

Biased-MPPI

Informing Sampling-Based Model Predictive Control by Fusing Ancillary Controllers

Trevisan, Elia; Alonso-Mora, Javier

DOI

[10.1109/LRA.2024.3397083](https://doi.org/10.1109/LRA.2024.3397083)

Publication date

2024

Document Version

Final published version

Published in

IEEE Robotics and Automation Letters

Citation (APA)

Trevisan, E., & Alonso-Mora, J. (2024). Biased-MPPI: Informing Sampling-Based Model Predictive Control by Fusing Ancillary Controllers. *IEEE Robotics and Automation Letters*, 9(6), 5871-5878.
<https://doi.org/10.1109/LRA.2024.3397083>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Biased-MPPI: Informing Sampling-Based Model Predictive Control by Fusing Ancillary Controllers

Elia Trevisan , *Graduate Student Member, IEEE*, and Javier Alonso-Mora , *Senior Member, IEEE*

Abstract—Motion planning for autonomous robots in dynamic environments poses numerous challenges due to uncertainties in the robot’s dynamics and interaction with other agents. Sampling-based MPC approaches, such as Model Predictive Path Integral (MPPI) control, have shown promise in addressing these complex motion planning problems. However, the performance of MPPI relies heavily on the choice of sampling distribution. Existing literature often uses the previously computed input sequence as the mean of a Gaussian distribution for sampling, leading to potential failures and local minima. We propose a novel derivation of MPPI that allows for arbitrary sampling distributions to enhance efficiency, robustness, and convergence while alleviating the problem of local minima. We present an efficient importance sampling scheme that combines classical and learning-based ancillary controllers simultaneously, resulting in more informative sampling and control fusion. Several simulated and real-world demonstrate the validity of our approach.

Index Terms—Motion and path planning, optimization and optimal control, collision avoidance, sampling-based MPC, MPPI.

I. INTRODUCTION

NAVIGATING autonomous robots through dense and dynamic environments poses a formidable challenge due to significant uncertainties, including the robot’s state, model, environmental conditions, and interactions with other agents. Achieving desired behaviors under such conditions often necessitates using intricate cost functions and constraints, resulting in complex, nonlinear, non-convex, and occasionally discontinuous problem formulations. The dynamic nature of the environment introduces potential unexpected changes, demanding rapid adaptability in the robot’s actions.

To address these challenges, one approach is to cast the problem in a stochastic optimal control setting, where they can be mathematically represented as stochastic Hamilton-Jacobi-Bellman (HJB) equations. However, solving these equations

numerically can be challenging due to the curse of dimensionality. Pioneering work demonstrated that the stochastic HJB equations can be linearized for control-affine systems, and their solution can be approximated through sampling using the path integral formulation [1]. Implemented in a receding horizon fashion, Model Predictive Path Integral (MPPI) control [2], [3], and its Information-Theoretic counterpart [4], [5] have been initially used for racing a small-scale rally car. MPPI has also been successfully applied to several other planning problems, such as for autonomous vehicles with dynamic obstacles [6], solving games [7], flying drones in partially observable environments [8], performing complex maneuvers [9] and used in combination with adaptive control schemes [10]. It has also been adapted to multi-agent systems for formation flying [11], cooperative behavior [12], and simultaneous prediction and planning [13]. Furthermore, MPPI has shown promise in manipulating objects with robot arms [14] including model uncertainties [15], in pushing tasks [16], [17] and planning motion for four-legged walking robots [18]. MPPI is a model-based approach that requires a model to forward simulate trajectories given sampled inputs. Recent work has utilized physics engines to simulate samples [19], [20], eliminating the need for explicitly defining the dynamics of agents and the environment, thus providing a significant advantage in contact-rich manipulation tasks.

One of the critical challenges in applying MPPI to dynamic environments is ensuring the algorithm’s performance and reliability. The success of MPPI heavily relies on the choice of sampling distribution, which is crucial, especially in real-time scenarios. Most existing literature uses the previously computed input sequence as the mean of a Gaussian distribution for sampling [2]. However, using the previous input sequence may trap the algorithm in local minima and can lead to catastrophic failures in the presence of unexpected disturbances or changes in the environment [21] (Fig. 1). This letter explores the application of MPPI in dynamic environments, emphasizing the need to improve its performance and reliability in the face of unexpected disturbances and rapidly changing conditions.

A. Previous Work

Several works tried to make the method more efficient or more robust. Early work [22] proposed using Expectation Propagation instead of Monte Carlo sampling, demonstrating better efficiency in scenarios with hard constraints. Other works instead accelerate the convergence of MPPI by leveraging gradient

Manuscript received 17 January 2024; accepted 27 April 2024. Date of publication 6 May 2024; date of current version 13 May 2024. This letter was recommended for publication by Associate Editor S. Zhao and Editor A. Bera upon evaluation of the reviewers’ comments. This work was supported in part by the Project Sustainable Transportation and Logistics over Water: Electrification, Automation and Optimization (TRiLOGy) of the Netherlands Organization for Scientific Research (NWO), domain Science (ENW), and in part by the Amsterdam Institute for Advanced Metropolitan Solutions (AMS) in the Netherlands. (Corresponding author: Elia Trevisan.)

The authors are with the Cognitive Robotics Department, TU Delft, 2628 CD Delft, The Netherlands (e-mail: e.trevisan@tudelft.nl; j.alonsomora@tudelft.nl).

Website: autonomousrobots.nl/paper_websites/biased-mppi

This letter has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2024.3397083>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2024.3397083

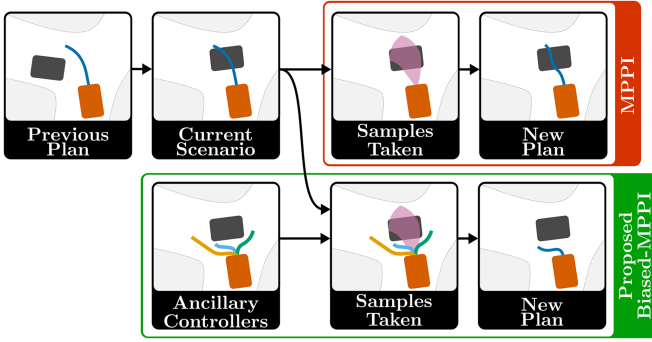


Fig. 1. *Top*: Usually, MPPI only takes samples around a previous plan. Here, the environment changes unexpectedly, and all the sampled trajectories are in collision, which leads to computing a new plan that also collides. *Bottom*: Our biased-MPPI adds ancillary controllers to the sampling distribution, quickly converging to a collision avoidance maneuver.

descent updates [23]. Another option to be more reactive to environmental changes is to iteratively converge to a solution through adaptive importance sampling [24]. This, however, requires multiple iterations between each planning time step, diminishing the parallelizability of MPPI. Many other works propose improving the algorithm's convergence by somehow changing its sampling distribution. This can be done by substituting the Gaussian used for sampling with a different hand-crafted distribution [25] or by directly learning a distribution from data [26], [27]. Given that MPPI allows for tuning the variance of the sampling distribution [3], some works sought to improve the efficiency of the scheme by adapting the covariance online via covariance steering [28], [29]. Other ways to improve efficiency can be to fit splines to the sampled inputs [14] or to constrain the distribution to sample areas that are known to contain low-cost trajectories [18]. Previous works have also experimented with ancillary controllers. In [30], authors propose to sample inputs around a path previously computed by RRT. Other works instead robustify MPPI by switching to an iLQG controller [21] or by integrating one into the system's model [31]. Previous work also compares an MPPI that samples around a previously computed input, an input sequence computed by a sequential linear-quadratic MPC, and a learned sampling policy [18]. In general, however, the original derivations of MPPI [5] only allow samples to be drawn from a uni-modal Gaussian distribution, usually centered around the previous control sequence, which can hamper performance and reduce reactivity to unexpected changes in the environment.

B. Contributions

We propose a Biased-MPPI, for which we provide mathematical derivations that allow for arbitrary changes to the sampling distribution. We discuss the impact of introducing biases in the sampling distribution on the overall method. We experiment with an importance sampler that utilizes multiple classical and learning-based ancillary controllers simultaneously to take more informative samples, which can be seen as a control fusion scheme. Through simulated and real-world experiments, we

demonstrate the impact of taking suggestions from several underlying controllers on robustness to model uncertainties and local minima, reactivity to unexpected events, and sampling efficiency.

II. PRELIMINARIES

In this section, we provide a concise introduction to the key concepts of MPPI within the Information-Theoretic framework. For more details, we direct the reader to prior research [5]. We begin by defining a function:

$$\mathcal{F}(S, \mathbb{P}, x_0, \lambda) = -\lambda \log \left(\mathbb{E}_{\mathbb{P}} \left[\exp \left(-\frac{1}{\lambda} S(V) \right) \right] \right) \quad (1)$$

which we will denote as the free energy of the system. Here, V represents a sequence of inputs, \mathbb{P} is a base measure, λ is a tuning parameter, $S(V)$ is a cost, and x_0 represents the system's initial state. It can be shown that:

$$\mathcal{F}(S, \mathbb{P}, x_0, \lambda) \leq \mathbb{E}_{\mathbb{Q}}[S(V)] + \lambda \text{KL}(\mathbb{Q}||\mathbb{P}). \quad (2)$$

Here, \mathbb{Q} represents a probability measure that characterizes the controlled input distribution, and $\text{KL}(\mathbb{Q}||\mathbb{P})$ denotes the KL-Divergence between the base measure and the controlled measure. Equation (2) signifies that the free energy serves as a lower bound for the expected cost under the controlled distribution plus a control cost represented by the KL-Divergence. Hence, determining a control distribution that achieves this lower bound minimizes the expected cost and control cost. We can define a control distribution \mathbb{Q}^* through its Radon-Nikodym derivative to the base measure:

$$\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \frac{\exp(-\frac{1}{\lambda} S(V))}{\mathbb{E}_{\mathbb{P}}[\exp(-\frac{1}{\lambda} S(V))]} \quad (3)$$

Substituting \mathbb{Q} with \mathbb{Q}^* in (2), we can prove that \mathbb{Q}^* is an optimal control distribution in the sense that it achieves the lower bound. The idea is now to align our control distribution \mathbb{Q} with the optimal distribution \mathbb{Q}^* through KL minimization, which results in the optimal input sequence U^* :

$$U^* = \arg \min_U \text{KL}(\mathbb{Q}^*||\mathbb{Q}). \quad (4)$$

Now, considering a discrete-time system:

$$x_{t+1} = F(x_t, v_t), \quad v_t \sim \mathcal{N}(u_t, \Sigma). \quad (5)$$

Here, $x_t \in \mathbb{R}^n$ represents the state vector at time step t , $F(\cdot)$ is the state transition model, $v_t \in \mathbb{R}^m$ denotes the noisy input, $u_t \in \mathbb{R}^m$ is the commanded input, and Σ corresponds to the natural input variance of the system. If \mathbb{P} and \mathbb{Q} are the uncontrolled and controlled measures, respectively, we can define them through their probability density functions:

$$p(V) = \prod_{t=0}^{T-1} \frac{1}{((2\pi)^m |\Sigma|)^{1/2}} \exp \left(-\frac{1}{2} v_t^T \Sigma^{-1} v_t \right)$$

$$q(V|U) = \prod_{t=0}^{T-1} \frac{1}{((2\pi)^m |\Sigma|)^{1/2}} \times \exp \left(-\frac{1}{2} (v_t - u_t)^T \Sigma^{-1} (v_t - u_t) \right).$$

It can be proven from (4) that the optimal control input at time t is the mean input under the optimal distribution:

$$u_t^* = \int_{\Omega_V} q^*(V) v_t dV. \quad (6)$$

We can estimate such mean sampling from our controlled distribution via importance sampling:

$$\begin{aligned} u_t^* &= \int \frac{q^*(V)}{q(V|U)} q(V|U) v_t dV \\ &= \mathbb{E}_{\mathbb{Q}}[\omega(V) v_t], \end{aligned} \quad (7)$$

with the importance sampling weight $\omega(V)$ being:

$$\begin{aligned} \omega(V) &= \left(\frac{q^*(V)}{q(V|U)} \right) = \left(\frac{q^*(V)}{p(V)} \right) \left(\frac{p(V)}{q(V|U)} \right) \\ &= \frac{1}{\eta} \exp \left(-\frac{1}{\lambda} \left(S(V) + \frac{\lambda}{2} \sum_{t=0}^{T-1} u_t^T \Sigma^{-1} u_t + 2u_t^T \Sigma^{-1} \epsilon_t \right) \right). \end{aligned} \quad (8)$$

We can, therefore, sample K noisy input sequences:

$$\begin{aligned} V^k &= [v_0^k, v_1^k, \dots, v_t^k, \dots, v_{T_H}^k] \\ v_t^k &\sim \mathcal{N}(u_t, \Sigma) \end{aligned} \quad (9)$$

where t is a time step and T_H is the planning horizon. A practical choice often made in MPPI is to take u_t as a time-shifted version of the previously computed approximation of the optimal control sequence. We roll out the sampled V^k into state trajectories using the system's model $F(\cdot)$, evaluate their cost $S(V)$, compute the weights $\omega(V)$, get a new estimate of the optimal input sequence U^* via (7) and iterate. In (8), the control cost is multiplied and divided by λ . Not having control over the magnitude of the terms at the exponential can cause numerical issues. A change of base measure \mathbb{P} can solve the problem [5]. One might also need a higher variance Σ_s for sampling compared to the natural variance of the system Σ [32]. This again introduces terms at the exponential independent from λ . Moreover, introducing an arbitrary, potentially multi-modal sampling distribution \mathbb{Q}_s is difficult. All these issues stem from the ratio $p(v)/q(V|U)$ in (8). Our approach addresses this by showing that accepting a bias in the solution can eliminate the ratio and allow for arbitrary sampling distributions.

III. PROPOSED APPROACH

A. Biased-MPPI

Let us first redefine the cost function as:

$$\tilde{S}(V) = S(V) + \lambda \log \left(\frac{p(V)}{q_s(V)} \right). \quad (10)$$

We then define the free-energy with this new cost:

$$\begin{aligned} \mathcal{F}(\tilde{S}, \mathbb{P}, x_0, \lambda) &= -\lambda \log \left(\mathbb{E}_{\mathbb{P}} \left[\exp \left(-\frac{1}{\lambda} \tilde{S}(V) \right) \right] \right) \\ &= -\lambda \log \left(\mathbb{E}_{\mathbb{Q}} \left[\exp \left(-\frac{1}{\lambda} \tilde{S}(V) \right) \frac{p(V)}{q(V)} \right] \right) \\ &\leq -\lambda \mathbb{E}_{\mathbb{Q}} \left[\log \left(\exp \left(-\frac{1}{\lambda} \tilde{S}(V) \right) \frac{p(V)}{q(V)} \right) \right] = * \end{aligned} \quad (11)$$

where, as in [5], we applied Jensen's inequality. We can simplify the right-hand side as follows:

$$\begin{aligned} * &= -\lambda \mathbb{E}_{\mathbb{Q}} \left[-\frac{1}{\lambda} \tilde{S}(V) + \log \left(\frac{p(V)}{q(V)} \right) \right] \\ &= -\lambda \mathbb{E}_{\mathbb{Q}} \left[-\frac{1}{\lambda} S(V) - \log \left(\frac{p(V)}{q_s(V)} \right) + \log \left(\frac{p(V)}{q(V)} \right) \right] \\ &= \mathbb{E}_{\mathbb{Q}} [S(V)] + \lambda \mathbb{E}_{\mathbb{Q}} \left[\log \left(\frac{p(V)}{q_s(V)} \frac{q(V)}{p(V)} \right) \right] \\ &= \mathbb{E}_{\mathbb{Q}} [S(V)] + \lambda \text{KL}(\mathbb{Q} || \mathbb{Q}_s). \end{aligned}$$

The free energy inequality is then:

$$\mathcal{F}(S, \mathbb{P}, x_0, \lambda) \leq \mathbb{E}_{\mathbb{Q}} [S(V)] + \lambda \text{KL}(\mathbb{Q} || \mathbb{Q}_s). \quad (12)$$

Thus, while we start with $\tilde{S}(V)$, the free energy serves as a lower bound for the expected original cost $S(V)$ under the controlled distribution plus lambda times the KL-Divergence between the controlled and sampling distribution. An optimal control distribution achieving the lower bound would minimize the original cost $S(V)$ while pushing the controlled distribution to align with the sampling distribution, effectively introducing a bias toward the sampling distribution. We define a controlled distribution \mathbb{Q}^* as:

$$\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \frac{\exp(-\frac{1}{\lambda} \tilde{S}(V))}{\mathbb{E}_{\mathbb{P}}[\exp(-\frac{1}{\lambda} \tilde{S}(V))]}.$$

Under \mathbb{Q}^* , the KL-Divergence becomes:

$$\begin{aligned} \text{KL}(\mathbb{Q}^* || \mathbb{Q}_s) &= \mathbb{E}_{\mathbb{Q}^*} \left[\log \left(\frac{q^*(V)}{q_s(V)} \right) \right] \\ &= \mathbb{E}_{\mathbb{Q}^*} \left[\log \left(\frac{q^*(V)}{p(V)} \right) \right] + \mathbb{E}_{\mathbb{Q}^*} \left[\log \left(\frac{p(V)}{q_s(V)} \right) \right] \\ &= -\frac{1}{\lambda} \mathbb{E}_{\mathbb{Q}^*} [\tilde{S}(V)] - \log \left(\mathbb{E}_{\mathbb{P}} \left[\exp \left(-\frac{1}{\lambda} \tilde{S}(V) \right) \right] \right) \\ &\quad + \mathbb{E}_{\mathbb{Q}^*} \left[\log \left(\frac{p(V)}{q_s(V)} \right) \right] \end{aligned}$$

Substituting into (12) and simplifying leads to:

$$\begin{aligned} \mathcal{F}(\tilde{S}, \mathbb{P}, x_0, \lambda) &\leq -\lambda \log \left(\mathbb{E}_{\mathbb{P}} \left[\exp \left(-\frac{1}{\lambda} \tilde{S}(V) \right) \right] \right) \\ &= \mathcal{F}(\tilde{S}, \mathbb{P}, x_0, \lambda). \end{aligned}$$

This proves that \mathbb{Q}^* is the optimal distribution in that it achieves the lower bound in (12). Following the steps in [5], we can align our controlled distribution \mathbb{Q} to \mathbb{Q}^* as in (6), except we can now

use our sampling distribution:

$$u_t^* = \mathbb{E}_{\mathbb{Q}_s}[\omega(V)v_t], \quad (13)$$

with importance sampling weights:

$$\begin{aligned} \omega(V) &= \frac{1}{\eta} \exp\left(-\frac{1}{\lambda} \tilde{S}(V)\right) \left(\frac{p(V)}{q_s(V)}\right) \\ &= \frac{1}{\eta} \exp\left(-\frac{1}{\lambda} \left(S(V) + \lambda \log\left(\frac{p(V)}{q_s(V)}\right)\right)\right) \\ &\quad \times \exp\left(\log\left(\frac{p(V)}{q_s(V)}\right)\right) \\ &= \frac{1}{\eta} \exp\left(-\frac{1}{\lambda} S(V)\right). \end{aligned} \quad (14)$$

Note that our change of cost (10) resulted in the optimal control being biased towards the sampling distribution, as shown in (12). However, this simplified the weights $\omega(V)$ and allowed us to design arbitrary sampling distributions \mathbb{Q}_s . In [5], $S(V)$ was defined as the state-dependent cost. However, this restriction was made to relate the approach to path integral control [1]. Such relation was only shown exactly when \mathbb{P} is the distribution induced by an uncontrolled continuous-time control-affine system. This restriction is not required in the Information-Theoretic framework, which allows for a larger class of systems, and one can add input costs in $S(V)$.

B. Sampling From Ancillary Controllers

There are several ways one could design an arbitrary sampling distribution. This letter focuses on taking most samples around a previously computed input distribution and some samples from hand-crafted policies.

In particular, we design a set of task-specific ancillary controllers, these being, e.g., open-loop motion primitives, reference tracking feedback controllers, or learning-based strategies to propose J input sequences $U^j = [u_0^j, u_1^j, \dots, u_t^j, \dots, u_{T_H}^j]$. These ancillary controllers are described for each experiment in Sections IV and V. We then choose the K sampled input sequences V_s^k as,

$$V_s^k = \begin{cases} U^j, & \text{with } j = k \quad \text{if } k \leq J \\ V^k, & \text{as in (9)} \quad \text{if } k > J, \end{cases} \quad (15)$$

meaning that, at each time step, we take one sample from each of the J ancillary controllers, and the remaining $K - J$ samples are taken according to the classical MPPI strategy.

C. Autotuning the Inverse Temperature

As in [20] and similarly to [18], we autotune the inverse temperature λ online based on the normalization factor η .

$$\lambda_{t+1} = \begin{cases} 0.9\lambda_t & \text{if } \eta > \eta_{\max} \\ 1.2\lambda_t & \text{if } \eta < \eta_{\min} \\ \lambda_t & \text{otherwise} \end{cases} \quad (16)$$

In all experiments, this can roughly keep the number of samples with a significant weight between η_{\min} and η_{\max} .

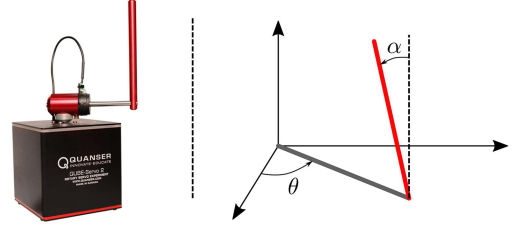


Fig. 2. Left, Quanser Qube-Servo, and right, its diagram. The arm's rotation, θ , is the actuated angle. The angle between the pendulum and the upright position, α , is not actuated.

IV. ILLUSTRATIVE EXPERIMENT

We apply our Biased-MPPI to a rotary inverted pendulum [33] (Fig. 2) in simulation to visualize its main features.

A. Swing-Up and Tracking

Starting at the bottom equilibrium with $\theta_0 = 0$ and $\alpha_0 = \pi$, the task is to swing up the pendulum to $\alpha_r = 0$ while keeping the arm close to $\theta_r = 1$. Thus, the running cost is:

$$C_p(x(t)) = 100((\theta_t - \theta_r)^2 + (\alpha_t - \alpha_r)^2) + \dot{\theta}_t^2 + 2\dot{\alpha}_t^2. \quad (17)$$

The system has dynamics $x(t+1) = F(x(t), u(t))$, where the state of the system at time-step t is denoted as $x(t) = [\theta_t, \alpha_t, \dot{\theta}_t, \dot{\alpha}_t]^T$, and u represents the system's input. The non-linear model is derived from the Lagrange equations. To design linear controllers, the model is linearized at the top equilibrium using Euler-Lagrange's method [34]. To showcase resilience against model uncertainties, the parameters of the simulation's pendulum model are multiplied by $1 + \gamma$ in each experiment, where $\gamma \sim \mathcal{N}(0, 0.05)$. The seed is consistent across methods. The system is discretized and controllers run at 50 Hz, the controller plans $T_H = 50$ steps ahead (1 s), covariance $\Sigma_s = 0.5$, $\eta_{\min} = 2$ and $\eta_{\max} = 5$.

1) *Ancillary Controllers*: We design three ancillary controllers as a baseline and to guide the sampling strategy.

a) *A linear quadratic regulator (LQR)*: designed using the `lqr` command in Matlab, stabilizes the pendulum at the top equilibrium.

b) *A linear quadratic integral (LQI)*: tracks the reference θ_r while maintaining the pendulum at the top equilibrium. It is synthesized with the `lqi` command in Matlab.

c) *A nonlinear energy-based controller (EBC)*: is designed as in [34] to swing up the pendulum to the top equilibrium by increasing the potential energy of the system [35].

2) *Switching Controller*: We introduce as baseline a switching strategy (18) that combines all ancillary controllers. It swings up the pendulum using the input from the ECB, u_{ebc} , until α is within $\alpha_{catch} = 0.2$ of the top equilibrium. The LQR controller, with u_{lqr} , then stabilizes the pendulum. Once the pendulum is close to the top equilibrium ($\alpha_{track} = 0.05$) with angular velocity below $\dot{\alpha}_{track} = 0.1$ rad/s, the LQI, with u_{lqi} , is engaged

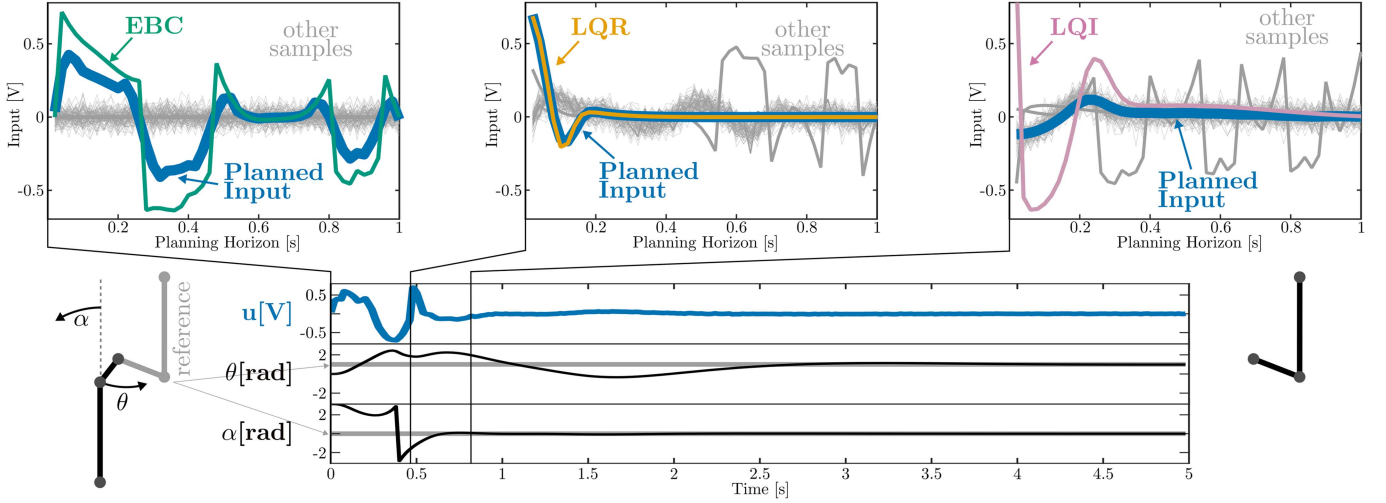


Fig. 3. Input and state evolution during a pendulum experiment with Biased-MPPI. We show the samples taken and the resulting planned input sequence over the planning horizon for three instances. While we sample all ancillary controllers in each instance, we highlight the one with the most influence on the planned input sequence.

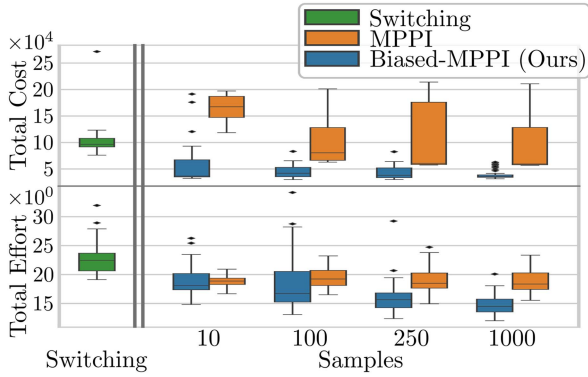


Fig. 4. Total cost and control effort over 50 pendulum swing-ups with randomized model parameters.

for reference tracking.

$$u = \begin{cases} u_{lqi}, & \text{if } (|\alpha| < \alpha_{track}) \cap (|\dot{\alpha}| < \dot{\alpha}_{track}) \\ u_{lqr}, & \text{if } (|\alpha| < \alpha_{catch}) \\ u_{ebc}, & \text{otherwise} \end{cases} \quad (18)$$

3) *Results:* Fig. 3 depicts a pendulum experiment's input and state evolution with Biased-MPPI, also showcasing the samples taken and the ancillary controllers' influence on the plan. At the beginning of the experiment, ECB rapidly swings up the pendulum, heavily influencing Biased-MPPI's planned input. Once near equilibrium, LQR provides a stabilizing sequence, closely tracked by Biased-MPPI. As stability is achieved, LQI suggests an input sequence swiftly bringing the arm towards the reference, albeit with high velocities. Hence, Biased-MPPI, while influenced by LQI, opts for a lower amplitude input sequence due to cost function (17).

Fig. 4 displays the distribution of total costs, defined as $\sum_{t=0}^{T_{end}} C_p(x(t))$ where $T_{end} = 250$ (5s) is the end of the episode, and the distribution of total efforts, defined as $\sum_{t=0}^{T_{end}} |u(t)|$, across 50 experiments. Biased-MPPI consistently outperforms both the switching strategy and the classic MPPI,

regardless of the number of samples used. Moreover, the results indicate that including ancillary controllers in the proposed Biased-MPPI vastly improves the sampling efficiency, requiring fewer samples for better performance and enhancing the algorithm's robustness to model uncertainties.

V. SIMULATED MOTION PLANNING EXPERIMENTS

Interaction-Aware (IA) MPPI [13] is a decentralized communication-free motion planning method that predicts short-term goals of other agents with a constant velocity model and, under homogeneity and rationality assumptions, each agent simultaneously plans and predicts motions for all agents. In its cost function, IA-MPPI encourages adherence to navigation rules, such as giving the right-of-way to agents from the right and preferring the right-hand side during head-on encounters. We will investigate the effects of biasing its sampling scheme with ancillary controllers. The agents are vessels modeled using Robot's model [36]. Controllers run at 10 Hz, plan $T_H = 100$ steps ahead (10 s), with $\Sigma_s = \text{diag}(6, 6, 0.12, 0.12)$, $\eta_{\min} = 5$ and $\eta_{\max} = 10$.

A. Solving an Intersection

An issue that can arise with classical MPPI formulation, which only takes samples around what was previously considered to be optimal, is the difficulty, once in one, of jumping out of local minima. This is particularly evident in IA-MPPI, especially in a crossing scenario. In this experiment, depicted in Fig. 5, two identical vessels start with zero velocity and have to cross each other's paths. In their cost function, described in previous work [13], the decentralized and communication-free IA-MPPI is encouraged to get each of the vessels across the intersection while being penalized for not yielding to the agent coming from the right-hand side.

1) *Ancillary Controllers:* To help switch out of local minima and improve sampling efficiency, four ancillary controllers are sampled using the proposed Biased-MPPI.

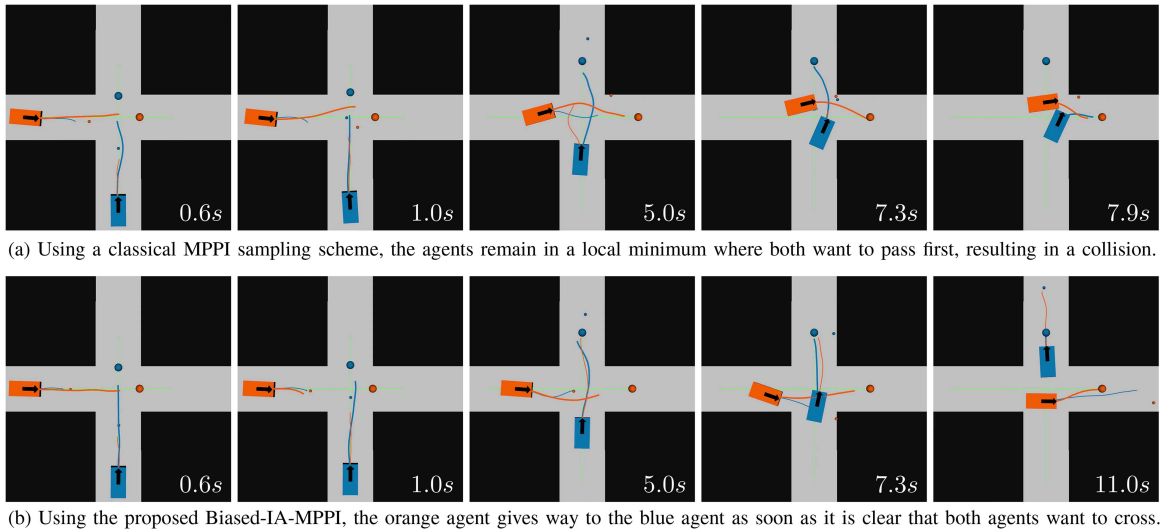


Fig. 5. Two vessels cross each other's path while penalized when not giving the right-of-way to agents coming from their right. The large circles are the agents' true local goals extracted from a global path. IA-MPPI is decentralized and communication-free, so the small dots are the goals vessels estimate of one another using constant velocity. The trajectories in blue are those the blue agent has planned for itself and predicted for the other, and the same goes for the orange agent.

- a) *Go-slow*: a sequence of inputs commanding a small amount of thrust to the vessel's side thrusters.
- b) *Go-fast*: commands a large thrust.
- c) *Braking*: gives a zero velocity reference.
- d) *Go-to-goal*: computes a velocity reference that takes each vessel towards its corresponding local goal at each time step of the planning horizon.

The velocity references proposed by the *Braking* and *Go-to-Goal* maneuvers are converted to input thrusts with a linear \mathcal{H}_∞ controller, which is robust to model non-linearities, designed using the `musyn` command in Matlab.

2) *Results*: With an initial velocity of zero, each agent anticipates an unobstructed intersection crossing. This expectation is based on a constant velocity prediction, as they assume the opposing agent will remain stationary. In Fig. 5(a), the classic IA-MPPI fails to switch from planning to cross first to a slower maneuver that yields since all of the samples are taken around the previous plan, leading to a collision. In Fig. 5(b), our Biased-IA-MPPI approach can swiftly switch between modes when it becomes evident that the vessel with the right-of-way will cross the intersection.

In Table I, we see that in 50 experiments, our Biased-IA-MPPI achieves zero collisions and rule violations for any number of samples, compared to the IA-MPPI based on the classical MPPI sampling scheme, which results in several. Thanks to the ancillary controllers, our Biased-IA-MPPI also travels straight to the goal, reducing the distance traveled. While our Biased-IA-MPPI has a lower variance in arrival times, it is not always faster on average. This confirms the results proved in (12), i.e. the *Braking* and *Go-Slow* maneuvers are biasing towards a slower trajectory.

B. Interaction-Aware Planning With Four Vessels

To further test Biased-IA-MPPI, we run 50 experiments with randomized initial conditions and goals, where four agents have

TABLE I
RESULTS OF 50 CROSSINGS FOR AN INCREASING NUMBER OF SAMPLES K

| K | Method | Collisions | Experiments With Rule Violations | Average Time to Arrival [s] | Average Distance Traveled [m] |
|------|----------------|------------|----------------------------------|-----------------------------|-------------------------------|
| 50 | IA-MPPI | 4 | 9 | 16.41 ± 10.10 | 21.89 ± 8.433 |
| | Biased-IA-MPPI | 0 | 0 | 17.64 ± 3.128 | 19.13 ± 2.466 |
| 200 | IA-MPPI | 10 | 4 | 12.77 ± 9.323 | 19.99 ± 10.16 |
| | Biased-IA-MPPI | 0 | 0 | 12.66 ± 1.902 | 18.07 ± 2.012 |
| 500 | IA-MPPI | 7 | 11 | 11.02 ± 2.731 | 18.70 ± 3.518 |
| | Biased-IA-MPPI | 0 | 0 | 11.43 ± 1.541 | 17.58 ± 1.880 |
| 1000 | IA-MPPI | 10 | 15 | 11.78 ± 3.823 | 19.31 ± 3.539 |
| | Biased-IA-MPPI | 0 | 0 | 11.00 ± 1.309 | 17.35 ± 1.625 |
| 2000 | IA-MPPI | 7 | 15 | 11.10 ± 4.101 | 19.72 ± 5.038 |
| | Biased-IA-MPPI | 0 | 0 | 10.68 ± 1.245 | 17.27 ± 1.716 |

Metrics are reported for successful runs.

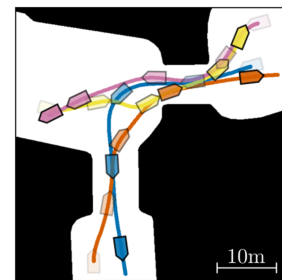


Fig. 6. Four agents navigating in the Herengracht. Video available on paper's website.

to navigate in cooperation in the Herengracht, an urban canal in Amsterdam, challenging due to its narrow sections under two bridges. The canal map and an example of successful navigation are shown in Fig. 6.

1) *Ancillary Controllers*: We use all of the ancillary controllers described in Section V-A1. Additionally, we use a learning-based trajectory prediction model adapted and trained

TABLE II
RESULTS FOR 50 RUNS OF FOUR-AGENT EXPERIMENTS IN THE HERENGCRACHT
WITH RANDOMIZED INITIAL CONDITIONS AND GOALS FOR AN INCREASING
NUMBER OF SAMPLES K

| K | Method | Successes | Deadlocks | Collisions | Experiments With Rule Violations |
|------|----------------|-----------|-----------|------------|--|
| 50 | IA-MPPI | 34 | 0 | 16 | 22 |
| | Biased-IA-MPPI | 40 | 10 | 0 | 18 |
| 200 | IA-MPPI | 43 | 1 | 6 | 34 |
| | Biased-IA-MPPI | 46 | 1 | 3 | 28 |
| 500 | IA-MPPI | 47 | 0 | 3 | 36 |
| | Biased-IA-MPPI | 49 | 0 | 1 | 35 |
| 1000 | IA-MPPI | 45 | 0 | 5 | 36 |
| | Biased-IA-MPPI | 50 | 0 | 0 | 33 |
| 2000 | IA-MPPI | 47 | 0 | 3 | 36 |
| | Biased-IA-MPPI | 49 | 0 | 1 | 34 |

for urban vessels [37]. However, we do not use this model for predictions. We track the trajectories it provides with an \mathcal{H}_∞ controller to generate input sequences, which Biased-MPPI can consider in its sampling scheme.

2) *Results:* In Table II, results from 50 experiments show that with 50 samples, our Biased-IA-MPPI is cautious, leading to 10 deadlocks, possibly biased by the *Braking* maneuver. In contrast, the conventional IA-MPPI approach, without the ancillary controller, results in 16 collisions.

As the number of samples increases, the bias from the ancillary controllers diminishes, causing Biased-IA-MPPI to behave less conservatively. Consequently, the number of deadlocks approaches zero, but a few collisions may occur. With both methods, over half of the successful experiments incur at least a rule violation. In these crowded scenes, violations are common, e.g., not stopping to yield to an agent with priority when it is still relatively far away. Still, in both collision counts and the number of experiments resulting in rule violations, our Biased-IA-MPPI consistently outperforms IA-MPPI using the traditional sampling method.

Fig. 7 displays both methods' quartiles, min, max, and outliers of successful experiments. The ancillary controllers direct the sampling distribution towards lower-cost areas of the state space, significantly reducing travel distances. Despite this, as predicted by (12), Biased-IA-MPPI also exhibits a bias towards slightly slower movement due to "Braking" and "Go-Slow" maneuvers, resulting in similar travel times as the regular IA-MPPI.

VI. REAL-WORLD MOTION PLANNING EXPERIMENT

A Clearpath Jackal robot attempts to drive to a goal as fast as possible (~ 2 m/s) while avoiding a box. Halfway through, the box is thrown in front of the robot. The position and the velocity of the box and the robot are estimated using information from a motion capture system. The velocity-controlled robot is modeled as a unicycle, and the box's position is propagated through the planning horizon using a constant velocity model. The cost function is defined as,

$$C_j(x(t)) = \|p_{t,r} - p_g\| + 100(\|p_{t,r} - p_{t,b}\| < 0.5) \quad (19)$$

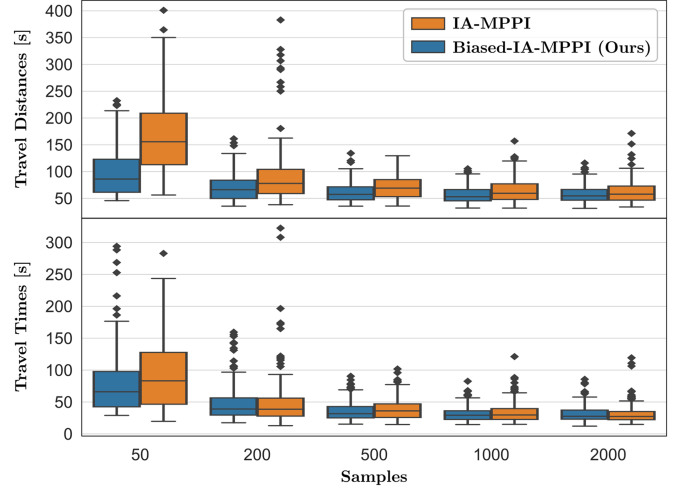


Fig. 7. Agents' traveled distance and travel time over 50 experiments in the Herengracht. Metrics are reported for experiments that were successful with both methods.

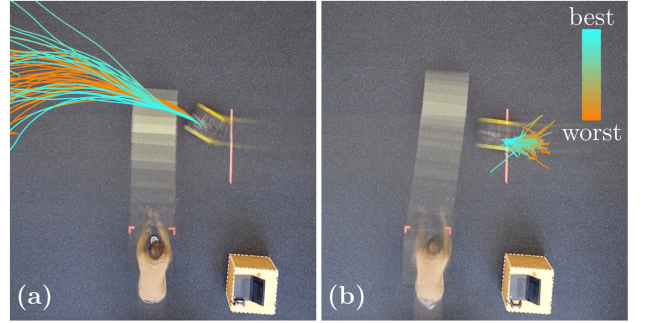


Fig. 8. Visualized are the top 50 sampled trajectories, color-graded by their cost. (a) Classic MPPI is about to crash. (b) Our Biased-MPPI avoids collision. Video and code are available on paper's website.

where $p_{t,r}$, p_g and $p_{t,b}$ are the position of the robot, the goal, and the box, respectively, at time t . Controllers run at 10 Hz, plan $T_H = 50$ steps ahead (5 s), with $K = 300$ samples, covariance $\Sigma_s = 0.5 \cdot I_{2 \times 2}$, $\eta_{\min} = 5$ and $\eta_{\max} = 10$.

1) *Ancillary Controllers:* We sample a *Braking* maneuver, i.e., a zero velocity reference throughout the horizon.

2) *Results:* Fig. 8 shows the top 50 sampled trajectories sampled by (a), MPPI, and (b), our proposed Biased-MPPI. When the box is unexpectedly thrown in front of the robot, MPPI only samples trajectories that collide with the box. Given the cost function, MPPI prefers the samples that remain in collision for the least time. On the other hand, sampling also a zero velocity reference, Biased-MPPI quickly converges to a braking maneuver, avoiding the collision altogether. MPPI resulted in six collisions over ten experiments, while Biased-MPPI resulted in none.

VII. CONCLUSION

In this letter, we have derived a sampling scheme for Model Predictive Path Integral (MPPI) control that removes computationally problematic terms and allows for the design of arbitrary sampling distributions as long as a bias in the solution

is allowed. We proposed using classical and learning-based ancillary controllers for several control and motion planning experiments to bias the sampling distribution and achieve more efficient sampling and better performances. We demonstrated how the proposed algorithm can act as a control fusion scheme, taking suggestions from an arbitrary number of controllers and improving upon them. The resulting Biased-MPPI was shown to be better performing and more robust to model uncertainties compared to classical controllers and the baseline MPPI method, achieving faster swing-ups for a rotational inverted pendulum as well as safer, closer to optimal trajectories in interaction-aware motion planning experiments in constrained multi-agent environments, all while requiring less samples. The overall gains in safety, performance, and sample efficiency come at the expense of a potentially harmful bias, as shown with the sampling of *Braking* and *Go-Slow* maneuvers, which can result in slower trajectories. In the future, our approach could be employed as a potential solution to complex multi-modal problems. For example, a higher-level task planner could propose several ancillary controllers and alternative plans to be sampled to achieve global solutions.

REFERENCES

- [1] H. J. Kappen, "Path integrals and symmetry breaking for optimal control theory," *J. Stat. Mechanics: Theory Experiment*, vol. 2005, no. 11, pp. 205–229, Nov. 2005.
- [2] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Aggressive driving with model predictive path integral control," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2016, pp. 1433–1440.
- [3] G. Williams, A. Aldrich, and E. A. Theodorou, "Model predictive path integral control: From theory to parallel computation," *J. Guid., Control, Dyn.*, vol. 40, no. 2, pp. 344–357, Feb. 2017.
- [4] G. Williams et al., "Information theoretic MPC for model-based reinforcement learning," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 1714–1721.
- [5] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and I. A. Theodorou, "Information-theoretic model predictive control: Theory and applications to autonomous driving," *IEEE Trans. Robot.*, vol. 34, no. 6, pp. 1603–1622, Dec. 2018.
- [6] D. Pérez-Morales and V. Fremont, "Information-theoretic sensor-based predictive control for autonomous vehicle navigation: A proof of concept," in *Proc. IEEE Int. Intell. Transp. Syst. Conf.*, 2021, pp. 879–884.
- [7] G. Williams, B. Goldfain, P. Drews, J. M. Rehg, and E. A. Theodorou, "Best response model predictive control for agile interactions between autonomous ground vehicles," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 2403–2410.
- [8] I. S. Mohamed, G. Allibert, and P. Martinet, "Model predictive path integral control framework for partially observable navigation: A quadrotor case study," in *Proc. 16th Int. Conf. Control, Automat., Robot. Vis.*, 2020, pp. 196–203.
- [9] J. Pravitra, E. A. Theodorou, and E. N. Johnson, "Flying complex maneuvers with model predictive path integral control," in *Proc. Amer. Inst. Aeronaut. Astronaut. Scitech Forum*, 2021, pp. 1–12.
- [10] J. Pravitra, K. A. Ackerman, C. Cao, N. Hovakimyan, and E. A. Theodorou, "L1-Adaptive MPPI architecture for robust and agile control of multirotors," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 7661–7666.
- [11] V. Gómez, S. Thijssen, A. Symington, S. Hailes, and H. J. Kappen, "Real-time stochastic optimal control for multi-agent quadrotor systems," in *Proc. Int. Conf. Automated Plan. Scheduling*, 2016, pp. 468–476.
- [12] N. Wan, A. Gahlawat, N. Hovakimyan, E. A. Theodorou, and P. G. Voulgaris, "Cooperative path integral control for stochastic multi-agent systems," in *Proc. Amer. Control Conf.*, 2021, pp. 1262–1267.
- [13] L. Streichenberg, E. Trevisan, J. J. Chung, R. Siegwart, and J. Alonso-Mora, "Multi-agent path integral control for interaction-aware motion planning in urban canals," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 1379–1385.
- [14] M. Bhardwaj et al., "STORM: An integrated framework for fast joint-space model-predictive control for reactive manipulation," in *Proc. 5th Annu. Conf. Robot Learn.*, 2021, pp. 750–759.
- [15] I. Abraham, A. Handa, N. Ratliff, K. Lowrey, T. D. Murphey, and D. Fox, "Model-based generalization under parameter uncertainty using path integral control," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 2864–2871, Apr. 2020.
- [16] E. Arruda, M. J. Mathew, M. Kopicki, M. Mistry, M. Azad, and J. L. Wyatt, "Uncertainty averse pushing with model predictive path integral control," in *Proc. IEEE-RAS 17th Int. Conf. Humanoid Robot.*, 2017, pp. 497–502.
- [17] L. Cong, M. Grner, P. Ruppel, H. Liang, N. Hendrich, and J. Zhang, "Self-adapting recurrent models for object pushing from learning in simulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 5304–5310.
- [18] J. Carius, R. Ranftl, F. Farshidian, and M. Hutter, "Constrained stochastic optimal control with learned importance sampling: A path integral approach," *Int. J. Robot. Res.*, vol. 41, pp. 189–209, 2022.
- [19] T. Howell, N. Gileadi, S. Tunyasuvunakool, K. Zakka, T. Erez, and Y. Tassa, "Predictive sampling: Real-time behaviour synthesis with MuJoCo," Dec. 2022, *arXiv:2212.00541*.
- [20] C. Pezzato, C. Salmi, M. Spahn, E. Trevisan, J. Alonso-Mora, and C. H. Corbato, "Sampling-based model predictive control leveraging parallelizable physics simulations," Jul. 2023, *arXiv:2307.09105*.
- [21] G. Williams, B. Goldfain, P. Drews, K. Saigol, J. Rehg, and E. Theodorou, "Robust sampling based model predictive control with sparse objective information," in *Proc. Robot.: Sci. Syst. XIV*, 2018.
- [22] T. Mensink, J. Verbeek, and B. Kappen, "EP for efficient stochastic control with obstacles," *Front. Artif. Intell. Appl.*, vol. 215, pp. 675–680, 2010.
- [23] M. Okada and T. Taniguchi, "Acceleration of gradient-based path integral method for efficient optimal and inverse optimal control," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 3013–3020.
- [24] D. M. Asmar, R. Senanayake, S. Manuel, and M. J. Kochenderfer, "Model predictive optimized path integral strategies," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 3182–3188.
- [25] I. S. Mohamed, K. Yin, and L. Liu, "Autonomous navigation of AGVs in unknown cluttered environments: Log-MPPI control strategy," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 10240–10247, Oct. 2022.
- [26] R. Kusumoto, L. Palmieri, M. Spies, A. Csiszar, and K. O. Arras, "Informed information theoretic model predictive control," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 2047–2053.
- [27] T. Power and D. Berenson, "Variational inference MPC using normalizing flows and out-of-distribution projection," in *Proc. Robot.: Sci. Syst. XVIII*, 2022.
- [28] I. M. Balci, E. Bakolas, B. Vlahov, and E. A. Theodorou, "Constrained covariance steering based Tube-MPPI," in *Proc. Amer. Control Conf.*, 2022, pp. 4197–4202.
- [29] J. Yin, Z. Zhang, E. Theodorou, and P. Tsotras, "Trajectory distribution control for model predictive path integral control using covariance steering," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 1478–1484.
- [30] O. Arslan, E. A. Theodorou, and P. Tsotras, "Information-theoretic stochastic optimal control via incremental sampling-based algorithms," in *Proc. IEEE Symp. Adaptive Dynamic Program. Reinforcement Learn.*, 2014, pp. 1–8.
- [31] M. S. Gandhi, B. Vlahov, J. Gibson, G. Williams, and E. A. Theodorou, "Robust model predictive path integral control: Analysis and performance guarantees," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 1423–1430, Apr. 2021.
- [32] G. R. Williams, "Model predictive path integral control: Theoretical foundations and applications to autonomous driving," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA, USA, Mar. 2019. [Online]. Available: <https://smartechn.gatech.edu/handle/1853/62666>
- [33] Q. Inc., "QUBE - servo 2 - quanser." Accessed: May 2024. [Online]. Available: <https://www.quanser.com/products/qube-servo-2/>
- [34] I. Tejado, D. Torres, E. Pérez, and B. M. Vinagre, "Physical modeling based simulators to support teaching in automatic control: The rotatory pendulum," in *Proc. 11th IFAC Symp. Adv. Control Educ.*, 2016, pp. 75–80.
- [35] K. J. Åström and K. Furuta, "Swinging up a pendulum by energy control," *Automatica*, vol. 36, no. 2, pp. 287–295, Feb. 2000.
- [36] W. Wang et al., "Design, modeling, and nonlinear model predictive tracking control of a novel autonomous surface vehicle," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 6189–6196.
- [37] W. Jansma, E. Trevisan, A. Serra-Gómez, and J. Alonso-Mora, "Interaction-aware sampling-based MPC with learned local goal predictions," in *Proc. Int. Symp. Multi-Robot Multi-Agent Syst.*, 2023, pp. 15–21.