Bicriteria Approximation of Chance Constrained Covering Problems

Weijun Xie
Department of Industrial & Systems Engineering, Virginia Tech, Blacksburg, VA 24061, wxie@vt.edu.

Shabbir Ahmed
School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, sahmed@isye.gatech.edu.

A chance constrained optimization problem involves constraints with random data which can be violated with probability bounded from above by a prespecified small risk parameter. Such constraints are used to model reliability requirements in a variety of application areas such as finance, energy, service and manufacturing. Except under very special conditions, chance constrained problems are extremely difficult. There has been a great deal of elegant work on developing tractable approximations of chance constraints. Unfortunately none of these approaches come with a constant factor approximation guarantee. We show that such a guarantee is impossible by proving an inapproximability result. On the other hand, for a large class of chance constrained covering problems, we propose a *bicriteria* approximation scheme. Our scheme finds a solution whose violation probability may be larger than, but is within a constant factor of, the specified risk parameter and whose objective value is within a constant factor of the true optimal value. Key to our developments is the construction of a tractable convex relaxation of a chance constrained problem and an appropriate scaling of a solution to this relaxation. We extend our approximation results to the setting when the underlying distribution of the constraint data is not known. That is, we consider distributionally robust chance constrained covering problems with convex moment and Wasserstein ambiguity sets, and provide bicriteria approximation results.

Key words: Chance constrained problem; approximation algorithm; distributionally robust optimization; convex relaxation.
History:

1. Introduction

A chance constrained optimization problem involves uncertain constraints (specified by stochastic data) which can be violated with at most a prescribed probability level. A typical formulation is:

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \; \boldsymbol{x} \in S, \; \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : \boldsymbol{x} \notin \mathcal{X}(\tilde{\boldsymbol{\xi}}) \right\} \leq \epsilon \right\}, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is a decision vector subject to a set of deterministic constraints $S \subseteq \mathbb{R}^n$, $\tilde{\boldsymbol{\xi}}$ is a random data vector or matrix with support Ξ and probability distribution $\mathbb{P}$, and $\mathcal{X}(\tilde{\boldsymbol{\xi}})$ denotes a system of stochastic constraints whose data is specified by the random vector $\tilde{\boldsymbol{\xi}}$. Problem (1) seeks a solution $\boldsymbol{x} \in S$ that minimizes the cost $\boldsymbol{c}^\top \boldsymbol{x}$ and is allowed to violate the uncertain constraints $\mathcal{X}(\tilde{\boldsymbol{\xi}})$ with probability at most ϵ, where $\epsilon \in (0,1)$ is a prespecified risk level. Often the distributional information of $\tilde{\boldsymbol{\xi}}$ might not be fully known, making it difficult to commit to a single $\mathbb{P}$. In this case, to hedge against distributional uncertainty, an alternative to (1) is a distributionally robust chance constrained problem of the form

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \in S, \; \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : \boldsymbol{x} \notin \mathcal{X}(\tilde{\boldsymbol{\xi}}) \right\} \leq \epsilon \right\}, \tag{2}$$

where the *ambiguity set* $\mathcal{P}$ denotes the set of all possible probability distributions subject to certain prespecified information.

The feasible regions of chance constrained optimization problems of the form (1) and (2) are typically nonconvex, making these problems extremely difficult. There are several special settings where the feasible regions of (1) and (2) are convex. For example, Kataoka (1963) showed that if set S is convex, support $\Xi = \mathbb{R}^d$, $\epsilon \leq 1/2$, $X(\tilde{\boldsymbol{\xi}})$ is defined by one uncertain linear inequality, and $\tilde{\boldsymbol{\xi}}$ has a symmetric and non-degenerate log-concave distribution, then the feasible region of (1) is convex; Calafiore and El Ghaoui (2006), El Ghaoui et al. (2003) proved that if set S is convex, $X(\tilde{\boldsymbol{\xi}})$ is defined by one uncertain linear inequality, and $\mathcal{P}$ consists of all probability distributions with given first and second moments, then the feasible region of (2) is convex. Due to intractability of chance constrained optimization problems, there has been a great deal of elegant work on developing tractable approximations of these problems. Unfortunately none of these approaches come with a constant factor approximation guarantee. We show that such a guarantee is impossible by proving an inapproximability result. On the other hand, for a large class of chance constrained problems involving uncertain *covering* constraints, we propose an approximation method with a provable performance guarantee. A covering constraint system has the property that any feasible solution to the system scaled by a factor $\lambda \geq 1$ remains feasible. Our proposed approximation method for chance constrained covering problems comes with a *bicriteria* performance guarantee. That is, the method finds a solution whose violation probability may be larger than, but is within a constant factor of, the specified risk parameter ϵ and whose objective value is within a constant factor of the true optimal value. Key to our developments is the construction of a tractable convex relaxation of a chance constrained optimization problem and an appropriate scaling of a solution to this relaxation. We extend our approximation results to distributionally robust chance constrained covering problems with convex moment and Wasserstein ambiguity sets, and provide analogous bicriteria approximation results.

The remainder of the paper is organized as follows. Section 2 details the problem setting and the underlying assumptions, reviews related literature, and summarizes our contributions. Section 3 exposes unbounded suboptimality of two well-known approximations of chance constraints, and provides a formal inapproximability result. Section 4 develops the bicriteria approximation results. Section 5 extends the complexity and approximation results to distributionally robust chance constrained covering problems. Section 6 presents a numerical illustration of the proposed approximation algorithms. Finally, Section 7 concludes the paper.

Notation: The following notation is used throughout the paper. We use bold-letters (e.g., $\boldsymbol{x}, \boldsymbol{A}$) to denote vectors or matrices, and use corresponding non-bold letters to denote their components. We let $\boldsymbol{e}$ be the all-ones vector, and let $\boldsymbol{e}_i$ be the ith standard basis vector. Given an integer n, we let $[n] := \{1, 2, \ldots, n\}$, and use $\mathbb{R}^n_+ := \{\boldsymbol{x} \in \mathbb{R}^n : x_j \geq 0, \forall j \in [n]\}$ and $\mathbb{R}^n_{++} := \{\boldsymbol{x} \in \mathbb{R}^n : x_j > 0, \forall j \in [n]\}$. Given a real number t, we let $(t)_+ := \max\{t, 0\}$, $\lceil t \rceil$ be its round-up and $\lfloor t \rfloor$ be its round-down. Given a finite set I, we let $|I|$ denote its cardinality. We let $\tilde{\boldsymbol{\xi}}$ denote a random vector or matrix with support Ξ and denote one of its realizations by $\boldsymbol{\xi}$. Given a probability distribution $\mathbb{P}$ on Ξ, we use $\mathbb{P}\{\mathcal{A}\}$ to denote $\mathbb{P}\{\tilde{\boldsymbol{\xi}} : \text{condition } \mathcal{A}(\tilde{\boldsymbol{\xi}}) \text{ holds}\}$ when $\mathcal{A}(\tilde{\boldsymbol{\xi}})$ is a condition on $\tilde{\boldsymbol{\xi}}$, and also to denote $\mathbb{P}\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}} \in \mathcal{A}\}$ when $\mathcal{A} \subseteq \Xi$ is $\mathbb{P}$-measurable. For matrix matrix $\boldsymbol{A}$, we let $\boldsymbol{A}_{i\bullet}$ denote the vector of ith row of $\boldsymbol{A}$ and $\boldsymbol{A}_{\bullet j}$ denote the vector of jth column of $\boldsymbol{A}$. Given a set $\mathcal{C}$, we let $\mathrm{int}(C)$ denote its interior. Additional notation will be introduced as needed.

2. Problem Setting, Related Literature and Contributions

2.1. Chance Constrained Covering Problems

A covering linear program (cf.,Vazirani 2001, Buchbinder and Naor 2009, Gupta and Nagarajan 2014) is of the form:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n_+} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \right\},$$

where $\boldsymbol{c} \in \mathbb{R}^n_+, \boldsymbol{A} \in \mathbb{R}^{m \times n}_+$ and $\boldsymbol{b} \in \mathbb{R}^m_+$. Note that without loss of generality, we can assume that $\boldsymbol{b} \in \mathbb{R}^m_{++}$, otherwise, if $b_i = 0$, then the ith constraint $\boldsymbol{A}_{i\bullet}^\top \boldsymbol{x} \geq b_i = 0$ is redundant. Therefore, since $\boldsymbol{b} \in \mathbb{R}^m_{++}$, we can scale the right-hand sides of the covering constraints $\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}$ to be the all-one vector

$\boldsymbol{e}$ by redefining $\boldsymbol{A}_{i\bullet} := \boldsymbol{A}_{i\bullet}/b_i$ for all $i \in [m]$. Hence, a covering linear program can be equivalently formulated as

$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+} \left\{\boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{e}\right\}.$$

Qiu et al. (2014) studied chance constrained linear covering problems of the form :

$$v^* = \min_{\boldsymbol{x}\in\mathbb{R}^n_+} \left\{\boldsymbol{c}^\top \boldsymbol{x} : \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}\boldsymbol{x} \geq \boldsymbol{e}\right\} \geq 1-\epsilon\right\}, \tag{3}$$

where $\tilde{\boldsymbol{\xi}}$ is an $m \times n$-dimensional nonnegative random matrix with support $\Xi \subseteq \mathbb{R}^{m\times n}_+$. Let us define a function $a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) = \min_{i\in[m]} \tilde{\boldsymbol{\xi}}_{i\bullet}^\top \boldsymbol{x}$. Thus, (3) can be reformulated as

$$v^* = \min_{\boldsymbol{x}} \left\{\boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \in S,\ \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) < 1\right\} \leq \epsilon\right\}. \tag{4}$$

Note that we have included the set of deterministic constraints S in the above formulation. In (3), $S = \mathbb{R}^n_+$. For the remainder of the paper we will consider this formulation (4) and refer to it as a *Chance constrained Covering Problem* (CCP). The distributionally robust counterpart

$$v^* = \min_{\boldsymbol{x}} \left\{\boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \in S,\ \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) < 1\right\} \leq \epsilon\right\}, \tag{5}$$

will be referred to as a *Distributionally Robust Chance constrained Covering Problem* (DRCCP).

Although we have motivated CCP and DRCCP from linear covering constraints, our results are applicable to slightly more general settings. Note that in the above formulations, v^* must be nonnegative, function $a : S \times \Xi \to \mathbb{R}_+$ is concave and positively homogeneous over the set S and when $S = \mathbb{R}^n_+$, it is nonempty, closed, convex and satisfies the so-called "scaling" property, i.e., for any $\boldsymbol{x} \in S$ and scalar $\eta \geq 1$, we must have $\eta\boldsymbol{x} \in S$. Therefore, we generalize the above CCP (4) and DRCCP (5) to a more general deterministic set S and function $a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})$ under the following assumptions.

(a) The deterministic constraint system S is nonempty, closed, convex, contained in a closed convex cone and is up-monotone. That is, $\emptyset \neq S \subseteq C$, where C is a closed pointed convex cone and for each $\boldsymbol{x} \in S$ and scalar $\eta \geq 1$, it holds that $\eta\boldsymbol{x} \in S$. Note that the set S itself need not be a cone, for example, it can be a deterministic covering system (i.e., $S = \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{e}\}$ with nonnegative matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}_+$).

(b) The objective cost vector $\boldsymbol{c} \in \mathrm{int}(C^*)$, where C^* is the dual cone of C and $\mathrm{int}(\cdot)$ denotes the interior.

(c) For each $\boldsymbol{\xi} \in \Xi$, $a(\cdot, \boldsymbol{\xi})$ is concave and positively homogeneous over S. That is, $\eta a(\boldsymbol{x}, \boldsymbol{\xi}) + (1-\eta)a(\boldsymbol{y}, \boldsymbol{\xi}) \leq a(\eta\boldsymbol{x} + (1-\eta)\boldsymbol{y}, \boldsymbol{\xi})$ for all $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\eta \in [0,1]$; and $a(\eta\boldsymbol{x}, \boldsymbol{\xi}) = \eta a(\boldsymbol{x}, \boldsymbol{\xi})$ for all $\boldsymbol{x} \in S$ and $\eta \geq 0$.

(d) The set of solutions is nonempty.

Note that Assumptions (a) and (c) together guarantee that any feasible solution to the problem (4) or (5) scaled by a factor $\eta \geq 1$ remains feasible. These two assumptions are necessary to prove the approximation guarantees for the Relax-and-Scale scheme in Section 4 (i.e., Theorem 5). Note that the results in Theorem 5 might still hold if the convexity of set S in Assumption (a) is relaxed, i.e., set S can be nonconvex instead of convex. However, by relaxing this condition, the relaxation schemes in Section 4 might not be tractable. Thus, for simplicity, we stick to Assumption (a). In Assumption (c), the concavity of $a(\cdot, \boldsymbol{\xi})$ is not necessary for the analysis of the Relax-and-Scale scheme in Section 4 but it is a sufficient condition for the relaxation sets proposed in Section 4.2 to be convex (i.e., Theorem 4), thus is necessary for the tractability of convex relaxations. Assumptions (b) and (d) together guarantee the existence of the optimal solution and nonnegativity of the optimal value v^*, which are necessary for the analysis of the Relax-and-Scale scheme in Section 4 (i.e., Theorem 5). Assumption (b) essentially tells that v^* is nonnegative, thus is bounded from below. If this assumption were relaxed, the chance constrained problems (1) and (2) might be unbounded, which leads to trivial cases.

2.2. Related Literature

There is a vast literature on general chance constrained optimization. Here we focus on the chance constrained covering problems and general distributionally robust chance constrained problems.

Applications of CCP: Various applications or their special cases of CCP of the form (4) have been studied in the literature. Next, we cite a few examples.

- Qiu et al. (2014) studied a chance constrained portfolio optimization problem of the form of

$$\min_{\boldsymbol{x} \in \mathbb{R}_+^n} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{e}^\top \boldsymbol{x} = 1, \mathbb{P}\{\tilde{\boldsymbol{a}}^\top \boldsymbol{x} \geq r\} \geq 1 - \epsilon \right\},$$

  where the constant $r > 0$, $\boldsymbol{c} \in \mathbb{R}_+^n, \tilde{\boldsymbol{a}} \in \mathbb{R}_+^n$ w.p.1. The problem seeks a minimum cost portfolio under a (normalized) budget constraint whose uncertain return is guaranteed to exceed the threshold r with probability at least $1 - \epsilon$. In the absence of the budget constraint, this chance constrained portfolio optimization problem is a special case of CCP (4).
- Beraldi and Ruszczyński (2002) studied the following probabilistic lot-sizing problem

$$\begin{aligned} \min_{\boldsymbol{x}, \boldsymbol{y}} \ & \sum_{t \in [T]} (k_t y_t + (c_t + H_t) x_t) \\ \text{s.t.} \ & \mathbb{P}\left\{ \sum_{t \in [\tau]} x_t \geq \sum_{t \in [\tau]} \tilde{\xi}_t, \forall \tau \in [T] \right\} \geq 1 - \epsilon \\ & x_t \leq M_t y_t, x_t \geq 0, y_t \in \{0, 1\}, \forall t \in [T], \end{aligned}$$

  where $\tilde{\boldsymbol{\xi}}$ a denote nonnegative random demand vector and $\boldsymbol{k}, \boldsymbol{c}, \boldsymbol{H}$ are nonnegative cost vectors. Here the problem is to decide a minimum cost production schedule, such that the cumulative production exceeds the uncertain cumulative demand with sufficiently high probability. The overall cost involves fixed and variable production costs and holding costs. The fact that the set-up variables $\boldsymbol{y}$ are binary violates Assumption (a). However, in case of zero fixed costs (i.e., $\boldsymbol{k} = \mathbf{0}$), one will always choose $\boldsymbol{y} = \boldsymbol{e}$; hence, under this condition, the lot-sizing problem is a special case of CCP (4).

- Talluri et al. (2006) considered a supply chain design problem under uncertainty of the form
$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^n_+,\theta} \left\{\theta : \mathbb{P}\left\{\tilde{\boldsymbol{Y}}_{n\bullet}^\top \boldsymbol{\lambda} < y_{no}\right\} \leq \epsilon, \forall n \in [N], \theta x_{mo} \geq \boldsymbol{X}_{m\bullet}^\top \boldsymbol{\lambda}, \forall m \in [M]\right\}$$
where $\tilde{\boldsymbol{Y}}$ is a random output matrix, $\boldsymbol{X}$ is a deterministic input matrix, and deterministic vectors $\boldsymbol{x}_o, \boldsymbol{y}_o$ denote the inputs and outputs of the unit under investigation, respectively. Here the problem is to minimizes the value of a relative efficiency score θ such that (1) the probability that observed outputs exceed the outputs $\boldsymbol{y}_o$ should be within a risk threshold, and (2) the total used inputs must not be more than $\theta \boldsymbol{x}_o$. In the setting where the matrices $\tilde{\boldsymbol{Y}}, \boldsymbol{X}$ and vectors $\boldsymbol{x}_o, \boldsymbol{y}_o$ are nonnegative, this problem is a special case of CCP (4).
- Dentcheva et al. (2000) studied a traffic assignment problem in telecommunication of the form
$$\min_{\boldsymbol{x}\in\mathbb{R}^n_+} \left\{\sum_{i\in[n]} x_i : \mathbb{P}\left\{\sum_{i\in[n]} \boldsymbol{Q}^{(i)} x_i \geq \tilde{\boldsymbol{D}}\right\} \geq 1-\epsilon, \boldsymbol{x} \in \mathbb{Z}^n_+\right\},$$
where $\tilde{\boldsymbol{D}} \in \mathbb{R}^{m\times m}_+$ denotes a random demand matrix, and $\boldsymbol{Q}^{(i)} \in \mathbb{R}^{m\times m}_+$ is a nonnegative transmission assignment matrix for each $i \in [n]$. Here the problem is to minimize the total time slots such that total transmission assignments from senders to receivers satisfy the demand with very high probability. The fact that variables $\boldsymbol{x}$ are integer violates Assumption (a). However, the continuous relaxation of this traffic assignment problem is a special case of CCP (4).

Solution approaches for CCP: CCP is known to be strongly NP-hard (Qiu et al. 2014). Exact solution methods rely on mixed integer programming (Dentcheva et al. 2000, Luedtke et al. 2010, Qiu et al. 2014) or global optimization (Cheon et al. 2006). A variety of approximation approaches for general chance constrained optimization as well as for the CCP have been developed. Calafiore et al. (2006), Nemirovski and Shapiro (2006) and related works developed convex restrictions of the feasible region of a chance constrained set, optimizing over which provides a feasible solution. As shown in Section 3, these approximations do not have any bounded approximation guarantees. To the best of our knowledge, all existing approximation algorithms with provable guarantees have been proposed for chance constrained *combinatorial* optimization problems. For example, Goyal and Ravi (2008) proposed constant factor approximation algorithms for certain classes of chance constrained combinatorial covering problems. Subsequently, Goyal and Ravi

(2010) developed a fully polynomial time approximation scheme for a chance constrained knapsack problem where the item sizes are drawn from independent normal distributions. Swamy (2011) studied a two-stage chance constrained set covering problem and proposed a polynomial time approximation algorithm by rounding linear program relaxation solutions.

General distributionally robust chance constrained problems: Recently, distributionally robust chance constrained optimization has attracted much attention (e.g., Calafiore and El Ghaoui 2006, Chen et al. 2010, Zymler et al. 2013, Hanasusanto et al. 2017, Jiang and Guan 2016, Xie and Ahmed 2018b, Yang and Xu 2016). In case of a single linear uncertain inequality constraint and an ambiguity set $\mathcal{P}$ defined by moment constraints, such problems are typically computationally tractable. This has been observed by El Ghaoui et al. (2003) and Calafiore and El Ghaoui (2006) for the Chebyshev ambiguity set (i.e., known first and second moments), by Hanasusanto et al. (2015) for conic moment constraints, and by Xie and Ahmed (2018b) for arbitrary convex moment constraints. However, for multiple uncertain constraints, tractability results are rare. As far as the authors are aware of, the only known results are by Hanasusanto et al. (2017) for the mean and dispersion ambiguity set, by Xie and Ahmed (2018b) for an ambiguity set with one moment inequality, by Xie and Ahmed (2018a) for two uncertain linear constraints and the Chebyshev ambiguity set, and by Xie et al. (2017) for separable uncertain constraints and ambiguity sets. In general, a distributionally robust chance constrained problem with multiple uncertain constraints and a moment ambiguity set is an NP-hard problem (c.f., Hanasusanto et al. 2017, Xie et al. 2017). When the ambiguity set $\mathcal{P}$ is specified by all distributions within a certain distance to a specified nominal distribution, there are even fewer tractability results. The only known results are when the probability distance metric is ϕ-divergence. In this case, a distributionally robust chance constraint is equivalent to a regular chance constraint under the nominal distribution with an adjusted risk parameter (Hu and Hong 2012, Jiang and Guan 2016, Shapiro 2017). Thus any tractability conditions for regular chance constraints carry over. All of the above mentioned results are for general distributionally robust chance constraints, and their specialization to the covering setting, i.e. DRCCP, has not been studied.

2.3. Contributions

We make several contributions to the complexity analysis and approximations of CCP and DRCCP. Our proposed approximations come with *bicriteria* approximation guarantees. Given a *violation ratio* $\sigma \geq 1$ and an *approximation ratio* $\gamma \geq 1$, a (σ, γ)-bicriteria approximation algorithm for CCP returns a solution $\widehat{\boldsymbol{x}} \in S$, such that $\mathbb{P}\{\tilde{\boldsymbol{\xi}} : \widehat{\boldsymbol{x}} \notin \mathcal{X}(\tilde{\boldsymbol{\xi}})\} \leq \sigma\epsilon$ and $\boldsymbol{c}^\top \widehat{\boldsymbol{x}} \leq \gamma v^*$, i.e., the solution violates the uncertain constraints with probability at most $\sigma\epsilon$ and has an objective value at most γ times the optimal value. Note that σ, γ may be dependent on the risk parameter ϵ and the underlying probability distribution $\mathbb{P}$. Following is a summary of our contributions.

1. We show that two well-known approximation approaches for CCP can have arbitrarily bad approximation ratios when no violation of the chance constraint is allowed.
2. We prove that, unless P=NP, it is impossible to obtain a polynomial time algorithm with a constant factor approximation for CCP if no violation of the chance constraint is allowed. This motivates the need for bicriteria approximations.
3. We propose and analyze a bicriteria approximation scheme for CCP that relies on solving a specific convex relaxation of the problem and appropriately scaling its solution. The violation and approximation ratios (σ, γ) of the scheme depends on the parameters of the convex relaxation solved.
4. We show that the performance guarantee of proposed algorithm can be improved by exploiting any structural independence among different uncertain constraints.
5. We extend the complexity results in the literature by showing that even in the covering setting, a distributionally robust chance constrained covering problem, i.e. DRCCP, with a convex moment or a Wasserstein ambiguity set is NP-hard.
6. We extend the bicriteria approximation analysis to DRCCP. In particular, we show that for a convex moment or a Wasserstein ambiguity set, given a violation ratio of $\sigma > 1$, the approach provides a solution within a factor of $\gamma = \frac{\sigma}{\sigma-1}$ of the optimal value.

3. Inapproximability

In this section, we first show that two well-known approximation approaches can have arbitrarily bad approximation ratios for CCP, and then provide a formal single criterion inapproximability result. This inapproximability result motivates our subsequent development of a bicriteria approximation algorithm.

3.1. $\mathbf{CVaR}$ Approximation

A well-known convex approximation of a chance constraint is to replace the nonconvex chance constraint by a convex constraint defined by the conditional value at risk or $\mathbf{CVaR}$ (see Nemirovski and Shapiro 2006 for details). For CCP as defined in (4), the resulting formulation of its $\mathbf{CVaR}$ approximation is

$$v^{\mathbf{CVaR}} = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \in S,\ \inf_{\beta} \left\{ -\epsilon\beta + \mathbb{E}\left[\left(1 - a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) + \beta \right)_+ \right] \right\} \leq 0 \right\}. \tag{6}$$

Problem (6) is a convex optimization problem and provides a feasible solution to CCP, thus $v^{\mathbf{CVaR}} \geq v^*$. The following result shows that the approximation quality from this approach can be arbitrarily bad.

Theorem 1 *Given a parameter $\kappa > 1$, consider the following instance of CCP:*

$$(P_\kappa): \quad v_\kappa^* = \min_{x \geq 0} \left\{ x : \mathbb{P}\{\tilde{\xi} : \tilde{\xi} x < 1\} \leq \epsilon \right\},$$

where $\tilde{\xi}$ has a finite distribution with N equiprobable realizations $\{\xi^i\}_{i \in [N]}$ such that ϵN is an integer, $\xi^i = 1$ for $i \in [\epsilon N]$ and $\xi^i = \kappa$ for $i \in [N] \setminus [\epsilon N]$. Let $v_\kappa^{\mathbf{CVaR}}$ denote the optimal value of the $\mathbf{CVaR}$ approximation (6) of (P_κ). Then

$$\lim_{\kappa \to \infty} \frac{v_\kappa^{\mathbf{CVaR}}}{v_\kappa^*} = \infty.$$

Proof: See e-companion EC.1.1. □

3.2. Scenario Approximation

The scenario approximation (SA) approach, proposed by Calafiore et al. (2006), uses $\bar{N}$ i.i.d. samples $\{\boldsymbol{\xi}^i\}_{i \in [\bar{N}]}$ from the distribution $\mathbb{P}$ and considers an optimization problem where the constraint corresponding to each sampled scenario is enforced. For CCP as defined in (4), its scenario approximation problem is

$$v^{SA} = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \in S, \ a(\boldsymbol{x}, \boldsymbol{\xi}^i) \geq 1, \ \forall i \in [\bar{N}] \right\}. \tag{7}$$

Calafiore et al. (2006) showed that when the sample size $\bar{N}$ satisfies

$$\bar{N} \geq \left\lceil \frac{2}{\epsilon} \log\left(\frac{1}{\delta}\right) + \frac{2n}{\epsilon} \log\left(\frac{2}{\epsilon}\right) + 2n \right\rceil,$$

then, with probability at least $(1-\delta)$ where $\delta \in (0,1)$ is a given confidence level, an optimal solution (if it exists) to the approximate problem (7) is a feasible solution to CCP, thus, $\mathbb{P}\{v^{SA} \geq v^*\} \geq 1-\delta$. The following result shows that the approximation quality from the SA approach can be arbitrarily bad by showing that (7) can be infeasible with high probability.

Theorem 2 *Consider the following instance of CCP:*

$$(P): \quad v^* = \min_{x \geq 0} \left\{ x : \mathbb{P}\{\tilde{\xi} : \tilde{\xi} x < 1\} \leq \epsilon \right\},$$

where $\tilde{\xi}$ is a Bernoulli random variable with probability $(1-\epsilon)$. Let v_δ^{SA} denote the optimal value of the scenario approximation problem (7) corresponding to (P) for a given confidence parameter $\delta \in (0,1)$ and sample size $\bar{N}_\delta = \left\lceil \frac{2}{\epsilon} \log\left(\frac{1}{\delta}\right) + \frac{2}{\epsilon} \log\left(\frac{2}{\epsilon}\right) + 2 \right\rceil$. Then

$$\mathbb{P}\left\{ \frac{v_\delta^{SA}}{v^*} = \infty \right\} \geq 1 - \delta^2 \epsilon^2.$$

Proof: See e-companion EC.1.2. □

3.3. Single-criterion Inapproximability

The previous two subsections demonstrated that two common approximation schemes for CCP can have arbitrarily bad approximation factors. Here we show that such approximation difficulty is a feature of this problem class rather than the specific approaches. In particular, we show that, unless $P = NP$, it is impossible to obtain a polynomial time algorithm for CCP with a constant factor single criteria approximation in terms of the approximation ratio or the violation ratio. This result motivates the need for our subsequent bicriteria approximation algorithms that relax both optimality and feasibility.

Theorem 3 *Suppose we have a polynomial time algorithm that returns a (σ, γ)-approximate solution to any CCP of the form* (4) *with a discrete distribution with N equiprobable realizations. Then, unless P=NP, the following holds:*

(i) if $\gamma = 1$, then we must have $\sigma = 1/\epsilon - f(N)(1-\epsilon)/\epsilon$ for some function f such that $f(N) \to 0$ as $N \to \infty$;

(ii) if $\sigma = 1$, then we must have $\gamma = g(N)$ for some function g such that $g(N) \to \infty$ as $N \to \infty$.

Proof: See e-companion EC.1.3. □

Remark: The proof of part (i) in Theorem 3 may also follow from Goyal and Ravi (2008) who show that the κ-edge dense graph problem (Alon et al. 2011) can be polynomially reduced to CCP of the form (4) with binary decision variables $\boldsymbol{x}$. However the approach of Goyal and Ravi (2008) does not directly work for establishing the inapproximability result in terms of both σ and γ in the presence of continuous variables.

4. Bicriteria Approximation of CCP

In this section we present an approximation scheme for CCP by first solving a convex relaxation and then appropriately scaling an optimal solution of this relaxation. We provide a bicriteria approximation analysis of this scheme. Finally, we show that the analysis can be improved by exploiting independence structure in the underlying distribution.

4.1. A Relax-and-Scale Scheme

For notational convenience, given $\alpha \in [0,1]$, we will use X_α to denote the following set:

$$X_\alpha := \left\{ \boldsymbol{x} \in S : \ \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) < 1 \right\} \leq \alpha \right\}. \tag{8}$$

With this notation, CCP of the form (4) can be written as

$$\min_{\boldsymbol{x}} \{ \boldsymbol{c}^\top \boldsymbol{x} : \ \boldsymbol{x} \in X_\epsilon \}.$$

Next, we will introduce a generic Relax-and-Scale algorithm (i.e., Algorithm 1) to construct an approximate solution to CCP, which comes with a bicriteria performance guarantee. The algorithm solves a convex relaxation of CCP defined by the set $\widehat{X}$, and then scales its solution by a scalar $u \in [1,\gamma]$ where $\gamma \geq 1$ is a specified approximation ratio parameter, to ensure that $u\widehat{\boldsymbol{x}} \in X_{\sigma\epsilon}$, where $\sigma \geq 1$ is a specified violation ratio parameter. That is, the scaled solution should satisfy the chance constraint at risk level of $\sigma\epsilon$. Here, we are assuming that checking if a given solution is in $X_{\sigma\epsilon}$ is easy; otherwise, we can set $u=\gamma$ and skip lines 4-7 of Algorithm 1.

Algorithm 1 : Relax-and-Scale $(\sigma, \gamma, \widehat{X})$

1: **input:** parameters $\sigma \geq 1$, $\gamma \geq 1$, and a convex relaxation $\widehat{X}$ of X_ϵ
2: **initialize:** set $l=1, u=\gamma$ and let $\delta > 0$ be a stopping tolerance
3: solve the convex relaxation: $\min_{\boldsymbol{x}}\{\boldsymbol{c}^\top \boldsymbol{x}: \ \boldsymbol{x} \in \widehat{X}\}$ and let $\widehat{\boldsymbol{x}}$ be an optimal solution
4: **while** $u - l > \delta$ **do**
5: $\quad \tau \leftarrow (l+u)/2$
6: $\quad$ **if** $\tau\widehat{\boldsymbol{x}} \in X_{\sigma\epsilon}$ set $u \leftarrow \tau$, **else** set $l \leftarrow \tau$
7: **end while**
8: **output:** $\bar{\boldsymbol{x}} = u\widehat{\boldsymbol{x}}$.

Next, we establish the choice of the convex relaxation $\widehat{X}$ and parameters σ, γ, such that Relax-and-Scale $(\sigma, \gamma, \widehat{X})$ returns a (σ,γ)-bicriteria approximate solution to CCP.

4.2. Convex Relaxations of CCP

We consider the convex relaxation scheme for chance constaints proposed by Ahmed (2014). Let Φ be a class of univariate "*generating*" functions $\phi: \mathbb{R} \to \mathbb{R}$ satisfying

(a) $\phi(s) \in [0,1]$ for all $s \in \mathbb{R}_+$;
(b) $\phi(s)$ is nondecreasing and strictly increasing in $[0,1]$; and
(c) $\phi(s)$ is concave in s.

Two typical generating functions are the *Markovian generator* $\phi(s) = \min\{1, s\}$ and the *Bernstein generator* $\phi(s) = 1 - e^{-\tau s}$ for some $\tau > 0$.

For any generating function $\phi \in \Phi$, consider the following set

$$X_\phi^{\text{rel}} = \left\{ \boldsymbol{x} \in S: \ \ \mathbb{E}[\phi(a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}))] \geq (1-\epsilon)\phi(1) \right\}, \tag{9}$$

and the optimization problem:

$$v_\phi^{\text{rel}} = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x}: \ \boldsymbol{x} \in X_\phi^{\text{rel}} \right\}. \tag{10}$$

The following result establishes that (10) is a valid convex relaxation of CCP. The result follows from Markov inequality and standard convex analysis.

Theorem 4 *(Ahmed 2014) For any $\phi \in \Phi$, $X_\epsilon \subseteq X_\phi^{\rm rel}$. Furthermore $X_\phi^{\rm rel}$ is a convex set, and so* (10) *is a convex relaxation of CCP.*

Remark: Note that for an arbitrary convex set S, the result of Theorem 4 still hold, i.e., in Assumption (a) of Section 2, the scaling property of the set S is not required for Theorem 4; however, Assumption (a) is necessary for the bicriteria approximation result to hold (i.e., Theorem 5) in the next section. Even though (10) is a convex optimization problem, it may not be tractable. Ahmed (2014) presents several generator functions ϕ and their associated conditions on the probability distribution that make the problem tractable. We discuss these in Section 4.4.

4.3. Bicriteria Analysis

In this section, we show that by choosing convex relaxations of the form (10) along with appropriate values of σ and γ, the Relax-and-Scale scheme correctly returns a (σ,γ)-bicriteria approximate solution to CCP. Before the main result, we state a simple lemma that we will make use of.

Lemma 1 *Let $\widehat{\boldsymbol{x}}$ be a solution satisfying $\mathbb{E}[\phi(a(\boldsymbol{x},\tilde{\boldsymbol{\xi}}))] \geq \alpha$ for some $\alpha \in [0,1]$. Given $r \in (0,1)$, let*

$$\Xi_r := \{\boldsymbol{\xi} \in \Xi : \ \phi(a(\widehat{\boldsymbol{x}},\boldsymbol{\xi})) \geq 1-r\}.$$

Then $\mathbb{P}\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}} \in \Xi_r\} \geq \frac{\alpha+r-1}{r}$.

Proof: We have

$$\alpha \leq \mathbb{E}[\phi(a(\widehat{\boldsymbol{x}},\tilde{\boldsymbol{\xi}}))] \leq \mathbb{P}\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}} \in \Xi_r\} + (1-\mathbb{P}\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}} \in \Xi_r\})(1-r)$$

where the second inequality is because $\phi(a(\widehat{\boldsymbol{x}},\boldsymbol{\xi})) \leq 1$ for all $\boldsymbol{\xi} \in \Xi$ and $\phi(a(\widehat{\boldsymbol{x}},\boldsymbol{\xi})) < 1-r$ for all $\Xi \setminus \Xi_r$. The result then follows. □

Now we are ready to present our main approximation results. The main proof idea is that for a given generating function $\phi \in \Phi$ and an optimal solution $\widehat{\boldsymbol{x}}$ of the convex relaxation (10), we first specify a violation ratio σ and an approximation ratio γ, then show that $\gamma\widehat{\boldsymbol{x}}$ is a (σ,γ)-bicriteria approximate solution to CCP.

Theorem 5 *Given a generating function $\phi \in \Phi$, choose*

- *violation ratio: $\sigma \in (\frac{1}{\epsilon} - (\frac{1}{\epsilon}-1)\phi(1), \frac{1}{\epsilon})$,*
- *approximation ratio: $\gamma = \left[\phi^{-1}\left(1-\frac{1}{\sigma\epsilon}+(\frac{1}{\sigma\epsilon}-\frac{1}{\sigma})\phi(1)\right)\right]^{-1}$, and*
- *relaxation: $\widehat{X} = X_\phi^{\rm rel}$.*

Then Relax-and-Scale $(\sigma,\gamma,\widehat{X})$ *returns a* (σ,γ)*-bicriteria approximate solution to CCP.*

Proof: Note that by Assumptions (c) and (d), there exists a feasible solution $\bar{\boldsymbol{x}}\neq \mathbf{0}$ to CCP (4), which is also feasible to the convex relaxation (10). Thus, let us define a set $\bar{\mathcal{C}}:=\left\{\boldsymbol{x}\in\mathcal{C}:\boldsymbol{c}^\top\boldsymbol{x}\leq\boldsymbol{c}^\top\bar{\boldsymbol{x}}\right\}$, then (10) is equivalent to

$$v_\phi^{\mathrm{rel}}=\min_{\boldsymbol{x}}\left\{\boldsymbol{c}^\top\boldsymbol{x}:\ \boldsymbol{x}\in X_\phi^{\mathrm{rel}}\cap\bar{\mathcal{C}}\right\}. \tag{*}$$

According to Assumption (b), we know that $\boldsymbol{c}\in\mathrm{int}(\mathcal{C}^*)$, i.e., that $\boldsymbol{c}^\top\boldsymbol{x}>0$ for all $\boldsymbol{x}\in\mathcal{C}\setminus\{\mathbf{0}\}$. Thus, set $\bar{\mathcal{C}}$ must be bounded; otherwise, suppose that there exists a sequence $\{\widehat{\boldsymbol{x}}^\tau\}_{\tau\geq 1}\subseteq\bar{\mathcal{C}}$ such that $\lim_{\tau\to\infty}\|\widehat{\boldsymbol{x}}^\tau\|=\infty$ but $\boldsymbol{c}^\top\widehat{\boldsymbol{x}}^\tau\leq\boldsymbol{c}^\top\bar{\boldsymbol{x}}$ for all $\tau\geq 1$. Now consider the normalized sequence $\{\widehat{\boldsymbol{x}}^\tau/\|\widehat{\boldsymbol{x}}^\tau\|\}_{\tau\geq 1}$, which is bounded, thus by Bolzano-Weierstrass Theorem, there exists a convergent subsequence $\{\widehat{\boldsymbol{x}}^{\tau_l}/\|\widehat{\boldsymbol{x}}^{\tau_l}\|\}_{l\geq 1}$ such that $\{\tau_l\}_{l\geq 1}\subseteq\{1,2,\ldots\}$ and $\lim_{l\to\infty}\widehat{\boldsymbol{x}}^{\tau_l}/\|\widehat{\boldsymbol{x}}^{\tau_l}\|=\tilde{\boldsymbol{x}}$. Clearly, $\tilde{\boldsymbol{x}}\in\mathcal{C}$, $\tilde{\boldsymbol{x}}\neq 0$ and $0<\boldsymbol{c}^\top\widehat{\boldsymbol{x}}^{\tau_l}/\|\widehat{\boldsymbol{x}}^{\tau_l}\|\leq\boldsymbol{c}^\top\bar{\boldsymbol{x}}/\|\widehat{\boldsymbol{x}}^{\tau_l}\|\to 0$ when $l\to\infty$. Thus, we must have $\boldsymbol{c}^\top\tilde{\boldsymbol{x}}=0$, a contradiction that $\boldsymbol{c}\in\mathrm{int}(\mathcal{C}^*)$. Thus, (*) is a convex minimization problem over a compact set, by Weierstrass Theorem, there exists an optimal solution to (*), i.e., there exists an optimal solution to the convex relaxation (10), denoted as $\widehat{\boldsymbol{x}}$.

Hence, $\widehat{\boldsymbol{x}}$ satisfies $\mathbb{E}[\phi(a(\widehat{\boldsymbol{x}},\tilde{\boldsymbol{\xi}}))]\geq(1-\epsilon)\phi(1)$. Let $r=\frac{1}{\sigma\epsilon}-(\frac{1}{\sigma\epsilon}-\frac{1}{\sigma})\phi(1)$. Note that $r\in(0,1)$. From Lemma 1 we then have $\mathbb{P}\{\tilde{\boldsymbol{\xi}}:\tilde{\boldsymbol{\xi}}\in\Xi_r\}\geq\frac{(1-\epsilon)\phi(1)+r-1}{r}=1-\sigma\epsilon$ by the definition of r.

Now, let $\tilde{\boldsymbol{x}}=\gamma\widehat{\boldsymbol{x}}=\left[\phi^{-1}(1-r)\right]^{-1}\widehat{\boldsymbol{x}}$. We first show that $\tilde{\boldsymbol{x}}\in X_{\sigma\epsilon}$. Consider any $\widehat{\boldsymbol{\xi}}\in\Xi_r$, then

$$\begin{aligned}
&\phi(a(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{\xi}}))\geq(1-r)\\
\Leftrightarrow\ & a(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{\xi}})\geq\phi^{-1}(1-r)\\
\Leftrightarrow\ & \left[\phi^{-1}(1-r)\right]^{-1}a(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{\xi}})\geq 1\\
\Leftrightarrow\ & \gamma a(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{\xi}})\geq 1\\
\Leftrightarrow\ & a(\gamma\widehat{\boldsymbol{x}},\widehat{\boldsymbol{\xi}})\geq 1\\
\Leftrightarrow\ & a(\tilde{\boldsymbol{x}},\widehat{\boldsymbol{\xi}})\geq 1
\end{aligned}$$

where the first equivalence is because $(1-r)\in(0,1)$ and ϕ is strictly increasing in $(0,1)$, and the second-to-last equivalence is due to the positive homogeneity of a (i.e., Assumption (c)). Thus

$$\mathbb{P}\{\tilde{\boldsymbol{\xi}}:a(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{\xi}})\geq 1\}\geq\mathbb{P}\{\tilde{\boldsymbol{\xi}}:\tilde{\boldsymbol{\xi}}\in\Xi_r\}\geq 1-\sigma\epsilon,$$

or $\tilde{\boldsymbol{x}}\in X_{\sigma\epsilon}$.

The output $\bar{\boldsymbol{x}}=u\widehat{\boldsymbol{x}}$ of Algorithm 1 for any $u\in[1,\gamma]$, is guaranteed to be contained in the set $X_{\sigma\epsilon}$, i.e., $\bar{\boldsymbol{x}}\in X_{\sigma\epsilon}$. Moreover,

$$\gamma v^*\geq\gamma v_\phi^{\mathrm{rel}}=\gamma\boldsymbol{c}^\top\widehat{\boldsymbol{x}}\geq u\boldsymbol{c}^\top\widehat{\boldsymbol{x}}=\boldsymbol{c}^\top\bar{\boldsymbol{x}},$$

where the first inequality is due to $v_\phi^{\mathrm{rel}}\leq v^*$ and the second one is because of $\gamma\geq u$ and $\boldsymbol{c}^\top\widehat{\boldsymbol{x}}\geq 0$. Thus $\bar{\boldsymbol{x}}$ is a (σ,γ)-bicriteria approximate solution. $\square$

Remark: Note that Theorem 5 does not depend on the underlying distribution. This suggests that the bicriteria analysis can be extended to when the distribution $\mathbb{P}$ is not fully specified, i.e. the distributionally robust setting. We do this in Section 5. Also note that in Theorem 5, the violation and approximation ratios depend on the chosen relaxation scheme (10).

4.4. Special Cases

From Theorem 5, we note that the appropriate choice of the violation and approximation ratios depend on the risk level and the relaxation used. Furthermore the tractability of the Relax-and-Scale scheme depends on the tractability of solving the associated relaxation. In this section, we mention two settings with tractable relaxations and their associated violation and approximation ratios.

Markovian Relaxation As shown by Ahmed (2014), the tightest generating function in the class Φ is the *Markovian* generator $\phi(s) = \min\{s, 1\}$. The corresponding convex relaxation (10) is called the Markovian relaxation. In this case, the following corollary to Theorem 5 is immediate by noting that $\phi(1) = 1$.

Corollary 1 *Given a Markovian generating function $\phi(s) = \min\{s, 1\}$, choose*
- *violation ratio: $\sigma \in (1, \frac{1}{\epsilon})$, and*
- *relaxation: $\widehat{X} = X_\phi^{\text{rel}}$.*

Then Relax-and-Scale $(\sigma, \frac{\sigma}{\sigma-1}, \widehat{X})$ returns a $(\sigma, \frac{\sigma}{\sigma-1})$-bicriteria approximate solution to CCP.

Remark: In this case, by relaxing the risk level of the chance constraint by a factor of 2, i.e. $\sigma = 2$, we can get a factor 2 approximately optimal solution.

In general, e.g. when $\tilde{\boldsymbol{\xi}}$ is high dimensional and has a continuous distribution $\mathbb{P}$, the Markovian relaxation can be intractable since it involves computing and optimizing over the expectation of a nonsmooth function. However, suppose that $\tilde{\boldsymbol{\xi}}$ has a finite support $\Xi = \{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^N\}$ where each realization $\boldsymbol{\xi}^i$ for $i \in [N] := \{1, \ldots, N\}$ has an associated probability mass $p_i \in \mathbb{R}_+$, then the convex relaxation (10) with the Markovian generator reduces to the deterministic convex optimization problem:

$$v_\phi^{\text{rel}} = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \ \boldsymbol{x} \in S, \ a(\boldsymbol{x}, \boldsymbol{\xi}^i) \geq z_i, \forall i \in [N], \ \sum_{i \in [N]} p_i z_i \geq 1 - \epsilon, \ \boldsymbol{z} \in [0,1]^N \right\}. \tag{11}$$

Note that for the chance constrained linear covering problems (3), i.e. when $a(\boldsymbol{x}, \boldsymbol{\xi}^i)$ is the minimum of a set of linear functions of $\boldsymbol{x}$, problem (11) is a tractable linear program.

Bernstein Relaxation Another generating function studied by Ahmed (2014) is the *Bernstein* generator $\phi(s)=1-e^{-\tau s}$ for some $\tau>0$. The corresponding convex relaxation (10) is called the Bernstein relaxation. In this case we have the following bicriteria guarantee.

Corollary 2 *Given a Bernstein generating function $\phi(s)=1-e^{-\tau s}$ for some $\tau>0$, choose*

- *violation ratio: $\sigma\in(1+\frac{1-\epsilon}{\epsilon e^{\tau}},\frac{1}{\epsilon})$,*
- *approximation ratio: $\gamma=\tau\left[\log(\sigma)-\log\left(1+\frac{1-\epsilon}{\epsilon e^{\tau}}\right)\right]^{-1}$, and*
- *relaxation: $\widehat{X}=X_{\phi}^{\mathrm{rel}}$.*

Then Relax-and-Scale $(\sigma,\gamma,\widehat{X})$ returns a (σ,γ)-bicriteria approximate solution to CCP.

Remark: Note that the best parameter τ^* can be chosen by optimizing the approximation ratio, i.e., $\tau^*\in\arg\min_{\tau}\left\{\tau\left[\log(\sigma)-\log\left(1+\frac{1-\epsilon}{\epsilon e^{\tau}}\right)\right]^{-1}:\tau>\log(1/\epsilon-1)-\log(\sigma-1)\right\}$. For instance, when $\epsilon=0.05$, the violation ratio $\sigma=3$, then the optimal $\tau^*\approx 4.32$ and the best approximation ratio is $\gamma\approx 4.95$. See Figure 1 for an illustration of different optimized (σ,γ) pairs corresponding to $\epsilon=0.05$.

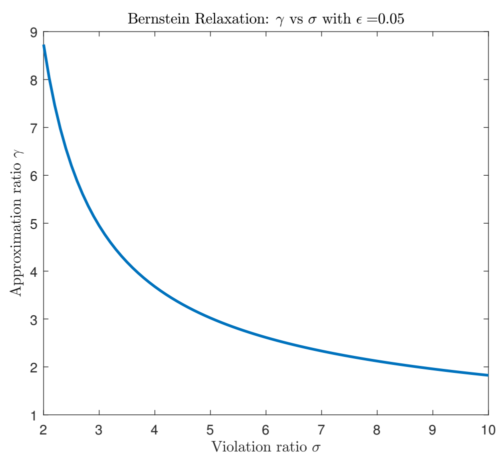


Figure 1 Illustration of optimized approximation and violation ratios for Bernstein relaxation with $\epsilon=0.05$.

Next we mention a setting in which the Bernstein relaxation is a tractable convex optimization problem. Consider instances of CCP where the left-hand-side of the uncertain constraint $a(\boldsymbol{x},\boldsymbol{\xi})\geq 1$ is of the form $a(\boldsymbol{x},\boldsymbol{\xi})=\sum_{j\in[d]}g_j(\xi_j)h_j(\boldsymbol{x})$ with $g_j:\mathbb{R}\to\mathbb{R}_+$ being arbitrary and $h_j:\mathbb{R}^n\to\mathbb{R}_+$ being concave and positively homogeneous in $\boldsymbol{x}$. Furthermore the random vector $\tilde{\boldsymbol{\xi}}$ has independently distributed components $\tilde{\boldsymbol{\xi}}_1,\ldots,\tilde{\boldsymbol{\xi}}_d$, with support $\Xi:=\prod_{j\in[d]}\Xi_j$. In this case the Bernstein relaxation (c.f., Ahmed 2014) is:

$$v_{\phi}^{\mathrm{rel}}=\min_{\boldsymbol{x}}\left\{\boldsymbol{c}^{\top}\boldsymbol{x}:\ \boldsymbol{x}\in S,\ \sum_{j\in[d]}\Lambda_j(-\tau h_j(\boldsymbol{x}))\leq\log\left(\epsilon+(1-\epsilon)e^{-\tau}\right)\right\},\tag{12}$$

where $\Lambda_j(t) = \log \mathbb{E}[e^{g_j(\tilde{\xi}_j)t}]$, i.e. the log moment generating functions of the random variable $g_j(\tilde{\xi}_j)$, for each $j \in [d]$. The problem (12) is a convex optimization problem and is tractable whenever the log moment generating function $\Lambda_j(t)$ is efficiently computable for all $t \in \mathbb{R}$ (c.f., Nemirovski and Shapiro 2006).

4.5. Exploiting Independence

In this section we consider CCP where the uncertainties associated with each uncertain inequalities are independent, i.e. left-hand-side of the uncertain constraint $a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \geq 1$ is of the form $a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) = \min_{l \in [m]}\{a_l(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_l)\}$, where the random vectors $\{\tilde{\boldsymbol{\xi}}_l\}_{l \in [m]}$ are independent. We will show that this independence structure allows us to improve the violation ratio for a given approximation ratio over that in Theorem 5.

First of all, in this independent uncertainty setting, CCP (4) can be reformulated as:

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \ \boldsymbol{x} \in S, \ \prod_{l \in [m]} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}}_l : a_l(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_l) \geq 1 \right\} \geq 1 - \epsilon \right\}. \tag{13}$$

For each $l \in [m]$, we introduce a new variable w_l such that $0 \leq 1 - w_l \leq \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}}_l : a_l(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_l) \geq 1 \right\}$, then the above formulation is equivalent to

$$v^* = \min_{\boldsymbol{x}, \boldsymbol{w}} \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{14a}$$

$$\text{s.t.} \quad \prod_{l \in [m]} (1 - w_l) \geq 1 - \epsilon, \tag{14b}$$

$$\mathbb{P}\left\{ \tilde{\boldsymbol{\xi}}_l : a_l(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_l) \geq 1 \right\} \geq 1 - w_l, \quad \forall l \in [m], \tag{14c}$$

$$\boldsymbol{x} \in S, w_l \in [0,1], \forall l \in [m], \tag{14d}$$

where (14b) can be reformulated as convex constraints by taking $\log(\cdot)$ on both sides or using second order cone constraints (c.f., Example 11 of Section 2.3.5 in Ben-Tal and Nemirovski 2001).

Now we apply the convex relaxation scheme of Section 4.2 to the chance constraints (14c) for each $l \in [m]$, and obtain the following relaxation to (14):

$$v_\phi^{\text{ind}} = \min_{\boldsymbol{x}, \boldsymbol{w}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \ \text{(14b)}, (\boldsymbol{x}, w_l) \in X_\phi^l, \forall l \in [m] \right\}, \tag{15}$$

where for each $l \in [m]$, X_ϕ^l is defined as

$$X_\phi^l = \left\{ (\boldsymbol{x}, w_l) : \boldsymbol{x} \in S, w_l \in [0,1], \ \mathbb{E}[\phi(a_l(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_l))] \geq (1 - w_l)\phi(1) \right\}. \tag{16}$$

The next theorem states that the violation ratio indeed can be improved by exploiting the independence structure.

Theorem 6 *Given a generating function $\phi \in \Phi$, choose*

- *violation ratio:* $\sigma \in (\frac{1}{\epsilon} - (\frac{1}{\epsilon} - 1)\phi(1), \frac{1}{\epsilon})$,
- *approximation ratio:* $\gamma = \left[\phi^{-1}\left(1 - \frac{1}{\sigma\epsilon} + (\frac{1}{\sigma\epsilon} - \frac{1}{\sigma})\phi(1)\right)\right]^{-1}$, *and*
- *relaxation:* $\widehat{X} =$ *the feasible region of* (15).

Then there exists a $\widehat{\sigma} \leq \sigma$, such that Relax-and-Scale $(\widehat{\sigma}, \gamma, \widehat{X})$ returns a $(\widehat{\sigma}, \gamma)$-bicriteria approximate solution to CCP of the form (13).

Proof: See e-companion EC.1.4. □

The following example illustrates Theorem 6 by showing that exploiting independence can lead to better approximation guarantees.

Example 1 Let $S \in \mathbb{R}_+, \Xi = \mathbb{R}_+^2, m = 2, \epsilon \in (0, 0.5), a_1(x, \tilde{\xi}_1) = \tilde{\xi}_1 x, a_2(x, \tilde{\xi}_1) = \tilde{\xi}_2 x$, $\tilde{\xi}_1, \tilde{\xi}_2$ are independent standard uniform random variables. In this setting, CCP (4) can be formulated as

$$v^* = \min\{x : x \geq 0, \mathbb{P}\left\{(\tilde{\xi}_1, \tilde{\xi}_2) : \tilde{\xi}_1 x \geq 1, \tilde{\xi}_2 x \geq 1\right\} \geq 1 - \epsilon\} = \min\left\{x : (1 - 1/x)^2 \geq 1 - \epsilon\right\}.$$

Consider the Markovian relaxation $\phi(s) = \min\{s, 1\}$. If we do not exploit the independence structure, then the Markovian relaxation of the above optimization problem is

$$v_\phi^{\mathrm{rel}} = \min\left\{x : x \geq 0, \mathbb{E}\left[\min\left\{\tilde{\xi}_1 x, \tilde{\xi}_2 x, 1\right\}\right] \geq 1 - \epsilon\right\},$$

while by considering the independence of random variables $\tilde{\xi}_1, \tilde{\xi}_2$, we can obtain the following relaxation

$$v_\phi^{\mathrm{ind}} = \min\left\{x : x \geq 0, \mathbb{E}\left[\min\left\{\tilde{\xi}_l x, 1\right\}\right] \geq 1 - w_l, l \in [2], (1 - w_1)(1 - w_2) \geq 1 - \epsilon\right\}.$$

By solving the above three optimization problems, we have $v^* = x^* = \frac{1+\sqrt{1-\epsilon}}{\epsilon}, v_\phi^{\mathrm{rel}} = x^{\mathrm{rel}} = \frac{1+\sqrt{1-4/3\epsilon}}{2\epsilon}, v_\phi^{\mathrm{ind}} = x^{\mathrm{ind}} = \frac{1+\sqrt{1-\epsilon}}{2\epsilon}$. Given an approximation ratio $\gamma > 1$, if we do not exploit the independence structure, then the violation ratio is $1 - \sigma\epsilon = (1 - 1/(\gamma x^{\mathrm{rel}}))^2$, i.e.,

$$\sigma = \frac{1}{\epsilon} - \frac{1}{\epsilon}\left[1 - \frac{2\epsilon}{\gamma(1 + \sqrt{1 - 4/3\epsilon})}\right]^2.$$

On the other hand, using the independence among different constraints, the violation ratio is $1 - \widehat{\sigma}\epsilon = (1 - 1/(\gamma x^{\mathrm{ind}}))^2$, i.e.,

$$\widehat{\sigma} = \frac{1}{\epsilon} - \frac{1}{\epsilon}\left[1 - \frac{2\epsilon}{\gamma(1 + \sqrt{1 - \epsilon})}\right]^2.$$

Clearly, we have $\widehat{\sigma} < \sigma$. ◇

Finally, we remark that we can further improve the violation ratio in Theorem 6 if for each separated relaxed set X_ϕ^i, we choose a different but a tighter generating function than the one used in the original convex relaxation set X_ϕ^{rel}.

5. Bicriteria Approximation of DRCCP

In this section, we will extend our bicriteria approximation results to distributionally robust chance constrained covering problems (DRCCP) of the form

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \in S, \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) < 1 \right\} \leq \epsilon \right\}. \tag{5}$$

In addition to Assumptions (a)-(d) in Section 2, hereon we will further assume that

(e) the left-hand-side of the uncertain covering constraints $a(\boldsymbol{x},\boldsymbol{\xi}) = \min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}$ are such that, for each $l \in [m]$, $a_l(\boldsymbol{x},\boldsymbol{\xi})$ is convex in $\boldsymbol{\xi}$ for any given $\boldsymbol{x} \in S$.

Note that DRCCP includes an optimization problem over measures and, as shown later in this section, this additional assumption is necessary for the tractability of the associated relaxations.

In general, DRCCP (5) is NP-hard. For example, if the ambiguity set $\mathcal{P}$ contains a single discrete distribution, then (5) reduces to a regular CCP, which is known to be NP-hard. However, the computational complexity of DRCCP for specific ambiguity sets requires further analysis. In the following we will consider two classes of commonly used ambiguity sets, show that the DRCCP is NP-hard in these settings as well, and provide tractable bicriteria approximation algorithms.

In the remainder of this section we revise the definition of X_α introduced in (8) to

$$X_\alpha := \left\{ \boldsymbol{x} \in S : \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) < 1 \right\} \leq \alpha \right\}, \tag{17}$$

for any $\alpha > 0$. Note that X_ϵ is the feasible region of (5). We propose to apply the Relax-and-Scale Algorithm 1 to DRCCP (5). Similar to Section 4.2 we consider the following convex relaxation based on a generating function $\phi \in \Phi$:

$$v_\phi^{\text{rel}} = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \in X_{\mathcal{P},\phi}^{\text{rel}} \right\}, \tag{18}$$

where

$$X_{\mathcal{P},\phi}^{\text{rel}} = \left\{ \boldsymbol{x} \in S : \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\phi(a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}))] \geq (1-\epsilon)\phi(1) \right\}. \tag{19}$$

Note that the left-hand-side of the inequality defining (19) is concave in $\boldsymbol{x}$, hence (18) is a convex optimization problem.

Recall that the bicriteria analysis in Section 4.3 is independent of the underlying distribution, thus we can easily adapt to the distributionally robust setting. The proof of the following result is completely analogous to the proof of Theorem 5, hence is omitted.

Theorem 7 *Given a generating function $\phi \in \Phi$, choose*

- *violation ratio:* $\sigma \in (\frac{1}{\epsilon} - (\frac{1}{\epsilon} - 1)\phi(1), \frac{1}{\epsilon})$,

- *approximation ratio:* $\gamma = \left[\phi^{-1}\left(1 - \frac{1}{\sigma\epsilon} + (\frac{1}{\sigma\epsilon} - \frac{1}{\sigma})\phi(1)\right)\right]^{-1}$, *and*
- *relaxation:* $\widehat{X} = X^{\mathrm{rel}}_{\mathcal{P},\phi}$, *as defined in* (19).

Then Relax-and-Scale $(\sigma, \gamma, \widehat{X})$ *returns a* (σ, γ)*-bicriteria approximate solution to DRCCP.*

Note that the tractability of Relax-and-Scale Algorithm 1 depends on the tractability of solving the convex relaxation (18). Next, we consider two specific classes of ambiguity sets in DRCCP (5) and discuss corresponding complexities and tractable approximations.

5.1. Moment Ambiguity Set

Following Xie and Ahmed (2018b), we consider the following ambiguity set specified by partial moment information

$$\mathcal{P}^M = \left\{\mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\psi_t(\tilde{\boldsymbol{\xi}})] = g_t, t \in \mathcal{T}_1, \ \mathbb{E}_{\mathbb{P}}[\psi_t(\tilde{\boldsymbol{\xi}})] \geq g_t, t \in \mathcal{T}_2, \right\} \tag{20}$$

where $\mathcal{P}_0(\Xi)$ denotes the set of all probability measures on Ξ with a sigma algebra $\mathcal{F}$, $\psi_t(\boldsymbol{\xi})$ is a real valued continuous function on $(\Xi, \mathcal{F})$ for each $t \in \mathcal{T}_1 \cup \mathcal{T}_2$. We further assume that

(i) Ξ is a closed convex set, and for each $t \in \mathcal{T}_1$, $\psi_t(\boldsymbol{\xi})$ is a linear function, and for each $t \in \mathcal{T}_2$, $\psi_t(\boldsymbol{\xi})$ is a concave function.

(ii) (Slater's condition, c.f., Theorem 5.99 in Bonnans and Shapiro 2000) suppose that $\mathrm{int}(\mathcal{P}^M) \neq \emptyset$, i.e., for any sufficiently small perturbation of $\{g_t\}_{t\in\mathcal{T}_1\cup\mathcal{T}_2}$, there exists a feasible probability measure such that the constraints in $\mathcal{P}^M$ hold.

The following result provides a deterministic characterization of the DRCCP set X_α in (17) with $\mathcal{P} = \mathcal{P}^M$. This characterization can be used to further reduce the computational approximation ratio in the Relax-and-Scale scheme, i.e., we can search for the best approximation ratio with lines 4-7 of Algorithm 1.

Theorem 8 *[Theorem 1 in Xie and Ahmed (2018b)]*
For any $\alpha \in (0,1)$*, the DRCCP set* X_α *defined in* (17) *with* $\mathcal{P} = \mathcal{P}^M$ *is equivalent to*

$$X_\alpha = \left\{ \boldsymbol{x} \in S : \begin{array}{ll} \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} g_t\eta_t \geq 1-\alpha, & \text{(21a)} \\ \lambda + \Psi_0(\boldsymbol{\eta}, \boldsymbol{0}) \leq 1, & \text{(21b)} \\ \lambda + \pi_l + \Psi_{a_l}(\boldsymbol{\eta}, \pi_l\boldsymbol{x}) \leq 0, \ \forall l \in I(\boldsymbol{x}), & \text{(21c)} \\ \eta_t \geq 0, \forall t \in \mathcal{T}_2, \pi_l \geq 0, \forall l \in [m], & \text{(21d)} \end{array} \right\}$$

where $I(\boldsymbol{x}) := \{l \in [m] : \exists \boldsymbol{\xi} \in \Xi, a_l(\boldsymbol{x}, \boldsymbol{\xi}) < 1\}$ *and* $\Psi_0(\cdot)$ *and* $\{\Psi_{a_l}(\cdot)\}_{l\in[m]}$ *are defined as*

$$\Psi_0(\boldsymbol{\eta}, \boldsymbol{x}) := \sup_{\boldsymbol{\xi}\in\Xi} \left(\sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} \psi_t(\boldsymbol{\xi})\eta_t \right), \tag{22a}$$

$$\Psi_{a_l}(\boldsymbol{\eta},\boldsymbol{x}) := \sup_{\boldsymbol{\xi}\in\Xi}\left(\sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2}\psi_t(\boldsymbol{\xi})\eta_t - a_l(\boldsymbol{x},\boldsymbol{\xi})\right), \tag{22b}$$

for each $l\in[m]$.

We next establish the complexity of DRCCP under the moment ambiguity set by a reduction to the well-known set covering problem.

Theorem 9 *DRCCP* (5) *with* $\mathcal{P}=\mathcal{P}^M$ *is NP-hard.*

Proof: See e-companion EC.1.5. □

Theorem 9 shows that DRCCP (5) with $\mathcal{P}=\mathcal{P}^M$ is difficult. Thus, we will explore its convex relaxations and study approximation algorithms. The next result provides a deterministic convex reformulation of the relaxation $X^{\text{rel}}_{\mathcal{P},\phi}$ when the ambiguity set is moment based and the generating function ϕ is Markovian.

Theorem 10 *If* $\mathcal{P}=\mathcal{P}^M$ *and* $\phi(s)=\min\{s,1\}$, *then* $X^{\text{rel}}_{\mathcal{P},\phi}$ *is equivalent to*

$$X^{\text{rel}}_{\mathcal{P},\phi} = \left\{ \boldsymbol{x}\in S: \begin{array}{ll} \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} g_t\eta_t \geq 1-\epsilon, & \text{(23a)}\\ \lambda + \Psi_0(\boldsymbol{\eta},\boldsymbol{0}) \leq 1, & \text{(23b)}\\ \lambda + \Psi_{a_l}(\boldsymbol{\eta},\boldsymbol{x}) \leq 0,\ \forall l\in[m], & \text{(23c)}\\ \eta_t \geq 0, \forall t\in\mathcal{T}_2, & \text{(23d)} \end{array} \right\}$$

where $\Psi_0(\cdot)$ *and* $\{\Psi_{a_l}(\cdot)\}_{l\in[m]}$ *are defined in* (22).

Proof: If $\mathcal{P}=\mathcal{P}^M$ and the generation function $\phi(s)=\min\{s,1\}$, then the set $X^{\text{rel}}_{\mathcal{P},\phi}$ in (19) is equivalent to

$$X^{\text{rel}}_{\mathcal{P},\phi} = \left\{\boldsymbol{x}\in S: \inf_{\mathbb{P}\in\mathcal{P}^M}\mathbb{E}_{\mathbb{P}}[\min\{1,a(\boldsymbol{x},\tilde{\boldsymbol{\xi}})\}] \geq (1-\epsilon)\right\}. \tag{24}$$

Using Lemma 1 in Xie and Ahmed (2018b), the inner infimum in (24) is equivalent to

$$\begin{aligned} \inf_{\mathbb{P}\in\mathcal{P}^M}\mathbb{E}_{\mathbb{P}}[\min\{1,a(\boldsymbol{x},\tilde{\boldsymbol{\xi}})\}] = \quad & \max_{\lambda,\boldsymbol{\eta}} \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} g_t\eta_t \\ \text{s.t.} \quad & \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2}\psi_t(\boldsymbol{\xi})\eta_t \leq \min\left\{1, \min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}\right\}, \forall\boldsymbol{\xi}\in\Xi,\\ & \eta_t\geq 0, \forall t\in\mathcal{T}_2. \end{aligned}$$

By replacing max operator with existence of $\lambda,\boldsymbol{\eta}$, the set $X^{\text{rel}}_{\mathcal{P},\phi}$ becomes

$$X^{\text{rel}}_{\mathcal{P},\phi} = \left\{ \boldsymbol{x}\in S: \begin{array}{ll} \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} g_t\eta_t \geq 1-\epsilon, & \text{(25a)}\\ \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2}\psi_t(\boldsymbol{\xi})\eta_t \leq \min\left\{1,\min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}\right\}, \forall\boldsymbol{\xi}\in\Xi, & \text{(25b)}\\ \eta_t\geq 0, \forall t\in\mathcal{T}_2. & \text{(25c)} \end{array} \right\}$$

Next in (25b), we note that the left-hand side should be no larger than any of $1, \{a_l(\boldsymbol{x},\boldsymbol{\xi})\}_{l\in[m]}$, thus, (25) is equivalent to

$$X^{\mathrm{rel}}_{\mathcal{P},\phi} = \left\{ \boldsymbol{x}\in S: \begin{array}{ll} \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} g_t\eta_t \geq 1-\epsilon, & \text{(26a)} \\ \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} \psi_t(\boldsymbol{\xi})\eta_t \leq 1, \forall \boldsymbol{\xi}\in\Xi, & \text{(26b)} \\ \lambda + \sum_{t\in\mathcal{T}_1\cup\mathcal{T}_2} \psi_t(\boldsymbol{\xi})\eta_t \leq a_l(\boldsymbol{x},\boldsymbol{\xi}), \forall \boldsymbol{\xi}\in\Xi, l\in[m], & \text{(26c)} \\ \eta_t \geq 0, \forall t\in\mathcal{T}_2. & \text{(26d)} \end{array} \right\}$$

Finally due to the definition of function $\Psi_0(\cdot)$ and $\{\Psi_{a_l}(\cdot)\}_{l\in[m]}$, we arrive at (23). □

Note that set $X^{\mathrm{rel}}_{\mathcal{P},\phi}$ in (23) is convex but optimization over it might not be tractable. Thus, the following result summarizes the conditions of tractability of convex relaxation (18) and approximation guarantees for DRCCP with moment ambiguity set. The result follows from Theorem 3.1 in Grötschel et al. (1981), and Theorems 7 and 10.

Corollary 3 *Suppose that*

- $\mathcal{P}=\mathcal{P}^M$,
- *the Markovian generating function* $\phi(s)=\min\{s,1\}$ *is chosen,*
- *the violation ratio is* $\sigma\in(1,\frac{1}{\epsilon})$,
- *the relaxation set is* $\widehat{X}=X^{\mathrm{rel}}_{\mathcal{P},\phi}$, *and*
- *separation over set* S *and the constraints* (23b) *and* (23c) *can be done in time that is polynomial in* n,m, *i.e. for any* $(\lambda,\boldsymbol{\eta},\boldsymbol{x})$ *such that* $\boldsymbol{x}$ *is not contained in set* S, *or* $(\lambda,\boldsymbol{\eta},\boldsymbol{x})$ *violates* (23b) *or* (23c), *one can find a hyperplane in polynomial time which separates* $\boldsymbol{x}$ *from set* S, *or* $(\lambda,\boldsymbol{\eta},\boldsymbol{x})$ *from constraints* (23b) *and* (23c).

Then

(i) the convex relaxation (18) *is tractable; and*

(ii) Relax-and-Scale $(\sigma,\frac{\sigma}{\sigma-1},\widehat{X})$ *returns a* $(\sigma,\frac{\sigma}{\sigma-1})$*-bicriteria approximate solution to DRCCP.*

The following example illustrates an application of Theorem 10.

Example 2 We consider the distributionally robust counterpart of chance constrained linear covering problem (3) as below:

$$v^* = \min_{\boldsymbol{x}\in\mathbb{R}^n_+} \left\{ \boldsymbol{c}^\top\boldsymbol{x} : \inf_{\mathbb{P}\in\mathcal{P}^M} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}\boldsymbol{x}\geq\boldsymbol{e}\right\} \geq 1-\epsilon \right\}. \tag{27a}$$

In the above problem, there are n items and m covering constraints. For each item $j\in[n]$, c_j represents its cost, and let $\tilde{\xi}_{lj}$ denote its random resource to cover lth constraint. Let x_j define the quantity of jth item being used. Suppose that we do not know the true probability distribution, but know the mean of each resource and the upper bound of their total deviation from the mean, thus $\mathcal{P}^M$ is defined as

$$\mathcal{P}^M = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\Xi\right) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[\|\tilde{\boldsymbol{\xi}} - \boldsymbol{\mu}\|_F] \leq g \right\}, \tag{27b}$$

where $\Xi = \mathbb{R}_+^{n\times m}$, $\boldsymbol{\mu} \in \mathbb{R}_{++}^{n\times m}$ and $\|\cdot\|_F$ denotes Frobenius norm.

In this example, the Markovian relaxation set $X^{\text{rel}}_{\mathcal{P},\phi}$ is equivalent to

$$X^{\text{rel}}_{\mathcal{P},\phi} = \left\{ \boldsymbol{x} \in \mathbb{R}^n_+ : \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}\left[\min\left\{ 1, \min_{l\in[m]} \{\tilde{\boldsymbol{\xi}}_{l\bullet}^\top \boldsymbol{x}\} \right\} \right] \geq 1-\epsilon \right\}, \tag{27c}$$

where $\boldsymbol{A}_{l\bullet}$ denotes the vector of lth row of matrix $\boldsymbol{A}$. By Theorem 10, the Markovian relaxation set $X^{\text{rel}}_{\mathcal{P},\phi}$ can be formulated as the following second order cone program (SOCP)

$$X^{\text{rel}}_{\mathcal{P},\phi} = \left\{ \boldsymbol{x} \in \mathbb{R}^n_+ : \begin{array}{l} \lambda + \langle \boldsymbol{\mu}, \boldsymbol{\eta}_1 \rangle - g\eta_2 \geq 1-\epsilon, \\ \lambda + \langle \boldsymbol{\mu}, \boldsymbol{\eta}_1 + \boldsymbol{\beta}_1 \rangle \leq 1, \\ \|\boldsymbol{\eta}_1 + \boldsymbol{\beta}_1\|_F \leq \eta_2, \\ \lambda + \langle \boldsymbol{\mu}, \boldsymbol{\eta}_1 + \boldsymbol{\beta}_2^l \rangle \leq \boldsymbol{\mu}_{l.}^\top \boldsymbol{x}, \forall l \in [m], \\ \sqrt{\sum_{\tau\in[m],\tau\neq l} \|(\boldsymbol{\eta}_1)_{\tau\bullet} + (\boldsymbol{\beta}_2^l)_{\tau\bullet}\|_2^2 + \|(\boldsymbol{\eta}_1)_{l\bullet} + (\boldsymbol{\beta}_2^l)_{l\bullet} - \boldsymbol{x}\|_2^2} \leq \eta_2, \forall l \in [m], \\ \boldsymbol{\eta}_1 \in \mathbb{R}^{m\times n}, \eta_2 \in \mathbb{R}_+, \boldsymbol{\beta}_1 \in \mathbb{R}_+^{m\times n}, \boldsymbol{\beta}_2^l \in \mathbb{R}_+^{m\times n}, \forall l \in [m], \end{array} \right\} \tag{27d}$$

where for any two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m,n}$, their inner product is defined as $\langle \boldsymbol{A}, \boldsymbol{B} \rangle := \text{trace}(AB^\top)$. Clearly, the Markovian relaxation of the DRCCP (27a), i.e., $\min_{\boldsymbol{x}\in X^{\text{rel}}_{\mathcal{P},\phi}} \boldsymbol{c}^\top \boldsymbol{x}$, is tractable. $\diamond$

5.2. Wasserstein Ambiguity Set

One disadvantage of a moment-based ambiguity set is that it might be unrealistic to estimate many moments required to describe the ambiguity set. Furthermore, even if all moment information for a "true" distribution is included, the corresponding ambiguity set is not guaranteed to recover the true distribution (c.f., Durrett 1996). Recently, data driven distributionally robust optimization problems with Wasserstein metric have been investigated (e.g., Esfahani and Kuhn 2015, Zhao and Guan 2018, and Gao and Kleywegt 2016). In this framework the ambiguity set considers all distributions within a Wasserstein ball of specified radius around an empirical distribution. It is known that as the sample size of the empirical distribution goes to infinity, the Wasserstein ambiguity set recovers the true generating distribution. The Wasserstein ambiguity set is

$$\mathcal{P}^W = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\Xi\right) : W_q\left(\mathbb{P}, \mathbb{P}_{\tilde{\zeta}}\right) \leq \delta \right\}, \tag{28}$$

where Wasserstein distance is defined as

$$W_q(\mathbb{P}_1,\mathbb{P}_2)=\inf\left\{\int_{\Xi\times\Xi}\|\boldsymbol{\xi}_1-\boldsymbol{\xi}_2\|_q\mathbb{Q}(d\boldsymbol{\xi}_1,d\boldsymbol{\xi}_2):\begin{array}{l}\mathbb{Q}\text{ is a joint distribution of }\tilde{\boldsymbol{\xi}}_1\text{ and }\tilde{\boldsymbol{\xi}}_2\\ \text{with marginals }\mathbb{P}_1\text{ and }\mathbb{P}_2\text{, respectively}\end{array}\right\},$$

and $\mathbb{P}_{\tilde{\zeta}}$ denotes a discrete empirical distribution of $\tilde{\boldsymbol{\zeta}}$ on the finite support $\mathcal{Z}=\{\boldsymbol{\zeta}_i\}_{i\in[N]}\subseteq\Xi$ with point mass function $\{p_i\}_{i\in[N]}$, and $q\geq 1$.

The following result provides a deterministic characterization of DRCCP set X_α in (17) with $\mathcal{P}=\mathcal{P}^W$. This characterization will be used in the subsequent developments and also to further reduce the computational approximation ratio in the Relax-and-Scale scheme, i.e., we can search for the best approximation ratio with steps 4-7 in Algorithm 1.

Theorem 11 *For any $\alpha\in(0,1)$, the DRCCP set X_α in (17) with $\mathcal{P}=\mathcal{P}^W$ is equivalent to*

$$X_\alpha=\left\{\boldsymbol{x}\in S:\begin{array}{l}\lambda\delta+\sum_{i\in[N]}p_i\eta_i\leq\alpha, \qquad (29a)\\ \eta_i+\inf_{\xi\in\Xi}\left[\lambda\|\boldsymbol{\xi}-\boldsymbol{\zeta}_i\|_q+\pi_{il}a_l(\boldsymbol{x},\boldsymbol{\xi})\right]\geq 1+\pi_{il},\forall i\in[N],l\in I(x) \qquad (29b)\\ \lambda\geq 0,\eta_i\geq 0,\pi_{il}\geq 0,\forall i\in[N],\forall l\in[m] \qquad (29c)\end{array}\right\}$$

where $I(\boldsymbol{x}):=\{l\in[m]:\exists\boldsymbol{\xi}\in\Xi,a_l(\boldsymbol{x},\boldsymbol{\xi})<1\}$.

Proof: See e-companion EC.1.6. □

Note that in (29), there are $O(Nm)$ variables and constraints. Thus, checking the feasibility of set X_α (i.e., steps 4-7 in Algorithm 1) might be difficult if the sample size N is very large. In this case, we might want to skip steps 4-7 in Algorithm 1.

We next establish the complexity of DRCCP under the Wasserstein ambiguity set by a reduction to the regular chance constrained linear covering problem (Qiu et al. 2014).

Theorem 12 *DRCCP (5) with $\mathcal{P}=\mathcal{P}^W$ is NP-hard.*

Proof: See e-companion EC.1.7. □

Theorem 12 suggests DRCCP (5) with $\mathcal{P}=\mathcal{P}^W$ is hard. Thus, in the remainder of this section, we will focus on deriving its convex relaxation. The next result provides a deterministic convex reformulation of the relaxation $X^{\text{rel}}_{\mathcal{P},\phi}$ for the Wasserstein ambiguity set and the Markovian generating function.

Theorem 13 *If $\mathcal{P}=\mathcal{P}^W$ and $\phi(s)=\min\{s,1\}$, then $X^{\text{rel}}_{\mathcal{P},\phi}$ is equivalent to*

$$X^{\text{rel}}_{\mathcal{P},\phi}=\left\{\boldsymbol{x}\in S:\begin{array}{l}\delta\lambda+\sum_{i\in[N]}p_i\eta_i\leq\epsilon, \qquad (30a)\\ \eta_i+\inf_{\boldsymbol{\xi}\in\Xi}\left[\lambda\|\boldsymbol{\xi}-\boldsymbol{\zeta}_i\|_q+a_l(\boldsymbol{x},\boldsymbol{\xi})\right]\geq 1,\forall i\in[N],l\in[m], \qquad (30b)\\ \lambda\geq 0,\eta_i\geq 0,\forall i\in[N]. \qquad (30c)\end{array}\right\}$$

Proof: If $\mathcal{P}=\mathcal{P}^W$ and the generation function $\phi(s)=\min\{s,1\}$, then the set $X^{\text{rel}}_{\mathcal{P},\phi}$ in (19) is equivalent to

$$X^{\text{rel}}_{\mathcal{P},\phi}=\left\{\boldsymbol{x}\in S:\ \sup_{\mathbb{P}\in\mathcal{P}^W}\mathbb{E}_{\mathbb{P}}[1-\min\{1,a(\boldsymbol{x},\tilde{\boldsymbol{\xi}})\}]\leq\epsilon\right\}. \tag{31}$$

By Theorem 1 in Gao and Kleywegt (2016) and $a(\boldsymbol{x},\boldsymbol{\xi})=\min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}$, the optimization problem $\sup_{\mathbb{P}\in\mathcal{P}^W}\mathbb{E}_{\mathbb{P}}[1-\min\{1,a(\boldsymbol{x},\tilde{\boldsymbol{\xi}})\}]$ is equivalent to

$$\min_{\lambda\geq0}\lambda\delta-\sum_{i\in[N]}p_i\inf_{\xi\in\Xi}\left[\lambda\|\boldsymbol{\xi}-\boldsymbol{\zeta}_i\|_q-[1-\min\{1,\min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}\}]\right].$$

By replacing min with existence of some $\lambda\geq0$, the set $X^{\text{rel}}_{\mathcal{P},\phi}$ becomes

$$X^{\text{rel}}_{\mathcal{P},\phi}=\left\{\boldsymbol{x}\in S:\begin{array}{ll}\lambda\delta-\sum_{i\in[N]}p_i\inf_{\xi\in\Xi}\left[\lambda\|\boldsymbol{\xi}-\boldsymbol{\zeta}_i\|_q-[1-\min\{1,\min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}\}]\right]\leq\epsilon, & \text{(32a)}\\ \lambda\geq0. & \text{(32b)}\end{array}\right\}$$

Next introduce new variable η_i and let $\eta_i\geq-\inf_{\xi\in\Xi}\left[\lambda\|\boldsymbol{\xi}-\boldsymbol{\zeta}_i\|_q-[1-\min\{1,\min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}\}]\right]$ for each $i\in[N]$. Then the set $X^{\text{rel}}_{\mathcal{P},\phi}$ becomes

$$X^{\text{rel}}_{\mathcal{P},\phi}=\left\{\boldsymbol{x}\in S:\begin{array}{ll}\lambda\delta+\sum_{i\in[N]}p_i\eta_i\leq\epsilon, & \text{(33a)}\\ \eta_i+\inf_{\xi\in\Xi}\left[\lambda\|\boldsymbol{\xi}-\boldsymbol{\zeta}_i\|_q+\min\{1,\min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}\}\right]\geq1,\forall i\in[N], & \text{(33b)}\\ \lambda\geq0. & \text{(33c)}\end{array}\right\}$$

By breaking down the "min" operators in (33b), we arrive at (30). □

The following result summarizes conditions of tractability of the convex relaxation (18) and approximation guarantees for DRCCP with Wasserstein ambiguity set. The result follows from Theorem 3.1 in Grötschel et al. (1981) and Theorem 7 and Theorem 13 in this paper.

Corollary 4 *Suppose that*

- $\mathcal{P}=\mathcal{P}^W$,
- *the Markovian generating function* $\phi(s)=\min\{s,1\}$ *is chosen,*
- *the violation ratio is* $\sigma\in(1,\frac{1}{\epsilon})$,
- *relaxation set is* $\widehat{X}=X^{\text{rel}}_{\mathcal{P},\phi}$, *and*
- *separation over the set* S *and constraints* (30b) *can be done in time that is polynomial in* n,m,N, *i.e. for any* $(\lambda,\boldsymbol{\eta},\boldsymbol{x})$ *such that* $\boldsymbol{x}$ *is not contained in set* S, *or* $(\lambda,\boldsymbol{\eta},\boldsymbol{x})$ *that violates constraints* (30b), *one can find a hyperplane in polynomial time which separates* $(\lambda,\boldsymbol{\eta},\boldsymbol{x})$ *from set* S *or constraints* (30b).

Then

(i) the convex relaxation (18) *is tractable; and*

(ii) Relax-and-Scale ($\sigma, \frac{\sigma}{\sigma-1}, \widehat{X}$) *returns a* ($\sigma, \frac{\sigma}{\sigma-1}$)*-bicriteria approximate solution to DRCCP.*

Recently Xie (2018) and Chen et al. (2018) proposed the following worst-case $\mathbf{CVaR}$ approximation of DRCCP set X_α with $\mathcal{P} = \mathcal{P}^W$.

Corollary 5 *(Xie 2018, Chen et al. 2018) The DRCCP set* X_α *with* $\mathcal{P} = \mathcal{P}^W$ *can be inner approximated by the following convex set*

$$X_\alpha^{\mathbf{CVaR}} = \left\{ \boldsymbol{x} \in S : \begin{array}{l} \lambda\delta + \sum_{i\in[N]} p_i \eta_i \leq \alpha\beta, \\ \eta_i + \inf_{\xi\in\Xi} [\lambda \|\boldsymbol{\xi} - \boldsymbol{\zeta}_i\|_q + a_l(\boldsymbol{x}, \boldsymbol{\xi})] \geq 1 + \beta, \forall i \in [N], l \in [I], \\ \lambda \geq 0, \eta_i \geq 0, \forall i \in [N]. \end{array} \right\} \begin{array}{r} (34a) \\ (34b) \\ (34c) \end{array}$$

Note that the above worst-case $\mathbf{CVaR}$ approach provides an inner approximation of the DRCCP and this approximation is tight under some conditions (Xie 2018, Chen et al. 2018). Thus it can lead to good-quality feasible solutions, while our proposed approach uses an outer approximation (a relaxation) and scales its solution to obtain a feasible solution. Next, we illustrate that the quality of solutions obtained by these two approaches is incomparable. In particular, Examples EC.1 and EC.2 in e-companion show that the approximation quality of the worst-case $\mathbf{CVaR}$ approximation can be arbitrarily bad while the theoretical approximation ratio of the proposed Relax-and-Scale scheme can be equal to 1, and vice versa. We also provide a numerical comparison of these approaches in Section 6.

6. Numerical Illustration

In this section, we provide numerical illustrations of the proposed Relax-and-Scale approximation algorithm. We consider a Chance Constrained Portfolio Optimization (CC-PO) problem studied in Pagnoncelli et al. (2009) and Qiu et al. (2014), which is to minimize the investment cost in a portfolio of n assets while guaranteeing a specified return level. CC-PO has been studied extensively in literature and has been reported to be very difficult to solve to the optimality even for small-scale instances (Qiu et al. 2014). To model CC-PO as CCP or DRCCP, we let $\tilde{\boldsymbol{\xi}} \in \mathbb{R}_+^n$ represent the (normalized) random return vector and x_j denote investment on asset $j \in [n]$. Then, its corresponding CCP formulation is

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \geq 0, \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}^\top \boldsymbol{x} \geq 1 \right\} \geq 1 - \epsilon \right\}, \tag{35}$$

and its DRCCP analog is

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \geq 0, \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}^\top \boldsymbol{x} \geq 1 \right\} \geq 1 - \epsilon \right\}. \tag{36}$$

In (35) and (36), the objective is to minimize the total investment cost, while the chance constraint is to guarantee that with probability at least $(1-\epsilon)$, the total return is at least 1. In the following two subsections, we use this example to illustrate the approximation ratios for CCP with discrete support and DRCCP under Wasserstein ambiguity set. All computations were executed on a laptop with a 2.67 GHz processor and 4GB RAM, with CPLEX 12.5.1 with default settings as the solver. The running time of Relax-and-Scale Algorithm 1 is within 60 seconds for all the instances.

6.1. CCP with Discrete Support

Here we assume that $\tilde{\boldsymbol{\xi}}$ has N equiprobable realizations denoted by $\{\boldsymbol{\xi}^i\}_{i\in[N]}$. Then the CCP (35) can be formulated as the following mixed integer program

$$v^* = \min_{\boldsymbol{x},\boldsymbol{z}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{x} \geq 0,\ \sum_{j=1}^{n} \xi_j^i x_j \geq z_i, \forall i \in [N],\ \boldsymbol{z} \in Z \right\}, \tag{37}$$

where $Z := \left\{ \boldsymbol{z} \in \{0,1\}^N :\ \sum_{i\in[N]} z_i \geq (1-\epsilon)N \right\}$. We generated the data in the similar way as Qiu et al. (2014), where $n = 50$, $N = 100$, $\{\tilde{\xi}_j\}_{j\in[n]}$ are i.i.d. uniform random variables in the range from 0.8 to 1.2 and the risk parameter $\epsilon \in \{0.05, 0.10\}$.

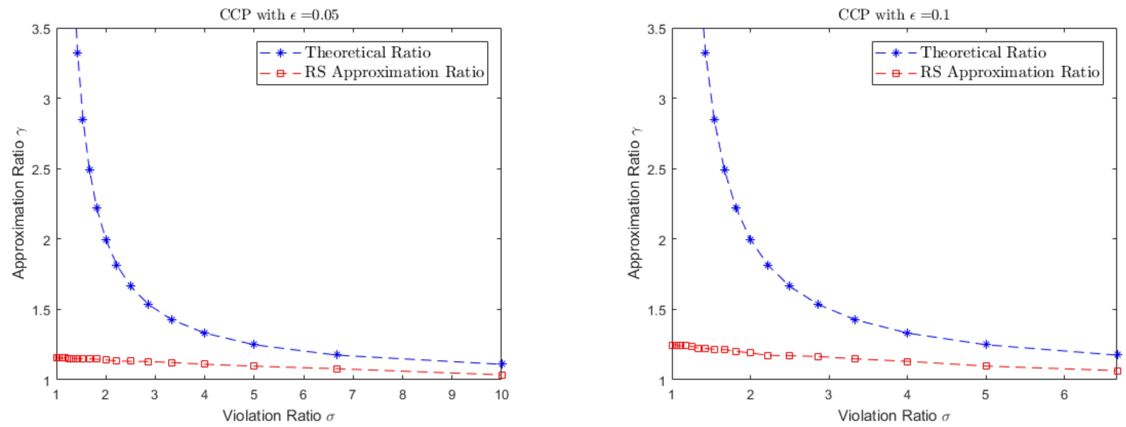
The computational results are shown in Figure 2. For any violation ratio $\sigma \in [1, 1/\epsilon)$, we let $\widehat{v}_\sigma$ denote the objective value of a solution from the Relax-and-Scale scheme (Algorithm 1) using the Markovian relaxation. Recall that in this case, the relaxation corresponds to the linear programming relaxation of the mixed integer program (37). Let $v^{\rm rel}$ denote the optimal value of the relaxation. In Figure 2, the blue starred curve denotes the theoretical approximation ratio $\gamma = \frac{\sigma}{\sigma-1}$ proposed in Corollary 1, and red squared curve denoted as RS (i.e., short for Relax-and-Scale) approximation ratio $\frac{\widehat{v}_\sigma}{v^{\rm rel}}$ is an upper bound on $\frac{\widehat{v}_\sigma}{v^*}$ since we could not solve the mixed integer program (37) to optimality within a 12-hour time limit.

We observe that as the violation ratio σ tends to 1, the theoretical approximation ratio (starred curve) tends to infinity. However, the RS approximation ratio is only off by a factor of at most 1.3 from the true optimal for both $\epsilon = 0.05$ and 0.1. When σ increases, the theoretical approximation ratio (starred curve) decreases dramatically. Typically, if we choose $\sigma = 2$, then the theoretical approximation ratio (starred curve) is around 2 but the RS ratio (squared curve) is around 1.2. However, when $\sigma > 2$, the RS approximation ratio does not improve too much. Thus, for these instances, we suggest to choose $\sigma \in (1, 2]$.

6.2. DRCCP under Wasserstein Ambiguity Set

Here we consider the DRCCP (36) with Wasserstein ambiguity set defined with respect to an empirical distribution. Using Theorem 11, this problem can be formulated as

$$v^* = \min_{\boldsymbol{x}\in\mathbb{R}^n_+} \quad \boldsymbol{c}^\top \boldsymbol{x},$$



(a) $\epsilon = 0.05$ (b) $\epsilon = 0.10$

Figure 2 Illustration of approximation ratios for CCP with discrete support.

$$\begin{aligned} \text{s.t.} \quad & \delta\lambda + \sum_{i\in[N]} p_i\eta_i \leq \epsilon, \\ & \eta_i + (\boldsymbol{\zeta}^i)^\top(\pi_i \boldsymbol{x} - \boldsymbol{\rho}_i) \geq 1 + \pi_i, \forall i \in [N], \\ & \|\pi_i \boldsymbol{x} - \boldsymbol{\rho}_i\|_2 \leq \lambda, \forall i \in [N], \\ & \lambda \geq 0, \eta_i \geq 0, \boldsymbol{\rho}_i \in \mathbb{R}^n_+, \pi_i \geq 0, \forall i \in [N], \end{aligned} \tag{38}$$

where in (28), we let $q = 2$. Note that (38) is a nonconvex optimization problem due to the bilinear terms $\{\alpha_i x\}_{i\in[N]}$, however, we can use this formulation in the feasibility check steps 4-7 in the Relax-and-Scale scheme (Algorithm 1). Using Theorem 12, the Markovian relaxation of (38) is formulated as

$$\begin{aligned} v^{\mathrm{rel}} = \min_{\boldsymbol{x}\in\mathbb{R}^n_+} \quad & \boldsymbol{c}^\top \boldsymbol{x}, \\ \text{s.t.} \quad & \delta\lambda + \sum_{i\in[N]} p_i\eta_i \leq \epsilon, \\ & \eta_i + (\boldsymbol{\zeta}^i)^\top(\boldsymbol{x} - \boldsymbol{\rho}_i) \geq 1, \forall i \in [N], \\ & \|\boldsymbol{x} - \boldsymbol{\rho}_i\|_2 \leq \lambda, \forall i \in [N], \\ & \lambda \geq 0, \eta_i \geq 0, \boldsymbol{\rho}_i \in \mathbb{R}^n_+, \forall i \in [N], \end{aligned} \tag{39}$$

which is a second order cone program (SOCP). From Corollary 5, the worst-case **CVaR** approximation is formulated as

$$\begin{aligned} v^{\mathbf{CVaR}} = \min_{\boldsymbol{x}\in\mathbb{R}^n_+} \quad & \boldsymbol{c}^\top \boldsymbol{x}, \\ \text{s.t.} \quad & \delta\lambda + \sum_{i\in[N]} p_i\eta_i \leq \epsilon\beta, \\ & \eta_i + (\boldsymbol{\zeta}^i)^\top(\boldsymbol{x} - \boldsymbol{\rho}_i) \geq 1 + \beta, \forall i \in [N], \end{aligned} \tag{40}$$

$$\|\boldsymbol{x}-\boldsymbol{\rho}_i\|_2 \leq \lambda, \forall i \in [N],$$
$$\lambda \geq 0, \eta_i \geq 0, \boldsymbol{\rho}_i \in \mathbb{R}^n_+, \forall i \in [N].$$

In the numerical study, we let $n=50$ and the risk parameter $\epsilon \in \{0.20, 0.25\}$. It turns out that if we choose a smaller risk parameter (e.g., $\epsilon = 0.05, 0.10$), the $\mathbf{CVaR}$ approximation (40) is infeasible. Thus, for the sake of comparison, we select a relatively larger risk parameter. The empirical distribution $\mathbb{P}_{\tilde{\zeta}}$ was constructed by $N=100$ i.i.d. samples from the random vector $\tilde{\boldsymbol{\xi}}$ (i.e., $p_i = \frac{1}{N}$ for all i), where $\{\tilde{\xi}_j\}_{j\in[n]}$ are i.i.d. uniform random variables in the range from 0.8 to 1.2. We computed the radius of Wasserstein ball via simulation as follows: we ran 100 simulations, and for each simulation, we generated 100 i.i.d. samples from the random vector $\tilde{\boldsymbol{\xi}}$, and computed its Wasserstein distance with the empirical distribution. Finally, we set δ to be 90th, 95th and 99th percentiles of the simulated Wasserstein distances, i.e., $\delta \in \{1.3811, 1.3837, 1.3855\}$.

The computational results are shown in Table 1 and Figure 3. For any violation ratio $\sigma \in [1, 1/\epsilon)$, we let $\widehat{v}_\sigma$ denote the objective value of the solution from the Relax-and-Scale scheme (Algorithm 1) using the Markovian relaxation (39), and let $v^{\mathbf{CVaR}}_\sigma$ denote the optimal value of the worst-case $\mathbf{CVaR}$ approximation (40) with the risk parameter $\sigma\epsilon$. In Table 1 and Figure 3, since it is difficult to obtain v^*, the RS (i.e., short for Relax-and-Scale) approximation ratio is computed by $\frac{\widehat{v}_\sigma}{v^{\mathrm{rel}}}$, while the worst-case $\mathbf{CVaR}$ approximation ratio is computed by $\frac{v^{\mathbf{CVaR}}_\sigma}{v^{\mathrm{rel}}}$. Moreover, in Figure 3, we let the blue starred curve denote the theoretical approximation ratio $\gamma = \frac{\sigma}{\sigma-1}$ from Corollary 4, let the red squared curve denote as RS approximation ratio $\frac{\widehat{v}_\sigma}{v^{\mathrm{rel}}}$, and let the black curve with triangular markers denote the worst-case $\mathbf{CVaR}$ approximation ratio $\frac{v^{\mathbf{CVaR}}_\sigma}{v^{\mathrm{rel}}}$.

We observe that when the violation ratio σ tends to 1, the theoretical approximation ratio tends to infinity, however, the RS approximation ratio or the worst-case $\mathbf{CVaR}$ approximation ratio is around 25 when the risk parameter $\epsilon = 0.20$, and is around 5 when the risk parameter $\epsilon = 0.25$. When σ increases, the theoretical approximation ratio, RS approximation ratio and the worst-case $\mathbf{CVaR}$ approximation ratio decrease dramatically. Typically, if we choose $\sigma = 2$, then the theoretical approximation ratio is 2 but the RS approximation ratio is around 1.1-1.3, which is much smaller than the one with $\sigma = 1$, while the worst-case $\mathbf{CVaR}$ approximation ratio is around 1.7-1.9. Indeed, we see that when σ grows, the RS approximation ratio is consistently smaller than the worst-case $\mathbf{CVaR}$ approximation ratio. This demonstrates the effectiveness of the proposed Relax-and-Scale scheme. When $\sigma > 2$, the RS approximation ratio can be even smaller than 1, which might be because the relaxed set $X_{\sigma\epsilon}$ is too large compared to the original feasible region X_ϵ. Thus, for these instances of DRCCP, it is reasonable to choose $\sigma \in (1, 2]$ where the chance constraint is relaxed a little bit and in practice, the proposed objective value is quite close to the true optimal one. We also notice that when ϵ becomes larger (i.e., with higher risks), both the

RS approximation ratio and the worst-case $\mathbf{CVaR}$ approximation ratio tend to be smaller. On the other hand, when δ grows (i.e., distance between empirical probability distribution and true one increases), both the RS approximation ratio and the worst-case $\mathbf{CVaR}$ approximation ratio increase as well, and are closer to the theoretical approximation ratio.

Table 1 Computational results for DRCCP with Wasserstein ambiguity set.

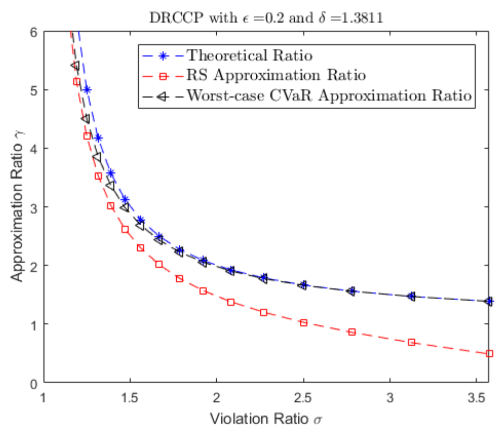
Violation Ratio σ			1.00	1.04	1.09	1.14	1.19	1.25	1.32	1.39	1.47	1.56	1.67	1.79	1.92	2.08	2.27	2.50	2.78	3.13	3.57
Theoretical Ratio $\gamma = \sigma/(\sigma-1)$			Inf	24.99	12.50	8.33	6.25	5.00	4.17	3.57	3.12	2.78	2.50	2.27	2.08	1.92	1.79	1.67	1.56	1.47	1.39
RS Approx. Ratio	$\delta = 1.3811$	$\epsilon = 0.20$	24.47	13.10	8.78	6.52	5.14	4.20	3.53	3.02	2.62	2.29	2.02	1.78	1.57	1.38	1.20	1.03	0.86	0.68	0.49
		$\epsilon = 0.25$	4.89	4.24	3.73	3.31	2.96	2.66	2.40	2.17	1.96	1.77	1.60	1.43	1.27	1.11	0.96	0.79	0.62	0.42	0.20
	$\delta = 1.3837$	$\epsilon = 0.20$	25.39	13.36	8.89	6.58	5.17	4.22	3.54	3.03	2.63	2.30	2.02	1.78	1.57	1.38	1.20	1.03	0.86	0.68	0.49
		$\epsilon = 0.25$	4.92	4.26	3.74	3.32	2.96	2.66	2.40	2.17	1.96	1.77	1.60	1.43	1.27	1.11	0.96	0.79	0.62	0.42	0.20
	$\delta = 1.3855$	$\epsilon = 0.20$	26.09	13.55	8.97	6.62	5.19	4.24	3.55	3.04	2.63	2.30	2.02	1.78	1.57	1.38	1.20	1.03	0.86	0.68	0.49
		$\epsilon = 0.25$	4.94	4.28	3.75	3.33	2.97	2.67	2.41	2.17	1.97	1.78	1.60	1.43	1.27	1.12	0.96	0.79	0.62	0.42	0.20
Worst-case $\mathbf{CVaR}$ Approx. Ratio	$\delta = 1.3811$	$\epsilon = 0.20$	24.07	13.17	8.97	6.76	5.41	4.50	3.85	3.36	2.98	2.68	2.43	2.23	2.05	1.90	1.78	1.66	1.56	1.47	1.39
		$\epsilon = 0.25$	4.84	4.26	3.81	3.44	3.14	2.88	2.67	2.48	2.32	2.17	2.05	1.93	1.83	1.74	1.66	1.58	1.51	1.44	1.38
	$\delta = 1.3837$	$\epsilon = 0.20$	24.98	13.44	9.09	6.82	5.45	4.52	3.87	3.37	2.99	2.69	2.44	2.23	2.06	1.91	1.78	1.66	1.56	1.47	1.39
		$\epsilon = 0.25$	4.87	4.28	3.82	3.45	3.15	2.89	2.67	2.48	2.32	2.18	2.05	1.94	1.83	1.74	1.66	1.58	1.51	1.44	1.38
	$\delta = 1.3855$	$\epsilon = 0.20$	25.66	13.63	9.17	6.87	5.47	4.54	3.88	3.38	3.00	2.69	2.44	2.23	2.06	1.91	1.78	1.66	1.56	1.47	1.39
		$\epsilon = 0.25$	4.89	4.30	3.83	3.46	3.15	2.90	2.68	2.49	2.32	2.18	2.05	1.94	1.84	1.74	1.66	1.58	1.51	1.44	1.38

7. Conclusion

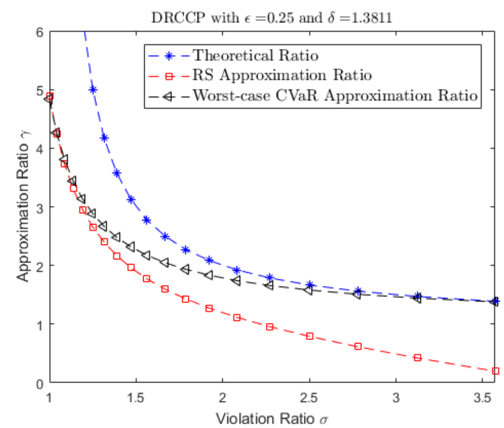
We show that while chance constrained and distributionally robust chance constrained optimization problems are in general inapproximable, their covering analogs admit bicriteria approximation guarantees. That is, by relaxing the chance constraint by a constant factor it is possible to obtain a solution which is within a constant factor of the optimal objective value. In particular we show that, for several important cases, a simple scheme that solves a tractable convex relaxation of the problem and then scales its solution, produces a solution that violates the chance constraint by at most a chosen factor $\sigma > 1$ and is guaranteed to be within a factor $\frac{\sigma}{\sigma-1}$ of the optimal value. It should be noted that if $\sigma = 1$, i.e. no violation of the chance constraint is allowed, the proposed approximation scheme can lead to an arbitrarily bad approximation ratio as predicted by the inapproximability result and the performance analysis. For future research, we would like to investigate if more sophisticated relaxations (e.g., the nonanticipative relaxation proposed in Ahmed et al. 2017) can be used to improve the approximation analysis. More generally, it would be valuable to extend the analysis to chance constraints beyond the covering systems considered here.

Acknowledgment

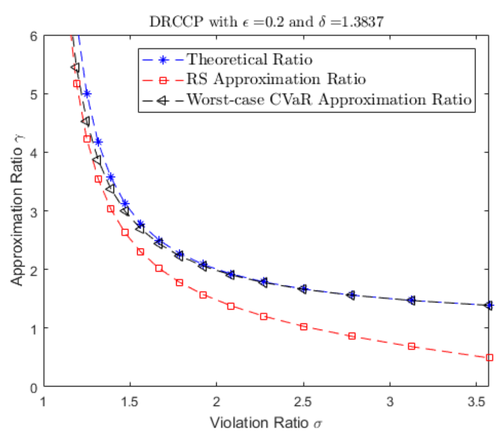
This research has been supported in part by the National Science Foundation grant 1633196 and the Office of Naval Research grant N00014-18-1-2075. Valuable comments from the editors and three anonymous reviewers are gratefully acknowledged.



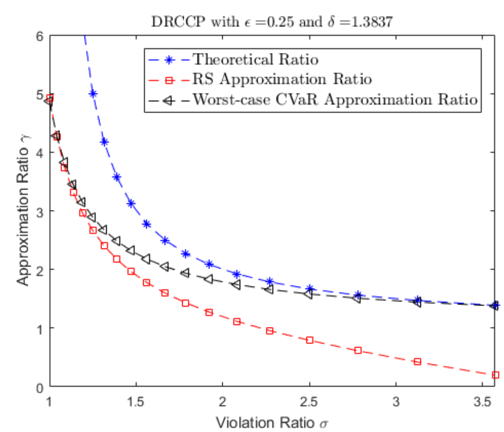
(a) $\epsilon = 0.20, \delta = 1.3811$



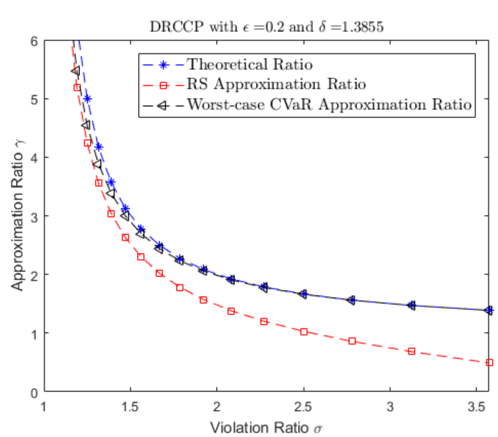
(b) $\epsilon = 0.25, \delta = 1.3811$



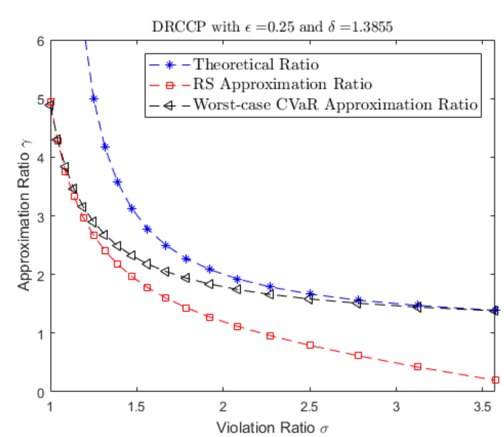
(c) $\epsilon = 0.20, \delta = 1.3837$



(d) $\epsilon = 0.25, \delta = 1.3837$



(e) $\epsilon = 0.20, \delta = 1.3855$



(f) $\epsilon = 0.25, \delta = 1.3855$

Figure 3 Illustration of approximation ratios for DRCCP with Wasserstein ambiguity set.

References

Ahmed, S. (2014). Convex relaxations of chance constrained optimization problems. *Optimization Letters*, 8(1):1–12.

Ahmed, S., Luedtke, J., Song, Y., and Xie, W. (2017). Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs. *Mathematical Programming*, 162(1-2):51–81.

Alon, N., Arora, S., Manokaran, R., Moshkovitz, D., and Weinstein, O. (2011). Inapproximability of densest κ-subgraph from average case hardness. *Unpublished manuscript*.

Ben-Tal, A. and Nemirovski, A. (2001). *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. SIAM.

Beraldi, P. and Ruszczyński, A. (2002). A branch and bound method for stochastic integer problems under probabilistic constraints. *Optimization Methods and Software*, 17(3):359–382.

Bonnans, J. F. and Shapiro, A. (2000). *Perturbation analysis of optimization problems*. Springer Science & Business Media.

Buchbinder, N. and Naor, J. (2009). Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research*, 34(2):270–286.

Calafiore, G. C., Campi, M. C., et al. (2006). The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753.

Calafiore, G. C. and El Ghaoui, L. (2006). On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22.

Chen, W., Sim, M., Sun, J., and Teo, C.-P. (2010). From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations research*, 58(2):470–485.

Chen, Z., Kuhn, D., and Wiesemann, W. (2018). Data-driven chance constrained programs over wasserstein balls. *arXiv preprint arXiv:1809.00210*.

Cheon, M.-S., Ahmed, S., and Al-Khayyal, F. (2006). A branch-reduce-cut algorithm for the global optimization of probabilistically constrained linear programs. *Mathematical Programming*, 108:617–634.

Dentcheva, D., Prékopa, A., and Ruszczynski, A. (2000). Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming*, 89:55–77.

Durrett, R. (1996). *Probability: theory and examples*. Cambridge university press.

El Ghaoui, L., Oks, M., and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556.

Esfahani, P. M. and Kuhn, D. (2015). Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. Available at Optimization Online.

Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*.

Goyal, V. and Ravi, R. (2008). Approximation algorithms for robust covering problems with chance constraints. Available at http://repository.cmu.edu/cgi/viewcontent.cgi?article=1365&context=tepper.

Goyal, V. and Ravi, R. (2010). A PTAS for the chance-constrained knapsack problem with random item sizes. *Operations Research Letters*, 38(3):161–164.

Grötschel, M., Lovász, L., and Schrijver, A. (1981). The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197.

Gupta, A. and Nagarajan, V. (2014). Approximating sparse covering integer programs online. *Mathematics of Operations Research*, 39(4):998–1011.

Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. (2015). A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151:35–62.

Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. (2017). Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767.

Hu, Z. and Hong, J. L. (2012). Kullback-leibler divergence constrained distributionally robust optimization. Available at Optimization Online.

Jiang, R. and Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming*, 158:291–327.

Kataoka, S. (1963). A stochastic programming model. *Econometrica: Journal of the Econometric Society*, pages 181–196.

Luedtke, J., Ahmed, S., and Nemhauser, G. L. (2010). An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122:247–272.

Nemirovski, A. and Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996.

Pagnoncelli, B. K., Ahmed, S., and Shapiro, A. (2009). Sample average approximation method for chance constrained programming: theory and applications. *Journal of optimization theory and applications*, 142(2):399–416.

Qiu, F., Ahmed, S., Dey, S. S., and Wolsey, L. A. (2014). Covering linear programming with violations. *INFORMS Journal on Computing*, 26(3):531–546.

Shapiro, A. (2017). Distributionally robust stochastic programming. *SIAM Journal on Optimization*, 27(4):2258–2275.

Swamy, C. (2011). Risk-averse stochastic optimization: probabilistically-constrained models and algorithms for black-box distributions. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1627–1646. SIAM.

Talluri, S., Narasimhan, R., and Nair, A. (2006). Vendor performance with supply risk: A chance-constrained dea approach. *International Journal of Production Economics*, 100(2):212–222.

Vazirani, V. V. (2001). *Approximation algorithms*. Springer Science & Business Media.

Xie, W. (2018). On distributionally robust chance constrained program with wasserstein distance. *arXiv preprint arXiv:1806.07418*.

Xie, W. and Ahmed, S. (2018a). Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation. *IEEE Transactions on Power Systems*, 33(2):1860–1867.

Xie, W. and Ahmed, S. (2018b). On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2):1151–1182.

Xie, W., Ahmed, S., and Jiang, R. (2017). Optimized bonferroni approximations of distributionally robust joint chance constraints. Available at Optimization Online.

Yang, W. and Xu, H. (2016). Distributionally robust chance constraints for non-linear uncertainties. *Mathematical Programming*, 155:231–265.

Zhao, C. and Guan, Y. (2018). Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters*, 46(2):262–267.

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

EC.1. Proofs

EC.1.1. Proof of Theorem 1

Theorem 1 *Given a parameter $\kappa > 1$, consider the following instance of CCP:*

$$(P_\kappa): \quad v_\kappa^* = \min_{x\geq 0}\left\{x : \mathbb{P}\{\tilde{\xi} : \tilde{\xi}x < 1\} \leq \epsilon\right\},$$

where $\tilde{\xi}$ has a finite distribution with N equiprobable realizations $\{\xi^i\}_{i\in[N]}$ such that ϵN is an integer, $\xi^i = 1$ for $i \in [\epsilon N]$ and $\xi^i = \kappa$ for $i \in [N] \setminus [\epsilon N]$. Let $v_\kappa^{\mathbf{CVaR}}$ denote the optimal value of the **CVaR** *approximation (6) of (P_κ). Then*

$$\lim_{\kappa\to\infty} \frac{v_\kappa^{\mathbf{CVaR}}}{v_\kappa^*} = \infty.$$

Proof: Here the **CVaR** approximation (6) of (P_κ) is

$$v_\kappa^{\mathbf{CVaR}} = \min_{x\geq 0}\left\{x : \inf_{\beta}\left[-\epsilon\beta + \epsilon\left(1 - x + \beta\right)_+ + \left(1-\epsilon\right)\left(1 - \kappa x + \beta\right)_+\right] \leq 0\right\},$$

where the optimal β in the constraint is nonnegative and upper bounded by $1 + \frac{\epsilon}{1-\epsilon}(x-1)_+ + (\kappa x - 1)_+$. Thus, we can replace the infimum by existence of β, i.e.,

$$v_\kappa^{\mathbf{CVaR}} = \min_{x\geq 0}\left\{x : -\epsilon\beta + \epsilon\left(1 - x + \beta\right)_+ + \left(1-\epsilon\right)\left(1 - \kappa x + \beta\right)_+ \leq 0, \beta \geq 0\right\}. \tag{EC.1}$$

We note that in (EC.1), $\beta = 0, x = 1$ is a feasible solution, thus $v_\kappa^{\mathbf{CVaR}} \leq 1$. On the other hand, since $1 - x + \beta \leq \left(1 - x + \beta\right)_+$ and $\left(1-\epsilon\right)\left(1 - \kappa x + \beta\right)_+ \geq 0$, hence,

$$-\epsilon\beta + \epsilon\left(1 - x + \beta\right) \leq -\epsilon\beta + \epsilon\left(1 - x + \beta\right)_+ + \left(1-\epsilon\right)\left(1 - \kappa x + \beta\right)_+ \leq 0$$

which implies that $x \geq 1$ in (EC.1), i.e., $v_\kappa^{\mathbf{CVaR}} = 1$.

On the other hand, by introducing binary variable z_i for each $i \in [N]$, the problem (P_κ) can be reformulated as

$$v_\kappa^* = \min_{x,\boldsymbol{z}}\left\{x : x \geq z_i, i \in [\epsilon N], \kappa x \geq z_i, i \in [N]\setminus[\epsilon N],\ \boldsymbol{z} \in Z \cap \mathbb{B}^N\right\},$$

where $\mathbb{B} = \{0,1\}$ and $Z := \left\{\boldsymbol{z} \in [0,1]^N : \sum_{i\in[N]} z_i \geq N - \epsilon N\right\}$. Clearly, the optimal value of the problem (P_κ) is $v_\kappa^* = \frac{1}{\kappa}$. Thus, $v_\kappa^{\mathbf{CVaR}}/v_\kappa^* = \kappa \to \infty$ when $\kappa \to \infty$. □

EC.1.2. Proof of Theorem 2

Theorem 2 *Consider the following instance of CCP:*

$$(P): \quad v^* = \min_{x\geq 0}\left\{x : \mathbb{P}\{\tilde{\xi} : \tilde{\xi}x < 1\} \leq \epsilon\right\},$$

where $\tilde{\xi}$ is a Bernoulli random variable with probability $(1-\epsilon)$. Let v_δ^{SA} denote the optimal value of the scenario approximation problem (7) corresponding to (P) for a given confidence parameter $\delta \in (0,1)$ and sample size $\bar{N}_\delta = \left\lceil \frac{2}{\epsilon}\log\left(\frac{1}{\delta}\right) + \frac{2}{\epsilon}\log\left(\frac{2}{\epsilon}\right) + 2 \right\rceil$. Then

$$\mathbb{P}\left\{\frac{v_\delta^{SA}}{v^*} = \infty\right\} \geq 1 - \delta^2\epsilon^2.$$

Proof: Clearly, in this example, the optimal value of (P) is $v^* = 1$. On the other hand, the SA problem (7) is equivalent to

$$v_\delta^{SA} = \min_x \left\{x : \xi^i x \geq 1, \forall i \in [\bar{N}_\delta]\right\}.$$

Note that for each $i \in [\bar{N}_\delta]$, $\xi^i \in \{0,1\}$ and the above minimization problem is infeasible if one of $\{\xi^i\}_{i\in[\bar{N}_\delta]}$ is equal to 0. The probability that none of $\{\xi^i\}_{i\in[\bar{N}_\delta]}$ is equal to 0 is upper bounded by

$$(1-\epsilon)^{\bar{N}_\delta} \leq (1-\epsilon)^{\frac{2}{\epsilon}\log\left(\frac{1}{\delta}\right)+\frac{2}{\epsilon}\log\left(\frac{2}{\epsilon}\right)+2} \leq \delta^2\epsilon^2.$$

Thus, with probability at least $1-\delta^2\epsilon^2$, we can get $v_\delta^{SA} = \infty$; i.e. $v_\delta^{SA}/v^* = \infty$. □

EC.1.3. Proof of Theorem 3

Theorem 3 *Suppose we have a polynomial time algorithm that returns a (σ,γ)-approximate solution to any CCP of the form* (4) *with a discrete distribution with N equiprobable realizations. Then, unless P=NP, the following holds:*

(i) if $\gamma = 1$, then we must have $\sigma = 1/\epsilon - f(N)(1-\epsilon)/\epsilon$ for some function f such that $f(N) \to 0$ as $N \to \infty$;

(ii) if $\sigma = 1$, then we must have $\gamma = g(N)$ for some function g such that $g(N) \to \infty$ as $N \to \infty$.

Proof:

(i) Consider the NP-complete problem κ- dense graph which asks

(κ nodes- dense graph): Given a graph $G(V,E)$ with nodes V with $|V| = n$ and edges E with $|E| = N$, does it contain a dense subgraph with κ nodes with number of edges at least $N(1-\epsilon)$?

This problem can be formulated as CCP (4) to minimize the number of selected nodes. In particular, we let $[n] = V, [N] = E$, $adj(i)$ denote the pair of nodes u,v related to an edge $i \in E$ such that $i = (u,v) \in E$ and binary variables $x_j = 1, z_i = 1$ denote node j and edge i are chosen respectively, and 0, otherwise. With the notation above, the problem of κ nodes-dense graph can be formulated as below:

$$v^* = \min_{\boldsymbol{x},\boldsymbol{z}} \quad \sum_{j\in[n]} x_j \tag{EC.2a}$$

$$\text{s.t. } x_j \geq z_i, \forall i \in [N], \forall j \in adj(i), \tag{EC.2b}$$

$$\sum_{i\in[N]} z_i \geq N - \epsilon N, \tag{EC.2c}$$

$$\boldsymbol{x} \in \mathbb{R}_+^n, z_i \in \{0,1\}, \forall i \in [N], \tag{EC.2d}$$

where constraints (EC.2b) enforce that z_i will equal to 1 only if both nodes of edge i are chosen, 0, otherwise. Note that (EC.2) is a special case of CCP (4) by letting $S = \mathbb{R}_+^n$ and $\mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) < 1\right\} = \frac{1}{N}\sum_{i\in[N]} \mathbb{I}(\min_{j\in adj(i)}\{x_j\} < 1)$, where $\mathbb{I}(\cdot)$ is an indicator function.

Now suppose we obtain an approximate solution, i.e. a subgraph with number of nodes $v^* = \kappa$ and number of edges $N - \sigma\epsilon N$. By Theorem 1.2 in Alon et al. (2011), unless P=NP, there cannot be a polynomial time approximation algorithm for κ nodes- dense graph with a constant factor, i.e. it must hold that

$$\frac{N - \sigma\epsilon N}{N - \epsilon N} = f(N),$$

for some function $f(\cdot)$ such that $\lim_{N\to\infty} f(N) = 0$. Thus, $\sigma = 1/\epsilon - f(N)(1-\epsilon)/\epsilon$ and $f(N) \to \infty$ as $N \to \infty$.

(ii) Now consider another variant of κ- dense graph which asks

($N(1-\epsilon)$ edges- dense graph) Given a graph with nodes V with $|V| = n$ and edges E with $|E| = N$, does it contain a dense subgraph with number of edges at least $N(1-\epsilon)$ and number of nodes at most κ?

This problem can be also formulated as (EC.2). Now suppose we get an approximate solution, i.e. a subgraph with number of nodes $v^* \leq \gamma\kappa$ and number of edges at least $N - \epsilon N$. We would like to show that γ cannot be a constant. We will prove it by contradiction, i.e., from now on, we assume that $\gamma \geq 2$ is a positive integer constant.

Next, we prove the following claim.

Claim: If there exists a γ ($\gamma \geq 2$ is a positive integer constant) approximation algorithm of $N(1-\epsilon)$ edges- dense graph, then there exists an $\frac{1}{2.5\gamma^2}$ approximation algorithm of κ nodes- dense graph.

Proof: Since by assumption, we know that there exists a polynomial-time algorithm that finds a subgraph G' of $G(V,E)$ with $\gamma\kappa$ nodes and $N(1-\epsilon)$ edges, thus it is sufficient to prove that we can find a subgraph G'' of G' with κ nodes and at least $\nu \in \mathbb{Z}_+$ edges in a polynomial time, where $\nu \geq N(1-\epsilon)/(2.5\gamma^2)$. We prove this statement by construction.

First, we partition the graph G' into arbitrary γ subgraphs $\{\mathcal{G}_i\}_{i\in[\gamma]}$, where each group has κ nodes. Next, we discuss the number of edges within each subgraph and across different subgraphs.

Case 1. If there exists a subgraph with at least $\nu \geq N(1-\epsilon)/(2.5\gamma^2)$ edges, then we are done.

Case 2. Suppose none of the subgraphs have no less than $N(1-\epsilon)/(2.5\gamma^2)$ edges. Note that there are $\binom{\gamma}{2} = \frac{\gamma(\gamma-1)}{2}$ pairs of subgraphs, therefore, by the pigeonhole principle, there must exist a pair of subgraphs such that the number of edges between these two subgraphs are at least

$$\frac{N(1-\epsilon)-\gamma N(1-\epsilon)/(2.5\gamma^2)}{\frac{\gamma(\gamma-1)}{2}} = 2N(1-\epsilon)\left(\frac{1-\frac{1}{2.5\gamma}}{\gamma(\gamma-1)}\right) \geq \frac{2N(1-\epsilon)}{\gamma^2}.$$

Suppose these two subgraphs are $\mathcal{G}_i, \mathcal{G}_j$ with $i,j \in [\gamma]$ but $i \neq j$.

Next, we are going to find a subgraph of $\mathcal{G}_i \cup \mathcal{G}_j$ with κ nodes and at least $\nu \geq N(1-\epsilon)/(2.5\gamma^2)$. To do so, we need to discuss whether κ is an even integer or not.

(i) If κ is an even integer, then we divide $\mathcal{G}_i$ and $\mathcal{G}_j$ into two separate subgraphs with $\kappa/2$ nodes, i.e., $\mathcal{G}_{i1}, \mathcal{G}_{i2}$ and $\mathcal{G}_{j1}, \mathcal{G}_{j2}$ in Figure EC.1, respectively. Thus, the number of edges between $\mathcal{G}_i$ and $\mathcal{G}_j$ is equal to the total number of the edges of the following four unions of two subgraphs, $\mathcal{G}_{i1} \cup \mathcal{G}_{j1}$, $\mathcal{G}_{i1} \cup \mathcal{G}_{j2}$, $\mathcal{G}_{i2} \cup \mathcal{G}_{j1}$,$\mathcal{G}_{i2} \cup \mathcal{G}_{j2}$, where the number of nodes of each union is equal to κ. Again, by the pigeonhole principle, there must exist a union of two subgraph with number of edges at least

$$\frac{2N(1-\epsilon)}{4\gamma^2} \geq \frac{N(1-\epsilon)}{2.5\gamma^2}.$$

(ii) If κ is an odd integer, then we divide $\mathcal{G}_i$ and $\mathcal{G}_j$ into three separate subgraphs, i.e., $\mathcal{G}_{i1}, s_i, \mathcal{G}_{i2}$ and $\mathcal{G}_{j1}, s_j, \mathcal{G}_{j2}$ in Figure EC.2, respectively, where there are $\lfloor \kappa/2 \rfloor$ nodes in $\mathcal{G}_{i1}, \mathcal{G}_{i2}, \mathcal{G}_{j1}, \mathcal{G}_{j2}$ and only 1 nodes in s_i, s_j. Let t_i be a subgraph with a single node from $\mathcal{G}_{i2}$. Thus, the number of edges between $\mathcal{G}_i$ and $\mathcal{G}_j$ is bounded by the total number of the edges of the following five unions of subgraphs, $\mathcal{G}_{i1} \cup \mathcal{G}_{j1} \cup s_j$, $\mathcal{G}_{i1} \cup \mathcal{G}_{j2} \cup s_i$, $\mathcal{G}_{i2} \cup \mathcal{G}_{j1} \cup s_i$, $\mathcal{G}_{i2} \cup \mathcal{G}_{j2} \cup s_j$ and $(\mathcal{G}_{i2} \setminus t_i) \cup \mathcal{G}_{j1} \cup s_i \cup s_j$, where the number of nodes of each union is equal to κ. Again, by the pigeonhole principle, there must exist a union of graph with number of edges at least

$$\frac{2N(1-\epsilon)}{5\gamma^2} = \frac{N(1-\epsilon)}{2.5\gamma^2}.$$

Note that the time complexity of the above procedure is $O(\gamma^2 N(1-\epsilon))$. Hence, we can find a subgraph G'' with κ nodes and at least $N(1-\epsilon)/(2.5\gamma^2)$ edges in polynomial time. Thus, there is a $\frac{1}{2.5\gamma^2}$ approximation algorithm of κ nodes- dense graph. $\diamond$

By Theorem 1.2 in Alon et al. (2011), there is no polynomial approximation algorithm for κ nodes- dense graph with constant factor. Therefore, by Claim, there is no polynomial time approximation algorithm of $N(1-\epsilon)$ edges- dense graph either, a contradiction that γ is constant. Thus, we must have $\gamma = g(N)$ with some function $g(\cdot)$ such that $\lim_{N\to\infty} g(N) = \infty$. $\square$

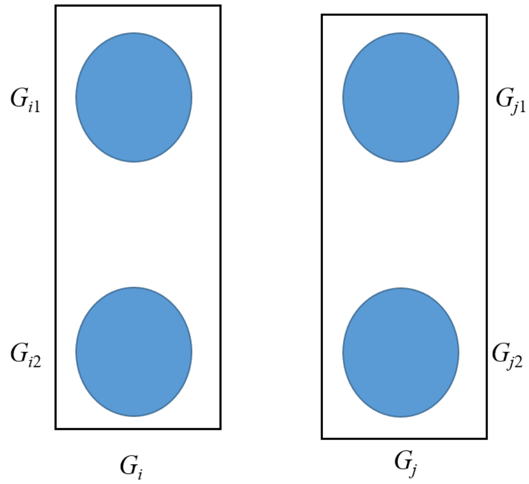


Figure EC.1 Illustration of Proof of Theorem 3: (a) κ is odd.

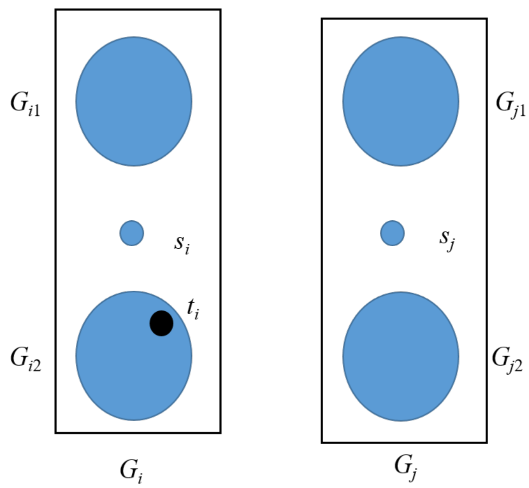


Figure EC.2 Illustration of Proof of Theorem 3: (b) κ is even

EC.1.4. Proof of Theorem 6

Theorem 6 *Given a generating function $\phi \in \Phi$, choose*

- *violation ratio: $\sigma \in (\frac{1}{\epsilon} - (\frac{1}{\epsilon} - 1)\phi(1), \frac{1}{\epsilon})$,*
- *approximation ratio: $\gamma = \left[\phi^{-1}\left(1 - \frac{1}{\sigma\epsilon} + (\frac{1}{\sigma\epsilon} - \frac{1}{\sigma})\phi(1)\right)\right]^{-1}$, and*
- *relaxation: $\widehat{X}$ = the feasible region of* (15).

Then there exists a $\widehat{\sigma} \leq \sigma$, such that Relax-and-Scale $(\widehat{\sigma}, \gamma, \widehat{X})$ returns a $(\widehat{\sigma}, \gamma)$-bicriteria approximate solution to CCP of the form (13).

Proof: Let $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{w}})$ be an optimal solution of (15) (note that the existence of an optimal solution to (15) directly follows from Theorem 5, thus is omitted here). From (14b), we have $\widehat{w}_l \in [0, \epsilon]$ for each $l \in [m]$. Given a proposed violation ratio $\sigma \in (\frac{1}{\epsilon} - (\frac{1}{\epsilon} - 1)\phi(1), \frac{1}{\epsilon})$, for simplicity, let us define

$r = \frac{1}{\sigma\epsilon}[1-(1-\epsilon)\phi(1)] \in (0,1)$, then approximation ratio can be rewritten as $\gamma = [\phi^{-1}(1-r)]^{-1}$. Also let us denote a set $\Xi_r^l := \{\boldsymbol{\xi}_l \in \Xi_l : \phi(a_l(\boldsymbol{x}, \boldsymbol{\xi}_l)) \geq 1-r\}$ for each $l \in [m]$.

Now let us define the new violation ratio $\widehat{\sigma} = \epsilon^{-1}\left[1 - \prod_{l\in[m]} \frac{(1-\widehat{w}_l)\phi(1)+r-1}{r}\right]$. We would like to show that (1) Relax-and-Scale $(\widehat{\sigma}, \gamma, \widehat{X})$ returns a $(\widehat{\sigma}, \gamma)$-bicriteria approximate solution to CCP of the form (13); and (2) $\widehat{\sigma} \leq \sigma$. Therefore, correspondingly, we will separate the proof into two steps.

(1) First, since $\widehat{\boldsymbol{x}}$ satisfies $\mathbb{E}[\phi(a_l(\widehat{\boldsymbol{x}}, \tilde{\boldsymbol{\xi}}_l))] \geq (1-\widehat{w}_l)\phi(1)$, thus from Lemma 1, we have

$$\mathbb{P}\{\tilde{\boldsymbol{\xi}}_l : \tilde{\boldsymbol{\xi}}_l \in \Xi_r^l\} \geq \frac{(1-\widehat{w}_l)\phi(1)+r-1}{r}. \tag{EC.3}$$

Now let $\tilde{\boldsymbol{x}} = \gamma\widehat{\boldsymbol{x}} := [\phi^{-1}(1-r)]^{-1}\widehat{\boldsymbol{x}}$. We need to show that $\tilde{\boldsymbol{x}} \in X_{\widehat{\sigma}\epsilon}$. Clearly, for each $\boldsymbol{\xi}_l \in \Xi_r^l$,

$$a_l(\tilde{\boldsymbol{x}}, \boldsymbol{\xi}_l) = \gamma a_l(\widehat{\boldsymbol{x}}, \boldsymbol{\xi}_l) \geq 1,$$

where the inequality follows from the definition of Ξ_r^l and positive homogeneity and monotonicity of function $a_l(\cdot,\cdot)$. Hence,

$$\begin{aligned}
\mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : a(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\xi}}) < 1\right\} &= 1 - \prod_{l\in[m]} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}}_l : a(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\xi}}_l) \geq 1\right\} \\
&\leq 1 - \prod_{l\in[m]} \mathbb{P}\{\tilde{\boldsymbol{\xi}}_l : \tilde{\boldsymbol{\xi}}_l \in \Xi_r^l\} \\
&\leq 1 - \prod_{l\in[m]} \frac{(1-\widehat{w}_l)\phi(1)+r-1}{r} := \widehat{\sigma}\epsilon,
\end{aligned}$$

where the first inequality is due to $\mathbb{P}\left\{\tilde{\boldsymbol{\xi}}_l : a(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\xi}}_l) \geq 1\right\} \geq \mathbb{P}\{\tilde{\boldsymbol{\xi}}_l : \tilde{\boldsymbol{\xi}}_l \in \Xi_r^l\}$ and the second inequality is because of (EC.3). Thus, $\tilde{\boldsymbol{x}} \in X_{\widehat{\sigma}\epsilon}$.

(2) It remains to show that $\widehat{\sigma}$ is no larger than $\sigma = (r\epsilon)^{-1}[1-(1-\epsilon)\phi(1)]$, i.e., the following claim.

Claim: $\widehat{\sigma}\epsilon \leq r^{-1}[1-(1-\epsilon)\phi(1)]$.

Proof: By the definition of $\widehat{\sigma}$, it is sufficient to show that the optimal value of

$$\min_{\boldsymbol{w}} \left\{ \prod_{l\in[m]} \left(\frac{(1-w_l)\phi(1)+r-1}{r}\right) : \prod_{l\in[m]} (1-w_l) \geq 1-\epsilon, \boldsymbol{w} \in [0,\epsilon]^m \right\} \tag{EC.4}$$

is no smaller than $1 - r^{-1}[1-(1-\epsilon)\phi(1)]$. First of all, we take the log of the objective function and constraint and let $y_l = \log(1-w_l)$ for each l, i.e. $w_l = 1-e^{y_l}$. Thus, it is equivalent to prove that the following optimization problem

$$\min_{\boldsymbol{y}} \left\{ \sum_{l\in[m]} \log\left(\frac{e^{y_l}}{r}\phi(1) - \frac{1}{r} + 1\right) : \sum_{l\in[m]} y_i \geq \log(1-\epsilon), \boldsymbol{y} \in [\log(1-\epsilon), 0]^m \right\} \tag{EC.5}$$

has objective value no smaller than $\log((1-\epsilon)\phi(1)/r + 1 - 1/r)$. Note that (EC.5) is a concave minimization over polyhedral set, hence the optimal value is achieved by one of its extreme points. Given an extreme point $\widehat{\boldsymbol{y}}$, there are two cases:

(i) if $\sum_{l\in[m]}\widehat{y}_l > \log(1-\epsilon)$, then we must have $\widehat{y}_l = 0$ for each $l\in[m]$, which yields that the objective of (EC.5) is equal to $\log(\phi(1)/r+1-1/r)$, and is thus greater than $\log((1-\epsilon)\phi(1)/r+1-1/r)$;

(ii) if $\sum_{l\in[m]}\widehat{y}_l = \log(1-\epsilon)$, then there exists an l_0 such that $\widehat{y}_l = \log(1-\epsilon)$, if $l = l_0$ and 0, otherwise. In this case, the objective of (EC.5) is equal to $\log((1-\epsilon)\phi(1)/r+1-1/r)$.

In both cases, $\widehat{\sigma}$ is no larger than $(r\epsilon)^{-1}[1-(1-\epsilon)\phi(1)]$. ◇

This completes the proof of Theorem 6. □

EC.1.5. Proof of Theorem 9

Theorem 9 *DRCCP* (5) *with* $\mathcal{P}=\mathcal{P}^M$ *is NP-hard.*

Proof:

Let us first consider the NP-complete problem - set covering which asks

(Set covering problem) Given a set of $[m]$, a family of subsets $\mathcal{S}=\{e_1,\ldots,e_n\}$ with $|\mathcal{S}|=n$, $e_j\subseteq[m]$ for each $j\in[m]$ and a positive integer $K\in\{1,\ldots,n\}$, does there exist a collection of subsets with size at most K in $\mathcal{S}$ whose union covers set $[m]$, i.e., $[m]\subseteq\cup_{j\in[K]}e_j$?

Let us define $a_{lj}=1$ if set e_j contains elements l, 0, otherwise, decision variable $x_j=1$ if set e_j is chosen, 0, otherwise. Then decision version of set covering problem can be formulated as

$$v^{sc} = \min_{\boldsymbol{x}} \quad 0, \tag{EC.6a}$$

$$\text{s.t.} \quad \sum_{j\in[n]} x_j \le K, \tag{EC.6b}$$

$$\sum_{j\in[n]} a_{lj}x_j \ge 1, \forall l\in[m], \tag{EC.6c}$$

$$\boldsymbol{x}\in\{0,1\}^n, \tag{EC.6d}$$

where there exists a covering with size at most K if and only if $v^{sc}=0$.

Consider the following DRCCP with moment ambiguity set $\mathcal{P}=\mathcal{P}^M$:

$$v^{sc}_{\epsilon,\mathcal{P}^M} = \min_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n_+} \sum_{j\in[n]}(x_j+y_j), \tag{EC.7a}$$

$$\text{s.t.} \quad \frac{1}{n-K}\sum_{j\in[n]} y_j \ge 1, \tag{EC.7b}$$

$$x_j+y_j\ge 1, \forall j\in[n], \tag{EC.7c}$$

$$\inf_{\mathbb{P}\in\mathcal{P}^M}\mathbb{P}\left\{\tilde{\boldsymbol{\xi}}:\tilde{\xi}_j x_j + y_j \ge 1, \tilde{\xi}_{n+j}y_j + x_j \ge 1, \forall j\in[n]\right\} \ge 1-\epsilon, \tag{EC.7d}$$

(EC.6c),

where $\epsilon \in (0, 0.5], \Xi = \mathbb{R}_+^{2n}$ and the moment ambiguity set is defined as follows:

$$\mathcal{P}^M = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}_+^{2n}\right) : \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = 2\boldsymbol{e}, \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{\xi}} - 2\boldsymbol{e}|] \leq \frac{2\epsilon}{n}\boldsymbol{e} \right\}. \tag{EC.7e}$$

Now we make the following claims.

Claim 1: Given an $\mathcal{M} \subseteq [2n]$ and $\mathcal{P}^M$ defined in (EC.7e), we have

$$\inf_{\mathbb{P}\in\mathcal{P}^M} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : \tilde{\xi}_j \geq 1, \forall j \in \mathcal{M} \right\} = 1 - \frac{|\mathcal{M}|}{n}\epsilon.$$

Proof: Let

$$q^* = \inf_{\mathbb{P}\in\mathcal{P}^M} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : \tilde{\xi}_j \geq 1, \forall j \in \mathcal{M} \right\}.$$

Since the support of random variable $\tilde{\xi}_j$ is $\mathbb{R}_+$ for each $j \in [2n]$, thus equivalently, we can have

$$q^* = \inf_{\mathbb{P}\in\mathcal{P}^M} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} : \tilde{\xi}_j \geq 1, \forall j \in \mathcal{M}, \tilde{\xi}_j \geq 0, j \in [2n] \setminus \mathcal{M} \right\}.$$

We observe that the uncertain constraints in the above worst-case probability are separable in random variables $\{\tilde{\xi}_j\}_{j\in[2n]}$, and the moment ambiguity set $\mathcal{P}^M$ in (EC.7e) is also separable, i.e., $\mathcal{P}^M = \prod_{j\in[2n]} \mathcal{D}_j$ with

$$\mathcal{D}_j = \left\{ \mathbb{P}_j \in \mathcal{P}_0\left(\mathbb{R}_+\right) : \mathbb{E}_{\mathbb{P}_j}[\tilde{\xi}_j] = 2, \mathbb{E}_{\mathbb{P}}[|\tilde{\xi}_j - 2|] \leq \frac{2\epsilon}{n} \right\}$$

for each $j \in [2n]$. Therefore, by the proof of Theorem 3 in Xie et al. (2017), q^* can be equivalently computed as

$$q^* = \left(\sum_{j\in\mathcal{M}} \inf_{\mathbb{P}_j\in\mathcal{D}_j} \mathbb{P}_j\left\{ \tilde{\xi}_j : \tilde{\xi}_j \geq 1 \right\} + \sum_{j\in[2n]\setminus\mathcal{M}} \inf_{\mathbb{P}_j\in\mathcal{D}_j} \mathbb{P}_j\left\{ \tilde{\xi}_j : \tilde{\xi}_j \geq 0 \right\} - 2n + 1 \right)_+.$$

Above, we notice that $\mathbb{P}_j\left\{ \tilde{\xi}_j : \tilde{\xi}_j \geq 0 \right\} = 1$ for all $\mathbb{P}_j \in \mathcal{D}_j$ and $j \in [2n] \setminus \mathcal{M}$, thus $\sum_{j\in[2n]\setminus\mathcal{M}} \inf_{\mathbb{P}_j\in\mathcal{D}_j} \mathbb{P}_j\left\{ \tilde{\xi}_j : \tilde{\xi}_j \geq 0 \right\} = 2n - |\mathcal{M}|$. Thus, we have

$$q^* = \left(\sum_{j\in\mathcal{M}} \inf_{\mathbb{P}_j\in\mathcal{D}_j} \mathbb{P}_j\left\{ \tilde{\xi}_j : \tilde{\xi}_j \geq 1 \right\} - |\mathcal{M}| + 1 \right)_+. \tag{EC.8}$$

Let us define

$$q_j^* = \inf_{\mathbb{P}_j\in\mathcal{D}_j} \mathbb{P}_j\left\{ \tilde{\boldsymbol{\xi}}_j : \tilde{\xi}_j \geq 1 \right\}$$

for each $j \in \mathcal{M}$. It remains to show that $q_j^* = 1 - \frac{\epsilon}{n}$. By the proof of Theorem 1 in Xie and Ahmed (2018b), the above worst-case probability is equivalent to

$$q_j^* = \max_{\lambda,\beta,\eta} \qquad \lambda - 2\eta - \frac{2\epsilon}{n}\beta,$$

$$\begin{aligned} \text{s.t.} \quad & \lambda - \eta\xi_j - \beta|\xi_j - 2| \le \mathbb{I}(\xi_j \ge 1), \forall \xi_j \in \mathbb{R}_+, \\ & \beta \ge 0. \end{aligned}$$

Breaking down the indicator functions, we have

$$\begin{aligned} q_j^* = \max_{\lambda,\beta,\eta} \quad & \lambda - 2\eta - \frac{2\epsilon}{n}\beta, \\ \text{s.t.} \quad & \lambda - \eta\xi_j - \beta|\xi_j - 2| \le 1, \forall \xi_j \in \mathbb{R}_+, \\ & \lambda - \eta\xi_j - \beta|\xi_j - 2| \le 0, \forall \xi_j \in \mathbb{R}_+, \xi_j < 1, \\ & \beta \ge 0. \end{aligned} \tag{EC.9}$$

By optimizing over ξ_j, (EC.9) is reduced to

$$\begin{aligned} q_j^* = \max_{\lambda,\beta,\eta} \quad & \lambda - 2\eta - \frac{2\epsilon}{n}\beta, \\ \text{s.t.} \quad & \lambda + \max(-2\eta, -2\beta) \le 1, \\ & \lambda + \max(-2\beta, -\beta - \eta) \le 0, \\ & \beta \ge 0, \beta + \eta \ge 0. \end{aligned}$$

By breaking down the max operator and introducing a new variable t, adding a constraint $t \le \lambda - 2\eta - \frac{2\epsilon}{n}\beta$ into the above optimization problem, we can further reformulate it as

$$\begin{aligned} q_j^* = \max_{\lambda,\beta,\eta,t} \quad & t, \\ \text{s.t.} \quad & t \le \lambda - 2\eta - \frac{2\epsilon}{n}\beta, \\ & \lambda - 2\eta \le 1, \\ & \lambda - 2\beta \le 0, \\ & \lambda - \beta - \eta \le 0, \\ & \beta \ge 0, \beta + \eta \ge 0. \end{aligned}$$

Using Fourier-Motzkin elimination method to project out variables λ, η, we can see that

$$\begin{aligned} q_j^* = \max_{\beta,t} \quad & t, \\ \text{s.t.} \quad & t \le 1 - \frac{2\epsilon}{n}\beta, \\ & t \le 2\beta - \frac{2\epsilon}{n}\beta, \\ & \beta \ge 0, \end{aligned}$$

where the optimal value $q_j^* = t^* = 1 - \frac{\epsilon}{n}$ with $\beta^* = \frac{1}{2}$. Therefore, from (EC.8), we have

$$q^* = \left(\sum_{j\in\mathcal{M}} q_j^* - |\mathcal{M}| + 1\right)_+ = 1 - \frac{|\mathcal{M}|\epsilon}{n}.$$

$\diamond$

Now we are ready to show our key result.

Claim 2: $v^{sc} = 0$ if and only if $v^{sc}_{\epsilon,\mathcal{P}^M} = n$.

Proof:

$\Rightarrow$) From (EC.7), by aggregating inequalities (EC.7c), it is easy to see that $v^{sc}_{\epsilon,\mathcal{P}^M} \geq n$ if feasible. Now suppose that $v^{sc} = 0$, then let $\boldsymbol{x}^*$ be its optimal solution. Now let $\boldsymbol{y}^* = \boldsymbol{e} - \boldsymbol{x}^*$. Clearly, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ satisfies (EC.6c),(EC.7c) and (EC.7b). And (EC.7d) is reduced to

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}_j \geq 1, \forall j \in \mathcal{M}\right\} \geq 1 - \epsilon$$

where $\mathcal{M} = \{i : x_i^* = 1\} \cup \{n + i : x_i^* = 0\}$. Since $|\mathcal{M}| = n$, thus by Claim 1, we have $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}_j \geq 1, \forall j \in \mathcal{M}\right\} = 1 - \epsilon$.

Thus, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is feasible to (EC.7) with objective value equal to n. Therefore, $v^{sc}_{\epsilon,\mathcal{P}^M} = n$.

$\Leftarrow$) if $v^{sc}_{\epsilon,\mathcal{P}^M} = n$, then let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be its optimal solution. Since $\sum_{j\in[n]}(x_j^* + y_j^*) = n$ and $y_j^* + x_j^* \geq 1$ for each $j \in [n]$. Thus, $\boldsymbol{y}^* = \boldsymbol{e} - \boldsymbol{x}^*$. Clearly, $\boldsymbol{x}^*$ satisfies (EC.6b), (EC.6c).

Now we see that $\boldsymbol{x}^* \in \{0,1\}^n$. Note that (EC.7d) is reduced to

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}_j \geq 1, \forall j \in \mathcal{M}\right\} \geq 1 - \epsilon$$

where $\mathcal{M} = \{j : x_j^* > 0\} \cup \{n + j : x_j^* < 1\}$. Suppose that there exists j_0 that $\boldsymbol{x}_{j_0}^* \in (0,1)$, then $|\mathcal{M}| \geq n + 1$, by Claim 1, we have

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}}_j \geq 1, \forall j \in \mathcal{M}\right\} = 1 - \frac{|\mathcal{M}|}{n}\epsilon \leq 1 - \frac{n+1}{n}\epsilon < 1 - \epsilon,$$

contradiction the feasibility of $(\boldsymbol{x}^*, \boldsymbol{y}^*)$.

Thus, $\boldsymbol{x}^*$ is feasible to (EC.6), thus $v^{sc} = 0$.

$\diamond$

Hence, the set covering problem can be reduced as a special case of DRCCP. This completes the proof. $\square$

EC.1.6. Proof of Theorem 11

Theorem 11 *For any $\alpha \in (0,1)$, the DRCCP set X_α in* (17) *with $\mathcal{P} = \mathcal{P}^W$ is equivalent to*

$$X_\alpha = \left\{ \boldsymbol{x} \in S : \begin{array}{l} \lambda\delta + \sum_{i\in[N]} p_i \eta_i \leq \alpha, \\ \eta_i + \inf_{\xi\in\Xi} [\lambda \|\boldsymbol{\xi} - \boldsymbol{\zeta}_i\|_q + \pi_{il} a_l(\boldsymbol{x}, \boldsymbol{\xi})] \geq 1 + \pi_{il}, \forall i \in [N], l \in I(x) \\ \lambda \geq 0, \eta_i \geq 0, \pi_{il} \geq 0, \forall i \in [N], \forall l \in [m] \end{array} \right\} \quad \begin{array}{r} (29a) \\ (29b) \\ (29c) \end{array}$$

where $I(\boldsymbol{x}) := \{l \in [m] : \exists \boldsymbol{\xi} \in \Xi, a_l(\boldsymbol{x}, \boldsymbol{\xi}) < 1\}$.

Proof: Using Theorem 1 in Gao and Kleywegt (2016) and $a(\boldsymbol{x},\boldsymbol{\xi}) = \min_{l\in[m]}\{a_l(\boldsymbol{x},\boldsymbol{\xi})\}$, the optimization problem $\sup_{\mathbb{P}\in\mathcal{P}^W} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : a(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) < 1\right\}$ is equivalent to

$$\min_{\lambda\geq 0} \lambda\delta - \sum_{i\in[N]} p_i \inf_{\xi\in\Xi} [\lambda\|\boldsymbol{\xi} - \boldsymbol{\zeta}_i\|_q - \mathbb{I}(a_l(\boldsymbol{x},\boldsymbol{\xi}) < 1, \exists l \in [m])].$$

By replacing $\min$ with existence of some $\lambda \geq 0$, the set X_α reduces to

$$X_\alpha = \left\{ \boldsymbol{x} \in S : \begin{array}{l} \lambda\delta - \sum_{i\in[N]} p_i \inf_{\xi\in\Xi} [\lambda\|\boldsymbol{\xi} - \boldsymbol{\zeta}_i\|_q - \mathbb{I}(a_l(\boldsymbol{x},\boldsymbol{\xi}) < 1, \exists l \in [m])] \leq \alpha, \\ \lambda \geq 0. \end{array} \right\} \quad \begin{array}{r} (EC.10a) \\ (EC.10b) \end{array}$$

Next introduce a new variable η_i and let $\eta_i \geq -\inf_{\xi\in\Xi} [\lambda\|\boldsymbol{\xi} - \boldsymbol{\zeta}_i\|_q - \mathbb{I}(a_l(\boldsymbol{x},\boldsymbol{\xi}) < 1, \exists l \in [m])]$ for each $i \in [N]$. Thus, the set X_α becomes

$$X_\alpha = \left\{ \boldsymbol{x} \in S : \begin{array}{l} \lambda\delta + \sum_{j\in[N]} p_i \eta_i \leq \alpha, \\ \eta_i + \inf_{\xi\in\Xi} [\lambda\|\boldsymbol{\xi} - \boldsymbol{\zeta}_i\|_q - \mathbb{I}(a_l(\boldsymbol{x},\boldsymbol{\xi}) < 1, \exists l \in [m])] \geq 0, \forall i \in [N], \\ \lambda \geq 0. \end{array} \right\} \quad \begin{array}{r} (EC.11a) \\ (EC.11b) \\ (EC.11c) \end{array}$$

Note that $\mathbb{I}(a_l(\boldsymbol{x},\boldsymbol{\xi}) < 1, \exists l \in [m]) = 1$ if there exists $l \in [m]$ such that $a_l(\boldsymbol{x},\boldsymbol{\xi}) < 1$, otherwise, 0. Thus, by breaking down the indicator function and using the fact that $\boldsymbol{\zeta}_i \in \Xi$ for each $i \in [N]$, we arrive at

$$X_\alpha = \left\{ \boldsymbol{x} \in S : \begin{array}{l} \lambda\delta + \sum_{i\in[N]} p_i \eta_i \leq \alpha, \\ \eta_i + \inf_{\xi\in\Xi, a_l(\boldsymbol{x},\boldsymbol{\xi})<1} [\lambda\|\boldsymbol{\xi} - \boldsymbol{\zeta}_i\|_q] \geq 1, \forall i \in [N], l \in [m] \\ \lambda \geq 0, \eta_i \geq 0, \forall i \in [N]. \end{array} \right\} \quad \begin{array}{r} (EC.12a) \\ (EC.12b) \\ (EC.12c) \end{array}$$

Finally, let $I(\boldsymbol{x}) := \{l \in [m] : \exists \boldsymbol{\xi} \in \Xi, a_l(\boldsymbol{x},\boldsymbol{\xi}) < 1\}$ and by Convex Theorem of Alternatives in Ben-Tal and Nemirovski (2001), we arrive at (29). □

EC.1.7. Proof of Theorem 12

Theorem 12 *DRCCP* (5) *with* $\mathcal{P} = \mathcal{P}^W$ *is NP-hard.*

Proof: Let us first consider the NP-hard problem - chance constrained linear covering problem in Qiu et al. (2014) (with single chance constraint) which asks

> (chance constrained linear covering problem) Given a set $[N]$ of covering constraints $\mathcal{X}^i := \{\boldsymbol{x} \in \mathbb{R}^n_+ : \boldsymbol{\zeta}_i^\top \boldsymbol{x} \geq 1\}$ with $\boldsymbol{\zeta}_i \in \mathbb{Q}^n_+$ for each $i \in [N]$, and a cost parameter $\boldsymbol{c} \in \mathbb{R}^n_+$ (suppose $\|\boldsymbol{c}\|_1 = 1$), what is the minimum cost if at least $N - \lfloor \epsilon N \rfloor$ covering constraints should be satisfied?

Under this setting, the chance constrained linear covering problem can be formulated as

$$v^{ccsc} = \min_{\boldsymbol{x},\boldsymbol{z},\alpha,\eta} \boldsymbol{c}^\top \boldsymbol{x}, \tag{EC.13a}$$

$$\text{s.t.} \sum_{i\in[N]} z_i \geq N - \lfloor \epsilon N \rfloor, \tag{EC.13b}$$

$$\boldsymbol{\zeta}_i^\top \boldsymbol{x} \geq z_i, \forall i \in [N], \tag{EC.13c}$$

$$z_i \in \{0,1\}, \forall i \in [N], \boldsymbol{x} \in \mathbb{R}^n_+, \tag{EC.13d}$$

where v^{ccsc} denotes the optimal value of above chance constrained linear covering problem. And without loss of generality, we also assume that $\|\boldsymbol{\zeta}_i\|_1 > 0$ for each $i \in [N]$ and choose ϵ such that $\epsilon N \notin \mathbb{Z}_+$.

Note that for the DRCCP of linear covering problem, the point mass function of the empirical distribution $\mathbb{P}_{\boldsymbol{\zeta}}$ is $p_i = \frac{1}{N}$ for each $i \in [N]$, and uncertain function is $a(\boldsymbol{x}, \boldsymbol{\xi}) = \boldsymbol{\xi}^\top \boldsymbol{x}$.

By relaxing $\Xi = \mathbb{R}^n$ and according to Theorem 12, the optimal value of DRCCP of linear covering problem $v^{sc}_{\epsilon,\mathcal{P}^W}(\delta)$ can be upper bounded by

$$v^{sc}_{\epsilon,\mathcal{P}^W}(\delta) \leq \bar{v}^{sc}_{\epsilon,\mathcal{P}^W}(\delta) = \min_{\boldsymbol{x},\boldsymbol{z},\alpha,\eta} \boldsymbol{c}^\top \boldsymbol{x}, \tag{EC.14a}$$

$$\text{s.t.} \sum_{i\in[N]} z_i \geq (1-\epsilon)N + \delta N \eta, \tag{EC.14b}$$

$$z_i \leq \pi_i(\boldsymbol{\zeta}_i^\top \boldsymbol{x} - 1), \forall i \in [N], \tag{EC.14c}$$

$$\pi_i \|\boldsymbol{x}\|_{\frac{q}{q-1}} \leq \eta, \forall i \in [N], \tag{EC.14d}$$

$$z_i \leq 1, \forall i \in [N], \tag{EC.14e}$$

$$\boldsymbol{x} \in \mathbb{R}^n_+, \boldsymbol{\pi} \in \mathbb{R}^N_+, \eta \geq 0. \tag{EC.14f}$$

We first observe that $\bar{v}^{sc}_{\epsilon,\mathcal{P}^W}(\delta)$ is monotonically nondecreasing over positive parameter δ since the feasible region (EC.14) grows as δ decreases. And since $\mathcal{P}^W$ contains the empirical distribution, then $\bar{v}^{sc}_{\epsilon,\mathcal{P}^W}(\delta) \geq v^{ccsc}$.

Now we make the following claim.

Claim: Suppose that $N\epsilon$ is not an integer, then for any positive number $\tau > 0$, there exists a $\delta > 0$

such that $\bar{v}^{sc}_{\epsilon,\mathcal{P}^W}(\delta) - v^{ccsc} \leq \tau$.

Proof: If chance constrained linear covering problem is infeasible, then we are done. Now suppose it is feasible and let $\boldsymbol{x}^*$ be one of its optimal solutions. First define set $I := \{i \in [N] : \boldsymbol{\zeta}_i^\top \boldsymbol{x}^* \geq 1, \forall i \in [N]\}$. Clearly, we have $|I| \geq N - \lfloor \epsilon N \rfloor$. Now let $\widehat{\boldsymbol{x}} = \boldsymbol{x}^* + \tau \boldsymbol{e}$ and $\widehat{\pi}_i = \frac{1}{\tau \|\boldsymbol{\zeta}_i\|_1}, i \in I$ and 0, otherwise, $\widehat{z}_i = 1, i \in I$ and 0, otherwise, while $\widehat{\eta} = \max_{i \in I} \|\widehat{\boldsymbol{x}}\|_{\frac{q}{q-1}} / (\tau \|\boldsymbol{\zeta}_i\|_1)$ and $\delta = \frac{N\epsilon - \lfloor \epsilon N \rfloor}{N \widehat{\eta}}$.

Clearly, $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\alpha}}, \widehat{\eta})$ is feasible to (EC.14) and

$$\bar{v}^{sc}_{\epsilon,\mathcal{P}^W}(\delta) \leq \boldsymbol{c}^\top \widehat{\boldsymbol{x}} = \boldsymbol{c}^\top \boldsymbol{x}^* + \|\boldsymbol{c}\|_1 \tau = v^{ccsc} + \tau.$$

◇

By the above Claim, we also have for any positive number $\tau > 0$, there exists a $\delta > 0$ such that

$$v^{sc}_{\epsilon,\mathcal{P}^W}(\delta) \leq \boldsymbol{c}^\top \widehat{\boldsymbol{x}} = \boldsymbol{c}^\top \boldsymbol{x}^* + \|\boldsymbol{c}\|_1 \tau = v^{ccsc} + \tau.$$

By choosing an appropriate δ, we can approximate chance constrained linear covering problem by solving $v^{sc}_{\epsilon,\mathcal{P}^W}(\delta)$ to arbitrary accuracy. Hence, the DRCCP with Wasserstein ambiguity set is also NP-hard. This completes the proof. □

EC.2. Examples

Example EC.1 Given a parameter $\kappa > 3$, consider the following instance of DRCCP:

$$(\mathrm{DRCCP}_\kappa): \quad v^*_\kappa = \min_{x \geq 0} \left\{ x : \sup_{\mathbb{P} \in \mathcal{P}^W} \mathbb{P}\{\tilde{\xi} : \tilde{\xi} x < 1\} \leq \epsilon \right\},$$

where $\mathcal{P}^W$ denotes Wasserstein ambiguity set (28) with $q = 2$ and in (28), the empirical distribution $\tilde{\zeta}$ has a finite support with $N = 2$ realizations $\{\zeta_i\}_{i \in [N]}$ such that $p_1 = \frac{1}{\kappa}, p_2 = \frac{\kappa - 1}{\kappa}$, $\zeta_1 = 1$ and $\zeta_2 = \kappa$. Suppose that $\Xi = \mathbb{R}_+$, $\epsilon = \frac{1}{\kappa} + \frac{1}{\kappa^2}$ and $\delta = \frac{1}{\kappa^2(\kappa - 1)}$. Under this setting, DRCCP_κ is equivalent to

$$v^*_\kappa = \min_{x \geq 0} \{x : x \in X_\epsilon\},$$

where

$$X_\epsilon = \left\{ x \in \mathbb{R}_+ : \begin{array}{l} \dfrac{1}{\kappa(\kappa-1)} \lambda + \eta_1 + (\kappa - 1)\eta_2 \leq 1 + \dfrac{1}{\kappa}, \\ \eta_1 + \min\{\lambda, \pi_1 x\} \geq 1 + \pi_1, \\ \eta_2 + \kappa \min\{\lambda, \pi_2 x\} \geq 1 + \pi_2, \\ \lambda, \eta_1, \eta_2, \pi_1, \pi_2 \geq 0. \end{array} \right\}$$

Let $v^{\mathbf{CVaR}}_\kappa, v^{\mathrm{rel}}_{\kappa,\phi}$ denote the optimal value of the worst-case **CVaR** approximation and the optimal value of Markovian relaxation of DRCCP_κ, respectively, i.e.,

$$v^{\mathbf{CVaR}}_\kappa = \min_{x \geq 0} \left\{ x : x \in X^{\mathbf{CVaR}}_\epsilon \right\}, \quad v^{\mathrm{rel}}_{\kappa,\phi} = \min_{x \geq 0} \left\{ x : x \in X^{\mathrm{rel}}_{\mathcal{P},\phi} \right\},$$

where

$$X_{\epsilon}^{\mathbf{CVaR}} = \left\{ x \in \mathbb{R}_+ : \begin{array}{l} \frac{1}{\kappa(\kappa-1)}\lambda + \eta_1 + (\kappa-1)\eta_2 \leq \left(1+\frac{1}{\kappa}\right)\beta, \\ \eta_1 + \min\{\lambda, x\} \geq 1+\beta, \\ \eta_2 + \kappa \min\{\lambda, x\} \geq 1+\beta, \\ \lambda, \eta_1, \eta_2 \geq 0, \end{array} \right\}$$

and

$$X_{\mathcal{P},\phi}^{\mathrm{rel}} = \left\{ x \in \mathbb{R}_+ : \begin{array}{l} \frac{1}{\kappa(\kappa-1)}\lambda + \eta_1 + (\kappa-1)\eta_2 \leq 1+\frac{1}{\kappa}, \\ \eta_1 + \min\{\lambda, x\} \geq 1, \\ \eta_2 + \kappa \min\{\lambda, x\} \geq 1, \\ \lambda, \eta_1, \eta_2 \geq 0. \end{array} \right\}$$

We illustrate how to compute the exact values or upper bound of $v_\kappa^*, v_\kappa^{\mathbf{CVaR}}, v_{\kappa,\phi}^{\mathrm{rel}}$ as below:

- Note that $v_\kappa^* \leq \frac{1}{\kappa-2} \leq 1$, where equality can be achieved with a feasible solution $x = \frac{1}{\kappa-2}, \lambda = \kappa - 1 - \frac{1}{\kappa}, \eta_1 = 1, \eta_2 = 0, \pi_1 = 0, \pi_2 = \left(\kappa - 1 - \frac{1}{\kappa}\right)(\kappa-2)$.
- To obtain $v_\kappa^{\mathbf{CVaR}}$, note that by breaking down the minimum operators, we have

$$\begin{aligned} v_\kappa^{\mathbf{CVaR}} = \min_{x\geq 0} \ & x \\ \text{s.t.} \quad & \frac{1}{\kappa(\kappa-1)}\lambda + \eta_1 + (\kappa-1)\eta_2 \leq \left(1+\frac{1}{\kappa}\right)\beta, \\ & \eta_1 + \lambda \geq 1+\beta, \\ & \eta_1 + x \geq 1+\beta, \\ & \eta_2 + \kappa\lambda \geq 1+\beta, \\ & \eta_2 + \kappa x \geq 1+\beta, \\ & \lambda, \eta_1, \eta_2 \geq 0. \end{aligned}$$

  By projecting out auxiliary variable β, the above model becomes

$$\begin{aligned} v_\kappa^{\mathbf{CVaR}} = \min_{x\geq 0} \ & x \\ \text{s.t.} \quad & 1+\frac{1}{\kappa} \leq \left(1+\frac{2}{\kappa}-\frac{1}{\kappa-1}\right)\lambda + \frac{\eta_1}{\kappa} - (\kappa-1)\eta_2, \\ & \left(1+\frac{1}{\kappa}\right)(1-x) \leq \left(\frac{1}{\kappa}-\frac{1}{\kappa-1}\right)\lambda + \frac{\eta_1}{\kappa} - (\kappa-1)\eta_2, \\ & 1+\frac{1}{\kappa} \leq \left(\kappa+1+\frac{1}{\kappa}-\frac{1}{\kappa-1}\right)\lambda - \eta_1 - \left(\kappa-2-\frac{1}{\kappa}\right)\eta_2, \\ & \left(1+\frac{1}{\kappa}\right)(1-\kappa x) \leq \left(\frac{1}{\kappa}-\frac{1}{\kappa-1}\right)\lambda - \eta_1 - \left(\kappa-2-\frac{1}{\kappa}\right)\eta_2, \\ & \lambda, \eta_1, \eta_2 \geq 0. \end{aligned}$$

  Due to $\eta_2 \geq 0$ and $\kappa > 3$, thus, we can further project out η_2 as below:

$$v_\kappa^{\mathbf{CVaR}} = \min_{x\geq 0} x$$

$$
\begin{aligned}
\text{s.t.}\quad & 1+\frac{1}{\kappa} \leq \left(1+\frac{2}{\kappa}-\frac{1}{\kappa-1}\right)\lambda+\frac{\eta_1}{\kappa},\\
& \left(1+\frac{1}{\kappa}\right)(1-x) \leq \left(\frac{1}{\kappa}-\frac{1}{\kappa-1}\right)\lambda+\frac{\eta_1}{\kappa},\\
& 1+\frac{1}{\kappa} \leq \left(\kappa+1+\frac{1}{\kappa}-\frac{1}{\kappa-1}\right)\lambda-\eta_1,\\
& \left(1+\frac{1}{\kappa}\right)(1-\kappa x) \leq \left(\frac{1}{\kappa}-\frac{1}{\kappa-1}\right)\lambda-\eta_1,\\
& \lambda,\eta_1 \geq 0.
\end{aligned}
$$

Next, projecting out variable η_1, we have

$$
\begin{aligned}
v_\kappa^{\mathbf{CVaR}} = & \min_{x\geq 0} x\\
\text{s.t.}\quad & \kappa^2-1 \leq \left(2\kappa^2-2\kappa-1\right)\lambda,\\
& (\kappa-1)(\kappa+1-\kappa x) \leq \left(\kappa^2-\kappa-1\right)\lambda,\\
& \lambda \leq (\kappa-1)(2\kappa x-\kappa-1),\\
& \lambda \geq 0.
\end{aligned}
$$

By solving the above 2-dimensional linear program, we can obtain the following optimal value $v_\kappa^{\mathbf{CVaR}} = \frac{1+\frac{1}{\kappa}}{2-\frac{1}{\kappa(\kappa-1)}}$ with an optimal solution $x=\lambda=\frac{1+\frac{1}{\kappa}}{2-\frac{1}{\kappa(\kappa-1)}}, \eta_1=\frac{\kappa-\frac{1}{\kappa}}{2-\frac{1}{\kappa(\kappa-1)}}, \eta_2=0, \beta=\frac{\kappa+1}{2-\frac{1}{\kappa(\kappa-1)}}-1$, and

- To obtain $v_{\kappa,\phi}^{\mathrm{rel}}$, first of all, breaking down the minimum operators yields

$$
\begin{aligned}
v_{\kappa,\phi}^{\mathrm{rel}} = & \min_{x\geq 0} x\\
\text{s.t.}\quad & \frac{1}{\kappa(\kappa-1)}\lambda+\eta_1+(\kappa-1)\eta_2 \leq 1+\frac{1}{\kappa},\\
& \eta_1+\lambda \geq 1\\
& \eta_1+x \geq 1,\\
& \eta_2+\kappa\lambda \geq 1,\\
& \eta_2+\kappa x \geq 1,\\
& \lambda,\eta_1,\eta_2 \geq 0.
\end{aligned}
$$

Next, by projecting out variable η_2, we have

$$
\begin{aligned}
v_{\kappa,\phi}^{\mathrm{rel}} = & \min_{x\geq 0} x\\
\text{s.t.}\quad & \frac{1}{\kappa(\kappa-1)}\lambda+\eta_1 \leq 1+\frac{1}{\kappa},\\
& \left(\frac{1}{\kappa(\kappa-1)}-\kappa(\kappa-1)\right)\lambda+\eta_1 \leq 2+\frac{1}{\kappa}-\kappa,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{\kappa(\kappa-1)}\lambda+\eta_1\le 2+\frac{1}{\kappa}-\kappa+\kappa(\kappa-1)x,\\
&\eta_1+\lambda\ge 1\\
&\eta_1+x\ge 1,\\
&\lambda,\eta_1\ge 0.
\end{aligned}
$$

Next, project out variable η_1 and we have

$$
\begin{aligned}
v^{\text{rel}}_{\kappa,\phi}=&\min_{x\ge 0} x\\
\text{s.t.}\quad &\frac{1}{\kappa(\kappa-1)}\lambda\le 1+\frac{1}{\kappa},\\
&\left(\frac{1}{\kappa(\kappa-1)}-\kappa(\kappa-1)\right)\lambda\le 2+\frac{1}{\kappa}-\kappa,\\
&\frac{1}{\kappa(\kappa-1)}\lambda\le 2+\frac{1}{\kappa}-\kappa+\kappa(\kappa-1)x,\\
&\left(\frac{1}{\kappa(\kappa-1)}-\kappa(\kappa-1)-1\right)\lambda\le 1+\frac{1}{\kappa}-\kappa,\\
&\left(\frac{1}{\kappa(\kappa-1)}-1\right)\lambda\le 1+\frac{1}{\kappa}-\kappa+\kappa(\kappa-1)x,\\
&\frac{1}{\kappa(\kappa-1)}\lambda\le \frac{1}{\kappa}+x,\\
&\left(\frac{1}{\kappa(\kappa-1)}-\kappa(\kappa-1)\right)\lambda\le 1+\frac{1}{\kappa}-\kappa+x,\\
&\frac{1}{\kappa(\kappa-1)}\lambda\le 1+\frac{1}{\kappa}-\kappa+\kappa(\kappa-1)x+x,\\
&\lambda\ge 0.
\end{aligned}
$$

By solving the above 2-dimensional linear program, we can have $v^{\text{rel}}_{\kappa,\phi}=\frac{\kappa-1-\frac{1}{\kappa}}{\kappa^2-\kappa+1-\frac{1}{\kappa(\kappa-1)}}$ with an optimal solution $x=\lambda=\frac{\kappa-1-\frac{1}{\kappa}}{\kappa^2-\kappa+1-\frac{1}{\kappa(\kappa-1)}},\eta_1=1-\frac{\kappa-1-\frac{1}{\kappa}}{\kappa^2-\kappa+1-\frac{1}{\kappa(\kappa-1)}},\eta_2=\frac{2-\frac{1}{\kappa(\kappa-1)}}{\kappa^2-\kappa+1-\frac{1}{\kappa(\kappa-1)}}$. Thus, in this example, since this DRCCP_κ is a univariate optimization, thus lines 4-7 of Algorithm 1 will always find the optimal value v^*_κ. Therefore, the "*theoretical*" approximation ratio γ_κ of the Relax-and-Scale scheme is bounded by

$$
1\le\gamma_\kappa=\frac{v^*_\kappa}{v^{\text{rel}}_{\kappa,\phi}}\le\frac{\kappa^2-\kappa+1-\frac{1}{\kappa(\kappa-1)}}{\left(\kappa-1-\frac{1}{\kappa}\right)(\kappa-2)}.
$$

On the other hand, for the **CVaR** approximation, its approximation quality is

$$
\frac{v^{\mathbf{CVaR}}_\kappa}{v^*_\kappa}\ge\frac{\kappa(\kappa-1)(\kappa-2)}{2\kappa^2-2\kappa-1}.
$$

Hence, we have

$$
\lim_{\kappa\to\infty}\frac{v^{\mathbf{CVaR}}_\kappa}{v^*_\kappa}=\infty,\lim_{\kappa\to\infty}\gamma_\kappa=1,
$$

i.e., the worst-case **CVaR** approximation can be arbitrarily bad but the theoretical approximation ratio of the Relax-and-Scale scheme tends to be close to 1. ◇

The above example demonstrates that our proposed Relax-and-Scale scheme can be significantly better than the worst-case **CVaR** approach. On the other hand, the next example shows that the worst-case **CVaR** approximation can be tight, while the Markovian relaxation can be arbitrarily bad, i.e., the Relax-and-Scale scheme might not find a feasible solution to a DRCCP under Wasserstein ambiguity set when the violation ratio $\sigma = 1$.

Example EC.2 Given a parameter $\kappa > 1$, consider the following instance of DRCCP:

$$(\mathrm{DRCCP}_\kappa): \quad v_\kappa^* = \min_{x \geq 0} \left\{ x : \sup_{\mathbb{P} \in \mathcal{P}^W} \mathbb{P}\{\tilde{\xi} : \tilde{\xi} x < 1\} \leq \epsilon \right\},$$

where $\mathcal{P}^W$ denotes Wasserstein ambiguity set (28) with $q = 2$, and in (28), the empirical distribution $\tilde{\zeta}$ has a finite support with $N = 2$ realizations $\{\zeta_i\}_{i \in [2]}$ such that $p_1 = \epsilon, p_2 = 1 - \epsilon$, $\zeta_1 = \frac{1}{\kappa}, \zeta_2 = 1$. Suppose that $\Xi = \mathbb{R}_+$ and $\delta = \frac{\epsilon}{2\kappa}$. Under this setting, DRCCP_κ is equivalent to

$$v_\kappa^* = \min_{x \geq 0} \{x : x \in X_\epsilon\},$$

where

$$X_\epsilon = \left\{ x \in \mathbb{R}_+ : \begin{array}{l} \dfrac{\epsilon}{2\kappa}\lambda + \epsilon \eta_1 + (1 - \epsilon)\eta_2 \leq \epsilon, \\ \eta_1 + \dfrac{1}{\kappa} \min\{\lambda, \pi_1 x\} \geq 1 + \pi_1, \\ \eta_2 + \min\{\lambda, \pi_2 x\} \geq 1 + \pi_2, \\ \lambda, \eta_1, \eta_2, \pi_1, \pi_2 \geq 0. \end{array} \right\}$$

Let $v_\kappa^{\mathbf{CVaR}}, v_{\kappa,\phi}^{\mathrm{rel}}$ denote the optimal value of the worst-case **CVaR** approximation and the optimal value of Markovian relaxation of DRCCP_κ, respectively, i.e.,

$$v_\kappa^{\mathbf{CVaR}} = \min_{x \geq 0} \left\{ x : x \in X_\epsilon^{\mathbf{CVaR}} \right\}, \quad v_{\kappa,\phi}^{\mathrm{rel}} = \min_{x \geq 0} \left\{ x : x \in X_{\mathcal{P},\phi}^{\mathrm{rel}} \right\},$$

where

$$X_\epsilon^{\mathbf{CVaR}} = \left\{ x \in \mathbb{R}_+ : \begin{array}{l} \dfrac{\epsilon}{2\kappa}\lambda + \epsilon \eta_1 + (1 - \epsilon)\eta_2 \leq \epsilon\beta, \\ \eta_1 + \dfrac{1}{\kappa} \min\{\lambda, x\} \geq 1 + \beta, \forall i \in [\tau], \\ \eta_2 + \min\{\lambda, x\} \geq 1 + \beta, \forall i \in [N] \setminus [\tau], \\ \lambda, \eta_1, \eta_2 \geq 0, \end{array} \right\}$$

and

$$X_{\mathcal{P},\phi}^{\mathrm{rel}} = \left\{ x \in \mathbb{R}_+ : \begin{array}{l} \dfrac{\epsilon}{2\kappa}\lambda + \epsilon \eta_1 + (1 - \epsilon)\eta_2 \leq \epsilon, \\ \eta_1 + \dfrac{1}{\kappa} \min\{\lambda, x\} \geq 1, \\ \eta_2 + \min\{\lambda, x\} \geq 1, \\ \lambda, \eta_1, \eta_2 \geq 0. \end{array} \right\}$$

Following the similar proof of Example EC.1, i.e., projecting out auxiliary variables via Fourier-Motzkin elimination method (for the page limit, we omit the derivation here), we see that

- $v_\kappa^* = 2\kappa$ with an optimal solution $x = 2\kappa, \lambda = \frac{2\kappa}{2\kappa-1}, \eta_1 = \frac{2\kappa-2}{2\kappa-1}, \eta_2 = 0, \pi_1 = \pi_2 = \frac{1}{2\kappa-1}$,
- $v_\kappa^{\mathbf{CVaR}} = 2\kappa$ with an optimal solution $x = 2\kappa, \lambda = 2\kappa, \eta_1 = 2\kappa - 2, \eta_2 = 0, \beta = 2\kappa - 1$, and
- $v_{\kappa,\phi}^{\text{rel}} = \frac{2\kappa(1-\epsilon)}{2\kappa(1-\epsilon)+\epsilon}$ with an optimal solution $x = \frac{2\kappa(1-\epsilon)}{2\kappa(1-\epsilon)+\epsilon}, \lambda = \frac{2\kappa(1-\epsilon)}{2\kappa(1-\epsilon)+\epsilon}, \eta_1 = \frac{2(\kappa-1)(1-\epsilon)+\epsilon}{2\kappa(1-\epsilon)+\epsilon}, \eta_2 = \frac{\epsilon}{2\kappa(1-\epsilon)+\epsilon}$.

Thus, in this example, since lines 4-7 of Algorithm 1 will always find the optimal value v_κ^*, thus the theoretical approximation ratio γ_κ of the Relax-and-Scale scheme is

$$\gamma_\kappa = \frac{v_\kappa^*}{v_{\kappa,\phi}^{\text{rel}}} = 2\kappa + \frac{\epsilon}{1-\epsilon}.$$

On the other hand, for the **CVaR** approximation, its approximation quality is

$$\frac{v_\kappa^{\mathbf{CVaR}}}{v_\kappa^*} = 1.$$

Thus, we have

$$\lim_{\kappa\to\infty} \frac{v_\kappa^{\mathbf{CVaR}}}{v_\kappa^*} = 1, \lim_{\kappa\to\infty} \gamma_\kappa = \infty,$$

i.e., the worst-case **CVaR** approximation is tight but the Relax-and-Scale scheme can be arbitrarily bad. ◇