Distributionally Robust Policy and Lyapunov-Certificate Learning

Kehan Long[1] (Student Member), Jorge Cortés[1] (Fellow), Nikolay Atanasov[1] (Senior Member)

[1]Contextual Robotics Institute, University of California San Diego, La Jolla, CA 92093, USA

CORRESPONDING AUTHOR: Kehan Long (e-mail: k3long@ucsd.edu)

ABSTRACT This article presents novel methods for synthesizing distributionally robust stabilizing neural controllers and certificates for control systems under model uncertainty. A key challenge in designing controllers with stability guarantees for uncertain systems is the accurate determination of and adaptation to shifts in model parametric uncertainty during online deployment. We tackle this with a novel distributionally robust formulation of the Lyapunov derivative chance constraint ensuring a monotonic decrease of the Lyapunov certificate. To avoid the computational complexity involved in dealing with the space of probability measures, we identify a sufficient condition in the form of deterministic convex constraints that ensures the Lyapunov derivative constraint is satisfied. We integrate this condition into a loss function for training a neural network-based controller and show that, for the resulting closed-loop system, the global asymptotic stability of its equilibrium can be certified with high confidence, even with Out-of-Distribution (OoD) model uncertainties. To demonstrate the efficacy and efficiency of the proposed methodology, we compare it with an uncertainty-agnostic baseline approach and several reinforcement learning approaches in two control problems in simulation. Open-source implementations of the examples are available at https://github.com/KehanLong/DR_Stabilizing_Policy.

INDEX TERMS Learning for control, Lyapunov methods, optimization under uncertainty, stability of nonlinear systems

I. Introduction

In control theory and robotics, the task of synthesizing stabilizing controllers with provable certificates for open-loop systems under model uncertainty remains an important challenge. A Lyapunov function (LF) [2] is a fundamental tool for asserting the asymptotic stability of ordinary differential equations (ODEs). Existing approaches in deriving stabilizing controllers for uncertain nonlinear systems usually rely on known bounds or probabilistic models of the uncertainty, which may be hard to obtain in practice. In this work, we address this challenge by exploring the potential of neural networks to synthesize distributionally robust stabilizing controllers and associated Lyapnuov stability certificates, providing a new approach for ensuring robustness to model uncertainty.

The use of Lyapunov functions for ensuring stability of nonlinear systems is well documented in the literature

[3]–[5]. The early 2000s marked a significant development in the field with the introduction of sum-of-square (SOS) polynomials to identify LFs through semi-definite programming (SDP) [6], [7]. These approaches, however, often necessitate polynomial approximations of the system dynamics and scale poorly in higher dimensional spaces. Moreover, traditional SOS methods face difficulties in the joint synthesis of stabilizers and LF certificates for open-loop systems [8]. With the remarkable advancements in using neural networks as function approximators, researchers [9]–[13] have started representing LFs and/or control policies using neural networks. This progress holds great potential for developing stabilizing controllers and certificates for complex nonlinear systems, offering a vastly expanded functional space compared to SOS polynomials. Despite these advancements, a critical gap persists in deriving controllers equipped with suitable stability guarantees for systems with model uncertainty, an aspect paramount to the practical application of control and robotic systems in real-world scenarios. To

---

A preliminary version of this paper appeared as [1] at the 2023 Learning for Dynamics and Control Conference.

address the challenges of ensuring stability and stabilization in systems with model uncertainty, a related class of works [14]–[18] has focused on adapting stability certificates for nominal systems by incorporating model uncertainty during deployment. These approaches commonly assume knowing the nominal certificates or controllers as well as the error bounds or distribution of the model uncertainty.

Our work aims to develop a distributionally robust Lyapunov function and controller pair for systems with model uncertainty. Instead of relying on knowledge of the distribution or error bounds of the model uncertainty, we seek to synthesize a controller that ensures the global asymptotic stability of the controlled system, even in the presence of distribution shifts during online deployment. To do this, we rely on recent advances in distributionally robust optimization (DRO) [19], [20]. Distributionally robust chance constraints handle uncertainty in constraints with a limited number of available samples. The core strategy involves constructing an ambiguity set around the empirical distribution derived from these samples, defined by a radius using a probability distance measure such as the Wasserstein distance [20], [21]. The objective is to ensure constraint satisfaction with high confidence for all distributions within this set. This key benefit of the approach is its robust performance guarantee for distributional uncertainty, i.e., uncertainty in the possible probability distribution of the samples.

This paper is an extended version of our previous conference publication [1], which introduced distributionally robust Lyapunov stability for closed-loop systems with model uncertainty. In this work, we extend our approach to the joint synthesis of distributionally robust controllers and Lyapunov certificates for nonlinear control systems. Our **contributions** include the following.

- Given samples of uncertain model parameters, we introduce a novel distributionally robust Lyapunov formulation for systems under model uncertainty.
- We identify a sufficient convex constraint that ensures the distributionally robust Lyapunov condition is satisfied and employ it to learn a stabilizing neural network controller and Lyapunov function pair;
- We show that the resulting neural network controller guarantees the global asymptotic stability in probability of the closed-loop system;
- We validate in both theory and experiments that uncertain systems governed by the learned controller are stable with high confidence, even under out-of-distribution model uncertainty.

II. Related Work

This section reviews related work on synthesizing Lyapunov-stable controllers and distributionally robust optimization.

Stabilizing Control Design. Lyapunov theory plays a pivotal role in synthesizing stabilizing controllers for nonlinear systems. It offers strong asymptotic stability guarantees for the controlled system, which are particularly useful in real-world applications. For linear systems, synthesizing Lyapunov-stable controllers can be efficiently accomplished by linear quadratic regulator techniques [22]. For control-affine systems with polynomial dynamics, researchers [8] have successfully obtained controller and certificate pairs by solving SOS programs. However, the absence of a valid SOS Lyapunov function does not necessarily imply the system's instability, as certain positive-definite functions are not SOS-representable [23]. Moreover, the computational complexity of SOS techniques limits their applicability to relatively simple polynomial systems.

Recently, there has been an increasing focus on leveraging neural networks to address the limitations of SOS-based methods. Neural networks offer more general function representations compared to SOS polynomials, thereby enhancing the capacity to handle a broader class of systems.

Richards *et al.* [10] introduced a neural network-based method to learn the region of attraction for given controllers in discrete-time systems. Chang *et al.* [9] focused on synthesizing controllers and neural network Lyapunov functions for nonlinear continuous-time systems, with verification and improvement of the learned functions by satisfiability modulo theories (SMT) solvers. Boffi *et al.* [24] improved the efficiency of learning Lyapunov functions by integrating positive-definiteness and equilibrium conditions directly into the network architecture. Dai *et al.* [11] developed an approach for discrete-time systems that synthesizes neural network stabilizing controllers and Lyapunov functions, refining controller performance using mixed integer programs (MIP) verifiers. Gaby *et al.* [12] theoretically certified neural network Lyapunov functions by analyzing the approximation power of networks with specific architectures. Zhou *et al.* [25] proposed a framework for learning nonlinear systems, stabilizing controllers, and Lyapunov functions with stability guarantees verified by SMT solvers. Dawson *et al.* [26] extended these ideas to learning safety certificates as control Lyapunov-barrier functions for control-affine systems with bounded convex-hull uncertainty. A more extensive survey by Dawson *et al.* [27] details recent advancements in this line of research. Another related line of research is the use of CLFs to improve the sample efficiency and robustness of reinforcement learning (RL) algorithms [28], [29].

Despite these developments, there has been limited focus on synthesizing Lyapunov-stable controllers with provable certificates for systems under model uncertainty, particularly in settings where bounds or the distribution of the model uncertainty are challenging to determine.

Distributionally Robust Optimization. Distributionally robust optimization has emerged as a powerful approach to tackle uncertainty in optimization problems, particularly advantageous when dealing with a limited number of samples. To account for the discrepancy between the empirical and true distributions of the uncertainty samples, researchers have introduced several uncertainty descriptors, such as moment ambiguity sets [30], Kullback–Leibler ambiguity

sets [31], and Wasserstein ambiguity sets [20], [21], [32]. The DRO framework has been increasingly recognized for its robust performance guarantees against distributional uncertainty, leading to its widespread application in machine learning [33], [34], uncertainty quantification [35], control theory [36]–[40], and robotics [41], [42].

Long *et al.* [43] proposed a distributionally robust formulation for safe stabilizing control under model uncertainty, assuming nominal safe and stable certificates are provided. Ren and Majumdar [41] introduced a DRO framework that enhances policy robustness by iteratively training with adversarial environments generated through a learned generative model. Liviu *et al.* [44] introduced Wasserstein tube MPC for stochastic systems, enhancing robustness and efficiency by effectively analyzing state trajectory uncertainties with limited noise samples. In [45], a distributionally robust model predictive control formulation with the Wasserstein metric is proposed, with its parameters efficiently optimized through reinforcement learning. Hakobyan and Yang [46] developed a tractable distributionally robust differential dynamical programming method for stochastic systems by utilizing the Kantrovich duality principle. Distributionally robust RRT for motion planning that addresses model and environmental uncertainties was developed using moment-based ambiguity sets by Summers [47] and using Wasserstein-metric-based ambiguity sets by Lathrop *et al* [48].

III. Background

This section introduces our notation and offers a brief review of Lyapunov functions.

A. Notation

We denote the sets of real, non-negative real, and natural numbers by $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{N}$, respectively. For $N \in \mathbb{N}$, we let $[N] := \{1, 2, \ldots N\}$. We use $\mathbf{0}_n$ to denote the n-dimensional zero vector. For a scalar x, we define $(x)_+ := \max(x, 0)$. We denote by $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ the identity matrix. For a vector $\mathbf{x}$ and a matrix $\mathbf{X}$, we use $\|\mathbf{x}\|$ and $\|\mathbf{X}\|$ to denote the L_2 norm and the spectral norm, respectively. We denote by $B(\mathbf{x}; \delta)$ the open ball centered at $\mathbf{x}$ with radius δ, and by $\overline{B(\mathbf{x}; \delta)}$ its closure. The gradient of a differentiable function V is denoted by ∇V. We consider a complete separable metric space Ξ with metric d and Borel σ-algebra $\mathcal{F}$. The set of Borel probability measures on Ξ is denoted as $\mathcal{P}(\Xi)$. For a random variable $\boldsymbol{\xi}$ with distribution supported on Ξ, we denote the distribution and expectation by $\mathbb{P}^*$ and $\mathbb{E}_{\mathbb{P}^*}(\boldsymbol{\xi})$, respectively.

B. System Stabilization Certified by Lyapunov Functions

Consider the controlled continuous-time system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), (1)$$

where $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ denotes the system state and $\mathbf{u} \in \mathcal{U} \subseteq \mathbb{R}^m$ is the control input. We assume $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ is locally Lipschitz, and the origin $\mathbf{x} = \mathbf{0}_n$ is the desired equilibrium of the unforced system, i.e., $f(\mathbf{0}_n, \mathbf{0}_m) = \mathbf{0}_n$.

Plugging a Lipschitz controller $\mathbf{u} = \boldsymbol{\pi}(\mathbf{x})$ in results in the closed-loop system

$$\dot{\mathbf{x}} = \mathbf{f}_{\text{cl}}(\mathbf{x}) := \mathbf{f}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x})).(2)$$

In this context, a Lyapunov function (LF) $V : \mathbb{R}^n \to \mathbb{R}$ is used to analyze the asymptotic stability of the equilibrium. An LF for (2) satisfies the following conditions:

$$V(\mathbf{0}_n) = 0, \ V(\mathbf{x}) > 0, \ \forall \mathbf{x} \neq \mathbf{0}_n, \tag{3a}$$

$$\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x})) = \nabla V(\mathbf{x})^\top \mathbf{f}_{\text{cl}}(\mathbf{x}) < 0, \ \forall \mathbf{x} \neq \mathbf{0}_n. \tag{3b}$$

The existence of a radially unbounded ($V(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$) LF implies global asymptotic stability [49] of the dynamical system (2).

When the system model (1) is perfectly known, classical work [4], [50] has shown how to synthesize stabilizing controllers when a control Lyapunov function is available. This has provided the theoretical basis for work aimed at learning Lyapunov-stable controllers [8], [9], [24]. The availability of an LF-controller pair provides a complete solution, as the controller ensures system stabilization and the LF certifies its validity. However, real-world scenarios often present uncertainty in the dynamics model (1) that vary based on different physical parameters or operational conditions. Our goal is to find an LF-controller pair when uncertainty is present in the model. In particular, we do not assume any known error bounds or known distribution for the system uncertainty and instead investigate how to handle it based on finitely many realizations.

IV. Problem Formulation

We aim to synthesize a Lyapunov-stable controller for the continuous-time system with model uncertainty:

$$\dot{\mathbf{x}} = \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u}, \boldsymbol{\xi}) := \mathbf{f}(\mathbf{x}, \mathbf{u}) + \mathbf{W}(\mathbf{x}, \mathbf{u})\boldsymbol{\xi}, (4)$$

where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n$ denotes the state vector, $\mathbf{u} \in \mathcal{U} \subset \mathbb{R}^m$ is the control input, and the functions $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{W} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n \times k}$ are locally Lipschitz. The set $\mathcal{X} \subset \mathbb{R}^n$ represents the domain of interest, which includes the origin, and $\mathcal{U} \subset \mathbb{R}^m$ denotes the set of admissible control inputs. Each column of $\mathbf{W}$ represents a specific direction of uncertainty (e.g., mass, friction). The vector $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_k]^\top$ characterizes the uncertainty in the system model, and the distribution $\mathbb{P}^*$ of $\boldsymbol{\xi}$ is supported on a compact set $\Xi \subseteq \mathbb{R}^k$. Additionally, we assume that $\mathbf{W}(\mathbf{0}_n, \mathbf{0}_m) = \mathbf{0}_{n \times k}$, which ensures the origin is the desired equilibrium.

With the aim of stabilizing a system with model uncertainty in (4), we review chance-constraint formulations and their convex approximations. The Conditional Value-at-Risk (CVaR) is particularly useful in this context. For a random variable Q with distribution $\mathbb{P}_q$, the Value-at-Risk (VaR) at confidence level $1 - \epsilon$ is defined as $\text{VaR}_{1-\epsilon}^{\mathbb{P}_q}(Q) := \inf_{t \in \mathbb{R}}\{t \mid \mathbb{P}_q(Q \leq t) \geq 1 - \epsilon\}$. Based on VaR, CVaR is defined [51] as $\text{CVaR}_{1-\epsilon}^{\mathbb{P}_q}(Q) = \mathbb{E}_{\mathbb{P}_q}[Q \mid Q \geq \text{VaR}_{1-\epsilon}^{\mathbb{P}_q}(Q)]$. Fig. 1 shows an illustration of VaR and CVaR.
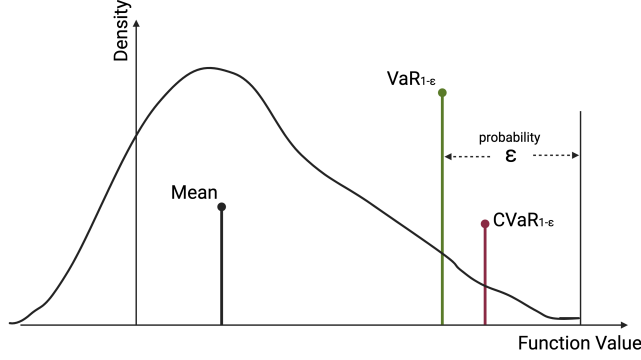
FIGURE 1: Illustration of VaR and CVaR within the distribution of a random variable. $\text{VaR}_{1-\epsilon}$ is the lower $1-\epsilon$ percentile of the random variable while $\text{CVaR}_{1-\epsilon}$ computes the expected value of the realizations above VaR.

Our goal is to find a function $V : \mathbb{R}^n \mapsto \mathbb{R}$ and a controller $\boldsymbol{\pi} : \mathbb{R}^n \mapsto \mathbb{R}^m$ that satisfy the Lyapunov conditions in (3). However, the uncertainty in the dynamical system (4), which manifests itself in the term $\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}))$, poses a challenge for ensuring that the condition on the derivative of the Lyapunov function (3b) are satisfied. Therefore, we consider the chance-constrained condition $\forall \mathbf{x} \neq \mathbf{0}_n$,

$$\mathbb{P}^*(\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}) + \gamma \|\mathbf{x}\| \leq 0) \geq 1 - \epsilon, \quad (5)$$

where $\gamma \in \mathbb{R}_{>0}$ and $\epsilon \in (0, 1)$ is a user-specified risk tolerance. The feasible set of the probabilistic constraint (5) is generally non-convex. To address this, Nemirosvski and Shapiro [52] proposed a convex CVaR approximation of the chance constraint in (5) as follows,

$$\text{CVaR}_{1-\epsilon}^{\mathbb{P}^*}(\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}) + \gamma \|\mathbf{x}\|) \leq 0, \qquad \forall \mathbf{x} \neq \mathbf{0}_n. (6)$$

This inequality implies the satisfaction of the original chance constraint in (5). As shown in [51], CVaR can also be formulated as the following convex program:

$$\begin{aligned}
&\text{CVaR}_{1-\epsilon}^{\mathbb{P}^*}(\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}) + \gamma \|\mathbf{x}\|) \\
&= \inf_{t \in \mathbb{R}} [\epsilon^{-1} \mathbb{E}_{\mathbb{P}^*}[((\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}) + \gamma \|\mathbf{x}\|) + t)_+] - t]. \quad (7)
\end{aligned}$$

Using the formulations in (5) and (6) requires knowledge of the distribution $\mathbb{P}^*$. However, in robotics and control applications typically $\mathbb{P}^*$ is unknown but a limited number of samples $\{\boldsymbol{\xi}_i\}_{i \in [N]}$ of the uncertain system parameters is available instead. Moreover, the distribution of $\boldsymbol{\xi}$ may shift at system deployment time. This motivates us to consider the following problem.

Problem 1 (Distributionally robust Lyapunov function and controller learning). *Consider the system* (4) *with nominal dynamics* $\mathbf{f}$ *and perturbation* $\mathbf{W}$. *Let* $\{\boldsymbol{\xi}_i\}_{i \in [N]}$ *be finitely many samples of the uncertainty* $\boldsymbol{\xi}$. *Taking potential distribution shift into account, design a control policy* $\boldsymbol{\pi} : \mathbb{R}^n \to \mathbb{R}^m$ *along with a Lyapunov certificate* $V : \mathbb{R}^n \to \mathbb{R}$ *that ensures that the equilibrium of the closed-loop system is globally asymptotically stable with high probability.*

V. Convex Distributionally Robust Stability Formulation

In this section, we develop a distributionally robust formulation for control synthesis and analyze the stability of the resulting closed-loop systems. Our approach builds upon the concepts of distributionally robust optimization [20], [32].

A. Distributionally Robust Lyapunov-Stable Constraint

Let $\mathbb{P}_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{\boldsymbol{\xi}_i}$ denote the discrete empirical distribution obtained from the samples $\{\boldsymbol{\xi}_i\}_{i \in [N]}$. We consider an ambiguity set of possible distributions for $\boldsymbol{\xi}$, which are similar to the empirical distribution $\mathbb{P}_N$. Let $\mathcal{P}_p(\Xi) \subseteq \mathcal{P}(\Xi)$ be the set of Borel probability measures with finite p-th moment for $p \geq 1$. The p-Wasserstein distance between two probability measures μ, ν in $\mathcal{P}_p(\Xi)$ is defined as [32]:

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \mathbb{Q}(\mu, \nu)} \left[\int_{\Xi \times \Xi} \|x - y\|^p \mathrm{d}\gamma(x, y) \right] \right)^{\frac{1}{p}}, (8)$$

where $\mathbb{Q}(\mu, \nu)$ is the set of measures on $\Xi \times \Xi$ with marginals μ and ν on the first and second factors. We define an ambiguity set $\mathcal{M}_N^r := \{\mu \in \mathcal{P}_p(\Xi) \mid W_p(\mu, \mathbb{P}_N) \leq r\}$ as a Wasserstein ball of distributions centered at $\mathbb{P}_N$ with radius r.

Remark V.1 (Choice of Wasserstein ball radius). There is a relationship between the sample size N and the Wasserstein radius r for constructing the ambiguity set $\mathcal{M}_N^r$. A distribution $\mathbb{P}$ is light-tailed if there exists an exponent ρ such that $A := \mathbb{E}_{\mathbb{P}}[\exp \|\boldsymbol{\xi}\|^\rho] = \int_\Xi \exp \|\boldsymbol{\xi}\|^\rho \mathbb{P}(d\boldsymbol{\xi}) < \infty$. If the true distribution $\mathbb{P}^*$ of $\boldsymbol{\xi}$ is light-tailed, the choice of radius $r = r_N(\bar{\epsilon})$ given in [20, Theorem 3.5],

$$r_N(\bar{\epsilon}) = \begin{cases} \left(\frac{\log(c_1 \bar{\epsilon}^{-1})}{c_2 N} \right)^{\frac{1}{\max\{k, 2\}}} & \text{if } N \geq \frac{\log(c_1 \bar{\epsilon}^{-1})}{c_2}, \\ \left(\frac{\log(c_1 \bar{\epsilon}^{-1})}{c_2 N} \right)^{\frac{1}{\rho}} & \text{else,} \end{cases} \quad (9)$$

where c_1 and c_2 are positive constants that depend on ρ, A and k (cf. [20, Theorem 3.4]), ensures that the ambiguity set $\mathcal{M}_N^{r_N(\bar{\epsilon})}$ contains $\mathbb{P}^*$ with probability at least $1 - \bar{\epsilon}$. $\bullet$

To account for the potential discrepancy between the empirical distribution $\mathbb{P}_N$ and the true distribution $\mathbb{P}^*$ of the system uncertainty $\boldsymbol{\xi}$ at run time, we aim to find a pair $(V^*, \boldsymbol{\pi}^*)$ that satisfies the following distributionally robust Lyapunov derivative constraint, $\forall \mathbf{x} \neq \mathbf{0}_n$,

$$\inf_{\mathbb{P} \in \mathcal{M}_N^r} \mathbb{P}(\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}) + \gamma \|\mathbf{x}\|) \leq 0) \geq 1 - \epsilon. (10)$$

Compared with the chance-constrained formulation in (5), the distributionally robust chance-constrained formulation (10) requires only a finite set of samples instead of the true distribution $\mathbb{P}^*$. Moreover, it offers robust constraint satisfaction guarantees against potential shifts of the uncertainty distribution within the constructed ambiguity set.

B. Distributionally Robust Stability Characterization

Next, we characterize the stability properties of the closed-loop system governed by a controller satisfying (10). If a pair $(V^*, \boldsymbol{\pi}^*)$ satisfies (10), the following result ensures that the closed-loop system satisfies a chance constraint under the true distribution.

Lemma V.2 (**Chance-constraint satisfaction under the true distribution**). *Assume the distribution $\mathbb{P}^*$ of $\boldsymbol{\xi}$ in (4) is light-tailed and the Wasserstein radius $r_N(\bar{\epsilon})$ is set according to (9). If the controller $\boldsymbol{\pi}^*(\mathbf{x})$ and Lyapunov function $V^*(\mathbf{x})$ pair satisfies (10) with $r = r_N(\bar{\epsilon})$, then $\forall \mathbf{x} \neq \mathbf{0}_n$,*

$$\mathbb{P}^*(\dot{V}^*(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi}) + \gamma\|\mathbf{x}\| \leq 0) \geq 1 - \epsilon - \bar{\epsilon}.(11)$$

Proof:
Consider the events $A := \{\mathbb{P}^* \in \mathcal{M}_N^{r_N(\bar{\epsilon})}\}$ and $B := \{\dot{V}^*(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi}) \leq 0 , \forall \mathbf{x} \neq \mathbf{0}_n\}$. Note that $\mathbb{P}^*(A \cap B) = \mathbb{P}^*(A) + \mathbb{P}^*(B) - \mathbb{P}^*(A \cup B) \geq \mathbb{P}^*(A) + \mathbb{P}^*(B) - 1$. From [20, Theorem 3.4], we have $\mathbb{P}^*(A) \geq 1 - \bar{\epsilon}$. From (10), we have that $\forall \mathbf{x} \neq \mathbf{0}_n$,

$$\inf_{\mathbb{P} \in \mathcal{M}_N^{r_N(\bar{\epsilon})}} \mathbb{P}(\dot{V}^*(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi}) + \gamma\|\mathbf{x}\| \leq 0) \geq 1 - \epsilon.$$

Hence, $\mathbb{P}^*(A \cap B) \geq 1 - \bar{\epsilon} + 1 - \epsilon - 1 = 1 - \epsilon - \bar{\epsilon}$. ∎

According to Lemma V.2, the closed-loop system satisfies the Lyapunov stability conditions pointwise in the state space with high probability. However, to analyze the global stability of the closed-loop system, it is necessary to extend this guarantee to trajectories of the system over time. To this end, the following result links pointwise Lyapunov stability to asymptotic stability in probability.

Lemma V.3 (**Global asymptotic stability in probability**). *Let $V^* : \mathbb{R}^n \rightarrow \mathbb{R}$ be a positive definite function with $V^*(\mathbf{0}_n) = 0$, and let V^* and the controller $\boldsymbol{\pi}^* : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a pair satisfying (11). Suppose $\boldsymbol{\xi} \sim \mathbb{P}^*$, then, the origin $\mathbf{0}_n$ of the closed-loop system $\dot{\mathbf{x}} = \tilde{\mathbf{f}}(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi})$ is globally asymptotically stable with probability at least $1 - \bar{\epsilon} - \epsilon$.*

Proof:
Define the sets

$$\mathcal{A} := \{\boldsymbol{\xi} \mid \dot{V}^*(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi}) + \gamma\|\mathbf{x}\| \leq 0 , \forall \mathbf{x} \neq \mathbf{0}_n\},$$

$$\mathcal{B} := \{\boldsymbol{\xi} \mid \text{the system is asymptotically stable at } \mathbf{x} = \mathbf{0}_n\}.$$

For $\boldsymbol{\xi} \in \mathcal{A}$, we have $\dot{V}^*(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi}) \leq -\gamma\|\mathbf{x}\|$, implying V^* decreases along the system trajectories, and leading to convergence to the equilibrium state $\mathbf{x} = \mathbf{0}$. Therefore, $\mathcal{A} \subseteq \mathcal{B}$, and we have $\mathbb{P}^*(\mathcal{B}) \geq \mathbb{P}^*(\mathcal{A}) \geq 1 - \epsilon - \bar{\epsilon}$, which concludes the result. ∎

Lemmas V.2 and V.3 establish global asymptotic stability in probability of the closed-loop system. Lemma V.2 demonstrates that if a pair $(V^*, \boldsymbol{\pi}^*)$ satisfies the distributionally robust derivative constraint (10), then the system fulfills the Lyapunov stability condition with high probability under the true distribution $\mathbb{P}^*$. Subsequently, Lemma V.3 extends this pointwise stability condition to the global asymptotic stability of the system. Therefore, satisfaction of the constraint in (10) ensures global asymptotic stability in probability.

Remark V.4 (**Exponential stability under additional conditions**). Under additional conditions, Lemma V.3 can ensure global exponential stability in probability. Consider the closed-loop system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi})$ with equilibrium at $\mathbf{0}_n$. Suppose there exists a positive constant α such that for all $\mathbf{x} \neq \mathbf{0}_n$, the pair $(V^*, \boldsymbol{\pi}^*)$ satisfies

$$\inf_{\mathbb{P} \in \mathcal{M}_N^r} \mathbb{P}(\dot{V}^*(\mathbf{x}, \boldsymbol{\pi}^*(\mathbf{x}), \boldsymbol{\xi}) + \gamma\|\mathbf{x}\| + \alpha V^*(\mathbf{x}) < 0) \geq 1 - \epsilon, (12)$$

and V^* satisfies

$$\alpha_1\|\mathbf{x}\|^p \leq V^*(\mathbf{x}) \leq \alpha_2\|\mathbf{x}\|^p, (13)$$

for some constants $\alpha_1, \alpha_2, p > 0$. Similar to the reasoning presented in Lemma V.2 and V.3, we conclude that the equilibrium of the closed-loop system is globally exponentially stable [49] with probability at least $1 - \epsilon - \bar{\epsilon}$. •

Remark V.5 (**Connections to other probabilistic stability notions**). In the literature, there exist other notions of stability in probability, especially for stochastic discrete-time systems, see e.g., [53]–[55]. For the stochastic discrete-time system:

$$\mathbf{x}_{i+1} = \tilde{\mathbf{f}}(\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\xi}_i), (14)$$

a popular stability notion refers to the probability that the state remains in a bounded region [55, Definition 6]. Formally, the system (14) is bounded in probability for some $K \in \mathbb{N}$ if there exists $M > 0$ and $\epsilon \in (0, 1)$ such that

$$\mathbb{P}^*\left\{ \max_{k \leq K} \|\mathbf{x}_k\| \leq M \right\} \geq 1 - \epsilon. (15)$$

This notion is widely used for ensuring finite-time stability and safety [56], [57] in stochastic systems. In typical stochastic systems, the parameter $\boldsymbol{\xi}$ varies as time progresses, which is a more general and challenging setting. In contrast, as outlined in (4), our work focuses on scenarios where $\boldsymbol{\xi}$ remains fixed over time, allowing us to derive stronger convergence results. Lemma V.3 not only certifies stability in probability over finite-time horizons but also extends to global asymptotic stability guarantees in probability. •

C. Convex Reformulation of Distributionally Robust Stability Constraint

Synthesizing a pair $(V^*, \boldsymbol{\pi}^*)$ that satisfies (10) is challenging because the constraint encompasses an infimum over a set of probability measures. We leverage recent advances [20], [21] in distributionally robust optimization to identify a sufficient condition in the form of a convex constraint.

Proposition V.6 (**Convex distributionally robust stability condition**). *Let $\epsilon \leq \frac{1}{N}$. Then, the following*

$$\frac{r}{\epsilon}\|\mathbf{W}^\top(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}))\nabla V(\mathbf{x})\| + \max_i(\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}_i))$$
$$+ \gamma\|\mathbf{x}\| \leq 0, \ \forall \mathbf{x} \neq \mathbf{0}_n \quad (16)$$

is a sufficient condition for (10) *to hold with ambiguity set corresponding to* 1-*Wasserstein distance.*

Proof:
Similar to the CVaR approximation (6) of the chance constraint (5), the following is a sufficient condition for (10):

$$\sup_{\mathbb{P} \in \mathcal{M}_N^r} \text{CVaR}_{1-\epsilon}^{\mathbb{P}}(\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}) + \gamma\|\mathbf{x}\|)) \leq 0.(17)$$

Using (7), we write (17) equivalently as

$$\sup_{\mathbb{P}\in\mathcal{M}_N^r}\inf_{t\in\mathbb{R}}[\mathbb{E}_\mathbb{P}[(\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi})+\gamma\|\mathbf{x}\|+t)_+]-t\epsilon]\le 0. \quad (18)$$

Based on [21, Lemma V.8] and [20, Theorem 6.3], with the 1-Wasserstein distance, the supremum over the ambiguity set can be written conservatively with the sample average over the empirical distribution $\mathbb{P}_N$ and a regularization term:

$$rL_V(\mathbf{x})+\inf_{t\in\mathbb{R}}\left[\mathbb{E}_{\mathbb{P}_N}[(\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi})+\gamma\|\mathbf{x}\|+t)_+]-t\epsilon\right]\le 0, \quad (19)$$

where $L_V(\mathbf{x}) : \mathcal{X} \mapsto \mathbb{R}_{>0}$ is the Lipschitz constant of $\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi})$ in $\boldsymbol{\xi}$. Note that $\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi}) = \nabla V(\mathbf{x})^\top(\mathbf{f}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}))+\mathbf{W}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}))\boldsymbol{\xi})$. Since the Lipschitz constant of a differentiable affine function equals the dual norm of its gradient, and the dual norm of the L_2-norm is itself, we have the convex function L_V as

$$L_V(\mathbf{x}) = \|\mathbf{W}^\top(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}))\nabla V(\mathbf{x})\|. \quad (20)$$

Next, we bound the second term in (19),

$$\inf_{t\in\mathbb{R}}\left[\mathbb{E}_{\mathbb{P}_N}[(\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi})+\gamma\|\mathbf{x}\|+t)_+]-t\epsilon\right]. \quad (21)$$

Observe that the function $\frac{1}{N}\sum_{i=1}^N(\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi}_i)+\gamma\|\mathbf{x}\|+t)_+ - t\epsilon$ is piecewise-linear in t with $N+1$ intervals and N breakpoints, given by $\{-\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi}_i) - \gamma\|\mathbf{x}\|\}_{i\in[N]}$ and the slope for the i-th interval is $\frac{i-1}{N} - \epsilon$. Thus, the optimal solution is $t^* = -\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi}_j) - \gamma\|\mathbf{x}\|$, where k satisfies $\frac{j-1}{N} - \epsilon < 0$ and $\frac{j}{N} - \epsilon \ge 0$. Since $\epsilon \le \frac{1}{N}$, only the first interval has negative slope and (21) can be written equivalently as

$$\max_i \epsilon(\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi}_i)+\gamma\|\mathbf{x}\|).$$

Therefore, given that $\epsilon \le \frac{1}{N}$, we write $\mathbf{W}^\top(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}))$ as $\mathbf{W}^\top(\mathbf{x})$ for brevity and conclude that

$$r\|\mathbf{W}^\top(\mathbf{x})\nabla V(\mathbf{x})\| + \max_i \epsilon(\dot{V}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}),\boldsymbol{\xi}_i)+\gamma\|\mathbf{x}\|)\le 0$$

is equivalent to (19), and thus sufficient for (10). ∎

Based on Proposition V.6, our goal is to find a controller $\boldsymbol{\pi}^*(\mathbf{x})$ and Lyapunov function $V^*(\mathbf{x})$ pair that fulfills condition (16). Once such a pair is determined, leveraging Lemmas V.2 and V.3, we can ascertain with high confidence that the system (4), governed by $\boldsymbol{\pi}^*(\mathbf{x})$, exhibits (exponential) asymptotic stability of the equilibrium.

VI. Distributionally Robust Lyapunov-Stable Controller Learning

In this section, we present a learning-based approach for finding a distributionally robust controller $\boldsymbol{\pi}^*(\mathbf{x})$ and Lyapunov function $V^*(\mathbf{x})$ pair satisfying (16). We make the following assumptions.

Assumption 1 (**Lipschitz Continuity and Boundedness**). *Assume $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{U} \subset \mathbb{R}^m$ are compact, and the nominal system dynamics $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and the perturbation function $\mathbf{W} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n\times k}$ are Lipschitz on $\mathcal{X} \times \mathcal{U}$ with*

constants L_f and L_W, respectively, i.e., for all $\mathbf{x},\mathbf{x}' \in \mathcal{X}$ and $\mathbf{u},\mathbf{u}' \in \mathcal{U}$,

$$\|\mathbf{f}(\mathbf{x},\mathbf{u}) - \mathbf{f}(\mathbf{x}',\mathbf{u}')\| \le L_f(\|\mathbf{x}-\mathbf{x}'\| + \|\mathbf{u}-\mathbf{u}'\|),$$
$$\|\mathbf{W}(\mathbf{x},\mathbf{u}) - \mathbf{W}(\mathbf{x}',\mathbf{u}')\| \le L_W(\|\mathbf{x}-\mathbf{x}'\| + \|\mathbf{u}-\mathbf{u}'\|).$$

Lipschitzness together with the compactness of the domain implies that both $\mathbf{f}$ and $\mathbf{W}$ are bounded on $\mathcal{X} \times \mathcal{U}$. Specifically, there exist constants $B_f, B_W \in \mathbb{R}_{>0}$ such that, for any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{u} \in \mathcal{U}$,

$$\|\mathbf{f}(\mathbf{x},\mathbf{u})\| \le B_f \quad \text{and} \quad \|\mathbf{W}(\mathbf{x},\mathbf{u})\| \le B_W.$$

The Lipschitz continuity assumption is particularly useful in our setting, as the neural networks are trained with discrete data, and we aim to certify stability for the entire state space. Lipschitzness ensures that the system dynamics do not change drastically for nearby states, which is crucial for the stability proofs presented later in this work.

A. Model Architecture and Loss Function

To parameterize the desired pair $(V^*(\mathbf{x}), \boldsymbol{\pi}^*(\mathbf{x}))$, we consider the following neural network representations. For the Lyapunov function, termed NN-LF, we set

$$V_{\boldsymbol{\theta}_1}(\mathbf{x}) := \|\phi_{\boldsymbol{\theta}_1}(\mathbf{x}) - \phi_{\boldsymbol{\theta}_1}(\mathbf{0}_n)\|^2 + \hat{\alpha}\|\mathbf{x}\|^2, \quad (22)$$

where $\phi_{\boldsymbol{\theta}_1} : \mathbb{R}^n \mapsto \mathbb{R}$ is a fully-connected neural network with parameters $\boldsymbol{\theta}_1$ and $\tanh$ activations, and $\hat{\alpha}$ is a user-chosen parameter [11], [12]. By construction, this function is positive definite and $V_{\boldsymbol{\theta}_1}(\mathbf{0}_n) = 0$. Furthermore, the use of $\tanh$ activations ensures continuous differentiability and bounded derivatives. Therefore, the gradient $\nabla V_{\boldsymbol{\theta}_1}(\mathbf{x})$ is Lipschitz with some constant $L_{\nabla V}$ and its norm $\|\nabla V_{\boldsymbol{\theta}_1}(\mathbf{x})\|$ is bounded within the compact set $\mathcal{X}$ by some constant B_V. As shown in [58], these constants can be estimated explicitly, but we will only rely on their existence for stability analysis.

For the controller, we set

$$\boldsymbol{\pi}_{\boldsymbol{\theta}_2}(\mathbf{x}) := \varphi_{\boldsymbol{\theta}_2}(\mathbf{x}) - \varphi_{\boldsymbol{\theta}_2}(\mathbf{0}_n), \quad (23)$$

where $\varphi_{\boldsymbol{\theta}_2}$ is a neural network with parameters $\boldsymbol{\theta}_2$ and $\tanh$ activations. By construction, this formulation guarantees $\boldsymbol{\pi}_{\boldsymbol{\theta}_2}(\mathbf{0}_n) = \mathbf{0}_m$ and ensures that $\boldsymbol{\pi}_{\boldsymbol{\theta}_2}(\mathbf{x})$ is locally Lipschitz in $\mathcal{X}$, with constant L_π.

When employing a neural-network Lyapunov function, the Lyapunov derivative condition in (3b) may not hold within a small neighborhood of the equilibrium due to numerical inaccuracies, see e.g., [9], [12]. These inaccuracies stem from a neural network's inability to precisely approximate the derivative's behavior near the equilibrium, where its values are close to zero. This observation leads us to focus on establishing the validity of learned controller and Lyapunov function pairs outside of a neighborhood close to the origin. Therefore, our analysis concentrates on the region $\mathcal{X}_\delta := \mathcal{X} \setminus \overline{B(\mathbf{0}_n;\delta)}$, which excludes a δ-radius ball around the equilibrium.

Definition VI.1 (δ-**accurate Lyapunov function over** $\mathcal{X}$). *A Lyapunov function $V(\mathbf{x})$ for a dynamical system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\boldsymbol{\pi}(\mathbf{x}))$ is δ-accurate if it is positive definite, $V(\mathbf{0}_n) = 0$,*

and satisfies the Lyapunov derivative conditions (3b) *every-where in $\mathcal{X}$ except possibly in the closed ball* $\overline{B(\mathbf{0}_n; \delta)}$.

The existence of a δ-accurate Lyapunov function implies the ultimate boundedness of the closed-loop system trajectories within a closed ball $\overline{B(\mathbf{0}_n; \delta)}$. For a neural network representation, $\delta > 0$ can be chosen arbitrarily small, potentially requiring a larger network size for $V_{\boldsymbol{\theta}_1}$ and $\boldsymbol{\pi}_{\boldsymbol{\theta}_2}$ and greater sampling density for the training set. For our distributionally robust formulation, we introduce a distributionally robust δ-accurate Lyapunov function as follows.

Definition VI.2 (**Distributionally robust δ-accurate Lyapunov function with ϵ margin of error over $\mathcal{X}$**). *A Lyapunov function $V(\mathbf{x})$ for a dynamical system $\dot{\mathbf{x}} = \tilde{\mathbf{f}}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi})$ in (4) is distributionally robust δ-accurate if it is positive definite, $V(\mathbf{0}_n) = 0$, and satisfies*

$$\inf_{\mathbb{P} \in \mathcal{M}_N^r} \mathbb{P}(\dot{V}(\mathbf{x}, \boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\xi}) + \gamma\|\mathbf{x}\|) \le 0) \ge 1 - \epsilon$$

everywhere in $\mathcal{X}$ except possibly in the closed ball $\overline{B(\mathbf{0}_n; \delta)}$.

Arguments analogous to the ones we employed in the proofs of Lemmas V.2 and V.3 show that the existence of a distributionally robust δ-accurate Lyapunov function implies the stability of the closed-loop system to the closed ball $\overline{B(\mathbf{0}_n; \delta)}$ with high probability.

To facilitate the training process of the pair $(V_{\boldsymbol{\theta}_1}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2})$, a training set $\mathcal{D}_{\text{LF}} := \{\mathbf{x}_i\}_{i \in [M]}$ is constructed by sampling $\{\mathbf{x}_i\}_{i \in [M]}$ uniformly from $\mathcal{X}_\delta$. We also assume the availability of the uncertainty training set $\mathcal{D}_{\boldsymbol{\xi}} := \{\boldsymbol{\xi}_i\}_{i \in [N]}$, collected by offline measurements of the system.

For a system without any uncertainty, we minimize the following empirical loss function that encourages the satisfaction of the Lyapunov derivative constraint (3b):

$$\ell_{\text{Nominal}}(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^M (\dot{V}_{\boldsymbol{\theta}_1}(\mathbf{x}_i, \boldsymbol{\pi}_{\boldsymbol{\theta}_2}(\mathbf{x}_i)) + \gamma\|\mathbf{x}_i\|)_+, (24)$$

where $\gamma \in \mathbb{R}_{>0}$ is a user-chosen parameter.

To address the model uncertainty in (4), we aim to find a distributionally robust (DR)-NN-LF $V_{\boldsymbol{\theta}_1}$ and a controller $\boldsymbol{\pi}_{\boldsymbol{\theta}_2}$ pair. For the DR formulation, using the uncertainty training set $\mathcal{D}_{\boldsymbol{\xi}}$ and assuming $\epsilon \le \frac{1}{N}$, as required by Proposition V.6, we design the following empirical loss function for learning the DR-NN-LF and controller,

$$\ell_{\text{DR}}(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^M (\frac{r}{\epsilon}\|\mathbf{W}^\top(\mathbf{x}_i, \boldsymbol{\pi}_{\boldsymbol{\theta}_2}(\mathbf{x}_i))\nabla V_{\boldsymbol{\theta}_1}(\mathbf{x}_i)\| +$$
$$\max_j (\dot{V}_{\boldsymbol{\theta}_1}(\mathbf{x}_i, \boldsymbol{\pi}_{\boldsymbol{\theta}_2}(\mathbf{x}_i), \boldsymbol{\xi}_j)) + \gamma\|\mathbf{x}_i\|)_+. \quad (25)$$

In the following, for a sufficiently large M, we aim to show that the minimizer of the loss function (25) yields a pair $(V_{\boldsymbol{\theta}_1^*}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*})$ satisfying the condition in (10).

B. Nominal Neural Lyapunov-Stable Control

First, we show based on results from [12] that for a sufficiently large M, the nominal stabilizing controller learned from (24) stabilizes the nominal system. We formalize the

necessary sampling density required to ensure stability using the learned LF.

Lemma VI.3. *([12, Lemma 4]). For any $\delta > 0$ and a chosen parameter $c > 0$, there exists a minimum number of samples $M(\delta, c) \in \mathbb{N}$ such that, for all $M \ge M(\delta, c)$, a uniformly sampled training dataset $\mathcal{D}_{LF} = \{\mathbf{x}_i\}_{i \in [M]} \subset \mathcal{X}_\delta$ ensures that the domain $\mathcal{X}_\delta$ can be covered by the union of balls $B(\mathbf{x}_i; c\|\mathbf{x}_i\|)$. Specifically, this guarantees that for any $\mathbf{x} \in \mathcal{X}_\delta$, there exists $\mathbf{x}_i \in \mathcal{D}_{LF}$ satisfying $\|\mathbf{x} - \mathbf{x}_i\| \le c\|\mathbf{x}_i\|$.*

This result underscores the relation between δ, c, and the sampling density M, crucial for approximating the controller and Lyapunov function effectively.

Lemma VI.4 (**Neural Lyapunov-stable control**). *Let $\mathcal{D}_{LF} = \{\mathbf{x}_i\}_{i \in [M]} \subset \mathcal{X}_\delta$ be a uniformly sampled training set, with $M \ge M(\delta, c)$ as defined in Lemma VI.3. Let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)$ be the trained parameters such that $\ell_{Nominal}(\boldsymbol{\theta}^*) = 0$. Denote by $L_{\nabla V}$ and L_π the Lipschitz constants of $\nabla V_{\boldsymbol{\theta}_1^*}$ and $\boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}$, respectively, and let B_V be the bound on $\|\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})\|$ for $\mathbf{x} \in \mathcal{X}$. Let L_f and B_f be as defined in Assumption 1. If $c > 0$ is sufficiently small to ensure $\gamma - (L_f(L_\pi + 1)B_V + L_{\nabla V}B_f)c > 0$, then the controller $\boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}$ stabilizes the nominal system $\mathbf{f}$ to the closed ball $\overline{B(\mathbf{0}_n; \delta)}$ as certified by $V_{\boldsymbol{\theta}_1^*}$, which is a δ-accurate Lyapunov function for the controlled system over $\mathcal{X}$.*

Proof:
Given that $\ell_{\text{Nominal}}(\boldsymbol{\theta}^*) = 0$, we have

$$\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}_i, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{x}_i)) + \gamma\|\mathbf{x}_i\| \le 0, (26)$$

for all $\mathbf{x}_i \in \mathcal{D}_{\text{LF}}$. Since $\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})$ and $\boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}$ are Lipschitz, based on the Assumption 1, we have for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\delta$,

$$\|\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x}) - \nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{y})\| \le L_{\nabla V}\|\mathbf{x} - \mathbf{y}\|, \quad (27)$$
$$\|\mathbf{f}(\mathbf{x}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{x})) - \mathbf{f}(\mathbf{y}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{y}))\| \le L_f(L_\pi + 1)\|\mathbf{x} - \mathbf{y}\|.$$

We denote $L_q := L_f(L_\pi + 1)$ and write $\mathbf{f}(\mathbf{x}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{x}))$ as $\mathbf{f}(\mathbf{x})$ for brevity. Based on (27), we have for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\delta$,

$$|\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{x})) - \dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{y}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{y}))| =$$
$$|[\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})]^\top \mathbf{f}(\mathbf{x}) - [\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{y})]^\top \mathbf{f}(\mathbf{y})| \le$$
$$\|\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})\|\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| + \|\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x}) - \nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{y})\|\|\mathbf{f}(\mathbf{y})\|$$
$$\le (L_q\|\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})\| + L_{\nabla V}\|\mathbf{f}(\mathbf{y})\|)\|\mathbf{x} - \mathbf{y}\|$$
$$\le (L_q B_V + L_{\nabla V}B_f)\|\mathbf{x} - \mathbf{y}\|. \quad (28)$$

This establishes that $L_{\dot{V}}$, the Lipschitz constant for the Lyapunov derivative, can be bounded by $L_q B_V + L_{\nabla V}B_f$. By Lemma VI.3, for any $\mathbf{x} \in \mathcal{X}_\delta$, there exists a nearby sampled point $\mathbf{x}_i$ such that $\|\mathbf{x} - \mathbf{x}_i\| \le c\|\mathbf{x}_i\|$. Using (26) and (28),

$$\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{x})) \le \dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}_i, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{x}_i)) + L_{\dot{V}}c\|\mathbf{x}_i\| \quad (29)$$
$$\le -\gamma\|\mathbf{x}_i\| + L_{\dot{V}}c\|\mathbf{x}_i\|$$
$$\le -(\gamma - L_{\dot{V}}c)\delta.$$

Since $\gamma - L_{\dot{V}}c > 0$, this ensures that $\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}, \boldsymbol{\pi}_{\boldsymbol{\theta}_2^*}(\mathbf{x}))$ satisfies the Lyapunov derivative condition for all $\mathbf{x} \in \mathcal{X}_\delta$, and thus

it is a δ-accurate Lyapunov function. Therefore, the learned controller provides δ-accurate stabilization of the nominal system $\mathbf{f}$ as certified by the Lyapunov function. ∎

The stability assurances provided by Lemma VI.4 may not hold in the presence of the uncertainties in (4), which motivates our study of the system under the distributionally robust controller learned from (25).

C. Neural Distributionally Robust Lyapunov-Stable Control

For a sufficiently large sampling density, we establish probabilistic δ-accurate stability guarantees for the uncertain system (4) governed by the controller learned from (25).

Proposition VI.5 (Distributionally robust neural Lyapunov-stable control). *Let $\mathcal{D}_{LF} = \{\mathbf{x}_i\}_{i \in [M]} \subset \mathcal{X}_\delta$ be a uniformly sampled training set, with $M \geq M(\delta, c)$ as defined in Lemma VI.3. Let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)$ be the trained parameters such that $\ell_{DR}(\boldsymbol{\theta}^*) = 0$. Let $L_{\nabla V}$ and L_π denote the Lipschitz constants of $\nabla V_{\boldsymbol{\theta}_1^*}$ and $\pi_{\boldsymbol{\theta}_2^*}$, respectively, and let B_V be the bound on $\|\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})\|$ for $\mathbf{x} \in \mathcal{X}$ and $\epsilon \leq \frac{1}{N}$. Let $c > 0$ be sufficiently small such that $\gamma - ((L_W(L_\pi + 1))B_V + L_{\nabla V}B_W)B_\xi \frac{r}{\epsilon} c - (L_f(L_\pi + 1)B_V + L_{\nabla V}B_f)c > 0$, where L_W, L_f, B_W, and B_f are as defined in Assumption 1, and $B_\xi \in \mathbb{R}_{>0}$ is such that $\|\boldsymbol{\xi}\| \leq B_\xi$ for all $\boldsymbol{\xi} \in \Xi$. Then, the controller $\pi_{\boldsymbol{\theta}_2^*}$ stabilizes the uncertain system (4) to the closed ball $\overline{B(\mathbf{0}_n; \delta)}$ with high probability, as certified by $V_{\boldsymbol{\theta}_1^*}$, which is a distributionally robust δ-accurate Lyapunov function for the controlled system over $\mathcal{X}$.*

Proof:
Given that $\ell_{\mathrm{DR}}(\boldsymbol{\theta}^*) = 0$, we have for all $\mathbf{x}_i \in \mathcal{D}_{\mathrm{LF}}$,

$$\frac{r}{\epsilon}\|\mathbf{W}^\top(\mathbf{x}_i, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}_i))\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x}_i)\|_\infty +$$
$$\max_j(\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}_i, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}_i), \boldsymbol{\xi}_j)) + \gamma\|\mathbf{x}_i\| \leq 0.$$

Based on Assumption 1 and the Lipschitz constant L_π of the controller $\pi_{\boldsymbol{\theta}_2^*}$, we have for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\delta$:

$$\|\mathbf{W}(\mathbf{x}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x})) - \mathbf{W}(\mathbf{y}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}))\| \leq L_W(L_\pi + 1)\|\mathbf{x} - \mathbf{y}\|. \quad (30)$$

For brevity, we denote $L_p = L_W(L_\pi + 1)$ and write $\mathbf{W}(\mathbf{x}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}))$ and $V_{\boldsymbol{\theta}_1^*}(\mathbf{x})$ as $\mathbf{W}(\mathbf{x})$ and $V(\mathbf{x})$, respectively. Using (27) and (30), we have for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\delta$:

$$\|\mathbf{W}^\top(\mathbf{x})\nabla V(\mathbf{x}) - \mathbf{W}^\top(\mathbf{y})\nabla V(\mathbf{y})\| \leq$$
$$\|(\mathbf{W}^\top(\mathbf{x}) - \mathbf{W}^\top(\mathbf{y}))\nabla V(\mathbf{x})\| + \|\mathbf{W}^\top(\mathbf{y})(\nabla V(\mathbf{x}) - \nabla V(\mathbf{y}))\|$$
$$\leq (L_p\|\nabla V(\mathbf{x})\| + L_{\nabla V}\|\mathbf{W}^\top(\mathbf{y})\|)\|\mathbf{x} - \mathbf{y}\|$$
$$\leq (L_p B_V + L_{\nabla V}B_W)\|\mathbf{x} - \mathbf{y}\|. \quad (31)$$

Thus, the term $\mathbf{W}^\top(\mathbf{x}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}))\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})$ is Lipschitz with constant $L_{\dot{W}}$, bounded by $L_p B_V + L_{\nabla V}B_W$. Moreover, as $\boldsymbol{\xi} \in \Xi$ and Ξ is compact, we know that $\|\boldsymbol{\xi}\| \leq B_\xi$ for some constant $B_\xi \in \mathbb{R}_{>0}$. Using (28) and (31), for the term $\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}), \boldsymbol{\xi})$, we have that, for each $\boldsymbol{\xi}$, and for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\delta$,

$$|\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}), \boldsymbol{\xi}) - \dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{y}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{y}), \boldsymbol{\xi})|$$

$$= |[\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})]^\top \mathbf{f}(\mathbf{x}) - [\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{y})]^\top \mathbf{f}(\mathbf{y})$$
$$+ ([\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{x})]^\top \mathbf{W}(\mathbf{x}) - [\nabla V_{\boldsymbol{\theta}_1^*}(\mathbf{y})]^\top \mathbf{W}(\mathbf{y}))\boldsymbol{\xi}|$$
$$\leq L_{\dot{V}}\|\mathbf{x} - \mathbf{y}\| + L_{\dot{W}}\|\boldsymbol{\xi}\|\|\mathbf{x} - \mathbf{y}\|$$
$$\leq (L_{\dot{V}} + L_{\dot{W}}B_\xi)\|\mathbf{x} - \mathbf{y}\|. \quad (32)$$

We denote $L_{\max} := L_{\dot{V}} + L_{\dot{W}}B_\xi$. We define $G(\mathbf{x}) = \max_j \dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}), \boldsymbol{\xi}_j)$. From (32), we deduce $|G(\mathbf{x}) - G(\mathbf{y})| \leq L_{\max}\|\mathbf{x} - \mathbf{y}\|$, and thus G is also Lipschitz with constant $L_{\max}$.

Lastly, following the covering argument in Lemma VI.3, we have that for any $\mathbf{x} \in \mathcal{X}_\delta$, there exists an $\mathbf{x}_i \in \mathcal{D}_{\mathrm{LF}}$ such that $\|\mathbf{x} - \mathbf{x}_i\| \leq c\|\mathbf{x}_i\|$. Therefore, $\forall \mathbf{x} \in \mathcal{X}_\delta$,

$$\frac{r}{\epsilon}\|\mathbf{W}^\top(\mathbf{x})\nabla V(\mathbf{x})\| + \max_j(\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}), \boldsymbol{\xi}_j))$$
$$\leq \frac{r}{\epsilon}\|\mathbf{W}^\top(\mathbf{x}_i)\nabla V(\mathbf{x}_i)\| + \max_j(\dot{V}_{\boldsymbol{\theta}_1^*}(\mathbf{x}_i, \pi_{\boldsymbol{\theta}_2^*}(\mathbf{x}_i), \boldsymbol{\xi}_j)) +$$
$$\left(\frac{r}{\epsilon}L_{\dot{W}}c + L_{\max}c\right)\|\mathbf{x}_i\| \leq -\left(\gamma - \frac{r}{\epsilon}L_{\dot{W}}c - L_{\max}c\right)\delta. \quad (33)$$

Note that $\gamma - \frac{r}{\epsilon}L_{\dot{W}}c - L_{\max}c > 0$. Thus, we have that $\frac{r}{\epsilon}\|\mathbf{W}^\top(\mathbf{x})\nabla V(\mathbf{x})\| + \max_j(\dot{V}(\mathbf{x}, \pi(\mathbf{x}), \boldsymbol{\xi}_j)) < 0$ holds for all $\mathbf{x} \in \mathcal{X}_\delta$.

Therefore, reasoning as in Proposition V.6, we conclude that, for any $\boldsymbol{\xi} \sim \mathbb{P}^*$, (10) is satisfied for all $\mathbf{x} \in \mathcal{X}_\delta$. This implies that $V_{\boldsymbol{\theta}_1^*}$ is a distributionally robust δ-accurate Lyapunov function. Therefore, the learned controller provides distributionally robust δ-accurate stabilization of the uncertain system, as certified by the Lyapunov function. ∎

The learning process for the DR-NN-LF and controller pair, as defined in (25), can be made more efficient by initializing the neural network weights with those obtained from the nominal training (24). This initialization strategy significantly reduces the required training time and computational resources for the DR formulation.

VII. Evaluation

This section evaluates our approach for synthesizing distributionally robust Lyapunov-stable controllers. We focus on two classic control problems from Gymnasium [59]: the Inverted Pendulum and the Continuous-time Mountain Car. For each system, we consider multiple instances with varying characteristics, each defined by its own physical parameters such as mass, length, and friction. These instances are used for training our distributionally robust Lyapunov-stable controller (25).

We use PyTorch [60] to train our neural network Lyapunov function and controller pair, both for the baseline case (24) and for the distributionally robust (DR) formulation (25). The models are trained using the Adam optimizer [61] with a learning rate of 0.003. The PyTorch clamp function is used to enforce input bounds for the control policy network, which is implemented with a smooth approximation and enables gradient back-propagation. The training was done on hardware with an Nvidia GeForce RTX 2070 Super GPU and an Intel i7 9700K CPU.

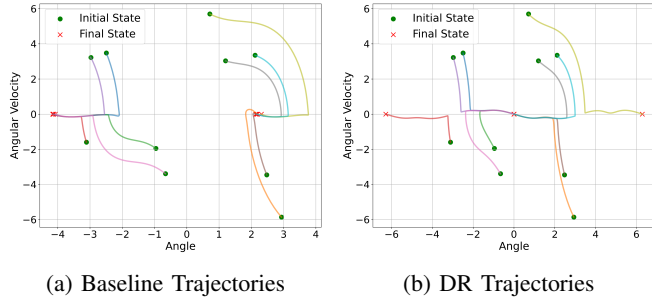(a) Baseline Trajectories   (b) DR Trajectories

FIGURE 2: Comparison of trajectories for the inverted pendulum system with test case parameters: mass $m = 1.1$, length $l = 1.0$, and damping $b = 0.18$. The 10 random sampled initial states are marked as green dots, while the final states are marked as red crosses. In (a), the baseline controller, trained using the average mass and damping from offline observations, fails to stabilize the pendulum to $(0, 0)$. In (b), the distributionally robust (DR) controller successfully stabilizing the pendulum to an upright position for every initial state, demonstrating improved robustness to distributional shifts in the system parameters.

During testing, we assume the physical parameters of each system may be drawn from distributions different from those of the training instances. This setup aims to assess the performance of the learned controller under realistic scenarios, where distributional shifts in the system parameters are common. By doing so, we effectively evaluate the robustness and adaptability of our controller in the presence of distributional uncertainty in the system models.

A. Inverted Pendulum
The inverted pendulum is a standard nonlinear control problem for testing control methods. The system consists of two state variables, angular position θ and angular velocity $\dot{\theta}$, and one control input u. The system dynamics are:

$$\underbrace{\begin{bmatrix} \dot{\theta} \\ \ddot{\theta} \end{bmatrix}}_{\dot{\mathbf{x}}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) := \begin{bmatrix} \dot{\theta} \\ \frac{mgl\sin\theta - b\dot{\theta}}{ml^2} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{ml^2} \end{bmatrix} u \quad (34)$$

where $g = 9.81$ m/s^2 is the gravity acceleration, $m = 1.0$ kg is the mass, $l = 1.0$ m is the length, and $b = 0.13$ N·m·s/rad is the damping coefficient. The equilibrium state is defined as $[\theta, \dot{\theta}] = [0, 0]$, which corresponds to the upright equilibrium position.

For training both the baseline and the DR controllers, we generate a training dataset by uniformly sampling $\{\mathbf{x}_i\}_{i \in [3600]}$ from the box region defined by $0 \le \theta \le 2\pi$ and $-8 \le \dot{\theta} \le 8$.

We assume uncertainties in mass and damping, represented by ξ_1 and ξ_2, respectively. The uncertain system dynamics are:

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\theta} \\ \frac{(m+\xi_1)gl\sin\theta - (b+\xi_2)\dot{\theta}}{(m+\xi_1)l^2} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{(m+\xi_1)l^2} \end{bmatrix} u. \quad (35)$$

Through first-order Taylor expansion around the nominal parameters, we derive the perturbation matrices $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2]$

for the inverted pendulum subject to model uncertainty:

$$\mathbf{w}_1(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} 0 \\ \frac{b\dot{\theta} - u}{m^2 l^2} \end{bmatrix}, \quad \mathbf{w}_2(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} 0 \\ -\frac{\dot{\theta}}{ml^2} \end{bmatrix}. \quad (36)$$

We assume that a set of offline samples $\{\boldsymbol{\xi}_i\}_{i=1}^5$ of the uncertainties $\boldsymbol{\xi} = [\xi_1, \xi_2]^\top$ are available for training the controller and certificate pairs. We take $\xi_1 \sim \mathcal{U}(-0.04, 0.08)$, $\xi_2 \sim \mathcal{N}(0.0, 0.02)$, where $\mathcal{U}$ and $\mathcal{N}$ denote uniform and normal distributions, respectively. However, during test time, the test uncertainty parameters are set to $\xi_1 = 0.1$ and $\xi_2 = 0.05$. These test values represent a significant deviation from the training distributions, allowing us to evaluate the controllers' robustness to distributional shifts in parameters.

To train the baseline controller, we compute the average mass $\tilde{m}$ and the average damping $\tilde{b}$ from the five offline samples. These average values are then used in the baseline system dynamics and the loss function (24) to learn the baseline controller and Lyapunov function pair.

In contrast, for training the DR controller, we set the Wasserstein radius to $r = 0.02$ and the risk tolerance to $\epsilon = 0.1$. The 5 uncertainty samples $\{\boldsymbol{\xi}_i\}_{i=1}^5$ are then used in the loss function (25) to learn the DR controller and its corresponding Lyapunov function.

Fig. 2 illustrates the performance of the baseline and DR controllers with the test uncertainty parameters $\xi_1 = 0.1$ and $\xi_2 = 0.05$. To assess the controllers' performance, we randomly sample 10 initial states within the box region defined by $0 \le \theta \le 2\pi$ and $-6 \le \dot{\theta} \le 6$. Each controller is applied to the system and the resulting trajectories are simulated.

In Fig. 2a, we observe that the baseline controller, trained using only the average mass and damping values, fails to stabilize the inverted pendulum system to the desired upright equilibrium. For all the 10 initial states, the baseline controller's trajectories converge to states that are not the desired equilibrium. This can be attributed to the increased damping ($\xi_2 = 0.05$) and the heavier pendulum mass ($\xi_1 = 0.1$) in the test scenario. The baseline controller, designed based on the average parameter values, lacks the necessary robustness to compensate for the increased damping and mass.

On the other hand, Fig. 2b shows that all trajectories converge to the desired equilibrium under the DR controller. Despite the presence of distributional shift in the model uncertainty, the DR controller successfully stabilizes the inverted pendulum in an upright position. This robustness can be attributed to the DR controller's training process, which explicitly takes into account the distributional information of the uncertainty and optimizes for worst-case performance within the constructed ambiguity set.

To further investigate the performance of the proposed approach, we compare it with state-of-the-art reinforcement learning (RL) algorithms, such as the Soft Actor-Critic algorithm (SAC) [62] and the Proximal Policy Optimization algorithm (PPO) [63]. The reward function used for training the RL algorithms is defined as $r = -(\theta^2 + 0.1\dot{\theta}^2 + 0.001u^2)$.

(a) Lyapunov Function Values   (b) Control Inputs   (c) RL Value Functions   (d) RL Control Inputs
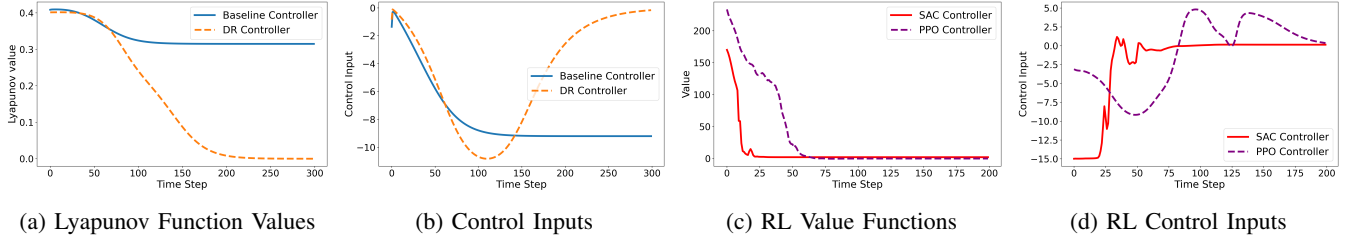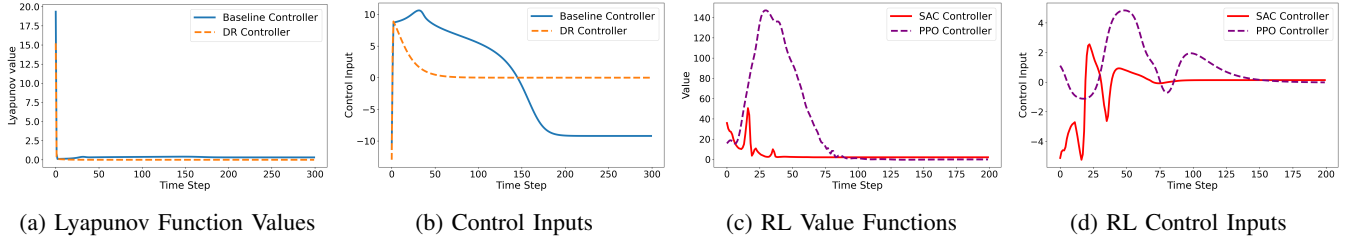
FIGURE 3: Comparison of Lyapunov function values, control inputs, and RL value functions for the initial state $(\pi, 0)$. (a) Lyapunov function values over time for the baseline and DR controllers. (b) Control inputs over time for the baseline and DR controllers. (c) Value function values over time for the SAC and PPO algorithms. (d) Control inputs generated by the SAC and PPO algorithms.

(a) Lyapunov Function Values   (b) Control Inputs   (c) RL Value Functions   (d) RL Control Inputs

FIGURE 4: Comparison of Lyapunov function values, control inputs, and RL value functions for the initial state $(-\pi/2, 5.5)$. (a) Lyapunov function values over time for the baseline and DR controllers. (b) Control inputs over time for the baseline and DR controllers. (c) Value function values over time for the SAC and PPO algorithms. (d) Control inputs generated by the SAC and PPO algorithms.

Figs. 3 and 4 present a qualitative comparison of the baseline controller, our proposed DR controller, and the two RL policies over time for two different initial states. In Fig. 3, the initial state is the downright position $(\theta, \dot{\theta}) = (\pi, 0)$, from which the pendulum has to swing up to the desired equilibrium. We observe that the Lyapunov function corresponding to the baseline controller fails to converge to zero, while the proposed DR Lyapunov function values successfully converge to zero, as illustrated in Fig. 3a. Although the value functions learned by the two RL algorithms eventually reach zero, they exhibit non-monotonic behavior, indicating that the stability of the system cannot be certified using these learned value functions (Fig. 3c).

In Fig. 4, we change the initial state to $(-\pi/2, 5.5)$. In this scenario, the non-monotonic behavior of the value functions learned by the RL algorithms becomes more apparent, as shown in Fig. 4c. This observation highlights the advantage of the proposed DR approach, which provides provable stability guarantees through the use of a certified Lyapunov function. The DR controller ensures that the Lyapunov function values monotonically decrease towards zero with high probability (Fig. 4a), thereby enabling robustness to distributional shift in the system parameters.

In conclusion, while the RL algorithms can learn control policies effectively, their lack of provable stability guarantees may limit their applicability in real-world control and robot systems. In contrast, the proposed DR approach offers a principled way to design controllers that are robust to distributional shift and provide certifiable stability guarantees.

Remark VII.1 (**Impact of Input Constraints on Controller Performance**). In practical applications, control inputs are often subject to physical limitations, such as actuator saturation. These constraints can significantly impact the performance and stability of the controlled system. In our implementation, we enforce control input constraints using the clamp$(\cdot, \mathbf{u}_{\min}, \mathbf{u}_{\max})$ function, which ensures that the output remains within the specified control limits. Specifically, for the inverted pendulum system, we constrain the input u within the bounds $-15 \leq u \leq 15$.

To investigate the effect of input constraints on our Lyapunov-based approach and on the reinforcement learning algorithms, we also considered a scenario where the control bounds are tightened to $-9 \leq u \leq 9$. Under these stricter input constraints, the Lyapunov-based approach failed to learn a controller with a valid Lyapunov function certificate. This failure occurs because the limited control authority hinders the ability of the Lyapunov-based controller to drive the system towards the desired equilibrium state, which is having the pendulum in the upright position.

Interestingly, some reinforcement learning algorithms, such as PPO [63], can still learn effective control policies even under tightened input constraints. These RL algorithms have the flexibility to discover alternative control strategies that exploit the system dynamics to achieve the desired goal. For instance, the RL controller may learn to swing the pendulum in one direction, leveraging gravity to gain acceleration, and then stabilize the pendulum in the upright position from the opposite direction. This swinging behavior

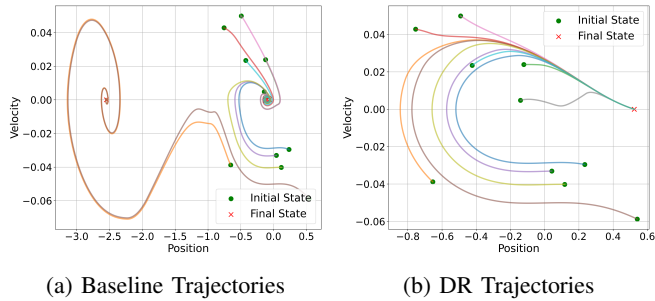(a) Baseline Trajectories   (b) DR Trajectories

FIGURE 5: Comparison of trajectories for the mountain car system with test case power parameter $p = 0.0012$. The 10 random sampled initial states are marked as green dots while the final states are markes as red crosses. In (a), the baseline controller, trained using the average power from offline observations, fails to stabilize the car to the desired equilibrium $(\pi/6, 0)$. Instead, (b) shows the distributionally robust (DR) controller successfully stabilizing the car to the top of the mountain for every initial state, demonstrating improved robustness to distributional shifts in system parameters.

allows the RL controller to overcome the limitations imposed by the reduced control bounds and successfully drive the system towards the equilibrium.   ∎

B. Mountain Car

We consider a car placed at the bottom of a sinusoidal valley. The goal is to drive the car to the top of the mountain. The system has two states, velocity v and position x of the car, and one control input u, representing the directional force applied on the car. The dynamics model is given by [64]:

$$\underbrace{\begin{bmatrix} \dot{x} \\ \dot{v} \end{bmatrix}}_{\dot{\mathbf{x}}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) := \begin{bmatrix} \dot{v} \\ -0.0025\cos(3x) \end{bmatrix} + \begin{bmatrix} 0 \\ p \end{bmatrix} u \quad (37)$$

where $p = 0.0015$ denotes the constant power of the car. The desired equilibrium is $(x, v) = (\pi/6, 0)$, corresponding to the peak of the sinusoidal valley. For training both the baseline and the DR controllers, we generate a training dataset by uniformly sampling $\{\mathbf{x}_i\}_{i\in[1600]}$ from the box region defined by $-2 \leq x \leq 2$ and $-0.4 \leq v \leq 0.4$. The control input is constrained within the bounds $-2 \leq u \leq 2$. We consider uncertainties in the car power p, represented by $\boldsymbol{\xi}$. The uncertain system dynamics can be expressed as:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) + \begin{bmatrix} 0 \\ u \end{bmatrix} \boldsymbol{\xi}. \quad (38)$$

We take $\{\boldsymbol{\xi}_i\}_{i=1}^3$ offline uncertainty samples from the normal distribution $\mathcal{N}(0.0, 0.0002)$ are available. However, during test time, the uncertainty parameters are set to $\boldsymbol{\xi} = -0.0003$.

For the baseline controller, we calculate the average power $\tilde{p}$ from the available power samples and incorporate it into the system dynamics (37) during the learning process of the baseline Lyapunov-stable controller. In contrast, for training the DR controller, we set the Wasserstein radius to $r = 0.0001$ and the risk tolerance to $\epsilon = 0.1$. The offline samples

$\{\boldsymbol{\xi}_i\}_{i\in[3]}$ are then utilized in the loss function (25) to learn the controller and Lyapunov function pair simultaneously.

To evaluate the performance of the baseline and DR controllers on the mountain car system, we consider a test scenario with uncertainty parameters set to $\boldsymbol{\xi} = -0.0003$, representing a decrease in the car's power. We randomly initialize 10 starting states within the box region defined by $-1 \leq x \leq 1$ and $-0.1 \leq v \leq 0.1$. The controllers are then applied to the system, and the resulting trajectories are simulated to assess their effectiveness.

Fig. 5a reveals that the baseline controller, which is trained on the average power values, fails to drive the car to the desired equilibrium at the top of the mountain. The trajectories corresponding to all 10 initial states fail to converge to the desired goal. The baseline controller's design, based on average parameter values, lacks the robustness necessary to adapt to the decreased power.

In contrast, Fig. 5b showcases the superior performance of the DR controller. Despite the presence of distributional shift in the system dynamics, the DR controller successfully drives all trajectories to the desired equilibrium at the mountain's peak. By explicitly considering the potential variations in the car's power during training, the DR controller is able to adapt and maintain its effectiveness for out-of-distribution parameter uncertainties.

VIII. Conclusions

In this article, we presented an approach for the joint synthesis of distributionally robust controllers and Lyapunov certificates suitable for nonlinear systems with model uncertainty. Our key contributions lie in the development of a distributionally robust Lyapunov stability formulation, leveraging finite samples of the model uncertainty parameters to address the challenge of stabilizing the uncertain system with guarantees. Through both theoretical analysis and experimental validation, we demonstrated that the system governed by the learned controller exhibits asymptotic stability with high probability, even in the presence of out-of-distribution model uncertainties.

Our approach contributes a new perspective to controlling systems with model uncertainty, offering a promising direction for certificate-based reinforcement learning in real systems. However, scalability challenges arise from the need to satisfy the Lyapunov condition for uniformly sampled data in the state space, leading to an exponential growth in the number of training samples as the dimension of the system state increases. Future research will focus on adapting our methods to more complex robotic systems. This includes exploring more efficient learning algorithms, considering relaxed stability notions, and establishing connections between Lyapunov theory and model-free reinforcement learning.

REFERENCES

[1] K. Long, Y. Yi, J. Cortes, and N. Atanasov, "Distributionally robust Lyapunov function search under uncertainty," in *Learning for Dynamics and Control Conference*, pp. 864–877, PMLR, 2023.

[2] J.-J. E. Slotine, W. Li, *et al.*, *Applied nonlinear control*, vol. 199. Prentice hall Englewood Cliffs, NJ, 1991.

[3] Z. Artstein, "Stabilization with relaxed controls," *Nonlinear Analysis-theory Methods & Applications*, vol. 7, pp. 1163–1173, 1983.

[4] E. Sontag, "A 'universal' construction of Artstein's theorem on nonlinear stabilization," *Systems & Control Letters*, vol. 13, no. 2, pp. 117–123, 1989.

[5] W. M. Haddad and V. Chellaboina, *Nonlinear dynamical systems and control: a Lyapunov-based approach*. Princeton university press, 2008.

[6] P. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. California Institute of Technology, 2000.

[7] A. Papachristodoulou and S. Prajna, "On the construction of Lyapunov functions using the sum of squares decomposition," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 3, pp. 3482–3487, 2002.

[8] Z. Jarvis-Wloszek, R. Feeley, W. Tan, K. Sun, and A. Packard, "Some controls applications of sum of squares programming," in *42nd IEEE International Conference on Decision and Control*, vol. 5, pp. 4676–4681, 2003.

[9] Y.-C. Chang, N. Roohi, and S. Gao, "Neural Lyapunov control," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[10] S. M. Richards, F. Berkenkamp, and A. Krause, "The Lyapunov neural network: Adaptive stability certification for safe learning of dynamical systems," in *Conference on Robot Learning*, pp. 466–476, PMLR, 2018.

[11] H. Dai, B. Landry, L. Yang, M. Pavone, and R. Tedrake, "Lyapunov-stable neural-network control," in *Proceedings of Robotics: Science and Systems*, (Virtual), July 2021.

[12] N. Gaby, F. Zhang, and X. Ye, "Lyapunov-net: A deep neural network architecture for Lyapunov function approximation," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 2091–2096, 2022.

[13] T. Li and N. Figueroa, "Task generalization with stability guarantees via elastic dynamical system motion policies," in *Conference on Robot Learning*, pp. 3485–3517, PMLR, 2023.

[14] A. Taylor, V. Dorobantu, H. Le, Y. Yue, and A. Ames, "Episodic learning with control Lyapunov functions for uncertain robotic systems," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6878–6884, 2019.

[15] F. Castañeda, J. Choi, B. Zhang, C. Tomlin, and K. Sreenath, "Gaussian process-based min-norm stabilizing controller for control-affine systems with uncertain input effects and dynamics," in *2021 American Control Conference (ACC)*, pp. 3683–3690, 2021.

[16] K. Long, V. Dhiman, M. Leok, J. Cortés, and N. Atanasov, "Safe control synthesis with uncertain dynamics and constraints," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7295–7302, 2022.

[17] P. Mestres and J. Cortés, "Feasibility and regularity analysis of safe stabilizing controllers under uncertainty," *Automatica*, 2023. Submitted.

[18] P. Mestres, K. Long, N. Atanasov, and J. Cortés, "Feasibility analysis and regularity characterization of distributionally robust safe stabilizing controllers," *IEEE Control Systems Letters*, vol. 8, pp. 91–96, 2024.

[19] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, 2009.

[20] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, pp. 115–166, 2018.

[21] A. R. Hota, A. K. Cherukuri, and J. Lygeros, "Data-driven chance constrained optimization under Wasserstein ambiguity sets," in *2019 American Control Conference (ACC)*, pp. 1501–1506, 2019.

[22] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics, 1994.

[23] D. Hilbert, "Ueber die darstellung definiter formen als summe von formenquadraten," *Mathematische Annalen*, vol. 32, pp. 342–350, September 1888.

[24] N. Boffi, S. Tu, N. Matni, J.-J. Slotine, and V. Sindhwani, "Learning stability certificates from data," in *Conference on Robot Learning*, pp. 1341–1350, PMLR, 2021.

[25] R. Zhou, T. Quartz, H. De Sterck, and J. Liu, "Neural Lyapunov control of unknown nonlinear systems with stability guarantees," in *Advances in Neural Information Processing Systems*, vol. 35, 2022.

[26] C. Dawson, Z. Qin, S. Gao, and C. Fan, "Safe nonlinear control using robust neural Lyapunov-barrier functions," in *Conference on Robot Learning*, vol. 164, pp. 1724–1735, PMLR, 2022.

[27] C. Dawson, S. Gao, and C. Fan, "Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1749–1767, 2023.

[28] T. Westenbroek, F. Castaneda, A. Agrawal, S. Sastry, and K. Sreenath, "Lyapunov design for robust and efficient robotic reinforcement learning," *arXiv preprint arXiv:2208.06721*, 2022.

[29] A. Lopez and D. Fridovich-Keil, "Decomposing control lyapunov functions for efficient reinforcement learning," *arXiv preprint arXiv:2403.12210*, 2024.

[30] B. P. G. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari, "Distributionally robust control of constrained stochastic systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 430–442, 2016.

[31] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Mathematical Programming*, vol. 158, pp. 291–327, 2016.

[32] W. Xie, "On distributionally robust chance constrained programs with Wasserstein distance," *Math. Program.*, vol. 186, pp. 115–155, 2021.

[33] S. Sagawa*, P. W. Koh*, T. B. Hashimoto, and P. Liang, "Distributionally robust neural networks," in *International Conference on Learning Representations*, 2020.

[34] S. Levine, A. Kumar, G. Tucker, and J. Fu, "Offline reinforcement learning: Tutorial, review, and perspectives on open problems," *arXiv preprint arXiv:2005.01643*, 2020.

[35] F. Boso, D. Boskos, J. Cortés, S. Martínez, and D. M. Tartakovsky, "Dynamics of data-driven ambiguity sets for hyperbolic conservation laws with uncertain inputs," *SIAM Journal on Scientific Computing*, vol. 43, no. 3, pp. A2102–A2129, 2021.

[36] D. Boskos, J. Cortés, and S. Martinez, "Data-driven ambiguity sets with probabilistic guarantees for dynamic processes," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 2991–3006, 2021.

[37] B. Li, Y. Tan, A.-G. Wu, and G.-R. Duan, "A distributionally robust optimization based method for stochastic model predictive control," *IEEE Transactions on Automatic Control*, vol. 67, no. 11, pp. 5762–5776, 2021.

[38] A. Cherukuri and A. R. Hota, "Consistency of distributionally robust risk- and chance-constrained optimization under Wasserstein ambiguity sets," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1729–1734, 2021.

[39] I. Yang, "Wasserstein distributionally robust stochastic control: A data-driven approach," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3863–3870, 2020.

[40] D. Boskos, J. Cortés, and S. Martínez, "High-confidence data-driven ambiguity sets for time-varying linear systems," *IEEE Transactions on Automatic Control*, vol. 69, no. 2, pp. 797–812, 2024.

[41] A. Z. Ren and A. Majumdar, "Distributionally robust policy learning via adversarial environment generation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1379–1386, 2022.

[42] J. Coulson, J. Lygeros, and F. Dörfler, "Distributionally robust chance constrained data-enabled predictive control," *IEEE Transactions on Automatic Control*, vol. 67, no. 7, pp. 3289–3304, 2021.

[43] K. Long, Y. Yi, J. Cortés, and N. Atanasov, "Safe and stable control synthesis for uncertain system models via distributionally robust optimization," in *2023 American Control Conference (ACC)*, pp. 4651–4658, 2023.

[44] L. Aolaritei, M. Fochesato, J. Lygeros, and F. Dörfler, "Wasserstein tube mpc with exact uncertainty propagation," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 2036–2041, IEEE, 2023.

[45] A. B. Kordabad, R. Wisniewski, and S. Gros, "Safe reinforcement learning using Wasserstein distributionally robust MPC and chance constraint," *IEEE Access*, vol. 10, pp. 130058–130067, 2022.

[46] A. Hakobyan and I. Yang, "Distributionally robust differential dynamic programming with Wasserstein distance," *IEEE Control Systems Letters*, vol. 7, pp. 2329–2334, 2023.

[47] T. Summers, "Distributionally robust sampling-based motion planning under uncertainty," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6518–6523, 2018.

[48] P. Lathrop, B. Boardman, and S. Martínez, "Distributionally safe path planning: Wasserstein safe RRT," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 430–437, 2021.

[49] S. Sastry, *Nonlinear systems: analysis, stability, and control*, vol. 10. Springer Science & Business Media, 2013.

[50] R. Freeman and P. Kototovic, *Robust Nonlinear Control Design*. Cambridge, MA, USA: Birkhäuser Boston Inc., 1996.

[51] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, pp. 21–41, 2000.

[52] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, pp. 969–996, 2006.

[53] H. J. Kushner and Kushner, *Stochastic stability and control*, vol. 33. Academic press New York, 1967.

[54] A. R. Teel, J. P. Hespanha, and A. Subbaraman, "A converse Lyapunov theorem and robustness for asymptotic stability in probability," *IEEE Transactions on Automatic Control*, vol. 59, no. 9, pp. 2426–2441, 2014.

[55] P. Culbertson, R. K. Cosner, M. Tucker, and A. D. Ames, "Input-to-state stability in probability," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 5796–5803, IEEE, 2023.

[56] J. Steinhardt and R. Tedrake, "Finite-time regional verification of stochastic non-linear systems," *The International Journal of Robotics Research*, vol. 31, no. 7, pp. 901–923, 2012.

[57] C. Santoyo, M. Dutreix, and S. Coogan, "A barrier function approach to finite-time stochastic system verification and control," *Automatica*, vol. 125, p. 109439, 2021.

[58] M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. Pappas, "Efficient and accurate estimation of lipschitz constants for deep neural networks," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[59] M. Towers, J. K. Terry, A. Kwiatkowski, J. U. Balis, G. d. Cola, T. Deleu, M. Goulão, A. Kallinteris, A. KG, M. Krimmel, R. Perez-Vicente, A. Pierré, S. Schulhoff, J. J. Tai, A. T. J. Shen, and O. G. Younis, "Gymnasium," Mar. 2023.

[60] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, vol. 32, 2019.

[61] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2015.

[62] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*, pp. 1861–1870, PMLR, 2018.

[63] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[64] A. W. Moore, "Efficient memory-based learning for robot control," tech. rep., University of Cambridge, 1990.