# Robustness Margins and Robust Guided Policy Search for Deep Reinforcement Learning

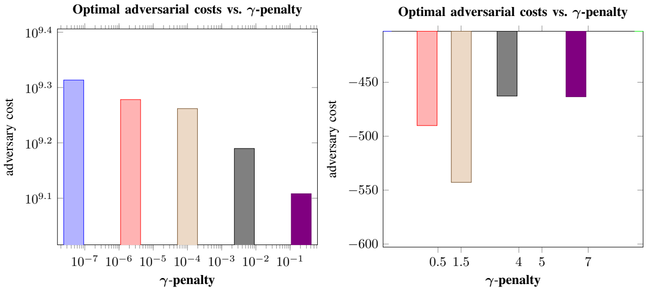
Tyler Summers, Olalekan Ogunmolu, Nicholas Gans

**Context:** Deep reinforcement learning methods aim to deliver high performance control policies on systems with complex difficult-to-model dynamics and with complex multi-modal sensory input [1]. However, many challenges remain, including data efficiency of the learning process and robustness of the resulting policies. Training robots in the real world can be expensive and hazardous, demanding highly data-efficient learning. This requirement can be alleviated by training in simulation, but this then demands robustness of the policy for real-world generalization.

Formal studies of robustness have a long history in feedback control theory and dynamic game theory. One effective way to both quantify and design robust feedback controllers is to consider the performance of the controller in the presence of an adversarial agent [2], [3], [4]. There have been some recent initial investigations into the robustness of deep reinforcement learning [5], though most focus on using model-free methods. However, much more research is required to fully understand the robustness of deep RL policies. Lack of robustness may be exacerbated by the complexity and opaqueness of policies represented by deep neural networks.

**Contributions:** We present methods for both quantifying and designing robust feedback control policies for robotic systems using deep reinforcement learning that can identify weaknesses and facilitate generalization from simulation to real world.

*a) Quantifying Robustness:* We quantify the robustness of a given fixed control policy by training adversarial agents against it. Multiple adversarial agents are trained with varying values of a scalar parameter $\gamma$ that limits capability by increasing the penalty on its input cost; if an adversary is able to cause unacceptable performance degradation with only small perturbations, then the given control policy is not robust and is unlikely to generalize beyond its training environment.



*b) Designing Robustness:* We also propose a Robust Guided Policy Search (R-GPS) framework for designing robust deep RL policies by simultaneously training a control policy and adversarial policy in a two-player, zero-sum Markov game setting. We adapt a recent Guided Policy Search framework [1] for training policies, which effectively combines optimal model-based trajectory generation and feedback control with data-based reinforcement learning and multi-layer neural networks for approximating control policies and value functions. Specifically, we develop an iterative Dynamic Game (iDG) method to generate a set of locally robust trajectories and local policies, and an alternating best response update of global control and adversary policies to obtain convergence to a saddle-point equilibrium. The basic ideas are essentially meta-algorithms that can in principle be extended to quantify and design robustness of policies for a variety of model-based and model-free reinforcement learning approaches.

*c) Numerical Results:* We implement our algorithms for a peg insertion task with a robotic arm that requires dexterous manipulation. A controller for this task is initially trained without an adversary using GPS. An adversary is then trained using the same GPS method against the system in closed-loop with the initial trained controller, with various values of $\gamma$ that effectively limit the torque it can apply. The figure above shows that the adversary causes a sharp degradation in the controller performance for $\gamma < 1.5$, as the arm is destabilized. For $\gamma \geq 1.5$, the adversary has a much smaller effect on controller performance, and this effect generally decreases for large $\gamma$. Videos illustrating the results are available at https://goo.gl/YmmdhC. Numerical experiments with our Robust GPS algorithm for simultaneous control and adversary training are forthcoming and will produce control policies that are significantly more robust than those of other state-of-the-art approaches.

## References

[1] S. Levine, C. Finn, T. Darrell, and P. Abbeel, “End-to-end training of deep visuomotor policies,” *Journal of Machine Learning Research*, vol. 17, no. 39, pp. 1–40, 2016.

[2] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.

[3] M. L. Littman, “Markov games as a framework for multi-agent reinforcement learning,” in *Proceedings of the eleventh international conference on machine learning*, vol. 157, 1994, pp. 157–163.

[4] J. Morimoto and K. Doya, “Robust reinforcement learning,” *Neural computation*, vol. 17, no. 2, pp. 335–359, 2005.

[5] L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, “Robust adversarial reinforcement learning,” *arXiv preprint arXiv:1703.02702*, 2017.

T. Summers is with the Department of Mechanical Engineering, University of Texas at Dallas. O. Ogunmolu and N. Gans are with the Department of Electrical Engineering, University of Texas at Dallas. {tyler.summers,olalekan.ogunmolu,ngans}@utdallas.edu