

# 基于深度强化学习的动基座双自由度 系统动力学控制方法\*

刘丰瑞<sup>1,2†</sup> 颜格<sup>3</sup> 张晓龙<sup>1,2</sup> 张文明<sup>3</sup> 王国鹏<sup>1,2</sup>

(1. 上海航天控制技术研究所, 上海 201109)

(2. 上海市空间智能控制技术重点实验室, 上海 201109)

(3. 上海交通大学 机械系统与振动国家重点实验室, 上海 200240)

**摘要** 具有硬边界的动基座双自由度动力学系统在受到基座周期性激励与随机冲击扰动时, 振子与边界碰撞后的强非线性特性导致系统产生复杂的混沌行为. 本文基于 Soft Actor-Critic 强化学习框架, 构建了同时实现振动控制与基座运动跟随的智能算法, 研究了宽频域范围内含有硬边界约束的动基座双自由度系统的动力学控制效果. 通过构建包含相对位移、控制力的复合奖励函数, 实现动力学系统精度较高的轨迹跟踪与振动抑制. 结果表明, 该算法可以实现频率范围跨 2 个数量级 (0.01 Hz 到 1 Hz) 的有效振动控制, 并通过与 PID 控制方法的比较, 展现了该方法在复杂环境中的稳定性与泛化性.

**关键词** 动力学, 振动控制, 强化学习, 非线性, 轨迹跟踪

中图分类号: O322

文献标志码: A

## Dynamic Control Method for Dual-Degree-of-Freedom Systems with Moving Base by Deep Reinforcement Learning\*

Liu Fengrui<sup>1,2†</sup> Yan Ge<sup>3</sup> Zhang Xiaolong<sup>1,2</sup> Zhang Wenming<sup>3</sup> Wang Guopeng<sup>1,2</sup>

(1. Shanghai Aerospace Control Technology Institute, Shanghai 201109, China)

(2. Shanghai Key Laboratory of Aerospace Intelligent Control Technology, Shanghai 201109, China)

(3. State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering,  
Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract** When a double-degree-of-freedom dynamical system with hard boundaries subjects to periodic excitations and random impulsive disturbances, the oscillator will collide with the boundaries and exhibit complex chaotic behavior. In this paper, an intelligent framework with the Soft Actor-Critic reinforcement learning algorithm is proposed to achieve both vibration control and base motion tracking. In addition, the dynamic control effect of the double-degree-of-freedom system on a moving base with hard boundary is also studied in a wide frequency range. By constructing a composite reward function that incorporates relative displacement and control force, high-precision trajectory tracking and vibration suppression in dynamic system are achieved. The results demonstrate that the algorithm can achieve effective vibration control in a wide frequency range from 0.01 Hz to 1 Hz. A comparison with the PID control method further exhibits the stability and generalizability of this approach in complex environments.

**Key words** dynamics, vibration control, reinforcement learning, nonlinearity, trajectory tracking

## 引言

浮动基动力学系统广泛存在于航空航天、海洋工程、工业制造等领域中,如带有柔性太阳能板或柔性机械臂的航天器、海上风力发电平台、作业船舶等都可以简化为浮动基的动力学系统<sup>[1-5]</sup>。这类浮动基动力学系统运行时往往既要实现系统轨迹的精确跟踪,也要能够有效抑制外界的未知扰动冲击<sup>[6]</sup>。目前来看,浮动基动力学控制仍然存在很多挑战,比如,动力学系统的碰撞接触导致了动力学响应的非光滑强非线性特性,容易出现分岔和混沌行为<sup>[7,8]</sup>;未知冲击扰乱了系统的动力学响应规律;与环境交互后,系统动力学参数[如构型改变导致的质量、刚度、阻尼改变(晃液、柔性绳索、包覆、太阳能板断裂等)]可能发生未知变化,导致原有控制策略失效等方面的问题,这些问题会导致系统的动力学响应复杂化,给相应的振动控制与轨迹跟踪带来困难<sup>[9,10]</sup>。

目前,动力学系统的振动控制方法主要分为被动控制、半主动控制、主动控制三类。其中,主动控制方法因为具有较强的灵活性和环境适应性成为近年来的研究热点。PID 控制方法作为一种经典的主动控制策略得到了广泛的应用,但难以实现含有碰撞、未知扰动、参数时变的强非线性动力学系统的有效控制<sup>[11-13]</sup>。随着现代控制理论的快速发展,一些智能控制策略被陆续提出,如自适应控制<sup>[14]</sup>、模糊控制<sup>[15]</sup>、滑模控制、神经网络控制<sup>[16]</sup>等方法。

自适应控制是一种实时调整控制器参数,以适应被控对象(也称为系统)的变化或未知动态特性的方法<sup>[3,17]</sup>。不过,仍然需要在运行中不断地对系统参数进行识别,修改自适应控制规则。采用的控制器初始参数需要根据经验或系统知识设定。模糊控制是一种基于模糊逻辑的控制方法,需要依据专家经验知识设计模糊控制规则、选择隶属函数、手动调整控制参数与模糊集合<sup>[18,19]</sup>。对系统的建模和规则的设计需要耗费一定的时间和精力,在处理高维大规模系统时容易出现规则爆炸的问题。此外,模糊控制算法对隶属函数和模糊规则的选择比较敏感,不合理的选择可能会导致控制性能下降。滑模控制是通过构建滑模面,调节控制量保证系统状态在面上滑动实现控制的方法<sup>[20,21]</sup>。该方法需要根

据经验设计以及调节滑模面与控制器,在系统状态接近滑模面时,可能会引起控制器切换,导致控制输入高频震荡,进而影响系统状态稳定性。这些智能控制策略仍然需要辨识系统动力学特性、依赖专家经验进行算法设计,并花费大量的时间调节算法参数。

随着神经网络技术的发展,基于神经网络的控制方法不断出现。与传统的智能控制技术相比,这类技术具有较强的非线性拟合能力,不需要系统模型参数,研究人员将神经网络与各种控制方法结合起来改善动力学系统的控制效果<sup>[22-26]</sup>。深度强化学习控制算法是一种通过神经网络智能体与动力学环境进行交互训练找到最优控制策略的方法,具有自主探索能力,并能够很方便地引入复杂约束条件与目标函数,并可以根据动力学环境的变化,动态调整控制模型<sup>[27-30]</sup>。和其他控制方法相比,深度强化学习算法表现出了显著的泛化性和稳定性,常见的深度强化学习算法有深度 Q-learning 算法、DDPG(深度确定性策略梯度)、TRPO(信赖域策略优化)算法、基于重要性采样技术的 PPO 算法和基于最大熵的 Soft Actor-Critic (SAC) 算法。其中,SAC 算法是一种 off-line 算法,在智能体与环境交互训练以优化控制策略的过程中,建立有样本回放缓冲区,能够反复利用历史样本,提高了数据利用率,同时采用最大熵探索策略对动作进行概率性选择,使智能体具有较强的探索性,对环境不确定扰动以及模型参数不确定的动力学性的适应性更强。

为了实现控制系统的泛化适应能力,本文基于 SAC 算法建立了硬碰撞边界的浮动基双自由度动力学系统的智能振动控制与轨迹跟踪策略。根据各振子位移与基座位移的偏离值的相对偏离程度,构建均衡考虑系统各自由度的奖励函数;通过引入指数函数,提高响应位移在目标位移附近的敏感性;引入惩罚项保证相对振幅不超过合理范围,减小算法训练搜索空间,提高算法收敛速度。对基座振动频率为 0.01 Hz~1 Hz 范围内的不同质量、刚度、阻尼的动力学系统进行了测试,验证基于该算法的双自由度动力学系统的振动控制与运动跟随效果。

## 1 动力学模型与控制算法框架

### 1.1 系统动力学模型

首先,建立含有硬边界的动基座双自由度系统

动力学模型.含质量块  $m_1$  和  $m_2$  的双自由度动力学系统连接在浮动基座上,质量块  $m_1$  与基座之间连接一组弹簧( $k_1$ )—阻尼( $c_1$ )结构,在质量块  $m_1$  左右两侧存在硬边界,边界与基座是一体的,左边界和右边界与质量块  $m_1$  的距离分别为  $d_1=0.3, d_2=0.3$ .质量块  $m_1$  与质量块  $m_2$  通过另一组弹簧( $k_2$ )—阻尼( $c_2$ )结构相连.基座在一维方向进行周期性运动  $x_0(t) = \sin(\omega t + \varphi)$  时,如果质量块与硬边界接触,将会发生硬碰撞,反弹系数为  $R=0.8$ .对该双自由度动力学系统施加控制策略时,控制力  $F_1$  和  $F_2$  分别施加在质量块  $m_1$  和  $m_2$  上.

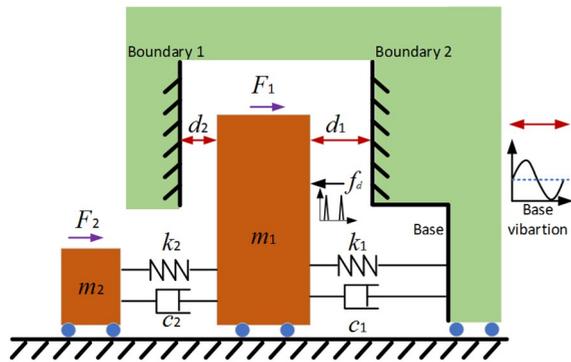


图1 含动基座和硬边界的双自由度动力学模型

Fig.1 Two-degree-of-freedom dynamic system with movable base and hard boundary

将强化学习智能控制策略用于动力学系统控制时,考虑到实际物理系统的刚度、质量、阻尼均可能有较大的变化,在仿真过程中,设定系统质量、刚度、阻尼在一定范围内随机变化.这里为便于归一化分析,将  $m_1, k_1, c_1$  设定为定值:  $m_1 = 1, k_1 = 2, c_1 = 0.1$ , 而  $m_2, k_2, c_2$  在一定范围内随机变化:  $m_2 = 1 \pm 0.2, k_2 = 2 \pm 0.4, c_2 = 0.1 \pm 0.02$ , 系统基座的激励频率在频率  $\omega$  在  $0.01\text{Hz}$  到  $1\text{Hz}$  之间随机取值.此外,存在随机脉冲扰动力  $f_d$  作用于质量块  $m_1$  上.

该动力学系统的运动微分方程为:

$$\begin{cases} m_1 \ddot{x}_1 + c_1(\dot{x}_1 - \dot{x}_b) + k_1(x_1 - x_b) = F_1(t) \\ m_2 \ddot{x}_2 + c_2(\dot{x}_2 - \dot{x}_1) + k_2(x_2 - x_1) = F_2(t) \end{cases} \quad (1)$$

$$(x_b - d_2 < x_1 < x_b + d_1)$$

式中,  $x_1$  和  $x_2$  分别为振子  $m_1$  和  $m_2$  的位移.发生硬碰撞时( $x_1 - x_b = d_1$  or  $-d_2$ ),振子  $m_1$  的速度变化为  $\dot{x}_{1+} = -R\dot{x}_{1-}$ .

## 1.2 马尔科夫决策过程

基于 SAC 算法的智能动力学控制过程可以看做马尔科夫决策过程,智能体输出的控制力只与系

统当前状态有关,相应的马尔科夫决策过程模型如图 2 所示.在动力学控制系统中,智能体 Agent 观察到动力学系统的状态  $S_n$  (振子的位移、速度),与目标状态比较后得到奖励值  $r_n$ , 经过计算输出动作策略  $a_n$  (控制力) 到动力学系统中,由此产生新的状态  $S_{n+1}$ , 智能体可以根据新的状态量计算下一步的控制策略.在智能体与环境的不断交互迭代过程中,智能体根据计算的奖励值不断调整控制策略,朝着奖励最大的方向优化,最终找到满足控制目标的最佳控制策略.

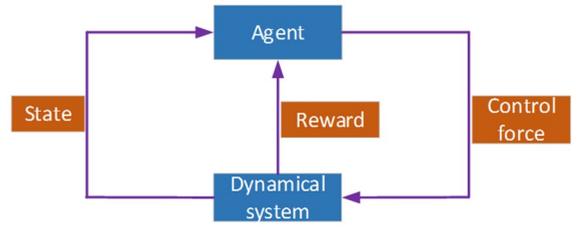


图2 马尔科夫决策过程模型  
Fig.2 Markov decision process model

智能体观察的动力学系统状态量有质量块的绝对位移  $x_1, x_2$ , 基座的相对位移  $x_{r1}, x_{r2}$ , 质量块与基座的相对速度  $v_{r1}, v_{r2}$ , 还有系统输出的控制力  $F_1, F_2$ . 考虑到实际控制系统的输出能力,作用在质量块  $m_1, m_2$  上的控制单元输出力幅值限定在  $-40 \leq F_1 \leq 40, -40 \leq F_2 \leq 40$  范围内,控制力输出采样率为  $0.04\text{s}$ .

为了引导控制策略朝着正确的方向优化,需要构建相应的奖励函数,输出能有效评价控制效果的奖励值.奖励函数需要让质量块  $m_1, m_2$  能够以尽可能小的误差跟随基座的运动,并具有克服外界冲击扰动的能力,避免与边界硬碰撞带来的混沌运动.此外,需要能够让控制力输出幅值尽可能小,减小控制系统的能量消耗.基于上述考虑,构建的奖励函数形式如下:

$$R = b_1 e^{-a_1 |x_{r1}|} + b_2 e^{-a_2 |x_{r2}|} + \Delta R_2 - c_1 \frac{1}{T} \int_0^T |F_1| dt - c_2 \frac{1}{T} \int_0^T |F_2| dt - g(x_{r1}, x_{r2}), \quad (2)$$

$$g(x_{r1}, x_{r2}) = \begin{cases} d_1, & (\text{other cases}) \\ 0, & (k_1 \leq x_{r2} \leq k_2) \end{cases} \quad (3)$$

式中,  $g(x_{r1}, x_{r2})$  是质量块  $m_2$  相对位移的惩罚项.本文构建了指数型的位移奖励子函数  $f_1 = b_1 e^{-a_1 |x_{r1}|}$  和  $f_2 = b_2 e^{-a_2 |x_{r2}|} + \Delta R_2$ , 其奖励函数与

质量块相对位移成指数关系,能够促进策略对目标位移更具有敏感性.针对质量块  $x_{r2}$  没有硬边界限制,容易出现较大振幅的情况,其相对位移奖励函数由指数项  $b_2 e^{-a_2 |x_{r2}|}$  和分段线性项  $\Delta R_2$  共同决定,其中分段线性项  $\Delta R_2$  为

$$\Delta R_2 = \begin{cases} -0.3(|x_{r2}| - x_{r0}), & |x_{r2}| > x_{r0} \\ 0, & |x_{r2}| \leq x_{r0} \end{cases} \quad (4)$$

在位移幅值较大(超过  $x_{r0}$ ) 的时候,奖励函数与位移不再是由小梯度的指数关系占主导关系,而是由线性关系项占主导地位,智能体不会失去梯度变化而停止优化.在相对位移幅值小于  $x_{r0}$  的时候,梯度较大的指数函数起主导地位,保证相对位移在小位移误差情况下具有较快的收敛性.为了鼓励智能体以较小的控制力幅值实现有效的振动控制与轨迹跟随,降低能量的消耗,所构建的控制力相关奖励函数为:  $f_3 = c_1 \cdot 1/T \cdot \int_0^T |F_1| dt$ ,  $f_4 = c_2 \cdot 1/T \cdot \int_0^T |F_2| dt$ , 函数值与一段时间内控制力的输出幅值积分正相关.此外,当质量块  $m_2$  的相对位移超过合理范围(比如  $\pm 10$ ),停止本次训练,并由惩罚项  $g(x_{r1}, x_{r2})$  返回一个较大的负奖励值  $d_1$ ,以促进智能体控制质量块  $m_2$  的相对位移不要超过合理范围.奖励函数设置不合理可能会影响收敛性,变动过于剧烈的奖励值可能导致学习不稳定,甚至无法收敛,变化过小则可能导致学习速度很慢,因此奖励函数的设计是强化学习类算法的一个重要组成部分.

## 1.2 智能控制算法框架

为实现宽频(0.01Hz~1Hz)范围内的振动控制与运动跟随,所建立的基于 Soft Actor-Critic (SAC)的控制算法框架如图 3 所示.仿真训练在连续动作空间中进行,算法中的神经网络结构由一个策略网络(Actor)、两个评价网络(Critic)、两个目标网络共同构成.训练时,智能体与动力学系统不断交互获取训练数据,根据得到的系统响应状态,不断更新各个网络的权重参数,直到输出的控制策略达到预期效果.

策略网络用于实现动力学系统控制力的输出,输入是动力学系统的状态(振子与基座的相对位移、相对速度等参量),输出为系统控制力的概率分

布(取高斯概率分布模型参数作为输出).本文根据动力学系统结构的复杂程度,输入量经过一个全连接层和激活函数后,分为两个分支,即均值网络和标准差网络,分别再经过若干全连接层和激活函数,得到作为高斯概率分布函数的参数.智能体根据高斯概率分布函数输出两组控制力到动力学系统中.

评价网络用于评估策略的价值,输入是系统状态和动作,输出是状态与动作的累积回报,本文采用两个初始化参数不同的评价网络,得到两组不同的回报值,为避免对回报值的过高估计,智能体将选取两组中较低的回报,作为对策略网络更新的依据.评价网络输入有两个分支,分别为状态评价网络和动作评价网络,均主要由多个全连接层和激活层构成,然后合并在一起进行综合评价.

目标网络用以提供稳定的目标 Q 值函数估计,每隔一定训练次数定期更新网络权重.

算法中还建立了一个经验回放缓冲区,用于实现策略的离线训练,提高数据利用效率.训练时,缓冲区中随机抽取固定数量的样本数据(智能体与环境交互产生的状态、动作、奖励、下一个状态等数据)对网络进行离线训练.

SAC 算法最大特点是通过最大熵目标学习随机策略.熵在策略函数被用于正则化项,鼓励智能体探索更多的动作空间,在网络中用于更新策略的目标值,提高策略的探索性能.算法中引入了调节熵正则化权重的自适应温度参数  $\alpha$ ,控制着策略对熵正则化的敏感程度, $\alpha$  较小时,策略更关注最大化累积回报, $\alpha$  较大时,策略更具有探索性,以提高策略的泛化与稳定性.训练过程中采用最优化方法自适应地更新  $\alpha$  的值,实现平衡探索性与利用效率之间的权衡.

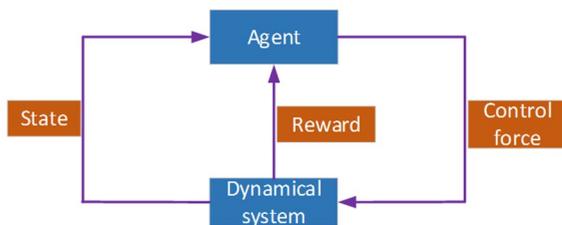


图 3 基于 SAC 的智能算法网络结构  
Fig.3 SAC network structure

值网络权重、策略网络权重和熵系数的更新采用梯度更新方式进行.不过,目标价值网络不采

用梯度方式,而是每隔一定训练次数将价值网络权重按照一定程度复制给目标价值网络.SAC算法通过最大化策略的期望累积回报实现策略网络优化,通过最小化值函数的均方根误差实现值函数网络优化.智能体根据策略网络输出的概率分布参数按照高斯分布概率性地输出控制力,以提升模型应对不确定环境的稳定性和适应性.

SAC算法的基本迭代过程如下:

- (1)初始化策略网络和评价网络,随机初始化参数;
- (2)重复执行以下步骤,直到达到停止条件:
  - a. 在当前策略下,收集一批样本轨迹.这些样本包括状态、动作、奖励和下一个状态.
  - b. 使用评价网络计算当前策略下的动作价值.
  - c. 使用策略网络计算当前状态下的动作概率.
  - d. 使用策略网络和采样策略(例如高斯策略)从当前状态中采样一个动作.
  - e. 使用评价网络评估此动作的价值,并计算其期望值.
  - f. 基于最大化策略网络的期望奖励和最小化策略网络的熵,计算SAC的目标函数.
  - g. 使用梯度下降方法更新策略网络和评价网络的参数.
  - h. 使用目标网络参数更新目标评价网络的参数.
  - i. 循环回到步骤a.
- (3)当控制效果满足预先设置的奖励阈值要求时,停止训练,以此时的策略网络作为控制策略网络.

## 2 控制性能仿真验证

如果对双自由度动力学系统不施加控制力,在基座进行周期性运动的时候,质量块1很容易与边界1和边界2发生硬碰撞,在受到随机脉冲扰动力 $f_d$ 作用后,很容易产生混沌的动力学响应位移,质量块2也会随之产生混沌运动.基座以0.01Hz,0.1Hz,1Hz三个频率振动时,系统动力学响应结果如图4所示.其中,图4(a)、4(b)、4(c)对应的基座频率分别为0.01Hz,0.1Hz,1Hz.

可以看出,在边界硬碰撞和随机扰动的双重作用下,两个质量块在较大频率范围内的响应位移是无规律的混沌运动,无法有效跟随基座运动.

为了实现振动控制与跟随运动,基于本文的算法框架对智能体进行了训练.训练时,策略网络的

均值网络和标准差网络分别含有三个全连接层,每层的网络参数为1200、800、2(输出信号数量),它们的第一个全连接层是共用参数的.评价网络的状态评价网络含有三个全连接层,每层网络节点数为1200、800、1,动作评价网络含有两个全连接层,每层网络节点数为800、1,它们的最后一个全连接层是共用参数的.目标网络与评价网络的结构和参数保持相同.训练过程中得到的奖励如图5所示.在训练达到200次之后,奖励值就已经基本稳定,为了保证控制能力的泛化性,继续训练到470次后停止.

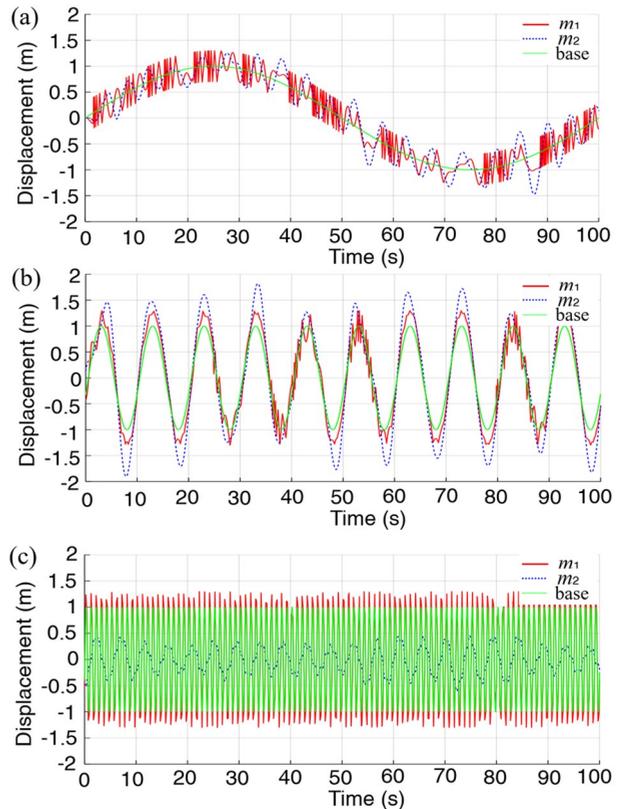


图4 基座频率等于0.01Hz,0.1Hz,1Hz时,系统双振子的位移-时间响应曲线

Fig.4 Displacement-time curves of  $x_1, x_2$  of the system, when the vibrational frequency of the base is 0.01Hz, 0.1Hz, and 1Hz

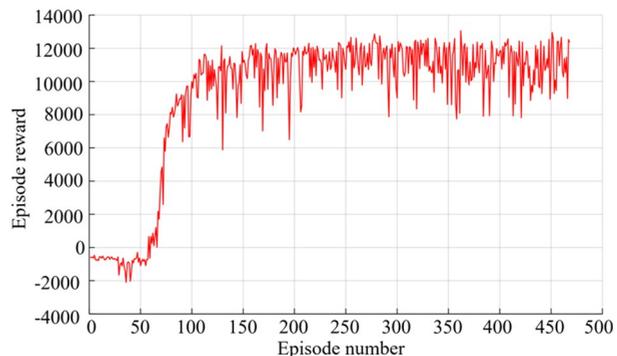


图5 训练过程中奖励函数值的变化  
Fig.5 Variation of episode reward value during training

采用训练好的控制策略对动力学系统进行控制,基座运动频率为 0.01Hz、0.1Hz、1Hz 三个频率时(算法网络参数和结构保持固定不变),系统振子响应位移曲线如图 6(a)、图 6(c)、图 6(e)所示.作为对比,对动力学系统每个自由度设置一个 PID 控制器,进行了对比测试,在控制力幅值范围为  $[-40, 40]$ 、输出时间间隔为 0.04s 的情况下, PID 算法的控制效果如图 6(b)、图 6(d)、图 6(f)所示.图 6(a)-(b)、图 6(c)-(d)、图 6(e)-(f)对应的基座频率分别为 0.01Hz、0.1Hz、1Hz.

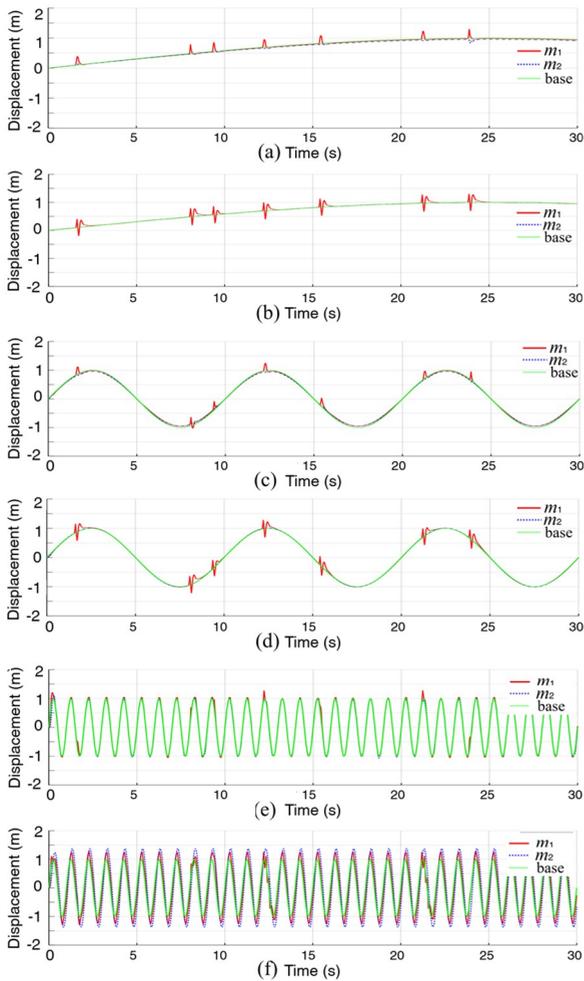


图 6 采用 SAC 智能控制策略与 PID 控制策略时,系统双振子的位移-时间响应曲线

Fig.6 Displacement-time curves of  $x_1, x_2$  of the system, when the SAC control method and PID method are applied

可以看到,在有限采样频率和有限出力情况下, PID 方法比 SAC 策略的控制效果差一些,在出现随机脉冲扰动后, SAC 策略能够迅速抑制振动,而 PID 需要更长的时间才能抑制振动,在此期间,质量块会发生硬碰撞,并且随着频率升高,运动跟随效果明显下降.如果扰动力的脉冲幅值增加一

倍,相同网络结构参数的智能体仍然能够有效控制振子的位移[图 7(a)],而采用 PID 控制时[图 7(b)],系统会出现明显的高频振动,这种高频振动是因为脉冲无法被有效控制,振子反复撞在左右两侧硬边界上,在 5 个高频周期之后, PID 算法才能有效控制住高频振动,表明之前选取的 PID 参数无法适用更强的脉冲扰动力.

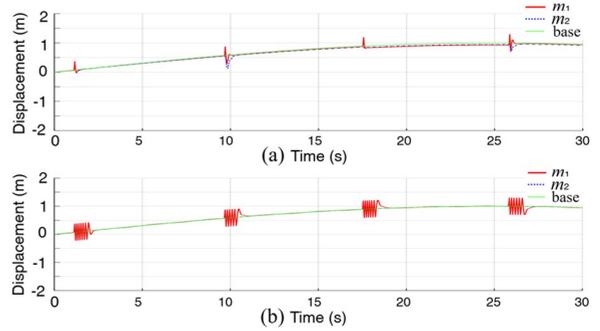


图 7 当基座频率为 0.01Hz,脉冲扰动力幅值增加 1 倍后,系统动力学响应的位移-时间曲线

Fig.5 Displacement-time curves of  $x_1, x_2$  of the system, when the force amplitude is increased twice. The vibrational frequency of the base is 0.01Hz

然后改变振子 2 的刚度和阻尼系数,测试了系统动力学参数变化对控制性能的影响,结果表明,基于 SAC 策略的系统控制效果基本没有变化(图 8).图 8(a)-(b)、8(c)-(d)、8(e)-(f)对应的基座频率分别为 0.01Hz、0.1Hz、1Hz.其中,图 8(a)、图 8(c)、图 8(e)的刚度等于 2.4,阻尼等于 0.1,图 8(a)、图 8(c)、图 8(e)的刚度等于 2.4,阻尼等于 0.12.

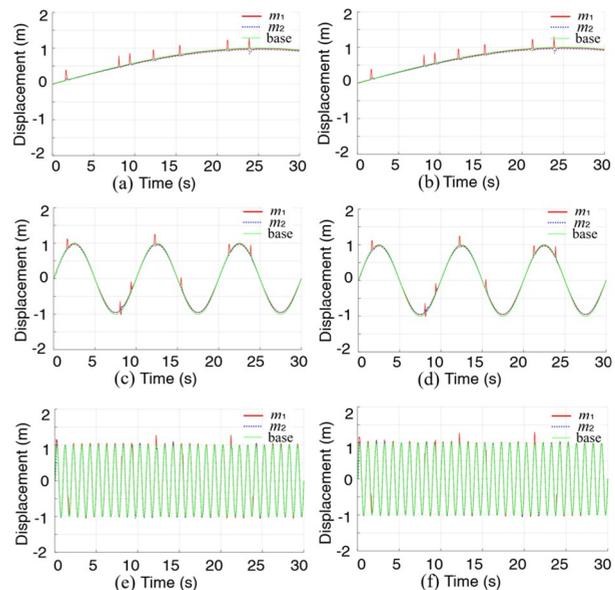


图 8 振子 2 的刚度  $k_2$  和阻尼  $c_2$  改变后,

基于 SAC 控制算法的系统响应的位移-时间曲线

Fig.8 Displacement-time curves of  $x_1, x_2$  of the system with SAC method, when the stiffness  $k_2$  and damping  $c_2$  is changed

通过不同刚度、阻尼以及基座频率的参数测试,并且与PID控制方法进行对比,可以看到本文构建的智能控制方法能够很好地适应动力学参数不确定、脉冲激励随机情况下的振动控制与轨迹跟踪效果。

### 3 总结

为了抑制含硬边界的双自由度非线性动力系统的混沌振动,实现宽频范围内(0.01Hz~1Hz)的轨迹跟随,本文基于SAC深度强化学习框架,搭建了适用动力学参数不确定的智能振动控制与轨迹跟踪算法.设计了能够分别表征系统的双振子位移误差和能量大小的复合奖励函数.在控制力输出幅值和输出频率有限的情况下,不同基座频率条件下的测试结果表明,训练后的控制模型能够对基座0.01Hz到1Hz范围内的脉冲激励产生的振动进行有效抑制,亦能实现有效的运动跟随.并且与PID控制方法进行了对比,表现出了更好的控制效果.实际系统中的动力学参数往往存在测量不准的情况,为了让控制系统能够具有较好的泛化性,随机改变了系统刚度、阻尼参数,测试表明算法依然能保持很好的振动抑制与运动跟随效果。

### 参考文献

- [1] JAHANI K, LANGLOIS R G, AFAGH F F. Structural dynamics of offshore wind turbines: a review [J]. *Ocean Engineering*, 2022,251:111136.
- [2] WEI J, CAO D, WANG L, et al. Dynamic modeling and simulation for flexible spacecraft with flexible jointed solar panels [J]. *International Journal of Mechanical Sciences*, 2017,130:558–570.
- [3] 陈志勇,张婷婷,郭益深. 弹性基和弹性关节空间机器人的自适应鲁棒抗扰控制及振动抑制 [J]. *自动化学报*, 2018,44(7):1271–1281.  
CHEN Z Y, ZHANG T T, GUO Y S. Adaptive robust anti-interference control and vibration suppression for an elastic-base elastic-joint space robot [J]. *Acta Automatica Sinica*, 2018,44(7):1271–1281. (in Chinese)
- [4] LIU L, WANG X D, SUN S P, et al. Dynamic characteristics of flexible spacecraft with double solar panels subjected to solar radiation [J]. *International Journal of Mechanical Sciences*, 2019,151:22–32.
- [5] FUNG P, GRIMBLE M. Dynamic ship positioning using a self-tuning Kalman filter [J]. *IEEE Transactions on Automatic Control*, 1983,28(3):339–350.
- [6] SERDUKOVA L, KUSKE R, YURCHENKO D. Fundamental competition of smooth and non-smooth bifurcations and their ghosts in vibro-impact pairs [J]. *Nonlinear Dynamics*, 2023,111(7):6129–6155.
- [7] LIU Y, PÁEZ CHÁVEZ J. Controlling multistability in a vibro-impact capsule system [J]. *Nonlinear Dynamics*, 2017,88(2):1289–1304.
- [8] FARID M. Dynamics of a hybrid vibro-impact oscillator: canonical formalism [J]. *Nonlinear Dynamics*, 2021,106(3):1769–1787.
- [9] 袁利,姜甜甜,魏春岭,等. 空间控制技术发展展望 [J]. *自动化学报*, 2023,49(3):476–493.  
YUAN L, JIANG T T, WEI C L, et al. Advances and perspectives of space control technology [J]. *Acta Automatica Sinica*, 2023,49(3):476–493. (in Chinese)
- [10] LI S, SHE Y C. Recent advances in contact dynamics and post-capture control for combined spacecraft [J]. *Progress in Aerospace Sciences*, 2021,120:100678.
- [11] LI J, LI Y. Dynamic analysis and PID control for a quadrotor [C]//2011 IEEE International Conference on Mechatronics and Automation, Beijing, China, 2011:573–578.
- [12] PARRA-VEGA V, ARIMOTO S, LIV Y H, et al. Dynamic sliding PID control for tracking of robot manipulators: theory and experiments [J]. *IEEE Transactions on Robotics and Automation*, 2003,19(6):967–976.
- [13] 仇智,苏琦,方梓帆,等. 一种基于状态反馈的比例伺服阀控制方法 [J]. *飞控与探测*, 2022,5(1):39–47.  
QIU Z, SU Q, FANG Z F, et al. Method of proportional servo valve based on state feedback [J]. *Flight Control & Detection*, 2022,5(1):39–47. (in Chinese)
- [14] Gan Z K, HILLIS A J, DARLING J. Adaptive control of an active seat for occupant vibration re-

- duction [J]. *Journal of Sound and Vibration*, 2015, 349:39–55.
- [15] NING D, SUN S, ZHANG F, et al. Disturbance observer based Takagi-Sugeno fuzzy control for an active seat suspension [J]. *Mechanical Systems and Signal Processing*, 2017, 93:515–530.
- [16] GOMI H, KAWATO M. Neural network control for a closed-loop System using *Feedback-error-learning* [J]. *Neural Networks*, 1993, 6(7):933–946.
- [17] ROHRS C E, VALAVANI L S, ATHANS M, et al. Robustness of continuous-time adaptive control algorithms in presence of unmodeled dynamics [J]. *IEEE Transactions on Automatic Control*, 1985, 30(9):881–889.
- [18] CASTILLO O, NEYOY H, SORIA J, et al. A new approach for dynamic fuzzy logic parameter tuning in Ant Colony Optimization and its application in fuzzy control of a mobile robot [J]. *Applied Soft Computing*, 2015, 28:150–159.
- [19] CAO S G, REES N W, GANG F. Analysis and design of fuzzy control systems using dynamic fuzzy-state space models [J]. *IEEE Transactions on Fuzzy Systems*, 1999, 7(2):192–200.
- [20] SIRA-RAMÍREZ H. On the dynamical sliding mode control of nonlinear systems [J]. *International Journal of Control*, 1993, 57(5):1039–1061.
- [21] LEE H, UTKIN V I. Chattering suppression methods in sliding mode control systems [J]. *Annual Reviews in Control*, 2007, 31(2):179–188.
- [22] CHENG L, HOU Z G, TAN M, et al. Neural-network-based adaptive leader-following control for multiagent systems with uncertainties [J]. *IEEE Transactions on Neural Networks*, 2010, 21(8):1351–1358.
- [23] SUN C, HE W, HONG J. Neural network control of a flexible robotic manipulator using the lumped spring-mass model [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, 47(8):1863–1874.
- [24] GAO H, HE W, ZHOU C, et al. Neural network control of a two-link flexible robotic manipulator using assumed mode method [J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(2):755–765.
- [25] YAO Q J, JAHANSHAHI H, MOROZ I, et al. Neural adaptive fixed-time attitude stabilization and vibration suppression of flexible spacecraft [J]. *Mathematics*, 2022, 10(10):1667.
- [26] 卫晓娟, 李宁洲, 张惠, 等. 一类含间隙碰撞振动系统混沌运动的 RBF 神经网络控制 [J]. *振动工程学报*, 2018, 31(2):336–342.
- WEI X J, LI N Z, ZHANG H, et al. Chaos control of a vibro-impact system with clearance based on RBF neural network [J]. *Journal of Vibration Engineering*, 2018, 31(2):336–342. (in Chinese)
- [27] ARULKUMARAN K, DEISENROTH M P, BRUNDAGE M, et al. Deep reinforcement learning: a brief survey [J]. *IEEE Signal Processing Magazine*, 2017, 34(6):26–38.
- [28] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. *Nature*, 2015, 518:529–533.
- [29] WURMAN P R, BARRETT S, KAWAMOTO K, et al. Outracing champion Gran Turismo drivers with deep reinforcement learning [J]. *Nature*, 2022, 602(7896):223–228.
- [30] 胥彪, 赵琛钰, 李爽, 等. 基于深度强化学习的高超声速飞行器动态面控制方法 [J]. *飞控与探测*, 2023, 6(1):15–23.
- XU B, ZHAO C Y, LI S, et al. Dynamic surface control method for hypersonic vehicle based on deep reinforcement learning [J]. *Flight Control & Detection*, 2023, 6(1):15–23. (in Chinese)