University of Groningen

Data-driven chance constrained optimization under wasserstein ambiguity sets

Hota, Ashish R.; Cherukuri, Ashish; Lygeros, John

[Link to publication in University of Groningen/UMCG research database](#)

Data-Driven Chance Constrained Optimization under Wasserstein Ambiguity Sets

Ashish R. Hota, Ashish Cherukuri and John Lygeros

Abstract— We present a data-driven approach for distributionally robust chance constrained optimization problems (DRCCPs). We consider the case where the decision maker has access to a finite number of samples or realizations of the uncertainty. The chance constraint is then required to hold for all distributions that are close to the empirical distribution constructed from the samples (where the distance between two distributions is defined via the Wasserstein metric). We first reformulate DRCCPs under data-driven Wasserstein ambiguity sets and a general class of constraint functions. When the feasibility set of the chance constraint program is replaced by its convex inner approximation, we present a convex reformulation of the program and show its tractability when the constraint function is affine in both the decision variable and the uncertainty. For constraint functions concave in the uncertainty, we show that a cutting-surface algorithm converges to an approximate solution of the convex inner approximation of DRCCPs. Finally, for constraint functions convex in the uncertainty, we compare the feasibility set with other sample-based approaches for chance constrained programs.

I. INTRODUCTION

Numerous engineering applications encounter optimization problems where constraints depend on uncertain parameters, and the goal is to compute a solution that satisfies the constraint with high probability. This class of problems, referred to as *chance constrained programs* (CCPs), are increasingly being relevant in many applications, such as stochastic model predictive control [1], [2], robotics [3], energy systems [4], and autonomous driving [5].

In order to solve a CCP, 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. *Scenario* [6], [7], [8] and *sample approximation* [9] approaches use this data to compute an approximate solution of the CCP. Their main advantage is that if the samples are drawn from a true underlying distribution and the number of samples is sufficiently large, the solutions are feasible for the original CCP with high probability. However, in practice, samples may be few and not be drawn from the true distribution. In such settings, it is desirable to find a solution that satisfies the chance constraint for a suitably defined family of distributions, or a so-called *ambiguity set*. This class of problems is known as *distributionally robust chance constrained programs* (DRCCPs).

Ashish R. Hota is with the Department of Electrical Engineering, Indian Institute of Technology (IIT), Kharagpur. Ashish Cherukuri is with the Engineering and Technology Institute Groningen, University of Groningen. John Lygeros is with the Automatic Control Laboratory, ETH Zürich. Email: ahota@ee.iitkgp.ac.in,a.k.cherukuri@rug.nl,lygeros@control.ee.ethz.ch.

In distributionally robust stochastic optimization (DRSO) in general and DRCCPs in particular, the ambiguity set is defined either as a set of probability distributions that satisfy certain moment constraints [10], [11], [12] or that are close under an appropriate distance function, such as the Prohorov metric [13] or ϕ-divergence [14]. Recent work in DRSO has shown that ambiguity sets based on Wasserstein distance [15] have desirable out-of-sample performance and asymptotic guarantees [16], [17]. DRSO with Wasserstein ambiguity sets were recently applied in optimal power flow problems [4] and uncertain Markov decision processes [18]. Motivated by these attractive features, we consider a data-driven approach for DRCCPs where the ambiguity set is defined as the set of distributions that are close (in the Wasserstein distance) to the empirical distribution induced by the observed samples.

The literature on DRCCPs with Wasserstein ambiguity sets is limited. The authors in [19] first showed that it is strongly NP-Hard to solve a DRCCP with Wasserstein ambiguity sets and proposed a bi-criteria approximation scheme for covering constraints. Two recent working papers presented reformulations and approximations of DRCCPs under Wasserstein ambiguity sets [20], [21] and for constraint functions that are affine in both the decision variable and the uncertainty. Both [20], [21] show that the exact feasibility set of DRCCPs with affine constraints can be reformulated as mixed integer conic programs. Xie [20] studies individual chance constraints and joint chance constraints with right hand side uncertainty, while Chen et. al., [21] consider general affine joint chance constraints. Both papers appeared subsequent to the appearance of the preprint of our work.

Summary of contributions: In this paper, we lay the foundations for tractable computation of (approximate) solutions of DRCCPs under data-driven Wasserstein ambiguity sets for a broader class of constraint functions. We first reformulate DRCCPs under Wasserstein ambiguity sets under general continuity and boundedness assumptions on the constraint functions (as opposed to the affine case studied in [20], [21]). We then focus on developing tractable reformulations and algorithms for DRCCPs. Since the feasibility set of (DR)CCPs is nonconvex except for restrictive special cases [22], we consider constraint functions that are convex in the decision variable, and replace the exact feasibility set of the DRCCP with its convex conditional value-at-risk (CVaR) approximation following [23]. We then present a tractable reformulation of the CVaR approximation when the constraint function is the maximum of functions that are affine in both the decision variable and the uncertainty, and the support of the uncertainty is a polyhedron. When

the constraint function is concave in the uncertainty, we show that a central cutting-surface algorithm [24], [25] can be used to compute an approximately optimal solution of the CVaR approximation of the DRCCP. Finally, when the constraint function is convex in the uncertainty, we compare the feasibility set of the CVaR approximation with those of the sample approximation approach [9] and the scenario approach [6], [7]. We omit the proofs of several of our results due to space constraints, and present them in [26].

Notation: The sets of real, positive real, non-negative real, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0}$, 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 $\mathrm{cl}(\mathcal{S})$. Feasibility sets constructed using data are denoted by $\widehat{\cdot}$. 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 collect preliminary notions and results on CCPs, conditional value-at-risk, and Wasserstein ambiguity sets.

A. Chance Constrained Programs and CVaR Approximation

Throughout we consider Ξ to be a complete separable metric space with metric d. Let $\mathcal{B}(\Xi)$ and $\mathcal{P}(\Xi)$ be the Borel σ-algebra and the set of Borel probability measures on Ξ, respectively. A canonical CCP is of the form

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

where $X \subseteq \mathbb{R}^n$ is a closed convex set, $c \in \mathbb{R}^n$, $\alpha \in (0, 1)$, $\mathbb{P} \in \mathcal{P}(\Xi)$, and $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$. With the exception of a restricted class of distributions and constraint functions, the feasibility set of (1) is nonconvex even when X is convex and F is convex in x for every ξ [22].

Several convex approximations exist that overcome this intractability. We now describe the approximation framework developed in [23] that plays a central role in our results. Consider the function $\psi(z) : \mathbb{R} \to \mathbb{R}$, given as $\psi(z) = \max(z + 1, 0)$. This function belongs to the class of moment generating functions defined in [23]. For a given $\mathbb{P} \in \mathcal{P}(\Xi)$, define $\Psi_{\mathbb{P}} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ as

$$\Psi_{\mathbb{P}}(x, t) := t \mathbb{E}_{\mathbb{P}}[\psi(t^{-1} F(x, \xi))]. \quad (2)$$

Note that if $x \mapsto F(x, \xi)$ is convex for every $\xi \in \Xi$, then $\Psi_{\mathbb{P}}$ is convex in x and t. Furthermore, we have

$$\inf_{t>0}[\Psi_{\mathbb{P}}(x, t) - t\alpha] \leq 0 \implies \mathbb{P}(F(x, \xi) \leq 0) \geq 1 - \alpha. \quad (3)$$

Therefore, replacing the chance constraint by $\inf_{t>0}[\Psi_{\mathbb{P}}(x, t) - t\alpha] \leq 0$ gives a convex conservative approximation of the CCP (1). This approximation is equivalent to replacing the probabilistic constraint with its conditional value-at-risk (CVaR). Formally, the CVaR of a random variable Z with distribution $\mathbb{P}$ at level α is [27]

$$\mathrm{CVaR}_{1-\alpha}^{\mathbb{P}}(Z) := \inf_{t \in \mathbb{R}}[\alpha^{-1} \mathbb{E}_{\mathbb{P}}[(Z + t)_+] - t]. \quad (4)$$

One can show (as done in [23]) that

$$\inf_{t>0}[\Psi_{\mathbb{P}}(x, t) - t\alpha] \leq 0 \iff \mathrm{CVaR}_{1-\alpha}^{\mathbb{P}}(F(x, \xi)) \leq 0. \quad (5)$$

We note that (5) is stronger than simply requiring $F(x, \xi) \leq 0$ with probability at least $1 - \alpha$ as in this case, $F(x, \cdot)$ could take arbitrarily large values for realizations of ξ with measure at most α. In contrast, (5) requires the expected value of $F(x, \cdot)$ for the worst possible realizations of ξ with measure α to be at most zero. We refer the program where the constraint (1) is replaced by (5), as its *CVaR approximation.*

B. Wasserstein ambiguity sets

Let $\mathcal{P}_p(\Xi) \subseteq \mathcal{P}(\Xi)$ be the set of Borel probability measures with finite p-th moment for $p \in [1, \infty)$. Recall that d is the metric on Ξ. Following [15], for $p \in [1, \infty)$, the p-Wasserstein distance between measures $\mu, \nu \in \mathcal{P}_p(\Xi)$ is

$$(W_p(\mu, \nu))^p := \min_{\gamma \in \mathcal{H}(\mu, \nu)} \left\{ \int_{\Xi \times \Xi} d^p(\xi, \omega) \gamma(d\xi, d\omega) \right\}, \quad (6)$$

where $\mathcal{H}(\mu, \nu)$ is the set of all distributions on $\Xi \times \Xi$ with marginals μ and ν. The minimum in (6) is attained because d is lower semicontinuous [16].

In this paper, we define the ambiguity set as the set of all distributions that are 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 the observed samples $\{\widehat{\xi}_i\}_{i \in [N]}$. We define the data-driven Wasserstein ambiguity set as

$$\mathcal{M}_N^\theta := \{\mu \in \mathcal{P}_p(\Xi) | W_p(\mu, \widehat{\mathbb{P}}_N) \leq \theta\}, \quad (7)$$

which contains all distributions that are within a distance $\theta \geq 0$ of $\widehat{\mathbb{P}}_N$. We now present a duality theorem for distributionally robust stochastic optimization over Wasserstein ambiguity sets from [16] that is central to proving our reformulations. Let $H : \Xi \to \overline{\mathbb{R}}$ and consider the following primal and dual problems

$$v_P := \sup_{\mu \in \mathcal{P}_p(\Xi)} \left\{ \int_\Xi H(\xi) \mu(d\xi) \,\Big|\, W_p(\mu, \widehat{\mathbb{P}}_N) \leq \theta \right\}, \quad (8a)$$

$$v_D := \inf_{\lambda \geq 0} \left[\lambda \theta^p + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} [H(\xi) - \lambda d^p(\xi, \widehat{\xi}_i)] \right]. \quad (8b)$$

Theorem II.1. *(Zero-duality gap [16]): Assume that H is upper semicontinuous and either Ξ is bounded, or there exists $\xi_0 \in \Xi$ such that*

$$\limsup_{d(\xi, \xi_0) \to \infty} \frac{H(\xi) - H(\xi_0)}{d^p(\xi, \xi_0)} < \infty.$$

Then, the dual problem (8b) always admits a minimizer λ^* and $v_p = v_D < \infty$.

III. DRCCPs AND EXACT REFORMULATION

In this section, we describe our problem of interest: distributionally robust chance constrained program (DRCCP) with Wasserstein ambiguity sets. Following that, we present two exact reformulations of the DRCCP that have simpler representations. Let $\{\widehat{\xi}_i\}_{i=1}^N$ be a set of N samples of ξ

available to the decision maker. Given this data and $\theta > 0$, the DRCCP for the Wasserstein ambiguity sets (7) is

$$\min\{c^\intercal x : x \in \widehat{X}_{\text{DCP}}\}, \qquad \text{where}$$
$$\widehat{X}_{\text{DCP}} := \left\{ x \in X \;\middle|\; \sup_{\mathbb{P} \in \mathcal{M}_N^\theta} \mathbb{P}(F(x,\xi) > 0) \leq \alpha \right\}. \qquad (9)$$

Note that if $F : \mathbb{R}^n \times \Xi \to \mathbb{R}^K$, then we can instead define F as the component-wise maximum of K constraints. We assume F to be continuous. The probabilistic constraint defining $\widehat{X}_{\text{DCP}}$ can be equivalently written as

$$\sup_{\mathbb{P} \in \mathcal{M}_N^\theta} \mathbb{P}(F(x,\xi) > 0) \leq \alpha \iff \inf_{\mathbb{P} \in \mathcal{M}_N^\theta} \mathbb{P}(F(x,\xi) \leq 0) \geq 1-\alpha.$$

Note that (9) involves optimization over a set of distributions. In order to get a handle on this infinite-dimensional optimization problem, we provide below exact reformulations that involve optimization over finite dimensions. The reformulations presented below were independently shown in [20] for F affine in both x and ξ. Here we note that the results hold more generally.

Theorem III.1. *(Exact reformulations of DRCCP): Let the function $G : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ be given as*

$$G(x,\widehat{\xi}) := \begin{cases} \inf_{\{\xi \mid F(x,\xi)>0\}} d^p(\xi,\widehat{\xi}), & \{\xi \mid F(x,\xi) > 0\} \neq \emptyset, \\ +\infty, & otherwise. \end{cases} \qquad (10)$$

Suppose $\Xi = \mathbb{R}^m$ and there exists $\xi_0 \in \Xi$ such that

$$\limsup_{d(\xi,\xi_0)\to\infty} \frac{F(x,\xi) - F(x,\xi_0)}{d^p(\xi,\xi_0)} < \infty, \quad \forall x \in X. \qquad (11)$$

Then, the feasibility set of the DRCCP (9) satisfies

$$\widehat{X}_{\text{DCP}} = \left\{ x \in X \;\middle|\; \begin{array}{l} \exists \lambda \geq 0, \lambda\theta^p + \frac{1}{N}\sum_{i=1}^N s_i \leq \alpha, \\ s_i = \max\{1 - \lambda G(x,\widehat{\xi}_i), 0\} \end{array} \right\}. \qquad (12)$$

In addition, if $\{\xi \mid F(x,\xi) > 0\}$ is nonempty for every $x \in X$, then

$$\widehat{X}_{\text{DCP}} = \left\{ x \in X \;\middle|\; \frac{\theta^p}{\alpha} + \text{CVaR}_{1-\alpha}^{\widehat{\mathbb{P}}_N}(-G(x,\xi)) \leq 0 \right\}. \qquad (13)$$

The proof is presented in the preprint [26]. The condition (11) is met if F is bounded or $\xi \mapsto F(x,\xi)$ is Lipschitz for every $x \in X$ with $p = 1$. In [19], authors show that DRCCPs under Wasserstein ambiguity sets (9) are strongly NP-Hard even for affine F. In light of this fact, we now focus on developing tractable approximations of DRCCPs using CVaR of the constraint function.

IV. CVaR Approximation of DRCCPs

When F is convex in x, the CVaR approach of [23] provides a convex inner approximation of the feasibility set of the original (DR)CCP (see Section II-A for details). In the remainder of the paper, we study this CVaR approximation of the DRCCP (9) under the following assumptions.

Assumption IV.1. *(F is convex-bounded): The set Ξ is a subset of $\mathbb{R}^m$. The function $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$ satisfies:*

(i) *for every $\xi \in \Xi$, $x \mapsto F(x,\xi)$ is convex on X,*
(ii) *for every $x \in X$, $\xi \mapsto F(x,\xi)$ is bounded on Ξ.*

Note that the second property in the above assumption implies (11). Following our earlier discussion in Section II-A, the CVaR approximation of the DRCCP (9) is

$$\min\{c^\intercal x : x \in \widehat{X}_{\text{CDCP}}\}, \qquad \text{where}$$
$$\widehat{X}_{\text{CDCP}} := \left\{ x \in X \;\middle|\; \sup_{\mathbb{P} \in \mathcal{M}_N^\theta} \inf_{t \in \mathbb{R}} [\mathbb{E}_{\mathbb{P}}[(F(x,\xi) + t)_+] - t\alpha] \leq 0 \right\}. \qquad (14)$$

We start by reformulating the expression of $\widehat{X}_{\text{CDCP}}$ and establishing its convexity.

Proposition IV.2. *(Convex reformulation of (14)): Under Assumption IV.1, the CVaR approximation of the DRCCP problem (14) is equivalent to the following convex program*

$$\min c^\intercal x$$
$$\text{s.t. } \lambda\theta^p + \frac{1}{N}\sum_{i=1}^N s_i \leq t\alpha, \qquad (15)$$
$$s_i \geq \sup_{\xi \in \Xi}[F(x,\xi) + t - \lambda d^p(\xi,\widehat{\xi}_i)], \forall i \in [N],$$
$$\lambda \geq 0, t \in \mathbb{R}, x \in X, s_i \geq 0, \forall i \in [N].$$

Specifically, x lies in the feasibility set of (14) if and only if there exists $(\lambda, t, \{s_i\}_{i=1}^N)$ such that $(x, \lambda, t, \{s_i\}_{i=1}^N)$ is a feasible point for (15).

In the interest of space, the proof is omitted, and can be found in [26]. The above result shows that the CVaR approximation of DRCCPs under Wasserstein ambiguity sets can be reformulated as a convex optimization problem. However, the constraints involving s_i in (15) involve supremum operators. In the remainder of the paper, we develop tractable reformulations and algorithms to solve (15) under suitable assumptions on the constraint function F.

V. Reformulations and Algorithms for Several Classes of Constraint Functions

A. F Piecewise Affine in Uncertainty

We now present a tractable reformulation (15) when F is the maximum of a set of functions that are affine in ξ. The analysis is inspired by a similar reformulation in [17] shown for distributionally robust stochastic optimization. The proof can be found in [26].

Proposition V.1. *(Reformulation of DRCCP for piecewise affine F): Let $\Xi = \{\xi \in \mathbb{R}^m \mid C\xi \leq h\}$ be compact, where $C \in \mathbb{R}^{p\times m}$ and $h \in \mathbb{R}^p$ for some $p > 0$. Suppose that for some positive integer K, $F(x,\xi) := \max_{k \leq K} x^\intercal A_k \xi + b_k(x)$, where $A_k \in \mathbb{R}^{n\times m}$ and $b_k : \mathbb{R}^n \to \mathbb{R}$ are convex functions for all $k \in [K]$. Let the ambiguity set $\mathcal{M}_N^\theta$ be defined using the 1-Wasserstein metric and d be the standard Euclidean distance. Then, the DRCCP (15) is equivalent to*

the following tractable convex optimization problem

$$\min \quad c^{\mathsf{T}}x$$

$$\text{s.t.} \; \lambda\theta + \frac{1}{N}\sum_{i=1}^{N} s_i \le t\alpha,$$

$$\left(b_k(x) + t + (x^{\mathsf{T}}A_k - C^{\mathsf{T}}\eta_{ik})^{\mathsf{T}}\widehat{\xi}_i + \eta_{ik}^{\mathsf{T}}h\right)_+ \le s_i,$$

$$\|x^{\mathsf{T}}A_k - C^{\mathsf{T}}\eta_{ik}\| \le \lambda, \eta_{ik} \ge 0,$$

$$x \in X, t \in \mathbb{R}, \lambda \ge 0,$$

where the inequality involving the set of variables η_{ik} hold for $i \in [N]$ and $k \in [K]$.

Remark V.2. *(Comparison with literature and exactness of CVaR approximation):* In [20], [21], authors derive the reformulation given in Proposition V.1 for the case when $\Xi = \mathbb{R}^m$. In addition, they show that when $\Xi = \mathbb{R}^m$ and $N\alpha \le 1$, the CVaR approximation is exact, i.e., $\widehat{X}_{\text{DCP}} = \widehat{X}_{\text{CDCP}}$. •

In the following subsection, we present an algorithm that solves the CVaR approximation of DRCCPs when the constraint function is concave in uncertainty.

B. F Concave in Uncertainty

Here we aim to develop an algorithm for (15) when F is concave in ξ. The roadblock in solving (15) is the supremum operator present in the constraint that makes implementing first- or second-order methods almost impossible. To construct the algorithm, we view (15) as a semi-infinite program and employ the central cutting surface algorithm proposed in [24]. The algorithm requires the feasibility set of the problem to be compact. Thus, as a first step, we identify a compact set which contains the optimizers of (15). Our results hold under the following assumption.

Assumption V.3. *(F concave in uncertainty and existence of robustly feasible point): The sets X and Ξ are compact. For every $x \in X$, the function $\xi \mapsto F(x,\xi)$ is concave. There exists $\bar{x} \in X$ such that $F(\bar{x},\xi) \le -\delta < 0$ for all $\xi \in \Xi$.*

The next result provides bounds on the optimizers of (15).

Lemma V.4. *(Optimizers of (15) belong to a compact set): Under Assumption IV.1 and V.3, the optimizers of (15) belong to the set $X \times [0,t^M] \times [0,\lambda^M] \times [0,\alpha N t^M]^N$, where*

$$t^M := \frac{1}{1-\alpha} \sup_{x\in X, \xi\in\Xi} -F(x,\xi), \quad \text{and} \quad \lambda^M = \frac{\alpha t^M}{\theta^p}.$$

The proof is relegated to [26] in the interest of space. Using the above result, one can restrict the feasibility set of (15) without disturbing its optimizers. We denote the decision variables of (15) as $y := (x,t,\lambda,\{s_i\}_{i=1}^N)$, and its feasibility set as the compact set $Y := X \times [0,t^M] \times [0,\lambda^M] \times [0,\alpha N t^M]^N$. The optimization problem (15) over

the restricted domain written as semi-infinite program is

$$\min \; c^{\mathsf{T}}x$$

$$\text{s.t.} \; \lambda\theta^p + \frac{1}{N}\sum_{i=1}^{N} s_i \le t\alpha, \tag{16}$$

$$s_i \ge F(x,\xi) + t - \lambda d^p(\xi,\widehat{\xi}_i), \forall \xi \in \Xi, \forall i \in [N],$$

$$(x,t,\lambda,\{s_i\}_{i=1}^N) \in Y.$$

Now, for each $i \in [N]$, we define the function

$$H_i(y,\xi) := F(x,\xi) + t - \lambda d^p(\xi,\widehat{\xi}_i) - s_i.$$

Next, set the parameter $B > 0$ satisfying

$$B > \|g^i(y,\xi)\|, \forall y \in Y, \forall \xi \in \Xi, \forall i \in [N]$$

where $g^i(y,\xi) = (g_y^i(y,\xi), g_\xi^i(y,\xi)) \in \partial_y H_i(y,\xi) \times \partial_\xi H_i(y,\xi)$. That is, B bounds the set of subgradients of H_i, for all i, over the feasibility set Y. Semi-infinite optimization problems are difficult to solve in general. Thus, our objective is to design an algorithm that can find an approximate solution to the problem (16). This is made precise below.

Definition V.5. *(Approximate feasibility and optimality of (16)): We say that a point $y = (x,t,\lambda,\{s_i\}_{i=1}^N) \in Y$ is η-feasible for the problem (16) if it satisfies*

$$\lambda\theta^p + \frac{1}{N}\sum_{i=1}^N s_i \le t\alpha,$$

$$s_i + \eta \ge F(x,\xi) + t - \lambda d^p(\xi,\widehat{\xi}_i), \forall \xi \in \Xi, \forall i \in [N],$$

Further, a point $(x_\eta^\star, t_\eta^\star, \lambda_\eta^\star, \{s_{i,\eta}^\star\}_{i=1}^N)$ is an η-optimal solution of (16) if it is η-feasible and $c^{\mathsf{T}}x_\eta^\star \le c^{\mathsf{T}}x^\star$ where $(x^\star, t^\star, \lambda^\star, \{s_i^\star\}_{i=1}^N)$ is an optimizer of (16).

We propose an algorithm that finds an η-optimal solution of (16). Our scheme involves solving a convex optimization problem, termed the master problem, at every iteration of the algorithm. The master problem for the kth iteration is

$$\max \; \sigma$$

$$\text{s.t.} \; c^{\mathsf{T}}x + \sigma \le M^{(k-1)},$$

$$\lambda\theta^p + \frac{1}{N}\sum_{i=1}^N s_i \le t\alpha, \tag{17}$$

$$H_i(y,\xi_i) + \sigma B \le 0, \forall \xi_i \in Q_i^{(k-1)},$$

$$(x,t,\lambda,\{s_i\}_{i=1}^N) \in Y.$$

Various terms of the above optimization are introduced below where we elaborate on the steps of Algorithm 1.

Each iteration k starts by solving (17). The aim of this step is to find $y^{(k)}$ that is *robustly* feasible to the constraints sampled till the kth iteration, $Q_i^{(k-1)}$, $i \in [N]$, and that also improves the upper bound on the objective value $M^{(k-1)}$. The variable $\sigma^{(k)}$ denotes this improvement. Upon solving (17), two cases arise. First, $y^{(k)}$ is η-feasible and so, there does not exist, for any i, a violating constraint $\xi_i^{(k)}$ that can be added to $Q_i^{(k-1)}$. In this case, we move to Step 5 where the constraint set is kept same, the best estimate of the optimizer $\tilde{y}^{(k-1)}$ is updated to the η-feasible solution found in this iteration, and the upper bound is updated. In the

---

Algorithm 1: A central cutting-surface algorithm for (16)

Input: Assumption V.3 holds. For a given y and $i \in [N]$, whenever $\sup_{\xi \in \Xi} H_i(y, \xi) > \eta$, then there exists an oracle that determines a point $\xi \in \Xi$ such that $H_i(y, \xi) > 0$.

Initialize: Set $k = 1$, $M^{(0)} = U := \max_{x \in X} c^\mathsf{T} x$, $Q_i^{(0)} = \emptyset$ for all $i \in [N]$, $\tilde{y}^{(0)} = 0$.

1 Determine the optimizer $(y^{(k)}, \sigma^{(k)})$ of the master problem (17)

2 If $\sigma^{(k)} = 0$, stop and return $\tilde{y}^{(k-1)}$

3 For each $i \in [N]$, find (if possible) $\xi_i^{(k)} \in \Xi$ such that $H_i(y^{(k)}, \xi_i^{(k)}) > 0$ and then go to Step 4; if no such point exists for any i, then go to Step 5

4 Set for each $i \in [N]$, $Q_i^{(k)} = Q_i^{(k-1)} \cup \{\xi_i^{(k)}\}$ whenever a point $\xi_i^{(k)}$ is found in Step 3, otherwise $Q_i^{(k)} = Q_i^{(k-1)}$; Set $\tilde{y}^{(k)} = \tilde{y}^{(k-1)}$ and $M^{(k)} = M^{(k-1)}$; Go to Step 6

5 Set $Q^{(k)} = Q^{(k-1)}$, $\tilde{y}^{(k)} = y^{(k)}$, and $M^{(k)} = c^\top x^{(k)}$

6 Increase k by one and go to Step 1

---

second case, a violating constraint is determined for each i (if possible) in Step 3. Subsequently, in Step 4, the constraint set is updated while the best estimate of the optimizer and the upper bound are kept the same. The algorithm converges when the objective value cannot be improved anymore over the set of all η-feasible solutions.

The next result states the correctness of Algorithm 1. The proof involves arguments similar in reasoning to those presented in [24]. An important ingredient is the compactness of the feasibility set which we achieved due to Lemma V.4.

Proposition V.6. *(Convergence guarantee of Algorithm 1): Let Assumptions IV.1 and V.3 hold. Consider the iterates $(\tilde{y}^{(k)})_{k=1}^\infty$ generated by Algorithm 1.*

(i) *If Algorithm 1 terminates in the kth iteration, then $\tilde{y}^{(k-1)}$ is an η-optimal solution to (16).*

(ii) *If Algorithm 1 does not terminate, then there exists an index $\hat{k}$ such that the sequence $\{\tilde{y}^{(\hat{k}+i)}\}_{i=1}^\infty$ consists entirely of η-feasible solutions of (16).*

(iii) *If Algorithm 1 does not terminate, then the sequence $\{\tilde{y}^{(k)}\}_{k=1}^\infty$ has an accumulation point, and each accumulation point is an η-optimal solution to (16).*

C. F Convex in Uncertainty

We now consider F to be convex in ξ. For this class of functions, unlike the case dealt in the previous section, the supremum present in the definition of the constraint set of (15) is nonconvex, as it involves maximizing a difference of convex functions. In this section, we provide a convex inner approximation of (15) which is computable using standard convex optimization tools. We then compare the feasibility set of this convex inner approximation with two other feasibility sets obtained from sample based approaches for CCPs. We consider $\Xi \subseteq \mathbb{R}^m$ and the 1-Wasserstein

distance in this section, i.e., $p = 1$. All proofs in this section are omitted in the interest of space, and can be found in [26]. The results rely on the following assumption.

Assumption V.7. *(F Lipschitz in uncertainty): For every $x \in X$, the function $\xi \mapsto F(x, \xi)$ is convex. Moreover, there exists a convex function $L_F : X \to \mathbb{R}_{>0}$, such that $\xi \mapsto F(x, \xi)$ is Lipschitz continuous with constant $L_F(x)$.*

Under the above assumption, we derive the following inner approximation of the feasibility set of the CVaR approximation of DRCCP $\widehat{X}_{\mathrm{CDCP}}$ given by (14).

Lemma V.8. *(Inner approximation of $\widehat{X}_{\mathrm{CDCP}}$): Let Assumptions IV.1 and V.7 hold. Define*

$$\widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}} := \Big\{ x \in X \, \Big| \, \theta L_F(x) + \inf_{t \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^{N} (F(x, \widehat{\xi}_i) + t)_+ - t\alpha \leq 0 \Big\}.$$

Then, $\widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}} \subseteq \widehat{X}_{\mathrm{CDCP}}$ and these sets are equal when $\Xi = \mathbb{R}^m$.

Observe that above, we upper bound the supremum over the Wasserstein ambiguity set in (14) with the sample average and a regularizer term. The proof is a consequence of [17, Theorem 6.3, Proposition 6.5]. The Lipschitz continuity of $\xi \mapsto F(x, \xi)$ is a sufficient condition for [17, Theorem 6.3], and thus, Lemma V.8 may indeed hold for a more general class of functions.

Due to Lemma V.8, instead of minimizing the objective over $\widehat{X}_{\mathrm{CDCP}}$, one could perform the minimization over $\widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}}$. The later problem is easier to deal with and the obtained solution will be feasible with respect to $\widehat{X}_{\mathrm{CDCP}}$ and hence $\widehat{X}_{\mathrm{DCP}}$. Consequently, the optimal value will provide an upper bound on the cost of (15). We now compare the set $\widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}}$ with the feasibility sets of the sample approximation approach [9], and the scenario approach [6]. Given $\delta \in [0, 1]$ and samples $\{\widehat{\xi}_i\}_{i=1}^N$, the sample approximation feasibility set is

$$\widehat{X}_{\mathrm{SA}, \delta} := \Big\{ x \in X \, \Big| \, \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{F(x, \widehat{\xi}_i) \leq 0\}} \geq 1 - \delta \Big\}. \quad (18)$$

Specifically, if $x \in \widehat{X}_{\mathrm{SA}, \delta}$, then at most δ fraction of samples $\{\widehat{\xi}_i\}$ violate the constraint $F(x, \xi) \leq 0$. Similarly, given $\delta \geq 0$ and samples $\{\widehat{\xi}_i\}_{i=1}^N$, we define

$$\widehat{X}_{\mathrm{SCP}, \delta} := \Big\{ x \in X \, \Big| \, F(x, \widehat{\xi}_i) + \delta \leq 0, i \in N \Big\}. \quad (19)$$

Note that the feasibility set of the scenario program is $\widehat{X}_{\mathrm{SCP}, 0}$. Thus, $\widehat{X}_{\mathrm{SCP}, \delta}$ defines a "robust" scenario program, and for any $\delta > 0$, $\widehat{X}_{\mathrm{SCP}, \delta} \subseteq \widehat{X}_{\mathrm{SCP}, 0}$. Also note that $\widehat{X}_{\mathrm{SCP}, 0} = \widehat{X}_{\mathrm{SA}, 0}$. The main result of this subsection is stated below.

Proposition V.9. *(Subset of $\widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}}$): Let Assumptions IV.1 and V.7 hold. Assume L_F is constant over X. Let $t^* := \sup_{x \in X, \xi \in \Xi} - F(x, \xi)$, $\delta_1 := \alpha - \frac{\theta L_F}{t^*}$, and $\delta_2 := \frac{\theta L_F}{\alpha}$. Then, $\widehat{X}_{\mathrm{SCP}, \delta_2} \subseteq \widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}} \subseteq \widehat{X}_{\mathrm{SA}, \delta_1}$.*

The above result shows that the feasibility set of the robust scenario program (19) is contained in the set $\widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}}$. Furthermore, by the definition of the sample approximation set (18),

the above result implies that if $\bar{x} \in \widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}}$, then at most $\delta_1 < \alpha$ fraction of samples violate the constraint $F(x, \xi) \le 0$. Both δ_1 and δ_2 depend on the Lipschitz constant, the probability of constraint violation α, the Wasserstein radius, and δ_1 depends additionally on t^*.

Independent of our work, [20] showed the above relationships between the feasibility sets $\widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}}$, $\widehat{X}_{\mathrm{SA},\alpha}$ and $\widehat{X}_{\mathrm{SCP},\delta}$ when the constraint function is affine in x and ξ. We show that the above comparison holds more generally when the constraint function is convex in both x and ξ.

We conclude with the following comparison between different feasibility sets studied in this paper. For $\delta_1 = \alpha - \frac{\theta L_F}{t^*}$ and $\delta_2 = \frac{\theta L_F}{\alpha}$, we have

$$
\begin{array}{ccc}
\widehat{X}_{\mathrm{SCP},0} & = & \widehat{X}_{\mathrm{SA},0} \\
\cup| & & |\cap \\
\widehat{X}_{\mathrm{SCP},\delta_2} \subseteq & \widehat{X}_{\mathrm{CDCP}}^{\mathrm{in}} \subseteq & \widehat{X}_{\mathrm{SA},\delta_1} \\
& |\cap & \\
& \widehat{X}_{\mathrm{CDCP}} \subseteq & \widehat{X}_{\mathrm{DCP}}.
\end{array}
$$

Note that $\widehat{X}_{\mathrm{SA},\delta_1}$ and $\widehat{X}_{\mathrm{SCP},0}$ are in general incomparable with $\widehat{X}_{\mathrm{CDCP}}$. Thus, the objective values obtained by optimizing over these sets are not necessarily upper or lower bounds on the optimal solution of (15).

VI. CONCLUSION

We studied distributionally robust chance constrained optimization under Wasserstein ambiguity sets defined as the set of all distributions that are close to the empirical distribution. We presented a convex reformulation of the program when the original chance constraint is replaced by its convex CVaR counterpart. We then showed the tractability of this convex reformulation for affine constraint functions. Furthermore, for constraint functions concave in the uncertainty, we presented a cutting-surface algorithm that converges to an approximately optimal solution of the CVaR approximation of the DRCCP. Finally, for constraint functions convex in the uncertainty, we compared the feasibility sets of DRCCP and its approximations with those of the scenario and sample approximation approaches. A rigorous comparison of DRCCPs and the scenario approach vis-a-vis finite sample guarantees and asymptotic convergence of optimal solutions remain as challenging open problems.

REFERENCES

[1] M. Farina, L. Giulioni, and R. Scattolini, "Stochastic linear model predictive control with chance constraints–a review," *Journal of Process Control*, vol. 44, pp. 53–67, 2016.

[2] G. Schildbach, L. Fagiano, C. Frei, and M. Morari, "The scenario approach for stochastic model predictive control with bounds on closed-loop constraint violations," *Automatica*, vol. 50, no. 12, pp. 3009–3018, 2014.

[3] L. Blackmore, M. Ono, and B. C. Williams, "Chance-constrained optimal path planning with obstacles," *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1080–1094, 2011.

[4] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. H. Summers, "Data-based distributionally robust stochastic optimal power flow, part I: Methodologies," *arXiv preprint arXiv:1804.06388*, 2018.

[5] A. Carvalho, S. Lefévre, G. Schildbach, J. Kong, and F. Borrelli, "Automated driving: The role of forecasts and uncertainty - A control perspective," *European Journal of Control*, vol. 24, pp. 14–32, 2015.

[6] G. C. Calafiore and M. C. Campi, "Uncertain convex programs: Randomized solutions and confidence levels," *Mathematical Programming*, vol. 102, no. 1, pp. 25–46, 2005.

[7] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1211–1230, 2008.

[8] G. C. Calafiore, "Random convex programs," *SIAM Journal on Optimization*, vol. 20, no. 6, pp. 3427–3464, 2010.

[9] J. Luedtke and S. Ahmed, "A sample approximation approach for optimization with probabilistic constraints," *SIAM Journal on Optimization*, vol. 19, no. 2, pp. 674–699, 2008.

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

[11] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Mathematical Programming*, vol. 137, pp. 167–198, 2013.

[12] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, "Ambiguous joint chance constraints under mean and dispersion information," *Operations Research*, vol. 65, no. 3, pp. 751–767, 2017.

[13] E. Erdoğan and G. Iyengar, "Ambiguous chance constrained problems and robust optimization," *Mathematical Programming*, vol. 107, no. 1-2, pp. 37–61, 2006.

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

[15] C. Villani, *Topics in optimal transportation*. American Mathematical Soc., 2003, no. 58.

[16] R. Gao and A. J. Kleywegt, "Distributionally robust stochastic optimization with Wasserstein distance," 2016, available online at https://arxiv.org/abs/1604.02199.

[17] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming. Ser. A*, 2017, to appear. Available at https://arxiv.org/abs/1505.05116.

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

[19] W. Xie and S. Ahmed, "Bicriteria approximation of chance constrained covering problems," 2018, available online at http://www.optimization-online.org/DB_HTML/2018/01/6411.html.

[20] W. Xie, "On distributionally robust chance constrained program with Wasserstein distance," 2018, available online at http://www.optimization-online.org/DB_FILE/2018/06/6662.pdf.

[21] Z. Chen, D. Kuhn, and W. Wiesemann, "Data-driven chance constrained programs over Wasserstein balls," 2018, available online at http://www.optimization-online.org/DB_FILE/2018/06/6671.pdf.

[22] A. Prekopa, "On probabilistic constrained programming," in *Proceedings of the Princeton symposium on mathematical programming*, vol. 113, 1970, p. 138.

[23] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2006.

[24] S. Mehrotra and D. Papp, "A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization," *SIAM Journal on Optimization*, vol. 24, no. 4, pp. 1670–1697, 2014.

[25] F. Luo and S. Mehrotra, "Decomposition algorithm for distributionally robust optimization using wasserstein metric," 2017, arXiv preprint arXiv:1704.03920.

[26] A. R. Hota, A. Cherukuri, and J. Lygeros, "Data-driven chance constrained optimization under wasserstein ambiguity sets," 2018, arXiv preprint arXiv:1805.06729.

[27] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.