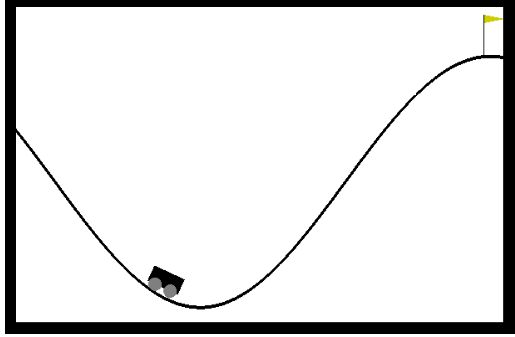
Advances in Reinforcement Learning Inspired by Statistical Mechanics

J. Adamczyk

Advisor: Prof. Rahul V. Kulkarni

Collaborators: Stas Tiomkin, Volodymyr Makarenko, Argenis Arriojas, Di Luo

Reinforcing Rewards

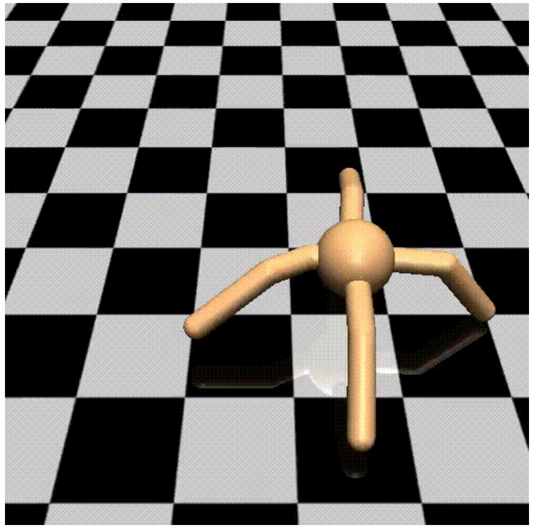


Mountain-Car environment

Reinforcement Learning (RL) is a method of solving sequential decision-making problems by interacting

<u>Basic Idea:</u>

- An agent interacts with the **environment**, by taking **actions**
- Positive behaviors are reinforced relative to undesirable behaviors
  - Reinforcement is implemented via a **reward function**
- The agent should learn to maximize rewards received
  - Specifically, average reward over (infinitely) long trajectories




184 1 1

Bountiful Bridge

<u>Statistical Physics</u>
- Free Energy
- Ground State
- Correlation Scales
- Hamiltonian MCMC
- Tightening Bounds

<u>Reinforcement Learning</u>
- Value Functions
- Max Reward-Rate
- Mixing Time
- Sampling Methods
- New Algorithms

Fresh Findings

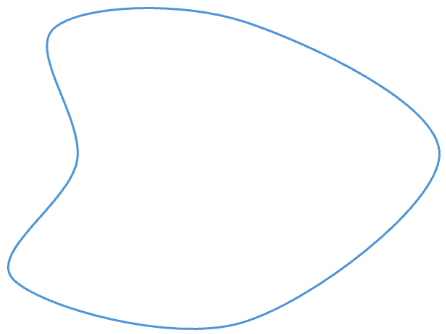
- New proof and perspective of Policy Improvement (PI)
- New algorithm for MaxEnt RL by bounding the Q function [1]
- Ground-state formulation of RL problem
- Extension of Donsker-Varadhan formula to value functions [2]
- Average-Reward Algorithms [3]

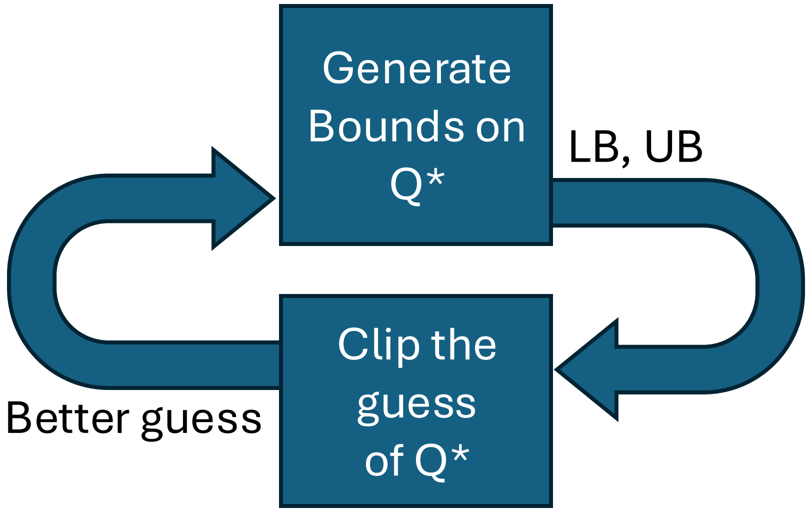
[1]: "Boosting Soft Q Learning by Bounding", Under review at RLC, 2024
[2]: "Eigenvector Based Average-Reward Learning" In Preparation for JMLR
[3]: "Off-Policy Algorithms for Entropy-Regularized Average-Reward RL", Under review at NeurIPS 2024

Bounds Abound

The *optimal* Q function can be bounded from arbitrary guess

Estimate the Q function





Friendly Free Energy

The free energy (Q-func) satisfies a variational form (for arb. p):

$$F \doteq -\frac{1}{\beta}\log\sum_x p_0(x)e^{-\beta E(x)} \leq \mathbb{E}_p\{E + \beta^{-1}D_{\mathrm{KL}}(p||p_0)\} \doteq F_p$$

How big is the gap between true free energy (F) and variational guess (F_p)?

$$Q^*(s,a) = Q^\pi(s,a) - D_{\mathrm{KL}}(\pi|\pi^*)$$

$$F = F_p - D_{\mathrm{KL}}(p|p^*)$$

$$p^*(x) \propto p_0(x)e^{-\beta E(x)}$$

Thank you!

Rahul Kulkarni

Argenis Arriojas

Di Luo

Stas Tiomkin

Vlad Makarenko