Delft University of Technology

Document Version
Final published version

Citation (APA)
Adams, S. J. L., Figueiredo Mota Diniz Costa, E., & Laurenti, L. (2025). Formal Uncertainty Propagation for Stochastic Dynamical Systems with Additive Noise. In *Proceedings of the IEEE 64th Conference on Decision and Control (CDC 2025)* (pp. 4896-4903). (Proceedings of the IEEE Conference on Decision and Control). IEEE. https://doi.org/10.1109/CDC57313.2025.11312337

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

Copyright
In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

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

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

This work is downloaded from Delft University of Technology.

Green Open Access added to TU Delft Institutional Repository as part of the Taverne amendment.

More information about this copyright law amendment can be found at https://www.openaccess.nl.

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

2025 IEEE 64th Conference on Decision and Control (CDC)
December 10-12, 2025. Rio de Janeiro, Brazil

Formal Uncertainty Propagation for Stochastic Dynamical Systems with Additive Noise

Steven Adams*, Eduardo Figueiredo* and Luca Laurenti

2025 IEEE 64th Conference on Decision and Control (CDC) | 979-8-3315-2627-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CDC57313.2025.11312337

***Abstract*— In this paper, we consider discrete-time non-linear stochastic dynamical systems with additive process noise in which both the initial state and noise distributions are uncertain. Our goal is to quantify how the uncertainty in these distributions is propagated by the system dynamics for possibly infinite time steps. In particular, we model the uncertainty over input and noise as ambiguity sets of probability distributions close in the ρ-Wasserstein distance and aim to quantify how these sets evolve over time. Our approach relies on results from quantization theory, optimal transport, and stochastic optimization to construct ambiguity sets of distributions centered at mixture of Gaussian distributions that are guaranteed to contain the true sets for both finite and infinite prediction time horizons. We empirically evaluate the effectiveness of our framework in various benchmarks from the control and machine learning literature, showing how our approach can efficiently and formally quantify the uncertainty in linear and non-linear stochastic dynamical systems.**

I. Introduction

Modern control systems are commonly uncertain. This uncertainty is due not only to the inherently stochastic dynamics [1], but also to the use of data-driven methods and statistical estimators [2], [3]. The models of such systems are consequently not only stochastic, but the probability distributions of the various random variables are themselves uncertain [4]. In many applications, where failures may have catastrophic consequences, these uncertainties cannot be ignored but must be formally accounted for during the temporal evolution of the system. Unfortunately, propagating a set of distributions through non-linear stochastic dynamics is computationally intractable [5]. This leads to the main question in this paper: Can we efficiently estimate the state of a non-linear stochastic dynamical system with uncertain initial and noise distributions over a possibly infinite time horizon while providing formal guarantees of correctness?

Various recent works have considered formal uncertainty propagation for dynamical systems. The more restrictive, but already intractable, problem of propagating known distributions over deterministic non-linear dynamics [5] has been well studied, using techniques such as moment matching [6], discretization-based methods [7], [8] or approximate numerical integration [9], [10]. However, these techniques commonly lack formal guarantees or are computationally demanding due to the need to discretize the full state space, and, critically, do not account for the fact that the actual underlying (state or noise) distributions are rarely known exactly. The problem becomes even more challenging for uncertain distributions, as it requires propagating a set of distributions through the system dynamics. While this setting is receiving increasing attention, most existing approaches are either restricted to linear dynamics, lack formal guarantees, or do not scale to large prediction horizons [11]–[13]. For example, [11] proposes a framework for propagating sets of distributions close in Wasserstein distance through deterministic dynamics over multiple time steps, but the method is effectively limited to linear systems due to numerical tractability issues. In the stochastic setting, [12] approximate the evolving state distributions with mixture distributions, while providing guarantees of correctness in the Total Variation metric. However, the resulting bounds generally become uninformative for small time horizons. Recently, [13] proposed propagating an uncertain distribution through non-linear dynamics by approximating it with discrete distributions with quantification of the resulting uncertainty in Wasserstein distance. While the method supports general stochastic dynamics, it requires constructing Cartesian products of state- and noise discretizations, limiting scalability.

In this paper, we present an algorithmic framework for formally propagating an uncertain input distribution through a non-linear dynamical system with possibly uncertain additive noise over a possibly infinite time horizon. Our approach relies on selecting a reference distribution from the input set, which is then propagated through the stochastic dynamics by approximating it with a mixture of Gaussian distributions. The resulting distribution is then used to construct a tractable convex set that is guaranteed to contain all possible distributions after propagation. To quantify uncertainty, we rely on the Wasserstein distance [14], which not only guarantees that distributions close in this distance have similar moments, but also enables the use of results from optimal transport [14], [15]. These results allow us to build a framework that not only efficiently propagates formal uncertainty sets but, for contracting systems, is even suitable for infinite-horizon propagation, as the radii of the resulting sets converge to a fixed point. We empirically illustrate the efficacy of our framework in various benchmarks. The experiments illustrate how our approach is able to successfully propagate uncertainty in various non-linear dynamical systems.

In summary, the main contributions of this work are:

- an algorithmic framework for obtaining tractable relaxations of sets of distributions propagated through non-linear dynamics with uncertain additive noise, with guarantees in Wasserstein distance,
- a proof that the Wasserstein bounds converge to a fixed

* Equal contribution
All authors are with the Delft Center for Systems and Control, Technical University of Delft, 2628 CD Delft, The Netherlands. Corresponding author: s.j.l.adams@tudelft.com

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.

point for contracting dynamics,

- extensive empirical validation of our approach on several benchmarks, including neural network dynamics with multiple hidden layers.

II. Notation

For a vector $x \in \mathbb{R}^n$ we use $x^{(i)}$ to denote the i-th element of x. Furthermore, for region $\mathcal{X} \subset \mathbb{R}^n$, the indicator function for $\mathcal{X}$ is denoted as $\mathbf{1}_{\mathcal{X}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{X} \\ 0, & \text{otherwise} \end{cases}$. Given a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with $\mathcal{B}(\mathcal{X})$ being the σ-algebra, we denote by $\mathcal{P}(\mathcal{X})$ the set of probability distributions on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. The set of probability distributions with finite moments up to order $\rho \in \mathbb{N}_{>0}$, $\mathcal{P}_\rho(\mathcal{X})$, is defined as the set of all $\mathbb{P} \in \mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} \|x\|^\rho \mathbb{P}(dx) < \infty$, where $\|\cdot\|$ is the L_ρ-norm. In this paper, for a metric space $\mathcal{X}$, $\mathcal{B}(\mathcal{X})$ is assumed to be the Borel σ-algebra of $\mathcal{X}$. For a random variable x taking values in $\mathcal{X}$, $\mathbb{P}_x \in \mathcal{P}(\mathcal{X})$ represents the probability measure associated to x. For $\rho \in \mathbb{N}_{>0}$ and $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_\rho(\mathcal{X})$, the *$\rho$-Wasserstein distance* between $\mathbb{P}$ and $\mathbb{Q}$ is defined as:

$$\mathbb{W}_\rho(\mathbb{P}, \mathbb{Q}) = \left(\inf_{\gamma \in \Gamma(\mathbb{P}, \mathbb{Q})} \int_{\mathcal{X} \times \mathcal{X}} \|x - x'\|^\rho \gamma(dx, dx') \right)^{\frac{1}{\rho}} \quad (1)$$

where $\Gamma(\mathbb{P}, \mathbb{Q}) \subset \mathcal{P}(\mathcal{X} \times \mathcal{X})$ represents the set of probability distributions with marginal distributions $\mathbb{P}$ and $\mathbb{Q}$. For $\mathbb{P} \in \mathcal{P}_\rho(\mathcal{X})$ and $\theta \geq 0$ the set of distributions closer than θ to $\mathbb{P}$ in the ρ-Wasserstein distance, also called *ρ-Wasserstein ambiguity set*, is denoted by

$$\mathbb{B}_\theta(\mathbb{P}) = \{\mathbb{Q} \in \mathcal{P}_\rho(\mathcal{X}) \mid \mathbb{W}_\rho(\mathbb{P}, \mathbb{Q}) \leq \theta\} \subset \mathcal{P}_\rho(\mathcal{X}). \quad (2)$$

For $N \in \mathbb{N}$, $\Pi^N = \{\pi \in \mathbb{R}^N_{\geq 0} \ : \ \sum_{i=1}^N \pi^{(i)} = 1\}$ is the N-simplex. A discrete probability distribution $\mathbb{D} \in \mathcal{P}(\mathcal{X})$ is defined as $\mathbb{D} = \sum_{i=1}^N \pi^{(i)} \delta_{c_i}$, where δ_c is the Dirac delta function centered at location $c \in \mathcal{X}$ and $\pi \in \Pi^N$. The set of discrete probability distributions on $\mathcal{X}$ with at most N locations is denoted as $\mathcal{D}_N(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$.

For measurable spaces $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, a probability distribution $\mathbb{P} \in \mathcal{P}(\mathcal{X})$ and a measurable function $g : \mathcal{X} \to \mathcal{Y} \subseteq \mathbb{R}^q$, we use $g\#\mathbb{P}$ to denote the push-forward measure of $\mathbb{P}$ by g, i.e., the measure on $\mathcal{Y}$ such that for all $\mathcal{A} \in \mathcal{B}(\mathcal{Y})$, $(g\#\mathbb{P})(\mathcal{A}) = \mathbb{P}(g^{-1}(\mathcal{A}))$. For probability distributions $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{X})$, we use $\mathbb{P} * \mathbb{Q} \in \mathcal{P}(\mathcal{X})$ to denote the convolution of $\mathbb{P}$ and $\mathbb{Q}$, i.e., $(\mathbb{P} * \mathbb{Q})(\mathcal{A}) = \int_{\mathcal{X} \times \mathcal{X}} \mathbf{1}_{\mathcal{A}}(x + x') \mathbb{P}(dx) \mathbb{Q}(dx')$ for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$.

III. Problem Formulation

We consider a non-linear stochastic process with additive noise described as:

$$x_{k+1} = f(x_k) + \omega_k, \quad x_0 \sim \mathbb{P}_{x_0}, \omega_k \sim \mathbb{P}_\omega \quad (3)$$

where, for state space $\mathcal{X} \subseteq \mathbb{R}^n$, $f : \mathcal{X} \mapsto \mathcal{X}$ is a possibly non-linear measurable piecewise Lipschitz continuous function representing the one-step dynamics of System 3. The initial distribution $\mathbb{P}_{x_0} \in \mathcal{P}(\mathcal{X})$ and the process noise distribution $\mathbb{P}_\omega \in \mathcal{P}(\mathcal{X})$ are assumed to be unknown, as formalized in Assumption 1 below. Intuitively, System 3 represents a general model of a n-dimensional discrete-time autonomous stochastic system with additive noise, but possibly non-linear dynamics, in which the knowledge of the probability distributions is uncertain. As a consequence, System 3 encompasses a large class of modern control systems commonly used in practice, such as Neural Network dynamical systems [16], [17], or systems with additive noise learned from data [4], [18].

Assumption 1. *We assume that $\mathbb{P}_{x_0}$ and $\mathbb{P}_\omega$ are uncertain. In particular, for $\rho \in \mathbb{N}_{>0}$, $\theta_{x_0}, \theta_\omega \geq 0$, and known mixture of Gaussian distributions $\bar{\mathbb{P}}_{x_0}, \bar{\mathbb{P}}_\omega \in \mathcal{P}_\rho(\mathcal{X})$, we assume that $\mathbb{P}_{x_0} \in \mathbb{B}_{\theta_{x_0}}(\bar{\mathbb{P}}_{x_0})$ and $\mathbb{P}_\omega \in \mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)$. That is, $\mathbb{P}_{x_0}$ and $\mathbb{P}_\omega$ are only known to lie in certain ρ-Wasserstein ambiguity sets.*

Assumption 1 guarantees that we can quantify the uncertainty in System (3) using the $\rho-$Wasserstein distance. This ensures that the resulting ambiguity sets could be estimated from data using existing data-driven approaches that guarantee closeness in the $\rho-$Wasserstein distance [18], [19]. Furthermore, one can rely on the various properties of the $\rho-$Wasserstein distance, such as the fact that in contrast to other metrics commonly used such as KL-divergence [20], closeness in the $\rho-$Wasserstein distance guarantees closeness in the first $\rho-$moments [21]. We should also stress that the assumption that $\mathbb{B}_{\theta_{x_0}}(\bar{\mathbb{P}}_{x_0})$ and $\mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)$ are centered in mixtures of Gaussian distributions is not limiting, as this class of distributions can approximate any distribution arbitrarily well[1] and the radius around it allows for distributions that are not mixtures nor Gaussian to be in the ambiguity sets.

Further, we note that if $\mathbb{P}_{x_0}$ and $\mathbb{P}_\omega$ were known, then the distribution of the system at time k, $\mathbb{P}_{x_k}$, could be computed recursively as:

$$\mathbb{P}_{x_k} = (f\#\mathbb{P}_{x_{k-1}}) * \mathbb{P}_\omega, \quad (4)$$

that is, $\mathbb{P}_{x_k}$ is obtained by first propagating the distribution at the previous time step through state dynamics f and then taking the convolution with the noise distribution, which arises because of the additivity of the noise [23]. However, in System (3) both $\mathbb{P}_{x_0}$ and $\mathbb{P}_\omega$ are uncertain. Furthermore, we should emphasize that, in practice, even for known distributions, the computation of Eqn (4) in closed form is generally infeasible [5]. This makes the setting considered in this paper particularly challenging.

A. Problem Statement

Given the ambiguity sets of initial and noise distributions introduced in Assumption 1, our goal is to quantify how this uncertainty is propagated by System (3) over time. In particular, we consider the following problem.

Problem 1. *Given time horizon $K \in \mathbb{N} \cup \{\infty\}$, compute sets $\mathbb{S}_{x_0}, ..., \mathbb{S}_{x_K}$ defined recursively as follows for $k < K$:*

$$\mathbb{S}_{x_0} = \mathbb{B}_{\theta_{x_0}}(\bar{\mathbb{P}}_{x_0}), \quad (5)$$
$$\mathbb{S}_{x_{k+1}} = \{(f\#\mathbb{P}_{x_k}) * \mathbb{P}_{\omega_k} \mid \mathbb{P}_{x_k} \in \mathbb{S}_{x_k}, \mathbb{P}_{\omega_k} \in \mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)\}.$$

[1]The space of mixtures of Gaussians is dense in $\mathcal{P}_\rho(\mathcal{X})$ for the metric $\mathbb{W}_\rho$ [14], [22].

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.

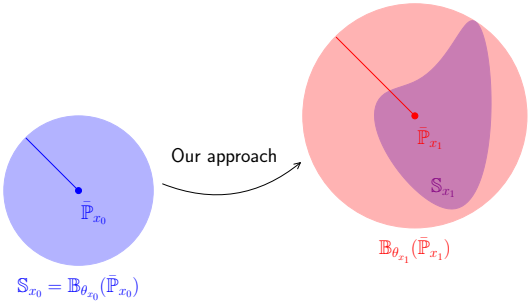


Fig. 1: Schematic representation of our proposed approach for $k = 1$. As $\mathbb{S}_{x_1}$ is generally intractable, we over-approximate it with a ρ-Wasserstein ambiguity ball centered at $\bar{\mathbb{P}}_{x_1} \approx (f\#\bar{\mathbb{P}}_{x_0}) * \bar{\mathbb{P}}_{\omega}$, that is, at the propagation of the center of the ambiguity ball at the previous time step through the system dynamics according to Eqn (4). The error in this approximation is formally accounted for in the choice of θ_{x_1}.

Following Eqn (4), sets $\mathbb{S}_{x_k}$ represent all the state distributions that System 3 can assume at time step k. Unfortunately, due to the potential nonlinearity of f, solving Problem 1 exactly is generally infeasible[2]. Consequently, in what follows, we aim to design ρ-Wasserstein ambiguity sets of distributions that are guaranteed to include the true set of distributions of System 3 over time. We should emphasize that a solution of Problem 1 would allow one not only to formally quantify the uncertainty in dynamical systems and to compute worst-case expectations of various quantities due to the properties of the ρ-Wasserstein distance (see, e.g., Theorem 1 in [25]), but it would also represent an important step towards developing distributionally robust control algorithms for non-linear stochastic systems.

Remark 1. *If both the initial distribution $\mathbb{P}_{x_0}$ and process noise $\mathbb{P}_{\omega}$ are known (i.e., $\mathbb{S}_{x_0} = \{\mathbb{P}_{x_0}\}$ and $\mathbb{S}_{\omega} = \{\mathbb{P}_{\omega}\}$), Problem 1 reduces to compute Eqn* (4)*, which, as mentioned earlier, is also object of research (e.g. [12]).*

a) Approach: For every $k \leq K$, our approach is illustrated in Figure 1 and is based on constructing center distributions $\bar{\mathbb{P}}_{x_k} \subset \mathcal{P}_{\rho}(\mathcal{X})$ and radius $\theta_{x_k} \geq 0$ such that $\mathbb{S}_{x_k} \subseteq \mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k})$ for any $k \in \{1, \cdots, K\}$. More specifically, $\bar{\mathbb{P}}_{x_k}$ is obtained in Section V-A by approximating the push-forward distribution of the previous center by the dynamics of System (3) with a mixture of distributions. Then, in Section V-B, we show how a radius guaranteeing enclosure can be computed by relying on recent results on uncertainty propagation in non-linear stochastic systems [13]. Convergence guarantees of our approach are provided in Section V-B. Lastly, in Section VI, we demonstrate the efficacy of our approach through numerical experiments.

[2]Note that given a convex set S, and a non-linear map $T : S \mapsto V$, $T(S) \subset V$ is not necessarily convex [24]. Consequently, sets $\mathbb{S}_{x_k}$ are generally non-convex.

IV. Preliminaries

Our approach relies on quantization of distributions and mixture distributions, which we recall in this Section.

A. Quantization of Probability Distributions

For $\mathcal{X} \subseteq \mathbb{R}^n$, a set of N-points $\boldsymbol{\mathcal{C}} = \{c_i\}_{i=1}^N$ in $\mathcal{X}$ called locations, let $\boldsymbol{\mathcal{R}} = \{\mathcal{R}_i\}_{i=1}^N$ be the Voronoi partition of $\mathcal{X}$ w.r.t. $\boldsymbol{\mathcal{C}}$, i.e., for each i, i.e., $\mathcal{R}_i = \{x \in \mathcal{X} \mid \|z - c_i\| \leq \|z - c_j\|, \forall i \neq j\}$. Then, the quantization operator $\Delta_{\boldsymbol{\mathcal{C}}} : \mathbb{R}^n \to \mathbb{R}^n$ is defined as $\Delta_{\boldsymbol{\mathcal{C}}}(x) := \sum_{i=1}^N c_i \mathbf{1}_{\mathcal{R}_i}(x)$. Intuitively, the quantization operator maps any point within the region $\mathcal{R}_i$ to the location c_i. Consequently, for any probability distribution $\mathbb{P} \in \mathcal{P}(\mathcal{X})$, it holds that $\Delta_{\boldsymbol{\mathcal{C}}}\#\mathbb{P} = \sum_{i=1}^N \mathbb{P}(\mathcal{R}_i)\delta_{c_i} \in \mathbb{D}_N(\mathcal{X})$. Here, $\Delta_{\boldsymbol{\mathcal{C}}}\#\mathbb{P}$ is called the *discretization* or *quantization* of $\mathbb{P}$, effectively converting a possibly continuous distribution into a discrete one with support of size N based on the specified locations.

Remark 2. *$\Delta_{\boldsymbol{\mathcal{C}}}\#\mathbb{P}$ can be computed particularly efficiently when the probability mass of each Voronoi region can be expressed in analytic form, typically via the distribution's cumulative distribution function. For instance, univariate Gaussian allow exact computation over intervals, while multivariate Gaussians admit closed-form expressions over regions aligned with the eigenbasis of the covariance matrix [21] (consequently, the same also holds for mixtures of Gaussians for which all components satisfy the same property).*

B. Mixture Distributions

A mixture distribution of size $M \in \mathbb{N}$, is a set of M distributions of the same type, also called components, averaged w.r.t. a probability vector $\pi \in \Pi_M$, also called the weights [26]. The convolution of a discrete distribution $\sum_{i=1}^N \pi^{(i)}\delta_{c_i}$ with distribution $\mathbb{P}$ results in a mixture distribution with components $\delta_{c_i} * \mathbb{P}$ and weights π, i.e., $\left(\sum_{i=1}^N \pi^{(i)}\delta_{c_i}\right) * \mathbb{P} = \sum_{i=1}^N \pi^{(i)}(\delta_{c_i} * \mathbb{P})$, where $(\delta_c * \mathbb{P})(dx) = \mathbb{P}(dx - c)$. When $\mathbb{P} = \mathcal{N}(m, \Sigma)$, i.e. a Gaussian distribution with mean m and covariance Σ, $\delta_{c_i} * \mathbb{P} = \mathcal{N}(m + c_i, \Sigma)$ and, thus, $\sum_{i=1}^N \pi^{(i)}(\delta_{c_i} * \mathbb{P}) = \sum_{i=1}^N \pi^{(i)}\mathcal{N}(m + c_i, \Sigma)$ is a Gaussian Mixture.

V. Approximation Scheme

In this section, we introduce our approach to solve Problem 1 by iteratively approximating the true sets of uncertain distributions $\mathbb{S}_{x_1}, \ldots, \mathbb{S}_{x_K}$ with a set of ρ-Wasserstein ambiguity balls $\mathbb{B}_{\theta_1}(\bar{\mathbb{P}}_{x_1}), \ldots, \mathbb{B}_{\theta_K}(\bar{\mathbb{P}}_{x_K})$. We first show that due to the additive nature of the process noise in System 3 a natural choice is to consider centers $\bar{\mathbb{P}}_{x_1}, \ldots, \bar{\mathbb{P}}_{x_K}$ as a mixture of Gaussian distributions approximating the pushforward distribution of the center at the previous time, that is, of the time evolution of $\bar{\mathbb{P}}_{x_0}$ through the dynamics in System (3). After that, we determine the radii $\theta_1, \ldots, \theta_K$ to ensure that $\mathbb{B}_{\theta_k}(\bar{\mathbb{P}}_{x_k})$ is guaranteed to include $\mathbb{S}_{x_k}$ for at all time steps k. Further, in Proposition 3, we show that, for contractive systems, the radius of the ambiguity ball converges to a fixed finite value as $k \to \infty$. Lastly, in Subsection V-C, we present the full algorithm for our approximation scheme.

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.

A. Ambiguity Ball Centers via Mixture Distributions

For each time step k, consider a set of locations $\mathcal{C}_k \subset \mathcal{X}$ that define a quantization operator $\Delta_{\mathcal{C}_k}$, where we refer to Subsection V-C for how to select $\mathcal{C}_k$. Given $\bar{\mathbb{P}}_{x_0}$ and $\bar{\mathbb{P}}_\omega$, which, for clarity of presentation and without loss of generality, we assume are Gaussian mixtures with equal covariance across components, we define the center of the ambiguity balls $\bar{\mathbb{P}}_{x_{k+1}}$ iteratively for each $k \in \{0, \dots, K-1\}$ as follows[3]:

$$\begin{aligned}\bar{\mathbb{P}}_{x_{k+1}} &= (f\#\Delta_{\mathcal{C}_k}\#\bar{\mathbb{P}}_{x_k}) * \bar{\mathbb{P}}_\omega \\ &= \sum_{i=1}^{|\mathcal{C}_k|} \bar{\mathbb{P}}_{x_k}(\mathcal{R}_{k,i})(\delta_{f(c_{k,i})} * \bar{\mathbb{P}}_\omega), \end{aligned} \tag{6}$$

where $\mathcal{R}_{k,i}$ is the region in the Voronoi partition of $\mathcal{X}$ w.r.t. $\mathcal{C}_k$ that corresponds to location $c_{k,i} \in \mathcal{C}_k$. Note that, since all components of mixture $\bar{\mathbb{P}}_\omega$ share the same covariance, the components of the resulting mixtures $\bar{\mathbb{P}}_{x_{k+1}}$ have identical covariances. As explained in Remark 2, this choice guarantees that $\bar{\mathbb{P}}_{x_k}(\mathcal{R}_{k,i})$ can be computed in closed form for a suitable choice of locations $\mathcal{C}_k$, thereby facilitating efficient quantization in Subsection V-C. Intuitively, at each time k, Eqn (6) builds a mixture approximation of $(f\#\bar{\mathbb{P}}_{x_k}) * \bar{\mathbb{P}}_\omega$ by averaging and shifting $\bar{\mathbb{P}}_\omega$ with the weights and locations resulting from the quantization of $\bar{\mathbb{P}}_{x_k}$ by $\Delta_{\mathcal{C}_k}$. Note that the mixture approximation consists of a straightforward application of state dynamics f to the support of the discrete approximation of the state dynamics at the previous time step, i.e., $c_{k,1}, \dots, c_{k,|\mathcal{C}_k|}$.

Example 1. *Consider $\bar{\mathbb{P}}_\omega$ to be a known Gaussian noise distribution with mean m and covariance Σ, i.e., $\bar{\mathbb{P}}_\omega = \mathcal{N}(m, \Sigma)$. The center distributions defined by Eqn* (6) *for each time step k are GMMs that can be written as*

$$\bar{\mathbb{P}}_{x_{k+1}} = \sum_{i=1}^{|\mathcal{C}_k|} \bar{\mathbb{P}}_{x_k}(\mathcal{R}_{k,i})\mathcal{N}(f(c_{k,i}) + m, \Sigma).$$

B. The Radii of the Ambiguity Balls

After defining the centers of the ambiguity balls, we now determine the radii θ_k. The following Proposition presents an upper bound for the radii that ensure that $\mathbb{S}_{x_k} \subseteq \mathbb{B}_{\theta_k}(\bar{\mathbb{P}}_{x_k})$ for all time steps $k \leq K$.

Proposition 1. *Let $\bar{\mathbb{P}}_{x_1}, \dots, \bar{\mathbb{P}}_{x_K}$ be defined according to Eqn* (6)*. For each $k \in \{0, \dots, K-1\}$, iteratively define*

$$\theta_{x_{k+1}} \geq \sup_{\substack{\mathbb{P}\in\mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k}),\\ \mathbb{P}_\omega\in\mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)}} \mathbb{W}_\rho(f\#\mathbb{P} * \mathbb{P}_\omega, \bar{\mathbb{P}}_{x_{k+1}}) \tag{7}$$

Then, $\forall k \in \mathbb{N}$ it holds that $\mathbb{S}_{x_k} \subseteq \mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k})$.

The proof of Proposition 1 is in Section VIII and relies on the fact that, after selecting a particular center distribution $\bar{\mathbb{P}}_{x_k}$, it is enough to take a radius larger than the worst distance of this center to a set of distributions containing the true ambiguity set $\mathbb{S}_{x_k}$. While computing the supremum in Eqn (7) is generally intractable, upper bounds can be computed by extending results in [13] as we show in Proposition 2.

Proposition 2. *Consider a set of locations $\mathcal{C} \subset \mathcal{X}$. Further, call*

$$\theta_\Delta = \left(\sum_{\ell=1}^{|\mathcal{C}|} \int_{\mathcal{R}_\ell} \|x - c_\ell\|^\rho \, d\bar{\mathbb{P}}_{x_k}(x)\right)^{\frac{1}{\rho}}, \tag{8}$$

and for $\ell \in \{1, \dots, |\mathcal{C}|\}$, let $\alpha_\ell, \beta_\ell \in \mathbb{R}_{\geq 0}$ be such that for $x \in \mathcal{X}$

$$\|f(x) - f(c_\ell)\|^\rho \leq \alpha_\ell \|x - c_\ell\|^\rho + \beta_\ell. \tag{9}$$

Then, for $\hat{\alpha} = \max_{\ell\in\{1,\dots,|\mathcal{C}|\}} \alpha_\ell$, it holds that

$$\begin{aligned}&\sup_{\substack{\mathbb{P}\in\mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k}),\\ \mathbb{P}_\omega\in\mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)}} \mathbb{W}_\rho(f\#\mathbb{P} * \mathbb{P}_\omega, \bar{\mathbb{P}}_{x_{k+1}}) \\ &\qquad \leq \theta_\omega + \left(\hat{\alpha}(\theta_{x_k} + \theta_\Delta)^\rho + \sum_{\ell=1}^{|\mathcal{C}|} \bar{\mathbb{P}}_{x_k}(\mathcal{R}_\ell)\beta_\ell\right)^{\frac{1}{\rho}}. \end{aligned} \tag{10}$$

Proposition 2 guarantees that the radius $\theta_{x_{k+1}}$ of our proposed ambiguity ball $\mathbb{B}_{\theta_{x_{k+1}}}(\bar{\mathbb{P}}_{x_{k+1}})$ can be taken as in Eqn (10). This bound crucially depends on θ_Δ, which represents a quantization penalty as explained in Remark 3 from [13] and the norm linearization of f around each point c_i in Eqn (9). Note that both the center and the radius of the ambiguity balls depend on the choice of quantization locations $\mathcal{C}_k$. The selection of the quantization operator, as well as the computation of the quantization penalty and the norm linearization, is discussed in the next subsection.

Before defining an algorithm for Problem 1, it is important to discuss how the radii of the ambiguity sets resulting from Proposition 2 evolve over time. In Proposition 3 below, we show that when System 3 is contractive (i.e. the global Lipschitz constant of f is smaller than one, i.e. $\mathcal{L}_f < 1$), the ρ-Wasserstein ambiguity radius in Eqn (10) converges to a fixed point, which allows us to obtain informative (i.e. finite radius) sets also for arbitrarily large K.

Proposition 3. *Let $f : \mathcal{X} \mapsto \mathcal{X}$ be a piecewise Lipschitz continuous function with global Lipschitz constant $\mathcal{L}_f < 1$. Given $\epsilon > 0$, let $\mathcal{C}_k \in \mathcal{X}$ be a set of points such that $\theta_{\Delta,k} = \left(\sum_{\ell=1}^{|\mathcal{C}_k|} \int_{\mathcal{R}_{k,\ell}} \|x - c_{k,\ell}\|^\rho \, d\bar{\mathbb{P}}_{x_k}(x)\right)^{\frac{1}{\rho}} \leq \epsilon$ for every k. Consider the following iterative process describing the evolution of the radius in Eqn* (10) *for $k \in \mathbb{N}_{>0}$, with $\theta_0 = \theta_{x_0}$ and*

$$\theta_{k+1} = \theta_\omega + \left(\hat{\alpha}_k(\theta_k + \theta_{\Delta,k})^\rho + \sum_{\ell=1}^{|\mathcal{C}_k|} \bar{\mathbb{P}}_{x_k}(\mathcal{R}_{k,\ell})\beta_{k,\ell}\right)^{\frac{1}{\rho}}. \tag{11}$$

Then,

$$\lim_{k\to\infty} \theta_k \leq \frac{\theta_\omega}{1 - \mathcal{L}_f} + \frac{\mathcal{L}_f}{1 - \mathcal{L}_f}\epsilon. \tag{12}$$

Proposition 3 is similar to Theorem 7 in [13] and it relies on the Banach Fixed Point Theorem [27]. The immediate

[3] In the general case of mixture of Gaussians with different covariance per component, following standard practice [21], one can apply a separate quantization operator to each mixture component. That is, for $\mathbb{P} = \sum_{i=1}^N \pi^{(i)}\mathbb{P}_i$, one can quantization locations $\mathcal{C}_i$ for each $\mathbb{P}_i$ and take $\sum_{i=1}^N \pi^{(i)}\Delta_{\mathcal{C}_i}\#\mathbb{P}_i$ as the quantization of $\mathbb{P}$.

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.

consequence of this result is that for contractive systems, the radius of $\mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k})$, θ_{x_k}, remains finite even for an infinite horizon. Also note that, while ϵ can be made arbitrarily small by controlling the quantization error, θ_w represents the uncertainty in the additive noise. Consequently, the radius of the $\rho-$Wasserstein ambiguity sets resulting from our approach will converge to 0 only if the additive noise distribution is not uncertain. This is intuitive because while the uncertainty in the initial distribution can be made arbitrarily small over time by the contractive nature of the dynamics, the uncertainty in the noise is added to the system at every time step.

C. Algorithm

We summarize our procedure to construct the ρ-Wasserstein ambiguity sets with System 3 in Algorithm 1. In line 2, given the center of the ambiguity set at time $k \geq 0$, $\bar{\mathbb{P}}_{x_k}$, we select the quantization locations $\boldsymbol{\mathcal{C}}_k$ according to Algorithm 2 in [21]. The resulting Voronoi partition $\{\mathcal{R}_{k,i}\}_{i=1}^{|\boldsymbol{\mathcal{C}}_k|}$ w.r.t. $\boldsymbol{\mathcal{C}}_k$, is guaranteed to align with the eigenbasis of the identical covariance shared by all components of $\bar{\mathbb{P}}_{x_k}$. This structure enables tractable quantization of $\bar{\mathbb{P}}_{x_k}$ and ensures a closed form expression for the quantization error θ_Δ in Eqn. (8). In line 3, the quantization of $\bar{\mathbb{P}}_{x_k}$ is propagated through the system dynamics, resulting in a discrete distribution $\mathbb{D}_{k+1}$. To control the number of components of the mixture, and thus the computational complexity of the subsequent quantizations, we include an optional compression step in line 4. This step approximates $\mathbb{D}_{k+1}$ by a discrete distribution with fewer elements. We implement this compression, referred to as `Compress`, by applying N-means clustering to the support of $\mathbb{D}_{k+1}$. The error introduced by the compression step is accounted for by computing the ρ-Wasserstein distance between two discrete distributions which can be solved using linear programming techniques, as shown in [15]. In line 5, the center of the ambiguity set at time $k+1$ is obtained by convoluting the compressed discrete distribution with $\bar{\mathbb{P}}_\omega$. Then, the radius of the ambiguity set is computed according to Proposition 2, the quantization error θ_Δ according to Corollary 10 in [21], and the norm linearization in Eqn. (9) as in Section V.A of [13], using bound propagation techniques [28].

VI. Experiments

In this section, we experimentally evaluate the effectiveness of Algorithm 1 in propagating uncertainty in various benchmarks. We begin with a qualitative evaluation of the ambiguity sets computed with our framework, highlighting their structure and evolution over time and illustrating how they can be used to derive bounds in expectation. Next, we analyze how the quantization and compression steps, as well as the uncertainty radii for the initial and noise distributions, impact the size of the resulting ambiguity sets over time[4]. For all experiments, we fix $\rho = 2$.[5] We consider

[4]Our code is available at `https://github.com/sjladams/DUQviaWasserstein/tree/additive_noise`

[5]We consider $\rho = 2$ since $\mathbb{W}_2$ ensures closeness in second moments and the quantization error θ_Δ in Eqn (8) admits a closed form.

Algorithm 1: Propagate ρ-Wasserstein ambiguity ball with System 3

Input: Function f from System 3, state and noise ambiguity balls, $\mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k}), \mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)$ at step k

Output: Ambiguity ball $\mathbb{B}_{\theta_{x_{k+1}}}(\bar{\mathbb{P}}_{x_{k+1}})$ according to approximation scheme in Section V

1 **function** `PropagateAmbiguityBall(`f, $\mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k})$, $\mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)$`):`

2 $\quad \boldsymbol{\mathcal{C}}_k \leftarrow$ `Algorithm 2 from [21]`

3 $\quad \mathbb{D}_{k+1} \leftarrow f \# \Delta_{\boldsymbol{\mathcal{C}}_k} \# \bar{\mathbb{P}}_{x_k}$

4 $\quad \mathbb{D}_{k+1}, \theta_{\text{compr}} \leftarrow$ `Compress(`$\mathbb{D}_{x_{k+1}}, N$`)`

5 $\quad \bar{\mathbb{P}}_{x_{k+1}} \leftarrow \mathbb{D}_{k+1} * \bar{\mathbb{P}}_\omega$

6 $\quad \theta_{x_{k+1}} \leftarrow$ `Prop. 2 with` $\theta = \theta_{x_k} + \theta_{\text{compr}}$

7 **return** $\mathbb{B}_{\theta_{x_{k+1}}}(\bar{\mathbb{P}}_{x_{k+1}})$

the following benchmarks: a 2-D piecewise linear system defined by $f(x) = \begin{cases} A_1 x & \text{if } x^{(1)} \leq 0 \\ A_2 x & \text{else} \end{cases}$, with rotation matrices $A_i = 0.8 \cdot \begin{bmatrix} \cos(\phi_i) & -\sin(\phi_i) \\ \sin(\phi_i) & \cos(\phi_i) \end{bmatrix}$ for $i \in \{1, 2\}$, and angles $\phi_1 = \frac{1}{8}\pi$, $\phi_2 = -\frac{1}{8}\pi$, referred to as the *Double Spiral* dynamics (Figure 2a); a 2-D switched system from [29] with five linear modes under an optimal reach-avoid strategy, resulting in *piecewise linear* dynamics (Figure 2d); a 2-D Neural Network (NN) dynamical system where f is an MLP with two hidden layers of width 64 and sigmoid activations trained to approximate the discrete-time dynamics of a pendulum [30], referred to as *NN Pendulum* (Figure 2g); and an instance of the 4-D *Quadruple-Tank* dynamics from [31]. For all benchmarks, the initial uncertainty sets are centered at $\bar{\mathbb{P}}_{x_0} = \mathcal{N}(\mu_x, \text{diag}(v_x))$. For the 2-D benchmarks, the noise uncertainty set is centered at $\bar{\mathbb{P}}_\omega = \mathcal{N}(\bar{0}, \text{diag}(v_\omega))$, and for the 4-D Quadruple-Tank $\bar{\mathbb{P}}_\omega = \frac{1}{2}\mathcal{N}(\mu_\omega, \text{diag}(v_\omega)) + \frac{1}{2}\mathcal{N}(-\mu_\omega, \text{diag}(v_\omega))$. The corresponding parameter values are listed in Table II.

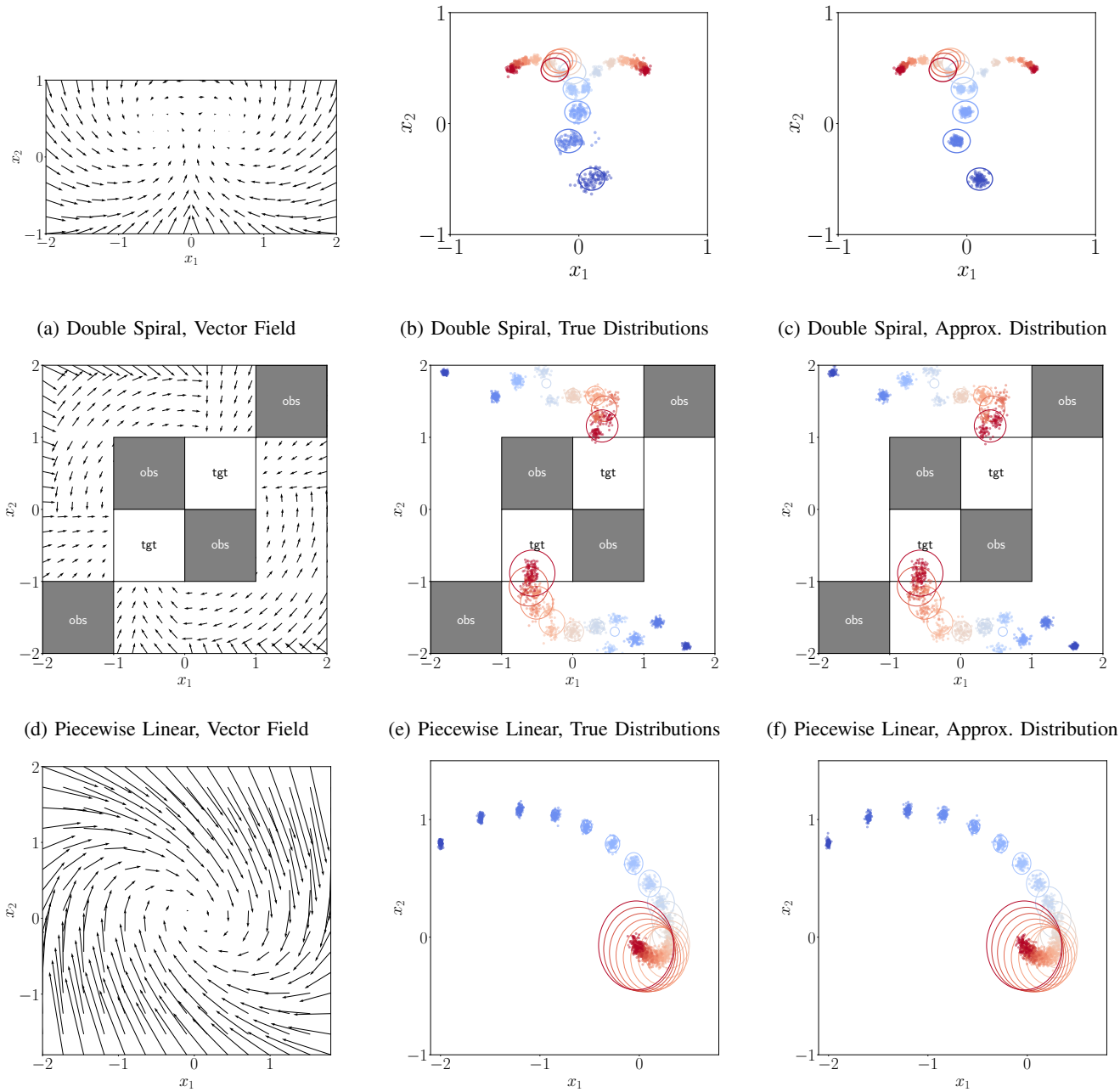
Dynamics	Distributions parameters
Double Spiral	$\mu_x = [0.1, -0.5]$, $v_x = 10^{-3} \cdot \bar{1}$ $v_\omega = 0.0001 \cdot \bar{1}$
Piecewise Linear	$\mu_x = [-1.8, 1.9]$ or $[1.6, -1.9]$, $v_x = 0.0005 \cdot \bar{1}$ $v_\omega = [0.0009, 0.0009]$
NN Pendulum	$\mu_x = [-2, 0.8]$, $v_x = [0.0001, 0.0005]$ $v = [0.0001, 0.0005]$
Quadruple -Tank	$\mu_x = [1.5, 2.5, 0.5, 1.0]$, $v_x = [0.001, 0.02, 0.4, 0.01]$ $\mu_\omega = \pm 0.01 \cdot \bar{1}$, $v_\omega = [0.01, 0.01, 0.0002, 0.001]$

TABLE I: Parameter values for the center distributions of the initial and noise uncertainty sets.

A. Qualitative analysis of the ambiguity set evolution

We qualitatively validate our ambiguity sets by analyzing the time evolution of both the true distributions, approximated using Monte Carlo samples, and the 2-Wasserstein ambiguity sets obtained with our framework for the 2-D benchmarks. From Figure 2, we first visually observe that the

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.



(a) Double Spiral, Vector Field (b) Double Spiral, True Distributions (c) Double Spiral, Approx. Distribution

(d) Piecewise Linear, Vector Field (e) Piecewise Linear, True Distributions (f) Piecewise Linear, Approx. Distribution

(g) NN Pendulum, Vector Field (h) NN Pendulum, True Distributions (i) NN Pendulum, Approx. Distribution

Fig. 2: Vector fields (left column), samples from the true uncertainty sets $\mathbb{S}_{x_k}$ obtained via Monte Carlo sampling (center column), and samples from the centers $\bar{\mathbb{P}}_{x_k}$ of the ambiguity sets obtained via Algorithm 1 (right column) for different benchmarks (by row), evaluated over 20 time-step. Sampled states are plotted as dots, with the blue-red gradient indicating time evolution from the initial to the final time step. The circles, centered at the mean of $\bar{\mathbb{P}}_{x_k}$ with radius θ_{x_k}, are guaranteed to contain the mean of the true distributions. The piecewise linear dynamics result from a switched system by the optimal switching strategy in [29], which steers the system toward the target region while avoiding the obstacles.

centers of our ambiguity sets closely resemble the true distributions. This is expected, as the centers of our ambiguity sets are obtained by propagating distribution $\bar{\mathbb{P}}_{x_0}$, which is known to lie within the true set of possible initial distributions through the system dynamics. Importantly, the bimodality of the Double Spiral in its converging sequence and the transient bimodal structure in the third time step of both paths of the Piecewise Linear system, underscore the effectiveness of mixture-based approximations. Such multimodal behavior cannot be easily captured with usual propagation techniques such as moment matching [6].

To illustrate the 2-Wasserstein ambiguity set obtained with

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1, we plot circles centered at the mean of $\bar{\mathbb{P}}_{x_k}$ with radii θ_{x_k}. Since the ρ-th Wasserstein distance implies closeness in the ρ-th moment (see e.g., Lemma 2 in [21]), these circles are guaranteed to include the mean of the true distributions. In Figure 2b, we observe that for the contracting Double Spiral dynamics, the ambiguity sets converge to a fixed point, as expected from Proposition 3. Interestingly, the position of the converging circles, which includes the means of the true bimodal true distributions, indicates that more mass is concentrated in the left mode than in the right. In contrast, for the NN Pendulum dynamics, we observe slowly expanding ambiguity sets. This can be attributed to the difficulty of computing tight, global-Lipschitz like, norm-linearizations for Neural Networks. Such computations require relaxing the network, which introduces conservatism and leads to expanding radii. Nevertheless, the ambiguity sets still provide informative insight on the closeness in expectation to the equilibrium of the true distribution expectation after 20 time steps. For the Piecewise linear system, which includes non-contracting modes, the ambiguity sets gradually grow over time, as expected. In this case, the circles enable us to conclude that in expectation none of the true distributions enters the obstacle regions before reaching the target regions.

B. Parametric analysis

We start our analysis from the upper part of Table II, which evaluates how the hyper-parameters of Algorithm 1, that is, the quantization size $|\mathcal{C}|$ and the number post-compression components N, influence the radius θ_{x_k} of the 2-Wasserstein ambiguity sets after 20 time steps. As expected, we observe a monotonic decrease in the ambiguity radius as $|\mathcal{C}|$ or N increases, which confirms that finer quantization and richer mixture approximations lead to tighter ambiguity sets. The relative minor improvement when increasing the signature size from 100 to 1000 suggests that a relatively small number of 100 quantization points suffices for accurate approximations. This also holds for the higher dimensional Quadruple Tank benchmark, indicating that our approach is able to obtain tight bounds also for higher dimensional systems, by only placing locations in regions with high probability mass. While one might expect only marginal gains from increasing the number of components from 5 to 10, given the unimodal and bimodal behavior of the systems in Figure 2, we observe a consistent decrease in radius. This effect arises because the compression step is applied to the discrete distribution $\mathbb{D}_{k+1}$ in line 4 of Algorithm 1, effectively reducing its support size and introducing approximation error. In the lower part of Table II, we analyze how the ambiguity sets evolve for different values of the initial and noise uncertainty radii θ_{x_0} and θ_ω. As expected, the final radius θ_{x_k} after 20 time steps is largely determined by the noise uncertainty. The exponential accumulation of θ_ω over time also explains the sharp increase in ambiguity radius when it increases from 0.01 to 0.1.

$\lvert\mathcal{C}\rvert$	N	Quadruple Tank	Double Spiral	NN Pendulum
10	1	0.524	0.115	0.554
	5	0.437	0.110	0.492
	10	0.400	0.105	0.447
100	1	0.456	0.103	0.466
	5	0.393	0.099	0.423
	10	0.354	0.096	0.391
1000	1	0.418	0.098	0.424
	5	0.369	0.095	0.392
	10	0.325	0.093	0.355
θ_{x_0}	θ_ω			
0.001	0.001	0.358	0.054	0.423
	0.01	0.393	0.099	0.610
	0.1	0.748	0.544	2.488
0.01	0.001	0.358	0.054	0.433
	0.01	0.393	0.099	0.620
	0.1	0.748	0.544	2.497
0.1	0.001	0.358	0.056	0.531
	0.01	0.394	0.100	0.718
	0.1	0.748	0.545	2.596

TABLE II: Radii of the 2-Wasserstein ambiguity balls, θ_{x_k}, obtained via Algorithm 1 after 20 time steps. The table has two parts: the upper part show results for different quantization ($|\mathcal{C}|$) and compression (N) sizes, with fixed uncertainty radii $\theta_{x_0} = \theta_\omega = 0.001$ for the NN Pendulum and $\theta_{x_0} = \theta_\omega = 0.01$ else; the lower part shows results for varying θ_{x_0} and θ_ω, with $|\mathcal{C}| = 100$ and $N = 10$.

VII. Conclusion

We introduced a novel framework to formally propagate uncertainty in non-linear stochastic dynamical systems with additive noise. Our approach relies on enveloping the uncertainty in the system in tractable convex ambiguity sets of probability distributions, described as ρ-Wasserstein ambiguity sets. By allowing for uncertain initial state and noise distributions, our formulation encompasses a large class of modern control problems, such as Neural Network dynamical systems or data-driven systems. On a set of numerical experiments based on benchmarks from the control community, we demonstrated the efficacy of our framework. As future research, we see at least two directions for further improvement of the current method: i) alternative methods for compression of Gaussian mixtures, needed to guarantee scalability to the framework, and ii) less conservative ways to compute the linearization coefficients in Eqn (9).

VIII. Proofs

a) Proof of Proposition 1: We will prove this Proposition by induction. First, we know that $\mathbb{S}_{x_0} \subseteq \mathbb{B}_{\theta_{x_0}}(\bar{\mathbb{P}}_{x_0})$ since $\mathbb{S}_{x_0} = \mathbb{B}_{\theta_{x_0}}(\bar{\mathbb{P}}_{x_0})$, which is our base case. Now, for $k \geq 0$ assume $\mathbb{S}_{x_k} \subseteq \mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k})$. Then, for any $\mathbb{Q} \in \mathbb{S}_{x_{k+1}}$,

$$\begin{aligned}
\mathbb{W}_\rho(\mathbb{Q}, \bar{\mathbb{P}}_{x_{k+1}}) &\leq \sup_{\mathbb{Q}\in\mathbb{S}_{x_{k+1}}} \mathbb{W}_\rho(\mathbb{Q}, \bar{\mathbb{P}}_{x_{k+1}}) \\
&= \sup_{\substack{\mathbb{P}\in\mathbb{S}_{x_k},\\ \mathbb{P}_\omega\in\mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)}} \mathbb{W}_\rho(f\#\mathbb{P} * \mathbb{P}_\omega, \bar{\mathbb{P}}_{x_{k+1}}) \\
&\leq \sup_{\substack{\mathbb{P}\in\mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k}),\\ \mathbb{P}_\omega\in\mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)}} \mathbb{W}_\rho(f\#\mathbb{P} * \mathbb{P}_\omega, \bar{\mathbb{P}}_{x_{k+1}}) \leq \theta_{x_{k+1}}.
\end{aligned}$$

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.

Thus, $\mathbb{S}_{x_{k+1}} \subseteq \mathbb{B}_{\theta_{x_{k+1}}}(\bar{\mathbb{P}}_{x_{k+1}})$, which concludes the proof. □

b) Proof of Proposition 2: First, we note that for any $\mathbb{P}, \hat{\mathbb{P}}, \mathbb{Q}, \hat{\mathbb{Q}} \in \mathcal{P}_\rho(\mathcal{X})$:

$$\begin{aligned}\mathbb{W}_\rho(\mathbb{P} * \mathbb{Q}, \hat{\mathbb{P}} * \hat{\mathbb{Q}}) &\leq \mathbb{W}_\rho(\mathbb{P} * \mathbb{Q}, \hat{\mathbb{P}} * \mathbb{Q}) + \mathbb{W}_\rho(\hat{\mathbb{P}} * \mathbb{Q}, \hat{\mathbb{P}} * \hat{\mathbb{Q}}) \\ &\leq \mathbb{W}_\rho(\mathbb{P}, \hat{\mathbb{P}}) + \mathbb{W}_\rho(\mathbb{Q}, \hat{\mathbb{Q}}),\end{aligned}$$

where the first inequality holds by the triangle inequality and the second one by Proposition 10 in [11]. Then, for any $\mathbb{P} \in \mathbb{B}_{\theta_{x_k}}(\bar{\mathbb{P}}_{x_k})$, $\mathbb{P}_\omega \in \mathbb{B}_{\theta_\omega}(\bar{\mathbb{P}}_\omega)$, the previous result yields:

$$\begin{aligned}&\mathbb{W}_\rho(f\#\mathbb{P} * \mathbb{P}_\omega, \bar{\mathbb{P}}_{x_{k+1}}) \\ &\quad = \mathbb{W}_\rho(f\#\mathbb{P} * \mathbb{P}_\omega, f\#\Delta_{\boldsymbol{\mathcal{C}}_k}\#\bar{\mathbb{P}}_{x_k} * \bar{\mathbb{P}}_\omega) \\ &\quad \leq \mathbb{W}_\rho(f\#\mathbb{P}, f\#\Delta_{\boldsymbol{\mathcal{C}}_k}\#\bar{\mathbb{P}}_{x_k}) + \mathbb{W}_\rho(\mathbb{P}_\omega, \bar{\mathbb{P}}_\omega) \\ &\quad \leq \mathbb{W}_\rho(f\#\mathbb{P}, f\#\Delta_{\boldsymbol{\mathcal{C}}_k}\#\bar{\mathbb{P}}_{x_k}) + \theta_\omega. \qquad (13)\end{aligned}$$

To conclude, we apply the sup operator on both sides of Eqn (13) and use Theorem 3 in [13] to the left term. □

c) Proof of Proposition 3: First, we note that by construction, as the coefficient pair $(\alpha_{k,\ell}, \beta_{k,\ell}) = (\mathcal{L}_f, 0)$ always satisfies Eqn (9) for any point $c_{k,\ell} \in \boldsymbol{\mathcal{C}}_k$, it holds that:

$$\begin{aligned}&\theta_\omega + \left(\hat{\alpha}_k(\theta_k + \theta_{\Delta,k})^\rho + \sum_{\ell=1}^{|\boldsymbol{\mathcal{C}}_k|} \bar{\mathbb{P}}_{x_k}(\mathcal{R}_{k,\ell})\beta_{k,\ell}\right)^{\frac{1}{\rho}} \\ &\leq \theta_\omega + \mathcal{L}_f(\theta_k + \theta_{\Delta,k}) \leq \theta_\omega + \mathcal{L}_f(\theta_k + \epsilon).\end{aligned}$$

where the second inequality comes from the fact that $\theta_{\Delta,k} \leq \epsilon$. Now, note that the map $T(\theta) = \theta_\omega + \mathcal{L}_f(\theta + \epsilon)$ is contractive for $\mathcal{L}_f < 1$ (i.e. $|T(\theta) - T(\tilde{\theta})| \leq |\theta - \tilde{\theta}|$). Then, by the Banach fixed point theorem [27], for the sequence $\theta_{k+1} = T(\theta_t)$, it holds that $\lim_{k\to\infty} \theta_k = \theta^*$, where θ^* is given by

$$\theta^* = T(\theta^*) \iff \theta^* = \frac{\theta_\omega}{1 - \mathcal{L}_f} + \frac{\mathcal{L}_f}{1 - \mathcal{L}_f}\epsilon,$$

which concludes the proof. □

IX. Acknowledgments

L.L. and E.F. are partially supported by the NWO (grant OCENW.M.22.056).

References

[1] R. F. Stengel, *Optimal control and estimation*. Courier Corporation, 1994.

[2] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[3] C. Williams and C. Rasmussen, "Gaussian processes for regression," *Advances in neural information processing systems*, vol. 8, 1995.

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

[5] D. Landgraf, A. Völz, F. Berkel, K. Schmidt, T. Specker, and K. Graichen, "Probabilistic prediction methods for nonlinear systems with application to stochastic model predictive control," *Annual Reviews in Control*, vol. 56, p. 100905, 2023.

[6] M. Deisenroth and C. E. Rasmussen, "Pilco: A model-based and data-efficient approach to policy search," in *Proceedings of the 28th International Conference on machine learning (ICML-11)*, 2011, pp. 465–472.

[7] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[8] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

[9] J. Duník, S. K. Biswas, A. G. Dempster, T. Pany, and P. Closas, "State estimation methods in navigation: Overview and application," *IEEE Aerospace and Electronic Systems Magazine*, vol. 35, no. 12, pp. 16–31, 2020.

[10] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE transactions on automatic control*, vol. 45, no. 5, pp. 910–927, 2000.

[11] L. Aolaritei, N. Lanzetti, H. Chen, and F. Dörfler, "Distributional uncertainty propagation via optimal transport," *arXiv preprint arXiv:2205.00343*, 2022.

[12] E. Figueiredo, A. Patane, M. Lahijanian, and L. Laurenti, "Uncertainty propagation in stochastic systems via mixture models with error quantification," *2024 IEEE 63rd Conference on Decision and Control (CDC)*, 2024.

[13] E. Figueiredo, S. Adams, P. Esfahani, and L. Laurenti, "Efficient uncertainty propagation with guarantees in wasserstein distance," 2025. [Online]. Available: https://mohajerinesfahani.github.io/Publications/journal/2025/Propagation.pdf

[14] C. Villani *et al.*, *Optimal transport: old and new*. Springer, 2008, vol. 338.

[15] G. Peyré, M. Cuturi *et al.*, "Computational optimal transport: With applications to data science," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[16] A. Nagabandi, G. Kahn, R. S. Fearing, and S. Levine, "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 7559–7566.

[17] S. Adams, M. Lahijanian, and L. Laurenti, "Formal control synthesis for stochastic neural network dynamic models," *IEEE Control Systems Letters*, vol. 6, pp. 2858–2863, 2022.

[18] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1, pp. 115–166, 2018.

[19] N. Fournier and A. Guillin, "On the rate of convergence in wasserstein distance of the empirical measure," *Probability theory and related fields*, vol. 162, no. 3, pp. 707–738, 2015.

[20] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.

[21] S. Adams, M. Lahijanian, L. Laurenti *et al.*, "Finite neural networks as mixtures of gaussian processes: From provable error bounds to prior selection," *arXiv preprint arXiv:2407.18707*, 2024.

[22] J. Delon and A. Desolneux, "A wasserstein-type distance in the space of gaussian mixture models," *SIAM Journal on Imaging Sciences*, vol. 13, no. 2, pp. 936–970, 2020.

[23] V. I. Bogachev, *Measure theory*. Springer, 2007, vol. 1, no. 1.

[24] S. Boyd, "Convex optimization," *Cambridge UP*, 2004.

[25] R. Gao and A. Kleywegt, "Distributionally robust stochastic optimization with wasserstein distance," *Mathematics of Operations Research*, vol. 48, no. 2, pp. 603–655, 2023.

[26] R. J. Adler and J. E. Taylor, *Random fields and geometry*. Springer Science & Business Media, 2009.

[27] K. Goebel and W. A. Kirk, *Topics in metric fixed point theory*. Cambridge university press, 1990, no. 28.

[28] F. B. Mathiesen, S. C. Calvert, and L. Laurenti, "Safety certification for stochastic systems via neural barrier functions," *IEEE Control Systems Letters*, vol. 7, pp. 973–978, 2022.

[29] I. Gracia, D. Boskos, M. Lahijanian, L. Laurenti, and M. Mazo Jr, "Efficient strategy synthesis for switched stochastic systems with distributional uncertainty," *Nonlinear Analysis: Hybrid Systems*, vol. 55, p. 101554, 2025.

[30] G. Oriolo, "Stability theory for nonlinear systems," 2025, sapienza University of Rome. [Online]. Available: http://www.diag.uniroma1.it/oriolo/amr/material/stability.pdf

[31] K. H. Johansson, "The quadruple-tank process: A multivariable laboratory process with an adjustable zero," *IEEE Transactions on control systems technology*, vol. 8, no. 3, pp. 456–465, 2000.

Authorized licensed use limited to: TU Delft Library. Downloaded on January 15,2026 at 13:53:02 UTC from IEEE Xplore. Restrictions apply.