Safe Planning in Interactive Environments via Iterative Policy Updates and Adversarially Robust Conformal Prediction

Omid Mirzaeedodangeh OMIRZAEEDODA@ETHZ.CH
Automatic Control Laboratory, ETH Zürich, Switzerland
Eliot Shekhtman SHEKHE@SEAS.UPENN.EDU
Computer and Information Science, University of Pennsylvania, USA
Nikolai Matni* NMATNI@SEAS.UPENN.EDU
Electrical and Systems Engineering, University of Pennsylvania, USA
Computer and Information Science, University of Pennsylvania, USA
Lars Lindemann LLINDEMANN@ETHZ.CH
Automatic Control Laboratory, ETH Zürich, Switzerland

Editors: G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

Abstract

Safe planning of an autonomous agent in interactive environments – such as the control of a self-driving vehicle among pedestrians – poses a major challenge as the behavior of the environment is unknown and reactive to the behavior of the autonomous agent. This coupling gives rise to interaction-driven distribution shifts where the autonomous agent's control policy may change the environment's behavior, thereby invalidating safety guarantees in existing work. Indeed, recent works have used conformal prediction (CP) to generate distribution-free safety guarantees using observed data of the environment. However, CP's assumption on data exchangeability is violated in interactive settings due to a circular dependency where a control policy update changes the environment's behavior, and vice versa. To address this gap, we propose an iterative framework that robustly maintains safety guarantees across policy updates by quantifying the potential impact of a planned policy update on the environment's behavior. We realize this via adversarially robust CP where we perform a regular CP step in each episode using observed data under the current policy, but then transfer safety guarantees across policy updates by analytically adjusting the CP result to account for distribution shifts. This adjustment is performed based on a policy-to-trajectory sensitivity analysis, resulting in a safe, episodic open-loop planner. We further conduct a contraction analysis of the system providing conditions under which both the CP results and the policy updates are guaranteed to converge. We empirically demonstrate these safety and convergence guarantees on a two-dimensional car-pedestrian and a high-dimensional quadcopter case study. To the best of our knowledge, these are the first results that provide safety guarantees in such interactive settings.

Keywords: Safe planning in interactive environments, distribution shifts, adversarially robust conformal prediction, iterative control policy updates.

1. Introduction

Autonomous agents, e.g., self-driving vehicles and service robots, are increasingly deployed in human-centric, multi-agent environments Mavrogiannis et al. (2023); Feng et al. (2025). The safe control of an autonomous agent is challenging as the behavior of uncontrollable agents, e.g., pedestrians, are unknown and interactive, i.e., they may react to the behavior of the autonomous agent. This creates an intricate coupling and interaction-driven distribution shift where the distribution of uncontrollable agent behaviors changes with the control policy of the autonomous agent. In this paper, we address this "chicken-and-egg" problem where changing the control policy changes the behavior of uncontrollable agents, and vice versa, see Figure 1.

* Eliot Shekhtman and Nikolai Matni were partially supported by AFOSR Award FA9550-24-1-0102, SF Award SLES-2331880, and NSF CAREER award ECCS-2045834.

Non-interactive approaches sequentially integrate predictions of uncontrollable agents into the design of the control policy Trautman and Krause (2010); Du Toit and Burdick (2011), while interactive approaches simultaneously predict uncontrollable agents and design control policies to take interactions directly into account Kretzschmar et al. (2016); Everett et al. (2021). Non-interactive approaches usually assume worst-case interaction bounds to provide safety guarantees, tending to be conservative. Interactive approaches perform better in practice, but fail to provide any safety guarantees. We instead propose a planning framework that takes interaction into account and enjoys safety guarantees via iterative policy updates and robust conformal prediction.

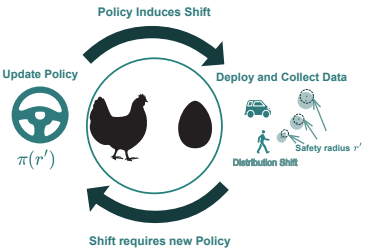

Figure 1: "Chicken-and-egg" problem: A policy change can induce a distribution shift in the environment, here a pedestrian. This shift results in a modified safety radius r' that captures the pedestrian's behavior under this policy. This modified safety radius in turn requires a policy update $\pi(r')$.

Safe planning with conformal prediction. Conformal prediction (CP) is an uncertainty quantification technique that provides statistical guarantees for test data after a one-time calibration on held-out calibration data Vovk et al. (2005); Angelopoulos and Bates (2021). CP has been used to construct prediction sets for uncontrollable agents that contain their true but unknown behavior with a user-specified probability Cleaveland et al. (2024); Sun and Yu (2022); Zecchin et al. (2024). These works learn a trajectory predictor for uncontrollable agents and perform calibration over a nonconformity score that measures the deviation between the predicted and the true motion on held-out trajectories. The resulting calibration result defines a set around new predictions that is probabilistically valid and thereby operational, i.e., suitable for motion planning. This has been explored via model predictive control Lindemann et al. (2023); Dixit et al. (2023); Shin et al. (2025), reinforcement learning Yao et al. (2024); Sun et al. (2023), control barrier functions Zhang et al. (2025); Hsu and Tsukamoto (2025), sampling-based search Sheng et al. (2024), and LLM planning Doula et al. (2025); Wang et al. (2024), see Lindemann et al. (2024) for a survey.

Interaction breaks conformal prediction guarantees. CP presumes that the data at test time is exchangeable with the data at calibration time.[1] Indeed, simply using non-exchangeable calibration data can break conformal prediction guarantees, see e.g., Tibshirani et al. (2019); Zhao et al. (2024). In interactive settings the premise of exchangeability is violated due to a circular dependency induced by interaction-driven distribution shifts: changing the control policy changes the data distribution at test time, while updating the held-out calibration data changes the control policy. Many CP extensions aim to mitigate this issue, e.g., via adaptive CP Gibbs and Candes (2021); Zaffran et al. (2022), CP for covariate shifts Tibshirani et al. (2019); Yang et al. (2024); Alijani and Najjaran (2025), or robust CP Cauchois et al. (2024); Aolaritei et al. (2025); Gendler et al. (2022). However, these extensions either require continual recalibration or additional estimators, yield undersirable time-averaged or worst-case guarantees, or do not close the loop with a planner whose policy changes drive the distribution shift. Our interactive setting demands a mechanism that (i) directly addresses policy-induced distribution shifts, and (ii) closes the loop with the planner.

Positioning and differences to prior work on planning with CP. Prior work embeds calibrated CP sets into the planner by either assuming exchangeability Lindemann et al. (2023); Tonkens et al. (2023); Yu et al. (2026) or by using adaptive CP Dixit et al. (2023); Yao et al. (2024); Shin et al. (2025); Sheng et al. (2024). Conceptually closest to our work are Wang et al. (2025) and Huang

1. We note that independent and identically distributed data is automatically also exchangeable.

et al. (2025), which also address agent interactions with CP; Wang et al. (2025) discretizes the joint agent space and uses CP to capture state-dependent interaction. However, data requirements grow exponentially with the state dimension without addressing the aforementioned core circular dependency. The work Huang et al. (2025) proposes iterative policy updates similar to our method, but without providing episodic safety guarantees as we enable via adversarially robust CP. Additionally, Huang et al. (2025) provide safety guarantees only after convergence, which is generally not guaranteed. In contrast, we provide: (i) episodic safety guarantees by analytically bounding the distribution shift, and (ii) explicit conditions on episodic convergence via a contraction analysis.

2 Problem Formulation

We consider the state of an ego agent $x_t \in \mathbb{R}^{d_x}$, its control input $u_t \in \mathbb{R}^{d_u}$, and the state of uncontrollable agents' $y_t \in \mathbb{R}^{d_y}$ at discrete times $t = 0, \dots, T$. Their coupled dynamics are

$$x_{t+1} = f_X(x_t, u_t), \qquad y_{t+1} = f_Y(y_t, x_t, u_t, \nu_t), \qquad t = 0, ..., T-1. \tag{2.1}$$

with exogenous noise ν_t, which is a random variable that models the uncontrollable agents' intentions. The mapping $(x_t, u_t) \mapsto f_Y(\cdot, x_t, u_t, \cdot)$ captures agent interaction. Here, safety is encoded by a safety function $H(x_{0:T}, y_{0:T})$ and defined over the entire trajectories $x_{0:T}$ and $y_{0:T}$. The system is considered safe if $H(x_{0:T}, y_{0:T}) \leq 0$. We assume that the system (2.1) is Lipschitz continuous.

Assumption 1 *There exist Lipschitz constants $L_{Xx}, L_{Xu}, L_{Yy}, L_{Yx}, L_{Yu} \geq 0$ such that $\|f_X(x,u) - f_X(x',u')\|_2 \leq L_{Xx}\|x-x'\|_2 + L_{Xu}\|u-u'\|_2$ and $\|f_Y(y,x,u,\nu) - f_Y(y',x',u',\nu)\|_2 \leq L_{Yy}\|y-y'\|_2 + L_{Yx}\|x-x'\|_2 + L_{Yu}\|u-u'\|_2$ for all $x, x' \in \mathbb{R}^{d_x}$, $u, u' \in \mathbb{R}^{d_u}$, and $y, y' \in \mathbb{R}^{d_y}$.*

Running Example. We use a running example of a self-driving vehicle and a pedestrian with states x_t and y_t (e.g., position, velocity) and controls u_t (e.g., acceleration, steering). The dynamics $f_Y(y_t, x_t, u_t, \nu_t)$ capture interaction as the pedestrian's future position y_{t+1} depends on their own state y_t, the vehicle's states x_t and actions u_t, and their own intentions ν_t. A common way to define H is as the maximum violation of a separation constraint over time, e.g., $H(x_{0:T}, y_{0:T}) := \max_{0 \leq t \leq T} \{ b_t - c(x_t, y_t) \} \leq 0$) where $c(x_t, y_t)$ represents a separation measure such as the distance between the vehicle's position and the pedestrian's position and $b_t \geq 0$ is a safety margin.

2.1 The Idealized Chance-Constrained Planning Problem

Our ideal (unfortunately unattainable) objective is to find a control policy $u_{0:T-1}$ that minimizes a performance cost $J(x_{0:T}, u_{0:T-1})$ while satisfying a probabilistic safety guarantee:

$$\min_{u_{0:T-1}} J(x_{0:T}, u_{0:T-1}) \quad \text{s.t.} \quad \begin{cases} \mathbb{P}\{H(x_{0:T}, y_{0:T}) \leq 0\} \geq 1 - \alpha \\ \text{Dynamics in equation (2.1)} \end{cases} \tag{2.2}$$

The problem (2.2) is generally intractable. First, the uncontrollable agents' dynamics f_Y and the noise distribution of ν_t are typically unknown, e.g., we cannot precisely model interaction and intentions, requiring a model- and distribution-free approach. Second, the high-dimensionality of the problem induces computational complexity: even if f_Y and ν_t were known, the chance constraint (2.2) would require solving a complex, high-dimensional integral over the distribution of $y_{0:T}$.

While the first two challenges have been addressed in the literature, the third and most difficult challenge arises due to **interaction-induced distribution shifts**. Effectively, the dynamics f_Y and the distribution of $y_{0:T}$ are policy-dependent, meaning they change when the ego-agent's states x_t and control inputs u_t change, e.g., if the car accelerates aggressively, the pedestrian's decision to

cross an intersection will change. Robust control, which enforces safety for all permissible trajectories, is overly conservative. Statistical uncertainty quantification techniques – such as in Lindemann et al. (2023) which use conformal prediction (CP) – fail as the policy-driven distribution shift breaks exchangeability assumptions needed in CP, e.g., data of a pedestrian crossing in front of a slow car is not exchangeable with the new scenario where the car accelerates.

2.2 Non-Interactive Planning with Distribution-Free Certificates via Conformal Prediction

We now summarize existing work on non-interactive planning – primarily following Lindemann et al. (2023) – in which case f_Y does not depend on x_t and u_t, i.e., not addressing the third of the aforementioned challenges. These techniques will serve as a starting point for our proposed method.

First, a pre-designed offline predictor is used to produce a single, nominal estimate $\hat{y}_{0:T} = (\hat{y}_0, \dots, \hat{y}_T)$ of the environment trajectory $y_{0:T} = (y_0, \dots, y_T)$. To assess the accuracy of this estimate, we define the nonconformity score and the induced trajectory tube, respectively, as

$$s(\hat{y}_{0:T}, y_{0:T}) := \max_{0\le t\le T} \|\hat{y}_t - y_t\|_2, \qquad \mathcal{C}_r(\hat{y}_{0:T}) := \{\, y_{0:T} : \; s(\hat{y}_{0:T}, y_{0:T}) \le r \,\}. \tag{2.3}$$

The threshold $r \ge 0$ is computed using CP to obtain probabilistic guarantees on the correctness of the set $\mathcal{C}_r(\hat{y}_{0:T})$. The main idea is simple: we use a set of N held-out calibration trajectories $\{y^{(i)}_{0:T}\}_{i=1}^N$ generated by the non-interactive dynamics $y^{(i)}_{t+1} = f_Y(y^{(i)}_t, \nu^{(i)}_t)$. We then compute r as the $(1-\alpha)$-quantile of the held-out nonconformity scores, denoted by $q_{1-\alpha}(\{s(\hat{y}_{0:T}, y^{(i)}_{0:T})\}_{i=1}^N \cup \{\infty\})$.[2] Since the test trajectory $y_{0:T}$ and the held-out trajectories $\{y^{(i)}_{0:T}\}_{i=1}^N$ are exchangeable, we have that $\mathbb{P}_{N+1}\{\, y_{0:T} \in \mathcal{C}_{q_{1-\alpha}}(\hat{y}_{0:T}) \,\} \ge 1-\alpha$, where now $\mathbb{P}_{N+1}\{\cdot\}$ is a product probability measure that captures the randomness in test $y_{0:T}$ and calibration data $\{y^{(i)}_{0:T}\}_{i=1}^N$, and as such can approximate the chance constraint in equation (2.2). This guarantee motivates the formulation of the following robust planning problem:

$$\min_{u_{0:T-1}} \quad J(x_{0:T}, u_{0:T-1}) \tag{2.4a}$$

$$\text{s.t.} \quad x_{t+1} = f_X(x_t, u_t), \quad t = 0, \dots, T-1, \tag{2.4b}$$

$$H(x_{0:T}, \zeta) \le 0 \quad \forall\, \zeta \in \mathcal{C}_{q_{1-\alpha}}(\hat{y}_{0:T}), \tag{2.4c}$$

which ensures $\mathbb{P}_{N+1}\{H(x_{0:T}, y_{0:T}) \le 0\} \ge 1-\alpha$ in the non-interactive case whenever (2.4c) is feasible Lindemann et al. (2023). For instance, for the collision avoidance-type safety constraints $H(x_{0:T}, y_{0:T}) = \max_{0\le t\le T}\{\, b_t - c(x_t, y_t)\}$, the constraint (2.4c) reduces to a step-wise tightening of the form $c(x_t, \hat{y}_t) \ge b_t + \; q_{1-\alpha}$ for all times $t = 0, \dots, T$, which can easily be implemented.

However, in the interactive case where the coupled agent dynamics are $y_{t+1} = f_Y(y_t, x_t, u_t, \nu_t)$ instead of $y_{t+1} = f_Y(y_t, \nu_t)$, a circular dependency – or "chicken-and-egg" problem – emerges. Changing the control inputs u_t and ego-agent trajectory x_t changes the distribution of $y_{0:T}$, violating the exchangeability assumption with the held-out trajectories $\{y^{(i)}_{0:T}\}_{i=1}^N$. On the other hand, updating the held-out trajectories $\{y^{(i)}_{0:T}\}_{i=1}^N$ via $y^{(i)}_{t+1} = f_Y(y^{(i)}_t, x_t, u_t, \nu^{(i)}_t)$, changes the quantile $q_{1-\alpha}$ and thereby u_t and x_t. To address this issue, we propose an episodic framework in which we iteratively compute new held-out trajectories while updating the control inputs.

Running Example (cont.). In our running example this non-interactive approach would first predict a nominal path $\hat{y}_{0:T}$ for the pedestrian (e.g., walking straight on the sidewalk).

2. The split-conformal quantile $q_{1-\alpha}$ is computed from the set of nonconformity scores $\{s(\hat{y}_{0:T}, y^{(i)}_{0:T})\}_{i=1}^N \cup \{\infty\}$. Let $s_{(k)}$ be the k-th smallest nonconformity score (the k-th order statistic). Then, the quantile is $q_{1-\alpha} = s_{(k)}$ with $k = \lceil (N+1)(1-\alpha) \rceil$.

The trajectory tube $\mathcal{C}_r$ represents a tube of radius $r = q_{1-\alpha}$ around this nominal path that guarantees the actual pedestrian path $y_{0:T}$ will be inside this tube with $1-\alpha$ probability, assuming the car's actions do not influence the pedestrian. The "chicken-and-egg" problem arises in the interactive case: the car's plan u_t depends on the size of the pedestrian's uncertainty set (the radius $q_{1-\alpha}$), but the pedestrian's actual behavior, which determines the size of that tube, depends on the car's plan u_t.

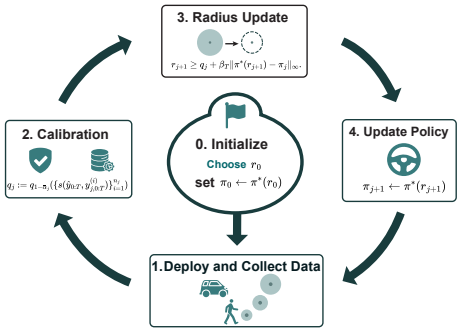

Figure 2: Our iterative algorithm: (0) initialize with a safe (yet conservative) radius r_0 and compute policy $\pi^\star(r_0)$; (1) deploy $\pi^\star(r_j)$ and collect data; (2) calibrate q_j; (3) compute new radius r_{j+1} from q_j; (4) update policy to $\pi^\star(r_{j+1})$; (5) repeat.

2.3 Episodic Problem Formulation and Design Goals

To address the aforementioned circular dependency, we reframe the problem into an iterative, episodic framework which we explain below and summarize in Figure 2. Planning proceeds in episodes $j = 0, 1, 2, \ldots$. We use a fixed nominal predictor $\hat{y}_{0:T} = (\hat{y}_0, \ldots, \hat{y}_T)$ to anchor the geometry of the uncertainty. Updating the predictor episodically is possible with minimal modifications, but omitted here for simplicity. At each episode j, we solve the following robust optimization problem, parameterized by an uncertainty radius $r_j \geq 0$:

$$\boxed{\begin{array}{lll} \mathbf{P}[j;\, r_j] & \min\limits_{u_{j,0:T-1}} & J(x_{j,0:T}, u_{j,0:T-1}) \\ & \text{s.t.} & x_{j,t+1} = f_X(x_{j,t}, u_{j,t}), \quad t = 0, \ldots, T-1, \\ & & H(x_{j,0:T}, \zeta) \leq 0 \quad \forall\, \zeta \in \mathcal{C}_{r_j}(\hat{y}_{0:T}). \end{array}} \tag{2.5}$$

Solving (2.5) yields the policy $\pi_j := u_{j,0:T-1}$ and the resulting optimal cost $J_j := J(x_{j,0:T}, u_{j,0:T-1})$. This policy is then executed on the physical system n_j times, which produces n_j i.i.d. rollouts of the environment's trajectories $\left\{y^{(i)}_{j,0:T}\right\}_{i=1}^{n_j} \sim \mathcal{D}(\pi_j)$ under the ego-agent's policy, as governed by (2.1) and denoted by $\mathcal{D}(\pi_j)$. The design of this iterative process is driven by three high-level objectives:

$$\textbf{Per-episode safety:} \quad \mathbb{P}\{\, H(x_{j,0:T}, y_{j,0:T}) \leq 0 \,\} \geq 1-\alpha \quad \text{for all } j, \tag{2.6}$$

$$\textbf{Performance improvement:} \quad J_{j+1} \leq J_j - \Delta_j \quad \text{with } \Delta_j \geq 0, \tag{2.7}$$

$$\textbf{Stability of uncertainty:} \quad r_{j+1} \approx r_j \quad \text{eventually.} \tag{2.8}$$

The key challenge is to determine how to update r_j to r_{j+1}. This update must leverage the new data $\{y^{(i)}_{j,0:T}\}$ to reduce conservatism, while also guaranteeing that the next policy π_{j+1} will be safe. Because r_j is implicitly tied to the policy π_j through interaction, the policy-induced distribution shift from j to $j+1$ must be accounted for to maintain (2.6). We make the following assumption.

Assumption 2 At iteration $j = 0$, we know r_0 s.t. π_0 satisfies $\mathbb{P}\{H(x_{0,0:T}, y_{0,0:T}) \leq 0\} \geq 1-\alpha$.

This assumption is natural to our problem formulation. In practice, we can typically find this constant by selecting a sufficiently large r_0 that is valid for any permissible ego agent policy.

Running Example (cont.). In episode $j = 0$, assume that we have selected a valid uncertainty radius r_0 for the ego agent's policy π_0, e.g., that the pedestrian will deviate at most $r_0 = 2$ meters from its predicted path under π_0 with probability of at least $1-\alpha$. We then observe realized pedestrian trajectories $\{y^{(i)}_{0,0:T}\}$ under π_0, where we may now notice that the pedestrian may deviate less than r_0 meters from the predicted path. The key challenge is hence to iteratively update the next radii $r_1, r_2, \ldots$ and compute the next policies π_1, π_2 while adhering to the objectives in (2.6)-(2.8).

3 An Iterative Planning Framework for Interactive Environments

In our interactive setting, the exchangeability assumption underlying CP is violated at every policy update. When we solve (2.5) and update the policy from π_j to π_{j+1}, the realized trajectory $y_{j+1,0:T}$ is drawn from a different distribution. This re-introduces the "chicken-and-egg" problem.

3.1 Adversarial Conformal Prediction for Policy-Induced Shifts

To carry safety certificates across such policy-induced shifts, our approach uses adversarial conformal prediction (ACP), which provides guarantees under bounded perturbations Gendler et al. (2022). Let $z \sim \mathcal{D}$ be a random variable and $\hat{z}$ be an estimate of z. Given a nonconformity score $s(\hat{z}, z)$ and a perturbation radius $\rho \geq 0$, ACP provides coverage for $s\big(\hat{z}, z + \Delta\big)$ where the random variable z may be perturbed by any Δ with $\|\Delta\| \leq \rho$. The next result follows by combining ACP with conditional CP Vovk (2012); Duchi (2025), as summarized in (Lindemann et al., 2024, Lemma 2).

Lemma 1 (Gendler et al. (2022) and Vovk (2012); Duchi (2025)) *Let $z, z^{(1)}, \ldots, z^{(N)} \sim \mathcal{D}$ be exchangeable random variables, $\rho > 0$ be a perturbation radius, $\alpha, \delta \in (0,1)$ be failure probabilities, and $r := q_{1-\bar{\alpha}}\big(\{s(\hat{z}, z^{(i)})\}_{i=1}^N\big) := s_{(\lceil N(1-\bar{\alpha})\rceil)}$ be the empirical $(1-\bar{\alpha})$-quantile of the unperturbed nonconformity scores $\{s(\hat{z}, z^{(i)})\}_{i=1}^N$, where $s_{(1)} \leq \cdots \leq s_{(N)}$ denote the associated order statistics and $\bar{\alpha} := \alpha - \sqrt{\ln(1/\delta)/2N}$.[3] Assume there exists a constant $M \geq 0$ such that for all z it holds that $\sup_{\|\Delta\| \leq \rho} s(\hat{z}, z + \Delta) \leq s(\hat{z}, z) + M$.[4] Then, we have that*

$$\mathbb{P}_N\Big\{ \mathbb{P}\big\{ \forall\, \Delta \text{ with } \|\Delta\| \leq \rho :\ z + \Delta \in \mathcal{C}_r^{adv}(\hat{z}; \rho) \big\} \geq 1 - \alpha \Big\} \geq 1 - \delta. \tag{3.1}$$

where $\mathcal{C}_r^{adv}(\hat{z}; \rho) := \big\{ z : s(\hat{z}, z) \leq r + M \big\}$ is a robustified prediction set.

We note that the guarantees in (3.1) are conditional in the sense that $\mathbb{P}_N\{\cdot\}$ is a product probability measure that captures randomness in the calibration data $\{z^{(i)}\}_{i=1}^N$, so that the inner probability measure $\mathbb{P}\{\cdot\}$ captures randomness over test data z that holds with probability no less than $1 - \delta$. To address the "chicken-and-egg" problem, we will apply robustification via ACP iteratively. At each episode j, we perform a recalibration step using the new data $\{y_{j,0:T}^{(i)}\}_{i=1}^{n_j} \sim \mathcal{D}(\pi_j)$ to compute the empirical quantile $q_j := q_{1-\bar{\alpha}_j}\big(\{s(\hat{y}_{0:T}, y_{j,0:T}^{(i)})\}_{i=1}^{n_j}\big)$ where $\delta_j \in (0,1)$ and $\bar{\alpha}_j := \alpha - \sqrt{\ln(1/\delta_j)/(2n_j)}$. By Lemma 1, this empirical quantile q_j satisfies $\mathbb{P}_{n_j}\big\{\mathbb{P}\big\{y_{j,0:T} \in \mathcal{C}_{q_j}(\hat{y}_{0:T})\big\} \geq 1-\alpha\big\} \geq 1-\delta_j$, where $y_{j,0:T} \sim \mathcal{D}(\pi_j)$ and $\mathcal{C}_{q_j}(\hat{y}_{0:T}) := \{y_{0:T} : s(\hat{y}_{0:T}, y_{0:T}) \leq q_j\}$. Thus, q_j is a valid safety radius for policy π_j. To maintain safety for the next policy π_{j+1}, we must now account for policy-induced distribution shifts caused by the change from π_j to π_{j+1}. We interpret and treat the policy update from π_j to π_{j+1} as the adversarial perturbation term ρ_j.

High-level intuition. Let us provide some intuition before we present the iterative algorithm and a detailed analysis in Sections 3.2 and 3.3, respectively. Under Assumption 1, we can bound the change in the uncontrollable agents trajectory by a constant $\beta_T \geq 0$ such that $\max_{0 \leq t \leq T} \big\|y_t(\pi_{j+1}) - y_t(\pi_j)\big\|_2 \leq \beta_T \|\pi_{j+1} - \pi_j\|_\infty$, where $\|\pi_{j+1} - \pi_j\|_\infty := \max_{0 \leq t \leq T-1} \|u_{j+1,t} - u_{j,t}\|_2$.[5] We formalize the existence of β_T next and provide the proof in Appendix A.1.

Lemma 2 (Episode Coupling Sensitivity) *Let the system in (2.1) be given and Assumption 1 hold. Then there exists a constant $\beta_T \geq 0$ which depends on the horizon T and the Lipschitz*

3. The empirical quantile $s_{(\lceil N(1-\bar{\alpha})\rceil)}$ here differs slightly from the empirical quantile $s_{(\lceil (N+1)(1-\alpha)\rceil)}$ from Section 2.
4. If the nonconformity score s is Lipschitz continuous in its second argument with constant L, then $M = L\rho$.
5. We use the notation of $y_t(\pi_j)$ instead of $y_{j,t}$ to highlight that the trajectory y_t is a function of π_j.

constants $L_{Xx}, L_{Xu}, L_{Yy}, L_{Yx}, L_{Yu}$ *such that for two policies* $\pi = \{u_t\}_{t=1}^{T-1}$ *and* $\pi' = \{u'_t\}_{t=1}^{T-1}$*:* $\|y_{0:T}(\pi') - y_{0:T}(\pi)\|_\infty \leq \beta_T \|\pi' - \pi\|_\infty$.

Next, note that our nonconformity score in (2.3) is 1-Lipschitz continuous. Therefore, we have $|s(\hat{y}_{0:T}, y_{j,0:T}) - s(\hat{y}_{0:T}, y_{j+1,0:T})| \leq \beta_T \|\pi_{j+1} - \pi_j\|_\infty$, i.e., a policy change can increase the nonconformity score by at most $M_{j+1} := \beta_T \|\pi_{j+1} - \pi_j\|_\infty$.

Using Lemma 1 and combining this term with the empirical quantile q_j yields a high-probability upper bound on the required radius for the next policy π_{j+1} that is of the form:

$$\mathbb{P}_{n_j}\big\{\, \mathbb{P}\big\{\, s(\hat{y}_{0:T}, y_{j+1,0:T}) \;\leq\; q_j + M_{j+1} \big\} \;\geq\; 1-\alpha \big\} \geq 1 - \delta_j. \tag{3.2}$$

This motivates our safe radius update rule $r_{j+1} := q_j + M_{j+1}$, which is constructed to be a high-probability upper bound on the true, unknown $(1-\alpha)$-quantile of the next distribution $\mathcal{D}(\pi_{j+1})$. Before deploying π_{j+1}, we solve the next planning problem $P[j{+}1; r_{j+1}]$ using r_{j+1} as the required safety tube radius. By construction, any policy that is feasible for $P[j{+}1; r_{j+1}]$ is guaranteed, with high probability, to remain safe against all environment trajectories in that inflated tube.

3.2 The Iterative Planning Algorithm

We follow the previously described "*recalibrate each episode, then transfer*" approach. Algorithmically, the central challenge is to compute the next radius r_{j+1}, which must ensure safety for the next policy π_{j+1} before we have even computed π_{j+1}. Formally, to compute π_{j+1} we need to know r_{j+1} for which we have to know M_{j+1}, which in turn depends on π_{j+1}, creating an implicit problem. We first define this problem and then present two ways to solve it: a computationally expensive but exact "implicit solver" and a computationally cheap "explicit solution" (used in our main analysis).

Implicit Safety. From (3.2), r_{j+1} must cover q_j and the robustification term M_{j+1} such that $r_{j+1} \geq q_j + M_{j+1}$. However, the robustification term M_{j+1} depends on the policy π_{j+1} that we are trying to find, since $\pi_{j+1} = \pi^\star(r_{j+1})$. This gives the true, implicit safety requirement:

$$r_{j+1} \geq q_j + \beta_T \|\pi^\star(r_{j+1}) - \pi_j\|_\infty . \tag{3.3}$$

We are now looking for the smallest $r_{j+1} \in \mathcal{R} := [r_{\min}, r_{\max}]$ that satisfies this inequality. We next present details for both solutions, which are summarized in Algorithm 1 in Appendix B.

Approach 1: The Implicit Solver. We enforce the implicit inequality in (3.3) directly, treating $\pi^\star(r)$ as a black box. At episode j, the quantities q_j, π_j, and β_T are fixed, so computing r_{j+1} reduces to a scalar one-dimensional program: we search for the smallest $r \in [q_j, r_{\max}]$ satisfying (3.3), which we solve by a bracketed line search and bisection or by a constrained solver. The procedure is detailed in Appendix C. The theoretical guarantees of this solver use the global sensitivity bound β_T from Lemma 2, but one may also substitute a data-driven estimate $\widehat{\beta}_T$ (see Appendix D).

Approach 2: The Tractable Explicit Solver. We here solve the implicit inequality (3.3) analytically. To get such an analytical bound, we assume that there exists a Lipschitz constant $L_U \geq 0$ that bounds how much the optimal policy $\pi^*(r)$ changes in response to a change in the radius $r \in \mathcal{R}$.

Assumption 3 There exists a constant $L_U > 0$ s.t. $\|\pi^*(r) - \pi^*(r')\|_\infty \leq L_U |r - r'|,\ \forall r, r' \in \mathcal{R}$.

The Lipschitz continuity property in Assumption 3 holds for many optimization-based planners. We provide a detailed analysis in Appendix A.2 for the common case in which the optimization problem $P[j; r]$ is feasible, convex, and has a unique, regular optimizer. From here, we now get

$$\beta_T \|\pi^\star(r_{j+1}) - \pi_j\|_\infty = \beta_T \|\pi^\star(r_{j+1}) - \pi^\star(r_j)\|_\infty \leq \beta_T L_U \, |r_{j+1} - r_j| = \kappa \, |r_{j+1} - r_j| \,,$$

where $\kappa := \beta_T L_U$ is a closed-loop gain that will help us analyze properties of our iterative planner. Instead of the implicit inequality (3.3), we can use this upper bound to get the sufficient inequality:

$$r_{j+1} \geq q_j + \kappa \, |r_{j+1} - r_j| \, . \tag{3.4}$$

The inequality (3.4) is a simple scalar inequality for r_{j+1}. As we will show in Section 3.3, this inequality has a unique, minimal (least conservative) solution, which is given by a closed-form expression. This is the algorithm we use in our analysis in the next section. Lastly, we remark that we can estimate L_U similarly to β_T (recall Appendix D).

3.3 Main Guarantees: Safety, Algorithmic Stability, Convergence, and Performance

Throughout this section, let the system in (2.1) be given and Assumptions 1 and 2 hold. We start our analysis by stating our episodic safety guarantees which are proven in Appendix A.3.

Theorem 1 (Per-Episode Safety Guarantee) *Let r_{j+1} follow the implicit safety requirement (3.3). Then $\mathbb{P}_{n_j}\{\mathbb{P}\{s(\hat{y}_{0:T}, y_{j+1,0:T}) \leq r_{j+1}\} \geq 1-\alpha\} \geq 1-\delta_j$. Furthermore, if $\mathbf{P}[j{+}1;\, r_{j+1}]$ is feasible, then $\mathbb{P}_{n_j}\{\mathbb{P}\{H(x_{j+1,0:T}, y_{j+1,0:T}) \leq 0\} \geq 1-\alpha\} \geq 1-\delta_j$.*

If H is Lipschitz continuous with Lipschitz constant L_H, we note that a sufficient condition for the feasibility of $\mathbf{P}[j{+}1;\, r_{j+1}]$ is that $H(x_{j+1,0:T}, \hat{y}_{0:T}) \leq -\, L_H \, r_{j+1}$, see Appendix A.4.

In the remainder of this section, let additionally Assumption 3 hold. We proceed our analysis using the explicit solver for the implicit inequality (3.3) which resulted in the tractable inequality in (3.4). We show its solution next and present the proof in Appendix A.5.

Lemma 3 (Safe Explicit Radius Update) *Let $\kappa < 1$. The minimal (i.e., least conservative) radius r_{j+1} that satisfies the tractable inequality (3.4) is given by the following closed-form solution:*

$$r_{j+1} = \begin{cases} \dfrac{q_j + \kappa r_j}{1+\kappa} & \text{if } q_j \leq r_j \quad (\textit{Shrinkage Branch}) \\ \dfrac{q_j - \kappa r_j}{1-\kappa} & \text{if } q_j > r_j \quad (\textit{Expansion Branch}) \end{cases} \tag{3.5}$$

If we limit r_{j+1} to an admissible interval $\mathcal{R} = [r_{\min}, r_{\max}]$ (e.g., as needed in Appendix A.2 to derive L_U), then we can project r_{j+1} onto $\mathcal{R}$ via the operator $\Pi_{\mathcal{R}}(r_{j+1}) = \min(r_{\max}, \max(r_{\min}, r_{j+1}))$. The rule (3.5) leads to the following guarantees, for which we provide the proof in Appendix A.6.

Theorem 2 (Episode-to-Episode Stability and Shrinkage) *Let $\kappa < 1$ and the radius r_{j+1} satisfy (3.5). Then, the per-episode change in the radius r_{j+1} is bounded by: $|r_{j+1} - r_j| \leq \frac{1}{1-\kappa} |q_j - r_j|$. Furthermore, if $q_j < r_j$, then it holds that $r_{j+1} < r_j$.*

Finally, we analyze conditions under which our algorithms converge. We study the true quantile of the nonconformity score $s(\hat{y}_{0:T}, y_{0:T})$ for $y_{0:T} \sim \mathcal{D}(\pi^\star(r))$. For simplicity, let $F_r(t) := \mathbb{P}\{s(\hat{y}_{0:T}, y_{j+1,0:T}) \leq t\}$ denote the cumulative distribution function (CDF) of the score and define the true quantile as $Q_{1-\alpha}(\pi^\star(r)) := \inf\{\, t \in \mathbb{R} : \; F_s(t) \geq 1-\alpha \,\}$. We now provide conditions for the convergence of r_j and q_j and provide the proof in Appendix A.7.

Theorem 3 (Convergence with Explicit Error Summation) *Let $T(r) := Q_{1-\alpha}\big(\pi^\star(r)\big)$ be the $(1-\alpha)$ true quantile of the nonconformity score. Assume that $T(r)$ is Lipschitz continuous so that*

$$|T(r) - T(r')| \leq \kappa \, |r - r'| \quad \textit{for all } r, r' \in \mathbb{R}_{\geq 0}, \qquad \kappa = \beta_T L_U \in (0,1).$$

Suppose further that there is a fixed point $r^\star \in \mathbb{R}_{\geq 0}$ with $T(r^\star) = r^\star$. At episode j, let the radius r_{j+1} satisfy (3.5). Define $e_j = |r_j - r^\star|$ and $\eta_j = q_j - T(r_j)$. Then, the following hold:

(P1) Finite-horizon error bound. *At each episode j, we are guaranteed that*

$$e_{j+1} \le \gamma_\kappa^{j+1} e_0 + B_\kappa \sum_{m=0}^{j} \gamma_\kappa^{j-m} |\eta_m|,\ \gamma_\kappa = \frac{2\kappa}{1-\kappa},\ B_\kappa = \frac{1}{1-\kappa},\ \gamma_\kappa < 1 \iff \kappa < \tfrac{1}{3}. \quad (3.6)$$

(P2) Closed form under a uniform perturbation bound. *If $|\eta_m| \le C$ for all $m \in \{0, \dots, j\}$, then*

$$e_{j+1} \le \gamma_\kappa^{j+1} e_0 + B_\kappa C \frac{1-\gamma_\kappa^{j+1}}{1-\gamma_\kappa},\ \text{and if } \kappa < \tfrac{1}{3}:\ \limsup_{j\to\infty} e_j \le \frac{B_\kappa}{1-\gamma_\kappa} C = \frac{1}{1-3\kappa} C. \quad (3.7)$$

(P3) High-probability control of η_j. *If the CDF $F_{r_j}(s)$ is differentiable and has density $f_{r_j}(s)$ no less than $f_\star > 0$ for all s in a sufficiently large neighborhood of its $(1-\alpha)$ quantile $Q_{1-\alpha}(\pi^\star(r_j))$, then we have $\mathbb{P}_{n_j}\{|\eta_j| \le A_j\} \ge 1-\delta_j$ and $\mathbb{P}_{\sum_{m=0}^{j} n_m}\{e_{j+1} \le \gamma_\kappa^{j+1} e_0 + B_\kappa \sum_{m=0}^{j} \gamma_\kappa^{j-m} A_m\} \ge 1 - \sum_{m=0}^{j} \delta_m$ where $A_m := \frac{|\alpha-\bar\alpha_m|}{f_\star} + \frac{1}{f_\star}\sqrt{\frac{\ln(2/\delta_m)}{2n_m}}$.*

(P4) Asymptotics. *If $n_j \to \infty$ and $\bar\alpha_j \to \alpha$, then $\mathbb{P}_{n_j}\{|\eta_j| \to 0\} \ge 1-\delta_j$, and for $\kappa < \frac{1}{3}$, $\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\{e_j \to 0\} \ge 1 - \sum_{m=0}^{\infty} \delta_m$. If instead $\bar\alpha_j \equiv \bar\alpha < \alpha$ and $n_j \to \infty$, then $\mathbb{P}_{n_j}\{|\eta_j| \to (\alpha-\bar\alpha)/f_\star\} \ge 1-\delta_j$, and $\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\{\limsup_{j\to\infty} e_j \le \frac{1}{1-3\kappa} \cdot \frac{\alpha-\bar\alpha}{f_\star}\} \ge 1 - \sum_{m=0}^{\infty} \delta_m$.*

Lastly, we study study convergence of the performance as measured via the function $J(\cdot)$. We note that $\mathbf{P}[j; r]$ in (2.5) depends on j only through r. Hence, let $V(r) := \min_\pi J(x_{0:T}(\pi), \pi)$ s.t. $H(x_{0:T}(\pi), \zeta) \le 0\ \forall\, \zeta \in \mathcal{C}_r(\hat y_{0:T})$ and $J_j := V(r_j)$. The next result is proven in Appendix A.9.

Theorem 4 (Performance convergence) *Let $V(r)$ have Lipschitz constant L_V on $\mathcal{R}$ so that $|V(r) - V(r')| \le L_V |r - r'|$ for all $r, r' \in \mathcal{R}$. Then, we have that $|J_{j+1} - J^\star| \le L_V\, e_{j+1}$ where $J^\star := V(r^\star)$ and $r^\star$, e_j, η_j, A_j are defined in Theorem 3. Furthermore, the following hold:*

(P1) Improvement over the initial policy. *If $r_0 > r^\star$ and $V(r)$ is strictly increasing on $[r^\star, r_0]$, then $J_j < J_0$ for any $r_j \in [r^\star, r_0)$. If there exists $m_V > 0$ such that $V(r) - V(r') \ge m_V(r - r')$ for all $r, r' \in [r^\star, r_0]$, then every $r_j \le r_0$ satisfies $J_0 - J_j \ge m_V(r_0 - r^\star - e_j)$.*

(P2) One-step improvement. *If $r_j \ge r^\star$ and $|\eta_j| < (1-\kappa)(r_j - r^\star)$, then $q_j < r_j$ and $J_{j+1} \le J_j$. If $A_j < (1-\kappa)(r_j - r^\star)$ for all $j \in \{J, \dots, K\}$, then $\mathbb{P}_{\sum_{j=J}^{K} n_j}\{J_{j+1} \le J_j\ , \forall\, j\} \ge 1 - \sum_{j=J}^{K} \delta_j$.*

(P3) Asymptotics. *Under the assumptions in Theorem 3(P4), $\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\{J_j \to J^\star\} \ge 1 - \sum_{m=0}^{\infty} \delta_m$. If $r_0 > r^\star$, then there exists j_0 such that $\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\{J_j \le J_0\ , \forall\, j \ge j_0\} \ge 1 - \sum_{m=0}^{\infty} \delta_m$.*

4 Case Studies

We consider two case studies: a two-dimensional car–pedestrian interaction (Section 4) and a multi-quadcopter navigation task in QuadSwarm Huang et al. (2023b) (in Appendix E). Both instantiate the program $\mathbf{P}[j; r_j]$ in (2.5) with $\alpha := 0.1$, $\delta_j := 0.05$, $r_0 = r_{\max}$, and $n_j := 1000$. We report the radius r_j, the cost $J_j := V(r_j)$, and the empirical coverages for $\mathbb{P}\{y_{j,0:T} \in \mathcal{C}_{r_j}(\hat y_{0:T})\}$ and $\mathbb{P}\{H(x_{j,0:T}, y_{j,0:T}) \le 0\}$ at target $1 - \alpha = 0.9$ to show Theorem 1. We compare five rules that differ in (i) whether they recalibrate q_j each episode, (ii) whether they transfer the certificate through the policy update via (3.3) or (3.5), and (iii) how the transfer is resolved: *Robust CP – implicit (ours)* enforces (3.3) directly (see Appendix C); *Robust CP – explicit (ours)* uses (3.5) with fixed $\kappa = \beta_T L_U$; *Naive CP* recalibrates but drops the transfer, setting $r_{j+1} = q_j$ and ignoring the perturbation M_{j+1} in (3.2); *One-time CP* fixes $q_j \equiv q_{\text{base}}$ from the first episode but keeps the transfer; *Split CP* computes q_{base} from non-interactive rollouts and sets $r_j \equiv q_{\text{base}}$.

4.1. Two-Dimensional Car–Pedestrian Interaction The ego state $x_t \in \mathbb{R}^2$ follows single-integrator dynamics $x_{t+1} = x_t + \Delta u_t$ with $\Delta := 0.1$ and $\|u_t\|_\infty \le 5$. The pedestrian state

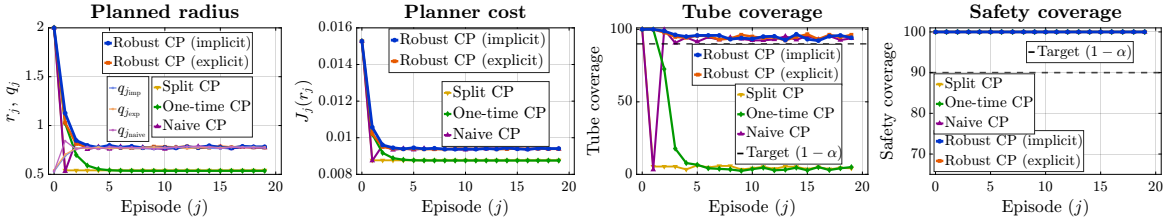

Figure 3: Car–pedestrian metrics. Left to right: r_j, J_j, empirical tube and safety coverages across episodes.

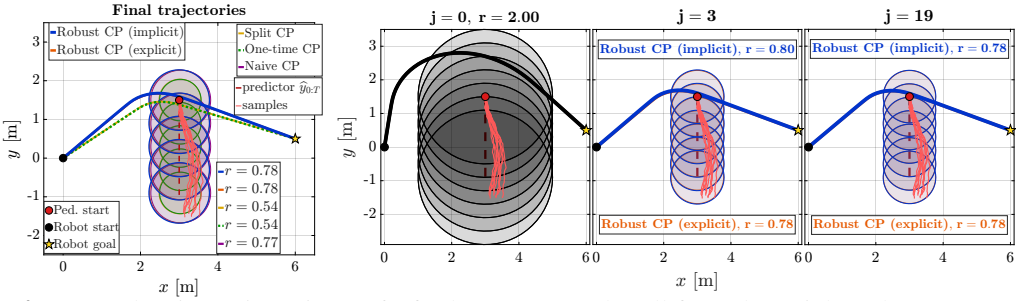

Figure 4: Car–pedestrian trajectories. Left: final ego agent paths, all five rules. Right: tubes at $j = 0, 3, 19$.

$y_t \in \mathbb{R}^2$ follows interactive dynamics $y_{t+1} = y_t + \Delta(v_0 + \phi(\|r_t\|_2)\, e_t) + w_t$ with relative position $r_t := y_t - x_t$, direction $e_t := r_t/\|r_t\|_2$, nominal velocity $v_0 := [-0.5, 0]^\top$, noise $w_t \sim \mathcal{N}(0, \Delta\sigma^2 I)$ for $\sigma := 0.05$, and repulsion $\phi(s) := v_{\max}\ell_c^2/(s^2 + \ell_c^2)$ with $v_{\max} := 1$ and $\ell_c := 1$. The predictor $\hat{y}_{t+1} = \hat{y}_t + \Delta v_0$ is intentionally misspecified: it omits $\phi(\cdot)$. The cost is $J(x, u) := \|x_T - x_{\text{goal}}\|_2^2 + 10^{-3}\sum_{t=0}^{T-1}\|u_t\|_2^2$ with $x_{\text{goal}} := [6, 0.5]^\top$ and $T := 50$; the safety function is $H(x, y) := \max_t(d_{\min} - \|x_t - y_t\|_2)$ with $d_{\min} := 0.8$. By the triangle inequality, the constraint $H(x, \zeta) \le 0\ \forall\, \zeta \in \mathcal{C}_{r_j}(\hat{y}_{0:T})$ reduces to $\|x_{j,t} - \hat{y}_t\|_2 \ge d_{\min} + r_j$ for all $t = 0, \ldots, T$.

Observations. Figure 3 shows that under both *Robust CP* variants the radius r_j contracts from $r_0 = r_{\max}$ and stabilizes near $r_j \approx 0.78$, the planner cost J_j decreases from J_0 and stabilizes, and both tube and safety coverages remain at the target $1 - \alpha$, in agreement with Theorem 1 and Theorem 4. Both solvers converge to the same steady-state radius and cost, with the implicit solver yielding a marginally larger radius during the transient ($r \approx 0.80$ vs. $r \approx 0.78$ at $j = 3$, see Figure 4). *Naive CP* sets $r_{j+1} = q_j$, omitting M_{j+1} in (3.2): tube coverage drops to near zero in the early episodes before recovering once the radius stabilizes, precisely the tube coverage failure that (3.2) is designed to prevent. *One-time CP* and *Split CP* hold q_j fixed and therefore cannot track $\mathcal{D}(\pi_j)$; their tube coverage collapses to near zero as the stale q_{base} does not account for the policy-induced distribution shift per (3.2), and while their smaller radii ($r_j \approx 0.54$) yield lower planner costs, these lack valid coverage guarantees. Only the two *Robust CP* variants simultaneously maintain both coverages and reduce J_j from the conservative initialization, matching the shrinkage branch of (3.5) under Theorem 4. Figure 4 visualizes the trajectories and tube contraction at episodes $j = 0, 3, 19$.

4.2 Discussion Across both case studies, the two *Robust CP* variants are the only rules that simultaneously (i) maintain empirical tube and safety coverages at the target $1 - \alpha$ and (ii) contract r_j from $r_0 = r_{\max}$, yielding a reduction in J_j in the car–pedestrian study and an increase in cumulative reward in the multi-quadcopter study. *Naive CP* omits M_{j+1} in (3.2) and exhibits early-episode tube undercoverage in the car–pedestrian study, recovering only after the radius stabilizes—precisely the failure mode that (3.2) is designed to prevent. *One-time CP* and *Split CP* cannot track $\mathcal{D}(\pi_j)$ and lose tube coverage; while their smaller radii yield lower planner costs, these lack valid coverage guarantees per (3.2). Per-episode safety is delivered by the transfer requirement (3.2) through Theorem 1, and the cost improvement is delivered by the shrinkage branch of (3.5) through Theorem 4.

References

Shadi Alijani and Homayoun Najjaran. Wqlcp: Weighted adaptive conformal prediction for robust uncertainty quantification under distribution shifts, 2025. URL https://arxiv.org/abs/2505.19587.

Aaron D. Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. *European Control Conference (ECC)*, pages 3420–3431, 2019.

Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

Liviu Aolaritei, Zheyu Oliver Wang, Julie Zhu, Michael I Jordan, and Youssef Marzouk. Conformal prediction under levy-prokhorov distribution shifts: Robustness to local and global perturbations. *arXiv preprint arXiv:2502.14105*, 2025.

Jan-Peter Calliess. Lazily adapted constant kinky inference for nonparametric regression and model-reference adaptive control. *arXiv preprint arXiv:1701.00178*, 2017. URL https://arxiv.org/abs/1701.00178.

Maxime Cauchois, Suyash Gupta, Alnur Ali, and John C Duchi. Robust validation: Confident predictions even when distributions shift. *Journal of the American Statistical Association*, 119 (548):3033–3044, 2024.

Matthew Cleaveland, Insup Lee, George J Pappas, and Lars Lindemann. Conformal prediction regions for time series using linear complementarity programming. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20984–20992, 2024.

Anushri Dixit, Lars Lindemann, Skylar X Wei, Matthew Cleaveland, George J Pappas, and Joel W Burdick. Adaptive conformal prediction for motion planning among dynamic agents. In *Proceedings of The 5th Annual Learning for Dynamics and Control Conference*, volume 211 of *Proceedings of Machine Learning Research*, pages 300–314. PMLR, 2023. URL https://proceedings.mlr.press/v211/dixit23a.html.

Achref Doula, Max Mühlhäuser, and Alejandro Sanchez Guinea. Safepath: Conformal prediction for safe llm-based autonomous navigation. *arXiv preprint arXiv:2505.09427*, 2025.

Noel E Du Toit and Joel W Burdick. Robot motion planning in dynamic, uncertain environments. *IEEE Transactions on Robotics*, 28(1):101–115, 2011.

John Duchi. Sample-conditional coverage in split-conformal prediction. In *Adv. Neural Inf. Process. Syst.*, 2025.

Michael Everett, Yu Fan Chen, and Jonathan P How. Collision avoidance in pedestrian-rich environments with deep reinforcement learning. *IEEE Access*, 9:10357–10377, 2021.

Zhaohan Feng, Ruiqi Xue, Lei Yuan, Yang Yu, Ning Ding, Meiqin Liu, Bingzhao Gao, Jian Sun, and Gang Wang. Multi-agent embodied ai: Advances and future directions, 2025. URL https://arxiv.org/abs/2505.05108.

Asaf Gendler, Tsui-Wei Weng, Luca Daniel, and Yaniv Romano. Adversarially robust conformal prediction. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=9L1BsI4wP1H. OpenReview ID: 9L1BsI4wP1H.

Isaac Gibbs and Emmanuel Candes. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems*, volume 34, pages 1660–1672, 2021.

Ting-Wei Hsu and Hiroyasu Tsukamoto. Statistical guarantees in data-driven nonlinear control: Conformal robustness for stability and safety. *IEEE Control Systems Letters*, 2025.

Julien Walden Huang, Stephen Roberts, and Jan-Peter Calliess. On the sample complexity of Lipschitz constant estimation. *Transactions on Machine Learning Research*, 2023a.

Zhe Huang, Tianchen Ji, Heling Zhang, Fatemeh Cheraghi Pouria, Katherine Driggs-Campbell, and Roy Dong. Interaction-aware conformal prediction for crowd navigation. *arXiv preprint arXiv:2502.06221*, 2025.

Zhehui Huang, Sumeet Batra, Tao Chen, Rahul Krupani, Tushar Kumar, Artem Molchanov, Aleksei Petrenko, James A. Preiss, Zhaojing Yang, and Gaurav S. Sukhatme. Quadswarm: A modular multi-quadrotor simulator for deep reinforcement learning with direct thrust control, 2023b. URL https://arxiv.org/abs/2306.09537.

Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The Int. Journal Robot. Research*, 35 (11):1289–1307, 2016.

Lars Lindemann, Matthew Cleaveland, Gihyun Shim, and George J. Pappas. Safe planning in dynamic environments using conformal prediction. *IEEE Robotics and Automation Letters*, 8(8): 5116–5123, 2023. doi: 10.1109/LRA.2023.3292071.

Lars Lindemann, Yiqi Zhao, Xinyi Yu, George J Pappas, and Jyotirmoy V Deshmukh. Formal verification and control with conformal prediction. *arXiv preprint arXiv:2409.00536*, 2024.

Christoforos Mavrogiannis, Francesca Baldini, Allan Wang, Dapeng Zhao, Pete Trautman, Aaron Steinfeld, and Jean Oh. Core challenges of social robot navigation: A survey. *ACM Transactions on Human-Robot Interaction*, 12(3):1–39, 2023.

Shili Sheng, Pian Yu, David Parker, Marta Kwiatkowska, and Lu Feng. Safe pomdp online planning among dynamic agents via adaptive conformal prediction. *IEEE Robotics and Automation Letters*, 2024.

Jaeuk Shin, Jungjin Lee, and Insoon Yang. Egocentric conformal prediction for safe and efficient navigation in dynamic cluttered environments. *arXiv preprint arXiv:2504.00447*, 2025.

Jiankai Sun, Yiqi Jiang, Jianing Qiu, Parth Nobel, Mykel J Kochenderfer, and Mac Schwager. Conformal prediction for uncertainty-aware planning with diffusion dynamics model. *Advances in Neural Information Processing Systems*, 36:80324–80337, 2023.

Sophia Sun and Rose Yu. Copula conformal prediction for multi-step time series forecasting. *arXiv preprint arXiv:2212.03281*, 2022.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32, 2019.

Sander Tonkens, Sophia Sun, Rose Yu, and Sylvia Herbert. Scalable safe long-horizon planning in dynamic environments leveraging conformal prediction and temporal correlations. In *Long-Term Human Motion Prediction Workshop, International Conference on Robotics and Automation*, 2023.

Peter Trautman and Andreas Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 797–803. IEEE, 2010.

Vladimir Vovk. Conditional validity of inductive conformal predictors. In *Asian conference on machine learning*, pages 475–490. PMLR, 2012.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

Jun Wang, Guocheng He, and Yiannis Kantaros. Probabilistically correct language-based multi-robot planning using conformal prediction. *IEEE Robotics and Automation Letters*, 2024.

Shuqi Wang, Yue Gao, and Xiang Yin. Learning-based conformal tube mpc for safe control in interactive multi-agent systems. *arXiv preprint arXiv:2504.03293*, 2025.

Yachong Yang, Arun Kumar Kuchibhotla, and Eric Tchetgen Tchetgen. Doubly robust calibration of prediction sets under covariate shift. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(4):943–965, 2024.

Jianpeng Yao, Xiaopan Zhang, Yu Xia, Zejin Wang, Amit K Roy-Chowdhury, and Jiachen Li. Sonic: Safe social navigation with adaptive conformal inference and constrained reinforcement learning. *arXiv preprint arXiv:2407.17460*, 2024.

Xinyi Yu, Yiqi Zhao, Xiang Yin, and Lars Lindemann. Signal temporal logic control synthesis among uncontrollable dynamic agents with conformal prediction. *Automatica*, 183:112616, 2026. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2025.112616. URL https://www.sciencedirect.com/science/article/pii/S0005109825005126.

Margaux Zaffran, Olivier Féron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive conformal predictions for time series. In *International Conference on Machine Learning*, pages 25834–25866. PMLR, 2022.

Matteo Zecchin, Sangwoo Park, and Osvaldo Simeone. Forking uncertainties: Reliable prediction and model predictive control with sequence models via conformal risk control. *IEEE Journal on Selected Areas in Information Theory*, 5:44–61, 2024.

Junhui Zhang, Bardh Hoxha, Georgios Fainekos, and Dimitra Panagou. Conformal prediction in the loop: Risk-aware control barrier functions for stochastic systems with data-driven state estimators. *IEEE Control Systems Letters*, 2025.

Yiqi Zhao, Bardh Hoxha, Georgios Fainekos, Jyotirmoy V Deshmukh, and Lars Lindemann. Robust conformal prediction for stl runtime verification under distribution shift. In *2024 ACM/IEEE 15th International Conference on Cyber-Physical Systems (ICCPS)*, pages 169–179. IEEE, 2024.

Appendix A. Deferred Proofs

A.1. Proof of Lemma 2

Let us first recall the dynamics of ego and uncontrollable agents from (2.1) as

$$x_{t+1} = f_X(x_t, u_t), \qquad y_{t+1} = f_Y(y_t, x_t, u_t, \nu_t),$$

under the noise sequence $\{\nu_t\}_{t=0}^{T-1}$. Under Assumption 1, i.e., Lipschitz continuity of f_X and f_Y, our goal is now to bound the deviation in the environment trajectory from $y_{0:T}$ to $y'_{0:T}$ caused by a change in the policy from $\pi = \{u_t\}_{t=1}^{T-1}$ to $\pi' = \{u'_t\}_{t=1}^{T-1}$. Indeed, we want to show that

$$\|x_{0:T} - x'_{0:T}\|_\infty \leq A_T\, \Delta u, \qquad A_T := L_{Xu} \sum_{t=0}^{T-1} (L_{Xx})^t, \tag{A.1}$$

$$\|y_{0:T} - y'_{0:T}\|_\infty \leq \beta_T\, \Delta u, \qquad \beta_T := (L_{Yx} A_T + L_{Yu}) \sum_{t=0}^{T-1} (L_{Yy})^t. \tag{A.2}$$

where $\Delta u := \|\pi - \pi'\|_\infty$ with $\|\pi - \pi'\|_\infty := \max_{0\leq t\leq N-1} \|u_t - u'_t\|_2$.

Proof Let $\Delta x_t := x_t - x'_t$, $\Delta y_t := y_t - y'_t$, and $\Delta u_t := u_t - u'_t$. From Assumption 1, we get

$$\|\Delta x_{t+1}\|_2 \leq L_{Xx} \|\Delta x_t\|_2 + L_{Xu} \|\Delta u_t\|_2. \tag{A.3}$$

Unrolling (A.3) until time t with $\Delta x_0 = 0$ and $\|\Delta u_t\|_2 \leq \Delta u$ then gives

$$\|\Delta x_t\|_2 \leq L_{Xu} \sum_{i=0}^{t-1} (L_{Xx})^i\, \Delta u. \tag{A.4}$$

Since all Lipschitz constants are nonnegative, the partial sum in (A.4) is monotone in t so that

$$\|x_{0:T} - x'_{0:T}\|_\infty = \max_{0\leq t\leq T} \|\Delta x_t\|_2 = \|\Delta x_T\|_2 \leq L_{Xu} \sum_{t=0}^{T-1} (L_{Xx})^t\, \Delta u,$$

which proves (A.1). Next, from Assumption 1, we get

$$\|\Delta y_{t+1}\|_2 \leq L_{Yy} \|\Delta y_t\|_2 + L_{Yx} \|\Delta x_t\|_2 + L_{Yu} \|\Delta u_t\|_2.$$

Using (A.4) and $\|\Delta u_t\|_2 \leq \Delta u$, it the follows that

$$\|\Delta y_{t+1}\|_2 \leq L_{Yy} \|\Delta y_t\|_2 + \big(L_{Yx} A_T + L_{Yu}\big) \Delta u. \tag{A.5}$$

Unrolling (A.5) until time t with $\Delta y_0 = 0$ then gives

$$\|\Delta y_t\|_2 \leq \big(L_{Yx} A_T + L_{Yu}\big) \sum_{i=0}^{t-1} (L_{Yy})^i\, \Delta u.$$

Again by nonnegativity of the Lipschitz constants, this sum is monotone in t so that

$$\|y_{0:T} - y'_{0:T}\|_\infty = \|\Delta y_T\|_2 \leq \big(L_{Yx} A_T + L_{Yu}\big) \sum_{t=0}^{T-1} (L_{Yy})^t\, \Delta u,$$

which proves (A.2), what was to be shown. ■

A.2. Derivation of Planner Sensitivity L_U

Proposition 1 (Planner Regularity & Lipschitzness) *Assume there exists a compact interval of radii $\mathcal{R} := [r_{\min}, r_{\max}]$ such that for all $r \in \mathcal{R}$, the robust planning problem $\mathbf{P}[j; r_j]$ in (2.5) is feasible, convex, and has a unique, regular optimizer $\pi^*(r)$ satisfying standard KKT conditions (LICQ, strict complementarity, and a nonsingular KKT matrix). Then the solution map $r \mapsto \pi^*(r)$ is locally Lipschitz on $\mathcal{R}$. In particular, there exists a uniform constant $L_U < \infty$ such that*

$$\|\pi^*(r) - \pi^*(r')\|_\infty \;\le\; L_U\,|r - r'|\,, \qquad \forall\, r, r' \in \mathcal{R}. \tag{A.6}$$

Proof We derive L_U via parametric sensitivity of $\mathbf{P}[j; r]$ in (2.5). Throughout, we assume the inequality constraints depend affinely on the radius,

$$a_k(\pi) + b_k\, r \le 0 \quad (k = 1, \dots, m),$$

and any equality constraints $h(\pi) = 0$ do not depend on r.[6][7]

Active set and multipliers. Fix $r \in \mathcal{R}$ and let $\pi^*(r)$ denote the unique optimizer. Define the active set

$$\mathcal{A}(r) \;:=\; \big\{\, k \;:\; a_k(\pi^*(r)) + b_k\, r = 0 \,\big\},$$

and let $\lambda_k^*(r) > 0$ be the corresponding Lagrange multipliers for $k \in \mathcal{A}(r)$ (strict complementarity). Let $\mu^*(r)$ be the multipliers for $h(\pi) = 0$ (if present).

Second-order objects. Define the Hessian of the Lagrangian, the Jacobian of active constraints, and the vector of active radius coefficients:

$$H(r) := \nabla^2_{\pi\pi} J(\pi^*(r)) + \sum_{k \in \mathcal{A}(r)} \lambda_k^*(r)\, \nabla^2_{\pi\pi} a_k(\pi^*(r)), \tag{A.7}$$

$$G_{\mathcal{A}}(r) := \big[\nabla_\pi a_k(\pi^*(r))\big]_{k \in \mathcal{A}(r)} \in \mathbb{R}^{|\mathcal{A}(r)| \times n}, \tag{A.8}$$

$$b_{\mathcal{A}}(r) := \big[b_k\big]_{k \in \mathcal{A}(r)} \in \mathbb{R}^{|\mathcal{A}(r)|}. \tag{A.9}$$

If equalities are present, let $E(r) := \nabla_\pi h(\pi^*(r))$.

KKT conditions. At $(\pi^*, \lambda^*, \mu^*; r)$ the KKT conditions read

$$\nabla_\pi J(\pi^*(r)) \;+\; \sum_{k \in \mathcal{A}(r)} \lambda_k^*(r)\, \nabla_\pi a_k(\pi^*(r)) \;+\; E(r)^\top \mu^*(r) \;=\; 0, \tag{A.10}$$

$$a_k(\pi^*(r)) + b_k\, r \;\le\; 0 \;\; (k = 1, \dots, m), \qquad h(\pi^*(r)) \;=\; 0, \tag{A.11}$$

$$\lambda_k^*(r) \;\ge\; 0 \;\; (k = 1, \dots, m), \qquad \lambda_k^*(r)\,\big(a_k(\pi^*(r)) + b_k\, r\big) \;=\; 0 \;\; (k = 1, \dots, m). \tag{A.12}$$

Under LICQ, strict complementarity, and the second-order condition, the KKT matrix is nonsingular (strong regularity), and the active set $\mathcal{A}(r)$ is locally constant.

6. For instance, this is the case for the program in (2.5) when H is Lipschitz continuous with Lipschitz constant L_H so that $H(x_{j,0:T}, \zeta) \le 0 \quad \forall\, \zeta \in \mathcal{C}_{r_j}(\hat{y}_{0:T})$ can be rewritten as $H(x_{j,0:T}, \hat{y}_{0:T}) \;\le\; -\,L_H\, r_j$ (see Appendix A.4).
7. If h (or the gradients of a_k) depend on r, the derivative formula below includes additional right-hand-side terms; see Remark 1. The Lipschitz conclusion (A.6) is unchanged under strong regularity.

Linearized KKT system. Differentiate (A.10)–(A.12) with respect to r over any subinterval where $\mathcal{A}(r)$ is fixed. With $E \equiv E(r)$ and dropping the explicit r for brevity,

$$\begin{bmatrix} H & G_{\mathcal{A}}^{\top} & E^{\top} \\ G_{\mathcal{A}} & 0 & 0 \\ E & 0 & 0 \end{bmatrix} \begin{bmatrix} \dfrac{d\pi^*}{dr} \\ \dfrac{d\lambda^*_{\mathcal{A}}}{dr} \\ \dfrac{d\mu^*}{dr} \end{bmatrix} = \begin{bmatrix} 0 \\ -\,b_{\mathcal{A}} \\ 0 \end{bmatrix}. \tag{A.13}$$

Nonsingularity of the KKT matrix (strong regularity) ensures a unique solution to (A.13).

Sensitivity formula. Eliminate multipliers in (A.13) via the Schur complement. If no equalities are present ($E = 0$), we obtain the classical expression

$$\frac{d\pi^*(r)}{dr} \;=\; -\,H(r)^{-1}\,G_{\mathcal{A}}(r)^{\top}\Big(G_{\mathcal{A}}(r)\,H(r)^{-1}\,G_{\mathcal{A}}(r)^{\top}\Big)^{-1}\,b_{\mathcal{A}}(r). \tag{A.14}$$

With equalities, replace $G_{\mathcal{A}}$ by $W(r) := \begin{bmatrix} G_{\mathcal{A}}(r) \\ E(r) \end{bmatrix}$ and $b_{\mathcal{A}}$ by $c_{\mathcal{A}}(r) := \begin{bmatrix} b_{\mathcal{A}}(r) \\ 0 \end{bmatrix}$ in (A.14).

Uniform bound and Lipschitz continuity. Define

$$L_U \;:=\; \sup_{r\in\mathcal{R}} \Big\| H(r)^{-1} G_{\mathcal{A}}(r)^{\top}\Big(G_{\mathcal{A}}(r) H(r)^{-1} G_{\mathcal{A}}(r)^{\top}\Big)^{-1} b_{\mathcal{A}}(r) \Big\|_{\infty}. \tag{A.15}$$

Continuity of the matrices in (A.7)–(A.9) and uniform nonsingularity on each fixed-active-set region imply the supremum in (A.15) is finite. The mean-value theorem then yields (A.6). ■

Remark 1 (Radius-dependent data) *If $h(\pi, r)$ depends on r or if $\nabla_\pi a_k(\pi, r)$ depends on r, the right-hand side of (A.13) becomes $\big[\, g^{\top},\, -b_{\mathcal{A}}^{\top},\, -d_h^{\top} \big]^{\top}$, where $g := \partial_r\big(\nabla_\pi J + \sum_{k\in\mathcal{A}} \lambda_k^* \nabla_\pi a_k + E^{\top}\mu^*\big)$ and $d_h := \partial_r h(\pi^*, r)$. The proof proceeds identically, and the Lipschitz property (A.6) continues to hold under strong regularity.*

A.3. Proof of Theorem 1

Let $y_{j+1,0:T}$ and $y_{j,0:T}$ be trajectories generated by the control policies π_{j+1} and π_j in episodes $j+1$ and j, respectively. Recall from our previous discussion and Lemma 2 that

$$\|y_{j+1,0:T} - y_{j,0:T}\|_{\infty} \;\le\; \beta_T\, \|\pi_{j+1} - \pi_j\|_{\infty}. \tag{A.16}$$

The nonconformity score s defined in (2.3) has Lipschitz constant one so that we have

$$|s(\hat{y}_{0:T}, y_{j,0:T}) - s(\hat{y}_{0:T}, y_{j+1,0:T})| \le \beta_T\, \|\pi_{j+1} - \pi_j\|_{\infty}.$$

Next, applying adversarial CP from Lemma 1 immediately results in the guarantee

$$\mathbb{P}_{n_j}\big\{\, \mathbb{P}\big\{\, s(\hat{y}_{0:T}, y_{j+1,0:T}) \;\le\; q_j + \beta_T\, \|\pi_{j+1} - \pi_j\|_{\infty} \big\} \;\ge\; 1-\alpha \big\} \ge 1 - \delta_j.$$

Since the implicit safety requirement in (3.3) holds by assumption, we know that $r_{j+1} \ge q_j + \beta_T\, \|\pi^\star(r_{j+1}) - \pi_j\|_{\infty}$ is satisfied. This in turn implies that

$$\mathbb{P}_{n_j}\big\{\, \mathbb{P}\big\{\, s(\hat{y}_{0:T}, y_{j+1,0:T}) \le r_{j+1} \big\} \ge 1-\alpha \,\big\} \;\ge\; 1 - \delta_j, \tag{A.17}$$

which is the first claim of Theorem 1.

Assume now that the optimization problem $\mathbf{P}[j{+}1;\, r_{j+1}]$ in (2.5) is feasible. This means that

$$H(x_{j+1,0:T}, \zeta) \le 0 \quad \forall\, \zeta \in \mathcal{C}_{r_{j+1}}(\hat{y}_{0:T}) \tag{A.18}$$

where we recall that the set $\mathcal{C}_{r_{j+1}}(\hat{y}_{0:T})$ is defined as $\mathcal{C}_{r_{j+1}}(\hat{y}_{0:T}) := \{\, y_{0:T} : s(\hat{y}_{0:T}, y_{0:T}) \leq r_{j+1} \,\}$. Combining (A.17) and (A.18) implies that

$$\mathbb{P}_{n_j}\{\, \mathbb{P}\{\, H(x_{j+1,0:T}, y_{j+1,0:T}) \leq 0 \,\} \geq 1-\alpha \} \;\geq\; 1-\delta_j,$$

which is the second claim of Theorem 1.

A.4. Sufficient Condition for Feasibility of $\mathbf{P}[j+1;\, r_{j+1}]$ in (2.5)

First note that feasibility of the optimization problem $\mathbf{P}[j;\, r_j]$ in (2.5) is guaranteed if the constraint

$$H(x_{j,0:T}, \zeta) \leq 0 \quad \forall\, \zeta \in \mathcal{C}_{r_j}(\hat{y}_{0:T}) \tag{A.19}$$

is satisfied by the trajectory $x_{j,0:T}$, where we recall that $\mathcal{C}_{r_j}(\hat{y}_{0:T}) = \{\, y_{0:T} : \|y_{0:T} - \hat{y}_{0:T}\|_\infty \leq r_j \,\}$.

Assume now that there exists a Lipschitz constant $L_H \geq 0$ such that

$$\big|\, H(x_{0:T}, y_{0:T}) - H(x_{0:T}, y'_{0:T}) \,\big| \;\leq\; L_H \, \|y_{0:T} - y'_{0:T}\|_\infty$$

for all permissible $x_{0:T}$ and all $y_{0:T}, y'_{0:T}$. Consequently, if the trajectory $x_{j,0:T}$ satisfies

$$H(x_{j,0:T}, \hat{y}_{0:T}) \;\leq\; -\, L_H \, r_j,$$

then, for any $y_{0:T} \in \mathcal{C}_{r_j}(\hat{y}_{0:T})$, we know that

$$\begin{aligned} H(x_{j,0:T}, y_{0:T}) &\leq H(x_{j,0:T}, \hat{y}_{0:T}) + \big|H(x_{j,0:T}, y_{0:T}) - H(x_{j,0:T}, \hat{y}_{0:T})\big| \\ &\leq -\, L_H r_j + L_H \, \|y_{0:T} - \hat{y}_{0:T}\|_\infty \;\leq\; -\, L_H r_j + L_H r_j \;=\; 0, \end{aligned}$$

by which we have shown that (A.19) is satisfied so that $\mathbf{P}[j;\, r_j]$ is feasible.

A.5. Proof of Lemma 3

Our goal is to derive the solution to the tractable inequality in (3.4), which we recall for convenience:

$$r_{j+1} \;\geq\; q_j \;+\; \kappa \, \big|r_{j+1} - r_j\big|. \tag{A.20}$$

By assumption, we have that $\kappa < 1$. Since the right-hand side of (A.20) is nondecreasing in r_{j+1}, the minimal solution is obtained by saturating (A.20). Due to the absolute value in $|r_{j+1} - r_j|$ in (A.20), we have to consider the two cases of shrinkage ($r_{j+1} \leq r_j$) and expansion ($r_{j+1} > r_j$).

Case 1: Shrinkage ($r_{j+1} \leq r_j$). We have that $|r_{j+1} - r_j| = r_j - r_{j+1}$ so that (A.20) becomes

$$r_{j+1} \;\geq\; q_j \;+\; \kappa\,(r_j - r_{j+1}) \;\Longleftrightarrow\; (1+\kappa)\, r_{j+1} \;\geq\; q_j + \kappa r_j.$$

Since $1 + \kappa > 0$, the minimal value of r_{j+1} saturates the inequality, i.e.,

$$r_{j+1} \;=\; \frac{q_j + \kappa r_j}{1+\kappa}. \tag{A.21}$$

Consistency with the shrinkage assumption requires

$$\frac{q_j + \kappa r_j}{1+\kappa} \;\leq\; r_j \;\Longleftrightarrow\; q_j \;\leq\; r_j.$$

Case 2: Expansion ($r_{j+1} > r_j$). We have that $|r_{j+1} - r_j| = r_{j+1} - r_j$ so that (A.20) becomes

$$r_{j+1} \;\geq\; q_j \;+\; \kappa\,(r_{j+1} - r_j) \;\Longleftrightarrow\; (1-\kappa)\, r_{j+1} \;\geq\; q_j - \kappa r_j.$$

Since $1 - \kappa > 0$, the minimal value of r_{j+1} saturates the inequality, i.e.,

$$r_{j+1} \;=\; \frac{q_j - \kappa r_j}{1-\kappa}. \tag{A.22}$$

Consistency with the expansion assumption requires

$$\frac{q_j - \kappa r_j}{1-\kappa} \;>\; r_j \iff q_j \;>\; r_j.$$

In conclusion, the above condition partition the line by $q_j \le r_j$ and $q_j > r_j$. Combining this with (A.21) and (A.22) yields the closed-form solution in (3.5), which was to be shown.

A.6. Proof of Theorem 2

We analyze the magnitude $|r_{j+1} - r_j|$ in the same two cases considered in Theorem 3.

- **Case 1: Shrinkage ($q_j \le r_j$).** We can derive that

$$|r_{j+1} - r_j| = r_j - r_{j+1} = r_j - \left(\frac{q_j + \kappa r_j}{1+\kappa}\right) = \frac{r_j(1+\kappa) - q_j - \kappa r_j}{1+\kappa} = \frac{r_j - q_j}{1+\kappa} = \frac{|q_j - r_j|}{1+\kappa}. \tag{A.23}$$

- **Case 2: Expansion ($q_j > r_j$).** We can derive that

$$|r_{j+1} - r_j| = r_{j+1} - r_j = \left(\frac{q_j - \kappa r_j}{1-\kappa}\right) - r_j = \frac{q_j - \kappa r_j - r_j(1-\kappa)}{1-\kappa} = \frac{q_j - r_j}{1-\kappa} = \frac{|q_j - r_j|}{1-\kappa}.$$

Since $\kappa < 1$ while being positive, we have $1-\kappa < 1+\kappa$. The denominator is smaller in the expansion case, so the worst-case bound for $|r_{j+1} - r_j|$ is given as

$$|r_{j+1} - r_j| \le \frac{1}{1-\kappa} |q_j - r_j|, \tag{A.24}$$

which was to be proven. Finally, in the case that $q_j < r_j$, we see that the numerator in (A.23) is positive, and since $1+\kappa > 0$, the change is positive, meaning $r_{j+1} < r_j$.

A.7. Proof of Theorem 3

Step 1: Setup and notation. For the convenience of the reader, we first recall our notation. Let

$$T(r) \;=\; Q_{1-\alpha}\big(\pi^\star(r)\big)$$

denote the $(1-\alpha)$ true population quantile under policy $\pi^\star(r)$. We assume the Lipschitz property

$$\big|T(r) - T(r')\big| \;\le\; \kappa\,|r - r'| \qquad \text{for all } r, r' \in \mathbb{R}_{\ge 0},$$

where $\kappa = \beta_T L_U \in (0,1)$ is the closed-loop gain. Suppose there exists a fixed point $r^\star \in \mathbb{R}_{\ge 0}$ with

$$T(r^\star) \;=\; r^\star.$$

Let the error and calibration perturbation at episode j be

$$e_j \;=\; |r_j - r^\star|, \qquad \eta_j \;=\; q_j - T(r_j), \tag{A.25}$$

where q_j is the empirical $(1-\bar\alpha_j)$ quantile formed with

$$\bar\alpha_j \;=\; \alpha \;-\; \sqrt{\frac{\ln(1/\delta_j)}{2\,n_j}} \in (0, \alpha).$$

Step 2: Branch-wise identities induced by the explicit update. Write $q_j = T(r_j) + \eta_j$ by (A.25). The explicit update from Theorem 3 gives us the following expressions:

$$\text{if } q_j \le r_j \text{ (shrinkage):} \quad (1+\kappa)\,\big(r_{j+1} - r^\star\big) = \big(T(r_j) - T(r^\star)\big) + \kappa\,(r_j - r^\star) + \eta_j, \tag{A.26}$$

$$\text{if } q_j > r_j \text{ (expansion):} \quad (1-\kappa)\,\big(r_{j+1} - r^\star\big) = \big(T(r_j) - T(r^\star)\big) - \kappa\,(r_j - r^\star) + \eta_j. \tag{A.27}$$

Derivation. In the shrinkage case, in order to obtain (A.26), we compute $r_{j+1}-r^\star$ while substituting $r_{j+1}=\frac{q_j+\kappa r_j}{1+\kappa}$ and $q_j=T(r_j)+\eta_j$ so that

$$(1+\kappa)(r_{j+1}-r^\star)=q_j+\kappa r_j-(1+\kappa)r^\star=T(r_j)-T(r^\star)+\kappa(r_j-r^\star)+\eta_j.$$

The expansion case is identical with $r_{j+1}=\frac{q_j-\kappa r_j}{1-\kappa}$.

Step 3: One-step bounds and contraction threshold. Taking absolute values in (A.26)–(A.27) and using the Lipschitz property

$$\big|T(r_j)-T(r^\star)\big| \;\le\; \kappa\,|r_j-r^\star| \;=\; \kappa\, e_j,$$

we obtain

$$|r_{j+1}-r^\star|\le\frac{1}{1+\kappa}\Big(\,|T(r_j)-T(r^\star)|+\kappa e_j+|\eta_j|\,\Big) \;\le\; \frac{2\kappa}{1+\kappa}\,e_j+\frac{1}{1+\kappa}\,|\eta_j|,$$
$$|r_{j+1}-r^\star|\le\frac{1}{1-\kappa}\Big(\,|T(r_j)-T(r^\star)|+\kappa e_j+|\eta_j|\,\Big) \;\le\; \frac{2\kappa}{1-\kappa}\,e_j+\frac{1}{1-\kappa}\,|\eta_j|.$$

The coefficients in the expansion case dominate those coefficients in the shrinkage case so for that for all episodes j we obtain the one-step recursion

$$e_{j+1} \;\le\; \gamma_\kappa\, e_j \;+\; B_\kappa\,|\eta_j|, \qquad \gamma_\kappa \;=\; \frac{2\kappa}{1-\kappa}, \quad B_\kappa \;=\; \frac{1}{1-\kappa}. \tag{A.28}$$

We specifically note that

$$\gamma_\kappa<1 \iff \frac{2\kappa}{1-\kappa}<1 \iff 3\kappa<1 \iff \kappa<\tfrac{1}{3}. \tag{A.29}$$

Step 4: Explicit finite-horizon error bound. We now unroll (A.28) over multiple episodes to obtain P1 of Theorem 3. For convenience, we recall this bound as

$$e_{j+1} \;\le\; \gamma_\kappa^{j+1}\,e_0 \;+\; B_\kappa\sum_{m=0}^{j}\gamma_\kappa^{j-m}\,|\eta_m|. \tag{A.30}$$

Base case $j=0$. From (A.28), $e_1\le\gamma_\kappa e_0+B_\kappa|\eta_0|$, which matches (A.30) with $j=0$.
Induction step. Assume (A.30) holds for some $j\ge 0$. Then

$$e_{j+2} \;\le\; \gamma_\kappa e_{j+1}+B_\kappa|\eta_{j+1}| \;\le\; \gamma_\kappa\left(\gamma_\kappa^{j+1}e_0+B_\kappa\sum_{m=0}^{j}\gamma_\kappa^{j-m}|\eta_m|\right)+B_\kappa|\eta_{j+1}|,$$

which simplifies to

$$e_{j+2} \;\le\; \gamma_\kappa^{j+2}e_0 \;+\; B_\kappa\sum_{m=0}^{j+1}\gamma_\kappa^{j+1-m}|\eta_m|,$$

i.e., (A.30) with $j\leftarrow j+1$. This completes the induction.

Step 5: Closed-form bounds. If $|\eta_m|\le C$ for all $m\in\{0,1,\dots,j\}$, then applying (A.30) and summing the geometric series gives

$$e_{j+1}\le\gamma_\kappa^{j+1}e_0+B_\kappa C\sum_{m=0}^{j}\gamma_\kappa^{j-m} \;=\; \gamma_\kappa^{j+1}e_0+B_\kappa C\sum_{\ell=0}^{j}\gamma_\kappa^{\ell}$$
$$=\gamma_\kappa^{j+1}e_0 \;+\; B_\kappa C\,\frac{1-\gamma_\kappa^{j+1}}{1-\gamma_\kappa}. \tag{A.31}$$

If moreover $\kappa < \frac{1}{3}$ (so that $\gamma_\kappa < 1$), then letting $j \to \infty$ in (A.31) yields the steady-state bound

$$\limsup_{j\to\infty} e_j \;\le\; \frac{B_\kappa}{1-\gamma_\kappa}\, C \;=\; \frac{1}{1-3\kappa}\, C.$$

This proves P2 of Theorem 3.

Step 6: High-probability control of η_j (level shift + empirical error). First, decompose

$$\eta_j \;=\; \underbrace{Q_{1-\bar\alpha_j}\big(\pi^\star(r_j)\big) - Q_{1-\alpha}\big(\pi^\star(r_j)\big)}_{\Delta_j^{\mathrm{lvl}}} \;+\; \underbrace{q_j - Q_{1-\bar\alpha_j}\big(\pi^\star(r_j)\big)}_{\varepsilon_j^{\mathrm{est}}}$$

into a level shift and empirical error component.

By assumption, the CDF $F_{r_j}(s)$ is differentiable and has density no less than $f_\star > 0$ in a sufficiently large neighborhood of its $(1-\alpha)$ quantile $Q_{1-\alpha}\big(\pi^\star(r_j)\big)$. Then, by Lemma 4 which we separately present in Appendix A.8, the inverse-CDF is $1/f_\star$-Lipschitz in the probability level, and therefore

$$\big|\Delta_j^{\mathrm{lvl}}\big| \;\le\; \frac{|\alpha-\bar\alpha_j|}{f_\star}. \tag{A.32}$$

Next, Lemma 5 which we separately present in Appendix A.8 and the inverse-CDF Lipschitz property imply

$$\mathbb{P}_{n_j}\left\{\big|\varepsilon_j^{\mathrm{est}}\big| \;\le\; \frac{1}{f_\star}\sqrt{\frac{\ln(2/\delta_j)}{2n_j}}\right\} \;\ge\; 1-\delta_j. \tag{A.33}$$

where $\delta_j \in (0,1)$ is any prescribed failure probability.

Combining (A.32)–(A.33) gives the per-episode high-probability bound

$$\mathbb{P}_{n_j}\left\{|\eta_j| \;\le\; \frac{|\alpha-\bar\alpha_j|}{f_\star} \;+\; \frac{1}{f_\star}\sqrt{\frac{\ln(2/\delta_j)}{2n_j}}\right\} \;\ge\; 1-\delta_j, \tag{A.34}$$

which proves the per-episode bound in P3 of Theorem 3.

Finally, applying equation (A.30) and a union bounding argument over the first $j+1$ episodes, we obtain the explicit finite-horizon control

$$\mathbb{P}_{\sum_{m=0}^{j} n_m}\left\{e_{j+1} \;\le\; \gamma_\kappa^{j+1} e_0 \;+\; \frac{B_\kappa}{f_\star}\sum_{m=0}^{j}\gamma_\kappa^{j-m}\,|\alpha-\bar\alpha_m| \;+\; \frac{B_\kappa}{f_\star}\sum_{m=0}^{j}\gamma_\kappa^{j-m}\sqrt{\frac{\ln(2/\delta_m)}{2n_m}}\right\} \;\ge\; 1-\sum_{m=0}^{j}\delta_m.$$

This shows the joint bound in P3 of Theorem 3.

Step 7: Asymptotic conclusions. By (A.34), $|\eta_j| \le A_j$ holds with $\mathbb{P}_{n_j}$-probability at least $1-\delta_j$. If $n_j \to \infty$, $\bar\alpha_j \to \alpha$, and $\delta_j \to 0$, then $A_j \to 0$, so $\mathbb{P}_{n_j}\{|\eta_j| \to 0\} \ge 1-\delta_j$. Since calibration samples at distinct episodes are independent, a union bound under the product measure $\mathbb{P}_{\sum_{m=0}^{\infty} n_m}$ gives $\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\{|\eta_j| \to 0\} \ge 1-\sum_{m=0}^{\infty}\delta_m$. If additionally $\kappa < \frac{1}{3}$, then $\gamma_\kappa < 1$ by (A.29) and (A.30) gives $e_j \to 0$ on the same event, so $\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\{e_j \to 0\} \ge 1-\sum_{m=0}^{\infty}\delta_m$.

If instead $\bar\alpha_j \equiv \bar\alpha < \alpha$, $n_j \to \infty$, and $\delta_j \to 0$, then $A_j \to (\alpha-\bar\alpha)/f_\star$, so $\mathbb{P}_{n_j}\{|\eta_j| \to (\alpha-\bar\alpha)/f_\star\} \ge 1-\delta_j$. Using (A.31) with $C = (\alpha-\bar\alpha)/f_\star$ and the same union bound yields

$$\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\Big\{\limsup_{j\to\infty} e_j \le \tfrac{1}{1-3\kappa}\cdot\tfrac{\alpha-\bar\alpha}{f_\star}\Big\} \;\ge\; 1-\sum_{m=0}^{\infty}\delta_m.$$

This completes the proof of Theorem 3.

A.8. Quantile perturbation lemmas (expanded and annotated)

Setup and notation. Recall that $F_r : \mathbb{R} \to [0,1]$ denotes the cumulative distribution function (CDF) of the nonconformity score $s(\hat{y}_{0:T}, Y_{0:T})$ for the trajectory $y_{0:T} \sim \mathcal{D}(\pi^\star(r))$. We have also defined the right-continuous quantile map as

$$Q_p(\pi^\star(r)) \;:=\; \inf\{\, s \in \mathbb{R} :\; F_r(s) \geq p \,\}, \qquad \alpha \in (0,1).$$

Assume that the CDF $F_r(s)$ is differentiable and has density $f_r(s)$ no less than $f_\star > 0$ for all s in a sufficiently large neighborhood of its p quantile $Q_p\big(\pi^\star(r)\big)$, i.e.,

$$f_r(s) \;\geq\; f_\star \;>\; 0 \qquad \text{for all } s \text{ in a sufficiently large neighborhood of } Q_p\big(\pi^\star(r)\big). \tag{A.35}$$

Condition (A.35) guarantees that the inverse-CDF is locally Lipschitz; geometrically, the CDF has a slope bounded away from 0 near the quantile, so small changes in probability level produce controlled changes in the quantile value.

Lemma 4 (Level-to-quantile Lipschitzness) *For any two probability levels $p, p' \in (0,1)$ and for any $Q_{p'}(\pi^\star(r))$ in the neighborhood of $Q_p(\pi^\star(r))$ from (A.35), we have that*

$$\big|\, Q_p(\pi^\star(r)) - Q_{p'}(\pi^\star(r)) \,\big| \;\leq\; \frac{|p - p'|}{f_\star}. \tag{A.36}$$

Proof ***1) Fix the two quantiles.*** *Let*

$$q \;=\; Q_p(\pi^\star(r)), \qquad q' \;=\; Q_{p'}(\pi^\star(r)).$$

By definition of Q_p, we have $F_r(q) \geq p$ and, by right continuity and strict monotonicity in our neighborhood, we can take $F_r(q) = p$ and $F_r(q') = p'$.

2) Apply the Mean Value Theorem (MVT). *Since F_r is differentiable on the open interval between q and q', there exists a point ξ between q and q' such that*

$$F_r(q') - F_r(q) \;=\; f_r(\xi)\,(q' - q).$$

3) Use the density lower bound. *Because $f_r(\xi) \geq f_\star$ by (A.35), we get*

$$|p' - p| \;=\; |F_r(q') - F_r(q)| \;=\; f_r(\xi)\,|q' - q| \;\geq\; f_\star\,|q' - q|.$$

4) Rearrange. *Hence*

$$|q' - q| \;\leq\; \frac{|p' - p|}{f_\star},$$

which is (A.36). ∎

Why we need Lemma 4. In the convergence analysis (Theorem 3), we compare the population quantiles at two *levels*, $1 - \bar{\alpha}_j$ (used for calibration) and $1 - \alpha$ (the target). Lemma 4 gives the clean bound

$$\big|\, Q_{1-\bar{\alpha}_j} - Q_{1-\alpha} \,\big| \;\leq\; \frac{|\alpha - \bar{\alpha}_j|}{f_\star},$$

which is the level-shift term in the perturbation η_j.

Lemma 5 (Empirical quantile error via DKW Inequality) *Let $F_{r,n}$ be the empirical CDF from n i.i.d. samples drawn from F_r, and let $q_{n,p}$ be the empirical p-quantile $q_{n,p} = \inf\{t :\; F_{r,n}(t) \geq p\}$. Then, for any outer tail level $\delta \in (0,1)$, we have*

$$\mathbb{P}_n\left\{ \sup_{t \in \mathbb{R}} \big|F_{r,n}(t) - F_r(t)\big| \leq \varepsilon_n(\delta) \right\} \;\geq\; 1 - \delta, \qquad \varepsilon_n(\delta) := \sqrt{\frac{\ln(2/\delta)}{2n}}. \tag{A.37}$$

Under the assumption in equation (A.35), the empirical quantile satisfies

$$\mathbb{P}_n\left\{ \left| q_{n,p} - Q_p(\pi^\star(r)) \right| \;\le\; \frac{\varepsilon_n(\delta)}{f_\star} \right\} \;\ge\; 1-\delta. \tag{A.38}$$

Proof ***1) DKW event (uniform CDF control).*** *By the DKW inequality, the statement in equation (A.37) immediately follows.*

2) Pin the target quantile and a local window. *Let*

$$q^\star \;=\; Q_p(\pi^\star(r)),$$

and pick a small $\Delta > 0$ that keeps $[q^\star - \Delta,\, q^\star + \Delta]$ inside the neighborhood where $f_r \ge f_\star$.

3) One-sided controls for the* true *CDF using the density lower bound. *By the Mean Value Theorem applied to F_r on $[q^\star,\, q^\star + \Delta]$ and $[q^\star - \Delta,\, q^\star]$ there exist points ξ_+, ξ_- in those intervals with*

$$F_r(q^\star + \Delta) - F_r(q^\star) \;=\; f_r(\xi_+)\,\Delta \;\ge\; f_\star \Delta, \qquad F_r(q^\star) - F_r(q^\star - \Delta) \;=\; f_r(\xi_-)\,\Delta \;\ge\; f_\star \Delta.$$

Since $F_r(q^\star) = p$, we get

$$F_r(q^\star + \Delta) \;\ge\; p + f_\star \Delta, \qquad F_r(q^\star - \Delta) \;\le\; p - f_\star \Delta. \tag{A.39}$$

4) Transfer these inequalities to the* empirical *CDF on the DKW event. *Using (A.37), we have*

$$\mathbb{P}_n\{F_{r,n}(q^\star + \Delta) \;\ge\; F_r(q^\star + \Delta) - \varepsilon \;\ge\; p + f_\star \Delta - \varepsilon \tag{A.40}$$

$$\text{and } F_{r,n}(q^\star - \Delta) \;\le\; F_r(q^\star - \Delta) + \varepsilon \;\le\; p - f_\star \Delta + \varepsilon\} \;\ge\; 1-\delta. \tag{A.41}$$

5) Choose Δ to make the inequalities straddle level p. *Set $\Delta \;=\; \frac{\varepsilon}{f_\star}$ so that*

$$\mathbb{P}_n\{F_{r,n}(q^\star + \Delta) \;\ge\; p \text{ and } F_{r,n}(q^\star - \Delta) \;\le\; p\} \;\ge\; 1-\delta.$$

6) Use the empirical quantile definition to trap $q_{n,p}$. *By definition, $q_{n,p} = \inf\{t :\; F_{r,n}(t) \ge p\}$. Since $F_{r,n}(q^\star - \Delta) \le p$ and $F_{r,n}(q^\star + \Delta) \ge p$, the monotonicity of $F_{r,n}$ implies*

$$\mathbb{P}_n\{q_{n,p} \;\in\; [\,q^\star - \Delta,\; q^\star + \Delta\,]\} \;\ge\; 1-\delta,$$

where we used the union bound. Therefore,

$$\mathbb{P}_n\left\{ |q_{n,p} - q^\star| \;\le\; \Delta \;=\; \frac{\varepsilon}{f_\star} \;=\; \frac{\varepsilon_n(\delta)}{f_\star} \right\} \;\ge\; 1-\delta.$$

This is exactly (A.38). ■

Why we need Lemma 5. In the convergence proof, the empirical quantile q_j is compared to the population quantile at the *same level* $1-\bar{\alpha}_j$. Lemma 5 converts the uniform CDF error (obtained using the DKW inequality) into an error on the quantile value, with the sharp factor $1/f_\star$. Concretely, we thereby get

$$\mathbb{P}_{n_j}\left\{ \left| q_j - Q_{1-\bar{\alpha}_j}(\pi^\star(r_j)) \right| \;\le\; \frac{1}{f_\star}\sqrt{\frac{\ln(2/\delta_j)}{2\,n_j}} \right\} \;\ge\; 1-\delta_j.$$

A.9. Proof of Theorem 4

The first result $|J_{j+1} - J^\star| = |V(r_{j+1}) - V(r^\star)| \le L_V\, e_{j+1}$ follows directly from the Lipschitz constant L_V of $V(r)$ and $e_{j+1} = |r_{j+1} - r^\star|$ from Theorem 3.

We next show that V is nondecreasing, which is used in all three parts below. For $r \in \mathcal{R}$, define the feasible policy set

$$\mathcal{F}(r) := \{\pi : \ H(x_{0:T}(\pi), \zeta) \le 0 \ \ \forall\, \zeta \in \mathcal{C}_r(\hat{y}_{0:T})\}. \tag{A.42}$$

For $r, r' \in \mathcal{R}$ with $r' \le r$, we have that $\mathcal{C}_{r'}(\hat{y}_{0:T}) \subseteq \mathcal{C}_r(\hat{y}_{0:T})$, so $\mathcal{F}(r) \subseteq \mathcal{F}(r')$. Taking infima of the same objective function over nested sets yields

$$V(r') = \inf_{\pi \in \mathcal{F}(r')} J\big(x_{0:T}(\pi), \pi\big) \ \le \ \inf_{\pi \in \mathcal{F}(r)} J\big(x_{0:T}(\pi), \pi\big) = V(r). \tag{A.43}$$

Part (P1): Improvement over the initial policy. Fix any $r_j \in [r^\star, r_0)$ with $r_0 > r^\star$. Since $r_j < r_0$, the monotonicity (A.43) gives $J_j = V(r_j) \le V(r_0) = J_0$, with $J_j = V(r_j) < V(r_0) = J_0$ when the function $V(r)$ is strictly increasing on $[r^\star, r_0]$.

Next, fix any $r_j \le r_0$. Assume additionally that there exists $m_V > 0$ such that $V(r) - V(r') \ge m_V(r - r')$ for all $r, r' \in [r^\star, r_0]$. If $r_j \ge r^\star$, then $r_j \in [r^\star, r_0]$ so that we directly obtain

$$J_0 - J_j \ = \ V(r_0) - V(r_j) \ \ge \ m_V(r_0 - r_j) \ = \ m_V(r_0 - r^\star - e_j)$$

where we recall that $e_j = r_j - r^\star$. If $r_j < r^\star$, then $e_j = r^\star - r_j \ge 0$ and the monotonicity property (A.43) implies $V(r_j) \le V(r^\star)$, so that we can derive

$$J_0 - J_j \ = \ V(r_0) - V(r_j) \ \ge \ V(r_0) - V(r^\star) \ \ge \ m_V(r_0 - r^\star) \ \ge \ m_V(r_0 - r^\star - e_j).$$

In both cases $J_0 - J_j \ge m_V(r_0 - r^\star - e_j)$.

Part (P2): One-step improvement. Recall from Theorem 3 that $\eta_j = q_j - T(r_j)$ where $T(r) := Q_{1-\alpha}(\pi^\star(r))$ satisfies $T(r^\star) = r^\star$ and has Lipschitz constant κ. For $r_j \ge r^\star$, this Lipschitz property $T(r)$ gives us that

$$T(r_j) \ \le \ T(r^\star) + \kappa\,(r_j - r^\star) \ = \ r^\star + \kappa\,(r_j - r^\star). \tag{A.44}$$

By the definition of η_j and (A.44), we have $q_j = T(r_j) + \eta_j \le r^\star + \kappa(r_j - r^\star) + |\eta_j|$. Under the condition $|\eta_j| < (1-\kappa)(r_j - r^\star)$ assumed in (P2), this yields

$$q_j \ < \ r^\star + \kappa(r_j - r^\star) + (1-\kappa)(r_j - r^\star) \ = \ r_j.$$

Since $q_j < r_j$, the update rule in (3.5) is on the shrinkage branch, and Theorem 2 (Episode-to-Episode Stability and Shrinkage) gives us that $r_{j+1} < r_j$. The monotonicity property (A.43) then yields $J_{j+1} = V(r_{j+1}) \le V(r_j) = J_j$.

For the last statement in P2, let $\mathcal{E}_j := \{J_{j+1} \le J_j\}$. By Theorem 3(P3), $\mathbb{P}_{n_j}\{|\eta_j| \le A_j\} \ge 1 - \delta_j$. On this event, the sufficient condition $A_j < (1-\kappa)(r_j - r^\star)$ from (P2) implies the deterministic condition above, so $\mathbb{P}_{n_j}(\mathcal{E}_j) \ge 1 - \delta_j$. Since calibration samples across episodes are independent, a union bound argument yields

$$\mathbb{P}_{\sum_{j=J}^K n_j}\Big\{\bigcap\nolimits_{j=J}^K \mathcal{E}_j\Big\} \ \ge \ 1 - \textstyle\sum_{j=J}^K \delta_j.$$

Part (P3): Asymptotics. By Theorem 3(P4), $e_j \to 0$ with probability at least $1 - \sum_{m=0}^\infty \delta_m$. On this event, $|J_j - J^\star| \le L_V\, e_j \to 0$ by the Lipschitz constant L_V of V, hence

$$\mathbb{P}_{\sum_{m=0}^\infty n_m}\{J_j \to J^\star\} \ \ge \ 1 - \sum_{m=0}^{\infty} \delta_m.$$

For the second claim, fix $r_0 > r^\star$. On the same event, $|r_j - r^\star| \to 0$, so $r_j \to r^\star$. Since $r^\star < r_0$, there exists j_0 such that $r_j < r_0$ for all $j \ge j_0$. The monotonicity (A.43) then gives

$J_j = V(r_j) \le V(r_0) = J_0$ for all $j \ge j_0$, hence

$$\mathbb{P}_{\sum_{m=0}^{\infty} n_m}\{\exists\, j_0 \text{ s.t. } J_j \le J_0\,, \forall\, j \ge j_0\} \;\ge\; 1 - \sum_{m=0}^{\infty} \delta_m,$$

which completes the proof.

Appendix B. The Iterative Policy Update Algorithm

We summarize the explicit and implicit solver in Algorithm 1.

Algorithm 1 Iterative Safe Policy Improvement (Explicit & Implicit Forms)

Input: confidence levels $1-\alpha$, $1-\delta$; initial safe policy π_0.
Input: closed-loop gain $\kappa = \beta_T L_U$; fixed predictor $\hat{y}_{0:T}$.
Input: solver choice: `SOLVER_TYPE` $\in$ {'EXPLICIT', 'IMPLICIT'}.
Initialize: choose $r_0 \in \mathcal{R}$; set $\pi_0 \leftarrow \pi^\star(r_0)$.
for $j = 0, 1, 2, \ldots$ **do**
  Execute π_j; collect rollouts $\{y^{(i)}_{j,0:T}\}_{i=1}^{n_j}$.
  Calibration level with δ inside: $\bar{\alpha}_j \leftarrow \alpha - \sqrt{\ln(1/\delta)/(2n_j)}$, so that $\bar{\alpha}_j \in (0, \alpha)$.
  Empirical quantile (scores at episode j): $q_j \leftarrow q_{1-\bar{\alpha}_j}\big(\{s(\hat{y}_{0:T}, y^{(i)}_{j,0:T})\}_{i=1}^{n_j}\big)$.
  if `SOLVER_TYPE`='EXPLICIT' **then**
    {use analytical solution from Lemma 3}
    if $q_j \le r_j$ **then**
      $r_{j+1} \leftarrow (q_j + \kappa r_j)/(1+\kappa)$
    else
      $r_{j+1} \leftarrow (q_j - \kappa r_j)/(1-\kappa)$
    end if
  else if `SOLVER_TYPE`='IMPLICIT' **then**
    {use iterative solver for equation (3.3), see Appendix C}
    $r_{j+1} \leftarrow \mathrm{FindRoot}\big(r \mapsto r - (q_j + \beta_T \|\pi^\star(r) - \pi_j\|_\infty)\big)$
  end if
  Certify and deploy: solve $\mathbf{P}[j{+}1; r_{j+1}]$ to get policy $\pi_{j+1} \leftarrow \pi^\star(r_{j+1})$.
  Monitor: record $|r_{j+1} - r_j|$ and $\|\pi_{j+1} - \pi_j\|_\infty$.
  if changes are below threshold **then break.**
end for

Appendix C. The Implicit Solver Algorithm

We here detail the implicit solver of Approach 1, as presented in Section 3.2. At episode j, the quantities q_j, π_j, and β_T are fixed, so the implicit safety requirement in equation (3.3) reduces to a scalar one-dimensional program with residual

$$g_j(r) \;:=\; r \;-\; \Big(q_j + \beta_T \big\|\pi^\star(r) - \pi_j\big\|_\infty\Big),$$

and we seek the smallest $r \in [q_j, r_{\max}]$ such that $g_j(r) \ge 0$. Each evaluation of $g_j(r)$ requires one call to the planner $\pi^\star(\cdot)$, i.e., one solve of $\mathbf{P}[j{+}1;\, r]$ in equation (2.5) (warm-started at π_j in our

implementation). In MATLAB, we handle this one-dimensional constrained problem using either (i) a bracketed line search on r to find a feasible point by increasing r until $g_j(r) \geq 0$ or $r = r_{\max}$, followed by bisection to a tolerance ε when the feasible set is observed to be an interval, or (ii) a generic constrained optimizer such as `fmincon` with objective $\min_r r$ and constraint $g_j(r) \geq 0$.

Appendix D. Practical Instantiations of Sensitivity Constants

Our analysis treats β_T from Lemma 2 and L_U from Assumption 3 as fixed global upper bounds so that the closed-loop gain $\kappa = \beta_T L_U$ is constant. The closed-form expressions for β_T in Appendix A.1 results from recursive unfolding of the coupled dynamics (2.1); our argument shows that β_T grows like $(1 + cL)^T$ in the horizon T, where c is a constant L is the largest Lipschitz constant in Assumption 1. For even moderate T, this bound can be overly conservative.

A practical alternative is to replace β_T and L_U by data-driven estimates: the maximum pairwise slopes of the maps $r \mapsto \pi^\star(r)$ and $r \mapsto y_{0:T}(\pi^\star(r))$ evaluated on a finite grid in $\mathcal{R}$. These are the kinky-inference-style Lipschitz estimators of Calliess (2017); Huang et al. (2023a). Under mild density assumptions, Huang et al. (2023a) shows these are certified upper bounds with confidence $1 - \delta_\beta$. Since the robustification term M_{j+1} then depends on estimation data, this substitution degrades the outer confidence in Theorem 1 from $1-\delta$ to $1-\delta-\delta_\beta$ by a union bounding argument. The same applies to the implicit solver and to Lemma 3.

Appendix E. Multi-Quadcopter Navigation

This appendix reports the results for the multi-quadcopter case study. All notation, targets ($1-\alpha = 0.9$, $\delta_j = 0.05$), and baselines for comparison are as defined in Section 4. Since the implicit solver requires multiple evaluations of $\pi^\star(r)$ per episode (see Appendix C), which is computationally expensive for the CBF-QP planner in (E.2), we employ only the explicit solver with update rule (3.5) and report four rules: *Robust CP – explicit (ours)*, *Naive CP*, *One-time CP*, and *Split CP*.

System and deployed policy. We consider one ego agent and $N_{\text{env}} = 5$ environment quadrotors, each modeled as a double integrator system operating in $\mathbb{R}^3$ (an instantiation of (2.1)), using the QuadSwarm simulator Huang et al. (2023b). The environment agents execute learned navigation policies trained with collision penalties and do not deploy a test-time safety filter. The ego agent uses nominal policy π^{RL} to track a goal position at every timestep without knowledge of the environment agents. We approximate the cost minimization in $\mathbf{P}[j; r]$ by minimizing the distance between the robust action and the nominal action at every timestep, subject to the safety constraints. Concretely, we define for each environment agent $i \in [N_{\text{env}}]$ and time t the barrier function

$$h_i(x_t, t) := \|x_t - \hat{y}_t^{(i)}\|_2 - d_{\text{coll}} - r_j, \tag{E.1}$$

which encodes separation from $\mathcal{C}_{r_j}(\hat{y}_{0:T})$. The ego agent then deploys the CBF-QP safety filter Ames et al. (2019)

$$\pi_j(x_t) := \arg\min_{u \in \mathcal{U}} \ \|u - \pi^{\text{RL}}(x_t)\|_2^2 \quad \text{s.t.} \ \ h_i\big(f_X(x_t, u),\, t{+}1\big) \geq (1{-}\gamma)\, h_i(x_t, t), \ \forall\, i \in [N_{\text{env}}], \tag{E.2}$$

where $\gamma \in (0, 1]$ is the CBF decay rate and $\mathcal{U}$ is the admissible control set. Since f_X is a double integrator (affine in u), (E.2) is a QP at each timestep. Hence π_j depends on j only through r_j, playing the role of $\pi^\star(r_j)$ in Section 3.2. A collision is declared when $\|x_{j,t} - y_{j,t}^{(i)}\|_2 < d_{\text{coll}} := 2.5\, \ell_{\text{arm}}$, with $\ell_{\text{arm}} = 0.046$ being the quadrotor arm length.

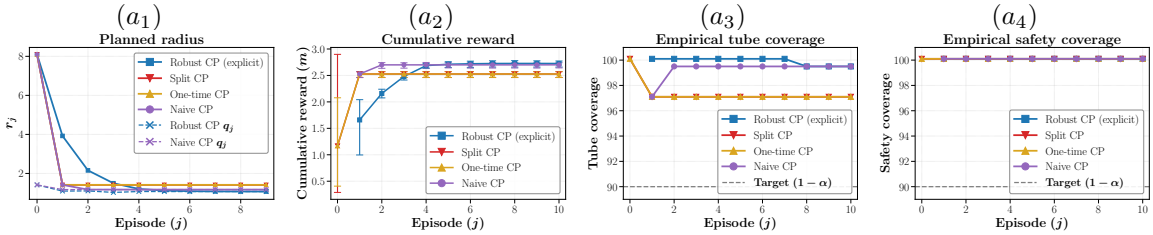

Figure 5: Multi-quadcopter quantitative results. Left to right: deployed radius r_j (with the current aggregate quantile $\hat{q}_j$ overlaid for the robust rule), cumulative reward, empirical tube coverage (target $1-\alpha = 0.9$), and empirical safety coverage. *Robust CP* contracts r_j and attains the highest cumulative reward while retaining both coverages at their target.

Calibration and update rule. We use (2.3) with the Euclidean norm in $\mathbb{R}^3$ and $n_j = 200$. We collect empirical quantiles $q_j^{(i)}$ defined in Section 3.1 for every agent $i \in [N_{\text{env}}]$, and aggregate them as $\hat{q}_j := \max_i q_j^{(i)}$ so that safety holds simultaneously for all agents, with $\hat{q}_j$ replacing q_j in (3.3) and (3.5). The admissible radius set is $\mathcal{R} = [2\,\ell_{\text{arm}}, 8\,\text{m}]$ and the sensitivity gain is $\kappa = 0.6$. All four update rules share $r_0 = r_{\max}$ and the same predictor $\hat{y}_{0:T}$, initialized from a non-interactive rollout. An additional 100 rollouts per episode estimate empirical tube coverage, empirical safety coverage, and the cumulative progress-to-goal reward.

Observations. Figure 5 reports the four metrics. Under *Robust CP*, r_j contracts from $r_0 = 8\,\text{m}$ toward $\approx 1\,\text{m}$ (e.g. $r_1 \approx 3.82\,\text{m}$, $r_5 \approx 1.02\,\text{m}$); the first step is a strict convex combination of r_0 and $\hat{q}_0$ via the shrinkage branch of (3.5). Cumulative reward increases from ≈ 1.6 to ≈ 2.68, empirical tube coverage remains above $1-\alpha = 0.9$, and empirical safety coverage stays at one, in agreement with Theorem 1. *Naive CP* attains lower cumulative reward than the robust rule, as the robust rule safely attains an equilibrium with a smaller final radius; *One-time CP* and *Split CP* hold r_j fixed at their first certificate $q_{\text{base}} \approx 1.31\,\text{m}$ and likewise attain lower cumulative reward at convergence. Only *Robust CP* converts the conservative initialization r_0 into a smaller stabilized operating radius without loss of either coverage, matching the shrinkage branch of (3.5) under Theorem 4. Figures 6 and 7 show the corresponding trajectories and tube evolution.

We note that all four rules maintain empirical safety coverage at 100% throughout (Figure $5(a_4)$). This is attributable to the experimental setup: the environment agents were trained jointly in a homogeneous multi-agent environment in which all agents shared the same collision-avoidant objective and policy. Their learned policies are therefore conditioned on collision-averse behavior from all agents, including the ego agent, so that the policy-induced distribution shift $\mathcal{D}(\pi_j) \to \mathcal{D}(\pi_{j+1})$ per Lemma 2 remains small even without the robustification term M_{j+1} in (3.2). This explains why *Naive CP* does not exhibit the tube coverage collapse observed in the car–pedestrian study (Section 4). Nevertheless, only *Robust CP* systematically reduces r_j while maintaining valid coverage guarantees per (3.2) and Theorem 1.

The qualitative panels in Figures 6 and 7 support this interpretation. As r_j contracts, the ego trajectory straightens and the tubes around representative environment agents shrink accordingly, yet continue to contain the realized trajectories. The baseline trajectories in Figure 7 remain visually similar across episodes, consistent with the small distribution shift in this setting, while operating at a fixed and larger safety margin than the converged *Robust CP* radius.

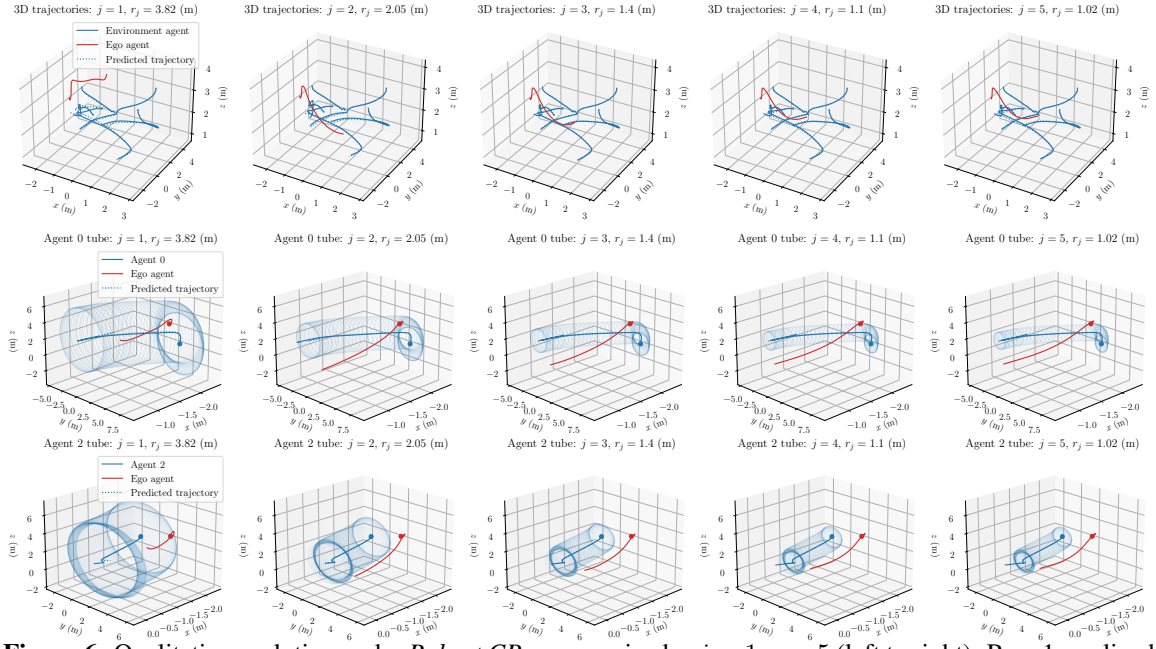

Figure 6: Qualitative evolution under *Robust CP* across episodes $j = 1, \ldots, 5$ (left to right). Row 1: realized trajectories (solid) and nominal predicted trajectories (dotted). Rows 2–3: conformal tubes $\mathcal{C}_{r_j}(\hat{y}_{0:T})$ for two representative environment agents. As r_j contracts, the ego agent's path straightens while the tubes continue to contain the realized environment trajectories.

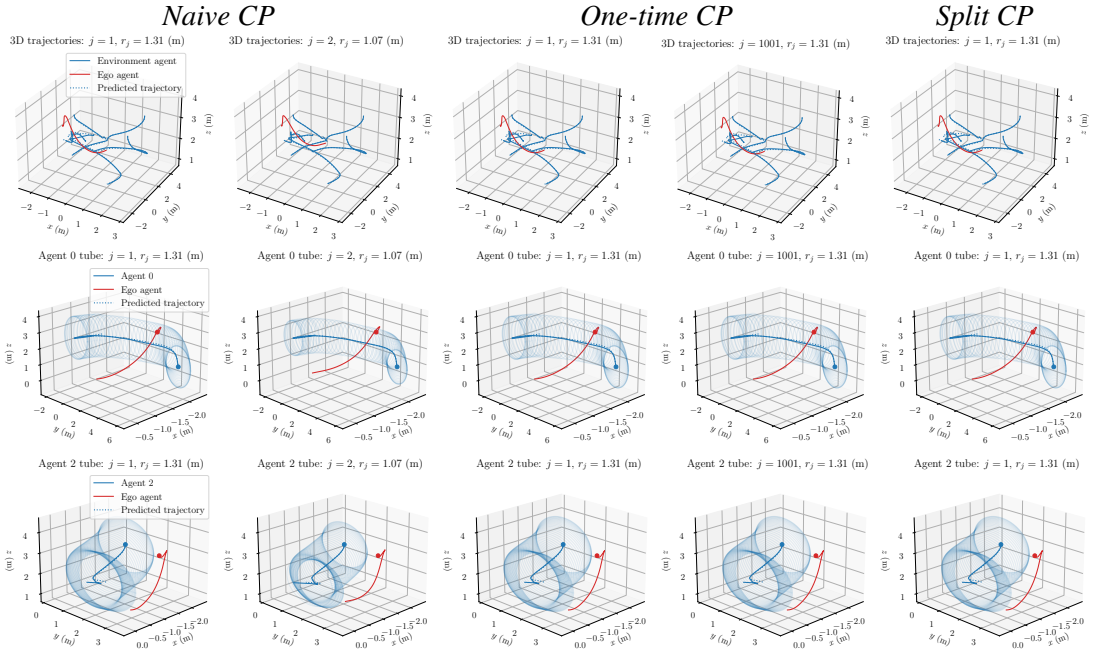

Figure 7: Qualitative baseline comparison. *Naive CP* at $j \in \{1, 2\}$; *One-time CP* at $j \in \{1, 1001\}$; *Split CP* at $j = 1$. Row 1: realized trajectories; rows 2–3: conformal tubes for two representative environment agents.