2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Comparing structured ambiguity sets for stochastic optimization: Application to uncertainty quantification

Lotfi M. Chaouach Tom Oomen Dimitris Boskos

Abstract— **The aim of this paper is to compare two classes of structured ambiguity sets, which are data-driven and can reduce the conservativeness of their associated optimization problems. These two classes of structured sets, coined Wasserstein hyperrectangles and multi-transport hyperrectangles, are explored in their trade-offs in terms of reducing conservativeness and providing tractable reformulations. It follows that multi-transport hyperrectangles lead to tractable optimization problems for a significantly broader range of objective functions under a decent compromise in terms of conservativeness reduction. The results are illustrated in an uncertainty quantification case study.**

I. Introduction

Decisions in the face of uncertainty are abundant across engineered systems, whose complexity and interaction with their environment introduce factors that cannot be exactly modeled. Taking further into account the rapidly increasing amount of sensing devices and available data explains the emerging use of stochastic optimization algorithms for system design and operation. The probabilistic models that capture the uncertainty often suffer from imperfections since they are usually inferred from data, which can only provide an approximation of the uncertainty.

Distributionally robust optimization (DRO) addresses the issue of imperfect distributional information by using an ambiguity set of distributions and minimizing the expected cost over the worst-case models from this set. This paradigm has found widespread applications that span across regularization for machine learning [24], portfolio optimization [1], power dispatch [19], and scheduling [17]. In addition, DRO is also employed to solve stochastic control and estimation problems. This includes distributionally robust linear quadratic regulator problems [30], [27], associated model predictive control algorithms [9], [21], [31], and more general distributionally robust dynamic programming formulations [20]. Further work considers also distributionally robust filtering in settings that simultaneously capture noise and model imperfections [32], [25].

The most common ambiguity set constructions group distributions based on moment constraints [11], [23], or their closeness to a nominal model. This closeness can be determined using various discrepancy notions like statistical divergences [6], [16], total variation metrics [26], or optimal transport discrepancies [22] that are typically captured through the Wasserstein distance [28].

All the authors are with the Delft Center for Systems and Control of TU Delft. Tom Oomen is also with the Department of Mechanical Engineering of TU Eindhoven, {L.Chaouach,D.Boskos}@tudelft.nl,T.A.E.Oomen@tue.nl.

Wasserstein ambiguity sets have received significant attention in data-driven problems [12], [18], which is in part also motivated by their statistical guarantees. These sets are balls in the Wasserstein space and contain all the distributions up to a certain distance from an empirical model constructed by the samples. The radius of the ambiguity set can be tuned in terms of the expected Wasserstein distance between the true distribution and the empirical model, which is known to converge to zero as the number of samples grows to infinity [13]. However, this convergence rate is typically exponentially slow with respect to the dimension of the uncertainty vector [7], [13], rendering the radii of such Wasserstein balls impractical from a statistical perspective.

One way to overcome this obstacle is to inform the ambiguity set by the optimization problem [2], [4], [14], [24]. However, there are problems like MPC or stochastic reachability, which entail the solution of multiple optimization problems, where this rationale cannot be directly applied. This motivates alternative approaches to mitigate the curse of dimensionality of Wasserstein balls. To this end, [7] exploits independence assumptions across lower-dimensional components of the random vector to build structured ambiguity sets of product distributions. The approach can alleviate the curse of dimensionality but admits tractable reformulations only for special cases of cost functions. This is resolved in [8], which considers an alternative construction of structured ambiguity sets that group distributions through couplings that respect multiple optimal transport constraints.

Although important developments have been made to construct ambiguity sets, at present there are many alternatives and it is not clear when to optimally use a specific ambiguity set for a specific purpose. The aim of this paper is to compare two recently developed structured ambiguity sets, namely the "Wasserstein hyperrectangles" provided in [7] and the "multi-transport hyperrectangles" introduced in [8]. Our first contribution is to establish that multi-transport hyperrectangles admit tractable DRO reformulations for a much broader class of cost functions under a mild increase of conservativeness when considering problems that can also be solved by Wasserstein hyperrectangles. The second contribution is the reformulation of distributionally robust uncertainty quantification problems into convex optimization problems using multi-transport hyperrectangles. Due to space constraints, the proofs are omitted and will appear elsewhere.

II. Preliminaries and notation

Throughout this paper, we use the following notation. We denote by $\|\cdot\|$ the Euclidean norm. The diameter of

$S \subset \mathbb{R}^d$ is $\mathrm{diam}(S) := \sup\{\|x-y\|_\infty \,|x, y \in S\}$, where $\|\cdot\|_\infty$ is the infinity norm. For $N \in \mathbb{N}\backslash\{0\}$, we denote $[N] := \{1, \ldots, N\}$. We denote by $\mathcal{S}^c$ the complement of a set $S \subset \mathbb{R}^d$. Given $\xi = (\xi_1, \ldots, \xi_d), \zeta = (\zeta_1, \ldots, \zeta_d) \in \mathbb{R}^d$, $\zeta \preceq (\prec)\xi$ if $\zeta_k \le (<)\xi_k$ holds for all $k \in [d]$. Given $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_n}$, we denote $\mathrm{pr}_k^{\boldsymbol{d}}(\xi) := \xi_k, k \in [n]$, where $\boldsymbol{d} := (d_1, \ldots, d_n)$, and omit the superscript $\boldsymbol{d}$ when it is clear form the context. Vectors will be interpreted as column vectors in linear algebra operations unless indicated by a transpose and as vectors of the same type (row or column) when appearing in an inner product.

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 the measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, a measurable map $\Psi : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ assigns to each measure μ in $(\Omega, \mathcal{F})$ the pushforward measure $\Psi_\#\mu$ in $(\Omega', \mathcal{F}')$ defined by $\Psi_\#\mu(B) := \mu(\Psi^{-1}(B))$ for all $B \in \mathcal{F}'$. The Dirac distribution at $\xi \in \mathbb{R}^d$ is denoted by δ_ξ. The indicator function $\mathbb{1}_\Xi$ of $\Xi \subseteq \mathbb{R}^d$ is $\mathbb{1}_\Xi(\xi) = 1$ if $\xi \in \Xi$ and 0 otherwise. Given $p \ge 1$, we denote by $\mathcal{P}_p(\mathbb{R}^d)$ the set of probability measures in $\mathcal{P}(\mathbb{R}^d)$ with finite pth moment. The pth Wasserstein distance of $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ is

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

(cf. [28]). Each $\pi \in \mathcal{M}(P, Q)$ is a transport plan, i.e., a distribution on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mathrm{pr}_{1\#}\pi = P$ and $\mathrm{pr}_{2\#}\pi = Q$. We denote by $P \otimes Q$ the product measure of P and Q.

III. Problem Formulation

We consider data-driven DRO problems of the form

$$\inf_{x \in \mathcal{X}} \sup_{P \in \mathcal{P}^N} \mathbb{E}_P[f(x, \xi)], \tag{1}$$

where f is the objective function, $x \in \mathcal{X}$ is the decision variable, $\xi \in \mathbb{R}^d$ is a random vector, and $\mathcal{P}^N$ is an ambiguity set of probability distributions that is inferred from data. Here, (1) represents a robust formulation of the problem

$$\inf_{x \in \mathcal{X}} \mathbb{E}_{P_\xi}[f(x, \xi)], \tag{2}$$

which one would solve if the distribution P_ξ of ξ were known. To compensate for the lack of knowledge of P_ξ, data-driven DRO formulations exploit i.i.d. samples $\xi^1, \ldots, \xi^N$ of ξ to build the ambiguity set $\mathcal{P}^N$ in a way that it contains P_ξ with high confidence.

A common way to construct data-driven ambiguity sets is to consider all the distributions up to a given distance ε from the empirical distribution $P_\xi^N := \frac{1}{N}\sum_{i=1}^N \delta_{\xi^i}$ using the pth Wasserstein metric. This yields the ball

$$\mathcal{B}_p(P_\xi^N, \varepsilon) := \{P \in \mathcal{P}_p(\mathbb{R}^d) : W_p(P_\xi^N, P) \le \varepsilon\},$$

which is centered at P_ξ^N and has radius ε. Wasserstein ambiguity balls are accompanied by finite-sample guarantees of containing the distribution P_ξ of the data [13] and enable the derivation of tractable reformulations of the DRO problem (1) with $\mathcal{P}^N \equiv \mathcal{B}_p(P_\xi^N, \varepsilon)$ [3], [12], [15].

A. Structured ambiguity set

The motivation to structure a data-driven ambiguity set is to facilitate including distributions that are closer to the true distribution P_ξ. This results in a smaller ambiguity set $\mathcal{P}^N$, which avoids including irrelevant distributions and reduces the gap between the solution of the original stochastic optimization problem (2) and its distributionally robust formulation (1).

To this end, [7] considers the case where the random vector $\xi = (\xi_1, \ldots, \xi_n)$ consists of n lower-dimensional independent components $\xi_k \in \mathbb{R}^{d_k}$, $k \in [n]$. This implies that P_ξ is a product measure, i.e., $P_\xi = P_{\xi_1} \otimes \ldots \otimes P_{\xi_n}$ and the ambiguity set $\mathcal{P}^N$ is built so that it only contains products of lower-dimensional probability distributions. In particular, using N i.i.d. samples of ξ and a vector of radii $\boldsymbol{\varepsilon} := (\varepsilon_1, \ldots, \varepsilon_n)$ (with positive entries), we consider the product empirical distribution

$$\boldsymbol{P}_\xi^N := P_{\xi_1}^N \otimes \ldots \otimes P_{\xi_n}^N, \tag{3}$$

where $P_{\xi_k}^N := \sum_{i=1}^N \delta_{\xi_k^i}$, and define the *Wasserstein hyperrectangle*

$$\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon}) := \{P_1 \otimes \ldots \otimes P_n : P_k \in \mathcal{B}_p(P_{\xi_k}^N, \varepsilon_k) \text{ for all } k \in [n]\}. \tag{4}$$

Essentially, the Wasserstein hyperrectangle is built by taking the products of the distributions across all lower-dimensional ambiguity balls $\mathcal{B}_p(P_{\xi_k}^N, \varepsilon_k)$ that are centered at the marginal empirical distributions $P_{\xi_k}^N$.

For the same type of distributions, [8] considers an alternative construction of structured ambiguity sets. Using again the product empirical distribution $\boldsymbol{P}_\xi^N$ as a reference measure and a vector $\boldsymbol{\varepsilon}$ as above, which represents available transport budgets, we define the *multi-transport hyperrectangle*

$$\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon}) := \big\{\mathrm{pr}_{2\#}\pi : \pi \in \mathcal{P}_p(\Xi \times \Xi),\ \mathrm{pr}_{1\#}\pi = \boldsymbol{P}_\xi^N, \text{ and } \int_{\mathbb{R}^{d_k}} \|\zeta_k - \xi_k\|^p d\pi(\zeta_k, \xi_k) \le \varepsilon_k^p \text{ for all } k \in [n]\big\}. \tag{5}$$

Namely, the multi-transport hyperrectangle consists of all distributions P for which there exists a transport plan π between P and the baseline distribution $\boldsymbol{P}_\xi^N$ that respects the n transport constraints in (5). This set has notable differences when compared to the Wasserstein hyperrectangle (4). First, since $\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ is built from lower dimensional Wasserstein balls, each marginal P_k of a distribution from $\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ is determined through a transport plan π_k that reallocates mass between P_k and its empirical counterpart $P_{\xi_k}^N$. Thus, by construction, the transport plans π_k are completely decoupled while in (5) all lower-dimensional transport constraints should be respected by the same transport plan. This impacts the type of distributions that are found in each set. For instance, $\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ contains only product measures while $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ also contains distributions that are not product measures. On the other hand, one can readily check from the definition of $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ that it is a convex set, which is a nice property that $\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ does not share.

B. Problem formulation

The two structured ambiguity set constructions provided above raise the question of which is more appropriate to use for a specific optimization problem. For instance, Wasserstein hyperrectangles contain only product distributions, which hints that they should be smaller in size, and thus, less conservative. On the other hand, multi-transport hyperrectangles are convex, which implies that they should admit tractable reformulations for a broader class of objective functions. Based on these observations, our goal in this paper is to compare the two structured ambiguity sets in terms of their size and the class of problems they can efficiently solve.

IV. STATISTICAL GUARANTEES FOR STRUCTURED AMBIGUITY SETS

The DRO formulation (1) seeks to provide a robust approximation of the stochastic optimization problem (2). In particular, if the true distribution P_ξ belongs to $\mathcal{P}^N$, the optimal value of (1) will always be an upper bound of the optimal value of (2), excluding overly-optimistic solutions. Therefore, if $\mathcal{P}^N$ in (1) contains P_ξ with a certain confidence, the value of any DRO problem that exploits this ambiguity set will represent an upper bound for the value of its associated stochastic optimization problem over the true distribution with the same confidence. We next review such guarantees for Wasserstein ambiguity balls and elaborate on how they can be improved when using structured ambiguity sets. To simplify the exposition, we will only consider compactly supported distributions.

A. Statistical guarantees for Wasserstein balls

Given N i.i.d. samples from the random vector ξ and a confidence level $1-\beta$, we can always select the radius ε of a Wasserstein ball so that

$$\mathbb{P}(P_\xi \in \mathcal{B}_p(P_\xi^N, \varepsilon)) \geq 1-\beta, \tag{6}$$

see, e.g. [13], [10], [29]. For a fixed confidence level, a radius ε which guarantees that $\mathcal{B}_p(P_\xi^N, \varepsilon)$ contains P_ξ with this confidence decreases with the number of samples. In fact, the radius can usually be determined by a bound of the form $\varepsilon(N,\beta) \leq K/N^{1/\max\{d,2p\}}$, see e.g., [7, Proposition 4.3], where d is the dimension of the vector ξ. Thus, for high-dimensional random vectors, the decrease rate of ε becomes excessively slow. This implies that exploiting further samples cannot guarantee any significant size reduction of the Wasserstein ball. Obtaining the specific constant K in the above bound for any possible values of Wasserstein exponents p and random vector dimensions d is in general hard. Nevertheless, [5, Proposition 24] provides explicit formulas to determine the ambiguity radius when $d \geq 2p+1$, which covers several cases of interest and allows us to compare the sizes of Wasserstein balls and structured hyperrectangles based on the upper bounds that define them.

B. Statistical guarantees for ambiguity hyperrectangles

Here, we provide statistical guarantees for both classes of structured ambiguity sets and elaborate on their relation to Wasserstein balls. The following result from [8] establishes that for the same Wasserstein radii/transport budget vectors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$, Wasserstein hyperrectangles are always contained inside multi-transport hyperrectangles when they are built around the same product empirical distribution.

Proposition 4.1: ***(Ambiguity hyperrectangle containment).*** Consider the Wasserstein hyperrectangle $\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ given by (4) and the multi-transport hyperrectangle $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ in (5). Then $\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon}) \subset \mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$.

The above containment implies that $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ inherits from $\mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ the statistical guarantees of containing the true distribution P_ξ. These guarantees are based on the following independence assumption.

Assumption 4.2: ***(Independence of random vector components).*** The components of $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_n} \equiv \mathbb{R}^d$ are independent random vectors.

Under this assumption, we next provide guarantees about how to tune the transport budgets of a multi-transport hyperrectangle so that it contains the true distribution with prescribed confidence. The result follows directly from Proposition 4.1 and [7, Proposition 4.4].

Corollary 4.3: ***(Multi-transport hyperrectangle probabilistic guarantees).*** Assume that the random vector ξ is supported on the compact set $\Xi \subset \mathbb{R}^d$ with $\rho := \mathrm{diam}(\Xi)$ and satisfies Assumption 4.2 with $d_k \geq 2p+1$ for each $k \in [n]$. For any confidence $1-\beta$, consider the multi-transport hyperrectangle $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ given by (5) with

$$\begin{aligned} \varepsilon_k &:= \rho\varepsilon_\star(\beta_k, p, d_k) N^{-1/d_k} \\ \varepsilon_\star(\beta, p, d) &:= \sqrt{d} 2^{1/2p} (C(d,p) + (\ln \beta^{-1})^{1/2p}) \\ C(d,p) &:= 2^{(d-2)/2p} \Big(\frac{1}{2^{1/2}-1} + \frac{1}{2^{1/2} - 2^{1/2-p}} \Big)^{1/p} \end{aligned}$$

and $\beta_k := \beta^{\frac{d_k}{d}}$ for each $k \in [n]$. Then the hyperrectangle $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ contains P_ξ with confidence $1-\beta$.

The next result from [8] illustrates how multi-transport hyperrectangles can mitigate the curse of dimensionality of monolithic Wasserstein balls.

Proposition 4.4: ***(Multi-transport hyperrectangle size reduction).*** Consider a confidence $1-\beta$ and the multi-transport hyperrectangle $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ with the transport budgets ε_k as given in Corollary 4.3. Then $\mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon}) \subset \mathcal{B}_p(\boldsymbol{P}_\xi^N, \varepsilon)$, with

$$\varepsilon := cn^{1/p+\max\{0,1/2-1/p\}} \rho \varepsilon_\star(\beta, p, d) N^{-1/\max\{d_1,\ldots,d_n\}},$$

$c := (\sqrt{5}+1)/\big(2e^{(\sqrt{5}+1)^2/4}\big) \approx 1.1043$, and $\varepsilon_\star$ as given in Corollary 4.3.

The same inclusion as in Proposition 4.4 is established for Wasserstein hyperrectangles in [7, Proposition 4.4], which for the same confidence level are contained in a Wasserstein ball of the exact same radius ε. Thus, although the multi-transport hyperrectangle always contains the Wasserstein hyperrectangle, it still enjoys the same guarantees in terms of size reduction with respect to the number of samples.

V. Tractable reformulations of DRO problems over structured ambiguity sets

This section provides tractable reformulations of problem (1) when the ambiguity set $\mathcal{P}^N$ is a structured hyperrectangle. The main difficulty in solving (1) relies on the fact that its inner maximization problem is infinite-dimensional. Since the main goal is to reformulate this inner problem, to facilitate notation, we fix the decision variable x and denote $h(\xi) := f(x, \xi)$. Therefore, we seek to reformulate the maximization problem

$$\sup_{P \in \mathcal{P}^N} \mathbb{E}_P[h(\xi)]. \tag{7}$$

Typical duality results and tractable reformulations are available for this inner problem when the ambiguity set is the ball $\mathcal{B}_p(P_\xi^N, \varepsilon)$ [12], [15] or when it is defined through a single transport cost [3].

Dual reformulations of (7) when $\mathcal{P}^N \equiv \mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ are provided in [7, Proposition 5.3]. This result, presented below, requires the objective function to be the sum or product of functions that depend only on one of the lower-dimensional components $\xi_1, \ldots, \xi_n$ of ξ.

Proposition 5.1: ***(DRO dual over Wasserstein hyperrectangles).*** Consider the problem (7) with $\mathcal{P}^N \equiv \mathcal{H}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ and assume that $h(\xi) = \sum_{k=1}^n h_k(\xi_k)$ or $h(\xi) = \prod_{k=1}^n h_k(\xi_k)$ and $h_k(\xi_k) \geq 0$. Then (7) admits the strong dual reformulations

$$\inf_{\boldsymbol{\lambda} \succeq 0} \sum_{k=1}^n \frac{1}{N} \sum_{i=1}^N \sup_{\xi_k \in \mathbb{R}^{d_k}} \{h_k(\xi_k) + \lambda_k(\varepsilon_k^p - \|\xi_k - \xi_k^i\|^p)\}$$

$$\inf_{\boldsymbol{\lambda} \succeq 0} \prod_{k=1}^n \frac{1}{N} \sum_{i=1}^N \sup_{\xi_k \in \mathbb{R}^{d_k}} \{h_k(\xi_k) + \lambda_k(\varepsilon_k^p - \|\xi_k - \xi_k^i\|^p)\}.$$

The limited applicability of Proposition 5.1 is due to the fact that Wasserstein hyperrectangles are non-convex sets, and thus, admit tractable reformulations only in special cases. On the other hand, by leveraging a duality analysis inspired by [3], the following result from [8] establishes that multi-transport hyperrectangles admit dual reformulations for a much broader class of objective functions.

Proposition 5.2: ***(DRO dual over multi-transport hyperrectangles).*** Assume that h is upper semicontinuous and the support of P_ξ is contained in Ξ. Then the problem (7) with $\mathcal{P}^N \equiv \mathcal{T}_p(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ admits the strong dual reformulation

$$\inf_{\boldsymbol{\lambda} \succeq 0} \frac{1}{N^n} \sum_{(i_1, \ldots, i_n) \in [N]^n} \sup_{\xi \in \Xi} \Big\{h(\xi) + \sum_{k=1}^n \lambda_k(\varepsilon_k^p - \|\xi_k^{i_k} - \xi_k\|^p)\Big\}. \tag{8}$$

The following set of virtual samples consists of the elements of the product empirical distribution (3).

Definition 5.3: ***(Set of virtual samples).*** Given N samples $\xi^1, \ldots, \xi^N$ of a random vector $\xi \in \mathbb{R}^d$ satisfying Assumption 4.2, we define the set $\widehat{\Xi} = \{\widehat{\xi}^l\}_{l=1}^{N^n}$ of all the vectors formed by the tuples $(\xi_1^{i_1}, \ldots, \xi_n^{i_n})$ for some indexing $l \in [N^n]$ of the elements $(i_1, \ldots, i_n) \in [N]^n$.

VI. Uncertainty quantification using structured ambiguity sets

In this section, we provide tractable reformulations of uncertainty quantification problems over multi-transport hyperrectangles. Such problems are for instance of great interest in applications where we seek to assess whether a physical or engineered system is safe or not.

We focus on scenarios where the probability distribution of the state is unknown and we can only use samples to bound the probability of being in a desired set with high confidence. To achieve this, we exploit the reformulations from [12, Corollary 5.3], which determine the highest probability that a random vector belongs to a polytope or its complement over all distributions in a Wasserstein ball. To overcome the potential conservativeness of these results for high-dimensional random vectors that satisfy Assumption 4.2, we exploit multi-transport hyperrectangles. Specifically, we solve the problem

$$\sup_{P \in \mathcal{T}_1(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})} P[\xi \in \mathbb{A}], \tag{9}$$

where $\mathbb{A} = \cup_{j=1}^m \mathbb{A}_j$ and $\mathbb{A}_j$ are convex polytopes.

Theorem 6.1: ***(Uncertainty quantification for unions of polytopes).*** Assume that the distribution P_ξ is supported on the polytope $\Xi := \{\xi \in \mathbb{R}^d : C\xi \preceq f\}$ and that $\mathbb{A}_j := \{\xi \in \mathbb{R}^d : A_j \xi \preceq b_j\}$, $j \in [m]$ are closed polytopes that have nonempty intersection with Ξ. Then (9) can be evaluated by solving the convex program

$$\begin{aligned} \inf_{\boldsymbol{\lambda}, s_l, \gamma_{lj}, \theta_{lj}} \ & \langle \boldsymbol{\lambda}, \boldsymbol{\varepsilon} \rangle + \frac{1}{N^n} \sum_{l=1}^{N^n} s_l \\ \text{s.t. } & 1 + \langle \theta_{lj}, b_j - A_j \widehat{\xi}^l \rangle + \langle \gamma_{lj}, f - C\widehat{\xi}^l \rangle \leq s_l \\ & \|\mathrm{pr}_k^{\boldsymbol{d}}[A_j^\top \theta_{lj} + C^\top \gamma_{lj}]\| \leq \lambda_k \\ & \gamma_{lj} \succeq 0, \ \theta_{lj} \succeq 0, \ s_l \geq 0 \\ & j \in [m], \ k \in [n], \ l \in [N^n], \end{aligned}$$

with $\widehat{\xi}^l \in \widehat{\Xi}$ as in Definition 5.3 and $\boldsymbol{d} := (d_1, \ldots, d_n)$.

Alternatively, we may also want to know with high confidence what is the highest probability that ξ lies outside $\mathbb{A}$, which corresponds to solving

$$\sup_{P \in \mathcal{T}_1(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})} P[\xi \notin \mathbb{A}], \tag{10}$$

where $\mathbb{A} = \cup_{j=1}^m \mathbb{A}_j$ and $\mathbb{A}_j$ are open convex polytopes. The next result exploits Theorem 6.1 to evaluate this probability.

Corollary 6.2: ***(Uncertainty quantification for complements of unions of polytopes).*** Assume that the distribution P_ξ is supported on the polytope $\Xi = \{\xi \in \mathbb{R}^d : C\xi \preceq f\}$ and consider the open polytopes $\mathbb{A}_j := \{\xi \in \mathbb{R}^d : \langle a_j^l, \xi \rangle < b_j^l \text{ for all } l \in [\alpha_j]\}$, $j \in [m]$. Then the value of the program

$$\begin{aligned} \inf_{\boldsymbol{\lambda}, s_l, \gamma_{lq}, \theta_{lq}} \ & \langle \boldsymbol{\lambda}, \boldsymbol{\varepsilon} \rangle + \frac{1}{N^n} \sum_{l=1}^{N^n} s_l \\ \text{s.t. } & 1 - \langle \theta_{lq}, b_q - A_q \widehat{\xi}^l \rangle + \langle \gamma_{lq}, f - C\widehat{\xi}^l \rangle \leq s_l \\ & \|\mathrm{pr}_k^{\boldsymbol{d}}(C^\top \gamma_{lq} - A_q^\top \theta_{lq})\| \leq \lambda_k \end{aligned}$$

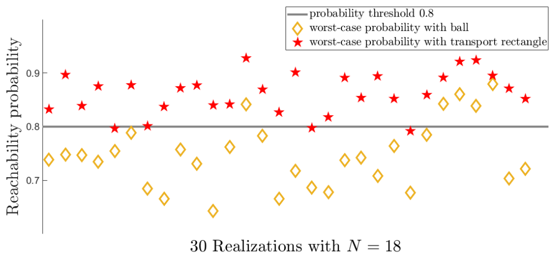

Fig. 1. The figure shows the worst-case probabilities of the event that at least one drone reaches the target across 30 realizations. The results obtained by using the monolithic ball are depicted by the diamonds while those obtained from the multi-transport hyperrectangle are depicted by the red stars. In all cases, the ambiguity sets are built using 18 samples. The results obtained by the hyperrectangle show a clear decrease of conservativeness. Moreover, the corresponding worst-case probabilities are most often above the minimum probability threshold, which happens rarely with the monolithic ball.

$$\gamma_{lq} \succeq 0,\ \theta_{lq} \succeq 0,\ s_l \geq 0$$
$$q \in Q,\ j \in [m],\ k \in [n],\ l \in [N^n],$$

is equal to the probability (10). Here $\boldsymbol{d} = (d_1, \ldots, d_n)$ and for any $q = (q_1, \ldots, q_m) \in \prod_{j=1}^{m} [\alpha_j]$, $A_q \in \mathbb{R}^{m \times d}$ is the matrix formed by concatenating the row vectors $(a_j^{q_j})^\top$, $j \in [m]$ and $b_q := (b_1^{q_1}, \ldots, b_m^{q_m})$. In addition, Q comprises of all indices $q \in \prod_{j=1}^{m} [\alpha_j]$ for which the sets $\{\xi \in \mathbb{R}^d : A_q \xi \succeq b_q\}$ have nonempty intersection with Ξ.

VII. Simulation Example

In this section, we solve an uncertainty quantification problem to illustrate the properties of each ambiguity hyperrectangle. We consider two drones that need to reach a region within a specific deadline to perform a search-and-rescue task. The probability that these drones succeed in reaching the region before the deadline determines whether a fallback plan for the mission has to be used or not.

The maximum velocity and the distance from the region are assumed to be random and independent across the drones. The probability distribution of these variables is unknown and we only consider historical data about them from previous deployments. Our goal is to build an ambiguity set from these data and determine a lower probability bound for two different events. The first is that at least one drone reaches the region before the deadline, whereas the second is that both drones reach it in time. To this end, we want to determine the worst-case probability of each event among all the distributions in the inferred ambiguity set.

Let τ denote the deadline and v_k, r_k denote the maximum velocity and distance of each drone from the target. Each drone k reaches the region iff

$$r_k - \tau v_k < 0 \iff a_k \xi < 0,\ k = 1, 2,$$

where $a_1 = (1, -\tau, 0, 0)$, $a_2 = (0, 0, 1, -\tau)$, and $\xi = (r_1, v_1, r_2, v_2)^\top$ represents the random vector of our problem. Denoting $\mathcal{R}_k$ the event that drone k reaches the target before the deadline, we get $\mathcal{R}_k = \{\xi \in \mathbb{R}^4 : a_k \xi < 0\}$, for $k = 1, 2$. Then the event that at least one drone reaches the region before the deadline is described by the set $\mathscr{E}_1 := \mathcal{R}_1 \cup \mathcal{R}_2$ and we seek to compute the worst-case probability

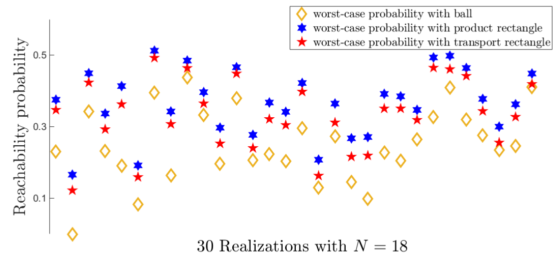

Fig. 2. The figure shows the worst-case probability of the event that both drones reach the target across 30 realizations for all three ambiguity sets. The results obtained by using the monolithic ball, the Wasserstein hyperrectangle, and the multi-transport hyperrectangle are depicted by the diamonds, the blue stars, and the red stars, respectively. In all cases, the ambiguity sets are built using 18 samples. The results obtained by both hyperrectangles outperform those obtained by the monolithic ball.

$$\min_{P \in \mathcal{P}^N} P[\xi \in \mathscr{E}_1] = 1 - \max_{P \in \mathcal{P}^N} P[\xi \notin \mathscr{E}_1], \tag{11}$$

where $\mathcal{P}^N$ is an ambiguity set that we infer from N i.i.d. samples of ξ. Since

$$\max_{P \in \mathcal{P}^N} P[\xi \notin \mathscr{E}_1] = \max_{P \in \mathcal{P}^N} \mathbb{E}[\mathbb{1}_{\mathscr{E}_1^c}(\xi)]$$

and the objective function $\mathbb{1}_{\mathscr{E}_1^c}(\xi)$ does not have the decoupled structure of Proposition 5.1, we cannot evaluate (11) using Wasserstein hyperrectanles. Nevertheless, since $\mathscr{E}_1$ is the union of two open polytopes, we can exploit Corollary 6.2 to evaluate the worst-case probability (11) over a multi-transport hyperrectangle, which we also compare to a Wasserstein ball.

Analogously, the event that both drones reach the region in due time is described by the set $\mathscr{E}_2 := \mathcal{R}_1 \cap \mathcal{R}_2$. Then, we aim to determine the worst-case probability

$$\begin{aligned} \min_{P \in \mathcal{P}^N} P[\xi \in \mathscr{E}_2] &= 1 - \max_{P \in \mathcal{P}^N} P[\xi \notin \mathscr{E}_2], \\ &= 1 - \max_{P \in \mathcal{P}^N} P[\xi \in \mathscr{E}_2^c], \end{aligned} \tag{12}$$

where $\mathscr{E}_2^c = \mathcal{R}_1^c \cup \mathcal{R}_2^c$. The set $\mathscr{E}_2^c$ is the union of two closed polytopes. Thus, we can evaluate (12) with $\mathcal{P}^N \equiv \mathcal{B}_1(P_\xi^N, \varepsilon)$ or $\mathcal{P}^N \equiv \mathcal{T}_1(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$ by exploiting the program of Theorem 6.1. In the case where $\mathcal{P}^N \equiv \mathcal{H}_1(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})$, we get in analogy to the derivation of (17) in [7] that

$$\min_{P \in \mathcal{H}_1(\boldsymbol{P}_\xi^N, \boldsymbol{\varepsilon})} P[\xi \in \mathscr{E}_2] = \prod_{k=1}^{2} \Big(1 - \max_{P_k \in \mathcal{B}_1(P_{\xi_k}^N, \varepsilon_k)} P_k[\mathcal{R}_k^c]\Big).$$

Consequently, this uncertainty quantification problem admits a tractable reformulation that hinges on solving two robust uncertainty quantification problems over Wasserstein balls.

For the simulations, the initial distances (in km) of the drones 1 and 2 follow the distributions $0.5\mathcal{U}[6, 10] + 0.5\mathcal{U}[10.1, 11.1]$ and $0.95\mathcal{U}[9, 10] + 0.05\mathcal{U}[10.1, 11.1]$, respectively, where $\mathcal{U}$ denotes the uniform distribution. All velocities (in m/sec) follow the distribution $\mathcal{U}[50, 50.5]$ and the deadline is set to $\tau = 200$sec. We also assume that the supports of these distributions are known and that this is the only knowledge that we have about the distributions besides the data. Using this information, and to avoid the potential conservativeness of confidence bounds as e.g., in [5, Proposition 24], we tune the radius ε of the monolithic ball and the radii ε_k of the hyperrectangle in relative terms by imposing the requirement $\varepsilon_k \leq c \frac{\rho_k}{\rho} N^{-1/2+1/4} \varepsilon$ (see [7]).

Figure 2 shows the solution of problem (11) across 30 realizations of the simulation that leverage 18 samples each. The multi-transport hyperrectangle exhibits superior performance compared to the Wasserstein ball since the worst-case probabilities are above the probability threshold of avoiding the fallback plan (set at 0.8) in 90% of the realizations in the former case compared to 16.67% in the latter. The true probability of at least one of the drones reaching the target equals 0.975, which implies that worst-case probabilities of the multi-transport hyperrectangle are much closer to the true probability compared to those of the ball. Figure 1 shows the solution of (12) across 30 realizations of the simulation, again using 18 samples each. This figure allows us to compare the worst-case probabilities across all three ambiguity sets. Again, the results from both hyperrectangles outperform those from the ambiguity ball in every realization. In addition, although the multi-transport hyperrectangle is more conservative than the Wasserstein one, it leads to an analogous improvement, providing a convenient tradeoff between tractability and conservativeness-reduction.

VIII. Conclusion

We compared two structured ambiguity sets for data-driven DRO problems. Both sets can be tuned to contain the true distribution with prescribed confidence and favor fast decay rates with the number of samples for high-dimensional uncertainty compared to traditional ambiguity balls. The Wasserstein hyperrectangle is the tightest ambiguity set but has limited practical applicability to DRO problems compared to the multi-transport hyperrectangle, which is slightly more conservative. Using duality of DRO problems over multi-transport hyperrectangles, we developed tractable reformulations of distributionally robust uncertainty quantification problems over unions or intersections of convex polytopes.

References

[1] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," *Mathematical Programming*, vol. 167, pp. 235–292, 2018.

[2] J. Blanchet, Y. Kang, and K. Murthy, "Robust Wasserstein profile inference and applications to machine learning," *Journal of Applied Probability*, vol. 56, no. 3, pp. 830–857, 2019.

[3] J. Blanchet and K. Murthy, "Quantifying distributional model risk via optimal transport," *Mathematics of Operations Research*, vol. 44, no. 2, pp. 565–600, 2019.

[4] J. Blanchet, K. Murthy, and N. Si, "Confidence regions in Wasserstein distributionally robust estimation," *Biometrika*, 2021, DOI: https://doi.org/10.1093/biomet/asab026.

[5] D. Boskos, J. Cortés, and S. Martínez, "High-confidence data-driven ambiguity sets for time-varying linear systems," 2021, arXiv preprint arXiv:2102.01142.

[6] G. C. Calafiore and L. E. Ghaoui, "On distributionally robust chance-constrained linear programs," *Journal of Optimization Theory & Applications*, vol. 130, no. 1, pp. 1–22, 2006.

[7] L. M. Chaouach, D. Boskos, and T. Oomen, "Uncertain uncertainty in data-driven stochastic optimization: towards structured ambiguity sets," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 4776–4781.

[8] L. M. Chaouach, T. Oomen, and D. Boskos, "Structured ambiguity sets for distributionally robust optimization," 2023, technical report, available at: https://sites.google.com/view/dimitris-boskos/publications.

[9] P. Coppens and P. Patrinos, "Data-driven distributionally robust MPC for constrained stochastic systems," *IEEE Control Systems Letters*, vol. 6, pp. 1274–1279, 2022.

[10] J. Dedecker and F. Merlevède, "Behavior of the empirical Wasserstein distance in R^d under moment conditions," *Electronic Journal of Probability*, vol. 24, 2019.

[11] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, p. 595–612, 2010.

[12] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1-2, pp. 115–166, 2018.

[13] 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, p. 707–738, 2015.

[14] R. Gao, "Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality," 2020, arXiv preprint arXiv:2009.04382.

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

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

[17] R. Jiang, M. Ryu, and G. Xu, "Data-driven distributionally robust appointment scheduling over Wasserstein balls," *arXiv preprint arXiv:1907.03219*, 2019.

[18] D. Kuhnl, P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh, "Wasserstein distributionally robust optimization: Theory and applications in machine learning," in *Operations research & management science in the age of analytics*. Informs, 2019, pp. 130–166.

[19] J. Liu, Y. Chen, C. Duan, J. Lin, and J. Lyu, "Distributionally robust optimal reactive power dispatch with Wasserstein distance in active distribution network," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 3, pp. 426–436, 2020.

[20] A. Nilim and L. E. Ghaoui, "Robust control of markov decision processes with uncertain transition matrices," *Operations Research*, vol. 53, no. 5, pp. 780–798, 2005.

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

[22] G. Pflug and D. Wozabal, "Ambiguity in portfolio selection," *Quantitative Finance*, vol. 7, no. 4, pp. 435–442, 2007.

[23] I. Popescu, "Robust mean-covariance solutions for stochastic optimization," *Operations Research*, vol. 55, no. 1, pp. 98–112, 2007.

[24] S. Shafieezadeh-Abadeh, D. Kuhn, and P. M. Esfahani, "Regularization via mass transportation," *Journal of Machine Learning Research*, vol. 20, no. 103, pp. 1–68, 2019.

[25] S. Shafieezadeh-Abadeh, V. A. Nguyen, D. Kuhn, and P. M. Esfahani, "Wasserstein distributionally robust Kalman filtering," in *Advances in Neural Information Processing Systems*, 2018, pp. 8474–8483.

[26] I. Tzortzis, C. D. Charalambous, and T. Charalambous, "Dynamic programming subject to total variation distance ambiguity," *SIAM Journal on Control and Optimization*, vol. 53, no. 4, pp. 2040–2075, 2015.

[27] I. Tzortzis, C. D. Charalambous, and C. N. Hadjicostis, "A distributionally robust LQR for systems with multiple uncertain players," in *IEEE Int. Conf. on Decision and Control*, 2021, pp. 3972–3977.

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

[29] J. Weed and F. Bach, "Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance," *Bernoulli*, vol. 25, no. 4A, pp. 2620–2648, 2019.

[30] I. Yang, "Wasserstein distributionally robust stochastic control: A data-driven approach," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3863–3870, 2021.

[31] Z. Zhong, E. A. del Rio-Chanona, and P. Petsagkourakis, "Data-driven distributionally robust mpc using the wasserstein metric," 2021.

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