Safe Reinforcement Learning in Robotics with Bayesian Models

Felix Berkenkamp, Matteo Turchetta, Angela P. Schoellig, Andreas Krause

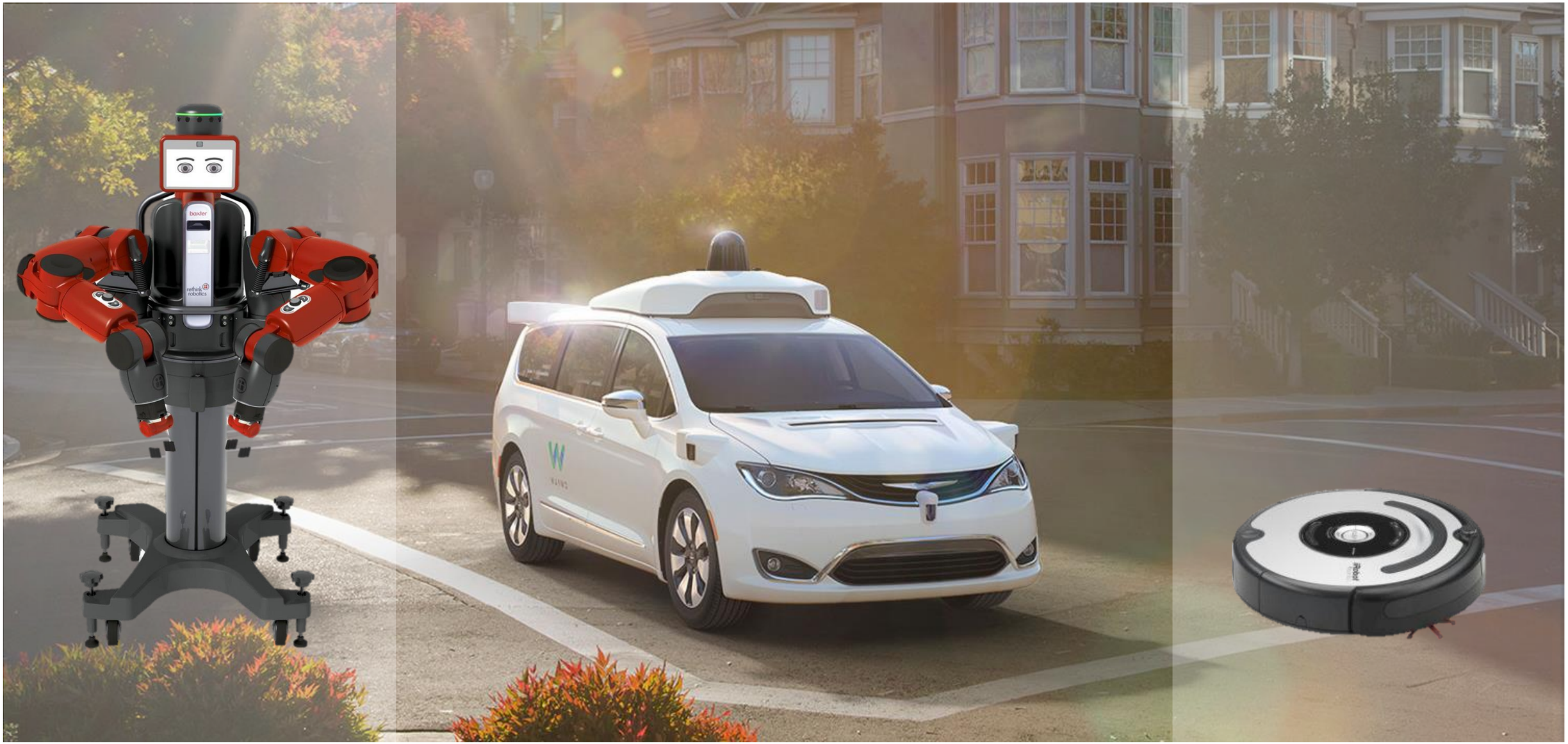
@Workshop on Reliable AI, October 2017

ETH zürich

DYNAMIC SYSTEMS LAB

Institute for Aerospace Studies
UNIVERSITY OF TORONTO

A new era of autonomy



Images: rethink robotics, Waymob, iRobot

Reinforcement learning

Policy

$\pi_\theta(\mathbf{x})$

Exploration

Policy update

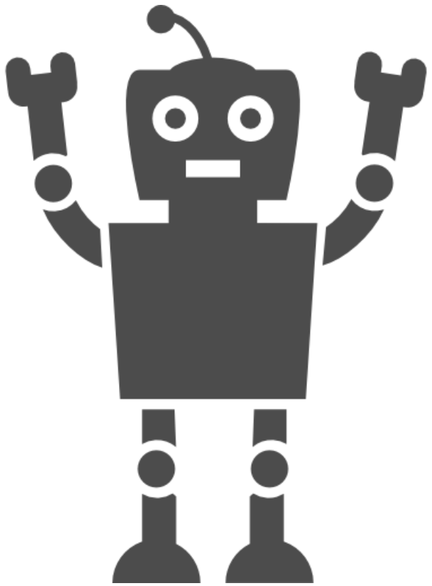


Image: Plainicon, https://flaticon.com

Dangers of autonomous learning

Safety despite uncertainty

Safe exploration

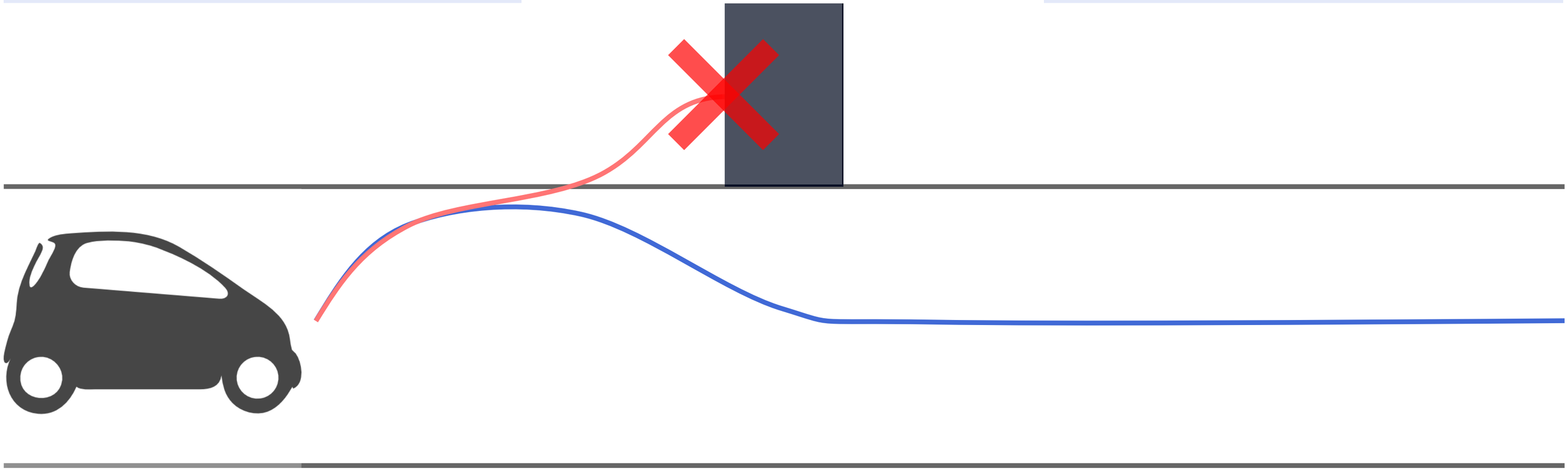


Image: Freepik, https://flaticon.com

Safe reinforcement learning

Bayesian models for safety

Model-free

Model-based

Policy

$\pi_\theta(\mathbf{x})$

Exploration

Policy update

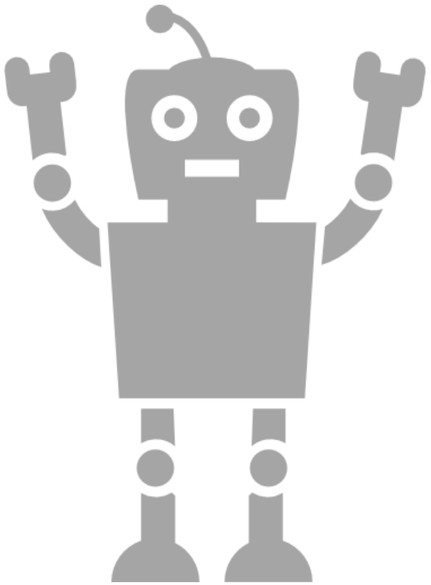
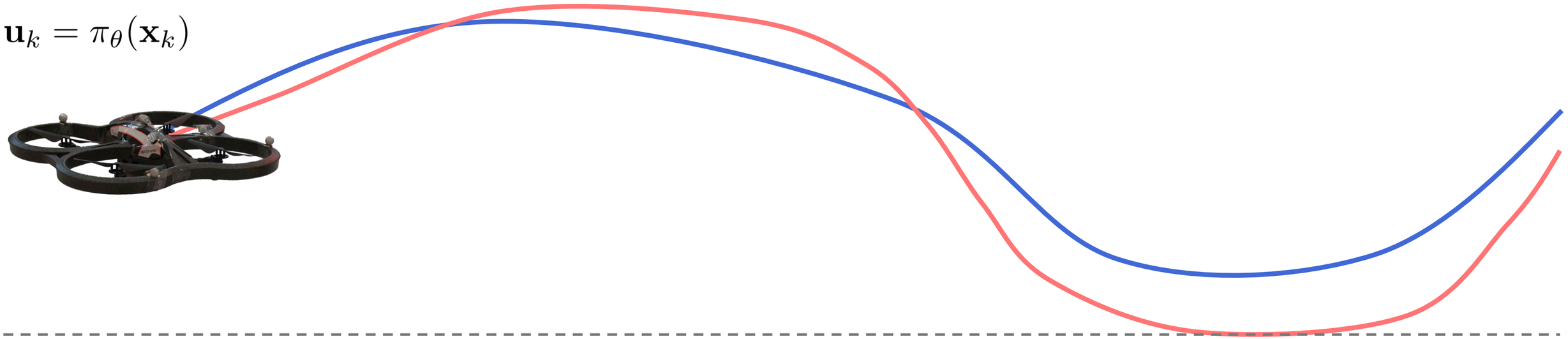


Image: Plainicon, https://flaticon.com

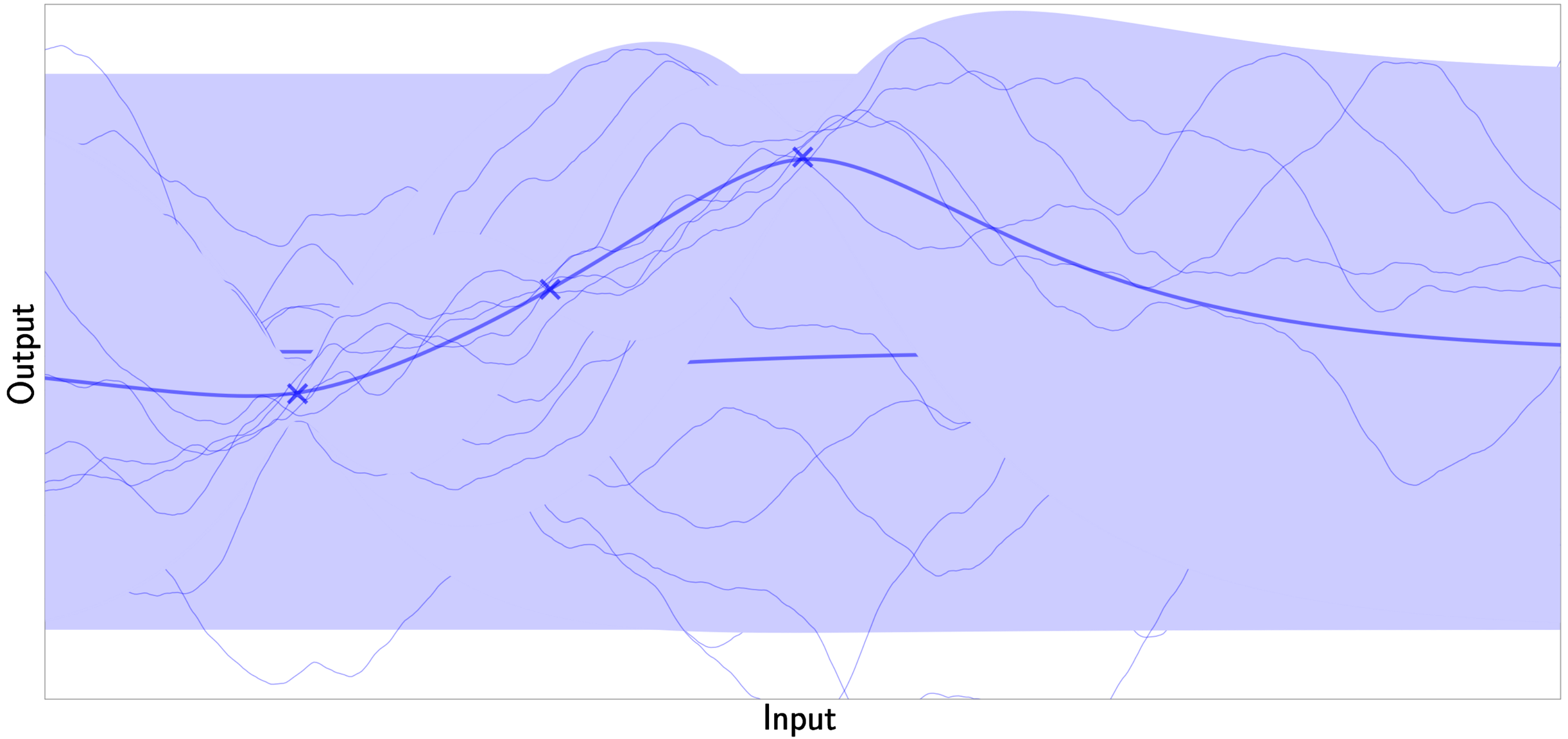
Model-free reinforcement learning

$$\mathbf{u}_k = \pi_\theta(\mathbf{x}_k)$$

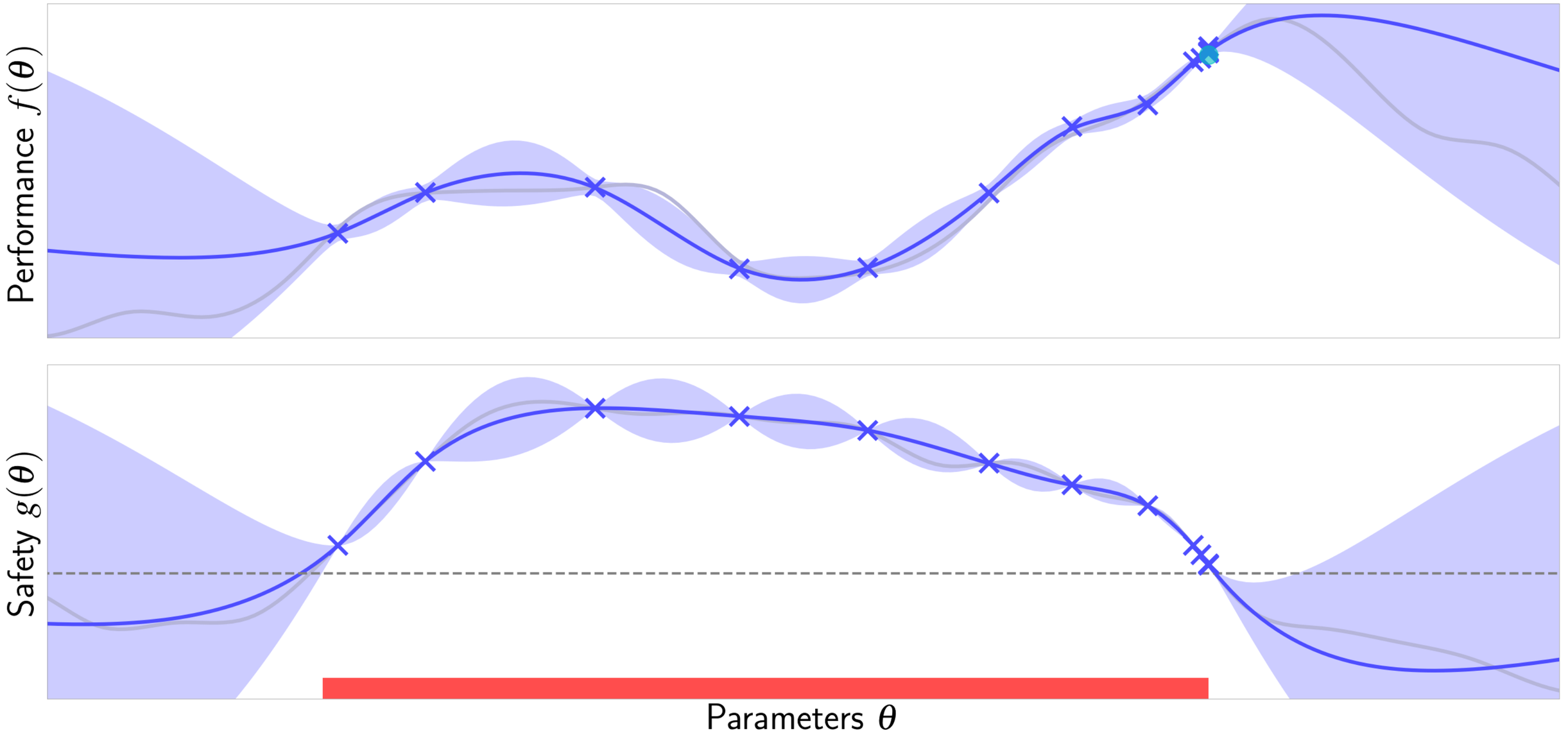


Tracking performance	$\max_\theta \ f(\theta)$	**Few experiments**
Safety constraint	$g(\theta) \geq 0$	**Safety for all experiments**

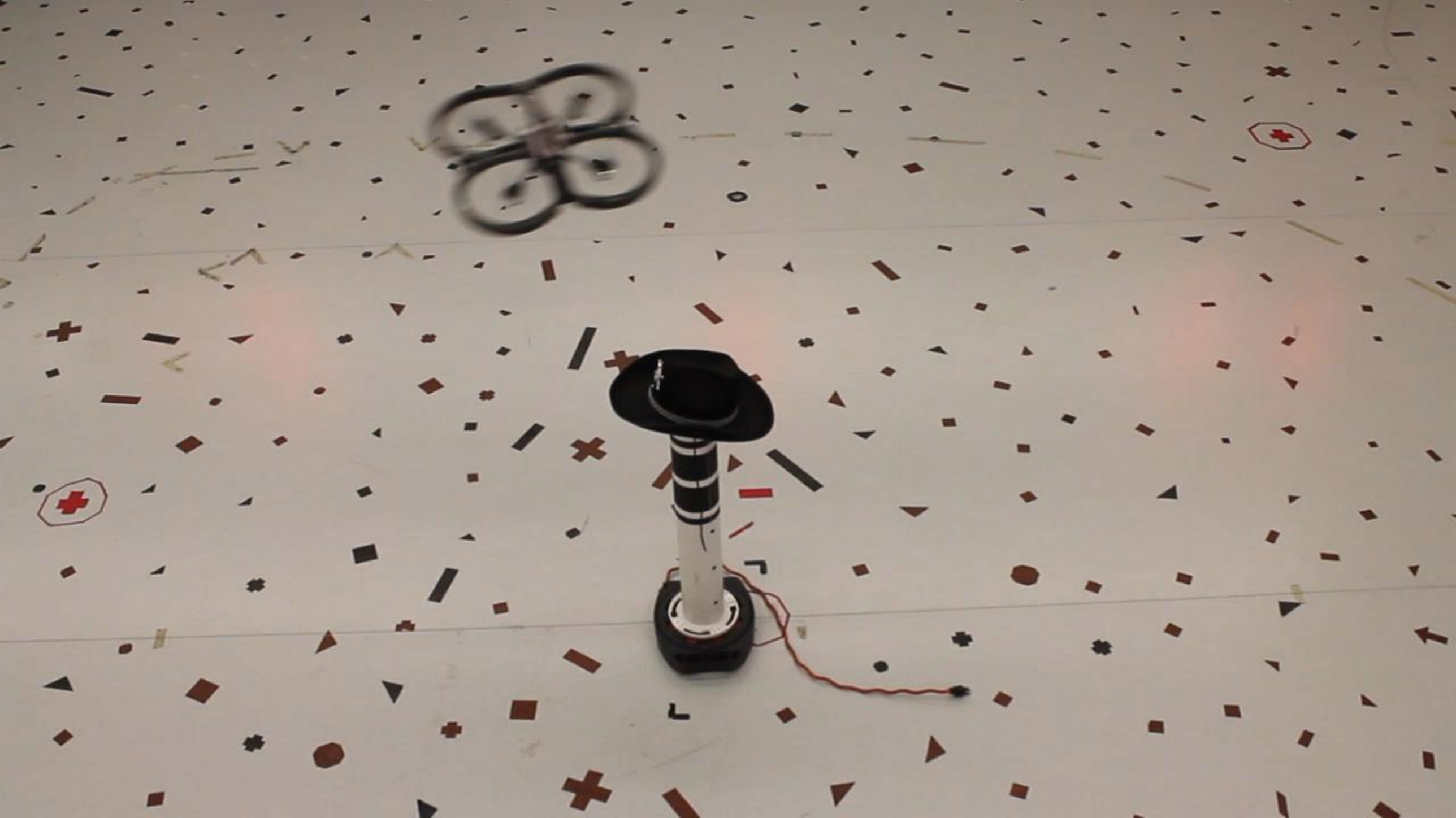
Gaussian process



Constrained Bayesian optimization



Video available at
http://tiny.cc/icra16_video



Safe reinforcement learning

Bayesian models for safety

Model-free

Model-based

Policy

$\pi_\theta(\mathbf{x})$

Exploration

Policy update

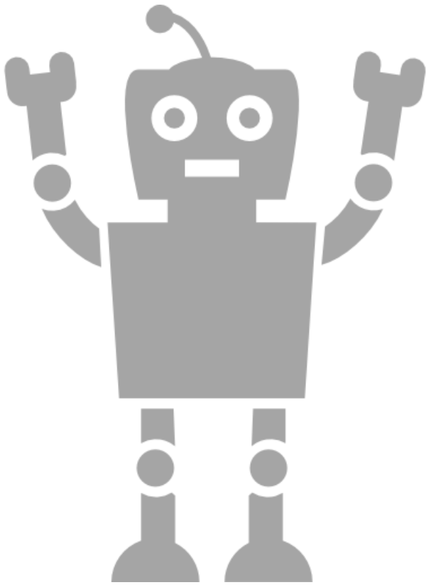


Image: Plainicon, https://flaticon.com

Model-based reinforcement learning

Modelling

$\mathbf{u}_k$ Model $\mathbf{x}_{k+1}$

Learning

Control Theory

Implement

$\mathbf{u}_k = \pi_\theta(\mathbf{x}_k)$

$\mathbf{u}_k = \pi_\theta(\mathbf{x}_k)$

Approximate dynamic programming

Dynamics

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k)$$

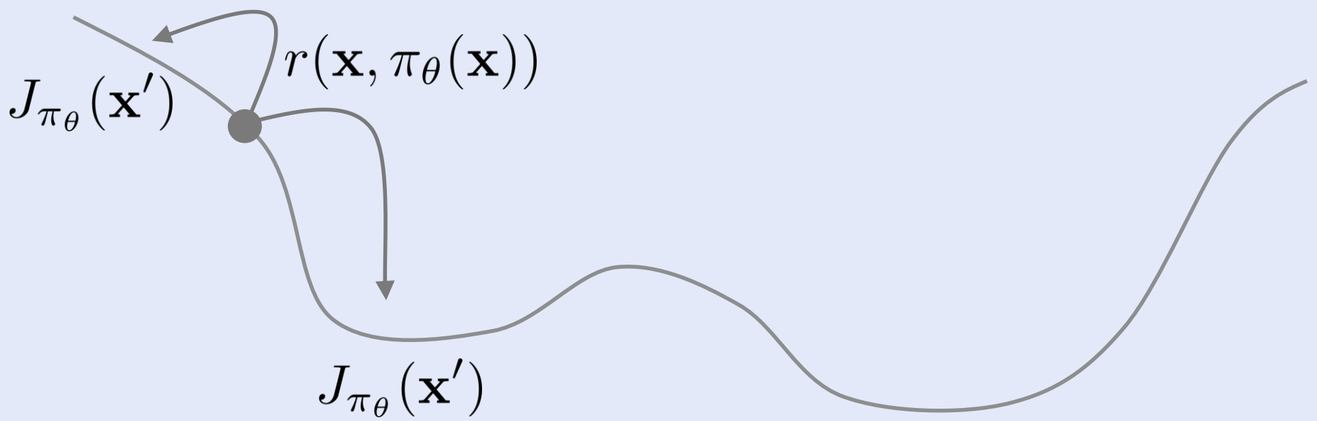
Expected cost

$$J_{\pi_\theta}(\mathbf{x}) = \mathbb{E}\left[\sum_k \gamma^k r(\mathbf{x}_k, \pi_\theta(\mathbf{x}_k))\right]$$

Policy update

$$\mathbf{u}_k = \pi_\theta(\mathbf{x}_k)$$

$$\min_\theta \int_{\mathcal{X}} r(\mathbf{x}, \pi_\theta(\mathbf{x})) + \gamma J_{\pi_\theta}(\mathbf{x}')$$

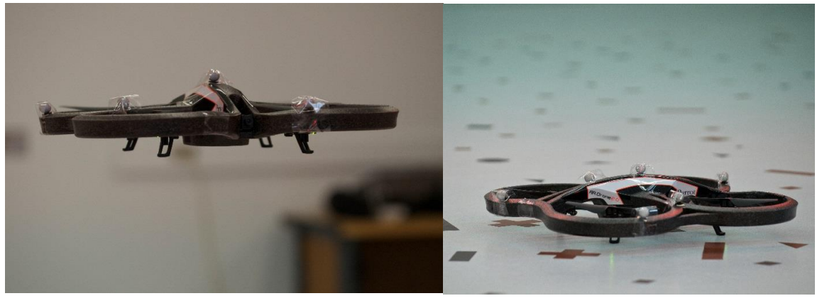


Uncertain dynamics

Dynamics model

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k)$$

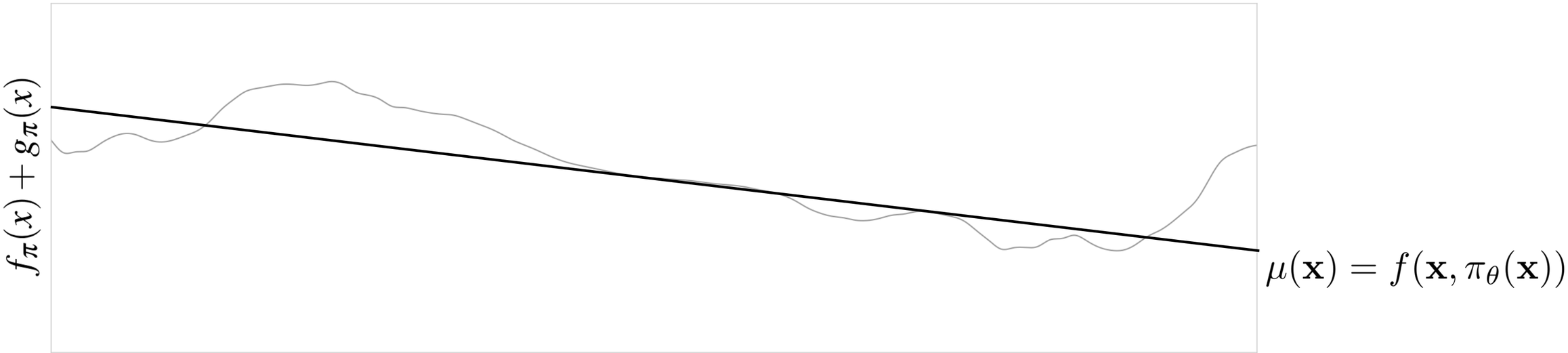
Safety-critical



Approximate dynamic programming

Dynamics

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



Reinforcement learning

Policy

$\pi_\theta(\mathbf{x})$

Safe exploration

Exploration

Safe policy update

Policy update

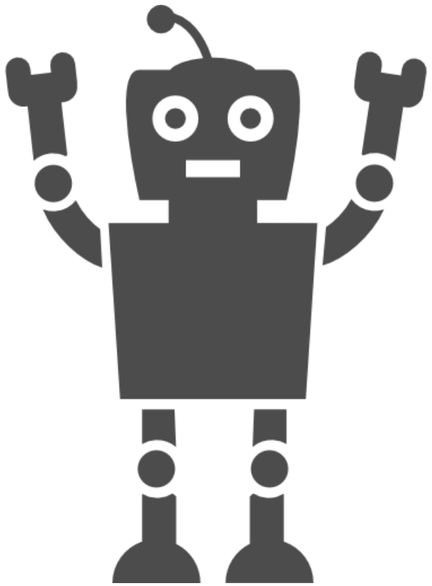
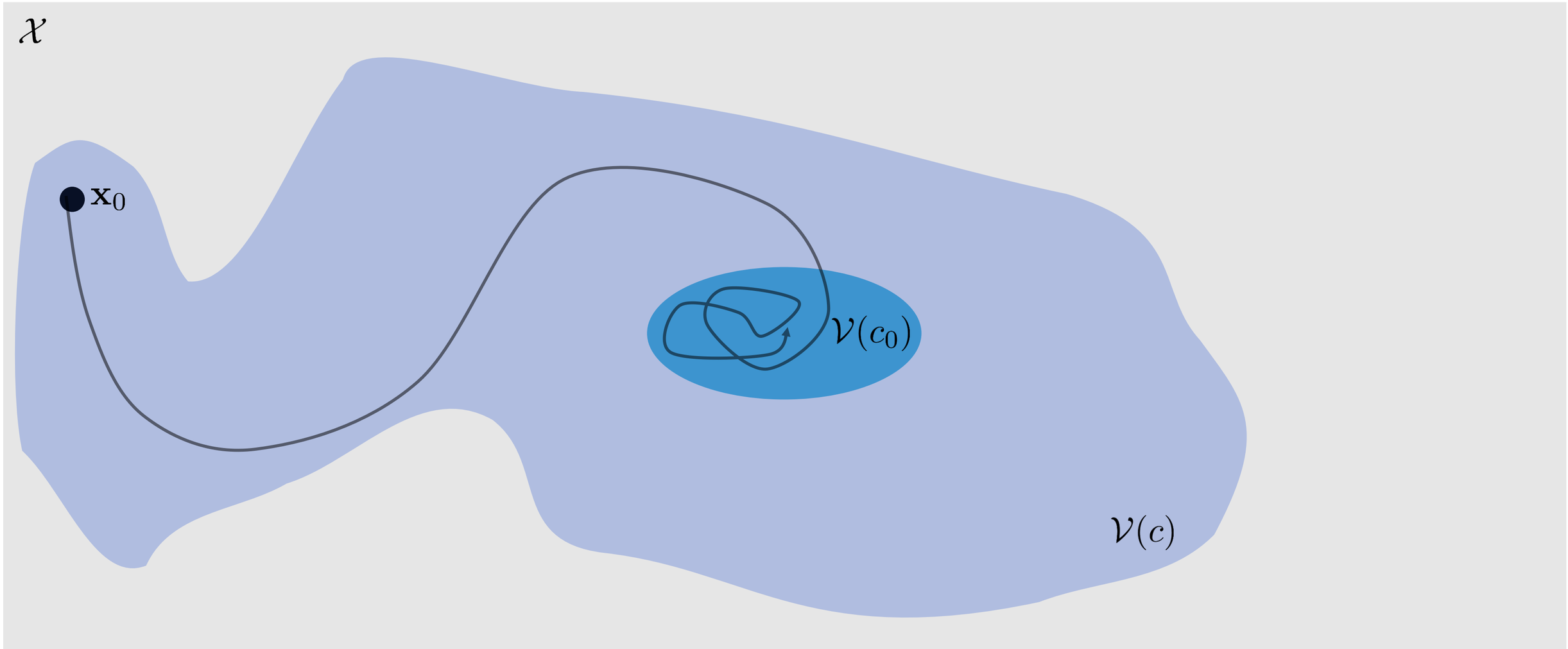


Image: Plainicon, https://flaticon.com

Region of attraction



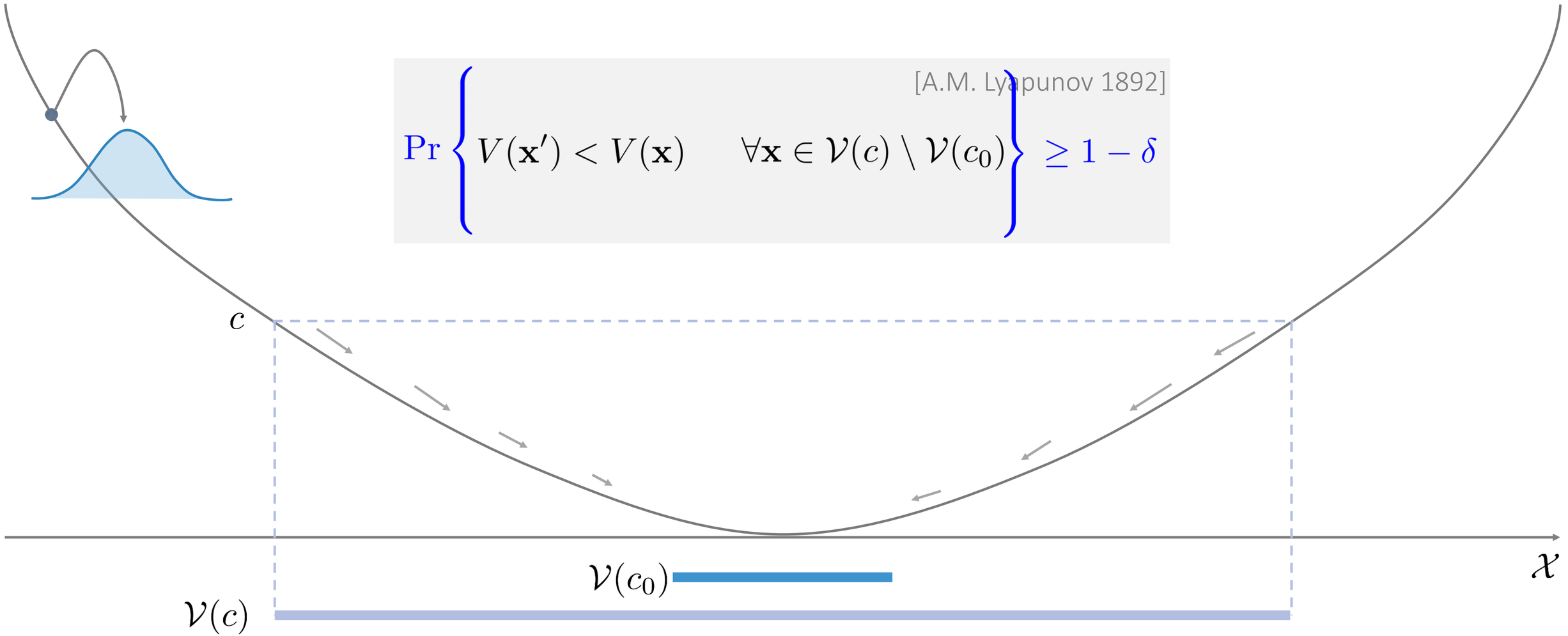
Lyapunov functions

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \pi_\theta(\mathbf{x}_k)) + g(\mathbf{x}_k, \pi_\theta(\mathbf{x}_k))$$

$V(\mathbf{x})$

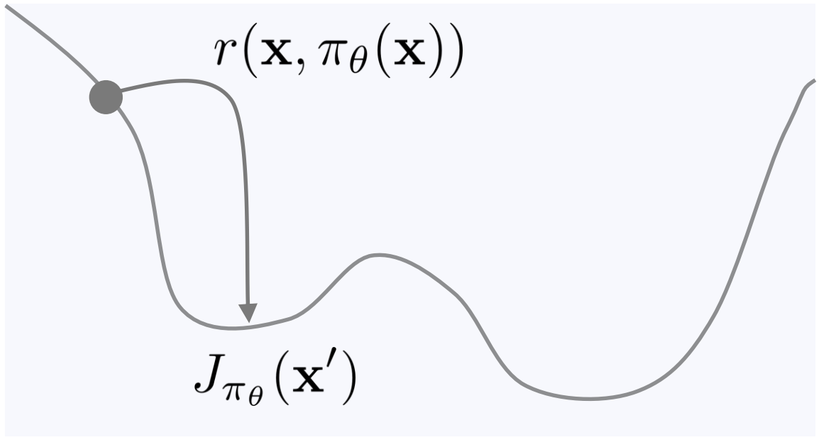
[A.M. Lyapunov 1892]

$$\Pr\left\{ V(\mathbf{x}') < V(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathcal{V}(c) \setminus \mathcal{V}(c_0) \right\} \geq 1 - \delta$$

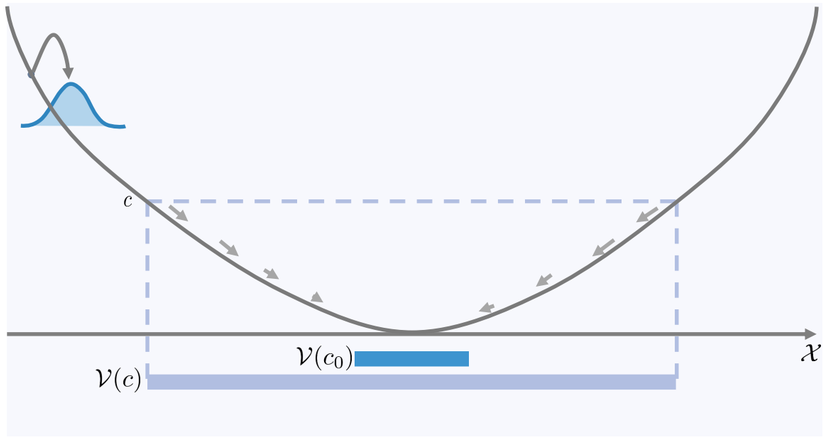


Safe policy optimization (NIPS 2017)

Optimize policy for performance



Determine safe region

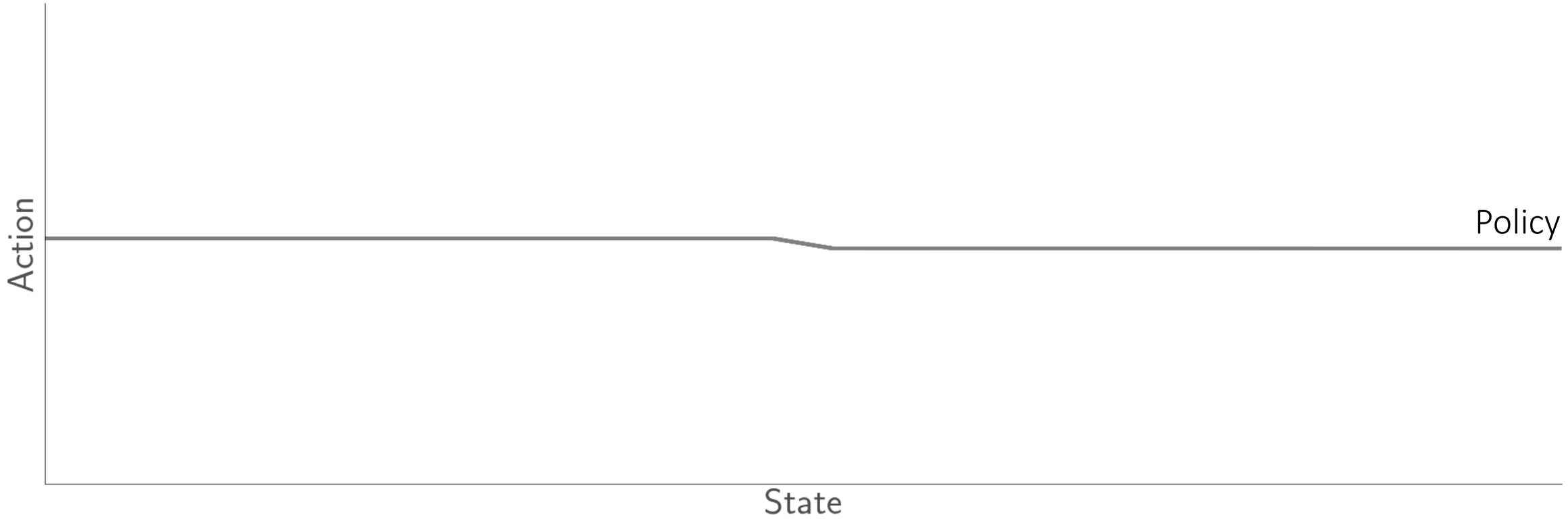


Policy update

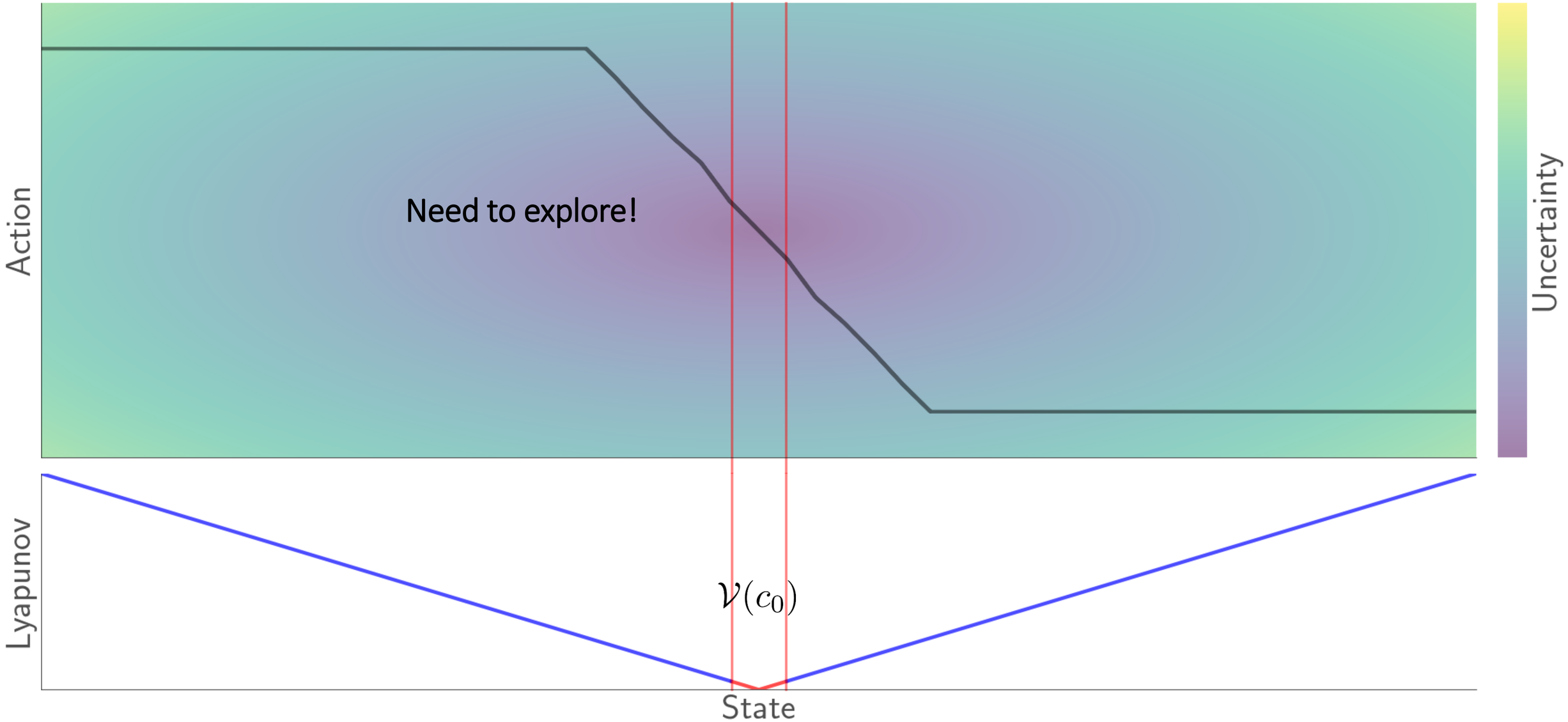
$$\min_\theta \int_{\mathcal{X}} r(\mathbf{x}, \pi_\theta(\mathbf{x})) + \gamma J_{\pi_\theta}(\mathbf{x}')$$

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

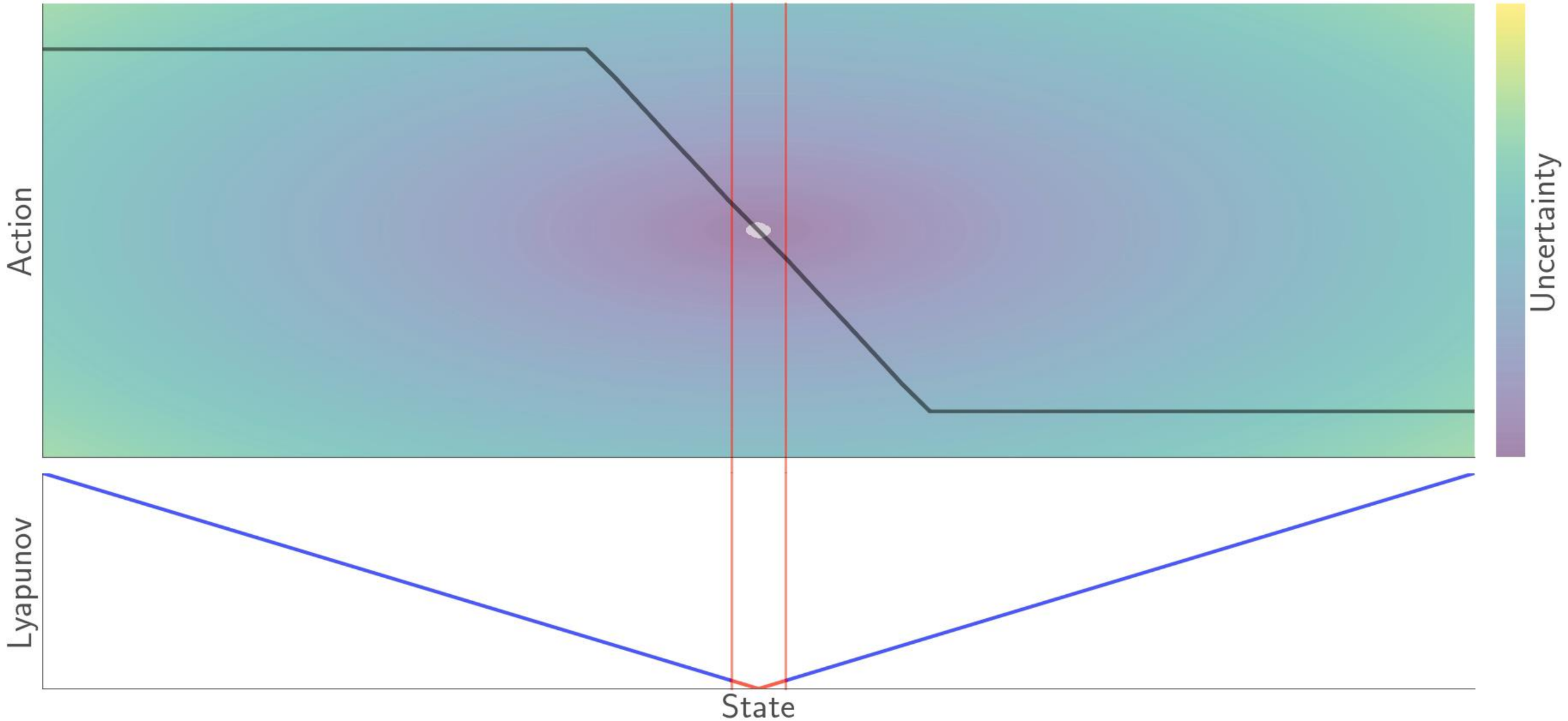
Policy optimization



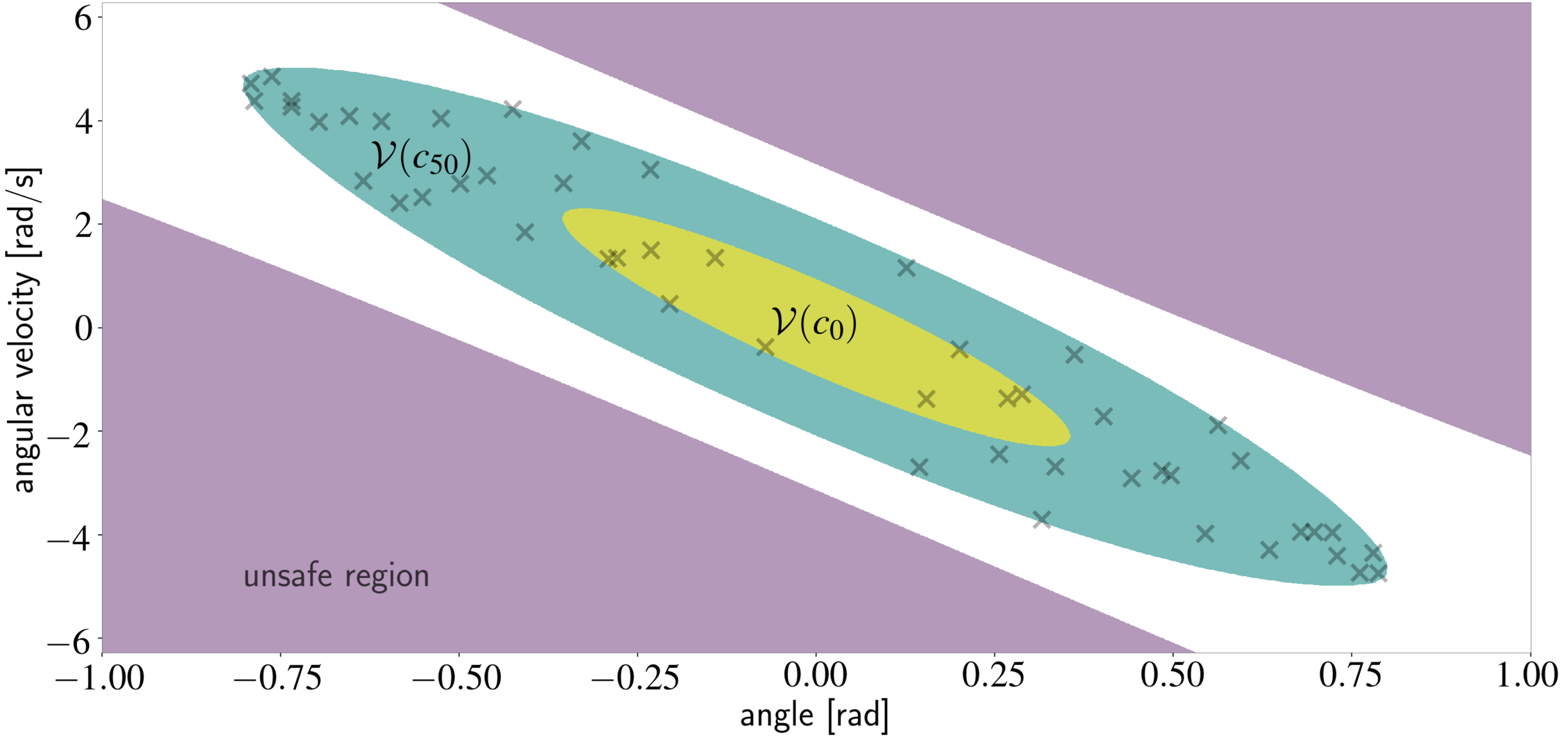
Policy optimization



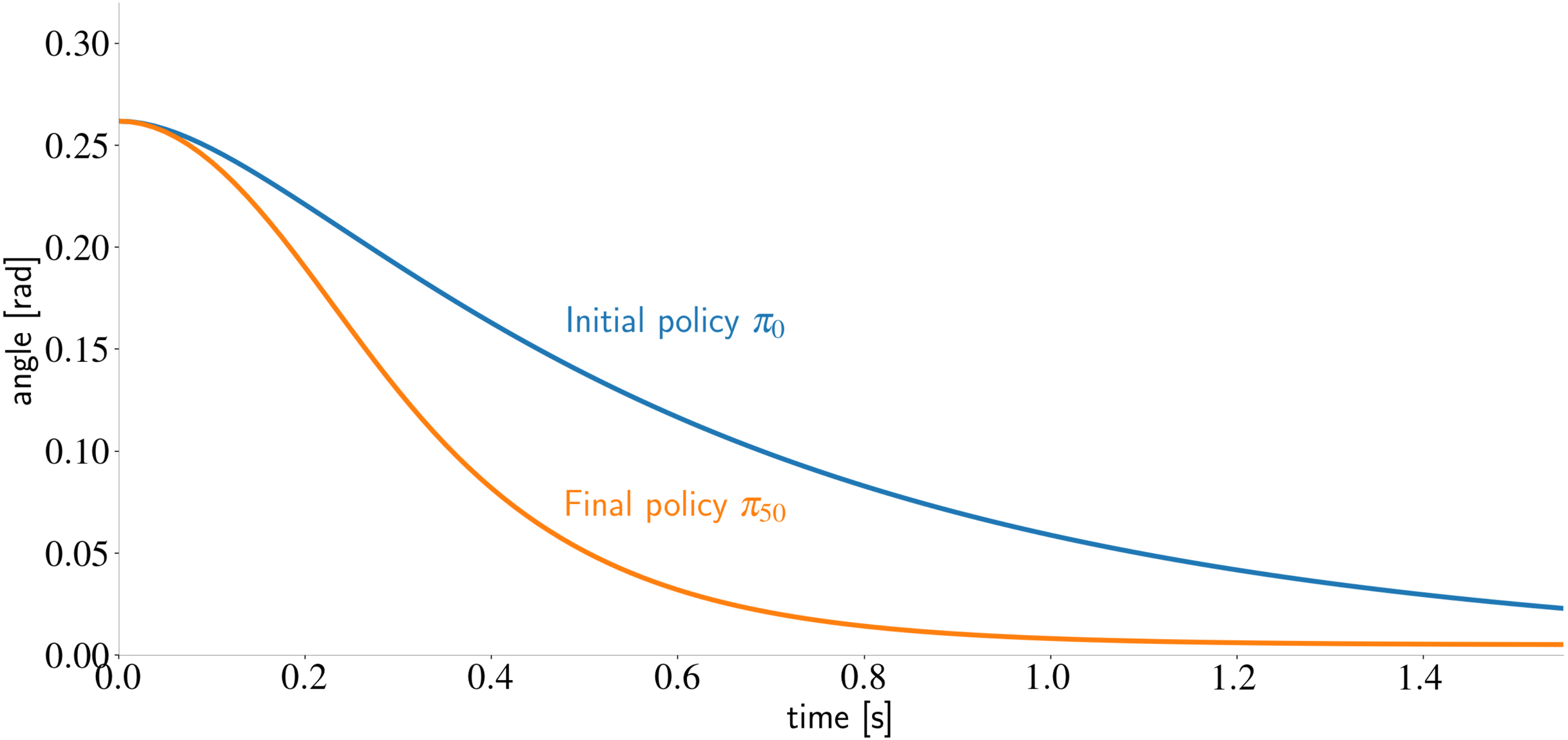
Obtaining data



Experimental results



Policy performance



Conclusion

Safe reinforcement learning!

Can use **statistical** models to give high-probability safety guarantees

Theoretical guarantees in the paper

Code at github.com/befelix

More safe learning at http://berkenkamp.me