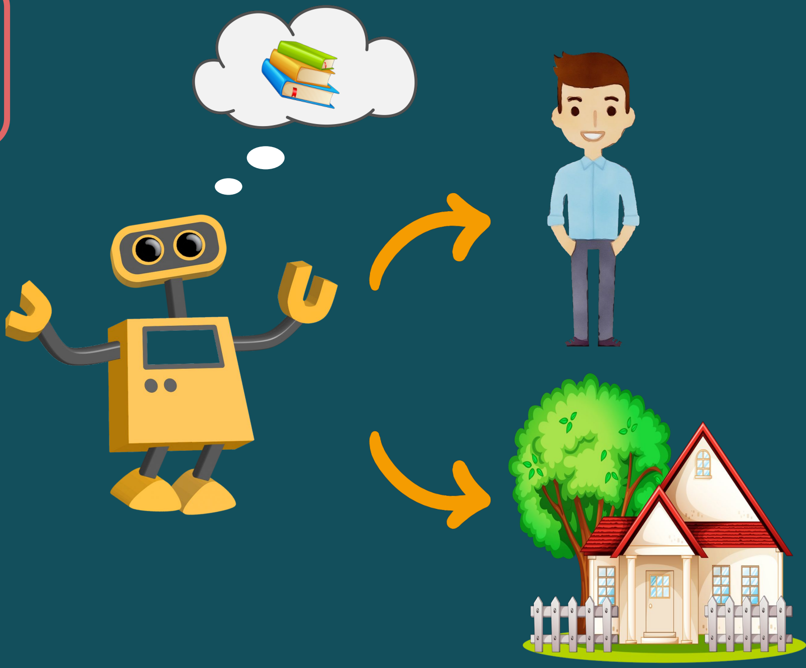
Robot Reinforcement Learning on the Constraint Manifold

Puze Liu[1], Davide Tateo[1], Haitham Bou-Ammar[2,3], Jan Peters[1]
[1]TU Darmstadt, [2]Huawei R&D UK, [3]University College London

TECHNISCHE UNIVERSITÄT DARMSTADT

Motivation

How can we ensure safety during the reinforcement learning process?



Safe Exploration

Problem Formulation

$$\max_\theta \ \mathbb{E}_{s_t,a_t}\left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)\right],$$

$$\text{s. t.} \quad f(q_t) = 0, \quad g(q_t) \leq 0$$

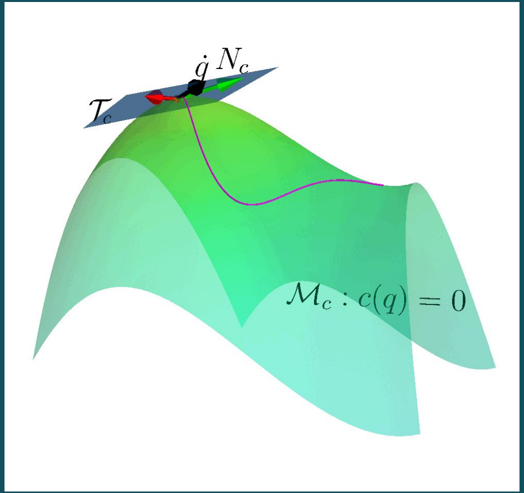
$$s_t = [q_t \ x_t]^{\mathrm{T}}$$

Idea

- Construct the constraint manifold

$$\mathcal{M}_c : c(q) = 0$$

- Determine the bases N_c of the tangent space $\mathcal{T}_c$
- Sample state velocity in the tangent space



ATACOM

Acting on the Tangent Space of the Constraint Manifold

Constraint Manifold

$$f(q_t) = 0 \qquad g(q_t) \leq 0 \quad \longrightarrow \quad c(q_t, \mu_t) = \begin{bmatrix} f(q_t) \\ g(q_t) + \frac{1}{2}\mu_t^2 \end{bmatrix} = 0$$

Tagent Space Velocity

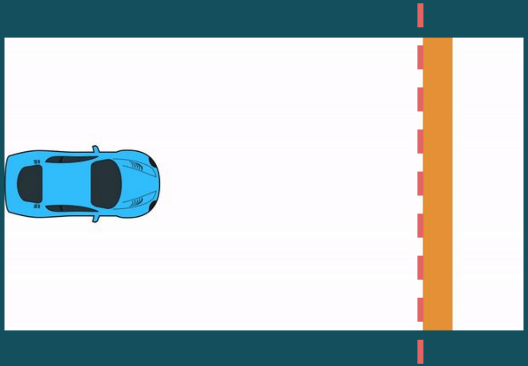
$$\dot{c}(q_t, \mu_t) = J_c \begin{bmatrix} \dot{q}_t \\ \dot{\mu}_t \end{bmatrix}, \qquad N_c = \mathrm{Null}(J_c),$$

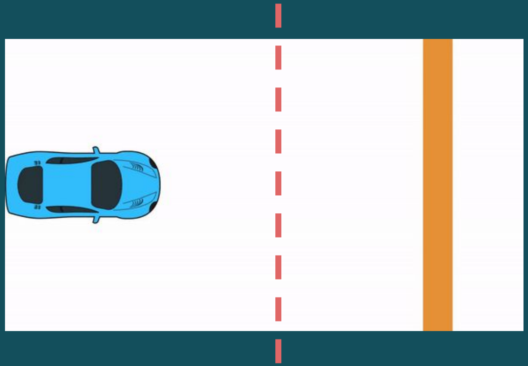
$$\begin{bmatrix} \dot{q}_t^{\mathcal{T}} \\ \dot{\mu}_t^{\mathcal{T}} \end{bmatrix} = N_c \alpha \,, \quad \dot{c}(q_t, \mu_t) = J_c N_c \alpha = 0$$

ATACOM

Acting on the Tangent Space of the Constraint Manifold

Viability Constraint



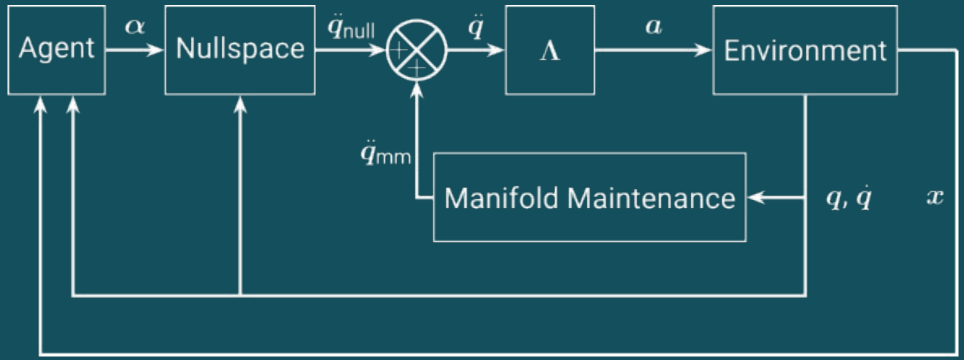


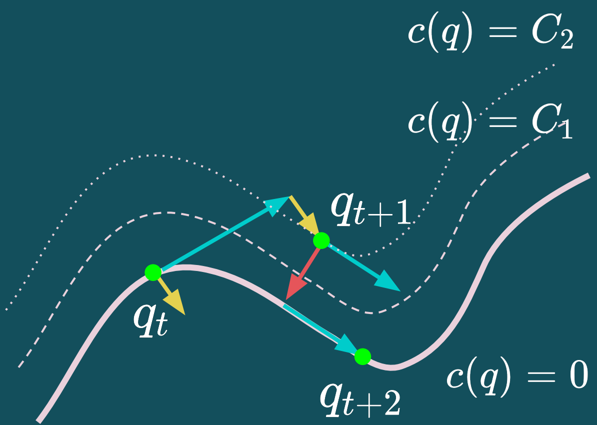
$$c(q_t, \dot{q}_t, \mu_t) = \begin{bmatrix} f(q_t) + K_f \dot{f}(q_t, \dot{q}_t) \\ g(q_t) + K_g \dot{g}(q_t, \dot{q}_t) + \frac{1}{2}\mu^2 \end{bmatrix} = 0$$

Overall Control Acceleration

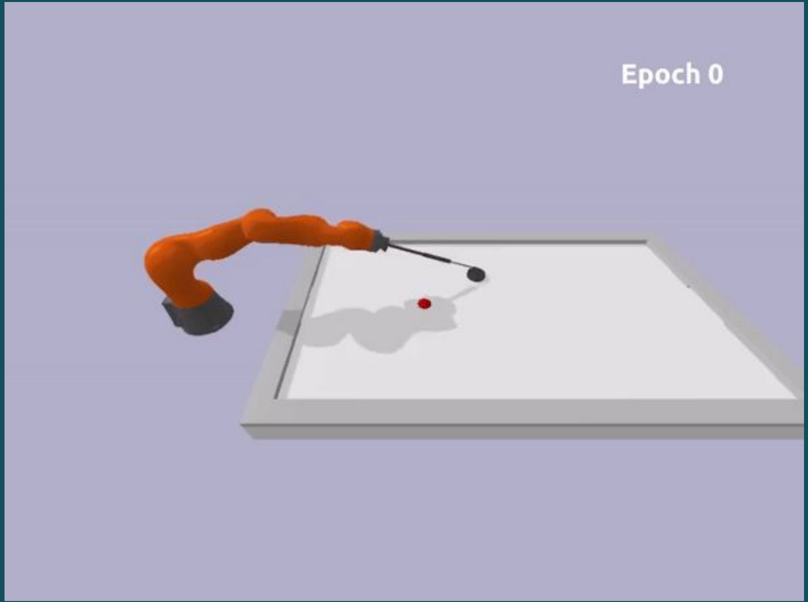
$$\begin{bmatrix} \ddot{q}_t \\ \dot{\mu}_t \end{bmatrix} = \underbrace{N_c(q_t, \mu_t)\alpha_t}_{\text{Tangent Space Action}} \underbrace{- J_c^{\dagger}(q_t, \mu_t)\psi(q_t, \dot{q}_t)}_{\text{Curvature Correction}}$$

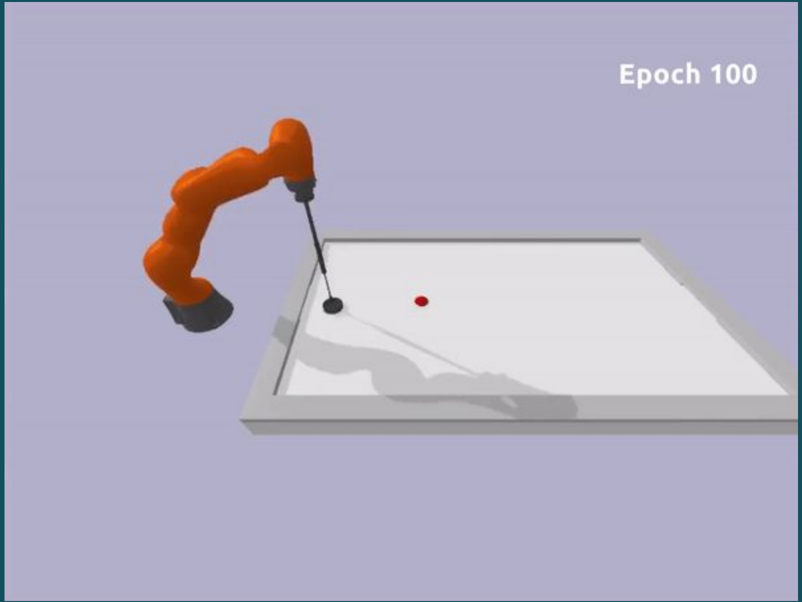
Control Action $a_t = \Lambda(\ddot{q}_t)$





Experiment



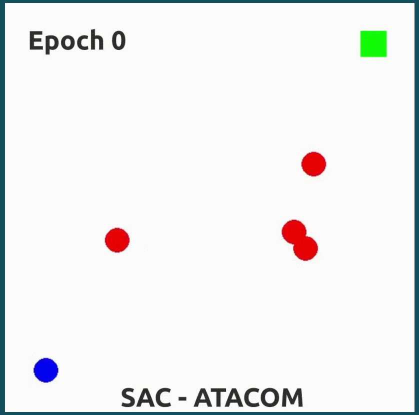


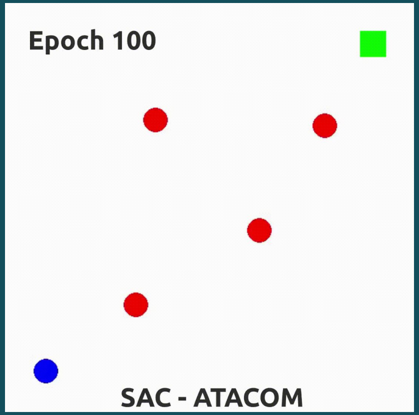
Partially Controllable Constraints

$$\max_\theta \; \mathbb{E}_{s_t,a_t}\left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)\right], \quad \text{s.t.} \quad f(q_t) = 0, \quad g(q_t) \le 0$$

→

$$\max_\theta \; \mathbb{E}_{s_t,a_t}\left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)\right], \quad \text{s.t.} \quad f(q_t, x_t) = 0, \quad g(q_t, x_t) \le 0$$





Conclusions

Our Method:

Acts on the tangent space of the constraint manifold

Handles equality constraint

Copes with partially controllable scenario