

# Adaptive Robust Model Predictive Control with Matched and Unmatched Uncertainty

Rohan Sinha, James Harrison, Spencer M. Richards, and Marco Pavone

**Abstract**—We propose a learning-based robust predictive control algorithm that compensates for significant uncertainty in the dynamics for a class of discrete-time systems that are nominally linear with an additive nonlinear component. Such systems commonly model the nonlinear effects of an unknown environment on a nominal system. We optimize over a class of nonlinear feedback policies inspired by certainty equivalent “estimate-and-cancel” control laws pioneered in classical adaptive control to achieve significant performance improvements in the presence of uncertainties of large magnitude, a setting in which existing learning-based predictive control algorithms often struggle to guarantee safety. In contrast to previous work in robust adaptive MPC, our approach allows us to take advantage of structure (i.e., the numerical predictions) in the a priori unknown dynamics learned online through function approximation. Our approach also extends typical nonlinear adaptive control methods to systems with state and input constraints even when we cannot directly cancel the additive uncertain function from the dynamics. Moreover, we apply contemporary statistical estimation techniques to certify the system’s safety through persistent constraint satisfaction with high probability. Finally, we show in simulation that our method can accommodate more significant unknown dynamics terms than existing methods.

## I. INTRODUCTION

Learning-based control offers promising methods to enable the deployment of autonomous systems in diverse, dynamic environments. Such methods learn from data to improve closed-loop performance over time. Upon deployment, these methods should provide safety guarantees and quickly adapt in the face of uncertainty; to this end, estimates of uncertainties in learned quantities must be maintained and updated as new data becomes available to ensure safety constraint satisfaction. However, many recently proposed learning-based control algorithms rely on uncertainty estimation methods that result in policies that are either too conservative (e.g., yielding limited performance to remain safe) or too fragile (e.g., infeasibility in the face of large uncertainties). In this work, we combine a simple nonlinear control law inspired by “estimate-and-cancel” methods in nonlinear adaptive control with robust model predictive control (MPC) techniques to control a system in an uncertain environment, represented using an unknown, nonlinear term in the dynamics. Our simulated examples show that this approach can both reduce the conservatism and fragility of existing methods.

**Related Work.** We briefly review two significant paradigms for the control of uncertain systems, namely *adaptive control* and *robust control*. We then discuss recent

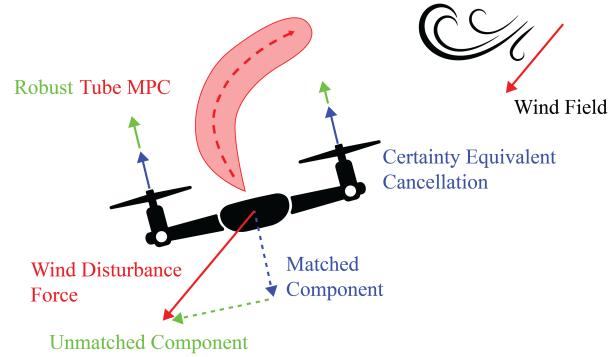


Fig. 1. Illustration of our high-level approach: We learn the effect of environmental uncertainty online and curb its influence through certainty equivalent adaptive control. Then, we account for any remaining uncertainty and guarantee constraint satisfaction using tube MPC.

works that combine ideas from both paradigms, oftentimes leveraging modern methods in machine learning.

*Adaptive control* concerns the joint design of a parametric feedback controller and a parameter adaptation law to improve closed-loop performance over time when the dynamics are (partially) unknown [1], [2]. Design of these components for nonlinear systems commonly relies on expressing unknown dynamics terms as linear combinations of *known basis functions*, i.e., *features* [1]. The adaptation law updates the feature weights online, and the controller applies part of the control signal towards cancelling the estimated term from the dynamics [1], [3]–[5]. These simple methods can achieve tracking convergence up to an error threshold that depends on the representation capacity of the features relative to the true dynamics [3], [4]. Recent works proposed combining high-capacity parametric and nonparametric models from machine learning with classical adaptive control designs. This includes deep neural networks via online back-propagation [6], Gaussian processes [7], and Bayesian neural networks [8], [9] via online Bayesian updates and meta-learned features [10], [11]. However, these approaches are fundamentally limited by common assumptions in classical adaptive control, namely that uncertain dynamics terms can be stably canceled by the control input in their entirety, i.e., that these terms are *matched uncertainties* [1]–[4]. Moreover, most of these works do not consider state and input constraints, essential to safe control in practice. We generalize these classical adaptive methods to incorporate safety constraints even if the uncertainty is not fully matched.

*Robust control* seeks consistent performance despite un-

The authors are with Stanford University, Stanford, CA. {rhnsinha, jharrison, spenrich, pavone}@stanford.edu

certainty in the dynamics. In this work, we consider the robust control of constrained discrete-time systems using tools from predictive control. In particular, robust MPC algorithms for linear systems consider the control of a system subject to bounded noise or uncertain dynamics terms, i.e. *disturbances*, as an optimization program with explicit state and input constraints. Some methods optimize the worst-case performance of the controller [12], while others tighten the constraints to accommodate the set of all possible trajectories induced by the disturbances and optimize the nominal predicted trajectory instead [13], [14]. To account for future information gain and reduce conservatism, these methods either fix a disturbance feedback policy [13] or optimize over state feedback policies [15].

*Adaptive robust MPC (ARMPC)*, often referred to as learning-based MPC, incorporates the online estimation (i.e., learning) of adaptive control methods into robust MPC to satisfy constraints in the presence of process noise and model uncertainty during learning. Recent years have seen a flurry of work on nonlinear predictive control methods that apply contemporary machine learning techniques to learn uncertain dynamics online [16]–[19]. These methods typically result in non-convex programs for trajectory optimization under the learned dynamics, while relying on conservative approximate methods for uncertainty propagation to guarantee constraint satisfaction. However, it is unclear how to construct the necessary components—the robust positive invariant and the terminal cost function—within predictive control to make claims of persistent constraint satisfaction (i.e., safety) or stability for arbitrary nonlinear systems. Some methods ignore these topics and do not make rigorous safety guarantees [17]. Other work, such as [18], [19], assumes these ingredients already exist, or considers only trajectory optimization tasks where a goal region needs to be reached in a finite number of timesteps [16]. Moreover, iterative methods used to solve for local minima of non-convex programs can be prohibitively computationally expensive and often have limited associated performance guarantees.

To make rigorous safety guarantees, we will focus on robust, adaptive methods for systems that are nominally linear as considered in [20]–[26]. A straightforward approach is to maintain an outer bound on an unknown, nonlinear term in the dynamics and use it as a disturbance bound in any chosen robust MPC scheme [20]–[23], [26]. These methods avoid some of the difficulties associated with trajectory optimization for nonlinear dynamics by ignoring the actual values of the nonlinear terms at any point in the state space. That is, these methods do not exploit the learned structure in the a priori unknown dynamics, often rendering them over-conservative or fragile.

**Contributions.** We present an ARMPC method for systems with an additive unknown nonlinear dynamics term, subject to state and input constraints. Rather than construct an outer envelope for such terms, as is normative in ARMPC literature for linear systems, we develop theoretical guarantees for a broad class of *function approximators*, including set membership and least-squares methods for certain noise

models. Our key idea is to decompose uncertain dynamics terms into a *matched* component that lies in a subspace that can be stably canceled by the control input, and an *unmatched* component that lies in an orthogonal complement to this subspace. We apply certainty equivalent adaptive control techniques to stably cancel the *matched* component from the dynamics and then apply robust MPC, considering the unmatched component as a bounded disturbance. Therefore, our method explicitly uses point estimates of the unknown term throughout the state space for control, i.e., it takes advantage of the learned structure in the dynamics.

Our approach can be viewed through the lens of both adaptive control and robust adaptive MPC: on one hand, we extend classical adaptive cancellation-based methods to a setting with uncertain, unmatched dynamics subject to state and input constraints. On the other hand, we introduce a simple nonlinear feedback law to construct an adaptive, robust MPC that reduces the conservatism of existing approaches by taking advantage of the learned structure in a priori unknown dynamics. We prove our method is recursively feasible and input-to-state stable. Moreover, we demonstrate on various simulated systems that our method reduces the conservatism and increases the feasible domain of the resulting robust MPC problem compared to typical adaptive robust MPC methods.

## II. PROBLEM FORMULATION

We consider the robust control of nonlinear discrete time systems of the form

$$x(t+1) = Ax(t) + Bu(t) + f(x(t)) + v(t), \quad (1)$$

where  $x(t) \in \mathbb{R}^n$  is the system state,  $u(t) \in \mathbb{R}^m$  is the control input,  $A \in \mathbb{R}^{n \times n}$  and  $B \in \mathbb{R}^{n \times m}$  are known constant matrices, and  $v(t) \in \mathcal{V}$  is a disturbance in a known convex, compact set  $\mathcal{V}$  containing the origin. In addition, an *unknown, nonlinear* dynamics term  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  acts on the system, representing the unmodelled influence of the environment on the nominally linear dynamics of system (1). For example,  $f(x)$  can model the effect that wind conditions have on the linearized dynamics of a drone. We assume the disturbances are zero mean and independent and identically distributed (i.i.d.), i.e.,  $v(t) \stackrel{\text{iid}}{\sim} p(v)$  and  $\mathbb{E}[v(t)] = 0$  for all  $t \geq 0$ . Our goal is to regulate the system close to zero according to the robust optimal control problem

$$\begin{aligned} & \underset{x,u}{\text{minimize}} \mathbb{E} \left[ \sum_{t=0}^{\infty} h(x(t), u(t)) \right] \\ & \text{subject to } x(t+1) = Ax(t) + Bu(t) + f(x(t)) + v(t) \\ & \quad u(t) \in \mathcal{U}, x(t) \in \mathcal{X}, v(t) \in \mathcal{V}, \quad \forall t \in \mathbb{N}_{\geq 0} \end{aligned} \quad (2)$$

where  $\mathcal{X} \subseteq \mathbb{R}^n$  and  $\mathcal{U} \subseteq \mathbb{R}^m$  are compact convex sets containing the origin, and  $h(x, u) = x^\top Qx + u^\top Ru$  is a quadratic stage cost parameterized by positive definite matrices  $Q \in \mathbb{S}_{>0}^n$  and  $R \in \mathbb{S}_{>0}^m$ . The problem (2) is computationally intractable to solve because the horizon is infinite and the nonlinear function  $f$  makes the problem non-convex. To approximately solve (2), we need additional

assumptions on the unknown, nonlinear dynamics term  $f$ . In particular, to derive a controller that is robust to any possible value of  $f(x)$ , we need  $f$  to be bounded on  $\mathcal{X}$ . Moreover, to construct guarantees on the online estimation of  $f$  and establish properties of a controller using this estimate, we also need to assume some structure of  $f$ . For these reasons, we make the following assumption.

**Assumption 1 (structure):** The nonlinear dynamics term  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is linearly parameterizable, i.e.,

$$f(x) = W\phi(x), \quad \forall x \in \mathbb{R}^n, \quad (3)$$

where  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^d$  is a *known* nonlinear feature map, and  $W \in \mathbb{R}^{n \times d}$  is an *unknown* weight matrix. Moreover,  $\|\phi(x)\| \leq 1$  for any  $x \in \mathcal{X}$ , where  $\|\cdot\|$  is the Euclidean norm.

Representing a nonlinear function using a feature map is common both in adaptive control [1], [2] and contemporary machine learning [9], [27], as they can represent arbitrary functions if properly designed. For simplicity, we assume the features satisfy a unit norm bound without loss of generality; this still allows for function classes like neural networks with scaled sigmoid outputs.

**Matched and Unmatched Uncertainty:** While it is common in adaptive control to assume the uncertain function  $f$  in (1) can be stably cancelled in its entirety [1], [4], we will generalize this approach to a setting where perfect cancellation is not possible. To distinguish between the components of the uncertain dynamics  $f$  that can and cannot be cancelled, we classify the dynamic uncertainty as follows:

**Definition 1 (matched and unmatched uncertainty):** The uncertain function  $f(x)$  in (1) is a *matched uncertainty* if  $f(x) \in \text{Range}(B)$  for all  $x \in \mathcal{X}$ . Conversely, if there exists an  $x \in \mathcal{X}$  such that  $f(x) \notin \text{Range}(B)$ , then  $f(x)$  is an *unmatched uncertainty*.

In this work, we assume that the  $B$  matrix in (1) has full column rank, i.e., there are no redundant actuators, thereby guaranteeing the existence of the Moore-Penrose pseudoinverse  $B^\dagger = (B^\top B)^{-1}B^\top$ . Therefore, if  $f(x)$  is a matched uncertainty, then the function  $g(x) = B^\dagger f(x)$  satisfies  $f(x) = Bg(x)$  for any  $x \in \mathcal{X}$ .

Controlling systems with matched uncertainty is a classical problem in the adaptive control literature, much of which relies on the observation that setting  $u(t) = \bar{u}(t) - g(x(t))$  in (1) would cancel the nonlinear term to yield linear dynamics with respect to the nominal input  $\bar{u}(t)$ . *Certainty equivalent* controllers that approximately cancel  $g(x)$  with an estimate  $\hat{g}(x)$  result in simple nonlinear adaptive laws that can achieve asymptotic tracking performance for matched systems. Even though systems are often designed to be easy to control, unmatched uncertainty affects many practical systems of interest, such as the dynamics of any underactuated robotic system (e.g., quadrotors and cars). We propose to decompose the uncertain function  $f$  into a matched and unmatched component, apply certainty equivalent cancellation to the matched component, and curb the impact of the unmatched component using robust MPC strategies. Applying part of the input to cancel matched uncertainty allows us to instantaneously prevent some components of  $f$  from leaking

into the dynamics, avoiding the need to react to large observed disturbances.

**ISS Stability.** Due to the disturbance  $v(t)$ , the system (1) cannot be regulated exactly to the origin even asymptotically. Therefore, robust control algorithms are typically analyzed using Input-to-State Stability (ISS) theory to establish more appropriate stability properties [14], [15], [22]. We briefly review relevant results in ISS theory for time-varying systems, since online adaptation of an estimate of the unknown function  $f$  results in a time-varying closed loop system.

**Definition 2 (ISS stability [28]):** The system  $x(t+1) = q(t, x(t), v(t))$  with disturbance  $v(t)$  is globally *Input-to-State Stable (ISS)* if there exist a class- $\mathcal{KL}$  function  $\beta : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and a class- $\mathcal{K}$  function  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that

$$\|x(t)\| \leq \beta(\|x(0)\|, t) + \gamma(\sup_{k \in \{0, \dots, t-1\}} \|v(k)\|), \quad (4)$$

for all  $x(0) \in \mathbb{R}^n$  and  $t \in \mathbb{N}_{\geq 0}$ .

In essence, ISS requires that the nominal system is asymptotically stable and the influence of the disturbances is bounded. This makes it a convenient framework to analyze the stability of systems subject to random disturbances. Similarly to regular nonlinear stability analysis, we can show a system is ISS if there exists an ISS-Lyapunov function.

**Definition 3 (ISS-Lyapunov function):** The function  $V : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$  is an *ISS-Lyapunov function* for the system  $x(t+1) = q(t, x(t), v(t))$  if it is continuous in  $x$ , continuous at the origin for all  $t \in \mathbb{N}_{\geq 0}$ , and there exist class- $\mathcal{K}_\infty$  functions  $\alpha_1, \alpha_2, \alpha_3$  and a class- $\mathcal{K}$  function  $\sigma$  such that

$$\begin{aligned} \alpha_1(\|x(t)\|) &\leq V(t, x(t)) \leq \alpha_2(\|x(t)\|) \\ V(t+1, x(t+1)) - V(t, x(t)) &\leq -\alpha_3(\|x(t)\|) + \sigma(\|v(t)\|) \end{aligned} \quad (5)$$

for all  $x(t) \in \mathbb{R}^n$ .

**Theorem 1 ([28]):** A time-varying system is globally ISS if it admits an ISS-Lyapunov function.

The above definitions naturally extend to local ISS stability; for a detailed discussion, we refer readers to [14], [28], [29].

### III. ADAPTIVE ROBUST MPC

In this section we first describe assumptions on and necessary features of the learning procedure in a way that is agnostic to the choice of learning algorithm. We then introduce our adaptive robust MPC approach, and prove stability of the combined learning and control framework.

**Learning Desiderata.** Since the nonlinear dynamics term  $f$  is unknown, our method takes a *certainty equivalent* approach by substituting an estimate  $\hat{f}$  that is refined online as more data becomes available. To guarantee the adaptive robust MPC framework satisfies state and input constraints for all time (i.e., safety), we make several assumptions on  $\hat{f}$ .

We maintain the estimate

$$\hat{f}(x, t) = \widehat{W}(t)\phi(x) \quad (6)$$

of  $f(x)$ , where  $\widehat{W}(t) \in \mathbb{R}^{n \times d}$  is our estimate of  $W$  at time  $t$ . To this end, we need bounds on our initial uncertainty, i.e., the difference between  $f(x)$  and  $\hat{f}(x, 0)$  for all  $x$ . For a general

statistical estimator, this entails specifying a risk tolerance  $\delta \in (0, 1)$  and computing confidence intervals on the estimate.

*Assumption 2 (prior knowledge):* Let  $w_i, \hat{w}_i(t)$ , and  $\tilde{w}_i(t)$  be the  $i$ -th rows of  $W$ ,  $\widehat{W}(t)$ , and  $\widetilde{W}(t) := \widehat{W}(t) - W$ , respectively, for  $i \in \{1, 2, \dots, n\}$ . At  $t = 0$ , we know an initial estimate  $\widehat{W}(0) \in \mathbb{R}^{n \times d}$  and bounded sets  $\{\mathcal{W}_i(0)\}_{i=1}^n$  with  $\mathcal{W}_i(0) \subset \mathbb{R}^d$ , such that  $\widetilde{W}(0) \in \mathcal{W}(0)$  with probability at least  $1 - \delta$ , where

$$\mathcal{W}(t) := \{\widetilde{W} \in \mathbb{R}^{n \times d} \mid \tilde{w}_i \in \mathcal{W}_i(t), \forall i \in \{1, \dots, n\}\}, \quad (7)$$

for all  $t \in \mathbb{N}_{\geq 0}$ .

*Assumption 2* provides only an initial bound on the error of the estimate, which we explicitly label as the estimate at  $t = 0$ . Later, we will define  $\mathcal{W}(t)$  for all  $t \in \mathbb{N}_{\geq 0}$  when we adaptively update our estimate  $\widehat{W}(t)$  and the bounds  $\{\mathcal{W}_i(t)\}_{i=1}^n$  online. Our approach leverages the certainty equivalent “estimate and cancel” control laws pioneered in classical unconstrained adaptive control [1], [4]. As such, *Assumption 1* and *Assumption 2* are necessary to bound our approximation error. However, since we specify the risk tolerance  $\delta$ , we cannot guarantee exact constraint satisfaction for all time. We instead relax our definition of safety to

$$\text{Prob}(x(t) \in \mathcal{X}, u(t) \in \mathcal{U}, \forall t \geq 0) \geq 1 - \delta, \quad (8)$$

which states that the probability of a constraint violation should be no more than  $\delta$  over the entire realized trajectory. Moreover, to guarantee closed loop safety, we assume we have an online adaptation strategy that ensures the quality of the estimate  $\widehat{W}(t)$  cannot get worse over time.

*Assumption 3 (online learning):* We have an *online parameter estimator* that maps an initial estimate  $\widehat{W}(0)$ , the associated  $1 - \delta$  confidence interval  $\mathcal{W}(0)$ , and the trajectory history  $\{x(k), u(k)\}_{k=0}^t$  to an *online estimate*  $\widehat{W}(t)$  and confidence interval  $\mathcal{W}(t)$  at time  $t$ , such that  $\hat{w}_i(t) - w_i \in \mathcal{W}_i(t)$  for all time  $t \in \mathbb{N}_{\geq 0}$  and  $i \in \{1, 2, \dots, n\}$  with probability at least  $1 - \delta$ . We assume the confidence intervals on  $\widehat{W}(t)$  are not growing with time, i.e.,

$$\mathcal{W}(t+1) \subseteq \mathcal{W}(t), \quad (9)$$

for all  $t \in \mathbb{N}_{\geq 0}$ .

Formulating separate confidence intervals for each row of  $\widehat{W}(t)$  is a natural approach, as fitting  $\widehat{W}(t)$  to historical data decomposes into  $n$  separate least-squares problems (one for each row). Crucially, *Assumption 3* allows us to treat the confidence intervals  $\mathcal{W}(t)$  as exact bounds in the control design, since a controller that guarantees constraint satisfaction conditioned on the event that  $\hat{w}_i(t) - w_i \in \mathcal{W}_i(t)$  for all time then satisfies (8). Therefore, we treat the chance constraint (8) as a proxy for robust constraint satisfaction and construct our approach for the remainder of §III conditioned on the event that  $\hat{w}_i(t) - w_i \in \mathcal{W}_i(t)$  for all time. This approach was also taken in [16], [30]. We discuss two commonplace estimators that satisfy our desiderata in §IV.

*Assumption 3* does not require the range of the estimated function to shrink over time. Therefore, our estimation procedure differs from methods such as [20], [22], [23] that

refine a shrinking bound exclusively on the range of the estimate without taking direct advantage of the structure in the nonlinear dynamics.

#### A. Certainty Equivalent Cancellation

We propose optimizing over feedback policies that cancel as much of the nonlinear term  $f(x)$  as possible.

*Definition 4:* The set of *matching Certainty Equivalent (CE) policies* is the time-varying function class whose elements  $\pi : \mathcal{X} \times \mathbb{N}_{\geq 0} \rightarrow \mathcal{U}$  are of the form

$$\pi(x(t), t) = u^*(x(t), t) - B^\dagger \hat{f}(x(t), t) \quad (10)$$

The matching CE policies simply project  $\hat{f}(x(t), t)$  onto  $\text{Range}(B)$ , and cancel out as much of the disturbance as we can in the Euclidean norm sense, since

$$B^\dagger \hat{f}(x(t), t) = \arg \min_z \|Bz - \hat{f}(x(t), t)\|.$$

The matching CE law (10) results in the closed-loop dynamics

$$\begin{aligned} x(t+1) &= Ax(t) + B\pi(x(t), t) + f(x(t)) + v(t) \\ &= Ax(t) + Bu^*(x(t), t) + d(t) \end{aligned}, \quad (11)$$

where we define the compound disturbance term as

$$\begin{aligned} d(t) &:= v(t) + f(x(t)) - BB^\dagger \hat{f}(x(t), t) \\ &= v(t) + BB^\dagger(f(x(t)) - \hat{f}(x(t), t)) \\ &\quad + (I - BB^\dagger)f(x(t)). \end{aligned} \quad (12)$$

We have written  $d(t)$  above with three terms to highlight that it is driven by the process disturbance  $v(t)$ , the estimation error  $f(x) - \hat{f}(x)$ , and the imperfect matching using  $B^\dagger$ .

We now introduce two simple polytopic approximations to bound the support of the cancellation term in (10) and the terms that make up the compound disturbance  $d(t)$ . This will allow us to apply the certainty equivalent cancellation in (10) and handle  $d(t)$  by appropriately selecting  $u^*(\cdot)$  using robust MPC. As mentioned in our learning desiderata, we assume  $\hat{w}_i(t) - w_i \in \mathcal{W}_i(t)$  for the rest of this section.

*Lemma 1:* Consider online approximation of  $f(x)$  with features satisfying *Assumption 1*, an estimator satisfying *Assumption 3*, and define the estimated support set as

$$\mathcal{F}(t) := \{z \in \mathbb{R}^n : |z_i| \leq \|\hat{w}_i(t)\| + 2 \max_{\tilde{w}_i \in \mathcal{W}_i(t)} \|\tilde{w}_i\|\}. \quad (13)$$

Then, for all  $x \in \mathcal{X}$  and  $t, k \in \mathbb{N}_{\geq 0}$  it holds that

$$\hat{f}(x, t+k), f(x) \in \mathcal{F}(t). \quad (14)$$

*Proof:* We show  $\mathcal{F}(t)$  over-approximates the range of values  $z = \hat{f}(x, t+k)$  can take for any  $x \in \mathcal{X}$  and  $k \geq 0$ . Let  $z = \hat{f}(x, t+k) = \widehat{W}(t+k)\phi(x)$ . Then,

$$\begin{aligned} |z_i| &= |\hat{w}_i(t+k)^\top \phi(x)| \\ &\leq \|\hat{w}_i(t+k)\| \\ &\leq \|\hat{w}_i(t)\| + \|\hat{w}_i(t+k) - \hat{w}_i(t)\| \\ &\leq \|\hat{w}_i(t)\| + \|\hat{w}_i(t+k) - w_i\| + \|\hat{w}_i(t) - w_i\|. \end{aligned} \quad (15)$$

The shrinking confidence interval property from *Assumption 3* gives  $\mathcal{W}_i(t+k) \subseteq \mathcal{W}_i(t)$  for  $k \geq 0$ , so

$$|z_i| \leq \|\hat{w}_i(t)\| + 2 \max_{\tilde{w}_i \in \mathcal{W}_i(t)} \|\tilde{w}_i\|, \quad (16)$$

which proves that  $\hat{f}(x, t+k) \in \mathcal{F}(t)$  for all  $k \geq 0$ . In addition, let  $y = f(x) = W\phi(x)$  for some  $x \in \mathcal{X}$ . Then,

$$\begin{aligned} |y_i| &= |w_i^\top \phi(x)| \\ &\leq \|w_i\| \\ &\leq \|\hat{w}_i(t)\| + \|\hat{w}_i(t) - w_i\| \\ &\leq \|\hat{w}_i(t)\| + \max_{\tilde{w}_i \in \mathcal{W}_i(t)} \|\tilde{w}_i\|. \end{aligned} \quad (17)$$

Hence,  $f(x) \in \mathcal{F}(t)$ .  $\blacksquare$

The set  $\mathcal{F}(t)$  in [Lemma 1](#) contains all possible values that our online estimate can take for all future times. It is not straightforward to create a tighter approximation (i.e., eliminate the factor of 2) without additional assumptions. To see this, consider a constant unit norm ball confidence interval. In the worst case, the true parameter lies on the boundary of the ball around the current estimate. This means all future estimates may lie a Euclidean distance of 2 units away from the current estimate, yielding the bound in [Lemma 1](#).

Moreover, [Lemma 1](#) does not require that  $\mathcal{F}(t+1) \subseteq \mathcal{F}(t)$ , so we provide the following corollary to help us create an approximation that is non-increasing in size.

*Corollary 1:* At time  $t$ , the sets  $\{\mathcal{F}(i)\}_{i=0}^t$  are known, so

$$\hat{f}(x, t+k), f(x) \in \bigcap_{i=0}^t \mathcal{F}(i) =: \hat{\mathcal{F}}(t) \quad (18)$$

for all  $x \in \mathcal{X}$  and  $k \in \mathbb{N}_{\geq 0}$ , where we define  $\hat{\mathcal{F}}(t)$  as the set

$$\{z : |z_i| \leq \min_{j \in \{0, \dots, t\}} [\|\hat{w}_i(j)\| + 2 \max_{\tilde{w}_i \in \mathcal{W}_i(j)} \|\tilde{w}_i\|]\}. \quad (19)$$

Note  $\hat{\mathcal{F}}(t)$  can be computed recursively in time.

To construct a robust MPC problem to optimize the CE policy [\(10\)](#), we need to account for the compound disturbance  $d(t)$ . We do this with the following lemma.

*Lemma 2:* Assume the online parameter estimator satisfies [Assumption 3](#) with features that satisfy [Assumption 1](#) and define the approximation error support  $\mathcal{D}(t)$  as the set

$$\{z \in \mathbb{R}^n \mid |z_i| \leq \max_{\tilde{w}_i \in \mathcal{W}_i(t)} \|\tilde{w}_i\|, \forall i \in \{1, \dots, n\}\}. \quad (20)$$

If we control the system [\(1\)](#) using the certainty equivalent control law [\(10\)](#), then at time  $t$  for all  $k \in \mathbb{N}_{\geq 0}$ , the compound disturbance  $d(t+k)$  in the dynamics [\(11\)](#) is contained in the set  $\hat{\mathcal{D}}(t) \subseteq \hat{\mathcal{D}}(t-1)$ , defined as

$$\hat{\mathcal{D}}(t) := (I - BB^\dagger) \hat{\mathcal{F}}(t) \oplus BB^\dagger \mathcal{D}(t) \oplus \mathcal{V}. \quad (21)$$

Here  $\oplus$  indicates the Minkowski sum and a matrix-set multiplication indicates a linear transformation of the set's elements.

*Proof:* At any state  $x \in \mathcal{X}$  and time  $t+k$ , let  $z = \hat{f}(x, t+k) - f(x) = (\hat{W}(t+k) - W)\phi(x) = \tilde{W}\phi(x)$  for some  $\tilde{W} \in \mathcal{W}(t+k) \subseteq \mathcal{W}(t)$ . Then

$$|z_i| = |\tilde{w}_i^\top \phi(x)| \leq \|\tilde{w}_i\| \leq \max_{\tilde{w}_i \in \mathcal{W}_i(t)} \|\tilde{w}_i\|.$$

So  $\hat{f}(x, t+k) - f(x) \in \mathcal{D}(t)$  for all  $k \in \mathbb{N}_{\geq 0}$ . Since  $\mathcal{W}_i(t) \subseteq \mathcal{W}_i(t-1)$  by [Assumption 3](#), this implies  $\mathcal{D}(t) \subseteq \mathcal{D}(t-1)$ . Then, by [Corollary 1](#),  $f(x) \in \hat{\mathcal{F}}(t)$ . Therefore,  $d(t)$  in the closed loop dynamics [\(11\)](#) is contained in  $\hat{\mathcal{D}}(t)$  since  $\mathcal{D}(t)$  is symmetric. In addition, since both  $\hat{\mathcal{F}}(t) \subseteq \hat{\mathcal{F}}(t-1)$  and  $\mathcal{D}(t) \subseteq \mathcal{D}(t-1)$ , the support of  $d(t)$  is nested over time, i.e.,  $\hat{\mathcal{D}}(t) \subseteq \hat{\mathcal{D}}(t-1)$ .  $\blacksquare$

*Remark 1:* We could consider multiple variations on the bounds in [Lemmas 1, 2](#) that would yield equivalent properties of the closed loop system. For example, if a bound on the true range of values of  $f$  is known a priori, we may project  $\hat{f}$  into a known box enclosing the support of  $f$ .

Rather than search over open-loop input sequences, which can incur issues with feasibility and stability under disturbances, we consider the standard practice of searching over closed-loop feedback policies [\[13\], \[15\], \[21\]](#). In particular, we follow [\[15\]](#) in optimizing over time-varying, causal, affine disturbance feedback policies of the form

$$u_{t+k|t} = \bar{u}_{t+k|t} + \sum_{j=0}^{k-1} K_{k,j|t} d_{t+j|t}, \quad (22)$$

to define the constraint-tightened robust MPC problem that we solve online as follows:

$$\begin{aligned} &\underset{\substack{\{K_{k,j|t}\}_{k=0, j=0}^{N-1, k-1}, \\ \{\bar{u}_{t+k|t}\}_{k=0}^{N-1}}}{\text{minimize}} \quad V_N(\bar{x}_{t+N|t}) + \sum_{k=0}^{N-1} h(\bar{x}_{t+k|t}, \bar{u}_{t+k|t}) \\ &\text{subject to} \quad \bar{x}_{t+k+1|t} = A\bar{x}_{t+k|t} + B\bar{u}_{t+k|t} \\ &\quad x_{t+k+1|t} = Ax_{t+k|t} + Bu_{t+k|t} + d_{t+k|t} \\ &\quad u_{t+k|t} = \bar{u}_{t+k|t} + \sum_{j=0}^{k-1} K_{k,j|t} d_{t+j|t} \\ &\quad x_{k|t} \in \mathcal{X}, u_{t+k|t} \in \mathcal{U} \ominus B^\dagger \hat{\mathcal{F}}(t) \\ &\quad \forall k \in \{0, 1, \dots, N-1\} \\ &\quad \bar{x}_{t|t} = x(t), x_{t|t} = x(t), x_{t+N|t} \in \mathcal{O}(t) \\ &\quad \forall \{d_{t+k|t}\}_{k=0}^{N-1} \subset \hat{\mathcal{D}}(t) \end{aligned} \quad (23)$$

The problem [\(23\)](#) optimizes a time-varying feedback policy with a cost on the nominal trajectory  $(\bar{x}, \bar{u})$  subject to state and input constraints on the realized trajectory. We use the subscript  $t+k|t$  for quantities at the  $k$ -th step of the prediction horizon when [\(23\)](#) is solved online at time  $t \in \mathbb{N}_{\geq 0}$ . If the function  $V_N : \mathcal{X} \rightarrow \mathbb{R}$ , the terminal set  $\mathcal{O}$ , and the disturbance set  $\hat{\mathcal{D}}$  are convex, then [\(23\)](#) is a convex problem and we refer the reader to [\[15\]](#) for implementation details. One might also consider a formulation of [\(23\)](#) where the feedback gains are fixed, yielding a more basic tube MPC as considered in [\[13\]](#).

Let the optimal value of [\(24\)](#) be  $J_N^*(t, x(t))$ , with associated optimal policy sequence  $[u_{t|t}^*, \dots, u_{t+N-1|t}^*]$ . We then choose the robust control term of the *certainty equivalent* control policy [\(10\)](#) as the receding horizon feedback law  $u^* : \mathcal{X} \times \mathbb{N}_{\geq 0} \rightarrow \mathcal{U}$  such that

$$u^*(x(t), t) = u_{t|t}^*. \quad (24)$$

In [\(23\)](#) we have tightened the input constraints to account for the matching term in the certainty equivalent policy [\(10\)](#) when compared to a more standard robust MPC problem (i.e. see [\[15\]](#)). As is standard in robust MPC, we assume we can both compute a robust control invariant set  $\mathcal{O}(t)$  and have access to a convex terminal cost function  $V_N$ .

**Assumption 4:** The terminal cost  $V_N : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is a continuous convex Lyapunov function for the nominal dynamics under a policy  $u_N(x) = -Kx$ . That is, there exists a class- $\mathcal{K}_\infty$  function  $\alpha_N(\|x\|) \geq h(x, u_N(x))$  for which

$$V_N((A - BK)x) - V_N(x) \leq -\alpha_N(\|x\|). \quad (25)$$

**Assumption 5:** For the policy  $u_N(x) = -Kx$  in **Assumption 4**, the terminal set  $\mathcal{O}(t) \subseteq \mathcal{X}$  is a maximal robust positive invariant set for the closed loop system  $x(t+1) = (A - BK)x(t) + d(t)$  for  $d(t) \in \widehat{\mathcal{D}}(t)$  subject to  $x(t) \in \mathcal{X}$  and  $u_N(x(t)) \in \mathcal{U} \ominus B^\dagger \widehat{\mathcal{F}}(t)$  for all  $t \geq 0$ .

**Assumption 4** and **Assumption 5** are standard and easily satisfied by taking  $u_N(x) = -Kx$  and  $V_N(x) = x^\top Px$  as the solution to an LQR problem with  $h$  as the stage cost. Then,  $\mathcal{O}(t)$  can be computed efficiently using the standard algorithms in [31].

**Lemmas 1** and **2** imply that  $\mathcal{O}(t-1) \subseteq \mathcal{O}(t)$  since  $\widehat{\mathcal{D}}(t) \subseteq \widehat{\mathcal{D}}(t-1)$  and  $\widehat{\mathcal{F}}(t) \subseteq \widehat{\mathcal{F}}(t-1)$ . Therefore, the terminal constraint becomes less conservative over time.

### B. Stability

We prove the stability of our algorithm through a recursive feasibility and input-to-state stability argument.

**Theorem 2:** Consider the system (1), a parameter estimator that satisfies **Assumption 3** with features that satisfy **Assumption 1** in closed loop feedback with the *matching certainty equivalent* control law (10),(24). If the tube MPC problem (23) is feasible at  $t = 0$ , then for all  $t \geq 0$  we have that (23) is feasible and the closed loop system (1),(10),(24) satisfies  $x(t) \in \mathcal{X}$ , and  $\pi(x(t), t) \in \mathcal{U}$ .

**Proof:** Suppose the optimal control problem (23) is feasible at time  $t$ , with solution  $[u_{t|t}^*(\cdot), \dots, u_{t+N-1|t}^*(\cdot)]$ . By **Lemma 1** and **Corollary 1** we have that  $\hat{f}(x, t+k) \in \widehat{\mathcal{F}}(t)$  for all  $k \in \mathbb{N}_{\geq 0}$ . Therefore, the time-varying CE control law

$$\pi_{t+k|t}(\cdot) = u_{t+k|t}^*(\cdot) - B^\dagger \hat{f}(x(t), t+k) \quad (26)$$

satisfies the input constraints for  $t \in [t, t+N-1]$ , since (23) then implies  $u_{t+k|t}^*(\cdot) \in \mathcal{U} \ominus B^\dagger \widehat{\mathcal{F}}(t)$ . Moreover, by **Lemma 2** the disturbance support shrinks in time, i.e.,  $\widehat{\mathcal{D}}(t+1) \subseteq \widehat{\mathcal{D}}(t)$ . Therefore, we have that under policy (26) the closed-loop trajectory formed by (1),(26) satisfies  $x(t+k) \in \mathcal{X}$  for all  $k \in [0, N]$  and that  $x(t+N) \in \mathcal{O}(t)$ . Hence, if we apply the CE policy (10),(24) at time  $t$ , then  $x(t+1) \in \mathcal{X}$  and  $\pi(x(t), t) \in \mathcal{U}$ .

By **Assumption 5**, for any  $x \in \mathcal{O}(t) \subseteq \mathcal{O}(t+1)$ , applying the policy  $u_N(x) \in \mathcal{U} \ominus B^\dagger \widehat{\mathcal{F}}(t)$  implies that  $Ax + Bu_N(x) + d \in \mathcal{O}(t+1)$  for any  $d \in \widehat{\mathcal{D}}(t)$ . Therefore, **Corollary 1** and **Lemma 2** imply the policy sequence  $[u_{t+1|t}^*(\cdot), \dots, u_{t+N-1|t}^*(\cdot), u_N(\cdot)]$  is feasible for the tube MPC problem (23) at time  $t+1$ . Therefore, if the MPC program (23) is feasible at time  $t = 0$ , it is also feasible for all  $t \geq 0$  and the closed-loop system formed by the matching CE law (1), (10), (24) must robustly satisfy state and input constraints by induction.  $\blacksquare$

**Remark 2:** **Theorem 2** and **Assumption 5** imply that the maximal RPI set  $\mathcal{O}(t)$  or disturbance sets  $\widehat{\mathcal{D}}(t)$ ,  $\widehat{\mathcal{F}}(t)$  need not be updated at every timestep to guarantee recursive feasibility

(nor stability), so we can apply this algorithm in an iterative setting and update the constraints only between episodes.

**Theorem 3:** Consider a system of the form in (1), a parameter estimator that satisfies **Assumption 3** with features that satisfy **Assumption 1** in closed-loop feedback with the *certainty equivalent* control law (10),(24). Let  $\mathcal{X}_N \subseteq \mathcal{X}$  denote the set of states for which the tube MPC problem (23) is feasible. Then the closed loop system is locally input-to-state stable with region of attraction  $\mathcal{X}_N$ .

**Proof:** Our proof closely resembles [21, Thm. 2]. We argue that the nominal system is stable by a standard MPC argument, and that the closed-loop system is ISS since the disturbances are bounded. Since we assume the stage cost is quadratic, there exist two class- $\mathcal{K}_\infty$  functions  $\alpha_1, \alpha_2$  such that for all  $t \geq 0$ ,  $\alpha_1(\|x\|) \leq J_N^*(t, x) \leq \alpha_2(\|x\|)$ , a class- $\mathcal{K}_\infty$  function  $\alpha_3$  such that  $h(x, u) \geq \alpha_3(\|x\|)$ , and  $J_N^*(t, 0) = 0$  (see [14, Prop. 1], [21, Thm. 2]). Let  $J_N^*(t, x(t))$  be the solution of (23) associated with the nominal prediction  $[\bar{x}_{t|t}^*, \dots, \bar{x}_{t+N|t}^*]$  and feedback policies  $[u_{t|t}^*(\cdot), \dots, u_{t+N-1|t}^*(\cdot)]$ . As in the proof of **Theorem 2**, we have that if we apply the CE control law (10),(24) at time  $t$ , then the policies  $[u_{t+1|t}^*, \dots, u_{t+N-1|t}^*, u_N]$  are a feasible solution for (23) at time  $t+1$ . Let  $\bar{J}(t, x)$  be the cost associated with forward simulating the nominal system using the policies  $[u_{t+1|t}^*, \dots, u_{t+N-1|t}^*, u_N]$  with  $x$  as initial condition. i.e., set  $u_{t+N|t}^* = u_N(\cdot)$  and let  $\bar{x}_{t+1|t+1} = x$ ,  $\bar{x}_{k+1|t+1} = \bar{A}\bar{x}_{k|t+1} + \bar{B}u_{k|t}^*(\bar{x}_{k|t+1})$  for  $k \in \{t+1, \dots, t+N\}$  so that

$$\bar{J}(t, x) = \sum_{k=t+1}^{t+N} h(\bar{x}_{k|t+1}, u_{k|t}^*(\bar{x}_{k|t+1})) + V_N(\bar{x}_{t+N+1|t+1}).$$

This gives that  $J_N^*(t+1, x(t+1)) \leq \bar{J}(t, x(t+1))$ . Moreover, since the stage cost is quadratic and by **Assumption 4**,  $\bar{J}(t, x)$  is uniformly continuous in  $x$  for all  $t \geq 0$  on the state space since the inputs are constrained in a compact set. It follows that for  $x_1, x_2 \in \mathcal{X}$ , there exists a  $\mathcal{K}_\infty$  function  $\alpha_J$  such that for all  $t \geq 0$ ,  $|\bar{J}(t, x_1) - \bar{J}(t, x_2)| \leq \alpha_J(\|x_1 - x_2\|)$  (see [14, Lem. 1]). Therefore,

$$\begin{aligned} & J_N^*(t+1, x(t+1)) - J_N^*(t, x(t)) \\ & \leq \bar{J}(t, x(t+1)) - J_N^*(t, x(t)) \\ & = \bar{J}(t, x(t+1)) - \bar{J}(t, \bar{x}_{t+1|t}) + \bar{J}(t, \bar{x}_{t+1|t}) - J_N^*(t, x(t)) \\ & \leq |\bar{J}(t, x(t+1)) - \bar{J}(t, \bar{x}_{t+1|t})| - h(x_t, \bar{u}_{t|t}^*(x(t))) \\ & \leq \alpha_J(\|d(t)\|) - \alpha_3(\|x(t)\|). \end{aligned}$$

So, the system is ISS by **Theorem 1**.  $\blacksquare$

**Remark 3:** The ISS result in **Theorem 3** does not explicitly show that improvements in the confidence of the model lead to better performance of the controller, since we only assume the model confidence is non-decreasing in **Assumption 3**. In adaptive control, stronger guarantees of performance improvement are typically made under *persistence of excitation* assumptions [1].

## IV. ADAPTATION LAWS & LEARNING ALGORITHMS

In this section we highlight two common and perhaps complimentary online function approximation schemes, one is

statistical and one is not, that satisfy the decaying confidence interval of [Assumption 3](#) that we used to construct our ARMPC algorithm.

**Set Membership Estimation.** A common approach in the adaptive MPC literature is to estimate constant, or slowly changing, disturbances through *set membership* estimation [20], [25]. These estimators maintain a feasible parameter set that is refined as more data becomes available. The feasible parameter set contains all credible model parameters that explain previous observations, which means that the feasible parameter sets are nested over time. We consider learning the parameters of a nonlinear uncertainty model of the form in (3) directly using set-membership estimation. Under the prior knowledge [Assumption 2](#), the initial feasible parameter set is given as  $\Theta(0) = \{\widehat{W}(0)\} \oplus \mathcal{W}(0)$  and the feasible parameter set at time  $t$  is obtained as

$$\Theta(t) = \{W \in \Theta : x(k+1) - Ax(k) - Bu(k) - W\phi(x(k)) \in \mathcal{V}, \forall k \in \{0, \dots, t-1\}\}. \quad (27)$$

When  $\mathcal{V}$  is a hyperbox, this estimator maintains independent feasible sets for each row of  $W$  and can be updated recursively in time with polytopical set intersections by rewriting (27) in terms of the row-wise vectorization of  $W$ . Clearly,  $\Theta(t) \subseteq \Theta(t-1)$ . As is common practice in the literature [22], we propose generating a point estimate of the parameters as the *Chebyshev center* of the feasible parameter set:

$$\widehat{W}(t) = \arg \min_{\widehat{W}} \max_{W \in \Theta(t)} \|\widehat{W} - W\|_F. \quad (28)$$

By definition, this approach minimizes the worst-case error of the point estimates and is typically straightforward to compute [32]. Denoting the Chebyshev radius for the feasible parameter set associated with the  $i$ -th row of  $W$  as  $r_i(t) = \min_w \max_{w_i \in \Theta_i(t)} \|w - w_i\|_2$ , we take the confidence interval on  $\widehat{w}_i(t) - w_i$  as  $\mathcal{W}_i(t) = \{\widehat{w}_i : \|\widehat{w}_i\|_2 \leq r_i(t)\}$ .

By definition, since  $\Theta(t) \subseteq \Theta(t-1)$ , the Chebyshev radii must be decreasing over time:  $r_i(t) \leq r_i(t-1)$ . Therefore, a set-membership estimator with point estimates as the Chebyshev center satisfies [Assumption 3](#).

**Recursive Bayesian Linear Regression (BLR).** In the case of Bayesian estimation, we can generate confidence intervals directly from the posterior distribution over parameters if we know the disturbance distribution. We outline this approach under a simple, standard assumption.

[Assumption 6](#): We assume that each entry of the process noise is bounded  $v(t) = [v_1(t), \dots, v_n(t)]^\top \in \mathcal{V} = \{v : |v| \leq \sigma_i\}$ , and that each entry  $v_i(t)$  is independent of the others. Hence,  $v_i(t)$  is sub-Gaussian with variance proxy  $\sigma_i^2$ . Under [Assumption 6](#), we can essentially treat the noise as both normally distributed for convenient analysis and provide safety guarantees for the algorithm proposed in [§III](#). If we place subjective priors over the rows of  $W$  of the form  $w_i \sim \mathcal{N}(\widehat{w}_i(0), \sigma_i^2 \Lambda_i^{-1}(0))$ , then the resulting posterior parameter distribution at time  $t$  is also Gaussian,  $w_i \sim \mathcal{N}(\widehat{w}_i(t), \sigma_i^2 \Lambda_i^{-1}(t))$ . We then use the mean of the posterior—also corresponding to the maximum a posteriori (MAP) estimate—as a point estimate for control:  $\hat{f}(x, t) =$

$\widehat{W}(t)\phi(x)$ . Defining the measurement and prediction at time  $t$  as  $y(t) := x(t+1) - Ax(t) - Bu(t)$  and  $\hat{y}(t) := \widehat{W}(t)\phi(t)$ , the MAP estimate for each row can then be updated with constant complexity in time using the recursive updates

$$\widehat{w}_i(t+1) = \widehat{w}_i(t) - \frac{(\hat{y}_i(t) - y_i(t))\phi(t)^\top \Lambda_i^{-1}(t)}{1 + \phi(t)^\top \Lambda_i^{-1}(t)\phi(t)} \quad (29)$$

$$\Lambda_i^{-1}(t+1) = \Lambda_i^{-1}(t) - \frac{\Lambda_i^{-1}(t)\phi(t)\phi(t)^\top \Lambda_i^{-1}(t)}{1 + \phi(t)^\top \Lambda_i^{-1}(t)\phi(t)}, \quad (30)$$

where we write each entry of  $\hat{y}(t)$  as  $\hat{y}_i(t) = \widehat{w}_i(t)^\top \phi(t)$  and  $\phi(t) := \phi(x(t))$ . We can recover the frequentist ordinary least-squares estimator if we assume a flat prior [33], which requires the availability of some amount of prior data to yield the initial values  $\widehat{W}(0)$ ,  $\Lambda(0)$ .

Taking a risk tolerance of  $\delta \in (0, 1)$ , we could naively define the confidence interval for the  $i$ -th row of  $\widehat{W}(t)$  as

$$\mathcal{W}_i^{\text{naive}}(t) := \{\tilde{w}_i \in \mathbb{R}^n : \tilde{w}_i^\top \Lambda_i(t) \tilde{w}_i \leq \sigma_i^2 \chi_n^2(1 - \frac{\delta}{n})\}, \quad (31)$$

where  $\chi_n^2(1 - \delta)$  is the  $1 - \delta$  quantile of the chi-square distribution with  $n$  degrees of freedom. However, the confidence interval in (31) does not capture the fact that we want to certify the safety of the policy for all time with high probability. We cannot achieve this with a single confidence interval of a point estimate at time  $t$ , as (31) ignores the correlations between the model estimates over time. For robust control, we instead desire confidence intervals such that

$$\widehat{W}(t) - W \in \mathcal{W}(t), \quad \forall t \in \mathbb{N}_{\geq 0}, \quad (32)$$

with probability at least  $1 - \delta$ .

Recent work applied a Martingale argument originating from the Bandits literature to generate such confidence intervals [16] by scaling the naive confidence intervals by a time-varying parameter<sup>1</sup>.

*Theorem 4:* [16, Thm 1] For the recursive Bayesian linear filter (29),(30), we have the estimation error  $\widehat{w}_i(t) - w_i \in \mathcal{W}_i(t)$  for all  $t \geq 0$  with probability at least  $1 - \delta$ , where

$$\mathcal{W}_i(t) := \{\tilde{w}_i : (\tilde{w}_i^\top \Lambda_i(t) \tilde{w}_i)^{\frac{1}{2}} \leq \sigma_i \beta_t(\delta/n)\}, \quad (33)$$

with  $\beta_t(\delta)$  equal to

$$\sqrt{2 \log \left( \frac{\det(\Lambda_i(t))^{1/2}}{\delta \det(\Lambda_i(0))^{1/2}} \right)} + \sqrt{\frac{\lambda_{\max}(\Lambda_i(0))}{\lambda_{\min}(\Lambda_i(t))}} \chi_n^2(1 - \delta). \quad (34)$$

The confidence intervals resulting from [Theorem 4](#) unfortunately do not immediately satisfy [Assumption 3](#) without a *persistence of excitation* or *active exploration* assumption as is made in [27]. A simple workaround is to update the estimate (29) fed into the controller only when the associated confidence intervals (33) have shrunk, effectively disregarding new data until the system has been excited sufficiently. This

<sup>1</sup>Subject to assumptions on the calibration of the prior, for which we refer the reader to [16, Assumption 3]. This assumption is trivially satisfied for flat priors, and we assume this assumption is satisfied for subjective priors.

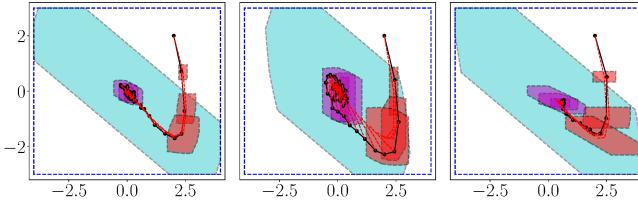


Fig. 2. Closed-loop trajectories of different systems. The true system evolution is in black, the predicted nominal trajectories are red, as are the predicted reachable sets at  $t = 0$ . The predicted reachable sets at  $t = 50$  are in purple. The dark blue line indicates the state constraints. The terminal invariant  $\mathcal{O}(50)$  is shown in cyan. Left: Adaptive CE MPC on the system with *matched* uncertainty (35). Middle: Benchmark Adaptive MPC on the *matched* system (35). Right: Adaptive CE MPC on the system with *unmatched* uncertainty (36).

approach was shown to perform well in practice [18]. Still, future work should explore strategies to guarantee confidence intervals constructed using [Theorem 4](#) (or other equivalent results) satisfy [Assumption 3](#) more naturally.

## V. EXPERIMENTS

To benchmark our adaptive, robust MPC (ARMPc) algorithm that *matches* as much of the uncertainty as possible, we compare it against the normative approach in ARMPc; we estimate the range of values that the uncertainty  $f$  can take and naively treat it as a disturbance using tube MPC as in [20], [22], [23], [25]. To the best of our knowledge, such algorithms have not been proposed for uncertain terms that satisfy [Assumption 1](#) in the literature. Therefore, we apply the tube MPC strategy in [15] online using the disturbance set  $\mathcal{V}'(t) = \hat{\mathcal{F}}(t) \oplus \mathcal{V}$ , since [Corollary 1](#) implies that both  $f(x(t)) + v(t) \in \mathcal{V}'(t)$  and  $\mathcal{V}'(t) \subseteq \mathcal{V}'(t-1)$  for all  $t \geq 0$ .

**Simple Matched System.** We illustrate the properties of our algorithm on a double-integrator system

$$x(t+1) = \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + f(x(t)) + v(t), \quad (35)$$

a typical example in the MPC literature that represents simplified second-order dynamics [31]. First, we consider the *matched* uncertainty  $f(x) = [0, w_1]^\top \tanh([0, 1]x)$ . We estimate the true parameter,  $w_1 = 0.5$ , online to improve performance. We take the disturbance as an isotropic Gaussian with  $\sigma^2 = 5 \times 10^{-3}$  clipped at its 95% confidence intervals. In addition, the system is subject to the state and input constraints  $(-4, -3) \preceq x \preceq (4, 3)$  and  $-2 \leq u \leq 2$ , respectively. We regulate the system to the origin from  $x(0) = [2, 2]^\top$  while minimizing the quadratic cost function  $h(x, u) = x^\top Qx + u^\top Ru$  over a horizon of length  $N = 3$ . We take  $Q = I_2$ ,  $R = 1$ . We use the BLR estimator with a flat prior and collect  $k = 45$  data points of the system evolution in a unit box near the origin to form an initial estimate of the model parameters and construct  $\mathcal{O}(t)$  using Alg. 10.5 in [31].

We plot the closed-loop system evolution in [Fig. 2](#) (left, middle). Our adaptive certainty-equivalent MPC algorithm is able to effectively control the system. In contrast, the benchmark ARMPc can only react to the learned dynamics after  $f$  enters the system as a disturbance, resulting in

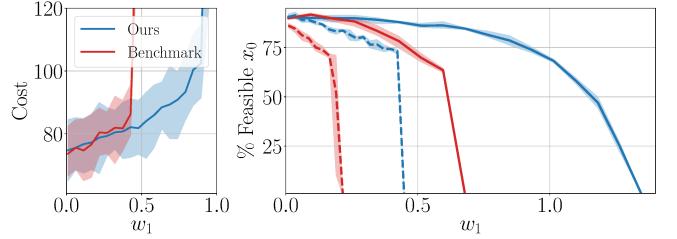


Fig. 3. Left: Closed-loop realized trajectory cost for the matched system (35) as a function of  $w_1$ . Exploding cost indicates infeasibility. The error bars indicate  $2\sigma$  bounds. Right: Solid lines indicate size of feasible envelopes as a function of  $w_1$  for a matched system. Dashed lines indicate the size of the feasible envelopes for the unmatched system (36) as a function of the magnitude of the unmatched dynamics  $w_1$  with  $w_2 = 0.5$ .

considerably larger closed-loop oscillations and uncertainty on the predicted trajectory. In [Fig. 2](#), we also plot the reachable sets associated with the predicted trajectory at the first and last timesteps of the control task. The reachable sets show that our adaptive MPC resolves prediction uncertainty in the system since they shrink over time. Moreover, online learning has little consequence for the benchmark, as increasing the confidence in the model does not significantly reduce the estimated range of values that the nonlinear function takes.

In addition, we compare the asymptotic performance of our algorithm with the benchmark ARMPc as a function of  $w_1$ , controlling the magnitude of the nonlinearity. To do this, we set the number of data points used to generate the initial model estimate to  $k = 10^3$  and collect trajectory rollouts for various values of  $w_1$  from the fixed initial condition. As [Fig. 3](#) (left) shows, our ARMPc is feasible for disturbances more than twice the magnitude of those the benchmark ARMPc can tolerate for the given initial condition. In addition, the realized control cost does not differ significantly from the benchmark for values of  $w_1$  when both controllers are feasible.

Finally, we illustrate how our algorithm tolerates larger dynamic uncertainty by comparing the size of the feasible envelope (i.e., the set of initial conditions for which the MPC problem is feasible) as a function of  $w_1$ . We set the number of data points to inform our prior to a modest  $k = 50$ , grid the state-space, and take the feasible region as the convex hull of the initial conditions for which the MPC problem is feasible. Then, we estimate the percentage of states  $x_0 \in \mathcal{X}$  in the feasible envelope as the ratio of volumes between the feasible envelope and the state space  $\mathcal{X}$ , illustrated with solid lines in [Fig. 3](#) (right). Our adaptive MPC algorithm can tolerate much larger disturbances than the benchmark ARMPc. In these experiments, the feasible envelope of the benchmark becomes empty when the maximal robust invariant is null ( $\mathcal{O}(t) = \emptyset$ ), indicating that there is no subset of  $\mathcal{X}$  in which the LQR policy associated with  $h$  results in provably safe behavior [31]. Hence, [Fig. 3](#) highlights the fragility of existing ARMPc approaches under large disturbances.

**Simple Unmatched System.** We now extend the simulations of the simple matched system to the unmatched case to understand the effect of additive nonlinear terms that cannot be canceled from the dynamics. We keep the nominal

dynamics identical to (35) and take the nonlinear dynamics

$$f(x) = \frac{1}{\sqrt{2}}[w_1 \sin(4x_1), w_2 \tanh(x_2)]^\top, \quad (36)$$

where  $w_1$  and  $w_2$  are unknown parameters. Similar to the previous experiments, we initialize the model with  $k = 45$  data points sampled around the origin. In this example, the certainty equivalent policy (10) can only compensate for the second component of the nonlinear dynamics (36). We set  $w_1 = 0.2$  and  $w_2 = 0.3$ . In Fig. 2 (right), the size of the reachable sets increases if we simulate the system with the unmatched dynamics (36). The benchmark ARMPC was infeasible from this initial condition, showing that our method still outperforms the benchmark. Next, we fix  $w_2 = 0.5$  and vary  $w_1$  to understand the impact of an estimated, unmatched dynamics component. Fig. 3 (dashed, right) shows that in our experiment, matching as much of the nonlinear dynamics as possible allows us to handle unmatched dynamics of about twice the magnitude as the benchmark. We conclude that our method is a more effective strategy even if the uncertainty is unmatched. Naturally, Fig. 3 (right) also shows that the benefit of our method diminishes as the proportion of the nonlinear dynamics  $f(x(t))$  in  $\text{Range}(B)$  becomes smaller.

**Planar Quadrotor.** Finally, we simulate a simplified example of a quadrotor in a windy environment. The force field induced by the wind varies spatially, modeling real-world scenarios such as down-wash from another quadrotor. We consider a planar version of the quadrotor for simplicity [34]. We model the wind disturbance as incident at a fixed angle with a velocity that drops off according to an inverse square exponential normal to the direction of incidence. We linearize the dynamics around  $x = 0$ ,  $u = \frac{mg}{2}[1, 1]^\top$  and discretize the simulation using Euler's method. We only set constraints on the pose of the drone  $(x, y, \theta)$ . Its linear and angular velocities are unconstrained. The quadrotor is an underactuated system, and therefore the discretized simulation has unmatched dynamics terms. Still, a drone controller can always match disturbance forces along the  $y$  axis. We take a Bayesian approach and model the unknown wind disturbance as a function of  $d = 20$  normalized random Fourier features  $\phi_i(x) = \frac{1}{\sqrt{d}} \cos(\alpha_i^\top x + \beta_i)$  where  $\alpha \sim \mathcal{N}(0, I_n)$  and  $\beta_i \sim \mathcal{U}[0, 2\pi]$ . To calibrate the Bayesian prior, we first select a zero-mean prior with a variance chosen such that the confidence interval reflects a conservative bound on the wind disturbance. We then calibrate the Bayesian prior using  $k = 100$  historical wind samples.

The support of the estimated wind disturbance was too large for our benchmark to be feasible in our experiments, so we compare our approach with a naive tube MPC that does not account for the wind disturbance at all. As shown in Fig. 4 (left), if the wind disturbance is axis aligned, our ARMPC learns to match the wind forces and reaches the origin quickly. In contrast, the naive tube MPC misses the origin and drifts significantly. In addition, if we set the angle of incidence of the wind as  $\theta_w = 22.5^\circ$ , Fig. 4 (right) shows that our approach still achieves decent control performance. This is because our certainty equivalent controller (10) cancels the  $y$ -component

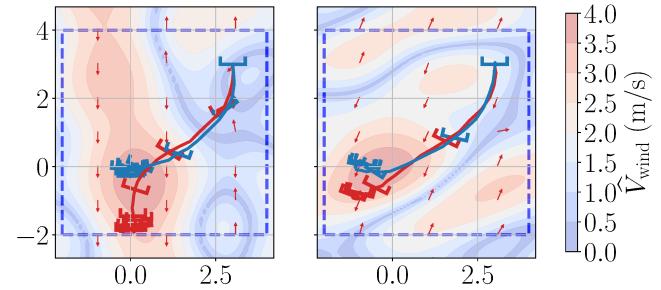


Fig. 4. Learned trajectories of our adaptive MPC (blue) compared to a naive tube MPC (red) in the  $xy$ -plane for a simulated planar quadrotor. The icons show the orientation of the quadrotor over time. The contours and colorbar indicate the learned wind speed in m/s. Left: Wind comes straight from above. Right: Wind comes at  $\theta_w = 22.5^\circ$ .

of the disturbance and converges to a small steady-state offset in the  $x$ -direction. In contrast, the benchmark ARMPC could not guarantee safety for any of the drone tasks, and a naive unsafe tube MPC that does not consider the wind disturbance at all performs poorly. Moreover, model adaptation reduced the realized closed-loop cost in the simulation Fig. 4 (left) by about 12% over 10 episodes from  $J^1 = 2570$  to  $J^{10} = 2259$ .

## VI. CONCLUSIONS AND FUTURE WORK

The simulations in §V show that our method can achieve substantial performance improvements compared to existing approaches, even when a certainty equivalent controller cannot cancel significant, unmatched components of the nonlinear dynamics. We conclude that we can reduce the conservatism of robust MPC approaches by extending certainty equivalent control laws from classical adaptive control. As a result, our method tolerates nonlinear terms of larger magnitude in the dynamics.

Since our control algorithm allows for adaptation laws based on statistical techniques that are more robust to misspecification or outliers than set membership estimation, future work should extend our simulations to hardware experiments. In addition, our method requires that the features are known a priori, an assumption that future work could relax by, for example, applying meta-learning algorithms [9].

## ACKNOWLEDGEMENTS

We thank Monimoy Bujarbarua for his thoughtful comments on an early manuscript. This research was supported in part by the NSF via CPS award #1931815, NASA ULI grant #80NSSC20M0163, and an Early Stage Innovations grant. S.M.R. and J.H. were also supported in part by NSERC. This article solely reflects our own opinions and conclusions, and not those of any NSF, NASA, or NSERC entity.

## REFERENCES

- [1] J.-J. E. Slotine and Weiping Li. *Applied nonlinear control*. 1991.
- [2] P. Ioannou and J. Sun. *Robust Adaptive Control*. 2012.
- [3] P. Ioannou and B. Fidan. *Adaptive Control Tutorial*. 2006.
- [4] K. Wise E. Lavretsky. *Robust and Adaptive Control with Aerospace Applications*. 2012.
- [5] K. J. Åström and B. Wittenmark. *Adaptive control*. Courier, 2013.

- [6] G. Joshi, J. Virdi, and G. Chowdhary. Asynchronous Deep Model Reference Adaptive Control. *arXiv:2011.02920*, 2020.
- [7] G. Chowdhary, H. A. Kingravi, J. P. How, and P. A. Vela. Bayesian Nonparametric Adaptive Control Using Gaussian Processes. *TNNLS*, 2015.
- [8] D. D. Fan, J. Nguyen, R. Thakker, N. Alatur, A.-a Agha-mohammadi, and E. A. Theodorou. Bayesian Learning-Based Adaptive Control for Safety Critical Systems. In *ICRA*, 2020.
- [9] J. Harrison, A. Sharma, and M. Pavone. Meta-learning Priors for Efficient Online Bayesian Regression. In *WAFR*, 2018.
- [10] J. Harrison, A. Sharma, R. Calandra, and M. Pavone. Control adaptation via meta-learning dynamics. In *NeurIPS Workshop on Meta-Learning*, 2018.
- [11] S. M. Richards, N. Azizan, J.-J. E. Slotine, and M. Pavone. Adaptive-control-oriented meta-learning for nonlinear systems. *RSS*, 2021.
- [12] P. O. M. Scokaert and D. Q. Mayne. Min-max feedback model predictive control for constrained linear systems. *TAC*, 1998.
- [13] D. Q. Mayne, M. M. Seron, and S. V. Raković. Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, 2005.
- [14] D. Limon, T. Alamo, D. M. Raimondo, D. M. De la Peña, J. M. Bravo, A. Ferramosca, and E. F. Camacho. Input-to-state stability: A unifying framework for robust model predictive control. In *Nonlinear Model Predictive Control*. 2009.
- [15] P. J. Goulart, E. C. Kerrigan, and J. M. Maciejowski. Optimization over state feedback policies for robust control with constraints. *Automatica*, 2006.
- [16] T. Lew, A. Sharma, J. Harrison, and M. Pavone. Safe Model-Based Meta-Reinforcement Learning: A Sequential Exploration-Exploitation Framework. *arXiv:2008.11700*, 2020.
- [17] L. Hewing, J. Kabzan, and M. N. Zeilinger. Cautious model predictive control using Gaussian process regression. *TAC*, 2017.
- [18] T. Koller, F. Berkenkamp, M. Turchetta, and A. Krause. Learning-Based Model Predictive Control for Safe Exploration. In *CDC*, 2018.
- [19] P.K. Mishra, M.V. Gasparino, A. Velsasquez, and G. Chowdhary. Deep model predictive control with stability guarantees, 2021.
- [20] M. Bujarbaruah, X. Zhang, U. Rosolia, and F. Borrelli. Adaptive MPC for Iterative Tasks. In *CDC*, 2018.
- [21] M. Bujarbaruah, X. Zhang, M. Tanaskovic, and F. Borrelli. Adaptive stochastic MPC under time varying uncertainty. *TAC*, 2021.
- [22] M. Bujarbaruah, S. H. Nair, and F. Borrelli. A Semi-Definite Programming Approach to Robust Adaptive MPC under State Dependent Uncertainty. In *ECC*, 2020.
- [23] R. Soloperto, M. A. Müller, S. Trimpe, and F. Allgöwer. Learning-Based Robust Model Predictive Control with State-Dependent Uncertainty. *IFAC NMPC*, 2018.
- [24] S. Di Cairano. Indirect adaptive model predictive control for linear systems with polytopic uncertainty. In *ACC*, 2016.
- [25] J. Köhler, E. Andina, R. Soloperto, M. Müller, and F. Allgöwer. Linear robust adaptive model predictive control: Computational complexity and conservatism. In *CDC*, 2019.
- [26] A. Aswani, H. Gonzalez, S. S. Sastry, and C. Tomlin. Provably safe and robust learning-based model predictive control. *Automatica*, 2013.
- [27] H. Mania, M.I. Jordan, and B Recht. Active learning for nonlinear system identification with guarantees, 2020.
- [28] H. Li, A. Liu, and L. Zhang. Input-to-state stability of time-varying nonlinear discrete-time systems via indefinite difference Lyapunov functions. *ISA Transactions*, 2018.
- [29] Z.-P. Jiang and Y. Wang. Input-to-state stability for discrete-time nonlinear systems. *Automatica*, 2001.
- [30] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. In *NeurIPS*, 2018.
- [31] F. Borrelli, A. Bemporad, and M. Morari. *Predictive control for linear and hybrid systems*. 2017.
- [32] S. Boyd and L. Vandenberghe. *Convex Optimization*. 2004.
- [33] M. R. Deisenroth, A. A. Faisal, and C. S. Ong. *Mathematics for machine learning*. Cambridge University Press, 2020.
- [34] R. Tedrake. *Underactuated Robotics: Algorithms for Walking, Running, Swimming, Flying, and Manipulation (Course Notes for MIT 6.832)*.