An iterative Dynamic Game Approach for Robust Deep Reinforcement Learning

Olalekan Ogunmolu[⋆] [*]

1 Introduction

Deep reinforcement learning methods aim to deliver high performance control policies on systems with complex or difficult-to-model dynamics as well as systems with complex multi-modal sensory input [1]. Despite its apparent success, many challenges remain, including data efficiency of the learning process and the robustness of the resulting policies. Training robots in the real world can be expensive and hazardous, often demanding highly data-efficient learning. The conventional way to circumvent hazardous gathering of data in the real-world is to provide the learning behavior for the robot by training in simulation. However, this demands robustness of the policy for real-world control or navigation and real-world generalization.

2 Related Works

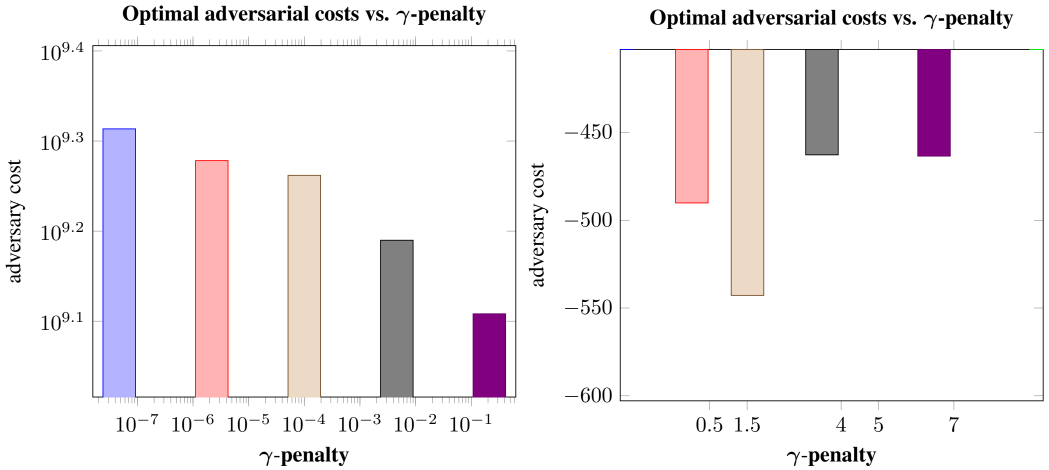
Formal studies of robustness have a long history in feedback control theory and dynamic game theory. One effective way to both quantify and design robust feedback controllers is to consider the performance of the controller in the presence of an adversarial agent [2, 3, 4]. There have been some recent 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. Lack of robustness can be an especially critical issue even in linear quadratic Gaussian settings [6], and this may be exacerbated by the complexity and opaqueness of policies represented by deep neural networks. There are fundamental tradeoffs between performance and robustness; the higher the expected reward an agents achieves, the more vulnerable to uncertainty it becomes. Robustness can be an especially critical issue in partially observable settings: even in the model-based linear quadratic Gaussian case, it has been long established that there are *zero* robustness margins even when using the globally optimal information feedback policy comprising a Kalman filter together with a linear state feedback controller [6]. Lack of robustness has also been observed and studied recently in static image classification contexts with Generative Adversarial Networks [7].

3 Lecture Contents

In this lecture, 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.

Quantifying Robustness To quantify the robustness of a given fixed control policy, we train adversarial agents against it. Multiple adversarial agents are trained with varying values of a scalar parameter γ 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.

[*][⋆]Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA 19104 {ogunmolo}@pennmedicine.upenn.edu



Designing Robustness We also present 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.

Numerical Results We implement our algorithms for a peg insertion task of [1] 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 γ 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 γ. 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] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.

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

[3] Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the eleventh international conference on machine learning*, volume 157, pages 157–163, 1994.

[4] Jun Morimoto and Kenji Doya. Robust reinforcement learning. *Neural computation*, 17(2):335–359, 2005.

[5] Lerrel Pinto, James Davidson, Rahul Sukthankar, and Abhinav Gupta. Robust adversarial reinforcement learning. *arXiv preprint arXiv:1703.02702*, 2017.

[6] John Doyle. Guaranteed margins for lqg regulators. *IEEE Transactions on Automatic Control*, 23(4): 756–757, 1978.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.