Distributionally Adaptive Meta Reinforcement Learning

Anurag Ajay*, Abhishek Gupta*, Dibya Ghosh, Sergey Levine, Pulkit Agrawal (*Equal Contribution)

Meta Reinforcement Learning resilient to task distribution shift via test-time *uncertainty set* adaptation

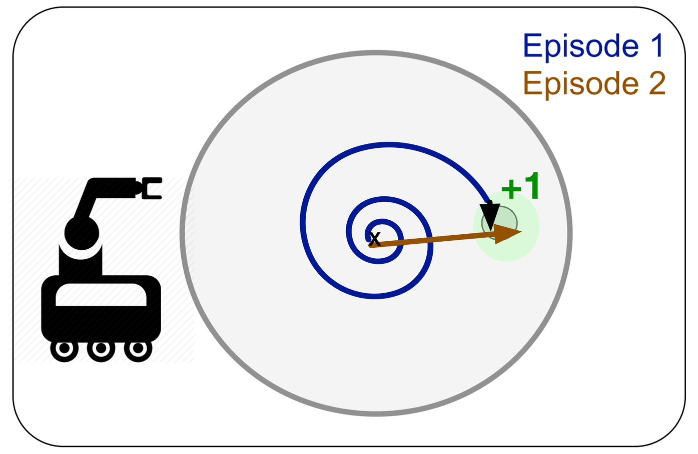
Failure of Meta Reinforcement Learning

$$\min_{\pi_{\text{meta}}} \mathbb{E}_{\mathcal{T}\sim p(\mathcal{T})}\left[\text{Regret}(\pi_{\text{meta}},\mathcal{T})\right]$$

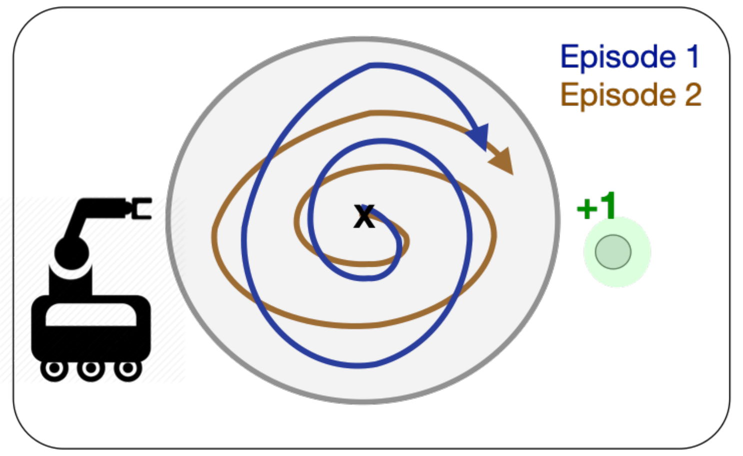
$$\text{Regret}(\pi_{\text{meta}},\mathcal{T}) = J(\pi^*_{\mathcal{T}}) - \mathbb{E}_{a_t^{(i)}\sim\pi_{\text{meta}}(\cdot|h_t^{(i)}),\mathcal{T}}\left[\frac{1}{k}\sum_{i=1}^{k}\sum_{t=1}^{T} r_t^{(i)}\right]$$

π_{meta} solves a new task T using history h of states, actions and rewards from *few* episodes (i.e. k)

But, what if $p_{\text{test}}(\mathcal{T}) \neq p(\mathcal{T})$ due to shift in reward or dynamic distribution?



$p_{train}(T) = p_{test}(T)$



$p_{train}(T) \neq p_{test}(T)$

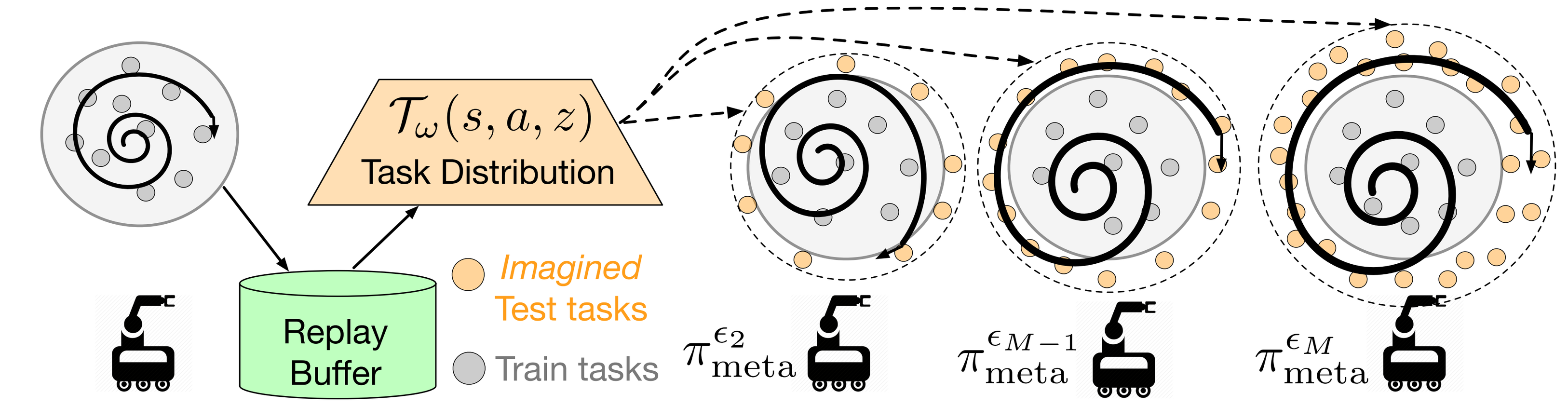
Known Level of Test-Time Distribution Shift

$$\min_{\pi_{\text{meta}}} \mathcal{R}(\pi_{\text{meta}}, p_{\text{train}}(\mathcal{T}), \epsilon) = \min_{\pi_{\text{meta}}} \max_{q(\mathcal{T})} \mathbb{E}_{\mathcal{T}\sim q(\mathcal{T})}\left[\text{Regret}(\pi_{\text{meta}},\mathcal{T})\right]$$

Uncertainty set $\boxed{\text{s.t. } D(p_{\text{train}}(\mathcal{T}) \| q(\mathcal{T})) \leq \epsilon}$

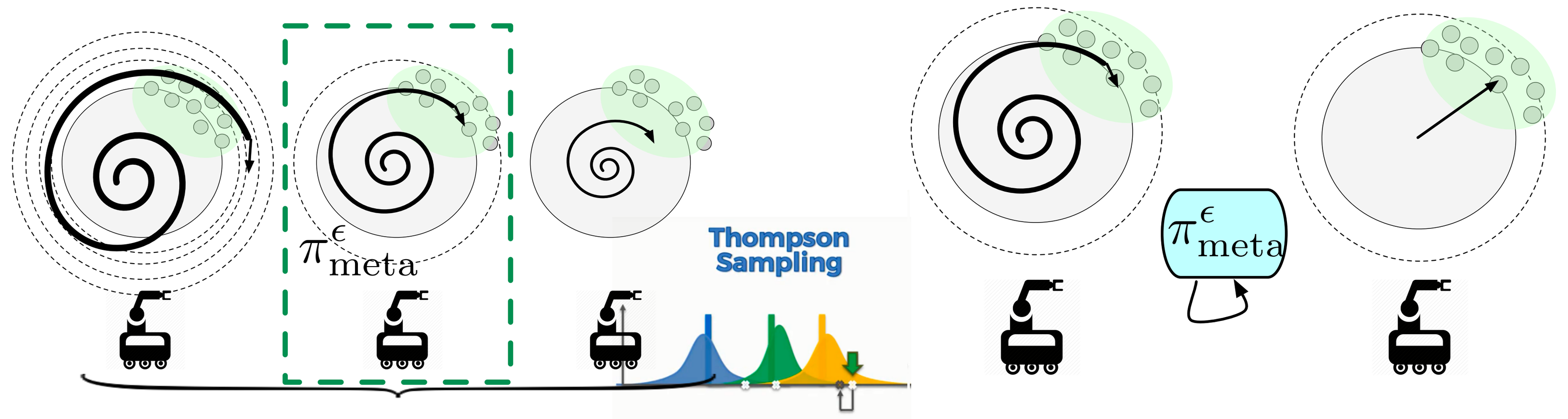
Makes π_{meta} robust to task distributions in the uncertainty set. However, a large uncertainty set can make π_{meta} too conservative.

Arbitrary Level of Distribution Shift



Meta-train on train-task distribution

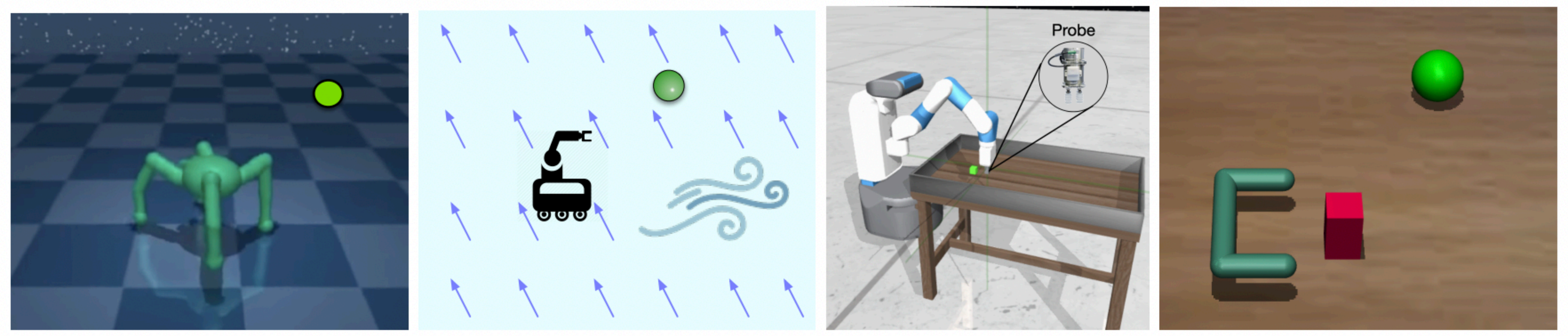
Meta-train on *imagined* test-task distributions



Meta-policy selection during meta-test

Test time Meta-policy adaptation

Experimental Setup



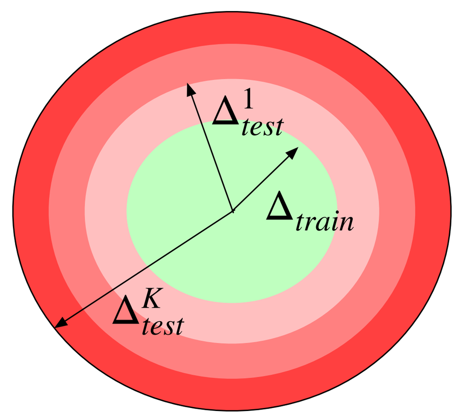
(a) Ant Navigation
Δ: target distance

(b) Wind Navigation
Δ: wind speed

(c) Object Localization
Δ: target distance

(d) Block Push
Δ: target distance

Train/test task parameter Δ distributions



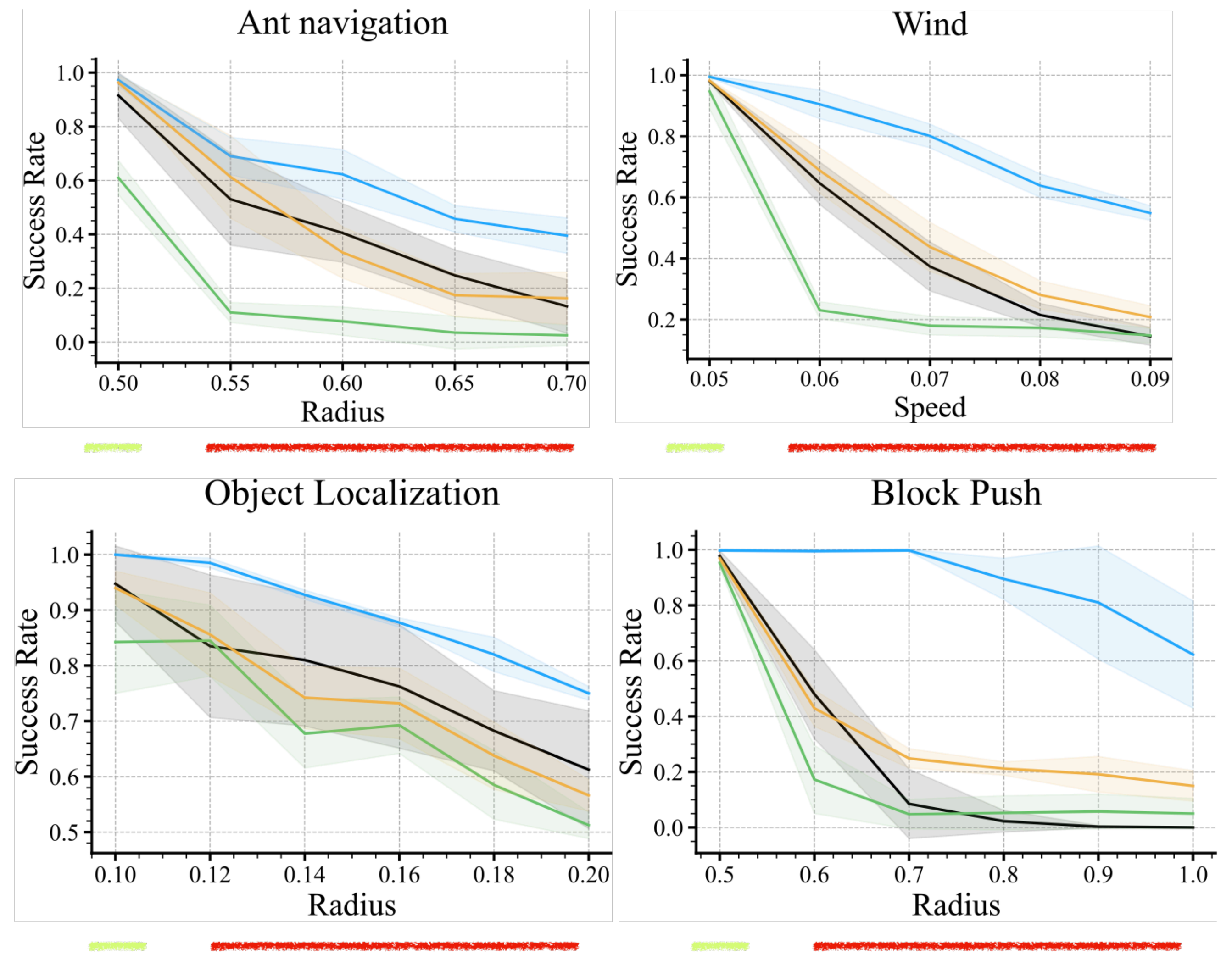
Baselines:
RL2 (Ni et al., 2021)
VariBAD (Zintgraf et al., 2019)
HyperX (Zintgraf et al., 2020)

Resilience to test-task distribution shift

RL2
DiAMetR (Ours)
VariBAD
HyperX

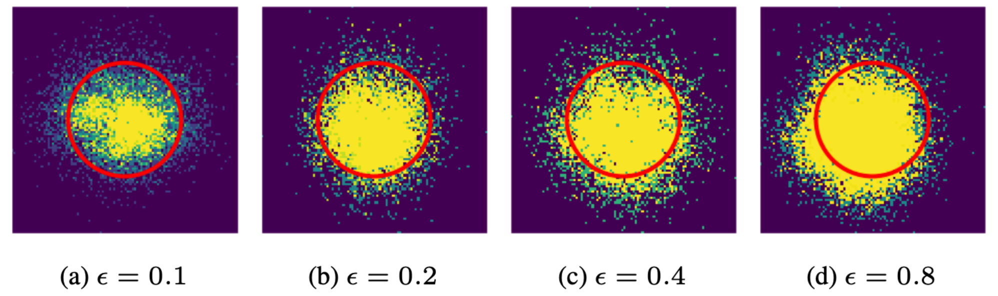
Train task distribution

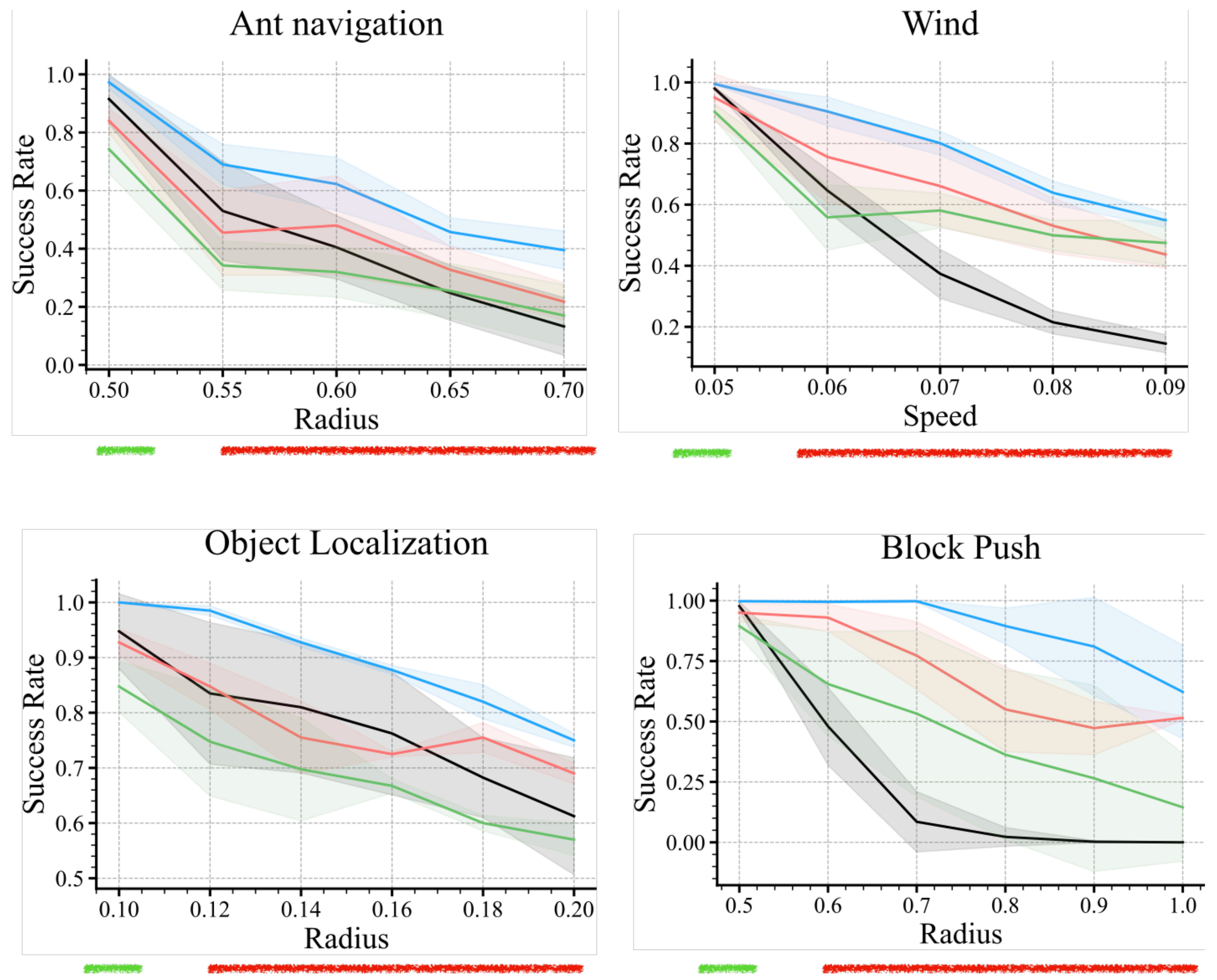
Test task distribution



Importance of Multiple Uncertainty Sets

Imagined target distributions from different uncertainty sets





Adapt infers uncertainty set during test time

Mid selects a mid-sized uncertainty set

Conservative selects the largest uncertainty set

Train task distribution

Test task distribution