On the Convexification of a Class of Mixed-Integer Conic Sets

Guxin Du, Rui Chen, and Linchuan Wei

School of Data Science, The Chinese University of Hong Kong, Shenzhen
guxindu@link.cuhk.edu.cn, {rchen,linchuanwei}@cuhk.edu.cn

Abstract. We investigate mixed-integer second-order conic (SOC) sets with a nonlinear right-hand side in the SOC constraint, a structure frequently arising in mixed-integer quadratically constrained programming (MIQCP). Under mild assumptions, we show that the convex hull can be exactly described by replacing the right-hand side with its concave envelope. This characterization enables strong relaxations for MIQCPs via reformulations and cutting planes. Computational experiments on distributionally robust chance-constrained knapsack variants demonstrate the efficacy of our reformulation techniques.

Keywords: Mixed-integer programming · Second-order cone · Ideal formulation.

1 Introduction

In this paper, we consider a mixed-integer nonlinear set specified by a second-order conic (SOC) constraint with a nonlinear right-hand side. Specifically, we consider

$$\mathcal{Z} := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathbb{R}^m : \|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \leq f(\boldsymbol{x}) \right\}, \tag{1}$$

where $\mathcal{X} \subsetneq \mathbb{R}^n$ is a compact set, A and B are matrices of appropriate dimensions, $\boldsymbol{d} \in \mathbb{R}^p$ is a constant vector, and $f : \mathcal{X} \to \mathbb{R}$ is a function that is upper semi-continuous on $\mathcal{X}$. We are particularly interested in the case when $\boldsymbol{x}$ are binary variables, i.e., $\mathcal{X} \subseteq \{0, 1\}^n$.

This general $\mathcal{Z}$ encompasses, as a special case, mixed-integer convex quadratically constrained sets, which form the core structure in mixed-integer chance-constrained and distributionally robust chance-constrained programming [38]. Such models have diverse applications, including surgery planning [16] and energy-efficient cloud computing [32]. For particular choices of A, B, d, and f, $\mathcal{Z}$ reduces to the following mixed-binary second-order conic set studied in [3]:

$$\left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \{0, 1\}^n \times \mathbb{R}^m : \sqrt{\sigma + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{m} b_i y_i^2} \leq T \right\},$$

whose convex hull description remains an open question.

Our contribution lies within the broader literature on strong formulations for mixed-integer nonlinear sets, which can be classified into three streams. One stream of research extends techniques from the linear to the nonlinear setting, including Gomory cuts [13], mixed-integer rounding cuts [6], lifting [7], intersection cuts [25], disjunctive programming [12], and lift-and-project cuts [34]. Another line of work exploits the submodularity [17] inherent in certain structures to derive tight relaxations [3–5]. Most relevant to our study is the third stream, which focuses on strong formulations for *structured* mixed-integer nonlinear sets. These include quadratically constrained sets [10, 11, 26, 35], quadratic programs with semicontinuous variables [1, 2, 14, 18, 21, 36], mixed-integer conic sets [3, 6, 20, 22, 31], and nonlinear disjunctive sets [23], among others.

In Section 2, we prove that under some mild assumptions, the convex hull of the set $\mathcal{Z}$ is given by

$$\mathcal{W} := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{conv}(\mathcal{X}) \times \mathbb{R}^m : \|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \leq \hat{f}(\boldsymbol{x}) \right\}, \qquad (2)$$

where $\hat{f}(\cdot)$ denotes the concave envelope of $f(\cdot)$ over the domain $\mathcal{X}$, i.e., $\hat{f} = -\mathrm{conv}(-f)$. Therefore, the task of convexifying $\mathcal{Z}$ reduces entirely to characterizing $\hat{f}$. In Section 3, we demonstrate how some mixed-integer quadratically constrained sets can be reformulated into $\mathcal{Z}$. Although obtaining a complete characterization of the concave envelope $\hat{f}$ is challenging, we show how a strong relaxation can be derived using the concave envelope of $f(\cdot)^2$ and establish an approximation gap. We assess the effectiveness of this reformulation technique through a computational study on a distributionally robust chance-constrained mixed-binary (multi-dimensional) knapsack problem. The results indicate that our reformulation yields a tighter bound and significantly reduces the solution time compared to a naive formulation.

Notations. Throughout this paper, we use boldface letters to denote vectors. Superscripts are used to distinguish between different vectors in a collection, while subscripts indicate specific components of a vector. Given a matrix A, $\mathrm{col}(A)$ denotes the linear space spanned by the column vectors of A. Given a set S, $\mathrm{Proj}_x(S)$ denotes its orthogonal projection onto the x-space. We let $\mathbb{N}$ denote the set of positive integers, i.e., $\mathbb{N} = \{1, 2, \ldots\}$. For $n \in \mathbb{N}$, we let Δ_n denote the standard n-simplex in $\mathbb{R}^{n+1}$, i.e., $\Delta_n = \{\boldsymbol{\lambda} \in \mathbb{R}_+^{n+1} : \sum_{k=1}^{n+1} \lambda_k = 1\}$.

2   Ideal Formulation for $\mathcal{Z}$

Without loss of generality, we may assume that $\mathcal{Z}$ satisfies the following assumptions, possibly after a reformulation.

Assumption 1 *For all $\boldsymbol{x} \in \mathcal{X}$, there exists $\boldsymbol{y} \in \mathbb{R}^m$, which may depend on $\boldsymbol{x}$, such that $\|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \leq f(\boldsymbol{x})$, i.e., $\mathrm{Proj}_{\boldsymbol{x}}(\mathcal{Z}) = \mathcal{X}$.*

Assumption 2 *All columns of A and the vector $\boldsymbol{d}$ can be expressed as linear combinations of columns of B, i.e., $\mathrm{col}(A) \subseteq \mathrm{col}(B)$, and $\boldsymbol{d} \in \mathrm{col}(B)$.*

Indeed, given $\mathcal{Z}$, we may redefine $\mathcal{X}$ as $\text{Proj}_{\boldsymbol{x}}(\mathcal{Z})$ so that Assumption 1 is satisfied. Regarding Assumption 2, given $(A, B, \boldsymbol{d})$ and $f(\cdot)$, one can decompose A and $\boldsymbol{d}$ into $A = \bar{A} + A^{\perp}$ and $\boldsymbol{d} = \bar{\boldsymbol{d}} + \boldsymbol{d}^{\perp}$ such that $\text{col}(\bar{A}) \subseteq \text{col}(B)$, $\bar{\boldsymbol{d}} \in \text{col}(B)$, columns of $A^{\perp}$ and $\boldsymbol{d}^{\perp}$ lie in the kernel of B by projecting the columns of A and $\boldsymbol{d}$ onto $\text{col}(B)$. Then we can rewrite $\mathcal{Z}$ equivalently as the set of $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathbb{R}^m$ such that

$$\left\| \bar{A}\boldsymbol{x} + B\boldsymbol{y} + \bar{\boldsymbol{d}} \right\|_2 \le f'(\boldsymbol{x}) := \sqrt{f^2(\boldsymbol{x}) - \left\| A^{\perp}\boldsymbol{x} + \boldsymbol{d}^{\perp} \right\|_2^2}, \tag{3}$$

in which case Assumption 2 is satisfied by $(\bar{A}, B, \bar{\boldsymbol{d}})$ and $f'(\cdot)$. Note that $f' : \mathcal{X} \to \mathbb{R}$ is well-defined and nonnegative on $\mathcal{X}$ under Assumption 1.

Under Assumption 1, $f(\cdot)$ is nonnegative over its domain $\mathcal{X}$, and so is $\hat{f}(\cdot)$ over $\text{conv}(\mathcal{X})$. Our next result shows that one can characterize the convex hull of the set $\mathcal{Z}$ given the concave envelope $\hat{f}(\cdot)$.

Theorem 1. *Suppose Assumptions 1 and 2 hold. Then the convex hull of $\mathcal{Z}$ can be characterized as*

$$\text{conv}(\mathcal{Z}) = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \text{conv}(\mathcal{X}) \times \mathbb{R}^m : \|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \le \hat{f}(\boldsymbol{x}) \right\}, \tag{4}$$

with $\text{conv}(\mathcal{Z})$ being a closed set.

Before proving the theorem, the following lemma on the property of concave envelopes over compact sets is needed. The proof is presented in Appendix A.

Lemma 1. *Let $\mathcal{X} \subsetneq \mathbb{R}^n$ be a compact set, $f : \mathcal{X} \to \mathbb{R}$ be an upper semi-continuous function, and $\hat{f} : \text{conv}(\mathcal{X}) \to \mathbb{R}$ be its concave envelope over $\text{conv}(\mathcal{X})$. Then, for any $\boldsymbol{x} \in \text{conv}(\mathcal{X})$, there exist $K \le n + 1$ points $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^K \in \mathcal{X}$ and coefficients $\lambda_1, \ldots, \lambda_K > 0$ with $\sum_{k=1}^K \lambda_k = 1$ such that $\boldsymbol{x} = \sum_{k=1}^K \lambda_k \boldsymbol{x}^k$ and $\hat{f}(\boldsymbol{x}) = \sum_{k=1}^K \lambda_k f(\boldsymbol{x}^k)$. Moreover, $\hat{f}(\cdot)$ is upper semi-continuous on $\text{conv}(\mathcal{X})$.*

Now we are ready to prove Theorem 1.

Proof (of Theorem 1). Let $\mathcal{W}$ be as defined in (2). Note that $\mathcal{W}$ is closed and convex since $\hat{f}(\cdot)$ is upper semi-continuous and concave, and $\text{conv}(\mathcal{X}) \times \mathbb{R}^m$ is closed and convex. Therefore, we only need to show that $\text{conv}(\mathcal{Z}) = \mathcal{W}$. We have $\mathcal{Z} \subseteq \mathcal{W}$ as by definition $\hat{f}(\boldsymbol{x}) \ge f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$. Then the inclusion $\text{conv}(\mathcal{Z}) \subseteq \mathcal{W}$ follows from convexity of $\mathcal{W}$. It remains to show that $\mathcal{W} \subseteq \text{conv}(\mathcal{Z})$.

We first consider the case when B has full column rank, i.e., $B\boldsymbol{y} = \boldsymbol{0}$ if and only if $\boldsymbol{y} = \boldsymbol{0}$. Then $\mathcal{Z}$ (resp. $\mathcal{W}$) is closed and bounded and hence compact, as $f(\cdot)$ (resp. $\hat{f}(\cdot)$) is bounded and upper semi-continuous. By [29][Corollary 18.5.1], in order to prove $\mathcal{W} \subseteq \text{conv}(\mathcal{Z})$, it suffices to show that every extreme point of $\mathcal{W}$ lies in $\mathcal{Z}$. Let $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ be an arbitrary extreme point of $\mathcal{W}$. Then by Lemma 1, there exist $1 \le K \le n + 1$ points $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^K \in \mathcal{X}$ and $\lambda_1, \ldots, \lambda_K > 0$ with $\sum_{k=1}^K \lambda_k = 1$, such that $\bar{\boldsymbol{x}} = \sum_{k=1}^K \lambda_k \boldsymbol{x}^k$ and $\hat{f}(\bar{\boldsymbol{x}}) = \sum_{k=1}^K \lambda_k f(\boldsymbol{x}^k)$. Then we have the following two cases.

(a) $K = 1$. In this case, we have $\bar{\boldsymbol{x}} = \boldsymbol{x}^1$ and $\hat{f}(\bar{\boldsymbol{x}}) = f(\boldsymbol{x}^1)$, implying $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \in \mathcal{Z}$.

(b) $K \geq 2$. We will prove by contradiction that $(\bar{x}, \bar{y})$ cannot be an extreme point of $\mathcal{W}$. The key idea is finding two distinct points whose convex combination is $(\bar{x}, \bar{y})$, and the discussion is separated into the following three subcases.

(i) $\|A\bar{x} + B\bar{y} + d\|_2 < \hat{f}(\bar{x})$. For any nonzero vector $u \in \mathbb{R}^m$, we can choose $\epsilon > 0$ small enough such that $\|A\bar{x} + B(\bar{y} \pm \epsilon u) + d\|_2 < \hat{f}(\bar{x})$. Thus $(\bar{x}, \bar{y} \pm \epsilon u) \in \mathcal{W}$, contradicting the fact that $(\bar{x}, \bar{y})$ is an extreme point of $\mathcal{W}$.

(ii) $\|A\bar{x} + B\bar{y} + d\|_2 = \hat{f}(\bar{x}) = 0$. Then we must have $f(x^k) = 0$ for $k = 1, ..., K$, since $f(x^k) \geq 0$ by Assumption 1 and $\sum_{k=1}^K \lambda_k f(x^k) = \hat{f}(\bar{x}) = 0$ in this case. Let $v = x^1 - \bar{x}$. Then we can perturb $\bar{x}$ along the direction and the reverse direction of v while keeping it in $\mathrm{conv}(\{x^1, \ldots, x^K\}) \subseteq \mathrm{conv}(\mathcal{X})$ since $K \geq 2$. Let $u \in \mathbb{R}^m$ satisfy $Bu = -Av$. Such u exists since $\mathrm{col}(A) \subseteq \mathrm{col}(B)$. For $\epsilon > 0$ small enough (so that $\bar{x} \pm \epsilon v$ are still in $\mathrm{conv}\{x^1, ..., x^K\}$)

$$\|A(\bar{x} \pm \epsilon v) + B(\bar{y} \pm \epsilon u) + d\|_2$$
$$= \|A\bar{x} + B\bar{y} + d\|_2$$
$$= 0$$
$$\leq \hat{f}(\bar{x} \pm \epsilon v).$$

The last inequality follows from nonnegativity of $\hat{f}$ (due to Assumption 1) and the fact that $\bar{x} \pm \epsilon v \in \mathrm{conv}(\{x^1, \ldots, x^K\}) \subseteq \mathrm{conv}(\mathcal{X})$. Hence we have $(\bar{x} \pm \epsilon v, \bar{y} \pm \epsilon u) \in \mathcal{W}$, contradicting to the fact that $(\bar{x}, \bar{y})$ is an extreme point.

(iii) $\|A\bar{x} + B\bar{y} + d\|_2 = \hat{f}(\bar{x}) > 0$. Let v be as defined above. But u in this case has to be chosen more carefully since the value of $\hat{f}(\cdot)$ may go down around $\bar{x}$ in this case. Specifically, by Assumption 2, we let $u \in \mathbb{R}^m$ be chosen such that

$$Bu = \frac{f(x^1) - \hat{f}(\bar{x})}{\hat{f}(\bar{x})}(A\bar{x} + B\bar{y} + d) - Av.$$

Then, for small enough $\epsilon > 0$, we have

$$\|A(\bar{x} + \epsilon v) + B(\bar{y} + \epsilon u) + d\|_2$$
$$= \left\|\frac{(1 - \epsilon)\hat{f}(\bar{x}) + \epsilon f(x^1)}{\hat{f}(\bar{x})}(A\bar{x} + B\bar{y} + d)\right\|_2$$
$$= (1 - \epsilon)\hat{f}(\bar{x}) + \epsilon f(x^1)$$
$$\leq (1 - \epsilon)\hat{f}(\bar{x}) + \epsilon \hat{f}(x^1) \leq \hat{f}(\bar{x} + \epsilon v), \tag{5}$$

where the second equality follows from positive homogeneity of the norm and the fact that we can choose ϵ small enough so that $(1 - \epsilon)\hat{f}(\bar{x}) + \epsilon f(x^1) > 0$, and the last inequality follows from concavity of $\hat{f}(\cdot)$. On

the other hand, for small enough $\epsilon > 0$, we have

$$\|A(\bar{\boldsymbol{x}} - \epsilon \boldsymbol{v}) + B(\bar{\boldsymbol{y}} - \epsilon \boldsymbol{u}) + \boldsymbol{d}\|_2$$

$$= \left\| \frac{(1+\epsilon)\hat{f}(\bar{\boldsymbol{x}}) - \epsilon f(\boldsymbol{x}^1)}{\hat{f}(\bar{\boldsymbol{x}})} (A\bar{\boldsymbol{x}} + B\bar{\boldsymbol{y}} + \boldsymbol{d}) \right\|_2$$

$$= (1+\epsilon)\hat{f}(\bar{\boldsymbol{x}}) - \epsilon f(\boldsymbol{x}^1)$$

$$= \left(1 + \epsilon - \frac{\epsilon}{\lambda_1}\right)\hat{f}(\bar{\boldsymbol{x}}) + \frac{\epsilon}{\lambda_1}\sum_{k=2}^{K}\lambda_k f(\boldsymbol{x}^k)$$

$$\leq \left(1 + \epsilon - \frac{\epsilon}{\lambda_1}\right)\hat{f}(\bar{\boldsymbol{x}}) + \frac{\epsilon}{\lambda_1}\sum_{k=2}^{K}\lambda_k \hat{f}(\boldsymbol{x}^k)$$

$$\leq \hat{f}\left(\left(1 + \epsilon - \frac{\epsilon}{\lambda_1}\right)\bar{\boldsymbol{x}} + \frac{\epsilon}{\lambda_1}\sum_{k=2}^{K}\lambda_k \boldsymbol{x}^k\right)$$

$$= \hat{f}(\bar{\boldsymbol{x}} - \epsilon \boldsymbol{v}), \tag{6}$$

where the second equality holds if we choose $\epsilon > 0$ small enough so that $(1+\epsilon)\hat{f}(\bar{\boldsymbol{x}}) - \epsilon f(\boldsymbol{x}^1) > 0$, and the third and the last equalities follow from the facts that $\hat{f}(\bar{\boldsymbol{x}}) = \sum_{k=1}^{K}\lambda_k f(\boldsymbol{x}^k)$ and $\bar{\boldsymbol{x}} = \sum_{k=1}^{K}\lambda_k \boldsymbol{x}^k$. To conclude, by inequalities (5) and (6), we have $(\bar{\boldsymbol{x}} \pm \epsilon \boldsymbol{v}, \bar{\boldsymbol{y}} \pm \epsilon \boldsymbol{u}) \in \mathcal{W}$ for small enough $\epsilon > 0$, whose midpoint is $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$.

In summary, we have demonstrated that any extreme point of $\mathcal{W}$ must belong to $\mathcal{Z}$, and thus $\mathcal{W} = \text{conv}(\mathcal{Z})$, when B has full column rank, i.e., $\text{rank}(B) = m$.

Now consider the case when $\text{rank}(B) = \hat{m} < m$. In this case, we can find a full-rank matrix $U \in \mathbb{R}^{m \times m}$ (corresponding to some elementary column operation) such that $BU = (\hat{B}, \boldsymbol{0})$, where $\hat{B}$ has $\hat{m}$ columns with full column rank. Define

$$\hat{\mathcal{W}} = \{(\boldsymbol{x}, \boldsymbol{w}) \in \text{conv}(\mathcal{X}) \times \mathbb{R}^m : \left\| A\boldsymbol{x} + \hat{B}\boldsymbol{w}_{1:\hat{m}} + \boldsymbol{d} \right\|_2 \leq \hat{f}(\boldsymbol{x})\},$$

$$\hat{\mathcal{Z}} = \left\{(\boldsymbol{x}, \boldsymbol{w}) \in \mathcal{X} \times \mathbb{R}^m : \left\| A\boldsymbol{x} + \hat{B}\boldsymbol{w}_{1:\hat{m}} + \boldsymbol{d} \right\|_2 \leq f(\boldsymbol{x})\right\},$$

where $\boldsymbol{w}_{1:\hat{m}}$ is the subvector of $\boldsymbol{w}$ consisting of the first $\hat{m}$ elements. The two sets are unconstrained in the last $m - \hat{m}$ dimensions. Let $\tilde{\mathcal{W}}$ (resp. $\tilde{\mathcal{Z}}$) denote the projection of $\hat{\mathcal{W}}$ (resp. $\hat{\mathcal{Z}}$) onto the first $(n+\hat{m})$-dimensional subspace. Then we have the following observations:

(a) $\begin{pmatrix} I_n & \boldsymbol{0} \\ \boldsymbol{0} & U^{-1} \end{pmatrix} : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ is a linear bijection from $\mathcal{W}$ (resp. $\mathcal{Z}$) to $\hat{\mathcal{W}}$ (resp. $\hat{\mathcal{Z}}$);

(b) $\hat{\mathcal{W}} = \tilde{\mathcal{W}} \times \mathbb{R}^{m-\hat{m}}$ and $\hat{\mathcal{Z}} = \tilde{\mathcal{Z}} \times \mathbb{R}^{m-\hat{m}}$;

(c) $\hat{\mathcal{W}}$ is convex and $\text{conv}(\hat{\mathcal{Z}}) = \text{conv}(\tilde{\mathcal{Z}}) \times \mathbb{R}^{m-\hat{m}}$;

(d) $\tilde{\mathcal{W}}$ and $\tilde{\mathcal{Z}}$ are compact sets respectively admitting characterization in the form of (2) and (1) with $\hat{B}$ having full column rank and Assumptions 1 and 2 holding.

Applying our proof for the full-rank B case, we know that $\tilde{\mathcal{W}} = \mathrm{conv}(\tilde{\mathcal{Z}})$ and they are compact. Therefore, $\hat{\mathcal{W}} = \mathrm{conv}(\hat{\mathcal{Z}})$ and they are closed. It then follows from observation (a) that $\mathcal{W} = \mathrm{conv}(\mathcal{Z})$ and they are closed, as linear bijection preserves closedness and convexity. □

Note that the above proof relies only on positive homogeneity and nonnegativity of the 2-norm, properties that hold for any norm. Therefore, Theorem 1 remains valid if the 2-norm in (1) and (4) is replaced by any other norm.

Corollary 1. *Let $\|\cdot\|$ be any norm defined in the Euclidean space. Define $\mathcal{Z}$ as*

$$\mathcal{Z} := \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathbb{R}^m : \|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\| \le f(\boldsymbol{x})\}.$$

Suppose Assumptions 1 and 2 hold. Then the convex hull of $\mathcal{Z}$ can be characterized as

$$\mathrm{conv}(\mathcal{Z}) = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{conv}(\mathcal{X}) \times \mathbb{R}^m : \|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\| \le \hat{f}(\boldsymbol{x})\right\},$$

with $\mathrm{conv}(\mathcal{Z})$ being a closed set.

However, we shall also note that if Assumption 2 is not originally satisfied by $\mathcal{Z}$, then in general there may not exist a reformulation of $\mathcal{Z}$ like (3) in the 2-norm case such that Assumption 2 holds after reformulation.

Assumption 2 ensures that, whichever direction we perturb $\boldsymbol{x}$ there exists a perturbation of $\boldsymbol{y}$ such that $A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}$ is scaled with an arbitrary positive factor we choose, while keeping the direction unchanged. That depicts the essential property of the conic set $\mathcal{Z}$ that underlies Theorem 1. However, if $\boldsymbol{y}$ is constrained, this property may no longer hold due to the reduced degrees of freedom in $\boldsymbol{y}$, resulting in the characterization (4) inexact. The following two examples illustrate the scenarios where Assumption 2 is violated and $\boldsymbol{y}$ is constrained, respectively.

Example 1. Let $m = n = 1$, $\mathcal{X} = \{0, 1\}$, $A = (1, 0)^\top$, $B = (0, 1)^\top$, $\boldsymbol{d} = \boldsymbol{0}$, and $f(x) = \sqrt{2}$ be a constant function over $\mathcal{X}$. Then set $\mathcal{Z}$ is given by

$$\mathcal{Z} = \left\{(x, y) \in \{0, 1\} \times \mathbb{R} : \sqrt{x^2 + y^2} \le \sqrt{2}\right\}.$$

For this particular $f(\cdot)$, we have $\hat{f}(x) = \sqrt{2}$ for all $x \in \mathrm{conv}(\mathcal{X}) = [0, 1]$ while Assumption 2 is not satisfied. In this case, $\mathcal{W}$ actually includes some extra region apart from $\mathrm{conv}(\mathcal{Z})$, as shown in Figure 1a. To obtain the $\mathrm{conv}(\mathcal{Z})$, following (3), one may first reformulate $\mathcal{Z}$ as

$$\mathcal{Z} = \left\{(x, y) \in \{0, 1\} \times \mathbb{R} : |y| \le f'(x) := \sqrt{2 - x^2}\right\}$$

so that Assumption 2 is satisfied, and note that $\hat{f}'(x) = (1 - \sqrt{2})x + \sqrt{2}$ over $\mathrm{conv}(\mathcal{X}) = [0, 1]$. Then by Theorem 1 we have

$$\mathrm{conv}(\mathcal{Z}) = \left\{(x, y) \in [0, 1] \times \mathbb{R} : |y| \le (1 - \sqrt{2})x + \sqrt{2}\right\}.$$
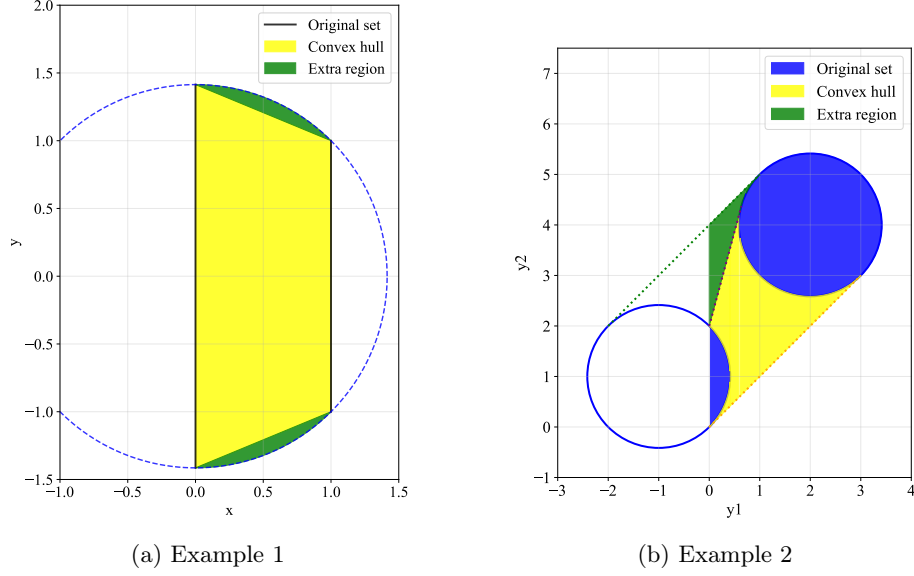
∎

(a) Example 1   (b) Example 2

Fig. 1: The convex hull of $\mathcal{Z}$ (a) when Assumption 2 does not hold, or (b) when $\boldsymbol{y}$ is further constrained.

Example 2. Let $m = 1$ and $n = 2$, and define $\mathcal{Z}$ as

$$\mathcal{Z} = \left\{ (x, y_1, y_2) \in \{0, 1\} \times \mathbb{R}^2_+ : \left\| \begin{pmatrix} 3 \\ 3 \end{pmatrix} x + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\|_2 \leq \sqrt{2} \right\}. \quad (7)$$

Note that, in this example, nonnegativity constraints are imposed on continuous variables (y_1, y_2). As shown in Figure 1b, the nonnegative constraint cuts off part of the left ball, and $\text{conv}(\mathcal{Z})$ is strictly smaller than the set obtained by directly convexifying $\mathcal{X}$ in (7). ∎

3 Applications to Mixed-Integer Quadratically Constrained Programming

One of the primary reasons for investigating $\mathcal{Z}$ is its frequent occurrence in Mixed-Integer Quadratically Constrained Programs (MIQCPs) as a substructure. Indeed, by properly choosing $(A, B, \boldsymbol{d})$ and $f(\cdot)$, any constraint of the form

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}^\top \begin{pmatrix} Q_{xx} & Q_{xy} \\ Q_{xy}^\top & Q_{yy} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} + \begin{pmatrix} \boldsymbol{a}_x \\ \boldsymbol{a}_y \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \leq g(\boldsymbol{x}) \quad (8)$$

with a positive definite Q_{yy} can be reformulated as $\|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \leq f(\boldsymbol{x})$, as demonstrated in Appendix B. In the most common case when $g(\cdot)$ is a constant, $f(\cdot)$ can be expressed as $\sqrt{q(\cdot)}$ for some quadratic function $q(\cdot)$. Then, by

applying Theorem 1, one can reduce the convexification of the substructure (8) (together with some constraints on $\boldsymbol{x}$) to the convexification of the hypograph of $f(\cdot) = \sqrt{q(\cdot)}$ (restricted to some domain). While the convexification of the epi-/hypo-graph of a quadratic function over integer or binary points has been widely studied in the literature, notably on the Boolean quadric polytope (BQP) [9, 15, 27], the convexification of the square root of a quadratic function over such points remains largely unexplored (with an exception in the submodular case [3, 4, 37, 38]). However, as shown in the next result, one can utilize the knowledge about the concave envelope $\hat{q}(\cdot)$ of a function $q(\cdot)$ to derive convex outer approximations for the concave envelope of $f(\cdot) = \sqrt{q(\cdot)}$. The proof appears in Appendix A.

Proposition 1. *Let $\mathcal{X} \subsetneq \mathbb{R}^n$ be a compact set, $f : \mathcal{X} \to \mathbb{R}_+$ be an upper semi-continuous function, and $\hat{f} : \mathrm{conv}(\mathcal{X}) \to \mathbb{R}_+$ be its concave envelope over $\mathrm{conv}(\mathcal{X})$. Let $q : \mathcal{X} \to \mathbb{R}_+$ be such that $q(\boldsymbol{x}) = f(\boldsymbol{x})^2$ for all $\boldsymbol{x} \in \mathcal{X}$, and $\hat{q} : \mathrm{conv}(\mathcal{X}) \to \mathbb{R}_+$ be its concave envelope over $\mathrm{conv}(\mathcal{X})$. Suppose $f(\boldsymbol{x}) \in [L, U]$ for all $\boldsymbol{x} \in \mathcal{X}$. Then $\sqrt{\hat{q}(\cdot)}$ is concave and upper semi-continuous over $\mathrm{conv}(\mathcal{X})$, and for all $\boldsymbol{x} \in \mathrm{conv}(\mathcal{X})$, we have*

$$0 \leq \sqrt{\hat{q}(\boldsymbol{x})} - \hat{f}(\boldsymbol{x}) \leq \frac{(U - L)^2}{4(L + U)}, \tag{9}$$

where $\frac{0}{0} := 0$.

Proposition 1 motivates the following convex relaxation of $\mathrm{conv}(\mathcal{Z})$:

$$\mathcal{R}(\mathcal{Z}) := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{conv}(\mathcal{X}) \times \mathbb{R}^m : \|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \leq \sqrt{\hat{q}(\boldsymbol{x})} \right\}. \tag{10}$$

In practice, in order to obtain $\mathcal{R}(\mathcal{Z})$ or approximations thereof, we propose the following reformulation of $\mathcal{Z}$ in the extended variable space,

$$\mathcal{Z} = \mathrm{Proj}_{(\boldsymbol{x}, \boldsymbol{y})} \left\{ (\boldsymbol{x}, \boldsymbol{y}, \eta, \tau) \in \mathcal{X} \times \mathbb{R}^m \times \mathbb{R}_+^2 : \right.$$
$$\left. \|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \leq \eta, \ \eta^2 \leq \tau, \ \tau \leq q(\boldsymbol{x}) \right\}, \tag{11}$$

to exploit standard MIP solvers' ability to approximate the concave envelope of $q(\cdot)$ using cutting planes in possibly an even further extended variable space, especially in the case when $q(\cdot)$ is quadratic. Note that Assumption 1 may not initially hold, in which case we may also rely on solver cutting planes to characterize $\mathrm{conv}(\mathrm{Proj}_{\boldsymbol{x}}(\mathcal{Z}))$.

4   Computational Study

In this section, we evaluate the effectiveness of the reformulation proposed in Section 3 on a distributionally robust chance-constrained mixed-binary (multi-

dimensional) knapsack problem (DRCC-MB-(M)KP), given by

$$\max \ \boldsymbol{p}_x^\top \boldsymbol{x} + \boldsymbol{p}_y^\top \boldsymbol{y} \tag{12a}$$

$$\text{s.t.} \ \inf_{\mathbb{P} \in \mathcal{D}_j} \left\{ \mathbb{P}((\boldsymbol{w}_x^j)^\top \boldsymbol{x} + (\boldsymbol{w}_y^j)^\top \boldsymbol{y} \le c^j) \right\} \ge 1 - \alpha, \quad j \in \mathcal{J} \tag{12b}$$

$$\boldsymbol{x} \in \{0,1\}^n, \boldsymbol{y} \in [0,1]^m, \tag{12c}$$

where n denotes the number of discrete items, m the number of continuous items, and $\mathcal{J}$ indexes the multiple resource types. The resource consumptions and profits of the discrete items are denoted by vectors $\boldsymbol{w}_x \in \mathbb{R}^n$ and $\boldsymbol{p}_x \in \mathbb{R}^n$, while those of the continuous items are denoted by $\boldsymbol{w}_y \in \mathbb{R}^m$ and $\boldsymbol{p}_y \in \mathbb{R}^m$, respectively. The capacity limit of each resource type $j \in \mathcal{J}$ is denoted by c^j. For each resource type $j \in \mathcal{J}$, the distributionally robust chance constraint (DRCC) (12b) enforces that the total consumption shall not exceed its capacity limit with probability at least $1 - \alpha$ under any distribution of resource consumption $(\boldsymbol{w}_x^j, \boldsymbol{w}_y^j)$ in the ambiguity set $\mathcal{D}_j$. The ambiguity set $\mathcal{D}_j$ is chosen to be a Chebyshev ambiguity set, i.e., the set of all distributions with some given mean and covariance matrix. Then, for each resource type $j \in \mathcal{J}$, it is known [19] that the DRCC (12b) is equivalent to

$$\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} + \sqrt{\frac{1-\alpha}{\alpha}} \sqrt{\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}^\top \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^\top & \Sigma_{yy} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}} \le c, \tag{13}$$

where the dependence on the resource type j is omitted for brevity, and $\boldsymbol{\mu}$ and Σ denote the given mean and covariance of resource consumptions, respectively. We may further reformulate constraint (13) as

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}^\top \begin{pmatrix} \tilde{\Sigma}_{xx} & \tilde{\Sigma}_{xy} \\ \tilde{\Sigma}_{xy}^\top & \tilde{\Sigma}_{yy} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} + 2c \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \le c^2, \tag{14a}$$

$$\boldsymbol{\mu}_x^\top \boldsymbol{x} + \boldsymbol{\mu}_y^\top \boldsymbol{y} \le c, \tag{14b}$$

where $\tilde{\Sigma}_{xx} = \tilde{\alpha} \Sigma_{xx} - \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top$, $\tilde{\Sigma}_{xy} = \tilde{\alpha} \Sigma_{xy} - \boldsymbol{\mu}_x \boldsymbol{\mu}_y^\top$, $\tilde{\Sigma}_{yy} = \tilde{\alpha} \Sigma_{yy} - \boldsymbol{\mu}_y \boldsymbol{\mu}_y^\top$, and $\tilde{\alpha} = (1 - \alpha)/\alpha$. Once $\Sigma_{yy} \succ 0$ and α is small enough, we have $\tilde{\Sigma}_{yy} \succ 0$. As discussed in Section 3, each constraint of (14a) can be reformulated as $\|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \le f(\boldsymbol{x})$, and hence into the form of (11). In the remainder of this section, we test the following two MIQCP formulations for DRCC-MB-(M)KP:

1. CCP: Model (12) with all DRCCs (12b) reformulated into the form of (14);
2. SOC: Model (12) with all DRCCs (12b) reformulated into the form of (11) (using auxiliary variables) and (14b).

Formulation CCP represents the vanilla MIQCP formulation of the original problem (12). Formulation SOC explicitly exposes the SOC structure and the BQP structure, which enables the solver to generate cutting planes to approximate the hypographs of quadratic functions. It is observed that the explicit hypograph structure $\eta \le q(\boldsymbol{x})$ is essential to trigger Gurobi's ability to generate such cuts.

Table 1: The average root gap, number of solved instances, and the average optimality gap over different types of DRCC-MB-KP problems with varying problem sizes.

(n, m)	Type	Root Gap		# Solved		Optimality Gap	
		CCP	SOC	CCP	SOC	CCP	SOC
(25, 25)	1	695.9%	**272.5%**	4/5	**5**/5	409.7%	**0.0%**
	2	450.2%	**113.7%**	3/5	**5**/5	22.7%	**0.0%**
	3	431.0%	**66.3%**	**5**/5	5/5	**0.0%**	**0.0%**
	4	303.0%	**43.2%**	**5**/5	5/5	**0.0%**	**0.0%**
(50, 50)	1	250.8%	**134.4%**	0/5	**4**/5	309.2%	**2.2%**
	2	427.4%	**273.3%**	0/5	**2**/5	411.6%	**3.3%**
	3	394.4%	**235.1%**	0/5	**2**/5	422.9%	**3.6%**
	4	260.8%	**109.4%**	2/5	**4**/5	14.2%	**1.0%**
(100, 100)	1	150.6%	**86.7%** (1)	0/5	0/5	160.2%	**63.5%**
	2	284.6%	**210.6%**	0/5	0/5	333.3%	**189.2%**
	3	294.8%	**210.9%** (3)	0/5	0/5	395.9%	**198.3%**
	4	229.4%	**175.6%**	0/5	0/5	263.1%	**154.1%**

The numbers in parentheses indicate the number of instances that terminate prematurely due to numerical issues.

We modify existing instances of deterministic (multi-dimensional) knapsack problems [24, 28] to construct our DRCC-MB-(M)KP instances. The construction of DRCC-MB-(M)KP instances and experiment settings are detailed in Appendix C and D, respectively.

We present the comparison of CCP and SOC on DRCC-MB-KP in Tables 1 and 2. For each type and size, we present the root gap, the number of instances solved to optimality, the ending optimality gap, the solution time, and the number of nodes explored before termination. The root gap is defined as the relative gap between the dual bound at the end of the root node (after cut generation) and the best objective value obtained by the two models within the time limit. Numbers in parentheses indicate the number of instances that terminate prematurely due to numerical issues. Each row is averaged over 5 instances.

As shown in Tables 1 and 2, SOC consistently yields tighter root bounds and smaller optimality gaps compared to CCP across all problem sizes and types. For small instances (i.e., with 50 items), SOC solves all instances within seconds, while CCP solves 17 out of 20 instances within an hour. On medium instances (i.e., with 100 items), SOC achieves optimality on more than half of the instances, whereas CCP generally fails to reach optimality, resulting in optimality gaps orders of magnitude larger than those of SOC. SOC demonstrates the ability to significantly tighten the optimality gap while exploring far fewer nodes. This efficiency stems from the automatic model lifting by linearization of the quadratic terms during the presolving phase and the generation of effective Gurobi cuts, primarily BQP cuts [27] and RLT cuts [33], which are absent in CCP. However, for

Table 2: The average times and the average numbers of explored nodes over different types of DRCC-MB-KP problems with varying problem sizes.

(n, m)	Type	Time (s)		# Nodes Explored	
		CCP	SOC	CCP	SOC
(25, 25)	1	787.6	**1.4**	9.7×10^4	6.6×10^2
	2	1479.4	**1.4**	1.2×10^5	8.8×10^2
	3	423.3	**5.3**	8.5×10^4	2.2×10^3
	4	744.2	**1.3**	7.9×10^4	8.3×10^2
(50, 50)	1	3600.2	**931.5**	3.4×10^5	1.1×10^4
	2	3600.1	**2562.7**	4.0×10^5	6.2×10^4
	3	3600.1	**2676.3**	4.2×10^5	5.7×10^4
	4	2196.8	**1172.3**	2.5×10^5	2.9×10^4
(100,100)	1	(1) 2882.1	3600.1	1.8×10^5	7.4×10^2
	2	3601.7	3600.1	2.5×10^5	6.1×10^2
	3	(3) 1441.0	3600.4	9.1×10^4	5.9×10^2
	4	3600.6	3600.1	3.5×10^5	5.9×10^2

The numbers in parentheses indicate the number of instances that terminate prematurely due to numerical issues.

large instances (i.e., with 200 items), the abundance of these cuts and auxiliary linearization variables can significantly slow down the branch-and-bound process (orders of magnitude branching nodes explored within the time limit), leading to much less competitive performance of SOC. Apart from solution efficiency, we observe that the reformulated model SOC exhibits better numerical stability, as CCP encountered four numerical issues while SOC encountered none.

Table 3 presents the computational results on seven DRCC-MB-MKP instances, where similar statistics are reported as Tables 1 and 2. The first three small-sized instances are solved to optimality by both methods within one second. For the relatively difficult instances (4 to 7), SOC demonstrates much better performance. It yields tighter root bounds and solves instances 4 to 6 to optimality with shorter time and fewer branching nodes needed to reach optimality. In instance 7, SOC terminates with an optimality gap 9.9%, significantly smaller than the gap 103.9% for CCP.

5   Conclusions

In this paper, we study a class of mixed-integer conic sets and established that, under mild assumptions, their convex hull can be fully characterized via the concave envelope of the nonlinear function appearing on the right-hand side of the conic constraint. Although explicitly deriving the convex hull is, in general, computationally challenging, the associated ideal formulation provides a foundation for strong relaxations through cutting-plane techniques. The effectiveness of these relaxations has been empirically validated in our numerical study. Building

Table 3: Root gaps, ending gaps, computation times, and number of branch-and-bound nodes over the seven DRCC-MB-MKP instances.

Instance	Root Gap		Optimality Gap		Time (s)		# Nodes Explored	
	CCP	SOC	CCP	SOC	CCP	SOC	CCP	SOC
1	688.3%	**0.0%**	*0.0%	*0.0%	0.0	0.0	1.7×10^1	1.0×10^0
2	**30.8%**	36.3%	*0.0%	*0.0%	0.1	0.3	1.2×10^2	1.2×10^2
3	20.5%	**11.8%**	*0.0%	*0.0%	0.3	0.5	7.6×10^2	2.1×10^2
4	29.6%	**17.2%**	*0.0%	*0.0%	166.6	**1.5**	4.0×10^4	4.9×10^2
5	3.9%	**2.0%**	*0.0%	*0.0%	31.4	**4.4**	1.9×10^4	1.2×10^3
6	99.1%	**20.9%**	140.2%	*0.0%	3600.2	967.1	2.7×10^5	7.0×10^4
7	70.2%	**19.8%**	103.9%	**9.9%**	3600.0	3600.0	1.8×10^5	9.6×10^4

* Indicates that the instance is solved to optimality.

on these findings, future research may focus on developing efficient procedures for generating cutting planes that approximate the convex envelope of square-root functions, particularly in the context of MIQCP, thereby enhancing solver performance on a broader class of mixed-integer nonlinear programs.

References

1. Aktürk, M.S., Atamtürk, A., Gürel, S.: A strong conic quadratic reformulation for machine-job assignment with controllable processing times. Operations Research Letters **37**(3), 187–191 (2009)
2. Anstreicher, K.M., Burer, S.: Quadratic optimization with switching variables: the convex hull for n= 2. Mathematical Programming **188**(2), 421–441 (2021)
3. Atamtürk, A., Gómez, A.: Submodularity in conic quadratic mixed 0–1 optimization. Operations Research **68**(2), 609–630 (2020)
4. Atamtürk, A., Narayanan, V.: Polymatroids and mean-risk minimization in discrete optimization. Operations Research Letters **36**(5), 618–622 (2008)
5. Atamtürk, A., Narayanan, V.: The submodular knapsack polytope. Discrete Optimization **6**(4), 333–344 (2009)
6. Atamtürk, A., Narayanan, V.: Conic mixed-integer rounding cuts. Mathematical Programming **122**(1), 1–20 (2010)
7. Atamtürk, A., Narayanan, V.: Lifting for conic mixed-integer programming. Mathematical Programming **126**(2), 351–363 (2011)
8. Bhatia, R., Davis, C.: A better bound on the variance. The American Mathematical Monthly **107**(4), 353–357 (2000)
9. Boros, E., Hammer, P.L.: Cut-polytopes, boolean quadric polytopes and nonnegative quadratic pseudo-boolean functions. Mathematics of Operations Research **18**(1), 245–253 (1993)
10. Burer, S., Kılınç-Karzan, F.: How to convexify the intersection of a second order cone and a nonconvex quadratic. Mathematical Programming **162**(1), 393–429 (2017)
11. Burer, S., Ye, Y.: Exact semidefinite formulations for a class of (random and nonrandom) nonconvex quadratic programs. Mathematical Programming **181**(1), 1–17 (2020)

12. Ceria, S., Soares, J.: Convex programming for disjunctive convex optimization. Mathematical Programming **86**(3), 595–614 (1999)
13. Çezik, M.T., Iyengar, G.: Cuts for mixed 0-1 conic programming. Mathematical Programming **104**(1), 179–202 (2005)
14. De Rosa, A., Khajavirad, A.: Explicit convex hull description of bivariate quadratic sets with indicator variables. Mathematical Programming pp. 1–43 (2024)
15. De Simone, C.: The cut polytope and the boolean quadric polytope. Discrete Mathematics **79**(1), 71–75 (1990)
16. Deng, Y., Shen, S., Denton, B.: Chance-constrained surgery planning under conditions of limited and ambiguous data. INFORMS Journal on Computing **31**(3), 559–575 (2019)
17. Edmonds, J.: Submodular functions, matroids, and certain polyhedra. In: Combinatorial Optimization—Eureka, You Shrink! Papers Dedicated to Jack Edmonds 5th International Workshop Aussois, France, March 5–9, 2001 Revised Papers. pp. 11–26. Springer (2003)
18. Frangioni, A., Gentile, C.: Perspective cuts for a class of convex 0–1 mixed integer programs. Mathematical Programming **106**(2), 225–236 (2006)
19. Ghaoui, L.E., Oks, M., Oustry, F.: Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. Operations Research **51**(4), 543–556 (2003)
20. Gómez, A.: Strong formulations for conic quadratic optimization with indicator variables. Mathematical Programming **188**(1), 193–226 (2021)
21. Günlük, O., Linderoth, J.: Perspective reformulations of mixed integer nonlinear programs with indicator variables. Mathematical programming **124**(1), 183–205 (2010)
22. Kılınç-Karzan, F., Küçükyavuz, S., Lee, D., Shafieezadeh-Abadeh, S.: Conic mixed-binary sets: Convex hull characterizations and applications. Operations Research **73**(1), 251–269 (2025)
23. Kılınç-Karzan, F., Yıldız, S.: Two-term disjunctions on the second-order cone. Mathematical Programming **154**(1), 463–491 (2015)
24. Martello, S., Pisinger, D., Toth, P.: Dynamic Programming and Strong Bounds for the 0-1 Knapsack Problem. Management Science **45**(3), 414–424 (1999). https://doi.org/10.1287/mnsc.45.3.414
25. Modaresi, S., Kılınç, M.R., Vielma, J.P.: Intersection cuts for nonlinear integer programming: convexification techniques for structured sets. Mathematical Programming **155**(1), 575–611 (2016)
26. Modaresi, S., Vielma, J.P.: Convex hull of two quadratic or a conic quadratic and a quadratic inequality. Mathematical Programming **164**(1), 383–409 (2017)
27. Padberg, M.: The boolean quadric polytope: some characteristics, facets and relatives. Mathematical Programming **45**(1), 139–172 (1989)
28. Petersen, C.C.: Computational Experience with Variants of the Balas Algorithm Applied to the Selection of R&D Projects. Management Science **13**(9), 736–750 (1967)
29. Rockafellar, R.T.: Convex Analysis, vol. 28. Princeton University Press (1997)
30. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, Grundlehren Der Mathematischen Wissenschaften, vol. 317. Springer, Berlin, Heidelberg (1998). https://doi.org/10.1007/978-3-642-02431-3
31. Shafiee, S., Kılınç-Karzan, F.: Constrained optimization of rank-one functions with indicator variables. Mathematical Programming **208**(1), 533–579 (2024)
32. Shen, S., Wang, J.: Stochastic modeling and approaches for managing energy footprints in cloud computing service. Service Science **6**(1), 15–33 (2014)

33. Sherali, H.D., Adams, W.P.: A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems, Nonconvex Optimization and Its Applications, vol. 31. Springer US, Boston, MA (1999). https://doi.org/10.1007/978-1-4757-4388-3
34. Stubbs, R.A., Mehrotra, S.: A branch-and-cut method for 0-1 mixed convex programming. Mathematical Programming **86**(3), 515–532 (1999)
35. Wang, A.L., Kılınç-Karzan, F.: On the tightness of sdp relaxations of qcqps. Mathematical Programming **193**(1), 33–73 (2022)
36. Wei, L., Atamtürk, A., Gómez, A., Küçükyavuz, S.: On the convex hull of convex quadratic optimization problems with indicators. Mathematical Programming **204**(1), 703–737 (2024)
37. Yu, J., Ahmed, S.: Polyhedral results for a class of cardinality constrained submodular minimization problems. Discrete Optimization **24**, 87–102 (2017)
38. Zhang, Y., Jiang, R., Shen, S.: Ambiguous Chance-Constrained Binary Programs under Mean-Covariance Information. SIAM Journal on Optimization **28**(4), 2922–2944 (Jan 2018). https://doi.org/10.1137/17M1158707

A   Omitted Proofs

Proof (of Lemma 1). By [29][Corollary 17.1.5], we have

$$\hat{f}(\boldsymbol{x}) = \sup \left\{ \sum_{k=1}^{n+1} \lambda_k f(\boldsymbol{x}^k) : \boldsymbol{\lambda} \in \Delta_n, (\boldsymbol{x}^k)_{k=1}^{n+1} \in \mathcal{X}^{n+1}, \sum_{k=1}^{n+1} \lambda_k \boldsymbol{x}^k = \boldsymbol{x} \right\}.$$

Note that the supremum is taken over a compact region, and the objective function is continuous in $(\boldsymbol{\lambda}, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^{n+1})$. The supremum is thus attained at a point $(\boldsymbol{\lambda}, \boldsymbol{x}^1, ..., \boldsymbol{x}^{n+1}) \in \Delta_n \times \mathcal{X}^{n+1}$, which satisfies $\boldsymbol{x} = \sum_{k=1}^{n+1} \lambda_k \boldsymbol{x}^k$. The first half of the statement can then be obtained by dropping points $\boldsymbol{x}^k$ associated with zero weights. The second half of the statement is equivalent to that the convex envelope of $-f(\cdot)$ over $\mathrm{conv}(\mathcal{X})$ is lower semi-continuous. Consider the extended real-valued extension $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of $-f(\cdot)$ to $\mathbb{R}^n$, for which $h(\boldsymbol{x}) = -f(\boldsymbol{x})$ if $\boldsymbol{x} \in \mathcal{X}$ and $h(\boldsymbol{x}) = +\infty$ otherwise. Note that $h(\cdot)$ is a proper, coercive (by compactness of $\mathcal{X}$), and lower semi-continuous function, with an epigraph identical to that of $-f(\cdot)$. By [30][Corollary 3.47], the convex envelope of $h(\cdot)$ is proper, coercive, and lower semi-continuous, and so is the convex envelope of $-f(\cdot)$ over $\mathrm{conv}(\mathcal{X})$, which completes the proof.   □

Proof (of Proposition 1). Concavity and upper semi-continuity of $\sqrt{\hat{q}(\cdot)}$ follow from concavity and monotonicity of $\sqrt{\cdot}$, and concavity and upper semi-continuity of $\hat{q}(\cdot)$. Then $\sqrt{\hat{q}(\cdot)}$ is a concave function majorizing $f(\cdot) = \sqrt{q(\cdot)}$ over $\mathcal{X}$, and thus majorizing the convave envelope $\hat{f}(\cdot)$ of $f(\cdot)$ over $\mathrm{conv}(\mathcal{X})$, which proves the left-hand side inequality of (9).

It then remains to prove the right-hand side inequality of (9). Given $\boldsymbol{x} \in \mathrm{conv}(\mathcal{X})$, by definition of the concave envelope, there exist $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^K \in \mathcal{X}$ and

$\boldsymbol{\lambda}^q, \boldsymbol{\lambda}^f \in \Delta^{K-1}$ such that

$$
\begin{aligned}
\sqrt{\hat{q}(\boldsymbol{x})} - \hat{f}(\boldsymbol{x}) &= \sqrt{\sum_{k=1}^{K} \lambda_k^q f(\boldsymbol{x}^k)^2} - \sum_{k=1}^{K} \lambda_k^f f(\boldsymbol{x}^k) \\
&= \sqrt{\sum_{k=1}^{K} \lambda_k^q f(\boldsymbol{x}^k)^2} - \max_{\lambda \in \Delta^{K-1}} \sum_{k=1}^{K} \lambda_k f(\boldsymbol{x}^k) \\
&\leq \max_{\lambda \in \Delta^{K-1}} \left(\sqrt{\sum_{k=1}^{K} \lambda_k f(\boldsymbol{x}^k)^2} - \sum_{k=1}^{K} \lambda_k f(\boldsymbol{x}^k) \right).
\end{aligned}
\tag{15}
$$

Given $\lambda \in \Delta^{K-1}$, let X denote a random variable which takes value $f(\boldsymbol{x}^k)$ with probability λ_k for $k = 1, \ldots, K$. Let μ denote its mean and σ^2 denote its variance, i.e.,

$$
\mu = \sum_{k=1}^{K} \lambda_k f(\boldsymbol{x}^k), \quad \sigma^2 = \sum_{k=1}^{K} \lambda_k f(\boldsymbol{x}^k)^2 - \mu^2.
\tag{16}
$$

Note that X is supported on $[L, U]$. Then by the Bhatia-Davis inequality [8], we have

$$
\sigma^2 \leq (U - \mu)(\mu - L).
\tag{17}
$$

Also note that $\mu \in [L, U]$. Combining (15)-(17), we have

$$
\sqrt{\hat{q}(\boldsymbol{x})} - \hat{f}(\boldsymbol{x}) \leq \max_{\mu \in [L, U]} \sqrt{(U - \mu)(\mu - L) + \mu^2} - \mu = \frac{(U - L)^2}{4(L + U)},
$$

where the last equality is obtained by maximizing the middle one-dimensional concave function of μ over $[L, U]$. □

B   SOC Reformulation of (8)

Since Q_{yy} is positive definite, there exists an invertable matrix $B \in \mathbb{R}^{m \times m}$ such that $Q_{yy} = B^\top B$. Then the inequality can be reformulated as

$$
\|A\boldsymbol{x} + B\boldsymbol{y} + \boldsymbol{d}\|_2 \leq f(x),
$$

where

$$
A = (B^\top)^{-1} Q_{xy}^\top, \quad \boldsymbol{d} = \frac{1}{2}(B^\top)^{-1} \boldsymbol{a}_y,
$$

$$
f(\boldsymbol{x}) = \sqrt{g(\boldsymbol{x}) + \boldsymbol{x}^\top (A^\top A - Q_{xx})\boldsymbol{x} + (2A^\top \boldsymbol{d} - \boldsymbol{a}_x)^\top \boldsymbol{x} + \|\boldsymbol{d}\|_2^2}.
$$

C   Instance Generation

C.1   Regular Knapsack

Following [24], we first generate four types of deterministic knapsack instances, which are characterized by different levels of correlation between weights and profits of N items. Within each type, the weight and the profit of each item is independently generated as follows:

1. Uncorrelated: The weight w_j and profit p_j are integers independently and uniformly distributed in $\{1, 2, \ldots, 10000\}$;
2. Weakly correlated: The weight w_j is an integer uniformly distributed in $\{1, 2, \ldots, 10000\}$, and the profit is defined as $p_j = \max\{w_j + \kappa_j, 1\}$, where κ_j is an integer uniformly distributed in $\{-1000, -999, \ldots, 1000\}$;
3. Strongly correlated: The weight w_j is an integer uniformly distributed in $\{1, 2, \ldots, 10000\}$, and $p_j = w_j + 1000$;
4. Inverse strongly correlated: The profit p_j is an integer uniformly distributed in $\{1, 2, \ldots, 10000\}$, and $w_j = p_j + 1000$.

For each instance type and each $N \in \{50, 100, 200\}$, 5 distinct instances are generated. Specifically, the capacity for the i-th instance (where $i = 1, ..., 5$) is set to $c = \sum_{j=1}^{N} w_j \cdot i/6$. The weights and profits for the five instances are generated independently, meaning they are not necessarily identical. Following the data generation, all weights and knapsack capacities are divided by 1000, to enhance the numerical stability.

To adapt these deterministic instances for DRCC-MB-KP, the following modifications are implemented. The last $N/2$ items are designated as continuous items, and their profits are divided by 5. The mean weight for each item is equal to its deterministic weight, i.e., $\boldsymbol{\mu} = \boldsymbol{w}$. The covariance matrix Σ is constructed as $\Sigma = U^{\top}\mathrm{diag}(\boldsymbol{w})\mathrm{diag}(\boldsymbol{w})U/4$, where U is an orthogonal matrix obtained from the decomposition of a randomly generated positive definite matrix. Finally, the knapsack capacity is multiplied by $3/2$, to accommodate a reasonable number of items, especially the discrete items, in the distributionally robust case.

C.2   Multi-Dimensional Knapsack

Seven multi-dimensional knapsack instances are adopted from [28] with the number of items $n \in \{6, 10, 15, 20, 28, 39, 50\}$. The number of knapsacks $|\mathcal{J}|$ is 10 for the first five instances and 5 for the last two. To each of these base instances, $m = \lceil n/2 \rceil$ continuous items are introduced. The unit weight for each continuous item is an integer sampled uniformly from $[w_{\min}/2, w_{\max}]$, where $w_{\min}$ (resp. $w_{\max}$) is the minimum (resp. maximum) weight of the discrete items in the base instance. The unit profit of each continuous item is an integer uniformly sampled from $[1, p_{\max}/10]$, where $p_{\max}$ is the maximum profit of discrete items in the base instance. To enhance numerical stability, the weights of all items (both discrete and continuous) are divided by 100. The mean weight in the stochastic model equals the scaled weight. The covariance matrix is constructed in the same

manner as in the regular knapsack case. Finally, the capacities of all knapsacks are divided by 10. We use 10 instead of 100 as the scaling factor to maintain sufficiently large capacities to accommodate a reasonable number of items, as the presence of multiple DRCCs can be quite restrictive.

D   Experiment Settings

All experiments were conducted on a Linux platform (Ubuntu 22.04.5 LTS) equipped with two Intel Xeon Platinum 8575C processors. Gurobi 12.0.1 was used to solve all optimization problems, with a time limit of 3600 seconds and up to 4 threads utilized. We set $\alpha = 0.5\%$ in all the experiments. Trivial initial solutions (setting x and y to 0) were used when solving SOC, as Gurobi sometimes fails to find feasible solutions for this model.