Reinforcement Learning for Active Damping of Harmonically Excited Pendulum with Highly Nonlinear Actuator

James D. Turner, Levi H. Manring, and Brian P. Mann

Department of Mechanical Engineering and Materials Science
Pratt School of Engineering

Duke University, Durham, NC 27708

ABSTRACT

Active vibration dampers can reduce or eliminate unwanted vibrations, but determining a good control policy can be challenging for highly nonlinear systems. For these types of systems, reinforcement learning is one method to optimize a control policy with only limited prior knowledge of the system dynamics. An experimental system was constructed by attaching a permanent magnet to the end of a pendulum and positioning an electromagnetic actuator below the resting position of the pendulum. The pendulum was excited with a sinusoidal force applied horizontally at the pivot point, and the control input was the applied voltage across the electromagnet. Due to the geometric arrangement and the strong dependence of magnetic force on distance, the relationship between the position of the pendulum and the actuation torque for any control input was highly nonlinear. A generalized version of the PILCO reinforcement learning algorithm was used to optimize a control policy for the electromagnet with the objective of minimizing the distance between the end of the pendulum and the downward position. After 16 s of interaction with the experimental system, the resulting learned policy was able to substantially reduce the amplitude of oscillation. This experiment illustrates the applicability of reinforcement learning to highly nonlinear active vibration damping problems.

Keywords: Reinforcement Learning, Active Damping, Nonlinear Dynamical System, Nonlinear Control, Vibration

INTRODUCTION

Vibration in mechanical structures such as machines, buildings, and vehicles can reduce user comfort and damage the structure due to high stresses or fatigue. One approach to reduce vibration amplitude is to add an active damper that provides controlled force inputs to the structure. Applications of active vibration dampers include helicopters, optical systems such as cameras, and manufacturing machinery [1]. Reinforcement learning (RL) provides an approach for controlling active dampers when traditional techniques are insufficient. For example, RL can handle cases where developing an analytical model of the system dynamics is difficult. Additionally, RL can handle optimal control problems such as minimizing total actuation energy to reach and maintain a desired amplitude of oscillation. RL does not require prior knowledge of the system dynamics; the control policy is improved based on information collected while interacting with the system.

Interacting with an experimental system can be time-consuming or expensive, so an important characteristic of RL algorithms is their sample efficiency, i.e. how much experience is required to learn a good policy. One RL algorithm known for its sample efficiency is the probabilistic inference for learning control (PILCO) algorithm [2, 3]. PILCO approximates the system dynamics with a Gaussian process (GP) model learned from data collected while interacting with the system, and then improves the policy by simulating the system with the GP model.

An experiment was conducted to determine the applicability of an extended version of the PILCO algorithm to the system illustrated in Fig. 1 with the goal of reducing the amplitude of oscillation of the pendulum. The pivot was fixed to a shaker table, a permanent magnet was attached to the end of the pendulum, and an electromagnet was placed directly under the pivot. The pendulum was driven by a 1 Hz sinusoidal force applied to the shaker table, and the system was controlled by applying voltage v across the electromagnet. Due to the geometric configuration and strong dependence of magnetic force on distance, the electromagnet could apply a noticeable torque on the pendulum only in a narrow band on either side of the downward position. The objective was to minimize the angle of the pendulum over an episode of 4 s, with the system starting from steady-state forced oscillation at $t = 0$. The control frequency was 20 Hz, and the applied voltage was constrained to $-v_{\max} \leq v \leq v_{\max}$.

METHODS

The PILCO algorithm optimizes the parameters for a control policy for an iterative dynamical system given no prior knowledge of the system dynamics. The PILCO algorithm assumes that the system is fully observable, the next state depends solely on the state $\boldsymbol{x}_t$ and control $\boldsymbol{u}_t$, and the dynamics are continuous and smooth. The objective is to find the parameters $\boldsymbol{\psi}$ for a control policy $\boldsymbol{u}_t = \boldsymbol{\pi}(\boldsymbol{x}_t, \boldsymbol{\psi})$ that minimize the total expected cost $J^{\boldsymbol{\pi}}(\boldsymbol{\psi}) = \sum_{t=0}^{T} \mathbb{E}_{\boldsymbol{x}_t}\left[c(\boldsymbol{x}_t)\right]$ of controlling the system for an episode of T time steps, where $c(\boldsymbol{x}_t)$ is a function of the difference between $\boldsymbol{x}_t$ and a target state $\boldsymbol{x}_{\text{target}}$. Briefly, the PILCO algorithm works by learning an approximate GP model of the dynamics from data, evaluates the policy by simulating the system with the GP and policy, and then improves the policy based on the simulation. The process of applying the policy to the system to collect data, updating the GP, simulating the system, and improving the policy repeats until convergence [2, 3].

For the forced pendulum problem, however, this approach is inadequate: the next state of the system depends not only on the state and control but also on the forcing $\boldsymbol{f}_t$. Assuming the forcing is the same for each episode such that $\boldsymbol{f}_t = \boldsymbol{f}(t)$, the PILCO algorithm can handle this type of problem with two modifications: (1) incorporate $\boldsymbol{f}_t$ as inputs to the GP model and policy, and (2) when simulating with the GP and policy, use the sequence of forcing $(\boldsymbol{f}_0, \ldots, \boldsymbol{f}_T)$ recorded during the most recent episode.

For the experimental system in Fig. 1, the state, forcing, and control were $\boldsymbol{x} = \left\{\theta\ \dot{\theta}\ x\ \dot{x}\ i\right\}^\top$, $\boldsymbol{f} = \left\{f\right\}^\top$, and $\boldsymbol{u} = \{v\}^\top$, respectively, where θ and $\dot{\theta}$ were the angle and angular velocity of the pendulum, x and $\dot{x}$ were the position and velocity of the shaker table, i was the current in the electromagnet, f was the external force applied to the shaker table, and v was the applied voltage across the electromagnet. However, the mass of the shaker table in the experimental system was so much larger than the mass of the pendulum that its motion could be approximated as being independent of the motion of the pendulum. This simplified the problem because the position and velocity of the shaker table could be handled as forcing terms. The new state and forcing vectors were then $\boldsymbol{x}_{\text{o}} = \left\{\theta\ \dot{\theta}\ i\right\}^\top$ and $\boldsymbol{f}_{\text{i}} = \left\{f\ x\ \dot{x}\right\}^\top$. It was also useful to split the angle into components for input to the GP and policy, so this state vector was $\boldsymbol{x}_{\text{i}} = \left\{\sin\theta\ \cos\theta\ \dot{\theta}\ i\right\}^\top$. The GP model $\mathcal{GP}$ was then $\boldsymbol{x}_{\text{o},\,t+1} - \boldsymbol{x}_{\text{o},t} = \mathcal{GP}\left(\boldsymbol{x}_{\text{i},t},\ \boldsymbol{f}_{\text{i},t},\ \boldsymbol{u}_t\right)$, and the policy function $\boldsymbol{\pi}$ was $\boldsymbol{u}_t = \boldsymbol{\pi}\left(\boldsymbol{x}_{\text{i},t},\ \boldsymbol{f}_{\text{i},t},\ \boldsymbol{\psi}\right)$. The target was $\theta = 0$.

RESULTS

Figure 2 shows the steady-state response of the system with $v = 0$. The amplitude was 0.27 rad, and the frequency was the same as the forcing frequency, 1 Hz. Figures 3 to 6 show the results of experimental episodes with various control policies, where each episode started at $t = 0$ at the point of the steady-state response with maximum θ.

Figure 4 shows the result of applying a constant voltage $v = -v_{\max}$. This corresponds to always attracting the pendulum to the electromagnet with as much force as possible. This policy immediately reduced the response amplitude to 0.18 rad. Then, the transient response gradually died out with damping, reaching a new steady-state response with amplitude 0.02 rad. This demonstrated that the amplitude could be kept small after dissipating as much energy as possible, but simply applying $v = -v_{\max}$ required the inherent damping of the system to reduce the amplitude to this level.

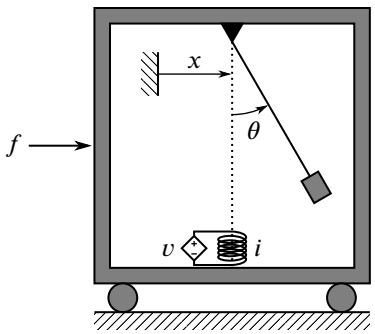

Fig. 1 Schematic of experimental system

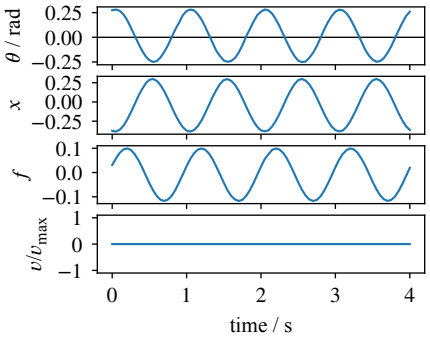

Fig. 2 Steady-state response with $v = 0$

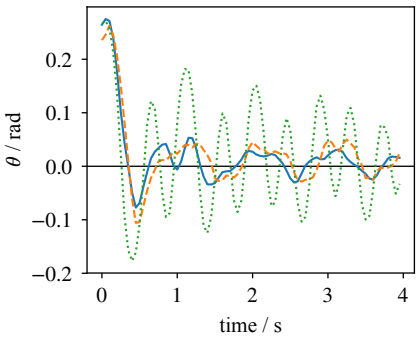

Fig. 3 Comparison of policy performance: $v = -v_{\max}$ (dotted green), bang–bang (dashed orange), PILCO-generated (solid blue)

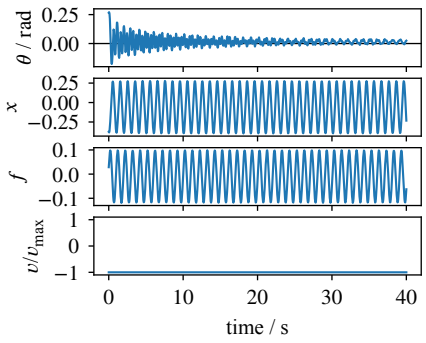

Fig. 4 Experimental episode when applying a constant $v = -v_{\max}$. Note that the time duration shown in the figure is longer than a single episode in order to illustrate the long-term behavior of this policy

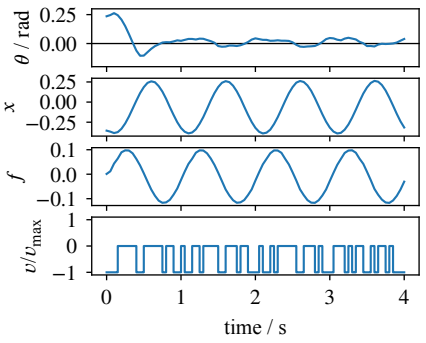

Fig. 5 Experimental episode when applying bang–bang control with control values $v \in \{-v_{\max}, 0\}$

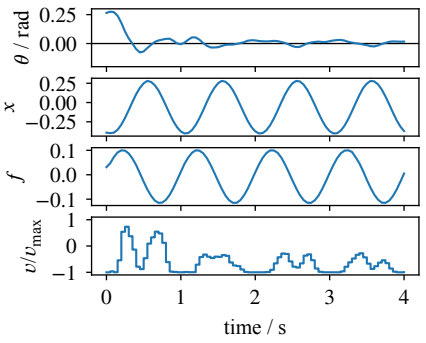

Fig. 6 Experimental episode when applying optimized policy generated by PILCO after 16 s of experience

Given prior knowledge of the dynamics, another approach was a bang–bang policy. This policy applied $v = 0$ while the pendulum was moving towards the target and $v = -v_{\max}$ while it was moving away from the target. Figure 5 shows that this policy rapidly brought the pendulum near the target and kept it there, substantially outperforming the constant voltage policy.

Figure 6 shows the result of a policy generated by the extended PILCO algorithm after 4 episodes of experience. The series of control values is more interesting than those of the other two policies. This policy repelled the pendulum as it moved toward the target in order to reduce overshoot, then mostly attracted it to keep its amplitude small.

Figure 3 shows a comparison of the pendulum angle for the episodes of all three policies. Both non-constant policies performed substantially better than the constant $v = -v_{\max}$ policy. The difference between the bang–bang and PILCO-generated policies was less significant, but the PILCO-generated policy exhibited less overshoot and overall 15% lower cost. Unlike the bang–bang policy, which was developed from intuition of the system's behavior, the PILCO-generated policy was based solely on 16 s of experience, with no prior knowledge of the dynamics.

CONCLUSION

A pendulum subject to sinusoidal external forcing and controlled by an electromagnet positioned below its pivot was constructed. An extended version of the PILCO reinforcement learning algorithm was used to optimize a policy with the goal of minimizing the angle of the pendulum from the downward position for a 4 s episode. The policy generated after 16 s of experience, with zero prior knowledge of the system dynamics, outperformed a simple human-generated bang–bang policy. This experiment demonstrated the applicability of the extended PILCO algorithm to highly nonlinear forced vibration problems.

ACKNOWLEDGMENTS

Funding was provided by Army Research Office (ARO) grant W911NF-17-0047 and the National Defense Science & Engineering Graduate (NDSEG) Fellowship.

REFERENCES

[1] Takács, G., and Rohal'-Ilkiv, B. *Model Predictive Vibration Control: Efficient Constrained MPC Vibration Control for Lightly Damped Mechanical Structures*. Springer, 2012. DOI: 10.1007/978-1-4471-2333-0.

[2] Deisenroth, M. P., and Rasmussen, C. E. "PILCO: A Model-based and Data-Efficient Approach to Policy Search". *Proceedings of the International Conference on Machine Learning*. 2011.

[3] Deisenroth, M. P., and Rasmussen, C. E. *A Practical and Conceptual Framework for Learning in Control*. Tech. rep. UW-CSE-10-06-01. Department of Computer Science & Engineering, University of Washington, June 2010.