Safe Exploration in Reinforcement Learning

Matteo Turchetta, Felix Berkenkamp, Andreas Krause

ETH zürich

Learning & Adaptive Systems

Safety in finite MDPs

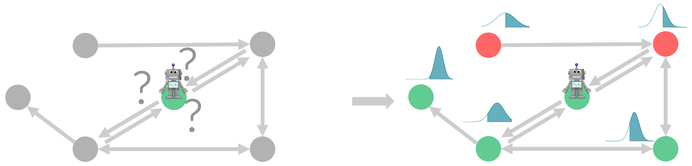
Problem

Learning agents accept short term loss for long term reward. This is unacceptable in safety critical applications.

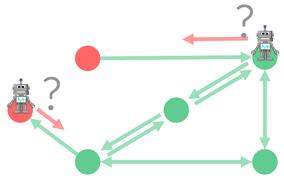
Goal: explore the environment without failure.

Formulation

The safety constraint over states, r(s)≥h, is unknown.
Gaussian process model to infer safety of unvisited states:



Need to consider safety in the long-term, not only per step:



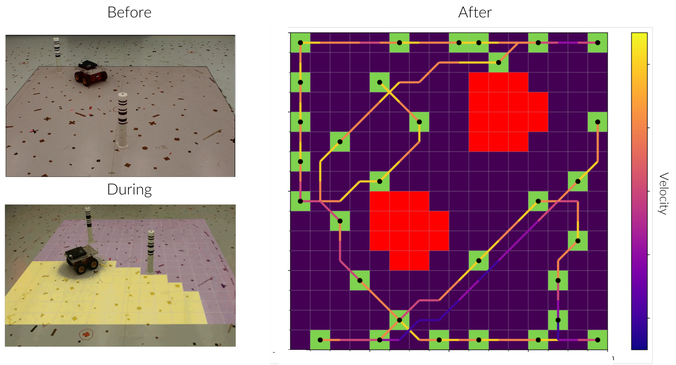
Intuition

1. Keep track of states that satisfy the safety constraint
2. Compute safe subset based on reachability
3. Ensure safety by never leaving the safe set
4. Actively gather data to expand the safe set

Theory

Safety: The safety constraint is not violated during exploration
Completeness: The largest safe reachable region is explored

Experiments

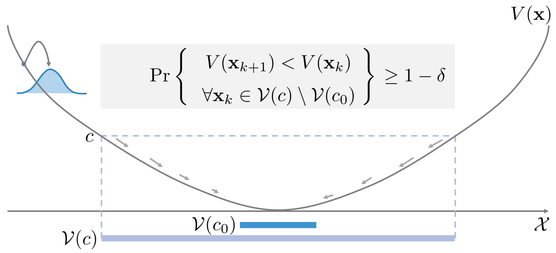


Safety through stability

Problem

Model $\mathbf{x}_{k+1} = \underbrace{f(\mathbf{x}_k, \mathbf{u}_k)}_{a\ priori\ \text{model}} + \underbrace{g(\mathbf{x}_k, \mathbf{u}_k)}_{\text{unknown model}}$ Requirement $\mathbf{x}_k \in \text{ROA}, \forall\, k$

Region of attraction



Policy update

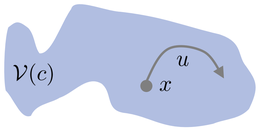
Expected performance + worst-case stability

$$\min_{\theta} \int_{\mathcal{X}} r(\mathbf{x}, \pi_\theta(\mathbf{x}_k)) + \gamma\, J_{\pi_\theta}(\mathbf{x}_{k+1})\, \partial \mathbf{x}_k$$

$$\text{s.t.} \quad \Pr\{\, V(\mathbf{x}') < V(\mathbf{x}) \,\} \geq 1-\delta$$

Exploration

Need to learn model, but cannot leave the region of attraction



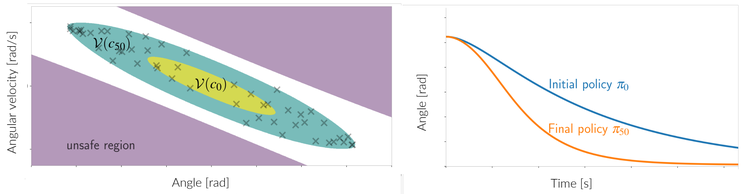
Drive the system to informative, safe states

$$(\mathbf{x}_n, \mathbf{u}_n) = \underset{\mathbf{x},\mathbf{u}}{\text{argmax}}\ \sigma_n(\mathbf{x}, \mathbf{u})$$
$$\text{s.t.} \quad \Pr\{f(\mathbf{x},\mathbf{u}) + g(\mathbf{x},\mathbf{u}) \in \mathcal{V}(c)\} \geq 1-\delta$$

Theory

Stability: Returned policy is stable
Safety: The agent never leaves the region of attraction
Completeness: The largest safe reachable region is explored

Experiments

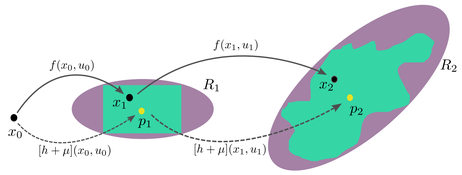


Safety via MPC

Problem

Model $\mathbf{x}_{k+1} = \underbrace{f(\mathbf{x}_k, \mathbf{u}_k)}_{a\ priori\ \text{model}} + \underbrace{g(\mathbf{x}_k, \mathbf{u}_k)}_{\text{unknown model}}$ Requirement $(\mathbf{x}_k, \mathbf{u}_k) \in \mathcal{X} \times \mathcal{U}, \forall\, k$

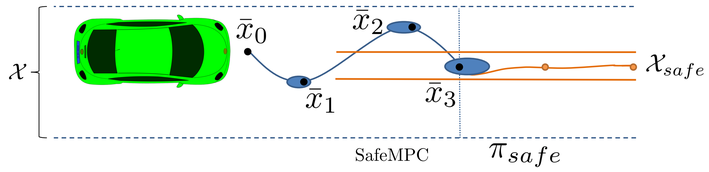
Uncertainty propagation



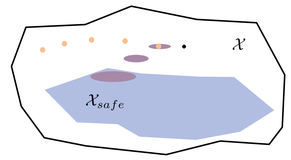
MPC

$$\begin{aligned} \underset{\pi_0,..,\pi_{T-1}}{\text{minimize}} \quad & J_t(R_0, .., R_T) \\ \text{subject to} \quad & R_{t+1} = \tilde{m}(R_t, \pi_t),\ t = 0, .., T-1 \\ & R_t \subset \mathcal{X},\ t = 1, .., T-1 \\ & \pi_t(R_t) \subset \mathcal{U}, t = 0, .., T-1 \\ & R_T \subset \mathcal{X}_{\text{safe}}, \end{aligned}$$

Intuition



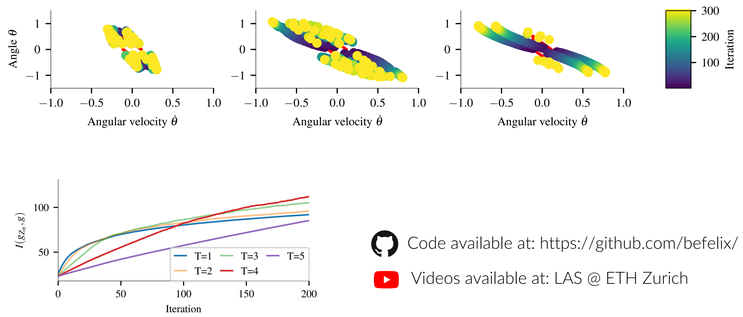
Including performance



Theory

Safety: The polytopic constraints are never violated

Experiments



Code available at: https://github.com/befelix/
Videos available at: LAS @ ETH Zurich