Trustworthy AI Autonomy

M3-2: Trustworthy RL-Safety

Ding Zhao

Assistant Professor

Carnegie Mellon University

Carnegie Mellon University

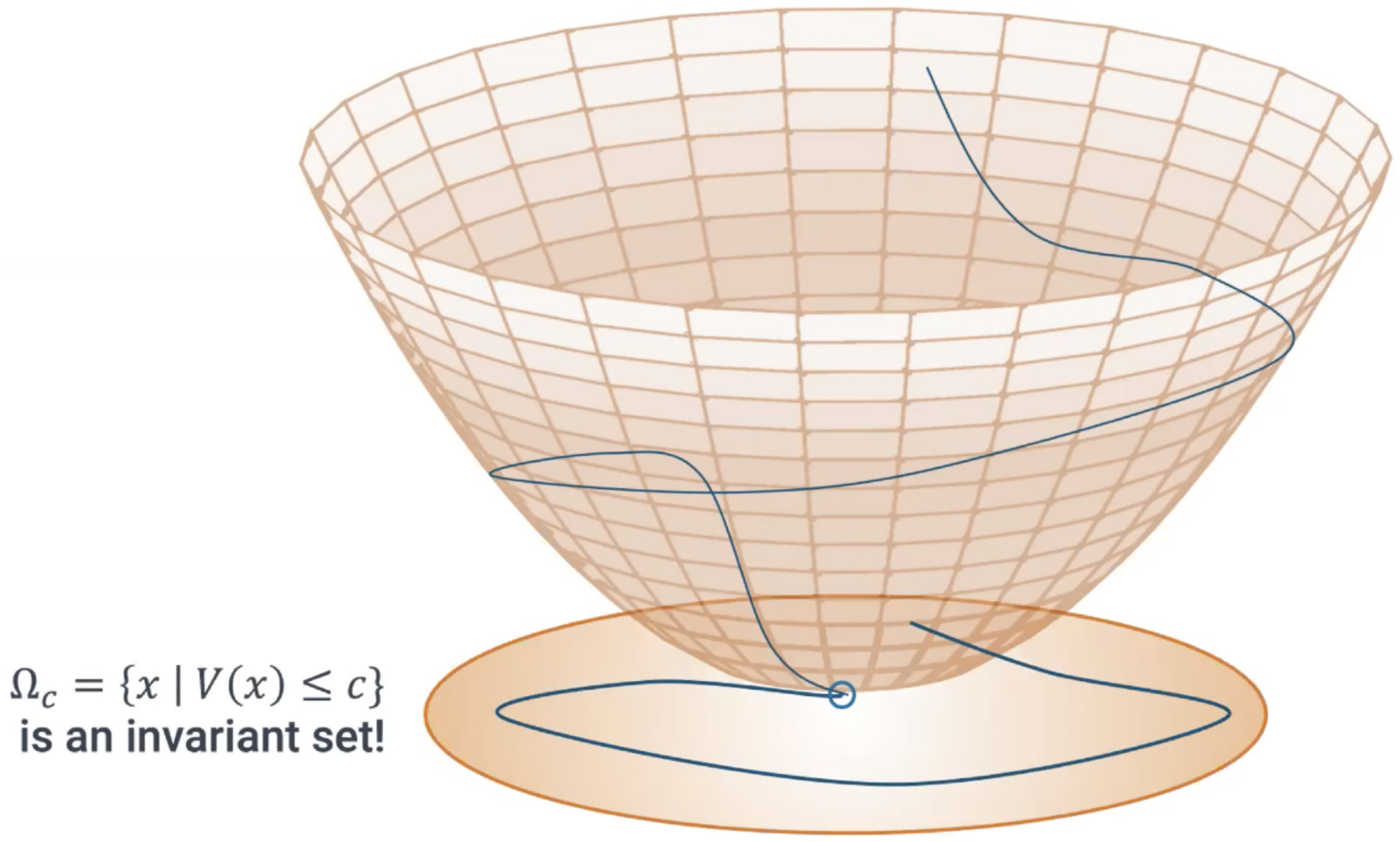
Safe AI Lab @CMU

Plan for today

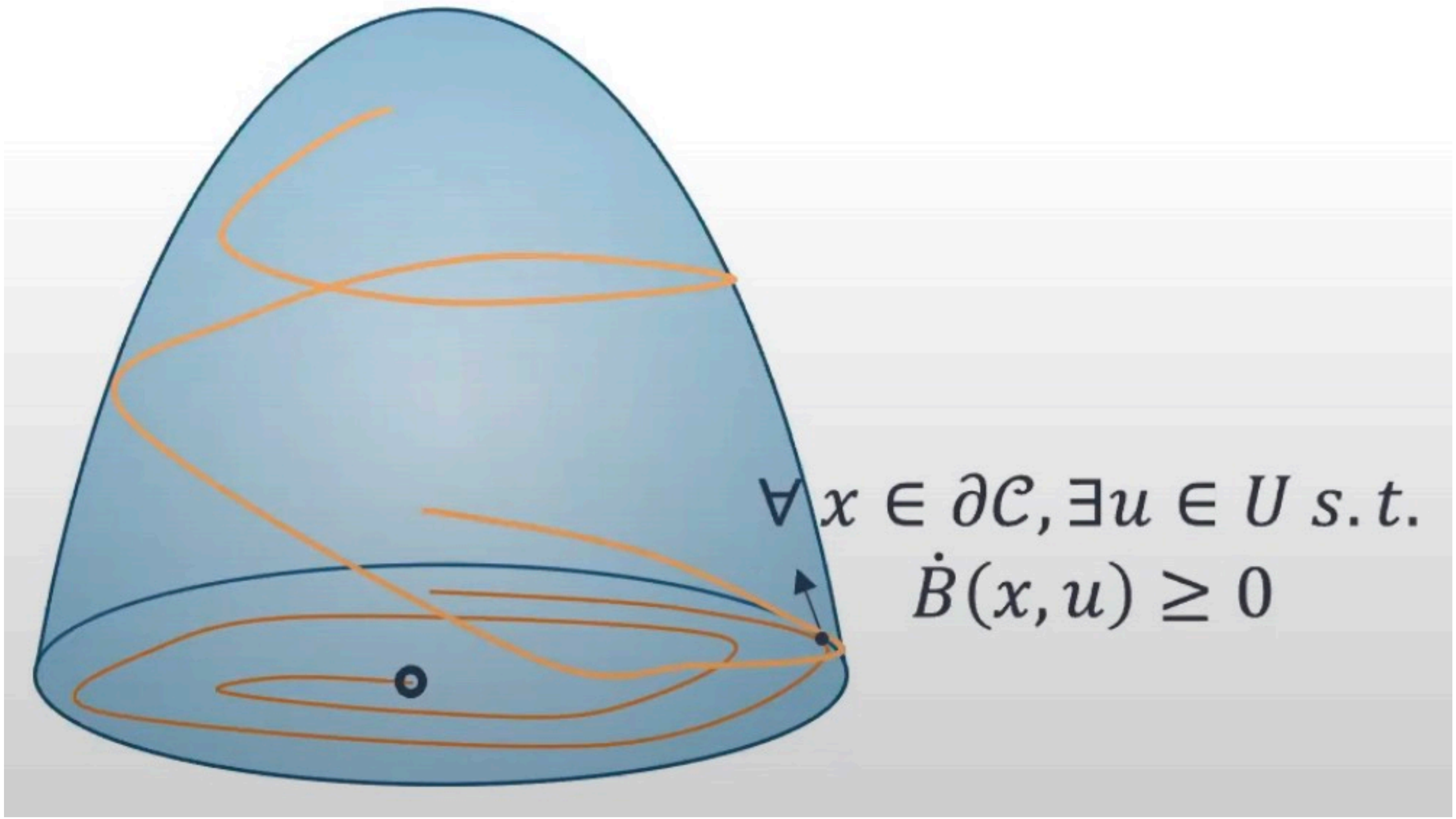
- Intuitions and definitions
    - Control Lyapunov Function (CLF)
    - Control Barrier Function (CBF)
- How they are used in safe RL
- Challenges and open problems

Intuitions

- Control Barrier Function (CBF) and Control Lyapunov Function (CLF) are commonly used in a control system to ensure safety
- Intuition: CLF is designed for reaching a target safe state in a stable way, while CBF is designed for avoiding a unsafe set.



An level set



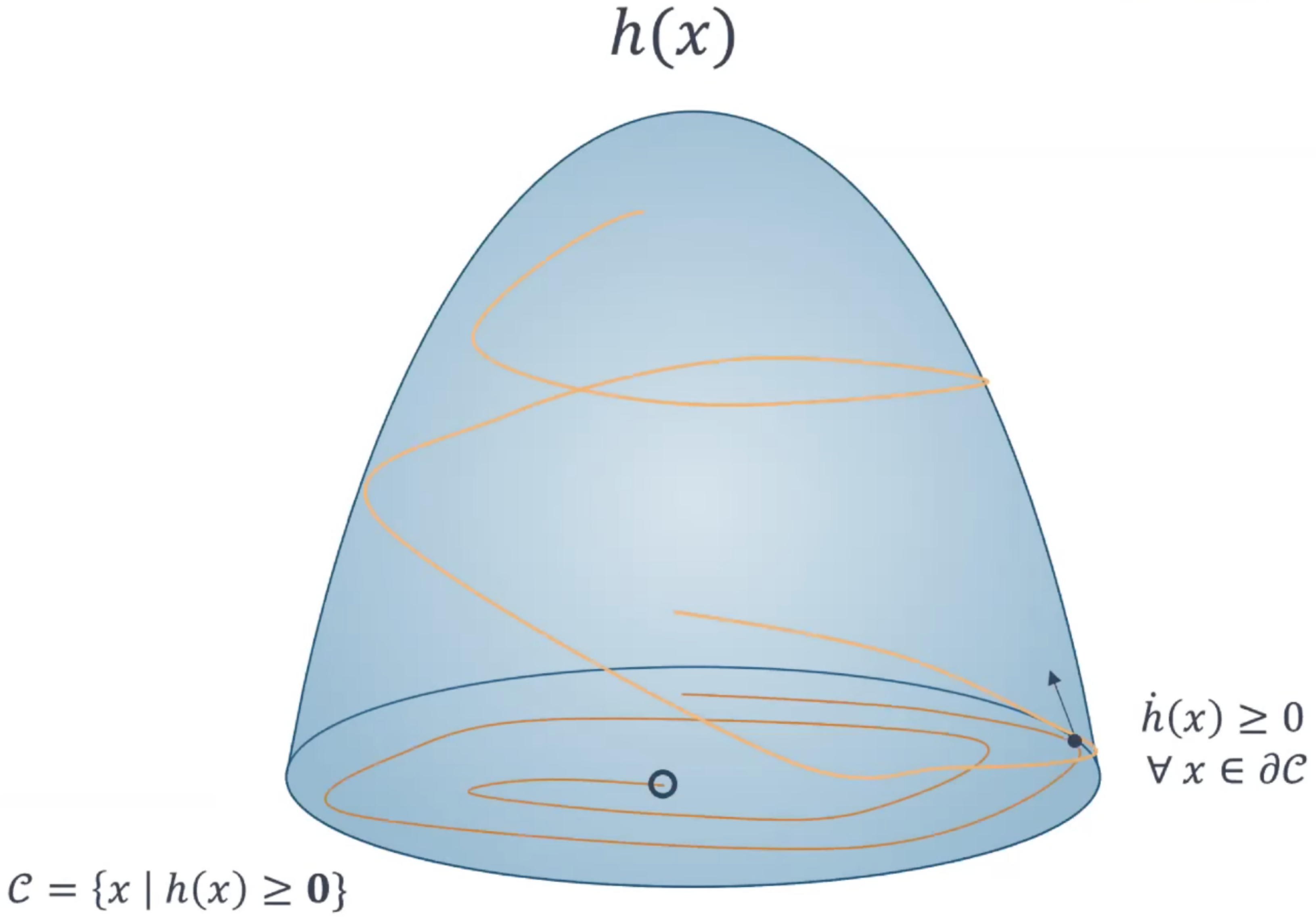
Control affine system

- Consider a non-linear time-invariant control affiliate system: $\dot{x} = f(x) + g(x)u$, where $x \in \mathbb{R}^n$ is the state and $u \in \mathbb{R}^m$ is the control input.

- f, g are Lipschitz continuous in x, and assume $x_e = 0$ is an equilibrium point.

- Denote the safe set as $\mathcal{S} = \{x \,|\, h(x) \geq 0\}$, where $h(x)$ is differentiable.

- The dynamical system should always be within the safe set



Marvi, Z., & Kiumarsi, B. (2021). Safe reinforcement learning: A control barrier function optimization approach. *International Journal of Robust and Nonlinear Control*, *31*(6), 1923-1940.

Control Lyapunov Function (CLF)

- Denote $V(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ as a continuous differentiable function.
- If there is a positive constant c such that:
  - An energy function: $V(x_e) = 0; \forall x \in \mathbb{R}^n \backslash \{x_e\}, V(x) > 0$
  - A level set: $\Omega_c = \{x \in \mathbb{R}^n : V(x) \leq c\}$
  - Energy decreasing over time: $\inf_{u \in U} \dot{V}(x, u) < 0, \forall x \in \Omega_c \backslash \{x_e\}$
- Then $V(x)$ is a local CLF, and Ω_c is the region of attraction (ROA), i.e. every state in ROA is asymptotically stabilizable to x_e.

Choi, Jason J., et al. "Robust control barrier-value functions for safety-critical control." *arXiv preprint arXiv:2104.02808* (2021).

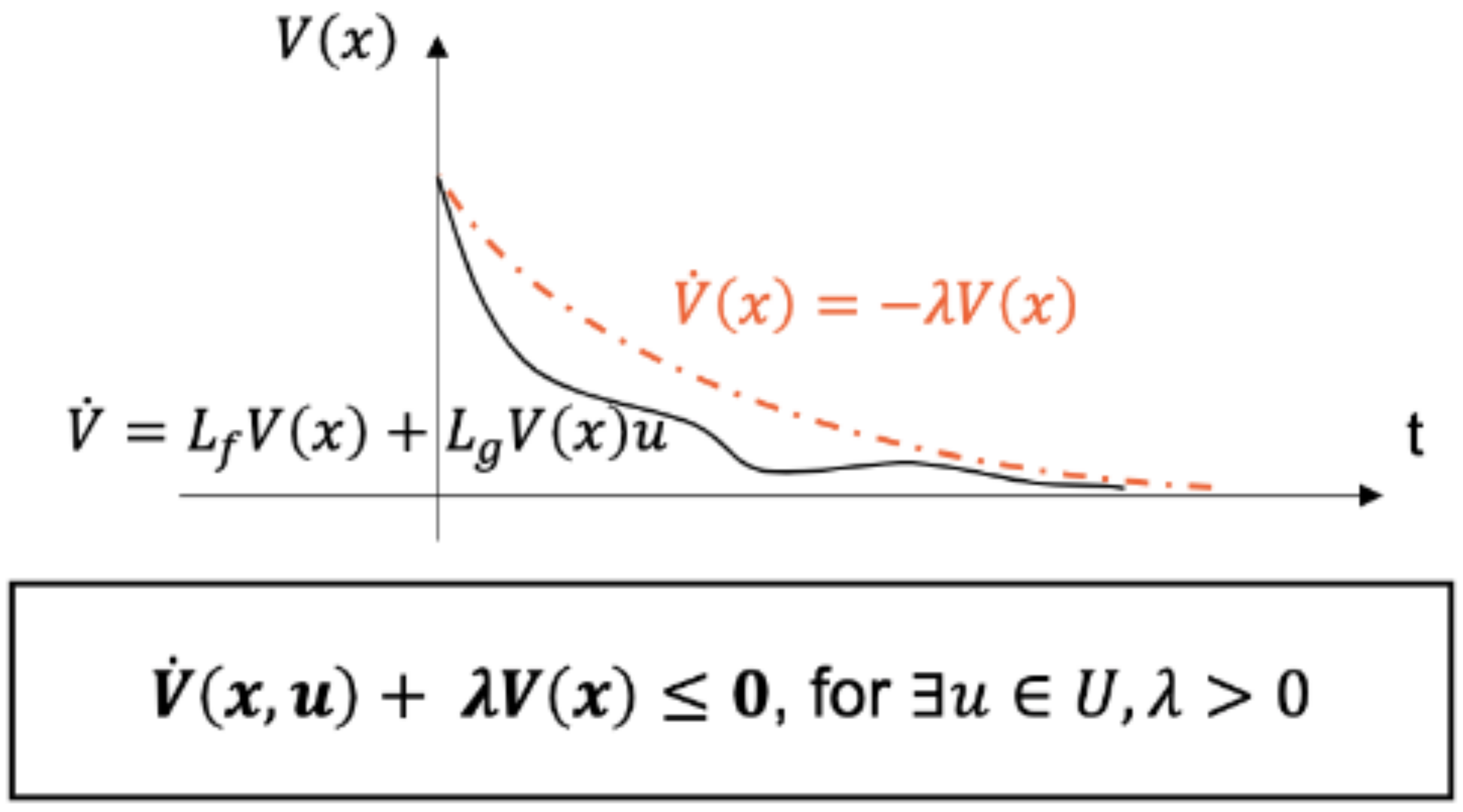
Control Lyapunov Function (CLF)

- Derivative of $V(x)$ along the dynamics is affine in u:
  - $\dot{V}(x, u) = \nabla V(x) f(x) + \nabla V(x) g(x) u = L_f V(x) + L_g V(x) u.$
- $L_p q(x) = \nabla q(x) p(x)$ is called the Lie derivative operator.
- Intuition behind CLF for safety: if the system starts near a safe equilibrium point (within ROA), then it will stay within the safety region (ROA) forever.

Choi, Jason J., et al. "Robust control barrier-value functions for safety-critical control." *arXiv preprint arXiv:2104.02808* (2021).

Exponential stability of CLF

- If there exists a positive constant λ such that:
    - $\inf_{u \in U} \dot{V}(x, u) + \lambda V(x) < 0$
- Then $V(x)$ is an exponentially stabilizing CLF, and any x is exponentially stabilizable to x_e.



Choi, Jason J., et al. "Robust control barrier-value functions for safety-critical control." *arXiv preprint arXiv:2104.02808* (2021).

Control Lyapunov Function (CLF)

- The CLF constraint is linear to u, so we can construct a quadratic programming formulation to track the reference control u_{ref}

$$\arg\min(u - u_{ref})^T H(u - u_{ref}) + p\delta^2$$

u : control input
δ : slack variable

subject to: $L_f V(x) + L_g V(x)u + \lambda V(s) \leq \delta$ CLF constraint

- H: objective matrix, p: weight for the slack variable
- It is a convex optimization problem, which can be solved efficiently.
- The slack variable is used to guarantee the feasibility of the QP problem.

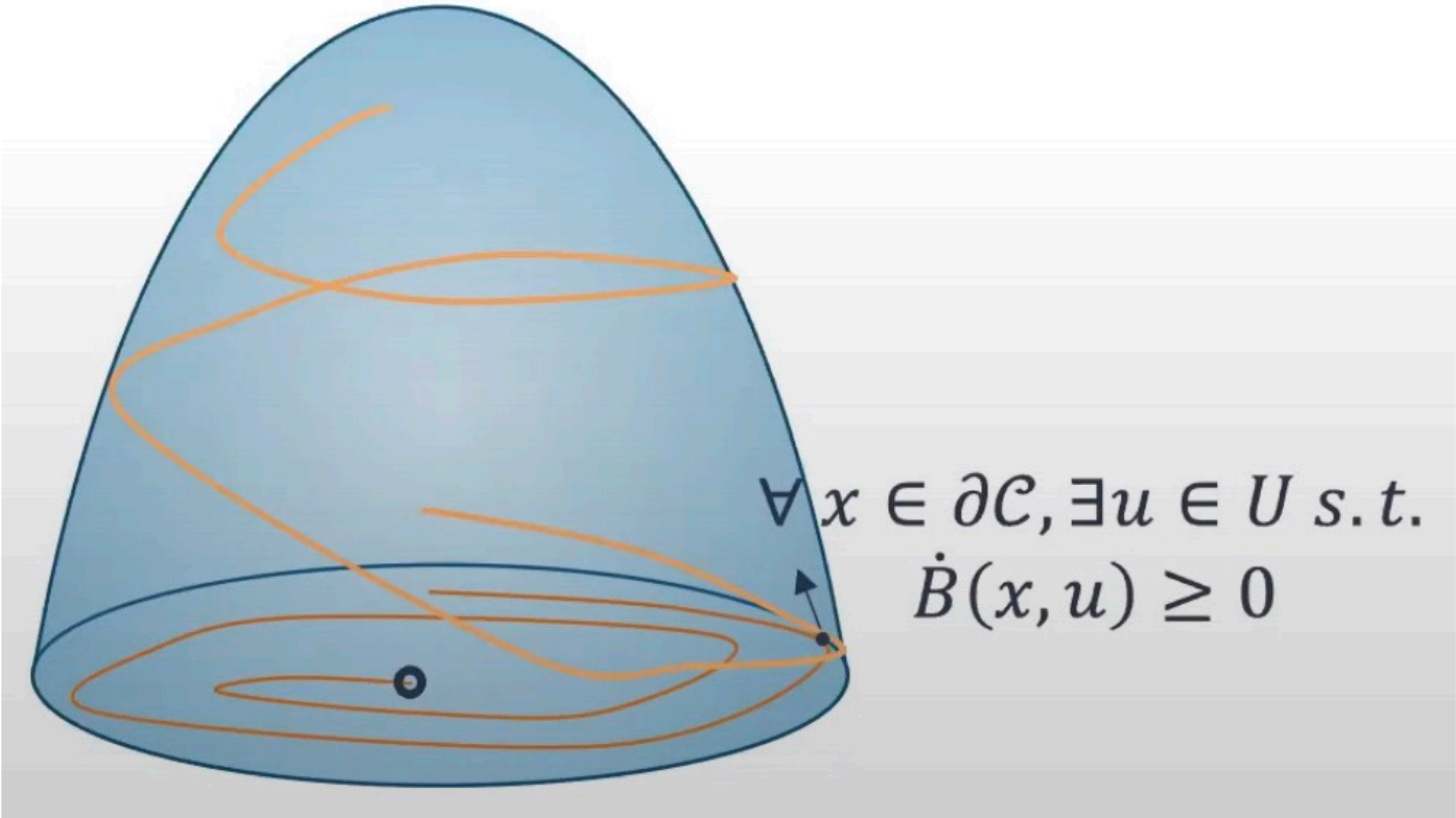
Choi, Jason J., et al. "Robust control barrier-value functions for safety-critical control." *arXiv preprint arXiv:2104.02808* (2021).

Control Barrier Function (CBF)

- A valid barrier function should be 1) positive within a set and reaches infinity at the boundary of the safe set; 2) has negative derivative in the vicinity of the boundary, and thus never reaches infinity.

- Forward invariance: A forward invariant set for a dynamical system is **a set that has solutions evolving within the set** (Nagumo's theorem).

- CBF can help to ensure the forward invariance within the safe set.



Marvi, Z., & Kiumarsi, B. (2021). Safe reinforcement learning: A control barrier function optimization approach. *International Journal of Robust and Nonlinear Control*, *31*(6), 1923-1940.

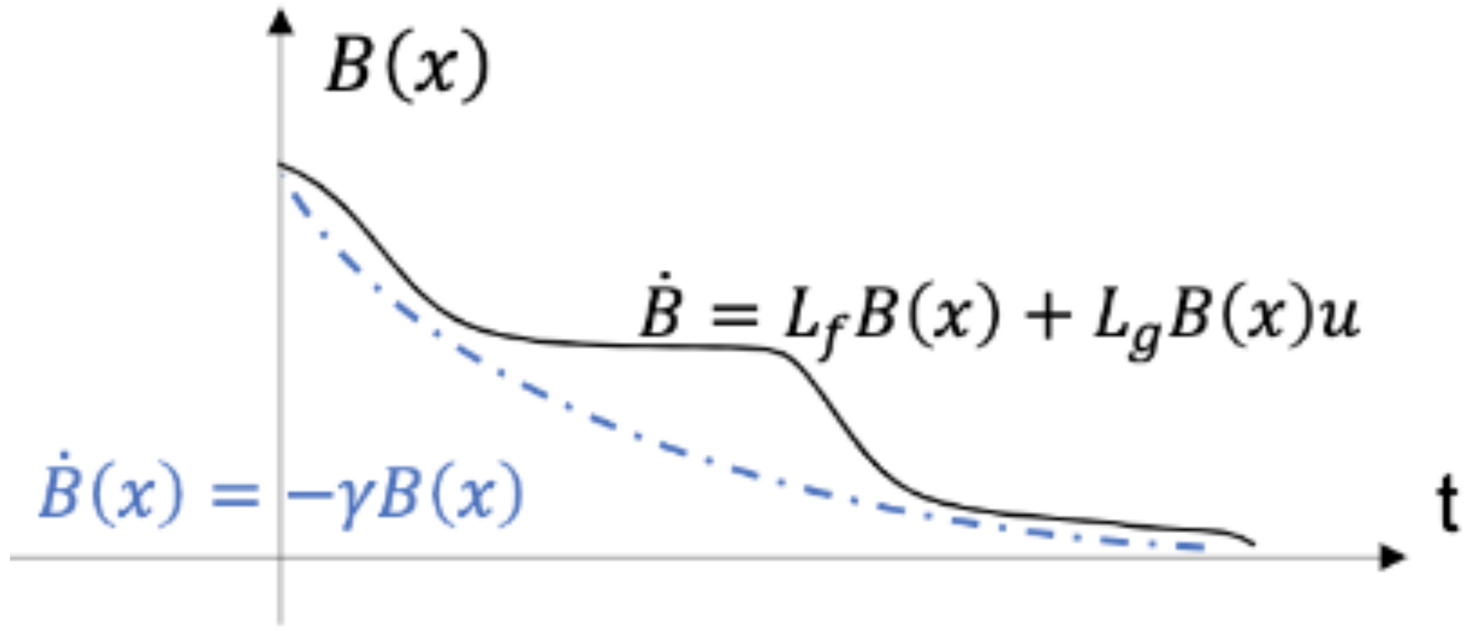
Control Barrier Function (CBF)

- Denote $B(x) : \mathbb{R}^n \to \mathbb{R}$ as a continuous differentiable function, if there exists $D, s.t. \mathcal{S} \subset D,$ and:

$$\sup_{u \in U} L_f B(x) + L_g B(x) u + \gamma B(x) \geq 0, \forall x \in D.$$

- Then $B(x)$ is a valid CBF, and any Lipschitz continuous control law that satisfies the above constraint will be within the safe set $\mathcal{S}$.

- γ serves as a decay rate.



$\dot{\boldsymbol{B}}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\gamma B}(\boldsymbol{x}) \geq \mathbf{0}$, for $\exists u \in U,\ \gamma > 0$

Choi, Jason, et al. "Reinforcement learning for safety-critical control under model uncertainty, using control lyapunov functions and control barrier functions." *arXiv preprint arXiv:2004.07584* (2020).

CBF-CLF

- Intuition: CLF is designed for reaching a target safe state in a stable way, while CBF is designed for avoiding a unsafe set

$$\arg\min (u - u_{ref})^T H (u - u_{ref}) + p\delta^2$$

u : control input
δ : slack variable

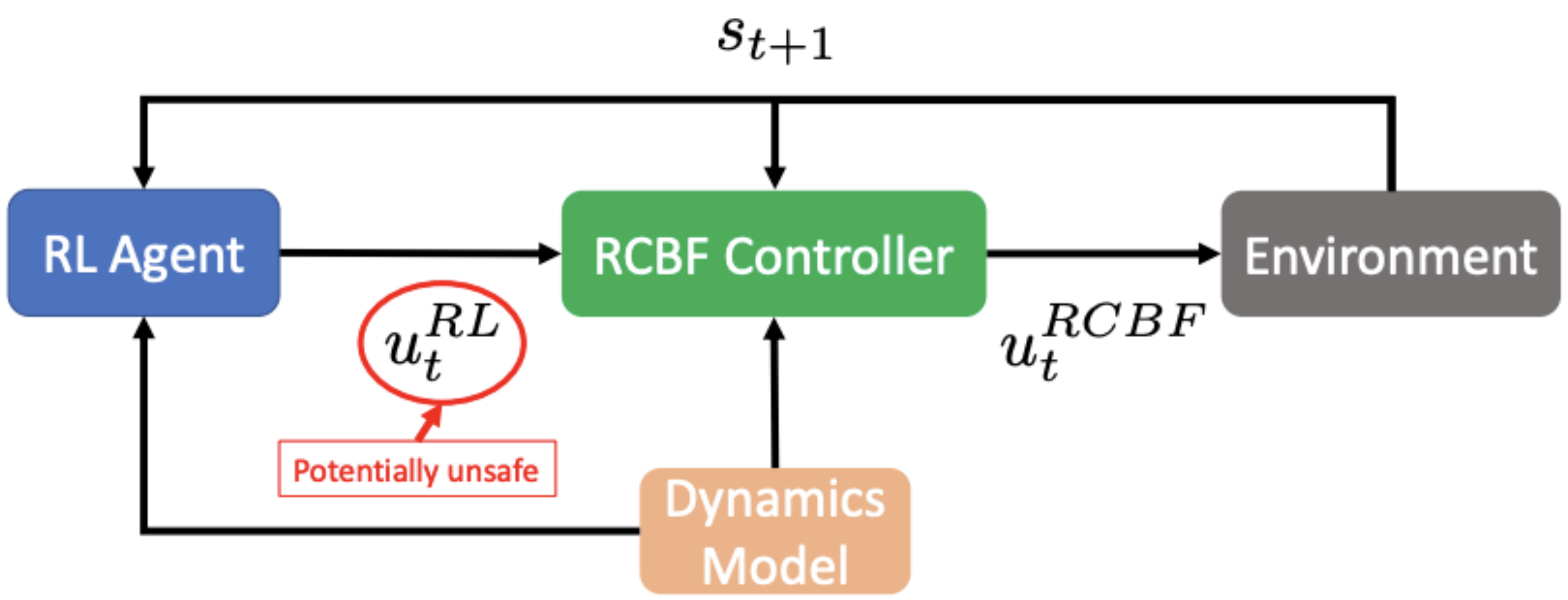
subject to: $L_f V(x) + L_g V(x) u + \lambda V(s) \leq \delta$ CLF constraint

$L_f B(x) + L_g B(x) u + \gamma B(s) \geq 0$ CBF constraint

Choi, Jason, et al. "Reinforcement learning for safety-critical control under model uncertainty, using control lyapunov functions and control barrier functions." *arXiv preprint arXiv:2004.07584* (2020).

CBF in RL

Use CBF as a post-process layer to guarantee safety



Prior knowledge of the system dynamics is required

Algorithm 1 SAC-RCBF

Require:
Dynamics prior $f(\cdot)$ and $g(\cdot)$ and RCBF $h(\cdot)$
for N iterations **do**
  Train GP models p_ψ on $\mathcal{D}_{\text{env}}$
  for E environment steps **do**
    Obtain action u_t^{RL} from π_ϕ
    Render action safe u_t^* using h and p_ψ ▷ (9)
    Take safe action u_t^* in environment
    Add transition to $\mathcal{D}_{\text{env}}$
  end for
  for M model rollouts **do**
    Sample x_t uniformly from $\mathcal{D}_{\text{env}}$
    for k model steps **do**
      Obtain action u_t^{RL} from π_ϕ
      Render action safe u_t^* using h and p_ψ ▷ (9)
      Generate synthetic transition using u_t^* and p_ψ
      Add transition to $\mathcal{D}_{\text{model}}$
    end for
  end for
  for G gradient steps **do**
    Update agent parameters (ϕ and θ) ▷ (6), (7)
  end for
end for

Synthetic data to increase efficiency

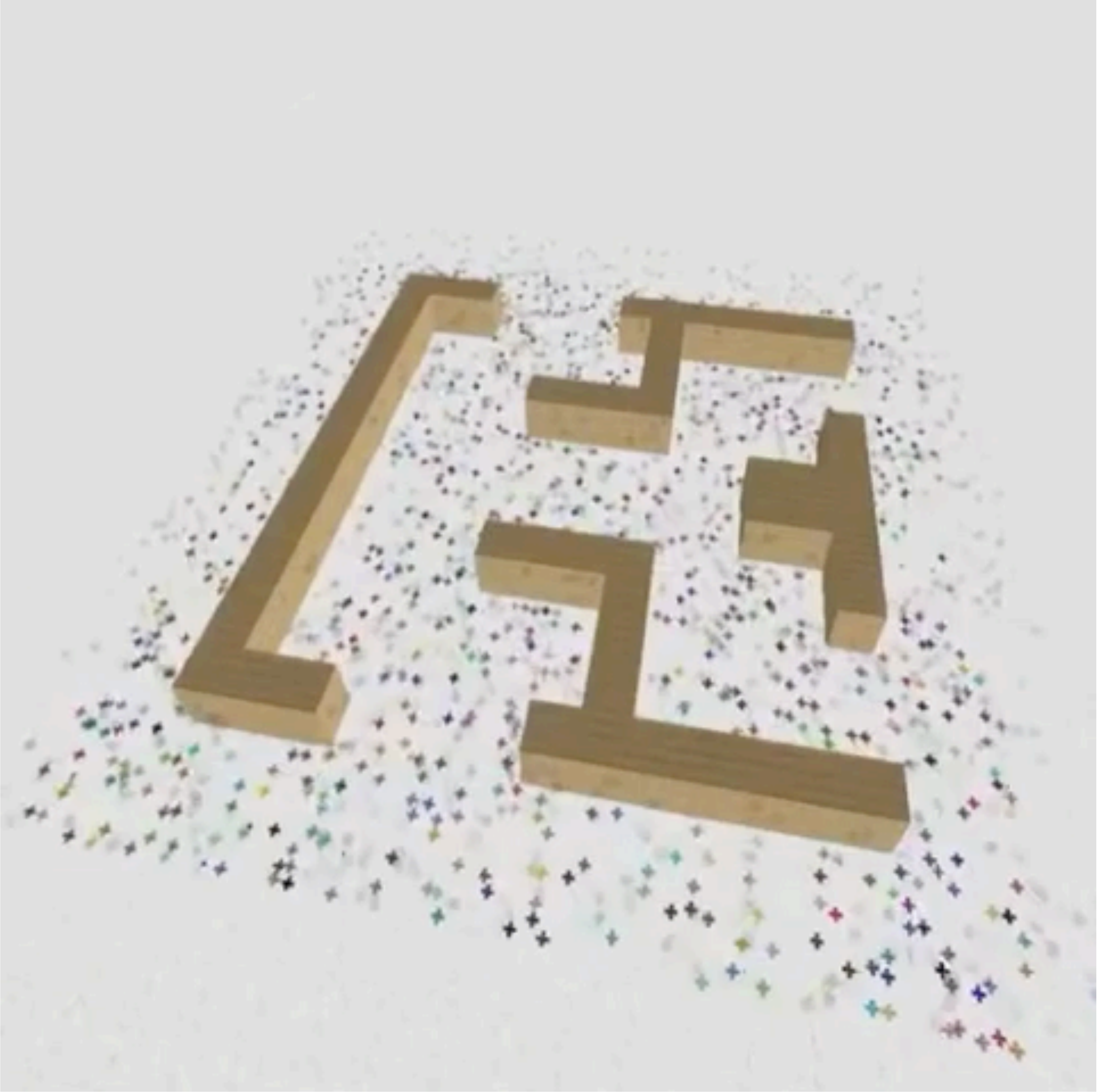
θ is the parameters of GP

Emam, Yousef, et al. "Safe model-based reinforcement learning using **robust control barrier functions**." *arXiv preprint arXiv:2110.05415* (2021).

Challenges

- Though CBF and CLF could provide a guarantee of safety, their limitations are also obvious. How to use CBF and CLF in the following situations are non-trivial:
  - Unknown/uncertain dynamics
  - Unknown/uncertain safe sets
  - High-dimensional state space
  - Non-stationary environments
- In addition, designing the CBF/CLF usually requires a lot of expert knowledge, which could be time-consuming and not scalable.

Deep NN-based CBF-CLF



Z Qin, K Zhang, Y Chen, J Chen, C Fan, Learning Safe Multi-Agent Control with Decentralized Neural Barrier Certificates, International Conference on Learning Representations (ICLR), 2021

Worthy Reading

- Chow, Yinlam, et al. "A Lyapunov-based approach to safe reinforcement learning." Advances in neural information processing systems 31 (2018).
- Choi, Jason J., et al. "Robust control barrier-value functions for safety-critical control." arXiv preprint arXiv:2104.02808 (2021).
- Z Qin, K Zhang, Y Chen, J Chen, C Fan, Learning Safe Multi-Agent Control with Decentralized Neural Barrier Certificates, International Conference on Learning Representations (ICLR), 2021
- Dawson, Charles, Sicun Gao, and Chuchu Fan. "Safe Control with Learned Certificates: A Survey of Neural Lyapunov, Barrier, and Contraction methods." arXiv preprint arXiv:2202.11762 (2022).