Consistency of Distributionally Robust Risk-and Chance-Constrained Optimization under Wasserstein Ambiguity Sets

Cherukuri, Ashish; Hota, Ashish R.

Consistency of Distributionally Robust Risk- and Chance-Constrained Optimization Under Wasserstein Ambiguity Sets

Ashish Cherukuri, *Member, IEEE*, and Ashish R. Hota, *Member, IEEE*

Abstract—**We study stochastic optimization problems with chance and risk constraints, where in the latter, risk is quantified in terms of the conditional value-at-risk (CVaR). We consider the distributionally robust versions of these problems, where the constraints are required to hold for a family of distributions constructed from the observed realizations of the uncertainty via the Wasserstein distance. Our main results establish that if the samples are drawn independently from an underlying distribution and the problems satisfy suitable technical assumptions, then the optimal value and optimizers of the distributionally robust versions of these problems converge to the respective quantities of the original problems, as the sample size increases.**

Index Terms—**Optimization, stochastic systems.**

I. INTRODUCTION

OPTIMIZATION problems under uncertain constraints are pervasive in engineering applications. In the paradigm of *chance-constrained programs* (CCPs), uncertain parameters are treated as random variables and the uncertain constraints are required to be satisfied with a high probability. However, the feasibility set of a CCP is in general non-convex [1]. Furthermore, although the probability of constraint violation is required to be small, the magnitude of constraint violation could potentially be unbounded which is not desirable.

Consequently, recent approaches model uncertain constraints via coherent risk measures that preserve analytical tractability; specifically the conditional value-at-risk (CVaR) [2], [3]. In contrast with chance constraints, (i) CVaR preserves the convexity of the feasibility set, (ii) it requires the magnitude of constraint violation to be bounded in expectation (to be made more precise in Section II-A), and (iii) CVaR constraints provide a convex inner approximation of chance constraints [1]. Accordingly, CVaR-constrained programs (referred to as *risk-constrained programs* (*RCPs*)) have seen widespread applications in financial engineering [4], stochastic optimal control [5]–[7], safety-critical control applications [8], robotics [9] and energy systems [10].

In order to solve stochastic optimization problems in general and CCPs and RCPs in particular, the decision maker needs to know the probability distribution of uncertain parameters. In practice, this information is often unavailable and instead, the decision maker has access to data about the uncertainty in the form of samples. Accordingly, recent work has focused on constructing a family of probability distributions or an *ambiguity set* from the observed samples followed by solving the uncertain optimization problem in a worst-case sense for all distributions in the ambiguity set. This approach is referred to as *distributionally robust* optimization. Within this paradigm, ambiguity sets defined via the Wasserstein distance (see Section II for the definition) have been shown to have desirable out-of-sample performance and analytical tractability [11]–[13]. Motivated by these attractive features, several recent works have proposed approximations and finite-dimensional reformulations of Wasserstein distributionally robust chance and CVaR constrained programs [13]–[16]. This class of problems have also been studied in the context of statistical learning [17], data-driven control [18], [19], and optimal power flow [20], among others.

Note that the Wasserstein ambiguity set is defined directly in terms of the available samples that are drawn from an underlying data-generating distribution. Consequently, the distributionally robust problem instance is a random instance of the original CCP (or RCP) defined in terms of the underlying distribution. Therefore, in addition to analytical tractability and finite sample guarantees, it is desirable to analyze how well the optimal solution of the (random) distributionally robust program approximates the optimal solution of the original CCP (or RCP); particularly in the regime when the number of samples grows to infinity. This property is termed as *asymptotic consistency* in stochastic programming. While *asymptotic consistency* has been established for Wasserstein distributionally robust optimization problems [12], analogous results for chance- and risk-constrained programs have not been explored in the prior work.

In this letter, we show under suitable assumptions that if the samples are being drawn from an underlying distribution $\mathbb{P}$, then the optimal solution and optimizers of the distributionally robust CCP or RCP converge to the corresponding quantities of the CCP or RCP (defined with respect to $\mathbb{P}$), as the number of samples increases and the size of the ambiguity set shrinks. We show that the convergence of the optimal values is from above if the rate at which the ambiguity set shrinks is chosen carefully. Our results provide the much needed asymptotic theoretical justification for Wasserstein distributionally robust constrained optimization programs.

Notation: The sets of real, positive real, non-negative real, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_{>0}$, $\mathbb{R}_{\geq0}$, and $\mathbb{N}$, respectively. The extended reals are $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. For $N \in \mathbb{N}$, we let $[N] := \{1, 2, \ldots, N\}$. For brevity, we denote $\max(x, 0)$ by x_+. The closure of a set $\mathcal{S}$ is denoted by $\mathrm{cl}(\mathcal{S})$. For a set $\mathcal{S}$ and $N \in \mathbb{N}$, we denote the N-fold cartesian product as $\mathcal{S}^N := \Pi_{i=1}^N \mathcal{S}$. Similar notation holds for the N-fold product of any probability distribution.

II. TECHNICAL PRELIMINARIES

Here we formally define the notion of CVaR, Wasserstein distance, and data-driven ambiguity sets.

A. (Conditional) Value-at-Risk

Let Y be a (real-valued) random variable with distribution $\mathbb{P}$. For a tolerance level $\alpha \in (0, 1)$, the value-at-risk (VaR) of Y at level α is

$$\mathrm{VaR}_\alpha^\mathbb{P}(Y) := \inf\{y \in \mathbb{R} \mid \mathbb{P}(Y \leq y) \geq 1 - \alpha\}. \quad (1)$$

That is, it is the $(1 - \alpha)$-quantile of the distribution of Y. The conditional value-at-risk (CVaR) of Y at level α is

$$\mathrm{CVaR}_\alpha^\mathbb{P}(Y) := \inf_{t \in \mathbb{R}} \{\alpha^{-1} \mathbb{E}_\mathbb{P}[(Y + t)_+] - t\}. \quad (2)$$

If Y has a continuous distribution, then $\mathrm{CVaR}_\alpha^\mathbb{P}(Y) = \mathbb{E}_\mathbb{P}[Y \mid Y \geq \mathrm{VaR}_\alpha^\mathbb{P}(Y)]$, i.e., it is the conditional expectation of Y given that Y exceeds $\mathrm{VaR}_\alpha^\mathbb{P}(Y)$.

B. Wasserstein Ambiguity Sets

Assume $\Xi \subseteq \mathbb{R}^m$ and d to be a complete metric on Ξ. Let $\mathcal{B}(\Xi)$ and $\mathcal{P}(\Xi)$ be the Borel σ-algebra and the set of Borel probability measures on Ξ, resp. Let $\mathcal{P}_1(\Xi) \subseteq \mathcal{P}(\Xi)$ be the set of measures with finite first moment. Following [12], the 1-Wasserstein distance between any two measures $\mu, \nu \in \mathcal{P}_1(\Xi)$ is

$$W_1(\mu, \nu) := \min_{\gamma \in \mathcal{H}(\mu,\nu)} \left\{ \int_{\Xi \times \Xi} d(\xi, \omega) \gamma(d\xi, d\omega) \right\}, \quad (3)$$

where $\mathcal{H}(\mu, \nu)$ is the set of all distributions on $\Xi \times \Xi$ with marginals μ and ν. The minimum in (3) is attained because the metric d is continuous [11].

We consider ambiguity sets containing distributions close to the empirical distribution induced by the observed samples. Specifically, let $\widehat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^N \delta_{\widehat{\xi}_i}$ be the empirical distribution constructed from samples $\{\widehat{\xi}_i\}_{i \in [N]}$, where $\delta_{\widehat{\xi}_i}$ is the unit point

mass at $\widehat{\xi}_i$. We define the data-driven Wasserstein ambiguity set as

$$\mathcal{M}_N^\theta := \{\mu \in \mathcal{P}_1(\Xi) \mid W_1(\mu, \widehat{\mathbb{P}}_N) \leq \theta\}, \quad (4)$$

which contains all distributions with finite first moment that are within a distance $\theta \geq 0$ of $\widehat{\mathbb{P}}_N$. In [21], it was shown that $\mathcal{M}_N^\theta$ is a weakly-compact subset of $\mathcal{P}_1(\Xi)$.

III. DISTRIBUTIONALLY ROBUST RISK-CONSTRAINED PROGRAMS AND THEIR CONSISTENCY

In this section, we introduce risk-constrained programs and their distributionally robust counterparts. We consider ambiguity sets defined by the Wasserstein metric and the empirical distribution as discussed above. Our main result establishes that as the number of samples increases, the optimizers and the optimal value of the distributionally robust problems converge, in an appropriate sense, to the corresponding quantities of the original (with respect to the true data-generating distribution) risk-constrained problem. Throughout we consider $\Xi \subset \mathbb{R}^m$ and d to be a complete metric. A canonical CVaR or *risk-constrained program* (RCP) is of the form

$$\begin{aligned} \min_{x \in X} \quad & c^\mathsf{T} x \\ \text{s.t.} \quad & \mathrm{CVaR}_\alpha^\mathbb{P}(F(x, \xi)) \leq 0, \end{aligned} \quad (5)$$

where $X \subseteq \mathbb{R}^n$ is a closed convex set (potentially defined via deterministic constraints), $c \in \mathbb{R}^n$, $\alpha \in (0, 1)$, $\mathbb{P} \in \mathcal{P}(\Xi)$ is the distribution of the uncertain parameter ξ (see Section II-B for notation), and $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$ is called the *constraint function*. Using (2), we can equivalently write the RCP as

$$\begin{aligned} \min_{x \in X, t \in \mathbb{R}} \quad & c^\mathsf{T} x \\ \text{s.t.} \quad & \mathbb{E}_\mathbb{P}[(F(x, \xi) + t)_+] - t\alpha \leq 0. \end{aligned} \quad (6)$$

By equivalent, we mean that x is a feasible point for (5) if and only if there exists t such that (x, t) is feasible for (6).

The distributionally robust version of the RCP (5), which we term as the *distributionally robust risk-constrained program* (DRRCP), is given by

$$\begin{aligned} \min_{x \in X} \quad & c^\mathsf{T} x \\ \text{s.t.} \quad & \sup_{\mathbb{Q} \in \mathcal{M}_N^\theta} \inf_{t \in \mathbb{R}} \mathbb{E}_\mathbb{Q}[(F(x, \xi) + t)_+ - t\alpha] \leq 0, \end{aligned} \quad (7)$$

where $\mathcal{M}_N^\theta$ is the data-driven Wasserstein ambiguity set defined in (4). In other words, we require the CVaR constraint to hold for all distributions that are within a distance $\theta \geq 0$ from the empirical distribution $\widehat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^N \delta_{\widehat{\xi}_i}$ induced by the samples $\{\widehat{\xi}_i\}_{i \in [N]}$, drawn independently from $\mathbb{P}$. This problem is of interest when the decision-maker does not know the distribution $\mathbb{P}$ of the uncertain parameters and instead has access to samples. Thus, the optimal solution of (7) is robust with respect to a family of distributions that are likely to have given rise to the observed samples.

We now present a set of general assumptions.

Assumption 1 (General assumptions on DRRCP): The following hold:
- (i) the function $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$ is continuous,
- (ii) for every $\xi \in \Xi$, $x \mapsto F(x, \xi)$ is convex on X,

(iii) for every $x \in X$, $\xi \mapsto F(x, \xi)$ is bounded on Ξ, and

(iv) F is uniformly Lipschitz over the set X, that is, there exists $L > 0$ such that

$$\left| F(x, \xi) - F(x, \xi') \right| \leq L \| \xi - \xi' \|,$$

for all $\xi, \xi' \in \Xi$ and all $x \in X$.

We first reformulate (7) into a form similar to (6). The below result shows that the inf and the sup operators in the constraint defining DRRCP (7) can be interchanged. The proof is an application of the min-max theorem due to [22] stated as Theorem 3 in the Appendix. The results hold under continuity, convexity, and boundedness conditions in Assumption 1 and the proof is presented in the Appendix.

Lemma 1 (Min-max equality for the constraint function): Suppose Assumption 1 1-1 hold. Then, for every $x \in X$, we have

$$\sup_{\mathbb{Q} \in \mathcal{M}_N^\theta} \inf_{t \in \mathbb{R}} \mathbb{E}_{\mathbb{Q}}[(F(x, \xi) + t)_+ - t\alpha]$$
$$= \inf_{t \in \mathbb{R}} \sup_{\mathbb{Q} \in \mathcal{M}_N^\theta} \mathbb{E}_{\mathbb{Q}}[(F(x, \xi) + t)_+ - t\alpha]. \quad (8)$$

As a consequence of the above result, we can write the DRRCP (7) equivalently as

$$\min_{x \in X, t \in \mathbb{R}} \quad c^\mathsf{T} x$$
$$\text{s.t.} \quad \sup_{\mathbb{Q} \in \mathcal{M}_N^\theta} \mathbb{E}_{\mathbb{Q}}[(F(x, \xi) + t)_+ - t\alpha] \leq 0. \quad (9)$$

That is, x is a feasible point for (7) if and only if there exists t such that (x, t) is feasible for (9). Having reformulated the DRRCP into (9), we move on to the consistency analysis. We require the following assumption throughout.

Assumption 2 (Sequence of finite-sample guarantees): Sequences $\{\beta_N\} \subset (0, 1)$ and $\{\epsilon_N\} \subset (0, \infty)$ are such that $\sum_{N=1}^\infty \beta_N < \infty$, $\lim_{N \to \infty} \epsilon_N = 0$, and the following finite-sample guarantee holds for each $N \in \mathbb{N}$,

$$\mathbb{P}^N(W_1(\mathbb{P}, \widehat{\mathbb{P}}_N) \leq \epsilon_N) \geq 1 - \beta_N. \quad (10)$$

The above assumption imposes that as the number of samples increases, the distance between the data-generating distribution and the empirical distribution becomes vanishingly small with higher confidence. Recent works have indeed established the existence of such sequences [12]. We start our analysis with some preliminary lemmas.

Lemma 2 (Uniform convergence of distributions [12, Lemma 3.7]): Under Assumption 2, we have

$$\mathbb{P}^\infty \left(\lim_{N \to \infty} \sup_{\mathbb{Q} \in \mathcal{M}_N^{\epsilon_N}} W_1(\mathbb{Q}, \mathbb{P}) = 0 \right) = 1.$$

The proof is analogous to the proof of [12, Lemma 3.7] and is omitted in the interest of space. The above result shows that if the Wasserstein radius decreases to zero in a carefully chosen manner, then any sequence of distributions drawn from the ambiguity sets converges to the true distribution.

Remark 1 (Comparison with [12]): Following the above lemma, [12] proves asymptotic consistency of the optimal value and optimizers of distributionally robust expected cost minimization programs under suitable boundedness and continuity assumptions on the cost function. While constrained optimization programs can be written equivalently as expected cost minimization problems via an indicator function on the feasibility set, the consistency results from [12] do not directly apply as the indicator function is not bounded for points that violate the constraints.

We now show that as the number of samples increases, the constraint function of the DRRCP's equivalent form (9) converges uniformly to that of the RCP (6). We first define

$$v(x, t) := \mathbb{E}_{\mathbb{P}}[(F(x, \xi) + t)_+ - t\alpha], \quad (11)$$
$$\widehat{v}_N(x, t) := \sup_{\mathbb{Q} \in \mathcal{M}_N^{\epsilon_N}} \mathbb{E}_{\mathbb{Q}}[(F(x, \xi) + t)_+ - t\alpha], \quad (12)$$

where note that $\widehat{v}_N$ is a random function as the ambiguity set depends on the samples. We now establish uniform $\mathbb{P}^\infty$-almost sure convergence of $\widehat{v}_N$ from above to v. For this, we require the constraint function to be uniformly Lipschitz continuous as stated in Assumption 1.

Lemma 3 (Uniform convergence of $\widehat{v}_N$ from above to v): Let Assumption 1 1, 1 and 1 hold. Further, suppose Assumption 2 holds. Then, the following hold

$$\mathbb{P}^\infty(v(x, t) \leq \widehat{v}_N(x, t) \text{ for all sufficiently large } N) = 1,$$
$$\mathbb{P}^\infty \left(\lim_{N \to \infty} \sup_{x \in X, t \in \mathbb{R}} |\widehat{v}_N(x, t) - v(x, t)| = 0 \right) = 1,$$

where the first equality is satisfied for all $(x, t) \in X \times \mathbb{R}$.

Proof: Fix any $(x, t) \in X \times \mathbb{R}$. From (10), we deduce that the following inequality holds with probability at least $1 - \beta_N$,

$$\mathbb{E}_{\mathbb{P}}[(F(x, \xi) + t)_+ - t\alpha] \leq \sup_{\mathbb{Q} \in \mathcal{M}_N^{\epsilon_N}} \mathbb{E}_{\mathbb{Q}}[(F(x, \xi) + t)_+ - t\alpha].$$

That is, $\mathbb{P}^N(v(x, t) \leq \widehat{v}_N(x, t)) \geq 1 - \beta_N$, for all $N \in \mathbb{N}$. Since $\sum_{N=1}^\infty \beta_N < \infty$, from Borel-Cantelli Lemma [23, Th. 2.3.6], we obtain the first assertion.

From the uniform Lipschitz condition on F stated in Assumption 1 1, we deduce that for any fixed $(x, t) \in X \times \mathbb{R}$ and any $\xi, \xi' \in \Xi$,

$$\left| ((F(x, \xi) + t)_+ - t\alpha) - ((F(x, \xi') + t)_+ - t\alpha) \right|$$
$$= \left| (F(x, \xi) + t)_+ - (F(x, \xi') + t)_+ \right|$$
$$\leq \left| F(x, \xi) - F(x, \xi') \right| \leq L \| \xi - \xi' \|,$$

where the first inequality holds because the operator $(\cdot)_+$ is Lipschitz with constant unity. The above reasoning implies that the map $\xi \mapsto (F(x, \xi) + t)_+ - t\alpha$ is uniformly Lipschitz over the set $X \times \mathbb{R}$. Using this fact in the dual form of the definition of the Wasserstein metric [12], we conclude that

$$\left| \mathbb{E}_{\mathbb{P}_1}[(F(x, \xi) + t)_+ - t\alpha] - \mathbb{E}_{\mathbb{P}_2}[(F(x, \xi) + t)_+ - t\alpha] \right|$$
$$\leq L W_1(\mathbb{P}_1, \mathbb{P}_2), \quad (13)$$

for any two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$. Consider now a sequence of positive real numbers δ_N, $N \in \mathbb{N}$ such that $\lim_{N \to \infty} \delta_N = 0$. For each $(x, t) \in X \times \mathbb{R}$, let $\mathbb{Q}_N^{(x,t)} \in \mathcal{M}_N^{\epsilon_N}$ be a δ_N-optimal distribution such that

$$\mathbb{E}_{\mathbb{Q}_N^{(x,t)}}[(F(x, \xi) + t)_+ - t\alpha] \leq$$
$$\widehat{v}_N(x, t) \leq \mathbb{E}_{\mathbb{Q}_N^{(x,t)}}[(F(x, \xi) + t)_+ - t\alpha] + \delta_N. \quad (14)$$

Existence of such a distribution is due to the fact that expectation is a linear operator. Next, we have

$$
\begin{aligned}
|\widehat{v}_N(x, t) - v(x, t)| &\leq |\mathbb{E}_{\mathbb{Q}_N^{(x,t)}}[(F(x, \xi) + t)_+ - t\alpha] \\
&\quad - \mathbb{E}_{\mathbb{P}}[(F(x, \xi) + t)_+ - t\alpha]| + \delta_N \\
&\leq L W_1(\mathbb{Q}_N^{(x,t)}, \mathbb{P}) + \delta_N \\
&\leq L \sup_{\mathbb{Q} \in \mathcal{M}_N^{\epsilon_N}} W_1(\mathbb{Q}_N, \mathbb{P}) + \delta_N.
\end{aligned}
\tag{15}
$$

The first inequality above uses (14), the second inequality follows from the condition (13), and the last inequality due to the fact that $\mathbb{Q}_N^{(x,t)} \in \mathcal{M}_N^{\epsilon_N}$. Since the right-hand side of (15) is independent of (x, t), we have

$$
\sup_{(x,t) \in X \times \mathbb{R}} |\widehat{v}_N(x, t) - v(x, t)| \leq L \sup_{\mathbb{Q} \in \mathcal{M}_N^{\epsilon_N}} W_1(\mathbb{Q}, \mathbb{P}) + \delta_N.
$$

The proof then concludes by invoking Lemma 2. ∎

We note here that the convergence from above of $\widehat{v}_N$ to v is due to summability of β_N in Assumption 2. If one only needs convergence, then ϵ_N tending to zero is sufficient. We now present our main result. We denote by J^{RCP} the optimal value of (5) and for a given N, we let $\mathsf{J}_N^{\text{DRRCP}}$ and $\{x_N^{\text{DRRCP}}\}_{N \in \mathbb{N}}$ denote the optimal value and an optimizer of (7), resp., where θ is set to ϵ_N satisfying Assumption 2, i.e., ϵ_N is chosen depending on N and β_N satisfying (10).

Theorem 1 [Asymptotic consistency of the DRRCP (7)]: Let Assumptions 1 and 2 hold. Assume that the feasibility set of (5) has a nonempty interior and that the optimizers of (5) belong to a compact set $\mathcal{Y} \subset X$. Moreover, assume that for sufficiently large N and any sequence of i.i.d samples $\{\widehat{\xi}_i\}_{i=1}^N$, optimizers of (7) with θ replaced with ϵ_N belong to $\mathcal{Y}$. Then, the following statements hold $\mathbb{P}^\infty$ - almost surely:

1) $\mathsf{J}^{\text{RCP}} \leq \mathsf{J}_N^{\text{DRRCP}}$ for all sufficiently large N,
2) $\mathsf{J}_N^{\text{DRRCP}} \to \mathsf{J}^{\text{RCP}}$ as $N \to \infty$, and
3) any accumulation point of any sequence of optimizers $\{x_N^{\text{DRRCP}}\}_{N \in \mathbb{N}}$ is an optimal solution of the problem (5).

Proof: The first statement here follows from the first assertion of Lemma 3. For the next two statements, the proof strategy is to show an analogous convergence argument: that the optima and optimizers of (9) approach (6). All convergence statements in this proof involving random quantities hold $\mathbb{P}^\infty$-almost surely and we omit restating this fact for the sake of brevity. Denote the feasibility sets of (6) and (9) as $\mathcal{F}^{\text{RCP}}$ and $\mathcal{F}_N^{\text{DRRCP}}$, respectively. Then, $\mathcal{F}^{\text{RCP}} = \{(x, t) \in X \times \mathbb{R} \mid v(x, t) \leq 0\}$ and $\mathcal{F}_N^{\text{DRRCP}} = \{(x, t) \in X \times \mathbb{R} \mid \widehat{v}_N(x, t) \leq 0\}$. Recall that the set $\mathcal{F}_N^{\text{DRRCP}}$ is random.

Step 1 Defining $\mathcal{W}$: Since F is continuous, $\mathcal{Y}$ is compact, and $F(x, \cdot)$ is bounded over Ξ for every $x \in \mathcal{Y}$, we deduce that the set $\{t \mid \mathbb{E}_{\mathbb{P}}[(F(x, \xi) + t)_+] - t\alpha \leq 0, x \in \mathcal{Y}\}$ is compact. Recall that optimizers of (5) belong to $\mathcal{Y}$. Thus, there exists a compact set $\mathcal{T} \subset \mathbb{R}$ such that optimizers of (6) belong to the set $\mathcal{W} := \mathcal{Y} \times \mathcal{T}$. Similarly, for all sufficiently large N and all sequence of N i.i.d samples, the set of optimizers of (9) (with θ replaced with ϵ_N) belong to the set $\mathcal{W}$. Since the intersection of $\mathcal{Y}$ and the feasibility set of (5) has a nonempty interior, one can assume, without loss of generality, that $\mathcal{W} \cap \mathcal{F}^{\text{RCP}}$ has a nonempty interior.

Step 2 Establishing $\mathcal{F}_N^{\text{DRRCP}} \to \mathcal{F}^{\text{RCP}}$: Following Lemma 3, we know that $\widehat{v}_N$ converges uniformly $\mathbb{P}^\infty$-almost surely to v.

Using this fact, one can establish convergence, defined in an appropriate sense, of $\mathcal{F}_N^{\text{DRRCP}}$ to $\mathcal{F}^{\text{RCP}}$. Specifically, we will show

$$
\lim_{N \to \infty} \sup_{(x,t) \in \mathcal{F}_N^{\text{DRRCP}} \cap \mathcal{W}} \text{dist}((x, t), \mathcal{F}^{\text{RCP}}) = 0,
\tag{16}
$$

where $\text{dist}((x, t), \mathcal{F}^{\text{RCP}})$ is the distance of the point (x, t) to the set $\mathcal{F}^{\text{RCP}}$, that is, $\text{dist}((x, t), \mathcal{F}^{\text{RCP}}) = \inf_{(x',t') \in \mathcal{F}^{\text{RCP}}} \|(x, t) - (x', t')\|$. We proceed with a contradiction argument to show (16). Recall the assertion that (16) holds $\mathbb{P}^\infty$-almost surely. Now, for the sake of contradiction, assume that there exists a set of sequence of i.i.d samples

$$
\mathcal{H} := \left\{ \{\widehat{\xi}_N(\sigma)\}_{N \in \mathbb{N}} \mid \sigma \in \Sigma \right\}
$$

that has finite measure under the distribution $\mathbb{P}^\infty$ and each element of $\mathcal{H}$ violates the limit (16). Here, Σ is some uncountable index set. To be more precise, $\mathcal{H}$ gives rise to a set of sequences $\{\{\mathcal{F}_N^{\text{DRRCP}}(\sigma)\}_{N \in \mathbb{N}} \mid \sigma \in \Sigma\}$ such that each element in this set violates (16). This in turn implies that for each $\sigma \in \Sigma$, one can assign a sequence $\{(x_N(\sigma), t_N(\sigma)) \in \mathcal{F}_N^{\text{DRRCP}}(\sigma) \cap \mathcal{W}\}_{N \in \mathbb{N}}$ and a constant $\gamma_\sigma > 0$ such that

$$
\text{dist}((x_N(\sigma), t_N(\sigma)), \mathcal{F}^{\text{RCP}}) > \gamma_\sigma, \quad \forall N \in \mathbb{N}.
\tag{17}
$$

Since $\mathcal{W}$ is compact, there exists a subsequence of $\{(x_N(\sigma), t_N(\sigma))\}$ that converges to some $(\bar{x}(\sigma), \bar{t}(\sigma)) \in \mathcal{W}$. We denote this subsequence by $\{(x_N(\sigma), t_N(\sigma))\}$ for convenience. Then, due to continuity of v, for any $\epsilon/2 > 0$, there exists $N_1(\sigma) \in \mathbb{N}$ such that

$$
\left| v(x_N(\sigma), t_N(\sigma)) - v(\bar{x}(\sigma), \bar{t}(\sigma)) \right| \leq \epsilon/2
$$

for all $N \geq N_1(\sigma)$. Moreover, by $\mathbb{P}^\infty$-almost sure uniform convergence of $\widehat{v}_N \to v$, for any $\epsilon/2 > 0$, for almost all $\sigma \in \Sigma$, there exists $N_2(\sigma) \in \mathbb{N}$ such that

$$
|\widehat{v}_N(x_N(\sigma), t_N(\sigma)) - v(x_N(\sigma), t_N(\sigma))| \leq \epsilon/2,
$$

for all $N \geq N_2(\sigma)$. Using the above two inequalities, we conclude that for almost all $\sigma \in \Sigma$, for any $\epsilon > 0$, there exists $\bar{N}(\sigma)$ such that

$$
\left| \widehat{v}_N(x_N(\sigma), t_N(\sigma)) - v(\bar{x}(\sigma), \bar{t}(\sigma)) \right| \leq \epsilon, \quad \forall N \geq \bar{N}(\sigma).
$$

This implies that $\lim_{N \to \infty} \widehat{v}_N(x_N(\sigma), t_N(\sigma)) = v(\bar{x}(\sigma), \bar{t}(\sigma))$ for almost all σ. Since $\widehat{v}_N(x_N(\sigma), t_N(\sigma)) \leq 0$ for all N, we get $v(\bar{x}(\sigma), \bar{t}(\sigma)) \leq 0$, that is, $(\bar{x}(\sigma), \bar{t}(\sigma)) \in \mathcal{F}^{\text{RCP}}$ for almost all $\sigma \in \Sigma$. This is in contradiction with (17). Hence, we have established (16).

Step 3 Convergence of optimizers and optimal values: Now let $(x_N, t_N) \in \mathcal{F}_N^{\text{DRRCP*}}$ for all N, where $\mathcal{F}_N^{\text{DRRCP*}}$ is the set of optimal solutions of (9). Since the sequence $\{(x_N, t_N)\}$ is contained in a compact set $\mathcal{W}$, by abuse of notation, we deduce that $(x_N, t_N) \to (\bar{x}, \bar{t})$ for some $(\bar{x}, \bar{t}) \in \mathcal{W}$. Since $\mathcal{F}^{\text{RCP}}$ is closed and (16) holds, we get $(\bar{x}, \bar{t}) \in \mathcal{F}^{\text{RCP}}$. By continuity,

$$
\lim_{N \to \infty} c^\mathsf{T} x_N = c^\mathsf{T} \bar{x} \geq \mathsf{J}^{\text{RCP}},
\tag{18}
$$

where J^{RCP} is the optimum value of (5).

Now, let $(x^*, t^*) \in \mathcal{F}^{\text{RCP*}}$, where $\mathcal{F}^{\text{RCP*}}$ is the set of optimal solutions of (6). Since $\mathcal{F}^{\text{RCP}}$ is convex and its interior is nonempty, there exists a sequence $\{(x_k, t_k)\}_{k \in \mathbb{N}}$ belonging to

the interior of $\mathcal{F}^{\mathrm{RCP}}$ such that $(x_k, t_k) \to (x^*, t^*)$. This implies that for any $\epsilon > 0$, there exists $\bar{k}$ satisfying

$$c^{\mathsf{T}} x_{\bar{k}} - \mathsf{J}^{\mathrm{RCP}} = c^{\mathsf{T}} x_{\bar{k}} - c^{\mathsf{T}} x^* \leq \epsilon. \tag{19}$$

Since $\{(x_k, t_k)\}$ belongs to the interior of $\mathcal{F}^{\mathrm{RCP}}$ and $\widehat{v}_N$ converges to v uniformly over $X \times \mathbb{R}$, we deduce that $(x_{\bar{k}}, t_{\bar{k}}) \in \mathcal{F}_N^{\mathrm{DRRCP}}$ for all sufficiently large N. For such N, optimality of x_N implies that $c^{\mathsf{T}} x_{\bar{k}} \geq c^{\mathsf{T}} x_N$. Using this fact in (19), we get $\mathsf{J}^{\mathrm{RCP}} \geq c^{\mathsf{T}} x_{\bar{k}} - \epsilon \geq c^{\mathsf{T}} x_N - \epsilon$. Taking $N \to \infty$ gives $\mathsf{J}^{\mathrm{RCP}} \geq c^{\mathsf{T}} \bar{x} - \epsilon$. Since ϵ can be chosen arbitrarily small, we obtain $\mathsf{J}^{\mathrm{RCP}} \geq c^{\mathsf{T}} \bar{x}$. Combined with (18), we conclude $c^{\mathsf{T}} \bar{x} = c^{\mathsf{T}} x^*$ and hence $\bar{x} \in \mathcal{F}^{\mathrm{RCP*}}$. Finally, the argument holds for any convergent subsequence of $\{(x_N, t_N)\}$. The convergence of the optimum values then follows by continuity. ∎

The first part of our result, that $\mathsf{J}^{\mathrm{RCP}} \leq \mathsf{J}_N^{\mathrm{DRRCP}}$ for all sufficiently large N, signifies that the solution of the DRRCP is a conservative approximation of the solution of the RCP in the asymptotic regime.

Remark 2 (Discussion on assumptions of Theorem 1): Our assumption on the interior of the feasibility set of (5) being nonempty is a fairly standard assumption in consistency analysis, e.g., [24, Th. 5.5, Prop. 5.30]. This ensures that the sample-based optimization problem (problem stated in (7)) is feasible for large N. A sufficient condition for this assumption to hold is the existence of $x \in X$ such that $F(x, \xi) < 0$ for all $\xi \in \Xi$; this condition can be checked without knowing $\mathbb{P}$ or samples.

Similarly, our assumption on the existence of a compact set $\mathcal{Y} \subset X$ containing the optimizers of (5) and (7) is also a standard one for consistency analysis [24, Th. 5.3, Proposition 5.3], and is required to establish the convergence $\mathcal{F}_N^{\mathrm{DRRCP}} \to \mathcal{F}^{\mathrm{RCP}}$. It is trivially satisfied if X is compact. If X is unbounded, this assumption holds if $x \mapsto \mathbb{E}_{\mathbb{P}}[F(x, \xi)]$ and $x \mapsto \mathbb{E}_{\mathbb{Q}}[F(x, \xi)]$ are coercive for some distribution $\mathbb{Q} \in \mathcal{M}_N^\theta$ (e.g., the empirical distribution).

Next, we analyze the consistency of distributionally robust chance-constrained programs.

IV. DISTRIBUTIONALLY ROBUST CHANCE-CONSTRAINED PROGRAMS AND THEIR CONSISTENCY

Consider the *chance-constrained program* (CCP),

$$\begin{aligned} \min_{x \in X} \quad & c^{\mathsf{T}} x \\ \text{s.t.} \quad & \mathbb{P}((F(x, \xi) \leq 0) \geq 1 - \alpha, \end{aligned} \tag{20}$$

where we borrow the notation from Section III. In comparison with the RCP (5), here, we require the uncertain constraint $F(x, \xi) \leq 0$ to hold with a high probability, i.e., at least $1 - \alpha$. Note that this constraint is equivalent to $\mathrm{VaR}_\alpha^{\mathbb{P}}(F(x, \xi)) \leq 0$ and in general, the set of points satisfying the constraint is non-convex.

The distributionally robust version of the CCP (20), which we term as the *distributionally robust chance-constrained program* (DRCCP), is given as

$$\begin{aligned} \min_{x \in X} \quad & c^{\mathsf{T}} x \\ \text{s.t.} \quad & \inf_{\mathbb{Q} \in \mathcal{M}_N^\theta} \mathbb{Q}((F(x, \xi) \leq 0) \geq 1 - \alpha, \end{aligned} \tag{21}$$

where $\mathcal{M}_N^\theta$ is the ambiguity set defined in (4). We next present the consistency analysis for the DRCCP. As explained before, the chance-constraint can render the feasibility set non-convex. Therefore, consistency requires the following conditions which are different from Assumption 1.

Assumption 3 (Regularity of CCP): The map F is uniformly continuous and either of the following holds:

(i) The distribution $\mathbb{P}$ satisfies

$$\mathbb{P}(\{\xi \in \Xi \mid F(x, \xi) = 0\}) = 0, \quad \text{for all } x \in X.$$

(ii) The set-valued map $H(x) := \{\xi \in \Xi \mid F(x, \xi) \leq 0\}$ is convex-valued and continuous over X (where continuity implies inner and outer semicontinuity of the set-valued map) and for any $x \in X$, $\mathbb{P}(\mathrm{bd}H(x)) = 0$, where $\mathrm{bd}H(x)$ denotes the boundary of the set $H(x)$.

We have the following consistency result. The proof is largely based on results from [25], where the consistency analysis was done for ambiguity sets that are not random. A key step in the proof is to establish almost sure convergence of the feasibility set of the DRCCP to the feasibility set of the CCP which requires the constraint function to be continuous. Assumption 3, inspired by [25], states complementary sufficient conditions which ensure this; [25, Example 4.3] illustrates how Assumption 3 (ii) holds in cases where Assumption 3 (i) fails.[1]

Theorem 2 [Asymptotic consistency of the DRCCP (21)]: Let Assumptions 2 and 3 hold. Assume that there exists a compact set $\mathcal{Y} \subset X$ such that the optimizers of (20) belong to $\mathcal{Y}$. Suppose there exists an optimizer x^* of (20) that belongs to the closure of the set $\{x \in X \mid \mathbb{P}(F(x, \xi) \leq 0) > 1 - \alpha\}$. Moreover, assume that for sufficiently large N and any sequence of i.i.d samples $\{\widehat{\xi}_i\}_{i=1}^N$, optimizers of (21) with θ replaced with ϵ_N belong to $\mathcal{Y}$. Then, the following hold $\mathbb{P}^\infty$-almost surely:

1) $\mathsf{J}^{\mathrm{CCP}} \leq \mathsf{J}_N^{\mathrm{DRCCP}}$ for all sufficiently large N,
2) $\mathsf{J}_N^{\mathrm{DRCCP}} \to \mathsf{J}^{\mathrm{CCP}}$ as $N \to \infty$, and
3) any accumulation point of any sequence of optimizers $\{x_N^{\mathrm{DRCCP}}\}_{N \in \mathbb{N}}$ is an optimizer of the problem (20).

Here, $\mathsf{J}^{\mathrm{CCP}}$ is the optimal value of (20) and for a given N, $\mathsf{J}_N^{\mathrm{DRCCP}}$ and x_N^{DRCCP} are the optimal value and an optimizer of (21), respectively, where θ is set to ϵ_N.

Proof: By assumption, without loss of generality, one can assume that $X = \mathcal{Y}$ is a compact set. Define

$$v^{\mathrm{CCP}}(x) := \mathbb{P}(F(x, \xi) \leq 0), \tag{22}$$

$$\widehat{v}_N^{\mathrm{DRCCP}}(x) := \inf_{\mathbb{Q} \in \mathcal{M}_N^{\epsilon_N}} \mathbb{Q}(F(x, \xi) \leq 0), \tag{23}$$

where $\{\epsilon_N\}_{N \in \mathbb{N}} \subset (0, \infty)$ is any sequence satisfying the hypotheses. Using Assumption 2 and following a similar reasoning as presented in the proof of Lemma 3, we conclude that for any $x \in X$,

$$\mathbb{P}^\infty\left(\widehat{v}_N^{\mathrm{DRCCP}}(x) \leq v^{\mathrm{CCP}}(x) \text{ for all sufficiently large } N\right) = 1.$$

Consequently, $\mathsf{J}^{\mathrm{CCP}} \leq \mathsf{J}_N^{\mathrm{DRCCP}}$ for all sufficiently large N. Regarding the convergence statements, note that from

---

[1] The assumption is satisfied for several classes of functions. For example, if the constraint function has an affine separable form $F(x, \xi) = Ax + B\xi$, B has full column rank, and $\mathbb{P}$ has a continuous distribution, then $\mathbb{P}(F(x, \xi) = 0) = 0$ for any x.

[25, Th. 4.9], Assumption 3 implies continuity of v^{CCP}. Further, from Lemma 2, we deduce that $\mathcal{M}_N^{\epsilon_N}$ converges weakly to $\mathbb{P}$ almost surely. That is, almost surely, any sequence $\{\mathbb{P}_N \in \mathcal{M}_N^{\epsilon_N}\}$ converges weakly to $\mathbb{P}$. Thus, from [25, Props. 5.2, 5.3, Th. 3.2], we obtain $\mathbb{P}^\infty(\lim_{N\to\infty} \sup_{x\in X} |\widehat{v}_N^{\text{DRCCP}}(x) - v^{\text{CCP}}(x)| = 0) = 1$. The proof concludes by applying [25, Th. 3.4]. ∎

V. CONCLUSION

We have studied the asymptotic consistency of data-driven distributionally robust risk- (captured by the CVaR) and chance-constrained optimization under Wasserstein ambiguity sets. As a consequence, under suitable assumptions on the problem data, the distributionally robust versions of the problems can be used as "robust approximators" of the original problems. In future, we plan to analyze the rate of convergence of the consistency arguments. Particularly, we wish to obtain confidence intervals for original optimizers of the problems using the solutions of the distributionally robust counterparts.

APPENDIX

The following result aids us in proving Lemma 1.

Theorem 3 (Stochastic min-max equality [22]): Let $\mathcal{M}$ be a nonempty and weakly compact set of probability measures on $(\Xi, \mathcal{B}(\Xi))$. Consider a function $g : \mathbb{R}^n \times \Xi \to \mathbb{R}$. Let $T \subseteq \mathbb{R}^n$ be a closed convex set. Assume that there exists a convex neighborhood V of T such that for all $t \in V$, the function $g(t, \cdot)$ is measurable, integrable with respect to all $\mathbb{P} \in \mathcal{M}$, and $\sup_{\mathbb{P}\in\mathcal{M}} \mathbb{E}_\mathbb{P}[g(t, \xi)] < \infty$. Further assume that $g(\cdot, \xi)$ is convex on V for all $\xi \in \Xi$. Let $\bar{t} \in \text{argmin}_{t\in T} \sup_{\mathbb{P}\in\mathcal{M}} \mathbb{E}_\mathbb{P}[g(t, \xi)]$. Assume that for every t in a neighborhood of $\bar{t}$, the function $g(t, \cdot)$ is bounded and upper semicontinuous on Ξ and the function $g(\bar{t}, \cdot)$ is bounded and continuous on Ξ. Then,

$$\inf_{t\in T} \sup_{\mathbb{P}\in\mathcal{M}} \mathbb{E}_\mathbb{P}[g(t, \xi)] = \sup_{\mathbb{P}\in\mathcal{M}} \inf_{t\in T} \mathbb{E}_\mathbb{P}[g(t, \xi)].$$

Proof of Lemma 1: We suppress the variable x in the proof for better readability. We verify that the hypotheses of the min-max theorem (Theorem 3) hold.

Drawing the parallelism in notation between our case and Theorem 3, note that here $\mathbb{R}$ plays the role of both T and V; $\mathcal{M}_N^\theta$ that of $\mathcal{M}$; and the function g is $g(t, \xi) := (F(\xi) + t)_+ - t\alpha$. Recall that $\mathcal{M}_N^\theta$ is weakly compact.

Note that g is continuous as F is so. Further since F is bounded, for every $t \in \mathbb{R}$, the function $\xi \mapsto g(t, \xi)$ is bounded and $\sup_{\mathbb{Q}\in\mathcal{M}_N^\theta} \mathbb{E}_\mathbb{Q}[g(t, \xi)] < \infty$. Finally, for every $\xi \in \Xi$, $t \mapsto g(t, \xi)$ is convex. Thus, to conclude the proof it remains to show that the infimum on the right-hand side of (8) is attained. To this end, define the function $h(t) := \sup_{\mathbb{Q}\in\mathcal{M}_N^\theta} \mathbb{E}_\mathbb{Q}[(F(\xi) + t)_+ - t\alpha]$. Now, for any $\mathbb{Q} \in \mathcal{M}_N^\theta$, the function $t \mapsto \mathbb{E}_\mathbb{Q}[(F(\xi) + t)_+ - t\alpha]$ is convex and real-valued. Since h is supremum over a family of such functions, h is convex, lower semicontinuous [26, Proposition 2.1.2]. Further, for any ξ, $(F(\xi) + t)_+ - t\alpha \to \infty$ as $|t| \to \infty$. This fact along with boundedness of F implies $h(t) \to \infty$ as $|t| \to \infty$. Thus, $\inf_{t\in\mathbb{R}} h(t)$ exists. ∎

REFERENCES

[1] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2006.

[2] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, no. 3, pp. 21–41, 2000.

[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Math. Finance*, vol. 9, no. 3, pp. 203–228, 1999.

[4] P. Krokhmal, J. Palmquist, and S. Uryasev, "Portfolio optimization with conditional value-at-risk objective and constraints," *J. Risk*, vol. 4, no. 2, pp. 43–68, 2002.

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

[6] P. Sopasakis, M. Schuurmans, and P. Patrinos, "Risk-averse risk-constrained optimal control," in *Proc. Eur. Control Conf.*, 2019, pp. 375–380.

[7] S. Singh, Y. Chow, A. Majumdar, and M. Pavone, "A framework for time-consistent, risk-sensitive model predictive control: Theory and algorithms," *IEEE Trans. Autom. Control*, vol. 64, no. 7, pp. 2905–2912, Jul. 2019.

[8] S. Samuelson and I. Yang, "Safety-aware optimal control of stochastic systems using conditional value-at-risk," in *Proc. Amer. Control Conf.*, 2018, pp. 6285–6290.

[9] A. Hakobyan, G. C. Kim, and I. Yang, "Risk-aware motion planning and control using CVaR-constrained optimization," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3924–3931, Oct. 2019.

[10] T. Summers, J. Warrington, M. Morari, and J. Lygeros, "Stochastic optimal power flow based on conditional value at risk and distributional robustness," *Int. J. Elect. Power Energy Syst.*, vol. 72, pp. 116–125, Nov. 2015.

[11] R. Gao and A. J. Kleywegt. (2016). *Distributionally Robust Stochastic Optimization With Wasserstein Distance*. [Online]. Available: https://arxiv.org/abs/1604.02199

[12] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Math. Program.*, vol. 171, nos. 1–2, pp. 115–166, 2018.

[13] A. R. Hota, A. Cherukuri, and J. Lygeros, "Data-driven chance constrained optimization under Wasserstein ambiguity sets," in *Proc. Amer. Control Conf.*, Jul. 2019, pp. 1501–1506.

[14] W. Xie, "On distributionally robust chance constrained program with Wasserstein distance," *Math. Program.*, to be published.

[15] Z. Chen, D. Kuhn, and W. Wiesemann. (2018). *Data-Driven Chance Constrained Programs Over Wasserstein Balls*. [Online]. Available: http://www.optimization-online.org/DB_FILE/2018/06/6671.pdf

[16] N. Ho-Nguyen, F. Kılınç-Karzan, S. Küçükyavuz, and D. Lee, "Distributionally robust chance-constrained programs with right-hand side uncertainty under Wasserstein ambiguity," 2020. [Online]. Available: arXiv:2003.12685.

[17] S. Shafieezadeh-Abadeh, D. Kuhn, and P. M. Esfahani, "Regularization via mass transportation," *J. Mach. Learn. Res.*, vol. 20, no. 103, pp. 1–68, 2019.

[18] I. Yang, "A convex optimization approach to distributionally robust markov decision processes with Wasserstein distance," *IEEE Control Syst. Lett.*, vol. 1, no. 1, pp. 164–169, Jul. 2017.

[19] J. Coulson, J. Lygeros, and F. Dörfler, "Distributionally robust chance constrained data-enabled predictive control," 2020. [Online]. Available: arXiv:2006.01702.

[20] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, and A. Cherukuri, "Wasserstein distributionally robust look-ahead economic dispatch," *IEEE Trans. Power Syst.*, early access, Oct. 28, 2020, doi: 10.1109/TPWRS.2020.3034488.

[21] A. Pichler and H. Xu, "Quantitative stability analysis for minimax distributionally robust risk optimization," *Math. Program.*, to be published.

[22] A. Shapiro and A. Kleywegt, "Minimax analysis of stochastic problems," *Optim. Methods Softw.*, vol. 17, no. 3, pp. 523–542, 2002.

[23] R. Durrett, *Probability: Theory and Examples*, 4th ed. New York, NY, USA: Cambridge Univ. Press, 2010.

[24] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA, USA: SIAM, 2009.

[25] S. Guo, H. Xu, and L. Zhang, "Convergence analysis for mathematical programs with distributionally robust chance constraint," *SIAM J. Optim.*, vol. 27, no. 2, pp. 784–816, 2017.

[26] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis* (Grundlehren Text Editions), 2nd ed. Berlin, Germany: Springer, 2004.