



Safe Control under Distribution Shifts with Robust Conformal Prediction

Lars Lindemann
Automatic Control Laboratory

Uncertainty quantification when the data changes

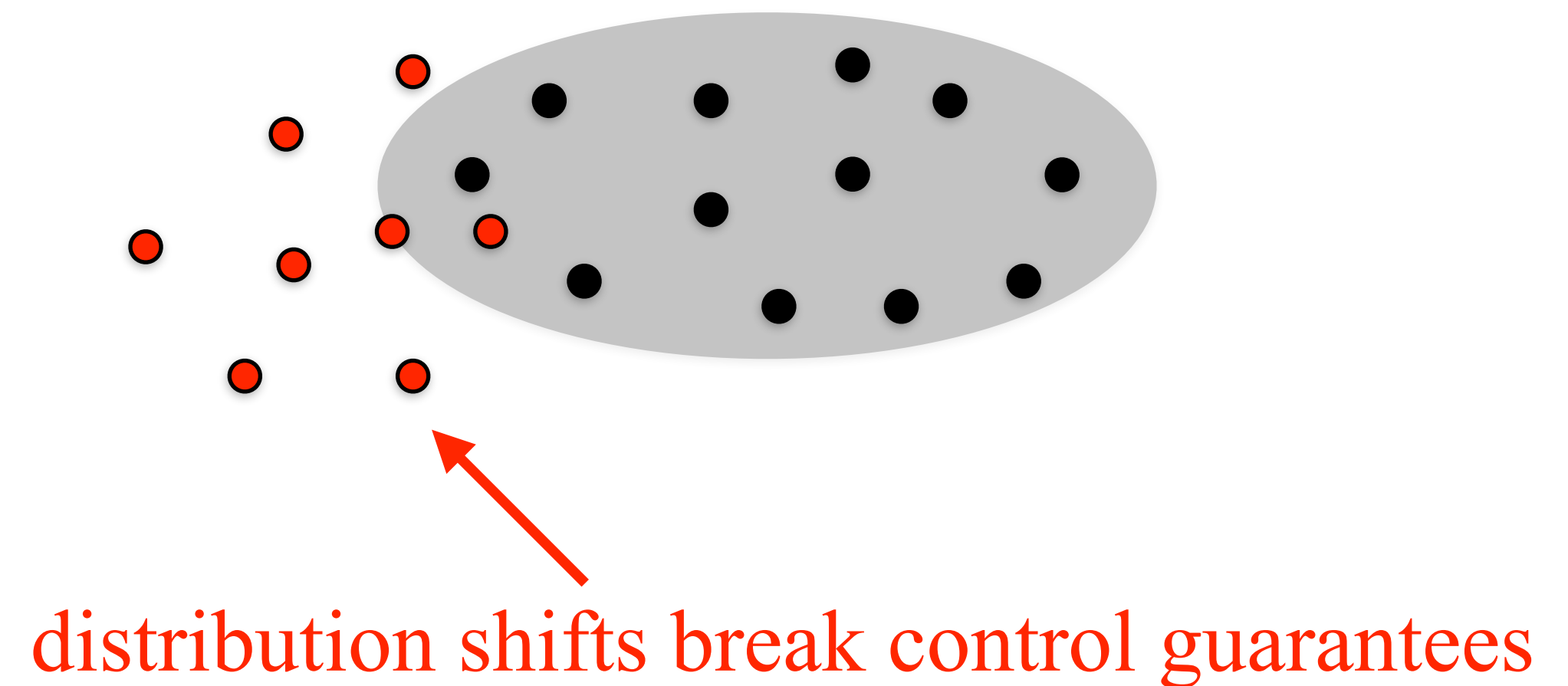
So far: uncertainty quantification and propagation is essential for control.

A **two-step procedure** for uncertainty quantification:

1. collect calibration data: $y^{(1)}, \dots, y^{(k)} \sim \mathcal{D}$
2. estimate uncertainty set: $C(y^{(1)}, \dots, y^{(k)})$

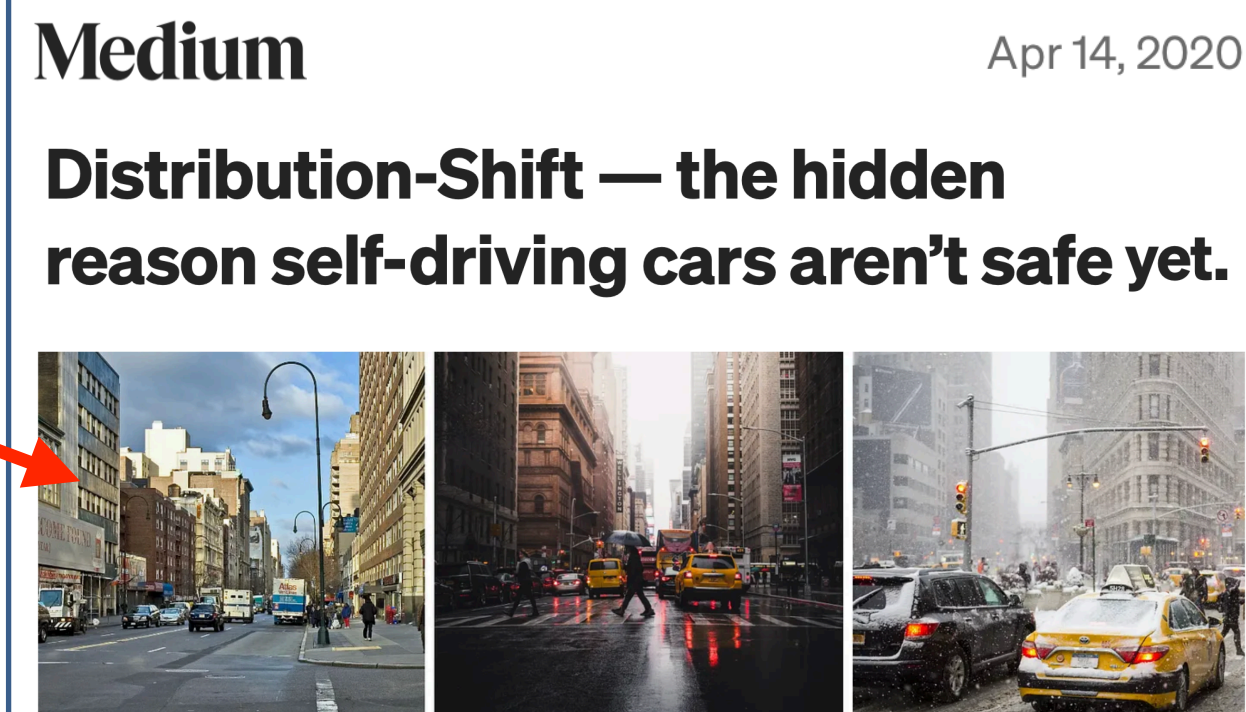
This procedure breaks if the **data distribution changes**:

3. data at test time: $y^{(0)} \sim \mathcal{D}_{test}$ for $\mathcal{D} \neq \mathcal{D}_{test}$



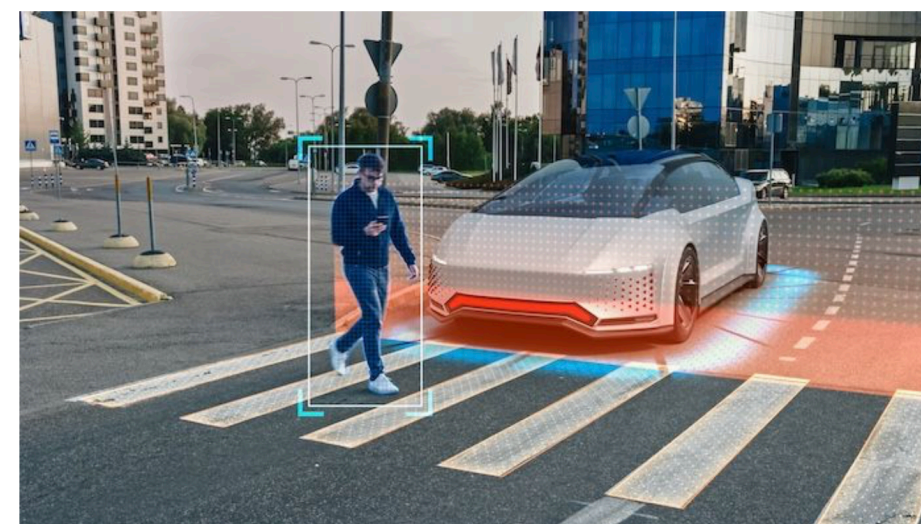
Two types of distribution shift:

uncontrollable
(exogenous factors)



New UK study reveals how pedestrians react to self-driving cars

October 30, 2025



controllable
(control-induced shift)

Safe Control in Dynamic Environments with Conformal Prediction

The safe control problem in dynamic environments

Compute **control inputs** so that the system avoids **dynamic agents** with a probability of at least $1 - \delta$.

Challenges:

- Uncertain and interactive agents
- Imperfect learning-enabled predictors

**uncertainty quantification
in control is challenging**

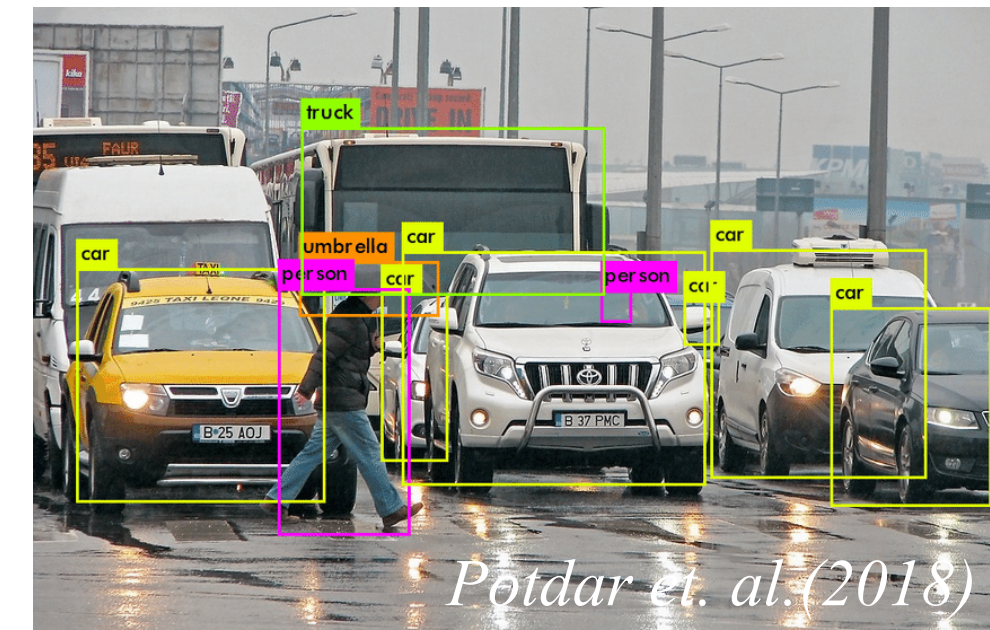
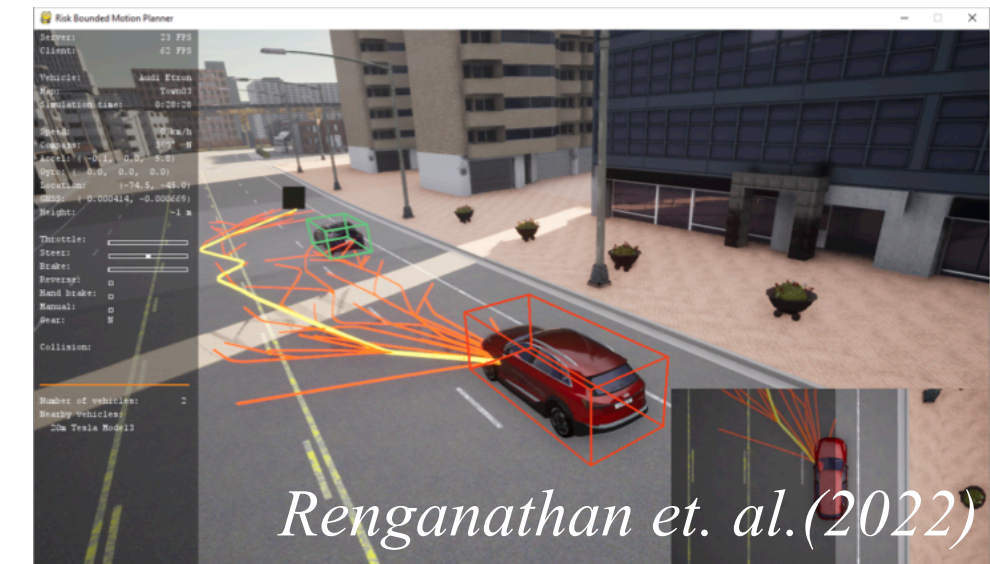
Existing work (references omitted):

- “Vanilla” motion planner (e.g., graph/sampling-based)
- Reactive control (e.g., navigation functions, DWA, CBF)
- Predictive interactive and non-interactive methods (e.g., deep RL, GP-based control)

ignores dynamic agents

**no predictions,
simple systems**

**no safety guarantees, or
simplifying assumptions**



Can we use **efficient statistical tools** for learning-enabled predictors and safe control?

Conformal Prediction

Conformal prediction in a nutshell

Uncertainty quantification under minimal assumptions:

- Distribution-free
- No assumptions on the predictor needed

$$(u, z) \sim \mathcal{D}$$

$$\hat{z} = \mu(u)$$

no assumptions needed

no assumptions needed

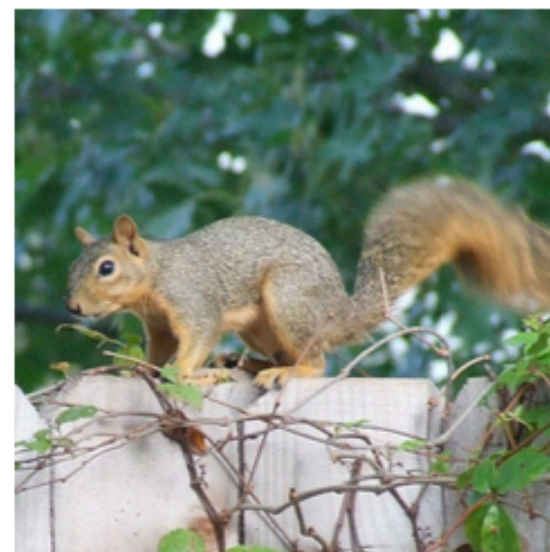
Assumption: Availability of i.i.d. calibration data $\{(u^{(i)}, z^{(i)})\}_{i=1}^k$

Goal^{1,2}:

For a failure probability of $\delta \in (0, 1)$, we want to obtain a prediction region C s.t.

$$\text{Prob}(z \in C(u)) \geq 1 - \delta$$

Example²:



$\left\{ \begin{array}{l} \text{fox} \quad \text{squirrel} \\ 0.99 \end{array} \right\}$



$\left\{ \begin{array}{l} \text{fox} \quad \text{gray} \quad \text{bucket} \quad \text{rain} \\ \text{squirrel}, \text{fox}, \text{barrel} \\ 0.82 \quad 0.03 \quad 0.02 \quad 0.02 \end{array} \right\}$



$\left\{ \begin{array}{l} \text{marmot}, \text{squirrel}, \text{mink}, \text{weasel}, \text{beaver}, \text{polecat} \\ 0.30 \quad 0.22 \quad 0.18 \quad 0.16 \quad 0.03 \quad 0.01 \end{array} \right\}$

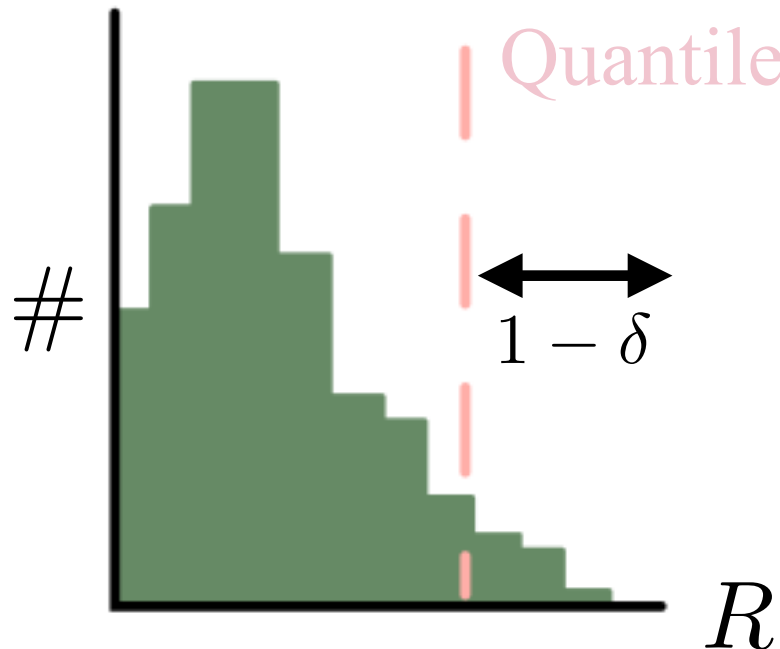
¹Vovk and Shafer, “A Tutorial on Conformal Prediction”, *JMLR*, 2008.

²Angelopoulos, “A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification”, *subm.*, 2021.

Conformal prediction in a nutshell

test data calibration data nonconformity score (e.g., prediction error)

Quantile lemma: Let $R^{(0)}, R^{(1)}, \dots, R^{(k)}$ be $k + 1$ i.i.d. random variables. It holds that

$$\text{Prob}(R^{(0)} \leq \text{Quantile}_{1-\delta}(R^{(1)}, \dots, R^{(k)}, \infty)) \geq 1 - \delta$$


The $p := \lceil (k + 1)(1 - \delta) \rceil$ -th smallest nonconformity score $\rightarrow$ computationally efficient

- Example (regression):** the nonconformity score $R^{(i)} := |z^{(i)} - \mu(u^{(i)})|$ results in

test data calibration data

$$\text{Prob}(|z - \mu(u)| \leq \text{Quantile}_{1-\delta}(R^{(1)}, \dots, R^{(k)}, \infty)) \geq 1 - \delta$$

$$\Leftrightarrow \text{Prob}(z \in [\mu(u) - \text{Quantile}_{1-\delta}, \mu(u) + \text{Quantile}_{1-\delta}]) \geq 1 - \delta$$

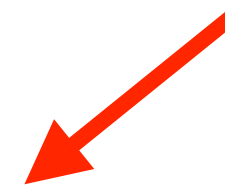
prediction region

- Simple, general, and efficient** with the right nonconformity score¹:



¹Tumu, Cleaveland, Mangharam, Pappas, and Lindemann, “Multi-Modal Conformal Prediction Regions by Optimizing Convex Shape Templates”, *L4DC*, 2024.

no distribution shift

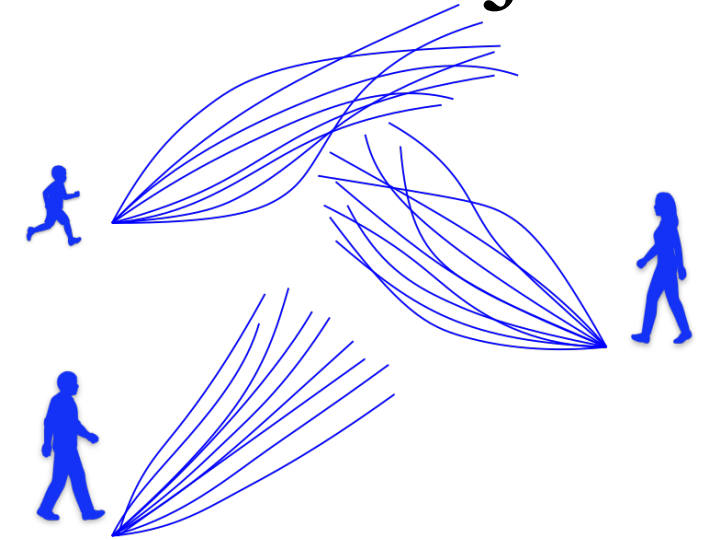


Safe Control in Dynamic Environments

Overview

no distribution shift for now

Offline trajectory dataset



$$(y_0, y_1, \dots) \sim \mathcal{D}$$

Goal: Design a probabilistically safe controller.

Conformal
Prediction

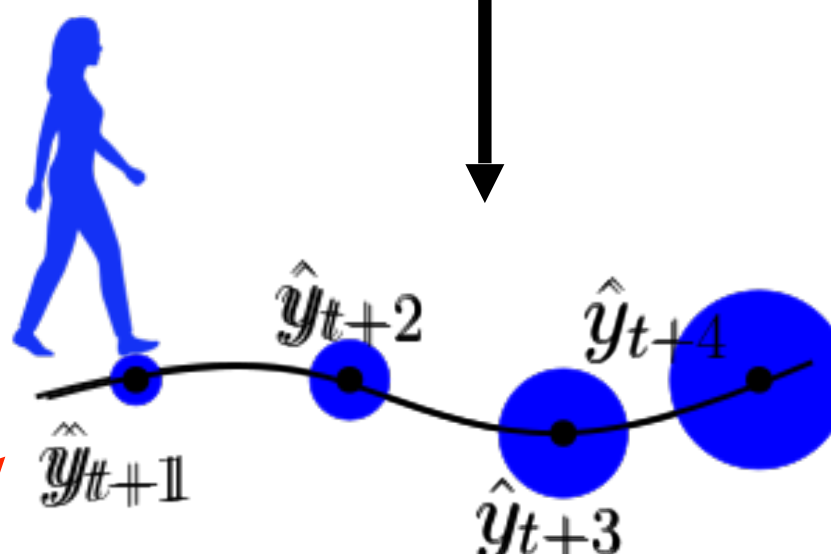
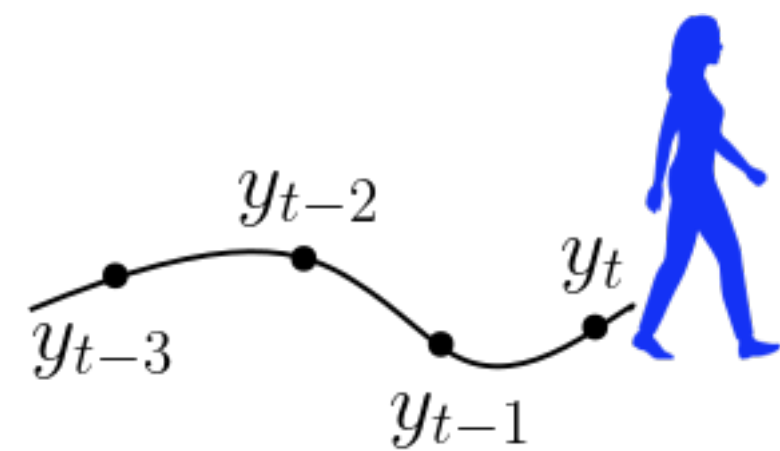
Dynamics

$$x_{t+1} = f(x_t, u_t)$$

Online observations

learned, e.g., RNN

Trajectory
predictor



Prediction region

$$\text{Prob}(y \in \bullet) \geq 1 - \delta$$

safety constraint

Model Predictive Control
 $\text{Prob}(c(x, y) \geq 0) \geq 1 - \delta$



How good are these estimates?

Key features:

- Computationally lightweight
- Probabilistic safety guarantees
- Distribution shift
- Temporal logic constraints

Uncertainty representation in dynamic environments

- Define the nonconformity score $R^{(i)} := \max_{\tau \in \{t+1, \dots, T\}} \alpha_{\tau} \|y_{\tau}^{(i)} - \hat{y}_{\tau}^{(i)}\|$

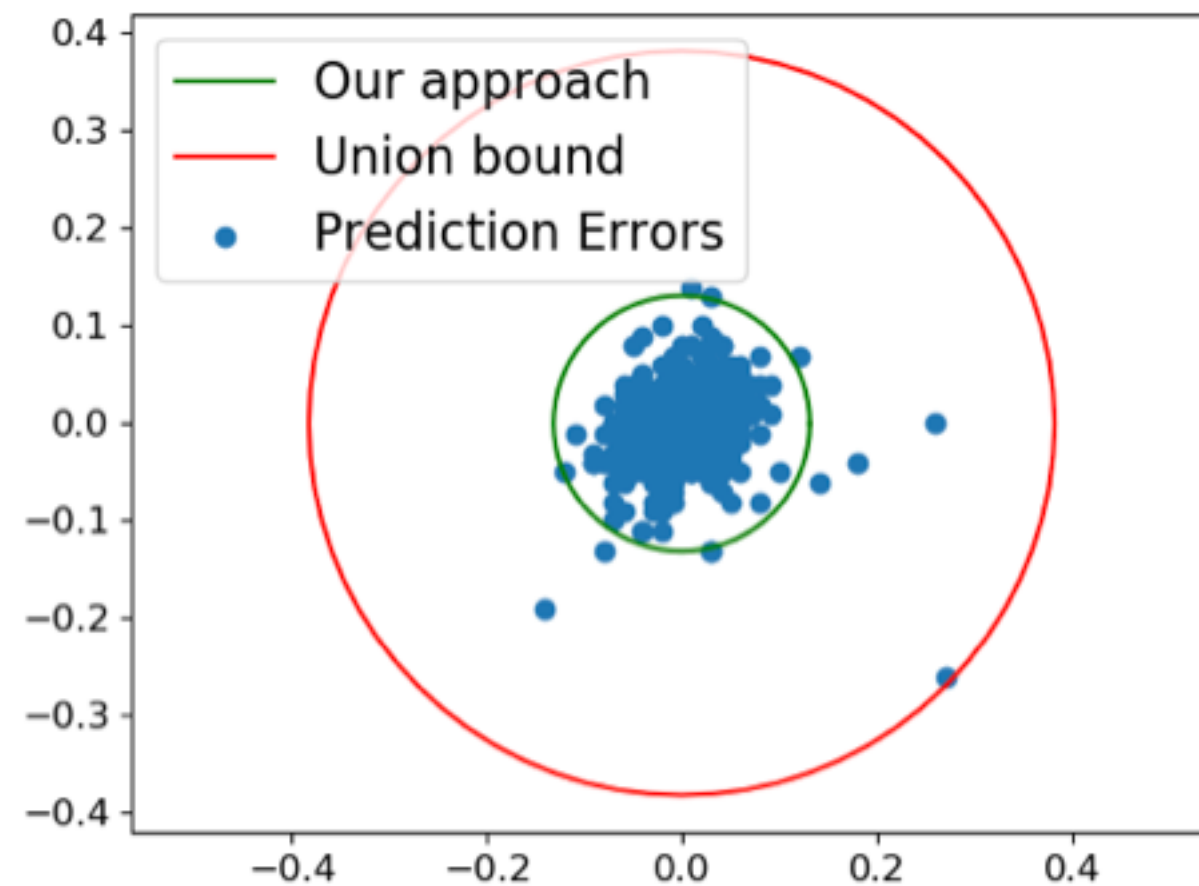
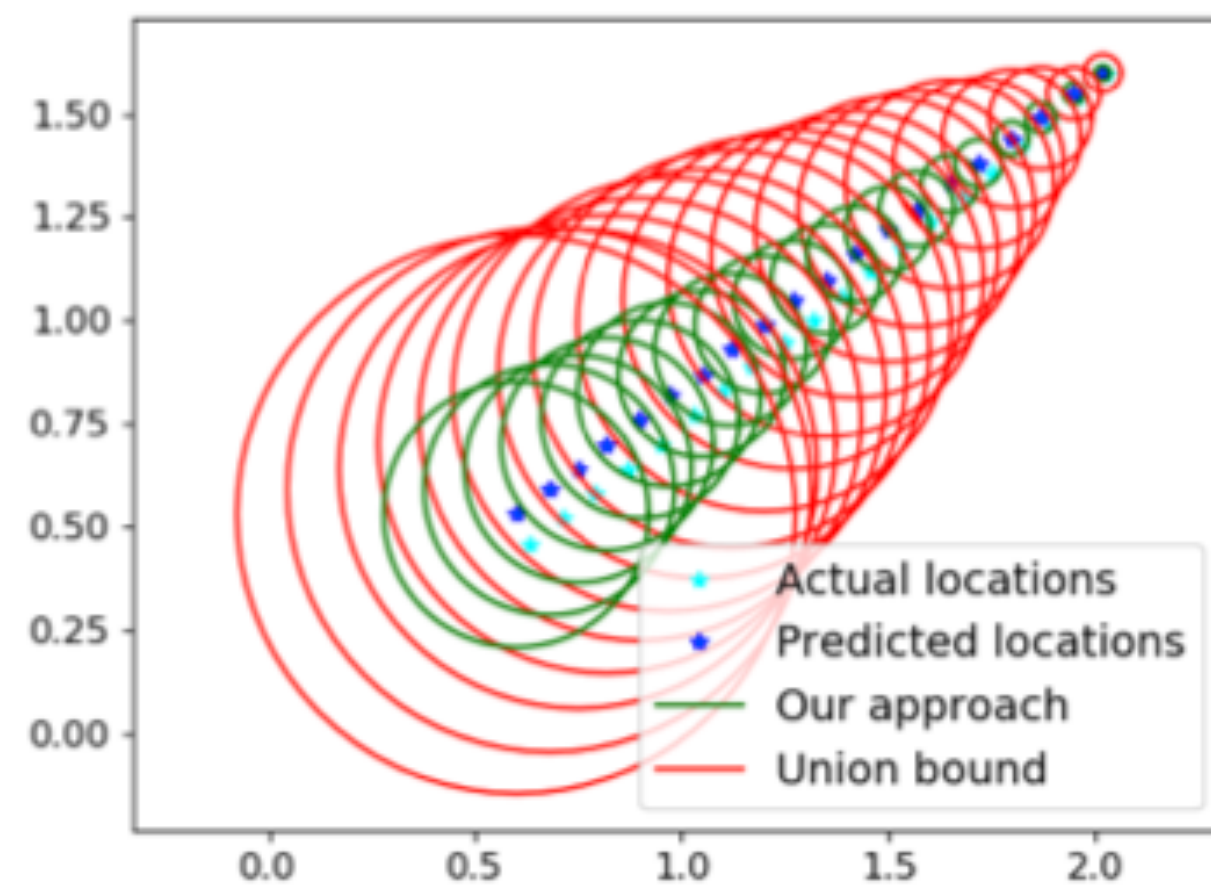
prediction error

normalization constant (see reference¹)

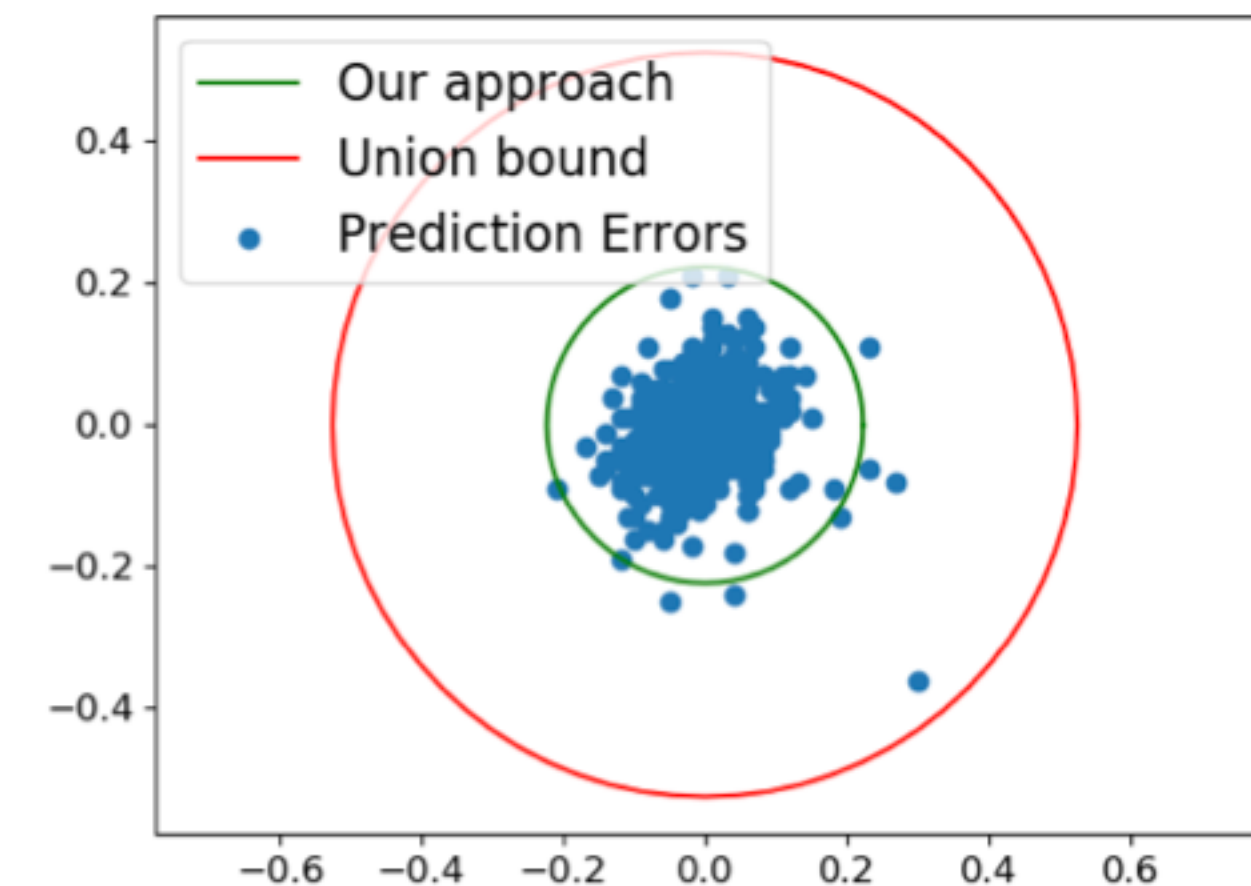
and apply conformal prediction $C = \text{Quantile}_{1-\delta}(R^{(1)}, \dots, R^{(k)}, \infty)$

Multi-step prediction regions: $\text{Prob}(\|y_{\tau} - \hat{y}_{\tau}\| \leq C/\alpha_{\tau}, \forall \tau \in \{t+1, \dots, T\}) \geq 1 - \delta$

Example:



1.25 second ahead prediction error



1.8 second ahead prediction error

¹Cleaveland, Lee, Pappas, and Lindemann, “Conformal Prediction for Time Series using Linear Complementarity Programming”, *AAAI*, 2024.

Uncertainty-aware predictive control

Model predictive control¹: $\min_{(u_t, \dots, u_{T-1})} J(x, u)$

s.t. $x_{\tau+1} = f(x_\tau, u_\tau), \quad \tau \in \{t, \dots, T-1\}$

enforced via $c(x_\tau, \hat{y}_\tau) \geq LC_\tau$

robust safety constraint

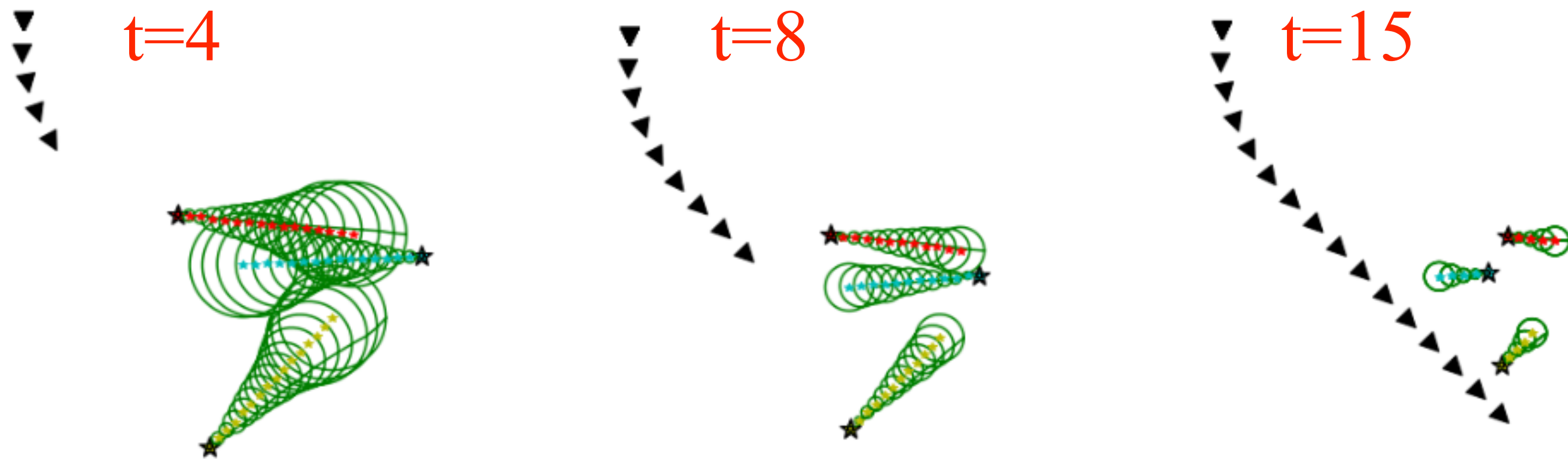
$\inf_{y_\tau \in \mathcal{B}_\tau} c(x_\tau, y_\tau) \geq 0, \quad \tau \in \{t+1, \dots, t+H\}$

$\mathcal{B}_\tau := \{y \text{ s.t. } \|y - \hat{y}_\tau\| \leq C_\tau\}$

$u_\tau \in \mathcal{U}, x_{\tau+1} \in \mathcal{X}, \quad \tau \in \{t, \dots, T-1\}$

If the MPC is recursively feasible^{2,3}, we find a safe controller with a probability of at least $1 - \delta$.

Toy example:



Animated:

¹Lindemann, Cleaveland, Shim, and Pappas, “Safe Planning in Dynamic Environments using Conformal Prediction”, *RA-L*, 2023.

²Stamouli, Lindemann, and Pappas, “Recursively Feasible Shrinking-Horizon MPC in Dynamic Environments with Conformal Prediction Guarantees”, *L4DC*, 2024.

³Yu, Zhao, Yin, and Lindemann, “STL Control Synthesis among Uncontrollable Agents with Conformal Prediction”, *Automatica*, 2026.

Case Study: CARLA

Results:

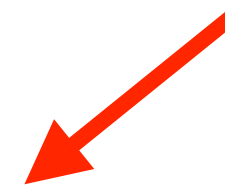
- 99.985 % correct one-step ahead predictions (at least 99.83 % expected)
- 99/100 test runs without constraint violation (at most 5 expected)



Comparison with Gaussian Process:

- Largely undercovers

controllable distribution shift



Safe Control in Interactive Environments

Interaction-induced distribution shifts

What if interaction between the control system and the dynamic agents changes the distribution?

- **Example:** self-driving vehicles among pedestrians

no safety guarantees



- **Chicken-and-egg problem:**

Dynamics
 $x_{t+1} = f(x_t, u_t)$



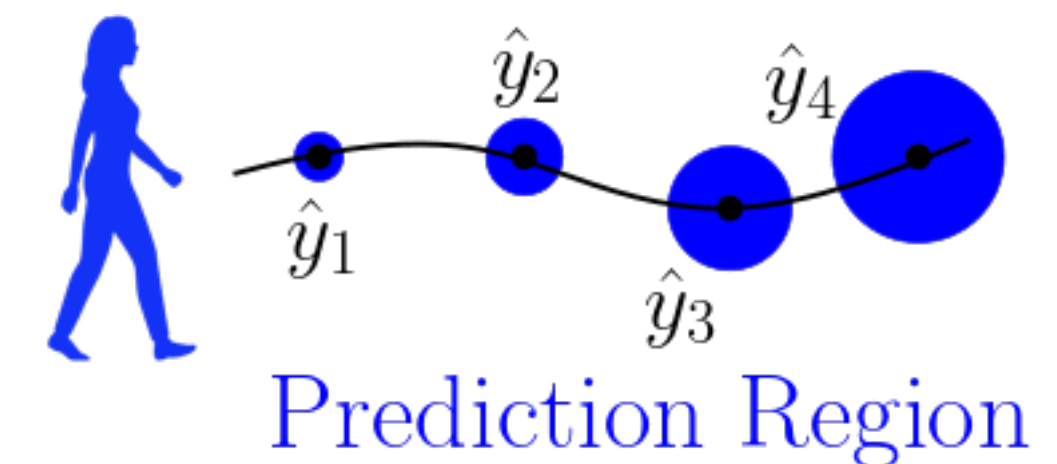
$$x = (x_1, x_2, \dots)$$

a function of x

$$u = \pi(\text{Pred.Region}(x))$$

Distribution
 $\mathcal{D}(x)$

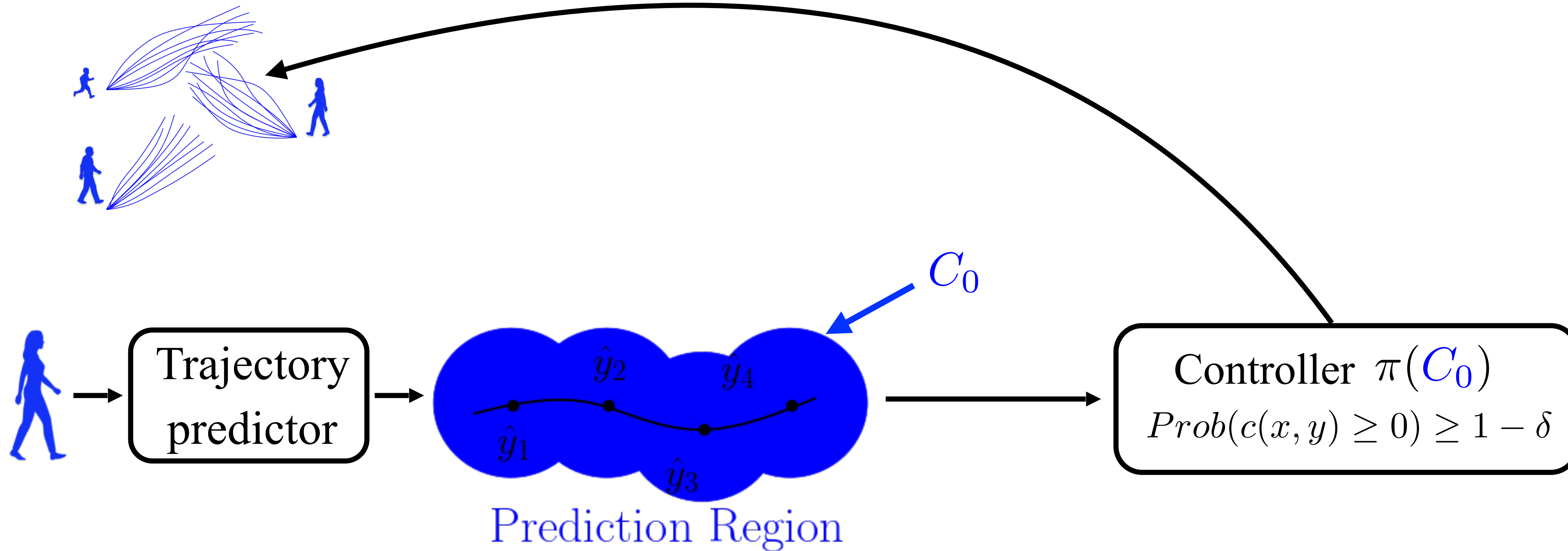
distribution shift



Can we iteratively update our controller to break circularity?

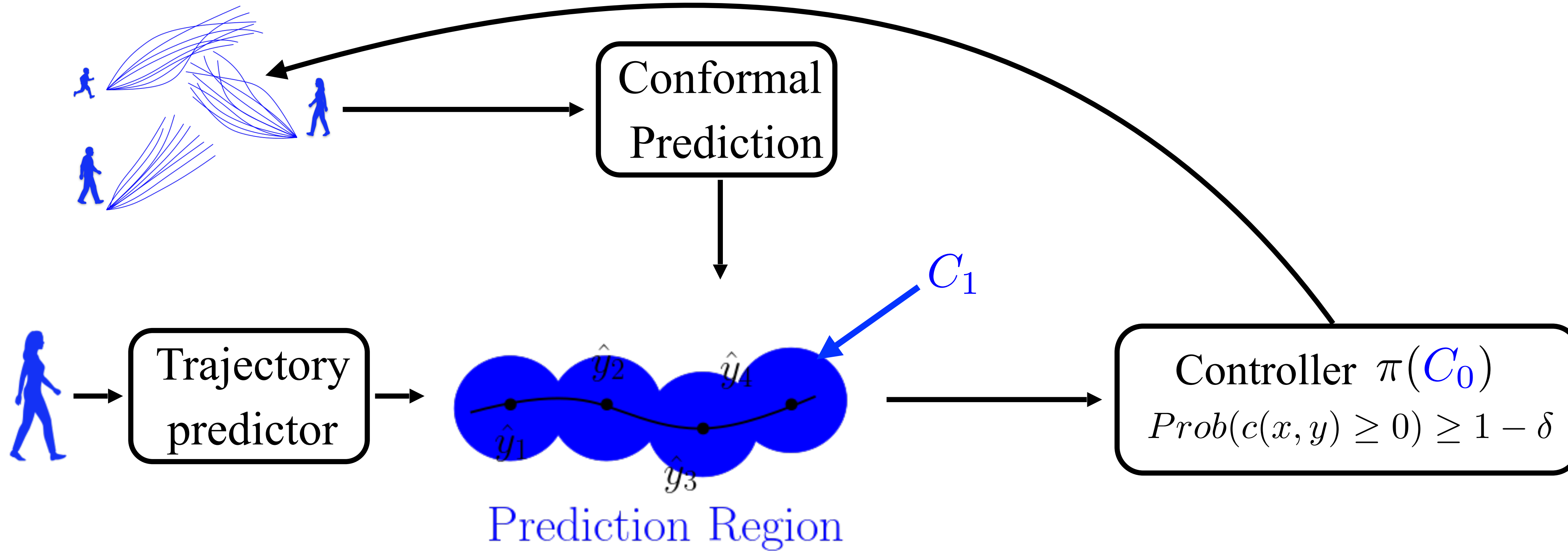
Iteratively breaking circularity

Collect data under $\pi(C_0)$



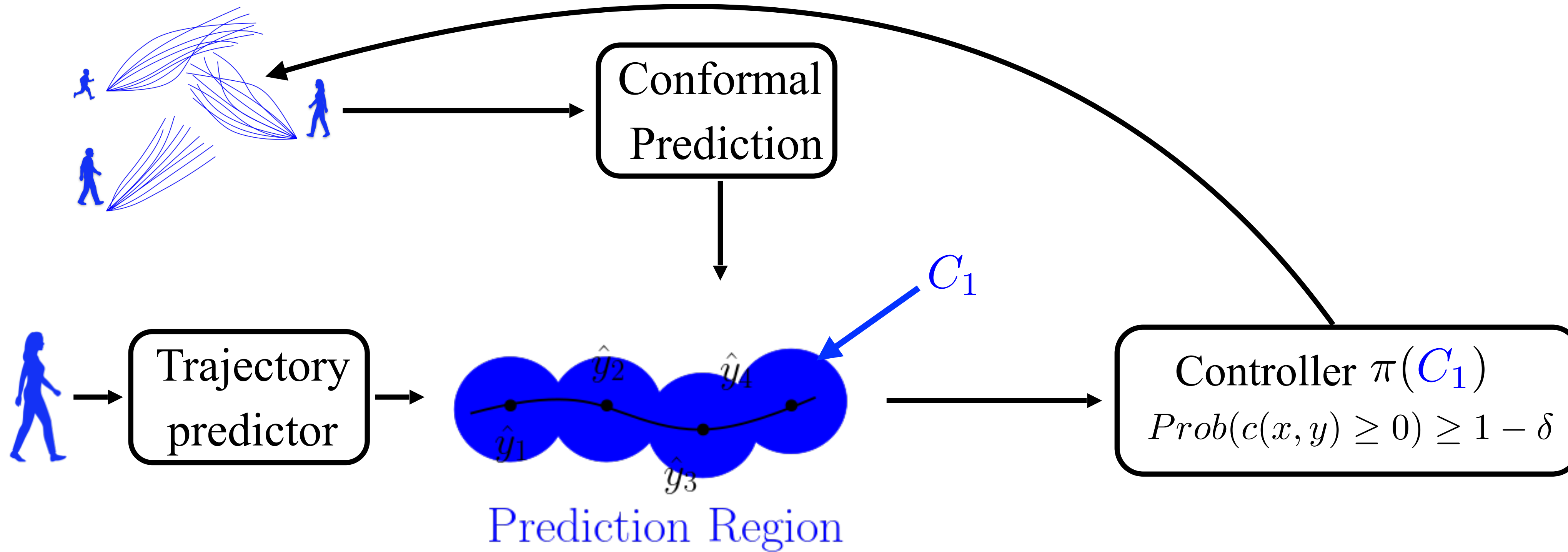
Iteratively breaking circularity

Collect data under $\pi(C_0)$



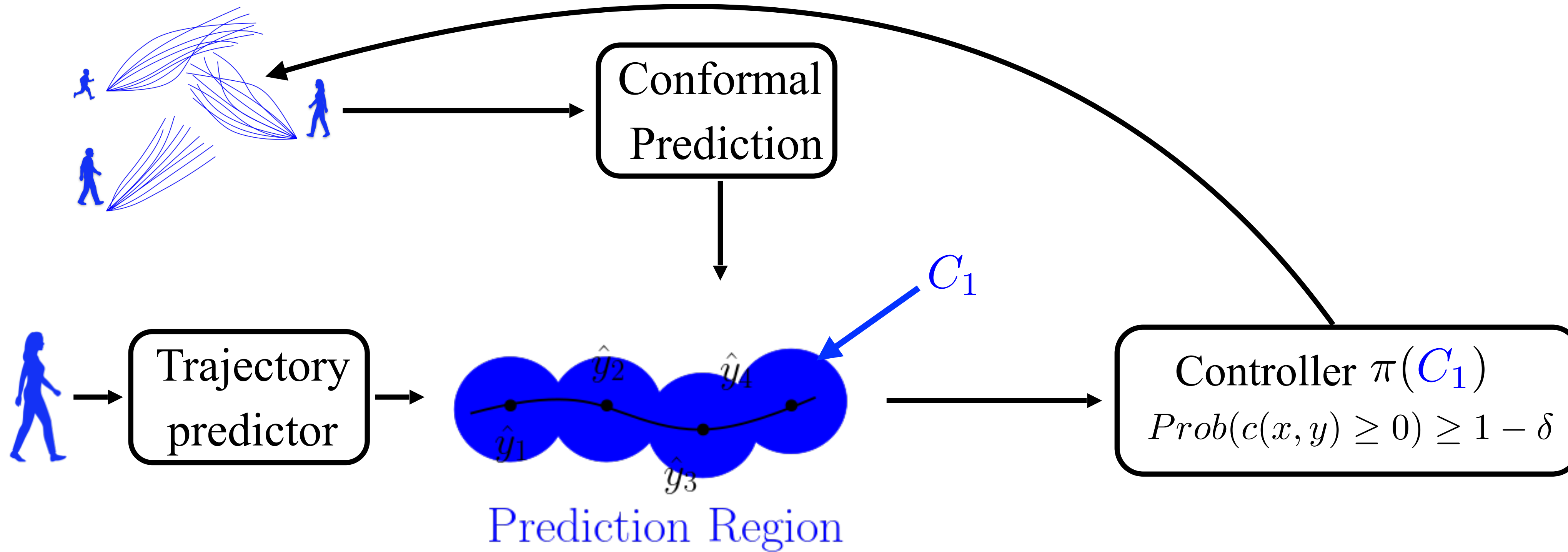
Iteratively breaking circularity

Collect data under $\pi(C_0)$



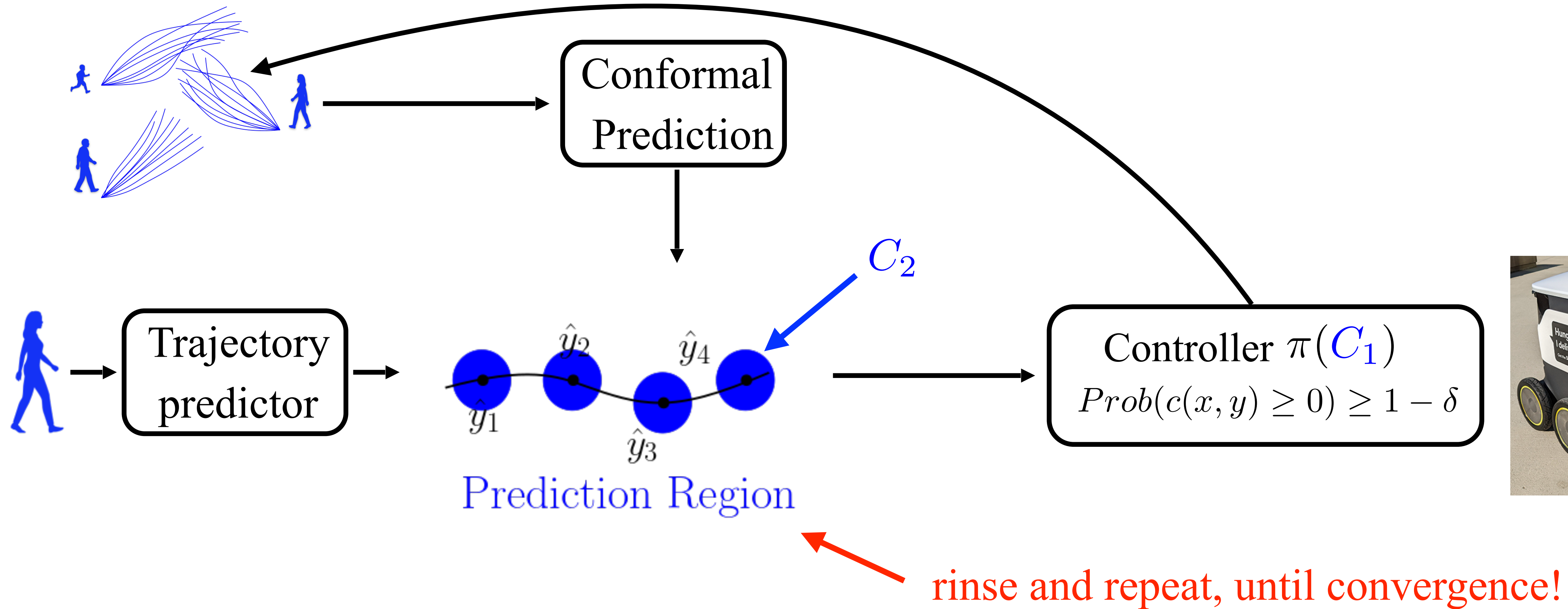
Iteratively breaking circularity

Collect data under $\pi(C_1)$



Iteratively breaking circularity

Collect data under $\pi(C_1)$



Safe iterative control design

- **Iterative controller updates:** $\pi(C_j) := \underset{(u_0, \dots, u_{T-1})}{\operatorname{argmin}} J(x, u)$

episode j

$$\text{s.t. } x_{\tau+1} = f(x_\tau, u_\tau),$$

iterative safety constraint

$$\inf_{y_\tau \in \mathcal{B}_{j,\tau}} c(x_\tau, y_\tau) \geq 0$$

Controller-agnostic: any controller $\pi(C_j)$ could be used!

$$\mathcal{B}_{j,\tau} := \{y \text{ s.t. } \|y - \hat{y}_\tau\| \leq C_j\}$$

- **Design goals:** update the radius C_j such that:

trajectories under controller $\pi(C_j)$

- (1) $\operatorname{Prob} \{ c(x_j, y_j) \geq 0 \} \geq 1 - \delta$ for all j , ← episodic safety
- (2) $J_{j+1} \leq J_j - \epsilon_j$ for all j , ← improved performance
- (3) $C_{j+1} \approx C_j$ as $j \rightarrow \infty$ ← convergence

Can we use **robust conformal prediction** to update the radius C_j safely?

Robust conformal prediction for non i.i.d. data

Let $R^{(0)}, R^{(1)}, \dots, R^{(k)}$ be $k + 1$ i.i.d. random variables. It holds that

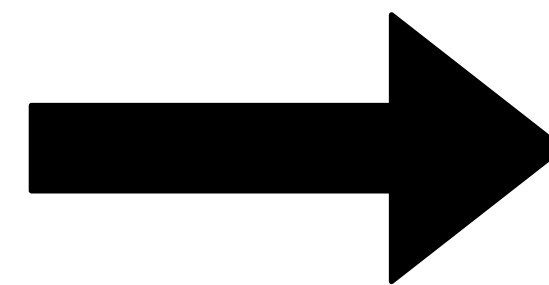
$$\text{Prob}(R^{(0)} \leq \text{Quantile}_{1-\delta}(R^{(1)}, \dots, R^{(k)}, \infty)) \geq 1 - \delta$$

What if the data is dependent or not identically distributed?

Conformal prediction in the non i.i.d. case:

- Adversarially robust conformal prediction
- Distributionally robust conformal prediction
- Adaptive conformal prediction
- Weighted conformal prediction
- ...

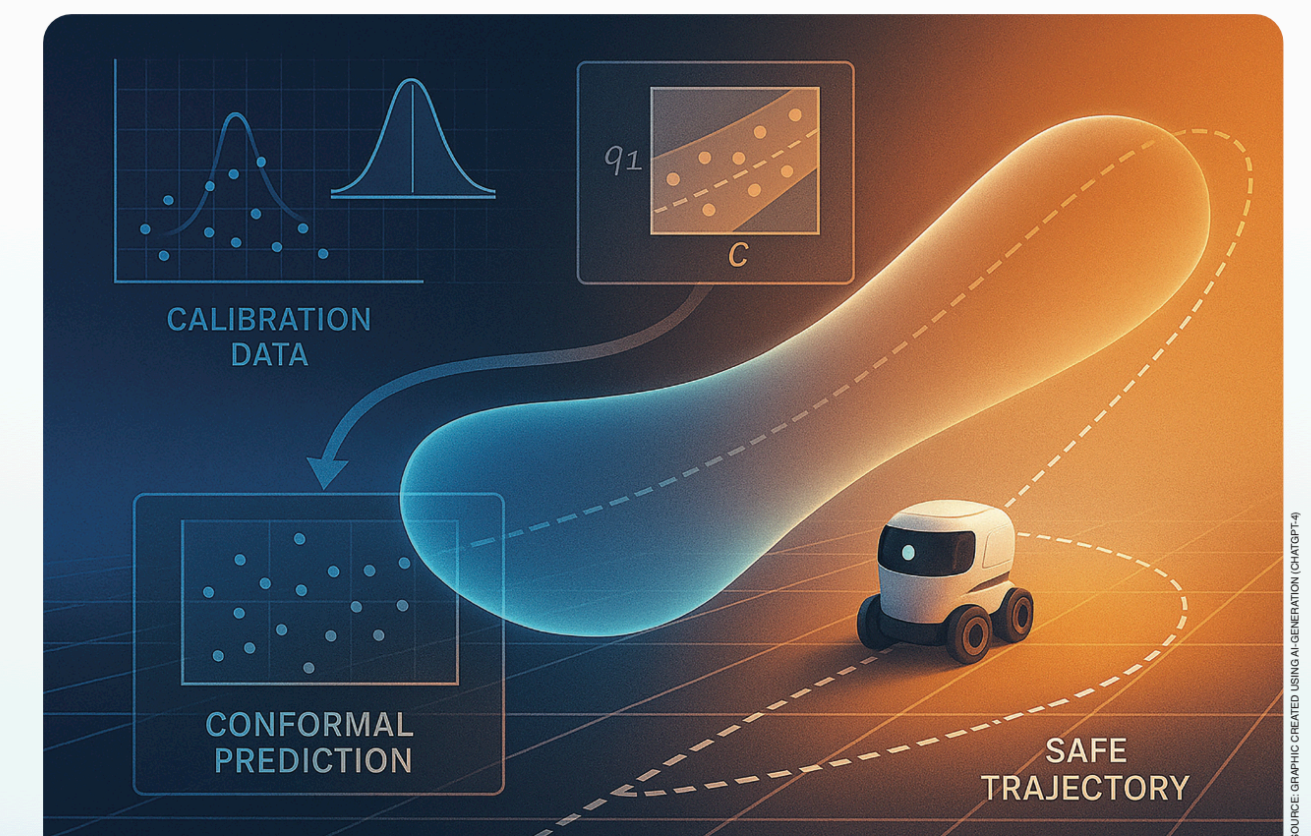
summary in



Formal Verification and Control With Conformal Prediction

PRACTICAL SAFETY GUARANTEES FOR AUTONOMOUS SYSTEMS

LARS LINDEMANN¹, YIQI ZHAO², XINYI YU², GEORGE J. PAPPAS², and JYOTIRMOY V. DESHMUKH²



Confidence meets control — verifiably safe systems with conformal prediction.

Today!

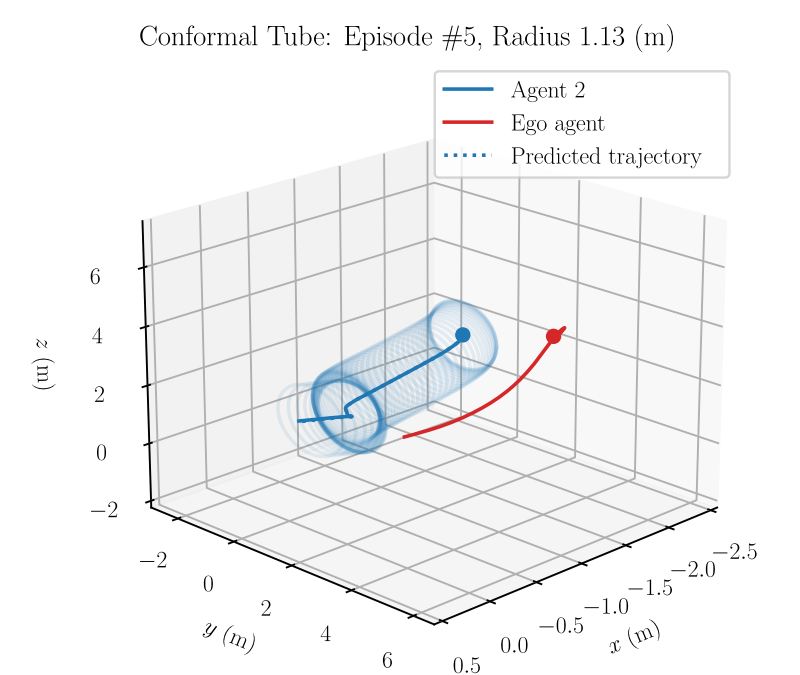
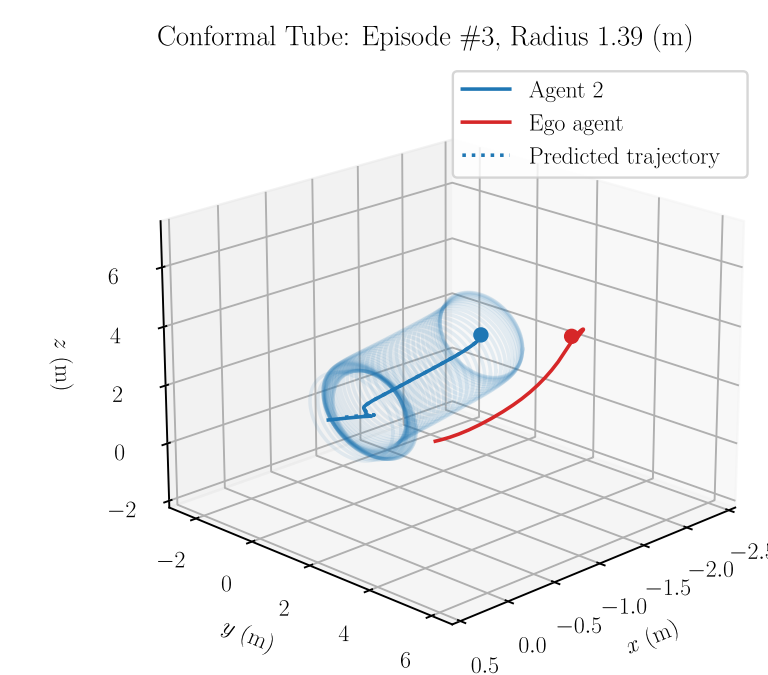
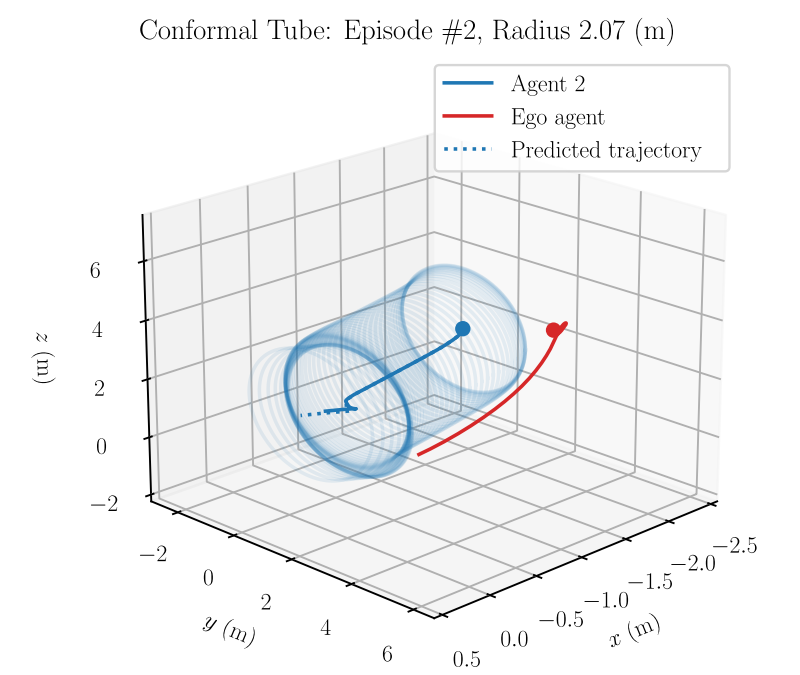
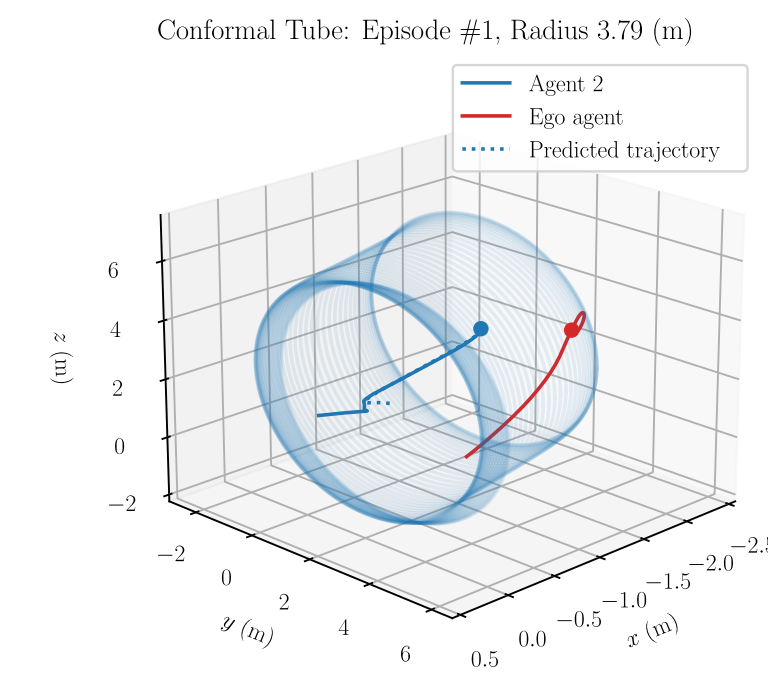
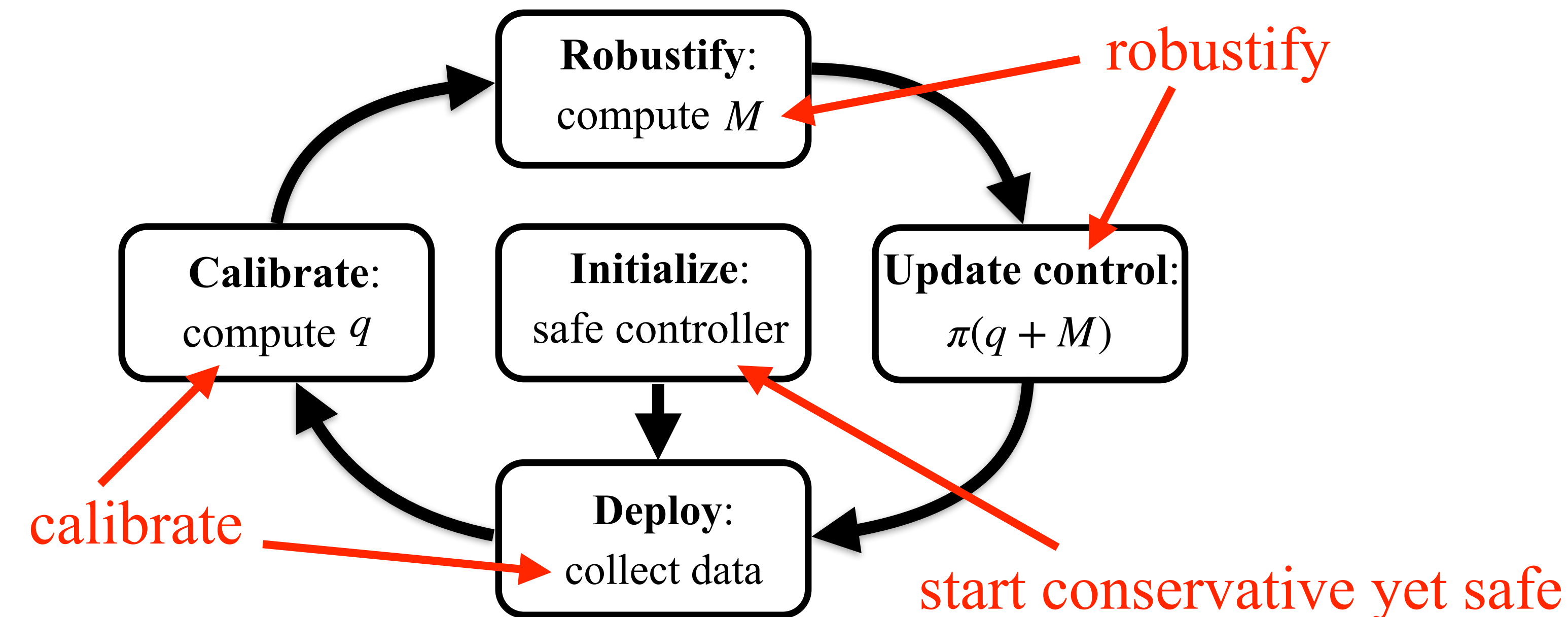
Safe control with robust conformal prediction

- **Adversarial conformal prediction¹:** Let $\sup_{\Delta} R(y + \Delta) \leq R(y) + M$ for all trajectories y .

Δ

It holds that $\text{Prob}(R(y^{(0)} + \Delta) \leq q + M, \forall \Delta) \geq 1 - \delta$ with $q := \text{Quantile}_{1-\delta}(R(y^{(1)}), \dots, R(y^{(k)}))$

- **Our approach²:** iterative robustification

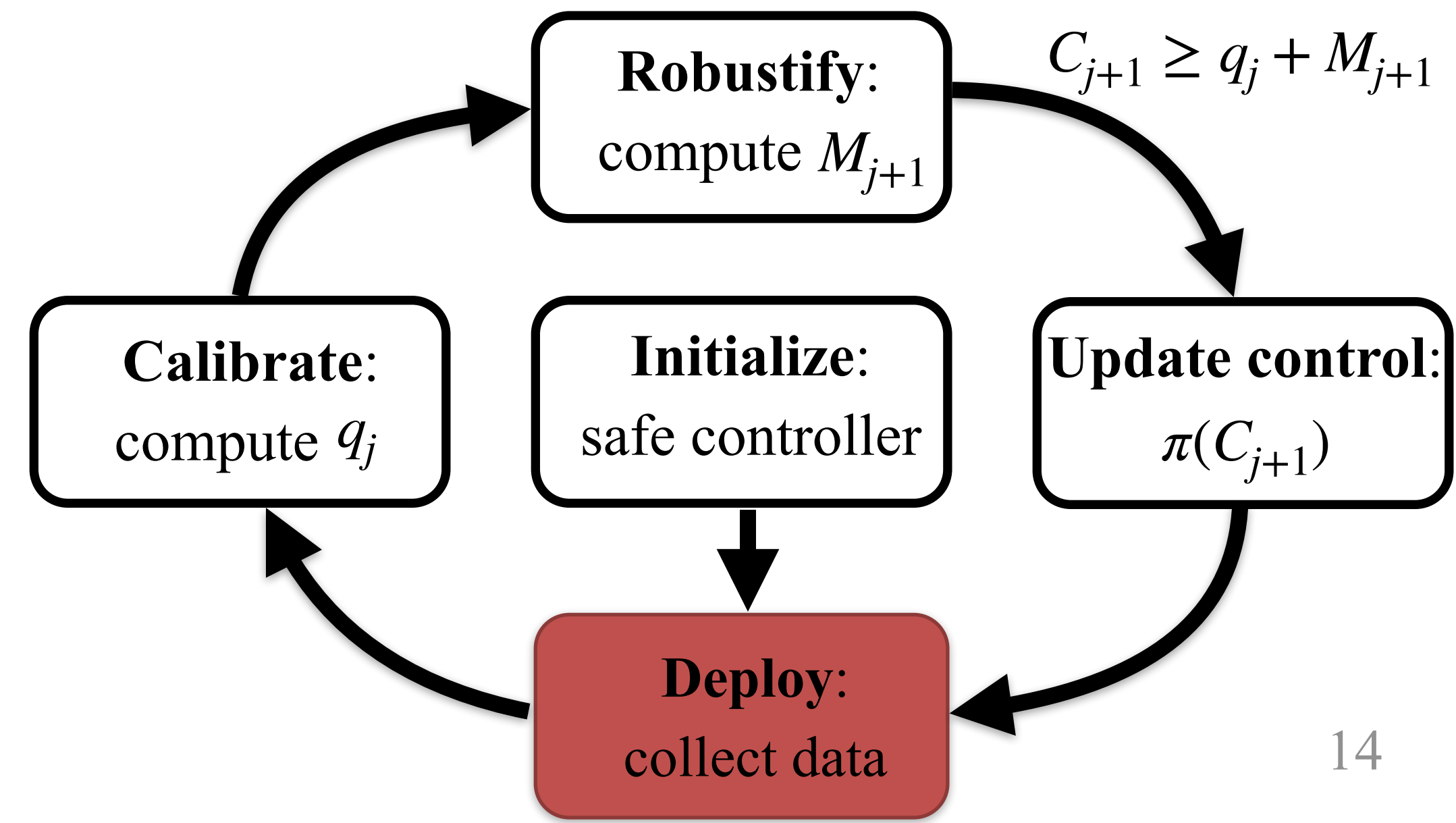


¹Gendler, Weng, Daniel, Romano, “Adversarially Robust Conformal Prediction”, ICLR, 2022.

²Mirzaeedodangeh, Shekhtman, Matni, **Lindemann**, “Safe Planning in Interactive Environments via Iterative Policy Updates and Adversarially Robust Conformal Prediction”, *L4DC*, 2026. 13

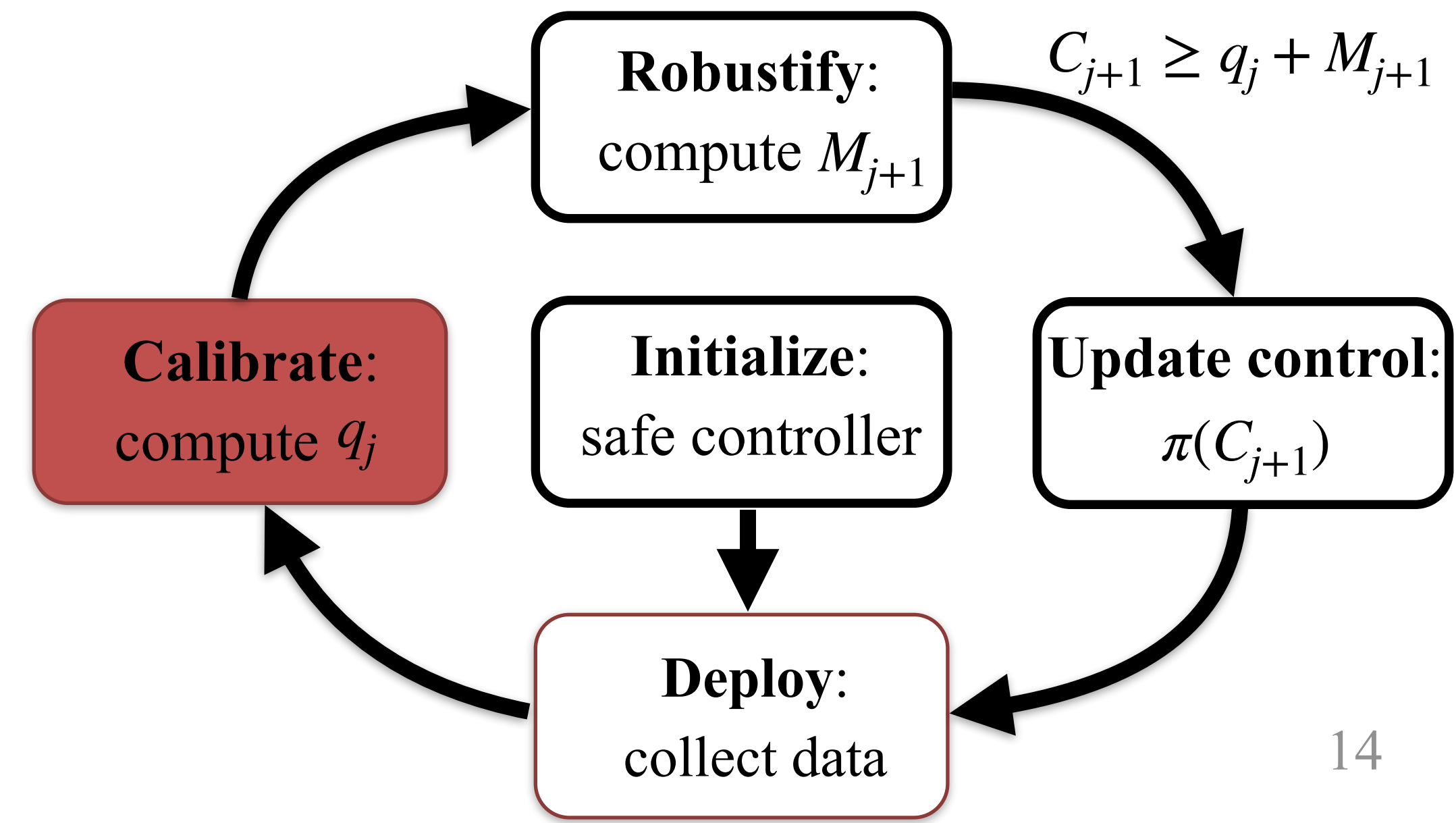
Algorithm for controller updates

- Collect new calibration data $y_j^{(1)}, \dots, y_j^{(k)} \sim \mathcal{D}(x_j)$



Algorithm for controller updates

- Collect new calibration data $y_j^{(1)}, \dots, y_j^{(k)} \sim \mathcal{D}(x_j)$

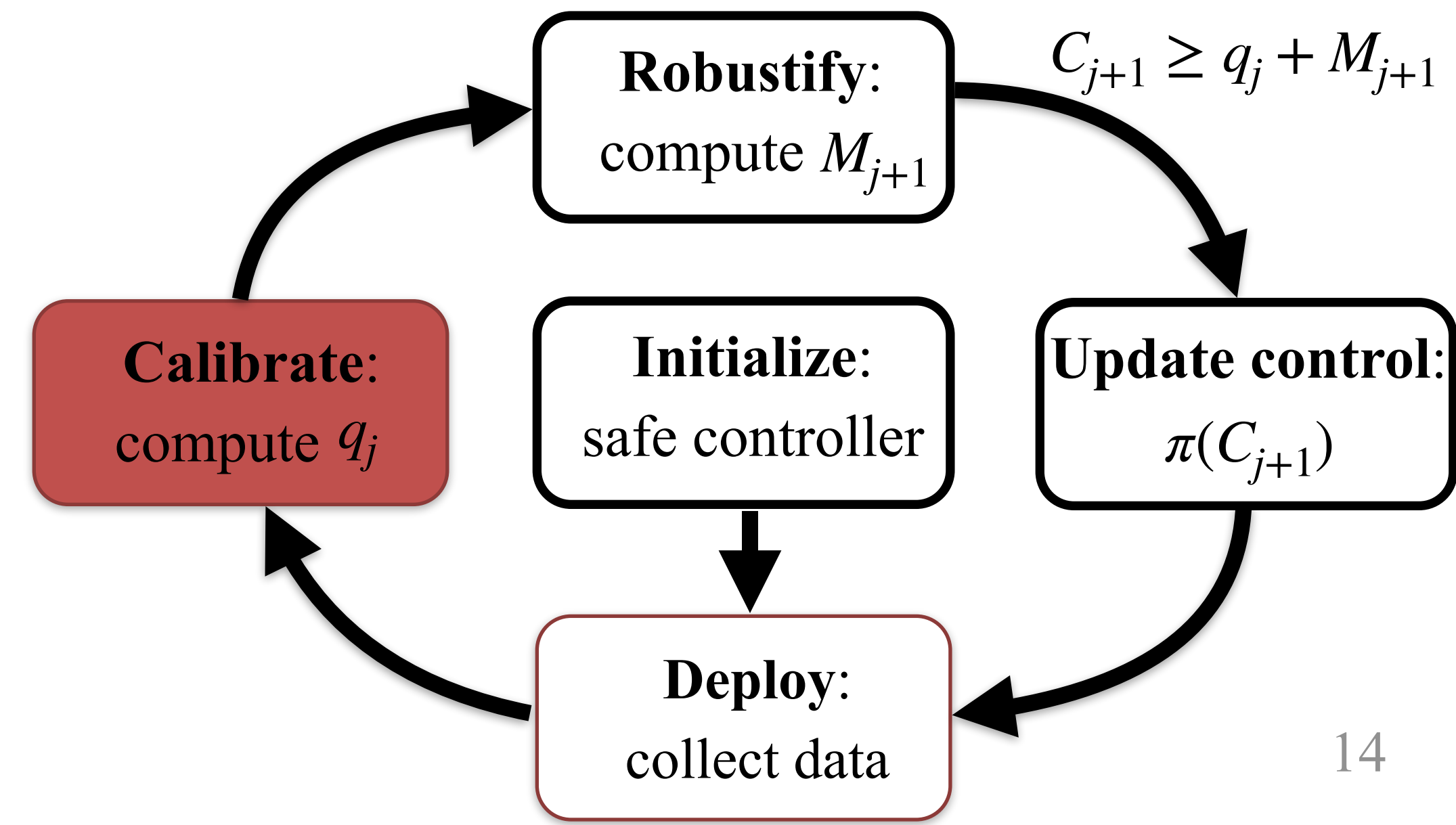


Algorithm for controller updates

- Collect new calibration data $y_j^{(1)}, \dots, y_j^{(k)} \sim \mathcal{D}(x_j)$
- Apply conformal prediction $q_j = \text{Quantile}_{1-\delta}(R_j^{(1)}, \dots, R_j^{(k)}, \infty)$

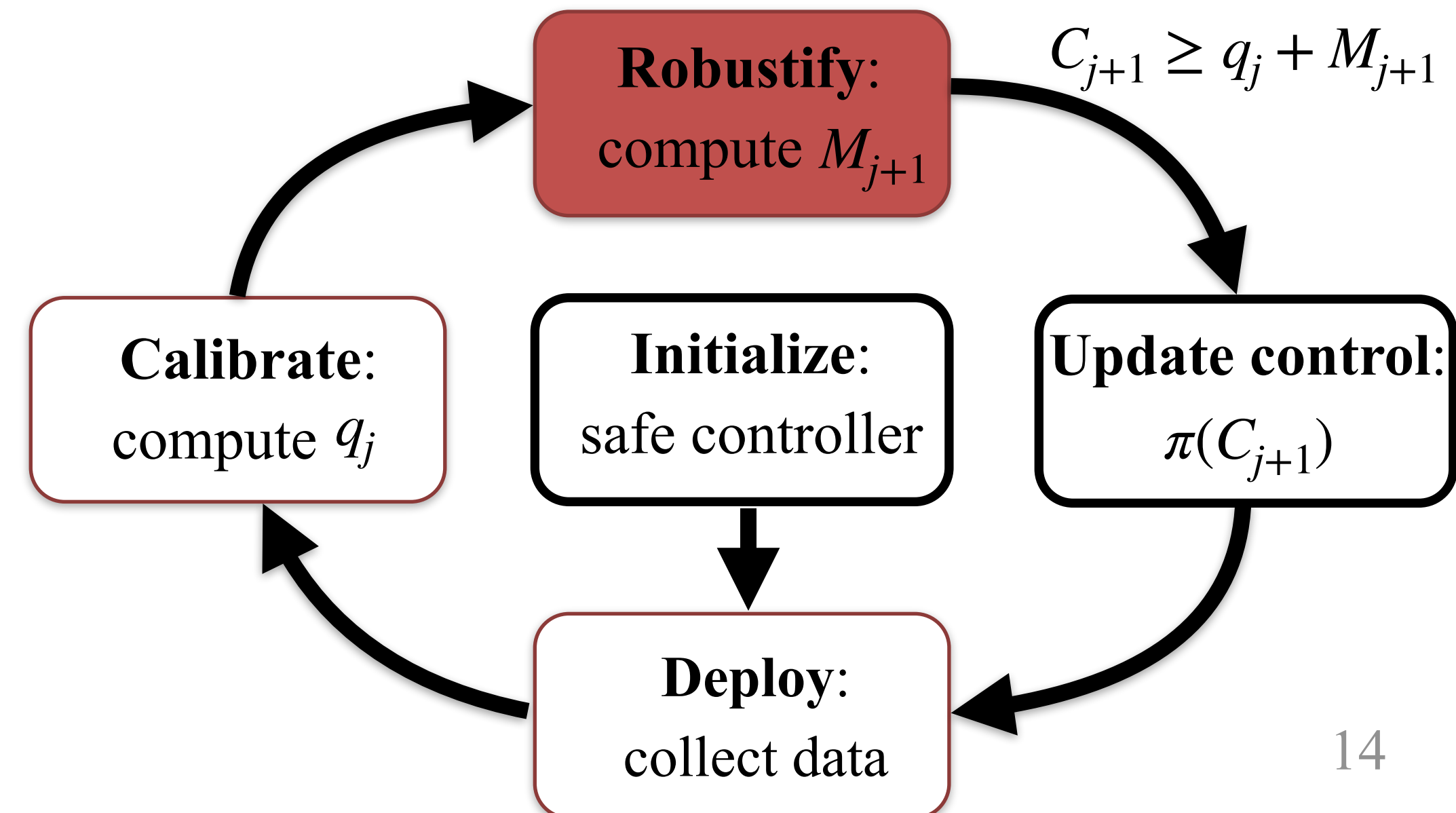
$$R_j^{(i)} := \max_{\tau \in \{1, \dots, T\}} \|y_{j,\tau}^{(i)} - \hat{y}_\tau^{(i)}\|$$

nonconformity score



Algorithm for controller updates

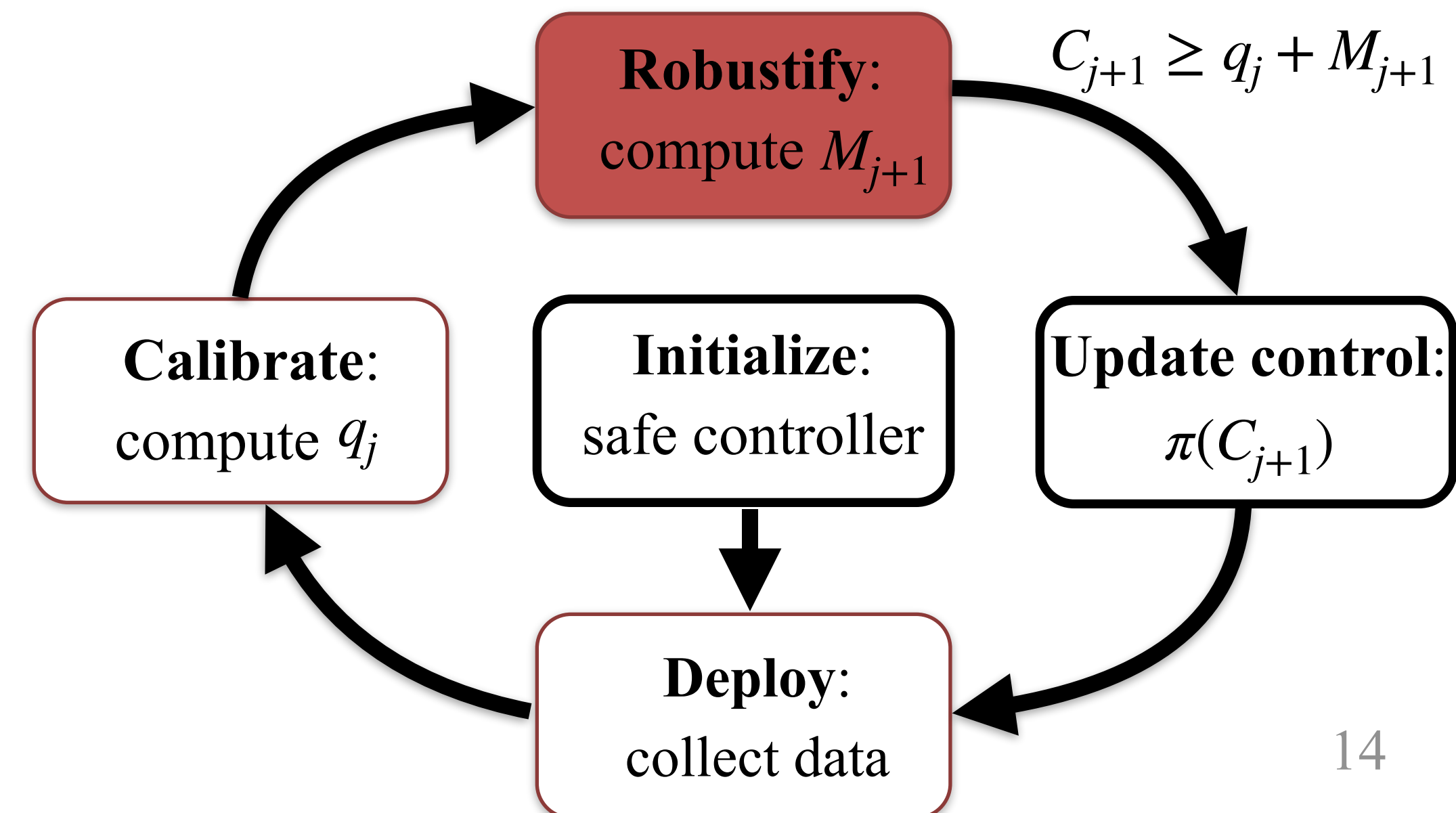
- Collect new calibration data $y_j^{(1)}, \dots, y_j^{(k)} \sim \mathcal{D}(x_j)$
 - Apply conformal prediction $q_j = \text{Quantile}_{1-\delta}(R_j^{(1)}, \dots, R_j^{(k)}, \infty)$
- $R_j^{(i)} := \max_{\tau \in \{1, \dots, T\}} \|y_{j,\tau}^{(i)} - \hat{y}_\tau^{(i)}\|$
nonconformity score



Algorithm for controller updates

- Collect new calibration data $y_j^{(1)}, \dots, y_j^{(k)} \sim \mathcal{D}(x_j)$
 - Apply conformal prediction $q_j = \text{Quantile}_{1-\delta}(R_j^{(1)}, \dots, R_j^{(k)}, \infty)$
 - Observe that $R_{j+1}^{(i)} - R_j^{(i)} \leq L_y \|\pi(C_{j+1}) - \pi(C_j)\|_\infty =: M_{j+1}$
- $R_j^{(i)} := \max_{\tau \in \{1, \dots, T\}} \|y_{j,\tau}^{(i)} - \hat{y}_\tau^{(i)}\|$
nonconformity score
- Lipschitz constant of dynamics

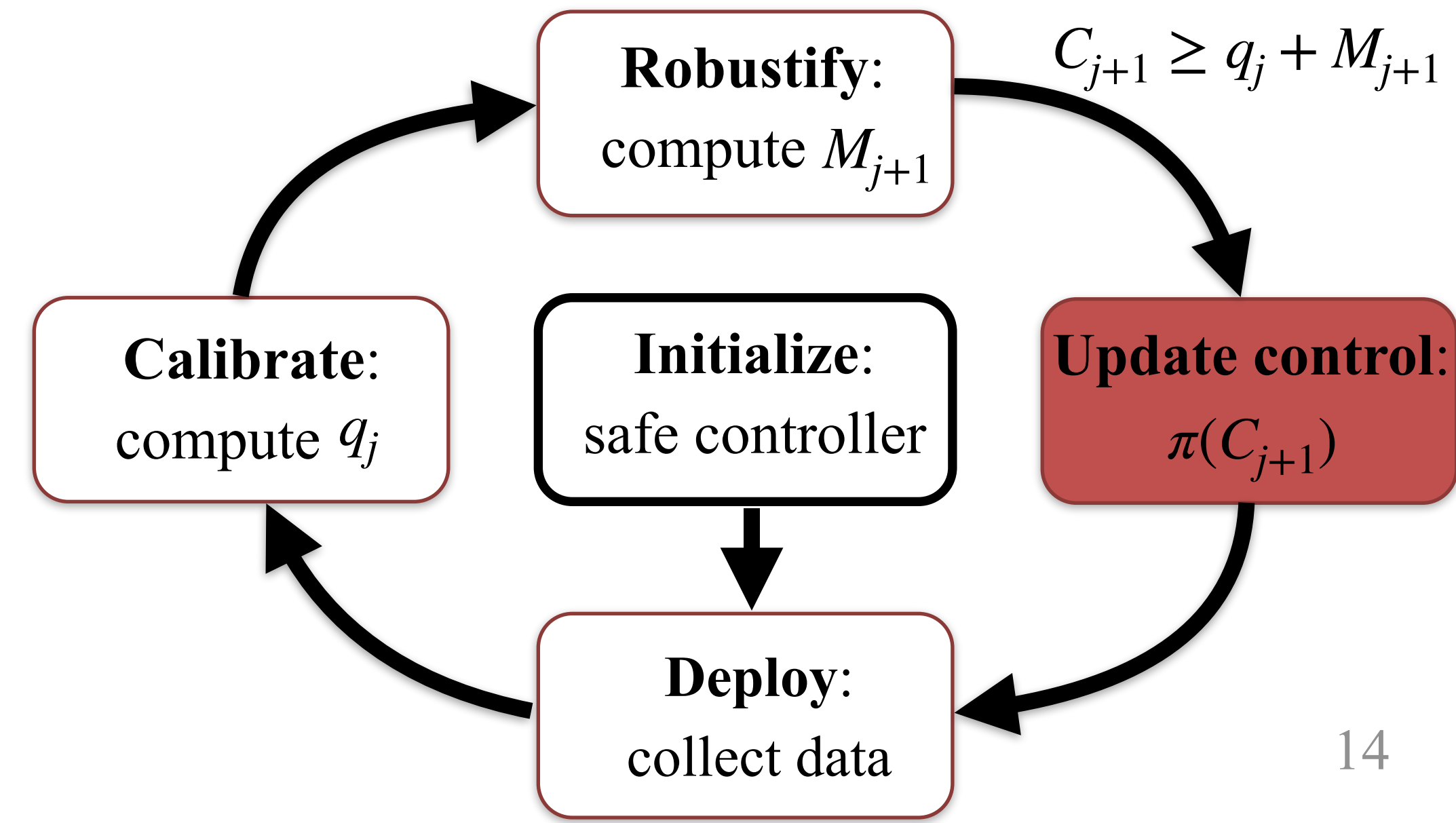
Adversarially conformal prediction guarantees that $\text{Prob}(R_{j+1} \leq q_j + M_{j+1}) \geq 1 - \delta$



Algorithm for controller updates

- Collect new calibration data $y_j^{(1)}, \dots, y_j^{(k)} \sim \mathcal{D}(x_j)$
 - Apply conformal prediction $q_j = \text{Quantile}_{1-\delta}(R_j^{(1)}, \dots, R_j^{(k)}, \infty)$
 - Observe that $R_{j+1}^{(i)} - R_j^{(i)} \leq L_y \|\pi(C_{j+1}) - \pi(C_j)\|_\infty =: M_{j+1}$
- $R_j^{(i)} := \max_{\tau \in \{1, \dots, T\}} \|y_{j,\tau}^{(i)} - \hat{y}_\tau^{(i)}\|$
nonconformity score
- Lipschitz constant of dynamics

Adversarially conformal prediction guarantees that $\text{Prob}(R_{j+1} \leq q_j + M_{j+1}) \geq 1 - \delta$



Algorithm for controller updates

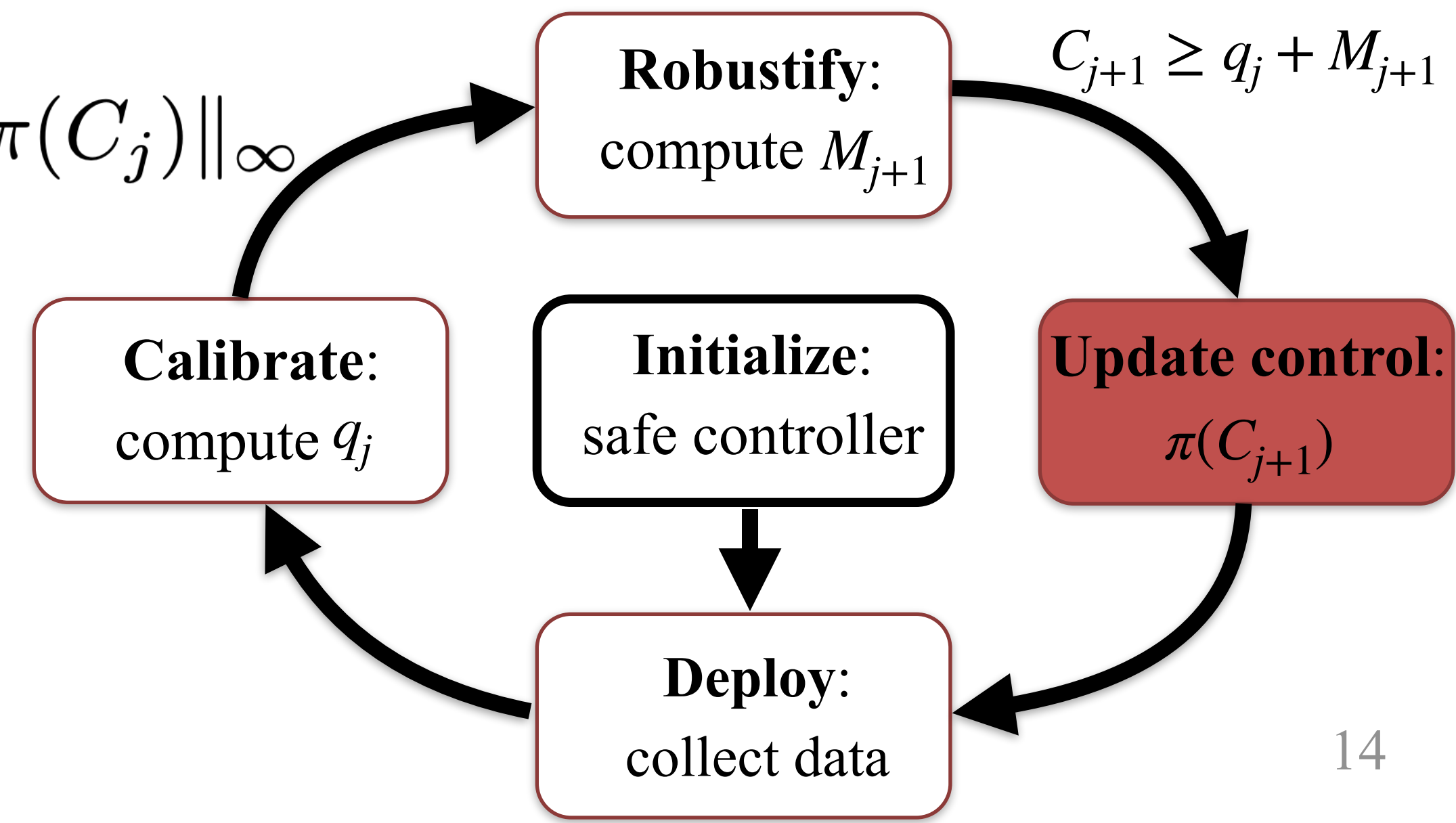
- Collect new calibration data $y_j^{(1)}, \dots, y_j^{(k)} \sim \mathcal{D}(x_j)$
 - Apply conformal prediction $q_j = \text{Quantile}_{1-\delta}(R_j^{(1)}, \dots, R_j^{(k)}, \infty)$
 - Observe that $R_{j+1}^{(i)} - R_j^{(i)} \leq L_y \|\pi(C_{j+1}) - \pi(C_j)\|_\infty =: M_{j+1}$
- $R_j^{(i)} := \max_{\tau \in \{1, \dots, T\}} \|y_{j,\tau}^{(i)} - \hat{y}_\tau^{(i)}\|$
nonconformity score
- Lipschitz constant of dynamics

Adversarially conformal prediction guarantees that $\text{Prob}(R_{j+1} \leq q_j + M_{j+1}) \geq 1 - \delta$

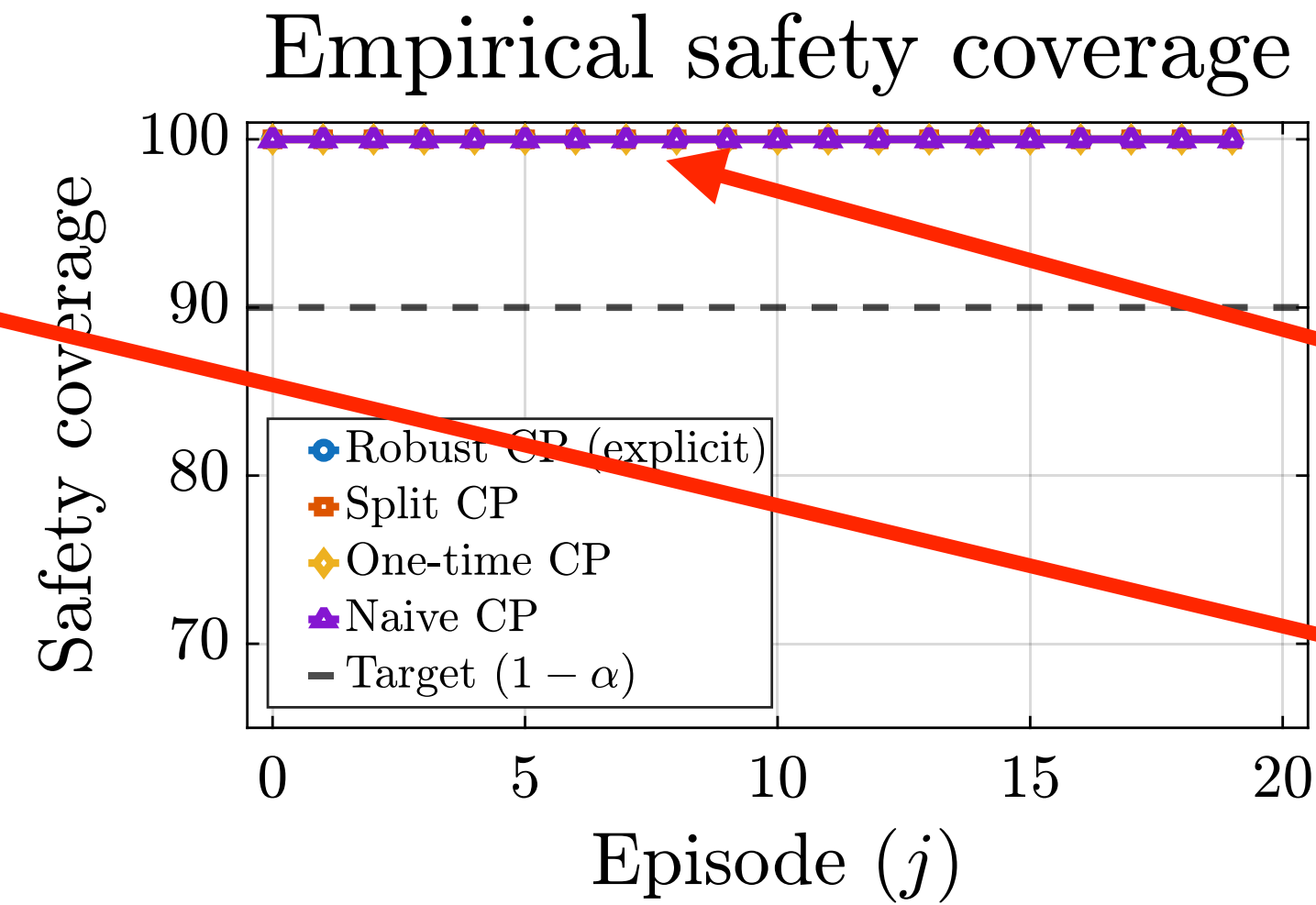
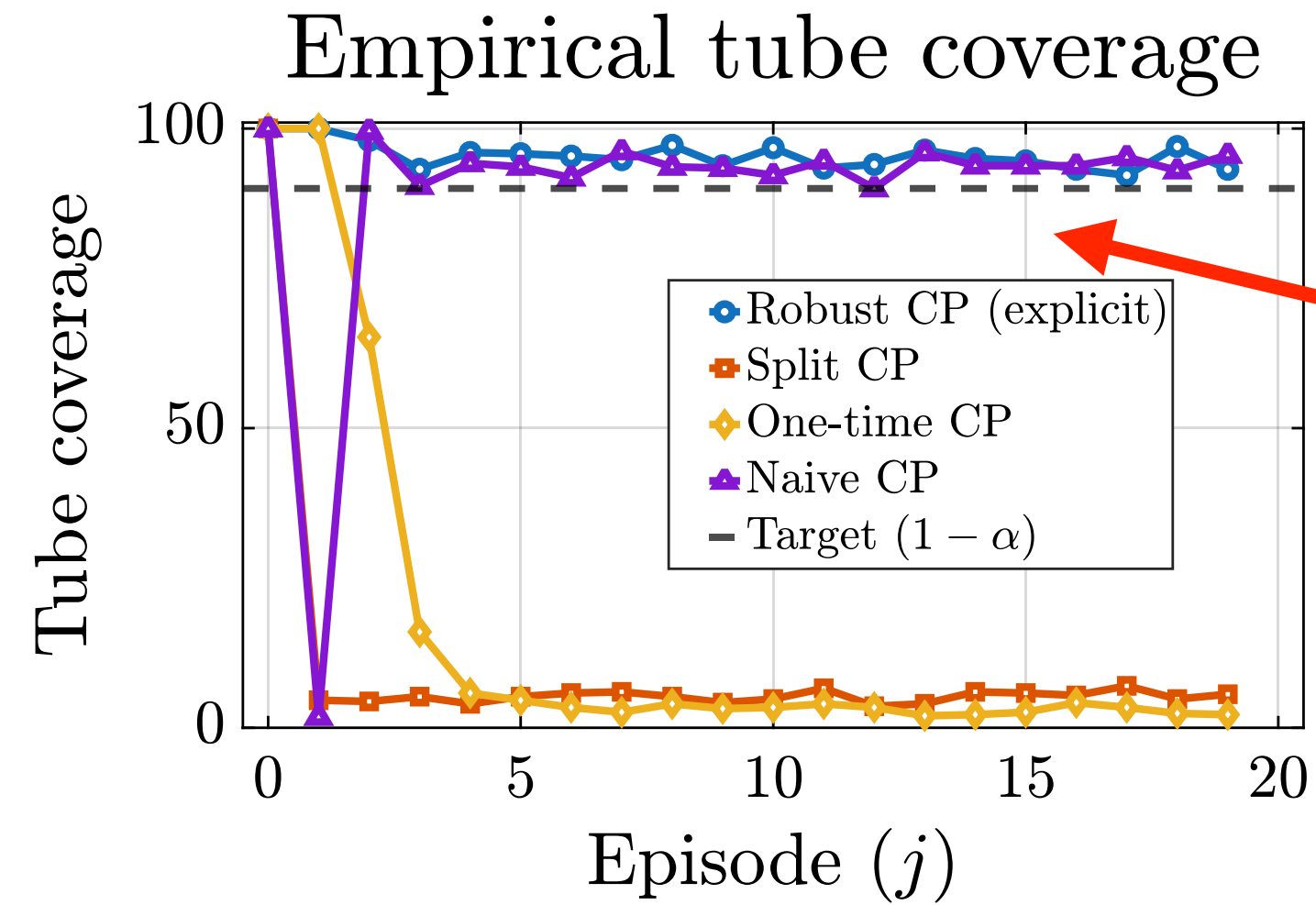
- This motivates the rule $C_{j+1} \geq q_j + M_{j+1}$

The policy $\pi(C_{j+1})$ guarantees episodic safety.

1. Solve via numerical solver (e.g., bisection)
2. Solve via $C_{j+1} \geq q_j + L_\pi L_y |C_{j+1} - C_j|$

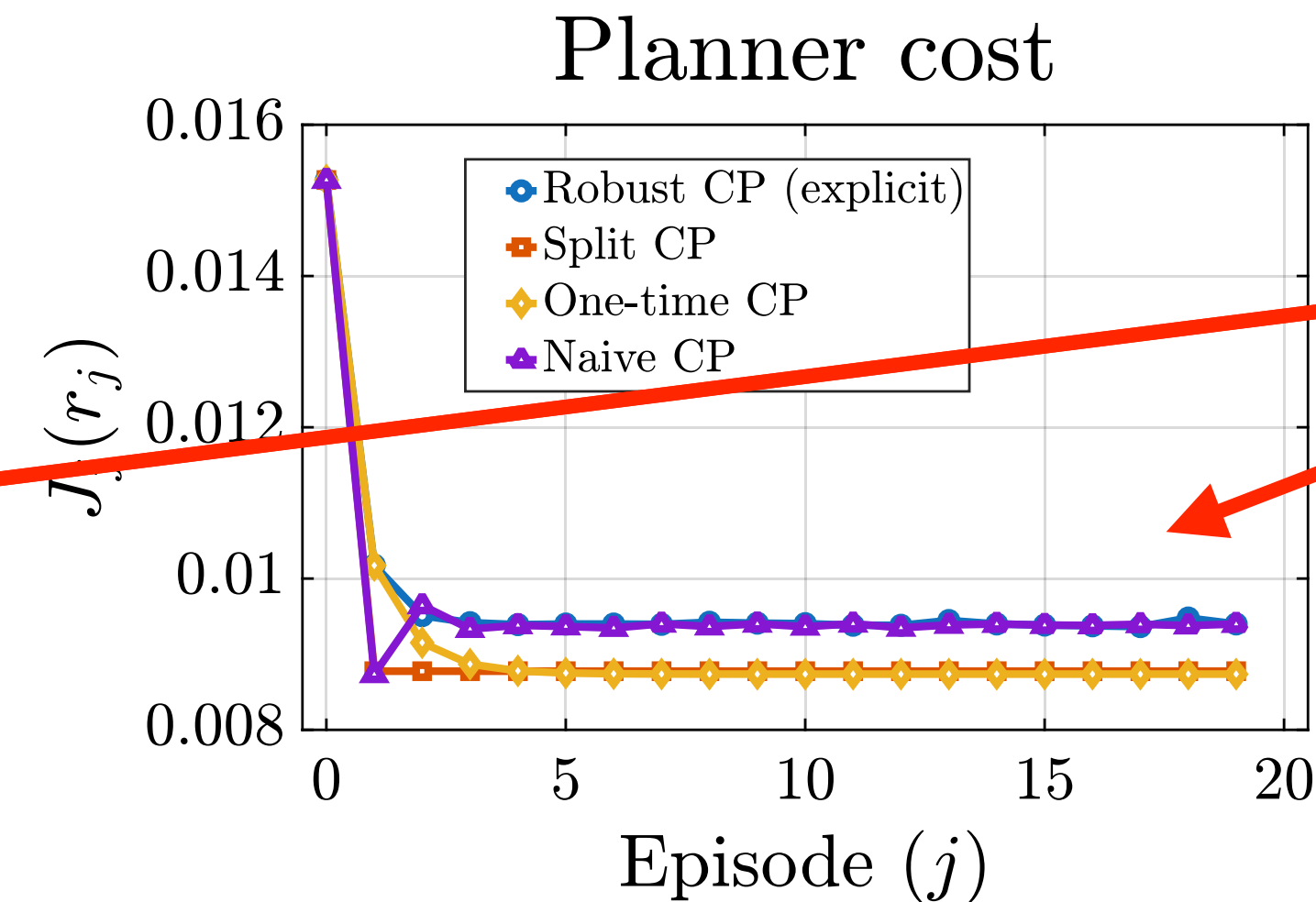
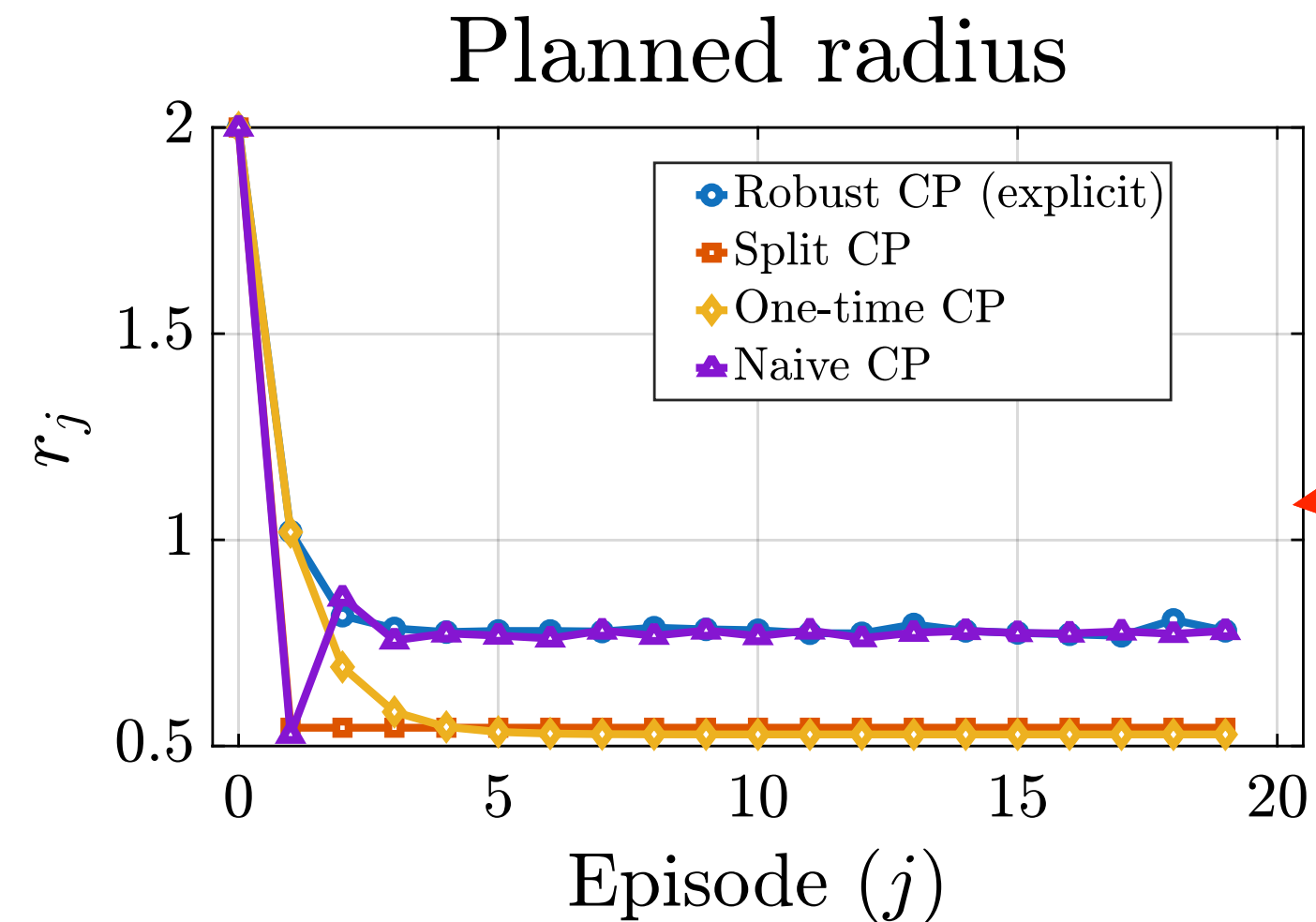


Academic example: car-pedestrian collision avoidance



$$\text{Prob}(c(x_j, y_j) \geq 0) \geq 1 - \delta$$

$$\text{Prob}\left(\max_{\tau} \|\hat{y}_{j,\tau} - y_j\| \leq C_j\right) \geq 1 - \delta$$

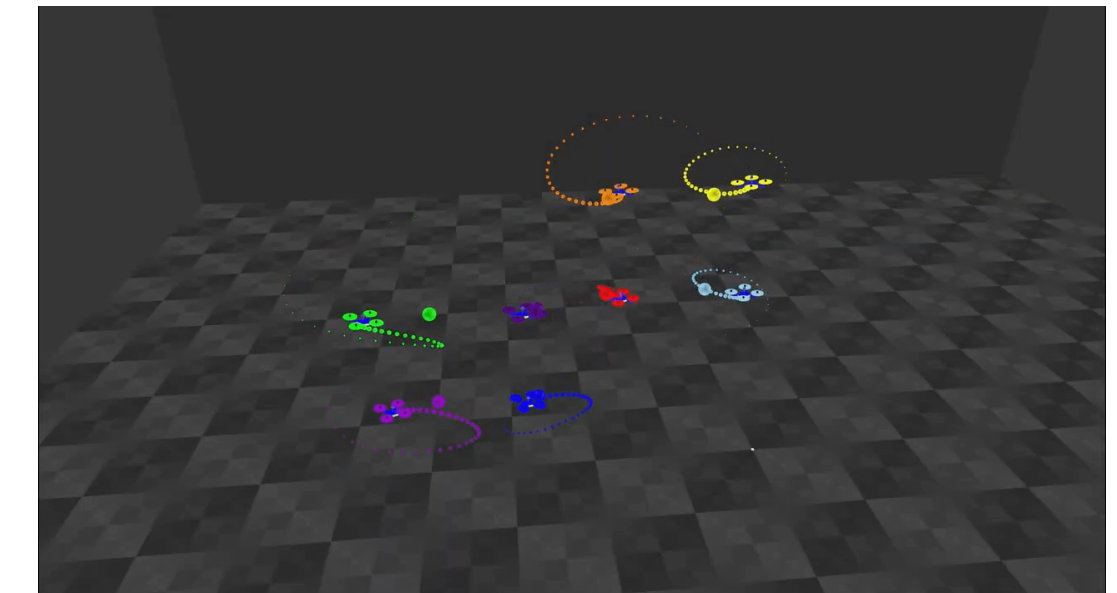


convergence conditions
presented in reference¹

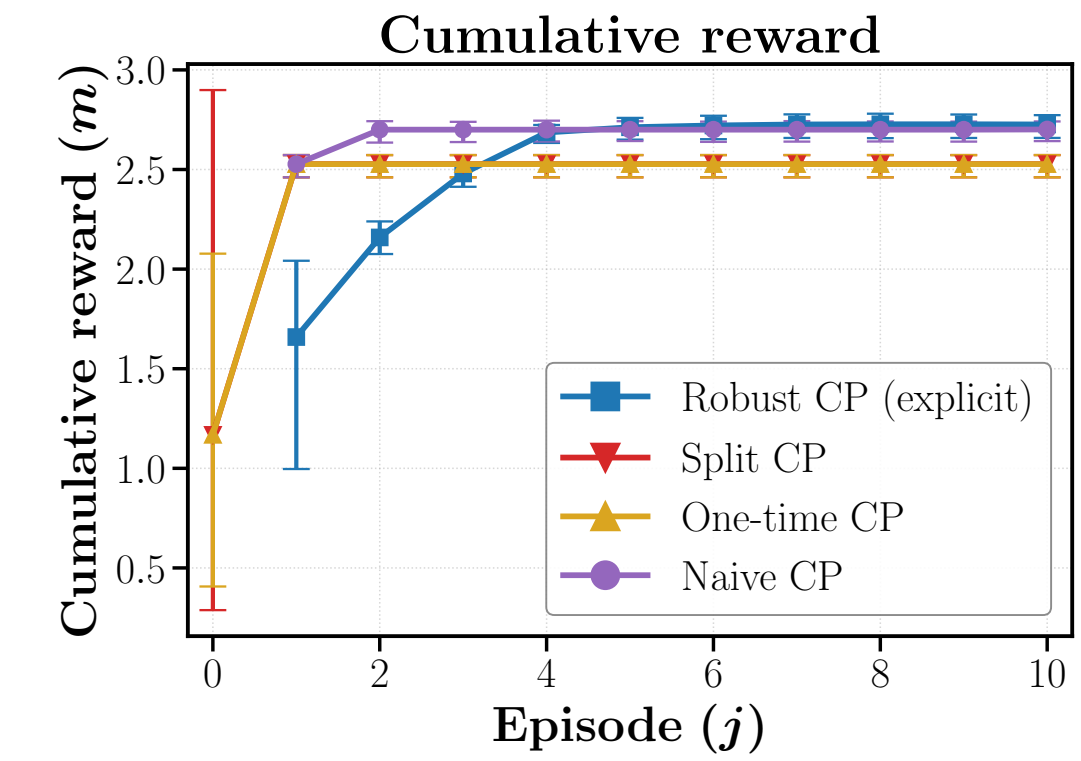
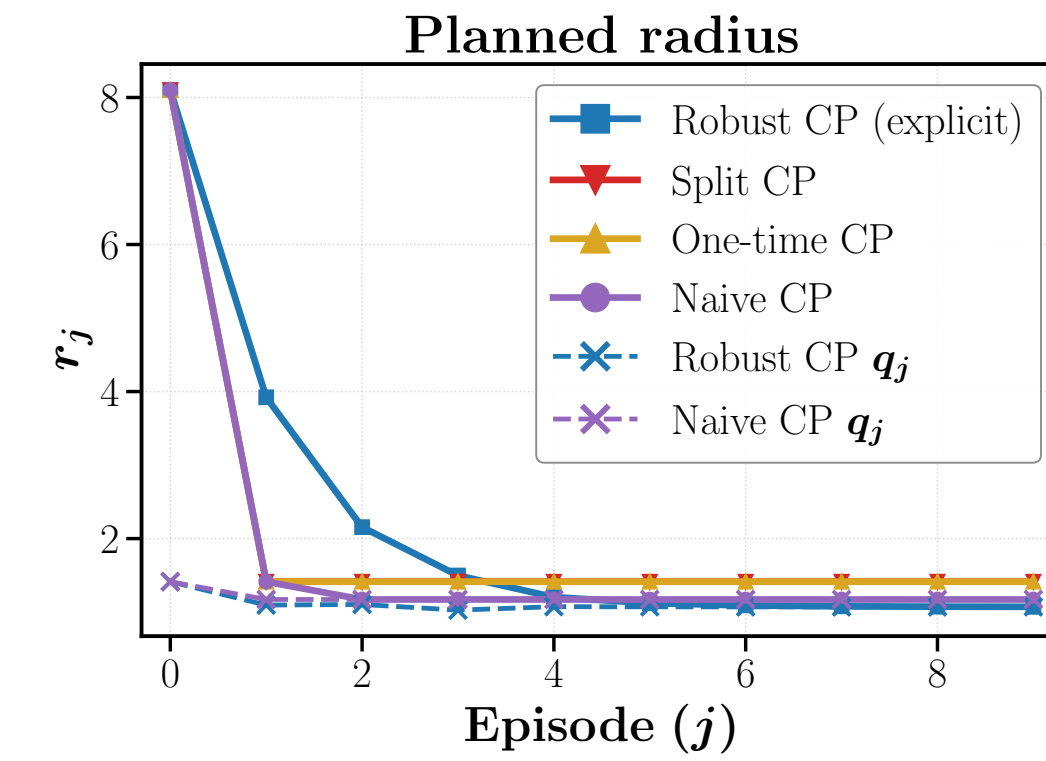
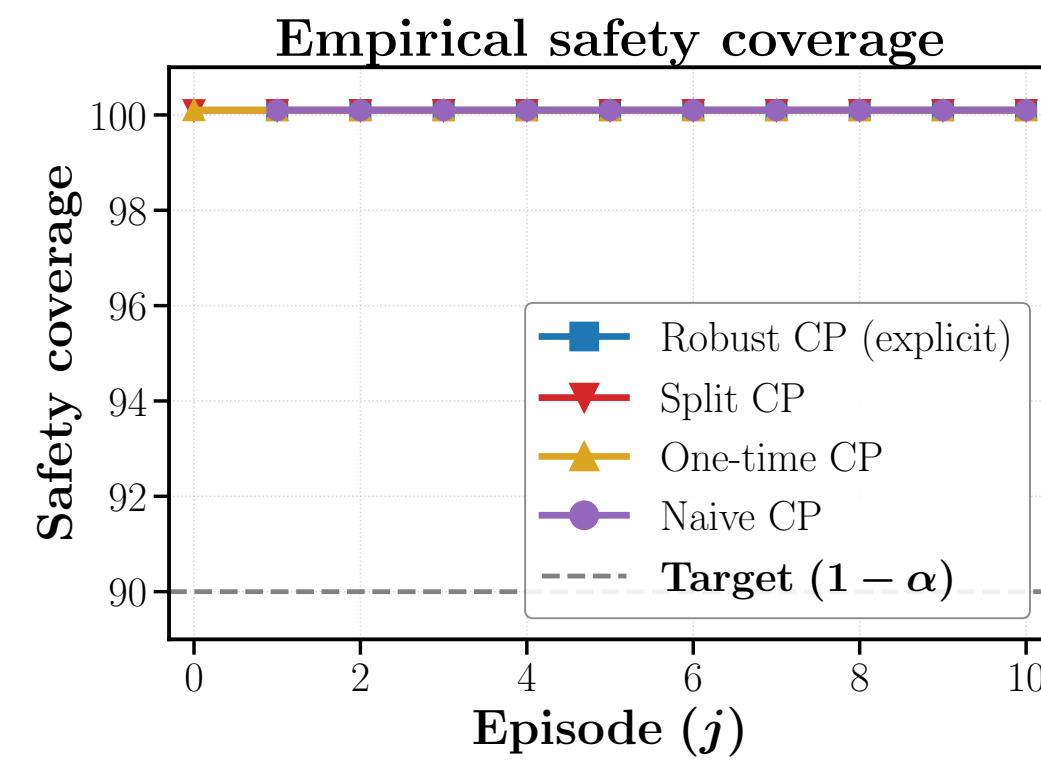
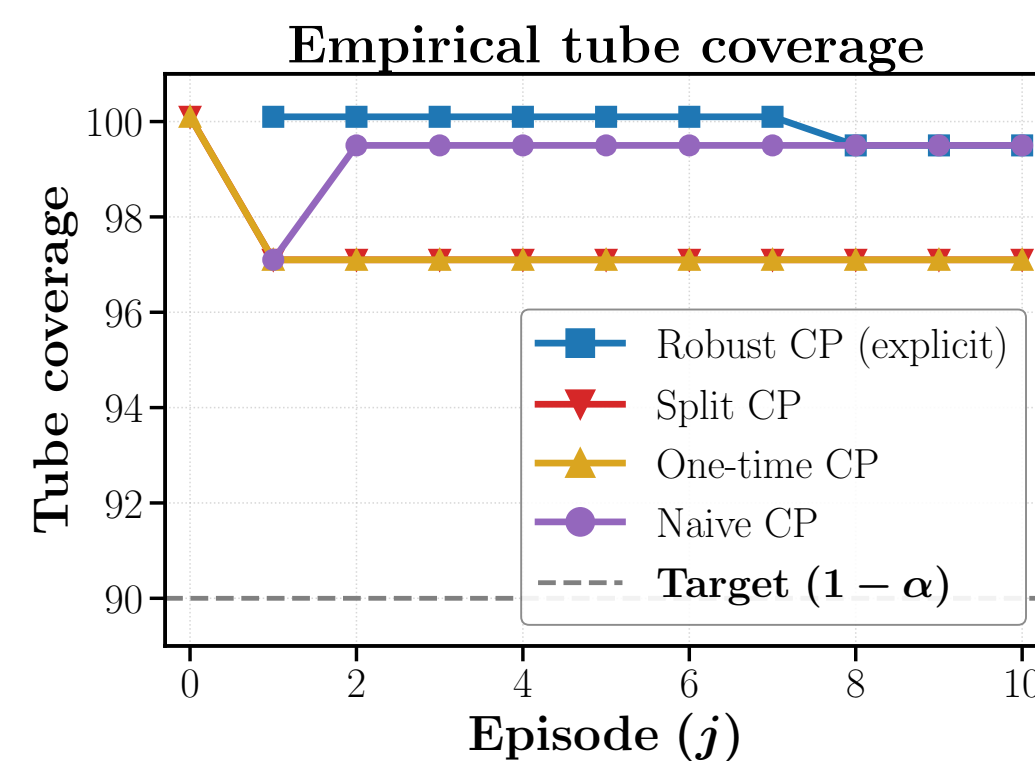
¹Mirzaeodang, Shekhtman, Matni, **Lindemann**, "Safe Planning in Interactive Environments via Iterative Policy Updates and Adversarially Robust Conformal Prediction", *L4DC*, 2026.

Quadswarm drone simulator

- We use Quadswarm¹ with five uncontrollable dynamic RL agents
- Ego drone deploys $\pi(C_j)$ filtered through a tracking CBF



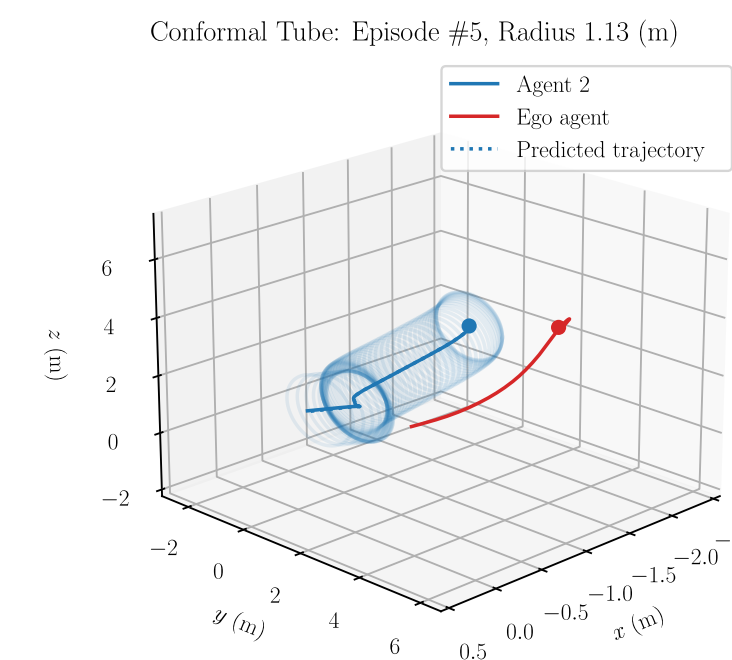
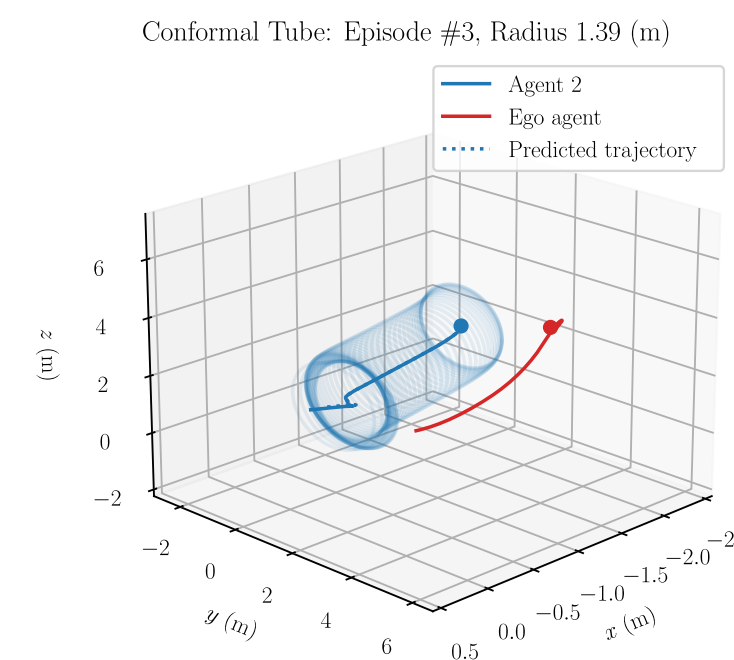
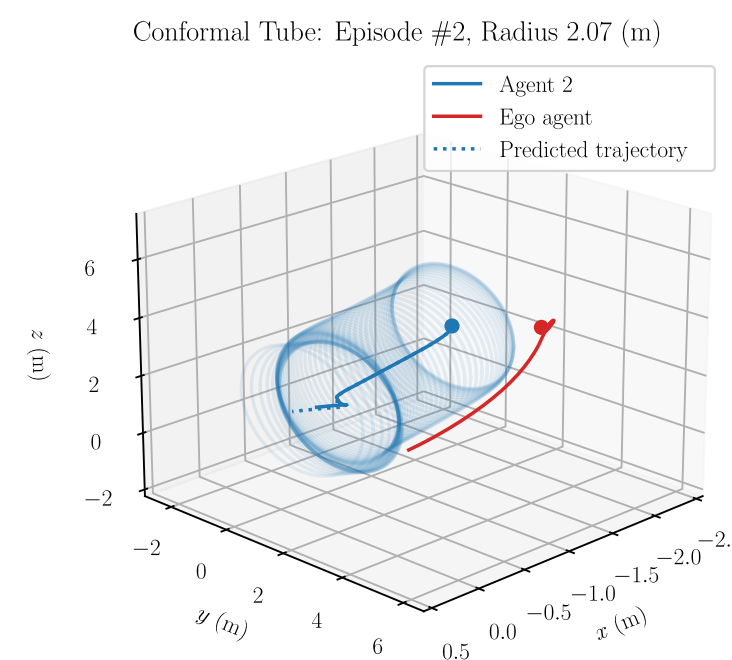
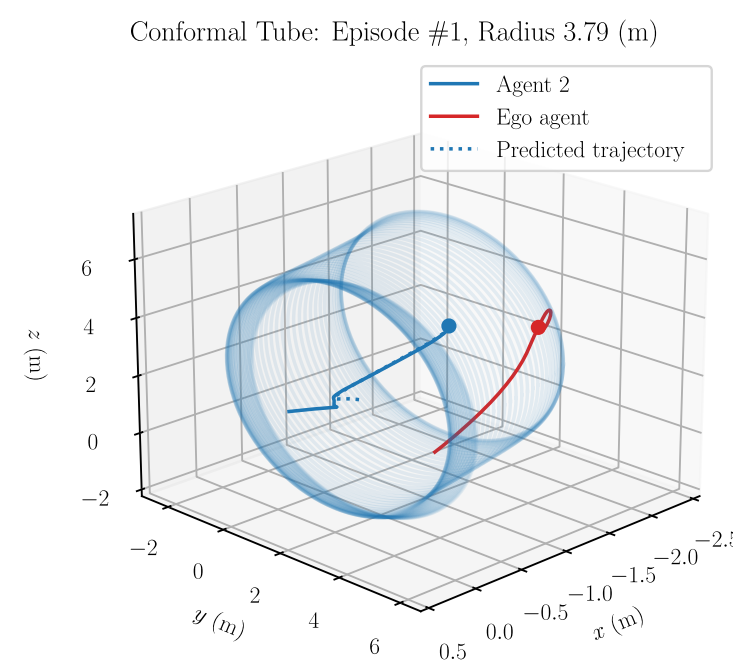
• Metrics:



$$\text{Prob}\left(\max_{\tau} \|\hat{y}_{j,\tau} - y_j\| \leq C_j\right) \geq 1 - \delta$$

$$\text{Prob}(c(x_j, y_j) \geq 0) \geq 1 - \delta$$

• Agent 2:



¹Huang et al., “QuadSwarm: A Modular Multi-Quadrotor Simulator for Deep Reinforcement Learning with Direct Thrust Control”, *ICRA (Workshop)*, 2023.

Safe Control in Changing Environments

uncontrollable distribution shift



Uncontrollable distribution shifts

- **Design and deployment conditions** may be different

Examples:

changing environments



data from simulators



← no safety guarantees

Assumption: design and deployment conditions are ϵ -close

- **Robust conformal prediction**^{1,2}: Let $R^{(1)}, \dots, R^{(k)} \sim \mathcal{R}$ and $R^{(0)} \sim \mathcal{R}_0$ as well as $D_f(\mathcal{R}_0, \mathcal{R}) \leq \epsilon$

It holds that $\text{Prob}(R^{(0)} \leq \tilde{C}) \geq 1 - \delta$ with $\tilde{C} := \text{Quantile}_{1-\tilde{\delta}(\epsilon)}(R^{(1)}, \dots, R^{(k)})$.

adjusted confidence level $\tilde{\delta}(\epsilon) < \delta$
(solution of convex optimization problem)

→ Increased data requirements

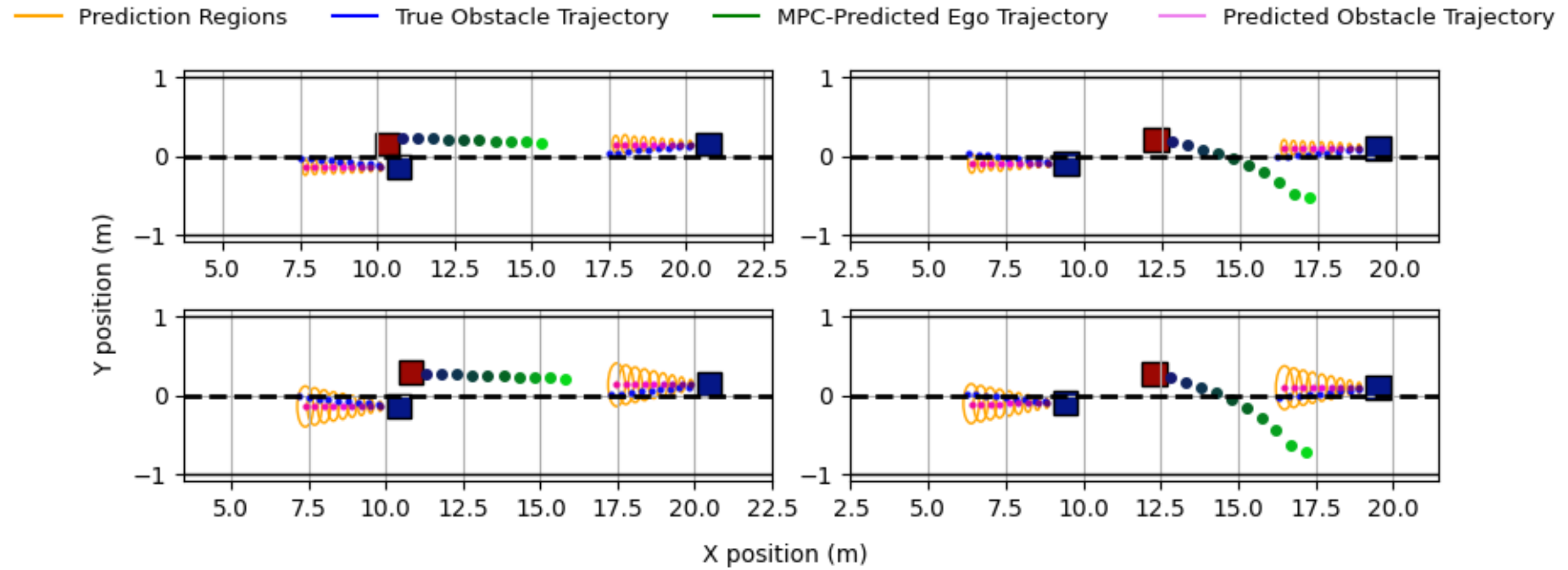
- **Idea:** perform control design with robust conformal prediction regions!

¹Cauchois, Gupta, Ali, and Duchi, “Robust validation: Confident predictions even when distributions shift”, JASA, 2024.

²Aolaritei, Wang, Zhu, Jordan, and Marzouk, “Conformal Prediction under Levy-Prokhorov Distribution Shift”, NeurIPS, 2025.

Safe control with generative priors

- **Example¹ (wrong-lane drivers):**



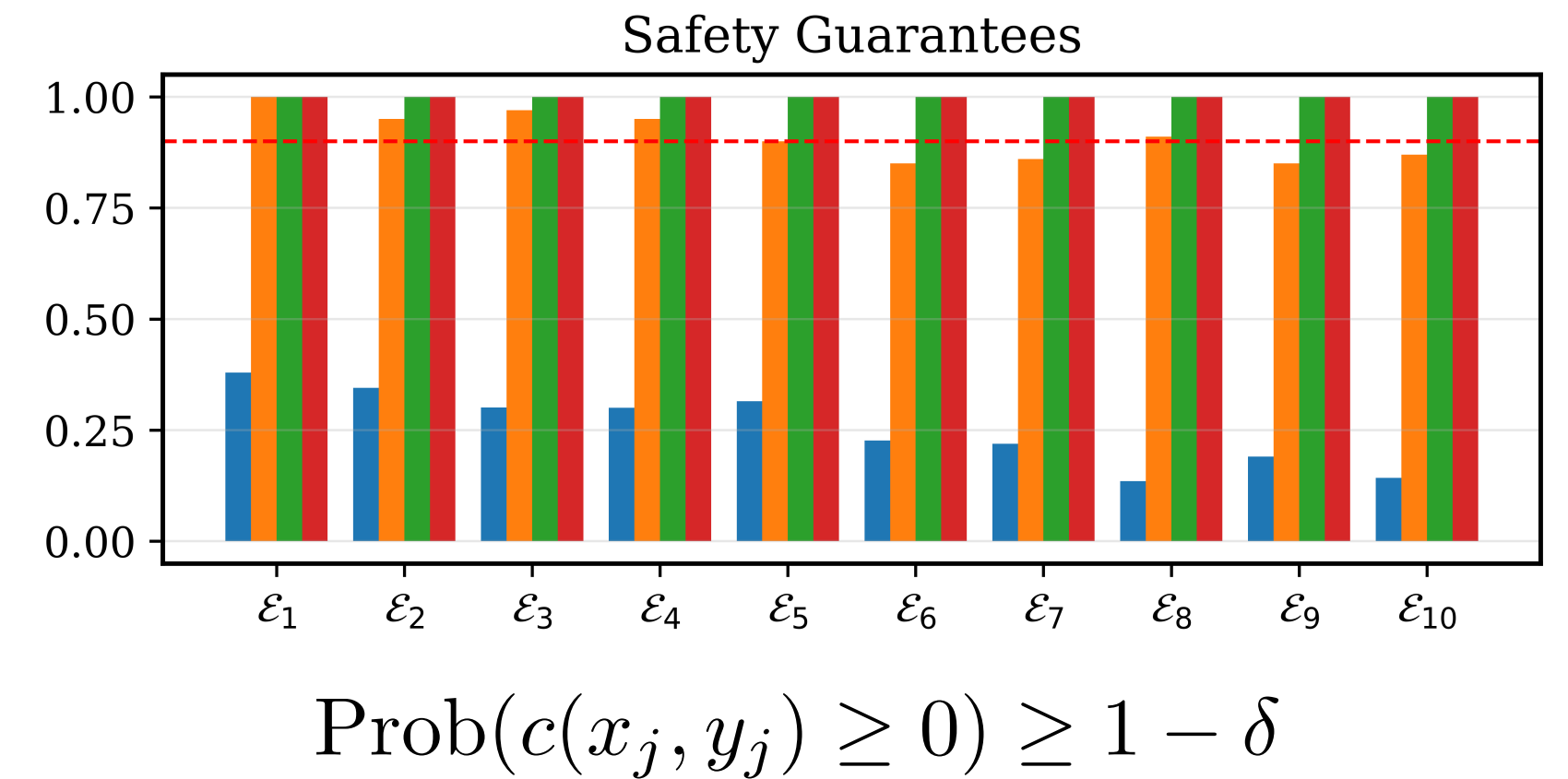
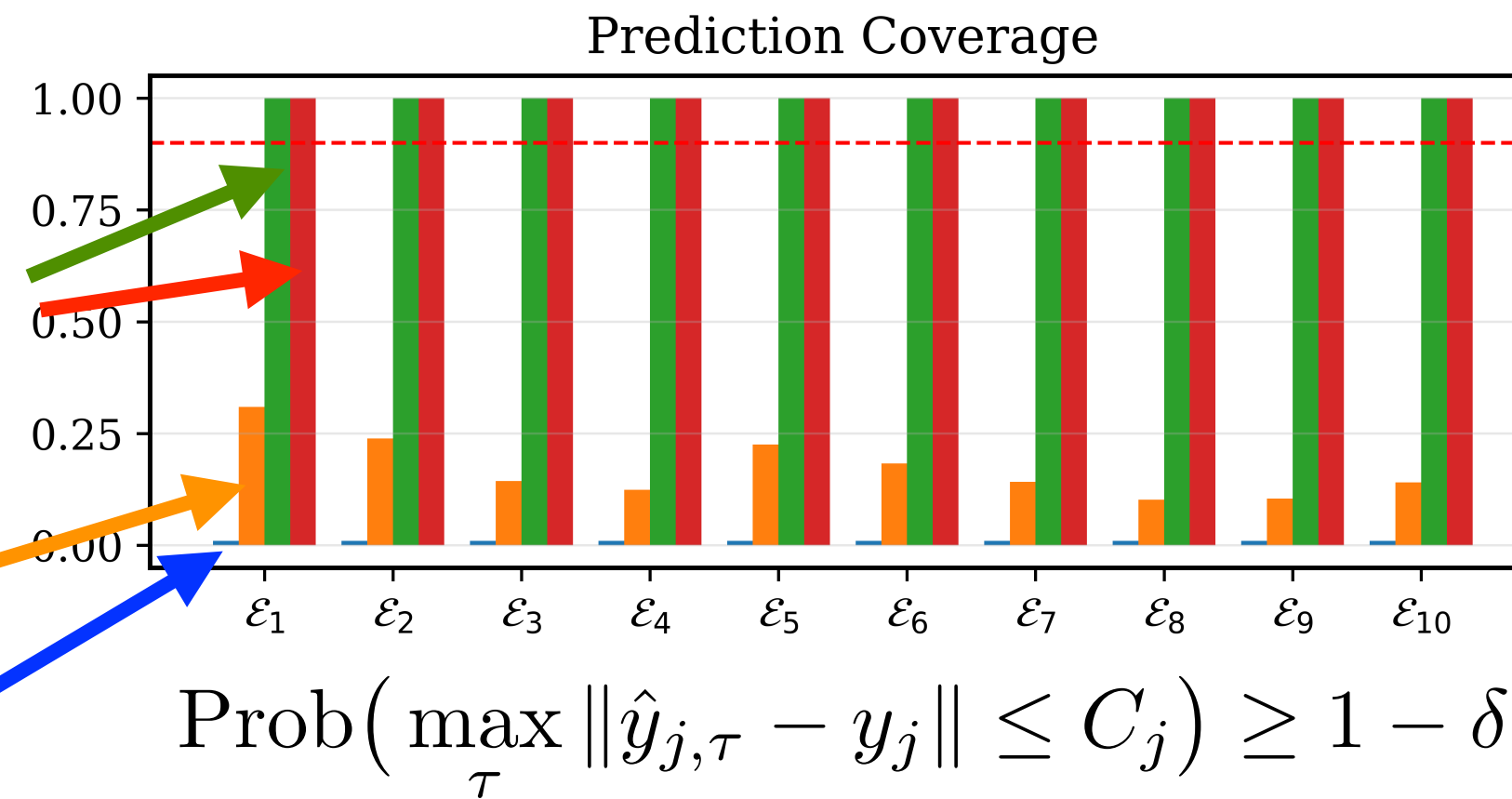
- We use conditional diffusion models to generate calibration data $R^{(1)}(\mathcal{E}), \dots, R^{(k)}(\mathcal{E}) \sim \mathcal{R}$
observable parameter (here averaged velocity) →

- **Result (summary):**

our method w. different ways of
estimating distribution shift

only synthetic data
(no robust CP)

only a static dataset



¹Rahaman, Deshmukh, Hota, and **Lindemann**, “When Environments Shift: Safe Planning with Generative Priors and Robust Conformal Prediction”, L4DC, 2026.

Adaptive control when the distribution shifts

- What if distributions shift “too much”?



→
adaptive CP
from online
datastream

$$\delta_{t+1} := \delta_t + \gamma(\delta - e_t)$$

learning rate γ desired coverage level δ empirical coverage e_t

$$e_t := \begin{cases} 0 & \text{if } r_t \leq C_t \\ 1 & \text{otherwise} \end{cases}$$

updated coverage level δ_{t+1}

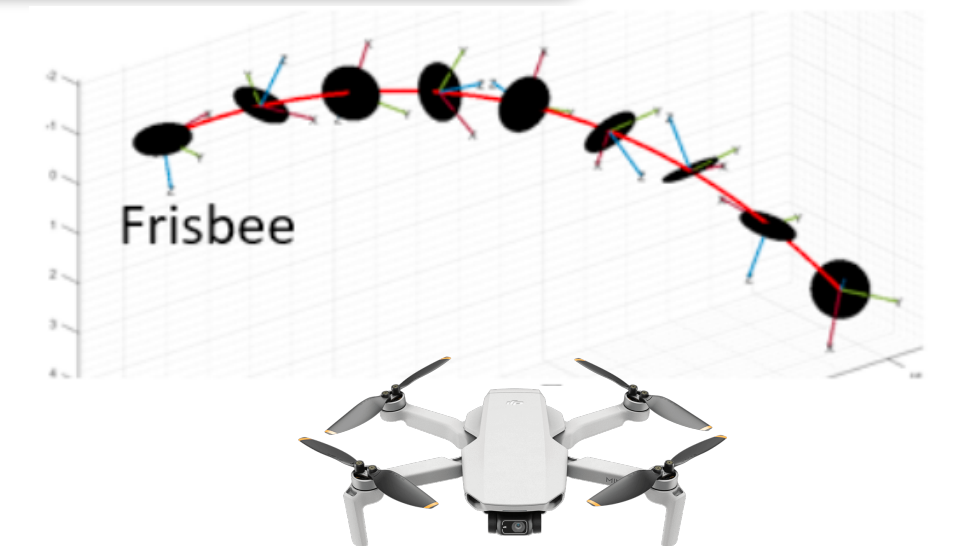
- We can design an MPC using adaptive conformal prediction regions

If the MPC is recursively feasible, we find an on average safe controller, i.e.,

$$\frac{1}{T} \sum_{t=0}^{T-1} \text{Prob}(c(x_{t+1}, y_{t+1}) \geq 0) \geq 1 - \delta - \frac{\delta_0 + \gamma}{T\gamma}$$

- Example:** Drone avoiding a flying frisbee

Case	δ UQ method	0.025			0.05		
		Proposed	Wei et al. (2022)	w/EKF	Proposed	Wei et al. (2022)	w/EKF
Frisbee	%Succ.	99.2	100	100	100	100	100
w/drag	$\bar{d}_{min}$	2.91	14.2	5.27	2.74	4.97	4.25



Selected References

Survey article:

- **L. Lindemann**, Y. Zhao, X. Yu, G. J. Pappas, and J. V. Deshmukh, “Formal Verification and Control with Conformal Prediction”, CSM, 2025

Main references:

- **L. Lindemann**, M. Cleaveland, G. Shim, and G. J. Pappas, “Safe Planning in Dynamic Environments using Conformal Prediction”, RA-L, 2023.
- M. Cleaveland, I. Lee, G. J. Pappas, and **L. Lindemann**, “Conformal Prediction Regions for Time Series using Linear Complementarity Programming”, AAI, 2024.
- O. Mirzaeedodangeh, E. Shekhtman, N. Matni, **L. Lindemann**, “Safe Planning in Interactive Environments via Iterative Policy Updates and Adversarially Robust Conformal Prediction”, L4DC, 2026.
- O. Mirzaeedodangeh, E. Shekhtman, N. Matni, **L. Lindemann**, “Robust Conformal CBF and CLF Controllers via Iterative Policy Updates”, CDC (subm.), 2026.
- K. Rahaman, J. V. Deshmukh, A. Hota, and **L. Lindemann**, “When Environments Shift: Safe Planning with Generative Priors and Robust Conformal Prediction”, L4DC, 2026.
- A. Dixit, **L. Lindemann**, S. Wei, M. Cleaveland, G.J. Pappas, and J. Burdick, “Adaptive Conformal Prediction for Motion Planning among Dynamic Agents”, *L4DC*, 2023.

Main Collaborators



Omid
Mirzaeedodangeh



Matthew
Cleaveland



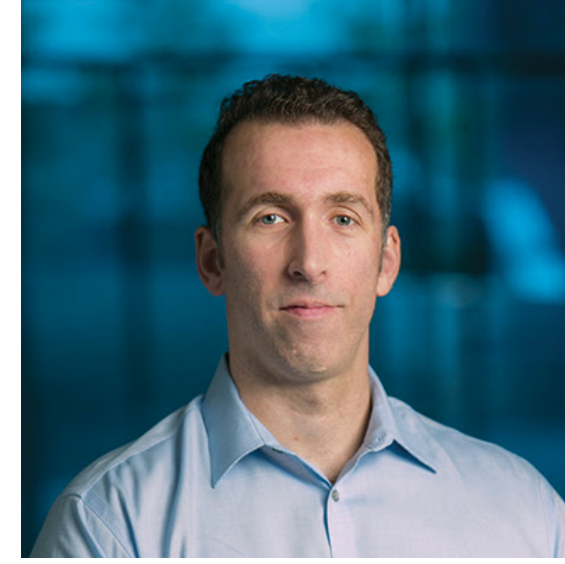
Eliot
Shekhtman



George Pappas



Jyotirmoy
Deshmukh



Nikolai Matni