Certified Learning of Safety Certificates

Jason Choi, Ian Chuang, Sampada Deglurkar, **Jingqi Li**, Ebonye Smith, **Chris Strong**
Claire Tomlin

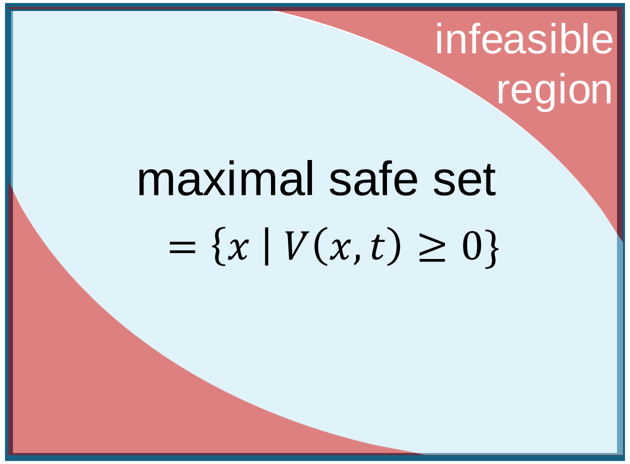
University of California, Berkeley
February 26 2025

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Hybrid Systems Lab

1. Data-driven safe set construction

Compute **maximal safe set**:

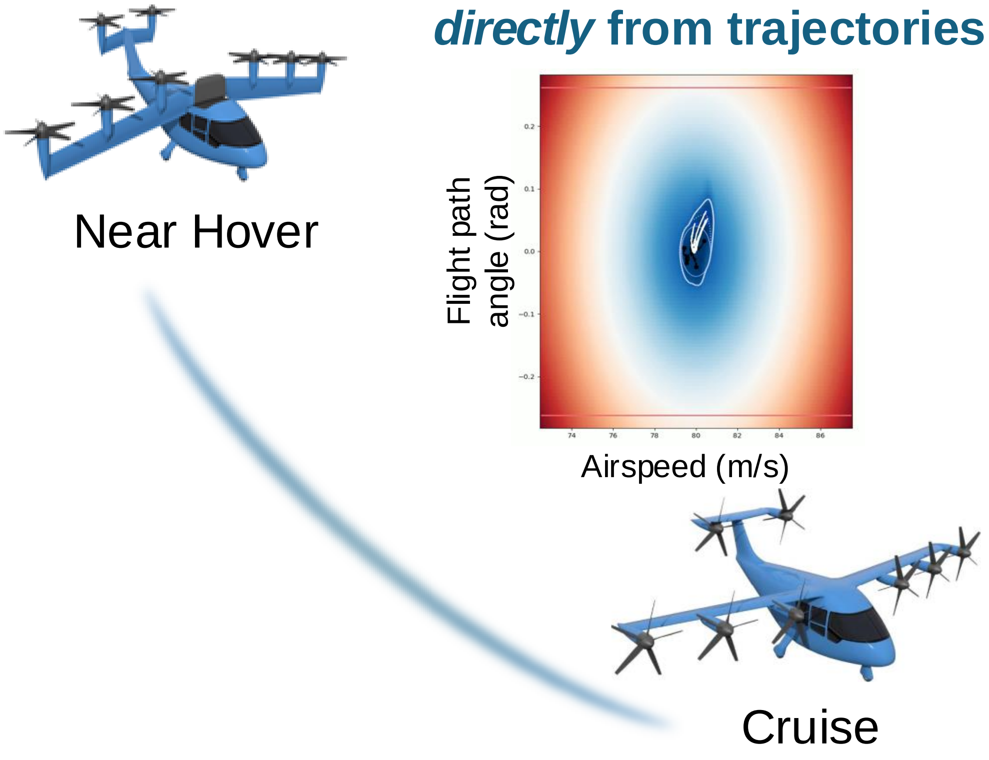


Solution to Hamilton-Jacobi PDE:

$$\min\left\{ l(x) - V(x,t), \frac{\partial V}{\partial t} + H\big(x, \nabla V(x,t)\big) \right\} = 0$$

$$V(x,0) = l(x)$$

Constructing safe operation region *directly* from trajectories



1. Data-driven safe set construction

$$\dot{x} = f(x,u),\ u \in U$$

$$\dot{x} = \underbrace{v}_{= f(x,u)}, \quad v \in \underbrace{F(x)}_{\text{vector field bound}}$$

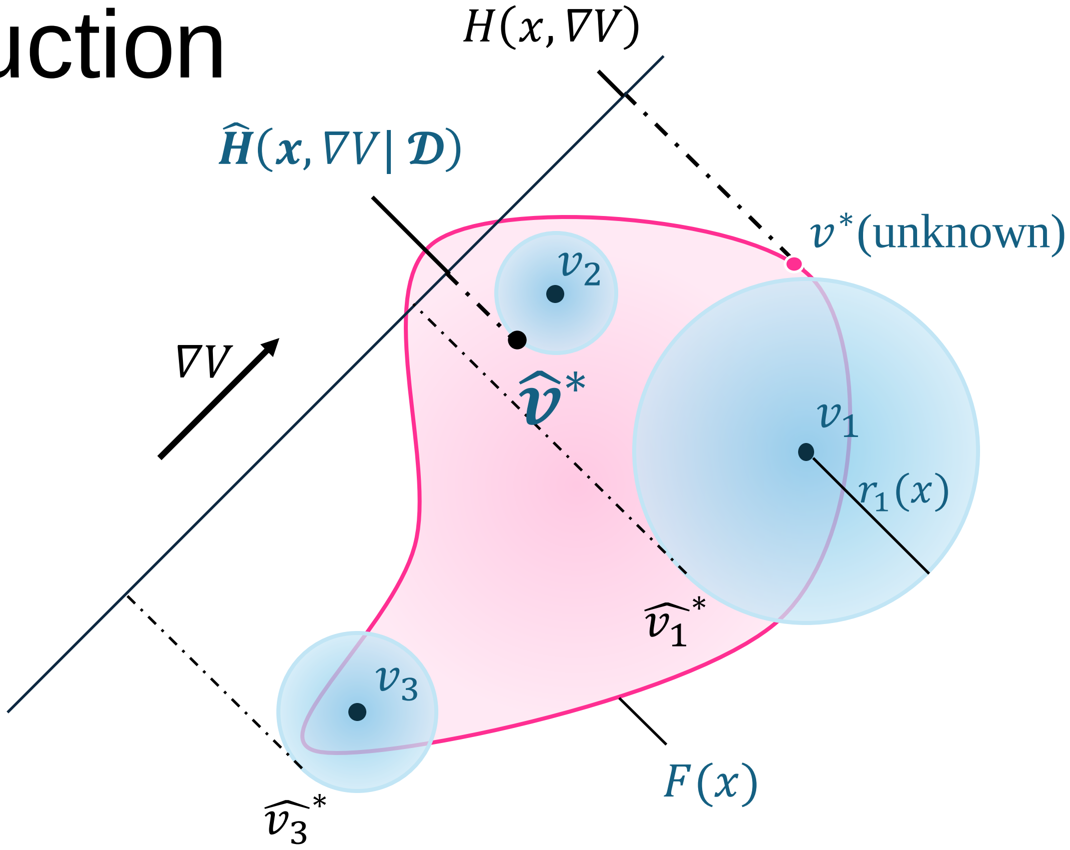
$$H(x, \nabla V) = \max_{u \in U} \nabla V \cdot f(x,u)$$

linear objective

$$H(x, \nabla V) = \max_{v \in \boxed{F(x)}} \boxed{\nabla V \cdot v}$$

nonconvex feasible set

1. Data-driven safe set construction



$$\hat{H}(x, \nabla V) := \max_{\hat{v} \in \{\hat{v}_i\}_{i=1}^N} \min_{\{\hat{v}_i\}_{i=1}^N} \nabla V \cdot \hat{v}$$

$$\text{s.t.} \quad \|\hat{v}_i - v_i\| \leq r_i(x),$$

$$\text{for } i = 1, \cdots, N$$

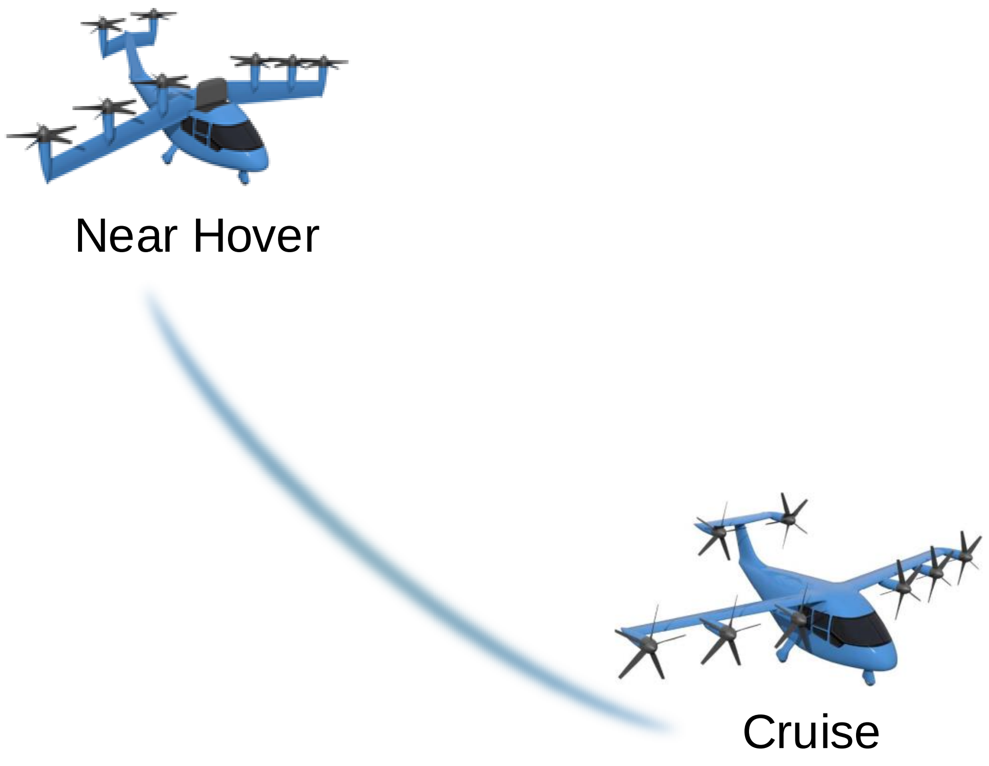
Data-driven Hamiltonian $\hat{H}(x, \nabla V | \mathcal{D})$:

- convex optimization problem with a closed-form solution
- guaranteed underapproximation of the ground-truth Hamiltonian

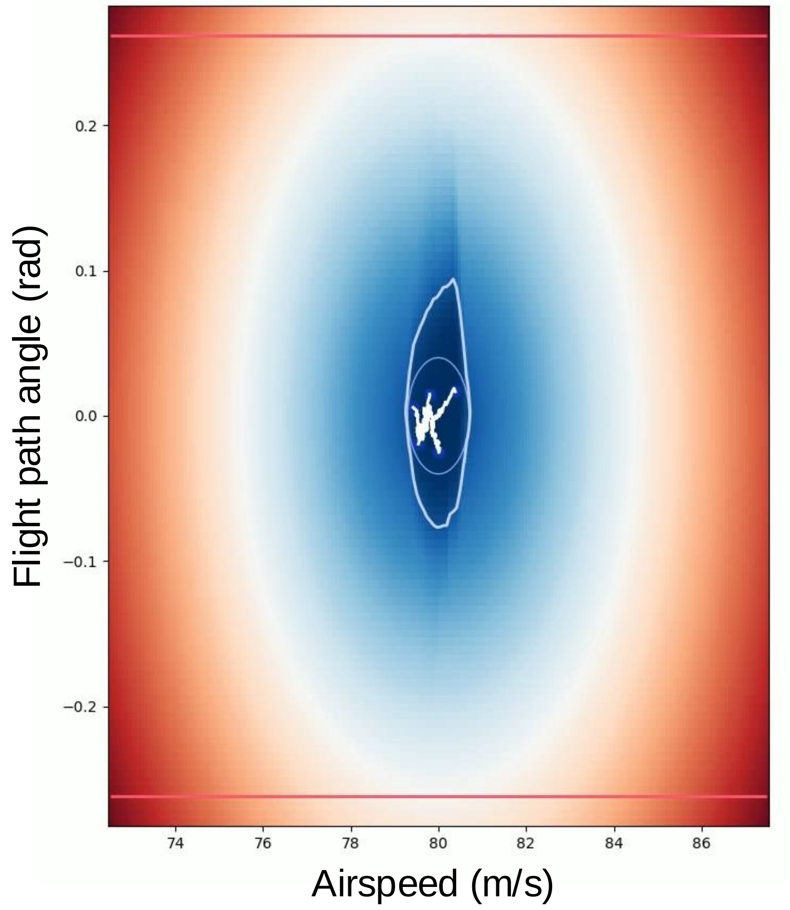
$$\hat{\boldsymbol{H}}(\boldsymbol{x}, \boldsymbol{\nabla V} | \mathcal{D}) \leq H(x, \nabla V) = \max_{u \in U} \nabla V \cdot f(x, u)$$

- results in a **guaranteed underapproximation** of the ground-truth safe set

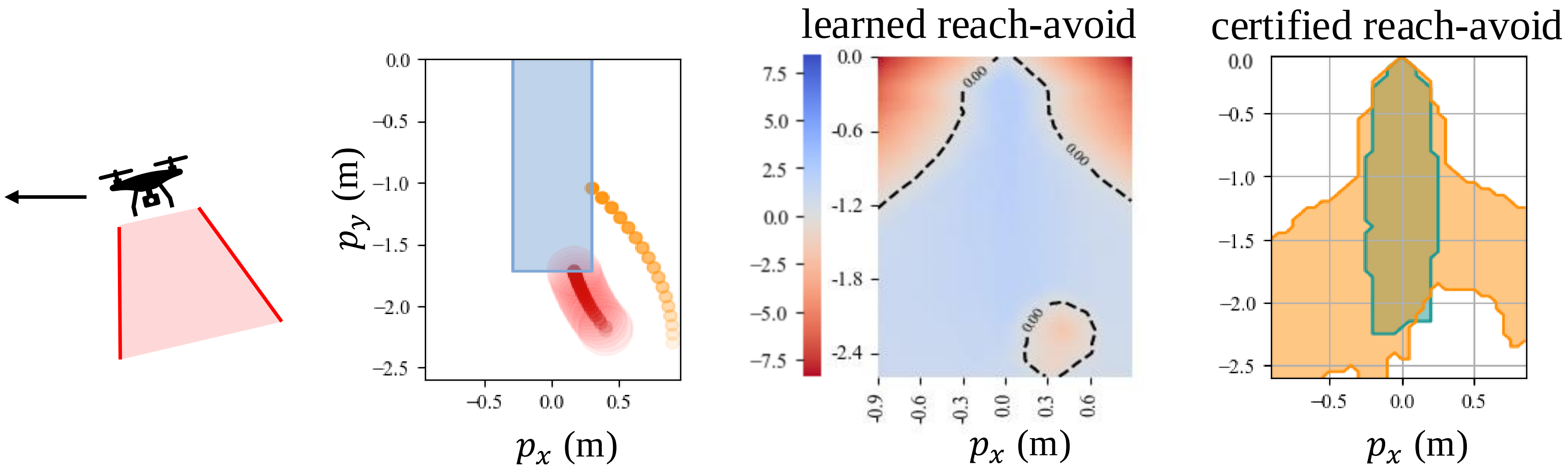
1. Data-driven safe set construction: Safe experiment design



Test iterations of eVTOL dynamics:



2. Certifying learned reach-avoid sets



2. Certifying learned reach-avoid sets: A new reach-avoid value function

$$V_{\gamma}(x_0) := \max_{\pi} \min_{\phi} \sup_{t=0,\dots} \min\left\{\gamma^t r(x_t), \min_{\tau=0,\dots,t} \gamma^\tau c(x_\tau)\right\}$$

Offline learning of a new reach-avoid value function and its policy

$\longrightarrow$

Post-learning certification of learned reach-avoid set

Deep Deterministic Policy Gradient (DDPG) to learn the value function and policy

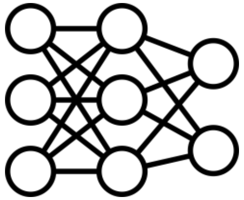
Control policy

x $\quad \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}$ $\quad \pi_{\theta_u}(x)$

$$\max_{\theta_u} \mathbb{E}_{x\sim\mathbb{P}}\; Q_{\theta_q}(x, \pi_{\theta_u}(x), \phi_{\theta_d}(x))$$

Disturbance policy

x

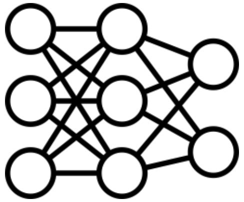


$\pi_{\theta_d}(x)$

$$\min_{\theta_d} \mathbb{E}_{x\sim\mathbb{P}}\; Q_{\theta_q}(x, \pi_{\theta_u}(x), \phi_{\theta_d}(x))$$

NN Q function

x
u
d



$Q_{\theta_q}(x, u, d)$

$$V_\theta(x) := Q_{\theta_q}(x, \pi_{\theta_u}(x), \phi_{\theta_d}(x))$$
$$\min_{\theta_q} \mathbb{E}_{x\sim\mathbb{P}} \| V_\theta(x) - B_\gamma[V_\theta(x)] \|_2^2$$

[DDPG: S. Li et al. AAAI 2019]

Certification method 1: Use suboptimal policy & Lipschitz constants to construct lower bound

$$\max_{\pi} \min_{\phi} \sup_{t=0,\dots} g_\gamma(\xi_x^{\pi,\phi}, t) = V_\gamma(x)$$

suboptimal policy π

$$\min_{\phi} \sup_{t=0,\dots} g_\gamma(\xi_x^{\pi,\phi}, t) \leq V_\gamma(x)$$

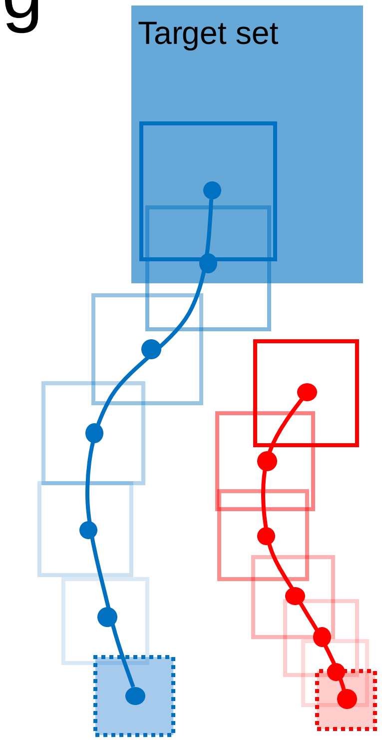
Lower bound using Lipschitz constants

$$\sup_{t=0,\dots} \check{g}_\gamma(\xi_x^{\pi,0}, t) \leq V_\gamma(x)$$

Finite horizon sim

$$\max_{t=0,\dots,T} \check{g}_\gamma(\xi_x^{\pi,0}, t) \leq V_\gamma(x)$$

Certification method 2: Use second order cone programming



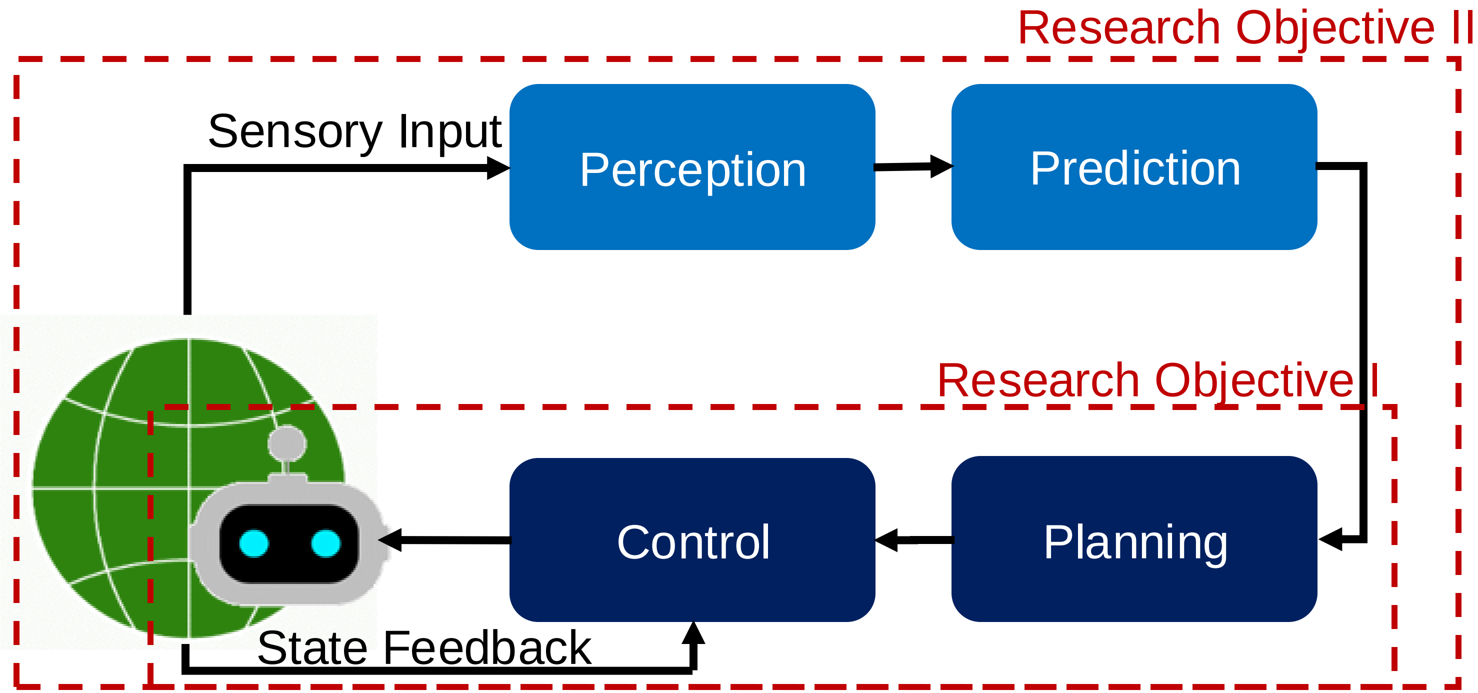
SOCP RA set verification:
We examine whether we can safely reach the target set under the worst-case disturbance by solving a sequence of SOCPs

Advantages, Limitations

- Compares favorably with baselines
- Can be used in real-time for local certification
- Lipschitz continuity appears to accelerate reachability learning
- Conservative by design
- Value function/policy learned offline, could be subject to distributional shift


H-4177S-Y

Certified learning of safety certificates



Thanks to Donggun Lee and Namhoon Cho