 ϑ 满足 $\mathcal{B}_i \vartheta(\mathcal{B}_i) \geq 0$; 3) 存在神经网络 $\mu(\cdot|\theta) \geq \mathbf{0}$, 其中 θ 为神经网络全部可调参数构成的列向量, 则式 (3) 神经网络嵌入控制器 \mathbf{u}_b^μ 对系统 S Lyapunov 稳定.

$$\mathbf{u}_b^\mu = \mathbf{u}_b + \vartheta(\mathcal{B}) \circ \mu(\cdot|\theta) \quad (3)$$

其中, \circ 为 Hadamard 积运算符.

证明. 将式 (3) 代入式 (2) 得:

$$\begin{aligned} \dot{V} &= \mathcal{B}^T (\mathcal{M} - \mathbf{F} - \mathbf{d} - \mathbf{b}(\mathbf{u}_b + \vartheta(\mathcal{B}) \circ \mu(\cdot|\theta))) + \varphi = \\ &= \underbrace{\mathcal{B}^T (\mathcal{M} - \mathbf{F} - \mathbf{d} - \mathbf{b}\mathbf{u}_b) + \varphi}_{\dot{V}_{\mathbf{u}_b}} - \underbrace{\mathcal{B}^T \mathbf{b} \vartheta(\mathcal{B}) \circ \mu(\cdot|\theta)}_{\dot{V}_\mu} = \\ &= \dot{V}_{\mathbf{u}_b} - \mathbf{v}^T \mu(\cdot|\theta) \end{aligned} \quad (4)$$

其中, 列向量 $\mathbf{v}_i = \mathcal{B}_i \mathbf{b}_i \vartheta(\mathcal{B}_i)$, 根据定理 1 条件 1) \mathbf{u}_b 使系统 Lyapunov 稳定, 因此 $\dot{V}_{\mathbf{u}_b} \leq 0$; 根据定理 1 条件 2) 与 \mathbf{b} 定义得 $\mathbf{v}_i \geq 0$; 又由定理 1 条件 3) 得出 $\dot{V}_\mu \geq 0$, 因此 $\dot{V} \leq 0$. \square

注 1. 由式 (3) 看出, 本文学习控制器是基于已有 Lyapunov 稳定的控制器建立, 因此在应用中本文方法的稳定性分析难度较低. 并且控制器对神经网络的形式无具体要求, 仅需满足网络输出非负. 确保该条件相对容易, 仅需将网络输出层激活函数设置为非负函数即可.

控制器 \mathbf{u}_b^μ 的意义在于, 在基础控制器中嵌入了神经网络项, 且由定理 1 可知网络参数向量空间任意值均可保证 \mathbf{u}_b^μ 对上述系统稳定. 因此可建立表征系统控制性能的目标函数, 并采用适当优化技术更新网络参数以调节控制器输出, 从而优化控制性能. 同时由于稳定性保证上述操作可在线执行, 对于优化性能的效果, 可给出一个保守但重要的命题加以解释.

定理 2. 对于系统 S , 设有给定基础控制器 \mathbf{u}_b 与神经网络控制器 $\mu(\cdot|\theta)$, 根据式 (3) 构成学习控制器 \mathbf{u}_b^μ 的函数空间为 U_b^μ , 则对于任意表征系统控制性能的度量 $\mathcal{L}(S, \mathbf{u})$ (小值更优), 均满足:

$$\mathcal{L}(S, \mathbf{u}_b^{*\mu}) \leq \mathcal{L}(S, \mathbf{u}_b) \quad (5)$$

其中, $\mathbf{u}_b^{*\mu} \in U_b^\mu$ 为空间中对 \mathcal{L} 最优控制器.

证明. 显然, 存在神经网络 μ , 满足置零其全部参数, 使 $\mu(\cdot|\mathbf{0}) \equiv \mathbf{0}$, 因此根据式 (3) 得: $\mathbf{u}_b^{\mu(\cdot|\mathbf{0})} = \mathbf{u}_b \in U_b^\mu$, 并根据命题条件得: $\mathcal{L}(S, \mathbf{u}_b^{*\mu}) \leq \mathcal{L}(S, \mathbf{u}_b^{\mu(\cdot|\mathbf{0})}) = \mathcal{L}(S, \mathbf{u}_b)$. \square

注 2. 本文主旨并非求解最优控制器 $\mathbf{u}_b^{*\mu}$, 而是

通过式 (3) 建立满足式 (5) 的学习控制器。

2 神经网络学习控制器性能优化

第 1 节论述方法将深度学习及优化技术稳定地引入到非线性控制中. 得益于深度学习中自动求导技术与优化方法, 使研究者摆脱了繁杂参数推导, 研究者可针对问题本身专注于网络结构与优化问题的设计中. 因此神经网络学习控制器性能优化可归结为对控制性能目标函数的设计问题. 为此建立能表征系统控制性能的目标函数. 跟踪误差是控制问题的首要指标, 因此需讨论以其为优化项的目标函数:

$$\mathcal{L}_S(\boldsymbol{\theta}) = \frac{1}{2} \|\mathbf{y}_d - \mathbf{y}\|^2 + \psi \quad (6)$$

其中, \mathbf{y}_d 为系统当前期望, \mathbf{y} 为系统当前实际输出, ψ 为表征其他控制指标正则项, 且对控制器 \mathbf{u}_b^μ 梯度已知.

目前主流的神经网络优化方法为基于梯度的方法, 式 (6) \mathcal{L}_S 对 $\boldsymbol{\theta}$ 的梯度为:

$$\frac{\partial \mathcal{L}_S}{\partial \boldsymbol{\theta}} = (\mathbf{y} - \mathbf{y}_d) \frac{\partial \mathbf{y}}{\partial \mathbf{u}_b^\mu} \frac{\partial \mathbf{u}_b^\mu}{\partial \boldsymbol{\mu}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} + \frac{\partial \psi}{\partial \boldsymbol{\theta}} \quad (7)$$

式 (7) 中 $\partial \mathbf{y} / \partial \mathbf{u}_b^\mu$ 通常需建立 \mathbf{y} 与 \mathbf{u} 的输入-输出关系模型, 或对其辨识. 这样将存在建模误差, 特别对 $\boldsymbol{\theta}$ 为高维向量时影响较大. 根据式 (1) 对系统描述可发现 $\partial \mathbf{y} / \partial \mathbf{u}_b^\mu$ 正比于某小范数常值向量 $\boldsymbol{\rho} > \mathbf{0}$:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{u}_b^\mu} \propto \boldsymbol{\rho} \quad (8)$$

因此将式 (6) 修改为:

$$\mathcal{L}_S^g(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{y}_d)^\top (\boldsymbol{\rho} \circ \mathbf{u}_b^\mu) + \psi \quad (9)$$

其中, $\boldsymbol{\rho} > \mathbf{0}$ 为可调小范数常值参数向量. 根据式 (8) 得出, 关于 $\boldsymbol{\theta}$ 的梯度仅相差常数向量 $\boldsymbol{\varsigma}$, 即:

$$\frac{\partial \mathcal{L}_S}{\partial \boldsymbol{\theta}} = \boldsymbol{\varsigma} \circ \frac{\partial \mathcal{L}_S^g}{\partial \boldsymbol{\theta}} \quad (10)$$

基于梯度的优化方法在更新迭代时通常对梯度乘以某小值正数, 即学习律. 因此可通过调节学习律来消除式 (10) 所述差异带来的影响. 特别是对单输入-单输出系统, 式 (10) 中的梯度差异问题可完全折算到学习律的调节. 综上得出, 对于梯度优化方法, 式 (6) 问题可等价求解式 (9) 问题.

式 (9) 相对于式 (6) 优势在于不需对输入输出进行建模或辨识, 直接利用观测输出 \mathbf{y} , 避免了估计误差. 同时由于改进后的优化问题仅需输出反馈并可单值优化, 因此控制器支持在线学习.

注 3. 根据文献 [32] 结论, 对于单值优化问题,

学习律非敏感参数, 因此在应用中若适当选定学习率, 即便对于多输入多输出问题, 式 (10) 影响也可忽略.

3 未知非线性与扰动等效值自适应

前两节分析了 Lyapunov 稳定条件下神经网络嵌入与性能优化方法, 可将任意结构神经网络嵌入基础控制器中并利用其优化控制性能. 但本文方法的应用前提是存在使系统稳定的基础控制器. 由于式 (1) 描述系统中存在未知非线性项与未知控制增益等不确定性与扰动问题, 因此基础控制器设计中需对上述问题加以处理以确保基础控制器满足稳定条件.

为此本节建立一种未知非线性与扰动的等效值自适应方法. 该方法的特点是无需逼近器模型, 仅用值更新即可实现对未知时变不确定项的自适应.

注 4. 本节目的是建立对不确定性与扰动具有补偿与抑制能力的基础控制器, 上述方法并不依赖于本节内容. 即任何保证式 (1) 描述系统的其他控制器都可以结合上述方法构成神经网络学习控制器.

首先将式 (1) 变换为:

$$\dot{\mathbf{x}}^n = \mathbf{F}(\mathbf{x}) + (\mathbf{b} - \mathbf{b}_0)\mathbf{u} + \mathbf{d} + \mathbf{b}_0\mathbf{u} \quad (11)$$

其中, \mathbf{b}_0 为 m 阶已知对角常值矩阵. 并定义:

$$\mathcal{F} = \mathbf{F}(\mathbf{x}) + (\mathbf{b} - \mathbf{b}_0)\mathbf{u} + \mathbf{d} \quad (12)$$

类似文献 [10] 处理, 将 \mathcal{F} 视为系统不确定与扰动值项的等效值加以估计. 设对该时变值的估计值为 $\hat{\mathcal{F}}$, 则其估计误差为:

$$\tilde{\mathcal{F}} = \mathcal{F} - \hat{\mathcal{F}} \quad (13)$$

向 Lyapunov 函数中引入该误差项: $1/2(\tilde{\mathcal{F}}^\top \tilde{\mathcal{F}})$, 因此式 (2) 被具体化为 ($\varphi = \tilde{\mathcal{F}}^\top \dot{\tilde{\mathcal{F}}}$):

$$\dot{V}_{\mathcal{F}} = \mathcal{B}^\top (\mathcal{M} - \mathbf{F} - \mathbf{d} - \mathbf{b}\mathbf{u}) + \tilde{\mathcal{F}}^\top \dot{\tilde{\mathcal{F}}} \quad (14)$$

将式 (12) 与式 (13) 代入式 (14) 得:

$$\begin{aligned} \dot{V}_{\mathcal{F}} &= \mathcal{B}^\top (\mathcal{M} - \hat{\mathcal{F}} - \tilde{\mathcal{F}} - \mathbf{b}_0\mathbf{u}) + \tilde{\mathcal{F}}^\top \dot{\tilde{\mathcal{F}}} = \\ &= \mathcal{B}^\top (\mathcal{M} - \hat{\mathcal{F}} - \mathbf{b}_0\mathbf{u}) - \mathcal{B}^\top \tilde{\mathcal{F}} + \tilde{\mathcal{F}}^\top \dot{\tilde{\mathcal{F}}} = \\ &= \mathcal{B}^\top (\mathcal{M} - \hat{\mathcal{F}} - \mathbf{b}_0\mathbf{u}) + \tilde{\mathcal{F}}^\top (\dot{\tilde{\mathcal{F}}} - \mathcal{B}) \end{aligned} \quad (15)$$

分别对式 (11) ~ 式 (13) 求导得:

$$\begin{cases} \dot{\tilde{\mathbf{x}}}^n = \dot{\mathcal{F}} + \mathbf{b}_0\dot{\mathbf{u}} \\ \dot{\tilde{\mathcal{F}}} = \dot{\mathcal{F}} - \dot{\hat{\mathcal{F}}} \end{cases} \quad (16)$$

根据式 (15) 前一项, 并选取向量函数 $\boldsymbol{\omega}$ 满足 $\mathcal{B}^\top \boldsymbol{\omega}(\mathcal{B}) \geq 0$ (本文采用 \tanh 函数为 $\boldsymbol{\omega}$ 函数), 可建立基于值自适应方法的基础控制器 $\mathbf{u}_b^{\mathcal{F}}$:

$$u_b^F = b_0^{-1} (\mathcal{M} - \hat{\mathcal{F}} + \varpi(\mathcal{B})) \quad (17)$$

式 (17) 使式 (15) 前项小于等于 0, 因此只需满足:

$$\dot{\hat{\mathcal{F}}} - \mathcal{B} = 0 \quad (18)$$

便可保证 $\dot{V}_F \leq 0$. 将式 (16) 代入式 (18) 并求解 $\dot{\hat{\mathcal{F}}}$, 得到 $\hat{\mathcal{F}}$ 的自适应律:

$$\dot{\hat{\mathcal{F}}} = \ddot{x}^n - b_0 \dot{u}_b^F - \mathcal{B} \quad (19)$$

注 5. 注意到 $\hat{\mathcal{F}}$ 自适应律中采用了信号微分项, 虽然文献 [33-34] 中证明了该方法的合理性, 但考虑实际信号中的噪声问题, 本文采用适当微分器加以处理. 关于微分器的收敛性可参见文献 [35].

综上, 本文提出的针对未知非线性动态与扰动系统的神经网络嵌入学习控制器表示如下:

$$\begin{cases} u_b^F(\cdot|\theta) = u_b^F + \vartheta(\mathcal{B}) \circ \mu(\cdot|\theta) \\ u_b^F = b_0^{-1} (\mathcal{M} - \hat{\mathcal{F}} + \varpi(\mathcal{B})) \\ \dot{\hat{\mathcal{F}}} = \ddot{x}^n - b_0 \dot{u}_b^F - \mathcal{B} \end{cases} \quad (20)$$

4 仿真分析

本节以带有三角函数、死区与摩擦特性的一、二阶非线性数值模型与实际物理模型进行仿真测试. 全部仿真基于 Python 框架, 采用 Pytorch 作为深度神经网络库. 本节与经典 RBF 自适应控制和文献 [13] 中自适应方法作对比, 以测试本文方法性能. 仿真测试中神经网络结构与相关参数设置见附录.

4.1 数值模型算例

算例 1. 考虑一阶非线性系统:

$$\dot{x} = a \sin(x)x^2 + bu \quad (21)$$

其中, $a = 3.0$, $b = 1.0$. 设计基础控制器:

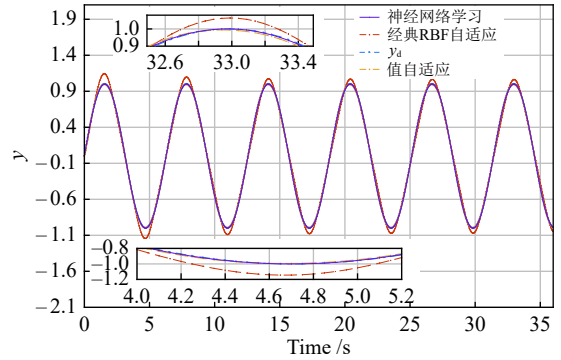
$$\begin{cases} u_b = \frac{\dot{y}_d - \hat{f} - ks}{b_0} \\ s = y_d - x_1 \end{cases} \quad (22)$$

其中, y_d 为控制期望.

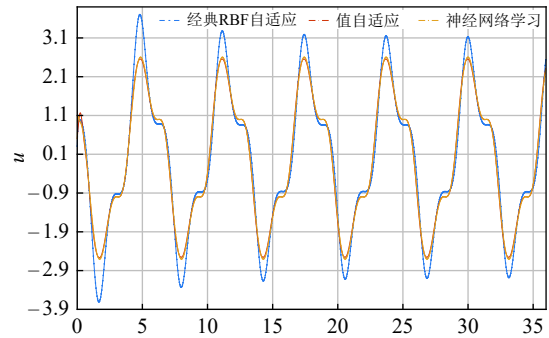
基于 u_b 分别设计经典 RBF 自适应、值自适应、神经网络学习三种控制器 (在值自适应基础上嵌入神经网络控制器), 并比较仿真结果. 其中统一取 $k = 36$, $b_0 = 2.0$, RBF 核函数 $\sigma = 1$, 在 $[-6, 6]$ 范围内等分取 100 个核函数中心值, 值自适应与神经网络学习控制采用自适应方法估计 \hat{f} 值, 经典 RBF 自适应方法则采用径向基函数估计.

图 1 为期望轨迹 $y_d = \sin(t)$ 三种控制方法的仿真对比结果, 以平均绝对误差 (Mean absolute error, MAE) 为评价指标. 图 1 结果显示: 1) 虽然经

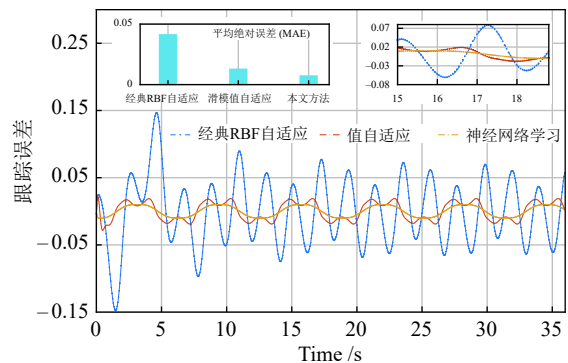
典 RBF 方法的输出响应与控制输入随时间增加逐渐趋于本文方法, 但本文的值自适应与神经网络学习方法的跟踪效果明显优于经典 RBF 方法. 原因在于未知非线性函数导致 RBF 初始估计与实际差别较大, 同时 RBF 估计收敛速度低于值自适应方法, 因此得出对于上述系统本文方法优于经典 RBF 方法. 2) 值自适应与神经网络学习控制方法的输出基本一致, 但后者的跟踪误差低于前者, 因此得出对于上述系统神经网络学习控制方法能根据



(a) 算例 1 系统响应曲线
(a) Responses of the Example 1



(b) 算例 1 控制输入
(b) Control inputs of the Example 1



(c) 算例 1 跟踪误差
(c) Tracking errors of the Example 1

图 1 算例 1 控制性能结果

Fig. 1 The controllers performances of the Example 1

跟踪误差目标函数调节神经网络, 实现控制性能优化.

算例 2. 再考虑带有三角函数与死区的二阶非线性系统:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f(x_1, x_2) + bu + d \\ f = \begin{cases} 0, & |x_1| < \delta \text{ or } |x_2| < \delta \\ a_1 \cos(x_1 x_2) + a_2 x_2 \sin(x_1), & \text{其他} \end{cases} \end{cases} \quad (23)$$

其中, $\delta = 0.1$, $a_1 = 3.0$, $a_2 = 6.0$, $b = 3.0$.

根据式 (20) 设计基础控制器:

$$\begin{cases} u_b = \frac{\ddot{y}_d - \hat{f} - ks}{b_0} \\ s = ke + \dot{e} \end{cases} \quad (24)$$

其中, $e = y_d - y$. 本算例中统一取 $k = 10$, $b_0 = 5$, RBF 核函数 $\sigma = 1$, 在 $[-12, 12]$ 范围内分别对 x_1 与 x_2 进行 10 等分取值, 即取 100 个 c 值, 其他参数同上例.

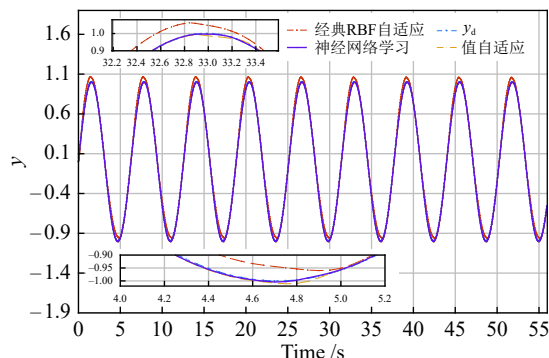
图 2 为 $y_d = \sin(t)$, $d = 0$ 三种控制方法的仿真对比结果. 图 2 结果显示: 1) 三种方法均能以较高精度实现式 (23) 系统的轨迹跟踪, 由图 2 (a) 看出不同于前一算例, RBF 方法在波峰处输出值高于期望, 在波谷处低于期望, 而其他两种方法则与其相反, 原因在于 RBF 对 \hat{f} 的估计方式与本文提出方法不同, 加之与前一算例的模型差异导致图 2 (a) 效果; 2) 图 2 (b) 中的控制输入 u 的峰值与频率均高于前一算例且出现震荡, 原因在于本算例系统阶数增高且多出死区非线性, 进而导致上述现象; 3) 三种方法的控制输出幅值大体一致, 跟踪误差排序仍为经典 RBF > 值自适应 > 神经网络学习控制. 综上得出, 对于带有三角函数与死区的二阶非线性系统, 本文方法效果优于经典 RBF 自适应方法, 且神经网络学习控制方法能在不明显提升控制输出条件下优化基准控制器性能.

算例 3. 上述模型加入如下扰动:

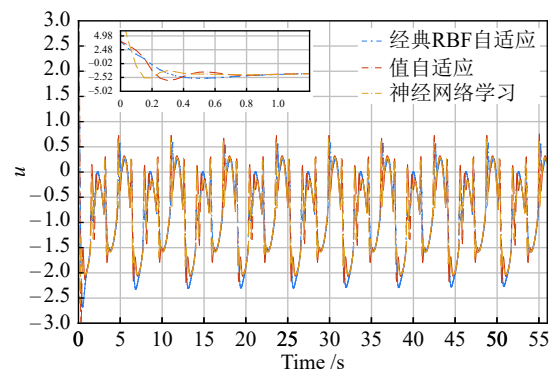
$$d(t) = G_d \sin(\omega_d t) + \xi(t) \quad (25)$$

其中, $G_d = 100.0$, $\omega_d = 10.0$, $\xi(t)$ 为 $[-30, 30]$ 随机量 (三种方法 $\xi(t)$ 相同). 沿用前算例控制器进行仿真测试以比较三种方法在扰动条件下的性能, 其中控制参数取 $k = 35$, $b_0 = 1.0$.

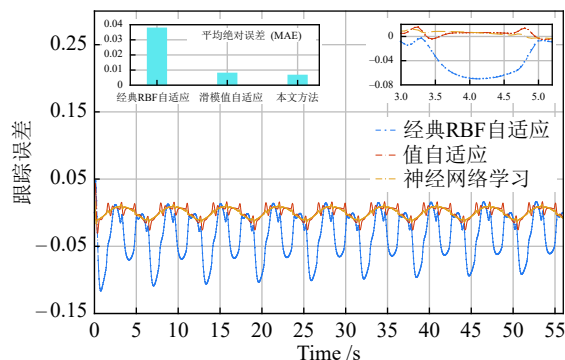
图 3 为 $y_d = \sin(t) + \sin(0.5t)$ 三种控制方法的仿真对比结果. 图 3 结果显示: 1) 在强扰动条件下三种方法均实现对 y_d 的稳定跟踪, 但三种控制输出均出现随机震荡, 如图 3 (a), 其原因在于施加的扰动中含有较大随机扰动; 2) 图 3 (b) 中的控制输入 u 幅值范围与频率均高于前一算例且震荡, 原因在于 y_d 高于前一算例且加入了高频扰动, 同时为了抑制强扰动, 控制增益高于前例; 3) 三种方法控制输出幅值大体一致, 跟踪误差排序仍为经典 RBF >



(a) 算例 2 系统响应曲线
(a) Responses of the Example 2



(b) 算例 2 控制输入
(b) Control inputs of the Example 2



(c) 算例 2 跟踪误差
(c) Tracking errors of the Example 2

图 2 算例 2 控制性能结果

Fig. 2 The controllers performances of the Example 2

值自适应 > 神经网络学习控制. 由此表明, 在强扰动条件下本文方法对不确定性与扰动的补偿和抑制能力优于经典 RBF 自适应方法, 同时本文提出的嵌入神经网络控制器仍能优化基准控制器性能.

4.2 物理模型仿真

本节采用与第 4.1 节不同结构的神经网络作为嵌入控制器 (详见附录描述), 对实际物理模型测试, 并与文献 [13] 方法 (一种基于 Backstepping 方法的神经网络自适应控制方法, 下称“对比方法”) 作

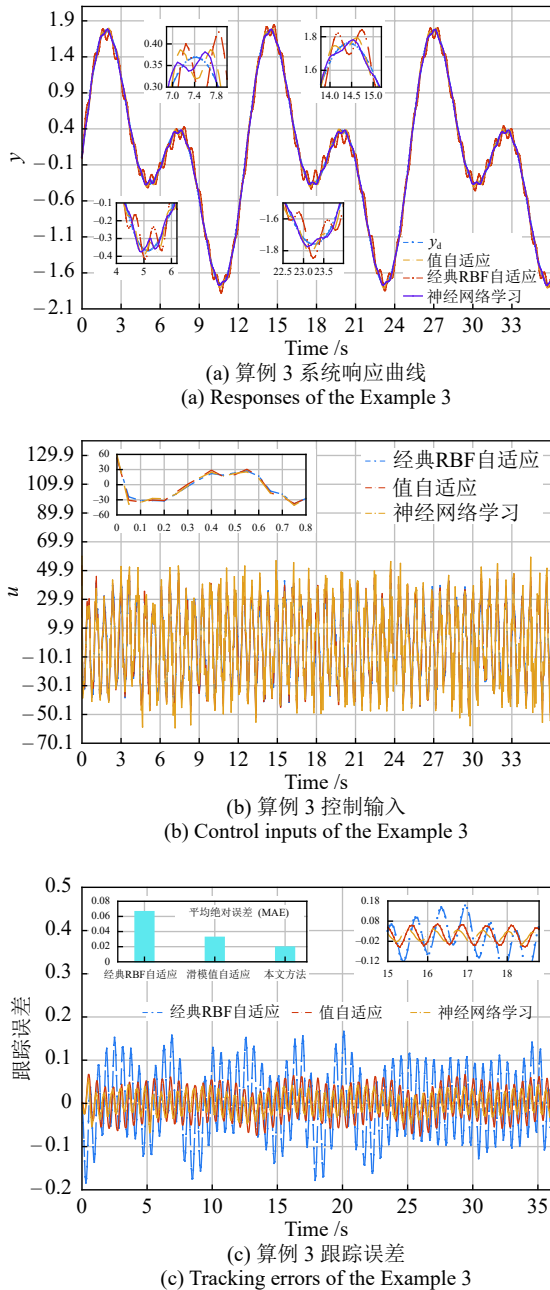


图 3 算例 3 控制性能结果

Fig. 3 The controllers performances of the Example 3

对比以验证本文方法的先进性.

算例 4. 指数摩擦特性的电机转矩控制模型如下:

$$\begin{cases} \dot{\theta} = \omega \\ \dot{\omega} = f_M(\theta, \omega) + \frac{K + K_\theta}{T} \tau \\ f_M(\theta, \omega) = \begin{cases} 0, & |\theta| < \sigma \text{ or } |\omega| < \sigma \\ -\frac{1}{T} \omega + \frac{K + K_\theta}{T} T_f, & \text{其他} \end{cases} \\ T_f = T_c \text{sgn}(\omega) + (T_s + T_c) e^{-\alpha|\omega|} \text{sgn}(\omega) \end{cases} \quad (26)$$

其中, θ 为电机转角 (单位 rad), ω 为其角速度 (单

位 rad/s), τ 为控制输入 (单位 N·m), f_M 为其非线性项, T_f 为摩擦力. 模型参数为: $K = 2.97$, $K_\theta = 0.25$, $T = 0.632$, $T_c = 0.2 \text{ N} \cdot \text{m}$, $T_s = 0.3 \text{ N} \cdot \text{m}$, $\alpha = 1.0$. 施加如式 (25) 扰动, $G_d = 10.0$, $\omega_d = 1.0$, ξ 为 $[-10, 10]$ 随机数.

沿用前节方法构造本文方法的基础控制器, $b_0 = 1.0$, $k = 50.0$ (根据对比方法设置的参数 k 值), 神经网络结构与参数见附录. 对于算例 4, 文献 [13] 方法性能趋于最佳的参数设置如下: 在 $[-10, 10]$ 范围内分别对 x_1, x_2 进行 36 等分取值, 即取 1296 个值作为对比方法中的 RBF 神经网络节点中心值, 设置对比方法的 α_1, α_2 分别为 30.0 与 50.0, 其余参数见文献 [13].

图 4 为 $y_d = \pi \sin(t)$ (rad) 两种方法的控制效果对比. 从中看出: 1) 图 4 (a) 显示两种方法均能快速稳定实现轨迹跟踪, 然而对比方法在初始阶段较本文方法有较大误差 (见第一次波峰波谷); 2) 图 4 (b) 显示两种方法的控制输入基本一致, 本文方法较对比方法无明显提升; 3) 图 4 (c) 显示本文方法的跟踪误差总体上低于对比方法. 因此可得出本文方法对于算例 4 的控制效果整体优于对比方法.

对比两种方法的最大误差 ($\max|e|$)、平均误差 ($\text{mean}|e|$)、最大控制输出 ($\max|u|$)、平均控制输出 ($\text{mean}|u|$)、最大不确定与扰动估计误差 ($\max|\tilde{f}|$) 与平均不确定与扰动估计误差 ($\text{mean}|\tilde{f}|$) 6 种量化指标, 如表 1. 从表 1 看出本文方法 $\max|e|$ 低于对比方法 41.63%, $\text{mean}|e|$ 低于对比方法 41.66%, $\max|u|$ 低于对比方法 2.63%, $\max|\tilde{f}|$ 低于对比方法 11.32%, $\text{mean}|\tilde{f}|$ 低于对比方法 11.14%, 仅 $\text{mean}|u|$ 略高于对比方法 2.58%.

以上分析得出本文方法的控制性能整体优于对比方法, 原因如下: 1) 对比方法 RBF 网络仅为对不确定项的逼近, 控制性能上限取决于 Backstepping 控制器, 同时初始时刻的未知动态与网络估计值有较大差距, 因此导致对比方法在第一次波峰、波谷阶段的误差较大; 2) 本文方法的值自适应方法能够较为快速地估计不确定性与扰动, 因此在初始时刻误差低于对比方法; 3) 本文方法的神经网络控制器中设计了前馈结构对控制性能有积极贡献, 本文优化方法能够有效地调节网络参数与输出从而优化控制性能.

综合以上算例得出, 本文提出的值自适应方法相比经典 RBF 方法对三角函数、死区与摩擦非线性和强扰动未知不确定系统具有更好的模型估计与扰动抑制能力, 嵌入神经网络控制器能在不提升控制输出幅值条件下优化基准控制器性能. 性能优化

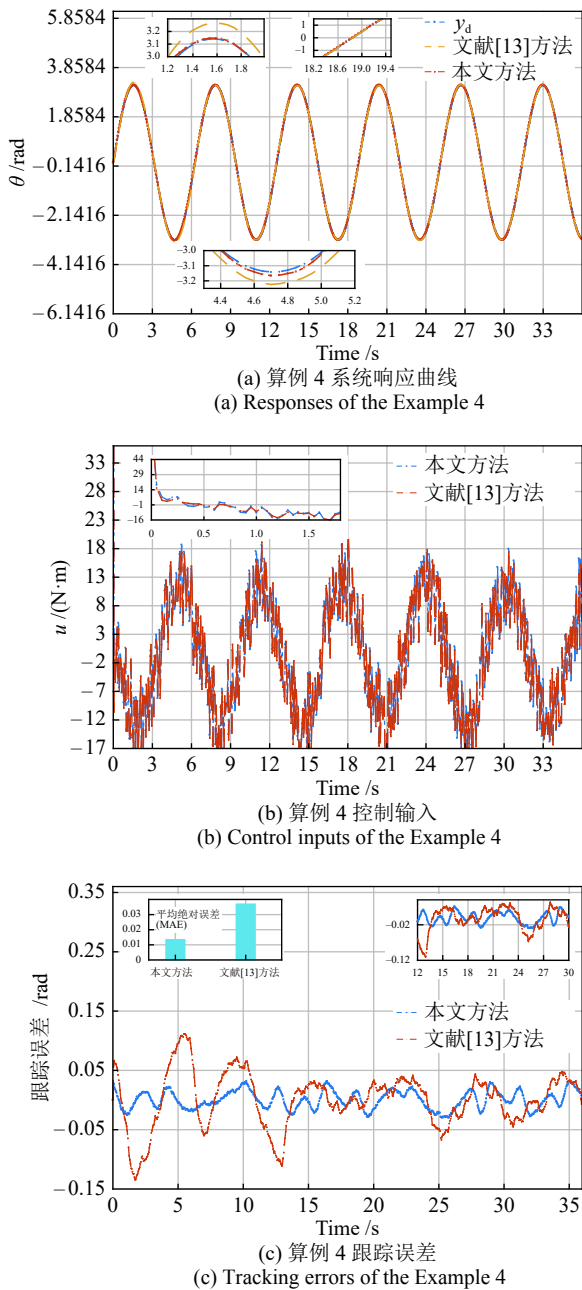


图 4 算例 4 对比实验控制性能结果

Fig. 4 The results for comparison test of control performances of the Example 4

表 1 算例 4 两种方法控制性能统计数据对比
Table 1 The comparison for control statistical indicators of two methods in the Example 4

方法	$\max e $	$\text{mean} e $	$\max u $
文献 [13]	0.1350	0.0360	83.6962
本文	0.0788	0.0210	81.4951
方法	$\text{mean} u $	$\max \dot{e} $	$\text{mean} \dot{e} $
文献 [13]	7.6009	15.8421	5.1999
本文	7.7972	14.0492	4.6208

结果充分证实了第 2 节中提出的梯度等效优化方法的合理性与有效性. 与对比方法的对比实验结果表明, 本文方法在对不确定与扰动的补偿与抑制和控制性能优化方面具有一定的先进性. 算例中采用不同结构的神经网络构建学习控制器, 体现出基于本文方法引入的深度学习及优化技术可针对具体控制问题灵活地设计与调整网络结构.

5 虚拟实验分析

前节已对本文方法控制性能作出较充分的比较分析, 因此本节专注验证本文方法解决实际问题的有效性. 上肢康复机器人控制问题中涉及不确定性、扰动与性能优化等方面, 因此选择该问题为应用实例, 针对神经网络的在线调节与训练后运行两种方式的控制效果进行对比分析. 本文基于 CoppeliaSim 物理模拟系统搭建虚拟实验平台, 如图 5 所示, 本文中仿真控制周期为 0.01 s, 物理引擎选择 Bullet 2.87, 精度设置为“最高精度”. 以下虚拟实验基于 Python 3.7 的控制周期均小于 10 ms, 即所有实验的控制频率可达到 100 Hz 以上.

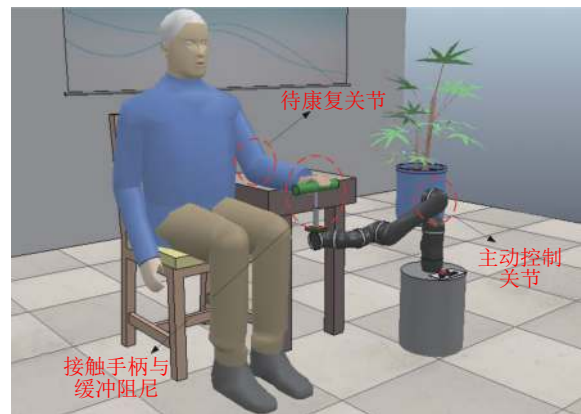


图 5 CoppeliaSim 虚拟实验示意图

Fig. 5 The demonstration of virtual experiment in CoppeliaSim

采用系统提供的 Python 远程同步 API (Application programming interface) 方式实现控制. 对于控制端物理虚拟系统模型及参数未知, 因此机械臂关节控制模型表示为:

$$\begin{cases} \dot{\theta}_j = \omega_j \\ \dot{\omega}_j = f_j + b_j \tau_j + d_j \end{cases} \quad (27)$$

其中, θ_j 为康复机器人主动关节角度 (单位 rad); ω_j 为其角速度 (单位 rad/s); f_j 为控制系统未知非线性动态项; b_j 为关节未知转动惯量 (单位 $\text{kg} \cdot \text{m}^2$); τ_j , d_j 为关节控制力矩与未知扰动项. 选定同前节的基础控制器, 分情况测试分析实际系统控制效果.

5.1 不同体重康复者测试

由于用户个体差异, 上肢体积与重量不尽相同, 以至施加于系统的负载不同. 分别选取偏瘦、中等、偏胖三种身形用户, 测试在不同负载条件下的控制性能. 设主动轴角度期望轨迹为 $\theta_1^d(t) = 0.3 \sin(t) + y_0$, $y_0 = 2.7$, 训练方式经 5 min 运行完成神经网络训练, 两种方式测试时间为 1 min.

本节基础控制器沿用式 (22), 其中 $k = 30.0$, $b_J = 16.0 \text{ kg} \cdot \text{m}^2$, 网络结构与参数见附录. 图 6 为不同身形用户两种方式的跟踪误差 MAE ($1/N \sum |e|$) 与控制输入幅值 MAE ($1/N \sum |u|$) 的统计图. 从中可得出: 1) 两种方式对于不同身形用户测试中均有较低的跟踪误差 MAE, 并且经训练后较在线方式有更低的控制误差; 2) 两种方式的控制输入幅值基本相等, 表明训练过程没有明显提升控制输入幅值.

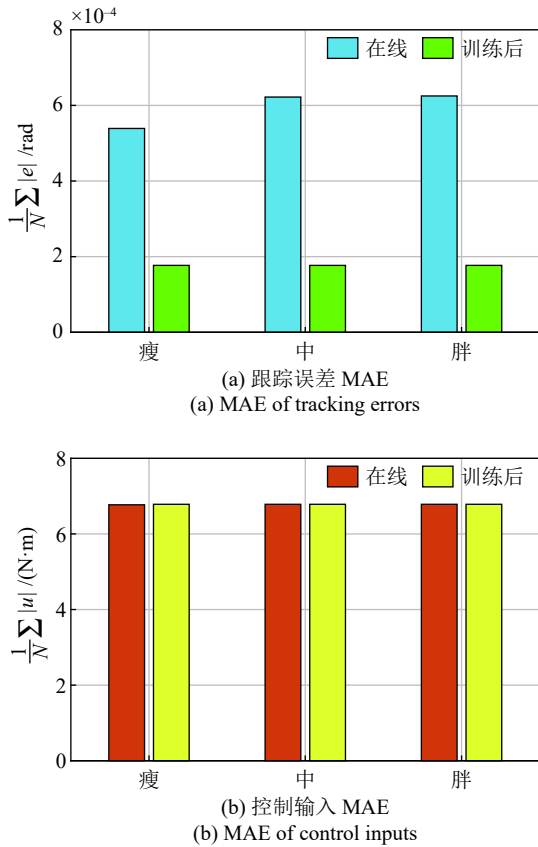


图 6 不同体重康复者测试跟踪误差与控制输入 MAE
Fig.6 The MAE of tracking errors and control inputs for tests to rehabilitation clients with different weights

5.2 不同康复任务性能测试

不同康复人群需制定不同的方案以达到康复目的的, 为此选择两种关节康复轨迹测试不同康复任务的系统控制性能. 设任务 1 的康复轨迹为 $\theta_{1,1}^d(t) =$

$0.2 \sin(t) + y_0$, $y_0 = 2.7$; 设任务 2 的康复轨迹为 $\theta_{1,2}^d(t) = \min(\max(\theta_{1,1}^d(t), \delta), -\delta)$, $\delta = 3.0$.

图 7 为不同任务轨迹的两种方式跟踪误差 MAE 与控制输入幅值 MAE 统计图. 对于两种任务轨迹, 从图中可得与前例类似结论. 此外两种方式对于任务 2 的误差高于任务 1, 原因之一在于任务 2 在任务 1 基础上增加了饱和条件, 因此饱和段跟踪误差会有所提高, 可加大训练时长提升控制精度.

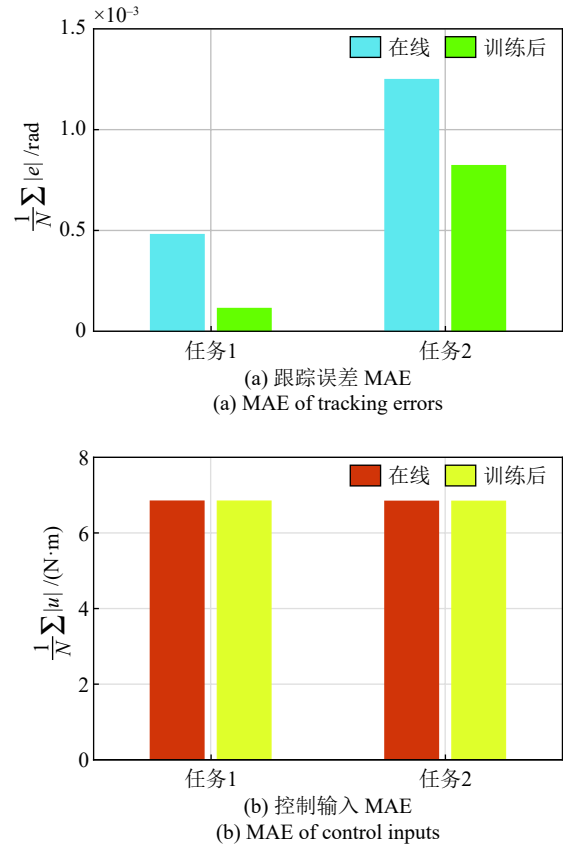


图 7 不同康复任务测试跟踪误差与控制输入 MAE
Fig.7 The MAE of tracking errors and control inputs for tests to different rehabilitation tasks

5.3 用户异步扰动测试

康复训练过程中, 用户可能自发运动上肢关节, 该运动可视为康复机器人系统扰动, 或导致与系统出现异步相位差. 本例在虚拟康复运动中施加用户关节动态力矩, 该力矩通过手柄传导作用于机器人主动关节. 设用户上肢关节力矩为 $\tau_d = 0.1 \sin(t)$ ($\text{N} \cdot \text{m}$), 期望轨迹为 $\theta_1^d = 0.2 \sin(t)$ (rad).

本节采用文献 [27] 中的基于卷积神经网络 (CNN) 控制和文献 [28] 中的深度强化学习控制 (Deep reinforcement learning, DRL) 方法作为对比方法, 测试用户异步扰动问题的控制性能.

为使对比结果更具说服力, 本测试中三种方法所需的神经网络被设置成相同的结构与参数. 均采用如文献 [27] 中描述的两层卷积神经网络. 输入统一设置成由状态向量时序组成的矩阵, 第 i 行状态向量为: $[\theta(t - \zeta_i), \dot{\theta}(t - \zeta_i), \theta_d(t - \zeta_i), \dot{\theta}_d(t - \zeta_i), \tau(t - \zeta_{i+1})]$, 其中 $\zeta_i = \Delta t(i - 1)(s)$, $\Delta t = 0.01 s$, $i = 1, 2, 3, \dots, N$ 为延迟时间常数. 本文 $N = 6$, 因此本文卷积网络的输入维数为 6×5 矩阵, CNN 两层卷积核数均为 10, 卷积核大小为 3×3 . 此外本文方法其他参数同前节.

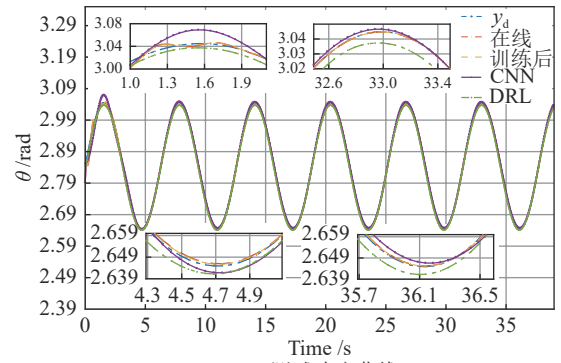
文献 [30] 中 DRL 方法的单次学习批数 $N_{\text{Batch}} = 50$, 奖励函数 γ 设计如式 (28), CNN 与 DRL 方法其他相关参数设置见文献 [27-28].

$$\gamma = \begin{cases} 1 - |e|, & |e| < 0.5 \\ -1, & |e| \geq 0.5 \end{cases} \quad (28)$$

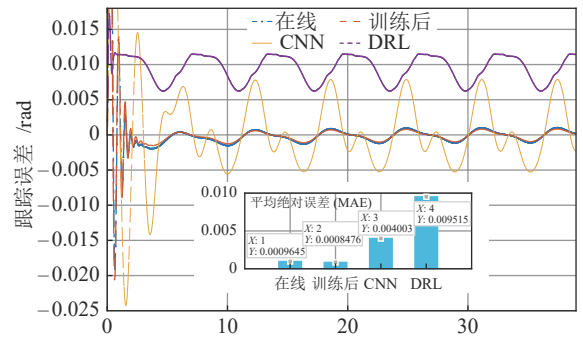
其中, $e = \theta^d - \theta_j$.

图 8 为带有康复者关节扰动的机器人控制对比实验结果. 需说明: 1) 因各方法的最大与平均控制转矩输出的绝对值基本相等, 因此未出图展示; 2) 图中显示的 DRL 控制曲线是经多次学习并收敛后 ($\text{mean}\gamma > 0.99$) 的效果. 从图 8 (a) 看出所有方法均能快速稳定地实现期望轨迹跟踪, 但从第一次波峰与波谷处曲线看出, CNN 与 DRL 方法相对本文方法具有较大误差, 虽然本文方法在第一次波峰处有震荡, 但跟踪误差均小于对比方法. 从末次波峰、波谷处曲线看出, DRL 方法的跟踪误差仍明显高于其他方法, 虽然 CNN 方法比第一次波峰降低了误差, 但相对本文方法仍有一定差距. 从图 8 (b) 看出, 由于初始时刻的系统不确定性与扰动作用, 所有方法均出现不同程度的抖动, 但本文方法整体误差最小, CNN 居中, DRL 方法则出现了较大的单向误差 (误差值全为正). 图 8 (b) 中的 MAE 指标 (单位 rad) 分别为: 本文方法在线效果 9.645×10^{-4} , 本文方法训练后效果 8.476×10^{-4} , CNN 方法 4.003×10^{-3} , DRL 方法 9.515×10^{-3} , 因此可看出本文方法控制精度明显优于对比方法, 且经训练后的控制精度高于在线精度.

上述对比实验结果分析如下: 1) DRL 方法的控制精度最低源于方法本身固有的优化瓶颈问题, 当达到一定精度后因平均奖励函数趋于上限 ($\text{mean}\gamma > 0.99$), 导致精度提升逐渐缓慢, 同时因缺乏稳定性保障, DRL 方法探索学习具有一定的随机性, 在实际应用中难免发生失控问题, 该随机性或造成了单向误差现象; 2) 虽然 CNN 相对 RBF 具有模型优势, 但文献 [28] 中的 CNN 方法仅用其逼近不确定项, 未充分发挥 CNN 模型上的优势, 因此 CNN 方法



(a) 测试响应曲线
(a) Responses of test



(b) 跟踪误差
(b) Tracking errors

图 8 带有康复者关节扰动的机器人控制对比实验结果
Fig.8 Comparison results of robot control methods for joint disturbances created by rehabilitation client

性能上限仅为基础控制器对确定性系统的控制性; 3) 本文方法具有稳定性保障和性能优化手段, 因此能在使用相同结构与参数的 CNN 网络构造控制器条件下, 呈现较高精度效果. 综上, 由对比实验结果与分析得出, 本文方法相较其他深度学习控制方法具有一定的先进性.

由本节三类虚拟实验结果得出, 对上肢康复机器人控制系统问题, 本文方法在系统非线性动态完全未知条件下对于用户体重、任务类型与关节扰动等不确定非线性问题具有良好的控制性能, 从而验证了本文方法应用于实际问题的有效性. 经过训练后的控制效果均优于在线更新, 进一步证实了本文提出的梯度等效优化方法的有效性. 因此可推测, 实际应用中本文方法可通过神经网络学习不断优化各实际工况的控制品质. 同时与其他相关深度学习控制方法的对比实验, 证实了本文提出的神经网络嵌入与优化方法对于实际系统控制具有一定的先进性.

6 结束语

通过分析测试结果得出如下结论: 1) 本文方

法通过神经网络控制器的嵌入在满足 Lyapunov 稳定条件下有效地优化了基准控制器性能; 2) 改进的优化目标函数能够有效地优化跟踪误差性能, 规避了辨识环节并支持在线学习; 3) 本文方法对函数时变与随机扰动具有较强的抑制能力, 对非线性系统的未知动态项具有良好的估计能力. 综上, 本文提出方法相对传统方法在性能优化、不确定性与扰动的补偿与抑制方面具有实用性和一定的先进性.

未来将从如下方面开展工作: 1) 将方法适用范围扩展至输出反馈系统; 2) 引入其他控制性能指标并设计相应的优化函数; 3) 扩展应用领域至多输入-多输出系统.

附录 A. 神经网络结构说明

本部分主要描述仿真与实验部分设计的神经网络结构、函数、参数等相关细节.

首先对于式 (3) 中涉及的 ϑ 函数, 本文采用改进的 sigmoid 函数:

$$\vartheta(x) = \frac{\alpha}{1 + e^{-\lambda x}} \quad (\text{A1})$$

其中, $\alpha \geq 1$ 为增益系数, $\lambda > 0$ 为缩放系数. 选择该函数理由为其导数相对平坦, 且增益与有效区域可调.

仿真与实验中采用了不同结构的神经网络以说明本文方法对于神经网络结构的广泛适用性.

图 A1 为仿真测试中算例 1~3 和 5.1 节、5.2 节中采用的神经网络结构. 该结构属于一种前向多层感知器, 其中 \mathbf{X} 为网络输入向量; $\boldsymbol{\mu}$ 为网络输出向量, 其具体含义为嵌入学习控制器的输出量; \mathbf{W} 与 \mathbf{G} 分别为网络可调权值矩阵 (以 \mathbf{W} 与 \mathbf{G} 全部参数构成的向量即式 (3) 中的 $\boldsymbol{\theta}$); \mathbf{H} 与 \mathbf{Y} 为网络中间向量. Γ 为激活函数, 本文采用 LeakyReLU 函数: $\Gamma(x) = \max(\beta x, x)$, $\beta = 0.18$.

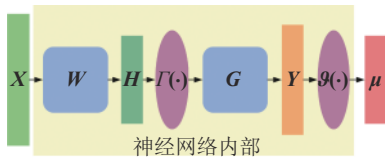


图 A1 算例 1~3 与 5.1 节、5.2 节学习控制器神经网络结构
Fig. A1 The architecture of neural network of learning controller in examples 1~3 and subsection 5.1~5.2

从输入 \mathbf{X} 到嵌入控制器输出 $\boldsymbol{\mu}$ 关系式如下:

$$\begin{cases} \mathbf{H} = \mathbf{W}^T \mathbf{X} \\ \mathbf{Y} = \mathbf{G}^T \Gamma(\mathbf{H}) \\ \boldsymbol{\mu} = \vartheta(\mathbf{Y}) \end{cases} \quad (\text{A2})$$

算例 1~3 网络输入 \mathbf{X} 为 $[x_1, x_2, y_d, \dot{y}_d, u(t-\zeta)]^T$, 其中 $x_1, x_2, y_d, \dot{y}_d, u$ 分别为算例 1~3 中的状态变量、期望输出及其导数和控制输出, ζ ($\zeta = 0.01$ s) 为延迟参数.

5.1 节和 5.2 节中, 网络输入 \mathbf{X} 为 $[\theta_j, \omega_j, \theta_j^d, \dot{\theta}_j^d, \tau_j(t-\zeta)]^T$, $\theta_j, \omega_j, \theta_j^d, \dot{\theta}_j^d, \tau_j$ 分别为机械臂关节角度、角速度、关节期望角度及其导数和关节控制转矩.

因此 \mathbf{W} , \mathbf{G} 维度分别为 $h_i \times 5$, $h_i \times 1$, h_i 表示第 4 节中算例 i 所对应的向量 \mathbf{H} 的维度, 分别为 18, 36, 30. 5.1 节和 5.2 节向量 \mathbf{H} 的维度分别为 36 与 30, 采用 Adamax 作为优化方法.

图 A2 为算例 4 中本文方法采用的嵌入神经网络, 该网络属于一种变形的 MLP (Multi-layer perceptron) 结构. 考虑了控制中的前馈机制, 该网络在图 A1 的网络内加入了一条前馈支路构成了分层 MLP 网络. 其中 \mathbf{X}_f 与 \mathbf{X}_b 为网络输入向量, $\boldsymbol{\mu}$ 为网络输出, 含义同上. \mathbf{W}_f , \mathbf{W}_b , \mathbf{G}_f 与 \mathbf{G}_b 为网络权值矩阵, 上述 4 个矩阵全部元素构成的向量即为网络的参数向量 $\boldsymbol{\theta}$. \mathbf{H}_f , \mathbf{H}_b , \mathbf{Y}_f , \mathbf{Y}_b 和 \mathbf{Y} 为网络中间向量. Γ 函数同上一模型.

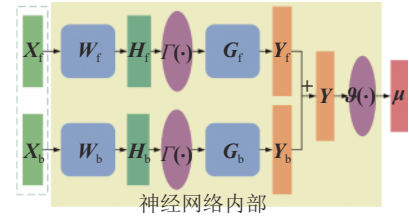


图 A2 算例 4 学习控制器神经网络结构

Fig. A2 The architecture of neural network of learning controller in the example 4

从输入 $\mathbf{X}_f, \mathbf{X}_b$ 到嵌入控制器输出 $\boldsymbol{\mu}$ 关系式如下:

$$\begin{cases} \mathbf{H}_f = \mathbf{W}_f^T \mathbf{X}_f \\ \mathbf{Y}_f = \mathbf{G}_f^T \Gamma(\mathbf{H}_f) \\ \mathbf{H}_b = \mathbf{W}_b^T \mathbf{X}_b \\ \mathbf{Y}_b = \mathbf{G}_b^T \Gamma(\mathbf{H}_b) \\ \mathbf{Y} = \mathbf{Y}_f + \mathbf{Y}_b \\ \boldsymbol{\mu} = \vartheta(\mathbf{Y}) \end{cases} \quad (\text{A3})$$

算例 4 中上述神经网络输入设置如下: $\mathbf{X}_f = [\theta_d^f, \dot{\theta}_d^f, \tau^f]^T$, $\theta_d^f, \dot{\theta}_d^f, \tau^f$ 分别为由电机期望角度及其导数和控制力矩构成的时序向量, 维度均为 N_f . 时序向量第 i 个元素分别表示为: $\theta_d^f(t-\zeta_i)$, $\dot{\theta}_d^f(t-\zeta_i)$, $\tau(t-\zeta_{i+1})$, $\zeta_i = \Delta t(i-1)$, $\Delta t = 0.01$ s. $\mathbf{X}_b = [\theta, \omega, \theta_d^b, \dot{\theta}_d^b]^T$, $\theta, \omega, \theta_d^b, \dot{\theta}_d^b$ 分别为由电机角度、角速度、期望角度及其导数构成的时序向量, 维度均为 N_b . 时序向量第 i 个元素分别表示为: $\theta(t-\zeta_i)$, $\omega(t-\zeta_i)$, $\theta_d^b(t-\zeta_i)$, $\dot{\theta}_d^b(t-\zeta_i)$.

算例 4 中取 N_f, N_b 分别为 6, 5, \mathbf{H}_f 与 \mathbf{H}_b 维度分别设为 12 与 18, 因此 $\mathbf{W}_f, \mathbf{W}_b, \mathbf{G}_f$ 与 \mathbf{G}_b 维度分别为 12×18 , 18×20 , 12×1 , 18×1 , 本例采用 SGD (Stochastic gradient descent) 作为优化方法.

附录 B. 关键变量与符号说明

表 B1
Table B1

第 1 节变量与符号	说明
S	n 阶 m 维状态反馈系统
F	未知非线性函数向量
d	有界未知扰动向量
b	m 阶未知可逆对角常数矩阵
V, \dot{V}	Lyapunov 函数及其导数
\mathcal{B}, \mathcal{M}	\dot{V} 中已知函数向量
φ	\dot{V} 中不含 u 的其余项之和
u_b	基础控制器
$\mu(\cdot \theta)$	神经网络嵌入控制器
u_b^μ	基于 u_b 与 $\mu(\cdot \theta)$ 的学习控制器
\circ	Hadamard 积运算符
$\vartheta(\cdot)$	嵌入控制器约束函数向量
\mathcal{L}	系统控制性能的量度

表 B2
Table B2

第 2 ~ 3 节变量与符号	说明
\mathcal{L}_S	系统 S 的性能优化目标函数
ψ	\mathcal{L}_S 中其他控制指标正则项
b_0	m 阶已知对角常值矩阵
$\mathcal{F}, \hat{\mathcal{F}}$	不确定与扰动值项的等效值与导数
$\hat{\mathcal{F}}, \dot{\hat{\mathcal{F}}}$	\mathcal{F} 的估计值及其导数
$\tilde{\mathcal{F}}, \dot{\tilde{\mathcal{F}}}$	\mathcal{F} 与 $\hat{\mathcal{F}}$ 的误差及其导数
$u_b^{\mathcal{F}}$	基于值自适应的基础控制器
$\varpi(\cdot)$	构造 $u_b^{\mathcal{F}}$ 所需函数

表 B3
Table B3

第 4 ~ 5 节变量与符号	说明
y_d	系统期望输出
b_0	已知系统增益 (b_0 的一维形式)
\hat{f}	待估计不确定项
k	控制器增益
$\xi(t)$	随机变量
θ, ω	电机角度与角速度
f_M	电机模型未知非线性项
θ_j, ω_j	机械臂关节角度与角速度
γ	DRL 方法的奖励函数

References

1 Wang H Q, Liu P X, Li S, Wang D. Adaptive neural output-feedback control for a class of nonlower triangular nonlinear systems with unmodeled dynamics. *IEEE Transactions on Neural Networks and Learning System*, 2018, **29**(8): 3658–3668

2 Cheng G Y, Yu W T, Hu J G. Improving the performance of motor drive servo systems via composite nonlinear control. *CES Transactions on Electrical Machines and Systems*, 2018, **2**(4): 399–408

3 Jamwal P K, Hussain S, Ghayesh M H, Rogozina S V. Impedance control of an intrinsically compliant parallel ankle rehabilitation robot. *IEEE Transactions on Industrial Electronics*, 2016, **63**(6): 3638–3647

4 Chen C, Modares H, Xie K, Lewis F L, Wan Y, Xie S. Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics. *IEEE Transactions on Automatic Control*, 2019, **64**(11): 4423–4438

5 Haddad W M, Chellaboina V. *Nonlinear Dynamical Systems and Control: A Lyapunov-based Approach*. Princeton: Princeton University Press, 2007. 135–198

6 Tian Fu-Qing, Jiang Shang, Liang Wei-Ge. Global backstepping fuzzy adaptive control for ammunition actuator with backlash. *Acta Automatica Sinica*, 2019, **45**(6): 1177–1185 (田福庆, 姜尚, 梁伟刚. 含齿隙弹载舵机的全局反步模糊自适应控制. *自动化学报*, 2019, **45**(6): 1177–1185)

7 Guerrero-Castellanos J F, Rifai H, Arnez-Paniagua V, Linares-Flores J, Mohammed S. Robust active disturbance rejection control via control lyapunov functions: Application to actuated-ankle foot-orthosis. *Control Engineering Practice*, 2019, **80**(18): 49–60

8 Ma Le, Liu Yue-Feng, Li Zhi-Wei, Xu Dong-Fu, Zhang Yu-Long. A framework of learning controller with Lyapunov-based constraint and application. *Chinese Journal of Scientific Instrument*, 2019, **40**(9): 189–198 (马乐, 刘跃峰, 李志伟, 徐东甫, 张玉龙. 一种基于Lyapunov约束的学习控制方法及应用. *仪器仪表学报*, 2019, **40**(9): 189–198)

9 Chen B, Zhang H G, Liu X P, Lin C. Neural observer and adaptive neural control design for a class of nonlinear systems. *IEEE Transactions on Neural Networks and Learning System*, 2018, **29**(9): 4261–4271

10 Li Y M, Tong S C. Adaptive neural networks prescribed performance control design for switched interconnected uncertain nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(7): 3059–3068

11 Fei J T, Lu C. Adaptive sliding mode control of dynamic systems using double loop recurrent neural network structure. *IEEE Transactions on Neural Networks and Learning System*, 2018, **29**(4): 1275–1286

12 Zhang X Y, Chen X K, Zhu G Q, Su C Y. Output feedback adaptive motion control and its experimental verification for time-delay nonlinear systems with asymmetric hysteresis. *IEEE Transactions on Industrial Electronics*, 2019, **67**(8): 6824–6834

13 Li Yang, Liu Ming-Yong, Zhang Xian-Jian. Adaptive RBF neural network-based backstepping control for supercavitating vehicles. *Acta Automatica Sinica*, 2020, **46**(4): 734–743 (李洋, 刘明雍, 张小件. 基于自适应RBF神经网络的超空泡航行体反演控. *自动化学报*, 2020, **46**(4): 734–743)

14 Park M, Chwa D, Eom M. Adaptive sliding-mode anti-sway control of uncertain overhead cranes with high-speed hoisting motion. *IEEE Transactions on Fuzzy Systems*, 2014, **22**(5): 1262–1271

15 Ouyang Hui-Min, Wang Jian, Zhang Guang-Ming, Mei Lei, Deng Xin. Tracking and anti-sway control for double-pendulum rotary cranes using novel sliding mode algorithm. *Acta Automatica Sinica*, 2019, **45**(7): 1344–1353 (欧阳慧珉, 王健, 张广明, 梅磊, 邓歆. 基于新型滑模算法的双摆旋转起重消摆跟踪控制. *自动化学报*, 2019, **45**(7): 1344–1353)

16 Chen W H, Yang J, Guo L, Li S H. Disturbance-observer-based control and related methods—An overview. *IEEE Transactions on Industrial Electronics*, 2016, **63**(2): 1083–1095

17 Zhong Q C, Kuperman A, Stobart R. Design of ude-based controllers from their two-degree-of-freedom nature. *International Journal of Robust and Nonlinear Control*, 2001, **21**(17):

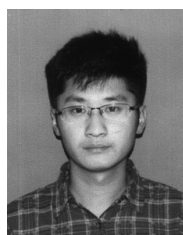
1994–2008

- 18 Lv M, Baldi S, Liu Z C. The non-smoothness problem in disturbance observer design: A set-invariance based adaptive fuzzy control method. *IEEE Transactions on Fuzzy Systems*, 2019, **27**(3): 598–604
- 19 Han J Q. From pid to active disturbance rejection control. *IEEE Transactions on Industrial Electronics*, 2009, **56**(3): 900–906
- 20 Li Jie, Qi Xiao-Hui, Wan Hui, Xia Yuan-Qing. Active disturbance rejection control: Theoretical results summary and future researches. *Control Theory and Applications*, 2017, **34**(3): 281–295
(李杰, 齐晓慧, 万慧, 夏元清. 自抗扰控制: 研究成果总结与展望. 控制理论与应用, 2017, **34**(3): 281–295)
- 21 Wen G X, Ge S S, Tu F W. Optimized backstepping for tracking control of strict-feedback systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(8): 3850–3862
- 22 Bu X H, Hou Z S. Adaptive iterative learning control for linear systems with binary-valued observations. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(1): 232–237
- 23 Meng T T, He W. Iterative learning control of a robotic arm experiment platform with input constraint. *IEEE Transactions on Industrial Electronics*, 2018, **65**(1): 664–672
- 24 Xu D Z, Liu J X, Yan X G, Yan W X. A novel adaptive neural network constrained control for a multi-area interconnected power system with hybrid energy storage. *IEEE Transactions on Industrial Electronics*, 2018, **65**(8): 6625–6634
- 25 Volodymyr M, Koray K, David S, Rusu A A, Joel V, Bellemare M G, et al. Human-level control through deep reinforcement learning. *Nature*, 2015, **518**(7540): 529–541
- 26 Kang Y, Chen S F, Wang X F, Cao Y. Deep convolutional identifier for dynamic modeling and adaptive control of unmanned helicopter. *IEEE Transactions on Neural Networks and Learning System*, 2019, **30**(2): 524–538
- 27 Zhou M, Feng Y, Xue C, Han F. Deep convolutional neural network based fractional-order terminal sliding-mode control for robotic manipulators. *Neurocomputing*, DOI: 10.1016/j.neucom.2019.04.087
- 28 Carlucho I, De Paula M, Wang S, Petillot Y, Acosta G G. Adaptive low-level control of autonomous underwater vehicles using deep reinforcement learning. *Robotics and Autonomous Systems*, 2018, **107**(2018): 71–86
- 29 Xiong H, Ma T Q, Zhang L, Diao X M. Comparison of end-to-end and hybrid deep reinforcement learning strategies for controlling cable-driven parallel robots. *Neurocomputing*, 2020, **377**(2020): 73–84
- 30 Yang Z Y, Merrick K E, Jin L W, Abbass H A. Hierarchical deep reinforcement learning for continuous action control. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(11): 5174–5184
- 31 Zhang Y Y, Li S, Liu X P. Neural network-based model-free adaptive near-optimal tracking control for a class of nonlinear systems. *IEEE Transactions on Neural Networks and Learning System*, 2018, **29**(12): 6227–6241
- 32 Liu L, Jiang H, He P, Chen W, Liu X, Gao J, et al. On the variance of the adaptive learning rate and beyond [Online], available: <https://arxiv.org/abs/1908.03265>, April 17, 2020
- 33 Vu V, Wang W. State/Disturbance observer and controller synthesis for the t-s fuzzy system with an enlarged class of disturbances. *IEEE Transactions on Fuzzy Systems*, 2018, **26**(6): 3645–3659
- 34 Li S, Wang H, Aitouche A, Tian Y, Christov N. Actuator fault and disturbance estimation using the t-s fuzzy model. *IFAC Papers Online*, 2017, **50**(1): 15722–15727
- 35 Guo B Z, Zhao Z L. On convergence of tracking differentiator. *International Journal of Control*, 2011, **84**(4): 693–701



马乐 东北电力大学自动化工程学院副教授. 主要研究方向为机器人学习、控制与视觉.

E-mail: male_robot_nedu@sina.com
(MA Le Associate professor at the School of Automation and Engineering, Northeast Electric Power University. His main research interest is robotics learning, controlling and vision.)



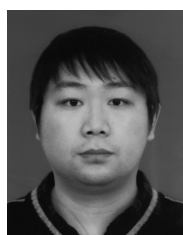
闫一鸣 东北电力大学自动化工程学院硕士研究生. 主要研究方向为神经网络学习控制.

E-mail: ddyym3914@163.com
(YAN Yi-Ming Master student at the School of Automation and Engineering, Northeast Electric Power University. His main research interest is neural network learning control.)



徐东甫 东北电力大学自动化工程学院副教授. 主要研究方向为机器人导航与控制. 本文通信作者.

E-mail: xu.dong.fu@163.com
(XU Dong-Fu Associate professor at the School of Automation and Engineering, Northeast Electric Power University. His main research interest is robot navigation and control. Corresponding author of this paper.)



李志伟 东北电力大学自动化工程学院副教授. 主要研究方向为非线性系统建模、控制及数值模拟.

E-mail: zhiwei.li@neepu.edu.cn
(LI Zhi-Wei Associate professor at the School of Automation and Engineering, Northeast Electric Power University. His main research interest is modeling, control and numerical simulation of nonlinear system.)



孙灵芳 东北电力大学自动化工程学院教授. 主要研究方向为热工过程先进控制.

E-mail: 15043283452@163.com
(SUN Ling-Fang Professor at the School of Automation and Engineering, Northeast Electric Power University. His main research interest is advanced control of thermal process.)