Dynamic Evolution of Distributional Ambiguity Sets and Precision Tradeoffs in Data Assimilation

Dimitris Boskos Jorge Cortés Sonia Martínez

Abstract— **This paper studies the evolution of ambiguity sets employed in distributionally robust optimization problems. We assume the unknown distribution of the random variable evolves according to a known deterministic dynamics. Assuming that the initial distribution of the data is compactly supported, we study how the assimilation of samples collected during some time interval evolution can be leveraged to make inferences about the unknown distribution of the process at the sampling horizon end. Under perfect knowledge of the dynamics' flow map, we provide sufficient conditions relating the solutions' growth and the sampling rate, which establish reduction of the ambiguity set size as the horizon increases. In the case where numerical errors are modeled during the computation of the flow, or the dynamics are subject to an unknown bounded disturbance, we also characterize the exploitable sample history that results in the guaranteed reduction of the ambiguity set.**

I. Introduction

Stochastic optimization constitutes a natural framework to address mathematical programming problems with probabilistic uncertainty, finding applications in numerous domains, including finance, control technology, and communication networks. To provide reliable results also in the case where the underlying probabilistic structure of the data is unknown, distributionally robust optimization (DRO) has emerged as a promising approach. A characteristic feature of DRO formulations is that the worst-case performance of the optimization problem is evaluated over a set of probability distributions containing the true one, termed an ambiguity set, which can be characterized through their appropriate closeness to the unknown distribution in a certain metric.

To a large extent, the formulation of DRO problems is performed under the assumption that the exploitable data is measured in a direct manner. However, there are several interesting problems where this is no longer the case, since the data can be governed by a time-varying process and is collected progressively. Thus, previous values of the assimilated samples are not directly exploitable to infer about the current status of the process. Our goal in this paper is to consider DRO problems when the time-varying data evolves according to some dynamic law. This motivates the study of how the associated ambiguity sets evolve in time and the identification of trade-offs between the amount of progressively assimilated data and its future adequacy, due to gradual precision loss in its predicted values.

Literature review: DRO optimization is an area of stochastic programming [1] which has gained significant recent research attention [9], [18], [4], in view of the progress on robust optimization during the last two decades [2]. A main characteristic of DRO is that worst-case decisions against model uncertainty can be quantified with performance guarantees, by considering a set of distributions up to a certain distance from a candidate model. There is an exhaustive number of choices for distances in spaces of probability distributions [16]. Among the most popular distance-type notions for DRO problems are ϕ-divergences [3], [12], and Wasserstein metrics [11], [6]. For data-driven problems where robustness is measured with respect to the empirical distribution, the Wasserstein distance becomes a suitable choice, since it does not require any absolute continuity conditions between the associated distributions. The work [9] leverages recent inequalities [10] on concentration of measure for the characterization of Wasserstein ambiguity sets around the empirical distribution and provides tractable reformulations of the resulting infinite-dimensional problem into finite-dimensional convex programs. These formulations with out-of-sample guarantees are employed in [7], [8], where a distributed reformulation of the min-max DRO problem is established via saddle-point dynamics, and in [15], where online sample assimilation is fused with an efficient optimization algorithm to provide on-the-fly data-driven DRO solutions. It is worthwhile mentioning also the work [5], where the notion of a robust Wasserstein profile function is leveraged, providing fast asymptotic convergence rates for high-dimensional samples. Recent work has considered distributionally robust Kalman filtering approaches for the state estimation of uncertain time-varying processes for the Kullback-Leibler [14], τ-divergences [20], and the Wasserstein [17] metrics.

Statement of contributions: Our motivation comes from the construction of ambiguity sets for DRO problems where the data generating process is time-varying and samples are collected in an online manner. To form the ambiguity sets we build on concentration of measure results, which establish appropriate closeness between the empirical and actual distribution in the Wasserstein metric. As discussed in [9], the choice of such a metric compared to divergences comes from the requirement to consider distributions admitting densities (for such distributions, the divergence from the empirical distribution is infinite, resulting in ambiguity sets that contain the true distribution with zero confidence). Regarding the data evolution, we consider dynamic processes with random initial conditions, and assume that samples are assimilated through multiple independent realizations of the

The authors are with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, {dboskos,cortes,soniamd}@ucsd.edu.

process, progressively in time. Samples collected up to a given time instant are used to infer about the corresponding unknown distribution of the process. Thus, we can draw conclusions about the expected state at this instant for another, unobserved independent realization of the system. This differentiates our approach from Kalman filtering, where the objective is to estimate the state of a single trajectory based on noisy observations of it. In particular, we focus on how such prior collected samples can be pushed forward in time and fused with newly assimilated data to provide information about the current distribution of the process. In addition, we identify the limitations of this fusion, in view of computation errors and disturbances. Our first contribution is to provide sufficient conditions relating the random elements' growth and the sampling rate, which establish convergence of the ambiguity radius. Our second contribution is the quantification of an effective sampling horizon over which the associated samples qualify for uncertainty reduction in the presence of model disturbances or computation errors during the pushforward. Due to space constraints, all proofs are omitted and will appear elsewhere.

Organization: Section II introduces notation and technical preliminaries. Section III presents the problem statement. In Section IV we present concentration of measure results for compactly supported distributions and leverage them to study the relation between the dynamics and asymptotic convergence of the ambiguity sets. Section V characterizes the effect of computation errors and disturbances on the ambiguity size. We gather our conclusions and ideas for future work in Section VI.

II. Preliminaries

In this section we present general notation and concepts from probability theory that will be used throughout the paper.

Notation: We denote by $\|\cdot\|$ and $\|\cdot\|_\infty$, the Euclidean and infinity norm in $\mathbb{R}^d$, respectively, and by $[n_1:n_2]$, the set of integers $\{n_1, n_1+1, \ldots, n_2\} \subset \mathbb{N}$. Given a differentiable function $G:\mathbb{R}^n \to \mathbb{R}^m$ we use the notation $DG(x)$ for its derivative at $x \in \mathbb{R}^n$.

Probability theory: We denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-algebra on $\mathbb{R}^d$, and by $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Given a real number $p \geq 1$, we denote by $\mathcal{P}_p(\mathbb{R}^d)$ the probability measures in $\mathcal{P}(\mathbb{R}^d)$ with finite p-th moment, i.e.,

$$\mathcal{P}_p(\mathbb{R}^d) := \Big\{\mu \in \mathcal{P}(\mathbb{R}^d) \mid \int_{\mathbb{R}^d} \|x\|^p d\mu < \infty\Big\}.$$

For any $p \geq 1$, and probability measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, their Wasserstein distance is defined as

$$W_p(\mu,\nu) := \Big(\inf_{\pi \in \mathcal{H}(\mu,\nu)} \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \pi(dx,dy) \Big\} \Big)^{1/p},$$

where $\mathcal{H}(\mu,\nu)$ is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, respectively. Also, given two measurable spaces $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$, and a measurable map Ψ from $(\Omega,\mathcal{F})$ to $(\Omega',\mathcal{F}')$, the pushforward map $\Psi_\#$ assigns to each measure μ in $(\Omega,\mathcal{F})$, a new measure ν in $(\Omega',\mathcal{F}')$, defined by $\nu := \Psi_\#\mu$ iff $\nu(B) = \mu(\Psi^{-1}(B))$ for all $B \in \mathcal{F}'$. The map $\Psi_\#$ is linear and satisfies $\Psi_\#\delta_\omega = \delta_{\Psi(\omega)}$, with δ_ω the Dirac measure centered at $\omega \in \Omega$.

III. Problem Formulation

Traditionally, a DRO Problem (DROP) is formulated as

$$\text{(P)} \qquad \inf_{x \in \mathcal{X}} \sup_{P \in \widehat{\mathcal{P}}^N} \mathbb{E}_P[f(x,\xi)],$$

where $x \in \mathcal{X} \subset \mathbb{R}^n$ is the decision variable and ξ represents a random variable distributed according to an unknown distribution $P_\xi \in \mathcal{P}(\mathbb{R}^d)$. Thus, a "worst-case" expectation problem is set up over an *ambiguity set* $\widehat{\mathcal{P}}^N$, which contains the true distribution P_ξ of ξ with confidence $1-\beta$. This ambiguity set is built based on N i.i.d. samples $\xi^1, \ldots, \xi^N$ of the unknown distribution $P \equiv P_\xi$ from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and which satisfies

$$\mathbb{P}(P_\xi \in \widehat{\mathcal{P}}^N) \geq 1-\beta.$$

We will also refer to this common formulation as a *static* DROP. Instead, we are interested in building this set online, as it may not be possible to process many samples instantaneously in view of the fact that the random element is dynamically varying.

A. Motivating example

Assume that the random data evolves according to certain dynamics over a given time horizon $[0,T]$. As a concrete example, let $\xi_t = (\xi_{1t}, \xi_{2t})$ describe the position and velocity of a unit acceleration particle, with dynamics

$$\dot{\xi}_{1t} = \xi_{2t}, \quad \dot{\xi}_{2t} = 1. \tag{1}$$

The initial position and velocity of the particle are considered random elements with an unknown probability distribution. Therefore, its state at each t is a random variable with law P_{ξ_t}. Our goal is to describe the trajectory of state distributions over time, motivated by solving a DROP problem at time T, i.e., at the end of the horizon. Given that these distributions are unknown, we will focus on specifying an ambiguity set $\widehat{\mathcal{P}}_T^N$ at time T which contains the true distribution P_{ξ_T} with a given confidence. For this, we assume that independent samples from the distribution of ξ_t are collected at time instants $0 \leq t_1 < \ldots < t_N = T$. That is, at each t_i we sample once from the distribution $P_{\xi_{t_i}}$ as shown in Figure 1. Based on the conventional approach of solving static DROPs of the form (P), a first attempt to solve a DROP with respect to P_{ξ_T} would be based on constructing an ambiguity set based on the single sample $\xi_{t_N}^N$. However, such an approach would result in poor performance, given that reliable ambiguity sets require a finite but sufficiently large amount of data. This motivates the problem of how to leverage the collection of samples at previous time instants to improve performance at time T. Note that each such previous sample can alternatively be considered as a realization of the state of system (1) at the corresponding time instant, as

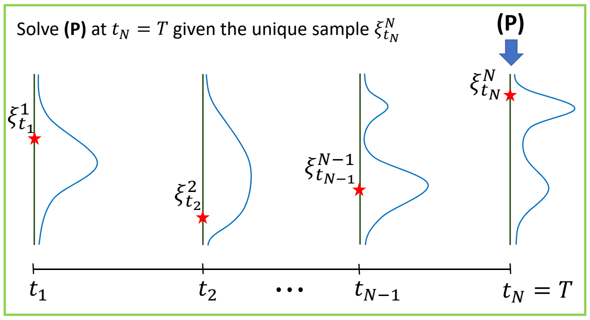

Fig. 1. Illustration of a time-varying distribution P_{ξ_t} from which samples are drawn at successive time instants $t_1, \ldots, t_N$.

shown in Figure 2. Thus, given the sampled value of (1) at t_i, for $i = 1, \ldots, N-1$, it is possible to determine its corresponding value at T by pushing $\xi^i_{t_i}$ forward through the dynamics' flow map Φ. In particular, we have that $\xi^i_T = \Phi_{T,t_i}(\xi^i_{t_i})$, where for any $0 \le s \le t \le T$, and state $\xi_s = (\xi_{1s}, \xi_{2s})$ of (1) at time s, the state at t is explicitly obtained as

$$\Phi_{t,s}(\xi_s) = (\xi_{1s} + (t-s)\xi_{2s} + (t-s)^2/2, \xi_{2s} + (t-s)).$$

Based on these considerations, assume that an observer collects samples from N independent particles evolving according to the same dynamics (1) and i.i.d. initial conditions ξ^i_0. Creating an ambiguity set by using all these samples as above enables the observer to make probabilistic inferences with high confidence at time T about an other unobserved particle, which has the same dynamics and unknown random characteristics. In this example, the dynamics are perfectly known and not subject to computational errors. A question of interest will be the analysis of the accuracy-horizon-length tradeoffs when the flow map can not accurately be computed.

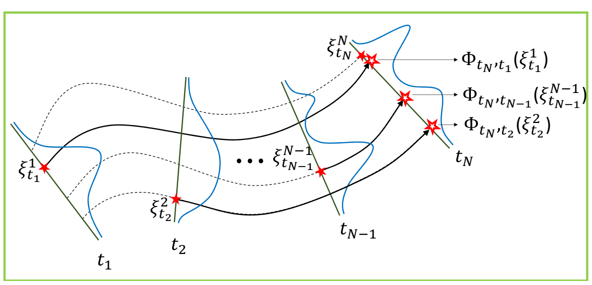

Fig. 2. Illustration of how samples collected in previous time instants can be pushed forward through the flow map to construct higher-confidence ambiguity sets.

B. Fixed horizon dynamic DROP

Building on the example in Section III-A, here we depart from the static DROP paradigm and formulate the *dynamic* DROPs considered in this paper, where the data evolves according to the dynamics

$$\dot{\xi}_t = F(t, \xi_t), \qquad \xi_t \in \mathbb{R}^d. \tag{2}$$

The initial condition ξ_0 is considered random with an unknown distribution P_{ξ_0}. Based on the evolution of (2) over a time horizon $[0, T]$, we are interested in solving a DROP with respect to the unknown distribution P_{ξ_T}. In analogy to the example, we make the following assumption for the sampling process.

Assumption 3.1: The data are collected from independent trajectories ξ^i, $i = 1, 2, \ldots$ of (2) with i.i.d. initial conditions ξ^i_0, over a strictly increasing sequence of sampling instants $0 \le t_1 < \cdots < t_{\bar{N}} = T$. At each t_i, a single sample $\xi^i_{t_i}$ is taken from trajectory ξ^i.

This hypothesis is made without any loss of generality to simplify the exposition. We will call any time length

$$\Delta \ge \max\{t_i - t_{i-1} \,|\, i \in [1:\bar{N}]\}$$

an *inter-sampling time bound*, and introduce the *effective sampling horizon* $[N^\flat : \bar{N}]$, $N^\flat \in [1 : \bar{N}]$, to denote that the data used for the ambiguity set construction is collected from time $t_{N^\flat}$ up to time $t_{\bar{N}}$. We now formulate the problem.

Problem formulation: Based on the collected data along $[0, T]$, determine an ambiguity set that contains the true distribution P_{ξ_T} with a given confidence. Also, in the presence of numerical errors or disturbances, whose effect on the ambiguity set accuracy integrates with time, quantify the effective sampling horizon length, up to which the ambiguity set is guaranteed to improve with the number of samples.

In Section IV we will study properties of dynamic ambiguity sets assuming perfect knowledge of the flow map, and data-assimilation-precision tradeoffs due to numerical errors or disturbances will be considered in Section V.

IV. Fixed-Horizon Ambiguity Sets Under Dynamic Data

In this section, we characterize ambiguity sets that appear in dynamic DROPs, as described in the problem formulation of Section III-B. Recall that according to Assumption 3.1, we consider a strictly increasing sequence of sampling times $0 \le t_1 < \cdots < t_{\bar{N}} = T$ where at each t_i, the sample $\xi^i_{t_i}$ is collected from a trajectory ξ^i of (2). The ambiguity set $\widehat{\mathcal{P}}^N_T$ is built based on the $N = \bar{N} - N^\flat + 1$ last effective samples by leveraging concentration of measure inequalities, which upper bound the distance between the empirical and actual distribution with a given confidence. In this way, using the Wasserstein metric W_p in the space of probability measures $\mathcal{P}_p(\mathbb{R}^d)$, *an ambiguity ball* is constructed, with a center $\widehat{P}^N_{\xi_T}$ that corresponds to the empirical distribution built from N independent samples ξ^i_T at time T, and a radius $\varepsilon_N(\beta)$ which depends on a selected confidence $1 - \beta$. More precisely, the true distribution P_{ξ_T} is $\varepsilon_N(\beta)$-close to the empirical $\widehat{P}^N_{\xi_T}$ with probability at least $1 - \beta$, namely,

$$\mathbb{P}(W_p(\widehat{P}^N_{\xi_T}, P_{\xi_T}) \le \varepsilon_N(\beta)) \ge 1 - \beta.$$

Recalling that in our problem formulation the samples $\xi^i_{t_i}$, $i \in [N^\flat : \bar{N}]$ are progressively collected prior to $t_{\bar{N}} = T$, we alternatively build the *cumulative empirical distribution*

$\bar{P}^N_{\xi_T}$, based on corresponding predicted values $\bar{\xi}^i_T$ of these samples at T.

To formalize the above discussion, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a finite sequence of i.i.d. $\mathbb{R}^d$-valued random variables $\{\xi^i_0\}_{i\in[1:\bar{N}]}$, with each ξ^i_0 representing the initial condition of a trajectory ξ^i. All random variables ξ^i_0 have the same law $P_{\xi_0} \equiv P$, since they are identically distributed. The corresponding trajectories evolve according to the dynamics (2). We assume that F in (2) is continuous and locally Lipschitz in ξ, and that the system is forward complete, namely, for each $s \geq 0$ and initial condition ξ the solution $\xi_t(s,\xi)$ is defined for all $t \geq s$. Then, the flow map $\Phi : \mathcal{D}_\Phi \to \mathbb{R}^d$, where

$$\mathcal{D}_\Phi := \{(t, s, \xi) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}^d : t \geq s\},$$

is defined by

$$\Phi(t, s, \xi) := \xi_t(s, \xi),$$

inducing a family of maps $\Phi_{t,s} : \mathbb{R}^d \to \mathbb{R}^d$. Assuming that all trajectories start at time $s = 0$ and using the notation $\Phi_t = \Phi_{t,0}$, the state of each trajectory ξ^i at certain $t \geq 0$ will be given by the random variable $\xi^i_t = \Phi_t \circ \xi^i_0$, with common law $P_{\xi_t} = \Phi_{t\#}P$. We next show that under perfect knowledge of the flow map, the cumulative empirical distribution at T formed by the predicted values $\bar{\xi}^i_T := \Phi_{T,t_i}(\xi^i_{t_i})$ of the progressively collected samples, produces an equivalent empirical distribution to that corresponding to samples extracted at the last time. This will be later leveraged to show that under appropriate relations among the growth of the system's solutions and the sampling rate, the ambiguity radius converges to zero as the horizon tends to infinity.

Lemma 4.1: (*Sample ideal pushforward*). Consider a sequence of trajectories ξ^i as in Assumption 3.1 and the empirical distribution $\widehat{P}^N_{\xi_T}$ formed by the samples of trajectories $\xi^{N^\flat}, \ldots, \xi^{\bar{N}}$ at T, namely,

$$\widehat{P}^N_{\xi_T} := \frac{1}{N} \sum_{i=N^\flat}^{\bar{N}} \delta_{\xi^i_T}.$$

Then, (i) all ξ^i_T are i.i.d.; and (ii) if we consider the cumulative empirical distribution

$$\bar{P}^N_{\xi_T} := \frac{1}{N} \sum_{i=N^\flat}^{\bar{N}} \delta_{\bar{\xi}^i_T}, \tag{3}$$

with $\bar{\xi}^i_T := \Phi_{T,t_i}(\xi^i_{t_i})$, $i \in [N^\flat : \bar{N}]$, it holds that $\bar{P}^N_{\xi_T} = \widehat{P}^N_{\xi_T}$.

A. Concentration of compactly supported measures

We next present results from concentration of measure to determine the radius of the ambiguity set $\widehat{\mathcal{P}}^N_T$ that contains the true distribution P_{ξ_T} of the data at T with a selected confidence. Our focus will be on the class of compactly supported distributions, which is preserved under the flow of forward complete systems. The following proposition is based on [10, Proposition 10] and provides a concentration inequality for such laws and its explicit dependence on the distribution's support.

Proposition 4.2: (*Concentration inequality*). Consider a sequence $(X_i)_{i\in\mathbb{N}}$ of i.i.d. $\mathbb{R}^d$-valued random variables with law μ, supported on the compact set K. Then, for any $p \geq 1$, $N \geq 1$, and $\varepsilon > 0$, it holds that

$$\mathbb{P}(W^p_p(\widehat{\mu}^N, \mu) \geq \varepsilon) \leq \chi_N(\varepsilon, \rho; p, d),$$
$$\chi_N(\varepsilon, \rho) := C \begin{cases} e^{-\frac{cN}{\rho^{2p}}\varepsilon^2}, & \text{if } p > d/2, \\ e^{-cN\frac{\varepsilon^2}{\rho^{2p}(\ln(2+\rho^p/\varepsilon))^2}}, & \text{if } p = d/2, \\ e^{-\frac{cN}{\rho^d}\varepsilon^{\frac{d}{p}}}, & \text{if } p < d/2, \end{cases}$$

where

$$\widehat{\mu}^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_i}, \quad \rho := \frac{1}{2}\sup\{\|x - y\|_\infty \,|\, x, y \in K\},$$

and the constants C and c depend only on p and d.

The following corollary to Proposition 4.2 characterizes the radius of the ambiguity balls in terms of the selected confidence and the support of the unknown distribution. It is leveraged to quantify how the flow map affects this radius, by pushing forward the initial distribution.

Corollary 4.3: (*Ambiguity radius*). Under the assumptions of Proposition 4.2, for any confidence $1 - \beta$, $\beta \in (0, 1)$, it holds that

$$\mathbb{P}(W_p(\widehat{\mu}^N, \mu) \leq \varepsilon_N(\beta, \rho)) \geq 1 - \beta,$$

where

$$\varepsilon_N(\beta, \rho) := \begin{cases} \left(\frac{\ln(C\beta^{-1})}{c}\right)^{\frac{1}{2p}} \frac{\rho}{N^{\frac{1}{2p}}}, & \text{if } p > d/2, \\ h^{-1}\left(\frac{\ln(C\beta^{-1})}{cN}\right)^{\frac{1}{p}} \rho, & \text{if } p = d/2, \\ \left(\frac{\ln(C\beta^{-1})}{c}\right)^{\frac{1}{d}} \frac{\rho}{N^{\frac{1}{d}}}, & \text{if } p < d/2, \end{cases} \tag{4}$$

with h^{-1} the inverse of $h(x) = \frac{x^2}{(\ln(2+1/x))^2}$, $x > 0$.

B. Growth conditions for ambiguity radius convergence

Here, we present sufficient conditions, related to the system's dynamics and the sampling rate, to guarantee that for any prescribed confidence, the ambiguity radius converges to zero as the horizon becomes infinitely large. We will therefore need the following Lyapunov-type result.

Proposition 4.4: (*Lyapunov-type growth rate condition*). Assume that F in (2) is continuous and locally Lipschitz in ξ, and that there exist a locally integrable function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}$, and a function $V \in C^1(\mathbb{R}^d; \mathbb{R})$, such that

$$a_1\|\xi\|^r \leq V(\xi) \leq a_2\|\xi\|^r, \quad \forall \xi \in \mathbb{R}^d,$$
$$DV(\xi)F(t,\xi) \leq \alpha(t)V(\xi) + M_1 V(\xi)^q, \quad \forall t \geq 0, \xi \in \mathbb{R}^d \setminus \{0\},$$
$$\int_{t_1}^{t_2} \alpha(t)dt \leq M_2, \quad \forall\, t_2 \geq t_1 \geq 0,$$

with $a_1, a_2 > 0$, $M_1, M_2 \geq 0$, $r > 1$, and $q \in (-\infty, 1)$. Then, for any initial condition $\xi_0 \in \mathbb{R}^d$, the solution of

(2) is defined for all $t \geq 0$ and there exist constants $\bar{c}, \bar{M}$, depending on $\|\xi_0\|$ and r, a_1, a_2, M_1, M_2, q, such that

$$\|\xi(t)\| \leq \bar{M}(1+\bar{c}t)^{\frac{1}{r(1-q)}}, \quad \forall t \geq 0.$$

We now provide the main result of this section, which shows that for bounded inter-sample durations and under the assumption that state measurements are pushed without errors forward in time, the ambiguity sets formed by dynamics which satisfy the growth conditions of Proposition 4.4 shrink as the DROP's horizon increases.

Proposition 4.5: (*Ambiguity radius convergence*). Assume that system (2) satisfies the assumptions of Proposition 4.4 and that P_{ξ_0} is supported on $K := \{\xi \in \mathbb{R}^d : \|\xi\|_\infty \leq \rho\}$, for certain $\rho > 0$. Select a confidence $1-\beta$, an exponent $p \geq 1$, and assume that

$$r(1-q) > \max\{2p, d\}, \tag{6}$$

with r, q as given in Proposition 4.4. For any horizon $[0, T]$, consider a sampling sequence as in Assumption 3.1 and a common inter-sampling time bound Δ. Let $\bar{P}^N_{\xi_T}$ be the cumulative empirical distribution in (3), with $N^\flat = 1$, $\bar{N} = N$, and

$$\rho_T := 1/2 \sup\{\|x-y\|_\infty \,|\, x, y \in \Phi_T(K)\}, \tag{7}$$

with $\Phi_T(K)$ the reachable states at T from K. Then, if we denote $\varepsilon_N(\rho_T) \equiv \varepsilon_N(\beta, \rho_T)$ with ε_N as in (4), it holds that

$$\mathbb{P}(W_p(\bar{P}^N_{\xi_T}, P_{\xi_T}) \leq \varepsilon_N(\rho_T)) \geq 1-\beta, \quad \lim_{T\to\infty} \varepsilon_N(\rho_T) = 0.$$

Note that under the conditions of Proposition 4.4 the system's state growth is at most of order $\sim t^{\frac{1}{r(1-q)}}$, i.e., lower than the typical exponential case. To additionally establish convergence of the ambiguity radius through Proposition 4.5, since $p \geq 1$, we deduce from (6) that the growth needs to be asymptotically dominated by the square root of t.

V. Ambiguity Sets under Pushforward Errors and Disturbances

Recalling the motivating example in Section III-A, it is worthwhile noting that in principle the dynamics (2) are not as convenient as in (1), where the flow map is computed explicitly. This requires the numerical integration of the system's solutions and results in an approximate model of the flow map, denoted by $\Phi^{\rm num}_{t,s}$, where $t \geq s \geq 0$. For instance, if a first order Euler scheme is used to numerically compute a trajectory of system (2) at time T, given its corresponding value ξ_τ at $\tau < T$ with $T - \tau = \kappa\Delta t$, we have that

$$\Phi^{\rm num}_{T,\tau}(\xi_\tau) = \phi^{\Delta t}_{\tau+(\kappa-1)\Delta t} \circ \cdots \circ \phi^{\Delta t}_{\tau}(\xi_\tau),$$

where

$$\phi^{\Delta t}_t(\xi) := \xi + \Delta t F(t, \xi).$$

In addition, it may happen that the dynamics of the process are subject to disturbances. Both cases suggest that the cumulative empirical distribution $\bar{P}^N_{\xi_T}$ in (3) will no longer coincide with the empirical distribution $\widehat{P}^N_{\xi_T}$ in Lemma 4.1, formed by N samples of independent trajectories at T. We proceed to quantify this difference and characterize the relevant ambiguity sets, focusing first on the case where the flow is numerically integrated. The dynamics (2) will be considered globally Lipschitz, namely,

$$\|F(t,\xi) - F(t,\xi')\| \leq L\|\xi - \xi'\|, \forall t \geq 0, \xi, \xi' \in \mathbb{R}^d, \tag{8}$$

for some $L > 0$. Thus, by invoking classical error quantification results from numerical integration [19, Theorem 3.4.7., Page 239], we have that

$$\|\Phi^{\rm num}_{t,s}(\xi) - \Phi_{t,s}(\xi)\| \leq \mathfrak{K}(e^{L(t-s)} - 1), \tag{9}$$

for all $t \geq s \geq 0$ and certain $\mathfrak{K} > 0$. To build the center of the ambiguity ball, we push the samples forward through the numerical flow and form a variant of the cumulative empirical distribution $\bar{P}^N_{\xi_T}$ in (3). Using this variant, we obtain the following characterization of the ambiguity radius.

Proposition 5.1: (*Non-ideal ambiguity radius*). Assume that system (2) satisfies (8) and $P_{\xi_{t_0}}$ is supported on the compact set K. Consider a sampling sequence as in Assumption 3.1 with inter-sampling time bound Δ and the cumulative empirical distribution $\bar{P}^N_{\xi_T}$ in (3), with $\bar{\xi}^i_T := \Phi^{\rm num}_{T,t_i}(\xi^i_{t_i})$, $i \in [N^\flat : \bar{N}]$. Then, for any confidence $1-\beta$ and $p \geq 1$ it holds that

$$\mathbb{P}(W_p(\bar{P}^N_{\xi_T}, P_{\xi_T}) \leq \psi_N(\beta, \Delta)) \geq 1-\beta, \tag{10a}$$
$$\psi_N(\beta, \Delta) := \varepsilon_N(\beta, \rho_T) + \bar{\varepsilon}_N(\Delta), \tag{10b}$$
$$\bar{\varepsilon}_N(\Delta) := \mathfrak{K}\Big(\frac{1}{N}\int_1^N (e^{L\Delta s} - 1)^p ds\Big)^{\frac{1}{p}}, \tag{10c}$$

with $\varepsilon_N(\beta, \rho_T)$ and ρ_T as given in (4) and (7), respectively.

We now quantify the effective sampling horizon up to which reduction of the ambiguity radius is guaranteed with respect to the number of samples.

Proposition 5.2: (*Effective sampling horizon*). Under the hypotheses of Proposition 5.1, assume additionally that $p \neq d/2$, and denote $\bar{p} := \max\{2p, d\}$ and

$$\bar{C} \equiv \bar{C}(\bar{p}, \beta, \rho_T) := \Big(\frac{\ln(C\beta^{-1})}{c}\Big)^{\frac{1}{\bar{p}}} \rho_T,$$

with C, c as in (4). Then, there exists $\Delta^* \equiv \Delta^*(\bar{C}, L, \mathfrak{K}) > 0$ such that for every $\Delta \in (0, \Delta^*)$, the set

$$\mathcal{N}(\Delta) := \{N \in \mathbb{N} \,|\, \bar{C}(\kappa^{-\frac{1}{\bar{p}}} - (\kappa+1)^{-\frac{1}{\bar{p}}}) > \bar{\varepsilon}_{\kappa+1}(\Delta) - \bar{\varepsilon}_\kappa(\Delta), \forall \kappa \in [1:N]\},$$

with $\bar{\varepsilon}_N$ as given in (10c), is nonempty and bounded. In addition, the ambiguity radius $\psi_N(\beta, \Delta)$ in (10b) is strictly decreasing with N, for $N \in [1 : N^*(\Delta)]$, with $N^*(\Delta) := \max(\mathcal{N}(\Delta)) + 1$.

Closed-form expressions for $\bar{\varepsilon}_N(\Delta)$ can be derived for integer values of p and facilitate the computation of $N^*(\Delta)$. For instance, when $p = 1$ we get that

$$\bar{\varepsilon}_N(\Delta) = \frac{\mathfrak{K}}{N}\Big(\frac{e^{L\Delta N} - e^{L\Delta}}{L\Delta} - (N-1)\Big).$$

It is noted that the results of this section are also applicable if we consider the sampled trajectories subject to the same

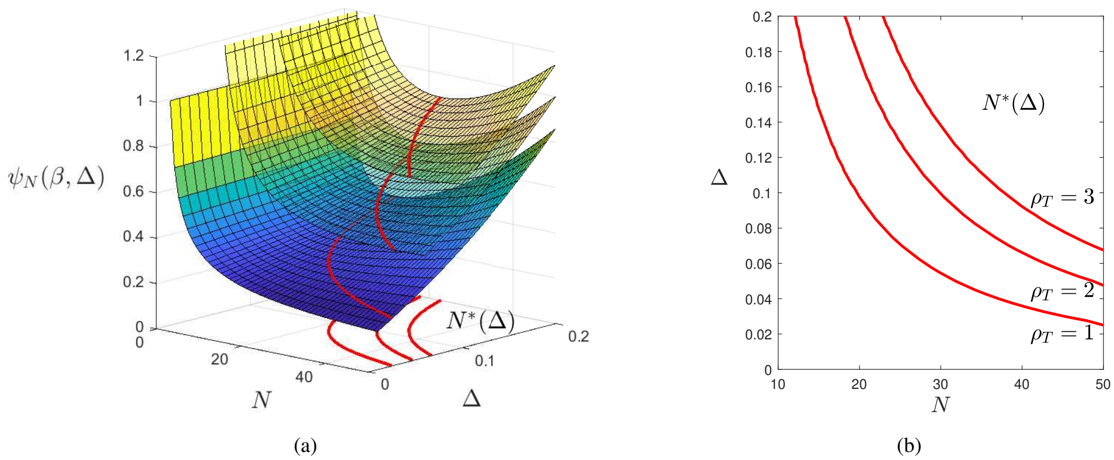

Fig. 3. (a) shows how the ambiguity radius $\psi_N(\beta, \Delta)$, cf. (10b), varies with respect to the sampling period and the number of samples in the presence of numerical errors. Given that the component ε_N of the radius, which is strictly decreasing with N, is proportional to the distributions' support size and that the effect of numerical errors is independent of ρ_T, the effective sampling horizon increases with ρ_T, as shown in (b).

unknown bounded disturbance, i.e., the dynamics are of form $\dot{\xi}_t = F(t, \xi_t, d_t)$. Then, if F is globally Lipschitz with respect to both ξ and d, it follows from the system's continuous dependence on parameters [13, Theorem 3.4], that an estimate of the form (9) will hold, with $\Phi_{t,s}$ denoting the undisturbed flow and $\Phi_{t,s}^{\text{num}}$ being replaced by the flow of $\dot{\xi}_t = F(t, \xi_t, d_t)$.

VI. Conclusions

We have provided a framework for the computation of ambiguity sets suitable for DROPs with dynamically varying data. For dynamic processes with random initial conditions from a compactly supported distribution, we have studied properties of the ambiguity sets computed at the end of the finite time horizon. Under appropriate growth conditions for the dynamics, we have shown that the ambiguity sets are guaranteed to shrink as the horizon increases. Furthermore, in the presence of numerical errors and disturbances, we have quantified the number of exploitable past samples which can be used to establish ambiguity reduction. Ongoing work includes the consideration of partial measurements when the full vector of the data is no longer available. In this case, the dynamic evolution of the process can facilitate recovery of the full data vector under appropriate observability assumptions. In addition, we aim to take into account data storage limitations in the modeling, and to employ the obtained results in receding horizon DRO formulations.

Acknowledgments

This work was supported by the DARPA Lagrange program through award N66001-18-2-4027.

References

[1] D. D. A. Shapiro *et al.*, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA: SIAM, 2014, vol. 16.

[2] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton University Press, 2009.

[3] A. Ben-Tal, D. D. Hertog, A. D. Waegenaere, B. Melenberg, and G. Rennen, "Robust solutions of optimization problems affected by uncertain probabilities," *Management Science*, vol. 59, no. 2, pp. 341–357, 2013.

[4] D. Bertsimas, M. Sim, and M. Zhang, "Adaptive distributionally robust optimization," *Management Science*, 2018.

[5] J. Blanchet, Y. Kang, and K. Murthy, "Robust Wasserstein profile inference and applications to machine learning," *arXiv preprint arXiv:1610.05627*, 2016.

[6] J. Blanchet and K. Murthy, "Quantifying distributional model risk via optimal transport," *arXiv preprint arXiv:1604.01446*, 2016.

[7] A. Cherukuri and J. Cortés, "Data-driven distributed optimization using Wasserstein ambiguity sets," in *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, 2017, pp. 38–44.

[8] ——, "Cooperative data-driven distributionally robust optimization," *IEEE Transactions on Automatic Control*, 2018, submitted.

[9] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations," *Mathematical Programming*, pp. 1–52, 2017.

[10] 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. 3-4, pp. 707–738, 2015.

[11] R. Gao and A. Kleywegt, "Distributionally robust stochastic optimization with Wasserstein distance," *arXiv preprint arXiv:1604.02199*, 2016.

[12] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Mathematical Programming*, vol. 158, no. 1-2, pp. 291–327, 2016.

[13] H. Khalil, *Nonlinear Systems*. Prentice Hall, 2002.

[14] B. C. Levy and R. Nikoukhah, "Robust state space filtering under incremental model perturbations subject to a relative entropy tolerance," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 682–695, 2013.

[15] D. Li and S. Martínez, "Online data assimilation in distributionally robust optimization," *arXiv preprint arXiv:180307984*, 2018.

[16] S. T. Rachev, L. Klebanov, S. V. Stoyanov, and F. Fabozzi, *The methods of distances in the theory of probability and statistics*. Springer, 2013.

[17] S. Shafieezadeh-Abadeh, V. A. Nguyen, D. Kuhn, and P. M. Esfahani, "Wasserstein distributionally robust Kalman filtering," *arXiv preprint arXiv:1809.08830*, 2018.

[18] A. Shapiro, "Distributionally robust stochastic programming," *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2258–2275, 2017.

[19] A. Stuart and A. R. Humphries, *Dynamical systems and numerical analysis*. Cambridge University Press, 1998, vol. 2.

[20] M. Zorzi, "Robust Kalman filtering under model perturbations," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2902–2907, 2017.