

Quantifying Action Uncertainty with Inaccurate Stochastic Dynamics through Conformalized Lie groups

Luís Marques, Maani Ghaffari, and Dmitry Berenson

Abstract—We propose a symmetry-aware conformal prediction-based algorithm that constructs, for a given action, a set guaranteed to contain the resulting system configuration at a user-defined probability. Our assurance holds under both aleatoric and epistemic uncertainty, non-asymptotically, and does not require strong assumptions about the true system dynamics, the uncertainty sources, or the quality of the approximate dynamics model. Typically, uncertainty quantification is tackled by making strong assumptions about the error distribution or magnitude, or by relying on uncalibrated uncertainty estimates — i.e., with no link to frequentist probabilities — which are insufficient for safe control. Recently, conformal prediction has emerged as a statistical framework capable of providing distribution-free probabilistic guarantees on test-time prediction accuracy. While current conformal methods treat robot configurations as Euclidean points, many systems have non-Euclidean configurations, e.g., some mobile robots have $SE(2)$. In this work, we rigorously analyze configuration errors using Lie groups, extending previous Euclidean space theoretical guarantees to $SE(2)$. Our experiments on a simulated JetBot, and on a real MBot, suggest that by considering the configuration space’s structure, our symmetry-informed nonconformity score leads to more volume-efficient prediction regions which represent the underlying uncertainty better than existing approaches.

I. INTRODUCTION

Robotic systems operate under both *aleatoric uncertainty* (e.g., external disturbances) and *epistemic uncertainty* (e.g., mismatch due to assumptions like “no-slip”), while using imperfect dynamics models for planning. Control methods often make strong distributional assumptions on the dynamics error (e.g., Gaussianity, known max bounds, determinism) [1], [2], [3], [4] which may not hold in practice. Conversely, data-driven approaches enable (almost) assumptionless estimation — be it via ensembles [5], posterior variances [6], reconstruction losses [7], etc. — typically at the cost of looser guarantees due to reliance on *uncalibrated uncertainty estimates*, i.e., that cannot be interpreted as likelihoods. Despite impressive results, providing rigorous test-time distribution-free *calibrated uncertainty predictions* (provably containing the true unobserved labels at the specified likelihood) when given inaccurate dynamics models and subject to uncharacterized external disturbances remains an active research problem.

Conformal Prediction (CP) is a data-driven uncertainty-quantification framework enabling the construction of prediction regions that contain the true unknown system state at a given user-set probability. However, current CP methods treat states as Euclidean vectors and consider point-prediction models [8], [9], [10], resulting in convex prediction regions that can be *volume-inefficient*, requiring large sets to achieve

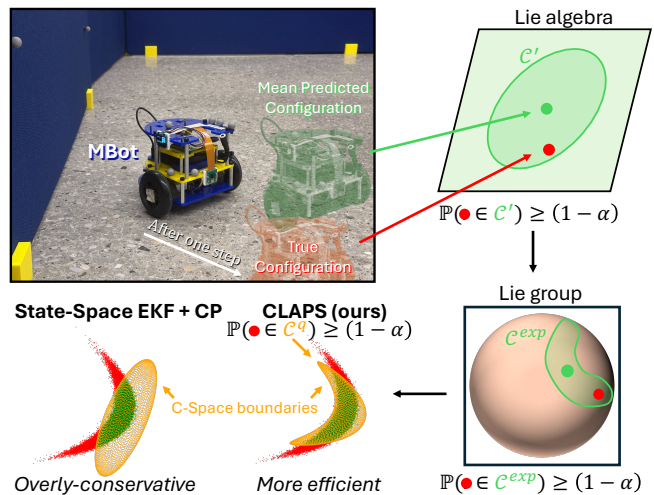


Fig. 1. Our proposed algorithm (CLAPS) constructs prediction regions C^q (in C-Space) that are marginally guaranteed to contain the next unknown system configuration at a user-set probability $(1 - \alpha)$. By considering the robot’s symmetry, we can construct more efficient prediction regions.

the target probability. This limits the downstream applicability of CP for safe control¹. We derive tighter prediction regions by accounting for the configuration space’s inherent symmetry in a theoretically grounded way. In state estimation, Lie groups have been used to represent and propagate robotic configuration uncertainty [11], [12], [13], [14], improving convergence guarantees, speed, and better representing the underlying uncertainty that state-space (SS) alternatives. However, the Invariant EKF (InEKF) uncertainty estimates are still *uncalibrated*, being insufficient for provably safe control.

We propose Conformal Lie-group Action Prediction Sets (CLAPS), a CP-based algorithm that uses a dataset of state transitions to calibrate the uncertainty estimates provided by approximate dynamics models. CLAPS can be applied as a *post-hoc calibration layer* on top of existing Lie-algebraic Gaussian uncertainty estimators, turning approximate covariances into *provably calibrated* ones. Our contributions are: 1) introducing an algorithm that, given an approximate dynamics model estimating prediction uncertainty as Gaussian, constructs state- and action-dependent *calibrated prediction sets* in $SE(2)$ that provably (marginally) contain the resulting configuration; 2) simulation and hardware experiments demonstrating an increase in prediction region *volume-efficiency* and *representation quality*.

¹If an action is deemed probabilistically safe when its resulting prediction region is collision-free, then region inefficiency reduces the set of safe actions, potentially also impacting task speed and other metrics

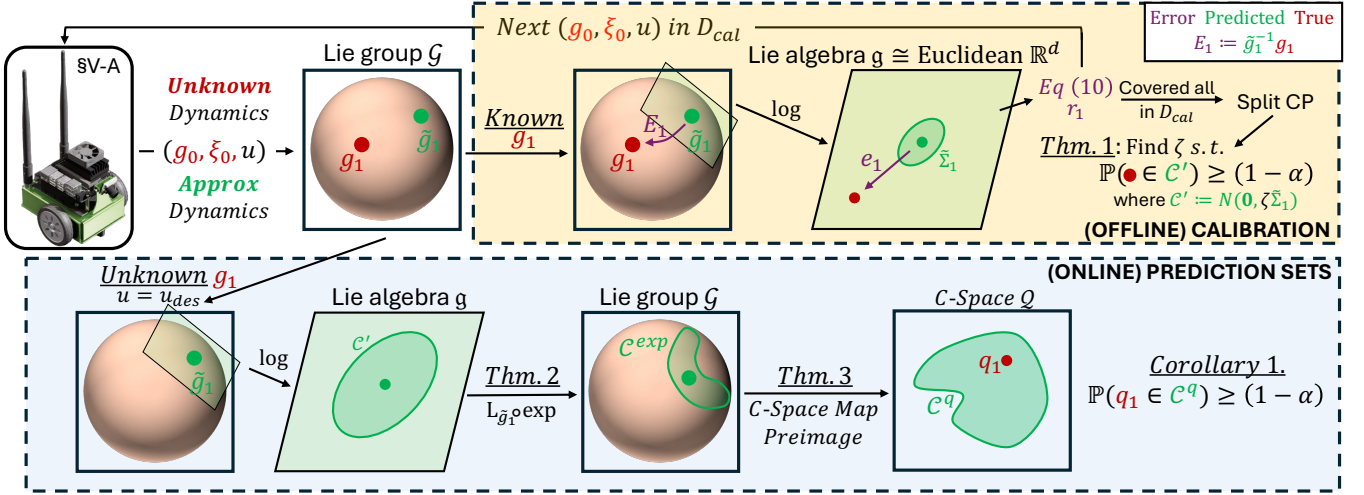


Fig. 2. Conformal Lie-group Action Prediction Sets | Offline: a dataset of state transitions is used jointly with an approximate dynamical model to derive a rigorous symmetry-aware probabilistic error bound on the configuration predictions. Online: our algorithm takes in a desired action u_{des} and computes a calibrated C-Space prediction region \mathcal{C}^q that is marginally guaranteed to contain the true configuration resulting from executing u_{des} .

II. PROBLEM STATEMENT

Let $q \in \mathcal{Q}$ be the C-Space configuration, $\dot{q} \in T_q \mathcal{Q}$ be the generalized velocity, and $s := (q, \dot{q}) \in T\mathcal{Q}$ the state. Let the *kinematics map* $K : \mathcal{Q} \rightarrow \mathcal{G}$ map generalized coordinates q to Lie group elements $g \in \mathcal{G}$. We use the standard notations [15], [16], [17] of Lie algebra \mathfrak{g} , vee operator $(\cdot)^\vee : \mathfrak{g} \rightarrow \mathbb{R}^d$, wedge operator $(\cdot)^\wedge : \mathbb{R}^d \rightarrow \mathfrak{g}$, left group multiplication $L_g : h \mapsto gh, \forall g, h \in \mathcal{G}$, matrix exponential $\exp_m : \mathfrak{g} \rightarrow \mathcal{G}$, matrix logarithm $\log_m : \mathcal{G} \rightarrow \mathfrak{g}$, exponential $\exp : \mathbb{R}^d \rightarrow \mathcal{G}, \xi \mapsto \exp_m(\xi^\wedge) = g$, and logarithm $\log : \mathcal{G} \rightarrow \mathbb{R}^d, g \mapsto \log_m(g)^\vee = \xi$. We write the *true unknown dynamics* as $s_{k+1} = f(s_k, u_k, w_k)$, $w_k \sim P_{noise}$, where f is an *unknown* deterministic function, w_k is a stochastic term drawn iid from an *unknown* distribution P_{noise} , and $u_k \in \mathbb{R}^m$ is the control. Without restricting w_k or f , we only consider approximate models \tilde{f} with prediction uncertainty modeled as Gaussian. To make uncertainty quantification tractable without imposing strong error assumptions, we assume access to an uncorrupted dataset of transitions.

Assumption 1: We are given dataset of transitions $D_{cal} = \{(s_k, u_k, s_{k+1})\}_{1:N}$, collected from the same transition distribution P_{data} observed at execution time.

D_{cal} is exchangeable with the test-time transitions $(s_k, u_k, s_{k+1})^2$, which is milder than iid (iid \Rightarrow exchangeable). While this implies access to test-time conditions, our theoretical guarantees are non-asymptotic. For a given admissible action u_{des} , we provide a C-Space prediction region $\mathcal{C}^q \subseteq \mathcal{Q}$ provably containing the resulting true *unknown* system configuration q_1 at a user-defined probability $(1 - \alpha)$, i.e. $\mathbb{P}(q_1 \in \mathcal{C}^q) \geq (1 - \alpha)$, where $\alpha \in (0, 1)$ is the user-set acceptable failure-probability. While achieving the above is trivial, e.g., by predicting $\mathcal{C}^q = \mathcal{Q}$, we want \mathcal{C}^q to be as *volume-efficient* as possible.

III. CLAPS

We construct *calibrated prediction regions* that provably contain the unknown robot configuration q_1 at the specified

likelihood (see Fig. 2). Given $s_0 = (g_0, \xi_0)$ and a commanded action u_{des} , \tilde{f} returns an expected next state $(\tilde{g}_1, \tilde{\xi}_1) := \mathbb{E}[\tilde{f}(g_0, \xi_0, u_0)]$ and an estimated uncertainty covariance $\tilde{\Sigma}_1$ on the Lie algebra. A natural error metric is the group difference $E_1 := \tilde{g}_1^{-1} g_1 \in \mathcal{G}$, or its projection to exponential coordinates $e_1 := \log(E_1) \in \mathbb{R}^d$. Yet, this point-wise metric that does not account for the estimated uncertainty. We thus define our symmetry-based uncertainty-aware non-conformity score r_1 as the Mahalanobis Distance between configuration g and the Gaussian Prediction $(\tilde{g}_1, \tilde{\Sigma}_1)$, i.e., $r_1(g; \tilde{g}_1, \tilde{\Sigma}_1) = \sqrt{\log(\tilde{g}_1^{-1} g)^\top \tilde{\Sigma}_1^{-1} \log(\tilde{g}_1^{-1} g)}$. By looping over each item in D_{cal} and calculating its score r_1 , we can build a set of scalars R_{cal} and calculate the CP threshold $\hat{q}_\alpha \in \mathbb{R}$, which is the $\lceil (1 - \alpha)(n + 1) \rceil$ largest element of R_{cal} . We then *calibrate the approximate uncertainty prediction* $N(0, \tilde{\Sigma}_1)$ in exponential coordinates, by scaling its covariance through the *conformal scaling factor* $\zeta := \hat{q}_\alpha^2 / \chi_\alpha^2(\dim \mathfrak{g}) \in \mathbb{R}$ into $N(0, \zeta \tilde{\Sigma}_1)^3$. Then the $(1 - \alpha)$ confidence region of $N(0, \zeta \tilde{\Sigma}_1)$ will contain at least $100(1 - \alpha)\%$ of the true unknown e_1 .

Theorem 1 (Thm 2 of [17]): Let \mathcal{C}' be the $100(1 - \alpha)\%$ confidence region of $N(0, \zeta \tilde{\Sigma}_1)$. Then $\mathbb{P}(e_1 \in \mathcal{C}') \geq (1 - \alpha)$.

While this guarantees the marginal probabilistic containment of the true unknown error vector e_1 in exponential coordinates, we aim to contain the true configuration q_1 in \mathcal{C}^q . To bridge this gap, we consider how e_1 propagates to C-Space: $\mathbb{R}^d \xrightarrow{(\cdot)^\wedge} \mathfrak{g} \xrightarrow{\exp_m} \mathcal{G} \xrightarrow{L_{\tilde{g}_1}} \mathcal{G} \xrightarrow{K^{-1}} \mathcal{Q}$. We can relate e_1 to group configurations using $\phi(e_1) = L_{\tilde{g}_1} \circ \exp(e_1) = \tilde{g}_1 \exp(\log(\tilde{g}_1^{-1} g_1)) = \tilde{g}_1 \tilde{g}_1^{-1} g_1 = g_1$. To ensure \mathcal{C}' lies in the bijective domain, we can clip the region as $\mathcal{C} := \mathcal{C}' \cap U^\vee$, leading to $\mathcal{C}^{\exp} := \phi(\mathcal{C})$. We can then use set relations to show:

Theorem 2: Let ϕ, \mathcal{C}^{\exp} be defined as above and $\mathcal{C} \subseteq \mathbb{R}^d$. It follows that $\mathbb{P}(g_1 \in \mathcal{C}^{\exp}) = \mathbb{P}(e_1 \in \mathcal{C})$.

The *preimage* K, K^{-1} , transports from \mathcal{G} to \mathcal{Q} . Thus:

Theorem 3: Let $g_1 := K(q_1)$. For $\mathcal{C}^{\exp} \subseteq \mathcal{G}$, its preimage is $\mathcal{C}^q := K^{-1}(\mathcal{C}^{\exp})$. Then $\mathbb{P}(q_1 \in \mathcal{C}^q) = \mathbb{P}(g_1 \in \mathcal{C}^{\exp})$.

²A random vector $D_{cal} \cup (s_k, u_k, s_{k+1}) := (s_k, u_k, s_{k+1})_{1:N+1}$ is exchangeable if its elements are equally likely to appear in any ordering.

³ $\chi_\alpha^2(\dim \mathfrak{g})$ denotes the $(1 - \alpha)$ -quantile of the χ^2 distribution of dimension $\dim \mathfrak{g}$. See [17] for more intuition into the *scaling factor* ζ .

IV. EXPERIMENTS & DISCUSSION

We evaluated our method on *second-order unicycles* with configuration in $SE(2)$. The true dynamics are unknown, inertial properties are estimated with standard system identification, and both systems are subject to *aleatoric disturbances*. Besides the errors introduced by sysID, *epistemic uncertainty* is present due to effects unmodeled by \hat{f} . We compare **CLAPS** with 7 baselines. All methods share the an *uncalibrated* initial uncertainty. Results show prediction regions and Monte Carlo (MC) particles after $\Delta t = 0.5$ sec. SS EKF uses State-Space dynamics, resulting in ellipsoidal \mathcal{C}^q . The InEKF propagates a Gaussian uncertainty on the Lie algebra. These baselines do not consider D_{cal} . InEKF+2M uses the uncentered second moment of the one-step configuration errors e_1 in D_{cal} as its uncertainty estimate, and InEKF+MLE fits both a bias correction and a centered covariance to the e_1 in D_{cal} . *None* of the four methods above provide guarantees on \mathcal{C}^q containing the future system configuration, thus being unsuitable for safety-critical control. SS PP + CP is a common approach [8], [9], [18], [19] using a point-prediction (PP) \tilde{q}_1 and using the L2 distance as score. Lie PP + CP is a naive extension, calculating the L2 distance in the Lie algebra instead of SS. SS EKF + CP [17] performs uncertainty-aware calibration in a Euclidean C-Space, using the Mahalanobis distance as score. Our proposed approach, can be interpreted as a *provably-correct symmetry-aware calibration of InEKF*. We use $\alpha = 0.1$ in all experiments.

Simulation (JetBot Fig. 2): The calibration dataset D_{cal} was collected by spanning a configuration-velocity-action grid, totalling $|D_{cal}| = 40,500$. The validation set spanned the same range with fewer elements, resulting in 625 cases. For each validation, we propagate $100k$ simulated robots to estimate *empirical coverage*, i.e., the probability that a system under the true unknown stochastic dynamics produces configurations in the prediction region \mathcal{C}^q . By averaging this metric over the 625 trials we estimate *marginal coverage*. Fig. 3 shows the workspace footprint of \mathcal{C}^q , which could be used for probabilistic obstacle avoidance. Both figures qualitatively demonstrate **CLAPS** ability to fit the underlying system uncertainty (represented by MC samples). This is supported **CLAPS**'

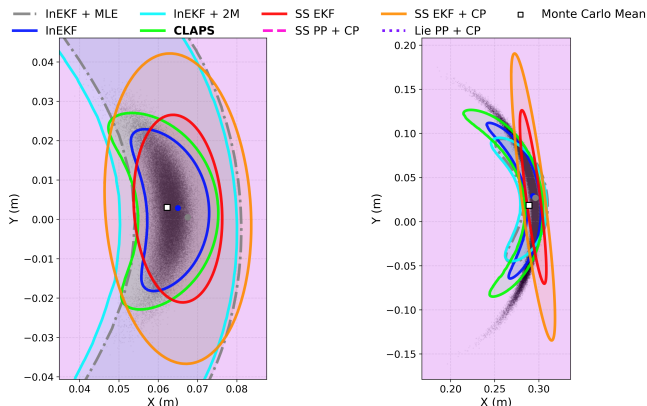


Fig. 3. Workspace (\mathbb{R}^2) Marginalization of the C-Space regions generated by all the methods, over two JetBot trials. InEKF+MLE has \tilde{q}_1 as gray dot, other methods as blue dot. PP methods generate large regions with boundaries outside margins. SS EKF, InEKF, InEKF+2M, and InEKF+MLE are not guaranteed to contain the resulting configuration at the user-set likelihood. Qualitatively, **CLAPS** appears to more accurately represent the underlying uncertainty distribution than the symmetry-unaware baselines.

TABLE I

JETBOT (SIMULATION) RESULTS (OVER 625 VALIDATION TRIALS)				
Algorithm	Marginal Coverage (%)	Avg. Volume Ratio ↓	Avg. Workspace IoU with Particles (%) ↑	Provable Guarantees?
SS EKF	78.7	0.63	35.0	✗
InEKF	82.7	0.47	41.8	✗
InEKF+2M	89.2	3.06	40.0	✗
InEKF+MLE	90.3	2.80	42.3	✗
SS PP + CP	89.9	2137	0.20	✓
Lie PP + CP	89.9	2138	0.20	✓
SS EKF + CP	91.2	2.86	30.4	✓
CLAPS	90.0	1.00	48.4	✓

red if coverage does not achieve ($<$) the user-set probability $(1 - \alpha) = 0.9$. The average volume ratio is reported relative to **CLAPS**.

larger Workspace Intersection-over-Union (IoU) with the true dynamics MC samples (see Table 1). The *uncalibrated* SS EKF and InEKF fail to satisfy the user-set specification (due to model mismatch). InEKF+2M and InEKF+MLE estimate the same uncertainty for all initial velocities and u_{des} , becoming volume inefficient. CP methods achieve at least $(1 - \alpha)\%$ coverage, as expected. Algorithms with L2 scores construct large ball-shaped \mathcal{C}^q , making them impractical. Our method produces efficient banana shaped regions containing a satisfactory probability mass of the future configurations – **CLAPS**' \mathcal{C}^q has smaller C-Space volumes than all calibrated baselines in the 625 validation trials we tested and achieves the highest average IoU with the MC Particles. These trials supported Theorems 1-3, with the MC particles satisfying $e_1 \in \mathcal{C} \Rightarrow g_1 \in \mathcal{C}^q$.

Hardware (MBot Fig. 1): The robot's pose and velocity were estimated using Motion Capture. Calibration and validation data were collected by randomly sampling u_{des} and holding it for Δt . We shuffled the data, allocating 5% for calibration ($|D_{cal}| = 237$), which corresponds to ≈ 2 min of driving, and leaving 4511 transitions for validation. Table 2 shows the results. In this low-data experiment, the approximate estimators failed to achieve the user-set requirement, while the CP methods satisfied it – as expected from the *finite-sample guarantees*. **CLAPS** produced a smaller average \mathcal{C}^q than all calibrated baselines, demonstrating its volume efficiency in real situations. Compared to SS EKF + CP, **CLAPS**' regions were on average 23% smaller and up to 75% smaller.

V. CONCLUSION

We proposed an algorithm enabling the construction *calibrated prediction regions* under *aleatoric* and *epistemic uncertainty*. By considering the robot's symmetry, our regions appear to be more volume-efficient and a better representation of the underlying uncertainty than existing approaches, extending CP guarantees from Euclidean space to $SE(2)$ robots.

TABLE II

MBOT (HARDWARE) COVERAGE, VOLUME (4511 VALIDATION TRIALS)

Algorithm	SS EKF	InEKF	InEKF+2M	InEKF+MLE	SS PP + CP	Lie PP + CP	SS EKF + CP	CLAPS
Marginal Coverage (%)	73.5	70.6	87.4	86.9	90.5	90.5	91.8	90.4
Avg. Volume Ratio ↓	0.32	0.27	0.08	0.05	2.09	2.09	1.30	1.0
Provable Guarantees?	✗	✗	✗	✗	✓	✓	✓	✓

red if coverage does not achieve ($<$) the user-set probability $(1 - \alpha) = 0.9$. The average volume ratio is reported relative to **CLAPS**.

REFERENCES

- [1] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, “Control barrier functions: Theory and applications,” in *ECC*, 2019.
- [2] M. P. Sumeet Singh and J.-J. Slotine, “Tube-based mpc: a contraction theory approach,” in *IEEE CDC*, 2016.
- [3] D. R. Agrawal and D. Panagou, “Safe and robust observer-controller synthesis using control barrier functions,” *Control Sys. Letters*, 2022.
- [4] C. E. Oestreich, R. Linares, and R. Gondhalekar, “Tube-based model predictive control with uncertainty identification for autonomous spacecraft maneuvers,” *Journ. of Guidance, Control, and Dyn.*, 2023.
- [5] H. Wang, J. Borquez, and S. Bansal, “Providing safety assurances for systems with unknown dynamics,” *Control Sys. Letters*, 2024.
- [6] M. Khan, T. Ibuki, and A. Chatterjee, “Safety uncertainty in control barrier functions using gaussian processes,” in *ICRA*, 2021.
- [7] K. Nagami and M. Schwager, “State estimation and belief space planning under epistemic uncertainty for learning-based perception systems,” *IEEE RA-L*, 2024.
- [8] L. Lindemann, M. Cleaveland, G. Shim, and G. Pappas, “Safe planning in dynamic environments using conformal prediction,” *RA-L*, 2023.
- [9] A. Dixit, L. Lindemann, S. X. Wei, M. Cleaveland, G. J. Pappas, and J. W. Burdick, “Adaptive conformal prediction for motion planning among dynamic agents,” in *LADC*, 2023.
- [10] S. Yang, G. J. Pappas, R. Mangharam, and L. Lindemann, “Safe perception-based control under stochastic sensor uncertainty using conformal prediction,” in *CDC*, 2023.
- [11] A. Long, K. Wolfe, M. Mashner, and G. Chirikjian, “The banana distribution is gaussian: A localization study with exponential coordinates,” *Robotics: Science and Systems VIII*, vol. 265, no. 1, 2013.
- [12] A. Barrau and S. Bonnabel, “The invariant extended kalman filter as a stable observer,” *IEEE TACON*, 2016.
- [13] M. Brossard, S. Bonnabel, and J.-P. Condomines, “Unscented kalman filtering on lie groups,” in *IROS*, 2017.
- [14] T. Barfoot, *State estimation for robotics*. Cambridge Uni Press, 2024.
- [15] A. M. Bloch, *Nonholonomic mechanics*. Springer, 2015.
- [16] J. Marsden and T. Ratiu, *Introduction to mechanics and symmetry*. Springer, 1999.
- [17] L. Marques and D. Berenson, “Quantifying aleatoric and epistemic dynamics uncertainty via local conformal calibration,” *WAFR*, 2024.
- [18] J. Sun, Y. Jiang, J. Qiu, P. Nobel, M. J. Kochenderfer, and M. Schwager, “Conformal prediction for uncertainty-aware planning with diffusion dynamics model,” *NeurIPS*, vol. 36, pp. 80324–80337, 2023.
- [19] K. J. Strawn, N. Ayanian, and L. Lindemann, “Conformal predictive safety filter for rl controllers in dynamic environments,” *RA-L*, 2023.