2024 American Control Conference (ACC)
July 8-12, 2024. Toronto, Canada

Distributionally Robust Path Integral Control

Hyuk Park* Duo Zhou* Grani A. Hanasusanto* Takashi Tanaka†

***Abstract*—We consider a continuous-time continuous-space stochastic optimal control problem, where the controller lacks exact knowledge of the underlying diffusion process, relying instead on a finite set of historical disturbance trajectories. In situations where data collection is limited, the controller synthesized from empirical data may exhibit poor performance. To address this issue, we introduce a novel approach named Distributionally Robust Path Integral (DRPI). The proposed method employs distributionally robust optimization (DRO) to robustify the resulting policy against the unknown diffusion process. Notably, the DRPI scheme shows similarities with risk-sensitive control, which enables us to utilize the path integral control (PIC) framework as an efficient solution scheme. We derive theoretical performance guarantees for the DRPI scheme, which closely aligns with selecting a risk parameter in risk-sensitive control. We validate the efficacy of our scheme and showcase its superiority when compared to risk-neutral and risk-averse PIC policies in the absence of the true diffusion process.**

***Index Terms*—stochastic optimal control, distributionally robust optimization, path integral method, risk-sensitive control**

I. Introduction

We consider a continuous-time and continuous-space stochastic control problem where the explicit representation of the system dynamics is unknown to the controller. In numerous real-world applications, the absence of true system dynamics is a common issue, as the system dynamics may be too complex to model, or the collection of historical data for the true dynamics may be limited. One approach to address unknown dynamics is constructing a simulator based on available data to serve as a proxy for the true environment and to test control policies in the simulator before real-world deployment.

However, when the available data is sparse, the simulator may not accurately capture the true characteristics of the system dynamics. In other words, the probability distribution of the disturbance in the simulator model may not faithfully represent the distribution of the true disturbance in the real-world system. Such a distributional mismatch may cause the simulator-based control policy to exhibit poor performance in the true system. To address this issue, we propose a distributionally robust (DR) control problem, using the emerging distributionally robust optimization (DRO) paradigm. The proposed approach constructs an ambiguity set that contains all possible distributions from the nominal distribution within a certain distance. In the literature, various measures have been used to define the distance between two distributions [1], [2], [3]. In this paper, we utilize the Kullback-Leibler (KL) divergence [4] for the ambiguity set. Subsequently, the DR control problem seeks a policy that performs optimally under the worst-case disturbance distribution taken from the KL divergence-based ambiguity set. Interestingly, our DR control problem shows similarities with the traditional risk-sensitive control problem [5]. While several papers have drawn a connection between risk-sensitive control and DR control in discrete-time settings [6], [7], there has not been prior work demonstrating this equivalence in continuous-time settings.

The similarities between the two problems allow us to efficiently solve the DR control problem using the path integral control (PIC) method. PIC has emerged as a promising stochastic optimal control framework for designing feedback control systems in the literature [8], [9]. It has recently found applications in diverse robotics domains, ranging from tasks such as aggressive driving [10], [11], off-road navigation [12] to multi-quadrotor control [13], and more [14], [15]. The key concept of PIC involves using the Feynman-Kac lemma [16] to convert the value function of the stochastic control problem to an expectation on all possible uncontrolled trajectories. This transformation allows for the approximation of the optimal control input using sample trajectories of uncontrolled system dynamics generated through the Monte Carlo method.

PIC offers several advantages over traditional optimal control techniques. One notable advantage is its capability to handle arbitrary nonlinear state-dependent cost functions and state-dependent dynamics. This flexibility makes PIC well-suited for systems with complex behaviors that are difficult to capture using traditional methods. In practice, to expedite the computation of the optimal control, PIC can harness modern graphics processing units (GPU) which can generate a large number of sample trajectories in parallel [13]. This parallelization significantly speeds up the computation process, enhancing the practical applicability of PIC for complex systems.

Recently, several works have proposed robust model predictive path integral control frameworks utilizing diverse methods, including covariance steering [17], conditional value-at-risk [18], control barrier functions [19], and robust sampling [20], [21]. However, to the best of our knowledge, our work is the first to combine PIC with the DRO method-

We would like to acknowledge the support of the National Science Foundation under grants 2342505 and 2343869.

*Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign. {hyukp2, duozhou2, gah}@illinois. edu

†Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin. ttanaka@utexas.edu

ology. The contributions of this paper are as follows:

1) We formulate a continuous-time and continuous-space stochastic control problem as the DR control problem and establish its equivalence with risk-sensitive control and H-infinity control.
2) We provide finite sample guarantees for the DR control policy under certain conditions. Our theoretical results serve as valuable guidance for the selection of the robust parameter.
3) In pursuit of an efficient solution scheme for the DR control problem, we propose a path integral-based algorithm and demonstrate its superior performance via numerical experiments in an online context.

The remaining sections of this paper are organized as follows. In Section II, we define the stochastic control problem of our interest. In Section III, we present our DR control problem and its reformulation as a single-level minimization problem. In Section IV, we introduce the data-driven approach to the DR control problem and analyze the theoretical performance guarantees. In Section V, the solution scheme based on path integral control is presented. In Section VI, we present numerical experiments to demonstrate the effectiveness of our scheme. We conclude the paper with Section VII, which summarizes the contributions of this work and suggests avenues for future research.

Notation

Bold letters represent vectors and matrices, while regular fonts indicate scalars. the identity matrix is denoted as $\mathbf{I}$—its dimension will be evident from the context. The partial derivatives with respect to the state $\boldsymbol{x}$ and time t are denoted by ∂_x and ∂_t, respectively. The trace operation of a square matrix $\boldsymbol{A}$ is denoted as $\mathrm{tr}(A)$. N sequences, each containing K vectors, are denoted as $\{\boldsymbol{x}^{(i)}(k)\}_{k=1}^{K}$, $\forall i = 1, 2, \ldots, N$.

II. Problem Statement

We consider a continuous-time and continuous-space stochastic dynamic $\boldsymbol{x}(t) \in \mathbb{R}^n$ that is affine in control and disturbance as follows:

$$\begin{aligned} d\boldsymbol{x}(t) =& f\left(\boldsymbol{x}(t), t\right) dt + \boldsymbol{G}\left(\boldsymbol{x}(t), t\right) \boldsymbol{u}(\boldsymbol{x}(t), t) dt + \\ & \boldsymbol{\Sigma}(\boldsymbol{x}(t), t) d\boldsymbol{\xi}(t). \end{aligned} \tag{1}$$

Here, $f\left(\boldsymbol{x}(t), t\right) \in \mathbb{R}^n$ is an arbitrary passive dynamic, $\boldsymbol{u}(\boldsymbol{x}(t), t) \in \mathbb{R}^k$ is a control input, $\boldsymbol{G}(\boldsymbol{x}(t), t) \in \mathbb{R}^{n \times k}$ is a full-rank control transition matrix function with $n \geq k$. In addition, $\boldsymbol{\xi}(t) \in \mathbb{R}^p$ is a diffusion process on a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\boldsymbol{\Sigma}(\boldsymbol{x}(t), t) \in \mathbb{R}^{n \times p}$ is a full-rank diffusion matrix function that maps the disturbance to the state. We define the diffusion process $\boldsymbol{\xi}(t)$ adapted to the filtration $\mathcal{F}_t$ as

$$d\boldsymbol{\xi}(t) = \boldsymbol{\mu}(t) dt + d\boldsymbol{w}. \tag{2}$$

Here, $\boldsymbol{\mu}(t)$ represents the drift, and $d\boldsymbol{w} = \{\mathbf{w}(t) : t \geq 0\}$ is a standard Brownian disturbance with respect to the probability law $\mathbb{P}^\star$. In this paper, we make the assumption that the controller lacks access to the true diffusion process (2) and $\mathbb{P}^\star$. The assumption aligns with many real-world applications where accurately characterizing the actual disturbance is challenging.

Given a finite time horizon $t \in [0, T]$, the initial state $\boldsymbol{x}(0)$, a running cost function $\mathcal{L}_{\boldsymbol{u}}(\boldsymbol{x}(t), t)$, and a terminal cost function $\psi(\boldsymbol{x}(T), T)$, the stochastic optimal control problem is formulated as

$$\inf_{\boldsymbol{u}} \mathbb{E}_{\mathbb{P}^\star}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\boldsymbol{x}(0), 0\right)\Big], \tag{3}$$

where $\mathcal{J}_{\boldsymbol{u}}(\cdot) = \psi\left(\boldsymbol{x}(T), T\right) + \int_0^T \mathcal{L}_{\boldsymbol{u}}\left(\boldsymbol{x}(s), s\right) ds$. Here, the dependence of $\mathcal{L}_{\boldsymbol{u}}(\cdot)$ and $\mathcal{J}_{\boldsymbol{u}}(\cdot)$ on $\boldsymbol{u}$ implies a certain policy $\boldsymbol{u}(\boldsymbol{x}(t), t)$ imposed for evaluating the costs and $\mathbb{E}_{\mathbb{P}^\star}[\cdot]$ is the expectation evaluated under $\mathbb{P}^\star$. We assume that the running cost function $\mathcal{L}_{\boldsymbol{u}}(\cdot)$ consists of an arbitrary state-dependent cost $q(\cdot)$ and a quadratic control cost with a positive definite weight matrix $\boldsymbol{R} \in \mathbb{R}^{p \times p}$, given by

$$\mathcal{L}_{\boldsymbol{u}}\left(\boldsymbol{x}(t), t\right) = q\left(\boldsymbol{x}(t), t\right) + \frac{1}{2} \boldsymbol{u}(\boldsymbol{x}(t), t)^\top \boldsymbol{R} \boldsymbol{u}(\boldsymbol{x}(t), t). \tag{4}$$

III. Distributionally Robust Control Problem

Designing the optimal policy for the stochastic control problem (3) clearly requires knowledge of the true diffusion process (2), which is unrealistic in many cases. Instead, we consider a scenario where the controller has access to an alternative nominal diffusion process $\widehat{\boldsymbol{\xi}}(t)$, denoted as $d\widehat{\boldsymbol{\xi}}(t) = \widehat{\boldsymbol{\mu}}(t) dt + d\boldsymbol{w}$, where $\boldsymbol{w}$ is a Brownian disturbance under the probability law $\mathbb{Q}$. In this case, a naïve approach to designing a control policy might involve constructing the optimal controller based on the dynamic system where the true diffusion process (2) is replaced by the nominal diffusion process. Nonetheless, this approach leads to sub-optimal performance if the nominal diffusion process fails to accurately represent the true system dynamics. In fact, this issue is common for simulator-based control policies, as the policies synthesized in erroneous simulator models often exhibit inferior performance when implemented in the actual system.

To address this issue, we adopt the emerging paradigm of distributionally robust optimization (DRO) to formulate the distributionally robust (DR) control problem for the true problem (3) as follows:

$$\inf_{\boldsymbol{u}} \sup_{\mathbb{P} \in P^\gamma(\mathbb{Q})} \mathbb{E}_{\mathbb{P}}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\boldsymbol{x}(0), 0\right)\Big]. \tag{5}$$

Here, the distributional (ambiguity) set $P^\gamma(\mathbb{Q})$ with robustness parameter $\gamma > 0$ is defined as

$$\begin{aligned} P^\gamma\left(\mathbb{Q}\right) = \Big\{ & \mathbb{P} \in D \ : \ \mathbb{D}\left(\mathbb{P} \| \mathbb{Q}\right) = \\ & \int_\Xi \log \frac{d\mathbb{P}}{d\mathbb{Q}}(\boldsymbol{\xi}) d\mathbb{P}(\boldsymbol{\xi}) \leq \gamma \Big\}. \end{aligned} \tag{6}$$

Here, D denotes the set of all probability laws of $\boldsymbol{\xi}(\cdot)$ and $\mathbb{D}(\mathbb{P} \| \mathbb{Q})$ denotes the Kullback-Leibler (KL) divergence from

$\mathbb{P}$ to $\mathbb{Q}$ where $d\mathbb{P}/d\mathbb{Q}$ is the likelihood ratio between $\mathbb{P}$ and $\mathbb{Q}$, also known as the Radon-Nikodym derivative.

The DR control problem (5) seeks a policy that performs best under the worst-case probability law $\mathbb{P}$ within the ambiguity set, thus providing robustness against the unknown true diffusion process (2). Compared to classical robust control, which is designed to optimize against worst-case disturbances, the DR control policy is less conservative, resulting in better performance across various applications, including robotics [7], control design [22], and power systems [23], etc. Furthermore, the DR control framework is well-suited for data-driven settings where a nominal diffusion process can be constructed based on available data. We will discuss how to construct the DR control policy in a data-driven manner in Section IV.

Tractable Reformulation

The min-max problem (5) is inherently difficult to solve since the cost function involves a maximization problem. To design an efficient solution scheme, we introduce an equivalent single-level problem for the DR control problem (5) by following the standard results of the convex analysis in [24] and duality between relative entropy and free energy in [6], [25]. We first make the following assumption

$$\sup_{\mathbb{P}\in D} \mathbb{E}_{\mathbb{P}}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\boldsymbol{x}(0),0\right)\Big] = \infty. \tag{7}$$

As discussed in [6], this assumption states that, without the KL divergence constraint, some arbitrary diffusion process can drive the expected cost to infinity. With this assumption, we present the single-level reformulation for the DR control problem in the following lemma.

Lemma 1. *For any given $\gamma > 0$, the DR control problem* (5) *can be equivalently reformulated as the following single-level minimization problem*

$$\inf_{\boldsymbol{u},\theta>0} \gamma\theta + \theta \log \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\frac{1}{\theta}\mathcal{J}_{\boldsymbol{u}}\left(\boldsymbol{x}(0),0\right)\right)\right]. \tag{8}$$

Proof. Dualizing the inner maximization problem in (5) with the ambiguity set (6) as a constraint, we have

$$\sup_{\mathbb{P}\in D} \inf_{\theta\geq 0} \mathbb{E}_{\mathbb{P}}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\cdot\right)\Big] - \theta\mathbb{D}\left(\mathbb{P}\|\mathbb{Q}\right) + \gamma\theta \tag{9}$$

$$\leq \inf_{\theta\geq 0} \gamma\theta + \sup_{\mathbb{P}\in D} \mathbb{E}_{\mathbb{P}}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\cdot\right)\Big] - \theta\mathbb{D}\left(\mathbb{P}\|\mathbb{Q}\right) \tag{10}$$

$$= \inf_{\theta>0} \gamma\theta + \sup_{\mathbb{P}\in D} \mathbb{E}_{\mathbb{P}}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\cdot\right)\Big] - \theta\mathbb{D}\left(\mathbb{P}\|\mathbb{Q}\right). \tag{11}$$

Here, the inequality between (9) and (10) holds due to weak duality, and the equality between (10) and (11) holds since the objective function in (10) evaluates to infinity for $\theta = 0$ by the assumption (7).

To derive the single-level problem (8), we first establish strong duality (i.e., the equality) between (9) and (10) by following [24, Theorem 1, Chapter 8]. It is clear that D in (9) is a convex set and the objective function in (9) is concave in $\mathbb{P}$ since $\mathbb{E}_{\mathbb{P}}[\mathcal{J}_{\boldsymbol{u}}(\cdot)]$ is linear in $\mathbb{P}$ and $\mathbb{D}\left(\mathbb{P}\|\mathbb{Q}\right)$ is convex in $\mathbb{P}$. Furthermore, we can show the existence of an interior point $\mathbb{P} \in D$: Let $\mathbb{P} = \mathbb{Q}$, then the following strict inequality holds:

$$\mathbb{D}\left(\mathbb{Q}\|\mathbb{Q}\right) = \int_{\Xi} \log \frac{d\mathbb{Q}}{d\mathbb{Q}}(\boldsymbol{\xi}) d\mathbb{P}(\boldsymbol{\xi}) = 0 < \gamma$$

for any $\gamma > 0$. Hence, by incorporating the minimization over $\boldsymbol{u}$ with (11), the DR control problem (5) becomes equivalent to

$$\inf_{\boldsymbol{u},\theta>0} \gamma\theta + \sup_{\mathbb{P}\in D} \mathbb{E}_{\mathbb{P}}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\cdot\right)\Big] - \theta\mathbb{D}\left(\mathbb{P}\|\mathbb{Q}\right).$$

Then, the remainder of the proof amounts to showing

$$\sup_{\mathbb{P}\in D} \mathbb{E}_{\mathbb{P}}\Big[\mathcal{J}_{\boldsymbol{u}}\left(\cdot\right)\Big] - \theta\mathbb{D}\left(\mathbb{P}\|\mathbb{Q}\right) = \theta \log \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\frac{1}{\theta}\mathcal{J}_{\boldsymbol{u}}\left(\cdot\right)\right)\right] \tag{12}$$

for all $\theta > 0$. This equality (12), known as the Legendre duality between the KL divergence and free energy, is already shown in [25, Section 4.6.3]. This completes the proof. □

Note that the reformulation (8) exhibits similarities with the risk-sensitive control problem [5]. In fact, if we consider θ as a parameter, the problem (8) is equivalent to the risk-sensitive control problem. This shows the intimate relationship between DR control and risk-sensitive control. A low value of the robustness parameter γ in (8) favors a high value of θ. In particular, as γ converges to 0, the problem becomes equivalent to the risk-neutral control problem. Conversely, a higher value of γ corresponds to a smaller value of θ. Specifically, DR control can be viewed as *risk-averse* control with an alternative parameter $\gamma > 0$ since the parameterization of γ excludes risk-seeking behaviors, i.e., $\theta < 0$.

While the connection between risk-sensitive control and DR control has been shown in several papers [6], [7], they have focused on discrete-time settings. More importantly, it remains an open question on how to determine the risk-sensitive parameter θ. In the literature, the selection of θ often involves a trial-and-error process where policies with different values of θ are tested in the actual environment until the resulting policy aligns with the modeler's risk profile. However, such experimentation can be costly in safety-critical tasks. On the other hand, as will be discussed in Section IV, our DR control framework offers theoretical performance guarantees in data-driven settings, guiding the selection of the robustness parameter γ based on available data. This choice corresponds to an appropriately selected θ in the risk-averse control framework.

IV. Data-Driven Approach

In this section, we explore the application of the DR control framework in a data-driven context. Specifically, we discuss the construction of the approximate diffusion process and the corresponding probability law $\mathbb{Q}$ using available data. We introduce an additional assumption in this section: the drift in the true diffusion process is time-invariant, i.e.,

$$d\boldsymbol{\xi} = \boldsymbol{\mu} dt + d\boldsymbol{w}. \tag{13}$$

While the DR control problem (5) discussed in Section III can accommodate a time-variant diffusion process (2), the assumption (13) enables us to construct an approximate diffusion process in a data-driven manner and establish theoretical performance guarantees for the DR control policy, as we will discuss in this section.

In the absence of the true diffusion process (13), we assume that the controller merely has access to N sequences of historical disturbance terms collected over time interval $\Delta t > 0$ denoted as $\{\Delta\boldsymbol{\xi}^{(i)}(k\Delta t)\}_{k=1}^{K},\ \forall i = 1, 2, \ldots, N$, where $K = T/\Delta t$ is the number of empirical disturbances for each sequence. Subsequently, we can construct the approximate diffusion process

$$d\widehat{\boldsymbol{\xi}} = \widehat{\boldsymbol{\mu}}dt + d\boldsymbol{w},\ \text{where}\ \widehat{\boldsymbol{\mu}} = \frac{1}{NK\Delta t}\sum_{i=1}^{N}\sum_{k=1}^{K}\Delta\boldsymbol{\xi}^{(i)}(k\Delta t) \tag{14}$$

and $d\boldsymbol{w}$ is a Brownian disturbance under the probability law $\mathbb{Q}$. Consequently, we can employ (14) as a nominal diffusion process in the ambiguity set $P^{\gamma}(\mathbb{Q})$ defined in (6), rendering a data-driven DR control model.

Performance Guarantees

We provide the finite sample guarantee of our DR control policy. The ambiguity set $P^{\gamma}(\mathbb{Q})$ centered at $\mathbb{Q}$ of the approximate diffusion process (14) can be viewed as a random object in a sense that different realizations of the empirical disturbance terms $\{\Delta\boldsymbol{\xi}^{(i)}(k\Delta t)\}_{k=1}^{K},\ \forall i = 1, 2, \ldots, N$, may result in a different approximate diffusion process. Intuitively, more data would provide a more reliable estimate of the true diffusion process (13). We can establish the following generalization bound by leveraging measure concentration theory and Girsanov's theorem [26].

Proposition 1. *Consider $\mathbb{P}^{\star}$ and $\mathbb{Q}$ as the probability laws of the true and approximate diffusion processes, given by* (13) *and* (14)*, respectively. Then, for any given value of the robust parameter γ , we have*

$$\mathbb{P}^{\star} \in P^{\gamma}(\mathbb{Q})\ \textit{w.p. at least}\ 1 - 2p\exp\left(-\frac{\gamma N}{\sqrt{p}}\right), \tag{15}$$

where p is the dimension of the diffusion process.

Proof. Suppose $\mathbb{P}^{\star} \in P^{\gamma}(\mathbb{Q})$ for any given value of γ. This implies

$$\mathbb{D}\left(\mathbb{P}^{\star}\|\mathbb{Q}\right) = \int_{\Xi}\log\frac{d\mathbb{P}^{\star}}{d\mathbb{Q}}(\boldsymbol{\xi})d\mathbb{P}^{\star}(\boldsymbol{\xi}) \leq \gamma. \tag{16}$$

Applying Girsanov's Theorem [26, Theorem 8.6.5], the KL divergence in (16) is equivalent to

$$\mathbb{D}\left(\mathbb{P}^{\star}\|\mathbb{Q}\right) = \mathbb{E}_{\mathbb{P}^{\star}}\left[-\int_0^T \boldsymbol{\mu}^{\top}d\boldsymbol{w} + \frac{1}{2}\int_0^T \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 ds\right],$$

where $\widehat{\boldsymbol{\mu}}$ is defined in (14). Since $\boldsymbol{w} \in \mathbb{R}^p$ is a Brownian disturbance under $\mathbb{P}^{\star}$, we can rewrite (16) as

$$\frac{T}{2}\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 \leq \gamma. \tag{17}$$

For any time step size $\Delta t > 0$, we know that the empirical disturbances $\Delta\boldsymbol{\xi}^{(i)}(k\Delta t)$, for all $i = 1, 2, \ldots, N$, and $k = 1, 2, \ldots, K$, are i.i.d. Gaussian samples drawn from $\mathcal{N}(\Delta t\boldsymbol{\mu}, \Delta t\mathbf{I})$. Therefore, using Hoeffding's inequality [27] and union bound, we have

$$\Pr\left[\frac{T}{2}\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 \leq \gamma\right] \geq 1 - 2p\exp\left(-\frac{\gamma N}{\sqrt{p}}\right).$$

□

Proposition 1 provides a lower bound on the probability that the unknown $\mathbb{P}^{\star}$ is contained in $P^{\gamma}(\mathbb{Q})$ for a fixed γ. Similarly, for any fixed $\epsilon \in (0, 1)$, we can obtain the smallest value of $\gamma(\epsilon)$ that guarantees $\mathbb{P}^{\star} \in P^{\gamma}(\mathbb{Q})$ w.p. $1 - \epsilon$: by setting $\epsilon = 2p\exp\left(-\gamma N/\sqrt{p}\right)$ in (15), we can rewrite (15) as

$$\mathbb{D}\left(\mathbb{P}^{\star}\|\mathbb{Q}\right) \leq \gamma(\epsilon) = \frac{\sqrt{p}}{N}\log\left(\frac{2p}{\epsilon}\right) \tag{18}$$

w.p. at least $1 - \epsilon$.

If $\mathbb{P}^{\star}$ generated by the true stochastic process (13) is indeed contained in the ambiguity set, the optimal value of the DR control problem serves as an upper bound on the true cost of implementing the DR control policy in the real environment. Finally, using (18), we propose the following finite sample guarantee for our DR control policy.

Theorem 1. *Suppose that $\widehat{\mathcal{J}}^N$ and $\widehat{\boldsymbol{u}}^N(\boldsymbol{x}(t), t),\ \forall t \in [0, T]$, represent the optimal value and the optimal policy of the DR control problem* (5) *with ambiguity set $P^{\gamma(\epsilon)}(\mathbb{Q})$ where we set $\gamma(\epsilon) = \sqrt{p}/N\log(2p/\epsilon)$ for a fixed value of $\epsilon \in (0, 1)$ as in* (18)*. Then, we have*

$$\mathbb{E}_{\mathbb{P}^{\star}}\left[\mathcal{J}_{\widehat{\boldsymbol{u}}^N}\left(\boldsymbol{x}(0), 0\right)\right] \leq \widehat{\mathcal{J}}^N \tag{19}$$

w.p. at least $1 - \epsilon$.

Proof. The claim immediately holds from (18) since

$$\mathbb{E}_{\mathbb{P}^{\star}}\left[\mathcal{J}_{\widehat{\boldsymbol{u}}^N}\left(\boldsymbol{x}(0), 0\right)\right] \leq \inf_{\boldsymbol{u}}\sup_{\mathbb{P}\in P^{\gamma}(\mathbb{Q})}\mathbb{E}_{\mathbb{P}}\left[\mathcal{J}_{\boldsymbol{u}}\left(\boldsymbol{x}(0), 0\right)\right],$$

whenever $\mathbb{P}^{\star} \in P^{\gamma}(\mathbb{Q})$. □

Theorem 1 provides valuable guidance for modelers on choosing $\gamma(\epsilon)$ that guarantees a prescribed confidence level $1-\epsilon$ prior to real-world implementation of the control policy. Furthermore, given that γ is at most $\mathcal{O}(1/N)$ as shown in (18), we can adjust its value at a rate of $1/N$ as more empirical data becomes available.

Remark 1. *The equivalence between risk-sensitive control and H-infinity control in the linear quadratic Gaussian (LQG) setting was initially noted in [28]. However, they highlighted the equivalence between two Riccati equations from the two different controls without any consideration of relative entropy. As demonstrated in the proof of Proposition 1, Girsanov's theorem transforms the ambiguity set* (6) *into the uncertainty set* (17) *concerning the unknown drift $\boldsymbol{\mu}$, bounded by the robustness parameter γ. This implies the equivalence between our DR control and the nonlinear*

generalization of H-infinity control, hence, expanding upon the earlier observation.

V. Solution Scheme

A. Decomposition

Using the equivalence between the risk-sensitive control and the DR control shown in Section III, we can decompose the reformulated DR control problem (8) into two minimization problems as follows: the master problem is

$$\inf_{\theta>0} \gamma\theta + g(\theta), \tag{20}$$

and the subproblem is

$$g(\theta) = \inf_{\boldsymbol{u}} \theta \log \mathbb{E}_{\mathbb{Q}} \left[\exp\left(\frac{1}{\theta} \mathcal{J}_{\boldsymbol{u}}(\boldsymbol{x}(0), 0) \right) \right]. \tag{21}$$

Note that the master problem (20) is merely an univariate optimization problem over $\theta > 0$ which can be solved by various methods. Then, for a fixed θ, the subproblem (21) becomes the risk-sensitive control problem where the expectation of the exponentiated cost function $\mathcal{J}_{\boldsymbol{u}}(\cdot)$ is evaluated under $\mathbb{Q}$.

B. Risk-Sensitive Path Integral Control

To efficiently solve the subproblem (21), we utilize the risk-sensitive path integral control framework proposed in [29]. They demonstrated that the standard (i.e., risk-neutral) path integral method [8] can be generalized to risk-sensitive control under the same assumption. In this section, we briefly restate their main results for clarity and completeness. Further details and derivations can be found in [29].

Solving the risk-sensitive control problem (21) involves setting up the following second-order partial differential equation (PDE) known as the stochastic Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} -\partial_t \mathcal{V}^\theta(\boldsymbol{x}(t), t) = \inf_{\boldsymbol{u}} \Bigg(& \partial_x \mathcal{V}^{\theta\top}(f + \boldsymbol{G}\boldsymbol{u} + \boldsymbol{\Sigma}\widehat{\boldsymbol{\mu}}) \\ & + \frac{1}{2} \operatorname{tr}\left(\partial_{\boldsymbol{x}}^2 \mathcal{V}^\theta \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \right) + q + \frac{1}{2} \boldsymbol{u}^\top \boldsymbol{R} \boldsymbol{u} \\ & + \frac{1}{2\theta} \left\| \boldsymbol{\Sigma}^\top \partial_{\boldsymbol{x}} \mathcal{V}^\theta \right\|^2 \Bigg)(\boldsymbol{x}(t), t) \end{aligned} \tag{22}$$

with boundary condition $\mathcal{V}^\theta(\boldsymbol{x}(T), T) = \psi(\boldsymbol{x}(T), T)$. Taking derivative with respect to $\boldsymbol{u}$ on the right-hand side in (22), one can derive the optimal control

$$\boldsymbol{u}(\boldsymbol{x}(t), t) = -\boldsymbol{R}^{-1} \boldsymbol{G}(\boldsymbol{x}(t), t)^\top \partial_{\boldsymbol{x}} \mathcal{V}^\theta(\boldsymbol{x}(t), t). \tag{23}$$

Substituting (23) into (22), we have

$$\begin{aligned} -\partial_t \mathcal{V}^\theta(\boldsymbol{x}(t), t) = \Bigg(& \partial_x \mathcal{V}^{\theta\top}(f + \boldsymbol{\Sigma}\widehat{\boldsymbol{\mu}}) + q \\ & + \frac{1}{2} \operatorname{tr}\left(\partial_{\boldsymbol{x}}^2 \mathcal{V}^\theta \left(\frac{1}{\theta} \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top - \boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G}^\top \right) \right) \Bigg)(\boldsymbol{x}(t), t). \end{aligned} \tag{24}$$

Note that the HJB equation (24) is generally nonlinear in $\mathcal{V}^\theta(\cdot)$ due to the last term on the right-hand side. The conventional solution method is to solve the HJB equation backward in time over the entire time horizon $[0, T]$ for all $\boldsymbol{x}(t)$. This recursive backward evaluation suffers from the curse of dimensionality, becoming intractable as the dimension of the state space increases.

Risk-sensitive path integral control can be used as an alternative solution approach to the backward recursion for the special case of (24) where there exists θ^* that satisfies the equation

$$\theta^* \boldsymbol{G}(\boldsymbol{x}, t) \boldsymbol{R}^{-1} \boldsymbol{G}(\boldsymbol{x}, t)^\top = \boldsymbol{\Sigma}(\boldsymbol{x}, t) \boldsymbol{\Sigma}(\boldsymbol{x}, t)^\top. \tag{25}$$

Note that the equality (25) implies that the HJB equation (24) becomes linearizable since the nonlinear term in $\mathcal{V}^\theta(\cdot)$ disappears if $\theta = \theta^*$. In a one-dimensional case, (25) trivially holds, while, in a higher dimensional space, it may impose constraints on the choice of the control dependent cost matrix $\boldsymbol{R}$. Therefore, in this paper, we make the assumption that θ satisfying (25) always exists. Therefore, if a chosen θ for the subproblem (21) coincides with θ^*, the HJB equation (24) is immediately linear. Otherwise, we can utilize a log transformation of the value function, defined as:

$$\mathcal{V}^\theta(\boldsymbol{x}(t), t) = -\frac{1}{\theta'} \log\left(\Psi(\boldsymbol{x}(t), t) \right), \tag{26}$$

where $\theta' = \theta/(\theta^*\theta - 1)$. This transformation renders the HJB equation linear in terms of $\Psi(\cdot)$. Subsequently, relying on the well-known Feynman-Kac lemma [16], we can derive the solution to the linearized HJB equation using the path integral control framework with the dynamic (1) being affine in control and disturbance, and the quadratic control cost in (3) as defined earlier. This framework enables the computation of the value function using all possible forward trajectories of the *uncontrolled* dynamics, i.e., $d\boldsymbol{x}(t) = f(\boldsymbol{x}(t), t)dt + \boldsymbol{\Sigma}(\boldsymbol{x}(t), t)d\boldsymbol{\xi}(t)$, as follows:

$$\mathcal{V}^\theta(\boldsymbol{x}(t), t) = \theta' \log \mathbb{E}_{\mathbb{Q}} \left[\exp\left(\frac{1}{\theta'} \mathcal{J}_{\boldsymbol{0}}(\boldsymbol{x}(t), t) \right) \right], \tag{27}$$

where $\mathcal{J}_{\boldsymbol{0}}(\cdot) = \psi(\boldsymbol{x}(T)) + \int_t^T q(\boldsymbol{x}(s))\, ds$ represents the cost of an uncontrolled trajectory. Furthermore, as demonstrated in [30], taking the derivative of $\Psi(\cdot)$ with respect to $\boldsymbol{x}$ yields the optimal control for the problem (3) at time t, as follows:

$$\begin{aligned} & \boldsymbol{u}(\boldsymbol{x}(t), t)\, dt = \\ & \boldsymbol{R}^{-1} \boldsymbol{G}_c^\top (\boldsymbol{G}_c \boldsymbol{R}^{-1} \boldsymbol{G}_c^\top)^{-1}(\boldsymbol{x}(t), t) \frac{\mathbb{E}_{\mathbb{Q}}\left[\exp\left(\mathcal{J}_{\boldsymbol{0}}/\theta' \right) \boldsymbol{\Sigma}_c d\boldsymbol{w}(t) \right]}{\mathbb{E}_{\mathbb{Q}}\left[\exp(\mathcal{J}_{\boldsymbol{0}}/\theta') \right]}. \end{aligned} \tag{28}$$

Here, $\boldsymbol{G}_c(\cdot) \in \mathbb{R}^{(n-l)\times p}$ and $\boldsymbol{\Sigma}_c(\cdot) \in \mathbb{R}^{(n-l)\times k}$ represent submatrices of the control transition and diffusion matrices $\boldsymbol{G}(\cdot)$ and $\boldsymbol{\Sigma}(\cdot)$, respectively. These submatrices correspond to the directly actuated states denoted as $\boldsymbol{x}_c(t) \in \mathbb{R}^l$ within the state vector $\boldsymbol{x}(t) = [\boldsymbol{x}_c(t),\ \boldsymbol{x}_p(t)]^\top$ where $l \leq n$ without loss of generality. The remaining part of the states, $\boldsymbol{x}_p(t)$, represents non-directly actuated states.

Algorithm 1: DRPI

Input: $\boldsymbol{x}(0)$: Initial state;
$f(\cdot), \boldsymbol{G}(\cdot), \boldsymbol{\Sigma}(\cdot)$: System dynamics;
$\boldsymbol{G}_c(\cdot)/\boldsymbol{\Sigma}_c(\cdot)$: Submatrix for control transition/diffusion matrix;
$q(\cdot), \psi(\cdot)/\boldsymbol{R}$: State/Control cost;
$\widehat{\boldsymbol{\mu}}(\cdot)$: Estimated drift term;
γ: Robustness parameter;
M: Number of sample trajectories;
Δt: Step size

Output: Control input $\boldsymbol{u}(k)$ for $k = 0, 1 \ldots, K-1$

1 **for** $k \leftarrow 0$ *to* $K-1$ **do**
2 Sample M trajectories of disturbance $\{\boldsymbol{\varepsilon}^{(i)}(\tau)\}_{i=1}^{M}$ for $\tau = k, k+1, \ldots, K-1$
3 **for** $i \leftarrow 1$ *to* M **do**
4 Initialize the i-th cost $\mathcal{J}_{\mathbf{0}}^{(i)} \leftarrow 0$
5 **for** $k' \leftarrow k$ *to* $K-1$ **do**
6 $\boldsymbol{x}^{(i)}(k'+1) = \boldsymbol{x}^{(i)}(k') + f(\boldsymbol{x}^{(i)}(k'), k')\Delta t + \boldsymbol{\Sigma}(\boldsymbol{x}^{(i)}(k'), k')(\widehat{\boldsymbol{\mu}}(\boldsymbol{x}^{(i)}(k'), k')\Delta t + \boldsymbol{\varepsilon}^{(i)}(k')\sqrt{\Delta t}\,)$
7 **if** $k' < K-1$ **then**
8 $\mathcal{J}_{\mathbf{0}}^{(i)} += q(\boldsymbol{x}^{(i)}(k'+1), k'+1)$
9 **else**
10 $\mathcal{J}_{\mathbf{0}}^{(i)} += \psi\left(\boldsymbol{x}^{(i)}(K), K\right)$
11 $\widehat{\theta} \in \arg\min_{\theta>0} \gamma\theta + \theta' \log\left(\frac{1}{M}\sum_{i=1}^{M} \exp\left(\frac{\mathcal{J}_{\mathbf{0}}^{(i)}}{\theta'}\right)\right)$
12 **for** $i \leftarrow 1$ *to* M **do**
13 $r^{(i)} = \exp\left(\frac{\mathcal{J}_{\mathbf{0}}^{(i)}}{\widehat{\theta}}\right) \Big/ \sum_{j=1}^{M} \exp\left(\frac{\mathcal{J}_{\mathbf{0}}^{(j)}}{\widehat{\theta}}\right)$
14 $\boldsymbol{u}_k \leftarrow \boldsymbol{R}^{-1}\boldsymbol{G}_c^{\top}(\boldsymbol{G}_c\boldsymbol{R}^{-1}\boldsymbol{G}_c^{\top})^{-1}(\boldsymbol{x}(k), k)\sum_{i=1}^{M} r^{(i)}\frac{\boldsymbol{\Sigma}_c\boldsymbol{\varepsilon}^{(i)}(k)}{\sqrt{\Delta t}}$
15 Send $\boldsymbol{u}(k)$ to the controller
16 Update the current state $\boldsymbol{x}(k+1)$

C. Numerical Method

The numerical implementation of the PIC requires two types of approximation, namely, time discretization and sampling trajectories. First, we can approximate the continuous-time dynamic system (1) using the Euler-Maruyama method [31]:

$$\boldsymbol{x}(k+1) = \boldsymbol{x}(k) + f(\boldsymbol{x}(k), k)\Delta t + \boldsymbol{G}(\boldsymbol{x}(k), k)\boldsymbol{u}(\boldsymbol{x}(k), k)\Delta t + \boldsymbol{\Sigma}(\boldsymbol{x}(k), k)\Delta\boldsymbol{\xi}(k)$$

for $k = 0, 1, \ldots, K-1$, where time step size $\Delta t > 0$ determines the total number of time steps, i.e., $K = T/\Delta t$ and $\Delta\boldsymbol{\xi}(k) = \boldsymbol{\mu}(k)\Delta t + \boldsymbol{\varepsilon}\sqrt{\Delta t}$ is the discrete-time diffusion process where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Subsequently, we can approximately compute the optimal control for the discrete-time version as we estimate the expectation in (28). This approximation involves utilizing a collection of M uncontrolled sample trajectories $\{\boldsymbol{x}^{(i)}(k)\}_{k=1}^{K}$ for $i = 1, \ldots, M$ generated via the Monte-Carlo method as follows:

$$\boldsymbol{u}(\boldsymbol{x}(k), k) = \boldsymbol{R}^{-1}\boldsymbol{G}_c^{\top}(\boldsymbol{G}_c\boldsymbol{R}^{-1}\boldsymbol{G}_c^{\top})^{-1}(\boldsymbol{x}(k), k)\frac{\sum_{i=1}^{M}\exp\left(\mathcal{J}_{\mathbf{0}}^{(i)}/\theta'\right)\boldsymbol{\Sigma}_c\frac{\boldsymbol{\varepsilon}}{\sqrt{\Delta t}}}{\sum_{j=1}^{M}\exp\left(\mathcal{J}_{\mathbf{0}}^{(j)}/\theta'\right)},$$

where $\mathcal{J}_{\mathbf{0}}^{(i)} = \psi(\boldsymbol{x}^{(i)}(K), K) + \sum_{s=k}^{K-1} q(\boldsymbol{x}^{(i)}(s), s)\Delta t$ is a cost associated with the i-th trajectory $\{\boldsymbol{x}^{(i)}\}$.

The proposed Distributionally Robust Path Integral (DRPI) Algorithm 1 makes use of the risk-sensitive path integral control to compute the optimal value of the subproblem $g(\theta)$ (21) for any $\theta > 0$. For the master problem (20), we can employ various line search methods in line 11 since it is an univariate optimization. Note that, in line 11, we reuse the costs $\mathcal{J}_{\mathbf{0}}^{(i)}$ for $i = 1, 2, \ldots, M$, to optimize over θ without sampling new trajectories. Hence, the scheme efficiently solves the master problem.

Remark 2. *For the special case where the stochastic control problem is convex, $g(\theta)$ becomes convex since it is a partial minimization of the convex function over $\boldsymbol{u}$. In this case, we can solve the master problem in polynomial time by a binary search over θ while $g(\theta)$ is solved by the path integral method for each θ. If we further simplify the problem to the case where the dynamic is linear and the cost function is quadratic in both $\boldsymbol{x}$ and $\boldsymbol{u}$, the subproblem $g(\theta)$ becomes the well-known linear exponential-of-quadratic Gaussian (LEQG) control problem, which can be analytically solved by the modified backward Riccati equation [28].*

VI. Numerical Experiments

In this section, we demonstrate the effectiveness of our DRPI scheme. To reflect limited data availability, we start each simulation run with an initial estimate $\widehat{\boldsymbol{\mu}}$ using only a single data point $\Delta\boldsymbol{\xi}^{(1)}$. Here we set the maximum time step to K. Then, for each time step $k = 1, 2, \ldots, K$, we refine $\widehat{\boldsymbol{\mu}}$ as we have access to a newly observed empirical disturbance term $\Delta\boldsymbol{\xi}^{(k)}$, as follows:

$$\widehat{\boldsymbol{\mu}} \leftarrow \widehat{\boldsymbol{\mu}} + \frac{\Delta\boldsymbol{\xi}^{(k)}/\Delta t - \widehat{\boldsymbol{\mu}}}{k}, k = 1, 2, \ldots, K.$$

As our estimate $\widehat{\boldsymbol{\mu}}$ improves with the increasing number of available data, we adjust $\gamma \leftarrow 1/k$ for each time step $k = 1, 2, \ldots, K$, based on the finite sample guarantees in Section IV.

We compare DRPI with risk-neutral and risk-averse path integral control. Risk-neutral path integral control (RN-PIC) amounts to DRPI with γ fixed at 0: as mentioned in Section III, if $\gamma = 0$, the ambiguity set (6) only includes the empirical probability law $\mathbb{Q}$, letting the inner maximization disappear in (5). Hence, RN-PIC ignores errors in the simulator model.

On the other hand, the drawback of risk-averse path integral control (RA-PIC) is the absence of a finite sample guarantee for a specified risk parameter θ, thereby lacking a clear method for determining θ in advance. To reflect this limitation, for RA-PIC, we parameterize $\theta = a \cdot 10^b$ where $a \in \{1, 2, \ldots, 9\}$ and $b \in \{-2, -1, 0, 1, 2\}$, and conduct experiments for each candidate value of θ to find the best-performing risk parameter.

A. Double Integrator Model

Consider the following double integrator model for a particle robot in a 2D plane as follows:

$$\boldsymbol{x}(k+1) = \begin{bmatrix} p_x(k+1) \\ p_y(k+1) \\ h(k+1) \\ v(k+1) \end{bmatrix} = \boldsymbol{x}(k) + \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{x}(k)\Delta t + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_h(k)\Delta t + \Delta\xi_h(k) \\ u_v(k)\Delta t + \Delta\xi_v(k) \end{bmatrix}. \tag{29}$$

Here, the states $p_x(k)$ and $p_y(k)$ represent positions in the horizontal and vertical directions, while $h(k)$ and $v(k)$ represent linear velocities in the corresponding directions. The control inputs $u_h(k)$ and $u_v(k)$ denote the acceleration in the horizontal and the vertical direction, respectively. The true disturbance $\boldsymbol{\xi}(k) = [\xi_h(k),\ \xi_v(k)]^\top$ is a Brownian disturbance with unknown drift. Given initial state $\boldsymbol{x}(0) = [-3.5,\ 2.5,\ 0.0,\ 0.0]^\top$, the objective is to design an optimal control policy $\boldsymbol{u}^*(\boldsymbol{x}(k), k) = [u_h^*(\boldsymbol{x}(k), k),\ u_v^*(\boldsymbol{x}(k), k)]^\top$ for the robot to arrive at the target position $[p_x^*(k),\ p_y^*(k)] = [0.0,\ 0.0]$ as soon as possible, while avoiding any collisions with both the inner square obstacle and the outer square boundary as shown in Figure 1(a).

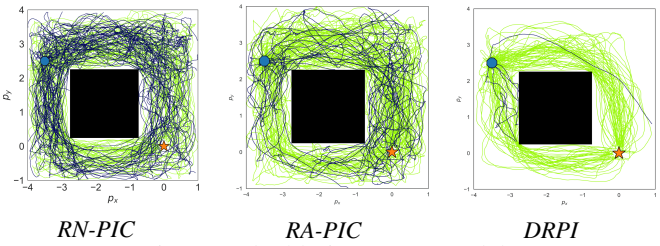
B. Unicycle Model

The second model is an unicycle model of a robot that can move forward and change its orientation in the 2D plane, but cannot move directly sideways as follows:

$$\boldsymbol{x}(k+1) = \begin{bmatrix} p_x(k+1) \\ p_y(k+1) \\ p_\theta(k+1) \end{bmatrix} = \boldsymbol{x}(k) + \begin{bmatrix} \cos(\theta) & 0 \\ \sin(\theta) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_v(k)\Delta t + \Delta\xi_v(k) \\ u_\omega(k)\Delta t + \Delta\xi_\omega(k) \end{bmatrix}. \tag{30}$$

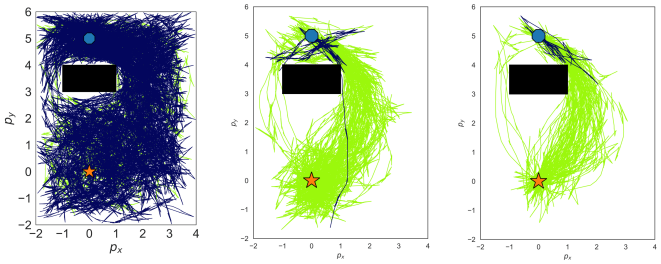
Here, the states $p_x(k)$ and $p_y(k)$ represent positions in the horizontal and vertical direction, respectively, and $p_\theta(k)$ represent the orientation of the unicycle. Control inputs $u_v(k)$ and $u_\omega(k)$ represent forward velocity and the rate of change of the orientation, respectively.

Similar to the first experiment, the true Brownian disturbance $\boldsymbol{\xi}(k) = [\xi_v(k),\ \xi_\omega(k)]^\top$ is unknown and the objective is to design an optimal control policy $\boldsymbol{u}^*(\boldsymbol{x}(k), k) = [u_v^*(\boldsymbol{x}(k), k),\ u_\omega^*(\boldsymbol{x}(k), k)]^\top$ for the robot to navigate to the target position $[p_x(k),\ p_y(k)] = [0.0,\ 0.0]$, without any collision as shown in Figure 1(b).



RN-PIC *RA-PIC* *DRPI*

Fig. 1(a) double integrator model



RN-PIC *RA-PIC* *DRPI*

Fig. 1(b) unicycle model

Fig. 1: The simulation results depict 100 trajectories for RN-PIC (left), RA-PIC (center), and DRPI (right). The simulations are conducted for two models: (a) the double integrator model and (b) the unicycle model. The initial positions are represented by blue octagons, while the target positions are denoted by orange stars. Trajectories are color-coded: deep blue trajectories collide with either inner or outer boundaries, whereas green trajectories represent successful simulations.

C. Experimental Design

For both the double integrator and unicycle models, we consider quadratic control-dependent cost with the weight matrix $\boldsymbol{R} = 10^{-3}\mathbf{I} \in \mathbb{R}^{2\times 2}$ and a nonlinear state-dependent cost function $q(\boldsymbol{x}(k))$ as follows:

$$q\left(\boldsymbol{x}(k)\right) = c_1 \cdot \|\boldsymbol{x}(k) - \boldsymbol{x}^*\|_2 + c_2 \cdot q_o\left(\boldsymbol{x}(k)\right) + c_3 \cdot q_b\left(\boldsymbol{x}(k)\right). \tag{31}$$

Here, the obstacle cost $q_o(\cdot)$ and the boundary cost $q_b(\cdot)$ are indicator functions that take the value 1 if the robot hits the obstacle and the boundary, respectively. The coefficient parameters in equation (31) are set to $c_1 = 10^{-2}$, $c_2 = c_3 = 10^2$. In the event of a collision with the inner or outer squares, the current simulation is immediately terminated and recorded as a failure.

We conducted 100 simulations for each model and recorded the trajectory for each simulation and the arrival times for the successful ones. The results summarized in Table I and Figure 1 show that the DRPI scheme offers significant advantages over both RN-PIC and RA-PIC, which include a higher success rate, faster arrive time on average, and more consistent and predictable performance due to

lower variability. In addition, it is worth noting that we only present the RA-PIC instance with the best performing θ while the majority of other instances of RA-PIC show significantly worse performances than DRPI. This demonstrates optimizing the risk parameter in real-time is indeed advantageous.

TABLE I: Performance of Different Schemes

Model	Scheme	Success Rate (%)	Arrive Time (s)		
			Mean	Std. Dev.	95 pct.
Double Integrator	RN-PIC	66	21.30	12.63	44.42
	RA-PIC	85	18.47	10.32	37.30
	DRPI	94	7.65	1.91	10.72
Unicycle	RN-PIC	19	25.89	11.13	44.52
	RA-PIC	88	7.41	4.28	14.41
	DRPI	89	3.40	0.99	4.96

VII. Conclusion

We introduced a novel data-driven approach to robustify policies for a broad class of stochastic control problems in the absence of the true diffusion process. Notably, we established an interesting connection between the DR control and the risk-sensitive control. Our theoretical results offered valuable insight into selecting the robust parameter, making simulator-based controllers more practical when dealing with unknown system dynamics. Furthermore, our proposed DRPI algorithm showcased outstanding performance. Future research avenues may include exploring extensions of our framework to more complex systems in various application domains.

References

[1] G. Bayraksan and D. K. Love, "Data-driven stochastic programming using phi-divergences," in *The Operations Research Revolution*. INFORMS, 2015, pp. 1–19.

[2] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[3] S. Mehrotra and H. Zhang, "Models and algorithms for distributionally robust least squares problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 123–141, 2014.

[4] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[5] P. Whittle, *Risk-Sensitive Optimal Control*, ser. Wiley Interscience Series in Systems and Optimization. Wiley, 1990. [Online]. Available: https://books.google.com/books?id=NvAZAQAAIAAJ

[6] I. R. Petersen, M. R. James, and P. Dupuis, "Minimax optimal control of stochastic uncertain systems with relative entropy constraints," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 398–412, 2000.

[7] H. Nishimura, N. Mehr, A. Gaidon, and M. Schwager, "RAT iLQR: A risk auto-tuning controller to optimally account for stochastic model mismatch," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 763–770, 2021.

[8] H. J. Kappen, "Path integrals and symmetry breaking for optimal control theory," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, p. P11011, 2005.

[9] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *The Journal of Machine Learning Research*, vol. 11, pp. 3137–3181, 2010.

[10] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Aggressive driving with model predictive path integral control," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1433–1440.

[11] G. Williams, N. Wagener, B. Goldfain, P. Drews, J. M. Rehg, B. Boots, and E. A. Theodorou, "Information theoretic MPC for model-based reinforcement learning," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 1714–1721.

[12] X. Cai, M. Everett, L. Sharma, P. R. Osteen, and J. P. How, "Probabilistic traversability model for risk-aware motion planning in off-road environments," *arXiv preprint arXiv:2210.00153*, 2022.

[13] G. Williams, A. Aldrich, and E. A. Theodorou, "Model predictive path integral control: From theory to parallel computation," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 2, pp. 344–357, 2017.

[14] A. Patil, A. Duarte, A. Smith, F. Bisetti, and T. Tanaka, "Chance-constrained stochastic optimal control via path integral and finite difference methods," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 3598–3604.

[15] A. Patil, M. O. Karabag, T. Tanaka, and U. Topcu, "Simulator-driven deceptive control via path integral approach," *arXiv preprint arXiv:2308.14092*, 2023.

[16] A. Friedman, "Stochastic differential equations and applications," in *Stochastic Differential Equations*. Springer, 1975, pp. 75–148.

[17] J. Yin, Z. Zhang, E. Theodorou, and P. Tsiotras, "Trajectory distribution control for model predictive path integral control using covariance steering," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 1478–1484.

[18] J. Yin, Z. Zhang, and P. Tsiotras, "Risk-aware model predictive path integral control using conditional value-at-risk," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 7937–7943.

[19] J. Yin, C. Dawson, C. Fan, and P. Tsiotras, "Shield model predictive path integral: A computationally efficient robust MPC approach using control barrier functions," *arXiv preprint arXiv:2302.11719*, 2023.

[20] G. Williams, B. Goldfain, P. Drews, K. Saigol, J. M. Rehg, and E. A. Theodorou, "Robust sampling based model predictive control with sparse objective information." in *Robotics: Science and Systems*, vol. 14, 2018, p. 2018.

[21] M. S. Gandhi, B. Vlahov, J. Gibson, G. Williams, and E. A. Theodorou, "Robust model predictive path integral control: Analysis and performance guarantees," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1423–1430, 2021.

[22] B. P. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari, "Distributionally robust control of constrained stochastic systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 430–442, 2015.

[23] K. Kim and I. Yang, "Minimax control of ambiguous linear stochastic systems using the wasserstein metric," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 1777–1784.

[24] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, 1997.

[25] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to The Theory of Large Deviations*. John Wiley & Sons, 2011.

[26] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media, 2013.

[27] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.

[28] D. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 124–131, 1973.

[29] B. v. d. Broek, W. Wiegerinck, and H. Kappen, "Risk sensitive path integral control," *arXiv preprint arXiv:1203.3523*, 2012.

[30] E. A. Theodorou, "Nonlinear stochastic control and information theoretic dualities: Connections, interdependencies and thermodynamic interpretations," *Entropy*, vol. 17, no. 5, pp. 3352–3375, 2015.

[31] P. E. Kloeden and E. Platen, *Stochastic Differential Equations*. Springer, 1992.