SIAM J. OPTIM.
Vol. 24, No. 3, pp. 1485–1506

DISTRIBUTIONALLY ROBUST STOCHASTIC KNAPSACK PROBLEM*

JIANQIANG CHENG†, ERICK DELAGE‡, AND ABDEL LISSER†

Abstract. This paper considers a distributionally robust version of a quadratic knapsack problem. In this model, a subsets of items is selected to maximizes the total profit while requiring that a set of knapsack constraints be satisfied with high probability. In contrast to the stochastic programming version of this problem, we assume that only part of the information on random data is known, i.e., the first and second moment of the random variables, their joint support, and possibly an independence assumption. As for the binary constraints, special interest is given to the corresponding semidefinite programming (SDP) relaxation. While in the case that the model only has a single knapsack constraint we present an SDP reformulation for this relaxation, the case of multiple knapsack constraints is more challenging. Instead, two tractable methods are presented for providing upper and lower bounds (with its associated conservative solution) on the SDP relaxation. An extensive computational study is given to illustrate the tightness of these bounds and the value of the proposed distributionally robust approach.

Key words. stochastic programming, chance constraints, distributionally robust optimization, semidefinite programming, knapsack problem

AMS subject classifications. 90C15, 90C22, 90C59

DOI. 10.1137/130915315

1. Introduction.

We consider the classical quadratic knapsack problem consisting of the decision to include or not each of a list of n items in a bag able to carry a certain maximum weight. The "multidimensional" version of this problem takes the form of the following optimization problem

$$(KP) \quad \underset{\mathbf{x}}{\text{maximize}} \quad \mathbf{x}^T\mathbf{R}\mathbf{x} \tag{1.1a}$$

$$\text{subject to} \quad \mathbf{w}_j^T\mathbf{x} \le d_j \ \ \forall\, j \in \{1,2,\dots,M\}, \tag{1.1b}$$

$$x_i \in \{0,1\} \ \forall\, i \in \{1,2,\dots,n\}, \tag{1.1c}$$

where $\mathbf{x}$ is a vector of binary values indicating whether each item is included in the knapsack, and $\mathbf{R} \in \Re^{n\times n}$ is a matrix whose (i,j)th term describes the linear contribution to reward of holding both items i and j. This allows us to model complementarity, when $R_{i,j} > 0$ or substitution, when $R_{i,j} < 0$, between items. Finally, for each j, $\mathbf{w}_j \in \Re^n$ is a vector of attributes (typically weights) whose total amount must satisfy some capacity constraint d_j. Constraints (1.1b) are typically called knapsack constraints and might serve to model the maximum weight or volume the bag is capable of carrying, or the fact that the total value of items that are carried must satisfy some given budget. There is a wide range of real life applications of the knapsack problem especially in the following topics: transportation, finance, e.g. the purchase of commodities or stocks with a limited budged, schedule planning [19]. The knapsack problem is also used as a subproblem in several combinatorial optimization problems,

*Received by the editors June 12, 2013; accepted for publication (in revised form) July 3, 2014; published electronically September 11, 2014.
http://www.siam.org/journals/siopt/24-3/91531.html
†LRI, Université Paris Sud, 91405, Orsay, France (cheng@lri.fr, lisser@lri.fr). The first author's research was supported by CSC (China Scholarship Council). The third author's research was supported by HADAMARD/PGMO grant.
‡HEC Montréal, H3T2A7, Montréal, Canada (erick.delage@hec.ca).

e.g., cutting problems, column generation, separation of cover inequalities, knapsack cryptosystems, and combinatorial auctions [18].

In practice, it is often the case that at the time of making the knapsack decision either the reward parameters or the weights parameter (or both) are not exactly known. In that case, one has the option to represent knowledge of these parameters through describing a measurable space of outcomes $(\Omega, \mathcal{F})$ and a probability measure F on this space. The knapsack problem thus becomes a stochastic problem where $\tilde{\mathbf{R}}$ and each $\tilde{\mathbf{w}}_j$ must be considered as a random matrix and a random vector, respectively. Specifically, $\tilde{\mathbf{R}} : \Omega \rightarrow \S_{\tilde{\mathbf{R}}}$ and $\tilde{\mathbf{w}}_j : \Omega \rightarrow \S_{\tilde{\mathbf{w}}}$, with $\S_{\tilde{\mathbf{R}}}$ and $\S_{\tilde{\mathbf{w}}}$ compact convex subsets of $\Re^{n\times n}$ and $\Re^n$, respectively.[1] In this context, it is natural to formulate the following stochastic program:

$$
\begin{aligned}
&\text{(1.2a)} \qquad (SKP) \quad \underset{\mathbf{x}}{\text{maximize}} \quad && \mathbb{E}_F[u(\mathbf{x}^T\tilde{\mathbf{R}}\mathbf{x})] \\
&\text{(1.2b)} \qquad \text{subject to} && \mathbb{P}_F(\tilde{\mathbf{w}}_j^T\mathbf{x} \le d_j \ \forall\, j \in \{1,\ldots,M\}) \ge 1-\eta, \\
&\text{(1.2c)} && x_i \in \{0,1\} \ \forall\, i \in \{1,2,\ldots,n\}
\end{aligned}
$$

for some concave utility function $u(\cdot)$ that captures risk aversion with respect to the total achieved reward. In this formulation, the knapsack constraints are required to be jointly met with probability larger than $1-\eta$ for some small η.

Under many circumstances, the assumption of full knowledge of the distribution F fails. For this reason, it can be necessary to consider that the only knowledge we have of a distribution is that it is part of some uncertainty set $\mathcal{D}$. Following a robust approach, in this context we might be interested in choosing items for our knapsack so that the value of the knapsack, as measured by the stochastic program, has the best worst-case guarantees under the choice of a distribution in this uncertainty set. Hence, our interest lies in solving:

$$
\begin{aligned}
&\text{(1.3a)} \qquad (DRSKP) \quad \underset{\mathbf{x}}{\text{maximize}} \quad && \inf_{F\in\mathcal{D}} \mathbb{E}_F[u(\mathbf{x}^T\tilde{\mathbf{R}}\mathbf{x})] \\
&\text{(1.3b)} \qquad \text{subject to} && \inf_{F\in\mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T\mathbf{x} \le d_j \ \forall\, j \in \{1,2,\ldots,M\}) \ge 1-\eta, \\
&\text{(1.3c)} && x_i \in \{0,1\} \ \forall\, i \in \{1,2,\ldots,n\}.
\end{aligned}
$$

In this paper, we consider a distributionally robust stochastic knapsack problem, where we assume that only part of the information on random data is known. For the knapsack problem with a single knapsack constraint, when information is limited to mean, covariance, and support of the uncertain data, we show that the (DRSKP) problem reduces to a semidefinite program (SDP) after applying the SDP based relaxation scheme to the binary constraints. To the best of our knowledge, this is the first time that an SDP representation is derived for distributionally robust chance constraints where both first and second moments and support are known. Unfortunately, this reformulation does not extend to multidimensional knapsack problems where $M > 1$. Therefore, for the more general form, after relaxing the binary constraints with some common linear matrix inequality, we study the case here information is limited to the distribution's first and second moments and some form of independence and propose efficient approximations that provide both a lower bound and an

[1] Note that d is assumed certain without loss of generality since one can always capture such uncertainty by including an additional decision x_0 that absolutely needs to be included in the knapsack and a set of random weights $(\tilde{\mathbf{w}}_j)_0$.

upper bound for the model. As a side product, a conservative solution to the relaxed problem is obtained. An extensive computational study is provided to illustrate the tightness of these bounds and the value of the proposed distributionally robust approach.

The rest of the paper is organized as follows. In section 2, we review recent work on the topic of stochastic and robust knapsack problems. In section 3, we describe in more details the distributionally robust stochastic knapsack problem (DRSKP) with single knapsack constraint and introduce its SDP relaxation. Later, in section 4, we propose two approximation methods for the DRSKP with multiple knapsack constraints, also referred to as a multidimensional knapsack problem. Finally, in section 5, we report our numerical experiments and section 6 summarizes our results.

Remark 1.1. Note that one can show that solving (DRSKP) is equivalent to solving

$$\underset{\mathbf{x}\in\{0,1\}^n}{\text{maximize}} \inf_{F\in\mathcal{D}} \text{SKP}(\mathbf{x}, F), \tag{1.4}$$

where $\text{SKP}(\mathbf{x}, F)$ refers to the objective function of this problem that is augmented with feasibility verification, *i.e.*,

$$\text{SKP}(\mathbf{x}; F) = \begin{cases} \mathbb{E}_F[u(\mathbf{x}^T\tilde{\mathbf{R}}x)] & \text{if } \mathbb{P}_F(\tilde{\mathbf{w}}_j^T x \le d_j \ \forall j \in \{1, 2, \ldots, M\}) \ge 1 - \eta, \\ \quad -\infty & \text{otherwise.} \end{cases}.$$

This is simply due to the fact that in order for $\inf_{F\in\mathcal{D}} \text{SKP}(\mathbf{x}, F)$ to be finite valued, $\mathbf{x}$ must satisfy constraint (1.3b).

2. Prior work. The binary quadratic knapsack problem (QKP) was first introduced by Gallo et al. [12]. Since then, there has been an extensive literature on QKP and its various aspects [3, 4, 6, 24]. Gallo, Hammer, and Simeone [12] invented QKP and gave a family of upper bounds based on upper planes. Martello and Toth [24] gave a pseudo-polynomial-time algorithm. Johnson, Mehrotra, and Nemhauser [16] considered a graph version of the QKP and solved it by a branch-and-cut system. Billionnet and Calmels [3] used a classical linearization technique of the objective function to obtain an integer linear program formulation. Due to different uncertainty factors, the parameters are not known in advance. Then, it is natural to model the problem as a stochastic problem. For more information about the stochastic knapsack problem, we refer the reader to [11, 19, 20, 23, 33] and references therein. The knapsack problem has also been studied under the lens of robust optimization. In [17], it was established that the problem is strongly NP-hard even when the uncertainty is limited to the objective function and the uncertainty set is composed of a finite number of scenarios. More recently, in [21] the authors considered a min-max knapsack problem that was specifically designed to solve the chance-constrained problem and proposed a solution method whose complexity again grew exponentially in the size of the scenario set. In an attempt to circumvent the computational challenge, Bertsimas and Sim [1] identified an uncertainty set, referred to as being "budgeted," that allowed a mixed integer linear programming reformulation for robust combinatorial problems like the knapsack, thus amenable to be solved by powerful available algorithms such as in CPLEX [15]. In [25], the authors provided tight bounds on the price of robustness of knapsack problems, i.e. relative cost increase for using a solution that was feasible for all weight perturbations. To the best of our knowledge, there is no prior work studying the application of distributionally robust optimization to the stochastic quadratic knapsack (or multidimensional knapsack) problem.

The concept of distributionally robust optimization was initially introduced in 1958 by Scarf [27] who named the method minimax stochastic programming. While the original application was a single item newsvendor problem, the range of applications of this method was rather limited due to numerical difficulties encountered when solving large instances. Recently, mainly due to the development of efficient interior point algorithms for solving SDPs and perhaps also because of the new "distributionally robust" branding, the method has become very popular. In [9], the authors showed that a large range of distributionally robust optimization models, where the distribution's mean and covariance matrix were constrained, can be solved in polynomial time, and were often equivalent to an SDP reformulation. In [5], the framework was applied to address chance constraints which were typically considered as an intractable construct when the distribution was assumed to be known. Both Chen et al. [7] and Zymler, Kuhn, and Rustem [34], extended this work to provide approximations when chance constraints were joint (i.e. constraints must be satisfied jointly with high probability). Recently, the work in [32] has shed even more light on the range of functions and uncertainty sets that can be addressed using SDP.

Unfortunately, to this day, the application of distributionally robust optimization to integer programs has been rather scarce. In [2], the authors sought bounds for the worst-case objective function value of stochastic integer programs when only moments were known about the distribution. In [22], the authors considered a scheduling problem where the integer decision variables were the number of employees assigned to different schedules. They considered uncertainty in scenario probabilities and used polyhedral uncertainty sets to preserve the mixed integer linear programming structure of the nominal problem. Most closely related to our work is the work by Wagner in [30] who studied a stochastic 0-1 linear program with chance constraints where distribution information was limited to moments up to order k. He presented how to address this problem using a hierarchy of SDPs with integer constraints. Comparatively, we address a stochastic 0-1 quadratic program and consider the joint chance constraint in the case of a multidimensional knapsack.

3. Distributionally robust knapsack problem. There are two difficulties that need to be addressed when searching for the optimal solution of the DRSKP problem. First, the worst-case analysis that is needed to evaluate the objective and verify the chance constraint involves optimization over an infinite dimensional decision space. Second, the problem is intrinsically a combinatorial one. In this section, we provide an approximation for the distributionally robust knapsack problem with a single knapsack constraint, i.e., $M = 1$. This approximation will be based on an exact reformulation of problem (1.3) as a finite dimensional SDP problem with binary variables. We will then relax the binary constraints using linear matrix inequalities.

3.1. Finite dimensional reformulation. We start our study by making a small set of assumptions. These will be needed to ensure that the approximation model we obtain can be solved efficiently.

DEFINITION 3.1. *Let ξ be a random vector in $\Re^m$ on which $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{w}}_1$ (referred to henceforth as $\tilde{\mathbf{w}}$) depend linearly:*

$$\tilde{\mathbf{R}} = \sum_{i=1}^{m} \mathbf{A}_i^{\tilde{\mathbf{R}}} \xi_i, \qquad\qquad \tilde{\mathbf{w}} = \mathbf{A}^{\tilde{\mathbf{w}}} \xi.$$

Assumption 3.2. The utility function $u(\cdot)$ is piecewise linear, increasing, and concave. In other words, it can be represented in the form

$$u(y) = \min_{k \in \{1,2,\ldots,K\}} a_k y + b_k,$$

where $\mathbf{a} \in \Re^K$ and $\mathbf{a} \geq 0$.

Assumption 3.3. The distributional uncertainty set accounts for information about the convex support $\S$, mean μ in the strict interior of $\S$, and an upper bound $\Sigma \succ 0$ on the covariance matrix of the random vector ξ:

$$\mathcal{D}(\S, \mu, \Sigma) = \left\{ F \left| \begin{array}{l} \mathbb{P}(\xi \in \S) = 1 \\ \mathbb{E}_F[\xi] = \mu \\ \mathbb{E}_F[(\xi - \mu)(\xi - \mu)^T] \preceq \Sigma \end{array} \right. \right\}.$$

Definition 3.1 is made without loss of generality since we can always consider ξ to contain all the terms of $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{w}}$. On the other hand, regarding Assumption 3.2, although piecewise-linear utility functions are not the common functions to use when applying expected utility theory, they provide enough flexibility to approximate to a high level of accuracy any function that one might be interested in using. Finally, Assumption 3.3 implies that one only knows the support of $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{w}}$ (or a set containing this support), the mean of these random terms, and some information about variance and covariance. We refer the reader to [9] for a full discussion about this choice of distributional uncertainty, in particular for the use of a linear matrix inequality to capture information about the covariance matrix.

THEOREM 3.4. *Under Assumptions* 3.2 *and* 3.3*, and given that* $\S$ *is compact and* $M = 1$*, then the following deterministic problem*

(3.1a)
$$\underset{\mathbf{x}, t, \mathbf{q}, \mathbf{Q}, \bar{t}, \bar{\mathbf{q}}, \bar{\mathbf{Q}}, s}{\text{maximize}} \quad t - \mu^T \mathbf{q} - (\Sigma + \mu\mu^T) \bullet \mathbf{Q}$$

(3.1b)
$$\text{subject to} \quad t \leq \sum_{j=1}^{m} a_k \bar{\xi}_j \mathbf{x}^T \mathbf{A}_j^{\tilde{\mathbf{R}}} \mathbf{x} + b_k + \bar{\xi}^T \mathbf{q} + \bar{\xi}^T \mathbf{Q} \bar{\xi} \;\forall \bar{\xi} \in \S \;\forall k = \{1, \ldots, K\},$$

(3.1c) $$\bar{t} + 2\mu^T \bar{\mathbf{q}} + (\Sigma + \mu\mu^T) \bullet \bar{\mathbf{Q}} \leq \eta s,$$

(3.1d) $$\bar{t} + 2\bar{\xi}^T \bar{\mathbf{q}} + \bar{\xi}^T \bar{\mathbf{Q}} \bar{\xi} \geq 0 \;\forall \bar{\xi} \in \S,$$

(3.1e) $$\bar{t} + 2\bar{\xi}^T \bar{\mathbf{q}} + \bar{\xi}^T \bar{\mathbf{Q}} \bar{\xi} - s + 2d - 2\bar{\xi}^T \mathbf{A}^{\tilde{\mathbf{w}}^T} \mathbf{x} \geq 0 \;\forall \bar{\xi} \in \S,$$

(3.1f) $$\mathbf{Q} \succeq 0, \quad \bar{\mathbf{Q}} \succeq 0, \quad s \geq 0,$$

(3.1g) $$x_i \in \{0, 1\} \;\forall i \in \{1, 2, \ldots, n\},$$

where t, $\bar{t} \in \Re$, $\mathbf{q}$, $\bar{\mathbf{q}} \in \Re^m$, $\mathbf{Q}$, $\bar{\mathbf{Q}} \in \Re^{m \times m}$, $s \in \Re$, *and* $\bullet$ *is the Frobenius inner product defined by* $A \bullet B = \sum_{i,j} A_{ij} B_{ij}$, *where* A *and* B *are two conformal matrices, identifies a solution* $\hat{\mathbf{x}}^*$ *that is guaranteed to be feasible for problem* (1.3) *and to perform at least as well in terms of the objective function as an optimal solution of problem* (1.3) *with* d_1 *replaced by* $d_1 - \epsilon$ *for any* $\epsilon > 0$.

Remark 3.5. In practice, one might actually consider that the solution of problem (3.1) is the solution that one is looking for given that d_j is a continuous parameter and could have easily been slightly overestimated (and similarly $\mathbb{E}[\tilde{\mathbf{w}}]$ might have been underestimated).

Proof. The proof relies here on applying the theory presented in [9] to convert the distributionally robust objective into its deterministic equivalent. This gives rise to the reformulated objective function in terms of auxiliary variables t, $\mathbf{q}$, and $\mathbf{Q}$, and constraint (3.1b). In a second step, one can adapt the ideas used to prove Theorem 1 of [10] to an uncertainty set that account for the first and second moments and the support constraints presented in Assumption 3.3. The key idea here will be to replace constraint (1.3b) with

$$\inf_{F\in(\S,\mu,\Sigma)} \mathbb{P}_F(\xi^T \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} < d) \geq 1-\eta,$$

which is both stricter than the original one and more relaxed than

$$\inf_{F\in(\S,\mu,\Sigma)} \mathbb{P}_F(\xi^T \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} \leq d-\epsilon) \geq 1-\eta$$

for any $\epsilon > 0$. This new constraint is equivalent to verifying that the optimal value of the following infinite dimensional problem is not greater than η:

$$\sup_{F\in(\S,\mu,\Sigma)} \mathbb{E}_F[\mathbb{1}\{\xi^T \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} \geq d\}],$$

where $\mathbb{1}\{\cdot\}$ is the indicator function that gives one if the statement is verified and zero otherwise. By duality theory (see [9]), this problem is equivalent to

$$\begin{aligned}
\underset{\bar{t},\bar{\mathbf{q}},\bar{\mathbf{Q}}}{\text{minimize}} \quad & \bar{t} + 2\mu^T\bar{\mathbf{q}} + (\Sigma+\mu\mu^T)\bullet\bar{\mathbf{Q}}\\
\text{subject to} \quad & -\mathbb{1}\{\bar{\xi}^T \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} \geq d\} + \bar{t} + 2\bar{\xi}^T\bar{\mathbf{q}} + \bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} \geq 0 \;\forall\, \bar{\xi}\in\S,\\
& \bar{\mathbf{Q}} \succeq 0,
\end{aligned}$$

where $\bar{t}\in\Re$, $\bar{\mathbf{q}}\in\Re^m$, and $\bar{\mathbf{Q}}\in\Re^{m\times m}$. One can show that duality is strict since the conditions laid out in Assumption 3.3 ensure that the Dirac measure δ_μ lies in the relative interior of the distributional set, hence the weaker version of Proposition 3.4 in [28] applies.

When $M=1$, we get that constraint (1.3b) can be reformulated as

$$\bar{t} + 2\mu^T\bar{\mathbf{q}} + (\Sigma+\mu\mu^T)\bullet\bar{\mathbf{Q}} \leq \eta, \tag{3.2a}$$

$$\bar{t} + 2\bar{\xi}^T\bar{\mathbf{q}} + \bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} \geq 0 \;\forall\, \bar{\xi}\in\S, \tag{3.2b}$$

$$-1 + \bar{t} + 2\bar{\xi}^T\bar{\mathbf{q}} + \bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} \geq 0 \;\forall\, \bar{\xi}\in\S\cap\{\bar{\xi}\in\Re^m | \bar{\xi}^T\mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} \geq d\}. \tag{3.2c}$$

Studying more closely the third constraint, one realizes that it can be expressed as

$$\min_{\bar{\xi}\in\S}\;\sup_{\lambda\geq 0}\; -1 + \bar{t} + 2\bar{\xi}^T\bar{\mathbf{q}} + \bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} + 2\lambda(d - \bar{\xi}^T\mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x}) \geq 0,$$

which by Sion's minimax theorem, since $\S$ is compact and $\bar{\mathbf{Q}}$ is positive semidefinite, is equivalent to

$$\sup_{\lambda\geq 0}\;\min_{\bar{\xi}\in\S} -1 + \bar{t} + 2\bar{\xi}^T\bar{\mathbf{q}} + \bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} + 2\lambda(d - \bar{\xi}^T\mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x}) \geq 0.$$

Again since $\S$ is compact, this constraint can be replaced by

$$\sup_{\lambda>0} \min_{\bar{\xi}\in\S} -1 + \bar{t} + 2\bar{\xi}^T\bar{\mathbf{q}} + \bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} + 2\lambda(d - \bar{\xi}^T\mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x}) \geq 0,$$

which can be reformulated as

$$\sup_{\lambda>0} \min_{\bar{\xi}\in\S} -1/\lambda + (1/\lambda)\bar{t} + 2(1/\lambda)\bar{\xi}^T\bar{\mathbf{q}} + (1/\lambda)\bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} + 2(d - \bar{\xi}^T\mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x}) \geq 0.$$

The three constraints can therefore be restated as

$$\begin{aligned} (1/\lambda)\bar{t} + 2(1/\lambda)\mu^T\bar{\mathbf{q}} + (\Sigma + \mu\mu^T)\bullet(1/\lambda)\bar{\mathbf{Q}} &\leq \eta(1/\lambda), \\ (1/\lambda)\bar{t} + 2(1/\lambda)\bar{\xi}^T\bar{\mathbf{q}} + (1/\lambda)\bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} &\geq 0\ \forall\,\bar{\xi}\in\S, \\ -(1/\lambda) + (1/\lambda)\bar{t} + 2(1/\lambda)\bar{\xi}^T\bar{\mathbf{q}} + (1/\lambda)\bar{\xi}^T\bar{\mathbf{Q}}\bar{\xi} + 2(d - \bar{\xi}^T\mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x}) &\geq 0\ \forall\,\bar{\xi}\in\S. \end{aligned}$$

A simple replacement of variables $s := (1/\lambda)$, $\bar{t}' := (1/\lambda)\bar{t}$, $\bar{\mathbf{q}}' := (1/\lambda)\bar{\mathbf{q}}$, and $\bar{\mathbf{Q}}' := (1/\lambda)\bar{\mathbf{Q}}$ leads to constraints (3.1c), (3.1d), (3.1e). ∎

It is important to realize that while many ideas behind this result are drawn from the existing litterature on distributionally robust optimization, to the best of our knowledge the proposed reformulation for a distributionally robust chance constraint that accounts for both first and second moment and support information appears for the first time and should be of interest in other fields of application. The key in obtaining this result is in exploiting the structure of $\mathbb{E}_F[(\xi-\mu)(\xi-\mu)^T] \preceq \Sigma$, as was done in [9], which ensures that $\mathbf{Q}$ and $\bar{\mathbf{Q}}$ are positive-semidefinite matrices. This property enables the use of conic duality to reformulate constraint (3.2c) instead of using an approximate version of the S-lemma.

While problem (3.1) still takes the form of an infinite dimensional problem, if the support set $\S$ is representable by a set of linear matrix inequalities, then one can easily obtain the robust counterparts of constraints (3.1b), (3.1d), and (3.1e). Next, we present three natural examples.

COROLLARY 3.6. *Given that the support of F is ellipsoidal, $\S = \{\xi | (\xi-\xi_0)^T\Theta(\xi-\xi_0) \leq 1\}$, problem (1.4) further reduces to the following problem*

$$\text{(3.3a)}\quad \underset{\mathbf{x},t,\mathbf{q},\mathbf{Q},\mathbf{v},\mathbf{s},\bar{t},\bar{\mathbf{q}},\bar{\mathbf{Q}},\bar{\mathbf{s}}}{\text{maximize}} \quad t - \mu^T\mathbf{q} - (\Sigma+\mu\mu^T)\bullet\mathbf{Q}$$

$$\text{(3.3b)}\quad \text{subject to} \quad \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q}+a_k\mathbf{v}}{2} \\ \frac{\mathbf{q}^T+a_k\mathbf{v}^T}{2} & b_k - t \end{bmatrix} \succeq -s_k \begin{bmatrix} \Theta & -\Theta\xi_0 \\ -\xi_0^T\Theta & \xi_0^T\Theta\xi_0 - 1 \end{bmatrix} \forall\, k,$$

$$\text{(3.3c)}\quad v_j = \mathbf{A}_j^{\tilde{\mathbf{R}}}\bullet(\mathbf{x}\mathbf{x}^T)\,\forall\, j\in\{1,2,\ldots,m\},$$

$$\text{(3.3d)}\quad \bar{t} + 2\mu^T\bar{\mathbf{q}} + (\Sigma+\mu\mu^T)\bullet\bar{\mathbf{Q}} \leq \eta\bar{s}_1,$$

$$\text{(3.3e)}\quad \begin{bmatrix} \bar{\mathbf{Q}} & \bar{\mathbf{q}} \\ \bar{\mathbf{q}}^T & \bar{t} \end{bmatrix} \succeq -\bar{s}_2 \begin{bmatrix} \Theta & -\Theta\xi_0 \\ -\xi_0^T\Theta & \xi_0^T\Theta\xi_0 - 1 \end{bmatrix},$$

$$\text{(3.3f)}\quad \begin{bmatrix} \bar{\mathbf{Q}} & \bar{\mathbf{q}} - \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} \\ (\bar{\mathbf{q}} - \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x})^T & \bar{t} + 2d - \bar{s}_1 \end{bmatrix} \succeq -\bar{s}_3 \begin{bmatrix} \Theta & -\Theta\xi_0 \\ -\xi_0^T\Theta & \xi_0^T\Theta\xi_0 - 1 \end{bmatrix},$$

$$\text{(3.3g)}\quad \mathbf{Q}\succeq 0,\ \ \bar{\mathbf{Q}}\succeq 0,\ \ \mathbf{s}\geq 0,\ \ \bar{\mathbf{s}}\geq 0,$$

$$\text{(3.3h)}\quad x_i\in\{0,1\}\,\forall i\in\{1,2,\ldots,n\}.$$

Alternatively, if the support of F is polyhedral, i.e., $\S = \{\xi | \mathbf{C}\xi \leq \mathbf{c}\}$ *with* $\mathbf{C} \in \Re^{p\times m}$ *and* $\mathbf{c} \in \Re^p$*, then problem* (1.4) *reduces to*

$$\begin{aligned}
\underset{\substack{\mathbf{x},t,\mathbf{q},\mathbf{Q},\mathbf{v},s,\bar{t},\bar{\mathbf{q}},\bar{\mathbf{Q}}\\ \lambda_1,\lambda_2,\ldots,\lambda_K+2}}{\text{maximize}} \quad & t - \mu^T\mathbf{q} - (\Sigma + \mu\mu^T) \bullet \mathbf{Q} \\
\text{subject to} \quad & \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q}+a_k\mathbf{v}+\mathbf{C}^T\lambda_k}{2} \\ \frac{\mathbf{q}^T+a_k\mathbf{v}^T+\lambda_k^T\mathbf{C}}{2} & b_k - t - \mathbf{c}^T\lambda_k \end{bmatrix} \succeq 0 \,\forall\, k \\
& v_j = \mathbf{A}_j^{\tilde{\mathbf{R}}} \bullet (\mathbf{x}\mathbf{x}^T) \,\forall\, j \in \{1,2,\ldots,m\}, \\
& \bar{t} + 2\mu^T\bar{\mathbf{q}} + (\Sigma + \mu\mu^T) \bullet \bar{\mathbf{Q}} \leq \eta s, \\
& \begin{bmatrix} \bar{\mathbf{Q}} & \bar{\mathbf{q}} + \frac{1}{2}\mathbf{C}^T\lambda_{K+1} \\ \bar{\mathbf{q}}^T + \frac{1}{2}\lambda_{K+1}^T\mathbf{C} & \bar{t} - \mathbf{c}^T\lambda_{K+1} \end{bmatrix} \succeq 0, \\
& \begin{bmatrix} \bar{\mathbf{Q}} & \bar{\mathbf{q}} - \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} + \frac{1}{2}\mathbf{C}^T\lambda_{K+2} \\ (\bar{\mathbf{q}} - \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x})^T + \frac{1}{2}\lambda_{K+2}^T\mathbf{C} & \bar{t} + 2d - s - \mathbf{c}^T\lambda_{K+2} \end{bmatrix} \succeq 0, \\
& \mathbf{Q} \succeq 0, \;\; \bar{\mathbf{Q}} \succeq 0, \;\; s \geq 0, \;\; \lambda_k \geq 0 \;\forall k \in \{1,2,\ldots,K+2\}, \\
& x_i \in \{0,1\} \,\forall i \in \{1,2,\ldots,n\},
\end{aligned}$$

where $\lambda_k \in \Re^p$ *for all* k *are the dual variables associated with the linear inequalities* $\mathbf{C}\xi \leq \mathbf{c}$ *for each infinite set of constraints. Finally, if the support of* F *is unbounded (i.e.* $\S = \Re^m$*), then problem* (1.4) *reduces to*

$$\underset{\mathbf{x},t,\mathbf{q},\mathbf{Q},\mathbf{v},\mathbf{z},\tau}{\text{maximize}} \quad t - \mu^T\mathbf{q} - (\Sigma + \mu\mu^T) \bullet \mathbf{Q} \tag{3.4a}$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q}+a_k\mathbf{v}}{2} \\ \frac{\mathbf{q}^T+a_k\mathbf{v}^T}{2} & b_k - t \end{bmatrix} \succeq 0 \,\forall\, k = \{1,2,\ldots,K\}, \tag{3.4b}$$

$$v_j = \mathbf{A}_j^{\tilde{\mathbf{R}}} \bullet (\mathbf{x}\mathbf{x}^T) \,\forall\, j \in \{1,2,\ldots,m\}, \tag{3.4c}$$

$$\mathbf{Q} \succeq 0, \quad \tau_k \geq 0 \,\forall\, k = \{1,2,\ldots,K\}, \tag{3.4d}$$

$$\begin{bmatrix} \mathbf{0}_{m,m} & \Sigma^{1/2}\mathbf{z} \\ \mathbf{z}^T\Sigma^{1/2} & 0 \end{bmatrix} \succeq \sqrt{\frac{\eta}{1-\eta}}(\mu^T\mathbf{z} - d)\mathbf{I}, \tag{3.4e}$$

$$\mathbf{z} = \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x}, \tag{3.4f}$$

$$x_i \in \{0,1\} \,\forall i \in \{1,2,\ldots,n\}. \tag{3.4g}$$

Proof. In the case of the spherical support, the proof simply relies on applying the S-Lemma (see [26] for details) on each infinite set of constraints (3.1b), (3.1d), and (3.1e). Otherwise, with the polyhedral set, since $\mathbf{Q}$ and $\bar{\mathbf{Q}}$ are positive semidefinite, one can easily apply duality theory to construct the so-called robust counterpart of each constraint that is indexed by $\xi \in \S$. Finally, although the support set is not compact when $\S = \Re^m$, duality theory can still be applied and the reductions needed for the worst-case expected cost and worst-case chance constraint have been presented in [9] and [10]. ∎

Example 3.7. For clarity, we present an example in which problem (3.1) returns a conservative solution for problem (1.3) that we argue might actually be the solution that a practitioner should implement. Consider a knapsack problem with a single item with weight in the range $[0,2]$ with an expected value estimated to be 1 and standard deviation of 1. The question is whether to put the item in the knapsack or not given that we wish to have at least 90% confidence that the total weight will be less than or equal to 2, i.e. the maximum weight the knapsack can hold. It is clear here that the optimal solution of problem (1.3) is to wrap the item in the knapsack thus giving full confidence that the total weight will be less than or equal to 2. Yet, since problem (3.1) is derived after changing the maximum total weight constraint to a strict inequality in the chance constraint, it will consider that the optimal solution is to leave the item out of the knapsack since in the worst case there is a 50% chance that the weight is exactly 2. This approximate solution is feasible in terms of problem (1.3) and performs at least as well as the optimal solution of all versions of the problem where the maximum total weight that is considered is strictly smaller than two. Since the worst-case distribution for the solution that wraps the item up has 50% probability that the knapsack will carry exactly its estimated maximum load, in practice the solution of leaving the item aside is potentially the better one considering that it is better protected against estimation error.

3.2. Semidefinite programming approximation. In order to obtain an approximation model with known polynomial convergence rate, we apply an SDP based relaxation scheme to the binary constraints. This scheme makes use of a lifting through the decision matrix $\mathbf{X}$:

$$\mathbf{X} = \mathbf{x}\mathbf{x}^T.$$

As what is commonly done for combinatorial problems, we start by imposing the redundant constraints

$$\mathbf{X}_{i,j} \geq 0 \,\forall\, i,j \quad \& \quad \mathbf{X}_{i,i} = x_i \,\forall\, i,$$

since $x_i x_j \geq 0$ and $x_i^2 = x_i$ when x_i is either zero or one.

We then relax the nonconvex constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^T$ to $\mathbf{X} \succeq \mathbf{x}\mathbf{x}^T$ which can take the form of a linear matrix inequality.

$$\begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix} \succeq 0.$$

When the support set is ellipsoidal, we have the following SDP problem:

(3.5a)

$$\underset{\mathbf{x},t,\mathbf{q},\mathbf{Q},\mathbf{v},\mathbf{s},\bar{t},\bar{\mathbf{q}},\bar{\mathbf{Q}},\bar{\mathbf{s}}}{\text{maximize}} \quad t - \mu^T\mathbf{q} - (\Sigma + \mu\mu^T) \bullet \mathbf{Q}$$

$$\text{(3.5b)} \quad \text{subject to} \quad \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q}+a_k\mathbf{v}}{2} \\ \frac{\mathbf{q}^T+a_k\mathbf{v}^T}{2} & b_k - t \end{bmatrix} \succeq -s_k \begin{bmatrix} \Theta & -\Theta\xi_0 \\ -\xi_0^T\Theta & \xi_0^T\Theta\xi_0 - 1 \end{bmatrix} \,\forall\, k,$$

$$\text{(3.5c)} \quad v_j = \bar{\mathbf{A}}_j^{\tilde{\mathbf{R}}} \bullet \mathbf{X} \,\forall\, j \in \{1,2,\ldots,m\},$$

$$\text{(3.5d)} \quad \bar{t} + 2\mu^T\bar{\mathbf{q}} + (\Sigma + \mu\mu^T) \bullet \bar{\mathbf{Q}} \leq \eta\bar{s}_1,$$

$$\text{(3.5e)} \quad \begin{bmatrix} \bar{\mathbf{Q}} & \bar{\mathbf{q}} \\ \bar{\mathbf{q}}^T & \bar{t} \end{bmatrix} \succeq -\bar{s}_2 \begin{bmatrix} \Theta & -\Theta\xi_0 \\ -\xi_0^T\Theta & \xi_0^T\Theta\xi_0 - 1 \end{bmatrix},$$

$$\text{(3.5f)} \quad \begin{bmatrix} \bar{\mathbf{Q}} & \bar{\mathbf{q}} - \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x} \\ (\bar{\mathbf{q}} - \mathbf{A}^{\tilde{\mathbf{w}}^T}\mathbf{x})^T & \bar{t} + 2d - \bar{s}_1 \end{bmatrix} \succeq -\bar{s}_3 \begin{bmatrix} \Theta & -\Theta\xi_0 \\ -\xi_0^T\Theta & \xi_0^T\Theta\xi_0 - 1 \end{bmatrix},$$

$$\mathbf{Q} \succeq 0, \quad \bar{\mathbf{Q}} \succeq 0, \quad \mathbf{s} \geq 0, \quad \bar{\mathbf{s}} \geq 0, \tag{3.5g}$$

$$\begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix} \succeq 0, \tag{3.5h}$$

$$\mathbf{X}_{i,i} = x_i \,\forall\, i, \quad \mathbf{X}_{i,j} \geq 0 \,\forall\, i, j, \tag{3.5i}$$

where we removed constraint $0 \leq \mathbf{x} \leq 1$ given that it is already imposed through constraints (3.5h) and (3.5i).

4. Multidimensional knapsack problem.

Here we consider a distributionally robust approach to stochastic multidimensional knapsack problems. To make this section self-contained, we recall that the DRSKP is formulated as

$$\underset{\mathbf{x}}{\text{maximize}} \quad \inf_{F \in \mathcal{D}} \mathbb{E}_F[u(\mathbf{x}^T \tilde{\mathbf{R}} \mathbf{x})] \tag{4.1a}$$

$$\text{subject to} \quad \inf_{F \in \mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \ldots, M) \geq 1 - \eta, \tag{4.1b}$$

$$x_i \in \{0, 1\} \,\forall\, i \in \{1, 2, \ldots, n\} \tag{4.1c}$$

for some piecewise linear concave increasing utility function $u(\cdot)$, where F now describes the joint distribution of all types of "weights" of all items, $\{\tilde{\mathbf{w}}_j\}_{j=1}^M \sim F$, and $\mathcal{D}$ now describes a set of such joint distributions.

Definition 4.1. *Without loss of generality, for all $j = 1, \ldots, M$, let ξ_j be a random vector in $\Re^m$ on which the $\tilde{\mathbf{w}}_j$ depend linearly and let $\tilde{\mathbf{R}}$ depend linearly on $\{\xi_j\}_{j=1}^M$:*

$$\tilde{\mathbf{R}} = \sum_{j=1}^{M} \sum_{i=1}^{m} \mathbf{A}_{ji}^{\tilde{\mathbf{R}}} (\xi_j)_i, \qquad \tilde{\mathbf{w}}_j = \mathbf{A}_j^{\tilde{\mathbf{w}}_j} \xi_j, j = 1, \ldots, M.$$

Assumption 4.2. The distributional uncertainty set accounts for information about the mean μ_j, and an upper bound Σ_j on the covariance matrix of the random vector ξ_j, for each $j = 1, \ldots, M$:

$$\mathcal{D}(\mu_j, \Sigma_j) = \left\{ F_j \left| \begin{array}{l} \mathbb{E}_{F_j}[\xi_j] = \mu_j \\ \mathbb{E}_{F_j}[(\xi_j - \mu_j)(\xi_j - \mu_j)^T] \preceq \Sigma_j \end{array} \right. \right\}.$$

Furthermore, the random vectors ξ_i and ξ_j are independent when $i \neq j$. Note that the support of F_j is unbounded, i.e. $\S = \Re^m$.

Theorem 4.3. *Under Assumptions* 3.2 *and* 4.2*, the following problem is a conservative approximation*[2] *of problem* (4.1)*:*

$$\underset{\mathbf{x}, t, \mathbf{q}, \mathbf{Q}, \mathbf{v}}{\text{maximize}} \quad t - \mu^T \mathbf{q} - (\Sigma + \mu \mu^T) \bullet \mathbf{Q} \tag{4.2a}$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q} + a_k \mathbf{v}}{2} \\ \frac{\mathbf{q}^T + a_k \mathbf{v}^T}{2} & b_k - t \end{bmatrix} \succeq 0 \,\forall\, k = \{1, 2, \ldots, K\}, \tag{4.2b}$$

[2] The concept of conservative approximation denotes the fact that the solution to the approximate model is guaranteed to achieve an expected utility that is at least as large as the obtained approximate optimal objective value.

$$v_{(j-1)M+i} = \mathbf{A}_{ji}^{\tilde{\mathbf{R}}} \bullet (\mathbf{x}\mathbf{x}^T)\, \forall j \in \{1,2,\ldots,M\}\ \forall i \in \{1,2,\ldots,m\}, \tag{4.2c}$$

$$\mathbf{Q} \succeq 0, \tag{4.2d}$$

$$\inf_{F\in\mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \le d_j, j=1,\ldots,M) \ge 1-\eta, \tag{4.2e}$$

$$x_i \in \{0,1\}\ \forall i \in \{1,2,\ldots,n\}, \tag{4.2f}$$

where $\mathbf{q}$, $\mathbf{v} \in \Re^{mM}$, *and* $\mathbf{Q} \in \Re^{mM\times mM}$, *and with*

$$\mu := \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{bmatrix} \qquad \textit{and} \qquad \Sigma = \begin{bmatrix} \Sigma_1 & \mathbf{0}_{m,m} & \mathbf{0}_{m,m} & \ldots & \mathbf{0}_{m,m} \\ \mathbf{0}_{m,m} & \Sigma_2 & \mathbf{0}_{m,m} & \ldots & \mathbf{0}_{m,m} \\ \mathbf{0}_{m,m} & \mathbf{0}_{m,m} & \Sigma_3 & \ldots & \mathbf{0}_{m,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m,m} & \mathbf{0}_{m,m} & \mathbf{0}_{m,m} & \ldots & \Sigma_M \end{bmatrix}.$$

Furthermore, this approximation is exact when $u(\cdot)$ *is linear, i.e. the attitude is risk neutral.*

Proof. The proof follows lines as presented in section 3 given that the assumption of independence implies that $\mathbb{E}_F[(\xi_i-\mu_i)(\xi_j-\mu_j)^T] = 0$, hence that $\mathbb{E}[(\xi-\mu)(\xi-\mu)^T] \preceq \Sigma$. Here however, we relax the constraint that F satisfies the independence assumption per se when addressing the objective function since this cannot be handled through the theory of infinite dimensional linear programming. Problem (4.2) is therefore a conservative approximation of problem (4.1). This approximation is necessarily tight when $u(\cdot)$ is linear since then the objective function (4.1a) reduces to

$$\sum_{j=1}^{M}\sum_{i=1}^{m} \mathbf{A}_{ji}^{\tilde{\mathbf{R}}}(\mu_j)_i. \qquad \square$$

Correspondingly, we have an SDP relaxation for this conservative approximation:

$$\underset{\mathbf{x},t,\mathbf{q},\mathbf{Q},\mathbf{v},\mathbf{X}}{\text{maximize}} \quad t - \mu^T\mathbf{q} - (\Sigma + \mu\mu^T)\bullet \mathbf{Q} \tag{4.3a}$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q}+a_k\mathbf{v}}{2} \\ \frac{\mathbf{q}^T+a_k\mathbf{v}^T}{2} & b_k - t \end{bmatrix} \succeq 0\ \forall k = \{1,2,\ldots,K\}, \tag{4.3b}$$

$$v_{(j-1)*M+i} = \mathbf{A}_{ji}^{\tilde{\mathbf{R}}} \bullet \mathbf{X}\ \forall j \in \{1,2,\ldots,M\}\ \forall i \in \{1,2,\ldots,m\}, \tag{4.3c}$$

$$(DRSKP-SDP) \quad \mathbf{Q} \succeq 0, \tag{4.3d}$$

$$\inf_{F\in\mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \le d_j, j=1,\ldots,M) \ge 1-\eta, \tag{4.3e}$$

$$\begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix} \succeq 0, \tag{4.3f}$$

$$\mathbf{X}_{i,i} = x_i\ \forall i, \quad \mathbf{X}_{i,j} \ge 0\ \forall i,j. \tag{4.3g}$$

Apart from the constraint (4.3e), (DRSKP-SDP) is an SDP problem. Thus, in this section, we investigate the distributionally robust joint chance constraint. First of all, we review two lower bounding (also referred to as "conservative") approximations that have recently been proposed in the literature. We then introduce two novel approximations for lower bounding and upper bounding the value of (DRSKP-SDP) which will exploit the structure of the joint chance constraint under our independence assumption.

4.1. Bonferroni's conservative bound. A popular approximation for joint chance-constrained problems is based on Bonferroni's inequality, which decomposes the joint constraint into M individual constraints. When $\sum_{j=1}^{M} \eta_j = \eta$ for any F, we have

$$\mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \le d_j) \ge 1 - \eta_j, j = 1, \ldots, M \Rightarrow \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \le d_j, j = 1, \ldots, M) \ge 1 - \eta.$$

One can already obtain from this fact a trivial conservative approximation for the (DRSKP-SDP) problem.

THEOREM 4.4. *The (DRSKP-SDP) problem with constraint* (4.3e) *replaced with*

$$\begin{bmatrix} \mathbf{0}_{m,m} & \Sigma_j^{1/2}\mathbf{z}_j \\ \mathbf{z}_j^T \Sigma_j^{1/2} & 0 \end{bmatrix} \succeq \sqrt{\frac{\eta_j}{1-\eta_j}} (\mu_j^T \mathbf{z}_j - d_j)\mathbf{I} \; \forall\, j = 1, \ldots, M,$$
$$\mathbf{z}_j = \mathbf{A}_j^{\tilde{\mathbf{w}}_j T} \mathbf{x} \; \forall\, j = 1, \ldots, M,$$

where $\mathbf{z}_j \in \Re^m$ are additional auxiliary decision variables, is an SDP problem. The optimal solution of this SDP is feasible according to the original (DRSKP-SDP) and its optimal value provides a lower bound on the value of the original problem.

4.2. Zymler, Kuhn, and Rustem's conservative bound. In [34], the authors address the distributionally robust chance constraint by introducing a scaling parameter $\alpha \in \mathcal{A} = \{\alpha \in \Re^M : \alpha > 0\}$ and reformulating it as a distributionally robust value-at-risk constraint. For any $\alpha \in \mathcal{A}$

$$\begin{aligned} &\inf_{F \in \mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \le d_j, j = 1, \ldots, M) \ge 1 - \eta \\ &\Leftrightarrow \inf_{F \in \mathcal{D}} \mathbb{P}_F\Big(\max_j \{\alpha_j(\tilde{\mathbf{w}}_j^T \mathbf{x} - d_j)\} \le 0\Big) \ge 1 - \eta \\ &\Leftrightarrow \inf_{F \in \mathcal{D}} \text{VaR-}\eta_F\Big(\max_j \{\alpha_j(\tilde{\mathbf{w}}_j^T \mathbf{x} - d_j)\}\Big) \le 0. \end{aligned}$$

The next step involves replacing the value-at-risk operator by an upper bounding conditional value-at-risk operation to make it both convex and tractable. For any $\alpha > 0$, this leads to the following conservative approximation of (DRSKP-SDP). We refer the reader to [34] for more details.

THEOREM 4.5. *Consider the (DRSKP-SDP) problem with constraint* (4.3e) *replaced by*

$$\beta + \frac{1}{\eta} \begin{bmatrix} \Sigma + \mu\mu^T & \frac{1}{2}\mu \\ \frac{1}{2}\mu^T & 1 \end{bmatrix} \bullet \mathcal{M} \le 0,$$
$$\mathcal{M} - \begin{bmatrix} \mathbf{0}_{mM,mM} & \frac{1}{2}\alpha_j \mathbf{y}_j \\ \frac{1}{2}\alpha_j \mathbf{y}_j^T & -\alpha_j d_j - \beta \end{bmatrix} \succeq 0 \; \forall j = 1, \ldots, M,$$
$$\mathbf{y}_j = \begin{bmatrix} \mathbf{0}_{(j-1)m}{}^T & \mathbf{x}^T \mathbf{A_j}^{\tilde{\mathbf{w}}_j} & \mathbf{0}_{(M-j)m}{}^T \end{bmatrix}^T,$$
$$\mathcal{M} \succeq 0,$$

where $\mathbf{0}_k$ and $\mathbf{0}_{k,k}$ are, respectively, the vector in $\Re^k$ and matrix in $\Re^{k\times k}$ containing all zeros, $\beta \in \Re$, $\mathcal{M} \in \Re^{(Mm+1)\times(Mm+1)}$, and $y_j \in \mathbb{R}^{mM}$ are auxiliary decision variables, which is an SDP problem for any $\alpha > 0$. The optimal solution of this SDP is feasible according to the original (DRSKP-SDP) and its optimal value provides a lower bound on the value of the original problem.

4.3. New bounds under independence assumption. Under Assumptions 3.2 and 4.2, one can actually show that the robust chance constraint (4.1b) is equivalent to

$$\prod_{j=1}^{M} \inf_{F\in\mathcal{D}(\Re^m,\mu_j,\Sigma_j)} \mathbb{P}_F((\mathbf{A}_j^{\tilde{\mathbf{w}}_j}\xi)^T\mathbf{x} \le d_j) \ge 1-\eta.$$

This constraint is therefore satisfied if and only if there exists a vector $\mathbf{y} \in \Re^M$ such that $\mathbf{y} \ge 0$, $\sum_{j=1}^{M} y_j = 1$ and

$$\inf_{F\in\mathcal{D}(\Re^m,\mu_j,\Sigma_j)} \mathbb{P}_F((\mathbf{A}_j^{\tilde{\mathbf{w}}_j}\xi)^T\mathbf{x} \le d_j) \ge (1-\eta)^{y_j}\ \forall j \in \{1,2,\dots,M\}.$$

Thus, based on [5], problem (DRSKP-SDP) is equivalent to the following deterministic problem:

$$\text{(4.4a)}\quad \underset{\mathbf{x},t,\mathbf{q},\mathbf{Q},\mathbf{v},\mathbf{X},\mathbf{y}}{\text{maximize}} \quad t-\mu^T\mathbf{q}-(\Sigma+\mu\mu^T)\bullet\mathbf{Q}$$

$$\text{(4.4b)}\quad \text{subject to} \quad \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q}+a_k\mathbf{v}}{2} \\ \frac{\mathbf{q}^T+a_k\mathbf{v}^T}{2} & b_k-t \end{bmatrix} \succeq 0\ \forall k=\{1,2,\dots,K\},$$

$$\text{(4.4c)}\quad v_{(j-1)*M+i} = \mathbf{A}_{ji}^{\tilde{\mathbf{R}}}\bullet\mathbf{X}\ \forall j\in\{1,2,\dots,M\}\forall i\in\{1,2,\dots,m\},$$

$$\text{(4.4d)}\quad \mathbf{Q}\succeq 0,$$

$$\text{(4.4e)}\quad \mu_j^T\mathbf{A}_j^{\tilde{\mathbf{w}}_j}\mathbf{x}+\sqrt{\frac{p^{y_j}}{1-p^{y_j}}}\left\|\Sigma_j^{1/2}\mathbf{A_j}^{\tilde{\mathbf{w}}_j\,T}\mathbf{x}\right\|_2 \le d_j,$$

$$\text{(4.4f)}\quad \sum_{j=1}^{M} y_j = 1, \quad y_j \ge 0,$$

$$\text{(4.4g)}\quad \begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix} \succeq 0,$$

$$\text{(4.4h)}\quad \mathbf{X}_{i,i} = x_i\ \forall i, \quad \mathbf{X}_{i,j} \ge 0\ \forall i,j,$$

where $p = 1-\eta$.

Remark 4.6. For the chance constraint (4.4e), if we take y_j as fixed parameters and transform them into the equivalent SDP constraint, the dimension of the corresponding linear matrix inequality is $(m+1)\times(m+1)$, compared to an $(Mm+1)\times(Mm+1)$ linear matrix inequality in the bound based on Zymler, Kuhn, and Rustem [34].

4.3.1. Conservative bound through sequential approximations. When we consider the variables y_j, $j = 1,\dots M$, to be fixed, the problem (DRSKP-SDP) becomes an SDP problem. Thus we propose a sequential approximation method that iteratively adjusts the parameters y and solves the obtained SDP until no further improvement is achieved.

Algorithm 1. Sequential Approximation Procedure.

- Initialization: Let $\mathbf{y}^1 \in R_+^M$ be scaling parameters, i.e., $\sum_{j=1}^M \mathbf{y}^1(j) = 1$, $\mathbf{y}^1 \geq 0$. Set the iteration counter to $t := 1$.
- Update: Solve problem (4.4) with $\mathbf{y}$ fixed to $\mathbf{y}^t$ and let $\mathbf{x}^t$ and f^t denote an optimal solution and the optimal value, respectively. Let $z_j = \frac{(d_j - \mu^T \mathbf{A}_j^{\tilde{\mathbf{w}}_j} \mathbf{x}^t)^2}{(\|\Sigma_j^{1/2} \mathbf{A_j}^{\tilde{\mathbf{w}}_j T} \mathbf{x}^t\|_2)^2}$ and $\tilde{y}_j^t = \log_p \frac{z_j}{z_j + 1}$. Set $\mathbf{y}^{t+1} \leftarrow U(\mathbf{y}; \tilde{\mathbf{y}})$ for some
$$U : \Delta^M \times [0,1]^M \to \Delta^M,$$
where $\Delta^M = \{\mathbf{y} \in \Re^M \,|\, \mathbf{y} \geq 0 \;\&\; \sum_j y_j = 1\}$ is the probability simplex, and "$\times$" is the Cartesian product.
- Stopping criterion: if $f^t - f^{t-1}$ is small enough, stop and return $\mathbf{x}^t$, f^t, and $\mathbf{y}^t$, otherwise, set $t := t+1$ and go to the "Update" step.

Theorem 4.7. *If the problem (DRSKP-SDP) is bounded and has a feasible solution for the initial values of $\mathbf{y}^1$, and given that $U(\mathbf{y}; \tilde{\mathbf{y}}) \geq \tilde{\mathbf{y}}$ for all $\mathbf{y} \geq \tilde{\mathbf{y}}$, then Algorithm 1 terminates in a finite number of steps and the returned value f^t is a lower bound for (DRSKP-SDP).*

Proof. We first show that the sequence of values f^t produced by the algorithm is nondecreasing. To do so, we simply need to verify that the solution $\mathbf{x}^t$ obtained at iteration t remains feasible in problem (4.4) when changing the fixed variables $\mathbf{y}$ from $\mathbf{y}^t$ to $\mathbf{y}^{t+1}$. Indeed, we have that the function $\sqrt{\frac{p^{y_j}}{1-p^{y_j}}}$ is decreasing on the interval $]0,1]$ and that $\tilde{y}_j^t$ captures the minimum amount that y_j should be to allow $\mathbf{x}^t$ to remain feasible with respect to the jth chance constraint. Since $\mathbf{x}^t$ is feasible, it must be that $\mathbf{y}^t \geq \tilde{\mathbf{y}}^t$, hence joint feasibility for all chance constraints is ensured by the condition that the update policy should then suggest a vector larger than $\tilde{\mathbf{y}}$.

We then confirm that every value of f^t is a lower bound on the optimal value of problem (4.4) since by the definition of the update policy each term of the sequence of $\mathbf{y}^t$ is feasible according to this problem. Hence, given that the sequence is nondecreasing, and that it is bounded by the optimal value of problem (4.4) which is itself bounded, this implies that the sequence $\{f^t\}$ will converge to a finite limit f^∞. We are therefore guaranteed that for any fixed tolerance level ϵ, after a finite number of steps T, the difference $f_{T+1} - f_T \leq f^\infty - f_T \leq \epsilon$, hence the algorithm should have terminated. ∎

In our implementation, we choose the following update policy:
$$U(\mathbf{y}; \tilde{\mathbf{y}}) = \tilde{\mathbf{y}} + \alpha(1 - (\mathbf{y} - \tilde{\mathbf{y}})),$$
where $\alpha = \frac{\sum_j (y_j - \tilde{y}_j)}{\sum_j (1 - (\mathbf{y}_j - \tilde{y}_j))}$. The reasoning behind this adjustment policy is that, under the current solution $\mathbf{x}^t$, it encourages giving more margin to the chance constraints that are the tightest at $\mathbf{x}^t$, effectively using $y_j^t - \tilde{y}_j^t$ as a measure of tightness. One can easily verify that it is indeed the case that this adjustment policy satisfies the required properties. First, if $\mathbf{y} \geq \tilde{\mathbf{y}}$ then one can easily show that both α and $1 - (\mathbf{y} - \tilde{\mathbf{y}})$ are nonnegative and thus we have that $U(\mathbf{y}; \tilde{\mathbf{y}}) \geq \tilde{\mathbf{y}} \geq 0$. Then, we can verify that
$$\mathbf{1}^T U(\mathbf{y}; \tilde{\mathbf{y}}) = \sum_j \tilde{y}_j + \alpha \sum_j (1 - (y_j - \tilde{y}_j)) = \sum_j \tilde{y}_j + \sum_j (y_j - \tilde{y}_j) = 1.$$
Since the vector generated through $U(\mathbf{y}; \tilde{\mathbf{y}}) \geq \tilde{\mathbf{y}}$ is nonnegative and sums to one, it necessarily lies on the probability simplex.

4.3.2. Upper bound through linearization. In order to estimate how conservative any of the above approximations are compared to the true solution of the (DRSKP-SDP), there is a need for a procedure that might provide an upper bound for the optimal value of the problem. In order to do so, we exploit the fact that the nonlinearity in constraint (4.4e) is due to the product of two special convex functions.

LEMMA 4.8. *Function* $f(y) = \sqrt{\frac{p^y}{1-p^y}}$ *is convex and decreasing for* $y \in]0,1]$ *when* $p \in]0,1[$. *Hence, given any set of values* $\{y_l\}_{l=1}^L$ *with each* $y_l \in]0,1]$, *then we have*

$$f(y) \geq \max_{l \in \{1,2,\ldots,L\}} \hat{a}_l y + \hat{b}_l, \quad y \in]0,1],$$

where

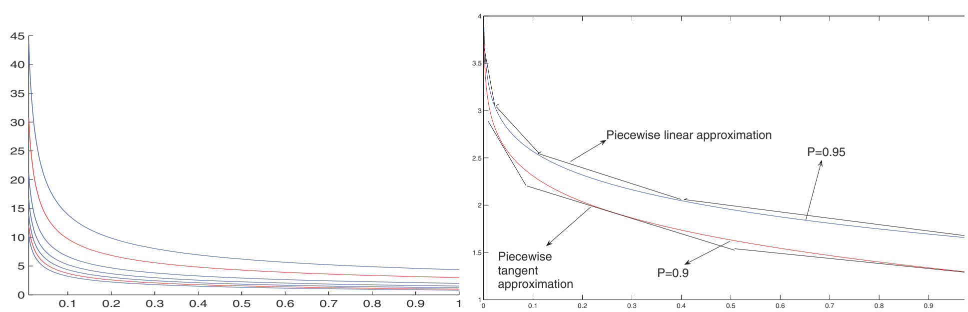
$$\hat{a}_l = \left. \frac{\partial \left(\sqrt{\frac{p^y}{1-p^y}}\right)}{\partial y} \right|_{y=y_l} \quad \textit{and} \quad \hat{b}_l = \sqrt{\frac{p^{y_l}}{1-p^{y_l}}} - \hat{a}_l y_l.$$

Proof. Since $(\sqrt{\frac{p^y}{1-p^y}})' = \frac{\ln p \sqrt{p^y}}{2(1-p^y)^{\frac{3}{2}}} \leq 0$, when $y \in]0,1], p \in]0,1[$, we have that $\sqrt{\frac{p^y}{1-p^y}}$ is decreasing. Furthermore, since $(\sqrt{\frac{p^y}{1-p^y}})'' = \frac{(\ln p)^2}{4} \frac{\sqrt{p^y}(1+2p^y)}{(1-p^y)^{\frac{5}{2}}} \geq 0$, when $y \in]0,1]$, $p \in]0,1[$, we confirm that $f(y)$ is convex. Finally, given any $y_l \in]0,1]$, convexity ensures that for all $y \in]0,1]$,

$$\begin{aligned} f(y) \geq f(y_l) + (y - y_l) f'(y_l) &\Rightarrow f(y) \geq f'(y_l) y + f(y_l) - y_l f'(y_l) \\ &\Rightarrow f(y) \geq \hat{a}_l y + \hat{b}_l. \end{aligned}$$

This concludes our proof. □

Figure 1 provides an illustration of this function for different values of p. It also illustrates how this function can be lower bounded by a piecewise linear convex function which is tangent to the actual function at a finite number of points for any value of $p \in]0,1]$. To get an upper bound for the value of our problem, our scheme



(a) (b)

FIG. 1. *Function* $f(y) = \sqrt{\frac{p^y}{1-p^y}}$ *is convex and decreasing in* y *for all* $p \in]0,1[$. (a) $f(y)$ *for* $p = 0.5, 0.55, 0.6, 0.7, 0.8, 0.9, 0.95$; (b) *shows how* $f(y)$ *can be lower bounded by a piecewise linear convex function that is tangent to* $f(y)$ *at a number of points.*

approximates the function $\sqrt{\frac{p^{y_j}}{1-p^{y_j}}}$ with a piecewise linear convex lower bounding function

$$\hat{f}(y) = \max_{l\in\{1,2,\ldots,L\}} \hat{a}_l y + \hat{b}_l \approx \sqrt{\frac{p^y}{1-p^y}}$$

and linearizes some quadratic terms that emerge after a series of algebraic manipulations. We obtain the following upper bounding SDP:

(4.5a)
$$\underset{\mathbf{x},t,\mathbf{q},\mathbf{Q},\mathbf{v},\mathbf{y},\mathbf{z},\bar{\mathbf{x}}}{\text{maximize}} \quad t - \mu^T\mathbf{q} - (\Sigma + \mu\mu^T)\bullet \mathbf{Q}$$

(4.5b)
$$\text{subject to} \quad \begin{bmatrix} \mathbf{Q} & \frac{\mathbf{q}+a_k\mathbf{v}}{2} \\ \frac{\mathbf{q}^T+a_k\mathbf{v}^T}{2} & b_k - t \end{bmatrix} \succeq 0 \;\forall\, k = \{1,2,\ldots,K\},$$

(4.5c) $$v_{(j-1)*M+i} = \mathbf{A}_{ji}^{\tilde{\mathbf{R}}} \bullet X \;\forall\, j\in\{1,2,\ldots,M\}\;\forall\, i \in\{1,2,\ldots,n\},$$

(4.5d) $$\mathbf{Q}\succeq 0,$$

(4.5e) $$\begin{bmatrix} \mathbf{0}_{m,m} & \Sigma_j^{1/2}\mathbf{A_j}^{\tilde{\mathbf{w}}_j T}\mathbf{z}_j \\ (\Sigma_j^{1/2}\mathbf{A_j}^{\tilde{\mathbf{w}}_j T}\mathbf{z}_j)^T & 0 \end{bmatrix}$$

(4.5f) $$\succeq (\mu_j^T\mathbf{A_j}^{\tilde{\mathbf{w}}_j T}\mathbf{x} - d_j t_j)\mathbf{I}\;\forall\, j\in\{1,2,\ldots,M\},$$

(4.5g) $$\mathbf{z}_j \geq \hat{a}_l\bar{\mathbf{x}}_j + \hat{b}_l\mathbf{x}\;\;\forall\, l\in\{1,2,\ldots,L\}\;\forall\, j\in\{1,2,\ldots,M\},$$

(4.5h) $$\sum_{j=1}^M \bar{\mathbf{x}}_j = \mathbf{x},$$

(4.5i) $$0\leq\bar{\mathbf{x}}_j\leq\mathbf{x},\quad (y_j-1)\mathbf{e}_n + \mathbf{x}\leq\bar{\mathbf{x}}_j\leq y_j\mathbf{e}_n\forall\, j\in\{1,2,\ldots,M\},$$

(4.5j) $$\sum_{j=1}^M y_j = 1,\quad y_j\geq 0\;\forall\, j\in\{1,2,\ldots,M\},$$

(4.5k) $$\begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix}\succeq 0,$$

(4.5l) $$\mathbf{X}_{i,i} = x_i\;\forall\, i\quad \mathbf{X}_{i,j}\geq 0\;\forall\, i,j,$$

where $\mathbf{z}_j\in\Re^n$, $\bar{\mathbf{x}}_j\in\Re^n$, and $\mathbf{e}_n$ is the vector in $\Re^n$ containing all ones.

THEOREM 4.9. *The optimal value of* (4.5) *is an upper bound for the optimal value of (DRSKP-SDP).*

Proof. The main idea of the proof relies on applying the theory presented in [8]. We focus on the constraint (4.4e). We can first show that

$$\left\{\mathbf{x}\in\Re^n \;\middle|\; \mu^T\mathbf{A}_j^{\tilde{\mathbf{w}}_j}\mathbf{x} + \sqrt{\frac{p^{y_j}}{1-p^{y_j}}}\left\|\Sigma_j^{1/2}\mathbf{A_j}^{\tilde{\mathbf{w}}_j T}\mathbf{x}\right\|_2 \leq d_j, j=1,\ldots,M\right\}$$

is equivalent to

$$\left\{x : \mu^T\mathbf{A}_j^{\tilde{\mathbf{w}}_j}\mathbf{x} + \left\|\Sigma_j^{1/2}\mathbf{A_j}^{\tilde{\mathbf{w}}_j T}\left(\sqrt{\frac{p^{y_j}}{1-p^{y_j}}}\mathbf{x}\right)\right\|_2 \leq d_j, j=1,\ldots,M\right\}.$$

Let $\mathbf{z}_j = f(y_j)\mathbf{x}$ and $\bar{\mathbf{x}}_j = y_j\mathbf{x}$, then by Lemma 4.8 we have

$$\mathbf{z}_j \geq \hat{f}(y)\mathbf{x} \geq \hat{a}_l y_j \mathbf{x} + \hat{b}_l \mathbf{x} \, \forall \, l \Rightarrow \mathbf{z}_j \geq \hat{a}_l \bar{\mathbf{x}}_j + \hat{b}_l \mathbf{x}, \, \forall \, l,$$

since $\mathbf{x}$ is nonnegative and

$$\sqrt{\frac{p^y}{1-p^y}} \geq \hat{f}(y) = \max_{l \in \{1,2,\ldots,L\}} \hat{a}_l y + \hat{b}_l, \quad y \in \,]0,1].$$

Constraints (4.5f) and (4.5g) follow directly. Since $\bar{\mathbf{x}}_j = y_j\mathbf{x}$ and since constraint (4.5j) holds, constraint (4.5h) must be satisfied. Finally we drop constraint $\mathbf{z}_j = f(y_j)\mathbf{x}$ and relax $\bar{\mathbf{x}}_j = y_j \cdot \mathbf{x}$ through constraint (4.5i). This leads to problem (4.5) and concludes this proof. ∎

5. Numerical results. In this section, we evaluate numerically the performance of our proposed methods. On one hand, we compare the different schemes that have been presented in section 4 to bound the (DRSKP-SDP) problem. We later illustrate how the solution from our proposed distributionally robust approach can protect against distribution ambiguity as opposed to solutions obtained through a stochastic programming approach. All the considered models were solved using Sedumi 1.3 [29] and CVX [13, 14] with their default parameters on an Intel Core 2 Duo @ 2.26 GHz with 4.0 GB RAM.

5.1. Numerical evaluation of (DRSKP-SDP) bounds. We focus on the SDP relaxation of the multidimensional knapsack problems, referred to as DRSKP-SDP in section 4. We compare the bound obtained using our sequential approximation approach, denoted by V^I to the bounds that are based on Bonferroni's inequality and on the work by Zymler, Kuhn, and Rustem denoted by V^B and V^Z, respectively. As all these bounds are lower bounds for (DRSKP-SDP), we compare their value with the proposed uppper bound based on linearization, denoted by V^U. In order to have a better idea of the quality of the fractional solution obtained from the (DRSKP-SDP) problem we apply a randomized rounding procedure to generate a feasible solution to the distributionally robust knapsack problem. The value of this rounded solution, denoted by V^R provides a lower bound for this problem.

We perform our tests on two different instance sizes characterized by the following parameters: number of items $n = \{10, 20\}$, the number of random variables $m = \{5, 6\}$, and number of joint chance constraints $M = \{4, 5\}$. For the sake of simplicity, we set the utility function $u(y) = y$ and the matrix $\mathbf{R}$ is deterministic and generated by the MATLAB function "gallery('randcorr',n)*10". The probabilistic capacity constraints are generated with vector means μ_j drawn from the uniform distribution on $[5, 10]$, and the covariance matrix Σ_j generated by MATLAB function "gallery('randcorr',n)*2". The capacity d_j is independently chosen from the $[200, 300]$ interval. The elements of $\mathbf{A_j}^{\tilde{\mathbf{w}}_j}$ are uniformly generated on the interval $[0, 1]$. Moreover, we choose three tangent points $z_1 = 0.01$, $z_2 = 0.1$, and $z_3 = 0.4$ for the piecewise tangent approximations.

Numerical results are given by Tables 1 and 2, where column one lists the values of η. The columns from two to five present the optimal value of the sequential approximation, the Bonferroni approximation, the approximation by Zymler, Kuhn, and Rustem and the piecewise tangent approximation, respectively, while their corresponding CPU time is given by the columns from nine to twelve. Notice that the CPU time is given in seconds. Columns six to eight present the gap between the

TABLE 1
Computational results of (DRSKP-SDP) when $n = 10$, $m = 5$, $M = 4$.

η	V^I	V^Z	V^B	V^U	$\frac{V^U-V^I}{V^I}$	$\frac{V^U-V^Z}{V^Z}$	$\frac{V^U-V^B}{V^B}$	CPU^I	CPU^Z	CPU^B	CPU^U	V^R
1%	46.67	46.65	36.04	46.76	0.19%	0.24%	29.74%	5.60	393.40	0.28	1.31	41.23
2%	54.78	54.74	43.44	54.94	0.29%	0.37%	26.47%	5.80	390.72	0.24	1.13	50.92
3%	59.26	59.20	47.80	59.43	0.29%	0.39%	24.33%	5.45	496.26	0.26	1.09	52.34
4%	62.25	62.18	50.83	62.47	0.35%	0.47%	22.90%	5.62	509.21	0.25	1.10	52.34
5%	64.46	64.38	53.16	64.70	0.37%	0.50%	21.71%	5.83	862.02	0.26	1.14	54.76
6%	66.19	66.10	55.02	66.39	0.30%	0.44%	20.67%	5.86	497.16	0.23	1.05	56.49
7%	67.59	67.50	56.56	67.77	0.27%	0.40%	19.82%	5.74	627.58	0.26	1.17	56.80
8%	68.76	68.66	57.88	68.93	0.25%	0.39%	19.09%	5.75	518.17	0.29	1.15	60.95
9%	69.76	69.66	59.02	69.93	0.24%	0.39%	18.49%	5.82	506.97	0.29	1.21	61.59
10%	70.63	70.53	60.02	70.81	0.25%	0.40%	17.98%	6.21	623.61	0.25	1.53	61.59

TABLE 2
Computational results of (DRSKP-SDP) when $n = 20$, $m = 6$, $M = 5$.

η	V^I	V^Z	V^B	V^U	$\frac{V^U-V^I}{V^I}$	$\frac{V^U-V^Z}{V^Z}$	$\frac{V^U-V^B}{V^B}$	CPU^I	CPU^Z	CPU^B	CPU^U	V^R
1%	60.96	60.94	46.57	61.19	0.38%	0.41%	31.39%	14.70	9311.99	0.61	5.40	50.19
2%	71.61	71.54	55.96	71.83	0.31%	0.41%	28.36%	17.32	8918.62	0.59	4.82	58.36
3%	77.39	77.29	61.50	77.65	0.34%	0.47%	26.26%	17.78	6831.69	0.58	5.18	69.26
4%	81.20	81.08	65.37	81.45	0.31%	0.46%	24.60%	18.52	12461.91	0.63	5.54	72.89
5%	83.95	83.83	68.31	84.23	0.33%	0.48%	23.31%	16.39	9536.62	0.71	5.41	76.61
6%	86.07	85.94	70.66	86.38	0.36%	0.51%	22.25%	17.12	9206.91	0.75	5.41	76.61
7%	87.78	87.64	72.61	88.11	0.38%	0.54%	21.35%	17.18	19749.16	0.72	5.52	76.61
8%	89.18	89.04	74.27	89.51	0.37%	0.53%	20.52%	17.30	9370.42	0.67	5.37	76.61
9%	90.37	90.22	75.70	90.70	0.37%	0.53%	19.82%	16.55	15440.99	0.65	5.22	76.61
10%	91.39	91.24	76.95	91.73	0.37%	0.54%	19.21%	16.98	15970.64	0.68	5.02	77.36

upper-bound approximation, the tangent approximation, and the three lower-bound approximations, i.e., the sequential approximation, the Bonferroni approximation, and the approximation by Zymler, Kuhn, and Rustem, respectively. The optimal value of the randomized rounding method is given in the last column.

From Tables 1 and 2, first of all, we can observe that the optimal values of all the four approximation methods, i.e., V^I, V^Z, V^B, and V^U, increase as η increases, which is true because the joint chance constraint becomes less restrictive as η increases. Second, based on the results of columns six and eight, the sequential approximation outperforms the Bonferroni approximation, as the largest gap for the sequential approximation is less than 0.40% while the smallest gap for the Bonferroni approximation is close to 18% for the two knapsack problems.[3] Admittedly, it takes more CPU time for the sequential approximation than for the Bonferrnoni approximation. Third, the sequential approximation outperforms the approximation by Zymler, Kuhn, and Rustem for two reasons: one is the gap of the sequential approximation is less than the one of the approximation by Zymler, Kuhn, and Rustem for all the instances. The second reason is that the approximation by Zymler, Kuhn, and Rustem takes much more CPU time than the sequential approximation. Especially for the instances of the second knapsack problem, the most CPU time that the sequential consumed is less than 20 seconds, but the greatest CPU time is more than 5 hours for the approximation by Zymler, Kuhn, and Rustem. In addition, column six of

[3]However, the Bonferroni approximation could be significantly improved using a sequential optimization method to select good η_j's (see, for example, [31]).

Tables 1 and 2 shows that the sequential approximation and the piecewise tangent approximation provide the competitive lower and upper bounds of (DRSKP-SDP), respectively, as shown by the small obtained gaps, i.e., less than 0.40%. Finally, it is easy to see that the randomized rounding procedure as well as our upper bound lead to small gaps for (DRSKP-SDP).

5.2. Numerical evaluation of distributionally robust framework. In this section, we compare the solution of our proposed distributionally robust approach with the solution of a stochastic programming approach. We compare the expected reward (i.e. under a risk neutral attitude) and the chance constraint performance obtained for solutions to the knapsack problem with a single capacity constraint when the real distribution is wrongly handled by different distributions for the weights although each of them display similar statistics regarding first and second moments. In particular, we consider the following problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathbb{E}_F[\xi^T \mathbf{A}^{\tilde{\mathbf{R}}^T} \mathbf{x}] \\ \text{subject to} \quad & \mathbb{P}_F(\tilde{\mathbf{w}}^T \mathbf{x} \leq d) \geq 1 - \eta, \\ & 0 \leq x_i \leq 1 \; \forall\, i \in \{1, 2, \ldots, n\}, \end{aligned}$$

where $0 < \eta < 0.5$, the expression $\xi^T \mathbf{A}^{\tilde{\mathbf{R}}^T} \mathbf{x}$ captures a linear reward function, and F has a well-known mean and covariance matrix structure but could be either one of the following three distributions: a normal distribution, a uniform distribution over an ellipsoidal ball, or a log-normal distribution.

In the case where the distribution is normal with mean μ and covariance matrix Σ, it is well known that the model reduces to the following problem [5]

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \nu^T \mathbf{x} \\ \text{subject to} \quad & \mu^T \mathbf{A}^{\tilde{\mathbf{w}}^T} \mathbf{x} + \Phi^{-1}(1-\eta) \|\Sigma^{1/2} \mathbf{A}^{\tilde{\mathbf{w}}^T} \mathbf{x}\|_2 \leq d \\ & 0 \leq x_i \leq 1, \forall\, i \in \{1, 2, \ldots, n\}, \end{aligned}$$

where $\nu = \mathbf{A}^{\tilde{\mathbf{R}}} \mu$ and $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of the standard normal distribution.

When ξ is uniformly distributed over the ellipsoid support $\S = \{\xi | (\xi - \mu)^T \Sigma^{-1} (\xi - \mu) \leq n + 3\}$, then $\mathbb{E}[(\xi - \mu)(\xi - \mu)^T] = \Sigma$ and the stochastic program is equivalent to the following problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \nu^T \mathbf{x} \\ \text{subject to} \quad & \mu^T \mathbf{A}^{\tilde{\mathbf{w}}^T} \mathbf{x} + \sqrt{(n+3)(\Psi^{-1}(1-2\eta))} \|\Sigma^{1/2} \mathbf{A}^{\tilde{\mathbf{w}}^T} \mathbf{x}\|_2 \leq d, \\ & 0 \leq x_i \leq 1 \; \forall\, i \in \{1, 2, \ldots, n\}, \end{aligned}$$

where $\Psi^{-1}(\cdot)$ is the inverse of the cumulative distribution of a beta(1/2;n/2+1) probability density function.

In what follows, we compare the performances of the solutions from these two stochastic-programming-based approaches to the solutions obtained by solving the distributionally robust models (3.3) and (3.4). In the case of the first model, we assume that the real support is given by the uniform distribution, i.e. $\S = \{\xi | (\xi - \mu)^T \Sigma^{-1} (\xi - \mu) \leq n + 3\}$.

Table 3

Comparison of average performance over 1000 *problem instances, with* $\eta = 0.01$ *and* $n = 10$*, for solutions obtained through stochastic programming (with assumed normal and uniform distribution) and distributionally robust solutions (with assumed mean, covariance information, and potentially support information). For each problem instance, the expected reward is evaluated relative to the value achieved by the solution of the stochastic program with uniform distribution while we measure the relative amount of additional capacity needed to satisfy the chance constraint under three different distributions with the same mean and covariance matrix.*

		Stochastic solutions		Robust solutions	
		Uniform	Normal	$\mathcal{D}(\S, \mu, \Sigma)$	$\mathcal{D}(\mu, \Sigma)$
Relative expected reward		-0%	-1%	-6%	-23%
Additional capacity needed	Uniform dist.	0%	0%	0%	0%
	Normal dist.	21%	0%	0%	0%
	Log-normal dist.	249%	225%	7%	0%

For the sake of comparing the robustness of the proposed solutions, we generate a random set of 1000 knapsack problem instances. For each instance, each term of the reward vector ν is generated uniformly from the interval $[0, 100]$, each term of the mean of ξ is uniformly drawn on the interval $[5, 10]$, and d is drawn on the interval $[20, 40]$. The terms of ξ are set to be uncorrelated with a variance uniformly distributed over $[0, \mu^2/(n + 3)]$; this choice ensures that the weights are necessarily positive when it is assumed that the distribution is uniform over an ellipsoid with covariance matrix Σ. Finally, for the sake of simplicity the matrix $\mathbf{A}^{\tilde{\mathbf{w}}}$ is chosen to be the identity matrix.

In Table 3, we compare the average performance over 1000 problem instances for the solutions obtained through stochastic programming, where the distribution is assumed normal or uniform, and through distributionally robust optimization. We set $\eta = 0.01$, $n = 10$, and assume that the mean and the covariance matrix are known with and without support information. For each problem instance, expected reward is evaluated relatively to the value achieved by the solution of the stochastic program with uniform distribution. For each of them, we also measure the relative amount of additional capacity that would be needed to satisfy the chance constraint under the three considered distributions. Note that while the normal and uniform distributions were uniquely defined based on the mean and covariance information, in the case of the log-normal distribution, we composed $\xi_i = 0.9\mu_i + \zeta_i$ with $\zeta_i \approx \ln \mathcal{N}\left(\ln\left(\frac{\bar{\mu}_i^2}{\sqrt{\bar{\mu}_i^2 + \sigma_i^2}}\right), \ln\left(1 + \frac{\sigma_i^2}{\bar{\mu}_i^2}\right)\right)$, where $\bar{\mu_i} = 0.1\mu_i$. The resulting distribution satisfies the mean and covariance information with heavier tail in the direction of large weight values.

Based on Table 3, we can infer that the distributionally robust approach is a conservative approximation of the stochastic program as it is willing to invest on average between 6% and 23% of the expected reward to protect against fluctuation of the distribution in the chance constraint. This is valuable since it is clearly the case that a stochastic program that assumes the distribution is uniform does not provide any protection against the changes of the other distribution that are studied. On the other hand, assuming the distribution is normal does provide more protection for a relatively small loss in reward. Yet, both stochastic-programming-based solutions fail significantly when the distribution ends up having heavier tails, as portrayed by log-normal distribution. In the latter situation, we remark that only the most robust approach still preserves the chance constraint requirement. The robust approach that assumes the ellipsoidal support might be thought of as a good compromise between stochastic and fully robust solutions. In particular, we see that the chance constraint is always satisfied for both uniform (as theory indicates) and normal distribution, while the capacity increase needed for the log-normal distribution appears reasonable.

6. Conclusions. In this paper, we studied a distributionally robust stochastic quadratic knapsack problem, where first and second moments, support, and independence information of the random variables are known. We prove that the single knapsack problem can be reformulated as an SDP after applying the SDP relaxation scheme to the binary constraints. Although this is not case for the multidimensional knapsack problem, it is still possible to efficiently approximate the relaxed version of the problem and obtain upper and lower bounds that appear numerically close to each other for a range of problem instances. Our numerical experiments also indicate that our proposed lower bounding approximation outperforms approximations that are based on Bonferroni's inequality and the work by Zymler, Kuhn, and Rustem [34]. Finally, an extensive set of experiments were conducted to illustrate how the conservativeness of the robust solutions does pay off in terms of ensuring the chance constraint is satisfied (or nearly satisfied) under a wide range of distribution fluctuations. Moreover, our approach can be applied to a large number of stochastic optimization problems with binary variables.

REFERENCES

[1] D. Bertsimas and M. Sim, *The price of robustness*, Oper. Res., 52 (2004), pp. 35–53.

[2] D. Bertsimas, K. Natarajan, and C.-P. Teo, *Probabilistic combinatorial optimization: Moments, semidefinite programming and asymptotic bounds*, SIAM J. Optim., 15 (2004), pp. 185–209.

[3] A. Billionnet and F. Calmels, *Linear programming for the* 0-1 *quadratic knapsack problem*, Eur. J. Oper. Res., 92 (1996), pp. 310–325.

[4] P. Brucker, *An O(n) algorithm for quadratic knapsack problems*, Oper. Res. Lett., 3 (1984), pp. 163–166.

[5] G. C. Calafiore and L. El Ghaoui, *On distributionally robust chance-constrained linear programs*, J. Optim. Theory Appl., 130 (2006), pp. 1–22.

[6] A. Caprara, D. Pisinger, and P. Toth, *Exact solution of the quadratic knapsack problem*, INFORMS J. Comput., 11 (1999), pp. 125–137.

[7] W. Chen, M. Sim, J. Sun, and C. P. Teo, *From CVaR to uncertainty set: implications in joint chance constrained optimization*, Oper. Res., 58 (2010), pp. 470–485.

[8] J. Cheng and A. Lisser, *A second-order cone programming approach for linear programs with joint probabilistic constraints*, Oper. Res. Lett., 40 (2012), pp. 325–328.

[9] E. Delage and Y. Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Oper. Res., 58 (2010), pp. 596–612.

[10] L. El Ghaoui, M. Oks, and F. Oustry, *Worst-case value-at-risk and robust portfolio optimization: a conic programming approach*, Oper. Res., 51 (2003), pp. 543–556.

[11] A. Gaivoronski, A. Lisser, and R. Lopez, *Knapsack problem with probability constraints*, J. Global Optim., 49 (2011), pp. 397–413.

[12] G. Gallo, P. Hammer, and B. Simeone, *Quadratic knapsack problems*, Math. Program. Stud., 12 (1980), pp. 132–149.

[13] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming, Version* 2.0 *Beta* http://cvxr.com/cvx (2012).

[14] M. Grant and S. Boyd, *Graph Implementations for Nonsmooth Convex Programs, Recent Advances in Learning and Control*, Lecture Notes in Control and Inform. Sci., Springer: Berlin, 2008.

[15] IBM ILOG, *CPLEX* Optimizer, http://www.ibm.com/software/integration/optimization/cplex/.

[16] E. Johnson, A. Mehrotra, and G. Nemhauser, *Min-cut clustering*, Math. Program., 62 (1993), pp. 133–152.

[17] P. Kouvelis and G. Yu, *Robust Discrete Optimization and its Application*, Kluwer Academic, Dordrecht, 1997.
[18] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*, Springer, Berlin, 2004.
[19] S. Kosuch and A. Lisser, *Upper bounds for the stochastic knapsack problem and a B&B algorithm*, Ann. Oper. Res., 176 (2010), pp. 77–93.
[20] S. Kosuch and A. Lisser, *On two-stage stochastic knapsack problems with or without probability constraint*, Discrete Appl. Math., 159 (2011), pp. 1827–1841.
[21] M. Kress, M. Penn, and M. Polukarov, *The minmax multidimensional knapsack problem with application to a chance-constrained problem*, Naval. Res. Logist., 54 (2007), pp. 656–666.
[22] S. Liao, C. van Delft, and J. P. Vial, *Distributionally robust workforce scheduling in call centers with uncertain arrival rates*, Optim. Methods Softw., 28 (2013), pp. 501–522.
[23] A. Lisser and R. Lopez, *Stochastic quadratic knapsack with recourse*, Electron. Notes Discrete Math., 36 (2010), pp. 97–104.
[24] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley, New York, 1990.
[25] M. Monaci and U. Pferschy, *On the robust knapsack problem*, SIAM J. Optim., 23 (2013), pp. 1956–1982.
[26] I. Pólik and T. Terlaky, *A survey of the S-Lemma*, SIAM Rev., 49 (2007), pp. 371–418.
[27] H. Scarf, *A min-max solution of an inventory problem,* in *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, CA, 1958, pp. 201–209.
[28] A. Shapiro, *On the duality theory of conic linear problems*, in Semi-Infinite Programming: Recent Advances, Kluwer Academic, Dordrecht, 2001, pp. 135–165.
[29] J. F. Sturm, *Using SeDuMi* 1.02*, a Matlab toolbox for optimization over symmetric cones*, Optim. Method. Softw., 11–12 (1999), pp. 625–653.
[30] M. R. Wagner, *Stochastic* 0–1 *linear programming under limited distributional information*, Oper. Res. Lett., 36 (2008), pp. 150–156.
[31] W. Wiesemann, D. Kuhn, and B. Rustem, *Multi-resource allocation in stochastic project scheduling*, Ann. Oper. Res., 193 (2012), pp. 193–220.
[32] W. Wiesemann, D. Kuhn, and M. Sim, *Distributionally robust convex optimization*, Oper. Res., to appear.
[33] K. Yoda and A. Prekopa, *Convexity and Solutions of Stochastic Multidimensional Knapsack Problems with Probabilistic Constraints*, Technical report Rutgers Center for Operations Research, New Brunswick, NJ, 2012.
[34] S. Zymler, D. Kuhn, and M. Rustem, *Distributionally robust joint chance constraints with second-order moment information*, Math. Program., 137 (2013), pp. 167–198.