# CONVERGENCE ANALYSIS AND A DC APPROXIMATION METHOD FOR DATA-DRIVEN MATHEMATICAL PROGRAMS WITH DISTRIBUTIONALLY ROBUST CHANCE CONSTRAINTS*

HAILIN SUN†, DALI ZHANG‡, AND YANNAN CHEN§

**Abstract.** In this paper, we consider the convergence analysis of data-driven mathematical programs with distributionally robust chance constraints (MPDRCC) under weaker conditions without continuity assumption of distributionally robust probability functions. Moreover, combining with the data-driven approximation, we propose a DC approximation method to MPDRCC without some special tractable structures. We also give the convergence analysis of the DC approximation method without continuity assumption of distributionally robust probability functions, and apply a recent DC algorithm to solve them. The numerical tests verify the theoretical results and show the effectiveness of the data-driven approximated DC approximation method.

**Key words.** Distributionally robust optimization, chance constraints, data-driven, convergence analysis, DC approximation

**AMS subject classifications.** 90C15

## 1. Introduction.

In this paper, we consider the mathematical programs with distributionally robust chance constraints (MPDRCC)

$$\begin{array}{ll} \min\limits_{x\in X} & f(x) \\ s.t. & \inf\limits_{P\in\mathcal{P}} P(g(x,\xi)\le 0)\ge 1-\alpha, \end{array} \tag{1}$$

and its data-driven formulation

$$\begin{array}{ll} \min\limits_{x\in X} & f(x) \\ s.t. & \inf\limits_{P\in\mathcal{P}_N} P(g(x,\xi)\le 0)\ge 1-\alpha, \end{array} \tag{2}$$

where $X$ is a closed subset of $\mathbb{R}^n$, $\xi:\Omega\to\Xi\subset\mathbb{R}^k$ is a random vector defined on measurable space $(\Omega,F)$ with support set $\Xi$, $f:\mathbb{R}^n\to\mathbb{R}$ is Lipschitz continuous function w.r.t. $x$, $g:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$ is Lipschitz continuous function w.r.t. $x$ for all $\xi\in\Xi$, $\mathscr{P}(\Xi)$ denotes the set of all probability measures defined on $\Xi$, $\mathcal{P}\subset\mathscr{P}(\Xi)$ is the *ambiguity set* which indicates ambiguity of the true probability distribution of $\xi$ and $\mathcal{P}_N\subset\mathscr{P}(\Xi)$ is a set of probability measures which approximates $\mathcal{P}$ in some sense (to be specified later) as $N\to\infty$. In the case when the ambiguity set $\mathcal{P}$ is singleton, the problem reduces to mathematical programs with chance constraints (MPCC)

$$\begin{array}{ll} \min\limits_{x\in X} & f(x) \\ s.t. & P(g(x,\xi)\le 0)\ge 1-\alpha \end{array} \tag{3}$$

*Submitted to the editors DATE.

**Funding:** This work was funded by National Natural Science Foundation of China #11871276, #11571056, #11571178 and #11771405. The authors certify that the general content of the manuscript, in whole or in part, is not submitted, accepted, or published elsewhere, including conference proceedings. The authors also certify that the general content of the manuscript, in whole or in part, will not be submitted, accepted, or published elsewhere, including conference proceedings, while it is under consideration or in production.

†Corresponding author. Jiangsu Key Laboratory for NSLSCS, School of Mathematical Sciences, Nanjing Normal University, Nanjing, 210023, China (hlsun@njnu.edu.cn).

‡Antai College of Economic and Management(SUGLI), Shanghai Jiao Tong University, 200030, Shanghai, China, (zhangdl@sjtu.edu.cn).

§School of Mathematical Sciences, South China Normal University, Guangzhou, 510631, China, (ynchen@scnu.edu.cn)

and $\mathbb{E}_P[\mathbb{1}_{\{g(x,\xi)\le 0\}}(g(x,\xi))] = P(g(x,\xi)\le 0)$ where

$$\mathbb{1}_{g<0}(g) = \begin{cases} 1 & g < 0 \\ 0 & g \ge 0 \end{cases}$$

is the index function.

The MPCC has wide applications in engineering design, supply chain management, production planing, and water management, see [34]. MPCC was first discussed by Charnes, Cooper, and Symonds [3], and then by Miller and Wagner [23] and Prékopa [27]. In recent years, there are many significant progress of MPCC from both a theoretical viewpoint and an algorithmic perspective. The convexity of chance constraints has been investigated by Henrion and Strugarek [14] and van Ackooij [38]. Under multivariate Gaussian distribution of random vector, van Ackooij and Henrion [39] propose an efficient method to compute the gradient and value of probability functions and then MPCC can be solved by existing solvers. Many approximation methods also proposed for solving MPCC such as sample approximation [22], constrained bundle method and other methods with regularization [6], inner-outer approximation[9], convex approximation [24], difference of two convex function (DC) approximation [15, 17, 32], sample average approximation (SAA) [25]. For more discussion on MPCC, we refer readers to [28, 34] and references therein.

In the case when the decision makers do not have full knowledge of the underlying probability distribution but still can obtain some partial information and use them to construct an ambiguity set of probability distributions which contains or approximates the true probability distribution, they may consider solving the MPDRCC. MPDRCC are important problems in distributionally robust optimization, a number of papers have appeared on this topic. In the case when the ambiguity set is characterized by its mean and variance and $g(\cdot,\cdot)$ has some special structures, Calafiore and El Ghaoui [2], Zymler, Kuhn, and Rustem [45] and Yang and Xu [43] prove that the MPDRCC can be tractable. Hanasusanto et al. [12] and Hanasusanto et al. [13] investigate tractability of a rich class of ambiguity sets defined through moment conditions and structural information for MPDRCC. In [42], Xie and Ahmed generalize the results in [2, 43, 45] and identified several sufficient conditions for convex reformulation of MPDRCC when ambiguity set is specified by moment constraints. Erdoğ̆an and Iyengar [7] construct an ambiguity set based on the Prohorov-metric and approximate MPDRCC by a set of sample-based robust optimization constraints. Jiang and Guan [18] and Hu and Hong [16] derive tractable reformulations for the family of $\phi$-diverge probability metric (e.g., Kullback-Leibler (KL) divergence, Hellinger distance, etc.). Xie [40], Xie and Ahmed [41] and Chen et al. [5] investigate convex reformulation of MPDRCC with Wasserstein distance. However, in general cases, the MPDRCC are still intractable and difficult to solve.

Our motivation starts from the convergence analysis of data-driven MPDRCC recently proposed by Guo et al. [11]. They consider the convergence analysis between (1) and (2) when $\mathcal{P}_N$ converges to $\mathcal{P}$ under a key assumption:

$$\inf_{P\in\mathcal{P}} P(g(x,\xi)\le 0) \text{ is continuous w.r.t. x.} \tag{4}$$

The continuity of the (distributional robust) probability functions is well documented in the literature of stochastic programming, see [11, 25, 37]. One sufficiently condition of (4) is

$$P\{g(x,\xi)=0\} = 0 \text{ for all } x\in X,\ \forall P\in\mathcal{P}. \tag{5}$$

Guo et al. [11] consider a weaker condition

$$P\{H(x)/\text{int}(H(x))\} = 0 \text{ for all } x \in X \text{ and } P \in \mathcal{P}, \tag{6}$$

to guarantee (4), where $H(x) := \{\xi : g(x,\xi) \le 0\}$. However, for a large class of ambiguity sets, (5) and (6) can not be satisfied. Moreover, from the motivated example in Section 2.1 (Example 2.1), it seems that when (4) is not satisfied, the convergence may still holds. This motivates us to investigate the convergence analysis of MPDRCC under weaker conditions without (4).

We also interesting in how to solve the data-driven MPDRCC. Different from tractable reformulation base on the special structure of MPDRCC in the literatures, we consider a DC approximation of MPDRCC which is inspired by the DC approximation proposed in Hong et al. [15] for MPCC. In [15], Hong et al. use condition

$$P\{g(x,\xi) = 0\} = 0 \text{ for all } x \in X \tag{7}$$

to guarantee the continuous assumption (4) and proof the convergence of DC approximation. Hu et al. [17] relax the condition (7) by using smoothing method. But they still require

$$\{x : P\{g(x,\xi) > 0\} \le \alpha\} = \text{cl}\{x : P\{g(x,\xi) > 0\} < \alpha\}. \tag{8}$$

By slightly changing their approach, we propose a different DC approximation which does not require condition (7) or (8), and apply the approach to MPDRCC. The contributions of the paper are as follows:

- (i) We study the data-driven MPDRCC and corresponding convergence analysis. This part can be considered as an extention of resent work by Guo et al. [11] but under weaker conditions without the continuity assumption (4).
- (ii) Moreover, for solving MPDRCC without tractable structures, we propose a DC approximation method for MPDRCC. The DC approximation method can be considered as an extension of DC approximation of MPCC in [15]. The convergence analysis has been investigated under weaker conditions without continuity assumption of probability functions. We also study the convergence analysis of data-driven DC approximated MPDRCC.
- (iii) With three kinds of ambiguity sets, we reformulate the data-driven DC approximated MPDRCC as optimization problems with DC constraints and apply the penalty and augmented Lagrangian methods proposed by Lu et. al. [21] to solve them. The numerical tests show that the correctness of our theoretical results and effectiveness of the DC approximation.

The paper is constructed as follows. Some preliminarise knowledge and a motivation example are given in Section 2. In Section 3, we study the data-driven MPDRCC and the corresponding convergence analysis without continuity condition (4). In Section 4, we propose a DC approximation of MPDRCC and investigate the convergence analysis. In Section 5, we reformulate the DC approximated MPDRCC as optimization problems with DC constraints under three kinds of ambiguity sets. The the penalty and augmented Lagrangian methods proposed by Lu et al. [21] has been applied to solve the optimization problems with DC constraints. Elementary numerical tests and applications are given in Section 6 to show the correctness of the theorems and the effectiveness of the DC approximation.

Throughout the paper, we use the following notation. $\mathbb{R}^n_+$ denotes the cone of vectors with non-negative components in $\mathbb{R}^n$. $\mathscr{P}(\Xi)$ denotes the set of all probability measures over $\Xi$. For a set $\mathcal{C} \subset \mathcal{Z}$, we use by convention "int $\mathcal{C}$" and "cl $\mathcal{C}$" to denote its interior and closure respectively.

## 2. A motivation examples and preliminaries .

**2.1. A motivation example.** We first give a motivation example as follows. The example shows that the continuity condition (4) is not a necessary condition of convergence analysis between MPDRCC (1) and (2) when the ambiguity set $\mathcal{P}$ is approximated by $\mathcal{P}_N$.

EXAMPLE 2.1. Consider

$$\begin{aligned} \min_{x\in X} \quad & x^2 \\ s.t. \quad & \inf_{P\in\mathcal{P}} \ \mathbb{E}_P[\mathbb{1}_{\{\xi-x\leq 0\}}(\xi)] \geq 1-\alpha, \end{aligned} \tag{1}$$

and its data-driven approximation problem

$$\begin{aligned} \min_{x\in X} \quad & x^2 \\ s.t. \quad & \inf_{P\in\mathcal{P}_N} \ \mathbb{E}_P[\mathbb{1}_{\{\xi-x\leq 0\}}(\xi)] \geq 1-\alpha, \end{aligned} \tag{2}$$

where $N \geq 1$, $\alpha = 0.5$,

$$\mathcal{P} := \left\{ P \in \mathscr{P} : \mathbb{E}_P[\xi] = 1, \xi \in \left[\frac{1}{2}, \frac{3}{2}\right] \right\},$$

and

$$\mathcal{P}_N := \left\{ P \in \mathscr{P} : \mathbb{E}_P[\xi] = 1 - \frac{1}{2N}, \xi \in \left[\frac{1}{2}, \frac{3}{2}\right] \right\}.$$

It is obvious that the feasible sets of problem (1) and (2) are $[\frac{3}{2}, +\infty)$ and $[\frac{3}{2} - \frac{1}{N}, +\infty)$. Then the optimal solutions of problem (1) and (2) are $x^* = \frac{3}{2}$ and $x_N = \frac{3N-2}{2N}$. So $x_N \to x^*$ as $N \to \infty$.

Let $P_0 \in \mathcal{P}$ with $P_0(\xi = \frac{1}{2}) = 0.5$, $P_0(\xi = \frac{3}{2}) = 0.5$. Then

$$\mathbb{E}_{P_0}[\mathbb{1}_{\{\xi-x\leq 0\}}(\xi - x)] = \begin{cases} 0 & x \leq \frac{1}{2} \\ 0.5 & x \in (\frac{1}{2}, \frac{3}{2}) \\ 1 & x \geq \frac{3}{2} \end{cases}$$

is not continuous at $x_1 = \frac{1}{2}$ and $x_2 = \frac{3}{2}$. Since $H(x_1) = \left[\frac{1}{2}, \frac{3}{2}\right]$, $\text{int}H(x_1) = \left(\frac{1}{2}, \frac{3}{2}\right)$ and $P_0\{H(x)/\text{int}(H(x))\} = P_0\left\{\xi \in \left\{\frac{1}{2}, \frac{3}{2}\right\}\right\} = 1$, then conditions (6) can not hold.

Moreover, it is easy to observe that

$$\inf_{P\in\mathcal{P}} \mathbb{E}_P\left[\mathbb{1}_{\{\xi-\frac{3}{2}\leq 0\}}(\xi)\right] = 1$$

while for any sufficiently small $\epsilon > 0$,

$$\inf_{P\in\mathcal{P}} \mathbb{E}_P\left[\mathbb{1}_{\{\xi-(\frac{3}{2}-\epsilon)\leq 0\}}(\xi)\right] \leq \frac{1}{2}.$$

Then the continuity assumption (4) can not hold.

The example shows that conditions (4) may not be necessary when we consider the convergence analysis of data-driven MPDRCC.

**2.2. Graphical convergence and metrics of probability measures.** Let $\mathbb{N} = \{1, 2 \ldots\}$ be the set of natural numbers, $\mathcal{N}_\infty^\# = \{\text{all subsequences of } \mathbb{N}\}$ and $\mathcal{N}_\infty = \{\text{all indexes} \geq \text{ some } \bar{k}\}$. For sequence $N$, we use $(x^k, \epsilon_k) \underset{N}{\to} (x, 0)$ to denote $\epsilon_k \downarrow 0$ and $x^k \to x$ when $k \in N$.

Definition 1. *[31, Definition 5.32] For the mappings $z^\epsilon : X \to \mathbb{R}^{n+1}$, the graphical outer limit, denoted by $g\text{-}\limsup_\epsilon z^\epsilon : X \rightrightarrows \mathbb{R}^{n+1}$, is the mapping having as its graph the set $\limsup_\epsilon \operatorname{gph} z^\epsilon$:*

$$g\text{-}\limsup_\epsilon z^\epsilon(x) = \{z \,|\, \exists N \in \mathcal{N}_\infty^\#,\ (x^k, \epsilon_k) \underset{N}{\to} (x, 0),\ z^{\epsilon_k}(x^k) \underset{N}{\to} z\}.$$

*The graphical inner limit, denoted by $g\text{-}\liminf_\epsilon z^\epsilon$, is the mapping having as its graph the set $\liminf_\epsilon \operatorname{gph} z^\epsilon$:*

$$g\text{-}\liminf_\epsilon z^\epsilon(x) = \{z \,|\, \exists\, N \in \mathcal{N}_\infty,\ (x^k, \epsilon_k) \underset{N}{\to} (x, 0),\ z^{\epsilon_k}(x^k) \underset{N}{\to} z\}.$$

*If the outer and inner limits coincide, the graphical limit $g\text{-}\lim_\epsilon z^\epsilon$ exists; thus, $Z^0 = g\text{-}\lim_\epsilon z^\epsilon$ if and only if $g\text{-}\limsup_\epsilon z^\epsilon \subseteq Z^0 \subseteq g\text{-}\liminf_\epsilon z^\epsilon$ and one writes $z^\epsilon \xrightarrow{g} Z^0$; the mappings $z^\epsilon$ are said to* converge graphically *to $Z^0$.*

**Metrics of probability measures:** We need appropriate metrics for the set in order to characterize convergence of $\mathcal{P}_N \to \mathcal{P}$. In this section, we will introduce $\zeta$-metrics and pseudometric, see [1, 10].

Definition 2. *A (semi-) $\zeta$-metric is defined by*

$$\mathsf{dl}_{\mathscr{G}}(P, Q) := \sup_{g \in \mathscr{G}} \left|\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]\right|, \tag{3}$$

*where $P, Q \in \mathscr{P}(\Xi)$ and $\mathscr{G}$ is a set of real-valued bounded measurable functions on $\Xi$.*

The $\zeta$-metrics cover a wide range of metrics in probability theory including the total variation metric, Kantorovich/ Wasserstein metric, bounded Lipschitz metric and some other metrics; see [10], [29] or [44] and references therein. Specifically, if

$$\mathscr{G} := \left\{ g : \Xi \to \mathbb{R} |\ g \text{ is } \mathscr{B} \text{ measurable},\ \sup_{\xi \in \Xi} |g(\xi)| \leq 1 \right\},$$

then $\mathsf{dl}_{\mathscr{G}}$ reduces to the *total variation metric*, denoted by $\mathsf{dl}_{TV}$. If $g$ is restricted further to be Lipschitz continuous with modulus bounded by 1, that is,

$$\mathscr{G} = \{g : g \text{ is Lipschitz continuous and Lipschtiz modulus } L_1(g) \leq 1\},$$

then we arrive at Kantorovich/ Wasserstein metric, denoted by $\mathsf{dl}_K$.

Under the $\zeta$-metric $\mathsf{dl}_{\mathscr{G}}$, we can define the distance from a distribution $Q$ to a set of distributions $\mathcal{C}$ as $\mathsf{dl}_{\mathscr{G}}(Q, \mathcal{C}) := \inf_{P \in \mathcal{C}} \mathsf{dl}_{\mathscr{G}}(Q, P)$, deviation from one set $\mathcal{C} \in \mathscr{P}(\Xi)$ to another $\mathcal{C}' \in \mathscr{P}(\Xi)$ as $\mathbb{D}(\mathcal{C}', \mathcal{C}; \mathsf{dl}_{\mathscr{G}}) := \sup_{Q \in \mathcal{C}'} \mathsf{dl}_{\mathscr{G}}(Q, \mathcal{C})$ and the Hausdorff distance between two subset $\mathcal{C}$ and $\mathcal{C}'$ in the space of probability measures $\mathscr{P}(\Xi)$ as $\mathbb{H}(\mathcal{C}', \mathcal{C}; \mathsf{dl}_{\mathscr{G}}) := \max\{\mathbb{D}(\mathcal{C}', \mathcal{C}; \mathsf{dl}_{\mathscr{G}}),\ \mathbb{D}(\mathcal{C}, \mathcal{C}'; \mathsf{dl}_{\mathscr{G}})\}$.

In the case when the set of function $\mathscr{G}$ is not large enough such that

$$\sup_{g \in \mathscr{G}} \left|\mathbb{E}_P[g] - \mathbb{E}_Q[g]\right| = 0$$

does not necessarily imply $P = Q$, the type of "metric" is not a $\zeta$-metric, and named by pseudometric. This type of pseudometric is widely used for stability analysis in stochastic programming; see an excellent review by Römisch [30].

We also introduce two definitions from [1]:

DEFINITION 3. *Let $\mathcal{A}$ be a set of probability measures on $(\Xi, \mathscr{B})$. $\mathcal{A}$ is said to be* tight *if for any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that $\inf_{P\in\mathcal{A}} P(\Xi_\epsilon) > 1-\epsilon$. In the case when $\mathcal{A}$ is a singleton, it reduces to the tightness of a single probability measure. $\mathcal{A}$ is said to be* closed *(under the weak topology) if for any sequence $\{P_N\} \subset \mathcal{A}$ with $P_N \to P$ weakly, we have $P \in \mathcal{A}$.*

DEFINITION 4. *Let $\{P_N\} \subset \mathscr{P}$ be a sequence of probability measures. $\{P_N\}$ is said to converge to $P \in \mathscr{P}$* weakly *if $\lim_{N\to\infty} \int_\Xi h(\xi) P_N(d\xi) = \int_\Xi h(\xi)P(d\xi)$ for each bounded and continuous function $h : \Xi \to \mathbb{R}$. Let $\mathcal{A} \subset \mathscr{A}$ be a set of probability measures. $\mathcal{A}$ is said to be weakly compact (under the weak topology) if every sequence $\{A_N\} \subset \mathcal{A}$ contains a subsequence $\{A_{N'}\}$ and $A \in \mathcal{A}$ such that $A_{N'} \to A$.*

**3. Convergence analysis in data-driven problem.** In this section, we consider the equivalent formulation of problem (1):

$$\begin{array}{ll} \min\limits_{x\in X} & f(x) \\ s.t. & \sup\limits_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x,\xi)>0}(\xi)] \leq \alpha, \end{array} \tag{1}$$

and its data-driven form:

$$\begin{array}{ll} \min\limits_{x\in X} & f(x) \\ s.t. & \sup\limits_{P\in\mathcal{P}_N} \mathbb{E}_P[\mathbb{1}_{g(x,\xi)>0}(\xi)] \leq \alpha. \end{array} \tag{2}$$

In what follows, we consider the convergence of the optimal solution sets of (2). To ease the exposition, for each fixed $x \in X$, let

$$v(x) := \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\mathbb{1}_{(g(x,\xi)>0)}\right], \quad v_N(x) := \sup_{P\in\mathcal{P}_N} \mathbb{E}_P\left[\mathbb{1}_{(g(x,\xi)>0)}\right], \tag{3}$$

$\mathcal{F}^*$, $\vartheta^*$, $\mathcal{S}^*$ and $\mathcal{F}_N$, $\vartheta_N$, $\mathcal{S}_N$ be the feasible solution set, the optimal value and the optimal solution set of problem (1) and (2) respectively. Let $\hat{\mathcal{P}} \subset \mathscr{P}$ denote a set of distributions such that

$$\mathcal{P}, \mathcal{P}_N \subset \hat{\mathcal{P}} \tag{4}$$

for $N$ sufficiently large. The existence of $\hat{\mathcal{P}}$ is trivial as we can take the union of $\mathcal{P}$ and $\mathcal{P}_N$, for $N = 1, 2, 3, \cdots$.

ASSUMPTION 3.1. *Let $\mathcal{P}, \mathcal{P}_N$ be nonempty and defined as in (1) and (2) respectively. Then*

*(a) there exists a weakly compact set (see Definition 4) $\hat{\mathcal{P}} \subset \mathscr{P}$ such that (4) holds;*

*(b) $\mathbb{H}(\mathcal{P}_N, \mathcal{P}; \mathsf{dl}_{\mathscr{L}_1}) \to 0$ almost surely as $N \to \infty$, where $\mathbb{H}(\cdot, \cdot; \mathsf{dl}_{\mathscr{L}_1})$ is defined as in Section 2.2 with $\mathscr{G}_1 := \{l(\cdot) = \mathbb{1}_{\{g(x,\xi)>0\}}(\xi) : x \in X\}$.*

In [11, Theorem 3.2], Guo et. al. investigate the convergence analysis between $v_N$ and $v$ under Assumption 3.1 and the assumption can hold in several cases, e.g., $\mathbb{H}(\mathcal{P}_N, \mathcal{P}; \mathsf{dl}_{TV}) \to 0$, see [35].

PROPOSITION 3.1. *[11, Theorem 3.2] Suppose Assumption 3.1 (b) holds, then $v_N(\cdot)$ converges uniformly to $v(\cdot)$ over $X$ as $N$ tends to infinity, that is,*

$$\lim_{N\to\infty} \sup_{x\in X} |v_N(x) - v(x)| = 0. \tag{5}$$

To avoid assuming continuity of $v(\cdot)$, we proof the lower semi-continuity of $v(\cdot)$.

LEMMA 3.1. *Suppose Assumption 3.1 (a) holds and $g(x,\xi)$ is continuous w.r.t. $x$ for all $\xi \in \Xi$, then the function $v(\cdot)$ defined in (3) is lower semi-continuous.*

*Proof.* Note that $g(\cdot,\xi)$ is continuous for all $\xi \in \Xi$, it is easy to observe that $\mathbb{1}_{\{g(x,\xi)>0\}}(g(x,\xi))$ is lower semi-continuous w.r.t. $x$ for all $\xi \in \Xi$.

Let $\{x_k\}$ be any sequence that converges to $x$. Then

$$\begin{aligned}\liminf_{k\to\infty} v(x_k) - v(x) &= \liminf_{k\to\infty}(v(x_k) - v(x)) \\ &= \liminf_{k\to\infty}\left(\sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x_k,\xi)>0}(g(x_k,\xi))] - \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x,\xi)>0}(g(x,\xi))]\right). \quad (6)\end{aligned}$$

Note that for any $P_0 \in \mathcal{P}$,

$$\begin{aligned}&\sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x_k,\xi)>0}(g(x_k,\xi))] - \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x,\xi)>0}(g(x,\xi))] \\ &\quad \geq \mathbb{E}_{P_0}[\mathbb{1}_{g(x_k,\xi)>0}(g(x_k,\xi))] - \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x,\xi)>0}(g(x,\xi))].\end{aligned}$$

Moreover, for any $\epsilon > 0$, there exists $P_\epsilon \in \mathcal{P}$ such that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x,\xi)>0}(g(x,\xi))] \leq \mathbb{E}_{P_\epsilon}[\mathbb{1}_{g(x,\xi)>0}(g(x,\xi))] + \epsilon.$$

Then

$$\begin{aligned}&\sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x_k,\xi)>0}(gx_k,\xi))] - \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{g(x,\xi)>0}(g(x,\xi))] \\ &\quad \geq \mathbb{E}_{P_\epsilon}[\mathbb{1}_{g(x_k,\xi)>0}(g(x_k,\xi))] - \mathbb{E}_{P_\epsilon}[\mathbb{1}_{g(x,\xi)>0}(g(x,\xi))] - \epsilon. \quad (7)\end{aligned}$$

Combine (6) and (7) and by Fatou's Lemma and the lower semi-continuous of $\mathbb{1}_{g(x,\xi)\geq 0}(g(x,\xi))$ w.r.t. $x$ for all $\xi \in \Xi$, we have

$$\begin{aligned}\liminf_{k\to\infty} v(x_k) - v(x) &\geq \liminf_{k\to\infty} \mathbb{E}_{P_\epsilon}\left[\mathbb{1}_{g(x_k,\xi)>0}(g(x_k,\xi)) - \mathbb{1}_{g(x,\xi)>0}(g(x,\xi))\right] - \epsilon \\ &\geq \mathbb{E}_{P_\epsilon}\left[\liminf_{k\to\infty}\left(\mathbb{1}_{g(x_k,\xi)>0}(g(x_k,\xi)) - \mathbb{1}_{g(x,\xi)>0}(g(x,\xi))\right)\right] - \epsilon \geq -\epsilon\end{aligned}$$

Moreover, by the arbitrariness of $\epsilon$, $\liminf_{k\to\infty} v(x_k) - v(x) \geq 0$ for any sequence $\{x_k\} \to x$ as $k \to \infty$, and then $v(\cdot)$ is lower semi-continuous w.r.t. $x$. ∎

Now we are ready to prove the convergence of data-driven MPDRCC, the main result in the section.

THEOREM 3.1. *Assume that $X$ is a compact set, $\mathcal{S}^*$ and $\mathcal{S}_N$ are nonempty for sufficiently large $N > 0$. Suppose (a) $g(x,\xi)$ is continuous w.r.t. $x$ for all $\xi \in \Xi$, (b) $cl\mathcal{F}_s \cap \mathcal{S} \neq \emptyset$, where $\mathcal{F}_s := \{x \in X : v(x) < \alpha\}$, (c) Assumption 3.1 holds. Then we have (i) $\lim_{N\to\infty} \vartheta_N = \vartheta^*$; (ii) $\lim_{N\to\infty} \mathbb{D}(\mathcal{S}_N, \mathcal{S}^*) = 0$.*

*Proof.* Let $\{x_N\}$ be any sequence of optimal solutions of problem (2) and $x^*$ be any accumulation point. Then without loss of generality, there exists a subsequence $\{N_k\}$ such that $\{x_{N_k}\}$ is the subsequence of the solutions of problem (2) and $\lim_{k\to\infty} x_{N_k} = x^*$. By the continuity of the objective function $f$, we only need to prove $x^* \in \mathcal{S}^*$.

We prove the feasibility of $x^*$ for problem (1) firstly. Note that $x_{N_k} \in \mathcal{S}_{N_k}$, $v_{N_k}(x_{N_k}) \leq \alpha$. Moreover, $v_N(\cdot)$ converges to $v(\cdot)$ uniformly, then for any $\delta_1 > 0$, there exists $k_0 > 0$ such that

$$\sup_{x\in X} |v_{N_k}(x) - v(x)| \leq \delta_1 \tag{8}$$

for all $k \geq k_0$. Then for any $\delta_1 > 0$, there exists $k_0 > 0$ such that $v(x_{N_k}) \leq \alpha + \delta_1$ for all $k \geq k_0$. Furthermore, by Lemma 3.1, $v(\cdot)$ is lower semi-continuous, then for all $k \geq k_0$, $v(x^*) \leq \liminf_{k\to\infty} v(x_{N_k}) \leq \alpha + \delta_1$. By the arbitrariness of $\delta_1$, $v(x^*) \leq \alpha$, which implies $x^* \in \mathcal{F}$.

Then we consider $\mathcal{F}_\delta := \{x \in X : v(x) \leq \alpha - \delta\}$ with $\delta > 0$ and problem

$$\vartheta_\delta := \min_{x\in X}\{f(x) \text{ s.t. } x \in \mathcal{F}_\delta\}. \tag{9}$$

Let $\{x_\delta\}$ be any sequence of optimal solutions of problem (9). By condition (b), we have $\lim_{\delta\downarrow 0} \mathcal{F}_\delta = cl\mathcal{F}_s$, then any accumulation point of the sequence $\bar{x} \in \mathcal{S}$. By (8), for any $\delta > 0$, there exists sufficiently large $k_\delta$ such that $\mathcal{F}_\delta \subseteq \mathcal{F}_{N_{k_\delta}}$ and $\{N_{k_\delta}\}$ is a subsequence of $\{N_k\}$. Then we have $\{x_{N_{k_\delta}}\} \to x^*$ and $f(x_{N_{k_\delta}}) \leq f(x_\delta)$, which implies $f(x^*) \leq f(\bar{x}) = \vartheta^*$. Combining this inequation with $x^* \in \mathcal{F}$, we have $x^* \in \mathcal{S}^*$. ∎

Condition (b) in Theorem 3.1 requires problem (1) to have a non-isolated optimal solution. It is fulfilled if the feasible set $\mathcal{F}$ is convex or connected and discussed in [11, Theorem 3.4]. Moreover, the following example shows when condition (b) in Theorem 3.1 fails, the convergence can not hold.

Example 3.1. *Consider*

$$\begin{array}{ll} \min\limits_{x\in X} & x^2 \\ s.t. & \sup\limits_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{\xi-x>0\}}(\xi)] \leq \alpha, \end{array} \tag{10}$$

*and its approximation problem*

$$\begin{array}{ll} \min\limits_{x\in X} & x^2 \\ s.t. & \sup\limits_{P\in\mathcal{P}_N} \mathbb{E}_P[\mathbb{1}_{\{\xi-x>0\}}(\xi)] \leq \alpha, \end{array} \tag{11}$$

*where*

$$\mathcal{P} = \left\{P \in \mathscr{P}(\Xi): \begin{array}{ll} \text{Prob}\{\xi=-1\} = 0.25, & \text{Prob}\{\xi=0\} = 0.25, \\ \text{Prob}\{\xi=1\} = 0.25, & \text{Prob}\{\xi=2\} = 0.25 \end{array}\right\}$$

*and*

$$\mathcal{P}_N = \left\{P \in \mathscr{P}(\Xi): \begin{array}{ll} \text{Prob}\{\xi=-1\} = 0.25 - \frac{1}{N}, & \text{Prob}\{\xi=0\} = 0.25, \\ \text{Prob}\{\xi=1\} = 0.25 + \frac{1}{N}, & \text{Prob}\{\xi=2\} = 0.25 \end{array}\right\},$$

$N \geq 4$, $\alpha = 0.5$ *and* $\Xi := \{-1, 0, 1, 2\}$. *It is obvious that the feasible sets of problem (10) and (11) are* $[0, +\infty)$ *and* $[1, +\infty)$. *Then the optimal solutions of problem (10) and (11) are* $x^* = 0$ *and* $x_N = 1$ *respectively. So* $x_N \nrightarrow x^*$ *as* $N \to \infty$.

**4. DC approximation methods of MPDRCC.** In the literatures, there are several tractable reformulation of MPDRCC when the problems satisfy some special structures. But in general cases, the MPDRCC are intractable. In this section, we consider a DC approximation method for general MPDRCC.

**4.1. DC Approximation Methods.** We consider the DC approximation of index function as follows:

DEFINITION 4.1. *The DC approximation function of index function $\mathbb{1}_{\{\cdot>0\}}(\cdot)$, denote by $\Psi^{DC}(\cdot,\epsilon)$, is defined as*

$$\Psi^{DC}(\cdot,\epsilon) := \frac{1}{\epsilon}\left((\cdot)_+ - (\cdot - \epsilon)_+\right), \tag{1}$$

*where $\epsilon > 0$.*

LEMMA 4.1. *The DC approximation function $\Psi^{DC}(\cdot,\epsilon)$ of index function $\mathbb{1}_{\{\cdot>0\}}(\cdot)$ is globally Lipschitz continuous and increaseing w.r.t. $\cdot$ for any $\epsilon > 0$ such that,*

*(1). $\Psi^{DC}(\cdot,\epsilon)$ pointwise converges to $\mathbb{1}_{\{\cdot>0\}}(\cdot)$;*

*(2). $\Psi^{DC}(\cdot,\epsilon) \xrightarrow{g} \Psi_0(\cdot)$ (see Definition 1), where* $\Psi_0(g) := \begin{cases} 1 & g > 0 \\ [0,1] & g = 0 \\ 0 & g < 0; \end{cases}$

*(3). $\Psi^{DC}(g,\epsilon) \geq 0$ and $\Psi^{DC}(g,\epsilon) = 0$ if $g = 0$, and $\Psi^{DC}(g,\epsilon) = 1$ if $g \geq \epsilon$.*

*Proof.* (1) and (3) are easy to observe, we only need to prove (2). For any $g \neq 0$, when $\epsilon$ is sufficiently small, both $\Psi^{DC}(g,\epsilon)$ and $\Psi_0(g)$ are single valued continuous function w.r.t. $g$ and

$$\limsup_{(g',\epsilon)\to(g,0)} \Psi^{DC}(g',\epsilon) = \Psi_0(g) = \liminf_{(g',\epsilon)\to(g,0)} \Psi^{DC}(g',\epsilon).$$

When $g = 0$, by [31, Proposition 5.33],

$$\limsup_{(g',\epsilon)\to(g,0)} \Psi^{DC}(g',\epsilon) = \Psi_0(g) = [0,1] = \liminf_{(g',\epsilon)\to(g,0)} \Psi^{DC}(g',\epsilon).$$

Then (2) holds. ∎

By using the DC approximation function $\Psi^{DC}(\cdot,\epsilon)$, the DC approximated MP-DRCC can be written as:

$$\begin{array}{ll} \min\limits_{x\in X} & f(x) \\ s.t. & \sup\limits_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(x,\xi),\epsilon)] \leq \alpha. \end{array} \tag{2}$$

PROPOSITION 4.1. *Consider constraint functions in the MPDRCC (1) and its DC approximation problem (2). For all $x_0 \in X$ and any sequence $\{(x',\epsilon)\} \to (x_0,0)$, the following inequality holds:*

$$\liminf_{(x',\epsilon)\to(x_0,0)} \sup_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)] \geq \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{g(x_0,\xi)>0\}}(\xi)]. \tag{3}$$

*Proof.* By Lemma 4.1, we have $\Psi(g(x,\xi),\epsilon) \xrightarrow{g} \Psi_0(g(x,\xi))$. Then by Aumann's (set-valued) expectation, we have g-$\lim_\epsilon \mathbb{E}[\Psi(g(x,\xi),\epsilon)] \subset \mathbb{E}[\Psi_0(g(x,\xi))]$. We also have $\mathbb{1}_{\{g(x,\xi)>0\}}(\xi) \subset \Psi_0(g(x,\xi))$ and $\varphi \geq \mathbb{1}_{\{g(x,\xi)>0\}}(\xi)$ for any $\varphi \in \Psi_0(g(x,\xi))$. Then, for any sequence $\{(x',\epsilon)\} \to (x_0,0)$ and $\xi \in \Xi$,

$$\liminf_{(x',\epsilon)\to(x_0,0)} \Psi^{DC}(g(x',\xi),\epsilon) \geq \mathbb{1}_{\{g(x_0,\xi)>0\}}(\xi).$$

By Fatou's lemma, for all distribution $P \in \mathcal{P}$,

$$\liminf_{(x',\epsilon)\to(x_0,0)} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)] \geq \mathbb{E}_P[\mathbb{1}_{\{g(x_0,\xi)>0\}}(\xi)].$$

Then

$$\sup_{P\in\mathcal{P}} \liminf_{(x',\epsilon)\to(x_0,0)} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)] \geq \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{g(x_0,\xi)>0\}}(\xi)].$$

Moreover, for any $P \in \mathcal{P}$,

$$\liminf_{(x',\epsilon)\to(x_0,0)} \sup_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)] \geq \liminf_{(x',\epsilon)\to(x_0,0)} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)],$$

and then

$$\liminf_{(x',\epsilon)\to(x_0,0)} \sup_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)] \geq \sup_{P\in\mathcal{P}} \liminf_{(x',\epsilon)\to(x_0,0)} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)],$$

which implies (3). □

Let $\mathcal{F} := \{x : \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{g(x,\xi)>0\}}(\xi)] \leq \alpha\}, \mathcal{F}^\epsilon := \{x : \sup_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(x,\xi),\epsilon)] \leq \alpha\}$, $\mathcal{S}$ and $\mathcal{S}^\epsilon$ be the solution sets of problem (1) and (2) respectively.

PROPOSITION 4.2. *Consider $\mathcal{F}^\epsilon$ and $\mathcal{F}$, we have*

$$\limsup_{\epsilon\downarrow 0} \mathcal{F}^\epsilon \subseteq \mathcal{F} \subseteq \liminf_{\epsilon\downarrow 0} \mathcal{F}^\epsilon. \tag{4}$$

*Proof.* First, we prove the left part of (4) $\limsup_{\epsilon\downarrow 0} \mathcal{F}^\epsilon \subseteq \mathcal{F}$. For any $x_0 \in \limsup_{\epsilon\downarrow 0} \mathcal{F}^\epsilon$, there exists a sequence $(x',\epsilon) \to (x_0,0)$ such that

$$\liminf_{(x',\epsilon)\to(x_0,0)} \sup_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(x',\xi),\epsilon)] \leq \alpha.$$

By Proposition 4.1, it is obvious that the above inequality implies

$$\inf_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{g(x_0,\xi)<0\}}(\xi)] \leq \alpha.$$

Then we have the left part of (4).

Moreover, by Lemma 4.1 (2) and (3), for any point $\bar{x} \in \mathcal{F}$ and any $\epsilon > 0$,

$$\inf_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(\bar{x},\xi),\epsilon)] \leq \inf_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{g(\bar{x},\xi)<0\}}(\xi)] \leq \alpha,$$

which implies $\bar{x} \in \mathcal{F}^\epsilon$ and $\mathcal{F} \subseteq \mathcal{F}^\epsilon$ for all $\epsilon > 0$. Then we have $\mathcal{F} \subseteq \liminf_{\epsilon\to 0} \mathcal{F}^\epsilon$, the right part of (4). □

THEOREM 4.1. *Let $\{x^\epsilon\}$ be any sequence of optimal solutions of problem (2) and $x^*$ be the cluster point of the sequence. Then $x^*$ is the optimal solution of problem (1). Moreover,* $\limsup_{\epsilon\to 0} \mathcal{S}^\epsilon \subseteq \mathcal{S}$.

*Proof.* By Proposition 4.2, $x^* \in \mathcal{F}$. Assume for a contradiction that $x^*$ is not the optimal solution of problem (1), then there exists $\bar{x} \in \mathcal{F}$ such that

$$f(\bar{x}) < f(x^*) \ \text{ and } \ \sup_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{g(\bar{x},\xi)>0\}}(\xi)] \leq \alpha.$$

By Lemma 4.1 (2) and (3), for any $\epsilon > 0$,

$$\mathbb{E}_P[\Psi^{DC}(g(\bar{x},\xi),\epsilon)] \leq \mathbb{E}_P[\mathbb{1}_{\{g(\bar{x},\xi)>0\}}(\xi)],$$

which implies

$$\inf_{P\in\mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(\bar{x},\xi),\epsilon)] \leq \inf_{P\in\mathcal{P}} \mathbb{E}_P[\mathbb{1}_{\{g(\bar{x},\xi)<0\}}(\xi)] \leq \alpha,$$

$\bar{x} \in \mathcal{F}^\epsilon$ and $f(\bar{x}) \geq f(x^\epsilon)$. Note that by the continuity of $f$, $f(x^\epsilon) \to f(x^*)$ as $\epsilon \downarrow 0$, then we have

$$f(\bar{x}) - f(x^*) = \lim_{\epsilon \to 0} \left((f(\bar{x}) - f(x^\epsilon)) + (f(x^\epsilon) - f(x^*))\right) \geq 0,$$

a contradiction. ∎

*Remark* 5. In [15], the DC approximation function is

$$\Psi(\cdot, \epsilon) := \frac{1}{\epsilon}\left((\cdot + \epsilon)_+ - (\cdot)_+\right), \tag{5}$$

which is different from $\Psi^{DC}$. From Fig 1- Fig 2, it is easy to find the difference between the two DC approximation functions. The advantage of (1) is it is a conservative approximation of the corresponding probability function. But it can not satisfy (1) and (3) in Lemma 4.1 and, when (1) is replaced by (5), the convergence results in Theorem 4.1 may not hold for problem (2) without continuity condition (4).

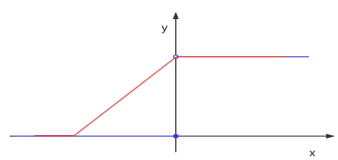


FIG. 1. *DC approximation function in [15]*

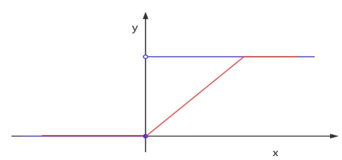


FIG. 2. *DC approximation function* (1)

**4.2. Convergence analysis in data-driven approximation problem of (2).** In this section, we consider the case when the real ambiguity set $\mathcal{P}$ is unknown, but we can construct an approximate set $\mathcal{P}_N$ from empirical data. Then we can construct data-driven approximation problem of (2) as follows:

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ s.t. \quad & \sup_{P \in \mathcal{P}_N} \mathbb{E}_P[\Psi^{DC}(g(x,\xi), \epsilon)] \leq \alpha, \end{aligned} \tag{6}$$

where $\mathcal{P}_N \subset \mathscr{P}$ is a set of probability measures which approximate $\mathcal{P}$ in some sense as $N \to \infty$. To ease the exposition, for each fixed $x \in X$, let

$$v_N^\epsilon(x) := \sup_{P \in \mathcal{P}_N} \mathbb{E}_P[\Psi^{DC}(g(x,\xi), \epsilon)],$$

and

$$\Phi_N^\epsilon(x) := \{P \in \mathrm{cl}\mathcal{P}_N : v_N^\epsilon(x) = \mathbb{E}_P[\Psi^{DC}(g(x,\xi), \epsilon)]\},$$

denote the optimal value of the inner maximization problem and the corresponding set of optimal solutions respectively, where "cl" denotes the closure of a set and the closure is defined in the sense of weak topology, see Definition 3. Likewise, we denote

$$v^\epsilon(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[\Psi^{DC}(g(x,\xi), \epsilon)], \; \Phi^\epsilon(x) := \{P \in \mathrm{cl}\mathcal{P} : v^\epsilon(x) = \mathbb{E}_P[\Psi^{DC}(g(x,\xi), \epsilon)]\}.$$

Consequently, we can write (2) and (6) respectively as

$$\vartheta^\epsilon := \min_{x \in X} \{f(x) \text{ s.t. } v^\epsilon(x) \leq \alpha\}$$

and

$$\vartheta_N^\epsilon := \min_{x\in X} \{f(x) \text{ s.t. } v_N^\epsilon(x) \le \alpha\},$$

where $\vartheta^\epsilon$ and $\vartheta_N^\epsilon$ denote the optimal values, $\mathcal{F}^\epsilon$ and $\mathcal{F}_N^\epsilon$ the sets of feasible solutions, and $\mathcal{S}^\epsilon$ and $\mathcal{S}_N^\epsilon$ the sets of optimal solutions of (2) and (6) respectively. Then we investigate the properties of $v_N^\epsilon$ and $v^\epsilon$ for any given $\epsilon > 0$.

ASSUMPTION 4.1. *Let $\Psi^{DC}(\cdot,\xi)$ be defined as in Definition 4.1, and $\hat{\mathcal{P}}$ be a set of probability measures satisfying (4), (a) for each fixed $\xi \in \Xi$, $g(\cdot,\xi)$ is Lipschitz continuous on $X$ with Lipschitz modulus being bounded by $\kappa(\xi)$, where $\sup_{P\in\hat{\mathcal{P}}} \mathbb{E}_P[\kappa(\xi)] < \infty$, (b) there exists $x_0 \in X$ such that $\sup_{P\in\hat{\mathcal{P}}} \|\mathbb{E}_P[g(x_0,\xi)]\| < \infty$.*

Assumption 4.1 guarantees that, for any $\epsilon > 0$, $\Psi^{DC}(g(\cdot,\xi),\epsilon)$ is Lipschitz continuous on $X$ with Lipschitz modulus being bounded by $\kappa_\epsilon(\xi)$, where $\sup_{P\in\hat{\mathcal{P}}} \mathbb{E}_P[\kappa_\epsilon(\xi)] < \infty$. And there exists $x_0 \in X$ such that $\sup_{P\in\hat{\mathcal{P}}} \|\mathbb{E}_P[\Psi^{DC}(g(x_0,\xi),\epsilon]\| < \infty$.

ASSUMPTION 4.2. *Let $\mathcal{P}, \mathcal{P}_N$ be nonempty and defined as in (2) and (6) respectively. Then (a) there exists a weakly compact set $\hat{\mathcal{P}} \subset \mathscr{P}$ such that (4) holds; (b) $\mathbb{H}(\mathcal{P}_N, \mathcal{P}; \mathsf{dl}_{\mathscr{G}^\epsilon}) \to 0$ almost surely as $N \to \infty$, where $\mathbb{H}(\cdot,\cdot;\mathsf{dl}_{\mathscr{G}^\epsilon})$ is defined as in Section 2.2 with $\mathscr{G}^\epsilon := \{l(\cdot) = \Psi^{DC}(g(x,\xi),\epsilon) : x \in X\}$.*

Assumption 4.2 is corresponding to Assumption 3.1 in Section 3 and the assumption can also hold in several cases, e.g., when $\mathbb{H}(\mathcal{P}_N, \mathcal{P}; \mathsf{dl}_K) \to 0$, see [4, 20, 26].

By [35, Theorem 1], we can investigate the convergence analysis of $\vartheta_N^\epsilon$ to $\vartheta^\epsilon$ and $\mathcal{S}_N^\epsilon$ to $\mathcal{S}^\epsilon$ as $N \to \infty$ for any $\epsilon > 0$.

PROPOSITION 4.3. *Suppose Assumption 4.1-4.2 hold. Then $v_N^\epsilon(x)$ converges uniformly to $v^\epsilon(x)$ over $X$ as $N \to \infty$, that is, $\lim_{N\to\infty} \sup_{x\in X} |v_N^\epsilon(x) - v^\epsilon(x)| = 0$.*

The result is direct corollary of [35, Theorem 1].

THEOREM 4.2. *Assume that $X$ is a compact set, $\mathcal{S}^\epsilon$ and $\mathcal{S}_N^\epsilon$ are nonempty for any $\epsilon > 0$ sufficiently small and $N$ sufficiently large. Suppose (a) Assumptions 4.1-4.2 hold; (b) $cl\mathcal{F}_s^\epsilon \cap \mathcal{S}^\epsilon \ne \emptyset$, where $\mathcal{F}_s^\epsilon := \{x \in X, v^\epsilon(x) < \alpha\}$. Then we have (i). $\lim_{N\to\infty} \mathbb{D}(\mathcal{F}_N^\epsilon, \mathcal{F}^\epsilon) = 0$; (ii). $\lim_{N\to\infty} \vartheta_N^\epsilon = \vartheta^\epsilon$; (iii). $\lim_{N\to\infty} \mathbb{D}(\mathcal{S}_N^\epsilon, \mathcal{S}^\epsilon) = 0$.*

Condition (b) is similar as Condition (b) in Theorem 3.1.

*Proof.* By [35, Proposition 2], $v^\epsilon(\cdot)$ and $v_N^\epsilon(\cdot)$ are continuous. By Proposition 4.3, the rest of the proof is similar to [11, Theorem 3.4], we omit the details. ∎

Then we can give the convergence analysis between (1) and (6). Let $\mathcal{F}^*$, $\mathcal{S}^*$ and $\vartheta^*$ denote the set of the feasible solutions, the optimal solutions and the optimal value of problem (1) respectively.

THEOREM 4.3. *Assume that $\mathcal{S}^*$ is nonempty. Suppose the conditions of Theorem 4.2 hold. Then we have (i) $\lim_{\epsilon\downarrow 0} \lim_{N\to\infty} \mathbb{D}(\mathcal{F}_N^\epsilon, \mathcal{F}) = 0$; (ii) $\lim_{\epsilon\downarrow 0} \lim_{N\to\infty} \vartheta_N^\epsilon = \vartheta^*$; (iii) $\lim_{\epsilon\downarrow 0} \lim_{N\to\infty} \mathbb{D}(\mathcal{S}_N^\epsilon, \mathcal{S}^*) = 0$.*

*Proof.* Combine the results from Theorem 4.1 and 4.2, we have the results. ∎

**5. Numerical formulations and algorithms.** In section 3.2, we prove the convergence analysis between (1) and (2). When (2) is tractable, we can solve (2) directly. But in the case when (2) is intractable, the DC approximation (6) becomes a good choice. In this section, we consider how to solve (6) with three kinds of ambiguity sets.

**5.1. MPDRCC with matrix moment constraints.** Consider the MPDRCC (1) with matrix moment constraints

$$\tilde{\mathcal{P}} := \left\{ P \in \mathscr{P}(\Xi) : \begin{array}{ll} \mathbb{E}_P[\Phi_i(\xi)] = \mu_i, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\Phi_i(\xi)] \preceq \mu_i, & \text{for } i = p+1, \cdots, q, \end{array} \right\}, \tag{1}$$

$\Xi \subset \mathbb{R}^k$ is compact, $\Phi_i$ can be vector valued and/or symmetric matrix valued functions and $\mu_i$ can be vectors and/or symmetric matrices. Then the DC approximation of the problem is problem (2) with ambiguity set $\tilde{\mathcal{P}}$, and its data-driven approxiamtion problem is

$$\begin{array}{ll} \min\limits_{x \in X} & f(x) \\ s.t. & \sup\limits_{P \in \tilde{\mathcal{P}}_N} \mathbb{E}_P[\Psi^{DC}(g(x,\xi), \epsilon)] \le \alpha, \end{array} \tag{2}$$

where

$$\tilde{\mathcal{P}}_N := \left\{ P \in \mathscr{P}(\Xi_N) : \begin{array}{ll} \mathbb{E}_P[\Phi_i(\xi)] = \mu_i^N, & \text{for } i = 1, \cdots, p \\ \mathbb{E}_P[\Phi_i(\xi)] \preceq \mu_i^N, & \text{for } i = p+1, \cdots, q \end{array} \right\},$$

$\Xi_N := \{\xi_1, \cdots, \xi_N\}$ is a discrete subset of $\Xi$ and

$$\beta_N := \max_{\xi \in \Xi} \min_{1 \le i \le N} \|\xi - \xi_i\|$$

such that $\beta_N \to 0$ as $N \to \infty$. Note that by convergence analysis between $\tilde{CP}$ and $\tilde{\mathcal{P}}_N$ in [4], Assumption 4.2 holds. Assume the Slater type condition (STC for short)

$$(1, 0) \in \text{int}\{(\mathbb{E}_P[1], \mathbb{E}_P[\Phi(\xi)]) + \{0_{p+1}\} \times \mathcal{K} : P \in \mathscr{M}_+\},$$

holds, where $\mathcal{K} := \mathcal{S}_+^{n_{p+1} \times n_{p+1}} \times \cdots \times \mathcal{S}_+^{n_q \times n_q}$; see [33, condition (3.12)] for general moment problems. Then the constraints of problem (2) can be reformulated as its dual form:

$$\min_{(\lambda_0, \Lambda) \in \bar{\Lambda}} \sup_{P \in \mathscr{P}(\Xi_N)} \mathbb{E}_P \left[ \Psi^{DC}(g(x,\xi), \epsilon) - \lambda_0 - \sum_{i=1}^q \langle \Lambda_i, \Phi_i(\xi) \rangle \right] + \lambda_0 + \sum_{i=1}^q \langle \mu_i^N, \Lambda_i \rangle \le \alpha$$

where $\bar{\Lambda} := \{(\lambda_0, \Lambda) : \lambda_0 \in \mathbb{R}, \Lambda_i \in \mathcal{S}^{n_i}, i = 1, \cdots, p, \text{ and } \Lambda_i \succeq 0, \text{ for } i = p + 1, \cdots, q\}$. Consequently, problem (2) can be reformulated as

$$\begin{array}{ll} \min\limits_{(x, \lambda_0, \Lambda) \in X \times \bar{\Lambda}} & f(x) \\ \text{s.t.} & \lambda_0 + \sum\limits_{i=1}^q \langle \mu_i^N, \Lambda_i \rangle \le \alpha, \\ & \Psi^{DC}(g(x,\xi), \epsilon) - \lambda_0 - \sum\limits_{i=1}^q \langle \Phi_i(\xi), \Lambda_i \rangle \le 0, \ \forall \xi \in \Xi_N. \end{array} \tag{3}$$

Note that if $g(x,\xi)$ is a convex function, then

$$\Psi^{DC}(g(x,\xi), \epsilon) = \frac{1}{\epsilon}((g(x,\xi))_+ - (g(x,\xi) - \epsilon)_+) = \Psi_1^{DC}(g(x,\xi), \epsilon) - \Psi_2^{DC}(g(x,\xi), \epsilon)$$

is a DC function, where

$$\Psi_1^{DC}(g(x,\xi), \epsilon) := \frac{1}{\epsilon}(g(x,\xi))_+ \ \text{ and } \ \Psi_2^{DC}(g(x,\xi), \epsilon) := \frac{1}{\epsilon}(g(x,\xi) - \epsilon)_+$$

are convex functions. Moreover, (3) is a DC constrained minimization problem and can be solved by the penalty and augmented Lagrangian method proposed in [21], see Section 5.4 for details.

**5.2. MPDRCC with ball constraints based on wasserstein distance.** In this subsection, we consider how to approximate MPDRCC (1) with ball constraints based on wasserstein distance: $\mathbb{B}_\delta(P_0) := \{P \in \mathscr{P}(\Xi) : \mathbb{D}(P_0, P; \mathsf{dl}_K) \leq \delta\}$.

By the DC approximation method proposed in Section 4, the MPDRCC (1) with ball constraints based on wasserstein distance can be approximated by

$$\begin{array}{ll} \min_{x\in X} & f(x) \\ s.t. & \sup_{P\in\mathbb{B}_\delta(P_0)} \mathbb{E}_P[\Psi^{DC}(g(x,\xi),\epsilon)] \leq \alpha. \end{array} \tag{4}$$

Note that since $P_0$ is the center distribution of the random variable and might be continuous distribution or unknown distribution, problem (4) is still not easy to solve. In the rest of this subsection, we consider a data-driven approximation to problem (4).

We use an emprical distribution $P_N = \frac{1}{N}\sum_{i=1}^N \delta_{\hat{\xi}^i}$ for $\hat{\xi}^i \in \Xi_N$ and $|\Xi_N| = N$ (constructed by historical data, Monte Carlo method, Quasi Monte Carlo method, etc.) to approximate $P_0$, and for any $\epsilon > 0$ and $\varepsilon > 0$, we can find $N_0$ such that for all $N \geq N_0$, $\text{Prob}\{\mathsf{dl}_K(P_0, P_N) \leq \epsilon\} \geq 1-\varepsilon$ and $\text{Prob}\{\beta_N \leq \epsilon\} \geq 1-\varepsilon$ ($\beta_N$ is defined in Section 5.1. Then the DC approximated data-driven MPDRCC is:

$$\begin{array}{ll} \min_{x\in X} & f(x) \\ s.t. & \sup_{P\in\mathbb{B}_\delta(P_N)} \mathbb{E}_P[\Psi^{DC}(g(x,\xi),\epsilon)] \leq \alpha, \end{array} \tag{5}$$

where $\mathbb{B}_\delta(P_N) := \{P \in \mathscr{P}(\Xi_N) : \mathsf{dl}_K(P_N, P) \leq \delta\}$ and $\Xi_N$ is the support set of $P_N$. Note since that $\mathbb{H}(\mathbb{B}_\delta(P), \mathbb{B}_\delta(P_N); \mathsf{dl}_K) \to 0$ as $\mathsf{dl}_K(P, P_N) \to 0$ and $\beta_N \to 0$, Assumption 4.2 holds, see [4].

By [8, Theorem 1], problem (5) equivalent to

$$\begin{array}{ll} \min_{x\in X,\lambda\geq 0} & f(x) \\ s.t. & \lambda\delta - \dfrac{1}{N}\displaystyle\sum_{i=1}^N (\lambda d(\xi^i, \hat{\xi}^i) - \Psi^{DC}(g(x,\xi^i),\epsilon)) \leq \alpha, \quad \forall \zeta \in \Pi_{i=1}^N \Xi_N^i, \end{array} \tag{6}$$

where $\Xi_N^i = \Xi_N$ and $\zeta = (\xi^1, \cdots, \xi^N)$.

Note that (6) is also a DC constrained minimization problem and can be solved by the penalty and augmented Lagrangian method proposed in [21] (see Section 5.4).

**5.3. MPDRCC with ambiguity set constructed through KL-divergence.** We also consider the MPDRCC (1) with ambiguity set constructed through KL-divergence. KL-divergence is introduced by Kullback and Leibler [19]. Let $p_0$ and $p$ denote the density functions of true probability measure $P_0$ and its perturbation $P$ respectively. Then KL-divergence can be used to measure the deviation of $p$ from $p_0$ as $\mathbb{D}_{KL}(P\|P_0) = \int_\Xi p(\xi)\log\left(\frac{p(\xi)}{p_0(\xi)}\right)d\xi$, and the ambiguity set can be constructed as

$$\mathcal{P}_{KL}^\eta(P_0) := \{P \in \mathscr{P} : \mathbb{D}_{KL}(P\|P_0) \leq \eta\}. \tag{7}$$

When $P_0$ is a discrete distribution, we understand $p_0(\xi)$ is the probability mass function and the integral as the summation.

By [16, 18], let $\bar{\alpha} := \sup_{t>0} \frac{e^{-\eta}(t+1)^\alpha - 1}{t}$, MPDRCC (1) with the ambiguity set (7) can be written as

$$\begin{array}{ll} \min_{x\in X} & f(x) \\ s.t. & \mathbb{E}_{P_0}[\mathbb{1}_{\{g(x,\xi)>0\}}(\xi)] \leq \bar{\alpha}. \end{array} \tag{8}$$

By the DC approximation method proposed in Section 4, the chance constraint of MPDRCC (8) can be approximated by $\mathbb{E}_{P_0}[\Psi^{DC}(g(x,\xi),\epsilon)] \le \bar{\alpha}$, and we can use sample average approximation (SAA) method to approximate the DC approximated (8) as follows:

$$\begin{array}{ll} \min\limits_{x\in X} & f(x) \\ s.t. & \frac{1}{N}\sum_{i=1}^{N}\Psi^{DC}(g(x,\xi^i),\epsilon) \le \bar{\alpha}, \end{array} \tag{9}$$

where $\{\xi^i\}_{i=1}^N$ are i.i.d. samples of $\xi$. Problem (9) is still a DC constrained problem and can be solved by the penalty and augmented Lagrangian method proposed by [21] (see Section 5.4). Note also that the convergence analysis of SAA problem (9) has been investigated in [36].

**5.4. Algorithm for DC constrained DC minimization.** In [21], Lu et al. propose a penalty and augmented Lagrangian method for solving DC constrained DC minimization. We will apply their method to solve our problems. For the complement of the paper, we write their algorithm as follows.

Consider DC constrained DC minimization

$$\begin{array}{ll} \min & \phi_0(x) - \psi_0(x) \\ \text{s.t.} & \phi_i(x) - \psi_i(x) \le 0, \ \forall i = 1, \cdots, l, \\ & x \in X, \end{array} \tag{10}$$

where $X \subseteq \mathbb{R}^n$ is a closed convex set, $\psi_i(x) = \max_{j\in\mathcal{J}_i}\psi_{i,j}(x)$, $\mathcal{J}_i = \{1, 2, \cdots, J_i\}$ for $i = 0, 1, \cdots, l$, $\phi_i$, $\psi_{i,j}$ are continuously differentiable and convex functions.

Given $\rho \ge 1$, define the penalty function:

$$F_\rho(x) = \phi_0(x) - \psi_0(x) + \rho\sum_{i=1}^{l}[\phi_i(x) - \psi_i(x)]_+.$$

Moreover, given $\bar{x} \in X$ and $\epsilon \ge 0$, we define

$$\mathcal{J}(\bar{x},\epsilon) = \{(j_0, j_1, \ldots, j_l) \,|\, j_i \in \mathcal{J}_i, \psi_{i,j_i}(\bar{x}) \ge \psi_i(\bar{x}) - \epsilon, \forall i = 0, 1, \ldots, l\}.$$

By choosing $\mathbb{J} = (j_0, j_1, \cdots, j_l) \in \mathcal{J}(\bar{x},\epsilon)$, we define

$$Q_\rho(x;\bar{x},\mathbb{J}) = \phi_0(x) - \psi_{0,j_0}(\bar{x}) - \nabla\psi_{0,j_0}(\bar{x})^\top(x-\bar{x}) + \rho\sum_{i=1}^{l}\left[\phi_i(x) - \psi_{i,j_i}(\bar{x}) - \nabla\psi_{i,j_i}(\bar{x})^\top(x-\bar{x})\right]_+.$$

Clearly, $Q_\rho(x;\bar{x},\mathbb{J})$ is a convex function. The penalty method for solving (10) is presented in Algorithm 1.

*Remark* 6. In practice, since $F_{\rho_k}(x) \le Q_{\rho_k}(x;\bar{x},\mathbb{J})$ for all $\bar{x} \in X$ and $Q_{\rho_k}(x;\bar{x},\mathbb{J})$ is a convex function, we only need to solve $\min_{x\in X} Q_{\rho_k}(x;\bar{x},\mathbb{J})$ for properly choosing $\bar{x} \in X$ and $\mathbb{J} \in \mathcal{J}(x^k,\epsilon)$. See [21] for more details.

**6. Numerical Experiments.** In this section, we show three numerical examples. The first and second examples are used to verify the convergence results in Section 3 and 4. The third example is used to test effectiveness of numerical formulations and algorithm in Section 5. In the numerical tests, we only focus on the case when ambiguity sets are constructed by moment information.

**Algorithm 1** A penalty method for solving DC minimization (10).

Choose $\epsilon > 0$, $\rho_0 \geq 1$, $\sigma > 1$, and a positive sequence $\eta_k \to 0$ as $k \to \infty$. Set $k \leftarrow 0$.
**while** not converged **do**
  Find an approximate solution $x^k$ of the penalty subproblem

$$\min_{x \in X} F_{\rho_k}(x)$$

  such that $x^k \in X$ and $F_{\rho_k}(x^k) \leq Q_{\rho_k}(x; x^k, \mathbb{J}) + \eta_k, \forall x \in X, \forall \mathbb{J} \in \mathcal{J}(x^k, \epsilon)$.
  Set $\rho_{k+1} \leftarrow \sigma \rho_k$ and $k \leftarrow k+1$.
**end while**

Example 6.1 (Academic example 1). *To verify Theorem 3.1, we design this experiment. Let $\vartheta$ and $\mathcal{S}$ be the optimal value and the optimal solution set of the program* (1) *respectively with ambiguity set*

$$\mathcal{P} = \{P \in \mathscr{P}(\mathbb{R}) : \mathbb{E}_P[\xi] = \mu, \mathbb{E}_P[\xi\xi^T] = \Sigma + \mu\mu^T\}.$$

*We consider a constraint function $g(x,\xi) = y^0(x) + y(x)^\top \xi$. Moreover, let $\vartheta_N$ and $\mathcal{S}_N$ be the optimal value and the optimal solution set of the data-driven approximation problem* (2) *with*

$$\mathcal{P}_N = \{P \in \mathscr{P}(\mathbb{R}) : \mathbb{E}_P[\xi] = \mu_N, \mathbb{E}_P[\xi\xi^T] = \Sigma_N + \mu_N\mu_N^T\}.$$

*According to [45], since $g(x,\xi) = y^0(x) + y(x)^\top \xi$ is linear in $\xi$, program* (1) *and* (2) *could be represented as tractable SDP formulas*

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } M \succeq 0, \quad \beta + \frac{1}{\alpha}\langle \Omega, M\rangle \leq 0, \\ &\quad M - \begin{bmatrix} 0 & \frac{1}{2}y(x) \\ \frac{1}{2}y(x)^\top & y^0(x) - \beta \end{bmatrix} \succeq 0, \\ &\quad x \in X, \quad \beta \in \mathbb{R}, \quad M \in \mathbb{S}^{n+1}, \end{aligned} \quad \text{and} \quad \begin{aligned} &\min\ f(x) \\ &\text{s.t. } M \succeq 0, \quad \beta + \frac{1}{\alpha}\langle \Omega_N, M\rangle \leq 0, \\ &\quad M - \begin{bmatrix} 0 & \frac{1}{2}y(x) \\ \frac{1}{2}y(x)^\top & y^0(x) - \beta \end{bmatrix} \succeq 0, \\ &\quad x \in X, \quad \beta \in \mathbb{R}, \quad M \in \mathbb{S}^{n+1}, \end{aligned}$$

*where* $\Omega = \begin{bmatrix} \Sigma + \mu\mu^\top & \mu \\ \mu^\top & 1 \end{bmatrix}$, and $\Omega_N = \begin{bmatrix} \Sigma_N + \mu_N\mu_N^\top & \mu_N \\ \mu_N^\top & 1 \end{bmatrix}$, $\mu, \mu_N \in \mathbb{R}^n$ *and* $\Sigma, \Sigma_N \in \mathbb{S}^n$.

*Consider a portfolio optimization problem over n products. The portfolio $x \in \mathbb{R}^n_+$ satisfies $\sum_i x_i = 1$. Our objective is the expectation of the difference of the total input and the real outcome $\xi^\top x$: $f(x) = 1 - \mu^\top x$, where $\xi \sim \mathcal{N}(\mu, \Sigma)$ is the rate of return. Let $g(x,\xi) = \eta - x^\top \xi$. Then the chance constraint is*

$$\inf_{P \in \mathcal{P}_N} P[\eta - x^\top \xi \leq 0] \geq 1 - \alpha.$$

*In our experiment, we set* $n = 10$, $\alpha = 0.05$, $\eta = 0.1$, $\mu = (0.6, 0.7, \ldots, 1.5)^\top$, *and*

$$\Sigma = \begin{pmatrix} 0.4 & -0.01 & 0 & 0.01 & 0 & -0.01 & 0.01 & 0.01 & 0 & -0.01 \\ -0.01 & 0.4 & -0.01 & -0.01 & 0 & 0.01 & -0.01 & -0.01 & 0.01 & -0.01 \\ 0 & -0.01 & 0.4 & -0.01 & 0.01 & 0 & 0 & 0.01 & 0 & 0 \\ 0.01 & -0.01 & -0.01 & 0.4 & 0.01 & -0.01 & 0.01 & -0.01 & 0 & 0.01 \\ 0 & 0 & 0.01 & 0.01 & 0.4 & -0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ -0.01 & 0.01 & 0 & -0.01 & -0.01 & 0.4 & 0.01 & 0 & -0.01 & -0.01 \\ 0.01 & -0.01 & 0 & 0.01 & 0.01 & 0.01 & 0.4 & -0.01 & 0.01 & 0 \\ 0.01 & -0.01 & 0.01 & -0.01 & 0.01 & 0 & -0.01 & 0.4 & 0.01 & 0 \\ 0 & 0.01 & 0 & 0 & 0.01 & -0.01 & 0.01 & 0.01 & 0.4 & -0.01 \\ -0.01 & -0.01 & 0 & 0.01 & 0.01 & -0.01 & 0 & 0 & -0.01 & 0.4 \end{pmatrix}.$$

*By experiments, the optimal value is* $\vartheta^* = -0.3243$, *which is illustrated as a green line in Figure 3. Associated optimal solution is marked by five-pointed stars in the right column of Figure 4.*

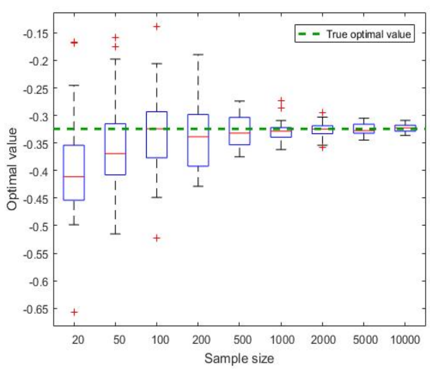


FIG. 3. *Optimal value v.s. the sample size.*

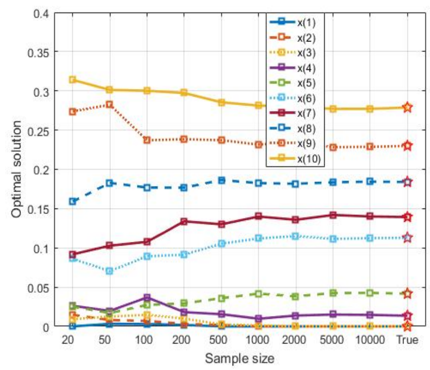


FIG. 4. *Optimal solution v.s. the sample size.*

*Whereafter, we sample random variables* $\xi$ *enjoying distribution* $\mathcal{N}(\mu, \Sigma)$ *with sample size* $N = 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000$, *respectively. For each* $N$, *we perform* 30 *tests. In each test, by estimating mean* $\mu_N$ *and covariance* $\Sigma_N$ *from associated sample, we compute estimated optimal value* $\vartheta_N$ *and estimated optimal solution* $x_N$. *By this means, for a fix sample size* $N$, *there are thirty optimal values, whose median is illustrated as a red short line in Figure 3. The top and button edges of the blue box are respectively the 25th and 75th percentiles of these estimated optimal values corresponding to a fixed* $N$, *and the whiskers extend to the most extreme estimated optimal values. Clearly, as the sample size* $N$ *increases, we see that the blue box shrinks and converges to the green line. Hence, we claim that the optimal value* $\vartheta_N$ *of data driven problems tends to* $\vartheta^*$ *as* $N \to \infty$.

*For each* $N$, *we also count the mean of estimated optimal solutions and illustrated components of the mean as squares in Figure 4. For each component, we also draw a line connecting estimated solutions to the true one. By Figure 4, we see these lines tend to the horizontal direction smoothly as* $N$ *increases. Therefore, the optimal solution set* $\mathcal{S}_N$ *of the data driven problems converges to the true solution set* $\mathcal{S}^*$ *as the sample size* $N$ *tends to infinity.*

EXAMPLE 6.2 (Academic example 2). *Our second example is used to verify the convergence analysis proposed in Section 4.2. We consider problem* (1) *with ambiguity set*

$$\mathcal{P} = \{P \in \mathscr{P}(\Xi) : \mathbb{E}_P[\xi] = \mu, \mathbb{E}_P[\xi\xi^T] = \Sigma + \mu\mu^T\}.$$

*By using the DC approximate scheme (1) with data-driven approximation, we construct data driven approximated DC approximation problem (6) with approximation parameter $\epsilon$ and the set of uncertain distributions characterized by statistics from historical samples:*

$$\mathcal{P}_N = \{P \in \mathscr{P}(\Xi_N) : \mathbb{E}_P[\xi] = \mu, \mathbb{E}_P[\xi\xi^T] = \Sigma + \mu\mu^T\},$$

*where $\Xi_N$ is a discrete approximation of $\Xi$ with $|\Xi_N| = N$ such that $\Xi_{N_1} \subset \Xi_{N_2} \subset \Xi$ for any $0 \le N_1 \le N_2$ and $\beta_N \to 0$ as $N \to \infty$. By [4, Theorem 3.1], $\mathbb{H}(\mathcal{P}, \mathcal{P}_N; \mathsf{dl}_{\mathscr{G}^\epsilon}) \to 0$ as $N \to \infty$.*

*In this example, we consider a portfolio $x \in [0,1]$ and a risk-free investment $(1-x)$. Parameters are set as $\mu = 1.1$, $\Sigma = 0.05$, $\Xi = [-50, 50]$, $\kappa = 2$, $y^0(x) = \frac{1}{3}$, and $\alpha = 0.1$, $y(x) = -x$, i.e., $g(x,\xi) = \frac{1}{3} - x^\top\xi$. We divide the support set $\Xi = [-50, 50]$ into $(N-1)$ equal parts and hence have $N$ end points of all small intervals. We solve problem (6) in this example by Algorithm 1 with different $N$ and $\epsilon$. Let $\vartheta_N^\epsilon$ denote the optimal value of (6) in this example with corresponding $\epsilon$ and $N$, the numerical results are shown as in Table 1. For a given $N$ and $\epsilon$, the first number in the associated cell is the optimal value $\vartheta_N^\epsilon$ of (6) in this example, and the vector in the second line of the cell is the optimal solution $(x_N^\epsilon, 1 - x_N^\epsilon)$.*

TABLE 1
*Optimal values and optimal solutions*

| Solving (6) in this example | $\epsilon = 0.01$ | $\epsilon = 0.001$ | $\epsilon = 0.0001$ |
|---|---|---|---|
| $N = 201$ | $-1.0250$<br>$(0.5000, 0.5000)^\top$ | $-1.0250$<br>$(0.5000, 0.5000)^\top$ | $-1.0250$<br>$(0.5000, 0.5000)^\top$ |
| $N = 501$ | $-1.0250$<br>$(0.5000, 0.5000)^\top$ | $-1.0250$<br>$(0.5000, 0.5000)^\top$ | $-1.0247$<br>$(0.5555, 0.4445)^\top$ |
| $N = 1001$ | $-1.0250$<br>$(0.5000, 0.5000)^\top$ | $-1.0247$<br>$(0.5543, 0.4457)^\top$ | $-1.0222$<br>$(0.6665, 0.3335)^\top$ |
| $N = 2001$ | $-1.0250$<br>$(0.5013, 0.4987)^\top$ | $-1.0239$<br>$(0.6045, 0.3955)^\top$ | $-1.0192$<br>$(0.7405, 0.2595)^\top$ |
| SDP upper bound: | $\bar\vartheta = -1.0173$ $(\bar{x}, 1-\bar{x})^\top = (0.7767, 0.2233)^\top$ | | |

*Let $\vartheta_\Xi$ be the optimal value of (1) in this example. Note that since*

$$\Psi^{DC}(g(x,\xi), \epsilon_1) \le \Psi^{DC}(g(x,\xi), \epsilon_2) \le \mathbb{1}_{\{g(x,\xi)>0\}}(g(x,\xi))$$

*for any $x \in X$, $\epsilon_1 \ge \epsilon_2 > 0$ and $\Xi_{N_1} \subset \Xi_{N_2} \subset \Xi$ for any $0 \le N_1 \le N_2$, then the feasible set of problem (6) in this example shrinks with $\epsilon \downarrow 0$ and $N \to \infty$, and we have $\vartheta_{N_1}^{\epsilon_1} \le \vartheta_{N_2}^{\epsilon_2} \le \vartheta_\Xi$. Moreover, let $\bar\vartheta$ denote the optimal value of problem (1) with ambiguity set*

$$\bar{\mathcal{P}} = \{P \in \mathscr{P}(\mathbb{R}) : \mathbb{E}_P[\xi] = \mu, \mathbb{E}_P[\xi\xi^T] = \Sigma + \mu\mu^T\}.$$

*Note that in this case, $\mathcal{P} \subset \bar{\mathcal{P}}$ and problem (1) can be solved by the SDP reformulation in [45] effectively. Then we have $\bar\vartheta \ge \vartheta_\Xi \ge \vartheta_{N_2}^{\epsilon_2} \ge \vartheta_{N_1}^{\epsilon_1}$. The optimal value $\bar\vartheta$ and the associated optimal solution are addressed in the last line of Table 1 as the "SDP upper bound". However, when $\epsilon$ sufficiently large and $N$ sufficiently large, we can see from the table that $\|(\bar\vartheta - \vartheta_N^\epsilon)\|$ decreases and becomes very small, note that $\vartheta_\Xi$ is between $\bar\vartheta$ and $\vartheta_N^\epsilon$, which shows the convergence result. Note also that when $\epsilon$ is not sufficiently large and $N$ is not sufficiently large, the feasible set of problem (6) in this example* *may be too large such that $(0.5, 0.5)^T$ is included in it, then the constraint is inactive and the optimal solution and optimal value are located on $(0.5, 0.5)^T$ and $-1.025$, see Table 1 for details.*

EXAMPLE 6.3. *In the numerical tests, we consider a portfolio optimization problem based on the data set (historical return rates) of the stocks in the NASDAQ-100 index (between January 2013 and January 2017). There are 100 stocks with 1000 historical return rates, i.e., $n = 100$ and the sample size $N = 1000$.*

*The portfolio optimization problem is constructed over a set of stocks $\{1, 2, \ldots, n(\leq 100)\}$ where we index the stocks in the subset of NASDAQ-100 index by $i = 1, 2, \ldots, n$. The loss function to the investor is the gap between the anticipated return $\eta_0 = 1$ and the real outcome $r^\top(\xi)x$ where $x$ is the normalized portfolio (i.e., $x := (x_1, \cdots, x_n)$, $x_i \geq 0$ for $i = 1, 2, \ldots, n$, and $x_1 + \cdots + x_n = 1$) into the stocks. Hence we have that the objective function in (6) being $f(x) = 1 - \frac{1}{N}\sum_{j=1}^N r(\xi^j)^\top x$. Whereafter, we define $g(x, \xi)$ by a nonlinear form $g(x, \xi) := \exp\left(\mu(\eta_1 - r(\xi)^\top x)\right) - 1$, where $g(x, \xi)$ is a nonlinear function of $x$. In the tests, we set $\mu = 0.1$ and $\eta_1 = 0.9$.*

*Let $\Xi_N := \{\xi^1, \cdots, \xi^N\}$ and $r_i^N(\xi^j)$ be the return rates of stock $i$ at historical trading day $j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, N$. The uncertain distribution set is defined as*

$$\mathcal{P}_N = \left\{ P \in \mathscr{P}(\Xi_N) : \begin{array}{l} \mathbb{E}_P\left[r_i(\xi)\right] = \overline{\mu}_i, \\ \mathbb{E}_P\left[(r_i(\xi) - \mu_i)^2\right] \leq \gamma\overline{\Sigma}_i, \end{array} \quad \forall i = 1, \ldots, n \right\},$$

*where $\gamma = 1.1$, $\overline{\mu}_i = \frac{1}{1000}\sum_{j=1}^{1000} r_i^N(\xi^j)$ and $\overline{\Sigma}_i = \frac{1}{1000}\sum_{j=1}^{1000}\left(r_i^N(\xi^j) - \overline{\mu}_i\right)^2$.*

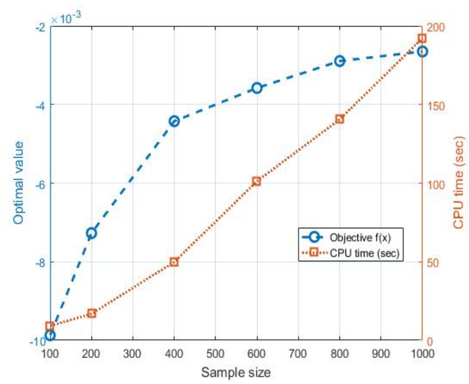


FIG. 5. *CPU time v.s. the sample size*

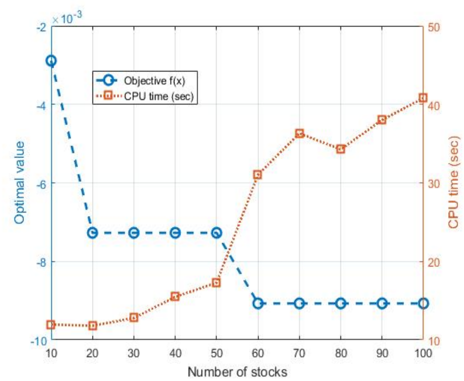


FIG. 6. *CPU time v.s. the number of stocks*

TABLE 2
*Optimal Values vs. Different Parameters*

| | $\alpha = 0.01$ | | | $\epsilon = 0.01$ | | |
|---|---|---|---|---|---|---|
| Parameters | $\epsilon = 0.1$ | $\epsilon = 0.01$ | $\epsilon = 0.001$ | $\alpha = 0.02$ | $\alpha = 0.01$ | $\alpha = 0.005$ |
| $N = 600$ | $-0.003731$ | $-0.003368$ | $-0.003177$ | $-0.003579$ | $-0.003368$ | $-0.003262$ |
| $N = 800$ | $-0.003034$ | $-0.002700$ | $-0.002526$ | $-0.002894$ | $-0.002700$ | $-0.002603$ |
| $N = 1000$ | $-0.002757$ | $-0.002504$ | $-0.002371$ | $-0.002651$ | $-0.002504$ | $-0.002430$ |

Note that although in [42], Xie and Ahmed propose a convex reformulation for DRMPCC with moment constraints when there is only a single uncertain constraint, they require the random function $-g(x, \xi)$ is concave w.r.t. $x$ and convex w.r.t. $\xi$, so their convex reformulation can not cover Example 6.3.

We implement the DC approximation method and Algorithm 1 to solve the problem. We perform the comparative analysis by varying parameter $\alpha$, sample size $N$ and the number of stocks. First we fix the number of stocks being 20, $\alpha = 0.02$ and $\epsilon = 0.01$, and increase the sample size $N$ from 100 to 1000 and record the optimal value and the CPU time used for solving the problem. The result is given in Figure 5. Note that with the increasing of $N$, the ambiguity set is enlarging, which leads to the increasing of the optimal values.

Second we fix the sample size being $N = 200$, $\alpha = 0.02$ and $\epsilon = 0.01$ and increase the number of stocks from 10 to 100 and record the optimal value and the CPU time used for solving the problem. The result is given in Figure 6. Note that with the increasing of number of stocks, the investigator has more choices, and this leads to the decreasing of the optimal values.

In the third set of test, we vary the parameters $\alpha$ from 0.02 to 0.005 and the DC approximate parameters in function $\Psi^{\mathrm{DC}}(\bullet, \xi)$ from 0.1 to 0.001 to verify the optimal values solved from (6) in this example. All other parameters are fixed as previous experiments. Resulting optimal values are reported in Table 2. We can see from Table 2 that with $\epsilon \downarrow 0$ and $\alpha \downarrow 0$, the optimal values increase, which is consistent with the fact that the feasible set of (6) in this example shrinks with $\epsilon \downarrow 0$ and $\alpha \downarrow 0$.

**7. Conclusion.** This paper investigates the convergence of optimal value and the optimal solutions of data-driven approximation of MPDRCC without assuming the continuity of distributional robust probability functions. One important issue to be handled is the lower semi-continuity of the distributionally robust probability functions, which is used to replace the continuity condition of distributionally robust probability functions.

Moreover, for the case when the MPDRCC is intractable, we propose a data-driven approximated DC approximation method for MPDRCC with the corresponding convergence analysis. The approximated problem can be solved by recent advances of DC algorithms effectively. One main advantage of the DC approximation method is the consistency of the feasible sets when the DC parameter $\epsilon \downarrow 0$, which allows us to avoid the continuity condition of distributionally robust probability functions. The preliminary numerical tests verify the convergence analysis and show the effectiveness of the proposed approximation methods.

## REFERENCES

[1] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory*, Springer Science & Business Media, New York, 2006.

[2] G. Calafiore and L. E. Ghaoui, *Distributionally robust chance-constrained linear programs with applications*, Journal of Optimization Theory and Application, 130 (2006), pp. 1–22.

[3] A. Charnes, W. Cooper, and G. H. Symonds, *Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil*, Management Science, 4 (1958), pp. 235–263.

[4] Y. Chen, H. Sun, and H. Xu, *Decomposition methods for solving two-stage distributionally robust optimization*, Optimization-online, (2019).

[5] Z. Chen, D. Kuhn, and W. Wiesemann, *Data-driven chance constrained programs over wasserstein balls*, 2018, https://arxiv.org/abs/1809.00210.

[6] D. Dentcheva and G. Martinez, *Regularization methods for optimization problems with probabilistic constraints*, Mathematical Programming, 138 (2013), pp. 223–251.

[7] E. Erdoğan and G. Iyengar, *Ambiguous chance constrained problems and robust optimization*, Mathematical Programming, 107 (2006), pp. 37–61.

[8] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with wasserstein distance*, 2016, https://arxiv.org/abs/1604.02199.
[9] A. Geletu, A. Hoffmann, M. Kloppel, and P. Li, *An inner-outer approximation approach to chance constrained optimization*, SIAM Journal on Optimization, 27 (2017), pp. 1834–1857.
[10] A. L. Gibbs and F. E. Su, *On choosing and bounding probability metrics*, International statistical review, 70 (2002), pp. 419–435.
[11] S. Guo, H. Xu, and L. Zhang, *Stability analysis for mathematical programs with distributionally robust chance constraint*, SIAM Journal on Optimization, 27 (2017), pp. 784–816.
[12] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, *A distributionally robust perspective on uncertainty quanlification and chance constrained programming*, Mathematical Programming, 151 (2015), pp. 427–439.
[13] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, *Ambiguous joint chance constraints under mean and dispersion information*, Operations Research, 65 (2017), pp. 751–767.
[14] R. Henrion and C. Strugarek, *Convexity of chance constraints with dependent random variables: The use of copulae, in Stochastic Optimization Methods in Finance and Energy: New Financial Products and Energy Market Strategies, M. Bertocchi, G. Consigli, and M. A. H. Dempster, eds.*, Springer, Now York, 2011.
[15] L. J. Hong, Y. Yang, and L. Zhang, *Sequential convex approximations to joint chance constrained programs: a monte carlo approach*, Operations Research, 59 (2011), pp. 617–630.
[16] Z. Hu and L. J. Hong, *Kullback-leibler divergence constrained distributionally robust optimization*, 2013, http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf.
[17] Z. Hu, L. J. Hong, and L. Zhang, *A smooth monte carlo approach to joint chance-constrained programs*, IIE Transactions, 45 (2013), pp. 716–735.
[18] R. Jiang and Y. Guan, *Data-driven chance constrained stochastic program*, Mathematical Programming, 158 (2016), pp. 291–327.
[19] S. Kullback and R. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics, 22 (1951), pp. 79–86.
[20] Y. Liu, A. Pichler, and H. Xu, *Discrete approximation and quantification in distributionally robust optimization*, Mathematics of Operations Research, 44 (2018), pp. 19–37.
[21] Z. Lu, Z. Sun, and Z. Zhou, *Penalty and augmented lagrangian methods for a class of structured nonsmooth dc constrained dc program*, Manuscript, (2019).
[22] J. Luedtke and S. Ahmed, *A sample approximation approach for optimization with probabilistic constraints*, SIAM Journal on Optimization, 19 (2008), pp. 674–699.
[23] L. B. Miller and H. Wagner, *Chance-constrained programming with joint constraints*, Operations Research, 13 (1965), pp. 930–945.
[24] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs*, SIAM Journal on Optimization, 17 (2006), pp. 969–996.
[25] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, *Sample average approximation method for chance constrained programming: Theory and applications*, Journal of Optimization Theory and Applications, 142 (2009), pp. 399–416.
[26] A. Pichler and H. Xu, *Quantitative stability analysis for minimax distributionally robust risk optimization*, Mathematical Programming, (2018), https://link.springer.com/article/10.1007/s10107-018-1347-4.
[27] A. Prékopa, *On probabilistic constrained programming. In: Proceedings of the Princeton Symposium on Mathematical Programming*, Princeton University Press, Princeton, 1970.
[28] A. Prékopa, *Probabilistic programming, in Stochastic Programming, A. Ruszczyński and A. Shapiro, eds.*, Elsevier, Amsterdam, 2003.
[29] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, John Wiley and Sons, West Sussex, England, 1991.
[30] W. Römisch, *Stability of stochastic programming problems, in Stochastic Programming, A. Ruszczyński and A. Shapiro, eds.*, Elsevier, Amsterdam, 2003.
[31] R. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, New York, 3th ed., 2009.
[32] F. Shan, X. T. Xiao, and L. W. Zhang, *Convergence analysis on a smoothing approach to joint chance constrained programs*, Optimization, 65 (2016), pp. 2171–2193.
[33] A. Shapiro, *On duality theory of conic linear problems, in Semi-Infinite Programming: Recent Advances, M. A. Goberna and M. A. Lopez, eds.*, Springer, New York, 2001.
[34] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.
[35] H. Sun and H. Xu, *Convergence analysis for distributionally robust optimization and equilibrium problems*, Mathematics of Operations Research, 41 (2016), pp. 377–401.

[36] H. Sun, H. Xu, and Y. Wang, *Asymptotic analysis of sample average approximation forstochastic optimization problems with joint chance constraints via conditional value at risk and difference of convex functions*, Journal of Optimization Theory and Applications, 161 (2014), pp. 257–284.

[37] W. van Ackooij, *Chance Constrained Programming: With Applications in Energy Management, , Ph.D. thesis,*, Ećole Centrale, Paris, 2013.

[38] W. van Ackooij, *Eventual convexity of chance constrained feasible sets*, Optimization, 64 (2015), pp. 1263–1284.

[39] W. van Ackooij and R. Henrion, *Gradient formulae for nonlinear probabilistic constraints with gaussian and gaussian-like distributions*, SIAM Journal on Optimization, 24 (2014), pp. 1864–1889.

[40] W. Xie, *On distributionally robust chance constrained programs with wasserstein distance*, Optimization-online, (2018), https://arxiv.org/pdf/1806.07418.pdf.

[41] W. Xie and S. Ahmed, *Bicriteria approximation of chance constrained covering problems*, Optimization-online, (2018), http://www.optimization-online.org/DB_FILE/2018/01/6411.pdf.

[42] W. Xie and S. Ahmed, *On deterministic reformulations of distributionally robust joint chance constrained optimization problems*, Mathematical Programminng, 28 (2018), pp. 1151–1182.

[43] W. Yang and H. Xu, *Distributionally robust chance constraints for non-linear uncertainties*, Mathematical Programminng, 155 (2016), pp. 231–265.

[44] V. M. Zolotarev, *Probability metrics*, Teoriya Veroyatnostei i ee Primeneniya, 28 (1983), pp. 264–287.

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