BCOL RESEARCH REPORT 17.01

Industrial Engineering & Operations Research
University of California, Berkeley, CA 94720-1777

Forthcoming in *Journal of Global Optimization*

LIFTED POLYMATROID INEQUALITIES FOR MEAN-RISK OPTIMIZATION WITH INDICATOR VARIABLES

ALPER ATAMTÜRK AND HYEMIN JEON

Abstract. We investigate a mixed $0-1$ conic quadratic optimization problem with indicator variables arising in mean-risk optimization. The indicator variables are often used to model non-convexities such as fixed charges or cardinality constraints. Observing that the problem reduces to a submodular function minimization for its binary restriction, we derive three classes of strong convex valid inequalities by lifting the polymatroid inequalities on the binary variables. Computational experiments demonstrate the effectiveness of the inequalities in strengthening the convex relaxations and, thereby, improving the solution times for mean-risk problems with fixed charges and cardinality constraints significantly.

Keywords: Risk, submodularity, polymatroid, conic integer optimization, valid inequalities.

March 2017; October 2017; May 2018

1. Introduction

Optimization problems with a conic quadratic objective arise often when modeling uncertainty with a mean-risk utility. We motivate such a model for an investment problem with a parametric Value-at-Risk (VaR) minimization objective. Given random variables $\ell_i,\ i \in N$, representing the uncertain loss in asset i, let y_i denote the amount invested in asset $i \in N$. Then, for small $\epsilon > 0$, minimizing the Value-at-Risk with confidence level $1-\epsilon$ is stated as

$$\text{(VaR)}\ \zeta(\epsilon) = \min \left\{ z : \mathbf{Prob}\left(\ell' y > z\right) \leq \epsilon, \quad y \in Y \right\},$$

where losses greater than $\zeta(\epsilon)$ occur with probability no more than ϵ. Here, Y represents the set of feasible investments. If ℓ_i's are independent normally distributed

A. Atamtürk: Industrial Engineering & Operations Research, University of California, Berkeley, CA 94720-1777. atamturk@berkeley.edu

H. Jeon: Industrial Engineering & Operations Research, University of California, Berkeley, CA 94720-1777. hyemin.jeon@berkeley.edu .

random variables with mean μ_i and variance σ_i^2, problem (VaR) is equivalent to the following mean-risk optimization problem:

$$\min \left\{ \mu' y + \Omega \sqrt{\sum_{i \in N} \sigma_i^2 y_i^2} : \ y \in Y \right\}, \tag{1}$$

where $\Omega = \Phi^{-1}(1-\epsilon)$ and Φ is the cumulative distribution function (c.d.f.) of the standard normal distribution [15]. If only the mean and variance of the distribution are known, one can write a robust version by letting $\Omega = \sqrt{(1-\epsilon)/\epsilon}$, which provides an upper bound on the worst-case VaR [13, 21]. Alternatively, if ℓ_i's are independent and symmetric with support $[u_i - \sigma_i, u_i + \sigma_i]$, then letting $\Omega = \sqrt{\ln(1/\epsilon)}$ gives an upper bound on on the worst-case VaR as well [11]. Hence, under different assumptions on the random variable ℓ, one arrives at different instances of the mean-risk model (1) with a conic quadratic objective. Ahmed [1] studies the complexity and tractability of various stochastic objectives for mean-risk optimization.

The objective of the mean-risk optimization problem (1) is a conic quadratic function in y, hence convex. If the feasible set Y is a tractable convex set as well, then (1) is an efficiently-solvable convex optimization problem [26]. In practice, though, most problems are accompanied with non-convex side constraints, such as a restriction on the maximum number of non-zero variables or fixed charges [3, 4, 7, 12, 14, 16, 18] that are needed to obtain more realistic and implementable solutions. To model such non-convexities it is convenient to introduce auxiliary binary variables $x_i, \ i \in N$, to indicate whether y_i is non-zero or not. The so-called *on-off constraints* $0 \leq y_i \leq u_i x_i$, where u_i is an upper bound on y_i, $i \in N$, model whether asset i is in the solution or not. By appropriately scaling y_i, we assume, without loss of generality, that $u_i = 1$ for all $i \in N$. The non-convexity introduced by the on-off constraints is a major challenge in solving practical mean-risk optimization problems. In order to address this difficulty, in this paper, we derive strong convex relaxations for the conic quadratic mixed-integer set with indicator variables:

$$F = \left\{ (x, y, z) \in \{0,1\}^N \times \mathbb{R}_+^N \times \mathbb{R}_+ : \sigma + \sum_{i \in N} a_i y_i^2 \leq z^2, \ \mathbf{0} \leq y \leq x \right\}, \tag{2}$$

where $\sigma \geq 0$ and $a_i > 0, \ i \in N$.

Problem (1) is a special case of the mean-risk optimization problem

$$\min \left\{ \mu' y + \Omega \sqrt{y' Q y} : y \in Y \right\} \tag{3}$$

with a positive semidefinite covariance matrix Q. By decomposing $Q = V + D$, where $V, D \succeq 0$ and D is a diagonal matrix, problem (3) is equivalently written as

$$\begin{aligned} \min \ & \mu y + \Omega z \\ \text{s.t.} \ & y' D y + s^2 \leq z^2 \\ & y' V y \leq s^2 \\ & y \in Y, \ z \in \mathbb{R}_+. \end{aligned}$$

Indeed, for high dimensional problems such a decomposition is readily available, as a low-rank factor covariance matrix V is estimated separately from the residual variance matrix D to avoid ill-conditioning [22]. Observe that the first constraint above is a conic quadratic with a diagonal matrix. Therefore, the valid inequalities derived here for the diagonal case can be applied more generally in the presence of correlations after constructing a suitable diagonal relaxation. We provide computational experiments on the application of the results for the general case with correlations as well.

Literature. Utilizing diagonal matrices is standard for constructing convex relaxations in binary quadratic optimization [5, 27]. In particular, for $x \in \{0,1\}^n$,

$$x'Qx \le z \iff x'(Q-D)x + \text{diag}(D)'x \le z$$

with a diagonal matrix D satisfying $Q - D \succeq 0$. This transformation is based on the ideal (convex hull) representation of the separable quadratic term $x'Dx$ as a linear term $\text{diag}(D)'x$ for $x \in \{0,1\}^n$.

A similar approach is also available for convex quadratic optimization with indicator variables. For $x \in \{0,1\}^n$ and $y \in \mathbb{R}^n$ s.t. $\mathbf{0} \le y \le x$, we have

$$y'Qy \le z \iff y'(Q-D)y + \text{diag}(D)'t \le z, \; y_i^2 \le x_i t_i$$

with $t \in \mathbb{R}^n_+$ [2, 23]. This transformation is based on the ideal representation of each quadratic term $D_{ii}y_i^2$ subject to on-off constraints as a linear term $D_{ii}t_i$ along with a rotated cone constraint $y_i^2 \le x_i t_i$. Decomposing Q for diagonalization is also studied for an effective application of linear perspective cuts [20].

For the conic quadratic constraint $\sqrt{x'Dx} \le z$, however, the terms are *not* separable even for the diagonal case, and simple transformations as in the quadratic cases above are not sufficient to arrive at an ideal convex reformulation. For the pure binary case, Atamtürk and Narayanan [9] exploit the submodularity of the underlying set function to describe its convex lower envelope via polymatroid inequalities. Atamtürk and Gómez [6] describe a variety of applications for this model and give strong valid inequalities for the mixed $0-1$ case without the on-off constraints. The ideal (convex hull) representation for the conic quadratic mixed $0-1$ set with indicator variables F remains an open question. We show, however, that exploiting the submodularity of the underlying set function for the $0-1$ restrictions is critical in deriving strong convex relaxations for F. Table 1 summarizes the results for the related sets described above. In addition, general conic mixed-integer cuts [8], lift-and-project cuts [17], disjunctive cuts [10, 24] are also applicable to the conic mixed-integer set F considered here.

Notation. Throughout, we denote by $\mathbf{0}$ the vector of zeroes, by $\mathbf{1}$ the vector of ones, and by e_i the ith unit vector. $N := \{1,2,\ldots,n\}$ and $[k] := \{1,2,\ldots,k\}$. For a vector $a \in \mathbb{R}^N$, let $a(S) = \sum_{i \in S} a_i$, $S \subseteq N$. We use $(\cdot)^+$ to denote $\max\{\cdot, 0\}$.

Outline. The remainder of the paper is organized as follows. In Section 2 we review the polymatroid inequalities for the binary restriction of the mean-risk problem and give a polynomial algorithm for an optimization problem over F. In Section 3 we

Table 1. Convex hull representations for $x \in \{0,1\}^n, y \in \mathbb{R}^n_+, z \in \mathbb{R}_+$.

	Separable Quadratic	Conic Quadratic
Pure $0-1$	$x'Dx \le z$: [5, 27]	$\sqrt{x'Dx} \le z$: [9]
Mixed $0-1$	$y'Dy \le z,\ \mathbf{0} \le y \le x$: [2, 23]	$\sqrt{y'Dy} \le z,\ \mathbf{0} \le y \le x$: ?

introduce three classes of convex valid inequalities for F that are obtained from binary restrictions of F through lifting the polymatroid inequalities. In Section 4 we present computational experiments performed for testing the effectiveness of the proposed inequalities in solving mean-risk optimization problems with on-off constraints. We conclude with a few final remarks in Section 5.

2. Preliminaries

2.1. **Polymatroid inequalities.** Given $\sigma \ge 0$ and $a_i > 0,\ i \in N$, consider the set

$$K_\sigma = \left\{ (x,z) \in \{0,1\}^N \times \mathbb{R}_+ : \sqrt{\sigma + \sum_{i \in N} a_i x_i} \le z \right\}. \tag{4}$$

Observe that K_0 is the binary restriction of F obtained by setting $y = x$. For a given permutation $((1),(2),\dots,(n))$ of N, let

$$\begin{aligned} \sigma_{(k)} &= a_{(k)} + \sigma_{(k-1)}, \text{ and } \sigma_{(0)} = \sigma, \\ \pi_{(k)} &= \sqrt{\sigma_{(k)}} - \sqrt{\sigma_{(k-1)}}, \end{aligned} \tag{5}$$

and define the *polymatroid inequality* as

$$\sum_{i=1}^n \pi_{(i)} x_{(i)} \le z - \sqrt{\sigma}. \tag{6}$$

Let Π_σ be the set of such coefficient vectors π for *all* permutations of N. The set function defining K_σ is non-decreasing submodular; therefore, Π_σ form the extreme points of a polymatroid [19] and the convex hull of K_σ is given by the set of all polymatroid inequalities [25].

Proposition 1 (Convex hull of K_σ)**.**

$$conv(K_\sigma) = \left\{ (x,z) \in [0,1]^N \times \mathbb{R}_+ : \pi' x \le z - \sqrt{\sigma}, \ \ \forall \pi \in \Pi_\sigma \right\}.$$

As shown by Edmonds [19], the maximization of a linear function over a polymatroid can be solved by the greedy algorithm; therefore, a point $(\bar{x}, \bar{z}) \in [0,1]^N \times \mathbb{R}_+$ can be separated from $\text{conv}(K_\sigma)$ via the greedy algorithm by sorting $\bar{x}_i,\ i \in N$ in non-increasing order in $O(n \log n)$ time.

Proposition 2 (Separation)**.** *A point* $(\bar{x}, \bar{z}) \notin conv(K_\sigma)$ *such that* $\bar{x}_{(1)} \ge \bar{x}_{(2)} \ge \dots \ge \bar{x}_{(n)}$ *is separated from* $\text{conv}(K_\sigma)$ *by inequality* (6).

Atamtürk and Narayanan [9] consider the mixed-integer version of K_σ:

$$L_\sigma = \left\{ (x, y, z) \in \{0,1\}^N \times [0,1]^M \times \mathbb{R}_+ : \sqrt{\sigma + \sum_{i \in N} a_i x_i + \sum_{i \in M} c_i y_i^2} \le z \right\},$$

where $c_i > 0, \; i \in M$ and give valid inequalities for L_σ based on the polymatroid inequalities. Without loss of generality, the upper bounds of the continuous variables in L_σ are set to one by scaling.

Proposition 3 (Valid inequalities for L_σ)**.** *For $T \subseteq M$ inequalities*

$$\pi' x + \sqrt{\sigma + \sum_{i \in T} c_i y_i^2} \le z, \quad \pi \in \Pi_{\sigma + c(T)} \tag{7}$$

are valid for L_σ.

Inequalities (7) are used to derive nonlinear valid inequalities for F in Section 3.

2.2. **Optimization.** In this section, we consider the optimization problem

$$\text{(OPT)} \qquad \min \left\{ c'x + d'y + \sqrt{\sigma + \sum_{i \in N} a_i y_i^2} : \mathbf{0} \le y \le x, \; x \in \{0,1\}^N, \; y \in \mathbb{R}_+^N \right\},$$

which will be useful in proving the validity of the inequalities for F. We characterize the optimal solutions and give a polynomial algorithm for (OPT). We assume that $\sigma \ge 0, \; a_i > 0, \; i \in N$ to ensure a real-valued objective. Without loss of generality, we assume that $c_i > 0, \; i \in N$, otherwise, we may set x_i to one; $d_i < 0, \; i \in N$, otherwise, we may set y_i to zero; and $c_i + d_i < 0, \; i \in N$, otherwise, we may set both x_i and y_i to zero.

Without loss of generality, assume that the variables are indexed so that

$$\frac{c_1 + d_1}{a_1} \le \frac{c_2 + d_2}{a_2} \le \cdots \le \frac{c_n + d_n}{a_n}.$$

The following proposition shows that the binary part of an optimal solution to (OPT) is a vector of consecutive ones, followed by consecutive zeroes.

Proposition 4. *If (x^*, y^*) is an optimal solution to (OPT), then $x_k^* = 1$ for some $k \in N$, implies $x_i^* = 1$ for all $i \in [k-1]$.*

Proof. Suppose for contradiction that $x_k^* = 1$, but $x_j^* = 0$ for some $j < k$. Consider two feasible points (x', y') and (x'', y'') with respective objective values z' and z'', constructed as:

$$(x', y') = (x^*, y^*) + (e_j, e_j),$$
$$(x'', y'') = (x^*, y^*) - (e_k, y_k^* e_k).$$

We will show that $z' < z^*$, contradicting the optimality of (x^*, y^*). To this end, let $\xi := \sigma + \sum_{i \in N} a_i y_i^{*2}$, and

$$\delta_1 := z^* - z'' = c_k + d_k y_k^* + \sqrt{\xi} - \sqrt{\xi - a_k y_k^{*2}},$$
$$\delta_2 := z' - z^* = c_j + d_j + \sqrt{\xi + a_j} - \sqrt{\xi}.$$

As (x'', y'') is a feasible solution, $\delta_1 \leq 0$. Also note that $y_k^* > 0$ as otherwise x_k^* would be zero in an optimal solution since $c_k > 0$. Now, we establish that

$$\frac{\delta_1}{a_k y_k^{*2}} - \frac{\delta_2}{a_j} = \left(\frac{c_k + d_k y_k^*}{a_k y_k^{*2}} - \frac{c_j + d_j}{a_j} \right) + \left(\frac{\sqrt{\xi} - \sqrt{\xi - a_k y_k^{*2}}}{a_k y_k^{*2}} - \frac{\sqrt{\xi + a_j} - \sqrt{\xi}}{a_j} \right) > 0,$$

from the inequality

$$\frac{c_k + d_k y_k^*}{a_k y_k^{*2}} \geq \frac{c_k + d_k}{a_k y_k^*} \geq \frac{c_j + d_j}{a_j},$$

which holds by the indexing assumption and that $0 < y_k^* \leq 1$, and from the inequality

$$\frac{\sqrt{\xi} - \sqrt{\xi - a_k y_k^{*2}}}{a_k y_k^{*2}} - \frac{\sqrt{\xi + a_j} - \sqrt{\xi}}{a_j} > 0,$$

which follows from the strict concavity of square root function. Therefore, we have $\frac{\delta_2}{a_j} < \frac{\delta_1}{a_k y_k^{*2}} \leq 0$, implying $\delta_2 < 0$, which contradicts the optimality of (x^*, y^*). □

Proposition 5. *There is an $O(n^2)$ time algorithm to solve (OPT).*

Proof. Proposition 4 implies that there exist only $n + 1$ possible candidates for optimal x, i.e., $\mathbf{0}$ and $\sum_{i=1}^{k} e_i$ for $k \in N$. After a single sort of the indices in $O(n \ log \ n)$ time, for each candidate x the resulting convex optimization problem in y can be solved in $O(n)$ time with Algorithm 1 in the Appendix . Therefore, an optimal solution to (OPT) can be found in $O(n^2)$ time. □

3. Lifted Polymatroid Inequalities

In this section, we derive three classes of valid inequalities for F by lifting the polymatroid inequalities (6) described in Section 2.1 from specific restrictions of the feasible set F. The first class of inequalities are linear, whereas the other two are nonlinear convex inequalities.

3.1. **Lifted Linear Polymatroid Inequalities.** Consider the restriction of F obtained by setting the continuous variables y to their binary upper bounds x. It follows from Section 2.1 that for any permutation ((1),(2), . . . , (n)) of N, the polymatroid inequality

$$\pi' x \leq z - \sqrt{\sigma} \tag{8}$$

with $\pi_{(i)} = \sqrt{\sigma_{(i)}} - \sqrt{\sigma_{(i-1)}},\ i = 1, 2, \ldots, n$, is valid for the restriction with $y = x$, but not necessarily for F. In this section, we lift inequality (8) to obtain the linear valid inequality

$$\pi' x \leq z + \alpha'(x - y) - \sqrt{\sigma}, \tag{9}$$

for F with coefficients $\alpha_{(i)} = a_{(i)} / \sqrt{\sigma_{(i)}},\ i = 1, 2, \ldots, n$.

Proposition 6. *Inequality* (9) *with α and π defined as above is valid for F.*

Proof. Consider the optimization problem over F:

$$\begin{aligned} \zeta = \max\ & \pi' x - \alpha'(x - y) - z + \sqrt{\sigma} \\ \text{s.t. }\ & \sigma + \sum_{i \in N} a_i y_i^2 \leq z^2 \\ & \mathbf{0} \leq y \leq x \\ & x \in \{0,1\}^N,\ y \in \mathbb{R}^N_+,\ z \in \mathbb{R}_+. \end{aligned}$$

Inequality (9) is valid for F iff $\zeta \leq 0$. By plugging in the values for π, α and eliminating z, the problem is equivalently written as

$$(V) \quad \begin{aligned} \zeta = \max\ & \sum_{i \in [n]} \left(\sqrt{\sigma_{(i)}} - \sqrt{\sigma_{(i-1)}} - \frac{a_{(i)}}{\sqrt{\sigma_{(i)}}} \right) x_{(i)} \\ & + \sum_{i \in [n]} \frac{a_{(i)}}{\sqrt{\sigma_{(i)}}} y_{(i)} - \sqrt{\sigma + \sum_{i \in [n]} a_{(i)} y_{(i)}^2} + \sqrt{\sigma} \\ \text{s.t. }\ & \mathbf{0} \leq y \leq x \\ & x \in \{0,1\}^N,\ y \in \mathbb{R}^N_+. \end{aligned}$$

Note that (V) is a special case of (OPT) with coefficients

$$c_{(i)} = -\left(\sqrt{\sigma_{(i)}} - \sqrt{\sigma_{(i-1)}} - \frac{a_{(i)}}{\sqrt{\sigma(i)}} \right), \text{ and } d_{(i)} = -\frac{a_{(i)}}{\sqrt{\sigma(i)}},\ i \in [n].$$

Then

$$\frac{c_{(i)} + d_{(i)}}{a_{(i)}} = -\frac{\sqrt{\sigma_{(i)}} - \sqrt{\sigma_{(i-1)}}}{a_{(i)}},\ i \in [n],$$

and, by concavity of the square root function, we have

$$\frac{c_{(i)} + d_{(i)}}{a_{(i)}} \leq \frac{c_{(j)} + d_{(j)}}{a_{(j)}}, \quad \text{for } i \leq j.$$

By Proposition 4, there exists an optimal solution (x^*, y^*) to (V) such that $x^* = \sum_{i=1}^m e_{(i)}$ for some $m \in [n]$. Then, y^* is an optimal solution to the following convex problem:

$$\max \sum_{i\in[m]} (\sqrt{\sigma_{(i)}} - \sqrt{\sigma_{(i-1)}}) - \sum_{i\in[m]} \frac{a_{(i)}}{\sqrt{\sigma_{(i)}}}(1-y_{(i)}) - \sqrt{\sigma + \sum_{i\in[m]} a_{(i)} y_{(i)}^2} + \sqrt{\sigma}$$
$$\text{s.t. } \mathbf{0} \le y \le \mathbf{1}.$$

This convex optimization problem over the continuous variables y is a special case of (COPT), considered in the Appendix, and its KKT conditions (following from (15a)–(15c)) are satisfied by (y, λ, μ) such that

$$y_i = 1, \; i \in [m],$$
$$\lambda_i = 0, \; i \in [m],$$
$$\mu_i = \frac{a_{(i)}}{\sqrt{\sigma_{(i)}}} - \frac{a_{(i)}}{\sqrt{\sigma_{(m)}}} \ge 0, \; i \in [m].$$

Therefore, there exists $(x^*, y^*) = (\sum_{i\in[m]} e_i, \sum_{i\in[m]} e_i)$ for some $m \in [n]$ with a binary y^*, implying $\zeta = 0$, i.e., the validity of (9). □

Remark 1. Observe that the proof of Proposition 6 implies that inequality (9) is tight for the following $n+1$ affinely independent points of F:

$$(x, y, z) = (\mathbf{0}, \mathbf{0}, \sqrt{\sigma});$$
$$(x, y, z) = \Big(\sum_{k\le i} e_{(k)}, \sum_{k\le i} e_{(k)}, \sqrt{\sigma_{(i)}}\Big), \quad i \in [n].$$

Example 1. Consider an instance of F with $a = [22, 18, 21, 19, 17]$, $\sigma = 0$ and the following fractional point is contained in its continuous relaxation:

$$\bar{x} = \bar{y} = \big[1, 0.3817, 0.6543, 0.3616, 0.8083\big], \bar{z} = 6.8705.$$

For the permutation $(1, 3, 5, 2, 4)$, π is computed as

$$\pi_{(1)} = \pi_1 = \sqrt{a_1} - \sqrt{0} = \sqrt{22} = 4.6904,$$
$$\pi_{(2)} = \pi_3 = \sqrt{a_1 + a_3} - \sqrt{a_1} = \sqrt{43} - \sqrt{22} = 1.8670,$$
$$\vdots$$
$$\pi_{(5)} = \pi_4 = \sqrt{a_1 + \cdots + a_5} - \sqrt{a_1 + a_2 + a_3 + a_5} = \sqrt{97} - \sqrt{78} = 1.0171.$$

The lifting coefficients α are computed accordingly, and we get inequality (9) with

$$\pi = \big[4.6904, 1.0858, 1.8670, 1.0171, 1.1885\big],$$
$$\alpha = \big[4.6904, 2.0381, 3.2025, 1.9292, 2.1947\big].$$

The fractional point $(\bar{x}, \bar{y}, \bar{z})$ is cut off by (9) as $\pi'\bar{x} - \alpha'(\bar{x} - \bar{y}) - \bar{z} = 0.7844 > 0$.

Although inequalities (9) cut off points of the continuous relaxation with fractional x, unlike for the binary case K_σ, adding all $n!$ inequalities (9) is not sufficient to describe conv(F) as conv(F) is not a polyhedral set. Therefore, in the next two subsections we present two nonlinear convex generalizations of inequalities (9).

3.2. **Lifted Nonlinear Polymatroid Inequalities I.** The second class of lifted inequalities is obtained by applying the procedure described in Section 3.1 for a subset of the variables. For $S \subseteq N$, introducing an auxiliary variable $t \in \mathbb{R}_+$, let us rewrite the conic constraint $\sum_{i\in N} a_i y_i^2 \le z^2$ as

$$t^2 + \sum_{i\in N\setminus S} a_i y_i^2 \le z^2,$$

$$\sigma + \sum_{i\in S} a_i y_i^2 \le t^2.$$

Applying Proposition 6 to the relaxation defined by constraints

$$\sigma + \sum_{i\in S} a_i y_i^2 \le t^2,\ \mathbf{0} \le y_S \le x_S,\ x_S \in \{0,1\}^S, y_S \in \mathbb{R}_+^S, t \in \mathbb{R}_+$$

for each permutation $((1),(2),\dots,(|S|))$ of S, we generate a lifted polymatroid inequality (9) of the form

$$\pi_S' x_S \le t + \alpha_S'(x_S - y_S) - \sqrt{\sigma}$$

where $\pi_{S(i)} = \sqrt{\sigma_{S(i)}} - \sqrt{\sigma_{S(i-1)}}$, $\alpha_{S(i)} = a_{(i)}/\sqrt{\sigma_{S(i)}}$, and the partial sums are defined as $\sigma_{S(i)} = a_{(i)} + \sigma_{S(i-1)}$ for $i = 1,2,\dots,|S|$ with $\sigma_{S(0)} = \sigma$. Eliminating the auxiliary nonnegative variable t, we obtain the following class of valid inequalities for F.

Proposition 7. *For $S \subseteq N$, the inequality*

$$\left((\pi_S' x_S + \sqrt{\sigma} - \alpha_S'(x_S - y_S))^+\right)^2 + \sum_{i\in N\setminus S} a_i y_i^2 \le z^2 \tag{10}$$

with π_S and α_S defined above is valid for F.

Note that inequality (10) is convex as it can be represented as conic quadratic by re-introducing the auxiliary variable $t \ge 0$. It is equivalent to (9) for $S = N$ and to the original constraint for $S = \emptyset$. Otherwise, it is distinct from both. It is differentiable at (x,y,z) with $\pi_S' x_S + \sqrt{\sigma} < \alpha_S'(x_S - y_S)$ or $\pi_S' x_S + \sqrt{\sigma} > \alpha_S'(x_S - y_S)$.

Remark 2. The following $n+1$ affinely independent points of F satisfy inequality (10) at equality:

$$(x,y,z) = (\mathbf{0},\mathbf{0},\sqrt{\sigma});$$

$$(x,y,z) = \Big(\sum_{k\le i} e_{(k)}, \sum_{k\le i} e_{(k)}, \sqrt{\sigma_{(i)}}\Big),\ i = 1,2,\dots,|S|;$$

$$(x,y,z) = (e_i, e_i, \sqrt{\sigma + a_i}),\ i \in N \setminus S.$$

The following example illustrates a point satisfying inequality (9), but cut off by inequality (10).

Example 2. Consider the instance in Example 1, and the fractional point

$$\bar{x} = \bar{y} = (1,0,0,0,0.8),\ \bar{z} = 5.7341.$$

This point satisfies inequality (9) generated in Example 1. Now letting $S = \{1, 2, 5\}$ and using the permutation (1,5,2), π_S is computed as

$$\begin{aligned}\pi_{S(1)} &= \pi_1 = \sqrt{a_1} - \sqrt{0} = \sqrt{22} = 4.6904, \\ \pi_{S(2)} &= \pi_5 = \sqrt{a_1 + a_5} - \sqrt{a_1} = \sqrt{39} - \sqrt{22} = 1.5546, \\ \pi_{S(3)} &= \pi_2 = \sqrt{a_1 + a_5 + a_2} - \sqrt{a_1 + a_5} = \sqrt{57} - \sqrt{39} = 1.3048.\end{aligned}$$

Consequently, we obtain inequality (10) with coefficients

$$\begin{aligned}\pi &= \left[4.6904, 1.3048, 0, 0, 1.5546\right], \\ \alpha &= \left[4.6904, 2.3842, 0, 0, 2.7222\right].\end{aligned}$$

Observe that the fractional point $(\bar{x}, \bar{y}, \bar{z})$ is cut off by inequality (10) as

$$\sqrt{\left((\pi'_S \bar{x} + \sqrt{\sigma} - \alpha'_S(\bar{x} - \bar{y}))^+\right)^2 + \sum_{i \in N \setminus S} a_i \bar{y}_i^2} - \bar{z} = 0.2 > 0.$$

3.3. **Lifted Nonlinear Polymatroid Inequalities II.** The third class of inequalities are derived from a partial restriction of F by setting a subset of the continuous variables to their upper bound. For $S \subseteq N$ and $T \subseteq N \setminus S$, consider the restriction of F with $y_i = x_i, \ i \in S$:

$$\begin{aligned}& t^2 + \sum_{i \in N \setminus (S \cup T)} a_i y_i^2 \leq z^2 \\ & \sigma + \sum_{i \in S} a_i x_i + \sum_{i \in T} a_i y_i^2 \leq t^2 \\ & y_i \leq x_i, \ i \in N \setminus S \\ & x \in \{0, 1\}^N, \ y \in \mathbb{R}_+^N, \ t \in \mathbb{R}_+.\end{aligned}$$

Applying the mixed-integer inequality (7) to the second constraint above, we obtain inequality

$$\pi'_S x_S + \sqrt{\sigma + \sum_{i \in T} a_i y_i^2} \leq t,$$

where $\pi_{S(i)} = \sqrt{\sigma_{S(i)}} - \sqrt{\sigma_{S(i-1)}}$ and $\sigma_{S(i)} = a_{(i)} + \sigma_{S(i-1)}$ for $i = 1, 2, \ldots, |S|$ with $\sigma_{S(0)} = \sigma + a(T)$. This inequality is valid for the restriction above, but not necessarily for F. Next, we lift it and eliminate the auxiliary nonnegative variable t, to obtain the third class of valid inequalities

$$\left(\left(\pi'_S x_S + \sqrt{\sigma + \sum_{i \in T} a_i y_i^2} - \alpha_S(x_S - y_S)\right)^+\right)^2 + \sum_{i \in N \setminus (S \cup T)} a_i y_i^2 \leq z^2, \tag{11}$$

for F with $\alpha_{(i)} = a_{(i)} / \sqrt{\sigma_{(i)}}, \ i = 1, 2, \ldots, |S|$.

Proposition 8. *Inequality* (11) *with* α_S *and* π_S *defined as above is valid for* F.

Proof. It suffices to prove the validity of inequality

$$\pi'_S x_S + \sqrt{\sigma + \sum_{i \in T} a_i y_i^2} - \alpha_S(x_S - y_S) \le t. \tag{12}$$

Consider the optimization problem:

$$\begin{aligned} \zeta = \max\ & \pi'_S x_S - \alpha'_S(x_S - y_S) + \sqrt{\sigma + \sum_{i \in T} a_i y_i^2} - t \\ \text{s.t.}\ & \sigma + \sum_{i \in N} a_i y_i^2 \le t^2 \\ & \mathbf{0} \le y \le x \\ & x \in \{0,1\}^N,\ y \in \mathbb{R}_+^N,\ t \in \mathbb{R}_+. \end{aligned}$$

Inequality (12) is valid for F iff $\zeta \le 0$. Observing that $x_i^* = y_i^* = 0$ for $i \in N \setminus (S \cup T)$ for an optimal solution (x^*, y^*) and eliminating t, the problem is written as

$$\begin{aligned} \zeta = \max\ & \pi'_S x_S - \alpha'_S(x_S - y_S) + \sqrt{\sigma + \sum_{i \in T} a_i y_i^2} - \sqrt{\sigma + \sum_{i \in S \cup T} a_i y_i^2} \\ \text{s.t.}\ & \mathbf{0} \le y \le x \\ & x \in \{0,1\}^N,\ y \in \mathbb{R}_+^N. \end{aligned}$$

Observe that by concavity of the square root function we have $y_i^* = 1$, $i \in T$. The validity of

$$\pi'_S x_S - \alpha_S(x_S - y_S) \le t - \sqrt{\sigma}$$

with $\bar{\sigma} = \sigma + a(T)$ and $t \ge \sqrt{\bar{\sigma} + \sum_{i \in S} a_i y_i^2}$ for this restriction implies that $\zeta \le 0$. □

Note that when $S = \emptyset$ and $T = N$, (11) is equivalent to the original constraint. When $S = N$, (11) is equivalent to (9). When $S \subseteq N$ and $T = \emptyset$, (11) is equivalent to (10). Otherwise, it is the distinct from the three.

Remark 3. The following $n + 1$ affinely independent points of F satisfy inequality (10) at equality:

$$(x, y, z) = (\mathbf{0}, \mathbf{0}, \sqrt{\sigma});$$

$$(x, y, z) = \Big(\sum_{k \le i} e_{(k)} + \sum_{j \in T} e_j, \sum_{k \le i} e_{(k)} + \sum_{j \in T} e_j, \sqrt{\sigma_{S(i)}}\Big),\ i = 1, 2, \ldots, |S|;$$

$$(x, y, z) = (e_i, e_i, \sqrt{\sigma + a_i}),\ i \in N \setminus S.$$

The following example illustrates inequality (11) cutting off a fractional point that is not cut by the previous inequalities.

Example 3. Consider again the instance in Example 1, and the fractional point

$$\bar{x} = \bar{y} = (0.8, 0.5, 1, 0, 1),\ \bar{z} = -0.1780.$$

Note that this point satisfies inequalities (9) and (10) generated in Example 1 and Example 2. Letting $S = \{1, 2\}$ and $T = \{3, 5\}$, we have $a(T) = a_3 + a_5 = 38$. For the permutation (1,2), π_S is computed as

$$\pi_{S(1)} = \pi_1 = \sqrt{a_1 + a(T)} - \sqrt{a(T)} = \sqrt{60} - \sqrt{38} = 1.5816,$$
$$\pi_{S(2)} = \pi_2 = \sqrt{a_1 + a_2 + a(T)} - \sqrt{a_1 + a(T)} = \sqrt{78} - \sqrt{60} = 1.0858$$

and we arrive at the corresponding inequality (11) with coefficients

$$\pi_S = \left[1.5816, 1.0858, 0, 0, 0\right],$$
$$\alpha_S = \left[2.8402, 2.0381, 0, 0, 0\right].$$

Observe the point $(\bar{x}, \bar{y}, \bar{z})$ is cut off by (11) as

$$\sqrt{\left(\left(\pi_S'\bar{x}_S + \sqrt{\sigma + \sum_{i \in T} a_i \bar{y}_i^2} - \alpha_S'(\bar{x}_S - \bar{y}_S)\right)^+\right)^2 + \sum_{i \in N \setminus (S \cup T)} a_i \bar{y}_i^2} - \bar{z} = 0.4506 > 0.$$

4. Computational Experiments

In this section, we report the result of computational experiments performed to test the effectiveness of inequalities (9), (10), and (11) in strengthening the continuous relaxation of mean-risk problems with on-off constraints. Three types of problems are used for testing: mean-risk problem with fixed-charges, mean-risk problem with a cardinality constraint, as well as the more general mean-risk problem with correlations and cardinality constraint.

All experiments are done using CPLEX 12.6.2 solver on a workstation with a 2.93GHz Intel R CoreTM i7 CPU and 8 GB main memory and with a single thread. The time limit is set to two hours and CPLEX' default settings are used with two exceptions: dynamic search is disabled to utilize the cut callbacks and the nodes are solved with the linear outer approximation for faster enumeration with node warm starts. The inequalities are added at nodes with depth less than ten.

Gradient cuts. Recall that inequalities (9) are linear; however, inequalities (10) and (11) are (convex) non-linear. Since only linear cuts can be added using CPLEX callbacks, at a differentiable point $(\bar{x}, \bar{y})$, instead of a nonlinear cut $f(x, y) \leq z$, we add the corresponding gradient cut

$$f(\bar{x}, \bar{y}) + [\nabla_x f(\bar{x}, \bar{y})]'(x - \bar{x}) + [\nabla_y f(\bar{x}, \bar{y})]'(y - \bar{y}) \leq z.$$

The gradient cut for inequality (10) at $(\bar{x}, \bar{y})$ with $\pi_S'\bar{x}_S + \sqrt{\sigma} > \alpha_S'(\bar{x}_S - \bar{y}_S)$ has the following form:

$$\kappa_1 + \frac{1}{f_1(\bar{x}, \bar{y})} \left[\tau_1(\bar{x}, \bar{y}) \left[\pi_S' x_S - \alpha_S'(x_S - y_S)\right] + \sum_{i \in N \setminus S} a_i \bar{y}_i y_i \right] \leq z, \tag{13}$$

where

$$f_1(x,y) = \sqrt{\tau_1(x_S,y_S)^2 + \sum_{i \in N \setminus S} a_i y_i^2},$$

$$\tau_1(x_S,y_S) = \pi'_S x_S + \sqrt{\sigma} - \alpha'_S(x_S - y_S),$$

and

$$\kappa_1 = f_1(\bar{x},\bar{y}) - \frac{\tau_1(\bar{x}_S,\bar{y}_S)\left[\tau_1(\bar{x}_S,\bar{y}_S) - \sqrt{\sigma}\right] + \sum_{i \in N \setminus S} a_i \bar{y_i}^2}{f_1(\bar{x},\bar{y})}.$$

Observe that κ_1 is a constant that equals to zero when $\sigma = 0$.

The gradient cut for inequality (11) at $(\bar{x},\bar{y})$ with $\pi'_S \bar{x}_S + \sqrt{\sigma + \sum_{i \in T} a_i \bar{y}_i^2} > \alpha'_S(\bar{x}_S - \bar{y}_S)$ has the form:

$$\kappa_2 + \frac{1}{f_2(\bar{x},\bar{y})}\left[\tau_2(\bar{x},\bar{y})\left[\pi'_S x_S - \alpha'_S(x_S - y_S) + \sum_{i \in T} \frac{a_i \bar{y}_i}{\nu(\bar{y})} y_i\right] + \sum_{i \in N \setminus (S \cup T)} a_i \bar{y}_i y_i\right] \leq z, \tag{14}$$

where

$$f_2(x,y) = \sqrt{\tau_2(x_S,y_{S\cup T})^2 + \sum_{i \in N \setminus (S \cup T)} a_i y_i^2},$$

$$\tau_2(x_S,y_{S\cup T}) = \pi'_S x_S + \nu(y_T) - \alpha'_S(x_S - y_S),$$

$$\nu(y_T) = \sqrt{\sigma + \sum_{i \in T} a_i y_i^2},$$

and

$$\kappa_2 = f_2(\bar{x},\bar{y}) - \frac{\tau_2(\bar{x}_S,\bar{y}_{S\cup T})\left[\pi'_S \bar{x}_S - \alpha_S(\bar{x}_S - \bar{y}_S) + \sum_{i \in T} \frac{a_i}{\nu(\bar{y}_T)} \bar{y}_i^2\right] + \sum_{i \in N \setminus (S \cup T)} a_i \bar{y}_i^2}{f_2(\bar{x},\bar{y})}.$$

Observe that κ_2 is a constant that equals to zero when $\sigma = 0$.

Separation. The separation problem for inequalities (6) and $\text{conv}(K_\sigma)$ is solved exactly and fast due to Edmond's greedy algorithm for optimization over polymatroids. We do not have such an exact separation algorithm for the lifted polymatroid inequalities and, therefore, use an inexact approach.

Given a point $(\bar{x},\bar{y},\bar{z})$, the separation for inequalities (9) and $\text{conv}(F)$ entails finding a permutation of N for which the violation is maximized. If $\bar{x} = \bar{y}$, as it is the case for optimal solutions of the continuous relaxation of (OPT) (see Appendix 5), inequality (9) coincides with the original polymatroid inequalitiy (6). Therefore, we check the violation of inequality (9) generated for a permutation $((1),\ldots,(n))$ satisfying $\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \ldots \geq \bar{x}_{(n)}$. If inequality (9) is violated, then it is added to the formulation. Otherwise, we attempt to find violated inequalities (10) and (11) for the same permutation.

For inequality (10), starting from $S = N$, we check for $i = (n), \ldots, (1)$ such that $\bar{x}_i - \bar{y}_i > 0$, whether moving i from S to $N \setminus S$ results in a violated inequality. If so, the corresponding gradient cut (13) is added to the formulation. Similarly, for inequality (11), starting from thus constructed S, we check for $i = (1), \ldots, (|S|)$ such that $\bar{x}_i - \bar{y}_i > 0$, whether moving i from S to T results in a violated inequality. If so, the corresponding gradient cut (14) is added to the formulation. This heuristic is repeated for two additional permutations of N: one such that $a_{(1)}\bar{x}_{(1)} \geq a_{(2)}\bar{x}_{(2)} \geq \cdots \geq a_{(n)}\bar{x}_{(n)}$, and the other such that $a_{(1)}/\bar{x}_{(1)} \geq a_{(2)}/\bar{x}_{(2)} \geq \cdots \geq a_{(n)}/\bar{x}_{(n)}$. Throughout the branch-and-bound algorithm, the entire cut generation process is applied up to $5,000$ times for the first permutation, and up to 500 times for the two additional permutations.

4.1. **Fixed-charge objective.** The first set of experiments are done on an optimization problem with fixed charges. Each non-zero y_i has a fixed-cost c_i, $i \in N$, which is modeled with cost vector c on the binary indicator variables x:

$$
(\mathrm{OPT}_f) \qquad
\begin{aligned}
\min\ & c'x + d'y + \Phi^{-1}(1-\epsilon)z \\
\text{s.t. } & \sum_{i \in N} a_i y_i^2 \leq z^2 \\
& \mathbf{0} \leq y \leq x \\
& x \in \{0,1\}^N, y \in \mathbb{R}_+^N, z \in \mathbb{R}_+.
\end{aligned}
$$

Five random instances are generated for each combination of confidence level $1 - \epsilon \in \{0.9, 0.95, 0.975\}$ and size $n \in \{100, 300, 500\}$. Coefficients $a_i,\ i \in N$, are drawn from integer uniform $[0.9n, 1.2n]$, $c_i,\ i \in N$, are drawn from integer uniform $[5, 20]$. Finally, for $i \in N$, d_i is set to $-c_i - h_i$, where h_i is drawn from integer uniform $[1, 4]$. The data used for the experiments is publicly available for download at `http://ieor.berkeley.edu/~atamturk/data/` .

We compare the original and the strengthened formulations in Table 2. Each row of the table presents the averages for five instances. We report the percentage integrality gap at the root node (rgap), solution time (time) in CPU seconds, the percentage gap between the best upper bound and lower bound at termination (egap), and the number of nodes explored (nodes). The number of cuts generated for each type is also reported. If there are instances not solved to optimality within the time limit, we report their number (#) next to egap.

One observes in Table 2 that the cuts have a profound effect in solving problem (OPT_f). With the default setting, only one of 60 instances is solved to optimality within two hours. The optimality gap reduces from 15.6% to 14.9% after exploring 142,016 nodes on average. On the other hand, when the cuts are added using the separation procedure outlined above, 50 of the 60 instances are solved at the root node without the need for enumeration. For the 10 instances that are not provably solved to optimality, the optimality gap is merely 0.2% compared to 21.6% with the default version for the same instances.

4.2. **Cardinality constraint.** The second problem type with binary indicator variables has a cardinality constraint on the maximum non-zero $y_i,\ i \in N$.

$$
(\mathrm{OPT}_c) \qquad
\begin{aligned}
\min\ & d'y + \ \Phi^{-1}(1-\epsilon)z \\
\text{s.t.}\ & \sum_{i \in N} a_i y_i^2 \le z^2 \\
& \sum_{i \in N} x_i \le \kappa n \\
& \mathbf{0} \le y \le x \\
& x \in \{0,1\}^N, y \in \mathbb{R}_+^N, z \in \mathbb{R}_+.
\end{aligned}
$$

Instances are tested with two cardinality levels ($\kappa = 0.2, 0.4$). Other parameters are generated as before.

The result of computations for (OPT_c) is summarized in Table 3. Although the root gap for this type of problem is smaller compared to the fixed-charge objective problem, only 27 out of 180 instances are solved to optimality using the default setting. When the cuts are utilized, the average root gap is reduced by 70% and 126 of the 180 instances are solved to optimality within the time limit. Accordingly, the average solution time as well as the number of nodes explored is reduced by orders of magnitude.

4.3. **Correlated case with cardinality constraint.** Finally, although the cuts are developed for the diagonal uncorrelated case, we test their effectiveness on the more general correlated case with a cardinality constraint. Using the reformulation introduced in Section 1, we state the problem as

TABLE 2. Computations with OPT_f.

		Default				With cuts						
n	$1-\epsilon$	rgap	time	egap (#)	nodes	rgap	time	egap (#)	nodes	cuts: (9)	(10)	(11)
100	0.9	3.0	6,099	1.4 (4)	237,470	0.0	0	0.0	0	14	0	0
	0.95	10.9	7,200	8.6 (5)	166,954	0.0	0	0.0	0	70	0	0
	0.975	30.7	7,200	26.2 (5)	226,365	0.0	0	0.0	0	82	15	0
300	0.9	4.0	7,200	3.8 (5)	125,005	0.0	1	0.0	0	49	0	0
	0.95	12.8	7,200	12.7 (5)	120,597	0.0	29	0.0	0	425	6	0
	0.975	31.4	7,200	31.4 (5)	134,433	0.0	23	0.0	0	437	36	1
500	0.9	3.9	7,200	3.8 (5)	108,684	0.0	5	0.0	0	83	0	0
	0.95	12.1	7,200	12.1 (5)	112,765	0.0	253	0.0	0	693	12	0
	0.975	31.3	7,200	31.3 (5)	177,891	0.0	913	0.0	0	1,119	158	2
1000	0.9	3.9	7,200	3.9(5)	87,363	0.0	48	0.0	0	197	0	0
	0.95	12.4	7,200	12.4(5)	96,394	0.1	7,200	0.1(5)	73	1,657	47	0
	0.975	30.9	7,200	30.9(5)	110,276	0.3	7,200	0.3(5)	2	1,631	25	0
avg		**15.6**	**7108**	**14.9**	**142016**	**0.0**	**1306**	**0.0**	**6**	**538**	**25**	**0**

Table 3. Computations with OPT_c.

			Default				With cuts						
n	κ	$1-\epsilon$	rgap	time	egap (#)	nodes	rgap	time	egap (#)	nodes	cuts: (9)	(10)	(11)
100	0.4	0.9	0.2	8	0.0	4,312	0.0	0	0.0	0	5	0	0
		0.95	0.4	148	0.0	23,884	0.0	0	0.0	0	16	0	0
		0.975	0.6	2,011	0.0	110,692	0.0	0	0.0	0	26	0	0
	0.2	0.9	1.1	718	0.0	104,967	0.0	0	0.0	0	58	0	0
		0.95	2.2	4,838	0.5(2)	357,476	0.0	1	0.0	1	116	0	0
		0.975	3.7	7,200	1.8(5)	417,491	0.2	2	0.0	2	225	0	0
	0.1	0.9	4.7	2,993	0.1(1)	247,352	1.8	2	0.0	12	326	0	0
		0.95	8.4	7,200	3.0(5)	461,193	3.8	8	0.0	26	695	0	0
		0.975	12.6	7,200	6.2(5)	512,712	6.1	26	0.0	56	1,614	0	0
300	0.4	0.9	0.4	7,200	0.3(5)	231,909	0.0	1	0.0	0	62	0	0
		0.95	0.7	7,200	0.6(5)	246,384	0.0	1	0.0	0	97	0	0
		0.975	1.0	7,200	1.0(5)	224,057	0.0	2	0.0	0	128	4	0
	0.2	0.9	1.9	7,200	1.7(5)	276,720	0.0	22	0.0	0	367	7	0
		0.95	3.2	7,200	3.1(5)	251,031	0.1	110	0.0	18	912	169	0
		0.975	4.8	7,200	4.6(5)	265,820	0.3	463	0.0	116	1,565	680	2
	0.1	0.9	5.9	7,200	5.3(5)	292,785	1.3	1,689	0.0(1)	10,338	2,779	461	1
		0.95	10.1	7,200	9.3(5)	274,213	3.2	6,193	0.5(4)	11,958	4,893	3,428	61
		0.975	14.7	7,200	13.9(5)	271,462	5.5	7,201	2.0(5)	5,235	5,000	3,604	70
500	0.4	0.9	0.5	7,200	0.4(5)	191,359	0.0	23	0.0	0	252	0	0
		0.95	0.8	7,200	0.8(5)	178,825	0.0	32	0.0	0	302	8	0
		0.975	1.1	7,200	1.1(5)	170,622	0.0	59	0.0	0	414	49	0
	0.2	0.9	2.0	7,200	2.0(5)	208,206	0.0	487	0.0	10	1,140	51	27
		0.95	3.4	7,200	3.3(5)	215,453	0.1	2,372	0.0 (1)	10,521	1,934	543	2
		0.975	4.9	7,200	4.9(5)	216,386	0.3	5,090	0.0 (2)	5,669	3,283	1,725	8
	0.1	0.9	6.2	7,200	6.1(5)	237,470	1.4	5,718	0.4(3)	1,113	4,520	2,606	2
		0.95	10.5	7,200	10.4(5)	232,484	3.5	7,200	2.3(5)	335	4,412	2,341	28
		0.975	15.2	7,200	15.1(5)	205,308	5.9	7,200	4.6(5)	646	4,250	2,092	5
1000	0.4	0.9	0.5	7,200	0.5(5)	102,180	0.0	311	0.0	0	509	43	0
		0.95	0.8	7,200	0.8(5)	99,684	0.0	550	0.0	0	762	0	0
		0.975	1.1	7,200	1.1(5)	116,666	0.0	1,265	0.0	0	1,094	0	0
	0.2	0.9	2.1	7,200	2.1(5)	142,802	0.0	6,792	0.0(3)	179	2,817	373	0
		0.95	3.5	7,200	3.5(5)	200,241	0.2	7,200	0.2(5)	28	2,680	368	0
		0.975	5.3	7,200	5.3(5)	187,518	0.8	7,200	0.8(5)	0	2,053	0	0
	0.1	0.9	6.8	7,200	6.8(5)	178,937	2.0	7,200	2.0(5)	0	1,938	0	0
		0.95	11.4	7,200	11.4(5)	205,205	4.5	7,200	4.5(5)	0	1,828	0	0
		0.975	16.6	7,200	16.6(5)	214,779	7.4	7,200	7.4(5)	0	1,915	0	0
avg			**4.6**	**6301**	**4.0**	**211091**	**1.4**	**2467**	**0.7**	**1285**	**1527**	**515**	**6**

$$
(\mathrm{OPT}_{corr}) \qquad
\begin{aligned}
\min\ & d'y + \Phi^{-1}(1-\epsilon) z \\
\text{s.t. } & y' V y \le s^2 \\
& s^2 + \sum_{i \in N} a_i y_i^2 \le z^2 \\
& \sum_{i \in N} x_i \le \kappa n \\
& \mathbf{0} \le y \le x \\
& x \in \{0,1\}^N,\ y \in \mathbb{R}_+^N, z \in \mathbb{R}_+.
\end{aligned}
$$

The covariance matrix $V \in \mathbb{R}^{n\times n}$ is computed using a factor model $V = \rho EFE'$, where $E \in \mathbb{R}^{n\times m}$ represents the exposures and $F \in \mathbb{R}^{m\times m}$ the factor covariance with $m = n/10$. We use a scaling parameter ρ to test the impact of the magnitude of correlations on the difficulty of the problem. Since the cuts are developed for the diagonal case, we expect them to perform well for small ρ. To ensure positive semidefiniteness, F is computed as $F = GG'$ for $G \in \mathbb{R}^{m\times m}$. Each G_{ij}, $i, j \in [m]$ is drawn from uniform $[-1, 1]$, and E_{ij}, $i \in [n]$, $j \in [m]$ is drawn from uniform $[0, 0.1]$ with probability 0.2 and set to 0 with probability 0.8. All other parameters are generated as before.

Table 4. Computations with OPT_{corr}.

			Default				With cuts						
n	ρ	$1-\epsilon$	rgap	time	egap (#)	nodes	rgap	time	egap (#)	nodes	cuts: (9)	(10)	(11)
100	0.1	0.9	1.2	3,039	0.1(2)	97,969	0.0	0	0.0	0	55	0	0
		0.95	2.3	6,814	0.7(4)	228,397	0.0	1	0.0	0	114	0	0
		0.975	3.7	7,200	2.2(5)	260,025	0.2	2	0.0	9	219	0	0
	1	0.9	1.1	3,028	0.1(2)	101,237	0.0	0	0.0	0	61	0	0
		0.95	2.3	7,200	0.8(5)	201,105	0.0	1	0.0	1	120	0	0
		0.975	3.6	7,200	2.2(5)	216,742	0.2	3	0.0	9	268	0	0
	10	0.9	1.1	3,017	0.1(2)	111,262	0.0	1	0.0	10	144	0	0
		0.95	2.2	7,200	0.74(5)	229,966	0.0	1	0.0	6	160	0	0
		0.975	3.5	7,200	2.16(5)	243,539	0.2	3	0.0	21	402	0	0
300	0.1	0.9	1.9	7,200	1.7(5)	163,249	0.0	72	0.0	33	695	125	0
		0.95	3.2	7,200	3.2(5)	154,394	0.1	271	0.0	65	1075	384	14
		0.975	4.8	7,200	4.7(5)	138,175	0.3	709	0.0	233	1921	991	10
	1	0.9	1.9	7,200	1.7(5)	129,379	0.0	2,034	0.0(1)	14,370	3,328	500	0
		0.95	3.2	7,200	3.2(5)	123,180	0.1	1,804	0.0(1)	8,926	2,099	810	35
		0.975	4.8	7,200	4.7(5)	132,002	0.3	2,579	0.0(1)	6,121	2,315	1,233	15
	10	0.9	1.8	7,200	1.6(5)	161,802	0.3	7,201	0.2(5)	31,917	3,893	485	1
		0.95	3.2	7,200	3.1(5)	143,816	0.3	7,201	0.2(5)	34,684	3,422	1,550	13
		0.975	4.8	7,200	4.6(5)	132,304	0.5	7,201	0.2(5)	37,645	3,639	2,592	23
avg			**2.8**	**6483**	**2.1**	**164919**	**0.1**	**1616**	**0.0**	**7447**	**1329**	**482**	**6**

Table 4 presents the results for confidence levels $1-\epsilon \in \{0.9, 0.95, 0.975\}$, problem sizes $n \in \{100, 300\}$, and scaling factors $\rho \in \{0.1, 1, 10\}$. As in the case of (OPT_c), the cuts result in significant improvements. Out of 90 instances, the number of unsolved instances is reduced from 80 to 18, and the average root gap is reduced by 96%. Especially, for instances with $\rho \in \{0.1, 1\}$, almost all instances are solved to optimality well within the time limit and the number of nodes is reduced by an order of magnitude. Even for $\rho = 10$, the end gap is reduced from 3.1% to only 0.2%. As expected, the computational results indicate that inequalities (9), (10), and (11) are more effective when the covariance matrix is more diagonal-dominant, i.e., for smaller values of ρ. The lifted polymatroid cuts are, nevertheless, valuable for the general correlated case as well.

5. Conclusion

In this paper we study a mixed 0-1 optimization with conic quadratic objective arising when modeling utilities with risk averseness. Exploiting the submodularity of the underlying set function for the binary restrictions, we derive three classes of strong convex valid inequalities by lifting of the polymatroid inequalities. Computational experiments demonstrate the effectiveness of the lifted inequalities in a cutting plane framework. The results indicate that the inequalities are very effective in strengthening the convex relaxations, and thereby, reducing the solution times problems with fixed charges and cardinality constraints substantially. Although the inequalities are derived for the diagonal case, they are also effective in improving the convex relaxations for the general correlated case.

Acknowledgement

This research is supported, in part, by grant FA9550-10-1-0168 from the Office of the Assistant Secretary of Defense for Research and Engineering.

References

[1] S. Ahmed. Convexity and decomposition of mean-risk stochastic programs. *Mathematical Programming*, 106:433–446, 2006.

[2] M. S. Aktürk, A. Atamtürk, and S. Gürel. A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Operations Research Letters*, 37:187–191, 2009.

[3] M. S. Aktürk, A. Atamtürk, and S. Gürel. Parallel machine match-up scheduling with manufacturing cost considerations. *Journal of Scheduling*, 13:95–110, 2010.

[4] M. S. Aktürk, A. Atamtürk, and S. Gürel. Aircraft rescheduling with cruise speed control. *Operations Research*, 62:829–845, 2014.

[5] K. M. Anstreicher. On convex relaxations for quadratically constrained quadratic programming. *Mathematical Programming*, 136:233–251, 2012.

[6] A. Atamtürk and A. Gómez. Submodularity in conic quadratic mixed 0-1 optimization. BCOL Research Report 16.02, IEOR, University of California–Berkeley, 2016. arXiv preprint arXiv:1705.05918.

[7] A. Atamtürk and O. Günlük. Network design arc set with variable upper bounds. *Networks*, 50:17–28, 2007.

[8] A. Atamtürk and V. Narayanan. Cuts for conic mixed integer programming. In M. Fischetti and D. P. Williamson, editors, *Proceedings of the 12th International IPCO Conference*, pages 16–29, 2007.

[9] A. Atamtürk and V. Narayanan. Polymatroids and risk minimization in discrete optimization. *Operations Research Letters*, 36:618–622, 2008.

[10] P. Belotti, J. C. Góez, I. Pólik, T. K. Ralphs, and T. Terlaky. A conic representation of the convex hull of disjunctive sets and conic cuts for integer second order cone optimization. In *Numerical Analysis and Optimization*, pages 1–35. Springer, 2015.

[11] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[12] H. Y. Benson and Ü. Sağlam. Mixed-integer second-order cone programming: A survey. In *INFORMS Tutorials in Operations Research*, pages 13–36. 2014.

[13] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15:780–804, 2005.

[14] D. Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, 74:121–140, 1996.
[15] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2011.
[16] P. Bonami and M. A. Lejeune. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations Research*, 57:650–670, 2009.
[17] M. T. Çezik and G. Iyengar. Cuts for mixed 0-1 conic programming. *Mathematical Programming*, 104:179–202, 2005.
[18] P. Damcı-Kurt, S. Küçükyavuz, D. Rajan, and A. Atamtürk. A polyhedral study of production ramping. *Mathematical Programming*, 158:175–205, 2016.
[19] J. Edmonds. Submodular functions, matroids, and certain polyhedra. In R. Guy *et al.*, editor, *Combinarotial Structures and Their Applications*, pages 69–87. Gordon and Breach, 1970.
[20] A. Frangioni and C. Gentile. SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Operations Research Letters*, 35:181–185, 2007.
[21] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51:543–556, 2003.
[22] R. C. Grinold and R. N. Kahn. *Active Portfolio Management*. McGraw-Hill, 1999.
[23] O. Günlük and J. Linderoth. Perspective relaxation of mixed integer nonlinear programs with indicator variables. *Mathematical Programming Series B*, 104:186–203, 2010.
[24] Fatma Kılınç-Karzan and Sercan Yıldız. Two-term disjunctions on the second-order cone. *Mathematical Programming*, 154:463–491, 2015.
[25] L. Lovász. Submodular functions and convexity. In Achim Bachem, Bernhard Korte, and Martin Grötschel, editors, *Mathematical Programming The State of the Art: Bonn 1982*, pages 235–257, Berlin, Heidelberg, 1983. Springer Berlin Heidelberg.
[26] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.
[27] S. Poljak and H. Wolkowicz. Convex relaxations of (0, 1)-quadratic programming. *Mathematics of Operations Research*, 20:550–561, 1995.

Appendix

In this appendix we characterize the solutions to the continuous relaxation of the optimization problem (OPT) and present an algorithm to solve it. Let $(\tilde{x}, \tilde{y})$ denote an optimal solution to the relaxation. As $c > 0$ it is clear that $\tilde{x}_i = \tilde{y}_i, i \in N$. Then letting $\tilde{c} = c + d$ (< 0 by assumption), we can reduce the continuous relaxation of (OPT) to the following convex optimization problem:

$$
\begin{aligned}
& \min \quad \tilde{c}'y + \sqrt{\sigma + \sum_{i \in N} a_i y_i^2} \\
\text{(COPT)} \quad & \text{s.t. } -y_i \leq 0, \ i \in N \qquad (\lambda_i) \\
& \qquad\ y_i \leq 1, \ i \in N. \qquad (\mu_i)
\end{aligned}
$$

Assume that the variables are indexed so that

$$\frac{\tilde{c}_1}{a_1} \leq \frac{\tilde{c}_2}{a_2} \leq \cdots \leq \frac{\tilde{c}_n}{a_n}.$$

Proposition 9. *If $\tilde{y}$ is an optimal solution to (COPT), then $\tilde{y}_i \geq \tilde{y}_k$, $1 \leq i < k \leq n$.*

Proof. Let λ, μ be the dual multipliers associated with the lower bound and upper bound constraints for y, respectively. Since (COPT) is a convex optimization problem with linear constraints, the KKT conditions

$$\tilde{c}_i + \frac{a_i}{\sqrt{\sigma + \sum_{j \in N} a_j y_j^2}} y_i - \lambda_i + \mu_i = 0, \ i \in N \tag{15a}$$

$$\begin{aligned} \mathbf{0} \leq y \leq \mathbf{1} \\ \lambda, \mu \geq \mathbf{0} \end{aligned} \tag{15b}$$

$$\begin{aligned} y_i \lambda_i = 0, \ i \in N \\ (1 - y_i)\mu_i = 0, \ i \in N \end{aligned} \tag{15c}$$

are necessary and sufficient for optimality.

By complementary slackness (c.s.), observe that λ_i and μ_i cannot be both positive. Therefore, there are three possible combinations of values for λ_i and μ_i, for $i \in N$:

1) $\lambda_i > 0, \ \mu_i = 0$: c.s. implies $y_i = 0$, which by (15a) implies $\lambda_i = \tilde{c}_i < 0$, violating $\lambda_i \geq 0$. Hence, this case is unattainable.
2) $\lambda_i = 0, \ \mu_i > 0$: c.s. implies $y_i = 1$. Then (15a) is written as
$$\mu_i = -\frac{a_i}{\sqrt{\sigma + \sum_{j \in N} a_j y_j^2}} - \tilde{c}_i,$$
which is dual feasible only if $-\frac{\tilde{c}_i}{a_i} \geq \frac{1}{\sqrt{\sigma + \sum_{j \in N} a_j y_j^2}}$.
3) $\lambda_i = \mu_i = 0$: In this case, (15a) reduces to
$$-\tilde{c}_i = \frac{a_i}{\sqrt{\sigma + \sum_{j \in N} a_j y_j^2}} y_i,$$
which is feasible only if $-\frac{\tilde{c}_i}{a_i} \leq \frac{1}{\sqrt{\sigma + \sum_{j \in N} a_j y_j^2}}$.

Hence, the assumption $\tilde{c} < 0$ implies $\tilde{y} > 0$ and, further, either one of the following holds for $i \in N$:

$$(a) \ \tilde{y}_i = 1 \text{ and } \frac{\tilde{c}_i}{a_i} \leq \frac{-1}{\sqrt{\tilde{\sigma}}},$$
$$(b) \ 0 < \tilde{y}_i < 1 \text{ and } \frac{\tilde{c}_i}{a_i} = \frac{-\tilde{y}_i}{\sqrt{\tilde{\sigma}}},$$

where $\tilde{\sigma} := \sigma + \sum_{i \in N} a_i \tilde{y}_i^2$. Then, if $\tilde{y}_k = 1$, since $\frac{\tilde{c}_i}{a_i} \leq \frac{\tilde{c}_k}{a_k} \leq \frac{-1}{\sqrt{\tilde{\sigma}}}$ for $i < k$, it follows that $\tilde{y}_i = 1$. If $\tilde{y}_k$ is fractional, for $i < k$

$$\tilde{y}_k = -\frac{\tilde{c}_k}{a_k}\sqrt{\tilde{\sigma}} \leq -\frac{\tilde{c}_i}{a_i}\sqrt{\tilde{\sigma}}$$

and, therefore, $\tilde{y}_k \leq \tilde{y}_i = \min\left\{1, -\frac{\tilde{c}_i}{a_i}\sqrt{\tilde{\sigma}}\right\}$. □

There are two simple special cases where we can generate a closed form solution for KKT points.

Remark 4. If $-\tilde{c}_i > \frac{a_i}{\sqrt{\sigma + \sum_{i \in N} a_i}}, \ i \in N$, then $\tilde{y}_i = 1, i \in N$.

Remark 5. If $\sigma + \sum_{i \in N} \frac{\tilde{c}_i^2}{a_i} = 1$ and $-\frac{\tilde{c}_i}{a_i} \leq 1,\ i \in N$, then $\tilde{y}_i = -\frac{\tilde{c}_i}{a_i},\ i \in N$.

In the remainder, we give an algorithm that constructs a KKT point for (COPT). Defining the sets $N_1 := \{i \in N : \tilde{y}_i = 1\}$ and $N_f := \{i \in N : 0 < \tilde{y}_i < 1\}$, we can express $\tilde{\sigma}$ as

$$\begin{aligned}
\tilde{\sigma} &= \sigma + \sum_{i \in N_1} a_i \tilde{y}_i^2 + \sum_{i \in N_f} a_i \tilde{y}_i^2 \\
&= \sigma + \sum_{i \in N_1} a_i + \sum_{i \in N_f} a_i \left(-\frac{\tilde{c}_i}{a_i} \sqrt{\tilde{\sigma}} \right)^2 \\
&= \sigma + \sum_{i \in N_1} a_i + \tilde{\sigma} \sum_{i \in N_f} \frac{\tilde{c}_i^2}{a_i}.
\end{aligned}$$

Therefore, given N_1 and N_f, one can compute

$$\begin{aligned}
\tilde{\sigma}(N_1, N_f) &= \frac{\sigma + \sum_{i \in N_1} a_i}{\left(1 - \sum_{i \in N_f} \frac{\tilde{c}_i^2}{a_i}\right)} \\
\tilde{y}_i &= -\frac{\tilde{c}_i}{a_i} \sqrt{\tilde{\sigma}(N_f, N_1)},\ i \in N_f, \\
\tilde{y}_i &= 1,\ i \in N_1.
\end{aligned}$$

Algorithm 1 describes how to construct N_1 and N_f. Initially, $N_f = \emptyset$ and $N_1 = N$, i.e., $y_i = 1$, for all $i \in N$. At each iteration of Algorithm 1, checks whether $\tilde{y}_p$ is fractional or one. Either p is moved from N_1 to N_f and the incumbent over-estimation $\tilde{\sigma}(N_1, N_f)$ on $\tilde{\sigma}$ is updated accordingly, or it is determined that $\tilde{y}_p = 1$ and the algorithm terminates as $\tilde{y}_i = 1,$ for all $i < p$ due to Proposition 9. Observe that if the indices satisfy non-decreasing order of $\tilde{c}_i/a_i$, Algorithm 1 runs in $O(n)$ time.

Algorithm 1: KKT point construction for (COPT).

`0. Initialize`

`Set` $N_f = \emptyset$, $N_1 = N$, $\tilde{\sigma} = \sigma + \sum_{i \in N} a_i$, $p = n$.

`1. Update`

if $-\frac{\tilde{c}_p}{a_p} \geq \frac{1}{\sqrt{\tilde{\sigma}}}$ *or* $p = 0$ **then**

  `go to step 2.`

else

  $N_f \leftarrow N_f \cup \{p\}$

  $N_1 \leftarrow N_1 \setminus \{p\}$

  $\tilde{\sigma} \leftarrow \tilde{\sigma}(N_1, N_f)$

  $p \leftarrow p - 1$

  `Repeat step 1.`

`2. Terminate`

`Return`

$\tilde{y}_i = -\frac{\tilde{c}_i}{a_i}\sqrt{\tilde{\sigma}},\ i \in N_f$

$\tilde{y}_i = 1,\ i \in N_1$