Towards Safe, Abstraction-based Online Learning and Synthesis for Unknown Systems

John Jackson

Smead Aerospace Engineering, University of Colorado Boulder

john.m.jackson@colorado.edu

Problem Overview

$$\mathbf{x}(k+1) = \overbrace{f_{\mathbf{u}(k)}(\mathbf{x}(k))}^{\textbf{Known}} + \overbrace{g_{\mathbf{u}(k)}(\mathbf{x}(k))}^{\textbf{Unknown}} + \underbrace{\mathbf{w}(k)}_{\text{sub-Gaussian noise}}$$

$$\mathbf{u}(k) \in \{u_1, \ldots, u_M\}$$

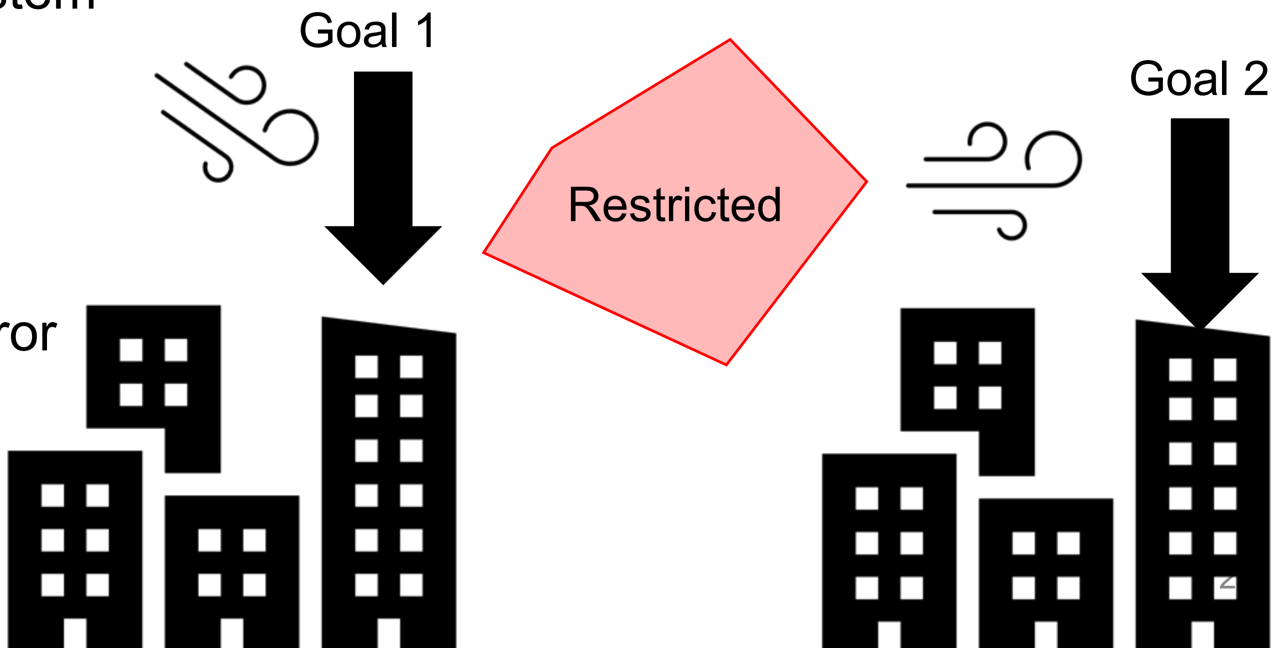
Mission: Go to Goal 1, Go to Goal 2, always avoid restricted airspace.

Given

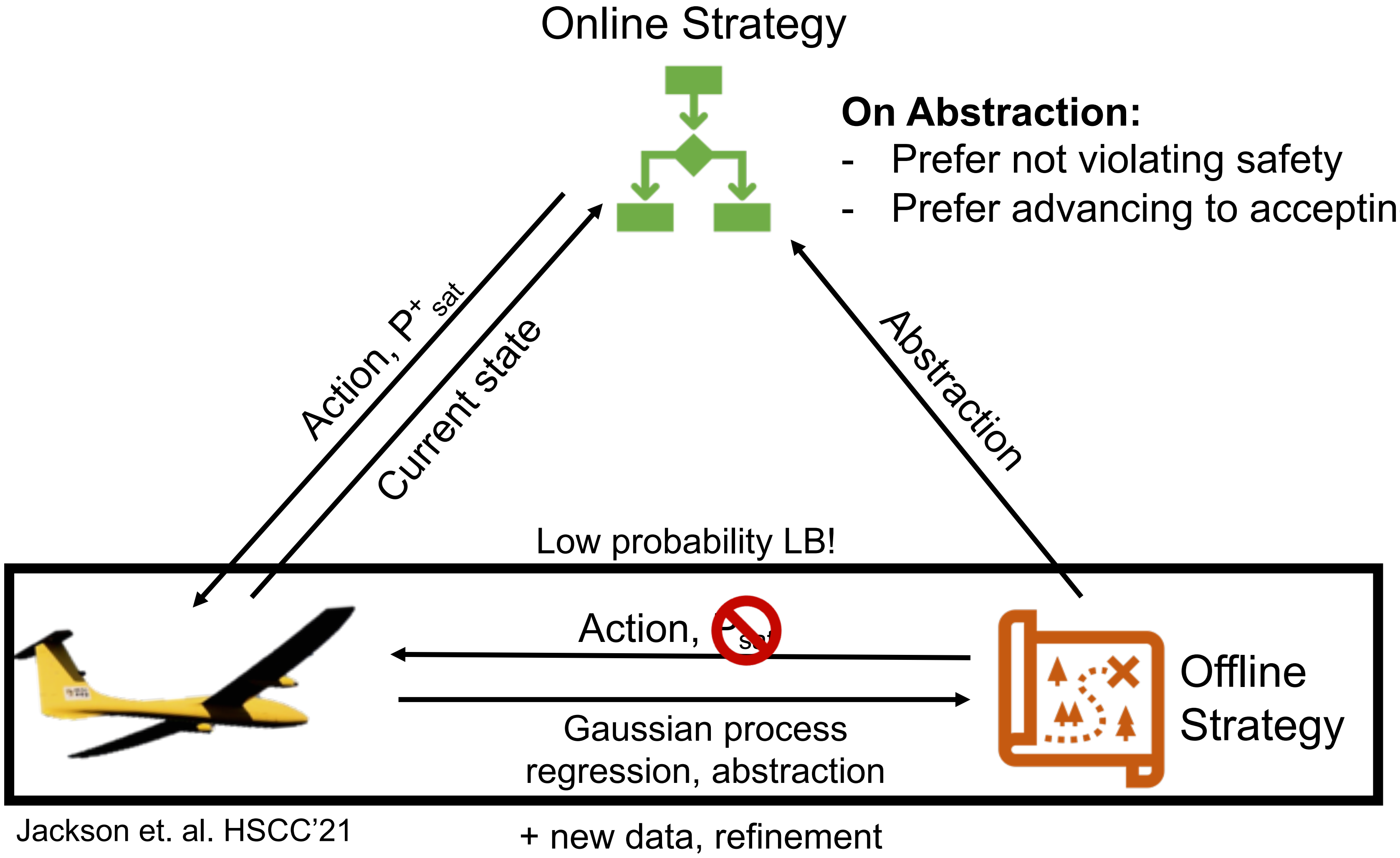
- Input-action-output dataset generated by the system
- Capability to deploy and get more data

Challenges

- *Rigorously* learn and reason about the model error
- *Quickly* synthesize a control strategy
- Provide *guarantees* on mission completion



Offline-Online Approach



On Abstraction:

- Prefer not violating safety
- Prefer advancing to accepting state

Jackson et. al. HSCC'21

Outcome: Improvements in probability of accomplishing the mission compared to the offline strategy.

Thank you!

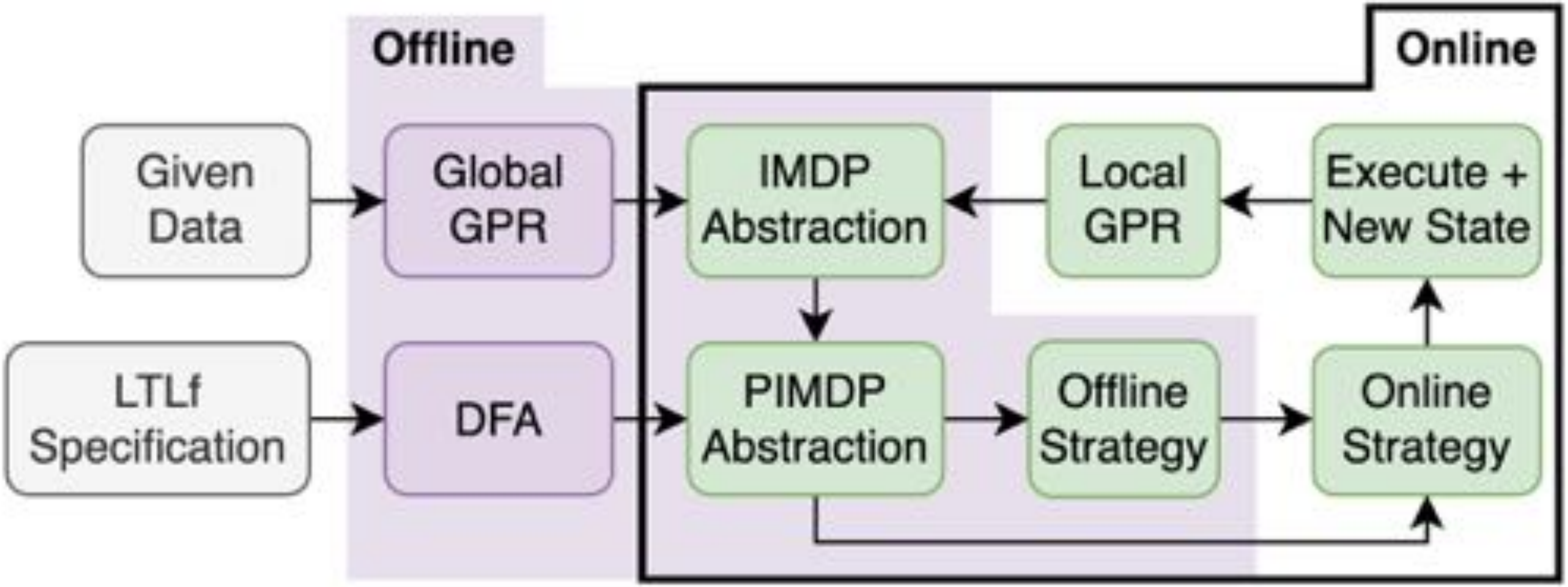
John Jackson

john.m.jackson@colorado.edu

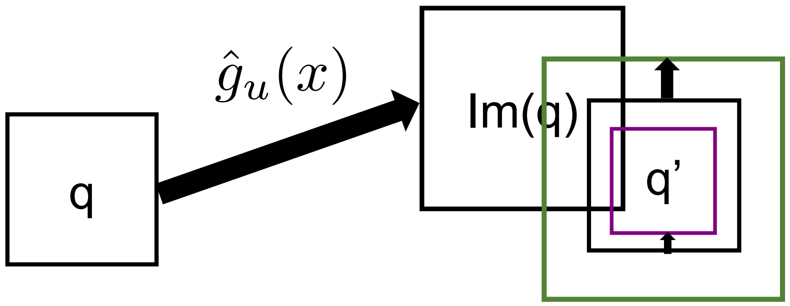
NSF

C-UAS

Framework Overview



Interval MDP Abstraction



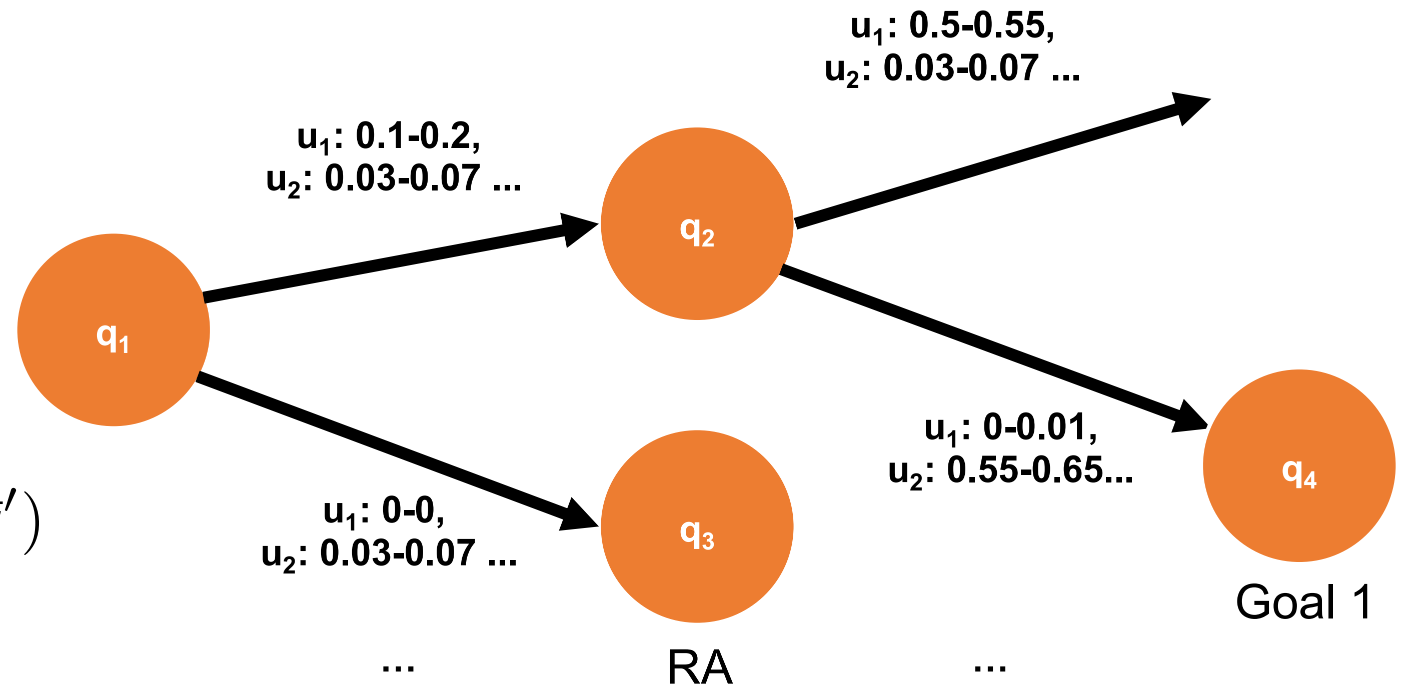
Uncertainties

1. Process Noise (Known) η, P_η
2. Learning Error – RKHS Analysis ϵ, P_ϵ
3. Discretization Error

$$\underline{\Pr}(q \to \underline{q}') \leq \Pr(q \to q') \leq \overline{\Pr}(q \to \overline{q}')$$

$$\overline{\Pr} = \mathbf{1}^{\overline{q}}_{Im(q)} P_\epsilon P_\eta (1 - P_\epsilon)$$

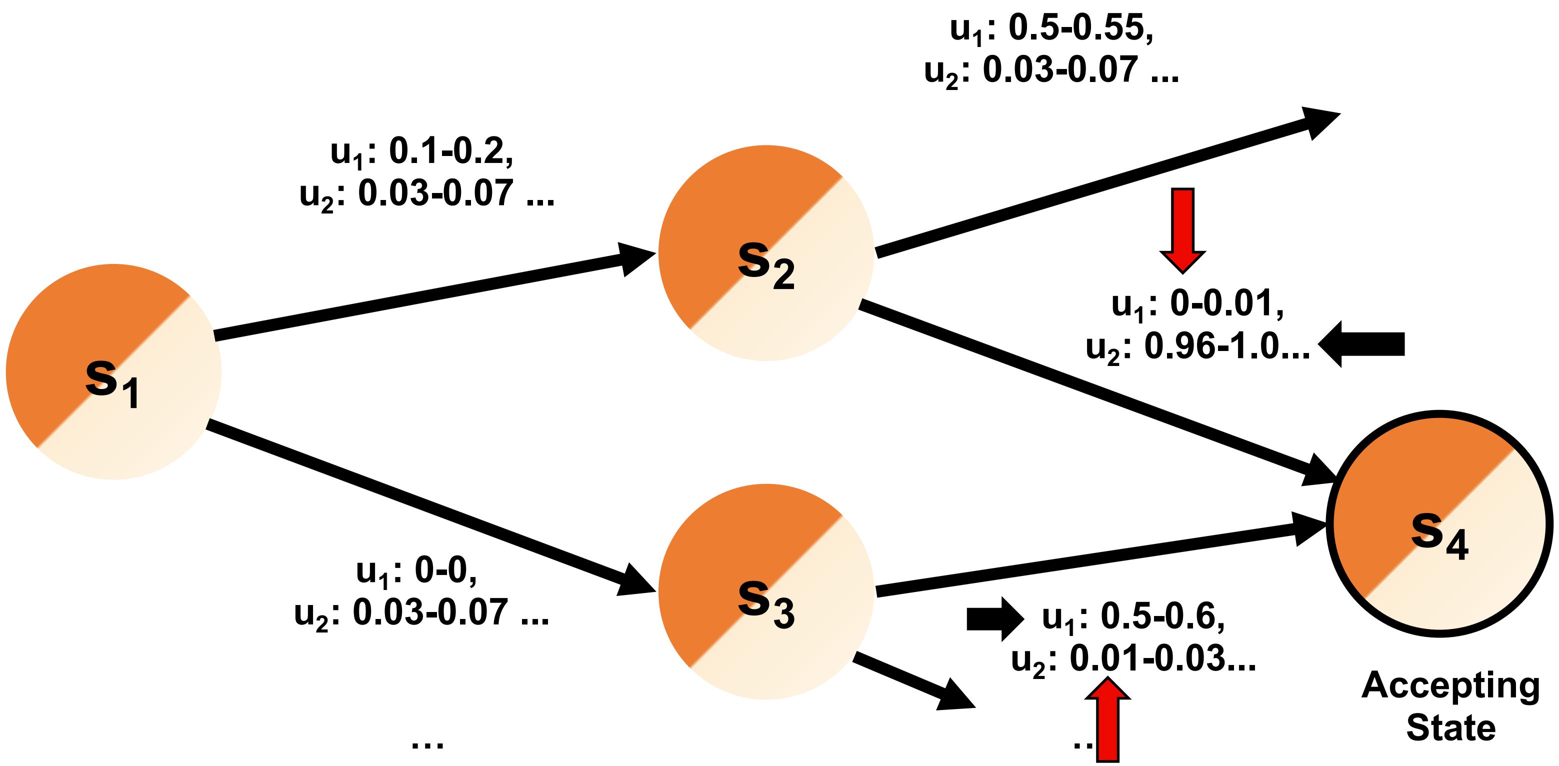
$$\underline{\Pr} = \mathbf{1}^{\underline{q}}_{Im(q)} P_\epsilon P_\eta$$



An IMDP defines a **space** of MDPs using **transition probability intervals** under each action.

Synthesis on an IMDP

- Synthesis on IMDP is a two-player game
- Efficient value iteration over strategies and transition distributions [Lahijanian TAC'15]



Player 1 chooses actions

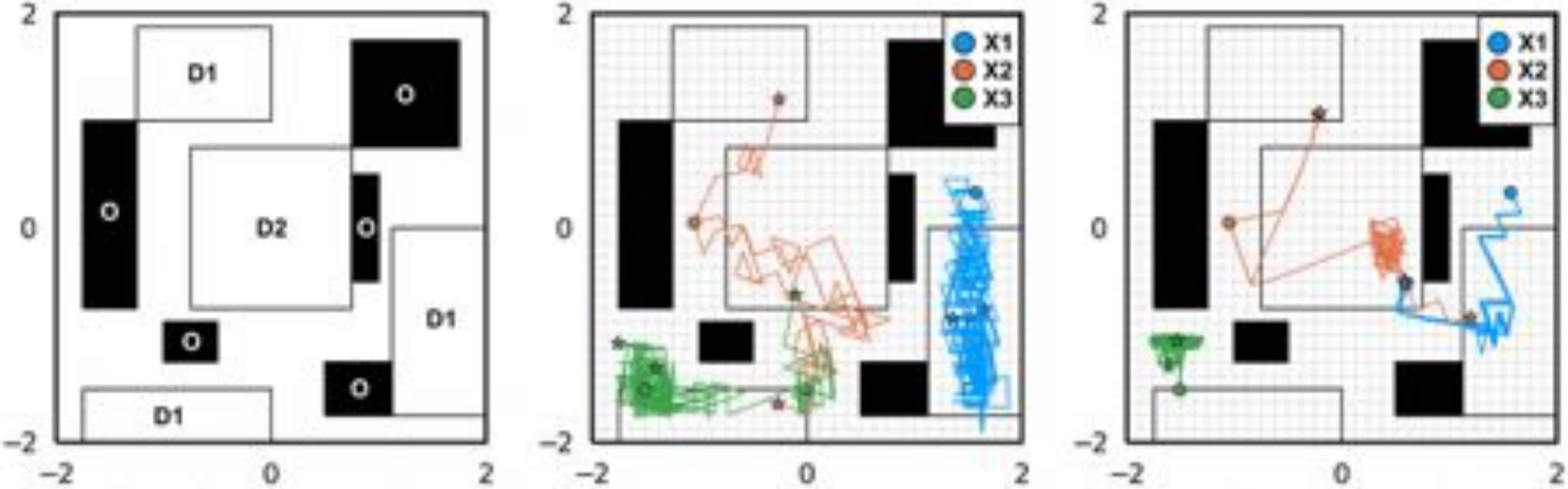
Player 2 chooses transition probability

$$p_L(s_0) = \max_{\pi} \min_{\gamma} p(s_0 \models \phi \mid \pi, \gamma)$$

$$p_U(s_0) = \max_{\pi} \max_{\gamma} p(s_0 \models \phi \mid \pi, \gamma)$$

Optimal Robust Strategy: Choose actions with the highest lower-bound of transition to the accepting state (accounting for player 2's worst actions)

Preliminary Results



(a) Regions of interests and obstacles (b) Sink metric (c) Sink+Prog metric

Preliminary Results

		Global GP (static)			Local GP (updates with $K = 75$ neighbors)					Local GP (static)		
x_0	Tertiary	P_{Violate}	P_{Satisfy}	Time	P_{Violate}	P_{Satisfy}	Time	# $\mathcal{P}$-updates	Val. $\mathcal{P}$-updates	P_{Violate}	P_{Satisfy}	Time
1	Offline	1.0	0.0	–	1.0	0.0	–	–	–	1.0	0.0	–
1	Sink+Prog	0.0	1.0	0.0002	0.0	1.0	0.0135	714.8	1.0	0.0	1.0	0.0003
1	Sink	0.778	0.216	0.0002	**0.326**	**0.322**	0.0323	6258	0.9996	0.77	0.22	0.0004
2	Offline	1.00	0.00	–	1.00	0.00	–	–	–	1.00	0.00	–
2	Sink+Prog	0.0	0.76	(0.0001)	0.0	**0.996**	0.0222	110.8	1.0	0.0	0.808	0.0001
2	Sink	0.0	1.0	0.0002	0.0	1.0	0.0115	521.6	0.9998	0.0	1.0	0.0002
3	Offline	0.348	0.652	–	0.348	0.652	–	–	–	0.348	0.652	–
3	Sink+Prog	0.098	0.0	0.0001	0.102	**0.864**	0.0130	1814	0.9957	0.074	0.0	0.0004
3	Sink	0.0	0.702	0.0001	0.088	0.6140	0.0120	5890	0.9970	0.00	0.704	0.0004