Particle MPC for Uncertain and Learning-Based Control

Robert Dyro, James Harrison, Apoorva Sharma, Marco Pavone

Abstract— As robotic systems move from highly structured environments to open worlds, incorporating uncertainty from dynamics learning or state estimation into the control pipeline is essential for robust performance. In this paper we present a nonlinear particle model predictive control (PMPC) approach to control under uncertainty, which directly incorporates any particle-based uncertainty representation, such as those common in robotics. Our approach builds on scenario methods for MPC, but in contrast to existing approaches, which either constrain all or only the first timestep to share actions across scenarios, we investigate the impact of a *partial consensus horizon*. Implementing this optimization for nonlinear dynamics by leveraging sequential convex optimization, our approach yields an efficient framework that can be tuned to the particular information gain dynamics of a system to mitigate both over-conservatism and over-optimism. We investigate our approach for two robotic systems across three problem settings: time-varying, partially observed dynamics; sensing uncertainty; and model-based reinforcement learning, and show that our approach improves performance over baselines in all settings.

I. Introduction

As autonomous decision-making agents such as robots move from narrowly tailored environments and begin operating in unstructured, uncertain worlds, incorporation of uncertainty into the decision-making pipeline is essential to intelligently trade off risk and reward. Uncertainty is ubiquitous in robotic systems; from state estimation, to fault detection, to dynamics learning, approximate Bayesian estimation and filtering methods hold central roles throughout the autonomy stack. Moreover, as robots regularly enter novel environments, interact with humans, or perform novel tasks, consideration of uncertainty is becoming increasingly consequential.

Modern robotic systems already quantify uncertainty in the state-estimation or dynamics learning process, through e.g. particle filtering or ensemble learning. Nevertheless, this uncertainty is often ignored in action selection: *certainty equivalent* control methods – which ignore uncertainty when choosing actions – are commonplace due to their ease of implementation, computational efficiency, and practical performance. Within the controls community, there exist several techniques for factoring in state and model uncertainty into an optimal control framework, yet these methods are often too conservative (failing to account for future information gain and/or assuming adversarial disturbances) [1] or too computationally intensive, limiting their applicability to systems with slower dynamics, such as process control [2].

In this work, we build on uncertainty-aware optimal control approaches developed in the controls community,

Authors are with Stanford University, Stanford CA 94305, USA. {rdyro, jharrison, apoorva, pavone}@stanford.edu

specifically scenario-based model predictive control (MPC), which optimizes a sequence of actions for sampled possible realizations of the uncertainty. Notably, in contrast to existing work which either constrains actions to be shared across scenarios for either the *full* planning horizon (leading to over-conservatism), or only the first time step (leading to over-optimism), we employ a *tunable consensus horizon* allowing a practitioner to tailor the controller to the information gain dynamics of their application. Our approach operates directly on the particle-based uncertainty representations commonplace in robotic systems, and, leveraging sequential convex programming, handles nonlinear dynamics while remaining computationally efficient.

A. Related Work

The problem of optimal control with uncertainty is the focus of the fields of stochastic and robust optimal control, each of which have long histories. A common approach is the class of robust methods—such as methods from classical robust control—which factor in uncertainty via assuming worst-case disturbances [3], [4]. However, if certain disturbances are less likely than others, this worst-case treatment of uncertainty can lead to overly conservative behavior, limiting practical use in robotic systems.

To handle probabilistic disturbances, work over the last several decades has focused on distributional approximations of uncertainties, typically approximating the disturbance distribution as a Gaussian. While these approaches, such as those of [5], [6], [7], yield tractable solutions, the Gaussian approximation can be poor for the arbitrary, potentially multi-modal uncertainties that are commonplace in robotics, e.g. a chance of actuator failure. Moreover, works which model the uncertainty as additive i.i.d. noise, such as [8], [6], ignore the temporal correlation of uncertainty. This is particularly problematic for dynamics uncertainty: if an unknown dynamics parameter is non-time-varying, the impact of this parameter will be different (and typically result in a wider distribution of outcomes) than if the same parameter is re-sampled at every time step [9]. However, consideration of temporal correlation in distributional approximations typically leads to computational difficulties.

Difficulties in handling temporal correlation of disturbances has motivated sampling-based (or similarly, particle methods or scenario methods) approaches to handling uncertainty. Several works have considered the case of linear systems, which yields a large convex optimization problem composed of all the scenarios [1], [10]. In the nonlinear setting, iterative importance sampling control schemes such as that of [11] have seen application in settings with model uncertainty [12], [13], [14]. These methods do not rely on the gradient of the cost function (or dynamics) for

optimization, and are strictly sampling-based. In contrast to these works, our framework is capable of handling arbitrary nonlinear systems, and uses available gradients to improve performance, as we demonstrate in our experiments.

Another family of approaches is multi-stage MPC, which consider a branching tree structure (corresponding to sampled scenarios or disturbances). Rather than share actions across all branches, as in robust control methods, multi-stage MPC imposes causality constraints by sharing actions only at each branch point, thus preventing the controller from anticipating a disturbance before it occurs [2]. The tree structure induced by these approaches has lead multi-stage methods to consider both a "robust" phase, in which branching occurs, and a tail phase in the control optimization in which branching no longer occurs and thus causality constraints are not imposed. While this approach is more reflective of the true information pattern available to the controller during online operation, it is either expensive due to an explosion in the number of sampled scenarios, or under-conservative [15]. Indeed, the computational complexity of these multi-stage methods has so far restricted their application to problems with slow dynamics, such as process control [15], [16]. Moreover, in the partially observed setting, reflecting the true information pattern requires either (differentiably) simulating the state estimator forward in time [17], [18], or may result in substantial under-conservatism due to assumed perfect state information after a small number of steps. These limitations stand in stark contrast to methods such as [11] that have seen widespread use in learning-based control due to their simplicity and broad applicability. In contrast to the literature on multi-stage MPC, we enforce a constraint requiring matching actions for an extended horizon while limiting the branching of the scenario tree, resulting in a simple, tunable, and effective approach.

B. Contributions

In this work, we propose a particle model predictive control (PMPC) formulation that leverages sequential convex programming (SCP) and a tunable consensus horizon to yield an efficient nonlinear control framework capable of handling arbitrary uncertainties. This particle MPC approach, in contrast to multi-stage or scenario MPC methods, minimizes branching and focuses on principled simplifications to make the approach computationally efficient, tunably conservative, and simple to implement. We argue that this formulation, based on Monte Carlo distributional approximations as opposed to methods based on set theoretic uncertainty or approximation with analytically tractable distributions, is a highly flexible and effective approach to control under uncertainty. We investigate the role of the consensus horizon both theoretically—showing that its optimal choice depends on the information gain rate in the system—as well as experimentally.

We investigate three problem settings: learning-based control (or equivalently, model-based reinforcement learning) in which an agents learns a model of the system dynamics with no prior information; time-varying parametric uncertainty such as faults or wind gusts; and sensing uncertainties such as those arising from particle filter state estimation. All three problems are investigated on two different systems—a free-flying space robot system and a planar quadrotor—and we find significant performance gains from our approach relative to the standard certainty equivalent and uncertainty-aware sampling-based control schemes.

II. PROBLEM FORMULATION

Our goal in this work is to control partially observed dynamical systems. This general formulation, a Partially Observable Markov Decision Process (POMDP), captures several sources of uncertainty in robotics. We denote the system state at time step j as $x^{(j)} \in \mathbb{R}^s$, the action taken from this state $u^{(j)} \in \mathbb{R}^a$, and the observation as $y^{(j)} \in \mathbb{R}^o$. The system has nonlinear, discrete time dynamics

$$x^{(j+1)} = f(x^{(j)}, u^{(j)}, w^{(j)}), \qquad y^{(j)} = g(x^{(j)}, v^{(j)})$$

where $f(\cdot)$ and $g(\cdot)$ are the state dynamics and observation function respectively, and $w^{(j)}, v^{(j)}$ are stochastic disturbances.

As solving POMDPs directly is intractable [19], in robotic systems this problem is typically decomposed into *state estimation*, which uses observations to estimate the current state $x^{(j)}$, and *control*, which aims to select the action $u^{(j)}$ given this estimate. The state can not be known exactly, robotic systems often turn to Bayesian filtering to maintain an update a *belief distribution* $\mathcal{X}_0$ over possible states. Note that this formulation can also characterize uncertain dynamics, where the unknown parameters of the dynamics can be considered as unobserved state elements which remain constant over time.

Our focus is control: choosing an optimal action to take given the current belief over state. Given information about the control task in the form of a stage-wise cost function $c^{(j)}(x^{(j)}, u^{(j)})$ and state and action constraints $x^{(j)} \in \mathbb{X}^{(j)}$, $u^{(j)} \in \mathbb{U}^{(j)}$, this gives rise to the finite horizon optimal control problem

$$
\begin{aligned}
\underset{\boldsymbol{u}}{\text{minimize}} \quad & \sum_{j=0}^{N} \mathbb{E}_{x^{(j)}} \left[c^{(j)}(x^{(j)}, u^{(j)}) \right] \\
\text{subject to} \quad & x^{(j+1)} = f(x^{(j)}, u^{(j)}, w^{(j)}) \\
& x^{(j)} \in \mathbb{X}^{(j)} \ \forall j \\
& u^{(j)} \in \mathbb{U}^{(j)} \ \forall j \\
& w^{(j)} \sim \mathcal{W}^{(j)} \ \forall j \\
& x^{(0)} \sim \mathcal{X}_0
\end{aligned}
\tag{1}
$$

We consider the cases where $\mathcal{X}_0$ is a discrete distribution over possible values of $x^{(0)}$, as is the case with a particle filtering approach to state estimation, or an ensemble or sampling based approach to quantifying uncertainty in learned dynamics models. Note that if the true system is not in the support of this discrete distribution, then guaranteeing (even probabilisitically) that the true system will also satisfy state constraints is an open research question which is outside the scope of this paper.

Algorithm 1 Receding horizon PMPC with learning

Require: State belief $\mathcal{X}_0$
1: Sample $\{x_i^{(0)}\}_{i=1}^M$, $\{w_i^{(j)}\}_{i=1}^M \sim (\mathcal{X}_0, \mathcal{W}_j)$ particles and disturbances from current belief (state, unknown disturbances)
2: Compute the PMPC plan $\{\boldsymbol{x}_i, \boldsymbol{u}_i\}$ on sampled particles.
3: Execute first action of plan $u^{(j=0)}$.
4: Observe transition $y \leftarrow g\big(f(x = x^{(0)}, u^{(j=0)}, w)\big)$
5: $\mathcal{X}^{(0)} \leftarrow$ Updated belief after observing $(y, u^{(j=0)})$

III. PARTICLE MODEL PREDICTIVE CONTROL

We propose to solve the optimal control problem (1) using a receding horizon approach; at each time step, we optimize a sequence of N control actions into the future. By replanning at every step, we are able to account for the evolution of the state estimate, which changes as new observations are made. Receding horizon control is a well established technique in applications of MPC and we make the approach explicit in Algorithm 1.

The fundamental challenges in solving (1) comes from (i) the presence of uncertainty and (ii) the nonlinear dynamics. In this section, we detail our algorithmic approach, first focusing on how we handle uncertainty, and then discussing how we deal with the nonlinear dynamics.

A. Handling Uncertainty through Particle Model Predictive Control

In designing a control strategy for this problem, we need an approach which can factor in the temporally correlated effects of state uncertainty. At the same time we need an approach which remains computationally efficient, so that we can replan online to handle deviations due to the inevitable plan mismatch with reality and factor in online updates in state belief updates coming from online estimation. In order to do so, we propose to operate directly on a particle-based representation of state uncertainty.

Formally, our particle model predictive control (PMPC) approach has the structure of the finite horizon optimal control problem (1). We approximate the uncertainty in the initial state and dynamics by a collection of particles. Each particle represents the evolution of the system under one realization of the initial state. That is, for particle i, the state evolves according to $x_i^{(j+1)} = f(x_i^{(j)}, u_i^{(j)}, w_i^{(j)})$, where $x_i^{(0)} \sim \mathcal{X}_0$. By propagating each particle separately in time, we account for the time correlated effects of dynamics and state uncertainty in contrast to approaches which handle uncertainty through uncorrelated disturbances at each time step. Importantly, for each particle we get a *separate trajectory* with potentially different states and control actions.

For applications in real systems, we employ the well known strategy of receding horizon control with replanning at every step. The overview of our approach is contained in Algorithm 1. We approximate the expectation in (1) via a (potentially weighted) sum of costs across these particles.

We enforce constraints uniformly over all sampled particles. This is only guaranteed to ensure the true system satisfies constraints if the set of particles includes the true system state and noise realization; if this is not the case then guaranteeing constraint satisfaction (even at some probability) is in general a hard set-theoretic problem in stochastic control, especially for arbitrary uncertainty distributions. In our experiments, we found that with a reasonable number of particles, our approach was sufficient, but developing practical guarantees on probabilistic constraint satisfaction remains an important direction of research. In our implementation, we handle state constraints with exact penalty functions [20], which ensures constraint satisfaction when the problem is feasible, and minimizes violation if the problem becomes infeasible.

Even as we simulate *several* possible future trajectories, or scenarios, in this optimization, we must choose a *single* action to execute.[1] This introduces a key question of how to enforce *consensus* among the different scenarios. Existing approaches have focused either on *full consensus*, where we ensure actions are shared across particles over all time steps $(u_i^{(j)} = u^{(j)}, \forall i = 1, \ldots, M, \quad j = 0, \ldots, N)$, or *one-step consensus* where only the first action (the one that will be executed) is shared across scenarios $(u_i^{(0)} = u^{(0), \forall i=1,\ldots,M})$. Full consensus ensures that actions perform well—open-loop—across all sampled scenarios. However, as robotic systems employ filtering techniques which reduce uncertainty at each timestep, this procedure can be over-conservative, and impractical for robotics applications. One-step consensus, on the other hand, allows choosing actions tailored to each scenario after the first time-step, and thus may be over-optimistic, and thus lead to suboptimal performance on the true uncertain system. A key contribution of this work is to consider interpolating between these two extremes, which each have their advantages and drawbacks, through a *tunable consensus horizon*. Specifically, we define the consensus horizon $N_{\rm c}$, as the number of timesteps for which actions must be shared across particles $(u_i^{(j)} = u_k^{(j)} \quad \forall j \in [0..N_{\rm c}] \quad i, k \in [1..M])$. We discuss the implications of this hyperparameter further in Section IV-A.

B. Combining Particle Representations with Sequential Convex Programming

The PMPC approach yields a large non-convex optimization problem. In order to approximately solve this problem efficiently, we employ sequential convex programming (SCP). SCP applies efficient, optimized solvers for convex optimization problems sequentially to locally optimize a nonlinear optimization problem. At each step of SCP, we first form a convex approximation of the non-convex problem around a reference setting of the optimization variables, and then solve this convex problem to get a new reference point—while attempting to stay close to the previous reference point. SCP can be shown to converge to a local optimum under some mild assumptions, for further discussion we refer the reader to [23].

---

[1]Even though some works suggest combining MPC with a feedback policy [21], [22], computed on- or offline, we experimentally found that to be computationally challenging in real-time applications and yielding little improvement in receding horizon control. Hence, we do not consider a feedback policy in our MPC problem.

In the case of PMPC, the optimization variables are the state and action trajectories of each particle $\{\boldsymbol{x}_i, \boldsymbol{u}_i\}_{i=1}^M$ where $\boldsymbol{x}_i = (x_i^{(0)}, \ldots, x_i^{(N+1)})$ and $\boldsymbol{u}_i = (u_i^{(0)}, \ldots, u_i^{(N)})$. To convexify the problem at a particular setting of these optimization variables, we replace the dynamics with a linear approximation and the cost with a linear approximation of its non-convex part, yielding the convex optimization problem:

$$\min_{\{\Delta \boldsymbol{x}_i, \Delta \boldsymbol{u}_i\}_{i=1}^M} \frac{1}{M} \sum_{i=1}^M \sum_{j=0}^N \Bigg(c_{\text{cvx}}(x_i^{(j)} + \Delta x_i^{(j)}, u_i^{(j)} + \Delta u_i^{(j)})$$
$$+ \nabla_x c_{\text{ncvx}}^{(j)} \Delta x_i^{(j)} + \nabla_u c_{\text{ncvx}}^{(j)} \Delta u_i^{(j)}$$
$$+ \rho_x \| \Delta x_i^{(j)} \|_2^2 + \rho_u \| \Delta u_i^{(j)} \|_2^2 \Bigg)$$
$$\text{s.t.} \quad \Delta x_i^{(j+1)} = \nabla_x f_i \Delta x_i^{(j)} + \nabla_u f_i \Delta u_i^{(j)}$$
$$\forall j \in [0..N] \quad i \in [1..M]$$
$$\bar{x}_i^{(j)} + \Delta x_i^{(j)} \in \mathbb{X}^{(j)} \quad \forall j \in [0..N] \quad i \in [1..M]$$
$$\bar{u}_i^{(j)} + \Delta u_i^{(j)} \in \mathbb{U}^{(j)} \quad \forall j \in [0..N] \quad i \in [1..M]$$
$$\underbrace{\Delta u_i^{(j)} = \Delta u_k^{(j)} \quad \forall j \in [0..N_{\text{c}}] \quad i, k \in [1..M]}_{\text{partial consensus}}$$
$$(2)$$

where this optimization problem is written in terms of the deviations $\{\Delta \boldsymbol{x}_i, \Delta \boldsymbol{u}_i\}_{i=1}^M$ from the previous solution $\{\bar{\boldsymbol{x}}_i, \bar{\boldsymbol{u}}_i\}_{i=1}^M$, and the gradients are evaluated at $\{\bar{\boldsymbol{x}}_i, \bar{\boldsymbol{u}}_i\}_{i=1}^M$. Note that we add two terms to the cost to minimize deviation from the linearization point, as the approximation is only good for small deviations. While there are many strategies for limiting this deviation in each SCP iteration, we choose the quadratic penalties as they work well in practice and introduce only two hyperparameters ρ_x and ρ_u. We find that the algorithm is generally not sensitive to the particular value of these hyperparamters, as long as they are above a certain threshold: Larger values of ρ_x, ρ_u slow the convergence of the SCP procedure, but too low a value leads to instability of the SCP procedure. In practice we assume a multiplicative relation between the two, i.e., $\rho_x = \alpha \rho_u$ and search for a good value ρ_x using a few steps of bisection. Both hyperparameters affect the optimization in the same manner—they penalize the variable deviation from the linearized trajectory—which justifies reducing them to a single hyperparameter. Explicitly recognizing separate penalization of state and control action deviation allows the system designer to account for difference in the order of magnitude of state and control actions, in cases where, for examples, control actions are confined between 0 and 1 and state between -100 and 100. SCP requires an initial trajectory guess, and choosing this initialization remains an open research problem [24]. In our experiments, we find that PMPC is robust to the initial trajectory guess; even simple infeasible guesses—such as static trajectory (repetition of the current state belief) and linear interpolation between initial and goal state—work well.

Algorithm 2 details the procedure for one planning iteration. Each iteration takes as input a set of particles, where each particle i has its own initial state, $x_i^{(0)}$, and dynamics deviation $w_i^{(j)}$. Note that in the case of a non-uniform belief

---

Algorithm 2 SCP PMPC

Require: Initial states $\{x_i^{(0)}\}_{i=1}^M$, dynamics models $\{f_i\}_{i=1}^M$, solution guess $\{\boldsymbol{x}_i, \boldsymbol{u}_i\}_{i=1}^M$
Require: Hyperparameters $\rho_x, \rho_u, N_{\text{c}}$
Require: Solution tolerance ϵ
1: **repeat**
2:     $\{\bar{f}_i^{(j)}, \nabla_x f_i, \nabla_u f_i\}_{i=1}^M \leftarrow$ Linearize dynamics around the trajectory guess $\{\boldsymbol{x}_i, \boldsymbol{u}_i\}_{i=1}^M$
3:     Split the cost into the convex and non-convex parts $\{c_{i,\text{cvx}}^{(j)}, c_{i,\text{ncvx}}^{(j)}\}_{i=1}^M$
4:     $\{\nabla_x c_{i,\text{ncvx}}^{(j)}, \nabla_u c_{i,\text{ncvx}}^{(j)}\}_{i=1}^M \leftarrow$ Linearize non-convex cost around $\{\boldsymbol{x}_i, \boldsymbol{u}_i\}_{i=1}^M$
5:     $\{\Delta x_i^{(j)}, \Delta u_i^{(j)}\}_{i=1}^M \leftarrow$ Solution to convex prob. (2)
6:     $\{\boldsymbol{x}_i, \boldsymbol{u}_i\}_{i=1}^M \leftarrow \{\boldsymbol{x}_i + \Delta \boldsymbol{x}_i, \boldsymbol{u}_i + \Delta \boldsymbol{u}_i\}_{i=1}^M$
7: **until** $\sum_{i=1}^M \sum_{j=0}^N \|\Delta x_i^{(j)}\| + \|\Delta u_i^{(j)}\| < \epsilon$
8: **return** $\{\boldsymbol{x}_i, \boldsymbol{u}_i\}_{i=1}^M$

---

over the particles, the relative weighting of the particles can be incorporated by scaling the particle's cost function by that weighting. SCP alternates between convexification of the dynamics and cost functions, and subsequently solving the convex problem (2) by leveraging an off-the-shelf efficient convex solver. If the algorithm converges, we are guaranteed to have a locally near-optimal solution. We detect convergence by evaluating the norm of the change in the solution trajectory, as at a locally optimal solution, solving (2) would yield a solution with deviations equal to zero. When stopped prior to convergence, the solution is approximate up to the error in the linear approximation of the dynamics and the cost.

IV. DISCUSSION

A. Choice of Consensus Horizon

A key hyperparameter in PMPC is the consensus horizon N_{c}, and the range from *one-step* to *full* consensus has a large impact on the closed loop performance of the receding horizon planning controller. One way to interpret the choice of consensus horizon is as an approximation of the information gain dynamics: uncertainty is present for N_{c} steps, after which it drops immediately to zero when full information is revealed. By enforcing control consensus for N_{c} steps, PMPC is assuming that the uncertainty will remain constant for this time period, and so must choose actions that work for all possibilities. After the consensus horizon, controls can be tailored to each particle, which is only possible when the true system is revealed. Much of the controls and decision making literature has focused on the two extremes for consensus horizon: one-step consensus corresponds to the QMDP approximation of POMDPs [25], and full consensus corresponds to robust control or scenario MPC. Instead, PMPC allows choosing any consensus horizon between these two extremes, interpolating between these two approaches. Full consensus chooses one sequence of actions which satisfy constraints across all possible values of the uncertain state. While this ensures safety, it can lead to

over-conservative solutions which fail to account for future information gain through the state estimation module of a robotic system. However, choosing a consensus horizon that is too short can lead to *over-optimism*, and result in actions that lead to constraint violations or large costs, as shown in the following theorem:

Theorem 1 (Consensus horizon & replanning): For any integers $N^\star \in [2..\infty)$, $K \in [1..N^\star - 1]$, there exists an optimal control problem (1) defined by $\mathcal{X}_0, f, c, \mathbb{X}, \mathbb{U}, \mathbb{W}$ for which running PMPC with a consensus horizon of $N_c = K < N^\star$ leads to constraint violation, but a consensus horizon of $N_c = N^\star$ does not.

The proof is available in the appendix, available at [26].

As this theorem shows, even though only the first action in the receding horizon plan is ever executed on the real system, limiting consensus to only this first action can lead to constraint violation and otherwise unsafe closed-loop behavior. This highlights the utility of choosing an intermediate value for the consensus horizon.

Our approach is similar in spirit to choosing a shorter robust horizon is multi-stage MPC [2], but there are several key differences. First, multi-stage MPC builds a scenario tree, branching at each timestep up to the robust horizon. Actions are only shared at branch points of the tree. In contrast PMPC creates a scenario tree that only branches at the root, but enforces control consensus over an tunable horizon N_c. Additionally, multi-stage MPC typically incorporates information gain over the robust horizon, in order to determine how to sample uncertainties and branch the tree. While factoring in information gain dynamics into action selection leads to closer-to-optimal performance, it comes at a steep computational expense. In multi-stage MPC, this adds a significant computational burden and the robust horizon is often set to 1 for computational efficiency. Indeed, considering this over the full planning horizon yields the *dual control* problem, which is computationally difficult for general nonlinear systems [19], [27]. In contrast, we sacrifice optimality and avoid the computationally expensive process of propagating belief dynamics. Instead, we adopt the simpler information gain dynamics of N_c-step consensus, enabling us to reap the safety benefits of choosing a consensus horizon greater than 1 without suffering a large computational burden associated with belief propagation.

B. Particle Model Predictive Control vs Certainty Equivalent Model Predictive Control

A baseline for control under uncertainty is to assume *certainty equivalence* (CE), i.e. choose a set of actions assuming a nominal value for $x^{(0)}$ is exactly correct. With the exception of systems with linear dynamics and Gaussian state uncertainty [28], this technique is sub-optimal, but nevertheless is easy to implement and commonly used in robotics. CE control exists as a special case of PMPC where we use only a single particle corresponding to the nominal setting of all uncertain quantities. CE control is appealing as a simpler, easier to implement solution, especially for low, unimodal uncertainty for which choosing a nominal value for $x^{(0)}$ it is significantly complicated when dealing with
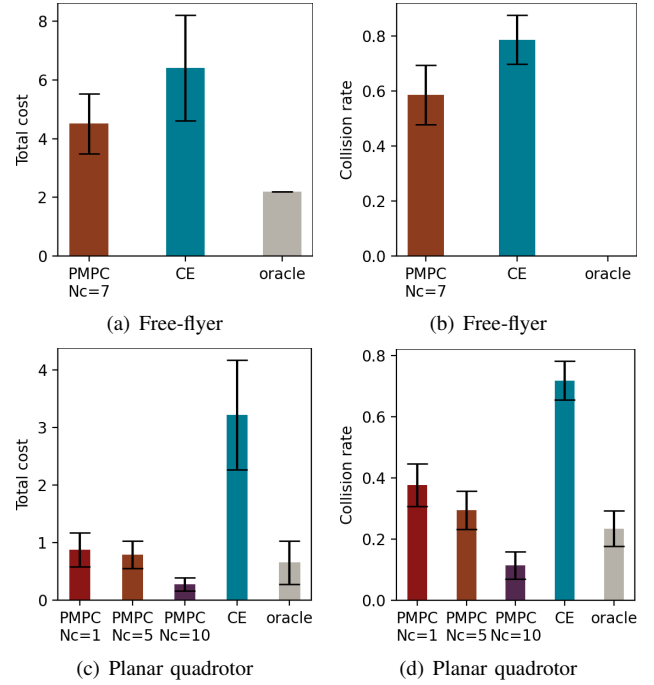

Fig. 1. Total cost and collision rate for the free-flyer and the planar quadrotor under **model uncertainty**.

multimodal uncertainties for which the mean may not be an accurate representation. For example, with multimodal sensing uncertainty, planning assuming the robot is at the mean state may lead to actions that are unsuitable for all possible scenarios. PMPC allows directly translating arbitrary uncertainty representations on state, dynamics, and cost function to the controller, sidestepping the often complex decision of reducing an uncertain system to a certain one for CE control. In our experiments, we demonstrate that PMPC outperforms CE control on common sources of uncertainty, both uni- and multi-modal.

V. EXPERIMENTS

We demonstrate the utility of PMPC by evaluating its performance on two nonlinear systems in three settings each, chosen to highlight distinct sources of uncertainty arising in robotics problems. Specifically, we use (1) a 6-D planar quadrotor (2-D action space), a common system for benchmarking highly dynamic control and reinforcement learning algorithms [29], [30], [31] and (2) a 6-D planar free-flyer, a free floating spacecraft constrained to a plane with actuation through gas thrusters and a reaction wheel (9-D action space) [32]. We use first principles continuous dynamics discretized with explicit Runge-Kutta 4 method.

For each of these systems, we consider three sources of uncertainty that are prevalent throughout robot autonomy: (i) changing system dynamics within an otherwise known dynamics model (e.g. actuator failures or external effects), (ii) state uncertainty, and (iii) epistemic uncertainty within black-box learning-based control (model-based reinforcement learning). In each of these settings, we execute a task corresponding to motion to a desired position while avoiding

obstacles. We encode this task using a cost function which generally takes the form of a quadratic cost on position to encourage movement towards the goal, with exact penalty for collision with obstacles which is zero clipped linear in the amount of violation. Note that the non-convexity of the free space makes this cost function non-convex. We also include experiments investigating the computational complexity of the method for varying numbers of particles, showing the approach is feasible in real time.

Because of the wide variety of test environments, we exclude details of each environment from the body of the paper. The details for all environments, including details of the design of the particle dynamics in the control scheme, are available in the appendix [26]. This appendix also includes additional experimental results. Throughout all figures in this section, we show 95% confidence intervals.

We solve the SCP problem using the open-source QP solver OSQP [33]. Except for dynamic uncertainty Free-flyer with $M = 16$, we choose $M = 10$ particles.

A. Dynamics Uncertainty

Dynamics uncertainty is commonplace in robotics. For one, robotic systems may evolve stochastically over time, e.g. having a nonzero chance of actuators failing. They can also be affected by similarly evolving external factors, e.g. wind disturbances. We use the two systems to investigate the performance of PMPC in these characteristic settings of dynamics uncertainty.

For the free-flyer problem, we consider a scenario in which the robot aims to regulate to a position near a wall, mimicking a docking scenario. For this system, each thruster has a random failure probability at each time step. The current failure status (a discrete random variable) is inferred online by a recursive Bayes filter. We compare to a certainty equivalent (CE) formulation that acts with respect to the maximum a posteriori (MAP) failure state estimate.

For the planar quadrotor, we assume a wind disturbance modeled by an Ornstein-Uhlenbeck process in both spatial dimensions. This process is, roughly, a Brownian motion process with a term that pulls the process toward 0 and thus captures the stochastic but temporally correlated nature of wind gusts. In this problem, the quadrotor aims to navigate a narrow passage while avoiding collision. The CE baseline assumes the current wind remains constant.

Results for both systems can be seen in Figure 1. In both experiments, we also compare to an oracle baseline which has perfect knowledge of future dynamics changes. For the free-flyer system, we plot only $N_c = 7$ as results were consistent across varying consensus horizons. For the planar quadrotor, we visualize performance for $N_c = 1, 5, 10$. In our experiments, all the PMPC approaches outperformed the CE approach, and avoided collisions substantially more frequently. Interestingly, PMPC with $N_c = 10$ actually outperformed the oracle model. While the oracle model has perfect knowledge of the time evolution of parameters governing stochastic disturbances *within the finite planning horizon*, it fails to account for sample values beyond it. Thus, the conservatism added by the consensus horizon improves
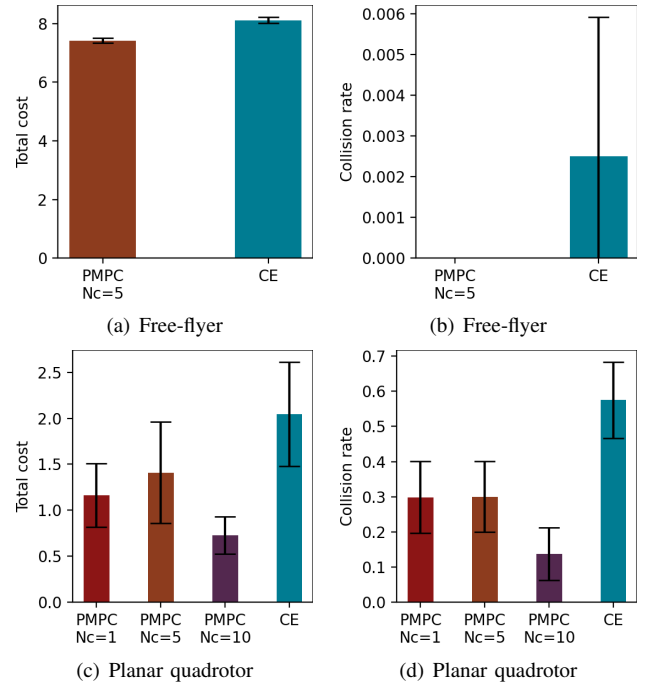

Fig. 2. Total cost and collision rate for the free-flyer and the planar quadrotor under **state uncertainty**.

performance due to additional robustness to these stochastic disturbances.

B. State Uncertainty

A common source of uncertainty in robotics comes from partial observability of the system's state. Typically, robotic systems employ online filtering to estimate their current state, for example, by using a particle filter [34]. We test how well SCP PMPC can address this form of uncertainty on both systems. In both cases, we assume we only have uncertainty on position, initialized as a unimodal distribution of particles. This distribution is updated online using a particle filter given noisy observations of distance from four laser range finder sensors aligned with cardinal directions in the vehicle frame. As the true observation noise is often not known exactly, we choose a different noise covariance for the particle filter based on performance of the downstream controller. These particles are directly fed in to the SCP PMPC controller. For the CE controller, we plan assuming the state is the expected position of the particles. In both settings, we evaluate performance over scenarios where the true position is sampled from initial belief over position.

Results for this setting are plotted in Figure 2. Again, we see uniformly better performance for the PMPC controller compared to the CE controller. We note that while the costs of the PMPC and CE controller for the free-flyer appear close, the difference in cost is highly significant, since it is due to collisions occurring in the CE, but not the PMPC controller.

C. Learning for Control

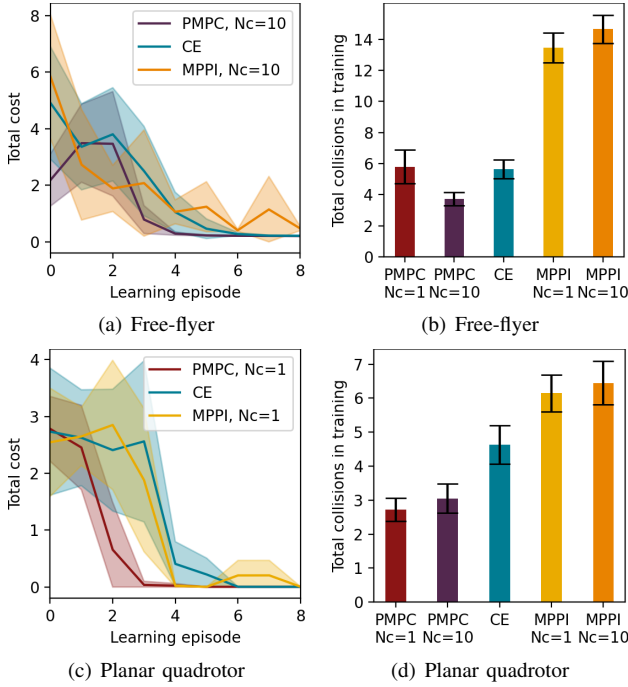A third, and increasingly common source of uncertainty in robotics arises from learned components. Any learned

Fig. 3. Learning curves (cost per iteration) and total number of collisions during training for the free-flyer and the planar quadrotor in a model-based reinforcement learning setting.

component has *epistemic uncertainty*, reflecting uncertainty in the underlying function as opposed to *aleatoric*, or irreducible, uncertainty which is an intrinsic feature of the environment. Characterization of this epistemic uncertainty has led to substantial improvements in learning-based control and reinforcement learning due to better robustness with respect to unknown dynamics and better exploration [12]. A common tool to quantify this uncertainty is using deep ensembles [35], wherein several neural networks are trained on the same data but starting from different initializations and regularized to different points. Following the problem setting of [12], we consider the problem of control with a learned dynamics model. Specifically, we consider a model-based RL framework, as the initial low-data regime highlights the need to factor in epistemic uncertainty.

We train D neural network dynamics models between every episode on the state, action, next-state tuples collected so far. We minimize a squared ℓ_2 loss, with each network regularized to its random initialization, following methods proposed by [36], [37]. During the episode, we use each network as the (stationary) dynamics for a single particle in the PMPC controller. We compare against a CE controller that uses a single network in the the ensembles (ignoring epistemic uncertainty) as well as an MPPI controller [11] which factors in this uncertainty but is a sampling-based, gradient free method (as used in [12]).

Our results are visualized in Figure 3. Figures (a) and (c) show learning curves for the reinforcement learning process: they show the cost per episode during learning. There are several things to note in these figures: (i) both the CE and PMPC controller outperform the MPPI controller and (ii) the

performance PMPC improves upon the CE method. (i) is likely the result of incorporating gradient-based information as opposed to using stochastic, sampling-based optimization schemes. (ii) shows the value of approximately incorporating uncertainty. Although the collisions in training are relatively close, the training curves for both systems show convergence 2-4 episodes before the CE approach. Additionally, the MPPI controller shows performance variation in later episodes due to its stochastic nature, while both the CE and PMPC controllers achieve consistent performance.

D. Computational Complexity

Figure 4 shows time per SCP iteration for a varying number of particles, for $N_c = 1, 5, 10$. The plots were generated on a Intel(R) Core(TM) i7-8559U CPU @ 2.70GHz. The slope of this log-log plot is approximately 1, showing approximately linear computational complexity in the number of particles. This is due to the utilization of sparse quadratic programming

Fig. 4. Time per SCP iteration of the nonlinear particle MPC scheme for varying number of particles, M.

solvers. Moreover, interestingly, the larger consensus period does not have a substantial impact on computational complexity until very large numbers of particles are used (≈ 1000), where there is an approximately $1.5\times$ performance difference for $N_c = 5, 10$ versus $N_c = 1$. The results show for an intermediate number of particles (e.g. approximately 20), operational frequencies of around 50 are possible even with off-the-shelf solvers and standard consumer CPUs. With GPU acceleration, we anticipate operational frequencies of 100Hz with up to 50 particles is achievable in the near term.

We highlight that the Monte Carlo methods we develop in this paper are currently becoming computationally feasible in real time due to advances in computational hardware for parallel processing and sparse solvers. Thus, we believe that effective design of Monte Carlo-based methods as opposed to imprecise approximations discussed in the introduction will be a fruitful avenue of research for uncertainty characterization in control in coming years.

VI. CONCLUSIONS

In this work we have presented a framework for particle-based nonlinear model predictive control enabling effective control under uncertainty, which offers tunable conservatism by using partial action consensus. We have demonstrated that on two different robot systems, accounting for three different uncertainties common in robotics, PMPC outperforms a certainty equivalent baseline. In the model-based RL (MBRL) problem setting, PMPC substantially outperforms uncertainty-aware MPPI [11], which has been shown to substantially improve MBRL relative to CE methods [12].
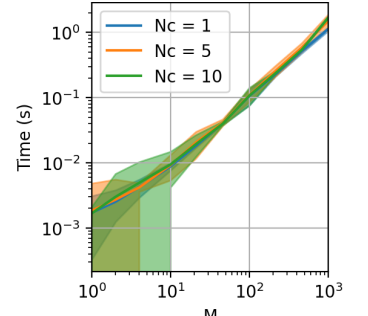
Finally, we have investigated the computational complexity of the proposed approach and found that the method is capable of real time operation, and will become increasingly capable for dynamic real time applications as computational hardware and sparse solvers improve. Thus, further algorithmic improvements in particle-based control methods represent a promising direction of future research for control under uncertainty.

REFERENCES

[1] Lars Blackmore, Masahiro Ono, Askar Bektassov, and Brian C. Williams. A probabilistic particle-control approximation of chance-constrained stochastic predictive control. *IEEE Transactions on Robotics*, 2010.

[2] Sergio Lucia, Tiago Finkler, and Sebastian Engell. Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty. *Journal of Process Control*, 2013.

[3] Mayuresh V. Kothare, Venkataramanan Balakrishnan, and Manfred Morari. Robust constrained model predictive control using linear matrix inequalities. *Automatica*, 1996.

[4] P.P. Khargonekar, I.R. Petersen, and M.A. Rotea. H$_\infty$-optimal control with state-feedback. *IEEE Transactions on Automatic Control*, 1988.

[5] Eugenio Cinquemani, Mayank Agarwal, Debasish Chatterjee, and John Lygeros. Convexity and convex approximations of discrete-time stochastic control problems with constraints. *Automatica*, 2011.

[6] Marc Deisenroth and Carl E Rasmussen. PILCO: A model-based and data-efficient approach to policy search. *International Conference on Machine Learning (ICML)*, 2011.

[7] Marc Peter Deisenroth. Learning to control a low-cost manipulator using data-efficient reinforcement learning. *Robotics: Science and Systems (RSS)*, 2012.

[8] Nikolas Kantas, J. M. Maciejowski, and A. Lecchini-Visintini. Sequential Monte Carlo for model predictive control. In *Nonlinear model predictive control*, pages 263–273. Springer, 2009.

[9] Andrew James McHutchon. *Nonlinear modelling and control using Gaussian processes*. PhD thesis, Cambridge University, 2015.

[10] Giuseppe C. Calafiore and Lorenzo Fagiano. Robust model predictive control via scenario optimization. *IEEE Transactions on Automatic Control*, 2012.

[11] Grady Williams, Nolan Wagener, Brian Goldfain, Paul Drews, James M Rehg, Byron Boots, and Evangelos A Theodorou. Information theoretic mpc for model-based reinforcement learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[12] Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In *Neural Information Processing Systems (NeurIPS)*, 2018.

[13] Ian Abraham, Ankur Handa, Nathan Ratliff, Kendall Lowrey, Todd D Murphey, and Dieter Fox. Model-based generalization under parameter uncertainty using path integral control. *IEEE Robotics and Automation Letters (R-AL)*, 2020.

[14] Ermano Arruda, Michael J Mathew, Marek Kopicki, Michael Mistry, Morteza Azad, and Jeremy L Wyatt. Uncertainty averse pushing with model predictive path integral control. In *IEEE-RAS International Conference on Humanoid Robots*, 2017.

[15] Sergio Lucia, Joel AE Andersson, Heiko Brandt, Moritz Diehl, and Sebastian Engell. Handling uncertainty in economic nonlinear model predictive control: A comparative case study. *Journal of Process Control*, 2014.

[16] Ruben Marti, Sergio Lucia, Daniel Sarabia, Radoslav Paulen, Sebastian Engell, and César de Prada. Improving scenario decomposition algorithms for robust nonlinear model predictive control. *Computers & Chemical Engineering*, 2015.

[17] S Thangavel, S Lucia, R Paulen, and S Engell. Dual robust nonlinear model predictive control: A multi-stage approach. *Journal of Process Control*, 2018.

[18] Elena Arcari, Lukas Hewing, Max Schlichting, and Melanie N Zeilinger. Dual stochastic mpc for systems with parametric and structural uncertainty. *Learning for Dynamics and Control (L4DC)*, 2020.

[19] Björn Wittenmark. Adaptive dual control methods: An overview. *Adaptive Systems in Control and Signal Processing*, 1995.

[20] Eric C Kerrigan and Jan M Maciejowski. Soft constraints and exact penalty functions in model predictive control. *UKACC Control Conference*, 2000.

[21] Masahiro Ono. Closed-loop chance-constrained mpc with probabilistic resolvability. In *IEEE Conference on Decision and Control (CDC)*, pages 2611–2618. IEEE, 2012.

[22] Jack Ridderhof, Kazuhide Okamoto, and Panagiotis Tsiotras. Nonlinear uncertainty control with iterative covariance steering. In *IEEE Conference on Decision and Control (CDC)*, pages 3484–3490. IEEE, 2019.

[23] Quoc Tran Dinh and Moritz Diehl. Local convergence of sequential convex programming for nonconvex optimization. In *Recent Advances in Optimization and its Applications in Engineering*, pages 93–102. Springer, 2010.

[24] Riccardo Bonalli, Abhishek Cauligi, Andrew Bylard, and Marco Pavone. Gusto: Guaranteed sequential trajectory optimization via sequential convex programming. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 6741–6747. IEEE, 2019.

[25] Michael L Littman, Anthony R Cassandra, and Leslie Pack Kaelbling. Learning policies for partially observable environments: Scaling up. In *International Conference on Machine Learning (ICML)*. 1995.

[26] Robert Dyro, James Harrison, Apoorva Sharma, and Marco Pavone. Particle MPC for uncertain and learning-based control (extended version). 2020. http://asl.stanford.edu/wp-content/papercite-data/pdf/Dyro.Harrison.ea.IROS21.pdf.

[27] AA Feldbaum. Dual control theory. I. *Avtomatika i Telemekhanika*, 1960.

[28] Dimitri P Bertsekas. *Dynamic programming and optimal control*. Number 1. 4 edition, 2012.

[29] Boris Ivanovic, James Harrison, Apoorva Sharma, Mo Chen, and Marco Pavone. BARC: Backward reachability curriculum for robotic reinforcement learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

[30] Jeremy H Gillula, Haomiao Huang, Michael P Vitus, and Claire J Tomlin. Design of guaranteed safe maneuvers using reachable sets: Autonomous quadrotor aerobatics in theory and practice. *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[31] Sumeet Singh, Anirudha Majumdar, Jean-Jacques Slotine, and Marco Pavone. Robust online motion planning via contraction theory and convex optimization. *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[32] Thomas Lew, Apoorva Sharma, James Harrison, and Marco Pavone. Safe model-based meta-reinforcement learning: A sequential exploration-exploitation framework. *arXiv:2008.11700*, 2020.

[33] Bartolomeo Stellato, Goran Banjac, Paul Goulart, Alberto Bemporad, and Stephen Boyd. OSQP: An operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 2020.

[34] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*. MIT Press, 2005.

[35] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Neural Information Processing Systems (NeurIPS)*, 2017.

[36] Ian Osband, John Aslanides, and Albin Cassirer. Randomized prior functions for deep reinforcement learning. In *Neural Information Processing Systems (NeurIPS)*, 2018.

[37] Tim Pearce, Felix Leibfried, and Alexandra Brintrup. Uncertainty in neural networks: Approximately bayesian ensembling. In *Artificial Intelligence and Statistics (AISTATS)*, 2020.

[38] Matt Wytock, Nicholas Moehle, and Stephen Boyd. Dynamic energy management with scenario-based robust MPC. In *American Control Conference (ACC)*, 2017.

[39] Kento Shimada and Takeshi Nishida. Particle filter-model predictive control of quadcopters. In *International Conference on Advanced Mechatronic Systems*, 2014.

[40] Martin A. Sehr and Robert R. Bitmead. Particle model predictive control: Tractable stochastic nonlinear output-feedback MPC. *IFAC*, 2017.

[41] Dominik Stahl and Jan Hauth. PF-MPC: Particle filter-model predictive control. *Systems & Control Letters*, 2011.

[42] Johan Pieter De Villiers, S. J. Godsill, and S. S. Singh. Particle predictive control. *Journal of statistical planning and inference*, 141(5), 2011.

[43] Kazuhide Okamoto, Maxim Goldshtein, and Panagiotis Tsiotras. Optimal covariance control for stochastic systems under chance constraints. *IEEE Control Systems Letters*, 2(2):266–271, 2018.

A. Proof of Theorem 1

We restate theorem 1 for completeness:

Theorem 1 (Consensus horizon & replanning): For any integers $N^\star \in [2..\infty)$, $K \in [1..N^\star - 1]$, there exists an optimal control problem (1) defined by $\mathcal{X}_0, f, c, \mathbb{X}, \mathbb{U}, \mathbb{W}$ for which running PMPC with a consensus horizon of $N_c = K < N^\star$ leads to constraint violation, but a consensus horizon of $N_c = N^\star$ does not.

Proof: Define an uncertain bimodal system f

$$x^{(0)} \sim \mathcal{X}_0 = \{0\}$$
$$\mathbb{X} = \{0, 1\}$$
$$\mathbb{U} = \{0, 1\}$$
$$\mathbb{W} = \{0, 1\}$$

where transitions $x^{(j+1)} = f(x^{(j)}, u^{(j)}, w^{(j)})$ and the state dependent cost are

$$f(x^{(j)}, u^{(j)}, w^{(j)}) = \begin{cases} \begin{cases} 0 \text{ if } u^{(j)} = 0 \\ 1 \text{ if } u^{(j)} = 1 \end{cases} & \text{if } x^{(j)} = 0 \\ \begin{cases} 1 \text{ if } u^{(j)} = w^{(j)} \\ 2 \text{ if } u^{(j)} = \neg w^{(j)} \end{cases} & \text{if } x^{(j)} = 1 \\ \begin{cases} 2 \text{ if } u^{(j)} = 0 \\ 2 \text{ if } u^{(j)} = 1 \end{cases} & \text{if } x^{(j)} = 2 \end{cases}$$

$$c(x^{(j)}) = \begin{cases} 1 & \text{if } x^{(j)} = 0 \\ 0 & \text{if } x^{(j)} = 1 \\ \infty \text{ (constraint violation)} & \text{if } x^{(j)} = 2 \end{cases}$$

Note that $x^{(j)} = 2 \notin \mathbb{X}$ implies a constraint violation.

For the bimodal parameter uncertainty distribution between w_1 or w_2

$$w_1^{(j)} = \begin{cases} 0 & \text{if } j < N^\star \\ j \bmod 2 & \text{if } j \geq N^\star \end{cases}$$
$$w_2^{(j)} = \begin{cases} 0 & \text{if } j < N^\star \\ (j+1) \bmod 2 & \text{if } j \geq N^\star \end{cases}$$

For receding planning at time $j = 0$, the control consensus of $N^\star$ suggests a control policy with $u^{(j)} = 0$ which *does not* lead to constraint violation.

Control consensus of $K < N^\star$ suggests a policy with $u^{(0)} = 1$ which leads to constraint violation, because too fast information gain is assumed, whereas in reality the system cannot be stabilized at $x^{(j)} = 1$.

This shows that for any integers $N^\star \in [2..\infty)$, $K \in [1..N^\star - 1]$, there exists an optimal control problem (1) with uncertainty such that running PMPC with consensus horizon K leads to constraint violation, but a consensus horizon of $N^\star$ correctly accounts for uncertainty, avoiding constraint violation.

∎

(a) Free-flyer, state uncertainty   (b) Free-flyer, model uncertainty

(c) Planar quadrotor, state uncertainty   (d) Planar quadrotor, model uncertainty
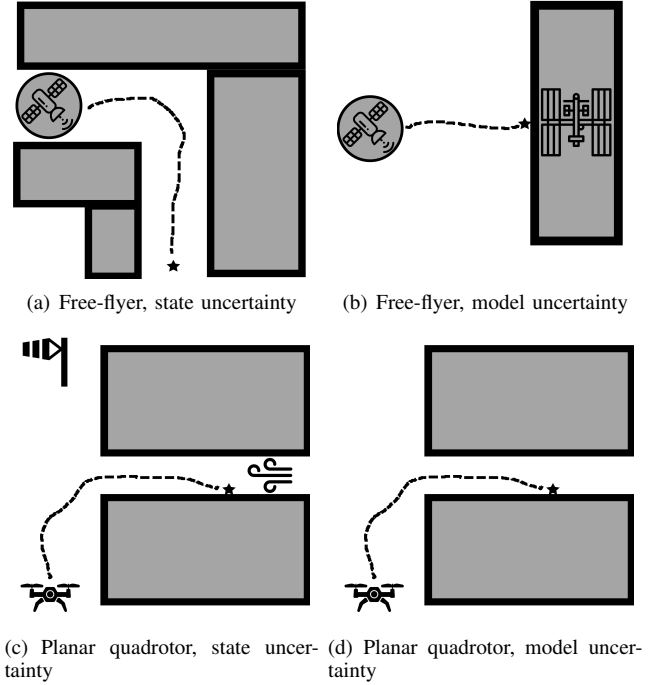
Fig. 5.   Experimental setup sketch

B. Experimental Details

1) Obstacles Implementation: In all experiments we enforce no obstacle violation via exact penalty functions in the objective. Such penalty specification has the property that it is enforced exactly for a sufficiently large penalty weight. All obstacles considered in the experiments are rectangles and, for the penalty function, we pick the distance to the nearest edge multiplied by a positive weight. We choose penalty approach instead of enforcing hard constraints to ensure plan feasibility: (i) during SCP convergence; (ii) in particle consensus planning, where a large number of different dynamics can include plans that cannot avoid obstacles. We choose the exact penalty approach because it results in the hard constraint solution at optimality and when final plan feasibility is possible.

2) MPPI Implementation: We implement MPPI following [13] where we take the weighting over control samples and sampled models. Since we use a consensus horizon shorter than the planning horizon, unlike [13], we weigh actions *separately for each model* past the consensus horizon. We attempted to automatically tune the MPPI parameters of (i) the number of iterations per time step, (ii) the number of random action sequences per iteration, (iii) the standard deviation of the normal sampling, (iv) the temperature of the algorithm. In both environments we choose these parameters by grid search to establish those which produced the lowest cost trajectories.

3) Dynamics Uncertainty:

a) Free-flyer.: We consider a setting in which the free-flyer must move to a location near a wall, but at each timestep, any of its four thrusters can fail with a probability of 5%.

Once an actuator fails, we assume it never recovers. In this setting, there are $2^4 = 16$ possible dynamics models corresponding to every combination of working and failed thrusters. Furthermore, dynamics are nonstationary, as we can probabilistically transition from working states to failure states over the course of time.

We assume the actuator failure is not directly observable and use a recursive Bayes filter to detect actuator failure. This filter maintains a belief over each of the 16 possible states of the system. For PMPC, each particle must represent one (time-varying) dynamics model. To obtain each particle i, we simulate the evolution of dynamics over time according to the probability of failure at each time step from the current state to obtain a time-varying dynamics function for each particle. We repeat this multiple times for each hypothesis in the particle filter to obtain M particles for PMPC. Each particle is weighted by the weight assigned to the root state by the particle filter. We compare to a certainty equivalent nonlinear MPC controller, in which we simulate actuator failure forward in time from the current state, and then select the maximum a posteriori (MAP) estimate for each time step. For the sake of comparison, we consider an oracle model which plans with access to the true dynamics evolution. All scenarios use a planning horizon $N = 20$, and PMPC uses a consensus horizon of $N_{\text{consensus}} = 7$ which was found to perform best.

The starting position is at $(x = -20, y = 0)$ and zero translational and angular velocity. The wall occupies $x > 0.0$ and its exact penalty weight was 10^2.

b) Planar quadrotor.: For the planar quadrotor setting, we consider wind disturbances that evolve according to an Ornstein-Uhlenbeck (OU) process independently in the x and y directions:

$$v_{\text{wind}}^{(j+1)} = \alpha v_{\text{wind}}^{(j)} + \epsilon \qquad \alpha = 0.9 \qquad \epsilon \sim \mathcal{N}(0, 2.0).$$

The goal is to fly the quadrotor into a narrow passageway, defined by $(x > 0.0, y \in [10, 13])$, and regulate to the lower wall at $(x = 15, y = 10)$ in the presence of these noisy, changing wind disturbances.

Here we assume the current wind is fully observed, so we do not require a particle filter. However, there is uncertainty over the future evolution of the wind, which we incorporate into PMPC. To sample the particles, we sample M trajectories of the OU process. Each of these define a particle of time-varying dynamics $\{f_i\}_{j=1}^N$. For a CE baseline, we consider planning assuming the current wind stays constant. Again, we compare to an oracle which knows the exact evolution of wind dynamics. We use a planning horizon of $N = 20$, $M = 10$ particles, and vary $N_c \in \{1, 5, 10\}$.

In all quadrotor settings the passageway is implemented by two obstacles, one above and one below with a penalty weight of 10^3. The starting position is at zero translation and angular velocity at $(x = -3, y = 12)$, to the left and slightly below the entrance to the passageway.

4) State Uncertainty:

a) Free-flyer.: For the free-flyer, the task is to navigate around a corner in a narrow corridor. We use a planning horizon of $N = 20$, $M = 10$ particles, and a consensus

horizon of $N_c = 5$ in PMPC. We see that PMPC in this setting does better than CE, leading to a lower chance of collision and overall lower cost.

We consider uncertainty only in the 2D position. To do so, at the beginning of each episode, we sample a fixed point cloud of position deviations with a standard deviation of 3 m; we maintain the same deviation points throughout the episode. The true position has zero deviation and is included as one of the particles in the point cloud.

Consequently, we use recursive Bayes filtering to estimate the true position. The free-flyer is equipped with 4 distance sensors along its two local axes, which obtain a noisy measurement of position with standard deviation of 3 m. In the filter, we assume a larger standard deviation of the observation noise of 10 m both because using the true observation noise led to near instantaneous adaptation and because the observation noise is often not known in practice. Using smaller than true observation noise can lead to spurious convergence to the wrong particle estimate and is not advisable.

In all free-flyer experiments the starting position is at $(x = 0, y = 0)$ and zero translational and angular velocity. The free space in the corridor spans $y = \in [-5, 5]$ meters vertically for $x < 20$ and $x \in [10, 20]$ for $y < 5$ to form an L-shaped corridor. The goal state is at $(x = 15, y = -20$; in the middle of the vertical corridor part. There are two obstacles outer to the L-turn with a penalty weight of 10^2 and one inner to the turn with a weight of 10^3.

b) Planar quadrotor.: For the planar quadrotor, the goal is to fly the quadrotor into a narrow passageway, defined by $(x > 0.0, y \in [10, 13])$, and regulate to the lower wall at $(x = 15, y = 10)$ in the presence of only position uncertainty. To do so, at the beginning of each episode, we sample a fixed point cloud of position deviations with a standard deviation of 1 m; we maintain the same deviation points throughout the episode. The true position has zero deviation and is included as one of the particles in the point cloud.

Consequently, we use a recursive Bayes filter to determine the true deviation. The free-flyer is equipped with 4 distance sensors along its two local axes, which obtain a noisy measurement of position with a standard deviation of 1 m. In the filter, we use the true observation noise. The adaptation with true noise is much slower than for the free-flyer in a corridor, likely because the quadrotor initially only has a wall to its right, unlike the free-flyer which typically registers walls on all of its sides and so has more measurements available.

In all quadrotor settings the passageway is implemented by two obstacles, one above and one below with a penalty weight of 10^3. The starting position is at zero translation and angular velocity at $(x = -3, y = 5)$, to the left and below the entrance to the passageway.

5) Learning for Control: For both environments we learn by fitting an ensemble network model of the dynamics after each episode—we train D neural networks the weights of each are regularized to a random initialization using weighted squared ℓ_2 distance to capture epistemic uncertainty. For the

CE case we train a single un-regularized neural network with the same architecture. We use a fixed learning rate of 10^{-3} and train until convergence on the noise-free dynamical transitions data obtained in all previous episodes.

a) Free-flyer.: We test the free-flyer in the same corridor navigation task as in VI-B.4.a. We use a 3-layer feedforward neural network model with 3 hidden layers of width 64. For the ensemble network we regularize the weights by 0.001 divided by the output dimension; we do this to capture epistemic uncertainty. The angle component of the input is processed into $(\sin(\theta), \cos(\theta))$ before being fed to the network to prevent wraparound issues. We use $M = 10$ particles, a planning horizon of $N = 20$, a consensus horizon of $N_{\text{consensus}} = 10$. We introduce normal control noise to all control strategies for the first 6 episodes to induce exploration, with a decaying sequence of standard deviations $(0.3, 0.3, 0.1, 0.1, 0.01, 0.01)$. The starting and goal position and the obstacles are identical to the one in VI-B.4.a. For the MPPI implementation we use 50 iterations per timestep, 60 action sequences, a standard deviation for sampling of 0.3 and a temperature of 10^{-3}.

b) Planar quadrotor.: The planar quadrotor is tested in the same narrow passage way task, but here without wind. The architecture of the network is identical to that for the free-flyer except of the width of 32. As learning happens faster for this environment, we found we only needed to add exploration control noise to the first 3 episodes with a standard deviation sequence of $(0.3, 0.1, 0.01)$. All controller parameters are identical to the free-flyer setup, except here we use $N_{\text{c}} = 1$.

In all quadrotor settings the passageway is implemented by two obstacles, one above and one below with a penalty weight of 10^3. The starting position is at zero translation and angular velocity at $(x = 0, y = 0)$, slightly to the left and significantly below the entrance to the passageway. For the MPPI implementation we use 30 iterations per timestep, 100 action sequences, a standard deviation for sampling of 0.2 and a temperature of 10^{-5}.