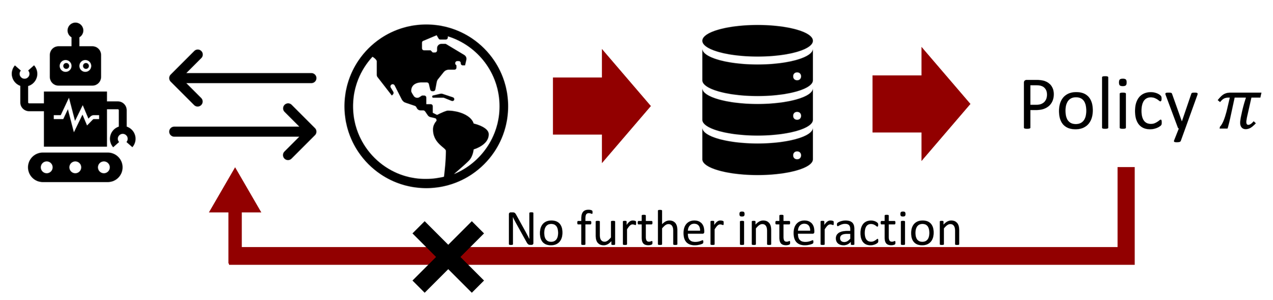
PLAS: Latent Action Space for Offline Reinforcement Learning

Wenxuan Zhou, Sujay Bajracharya, David Held

Carnegie Mellon University

Offline Reinforcement Learning



Offline Reinforcement Learning studies the problem of learning a policy from a static dataset.

Off-policy algorithms cannot be directly applied to offline RL problems due to overestimation bias caused by out-of-distribution actions.

Bellman operator:

$$\mathcal{T}\hat{Q}^{\pi}(s_t, a_t) = \mathbb{E}_{r_t, s_{t+1}}[r_t + \gamma \underline{\hat{Q}^{\pi}(s_{t+1}, \pi(s_{t+1}))}]$$

Avoiding out-of-distribution actions

Objective 1:

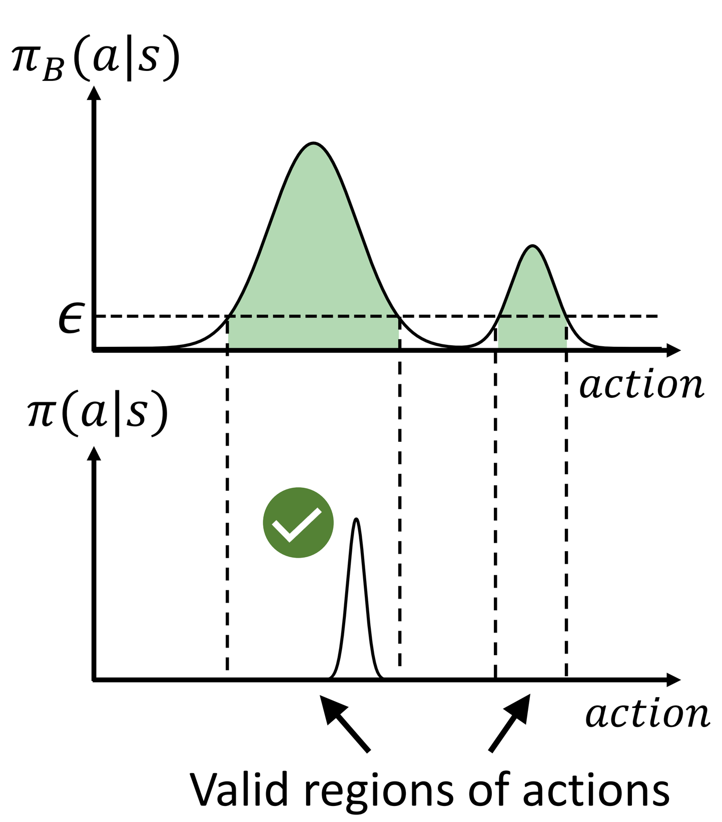
Constrain the policy to select actions within the support of the dataset π_B:

$$\max_{a\sim\pi(\cdot|s)} \mathbb{E}[G_t]$$

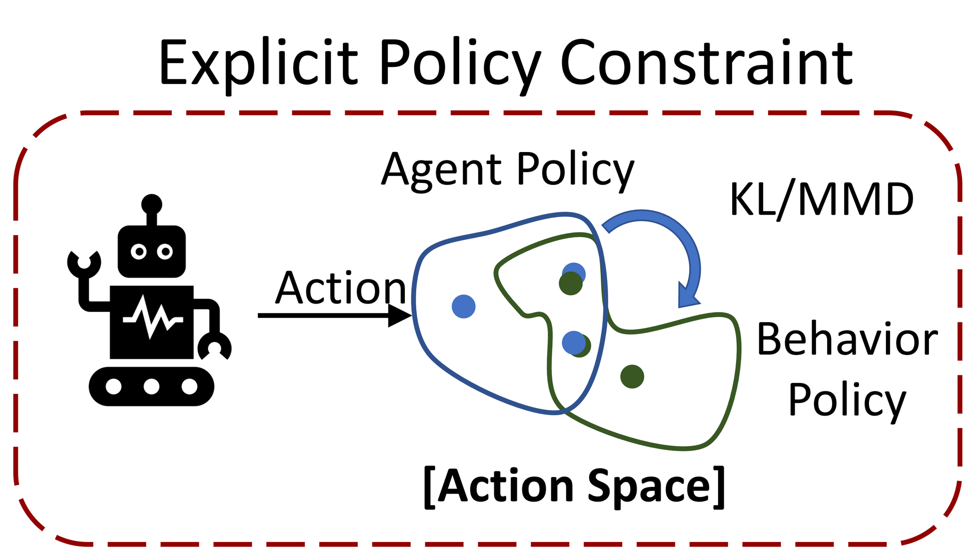
subject to $\pi_B(a|s) > \epsilon$

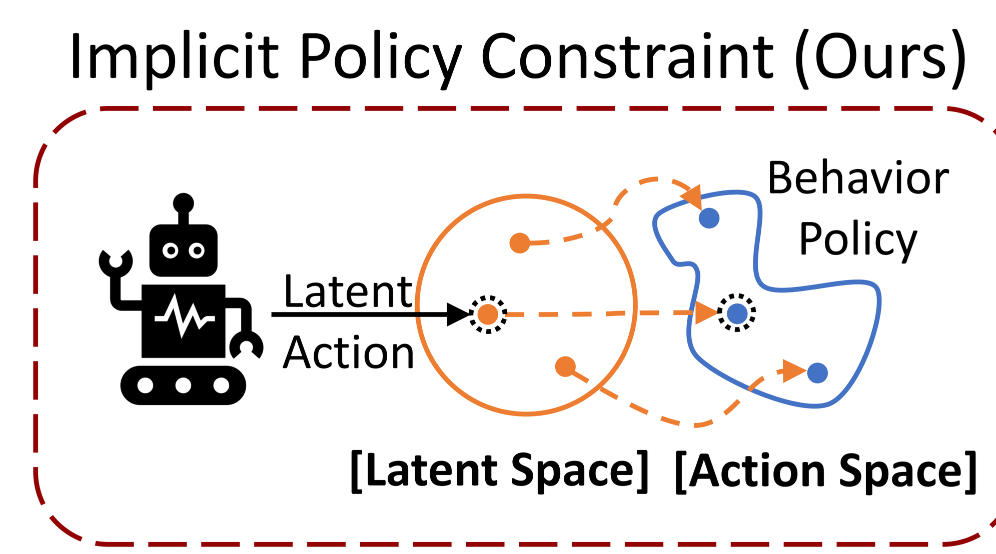
Objective 2:

The policy should not be affected by the density of the dataset π_B.



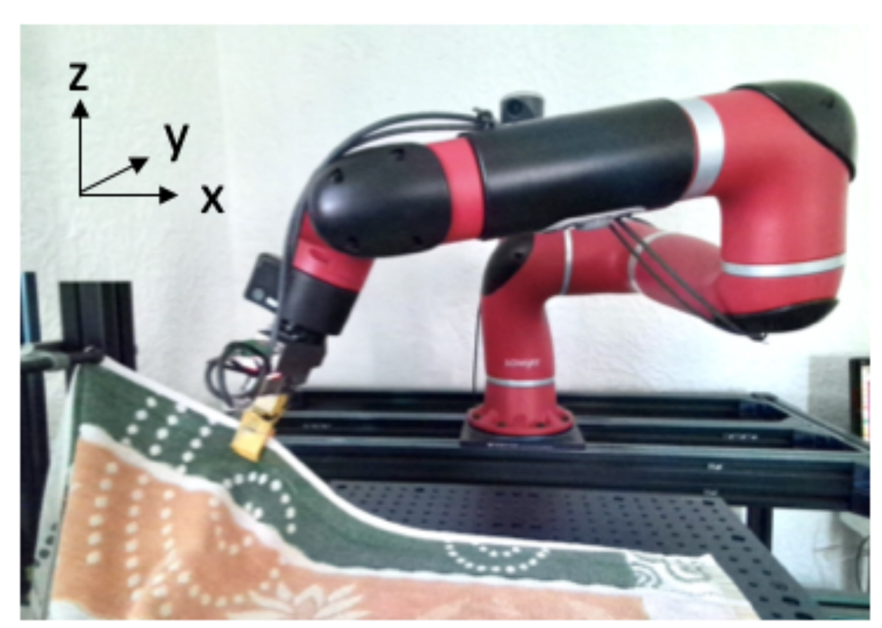
PLAS: Policy in Latent Action Space

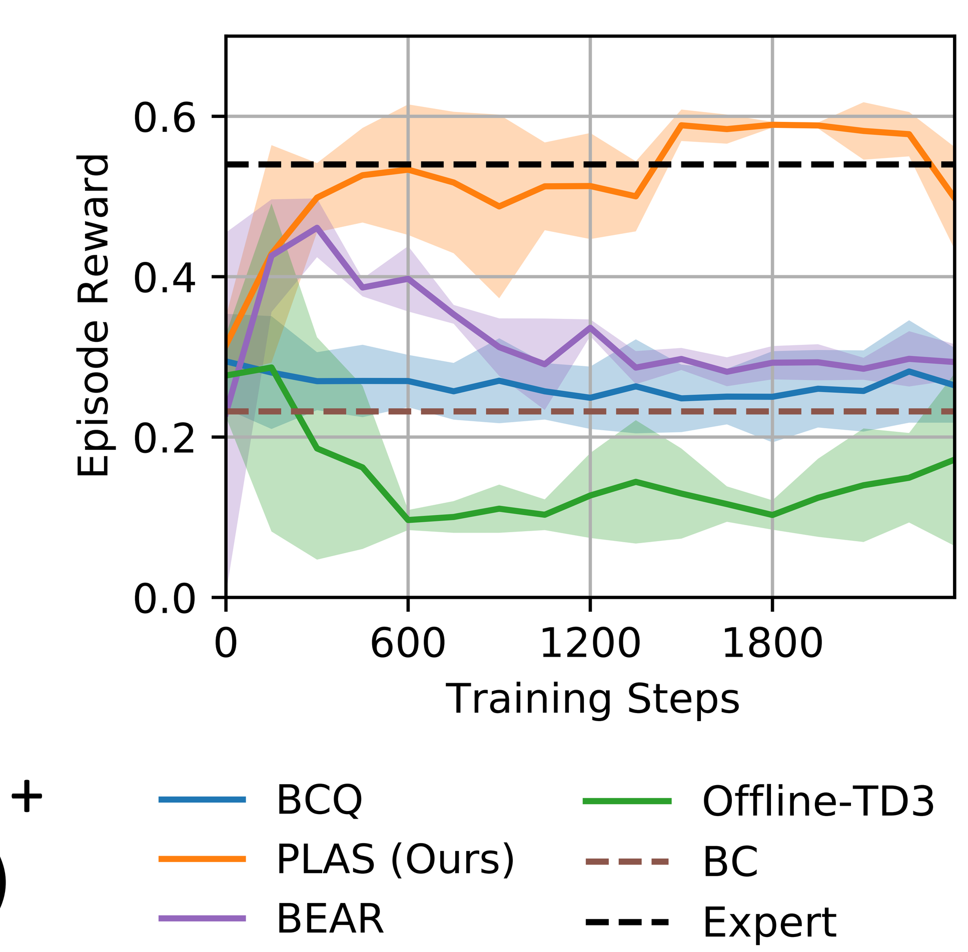




- Model the dataset using Conditional Variational Autoencoder (CVAE)
- Train a policy that outputs in the latent action space of the CVAE and then use the pretrained decoder to output an action in the original action space → Implicitly satisfies the constraint

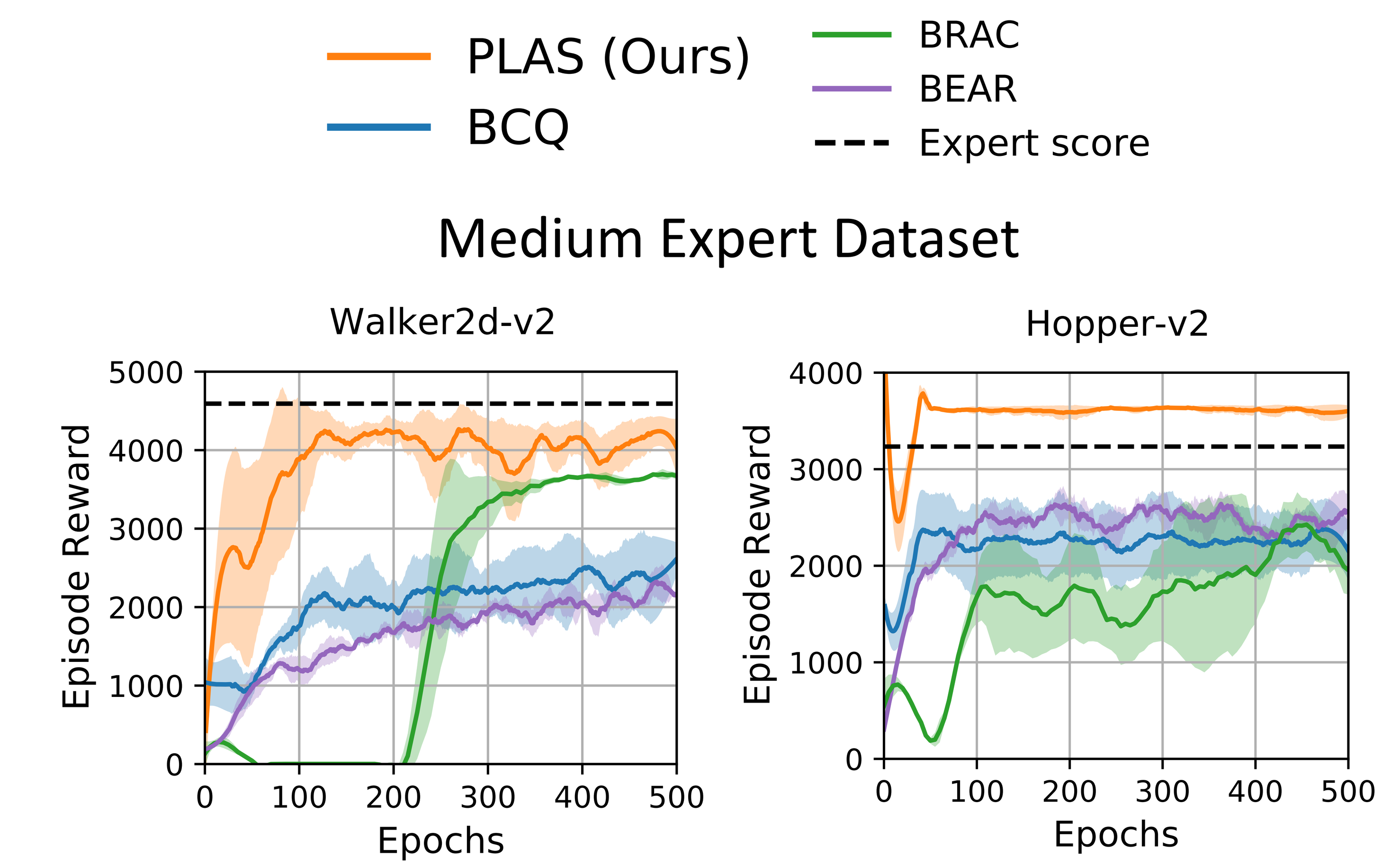
Robot Experiments: Cloth Sliding





Dataset:
replay buffer (7000 timesteps) + expert rollouts (300 timesteps)

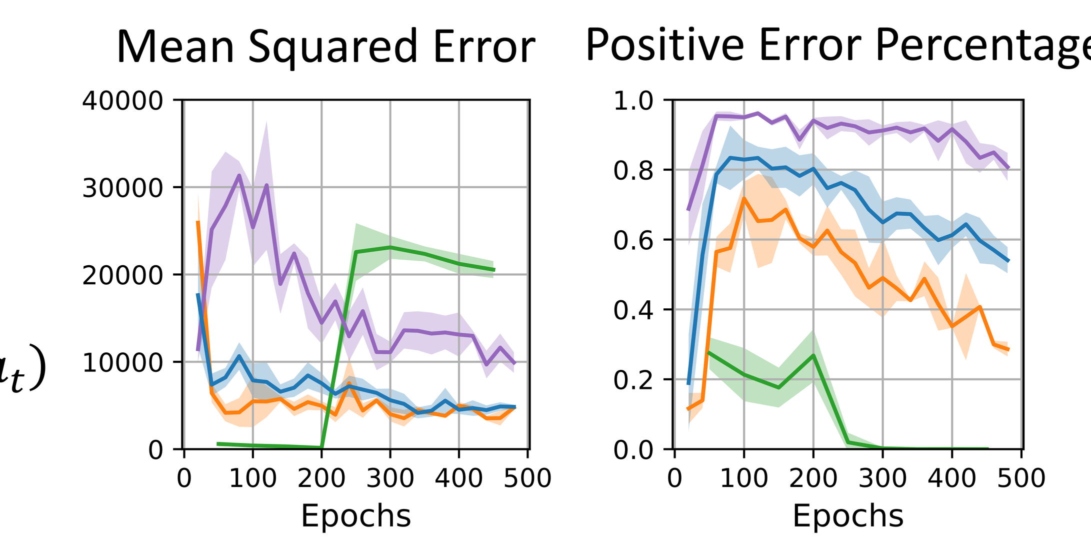
Experiments: D4RL Benchmark



Analysis: Q-function Estimation Error

Walker2d medium expert dataset

Estimation Error: $\hat{Q}(s_t, a_t) - G(s_t, a_t)$



Conclusion

We propose a simple and effective approach to implicitly constrains the policy to be within the support of the dataset without being restricted by the density of the dataset distribution. It achieves competitive performance on both real robot experiments and D4RL benchmark.