

Distributionally robust stochastic knapsack problem

Jianqiang Cheng¹, Erick Delage², Abdel Lisser¹

¹Laboratoire de Recherche en Informatique (LRI), Université Paris Sud - XI

²Department of Management Sciences, HEC Montréal

ICSP2013
8 - 12 July 2013
Bergamo, Italy

Outlines

① Introduction

Outlines

- ① Introduction
- ② Distributionally Robust SKP

Outlines

- ① Introduction
- ② Distributionally Robust SKP
- ③ Solution Methods

Outlines

- ① Introduction
- ② Distributionally Robust SKP
- ③ Solution Methods
- ④ Numerical study

Quadratic knapsack problem

The *multidimensional* version of quadratic knapsack problem:

$$(KP) \quad \underset{\mathbf{x}}{\text{maximize}} \quad \mathbf{x}^T \mathbb{R} \mathbf{x} \quad (1a)$$

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

$$x_i \in \{0, 1\}, \quad \forall i \in \{1, 2, \dots, n\} \quad (1c)$$

- \mathbf{x} is a vector of binary values indicating whether each item is included in the knapsack
- $\mathbb{R} \in \Re^{n \times n}$ is a matrix which (i, j) -th term describes the linear contribution to reward of holding both items i and j .
- $\mathbf{w}_j \in \Re^n$ is a vector of attributes (typically weights) which total amount must satisfy the knapsack capacity d_j .

Stochastic quadratic knapsack problem

$$(SKP) \quad \underset{\mathbf{x}}{\text{maximize}} \quad \mathbf{x}^T \mathbb{R} \mathbf{x} \quad (2a)$$

$$\text{subject to} \quad \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \dots, M) \geq 1 - \eta \quad (2b)$$

$$x_i \in \{0, 1\}, \forall i \in \{1, 2, \dots, n\} \quad (2c)$$

for some $0 < \eta < 1$.

Stochastic quadratic knapsack problem

$$(SKP) \quad \underset{\mathbf{x}}{\text{maximize}} \quad \mathbf{x}^T \mathbb{R} \mathbf{x} \quad (2a)$$

$$\text{subject to} \quad \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \dots, M) \geq 1 - \eta \quad (2b)$$

$$x_i \in \{0, 1\}, \forall i \in \{1, 2, \dots, n\} \quad (2c)$$

for some $0 < \eta < 1$.

What happens if the probability distribution is unknown ?

Stochastic quadratic knapsack problem

$$(SKP) \quad \underset{\mathbf{x}}{\text{maximize}} \quad \mathbf{x}^T \mathbb{R} \mathbf{x} \quad (2a)$$

$$\text{subject to} \quad \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \dots, M) \geq 1 - \eta \quad (2b)$$

$$x_i \in \{0, 1\}, \forall i \in \{1, 2, \dots, n\} \quad (2c)$$

for some $0 < \eta < 1$.

What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

Distributionally robust problem

$$\begin{aligned}
 (DRSKP) \quad & \underset{\mathbf{x}}{\text{maximize}} && \mathbf{x}^T \mathbb{R} \mathbf{x} \\
 & \text{subject to} && \inf_{F \in \mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \dots, M) \geq 1 - \eta \\
 & && x_i \in \{0, 1\}, \forall i \in \{1, 2, \dots, n\}.
 \end{aligned}$$

where \mathcal{D} is a set of distributions.

Distributionally Robust SKP

Semidefinite programming approximation

We 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.

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

Distributionally Robust SKP

Semidefinite programming approximation

We 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.

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

Accordingly, we have the following semidefinite program (SDP) approximation relaxation problem:

$$(DRSKP - SDP) \quad \underset{\mathbf{x}, \mathbf{X}}{\text{maximize}} \quad \mathbb{R} \bullet \mathbf{X} \quad (4a)$$

$$\text{subject to} \quad \inf_{F \in \mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \dots, M) \geq 1 - \eta \quad (4b)$$

$$\begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix} \succeq 0 \quad (4c)$$

$$\mathbf{X}_i = x_i, \forall i, \quad \mathbf{X}_{i,j} \geq 0, \forall i, j \quad (4d)$$

Distributionally Robust SKP

Definition

Without loss of generality, for all $j = 1, \dots, M$, let ξ_j be a random vector in \mathbb{R}^m from which $\tilde{\mathbf{w}}_j$ depends linearly and

$$\tilde{\mathbf{w}}_j = \mathbf{A}_j^{\tilde{\mathbf{w}}_j} \xi_j, \quad j = 1, \dots, M.$$

where $\mathbf{A}_j^{\tilde{\mathbf{w}}_j}$ is a deterministic matrix.

Distributionally Robust SKP

Assumption

- *The distributional uncertainty set accounts for information about the support \mathcal{S} , mean μ_j , and an upper bound Σ_j on the covariance matrix of the random vector $\xi_j, j = 1, \dots, M$*

$$\mathcal{D}(\mathcal{S}, \mu_j, \Sigma_j) = \left\{ F_j \left| \begin{array}{l} \mathbb{P}(\xi_j \in \mathcal{S}) = 1 \\ \mathbb{E}_F[\xi_j] = \mu_j \\ \mathbb{E}_F[(\xi_j - \mu_j)(\xi_j - \mu_j)^T] \preceq \Sigma_j \end{array} \right. \right\}.$$

Distributionally Robust SKP

Assumption

- *The distributional uncertainty set accounts for information about the support \mathcal{S} , mean μ_j , and an upper bound Σ_j on the covariance matrix of the random vector $\xi_j, j = 1, \dots, M$*

$$\mathcal{D}(\mathcal{S}, \mu_j, \Sigma_j) = \left\{ F_j \left| \begin{array}{l} \mathbb{P}(\xi_j \in \mathcal{S}) = 1 \\ \mathbb{E}_F[\xi_j] = \mu_j \\ \mathbb{E}_F[(\xi_j - \mu_j)(\xi_j - \mu_j)^T] \preceq \Sigma_j \end{array} \right. \right\}.$$

- *The random vectors ξ_i and ξ_j are independent when $i \neq j$. The support of F_j is unbounded, i.e. $\mathcal{S} = \mathbb{R}^m$.*

Solution Methods

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 $\sum_{j=1}^M \eta_j = \eta$, we have

$$\mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j) \geq 1 - \eta_j, j = 1, \dots, M$$

$$\Rightarrow \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \dots, M) \geq 1 - \eta$$

- Notice that it is easy to prove that the optimal values of the Bonferroni approximations are lower bounds of DRSKP-SDP.

Solution Methods

Zymler et al. (2013) conservative bound

- Zymler et al. address the distributionally robust chance constraint by introducing a scaling parameter $\alpha \in \mathcal{A} = \{\alpha \in \mathbb{R}^M : \alpha > 0\}$ and reformulating it as a distributionally robust conditional value at risk constraint. For any $\alpha \in \mathcal{A}$

$$\begin{aligned} & \inf_{F \in \mathcal{D}} \mathbb{P}_F(\tilde{\mathbf{w}}_j^T \mathbf{x} \leq d_j, j = 1, \dots, M) \geq 1 - \eta \\ \Leftrightarrow & \inf_{F \in \mathcal{D}} \mathbb{P}_F(\max_j \{\alpha_j(\tilde{\mathbf{w}}_j^T \mathbf{x} - d_j)\} \leq 0) \geq 1 - \eta \\ \Leftrightarrow & \inf_{F \in \mathcal{D}} \text{CVaR-}\eta_F(\max_j \{\alpha_j(\tilde{\mathbf{w}}_j^T \mathbf{x} - d_j)\}) \leq 0 \end{aligned}$$

- The next step involves replacing the value at risk operator by an upper bounding CVAR operator to make the both convex and tractable.
- For any $\alpha > 0$, this leads to the following conservative approximation of the DRSKP-SDP.

Solution Methods

Theorem

The DRSKP-SDP problem with constraint (4b) replaced by,

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

where $\mathbf{0}_k$ and $\mathbf{0}_{k,k}$ are all zeros vector in \mathbb{R}^k and matrix in $\mathbb{R}^{k \times k}$ respectively, $\beta \in \mathbb{R}$, $\mathcal{M} \in \mathbb{R}^{(Mm+1) \times (Mm+1)}$, and $\mathbf{y}_j \in \mathbb{R}^{mM}$ 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.

Solution Methods

Deterministic formulation

- Under Assumption 1, one can actually show that the robust chance constraint (4b) is equivalent to:

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

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

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

Solution Methods

Deterministic formulation

Problem DRSKP-SDP is equivalent to the following deterministic problem

$$\begin{array}{ll} \underset{\mathbf{x}, \mathbf{y}}{\text{maximize}} & \mathbb{R} \bullet \mathbf{X} \end{array} \quad (5a)$$

$$\text{subject to} \quad \mu_j^T \mathbf{A}_j^{\tilde{\mathbf{w}}_j} \mathbf{x} + \sqrt{\frac{p^{y_j}}{1 - p^{y_j}}} \|\Sigma_j^{1/2} \mathbf{A}_j^{\tilde{\mathbf{w}}_j T} \mathbf{x}\|_2 \leq d_j \quad (5b)$$

$$\sum_{j=1}^M y_j = 1, \quad y_j \geq 0 \quad (5c)$$

$$\begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{bmatrix} \succeq 0 \quad (5d)$$

$$\mathbf{X}_{i,i} = x_i, \quad \forall i, \quad \mathbf{X}_{i,j} \geq 0, \quad \forall i, j \quad (5e)$$

Solution Methods

Remark

- For the chance constraint (5b), if we take y_j as fixed parameters and transform it 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 et al. ([Zymler et al 2013]).

Solution Methods

Sequential approximations

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

Solution Methods

Sequential approximations

Algorithm 1: Sequential Approximation Procedure

Initialization: Let $\mathbf{y}^1 \in \mathbb{R}_+^M$ 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 (5) 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

$$U : \Delta^M \times [0, 1]^M \rightarrow \Delta^M,$$

where $\Delta^M = \{\mathbf{y} \in \mathbb{R}^M \mid \mathbf{y} \geq 0 \text{ \& } \sum_j y_j = 1\}$ is the probability simplex, and $\times_{j=1}^M [\tilde{y}_j^t, 1]$ is the cartesian product of each $[\tilde{y}_j^t, 1]$ interval.

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**.

Sequential approximations

Theorem

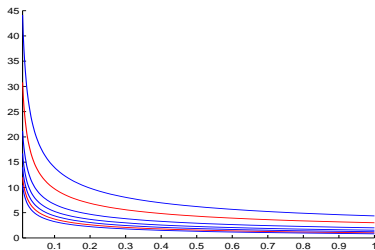
If the problem DRSKP-SDP is bounded and has a feasible solution for the initial values of \mathbf{y}^1 , and given that $U(\mathbf{y}; \tilde{\mathbf{y}}) \geq \tilde{\mathbf{y}}$ for all $\mathbf{y} \geq \tilde{\mathbf{y}}$, then Algorithm 1 terminates in a finite number of steps and the returned value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

Sequential approximations

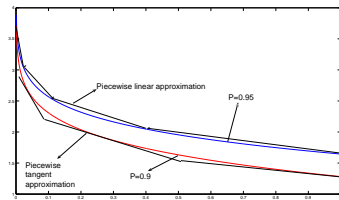
Remark

- In our implementation, for the initial parameter y^1 , all the elements are set to $\frac{1}{M}$.
- We choose the following update policy:
$$U(\mathbf{y}; \tilde{\mathbf{y}}) = \tilde{\mathbf{y}} + \alpha(1 - (\mathbf{y} - \tilde{\mathbf{y}}))$$
, where $\alpha = \frac{\sum_j (y_j - \tilde{y}_j)}{\sum_j (1 - (y_j - \tilde{y}_j))}$.
- The idea 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.

Upper bound through tangent approximation



(a)



(b)

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

Piecewise tangent approximation

Lemma

Function $\sqrt{\frac{p^y}{1-p^y}}$ is convex and decreasing in the region $(0, 1]$, when $p \in (0, 1)$.

$$\begin{array}{ll} \text{maximize} & \mathbb{R} \bullet \mathbf{X} \\ \text{subject to} & \end{array} \quad (6)$$

$$\left[\begin{array}{cc} \mathbf{0}_{m,m} & \Sigma_j^{1/2} \mathbf{A}_j \tilde{\mathbf{w}}_j^T \mathbf{z}_j \\ (\Sigma_j^{1/2} \mathbf{A}_j \tilde{\mathbf{w}}_j^T \mathbf{z}_j)^T & 0 \end{array} \right] \succeq (\mu_j^T \mathbf{A}_j \tilde{\mathbf{w}}_j^T \mathbf{x} - d_j) \mathbf{I} \quad (7)$$

$$\mathbf{z}_j \geq \hat{\mathbf{a}}_l \bar{\mathbf{x}}_j + \hat{\mathbf{b}}_l \mathbf{x}, \forall l \in \{1, 2, \dots, L\} \forall j \in \{1, 2, \dots, M\} \quad (8)$$

$$\sum_{j=1}^M \bar{\mathbf{x}}_j = \mathbf{x}, \forall i \quad (9)$$

$$0 \leq \bar{\mathbf{x}}_j \leq \mathbf{x}, \quad y_j - (1 - \mathbf{x}) \leq \bar{\mathbf{x}}_j \leq y_j, \forall j \in \{1, 2, \dots, M\} \quad (10)$$

$$\sum_{j=1}^M y_j = 1, \quad y_j \geq 0, \forall j \in \{1, 2, \dots, M\} \quad (11)$$

$$\left[\begin{array}{cc} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{array} \right] \succeq 0, \quad \mathbf{x}_{i,i} = x_i, \forall i, \quad \mathbf{x}_{i,j} \geq 0, \forall i, j, \quad (12)$$

Theorem

Problem DRSKP-SDP1 is a relaxed approximation of Problem DRSKP-SDP, i.e., the optimal value of DRSKP-SDP1 is an upper bound of DRSKP-SDP.

Numerical Results

Details of experiments

- We study the computational performance of our proposed methods for solving the DRSKP-SDP.
- The sequential approximation and the tangent approximation are compared with Bonferroni's approximation and Approximation by Zymmler et al..
- All the considered models are generated using MATLAB environment and solved by Sedumi with default parameters on an Intel(R)D @ 2.00 GHz with 4.0 GB RAM.

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

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

Table: Computational results of DRSKP-SDP when
 $n = 20, m = 6, M = 5$

Conclusions

- We solve the multidimensional quadratic optimization problem with uncertainty and joint probabilistic constraints.
- We consider that the uncertainty is defined by an uncertain set and solve the corresponding problem using a distributionally robust approach.
- We show that our approach outperforms the existing methods.
- Our approach can be easily extended to several stochastic combinatorial optimization problems.