Implicit Bias of Policy Gradient in Linear Quadratic Control: Extrapolation to Unseen Initial States

Noam Razin*

Joint work with Yotam Alexander*, Edo Cohen-Karlik, Raja Giryes, Amir Globerson, Nadav Cohen

ICML 2024

Policy Gradient in Optimal Control

Optimal Control Problem

System: Starting from an initial state $\mathbf{x}_0$

$$\mathbf{x}_{h+1} = f(\underbrace{\mathbf{x}_h}_{\text{state}}, \underbrace{\mathbf{u}_h}_{\text{control}}) \qquad h = 0, \ldots, \underbrace{H}_{\text{time horizon}} - 1$$

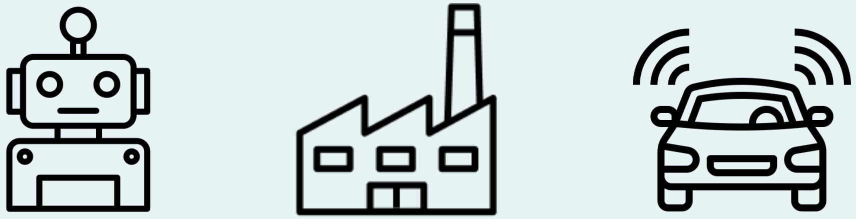
Goal: Choose controls that minimize the cost $\sum_{h=0}^{H} c(\mathbf{x}_h, \mathbf{u}_h)$

Policy Gradient

- Parameterize controller (e.g. as neural network)
- Minimize cost via **gradient descent** w.r.t. controller parameters

Extrapolation to Unseen Initial States

Issue of Prime Importance: Extrapolation to **initial states unseen in training**



Often multiple controllers minimize cost for **initial states seen in training**

→ Extrapolation is determined by the **implicit bias** of policy gradient

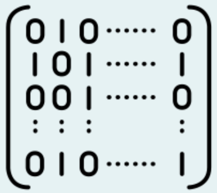
Effect of implicit bias on extrapolation was theoretically studied in supervised learning

(Xu et al. 2021, Abbe et al. 2022/23, Cohen-Karlik et al. 2022/23)

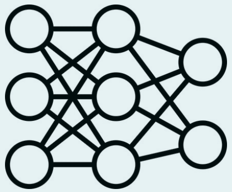
limited understanding in control

Main Contributions: Effect of Implicit Bias on Extrapolation

Q: To what extent does the implicit bias of policy gradient lead to extrapolation to initial states unseen in training?



Theory for the Linear Quadratic Regulator (LQR) Problem:
Extrapolation depends on exploration induced by the system from initial states seen in training



Experiments:
Support theory for LQR and demonstrate its conclusions apply to non-linear systems and neural network controllers

Going Forward:

- We hope our work will encourage further research on implicit bias in control
- Potential practical application: Algorithms for selecting initial states to train on