RESEARCH ARTICLE

Conic reformulations for Kullback-Leibler divergence constrained distributionally robust optimization and applications

Burak Kocuk[*]

Industrial Engineering Program, Faculty of Engineering and Natural Sciences, Sabancı University, Istanbul, Turkey
burak.kocuk@sabanciuniv.edu

ARTICLE INFO	ABSTRACT
	In this paper, we consider a Kullback-Leibler divergence constrained distributionally robust optimization model. This model considers an ambiguity set that consists of all distributions whose Kullback-Leibler divergence to an empirical distribution is bounded. Utilizing the fact that this divergence measure has an exponential cone representation, we obtain the robust counterpart of the Kullback-Leibler divergence constrained distributionally robust optimization problem as a dual exponential cone constrained program under mild assumptions on the underlying optimization problem. The resulting conic reformulation of the original optimization problem can be directly solved by a commercial conic programming solver. We specialize our generic formulation to two classical optimization problems, namely, the Newsvendor Problem and the Uncapacitated Facility Location Problem. Our computational study in an out-of-sample analysis shows that the solutions obtained via the distributionally robust optimization approach yield significantly better performance in terms of the dispersion of the cost realizations while the central tendency deteriorates only slightly compared to the solutions obtained by stochastic programming.

1. Introduction

Decision making under uncertainty is one of the most challenging tasks in operations research. Two paradigms are predominantly used in the literature to address uncertainty: stochastic programming and robust optimization. In the classical stochastic programming [1, 2], a predefined set of scenarios (or samples) are determined, either taken directly from observed data or after fitting an appropriate distribution. Then, the objective function is replaced with an expectation taken with respect to the random elements, and constraints are copied for each scenario. In addition to the assumption about knowing the underlying distribution, this basic stochastic programming approach has some limitations: Firstly, the size of the resulting deterministic equivalent formulation grows larger with the size of the scenarios. Secondly, an expectation may not be an appropriate performance measure for risk-averse decision makers. Thirdly, satisfying constraints for each scenario might be too restrictive. The respective remedies for these shortcomings are proposed such as sample average approximation to limit the problem size, risk-averse objective function for a more appropriate performance measure and chance constraints to allow constraint satisfaction with high probability. However, the implicit assumption of stochastic programming remains, which is the need to *assume* a distribution by analyzing the data or *fitting* one. Unfortunately, this step may not be performed satisfactorily in all cases.

---

*Corresponding author

In robust optimization [3–5], a predefined uncertainty set, which includes all possible values of the uncertain elements, is used. Then, the optimization is performed with the aim of optimizing with respect to the worst possible realization from the uncertainty set. There are two main advantages of using robust optimization. Firstly, the decision maker does not need to make any assumptions about the distribution of the uncertain elements in the problem as opposed to the stochastic programming approach. Secondly, the deterministic equivalent (or so-called the *robust counterpart*) formulation of the robust optimization problem typically has the same computational complexity as the deterministic version of the problem under reasonable assumptions on the uncertainty sets. On the other hand, the main disadvantage of the robust optimization approach is that depending on the construction of the uncertainty set, it might lead to overly conservative solutions, which might have poor performance in central tendency such as expectation.

Distributionally robust optimization (DRO) is a relatively new paradigm that aims to combine stochastic programming and robust optimization approaches. The main modeling assumption of DRO is that some *partial* information about the distribution governing the underlying uncertainty is available, and the optimization is performed with respect to the worst distribution from an *ambiguity set*, which contains all distributions consistent with this partial information. There are mainly two streams in the DRO literature based on how the ambiguity set is defined: moment-based and distance-based.

In moment-based DRO, ambiguity sets are defined as the set of distributions whose first few moments are assumed to be known or constrained to lie in certain subsets. If certain structural properties hold for the ambiguity sets such as convexity (or conic representability), then tractable convex (or conic reformulations) can be obtained [6–8]. In distance-based DRO, ambiguity sets are defined as the set of distributions whose distance (or divergence) from a reference distribution is constrained. For Wasserstein distance [9–12] and ϕ−divergence [13–17] constrained DRO, tractable convex reformulations have been proposed. Recently, chance constrained DRO problems have also drawn attention [18–21].

As summarized above, in many cases, tractable convex robust counterparts or reformulations can be obtained for robust and distributionally robust (DR) optimization problems. However, an even more special structure such as conic representability can be preferred whenever available.

Especially, if the robust counterpart can be expressed as a conic program for which the underlying cone admits a self-concordant barrier, then efficient polynomial-time interior point methods can be applied directly [22]. This desired property holds for linear programs, second-order cone programs and semidefinite programs, which appear extensively in both robust and DR optimization literature. We note that the efficiency of the conic programming solvers specialized in these three problem classes has improved considerably.

There is some recent interest in conic programs for which the underlying cone is not self-dual, such as exponential cone. There are two main reasons: i) exponential cone has extensive expressive power that is useful to model optimization problems involving the exponential and logarithm functions (see e.g. [23]), and ii) a practical implementation of a primal-dual interior point method is developed [24] although its polynomial-time complexity has not been proven yet. Our paper will exploit both the expressive power of the exponential cone and the practical implementation that can be used to solve the resulting optimization problem, as detailed below.

The Kullback-Leibler (KL) divergence [25] is a popular divergence measure in information theory that can be used to quantify the divergence of one distribution from another (see Definition 8) and we prove that it is exponential cone representable (see Definitions 3 and 6, and Proposition 2). Although the robust counterpart of KL divergence constrained DRO is proven to be a tractable convex program [26], to the best of our knowledge, its exponential cone representability has not been exploited in the literature before. Also, its practical performance against stochastic programming has not been analyzed in detail except for a limited number of applications from power systems [27, 28].

In this paper, we consider KL divergence constrained DRO problems and propose their dual exponential cone constrained reformulation under the mild assumption of conic representability. This allows us to solve the corresponding robust counterpart using a conic programming solver such as MOSEK [29]. We also present how the generic formulation can be specialized for two classical problems: Newsvendor and Uncapacitated Facility Location. Although the DRO methodology has been applied to variations of these problems [30–35], to the best of our knowledge, their KL divergence constrained versions have not been studied in detail. Our computational results suggest that solutions obtained via

a DR approach give slightly higher cost realizations when central tendencies such as mean and median are considered compared to solutions obtained via stochastic programming in an out-of-sample analysis. However, the dispersion (measured by the standard deviation and range of the cost realizations) and the risk (measured by the average of worst cost realizations and the third quartile) metrics improve significantly with solutions obtained via a DR approach.

Our main contribution in this paper is to exploit the exponential cone representation of the KL-divergence to solve DRO problems with ambiguity sets defined using this measure[1]. Unlike the previous literature, which treats such problems as "general convex programs" (see, for instance, [13]), we utilize the general-purpose conic programming solver MOSEK in our computations. The conic representability property can be quite useful in practice since one can then include other (mixed-integer) conic representable sets and functions to the underlying optimization problem and, hence, model a wide variety of real-life situations.

Although we carry out the computational experiments for two classical problems, we remark that our approach is rather general and it can be adapted to various settings. To give a few examples, one can use our framework in applications from portfolio optimization [7,10], image processing [9,16], asset pricing [13], multidimensional knapsack problem [12,20] and logistics [35,36].

The rest of the paper is organized as follows: In Section 2, we review basic concepts from convex analysis and probability theory which serve as the basis of our main result about conic reformulation of KL divergence constrained DRO problems in Section 3. Then, we analyze two applications, namely, the Newsvendor Problem in Section 4 and the Uncapacitated Facility Location Problem in Section 5, and present the results of our computational study. Finally, we conclude our paper in Section 6.

2. Preliminaries

Before stating our main result in Section 3, we will first review some important concepts from convex analysis in Section 2.1 and probability theory in Section 2.2.

2.1. Convex analysis

For a set $X \subseteq \mathbb{R}^m$, we denote its interior as $\text{int}(X)$, its relative interior as $\text{ri}(X)$ and its closure as $\text{cl}(S)$. We use the shorthand notation $[n]$ for the set $\{1, \ldots, n\}$.

We will first review some basic concepts from convex analysis related to cones.

Definition 1 (Regular cone). *A cone $K \subseteq \mathbb{R}^m$ is called regular if it is closed, convex, pointed and full-dimensional.*

Examples of regular cones include the nonnegative orthant, Lorentz (or second-order) cone and the cone of positive semidefinite matrices. We will refer to these cones as *canonical cones* in this paper.

Definition 2 (Dual cone). *The dual cone to a cone $K \subseteq \mathbb{R}^m$ is defined as $K_* = \{y \in \mathbb{R}^m : x^T y \geq 0, \ \forall x \in K\}$.*

It is well-known that the dual cone to a regular cone is also regular. In addition, the three canonical cones mentioned above are self-dual.

We will now define the exponential cone, which is the key ingredient of this paper.

Definition 3 (Exponential cone). *The exponential cone, denoted as K_{exp}, is defined as*

$$K_{\exp} = \text{cl}\big(\{x \in \mathbb{R}^3 : x_1 \geq x_2 e^{x_3/x_2}, \ x_2 > 0\}\big).$$

As opposed to the three canonical cones mentioned above, the exponential cone is not self-dual although it is a regular cone.

Proposition 1 (See e.g. [23]). *The dual cone to the exponential cone (or simply the dual exponential cone) is given as*

$$(K_{\exp})_* =$$
$$\text{cl}\big(\{s \in \mathbb{R}^3 : s_1 \geq -s_3 e^{(s_2-s_3)/s_3}, \ s_3 < 0\}\big).$$

The following definitions are instrumental in the description of conic programming problems:

Definition 4 (Conic inequality). *A conic inequality with respect to a regular cone K is defined as $x \succeq_K y$, meaning that $x - y \in K$. We will denote the relation $x \in \text{int}(K)$ alternatively as $x \succ_K 0$.*

Definition 5 (Conic representability of a set). *A set $X \subseteq \mathbb{R}^n$ is called conic representable if it can be expressed as*

$$X = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^k : Ax + By \succeq_K b\},$$

for some appropriately chosen regular cone K.

---

[1]We also note that two other $\phi-$divergence measures called Burg entropy and $J-$divergence have the same representation property.

If any of the variables in this representation is integer, then X is called a mixed-integer conic representable set.

Definition 6 (Conic representability of a function). *A function is called conic representable if its epigraph is a conic representable set.*

In this paper, when we say that "a set or function is conic representable", we will implicitly assume that the cone used in its representation is either one of the three canonical cones or the (dual) exponential cone.

2.2. Probability theory

Definition 7 (Probability simplex). *The probability simplex in dimension S is denoted as*

$$\Delta^S := \Big\{ p \in \mathbb{R}^S_+ : \sum_{s=1}^S p_s = 1 \Big\}.$$

The following function can be used to measure the "divergence" of one distribution from another.

Definition 8 (KL Divergence). *For two discrete probability distributions $p \in \Delta^S$ and $q \in \mathrm{ri}(\Delta^S)$, the KL divergence from p to q is defined as*

$$D_{KL}(p||q) := \sum_{s=1}^S p_s \log(p_s/q_s).$$

We note that the KL Divergence does not define a distance metric between two probability distributions since it is not symmetric. However, it has the following useful property.

Proposition 2. *Let $p \in \Delta^S$ and $q \in \mathrm{ri}(\Delta^S)$. Then, the function $D_{KL}(p||q)$ is exponential cone representable.*

Proof. Due to Definition 6, it suffices to show that the epigraph of the function $D_{KL}(p||q)$ is an exponential cone representable set. Since the set $\{(x, y, t) \in \mathbb{R}^3 : t \geq x \log(x/y)\}$ has the exponential cone representation $(y, x, -t) \in K_{\exp}$ [37], we obtain an exponential cone representation for the function $D_{KL}(p||q)$ as follows:

$$\{(p, q, \epsilon) \in \Delta^S \times \mathrm{ri}(\Delta^S) \times \mathbb{R} : D_{KL}(p||q) \leq \epsilon\}$$

$$= \Big\{(p, q, \epsilon) : \exists \delta \in \mathbb{R}^S : \sum_{s=1}^S \delta_s \leq \epsilon,$$

$$(q_s, p_s, -\delta_s) \in K_{\exp}, \ s \in [S]\Big\}. \qquad \square$$

The following proposition gives an upper bound on the KL-divergence of a given distribution from any other distribution.

Proposition 3. *Let $q \in \mathrm{ri}(\Delta^S)$. Then,*

$$\bar{\epsilon}(q) := \sup_{p \in \Delta^S} \{D_{KL}(p||q)\} = \log(1/\min_{s \in [S]}\{q_s\}).$$

Proof. Notice that the objective function of the optimization problem $\sup_{p \in \Delta^S}\{D_{KL}(p||q)\}$ is convex and its feasible region is a polytope. Therefore, there exists an optimal solution which is an extreme point of Δ^S. Observe that the extreme points of Δ^S are the unit vectors in $\mathbb{R}^S$, denoted by $\tilde{e}^s$ for $s \in [S]$ (note that $\tilde{e}^s_{s'} = 1$ for $s = s'$, and $\tilde{e}^s_{s'} = 0$ otherwise).
Let us now compute $D_{KL}(\tilde{e}^s||q)$ for some $s \in [S]$. In fact, we have

$$D_{KL}(\tilde{e}^s||q) = \sum_{s'=1}^S \tilde{e}^s_{s'} \log(\tilde{e}^s_{s'}/q_s)$$

$$= \tilde{e}^s_s \log(\tilde{e}^s_s/q_s) + \sum_{\substack{s'=1 \\ s' \neq s}}^S \tilde{e}^s_{s'} \log(\tilde{e}^s_{s'}/q_s)$$

$$= \log(1/q_s).$$

Here, we use the fact that $\lim_{x \to 0^+} x \log(x/y) = 0$ for $y > 0$ in the last equality (recall that $q \in \mathrm{ri}(\Delta^S)$, which implies that $q_s > 0$, $s \in [S]$).
Finally, we have

$$\bar{\epsilon}(q) = \sup_{p \in \Delta^S} \{D_{KL}(p||q)\}$$

$$= \max_{s \in [S]}\{D_{KL}(\tilde{e}^s||q)\}$$

$$= \max_{s \in [S]}\{\log(1/q_s)\}$$

$$= \log(1/\min_{s \in [S]}\{q_s\}).$$

$$\square$$

Proposition 3 is useful to quantify the ambiguity sets in KL divergence constrained DRO problems as we will see later.

3. Main results

In this section, we present our main result about the reformulation of a KL divergence constrained DRO problem as a conic program under mild conditions.

3.1. Generic problem formulation

We first give the generic problem setting considered in this paper. Suppose that there are m random variables $\xi^i \in \mathbb{R}$, $i \in [m]$, each with a discrete distribution $q^i \in \mathrm{ri}(\Delta^{S_i})$ estimated from the historical data as

$$\Pr(\xi^i = d^i_s) = q^i_s \quad s \in [S_i],$$

where $\{d^i_s : s \in [S_i]\}$ is the set of observed realizations of ξ^i, $i \in [m]$. Under this probabilistic

setting, we define the ambiguity set

$$\mathcal{P}^i(q^i, \epsilon^i) := \{p^i \in \Delta^{S_i} : \ D_{KL}(p^i \| q^i) \le \epsilon^i\},$$

for $i \in [m]$, where $\epsilon^i \in \mathbb{R}_+$ controls the divergence from the historical data (or robustness level). Then, we consider the following KL divergence constrained DRO problem,

$$\min_{y \in \mathcal{Y}} \Big\{ h(y) + \sum_{i=1}^m \max_{p^i \in \mathcal{P}^i(q^i, \epsilon^i)} \mathbb{E}_{\xi^i}[H^i(y, \xi^i)] \Big\}, \quad (1)$$

where each expectation is taken with respect to an ambiguous distribution $p^i \in \mathcal{P}^i(q^i, \epsilon^i)$. In problem (1), h is a real-valued function defined on $\mathbb{R}^n$; H^i is a real-valued function defined on $\mathbb{R}^n \times \mathbb{R}$; and $\mathcal{Y}$ is a subset of $\mathbb{R}^n$. Observe that given y decisions, the inner maximization problem is decomposable over the random elements ξ^i, $i \in [m]$. Although we are assuming discrete probability distributions in this paper, our framework can be used to solve problems involving continuous distributions via a finite support approximation.

3.2. Robust counterpart and conic reformulation

We will now obtain the robust counterpart [4] of problem (1) utilizing Conic Duality under mild conditions.

Theorem 1. *Consider the KL divergence constrained DRO problem* (1) *as described in Section 3.1, and assume that $\epsilon^i > 0$, $i \in [m]$. Then, the robust counterpart is given as follows:*

$$\min \ h(y) + \sum_{i=1}^m \left[\alpha^i + \epsilon^i \beta^i + \sum_{s=1}^{S_i} q_s^i u_s^i \right] \quad (2a)$$

$$\text{s.t. } \alpha^i - v_s^i \ge H^i(y, d_s^i) \quad i \in [m]; s \in [S_i] \quad (2b)$$

$$\beta^i + w_s^i = 0 \qquad i \in [m]; s \in [S_i] \quad (2c)$$

$$\alpha^i \in \mathbb{R}, \beta^i \in \mathbb{R}_+ \qquad i \in [m] \quad (2d)$$

$$(u_s^i, v_s^i, w_s^i) \in (K_{\exp})_* \ i \in [m]; s \in [S_i] \quad (2e)$$

$$y \in \mathcal{Y}. \quad (2f)$$

Proof. We will start the proof by analyzing the inner maximization problems. Given a vector $y \in \mathcal{Y}$, let us write the i-th inner maximization problem explicitly as the following exponential cone constrained program:

$$\max \ \sum_{s=1}^{S_i} H^i(y, d_s^i) p_s^i \quad (3a)$$

$$\text{s.t. } \sum_{s=1}^{S_i} p_s^i = 1 \quad (3b)$$

$$\sum_{s=1}^{S_i} \delta_s^i \le \epsilon^i \quad (3c)$$

$$\begin{bmatrix} 0 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_s^i \\ \delta_s^i \end{bmatrix} \preceq_{K_{\exp}} \begin{bmatrix} q_s^i \\ 0 \\ 0 \end{bmatrix} \quad s \in [S_i] \quad (3d)$$

$$p_s^i \in \mathbb{R}_+, \ \delta_s^i \in \mathbb{R} \qquad s \in [S_i]. \quad (3e)$$

Here, constraints (3b)-(3e) model the relation $p^i \in \mathcal{P}^i(q^i, \epsilon^i)$, as stated in Proposition 2.

Recall that $\epsilon^i > 0$ for each $i \in [m]$. Then, each inner maximization problem (3) satisfies essential strict feasibility [38] (e.g. consider $p_s^i = q_s^i$ and $\delta_i^s = \epsilon^i / |S_i|$ for $s \in [S_i]$), and its optimal value is bounded above (e.g. by $\max_{s \in [S_i]} H^i(y, d_s^i)$). Therefore, strong duality holds between problem (3) and its conic dual given as follows:

$$\min \ \alpha^i + \epsilon^i \beta^i + \sum_{s=1}^{S_i} q_s^i u_s^i \quad (4a)$$

$$\text{s.t. } \alpha^i - v_s^i \ge H^i(y, d_s^i) \qquad s \in [S_i] \quad (4b)$$

$$\beta^i + w_s^i = 0 \qquad s \in [S_i] \quad (4c)$$

$$\alpha^i \in \mathbb{R}, \beta^i \in \mathbb{R}_+ \quad (4d)$$

$$(u_s^i, v_s^i, w_s^i) \in (K_{\exp})_* \qquad s \in [S_i]. \quad (4e)$$

Here, α^i, β^i and (u_s^i, v_s^i, w_s^i) are the dual variables associated with the primal constraints (3b), (3c) and (3d), respectively. Notice that problem (4) is a dual exponential cone constrained program.

As the final step in the proof, we write the dual of each inner maximization problem and obtain the robust counterpart of problem (1) as problem (2). □

We will now discuss the consequences of Theorem 1 under additional structural properties such as convexity and conic representability.

Corollary 1. *Consider the KL divergence constrained DRO problem* (1) *as described in Theorem 1. In addition, let us assume that $\mathcal{Y}$ is a convex set, $h(y)$ and $H^i(y, \xi^i)$ are convex functions in y, $i \in [m]$. Then, the robust counterpart* (2) *is a convex program.*

Corollary 2. *Consider the KL divergence constrained DRO problem* (1) *as described in Theorem 1. In addition, let us assume that $\mathcal{Y}$ is a (mixed-integer) conic representable set, $h(y)$ and $H^i(y, \xi^i)$ are conic representable functions, $i \in [m]$. Then, the robust counterpart* (2) *is a dual exponential cone constrained (mixed-integer) program.*

As an application of Corollary 2, we will consider the Newsvendor Problem in Section 4 and the Uncapacitated Facility Location Problem in Section 5. The common characteristic of these two problems is that the set $\mathcal{Y}$ is a mixed-integer linear set, the function h is a linear function and the

functions H^i are the maxima of linear functions (hence, they are polyhedrally representable).

3.3. Extension to joint discrete probability distributions

We will now extend our analysis to the case of joint discrete probability distributions with a finite support. Suppose that we have a vector of random variables $\Xi \in \mathbb{R}^m$ with a joint probability distribution $q \in \mathrm{ri}(\Delta^S)$ estimated from the historical data as

$$\Pr(\Xi = \tilde{D}_s) = q_s \quad s \in [S],$$

where $\{\tilde{D}_s : s \in [S]\} \subseteq \mathbb{R}^m$ is the set of observed realizations of Ξ.

Consider the ambiguity set

$$\mathcal{P}(q, \epsilon) := \{p \in \Delta^S : D_{KL}(p||q) \leq \epsilon\},$$

where $\epsilon \in \mathbb{R}_+$, and the following KL divergence constrained DRO problem:

$$\min_{y \in \mathcal{Y}} \left\{ h(y) + \max_{p \in \mathcal{P}(q,\epsilon)} \mathbb{E}_{\Xi}[H(y,\Xi)] \right\}. \quad (5)$$

Here, h is a real-valued function defined on $\mathbb{R}^n$; H is a real-valued function defined on $\mathbb{R}^n \times \mathbb{R}^m$; and $\mathcal{Y}$ is a subset of $\mathbb{R}^n$. Then, we have the following result:

Theorem 2. *Consider the KL divergence constrained DRO problem* (5) *as described above, and assume that $\epsilon > 0$. Then, the robust counterpart is given as follows:*

$$\min \ h(y) + \left[\alpha + \epsilon\beta + \sum_{s=1}^{S} q_s u_s\right]$$

$$s.t. \ \alpha - v_s \geq H(y, \tilde{D}_s) \qquad s \in [S]$$
$$\beta + w_s = 0 \qquad\qquad s \in [S]$$
$$\alpha \in \mathbb{R}, \beta \in \mathbb{R}_+$$
$$(u_s, v_s, w_s) \in (K_{\exp})_* \qquad s \in [S]$$
$$y \in \mathcal{Y}.$$

We will omit the proof of Theorem 2 due to its similarity to the proof of Theorem 1. We remark that results similar to Corollaries 1 and 2 can be also obtained in this case under the assumptions of convexity and conic representability, respectively.

4. Application to the Newsvendor problem

In this section, we analyze a toy example, the KL divergence constrained DR version of the single-period, single-product Newsvendor Problem. In this case, since there is only one random variable ξ (that is, $m = 1$), representing the unknown demand, we will omit the superscript i for convenience.

4.1. Problem formulation

Consider the generic formulation (1) with the following specifications: We let $y \in \mathcal{Y} := \mathbb{Z}_+$ be the order quantity, and consider functions

$$h(y) := cy,$$

where c is the variable order cost, and

$$H(y, \xi) := c_b \max\{\xi - y, 0\} + c_h \max\{y - \xi, 0\},$$

where c_b is the back-order penalty for unsatisfied demand and c_h is the inventory cost. Notice that $H(y, \xi)$ is a piecewise linear convex function in y and can be rewritten as

$$H(y, \xi) = \max\{-c_b y + c_b \xi, c_h y - c_h \xi\}.$$

This observation will be useful to linearize constraint (2c).

By omitting i indices and simplifying the notation of problem (2) by taking into account the special structure of the newsvendor problem, we obtain the following dual exponential cone constrained MIP as its robust counterpart:

$$\min \ cy + \left[\alpha + \epsilon\beta + \sum_{s=1}^{S} q_s u_s\right] \qquad (6a)$$

$$s.t. \ \alpha - v_s \geq -c_b y + c_b d_s \qquad s \in [S] \qquad (6b)$$
$$\alpha - v_s \geq c_h y - c_h d_s \qquad s \in [S] \qquad (6c)$$
$$\beta + w_s = 0 \qquad\qquad s \in [S] \qquad (6d)$$
$$y \in \mathbb{Z}_+, \alpha \in \mathbb{R}, \beta \in \mathbb{R}_+ \qquad (6e)$$
$$(u_s, v_s, w_s) \in (K_{\exp})_* \qquad s \in [S]. \qquad (6f)$$

4.2. Computations

4.2.1. Experimental setup

To compare the effect of robustness level of KL divergence constrained DR version of the Newsvendor Problem, we propose Algorithm 1. Note that setting $\epsilon = 0$ in problem (6) reduces it to the stochastic programming approach while larger values of ϵ lead to more robustness (and conservativeness) in solutions.

Algorithm 1. *Input: A probability distribution $\mathcal{D}$, the number of samples R, the set of robustness levels $\mathcal{T}$.*

1: Sample R random variates from $\mathcal{D}$ for training, and obtain the empirical distribution q and the maximum KL divergence $\bar{\epsilon}(q)$ as computed in Proposition 3.

2: Solve problem (6) *with $\epsilon := \theta\bar{\epsilon}(q)$ for each $\theta \in \mathcal{T}$ to obtain a decision $y^*(\theta)$.*

3: Sample R random variates from $\mathcal{D}$ for testing, and then compute the cost realizations for each realization under the decision $y^*(\theta)$.

We implement Algorithm 1 in the Python programming language and use MOSEK 9.2 [29] to solve the dual exponential cone constrained MIP problem (6).

4.2.2. Results

For this illustration, we choose the following cost coefficients:

$$c = 1, \ c_b = 2, \ c_h = 1.$$

We will now specify the parameters of Algorithm 1. Firstly, we experiment with three different discrete distributions:

- Discrete Uniform Distribution with parameters 0 and 10, denoted as `Uniform(0, 10)`.
- Binomial Distribution with parameters 10 and 0.5, denoted as `Binomial(10, 0.5)`.
- Poisson Distribution with parameter 5, denoted as `Poisson(5)`.

We sample $R = 100$ random variates separately to obtain "training" and "test" datasets. Then, we repeat the experiments for the following "robustness" levels:

$$\mathcal{T} := \{0.00, 0.05, 0.10, 0.15, 0.20, 0.25\}.$$

The summary statistics of our experiments are reported in Tables 1-3 for Uniform, Binomial and Poisson distributions, respectively. In particular, we report the average and standard deviation of the cost realizations, abbreviated as "Avg." and "St. Dev.", respectively. In addition, we compute the average of the worst 10% of the realizations, abbreviated as "Worst 10%", to quantify the risk. We observe that as the robustness level θ increases, the optimal order quantity y^* increases (recall that $\theta = 0.00$ corresponds to the stochastic programming approach). Moreover, with increasing θ, the average cost increases while the standard deviation and the average of worst realizations decrease for each distribution. This is an expected behavior when robust optimization is utilized. We note that Binomial distribution is the least sensitive with respect to θ as the order quantity (and performance measures) do not change after $\theta \geq 0.05$. On the other hand, Uniform and Poisson distributions are more sensitive with respect to this parameter.

We also repeat the experiments with even higher values of θ and observe that only the results corresponding to the Poisson distribution changes,

which we attribute to its right-skewness. However, the order quantities in those cases are very high, which result in overly conservative policies and deteriorated performance measures.

Table 1. Summary results for the Newsvendor Problem with `Uniform(0, 10)` and $R = 100$.

θ	y^*	Avg.	St. Dev.	Worst 10%
0.00	4	8.08	2.92	13.80
0.05	4	8.08	2.92	13.80
0.10	5	8.67	2.22	12.80
0.15	5	8.67	2.22	12.80
0.20	5	8.67	2.22	12.80
0.25	6	9.53	1.94	12.00

Table 2. Summary results for the Newsvendor Problem with `Binomial(10, 0.5)` and $R = 100$.

θ	y^*	Avg.	St. Dev.	Worst 10%
0.00	4	6.76	2.38	11.40
0.05	5	7.02	1.67	10.40
0.10	5	7.02	1.67	10.40
0.15	5	7.02	1.67	10.40
0.20	5	7.02	1.67	10.40
0.25	5	7.02	1.67	10.40

Table 3. Summary results for the Newsvendor Problem with `Poisson(5)` and $R = 100$.

θ	y^*	Avg.	St. Dev.	Worst 10%
0.00	4	7.59	3.04	13.60
0.05	5	7.73	2.40	12.60
0.10	5	7.73	2.40	12.60
0.15	5	7.73	2.40	12.60
0.20	6	8.44	1.86	12.00
0.25	6	8.44	1.86	12.00

In addition to the summary statistics, we also provide the box plots of the cost realizations in Figures 1-3 for Uniform, Binomial and Poisson distributions, respectively. We observe that as the robustness level θ increases, the median of the cost realizations increases while the range shrinks for each distribution. We also note that the maximum and upper quartile values decrease for $\theta \in [0.05, 0.15]$. This is a desired property since it implies that the risk of the stochastic programming approach ($\theta = 0.00$) can be lowered.
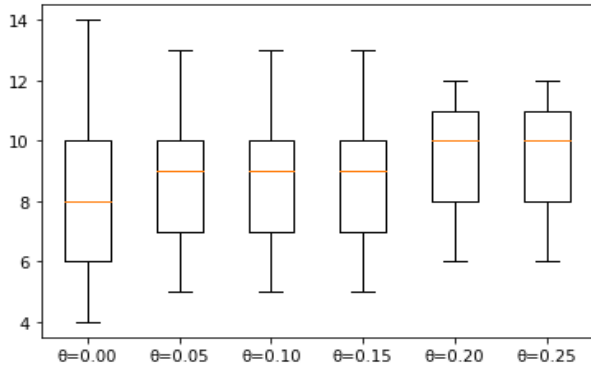
Figure 1. Box plot of the results for the Newsvendor Problem with `Uniform(0, 10)` and $R = 100$.
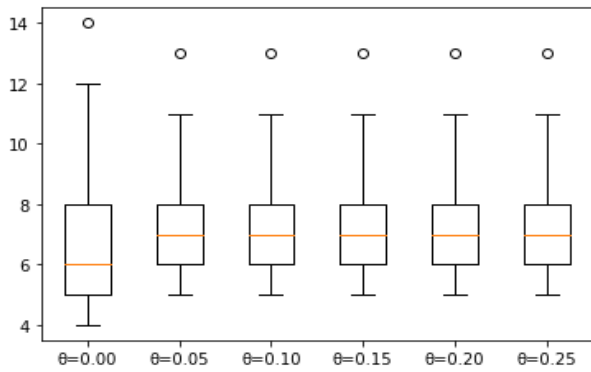

Figure 2. Box plot of the results for the Newsvendor Problem with `Binomial(10, 0.5)` and $R = 100$.
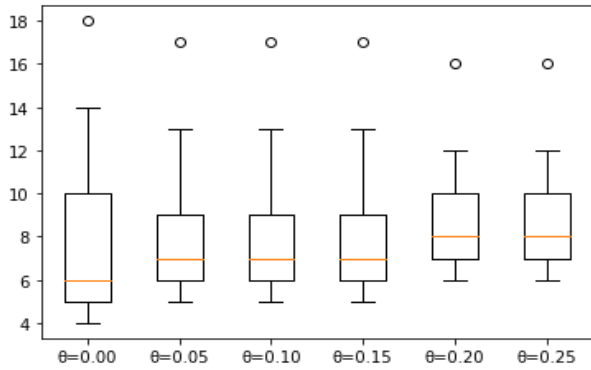

Figure 3. Box plot of the results for the Newsvendor Problem with `Poisson(5)` and $R = 100$.

As a final comparison, we formulate the DR version of the Newsvendor Problem assuming that the first and second moments are known (in our case, estimated from the training data) following [39]. We observe that the optimal order quantity y^* obtained from this approach turns out to be identical to the optimal order quantity obtained from the stochastic programming approach in our experimental setting.

5. Application to the uncapacitated facility location Problem

In this section, we analyze the KL divergence constrained DR version of the Uncapacitated Facility Location (UFL) Problem.

5.1. Deterministic version

We first remind the reader the deterministic version of the well-known UFL Problem. Suppose that we have m customers, each with demand d^i, $i \in [m]$. The demand must be satisfied by opening new facilities. There are n potential facilities, each with a fixed cost of f_j, $j \in [n]$. The unit transportation cost between each customer i and facility j is given as t_{ij}, $i \in [m]$, $j \in [n]$. The objective is to minimize the total fixed cost and transportation cost.

The UFL Problem can be modeled as an integer program by defining two sets of binary decision variables. The first set of decision variables, denoted as y_j, represent the status of each facility j, and the second set of decision variables, denoted as x_{ij}, represents the assignment of a customer i to a facility j. The complete model is given as follows:

$$\min \quad \sum_{j=1}^{n} \left[f_j y_j + \sum_{i=1}^{m} d^i t_{ij} x_{ij} \right] \tag{7a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = 1 \qquad i \in [m] \tag{7b}$$

$$x_{ij} \leq y_j \qquad i \in [m]; j \in [n] \tag{7c}$$

$$x_{ij} \in \{0, 1\} \qquad i \in [m]; j \in [n] \tag{7d}$$

$$y_j \in \{0, 1\} \qquad i \in [m]; j \in [n]. \tag{7e}$$

Here, constraint (7b) guarantees that each customer is served by exactly one facility while constraint (7c) ensures that each customer is served by an open facility.

We point out two useful observations about the UFL Problem. Firstly, in any feasible solution to problem (7), at least one facility must be opened. Therefore, we must have

$$\sum_{j=1}^{n} y_j \geq 1. \tag{8}$$

Secondly, given any optimal y^* vector, the optimal objective function value can be computed as

$$\sum_{j=1}^{n} f_j y_j^* + \sum_{i=1}^{m} d^i \min_{j:y_j^*=1} \{t_{ij}\}, \tag{9}$$

since each customer can be served by the closest open facility.

5.2. Distributionally robust version

Now, suppose that we replace the deterministic demand d^i with a random variable ξ^i having an empirical distribution $q^i \in \text{ri}(\Delta^{S_i})$, with realizations d^i_s, $s \in [S_i]$. Then, the DR version of the UFL Problem can be modeled as an instance of the generic model (1) with $\mathcal{Y} := \{y \in \{0,1\}^n : (8)\}$ as follows: We choose functions

$$h(y) := \sum_{j=1}^{n} f_j y_j$$

and

$$H^i(y, \xi^i) := \xi^i \min_{j:y_j=1} \{t_{ij}\}, \ i = 1, \dots, m,$$

due to (9). In the remainder of this subsection, we will obtain the robust counterpart of the KL divergence constrained DR UFL Problem as a dual exponential cone constrained MIP by the help of Lemma 1.

5.2.1. A lemma

The following lemma will be critical to linearize constraint (2c).

Lemma 1. *Let $t \in \mathbb{R}^n_+$ be a given vector and consider the function $g(y) : \{0,1\}^n \to \mathbb{R}$ defined as $g(y) := \min\{t_j : y_j = 1\}$. Then, for any $y \in \{0,1\}^n$ satisfying (8), we have*

$$g(y) = \max_{l=1,\dots,n} \left\{ t_l - \sum_{j=1}^{n} y_j \max\{t_l - t_j, 0\} \right\}. \quad (10)$$

Proof. Let $y \in \{0,1\}^n$ satisfying (8) be given. We define the following nonempty sets $T := \{j : y_j = 1\}$ and $T^* := \text{argmin}\{t_j : j \in T\}$. Notice that $g(y) = t_l$ for $l \in T^*$. Also, let us define the quantity

$$z_l := t_l - \sum_{j=1}^{n} y_j \max\{t_l - t_j, 0\}, \ l \in [n],$$

for convenience. Notice that we have

$$y_j \max\{t_l - t_j, 0\} =$$

$$\begin{cases} 0 & \text{if } j \notin T \\ 0 & \text{if } j \in T \text{ and } t_{il} \le t_{ij} \\ t_{il} - t_{ij} & \text{if } j \in T \text{ and } t_{il} > t_{ij} \end{cases}.$$

This observation helps us to rewrite z_l as

$$z_l = t_l - \sum_{j \in T : t_l > t_j} (t_l - t_j), \ l \in [n].$$

Now, we will look at the following cases to compute or bound z_l:

Case 1: Let $l^* \in T^*$. Then, we have $z_{l^*} = t_{l^*}$.

Case 2: Let $l \notin T^*$, and choose any $j^* \in T^*$. Then, we have

$$z_l = t_l - (t_l - t_{j^*}) - \sum_{j \in T \setminus \{j^*\} : t_{il} > t_{ij}} (t_l - t_j)$$

$$= t_{j^*} - \sum_{j \in T \setminus \{j^*\} : t_{il} > t_{ij}} (t_l - t_j)$$

$$\le t_{j^*}.$$

This analysis indicates that

$$\max_{l \in [n]} \left\{ t_l - \sum_{j=1}^{n} y_j \max\{t_l - t_j, 0\} \right\} = \max_{l \in [n]} z_l = t_{l^*},$$

where $l^* \in T^*$. Hence, we conclude that equation (10) holds true. $\square$

An alternative proof of Lemma 1 can be obtained via LP duality: First, one would write the problem $\min\{t_j : y_j = 1\}$ as an IP by introducing additional binary variables x_j. Secondly, this IP can be relaxed as an LP due to the totally unimodular structure. Then, the extreme points of the feasible region of the dual LP can be characterized, enabling the dual LP to be solved in closed form (see dual based arguments in [40, 41]).

5.2.2. The final formulation

Taking into account the special structure of the UFL Problem and utilizing Lemma 1 by setting $g := H^i$ for each $i \in [m]$, we obtain the following dual exponential cone constrained MIP:

$$\min \ \sum_{j=1}^{n} f_j y_j + \sum_{i=1}^{m} \left[\alpha^i + \epsilon^i \beta^i + \sum_{s=1}^{S_i} q^i_s u^i_s \right] \quad (11a)$$

$$\text{s.t. } \alpha^i - v^i_s \ge d^i_s \left(t_{il} - \sum_{j=1}^{n} y_j \max\{t_{il} - t_{ij}, 0\} \right)$$

$$i \in [m]; s \in [S_i]; l \in [n] \quad (11b)$$

$$(2c) - (2e), \ (7e), \ (8).$$

5.3. Computations

5.3.1. Experimental setup

We utilize Algorithm 2 to compare the effect of robustness level to KL divergence constrained DR version of the UFL Problem. This algorithm is quite similar to Algorithm 1 used for the analysis of the Newsvendor Problem.

Algorithm 2. **Input:** *A probability distribution $\mathcal{D}$, the number of samples R, the set of robustness levels $\mathcal{T}$.*

1: Sample R random variates from $\mathcal{D}$ for each customer $i \in [m]$ for training, and obtain the empirical distribution q^i and the maximum KL divergence $\bar{\epsilon}(q^i)$ for each $i \in [m]$.

2: *Solve problem (11) with $\epsilon^i := \theta\bar{\epsilon}(q^i)$ for each $i \in [m]$ and $\theta \in \mathcal{T}$ to obtain a decision vector $y^*(\theta)$.*

3: *Sample R random variates from $\mathcal{D}$ for each $i \in [m]$ for testing, and then compute the cost realizations for each realization under the decision vector $y^*(\theta)$.*

5.3.2. Results

For this illustration, we assume that there are 12 customers and three potential facilities located in the unit interval. Their precise locations are given as

$$\left\{ \frac{2\omega - 1}{36} : \omega \in [6] \right\} \cup \left\{ \frac{35 - 2\omega}{36} : \omega \in [6] \right\},$$

and

$$\left\{ \frac{2\Omega - 1}{6} : \Omega \in [3] \right\},$$

respectively, and are shown in Figure 4.

Figure 4. Locations of the customers (circles) and potential facilities (diamonds).

As evident from the figure, potential facilities are located evenly across the unit interval and there are two clusters of customers which are also distributed evenly in their respective regions. The fixed cost of opening facilities are given as $f_1 = f_3 = 10$ for the two facilities in the middle of these clusters (marked by a large diamond), and $f_2 = 5$ for the other facility (marked by a small diamond). Finally, the unit transportation cost between a facility-customer pair is assumed to be equal to the their distance from each other.

We specify the parameters of Algorithm 2 regarding the generation of random variates similar to that of Algorithm 1 as described in Section 4.2.2. The summary statistics of our experiments are reported in Tables 4-6 for Uniform, Binomial and Poisson distributions, respectively. We first observe that the optimal solutions and the performance measures are similar for every distribution, therefore, we will summarize our observations together. Due to the choice of parameters and the locations of the facilities and customers as can be seen from Figure 4, there is a fundamental trade-off in this instance: We can i) either open a single facility at the middle of the line segment with the lower fixed cost and serve customers via longer distances, or ii) open two facilities at the middle of two customer clusters with higher fixed cost and serve customers via shorter distances. In the

stochastic programming approach ($\theta = 0.00$), the first policy becomes optimal whereas in the DR approach ($\theta \geq 0.05$), the second policy becomes optimal. We note that considering the ambiguity of the demand distributions increases the average cost only slightly whereas both the standard deviation and the average of the worst 10% of the realizations decrease significantly. We remind the reader that the total fixed cost of the stochastic programming approach is only 5 while the fixed cost of the DR approach is 20. This also shows that the corresponding transportation cost, which is affected by the random uncertainty, is significantly smaller in the DR approach.

Table 4. Summary results for the UFL Problem with `Uniform(0, 10)` and $R = 100$.

θ	y^*	Avg.	St. Dev.	Worst 10%
0.00	0,1,0	23.11	3.47	29.19
0.05	1,0,1	24.52	0.92	26.10
0.10	1,0,1	24.52	0.92	26.10
0.15	1,0,1	24.52	0.92	26.10
0.20	1,0,1	24.52	0.92	26.10
0.25	1,0,1	24.52	0.92	26.10

Table 5. Summary results for the UFL Problem with `Binomial(10, 0.5)` and $R = 100$.

θ	y^*	Avg.	St. Dev.	Worst 10%
0.00	0,1,0	24.87	1.88	28.15
0.05	1,0,1	24.97	0.52	25.86
0.10	1,0,1	24.97	0.52	25.86
0.15	1,0,1	24.97	0.52	25.86
0.20	1,0,1	24.97	0.52	25.86
0.25	1,0,1	24.97	0.52	25.86

Table 6. Summary results for the UFL Problem with `Poisson(5)` and $R = 100$.

θ	y^*	Avg.	St. Dev.	Worst 10%
0.00	0,1,0	24.96	2.67	29.89
0.05	1,0,1	24.99	0.74	26.34
0.10	1,0,1	24.99	0.74	26.34
0.15	1,0,1	24.99	0.74	26.34
0.20	1,0,1	24.99	0.74	26.34
0.25	1,0,1	24.99	0.74	26.34

In addition to the summary statistics, we also provide the box plots of the cost realizations in Figures 5-7 for Uniform, Binomial and Poisson distributions, respectively. We observe that the median of the cost realizations either stays the same or increases slightly in the DR approach while the range shrinks significantly compared to the

stochastic programming approach for each distribution. We also note that the maximum and upper quartile values decrease considerably with the DR approach (especially for Binomial and Poisson distributions).
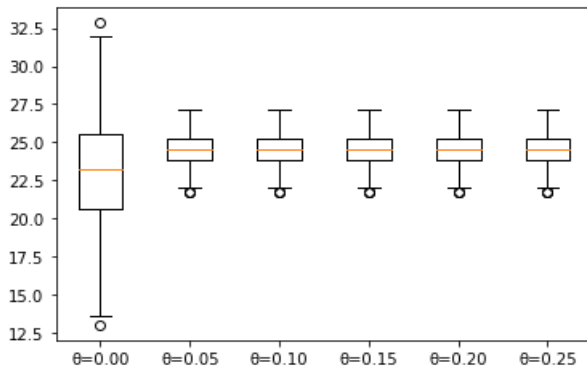

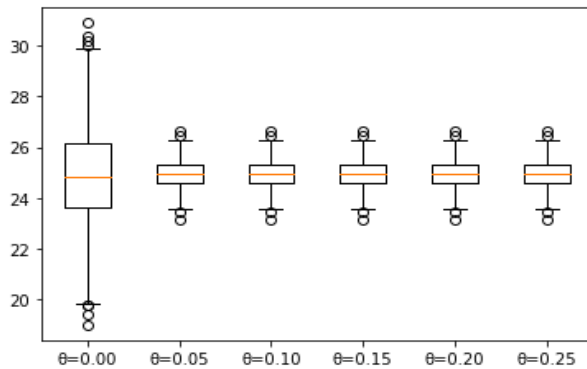**Figure 5.** Box plot of the results for the UFL Problem with `Uniform(0, 10)` and $R = 100$.

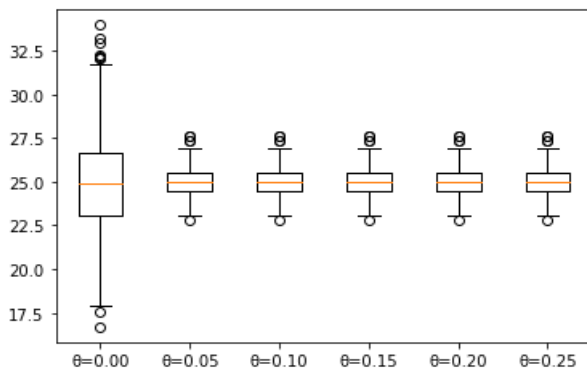**Figure 6.** Box plot of the results for the UFL Problem with `Binomial(10, 0.5)` and $R = 100$.

Figure 7. Box plot of the results for the UFL Problem with `Poisson(5)` and $R = 100$.

6. Conclusion

In this paper, we analyzed the KL divergence constrained DRO problems and proposed their dual exponential cone constrained reformulations utilizing the exponential cone representability property of KL divergence and Conic Duality. The resulting robust counterpart can be solved by a commercial conic programming solver directly. We specialized our results to the Newsvendor and UFL Problems by providing problem specific reformulations, and conducted a computational analysis comparing the performance of the solutions obtained via DR approach and stochastic programming from different aspects. We observed that although the mean and median of the cost realizations deteriorate slightly when the DR approach is preferred; the range, standard deviation and worst case values of the cost realizations improve significantly compared to stochastic programming approach.

Some future research directions seem promising. Firstly, by utilizing the semidefinite programming approximations of the matrix logarithm [42], we can try to formulate and solve KL-divergence constrained DRO problems that involve multivariate normal distributions. Secondly, we would like to test the success of the proposed method on different problems with real datasets. Lastly, we may adapt our results to the decision-dependent setting, which is a recent active research area in the DRO literature [35, 36, 43].

References

[1] Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

[2] Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2014). *Lectures on stochastic programming: modeling and theory*. Society for Industrial and Applied Mathematics.

[3] Ben-Tal, A., & Nemirovski, A. (2002). Robust optimization–methodology and applications. *Mathematical Programming*, 92(3), 453-480.

[4] Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.

[5] Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3), 464-501.

[6] Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1), 98–112.

[7] Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3), 595-612.

[8] Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6), 1358-1376.

[9] Gao, R. & Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *Optimization Online.*

[10] Mohajerin Esfahani, P. & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1), 115-166.

[11] Hanasusanto, G. A., & Kuhn, D. (2018). Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research*, 66(3), 849-869.

[12] Xie, W. (2019). On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, 1-41.

[13] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2), 341-357.

[14] Klabjan, D., Simchi-Levi, D., & Song, M. (2013). Robust stochastic lot-sizing by means of histograms. *Production and Operations Management*, 22(3), 691-710.

[15] Jiang, R., & Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming*, 158(1-2), 291-327.

[16] Yanıkoğlu, İ., den Hertog, D., & Kleijnen, J. P. (2016). Robust dual-response optimization. *IIE Transactions*, 48(3), 298-312.

[17] Lam, H. (2019). Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research*, 67(4), 1090-1105.

[18] Zymler, S., Kuhn, D., & Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2), 167-198.

[19] Yanıkoğlu, İ., & den Hertog, D. (2013). Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing*, 25(4), 666-681.

[20] Xie, W., & Ahmed, S. (2018). On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2), 1151-1182.

[21] Yanıkoğlu, İ. (2019). Robust reformulations of ambiguous chance constraints with discrete probability distributions. *An International Journal of Optimization and Control: Theories & Applications*, 9(2), 236-252.

[22] Nesterov, Y., & Nemirovski, A. (1994). *Interior-point polynomial algorithms in convex programming.* Society for Industrial and Applied Mathematics.

[23] Serrano, S. A. (2015). Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone. PhD Thesis. Stanford University.

[24] Dahl, J., & Andersen, E. D. (2019). A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization. *Optimization Online.*

[25] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1), 79-86.

[26] Hu, Z., & Hong, L. J. (2013). Kullback-Leibler divergence constrained distributionally robust optimization. *Optimization Online.*

[27] Chen, Y., Guo, Q., Sun, H., Li, Z., Wu, W., & Li, Z. (2018). A distributionally robust optimization model for unit commitment based on Kullback–Leibler divergence. *IEEE Transactions on Power Systems*, 33(5), 5147-5160.

[28] Li, Z., Wu, W., Zhang, B., & Tai, X. (2018). Kullback–Leibler divergence-based distributionally robust optimisation model for heat pump day-ahead operational schedule to improve PV integration. *IET Generation, Transmission & Distribution*, 12(13), 3136-3144.

[29] MOSEK ApS. (2020). MOSEK optimizer API for Python.

[30] Hanasusanto, G. A., Kuhn, D., Wallace, S. W., & Zymler, S. (2015). Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming*, 152(1-2), 1-32.

[31] Natarajan, K., Sim, M., & Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. *Management Science*, 64(7), 3146-3167.

[32] Lee, S., Kim, H., & Moon, I. (2020). A data-driven distributionally robust newsvendor model with a Wasserstein ambiguity set. *Journal of the Operational Research Society*, 1-19.

[33] Lu, M., Ran, L., & Shen, Z. J. M. (2015). Reliable facility location design under uncertain correlated disruptions. *Manufacturing & Service Operations Management*, 17(4), 445-455.

[34] Santiváñez, J. A., & Carlo, H. J. (2018). Reliable capacitated facility location problem with service levels. *EURO Journal on Transportation and Logistics*, 7(4), 315-341.

[35] Basciftci, B., Ahmed, S., & Shen, S. (2020). Distributionally robust facility location problem under decision-dependent stochastic demand. *European Journal of Operational Research.* doi:10.1016/j.ejor.2020.11.002.

[36] Noyan, N., Rudolf, G., & Lejeune, M. (2018). Distributionally robust optimization with decision-dependent ambiguity set. *Optimization Online*.

[37] MOSEK ApS. (2020). MOSEK modeling cookbook.

[38] Ben-Tal, A., & Nemirovski, A. (2001). *Lectures on modern convex optimization: analysis, algorithms, and engineering applications.* Society for Industrial and Applied Mathematics.

[39] Scarf, H. A. (1958). A min-max solution of an inventory problem. In: K. J. Arrow, S. Karlin and H. E. Scarf, Eds., *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, California, 201-209.

[40] Erlenkotter, D. (1978). A dual-based procedure for uncapacitated facility location. *Operations Research*, 26(6), 992-1009.

[41] Conn, A. R., & Cornuejols, G. (1990). A projection method for the uncapacitated facility location problem. *Mathematical Programming*, 46(1-3), 273-298.

[42] Fawzi, H., Saunderson, J., & Parrilo, P. A. (2019). Semidefinite approximations of the matrix logarithm. *Foundations of Computational Mathematics*, 19(2), 259-296.

[43] Luo, F., & Mehrotra, S. (2020). Distributionally robust optimization with decision dependent ambiguity sets. *Optimization Letters*, 14, 2565-2594.

Burak Kocuk *is an assistant professor at the Industrial Engineering Program, Sabancı University. He obtained his BS degrees in Industrial Engineering and Mathematics, and MS degree in Industrial Engineering from Boğaziçi University. He obtained his PhD degree of Operations Research at the School of Industrial and Systems Engineering, Georgia Institute of Technology. Before joining Sabancı University, he was a postdoctoral fellow at the Tepper School of Business, Carnegie Mellon University. His current research focuses on mixed-integer nonlinear programming and stochastic optimization problems, from both theoretical and methodological aspects.*
https://orcid.org/0000-0002-4218-1116