



To

Editorial Board Member

23 Aug 2022

Changyi Lei

Tel: +4407422506837

Email: changyi.lei@kcl.ac.uk

Dear Prof Khan Abdul Qadeer,

**Re: U-model-based Adaptive Sliding Mode Control Using Deep Deterministic Policy Gradient (ID: 8980664) by Changyi Lei, Quanmin Zhu**

Many thanks for your instruction for the revision of this study. We have revised the paper in line with the editor and reviewer recommendations and corrections, and positively responded to all of the comments/questions raised. Please find the replies to the editor and the reviewers attached to the cover letter.

In addition, we have shown our gratitude to the editors and the anonymous reviewers for their helpful comments and constructive suggestions with regard to the revision of the paper in the Acknowledgements.

Should you have any further advice with regard to this paper, please contact us again.

Yours sincerely,

*Changyi Lei*

## **Comment from Editor**

Major Revision Requested

Authors should revise the paper per reviewers' comments.

## **Comments from Reviewer #1**

This manuscript combines U-model framework and sliding mode control to solve trajectory tracking problem. Reinforcement learning is implemented to attenuate the chattering of SMC while maintaining tracking accuracy and convergence speed. The main contribution lies in introducing reinforcement learning to solve the chattering problem, providing a novel perspective to researchers in this field. Also, it is an extension and successful supplement to existing U-model-based control system design.

Advantages:

The methodology is novel.

The working flow of mathematic derivation is clear.

The work is complete, and simulation result looks good.

Description of methodology is comprehensive for the reproduction of the work.

Disadvantages:

- 1) The dynamic plant used for simulation is relatively simple, not only the dynamics itself, but also the way assigning noise.

### **Response**

Indeed, the dynamics we used is a single pendulum, and the noise assigned is random value within certain range. However, due to limited time and resources, this is only preliminary research into the combination of RL and SMC in dealing with chattering problem, which is a novel idea as far as we know. We hope to show that RL works well in terms of achieving adaptive SMC controller, which serves as an alternative to the other adaptive SMC methods, and can potentially applied to complement the other adaptive controllers.

About the random noise added, it is also a common practice to add total disturbance to the system with certain bounds. As discussed in Assumption 1 and Remark 1 under section 3.2.1. Besides, random noise is discontinuous signal with sudden change, which poses great challenge on the controller. Therefore, the robustness against random noise is not an easy requirement on the controllers.

- 2) Fully connected network is computationally expensive. An RBF network maybe better.

### **Response**

Indeed, there are many calculations in fully connected layer. However, the deep learning library has been optimized over the years. The forward calculation time of our model is only about 600 microseconds, which satisfies real-time requirements. The following sentence is added to the 1st paragraph of section 4.4:

“Through testing, the forward calculation time of our deep learning model is only about 600 microseconds, which satisfies real-time requirements.”

Revision should be considered before next round review

3) In algorithm 1, some inconsistent writing of  $\pi$ .

**Response**

Yes, we have unified the writing of  $\pi$  and  $Q$  in algorithm 1.

4) In 3.4, it should be critically damped instead of overdamped.

**Response**

Yes, we have corrected the typo.

5) Double check other formulations and typos.

**Response**

Yes, we have double checked.

6) An introduction about the future researches in this field should be included in the last of conclusion.

**Response**

We think it will be too bulky if we further add an introduction about the future researches in this field. For one thing, along with suggestions from other reviewers, we have added the challenges of proposed method, the merits of combining RL and SMC to the conclusion part. On the other hand, we have included future work with respect to our proposed method. This is our personal opinion, but it is a very good advice.

7) English reads fine but could be further fine-tuned.

**Response**

Yes, we have polished the writing in many places.

8) The challenges of the proposed method should be discussed in the revised paper.

**Response**

The following paragraph has been added to the conclusion section:

“However, compared with the other adaptive methods, there are some challenges exist for RL-based SMC. Firstly, while decreasing human craftsmanship, RL algorithm requires large amount of data for training, which may wear and tear the machines in practice. Many methods can be implemented to increase the data efficiency and accelerate convergence, e.g., model-based reinforcement learning, imitation learning to warm up. Secondly, the generalization problem. RL algorithms tend to overfit to specific scenarios. How to bridge the gap between simulation and reality, as well as how to transfer the model to another scenario, are open questions.”

## Comments from Reviewer #2

The manuscript proposes a kind of adaptive sliding mode control (SMC) based the framework of so-called U-model. The adaptive scheme adopts the Deep Deterministic Policy Gradient (DDPG) algorithm with the two networks, which are the actor network and the critic network. The simulation results of the single pendulum system (SPS)

are presented to validate their controller design. The manuscript was written well. However, I cannot endorse the idea of the control scheme for the overall system. The concerns are as follows.

- 1) First, the adaptive SMC idea is a traditional idea in the SMC realm, which were developed in many plentiful forms with neural networks (NNs), fuzzy logic systems (FLSs) and genetic algorithm (GA), and so on. The baseline to be compared should be a typical intelligent adaptive SMC with these methods, but not the so-called "SMC optimized by Particle Swarm Optimization (PSO) algorithm", because the author didn't give the mechanism in details about the PSO method. For instance, what did the PSO optimized? How did it optimize? The reader cannot see.

#### **Response**

Like GA, PSO borrows idea from swarm intelligence to find global optimum of the cost function. PSO is also widely used in SMC for optimal design. The detailed implementation of PSO has been added in section 4.3.

In this paper, the baseline to compare is not selected as adaptive method but PSO algorithm. It is mainly because this is only application-oriented preliminary research. The implementation of DDPG, as well as many details are simple version. The main goal is to illustrate the potential of DDPG in solving chattering problem, and of course, many delicate designs still need to be considered to maximize its outcome. This has been added to the conclusions as future work. Specifically, the following sentences are added to the last of the conclusion section:

"Last but not least, the implementation of DDPG in this paper is only a simple version. More delicate considerations need to be implemented to maximize the potential of reinforcement learning in the future research. And then, the DDPG-based method should be compared with other typical adaptive SMC controllers."

- 2) By my individual opinion, the proposed adaptive scheme, which uses the DDPG algorithm to tune the parameters  $\eta$  and  $\delta$  in the SMC controller, is the same essence with the traditional intelligent adaptive SMC. Therefore, I don't think the authors have the great contribution. This is my opinion.

#### **Response**

Indeed, as you point out in your fourth comment, the merit of this paper is the application of deep reinforcement learning algorithm, but not in theoretical derivations. The motivation of this paper lies in using DDPG to find optimal mapping from state variables to the SMC parameters. In comparison, there are many adaptive methods exist for SMC, but most of them require tedious manual design and cannot ensure optimality. We hope to utilize the advantages of DDPG to reduce human craftsmanship and increase the optimality of the controllers.

On the other hand, besides the implementation of DDPG, this paper also includes

detailed explanation of the model behaviour, using output visualization and SHAP (just added in section 4.7). Note that the transparency and explainability determines human's trust of the AI model, and consequently the implementation in practice. However, as far as we are concerned, few papers of AI controllers render explanation of the model behaviour.

- 3) I don't agree with the U-model Control Framework. This kind of framework is not a good theoretical framework. The reason is, the accurate  $G_p^{-1}$  in the loop (Fig.1) is very difficult to find from the robust control point of view. Then, what's your advantage? I cannot think about it. In the manuscript, the SMC controller seems to be considered as the dynamic inverter  $G_p^{-1}$ ? This is a kind of unnecessary actions. The dynamic can be designed by the sliding mode equation in the SMC theory, we don't need the invariant part  $G_I$ . This point of view is my individual.

**Response**

Indeed, the accurate  $G_p^{-1}$  in the loop (Fig.1) is very difficult to find, therefore a robust dynamic inverter is required. That is why we used SMC to be the dynamic inverter.

The usage of the invariant part  $G_I$  is combined with the design of unit feedback loop to increase the ability to reject disturbance. We can prove from simple calculation that any disturbance added to the system will converge to zero exponentially with the passage of time, and the convergence speed is also related to the parameters of the invariant part. Besides, this structure enables us to design the controller and the performance separately. So that the controller only needs to ensure stability, while the performance assignment can be done apart.

- 4) The merit of the manuscript is the application of DDPG algorithm. However, I don't support the application in such a way. The maximum advantage of these artificial intelligent (AI) method is the "black box" consideration. The authors should consider how to use these AI methods in a model-free way, not as now the math model still be required. Therefore, I don't agree with the superiority of the proposed methodology.

**Response**

We think the combination of AI method and conventional controllers like SMC is of some value, and there are indeed many papers integrating AI and conventional controllers, such as RL-PID. Specifically, in this paper, we think RL and SMC complements each other:

Firstly, from the controller point of view, the implementation of AI is just an alternative method to transform original controllers into adaptive controllers. Compared with typical adaptive controller design methods, AI methods can reduce manual efforts and find optimum automatically.

Secondly, combining with SMC enables the reservation of some explainability and robustness for RL. On the one hand, while we cannot understand the DDPG output directly, we can have an intuitive understanding if combining with the formula of SMC. On the other hand, RL algorithm tends to overfit, and the robustness of SMC can help overcome this advantage. Even when the real environment and dynamics is not completely the same with training (due to wear, tear, hysteresis etc), SMC maintains a certain range of stability, which prevents the RL from failure.

The following sentences are added to the last of the 1<sup>st</sup> paragraph of conclusion section.

- 5) Some minor problems please pay attention. For example, after the equation (3), "Take the derivative of (3) and integrate with (9)", here must not be the equation (9).

#### **Response**

Thank you for pointing it out. We have double checked other writing, and corrected this sentence to:

"Take the derivative of (3) and integrate with (2)"

### **Comments from Reviewer #3**

This paper presents a U-model-based adaptive Sliding Mode Control (SMC) using Deep Deterministic Policy Gradient (DDPG) for uncertain nonlinear systems. The configuration of the proposed methodology is consisted of a U-model framework and an SMC with variable boundary layer. The U-model framework forms the outer feedback loop that adjusts the overall performance of the nonlinear system, while SMC serves as a robust dynamic inverter that cancels the nonlinearity of the original plant. Besides, to alleviate the chattering problem while maintaining the intrinsic advantages of SMC, a DDPG network is designed to adaptively tune the boundary and switching gain. From the control perspective, this controller combines the interpretability of U-model and the robustness of SMC. From the deep reinforcement learning (DRL) point of view, the DDPG calculates nearly optimal parameters for SMC based on current states and maximizes its favourable features while minimizing the unfavourable ones. The simulation results of the single-pendulum system is compared with U-model-based SMC optimized by Particle Swarm Optimization (PSO) algorithm. The comparison as well as model visualization demonstrates the superiority of the proposed methodology. Generally, this is a quite interesting work. It can be accepted if the authors can consider the following issues such as:

- 1) What is the main motivation of the work? Why did the authors combine the SMC and DDPG?

**Response**

As discussed in section 1, the motivation of the work is to combine SMC and U-model, using SMC to increase the robustness of U-model control framework. In the meantime, to solve the chattering problem of SMC while maintaining its convergence speed, we utilized DDPG for the online tuning of SMC. Compared with other adaptive methods, DDPG is nearly optimal, data-driven, requires minimal human craftsmanship, and do not increase the order of the systems. Besides, a combination of SMC and reinforcement learning is still rare as far as we are concerned. Therefore, we combined SMC and DDPG.

- 2) Is the proposed method applicable for other kind of systems?

**Response**

Yes, the proposed method can be applied to other kind of systems. There are two core procedures for applying this method. Firstly, derivation of conventional SMC controller. A standard process exists for the calculation of SMC controller using backstepping. Secondly, the training of DDPG. DDPG is a data-driven method, so users only need to tune the hyperparameters of the network and the reward function. This requires minimal prior knowledge and human craftsmanship. It is really applicable when the tuning is non-intuitive or complicated.

- 3) Is the DDPG running in real time condition?

**Response**

Yes. Indeed, there are many calculations in fully connected layer. However, the deep learning library has been optimized over the years. The forward calculation time of our model is only about 600 microseconds, which satisfies real-time requirements. The following sentence is added to the 1st paragraph of section 4.4:

“Through testing, the forward calculation time of our deep learning model is only about 600 microseconds, which satisfies real-time requirements.”

- 4) Normally, the chattering problem of SMC is a big issue. How did the authors deal with the chattering problem.

**Response**

There are two things we did to deal with the chattering problem. Firstly, as discussed in section 3.2.2, we used variable boundary layer for conventional SMC. Secondly, in section 4.2, we trained a DDPG network to tune the parameters of the boundary layer, with penalization on both the tracking accuracy and the thickness of boundary layer. In this way, the network can ensure tracking accuracy while mitigating the chattering problem.

- 5) More related works on the topic are welcome to enrich the literature review such as Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets; Drag Coefficient Modeling of Heterogeneous Connected Platooning Vehicles via BP Neural Network and PSO Algorithm; Learning-based path following controller design for autonomous ground vehicles subject to stochastic delays and actuator constraints

**Response**

The second paper is most relevant to my article, so it is cited at the end of the 2<sup>nd</sup> paragraph in section 1. The following sentence is added:

“(high-order SMC \cite{Amrr2021RobustCD}, )neural network and PSO \cite{LUO2022117} etc.”

- 6) More words are welcome for the general U-model Control Framework. Normally, for a nonlinear system, what is the inverse of GP.

**Response**

Section 2 has been revised to provide more explanation of conventional U-model.

- 7) More examples and comparisons are welcome.

**Response**

After careful consideration, we decide to pay more attention to the explanation of the model. Therefore, section 4.7 is added using XAI method – SHAP, to give detailed explanation of the model behaviour. Besides, current reference signals (a step input and a sine wave) represent two typical reference targets, which we think is enough to show the fundamental outcome of the proposed method.

At last, thank you very much for all the suggestions given. It really helps a great deal to make this article better.