AN OUTER APPROXIMATION METHOD FOR SOLVING MIXED-INTEGER CONVEX QUADRATIC PROGRAMS WITH INDICATORS

LINCHUAN WEI, SIMGE KÜÇÜKYAVUZ

ABSTRACT. Mixed-integer convex quadratic programs with indicator variables (MIQP) encompass a wide range of applications, from statistical learning to energy, finance, and logistics. The outer approximation (OA) algorithm has been proven efficient in solving MIQP, and the key to the success of an OA algorithm is the strength of the cutting planes employed. In this paper, we propose a new technique for deriving cutting planes for MIQP from various convex relaxations, and, as a result, we develop new OA algorithms for solving MIQP at scale. The contributions of our work are two-fold: (1) we bridge the work on the convexification of MIQP and the algorithm design to solve large-scale problems, and (2) we demonstrate through a computational study on the sparse portfolio selection problem that our algorithms give rise to significant speedups compared with the state-of-the-art methods in the literature.

1. INTRODUCTION

Mixed-integer convex quadratic programs with indicators (MIQP) constitute a broad class of optimization problems formulated as:

$$\min_{x,y} \quad y^\top Q y + g^\top y + h^\top x \tag{1a}$$

$$\text{s.t.} \quad Ay \le b \tag{1b}$$

$$\text{(MIQP)} \qquad Cy \le Dx \tag{1c}$$

$$y_i(1 - x_i) = 0, \quad \forall i = 1, \dots, n \tag{1d}$$

$$x \in X \subseteq \{0,1\}^n, \tag{1e}$$

where $Q \in \mathbb{R}^{n \times n}$ is a positive definite Hessian matrix, and X denotes an arbitrary set over the binary variable vector x. The complementarity constraint (1d) encodes the relation that the indicator variable x_i activates/deactivates the associated continuous variable $y_i, i = 1, \dots, n$. We also consider arbitrary linear constraints over y and linear linking constraints over x and y. Examples of constraints (1b)–(1c) include the nonnegativity constraints ($y_i \ge 0$) and the semi-continuity constraints ($y_i = 0$ if $x_i = 0$, and $y_i \in [l_i, u_i]$ if $x_i = 1$, which can be modeled as $l_i x_i \le y_i \le u_i x_i$). Examples of X include the cardinality constraint, i.e., $X = X_k := \{x \in \{0,1\}^n : \sum_{i=1}^n x_i \le k\}$, for a given $k < n$. MIQP finds a wide range of applications in statistical learning with sparsity [5, 7, 32, 34, 35], portfolio optimization [3, 8], energy production [18], and logistics [15, 27]. On the other hand, MIQP is $\mathcal{NP}$-hard even when x and y are unconstrained [11].

1.1. **Background and Literature review.** By introducing an auxiliary variable η, we can write problem (1) equivalently as follows

$$\min \quad \eta$$

$$\text{s.t.} \quad \eta \ge y^\top Q y + g^\top y + h^\top x$$

$$Ay \le b$$

$$Cy \le Dx$$

$$y_i(1 - x_i) = 0, \quad \forall i = 1, \dots, n$$

$$x \in X \subseteq \{0,1\}^n,$$

L. Wei: Department of Industrial Engineering and Management Sciences, Northwestern University, linchuanwei2022@u.northwestern.edu.

S. Küçükyavuz: Department of Industrial Engineering and Management Sciences, Northwestern University, simge@northwestern.edu.

and we define an epigraph accordingly

$$\mathcal{Z} = \{(x, y, \eta) \in X \times \mathbb{R}^{n+1} \mid \eta \geq y^\top Q y + g^\top y + h^\top x,$$
$$Ay \leq b, Cy \leq Dx, y_i(1 - x_i) = 0, \ \forall i = 1, \ldots, n\}.$$

Note that minimizing η over $(x, y, \eta) \in \mathcal{Z}$ for some x and y is equivalent to minimizing over the closure of the convex hull of $\mathcal{Z}$ denoted as cl conv($\mathcal{Z}$). In recent years, there has been a growing interest in studying the convexification of $\mathcal{Z}$ via decomposition. For example, the Hessian matrix Q can be decomposed into $\sum_i \Gamma_i + R$, where each Γ_i exhibits some 'simple' structure. Then, a convex relaxation of $\mathcal{Z}$ can be derived by studying the mixed-integer quadratic epigraphs where each Γ_i is the Hessian of the associated quadratic term. Specifically, the perspective reformulation [10, 18] is obtained when we decompose the quadratic function into a sum of one-dimensional quadratic terms and a remainder ($Q = \mathrm{diag}(\delta) + R$ such that $\delta \geq 0$ and $R \succeq 0$) and replace $\delta_i y_i^2$ with $\delta_i \frac{y_i^2}{x_i}$, and $\frac{y_i^2}{x_i}$ is known as the perspective function of y_i^2 [30]. Other convexification work considers decomposing Q into rank-one matrices [2, 26, 40, 41], two-by-two matrices [21, 25, 28, 31], and tridiagonal matrices [33]. In principle, cl conv($\mathcal{Z}$) can be obtained directly using the disjunctive programming approach [10, 21]; however, a straightforward disjunctive formulation requires a copy of variables for each configuration of indicators in X, which results in an exponential number of auxiliary variables. Wei et al. [42] give a more compact description of cl conv($\mathcal{Z}$) in the absence of the linear constraints (1b)–(1c). Another related strand of research concerns the convexification of the lifted set $S = \{(x, y, y^\top y) \in \{0, 1\}^n \times \mathbb{R}^n \times \mathbb{R}^{n \times n} : y_i(1 - x_i) = 0, \forall i = 1, \ldots, n\}$, and S stems from the more general mixed-integer quadratic programs with indicators. Anstreicher and Burer [1] deliver a convex hull description for S with bound constraints $0 \leq y \leq x$ when $n = 2$. De Rosa and Khajavirad [12] give the convex hull of S with nonnegativity constraints for $n = 2$. Although there are many strong theoretical results on the convexification of MIQP, a stronger reformulation usually involves a larger set of auxiliary variables, more constraints, or more nonlinearity in the objective or constraints, which may hinder the branch-and-bound process.

Regarding the exact solution methods for MIQP, substantial research emerged in recent years which can be categorized into (i) branch-and-bound methods that solve a sequence of subproblems and (ii) outer-approximation (OA) algorithms that recursively solve polyhedral approximations of $\mathcal{Z}$.

Branch-and-bound: In the literature, various branch-and-bound methods using different reformulations have been proposed for solving MIQP (or its special cases) to optimality. Bienstock [8] first proposed solving MIQP via branch-and-bound, and, instead of using indicators, their algorithm directly branches on the continuous variables by enforcing bound constraints $y_i \leq 0$ or $y_i \geq \alpha_i$. Bertsimas et al. [7] modeled the complementarity constraints (1d) using Specially Ordered Sets of Type 1 (SOS-1) [6] and solved the best subset selection problem by branch-and-bound. A classic modeling practice is the so-called big-M reformulation [23], which replaces the nonconvex complementarity constraints (1d) with the constraints $-Mx_i \leq y_i \leq Mx_i, \ \forall i = 1, \ldots, n$. The big-M value is chosen so that the true optimal solution is not eliminated; on the other hand, the strength of the big-M reformulation relies heavily on the value of M, and a loose M will result in a weak relaxation of MIQP. MIQP formulated by big-M constraints can be equivalently written as mixed-integer second-order conic programs. Vielma et al. [39] solve the resulting formulations using lifted polyhedral approximations for second-order cones. Another line of branch-and-bound methods builds upon the perspective reformulation. Frangioni and Gentile [18] solve MIQP using perspective reformulation and a nonlinear branch-and-bound scheme. Subsequently, Frangioni and Gentile [19] and Zheng et al. [44] extend the work of Frangioni and Gentile [18] by designing different approaches for diagonal decomposition, and [20] propose an approximate perspective reformulation through a project-and-lift approach. Another line of work accelerates the branch-and-bound procedure through specialized nonlinear programming algorithms with 'warm start' capability [4, 29].

Outer approximation: The OA algorithm [14, 16] solves general mixed-integer nonlinear programming (MINLP) programs to optimality. It consists of recursively solving a master mixed-integer linear (MILP) approximation of MINLP and nonlinear subproblems to generate cutting planes. The advantages of OA methods compared with nonlinear branch-and-bound methods are (i) the reduced effort of solving linear programming relaxations than nonlinear programming relaxations, (ii) the ability to utilize the 'warm start' capability of linear programming solvers, the heuristics of MILP solvers, and

MILP cuts. On the other hand, the strength of the cutting planes employed is crucial to the success of OA methods. In early work [9, 17], the computational results show that nonlinear branch-and-bound outperforms OA for MIQP. The initiative to redesign OA to solve MIQP faster was taken by Fischetti et al. [15], in which an OA algorithm using the cutting planes of a perspective reformulation of the quadratic uncapacitated facility location problem [27] was proposed. Since then, significant advances have been accelerating the OA algorithm to solve MIQP. For example, Bertsimas and Van Parys [5] propose an OA algorithm based on the perspective reformulation of the best subset selection problem with an ℓ_2-norm penalty. Bertsimas and Cory-Wright [3] extend the work of Bertsimas and Van Parys [5] and give an OA algorithm for solving MIQP in its general form; however, their algorithm also alters the original problem by introducing an ℓ_2-norm penalty; and the cutting planes there are generated by solving a quadratic programming (QP) problem of size n, which could be computationally expensive. Friedrich and Kreber [22] remedy this issue by replacing constraints (1b) with a quadratic penalty term into the objective, which allows the use of the more efficient cut generation scheme in [5] (for unconstrained MIQP), but their approach does not handle the linear linking constraints (1c). The cutting planes used in [3, 5, 15, 22] are 'projective' cuts in the (x, η) space, and they are derived from the subgradient cuts of the marginal function of a convex envelope of $\mathcal{Z}$, i.e., a closed convex function f_{env} such that $\mathcal{Z} \subset \{(x, y, \eta) : \eta \geq f_{\text{env}}(x, y)\} = \text{epi}(f_{\text{env}})$. As a result, the master problem in the OA algorithm is a mixed-integer linear programming problem over the binary vector x, and an auxiliary continuous variable η capturing the objective function value. The OA approach using projective cutting planes has several algorithmic advantages. First, the problem size is reduced from $2n$ to n. Second, MIQP problems often have cardinality constraints on x, and such sparsity can be leveraged by commercial MILP solvers.

In the present paper, we fill the gap between the convexification theory of MIQP and the design of algorithms to solve MIQP. We develop a unifying framework to derive projective cutting planes in (x, η) space based on different reformulations of MIQP. As a special case, we give a formula for computing the projective cutting planes for the perspective reformulation of MIQP. Our formula is more efficient than [5] and can handle MIQP with linear linking constraints (1c).

1.2. **Structure.** The rest of this paper is organized as follows:
• In §3, we first lay the theoretical foundation for our approach by appealing to variational analysis. Later, we derive cutting planes for the perspective reformulation of MIQP. A key observation is that the cutting planes can be computed by solving a QP (with a size much smaller than n when a strong sparsity constraint is enforced) plus a few linear algebraic operations.
• In §4, we derive cutting planes based on the perspective reformulation strengthened by the conic-quadratic inequalities of rank-one quadratics [2]. In addition, we show rigorously that we can obtain a class of stronger cutting planes with rank-one strengthening.
• In §5, we design an OA procedure with the cutting planes derived in §3 and §4 and apply it to the sparse portfolio selection problem studied in [3, 18]. The computational results reveal that our algorithms lead to significant speedups compared with other state-of-the-art methods in the literature.

2. NOTATION

We denote the vector of all ones by $\mathbf{1}$. For a matrix H, we use the upper case H_i to denote the ith column of H and lower case h_i to denote the ith row of H. We use $\succeq$ to denote the partial order in the space of symmetric matrices defined via the cone of positive semidefinite matrices. For a convex set $\mathcal{C}$, we adopt the conventional notations $\text{cl}(\mathcal{C})$ and $\text{ri}(\mathcal{C})$ to denote its closure and relative interior. The closure of the convex hull of $\mathcal{C}$ is written as $\text{cl conv}(\mathcal{C})$. For a convex function $f(x)$, its epigraph is $\text{epi}(f) = \{(x, \eta) : \eta \geq f(x)\}$, and its Fenchel conjugate is $f^*(\mu) = \sup_x \{\langle \mu, x \rangle - f(x)\}$. The effective domain of f is denoted as $\text{dom}(f) = \{x \mid f(x) < +\infty\}$. The subdifferential set of f at x is written as $\partial f(x)$. We define the indicator function of set $\mathcal{C}$ as $\mathcal{I}_\mathcal{C}(x) = 0$ if $x \in \mathcal{C}$ and $+\infty$ otherwise. We define the projection operator $\text{Proj}_{x, \eta}(\mathcal{C})$ as $\text{Proj}_{x, \eta}(\mathcal{C}) = \{(x, \eta) : (x, y, \eta) \in \mathcal{C} \text{ for some } y\}$.

3. CUTTING PLANES FOR PERSPECTIVE REFORMULATION

For an arbitrary closed convex envelope f_{env} of $\mathcal{Z}$, we have, by definition, $\mathcal{Z} \subset \text{epi}(f_{\text{env}})$. It is known from convex analysis that the marginal function $\bar{f}_{\text{env}}$ obtained by partially minimizing over $y \in \mathbb{R}^n$, i.e.,

$$\bar{f}_{\text{env}}(x) = \inf_y f_{\text{env}}(x, y),$$

is also convex.

Lemma 1. *For any $t \in \partial \bar{f}_{env}(x)$, the following inequality*

$$\tilde{\eta} \geq \langle t, \tilde{x} - x \rangle + \bar{f}_{env}(x) \tag{2}$$

holds for any $(\tilde{x}, \tilde{y}, \tilde{\eta}) \in cl\ conv(\mathcal{Z})$.

Proof. Because $\bar{f}_{\mathrm{env}}$ is a convex function, the subgradient inequality (2) is valid for $\mathrm{epi}(\bar{f}_{\mathrm{env}})$. Moreover, we have the sequence of inclusion relations that $\mathrm{Proj}_{x,\eta}(\mathrm{cl\ conv}(\mathcal{Z})) \subset \mathrm{Proj}_{x,\eta}(\mathrm{epi}(f_{\mathrm{env}})) \subset \mathrm{epi}(\bar{f}_{\mathrm{env}})$. The first inclusion follows from the fact that $\mathcal{Z} \subset \mathrm{epi}(f_{\mathrm{env}})$ and $\mathrm{cl\ conv}(\mathcal{Z}) \subset \mathrm{epi}(f_{\mathrm{env}})$ because $\mathrm{epi}(f_{\mathrm{env}})$ is a convex set. For any y, $\eta \geq f_{\mathrm{env}}(x,y) \geq \bar{f}_{\mathrm{env}}(x) = \inf_{y \in \mathbb{R}^n} f_{\mathrm{env}}(x,y)$, by definition, thus the second inclusion holds. Based on the inclusion relations, we can see that the subgradient inequality (2) is valid for $\mathrm{Proj}_{x,\eta}(\mathrm{cl\ conv}(\mathcal{Z}))$, and we write the projection in an explicit form as $\mathrm{Proj}_{x,\eta}(\mathrm{cl\ conv}(\mathcal{Z})) = \{(x,\eta) : (x,y,\eta) \in \mathrm{cl\ conv}(\mathcal{Z})$ for some $y \in \mathbb{R}^n\}$. Now, it becomes clear that the subgradient inequality is also valid for $\mathrm{cl\ conv}(\mathcal{Z})$. $\qquad\square$

One insight that can be drawn from Lemma 1 is that we can construct projective cutting planes for $\mathcal{Z}$ using the subgradients of $\bar{f}_{\mathrm{env}}$. Our goal is to design a unifying procedure for computing the subgradients regardless of the explicit form of $\bar{f}_{\mathrm{env}}$, so that we can derive a rich family of projective cutting planes via various convex envelopes available in the literature. Toward this end, we will borrow tools from variational analysis.

In what follows, for a proper convex function f whose definition is not clear over the relative boundary of $\mathrm{dom}(f)$, we assume that we naturally extend it to $\mathrm{cl}(\mathrm{dom}(f))$ by the limiting argument

$$f(x) = \lim_{\lambda \to 0^+} f(x + \lambda(\tilde{x} - x)) \qquad \text{for some } \tilde{x} \in \mathrm{ri}(\mathrm{dom}(f)).$$

For example, for $\phi(x,y) = \frac{y^2}{x}$,

$$\phi(x,y) = \begin{cases} \frac{y^2}{x} & \text{if } x > 0 \\ 0 & \text{if } y = x = 0 \\ +\infty & \text{otherwise.} \end{cases}$$

We also assume that problem (1) is feasible.

Assumption 1. *There exists $(\tilde{x}, \tilde{y})$ satisfying constraints (1b)–(1e).*

Checking the feasibility of the system (1b)–(1e) is $\mathcal{NP}$-hard in general because it includes the feasibility problem of a mixed-integer linear set as a special case. However, for many interesting applications, i.e., sparse regression, best subset selection, and sparse portfolio selection, it is without loss of generality to assume that the MIQP problem always has a feasible point.

Lemma 2. *For $f(x,y) = \frac{y^2}{x} : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^+$, its subdifferential $\partial f(x,y)$ is as follows*

$$\partial f(x,y) = \begin{cases} \{(\Phi, \Psi) \mid \Phi \leq -\frac{1}{4}\Psi^2\} & \text{if } x = 0 \text{ and } y = 0 \\ (-\frac{y^2}{x^2}, \frac{2y}{x},) & \text{if } x > 0 \\ \emptyset & \text{otherwise.} \end{cases}$$

Proof. The result follows from the Fenchel-Young inequality [37]:

$$f(x,y) + f^*(\Phi, \Psi) \geq x\Phi + y\psi$$

and the inequality holds at equality if and only if

$$(\Phi, \Psi) \in \partial f(x,y).$$

The conjugate function $f^*(\Phi, \Psi)$, by definition, equals

$$f^*(\Phi, \Psi) = \sup_{x,y} \left\{ \Phi x + \Psi y - \frac{y^2}{x} \right\} = \mathcal{I}_{\{(\Phi,\Psi) \mid \Phi \leq -\frac{1}{4}\Psi^2\}}(\Phi, \Psi).$$

Note that the perspective function f is differentiable in its relative interior $x > 0$, and the subdifferential collapses to a single point, and when $x = 0$ but $y \neq 0$, $f(0,y) = +\infty$, so no subgradient exists. The only nontrivial case is when $x = 0$ and $y = 0$. By Fenchel-Young inequality, we know that $(\Phi, \Psi) \in \partial f(0,0)$ if and only if

$$\mathcal{I}_{\{(\Phi,\Psi) \mid \Phi \leq -\frac{1}{4}\Psi^2\}}(\Phi, \Psi) = 0 \iff \Phi \leq -\frac{1}{4}\Psi^2.$$

☐

Before we proceed with deriving the first class of cutting planes for $\mathcal{Z}$, we recall a result regarding the subdifferential of a marginal function from variational analysis, which is our main tool for computing projective cutting planes.

Theorem 1. *[38] For any proper closed convex function $f(x,y) : \mathbb{R}^{2n} \to \mathbb{R} \bigcup \{\infty\}$, suppose that there exists $\bar{x}, \bar{r}$ such that the level set*

$$\{y \mid f(\bar{x}, y) \leq \bar{r}\}$$

is compact, then

$$(i) \qquad \bar{f}(\cdot) : \mathbb{R}^n \to \mathbb{R} \text{ is finite and lower semicontinuous,}$$
$$(ii) \qquad \partial \bar{f}(x) = \{t \; : \; (t,0) \in \partial f(x,y) \text{ for some } y \in \mathbb{R}^n\}.$$

Here, we provide a simplification of the condition of Theorem 10.31 in [38] because when f is a closed convex function, the boundedness of the level set $\{y \mid f(\bar{x}, y) \leq \bar{r}\}$ at some point $\bar{x}$ for some value $\bar{r}$ implies the uniform level boundedness (Definition 1.16 [38]).

The level set boundedness condition in Theorem 1 is necessary to guarantee the validity of the subdifferential formula in Theorem 1 for the marginal function $\bar{f}(\cdot)$. Next, we will see an example in which the formula does not hold in the absence of the level set boundedness condition.

Example 1. We define function $f_1(x,y) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^+$ piecewise as follows:

$$f_1(x,y) = \begin{cases} x \exp\left(-\frac{y}{x}\right) & x > 0 \\ 0 & x = 0, y = 0 \\ 0 & x = 0, y > 0 \\ +\infty & x = 0, y < 0. \end{cases}$$

Note that on $\mathrm{ri}(\mathrm{dom}(f_1))$, f_1 coincides with the perspective function of a simple exponential, and on the boundary, it is not hard to check that the function values can be obtained via a limiting argument. Thus, f_1 is a closed convex function. It is easy to see that the level set boundedness condition does not hold for f_1. The marginal function $\bar{f}_1$ after partially minimizing over y is

$$\bar{f}_1(x) = 0,$$

which is a constant function over $[0, +\infty]$. Therefore, $\bar{f}_1$ is also a closed convex function whose subdifferential set $\partial \bar{f}_1(x)$ is nonempty and compact for any $x \geq 0$; however, a simple calculation reveals that $\partial f_1(x,y) = \{(\exp(-\frac{y}{x}) + \frac{y^2}{x} \exp(-\frac{y}{x}), -\exp(-\frac{y}{x}))\}$ for $x > 0$, and the second coordinate $-\exp(-\frac{y}{x})$ is always nonzero. This means there does not exist a point y such that $(t,0) \in \partial f_1(x,y)$ for any $x > 0$. Figure 1 provides a visualization of f_1.
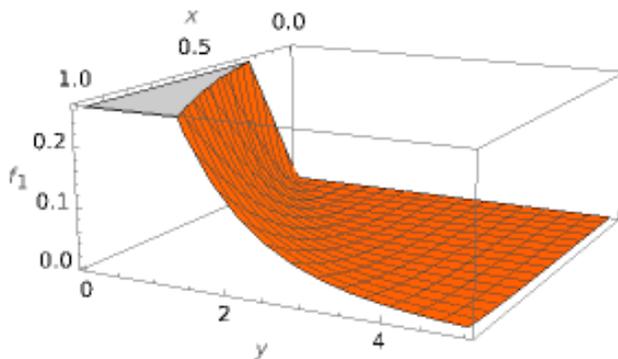

FIGURE 1. f_1.

We introduce the notations F_1 and F_2 to denote the sets of feasible points of the linear inequalities (1b) and (1c), respectively, i.e., $F_1 = \{y : Ay \leq b\}$ and $F_2 = \{(x,y) : Cy \leq Dx\}$. Starting from the perspective reformulation, we first give a corresponding convex envelope function of $\mathcal{Z}$. Suppose we

decompose Q as $Q = \text{diag}(\delta) + R$ such that $\delta \geq 0$ and $R \succcurlyeq 0$. The perspective reformulation is as follows:

$$\min_{x,y} \quad y^\top R y + \sum_{i=1}^n \delta_i \frac{y^2}{x_i} + g^\top y + h^\top x$$
$$\text{s.t.} \quad Ay \leq b$$
$$Cy \leq Dz$$
$$x \in X.$$

Alternatively, we can make use of the indicator function, and replace the constraints on the continuous variables with indicator functions in the objective function, i.e.

$$\min_{x,y} \quad y^\top R y + \sum_{i=1}^n \delta_i \frac{y^2}{x_i} + g^\top y + h^\top x + \mathcal{I}_{F_1}(y) + \mathcal{I}_{F_2}(x,y)$$
$$\text{s.t.} \quad x \in X.$$

Let us denote the objective function above as $f_{\text{persp}}(x,y)$. Note that the perspective relaxation is obtained when we replace the binary restriction $x \in X$ with $x \in \text{conv}(X)$. We have the relation that $(x,y,\eta) \in \mathcal{Z} \iff \eta \geq f_{\text{persp}}(x,y)$ and $x \in X$ (if x_i is zero for some $i \in [n]$, then to avoid f_{persp} being infinity, we must have y_i equal zero as well, and the complementarity constraints are satisfied), thus f_{persp} is a valid convex envelope function of $\mathcal{Z}$. Here, we ignore the subtle case where δ_i may be zero for some $i \in [n]$ since it is not hard to see that f_{persp} is a convex envelope function of $\mathcal{Z}$ under this scenario as well. As we have argued before, for any $x \in \text{conv}(X)$ and $t \in \partial \bar{f}_{\text{persp}}(x)$, the subgradient inequality $\tilde{\eta} \geq \langle t, \tilde{x} - x \rangle + \bar{f}_{\text{persp}}(x)$ is valid for any $(\tilde{x}, \tilde{y}, \tilde{\eta}) \in \mathcal{Z}$. In Proposition 1, we will characterize $\partial \bar{f}_{\text{persp}}$.

Proposition 1. *For any $x \in [0,1]^n$, suppose that S is the set of indices of nonzero entries of x, and S^C denotes its complement. Let A_S and C_S denote the submatrices where the columns are in S. For $y \in \mathbb{R}^n, \lambda \in \mathbb{R}^{m_1}$ and $\mu \in \mathbb{R}^{m_2}$, let $y_i = 0$ for $i \in S^C$ and y_S, λ, and μ be the primal-dual optimal solution of the following quadratic program (QP) in the reduced space indexed by S*

$$\min_{\tilde{y} \in \mathbb{R}^{|S|}} \quad \tilde{y}^\top \left(R_S + diag\left(\left\{ \frac{\delta_i}{x_i} \right\}_{i \in S} \right) \right) \tilde{y} + g_S^\top \tilde{y} \tag{3a}$$
$$\text{s.t.} \quad A_S \tilde{y} \leq b \tag{3b}$$
$$C_S \tilde{y} \leq Dx. \tag{3c}$$

Then $t \in \partial \bar{f}_{persp}(x)$ if and only if there exists $\Psi \in \mathbb{R}^{|S^C|}$ such that

$$t_i = -\delta_i \frac{y_i^2}{x_i^2} - \mu^\top D_i + h_i, \quad \forall i \in S \tag{4}$$

$$\delta_i \Psi_i = -(2R_i^\top y + g_i + \lambda^\top A_i + \mu^\top C_i), \quad \forall i \in S^C \tag{5}$$

$$t_i \leq -\frac{1}{4} \delta_i \Psi_i^2 - \mu^\top D_i + h_i, \quad \forall i \in S^C. \tag{6}$$

Proof. Given $x \in [0,1]^n$, let S denote the support set of x (i.e., $S = \{i \in [n] \mid x_i \neq 0\}$). By Assumption 1, there exists a feasible point $(\tilde{x}, \tilde{y}) \in X \times \mathbb{R}^n$ for (1). We have the inequality $f_{\text{persp}}(\tilde{x}, y) \geq y^\top Q y + g^\top y + h^\top \tilde{x}$ since $\delta_i \frac{y_i^2}{\tilde{x}_i} \geq \delta_i y_i^2$ for any $\tilde{x}_i \in [0,1]$. The strongly convex quadratic function $y^\top Q y + a^\top y$ has bounded level sets, which implies that any level set of $f_{\text{persp}}(\tilde{x}, \cdot)$ as a function of y is also bounded. Note, in particular, that the level set $\{y \mid f_{\text{persp}}(\tilde{x}, y) \leq f_{\text{persp}}(\tilde{x}, \tilde{y})\}$ is nonempty as it contains $\tilde{y}$, and it satisfies the level set boundedness condition in Theorem 1. Since we extend the perspective function $\frac{y_i^2}{x_i}$ to its closure, we have $\frac{y_i^2}{0} = +\infty$ if $y_i \neq 0$. Then for the subdifferential $\partial f_{\text{persp}}(x,y)$ to exist for some $y \in \mathbb{R}^n$, we must have $y_i = 0$ for any $i \in S^C$ and (x,y) is feasible for (1) (one of the indicator functions $\mathcal{I}_{F_1}$ or $\mathcal{I}_{F_2}$ will evaluate to $+\infty$ if otherwise). Recall that the subdifferential of the indicator function $\mathcal{I}_C(\cdot)$ of a closed convex set is its normal cone $\mathcal{N}_C(\cdot)$ [30], and the normal cones of polyhedral sets F_1 and F_2 are:

$$\mathcal{N}_{F_1}(y) = \{(0, \tilde{\lambda}^\top A) \mid \tilde{\lambda} \geq 0, \tilde{\lambda}_i = 0, \ \forall i \text{ such that } \langle a_i, y \rangle < b_i\},$$
$$\mathcal{N}_{F_2}(x,y) = \{(-\tilde{\mu}^\top D, \tilde{\mu}^\top C) \mid \tilde{\mu} \geq 0, \tilde{\mu}_i = 0, \ \forall i \text{ such that } \langle c_i, y \rangle < \langle d_i, x \rangle\}.$$

The indicator function of a polyhedron is a polyhedral function [37], and by Theorem 23.8 in [37] (in which the constraint qualification holds), $\partial f_{\mathrm{persp}}(\tilde{x}, \tilde{y})$ equals the sum of the subdifferentials of the components of f_{persp}, i.e., $y^\top R y$, $\delta_i \frac{y_i^2}{x_i}$, $\mathcal{I}_{F_1}$, and $\mathcal{I}_{F_2}$. By Theorem 1, $t \in \partial \bar{f}_{\mathrm{persp}}(x)$ if and only if t solves the first-order equation for some $y \in \mathbb{R}^n$:

$$(t, 0) \in \partial f_{\mathrm{persp}}(x, y).$$

We will reveal the equation above entry-by-entry in the order of (i) y_i for $i \in S$, (ii) y_i for $i \in S^C$, (iii) x_i for $i \in S$, and (iv) x_i for $i \in S^C$.
The first set of equations is for $i \in S$:

$$2 R_i^\top y + 2 \delta_i \frac{y_i}{x_i} + g_i + \tilde{\lambda}^\top A_i + \tilde{\mu}^\top C_i = 0. \tag{7}$$

The second set of equations is that there exists $\Psi \in \mathbb{R}^{|S^C|}$, where Ψ_i is the second entry of the subgradients in $\partial \frac{y_i^2}{x_i}$ at $(0, 0)$, such that for $i \in S^C$:

$$\delta_i \Psi_i + 2 R_i^\top y + g_i + \tilde{\lambda}^\top A_i + \tilde{\mu}^\top C_i = 0, \tag{8}$$

which is nothing but (5). The third set of equations is for $i \in S$:

$$t_i = -\delta_i \frac{y_i^2}{x_i^2} - \tilde{\mu}^\top D_i + h_i, \tag{9}$$

which is nothing but (4). The last set of equations is that there exists $\Phi \in \mathbb{R}^{|S^C|}$, where Φ_i is the first entry of the subgradients in $\partial \frac{y_i^2}{x_i}$ at $(0, 0)$, such that for $i \in S^C$:

$$t_i = -\delta_i \Phi_i - \tilde{\mu}^\top D_i + h_i. \tag{10}$$

In equations (7)–(10), by Lemma 2, Ψ_i, Φ_i for $i \in S^C$ satisfy the relations:

$$\Phi_i \leq -\frac{1}{4} \Psi_i^2, \quad \forall i \in S^C. \tag{11}$$

Thus far, we have argued that equations (7)–(11) must hold, (x, y) must be feasible to (1), and that $y_i = 0$ for all $i \in S^C$ for $\partial f_{\mathrm{persp}}(x, y)$ to exist. Next, we simplify these equations to obtain the desired result. Note that equations (7) where $\tilde{\lambda}$ and $\tilde{\mu}$ satisfy the complementarity condition together with the conditions that the given (x, y) satisfies constraints (1b)–(1c) and $y_i = 0$ for any $i \in S^C$ are equivalent to the Karush–Kuhn–Tucker (KKT) conditions of the QP problem (3). Hence, $(y_S, \tilde{\lambda}, \tilde{\mu})$ is the primal-dual optimal solution for the QP problem (3). Furthermore, we can combine equations (10) and (11) by plugging (11) into (10) which then gives (6). $\qquad \square$

By substituting (5) into (6), we can see that we can choose any t satisfying (4) and the following inequality:

$$t_i \leq -\frac{1}{4 \delta_i} (2 R_i^\top y + g_i + \lambda^\top A_i + \mu^\top C_i)^2 - \mu^\top D_i + h_i, \quad \forall i \in S^C, \tag{12}$$

in deriving the subgradient cuts for $\mathcal{Z}$. Nevertheless, the strongest cuts are those that attain equality. To see this, if we add the cut

$$\tilde{\eta} \geq \langle t, \tilde{x} - x \rangle + \bar{f}_{\mathrm{persp}}(x),$$

where $x_i = 0$, $\forall i \in S^C$, and expand it as $\tilde{\eta} \geq \sum_{i \in S} t_i (\tilde{x}_i - x_i) + \sum_{i \in S^C} t_i \tilde{x}_i + \bar{f}_{\mathrm{persp}}(x)$, then, a cut with smaller t_i for $i \in S^C$ is implied by the one with larger t_i for $i \in S^C$ and the bounds $0 \geq -x_i$ for $i \in [n]$. In Algorithm 1, we outline the procedure for obtaining the strongest cutting plane from the perspective reformulation.

Algorithm 1 A cut-generating procedure based on the perspective reformulation ($\mathrm{CUT}_{\mathrm{persp}}$)

1: **INPUT** $R, \delta, g, h, (A, b), (C, D), x$
2: Set S to be the set of indices such that $x_i > 0$
3: Set $y = 0$
4: Set (y_S, λ, μ) to be the optimal primal-dual solution of QP (3)
5: Set $t = 0$
6: **for** $i \in S$ **do**
7:     Set $t_i = -\delta_i \frac{y_i^2}{x_i^2} - \mu^\top D_i + h_i$
8: **end for**
9: **for** $i \in S^C$ **do**
10:     Set $\Psi_i = -\frac{1}{\delta_i}(2 \sum_{j \in S} R_{ij} y_j + g_i + \lambda^\top A_i + \mu^\top C_i)$
11: **end for**
12: **for** $i \in S^C$ **do**
13:     Set $t_i = -\frac{\delta_i}{4} \Psi_i^2 - \mu^\top D_i + h_i$
14: **end for**
15: **return** t

In some cases, e.g., in the presence of bound constraints, some constraints active at y may not involve y_i for $i \in S$, and the corresponding multipliers λ_i and μ_i do not appear in the optimal primal-dual solution of the QP (3). Consequently, such λ_i and μ_i are 'free' variables. As we have argued before, it would be beneficial to maximize the right-hand side of (12); however, doing so will require solving a QP. One simple strategy is to set these λ_i and μ_i to zero, which is how we implement the cutting planes in the numerical experiments.

One insight we can draw from Proposition 1 is that computing a cutting plane can be very efficient if only a small proportion of the entries of x are nonzero. When we enforce a sparsity constraint, i.e., $X \subseteq X_k$, at any binary point, the computational cost equals the solution time of a k-dimensional quadratic program (3) plus $\mathcal{O}(n(k + m_1 + m_2))$ floating-point operations. Next, we give a corollary of Proposition 1 for the case in which we do not have constraints over y.

Corollary 1. *For MIQP with no constraints over y, given $x \in [0, 1]^n$, suppose S is the set of indices of nonzero entries of x; for $y \in \mathbb{R}^n$, let $y_i = 0$ for $i \in S^C$ and y_S be the unconstrained optimal solution in the reduced subspace indexed by S, i.e., $y_S = -\frac{1}{2} Q_S^{-1} g_S$. Then $t \in \partial \bar{f}_{perps}(x)$ if and only if*

$$t_i = -\delta_i \frac{y_i^2}{x_i^2} + h_i \quad i \in S$$

$$\delta_i \Psi_i = -(2R_i^\top y + g_i) \quad \forall i \in S^C$$

$$t_i \leq -\frac{1}{4} \delta_i \Psi_i^2 + h_i \quad \forall i \in S^C.$$

Proof. The proof follows the same procedure as the proof of Proposition 1 by ignoring the constraints (1b) and (1c). □

Corollary 1 gives a formula for computing the subgradients of $\bar{f}_{\mathrm{persp}}(x)$ in the absence of constraints on y for an arbitrary diagonal decomposition ($Q = R + \mathrm{diag}(\delta)$ such that $R \succeq 0$ and $\delta \geq 0$), and it generalizes the result in [5] by allowing the remainder term R to be rank-deficient.

Bertsimas and Cory-Wright [3] proposed a class of cutting planes for $\mathcal{Z}$ when $X = X_k$ based on a convex Boolean reformulation of (1). Their result requires first reformulating (1) equivalently as a sparse regression problem, i.e.,

$$\min_{x,y} \quad \sum_{i=1}^{n} \delta_i y_i^2 + \|\beta - Ey\|_2^2 + \tilde{g}^\top y \tag{13a}$$

$$\text{s.t.} \quad Ay \leq b \tag{13b}$$

$$Cy \leq Dx \tag{13c}$$

$$y_i(1 - x_i) = 0, \quad \forall i = 1, \dots, n \tag{13d}$$

$$x \in X_k, \tag{13e}$$

where $R = E^\top E$ for some full row-rank $E \in \mathbb{R}^{k_1 \times n}$, $\beta = -\frac{1}{2}(EE^\top)^{-1}Eg$, and $\tilde{g} = (I - E^\top(EE^\top)^{-1}E)g$, and then applying the convex Boolean reformulation akin to [5, 36]. The perspective reformulation of (13) is written as:

$$\min_{x,y} \quad \sum_{i=1}^{n} \delta_i \frac{y_i^2}{x_i} + \|r\|_2^2 + \tilde{g}^\top y + c^\top x \tag{14a}$$

$$\text{s.t.} \quad r = \beta - Ey \tag{14b}$$

$$Ay \leq b \tag{14c}$$

$$Cy \leq Dz \tag{14d}$$

$$x \in X_k. \tag{14e}$$

We introduce another indicator function $\mathcal{I}_{F_3}(\cdot)$ with $F_3 \equiv \{(y, r) : r = \beta - Xy\}$, and define

$$f_{\text{persp}_2}(x, y, r) = \sum_{i=1}^{n} \delta_i \frac{y_i^2}{x_i} + \|r\|_2^2 + \tilde{g}^\top y + c^\top x + \mathcal{I}_{F_3}(y, r) + \mathcal{I}_{F_1}(y) + \mathcal{I}_{F_2}(x, y).$$

In Appendix A, we prove that $\partial \bar{f}_{\text{persp}_2}(x)$ coincides with the cutting planes in [3] and since f_{persp_2} only differs from f_{persp} by a constant, $\partial \bar{f}_{\text{persp}} = \partial \bar{f}_{\text{persp}_2}$. However, the cost of computing a cutting plane in [3] includes the cost of solving a QP of size k_1 plus $\mathcal{O}(n(k_1 + m_1 + m_2))$ floating-point operations, where k_1 (the rank of R) can be of the same order as n; on the contrary, the formula in Proposition 1 only requires the solution of a QP of size k in addition to $\mathcal{O}(n(k + m_1 + m_2))$ floating-point operations. Observe that k can be much smaller than n in some applications.

Remark 1. At any binary point $x \in X$, $\bar{f}_{\text{persp}}(x)$ is the optimal value of MIQP (1) given x.

When computing $\bar{f}_{\text{persp}}(x) = \min_{y \in \mathbb{R}^n} f_{\text{persp}}(x, y)$, to avoid the term $\delta_i \frac{y_i^2}{x_i}$ becoming $+\infty$, we must have $y_i = 0$ if $x_i = 0$ and thus the complementarity constraint (1d) holds. On the other hand, for any y satisfying the complementarity constraint (1d), $y^\top \left(\text{diag}\left(\left\{ \frac{\delta_i}{x_i} \right\}_{i \in [n]} \right) + R \right) y = y_S^\top \left(\text{diag}(\delta_S) + R_S \right) y_S = y_S^\top Q_S y_S$ since the term $\delta_i \frac{y_i^2}{x_i} = \frac{0}{0} = 0$ for any $i \in S^C$. Hence, the partial minimization problem $\min_{y \in \mathbb{R}^n} f_{\text{persp}}(x, y)$ is equivalent to

$$\min_{\tilde{y} \in \mathbb{R}^{|S|}} \quad \tilde{y}^\top Q_S \tilde{y} + g_S^\top \tilde{y} + h^\top x$$

$$\text{s.t.} \quad A_S \tilde{y} \leq b$$

$$C_S \tilde{y} \leq Dx,$$

which is exactly MIQP (1) given x.

So far, we have seen how to derive cutting planes for $\mathcal{Z}$ in the space of (x, η) using Theorem 1 and the perspective reformulation. It is also interesting to see how we can use the same framework to derive cutting planes for $\mathcal{Z}$ based on other reformulations in the literature. In the next section, we focus on one reformulation that uses the ideal formulation of the epigraph of a rank-one quadratic function with indicator variables and no constraints on y.

4. Cutting planes for perspective reformulation with rank-one inequalities

Let

$$X_{R1} \equiv \{(x, y, \eta) \in \{0, 1\}^n \times \mathbb{R}^{n+1} \mid \eta \geq (y^\top u)^2, \ y_i(1 - x_i) = 0, \ \forall i \in [n]\}$$

denote the epigraph of a rank-one quadratic function with indicator variables, and y is unconstrained. Atamtürk and Gómez [2] show that cl conv(X_{R1}) is obtained by simply adding the inequality

$$\eta \geq \max\left\{ (u^\top y)^2, \frac{(u^\top y)^2}{\sum_{i=1}^{n} x_i} \right\}, \tag{15}$$

and the closure of the convex hull of X_{R1} admits the following description:

$$\text{cl conv}(X_{R1}) \equiv \left\{ (x, y, \eta) \in [0, 1]^n \times \mathbb{R}^{n+1} \mid \eta \geq \max\left\{ (u^\top y)^2, \frac{(u^\top y)^2}{\sum_{i=1}^{n} x_i} \right\} \right\}.$$

The validity of (15) follows from the fact that for any $(x, y, \eta) \in X_{R1}$ if $x_i = 1$ for some $i \in [n]$, then $\eta \geq (u^\top y)^2 = \max \left\{ (u^\top y)^2, \frac{(u^\top y)^2}{\sum_{i=1}^n x_i} \right\}$, and if $x_i = 0$ for all $i \in [n]$, then $(u^\top y)^2 = \frac{(u^\top y)^2}{\sum_{i=1}^n x_i} = 0$ by the complementarity constraint. The convexity of the constraint (15) follows from the fact that its right-hand side is the maximum of two convex functions.

Inequality (15) is easy to implement with a decomposition of Q as the sum of rank-one quadratic matrices. In what follows, we will see that the conic-quadratic inequality (15) together with the perspective reformulation gives a class of stronger subgradient cuts than those in Proposition 1.

Before we derive our new cutting planes, we introduce the following lemma regarding the subdifferential of the composite of a perspective function and a linear mapping.

Lemma 3. *Let* $\phi(x, y) = \max \left\{ (u^\top y)^2, \frac{(u^\top y)^2}{\sum_{i=1}^n x_i} \right\}$, *then* $\partial \phi(x, y) =$

$$
\begin{cases}
(0, 2(u^\top y)u) & \sum_{i=1}^n x_i > 1 \\
conv(\{(0, 2(u^\top y)u), (-(u^\top y)^2 \mathbf{1}, 2(u^\top y)u)\}) & \sum_{i=1}^n x_i = 1 \\
\left(-\frac{(u^\top y)^2}{(\sum_{i=1}^n x_i)^2} \mathbf{1}, 2 \frac{1}{\sum_{i=1}^n x_i} (u^\top y)u \right) & 0 < \sum_{i=1}^n x_i < 1 \\
\{(\Phi \mathbf{1}, \Psi u) : \Phi \leq -\frac{1}{4} \Psi^2\} & x = 0.
\end{cases}
$$

Proof. There are four cases to consider:

(i) $\sum_{i=1}^n x_i > 1$. In a small neighborhood of such (x, y), we have $\sum_{i=1}^n x_i > 1$, thus $(u^\top y)^2 \geq \frac{(u^\top y)^2}{\sum_{i=1}^n x_i}$. Hence, $\phi(x, y)$ equals $(u^\top y)^2$ locally. Since the subdifferential is determined by the local geometry of the epigraph of $\phi(\cdot)$, $\partial \phi(x, y) = \partial (u^\top y)^2$.

(ii) $0 < \sum_{i=1}^n x_i < 1$. Similarly, when $0 < \sum_{i=1}^n x_i < 1$, $\phi(x, y) = \frac{(u^\top y)^2}{\sum_{i=1}^n x_i}$ in a small neighborhood, and $\partial \phi(x, y) = \partial \frac{(u^\top y)^2}{\sum_{i=1}^n x_i}$ which shrinks to a single point since $\frac{(u^\top y)^2}{\sum_{i=1}^n x_i}$ is differentiable.

(iii) $\sum_{i=1}^n x_i = 1$. In this case, then $(u^\top y)^2$ and $\frac{(u^\top y)^2}{\sum_{i=1}^n x_i}$ coincide, and they are finite in a small neighborhood of (x, y). Then, according to the rule of the subdifferential of the supremum of finite convex functions (Theorem 4.4.2 [30]), $\partial \phi(x, y) = \text{conv}(\partial (u^\top y)^2 \bigcup \partial \frac{(u^\top y)^2}{\sum_{i=1}^n x_i})$.

(iv) $x = 0$. In this case, $\phi(x, y)$ is locally equal to $\frac{(u^\top y)^2}{\sum_{i=1}^n x_i}$, the composite of a perspective function and a linear mapping. Then, $\partial \frac{(u^\top y)^2}{\sum_{i=1}^n x_i}$ can be derived via the chain rule of subdifferentials and Lemma 2.

$\square$

In what follows, we assume that we are given a decomposition $Q = LL^\top + \text{diag}(\delta) + N$ such that $\delta \geq 0$ and $N \succeq 0$, and $L \in \mathbb{R}^{n \times K}$ for some $1 \leq K \leq n$ (we will discuss how to obtain such a decomposition later). Then, we can apply inequality (15) to the terms $(L_i^\top y)^2$ for $i \in 1, \ldots, K$ and arrive at the following convex envelope function for $\mathcal{Z}$:

$$
f_{\text{persp+ro}}(x, y) \equiv \sum_{i=1}^K \max \left\{ (L_i^\top y)^2, \frac{(L_i^\top y)^2}{\sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j} \right\} + y^\top N y + \sum_{i=1}^n \delta_i \frac{y_i^2}{x_i} + g^\top y + h^\top x
$$
$$
+ \mathcal{I}_{F_1}(y) + \mathcal{I}_{F_2}(x, y).
$$

Proposition 2. *For any* $x \in [0, 1]^n$, S *is the set of indices of nonzero x entries, and S^C denotes its complement. Suppose*

$$
I^> = \{i : \sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j > 1, i = 1, \ldots, K\},
$$
$$
I^= = \{i : \sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j = 1, i = 1, \ldots, K\},
$$
$$
I^< = \{i : 0 < \sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j < 1, i = 1, \ldots, K\},
$$
$$
I^0 = \{i : \sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j = 0, i = 1, \ldots, K\}.
$$

Let $y \in \mathbb{R}^n$, $y_i = 0$ for $i \in S^C$, and (y_S, λ, μ) be the primal-dual optimal solution of the following QP:

$$\min_{\tilde{y} \in \mathbb{R}^{|S|}} \quad \tilde{y}^\top \left(\sum_{i \in I^> \bigcup I^=} (L_i)_S (L_i)_S^\top + \sum_{i \in I^<} \frac{1}{\sum_{1 \le l \le n, L_{li} \neq 0} x_l} (L_i)_S (L_i)_S^\top + diag\left(\left\{ \frac{\delta_i}{x_i} \right\}_{i \in S}\right) + N_S \right) \tilde{y} + g_S^\top \tilde{y} \tag{16a}$$

$$s.t. \quad A_S \tilde{y} \le b \tag{16b}$$

$$C_S \tilde{y} \le Dx, \tag{16c}$$

Then, $t \in \partial \bar{f}_{persp+ro}(x)$ if and only if there exist $\phi \in \mathbb{R}^{|I^=|+|I^0|}$, $\psi \in \mathbb{R}^{|I^0|}$, $\Psi \in \mathbb{R}^{|S^C|}$, and $\Phi \in \mathbb{R}^{|S^C|}$ such that

$$0 = \sum_{i \in I^> \bigcup I^=} 2L_{ji}(L_i^\top y) + \sum_{i \in I^<} 2 \frac{1}{\sum_{1 \le j \le n, L_{ji} \neq 0} x_j} L_{ji}(L_i^\top y) + \sum_{i \in I^0} L_{ji}\psi_i$$
$$+ \delta_j \Psi_j + 2N_i^\top y + g_j + \lambda^\top A_j + \mu^\top C_j = 0, \quad \forall j \in S^C \tag{17}$$

$$t_j = \sum_{i \in I^=, L_{ji} \neq 0} \phi_i - \sum_{i \in I^<, L_{ji} \neq 0} \frac{(L_i^\top y)^2}{(\sum_{1 \le l \le n, L_{li} \neq 0} x_l)^2} - \delta_j \frac{y_j^2}{x_j^2} - \tilde{\mu}^\top D_j + h_j, \quad \forall j \in S \tag{18}$$

$$t_j = \sum_{i \in I^=, L_{ji} \neq 0} \phi_i - \sum_{i \in I^<, L_{ji} \neq 0} \frac{(L_i^\top y)^2}{(\sum_{1 \le l \le n, L_{li} \neq 0} x_l)^2} + \sum_{i \in I^0, L_{ji} \neq 0} \phi_i + \delta_j \Phi_j - \mu^\top D_j + h_j, \quad \forall j \in S^C \tag{19}$$

$$\phi_i \ge -(L_i^\top y)^2, \quad \forall i \in I^= \tag{20}$$

$$\phi_i \le -\frac{1}{4}\psi_i^2, \quad \forall i \in I^0 \tag{21}$$

$$\Phi_i \le -\frac{1}{4}\Psi_i^2, \quad \forall i \in S^C. \tag{22}$$

Proof. Note that $f_{persp+ro}(x, y) \ge \sum_{i=1}^n \delta_i \frac{y_i^2}{x_i} + y^\top(LL^\top)y + y^\top Ny + g^\top y + h^\top x$, and the right-hand side can be considered as a perspective reformulation. Since we have shown in the proof of Proposition 1 that the level set boundedness condition in Theorem 1 holds for f_{persp}, it also holds for $f_{persp+ro}$. Then, the subdifferential of the marginal function $\bar{f}_{persp+ro}(x)$ can be obtained by solving for the first-order condition in Theorem 1. By the same argument as in the proof of Proposition 1, we can compute $\partial f_{persp+ro}(x, y)$ as the sum of the subdifferentials of the components of $f_{persp+ro}$. The only difference between $f_{persp+ro}$ and f_{persp} is the appearance of the term $\sum_{i=1}^K \max\left\{ (L_i^\top y)^2, \frac{(L_i^\top y)^2}{\sum_{1 \le j \le n, L_{ji} \neq 0} x_j} \right\}$, by Lemma 3 and the calculus of subdifferentials, its subdifferential set equals

$$\sum_{i \in I^>} (0, 2(L_i^\top y)L_i) +$$
$$\sum_{i \in I^=} conv\{(0, 2(L_i^\top y)L_i), (\underbrace{-(L_i^\top y)^2, \dots, -(L_i^\top y)^2}_{\text{indices where } L_{li} \neq 0}, \underbrace{0, \dots, 0}_{\text{indices where } L_{li} = 0}, 2(L_i^\top y)L_i)\} +$$
$$\sum_{i \in I^<} (\underbrace{-\frac{(L_i^\top y)^2}{(\sum_{1 \le j \le n, L_{ji} \neq 0} x_j)^2}, \dots, -\frac{(L_i^\top y)^2}{(\sum_{1 \le j \le n, L_{ji} \neq 0} x_j)^2}}_{\text{indices where } L_{li} \neq 0}, \underbrace{0, \dots, 0}_{\text{indices where } L_{li} = 0}, 2\frac{1}{\sum_{1 \le j \le n, L_{ji} \neq 0} x_j}(L_i^\top y)L_i) +$$
$$\sum_{i \in I^0} \{(\underbrace{\phi_i, \dots, \phi_i}_{\text{indices where } L_{li} \neq 0}, \underbrace{0, \dots, 0}_{\text{indices where } L_{li} = 0}, \psi_i L_i) : \phi_i \le -\frac{1}{4}\psi_i^2\}.$$

The subdifferential set of $\max\left\{ (L_i^\top y)^2, \frac{(L_i^\top y)^2}{\sum_{1 \le j \le n, L_{ji} \neq 0} x_j} \right\}$ for some $i \in I^=$ can be written more compactly as $\{(\underbrace{\phi_i, \dots, \phi_i}_{\text{indices where } L_{li} \neq 0}, \underbrace{0, \dots, 0}_{\text{indices where } L_{li} = 0}, 2(L_i^\top y)L_i) : 0 \ge \phi_i \ge -(L_i^\top y)^2\}$. The subdifferentials of other terms in $f_{persp+ro}$ are the same as f_{persp}. Now, we state again that $t \in \partial \bar{f}_{persp+ro}(x)$ if and only if there exists $y \in \mathbb{R}^n$ such that

$$(t, 0) \in \partial f_{persp+ro}(x, y). \tag{23}$$

Similarly, for the subdifferential set $\partial f_{\text{persp+ro}}(x, y)$ to exist, we must have $y_i = 0$ for $i \in S^C$ since otherwise, $\delta_i \frac{y_i^2}{x_i} = +\infty$, and (x, y) must satisfy constraints (1b)–(1c). Recall that the subdifferential sets of the indicator functions $\mathcal{I}_{F_1}$ and $\mathcal{I}_{F_2}$ are

$$\mathcal{N}_{F_1}(y) = \{(0, \tilde{\lambda}^\top A) \mid \tilde{\lambda} \geq 0, \tilde{\lambda}_i = 0, \ \forall i \text{ such that } \langle a_i, y \rangle < b_i\},$$

$$\mathcal{N}_{F_2}(x, y) = \{(-\tilde{\mu}^\top D, \tilde{\mu}^\top C) \mid \tilde{\mu} \geq 0, \tilde{\mu}_i = 0, \ \forall i \text{ such that } \langle c_i, y \rangle < \langle d_i, x \rangle\}.$$

Again, we uncover equation (23) entry by entry in the order of (i) y_j for $j \in S$, (ii) y_j for $j \in S^C$, (iii) x_j for $j \in S$, and (iv) x_j for $j \in S^C$. The first set of equations is for $j \in S$:

$$\sum_{i \in I^> \bigcup I^=} 2 L_{ji}(L_i^\top y) + \sum_{i \in I^<} 2 \frac{1}{\sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j} L_{ji}(L_i^\top y) + 2\delta_j \frac{y_j}{x_j} + 2 N_i^\top y + g_j + \tilde{\lambda}^\top A_j + \tilde{\mu}^\top C_j = 0. \quad (24)$$

The second set of equations is for $j \in S^C$:

$$\sum_{i \in I^> \bigcup I^=} 2 L_{ji}(L_i^\top y) + \sum_{i \in I^<} 2 \frac{1}{\sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j} L_{ji}(L_i^\top y) + \sum_{i \in I^0} L_{ji} \psi_i + \delta_j \Psi_j + 2 N_i^\top y + g_j + \tilde{\lambda}^\top A_j + \tilde{\mu}^\top C_j = 0, \quad (25)$$

which is exactly (17). The third set of equations is for $j \in S$:

$$t_j = \sum_{i \in I^=, L_{ji} \neq 0} \phi_i - \sum_{i \in I^<, L_{ji} \neq 0} \frac{(L_i^\top y)^2}{(\sum_{1 \leq l \leq n, L_{li} \neq 0} x_l)^2} - \delta_j \frac{y_j^2}{x_j^2} - \tilde{\mu}^\top D_j + h_j, \quad (26)$$

which is nothing but (18). The last set of equations is for $j \in S^C$:

$$t_j = \sum_{i \in I^=, L_{ji} \neq 0} \phi_i - \sum_{i \in I^<, L_{ji} \neq 0} \frac{(L_i^\top y)^2}{(\sum_{1 \leq l \leq n, L_{li} \neq 0} x_l)^2} + \sum_{i \in I^0, L_{ji} \neq 0} \phi_i + \delta_j \Phi_j - \tilde{\mu}^\top D_j + h_j, \quad (27)$$

which is equivalent to (19). In equations (25)–(27), ψ, Ψ, ϕ, and Φ satisfy the relations (20)–(22). Note that equations (24) where $\tilde{\lambda}$ and $\tilde{\mu}$ satisfy the complementarity condition together with the conditions that the given (x, y) satisfies the constraints (1b)–(1c) and $y_i = 0$ for all $i \in S^C$ are equivalent to saying that the triple $(y_S, \tilde{\lambda}, \tilde{\mu})$ satisfies the KKT condition for the QP (16). $\qquad \square$

Now, a question to ask is how to decompose Q as $\sum_{i=1}^K L_i L_i^\top + \text{diag}(\delta) + N$? Atamtürk and Gómez [2] propose a semidefinite programming (SDP) approach for decomposing Q into the sum of low-dimensional rank-one quadratic functions and one-dimensional quadratic functions for the sake of maximizing the relaxation lower bound; however, the computational cost of solving the SDP problem using interior-point methods could be prohibitive. A much cheaper option is the Cholesky decomposition, and we refer interested readers to [24] for a complete discussion on variants of the Cholesky decomposition and their numerical stability. We can divide the task into (i) first extract a diagonal term $\text{diag}(\delta)$ from Q, and then (ii) do a Cholesky decomposition on the remaining positive semidefinite matrix (the Cholesky decomposition is applicable for both positive definite and positive semidefinite matrices).

The subgradient cuts of $\bar{f}_{\text{persp+ro}}$ have a much simpler form at binary points. In what follows, we will present an algorithm for computing cutting planes in this case.

Proposition 3. *Given* $x \in X \subset \{0, 1\}^n$, *suppose we decompose* Q *as* $Q = \sum_{i=1}^K L_i L_i^\top + diag(\delta) + N$ *such that* $\delta \geq 0$, $N \succcurlyeq 0$ *and* $R = \sum_{i=1}^K L_i L_i^\top + N$. *Let* t *be the vector returned by Algorithm 2, then* $t \in \partial \bar{f}_{persp+ro}(x)$ *and* $\tilde{\eta} \geq \langle t, \tilde{x} - x \rangle + \bar{f}_{persp+ro}(x)$ *is a valid cut for* $\mathcal{Z}$.

Proof. Since $x \in X$ is binary, the index set $I^<$ does not exist. Thus, QP (16) reduces to:

$$\min_{\tilde{y} \in \mathbb{R}^{|S|}} \quad \tilde{y}^\top \left(\sum_{i \in I^> \bigcup I^+} (L_i)_S (L_i)_S^\top + \text{diag}\left(\left\{ \frac{\delta_i}{x_i} \right\}_{i \in S} \right) + N_S \right) \tilde{y} + g_S^\top \tilde{y} \quad (28a)$$

$$\text{s.t.} \quad A_S \tilde{y} \leq b \quad (28b)$$

$$\quad C_S \tilde{y} \leq Dx. \quad (28c)$$

Algorithm 2 A cut-generating procedure based on the perspective reformulation and rank-one inequalities ($\text{CUT}_{\text{persp+ro}}$)

1: **INPUT** $R, \delta, L, g, h, (A, b), (C, D), x$
2: Set S to be the set of indices such that $x_i > 0$
3: Set I^0 to be the set of indices such that $\sum_{1 \le l \le n, L_{li} \ne 0} x_l = 0$
4: Set $y = 0$
5: Set (y_S, λ, μ) to be the optimal primal-dual solution of QP (3)
6: Set $t = 0$
7: **for** $i \in S$ **do**
8:     Set $t_i = -\delta_i \dfrac{y_i^2}{x_i^2} - \mu^\top D_i + h_i$
9: **end for**
10: **for** $i \in I^0$ **do**
11:     Set $n_i = |\{j \ : \ L_{ji} \ne 0\}|$
12: **end for**
13: **for** $i \in S^C$ **do**
14:     Set $r_i = -2 \sum_{j \in S} R_{ij} y_j - g_i - \lambda^\top A_i - \mu^\top C_i$
15: **end for**
16: Solve the system $\left(\text{diag}\left(\{\delta_i\}_{i \in S^C} \right) + \sum_{i \in I^0} \frac{1}{n_i} (L_i)_{S^C} (L_i)_{S^C}^\top \right) \beta = 2r$
17: **for** $i \in S^C$ **do**
18:     Set $\Psi_i = \frac{1}{2} \beta_i$
19: **end for**
20: **for** $i \in I^0$ **do**
21:     Set $\psi_i = \frac{1}{2n_i} \sum_{j \in S^C} L_{ji} \beta_j$
22: **end for**
23: **for** $i \in S^C$ **do**
24:     Set $t_i = -\frac{\delta_i^2}{4} \Psi_i^2 - \frac{1}{4} \sum_{j \in I^0, L_{ij} \ne 0} \psi_j^2 - \mu^\top D_i + h_i$
25: **end for**
26: **return** t

Since each L_i for $i \in I^0$ is supported only on S^C, the Hessian of the objective equals

$$\sum_{i \in I^> \bigcup I^= \bigcup I^0} (L_i)_S (L_i)_S^\top + \text{diag}\left(\left\{ \frac{\delta_i}{x_i} \right\}_{i \in S} \right) + N_S,$$

which by definition is also equal to $\text{diag}\left(\left\{ \frac{\delta_i}{x_i} \right\}_{i \in S} \right) + R_S$. If we take $\phi_i = 0$ for $i \in I^=$ in equations (18)–(20), then equations (17)–(22) further simplify to

$$t_i = -\delta_i \frac{y_i^2}{x_i^2} - \mu^\top D_i + h_i, \quad \forall i \in S \tag{29}$$

$$\delta_i \Psi_i + \sum_{j \in I^0, L_{ij} \ne 0} L_{ij} \psi_j = -(2N_i^\top y + \sum_{j \in I^> \bigcup I^=} 2L_{ij}(L_j^\top y) + g_i + \lambda^\top A_i + \mu^\top C_i), \quad \forall i \in S^C \tag{30}$$

$$t_i = \delta_i \Phi_i + \sum_{j \in I^0, L_{ij} \ne 0} \phi_j - \mu^\top D_i + h_i, \quad \forall i \in S^C \tag{31}$$

$$\phi_i \le -\frac{1}{4} \psi_i^2, \qquad \forall i \in I^0 \tag{32}$$

$$\Phi_i \le -\frac{1}{4} \Psi_i^2, \qquad \forall i \in S^C. \tag{33}$$

We can substitute (32)–(33) into (31), to obtain

$$t_i \le -\frac{\delta_i}{4} \Psi_i^2 - \frac{1}{4} \sum_{j \in I^0, L_{ij} \ne 0} \psi_j^2 - \mu^\top D_i + h_i, \quad \forall i \in S^C. \tag{34}$$

The free variables are Ψ_i for $i \in S^C$, ψ_i for $i \in I^0$, and t_i for $i \in S^C$. Note that a cut $\tilde{\eta} \ge \langle \bar{t}, \tilde{x} - x \rangle + \bar{f}_{\text{persp+ro}}(x)$ 'dominates' another cut $\tilde{\eta} \ge \langle \tilde{t}, \bar{x} - x \rangle + \bar{f}_{\text{persp+ro}}(x)$ if $\bar{t}_i \ge \tilde{t}_i, \ \forall i \in S^C$

(because the latter is implied by the former together with the bounds $-x_i \leq 0$), thus one strategy is to select Ψ, ψ, t such that the summation $\sum_{i \in S^C} t_i$ is maximized, and this guarantees that the cut is non-dominated. Let us denote the right-hand side of (30) as r_i for all $i \in S^C$. Again, since each L_i for $i \in I^0$ is only supported on S^C, the right-hand side of (30) equals $-(2R_i^\top y + h_i + \lambda^\top A_i + \mu^\top C_i)$. The QP problem for maximizing $\sum_{i \in S^C} t_i$ is as follows:

$$\min \quad \sum_{i \in S^C} \delta_i \Psi_i^2 + \sum_{i \in I^0} \sum_{j \in S^C, L_{ji} \neq 0} \psi_i^2 \tag{35a}$$

$$\text{s.t.} \quad \delta_i \Psi_i + \sum_{j \in I^0} L_{ij} \psi_j = r_i \quad [\beta_i] \quad \forall i \in S^C. \tag{35b}$$

We associate each constraint with Lagrangian multiplier β_i, and then the KKT conditions of (35) are:

$$\Psi_i = \frac{1}{2} \beta_i, \quad \forall i \in S^C \tag{36}$$

$$\psi_i = \frac{\sum_{j \in S^C} L_{ji} \beta_j}{2n_i}, \qquad \forall i \in I^0 \tag{37}$$

$$\delta_i \Psi_i + \sum_{j \in I^0} L_{ij} \psi_j = r_i, \qquad \forall i \in S^C. \tag{38}$$

By substituting equations (36) and (37) into (38), we can solve for the multiplier β via

$$\left(\text{diag} \left(\{\delta_i\}_{i \in S^C} \right) + \sum_{i \in I^0} \frac{1}{n_i} (L_i)_{S^C} (L_i)_{S^C}^\top \right) \beta = 2r. \tag{39}$$

Since the coefficient matrix is a diagonal matrix plus a rank $|I^0|$ term, its inverse can be computed efficiently as the sum of a diagonal matrix and rank $|I^0|$ matrix using the Woodbury matrix identity.

Finally, we can use (36), (37), and (34) to obtain t_i for $i \in S^C$. $\qquad \square$

Remark 2. If we decompose Q as $Q = \sum_{i=1}^{K} L_i L_i^\top + \text{diag}(\delta) + N$ such that $\delta \geq 0, N \succcurlyeq 0$, and $R = \sum_{i=1}^{K} L_i L_i^\top + N$, then at any binary point $x \in X$, $\partial \bar{f}_{\text{persp}}(x) \subsetneqq \partial \bar{f}_{\text{persp+ro}}(x)$.

This strict inclusion relation can be seen from the fact that if we set $\psi_i = 0$ for all $i \in I^= \bigcup I^0$ and $\phi_i = 0$ for all $i \in I^0$, then equations (30) and (31) are identical to equations (5) and (6), and thus any t feasible for the system (4)–(6) is also feasible for the system (29)–(33).

Remark 3. If we decompose Q as $Q = \sum_{i=1}^{K} L_i L_i^\top + \text{diag}(\delta) + N$ such that $\delta \geq 0, N \succcurlyeq 0$, and $R = \sum_{i=1}^{K} L_i L_i^\top + N$, then at any binary point $x \in X$, $\bar{f}_{\text{persp}}(x) = \bar{f}_{\text{persp+ro}}(x)$. For any two t and $\tilde{t}$ returned by Algorithm 1 and Algorithm 2 respectively, we always have $\sum_{i \in S^C} \tilde{t}_i \geq \sum_{i \in S^C} t_i$.

For any $i \in I^> \bigcup I^=$, the term $\max \left\{ (L_i^\top y)^2, \frac{(L_i^\top y)^2}{\sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j} \right\}$ equals $(L_i^\top y)^2$. For any $i \in I^0$, the fact $\sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j = 0$ implies that $L_{ji} \neq 0$ only if $x_j = 0$. If there exists some $j \in S^C$ such that $y_j \neq 0$, then both $f_{\text{persp}}(x,y)$ and $f_{\text{persp+ro}}(x,y)$ will be $+\infty$ as they have the term $\delta_j \frac{y_j^2}{x_j}$. On the other hand, if $y_j = 0$ for all $j \in S^C$, then the term $\max \left\{ (L_i^\top y)^2, \frac{(L_i^\top y)^2}{\sum_{1 \leq j \leq n, L_{ji} \neq 0} x_j} \right\}$ vanishes for any $i \in I^0$, and $y^\top (\sum_{i \in I^> \bigcup I^=} L_i L_i^\top + \text{diag}(\delta) + N)y = y^\top (\text{diag}(\delta) + R)y = y^\top Q y$. Overall, we have $f_{\text{persp}}(x,y) = f_{\text{persp+ro}}(x,y)$ for any $y \in \mathbb{R}^n$, and thus $\bar{f}_{\text{persp}}(x) = \bar{f}_{\text{persp+ro}}(x)$. The second part holds since by Remark 2, any $t \in \partial \bar{f}_{\text{persp}}(x)$ is also a subgradient of $\bar{f}_{\text{persp+ro}}(x)$, and Algorithm 2 outputs the subgradient vector $\tilde{t}$ of $\bar{f}_{\text{persp+ro}}(x)$ such that $\sum_{i \in S^C} \tilde{t}_i$ is maximized. Remark 3 confirms the intuition that the tighter convex envelope $f_{\text{persp+ro}}$ leads to stronger cutting planes of $\mathcal{Z}$.

The computational cost of Algorithm 2 equals the cost of solving QP (3) plus $\mathcal{O}(|I^0|^2 n + |I^0|^3)$ floating-point operations (the cost of solving the linear system (39)). Again, the procedure can be very efficient if we are at a binary point x satisfying a strong sparsity constraint. On the other hand, we will only generate cuts at binary points when we utilize Algorithm 2 as the cut-generating procedure in an outer approximation algorithm.

In the next section, we will demonstrate the effectiveness of the cutting planes in Proposition 1 and Proposition 3 on sparse portfolio selection problems with minimum and maximum investment

constraints. We will benchmark the OA algorithms with Algorithm 1 and Algorithm 2 as the cut-generation procedure against the state-of-the-art methods in the literature.

5. Computational experiments

In this section, we demonstrate the practical performance of the cutting planes in Proposition 1 and Proposition 3 through a computational study on the sparse portfolio selection problem. We solve the corresponding MIQP by an OA algorithm in which we iteratively refine the polyhedral approximation for $\mathcal{Z}$ using cutting planes in Proposition 1 and Proposition 3. An outline of the OA algorithm is given in Algorithm 3. All the experiments are conducted on a laptop with a 12th Generation Intel(R) Core(TM) i7-12700H CPU and 16 GB RAM. The quadratic programming subproblems are modeled by CVX 2.2 and solved using MOSEK 10.1.13, and the mixed-integer programming (MIP) master problem is solved by Gurobi 10.0.1. We adapt the default settings of Gurobi, except that the number of threads in use is forced to be one. The MIP solver terminates when a solution of relative optimality gap $< 0.01\%$ is found. All the algorithms and mathematical optimization models are implemented in Python.

5.1. **Outer approximation.** The OA algorithm is formalized in Algorithm 3.

Theorem 2. *Algorithm 3 solves MIQP* (1) *exactly in a finite number of iterations.*

Proof. Let us denote the optimal value of MIQP (1) as **OPT**. The finite convergence of Algorithm 3 follows from the fact that each binary solution is visited at most once, and only a finite number of cutting planes are added. Suppose Algorithm 3 returns a solution (x^T, η^T), then because $\mathcal{Z}_t$ contains $\mathcal{Z}$ at iteration t (the cutting planes are valid for $\mathcal{Z}$), we have $\eta^T \leq \text{OPT}$. By the stopping criteria, we have $\eta^T \geq \bar{f}_{\text{persp}}(x^T)$. Remark 1 tells that $\bar{f}_{\text{persp}}(x^T)$ is the optimal value of MIQP when x is fixed to x^T, and hence upper bounds **OPT**. Putting these together, we have $\eta^T = \text{OPT}$. $\square$

Instead of rebuilding a branch-and-bound tree each time a violated cut is added, we adapt the single-tree strategy through the *lazy constraints*. This means that the cutting planes are added to the MIP solver when an integer solution is found, and it violates some cutting plane. The solver terminates when no more violated cutting planes are added, and the current integer solution is optimal.

5.1.1. *Incumbent solutions:* High-quality incumbent solutions help prune the branch-and-bound nodes and save efforts on exploring suboptimal regions. In the literature, heuristics exist under some particular constraints of (1), i.e., [3, 7, 43]; however, there does not exist a heuristic that can generate provably near-optimal solutions for MIQP in its general form. Another strategy is providing the visited binary points $(x^t, \bar{f}_{\text{persp}}(x^t))$ as incumbent solutions to the MIP solver (we have established in Theorem 2 that $\bar{f}_{\text{persp}}(x^t)$ upper bounds the optimal solution). Implementing this strategy takes no additional cost to the OA procedure as the computation of $\bar{f}_{\text{persp}}(x^t)$ is included in the cut generation, and we can set incumbent solutions through a **user callback** function for commercial solvers such as CPLEX and Gurobi.

5.1.2. *Copy variables:* We add a copy of the constraints $Ay \leq b$ and $Cy \leq Dx$ to the initial polyhedral approximation $\mathcal{Z}_1$, i.e., $Z_1 = \{(x, y, \eta) : Ay \leq b, Cy \leq Dx, x \in \text{conv}(X)\}$. Doing so guarantees that the quadratic programming subproblem (3) is always feasible when we generate cutting planes.

5.1.3. *Improve the lower bound:* Note that the perspective relaxation provides a valid lower bound on the optimal value of the MIQP (1). We denote the optimal value of the perspective relaxation as η_{persp}; then we can add the inequality $\eta \geq \eta_{\text{persp}}$ to the MIP solver. In some scenarios, we must obtain a high-quality solution before the MIP solver terminates. A near-optimal lower bound, such as η_{persp}, will help close the optimality gap of the best integer solution found. A stronger relaxation lower bound can be obtained via SDP relaxations, i.e., [2, 13, 28, 44], and we do not exploit these options due to the high computational cost of solving SDPs.

5.1.4. *Decomposition:* In terms of the diagonal decomposition, i.e., $Q = R + \mathrm{diag}(d)$ such that $R \succeq 0$ and $d \geq 0$, several approaches exist. For example, Frangioni and Gentile [19] propose solving a simple SDP while Zheng et al. [44] propose solving a larger-sized SDP, which seeks the diagonal decomposition such that the lower bound of the resulting perspective relaxation is maximized. Zheng et al. [44] also test the convex combination of those two. In our experiments, the decomposition method in [19] attains a better solution time for almost all the instances and solution methods. Thus, we only present the results using the method in [19] to extract a diagonal term. The rank-one decomposition is obtained by doing Cholesky decomposition on the remainder term R, i.e., $Q = \sum_{i=1}^{K} L_i L_i^{\top} + \mathrm{diag}(\delta)$ such that $\sum_{i=1}^{K} L_i L_i^{\top} = R$.

Algorithm 3 An outer approximation algorithm (OA)

1: Set $t = 1$
2: Set $\mathcal{Z}_1 = \{(x, y, \eta) : Ay \leq b, Cy \leq Dx, x \in \mathrm{conv}(X), \eta \geq \eta_{\mathrm{persp}}\}$
3: **while** $\eta^t < \bar{f}_{\mathrm{persp}}(x^t)$ **do**
4:    Compute the solution (x^t, η^t) of

$$\min_{x \in X, \eta} \quad \eta \quad \text{s.t.} \quad (x, y, \eta) \in \mathcal{Z}_t \text{ for some } y$$

5:    Add a cut t using Algorithm 1 or 2

$$\mathcal{Z}_t = \mathcal{Z}_t \bigcap \{(x, y, \eta) : \eta \geq \langle t, x - x^t \rangle + \bar{f}_{\mathrm{persp}}(x^t)\}$$

6:    Set $t = t + 1$
7: **end while**

5.2. **Portfolio selection problems.** Given a universe of n risky assets, we denote the mean return vector as μ and the covariance matrix as Q. The mean-variance portfolio selection problem with sparsity, minimum investment, maximum investment, and minimum return constraints can be formulated as follows:

$$\min_{x,y} \quad y^{\top} Q y \tag{40a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} y_i = 1 \tag{40b}$$

$$\mu^{\top} y \geq \rho \tag{40c}$$

$$\alpha_i x_i \leq y_i \leq u_i x_i, \quad \forall i \in [n] \tag{40d}$$

$$\sum_{i=1}^{n} x_i \leq k \tag{40e}$$

$$x \in \{0, 1\}^n. \tag{40f}$$

The random instances we use are generated the same way as the diagonal-dominant instances in [19]. The expected return (μ_i), minimum investment (α_i), and maximum investment (u_i) are drawn from uniform distributions with ranges [0.002, 0.01], [0.075, 0.125], and [0.375, 0.425], for each $i = 1, \ldots, n$, respectively. The minimum return value ρ is also from a uniform distribution over the interval [0.002, 0.01]. The off-diagonal entries of Q are drawn from a discrete uniform distribution between 1 and 10, and the diagonal entries are drawn from a discrete uniform distribution between $10n$ and $20n$. Since $\alpha_i \geq 0.075$, the number of nonzeros in the vector y is at most 13. We test with cardinality constraints when $k = 6, 8, 10$ and the case without a cardinality constraint, which we denote by 'nc'. For $n = 300, 400$, we reuse these publicly available instances at https://commalab.di.unipi.it/datasets/MV/ for benchmarking, and for larger problem sizes, which are not available in the public data set, we generate new random instances following the same scheme.

We compare the following solution methods.

- MISOCP: the mixed-integer second-order conic reformulation of the perspective reformulation of (40) solved by Gurobi.

- OA-BC: the OA method proposed by Bertsimas and Cory-Wright [3].

- OA-persp: Algorithm 3 with Algorithm 1 as the cut generating subroutine.

- OA-persp+ro: Algorithm 3 with Algorithm 2 as the cut generating subroutine.

Note that the original OA algorithm in [3] is implemented in Julia, and here, we re-implement it in Python for a fair comparison with the other methods implemented in Python. Authors in [3] also supply additional cuts at the root node to strengthen the original polyhedral approximation of $\mathcal{Z}$. We also test the OA algorithm with additional cuts generated by an in-out bundle method at the root node; however, although this implementation improves the solution time in the OA procedure, it takes more time overall because the cost of generating additional cuts is substantial. Another difference in the experiment settings is the choice of MIP solver. In [3, 19, 44], the MIP solver in use is CPLEX, whereas here we use Gurobi.

In Table 1, we present the comparison between MISOCP, OA-BC, OA-persp, and OA-persp+ro, and each entry represents an average of ten instances. A time limit of ten minutes is enforced, and the parenthesis followed by the solution time indicates the number of instances that hit the time limit before reaching an optimal solution. The method that attains the best solution time is shown in bold. We provide instance-wise performance in Appendix B for a more detailed comparison.

By comparing Table 1 with the tables in [3] (the most recent computational study on the sparse portfolio selection problem), we discover that with the recent developments of Gurobi in mixed-integer conic programming, we are now able to explore more than twice as many nodes as CPLEX version 12.8.0. run on a machine with a better configuration within the same amount of time. Also, the comparison between MISOCP and OA-BC is reversed in some instances (MISOCP now attains a better end gap than OA-BC in some instances when they both reach the time limit). Consistent with the findings in [3], MISOCP performs worst in the case where there is no explicit sparsity constraint but an implicit sparsity constraint because of the minimum investment constraints, and its performance significantly deteriorates (with an end gap $\approx 10\%$) when the problem size scales up to 500. This phenomenon suggests that OA methods are better for the sparse portfolio selection problem in large instances.

Another observation we make is that OA-BC hits the time limit in most instances. An even closer look at Table 1 and the tables in [3] reveals that for both the original implementation of OA-BC and our re-implementation, it takes roughly the same amount of branch-and-bound nodes and cutting planes to obtain an optimal solution; however, on our computing environment, we are only able to explore less than half of the branch-and-bound nodes within ten minutes.

As we have shown, the cuts generated by OA-BC and OA-persp are the same, but the cut-generating procedure in the latter one is much more efficient (it solves a QP subproblem of size k as opposed to the QP subproblem of size n in OA-BC). The computational results further confirm that OA-persp significantly outperforms OA-BC. OA-persp solves most instances within the time limit, and its improved efficiency against OA-BC allows the exploration of more branch-and-bound nodes and the generation of more cuts. For many instances, both the smaller ones ($n = 300$) and the larger ones ($n = 500$), OA-persp reduces the solution time by one-half or more (this reduction is underestimated as OA-BC fails to solve a large proportion of instances).

Finally, even though OA-persp is the winning algorithm in most cases, the performances of OA-perp and OA-persp+ro tie closely (with only a few instances in which OA-persp+ro takes approximately 10% extra time than OA-persp). Although OA-persp+ro generates stronger cuts than OA-persp, and accordingly maintains a stronger linear programming relaxation, the additional cost of solving the linear system in line 16 of Algorithm 2 may slow down the branch-and-bound process. One trend we can observe from Table 1 is that the gap between the solution times of OA-persp and OA-persp+ro diminishes as the problem size scales up.

TABLE 1. Comparison results for portfolio selection problems with minimum/maximum return and investment constraints.

Size	k	MISOCP			OA-BC				OA-persp				OA-persp+ro			
		Gap(%)	Time(s)	Node	Gap	Time	Node	Cut	Gap	Time	Node	Cut	Gap	Time	Node	Cut
300	6	0.13	413.57(3)	561	0.12	360.43(3)	6337	847	0.00	**31.60**	7330	945	0.00	33.62	7193	938
	8	0.16	470.77(5)	571	0.23	435.59(4)	10760	1081	0.00	**75.52**	18485	1566	0.00	81.36	18191	1551
	10	0.11	466.51(4)	571.	0.32	463.79(6)	17337	1081	0.09	**250.33(2)**	43794	2144	0.09	286.12(2)	47414	2092
	nc	1.81	600.14(10)	533	0.12	231.94(2)	250993	239	0.00	139.36	278212	251	0.00	**134.59**	247410	253
400	6	0.53	519.15(8)	523	0.47	405.62(6)	5447	750	0.00	**72.24**	11766	1699	0.00	86.53	11205	1700
	8	0.39	508.30(8)	521	0.41	431.89(6)	5869	780	0.00	**225.35**	30389	2855	0.00	247.60	30171	2829
	10	0.20	504.27(8)	518	0.26	442.80(6)	13698	743	0.08	**253.89(3)**	38300	1621	0.08	264.43(3)	41238	1587
	nc	1.49	540.83(9)	503	0.41	276.14(4)	421272	104	0.37	250.31(4)	473113	106	0.41	**250.14(4)**	478389	107
500	6	0.73	531.00(9)	482.57	0.34	451.93(8)	5283	638	0.00	**111**	15414	1843	0.02	129.28(1)	16171	1807
	8	9.93	534.05(9)	466	0.32	487.36(6)	6130	649	0.07	**167.37(1)**	26360	2227	0.07	190.01(1)	28313	2145
	10	9.95	516.10(9)	529	0.37	516.44(9)	8534	648	0.11	349.92(2)	54403	1918	0.11	**347.95(2)**	59129	1910
	nc	14.22	515.54(9)	441	1.12	391.93(7)	270996	161	1.05	**352(4)**	363384	184	0.93	357.82(4)	380651	161

6. Conclusion

In this paper, we consider constrained mixed-integer convex quadratic programs (MIQP) with indicators. The motivation for this work arises from a sequence of convexification results on MIQP and the scalability of outer approximation algorithms. We develop a framework to derive projective cutting planes for MIQP, and as special cases, we derive the cutting planes based on perspective reformulation and perspective reformulation with rank-one inequalities. Our approach has some apparent advantages over existing ones. First, we can handle different constraints (potentially nonlinear constraints) by representing them as indicator functions in the objective as long as we can readily compute the resulting subdifferentials. Second, our result applies to general mixed-integer convex programs (MICP) with indicators. Third, an explicit form of the marginal function is not required to derive cutting planes. Last, it can be easily adapted to other strong reformulations in the literature. The computational results in §6 show that the theory translates into promising algorithmic improvements.

The theoretical framework is not restricted to the problem formulation (MIQP) we study here. One possible extension is applying the framework to other MICP problems, i.e., when the objective is separable convex or the composite of a convex function and a linear mapping. It is also interesting for future research to study the cutting planes for problems with nonlinear constraints. A theoretical question to answer is, can we relax to the level set boundedness condition? By answering it, we can obtain cutting planes for a broader class of functions. From a computational perspective, an idea for improving the algorithm is to exploit the sparsity of the Hessian. For example, intuitively, the rank-one inequality (15) is stronger when it involves a small proportion of variables, and we can use the sparse Cholesky decomposition to decompose the Hessian into sparse rank-one quadratics.

Acknowledgement

References

[1] K. M. Anstreicher and S. Burer. Quadratic optimization with switching variables: The convex hull for $n = 2$. *Mathematical Programming*, 188:421–441, 2021.

[2] A. Atamtürk and A. Gómez. Rank-one convexification for sparse regression. *arXiv preprint arXiv:1901.10334*, 2019.

[3] D. Bertsimas and R. Cory-Wright. A scalable algorithm for sparse portfolio selection. *INFORMS Journal on Computing*, 34(3):1489–1511, 2022.

[4] D. Bertsimas and R. Shioda. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43(1):1–22, 2009.

[5] D. Bertsimas and B. Van Parys. Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Annal of Statistics*, 48(1):300–323, 2020.

[6] D. Bertsimas and R. Weismantel. *Optimization over integers*. Dynamic Ideas Belmont, 2005.

[7] D. Bertsimas, A. King, and R. Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44:813–852, 2016.

[8] D. Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, 74(2):121–140, 1996.

[9] B. Borchers and J. E. Mitchell. A computational comparison of branch and bound and outer approximation algorithms for 0–1 mixed integer nonlinear programs. *Computers & Operations Research*, 24(8):699–701, 1997.

[10] S. Ceria and J. Soares. Convex programming for disjunctive convex optimization. *Mathematical Programming*, 86:595–614, 1999.

[11] X. Chen, D. Ge, Z. Wang, and Y. Ye. Complexity of unconstrained $l_2 - l_p$ minimization. *Mathematical Programming*, 143(1):371–383, 2014.

[12] A. De Rosa and A. Khajavirad. Explicit convex hull description of bivariate quadratic sets with indicator variables. *arXiv preprint arXiv:2208.08703*, 2022.

[13] H. Dong, K. Chen, and J. Linderoth. Regularization vs. relaxation: A conic optimization perspective of statistical variable selection. *arXiv preprint arXiv:1510.06083*, 2015.

[14] M. A. Duran and I. E. Grossmann. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36:307–339, 1986.

[15] M. Fischetti, I. Ljubić, and M. Sinnl. Redesigning benders decomposition for large-scale facility location. *Management Science*, 63(7):2146–2162, 2017.

[16] R. Fletcher and S. Leyffer. Solving mixed integer nonlinear programs by outer approximation. *Mathematical Programming*, 66:327–349, 1994.

[17] R. Fletcher and S. Leyffer. Numerical experience with lower bounds for miqp branch-and-bound. *SIAM Journal on Optimization*, 8(2):604–616, 1998.

[18] A. Frangioni and C. Gentile. Perspective cuts for a class of convex 0–1 mixed integer programs. *Mathematical Programming*, 106:225–236, 2006.

[19] A. Frangioni and C. Gentile. SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Operations Research Letters*, 35:181–185, 2007.

[20] A. Frangioni, F. Furini, and C. Gentile. Approximated perspective relaxations: A project and lift approach. *Computational Optimization and Applications*, 63:705–735, 2016.

[21] A. Frangioni, C. Gentile, and J. Hungerford. Decompositions of semidefinite matrices and the perspective reformulation of nonseparable quadratic programs. *Mathematics of Operations Research*, 45(1):15–33, 2020.

[22] U. Friedrich and D. Kreber. Projective cutting planes for general qp with indicator constraints.

[23] F. Glover. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22(4):455–460, 1975.

[24] G. H. Golub and C. F. Van Loan. *Matrix computations*. JHU press, 2013.

[25] A. Gómez. Strong formulations for conic quadratic optimization with indicator variables. *Mathematical Programming*, pages 193–226, 2021.

[26] A. Gómez and A. Atamtürk. Supermodularity and valid inequalities for quadratic optimization with indicators. *Mathematical Programming*, 201:295–338, 2023.

[27] O. Günlük and J. Linderoth. Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Mathematical Programming*, 124:183–205, 2010.

[28] S. Han, A. Gómez, and A. Atamtürk. 2x2-convexifications for convex quadratic optimization with indicator variables. *Mathematical Programming*, pages 95–134, 2023.

[29] H. Hazimeh, R. Mazumder, and A. Saab. Sparse regression at scale: Branch-and-bound rooted in first-order optimization. *Mathematical Programming*, 196(1-2):347–388, 2022.

[30] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I: Fundamentals*, volume 305. Springer Science & Business Media, 2013.

[31] H. Jeon, J. Linderoth, and A. Miller. Quadratic cone cutting surfaces for quadratic programs with on–off constraints. *Discrete Optimization*, 24:32–50, 2017.

[32] S. Küçükyavuz, A. Shojaie, H. Manzour, and L. Wei. Consistent second-order conic integer programming for learning Bayesian networks. *Journal of Machine Learning Research*, 2023. Forthcoming.

[33] P. Liu, S. Fattahi, A. Gómez, and S. Küçükyavuz. A graph-based decomposition method for convex quadratic optimization with indicators. *Mathematical Programming*, 200:669–701, 2023.

[34] H. Manzour, S. Küçükyavuz, H.-H. Wu, and A. Shojaie. Integer programming for learning directed acyclic graphs from continuous data. *INFORMS Journal on Optimization*, 3(1):46–73, 2021.

[35] A. Miller. *Subset selection in regression*. CRC Press, 2002.

[36] M. Pilanci, M. J. Wainwright, and L. El Ghaoui. Sparse learning via boolean relaxations. *Mathematical Programming*, 151(1):63–87, 2015.

[37] R. T. Rockafellar. *Convex analysis*, volume 18. Princeton university press, 1970.

[38] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[39] J. P. Vielma, S. Ahmed, and G. L. Nemhauser. A lifted linear programming branch-and-bound algorithm for mixed-integer conic quadratic programs. *INFORMS Journal on Computing*, 20:438–450, 2008.

[40] L. Wei, A. Gómez, and S. Küçükyavuz. On the convexification of constrained quadratic optimization problems with indicator variables. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 433–447. Springer, 2020.

[41] L. Wei, A. Gómez, and S. Küçükyavuz. Ideal formulations for constrained convex optimization problems with indicator variables. *Mathematical Programming*, 192(1-2):57–88, 2022.

[42] L. Wei, A. Atamtürk, A. Gómez, and S. Küçükyavuz. On the convex hull of convex quadratic optimization problems with indicators. *Mathematical Programming*, pages 1–35, 2023.

[43] W. Xie and X. Deng. Scalable algorithms for the sparse ridge regression. *SIAM Journal on Optimization*, 30:3359–3386, 2020.

[44] X. Zheng, X. Sun, and D. Li. Improving the performance of miqp solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach. *INFORMS Journal on Computing*, 26(4):690–703, 2014.

Appendix A. Equivalence of Cutting Planes

Following Bertsimas and Cory-Wright [3], and for ease of exposition, we choose $\mathrm{diag}(\delta) = \frac{1}{\gamma} I$ for some $\gamma > 0$ (the proof of the general δ case is similar). It's easy to verify that the level set boundedness condition holds for f_{persp_2}, and by mimicking the steps of the proof of Proposition 1, we conclude that $t \in \partial \bar{f}_{\mathrm{persp}_2}(x)$ at some x if and only if

$$\exists y \in \mathbb{R}^n, \alpha \in \mathbb{R}^{k_1}, \lambda \in \mathbb{R}^{m_1}, \mu \in \mathbb{R}^{m_2} \tag{41a}$$

$$y_i = 0 \quad \forall i \in S^C \tag{41b}$$

$$r = \beta - Ey \tag{41c}$$

$$Ay \le b \tag{41d}$$

$$Cy \le Dx \tag{41e}$$

$$\lambda \ge 0, \mu \ge 0 \tag{41f}$$

$$\lambda_i(a_i^\top y - b_i) = 0, \qquad \forall i \in [m_1] \tag{41g}$$

$$\mu_i(c_i^\top y - d_i^\top x) = 0, \quad \forall i \in [m_2] \tag{41h}$$

$$2\mathrm{diag}\left(\left\{\frac{1}{\gamma x_i}\right\}_{i \in S}\right) y_S + E_S^\top \alpha + A_S^\top \lambda + C_S^\top \mu + \tilde{g}_S = 0 \tag{41i}$$

$$2r + \alpha = 0 \tag{41j}$$

$$t_i = -\frac{1}{\gamma}\frac{y_i^2}{x_i^2} - D_i^\top \mu + c_i \quad \forall i \in S \tag{41k}$$

$$t_i \le -\frac{\gamma}{4}(X_i^\top \alpha + A_i^\top \lambda + C_i^\top \mu + g_i)^2 - D_i^\top \mu + c_i. \quad \forall i \in S^C. \tag{41l}$$

Conditions (41a)–(41j) are the KKT conditions for the following regression problem in a reduced space:

$$\min_y \quad \sum_{i \in S} \frac{1}{\gamma}\frac{y_i^2}{x_i} + \|r\|_2^2 + \tilde{g}_S^\top y \tag{42}$$

$$\text{s.t.} \quad r = \beta - E_S y$$

$$A_S y \le b$$

$$C_S y \le Dx,$$

and y_S, α, λ, and μ are the optimal primal and dual solutions. By comparing with Theorem 1 in [3], we can see that $\partial \bar{f}_{\mathrm{persp}_2}$ attains the same set of cutting planes.

Appendix B. Supplementary computational results

We give instance-wise comparisons for the instances generated by Frangioni and Gentile [19].

Table 2. Instance-wise comparison results for portfolio selection problems when $n = 300$.

Problem	k	MISOCP			OA-BC				OA-persp				OA-persp+ro			
		Gap	Time	Node	Gap	Time	Node	Cut	Gap	Time	Node	Cut	Gap	Time	Node	Cut
pard300-0	6	0.16	600.04	654	0.45	600.38	9427	1513	0.00	63.75	14554	1992	0.00	79.19	16502	1993
pard300-0	8	0.30	600.07	635	0.65	600.01	10587	1503	0.00	190.04	39244	3466	0.00	202.97	36332	3416
pard300-0	10	0.14	600.02	564	0.45	600.61	20672	1358	0.00	233.07	54078	2314	0.00	387.61	48568	2263
pard300-0	nc	0.48	600.18	498	0.64	600.70	508165	393	0.00	387.09	609246	475	0.00	471.19	634535	472
pard300-1	6	0.00	358.01	551	0.14	600.34	7244	891	0.00	51.50	6644	915	0.00	31.37	6003	925
pard300-1	8	0.06	600.07	628	0.36	600.34	14410	1493	0.00	80.85	21515	1878	0.00	90.29	21973	1829
pard300-1	10	0.00	486.69	622	0.22	600.21	30051	1354	0.00	129.66	39900	1510	0.00	164.98	52168	1468
pard300-1	nc	0.17	600.11	577	0.00	107.59	237566	116	0.00	67.04	252109	117	0.00	64.63	253601	117
pard300-2	6	0.37	600.10	518	0.00	460.12	12199	1160	0.00	36.85	12538	1149	0.00	41.39	11514	1155
pard300-2	8	0.20	600.19	475	0.00	585.66	29930	1400	0.00	76.23	29629	1380	0.00	80.21	29450	1398
pard300-2	10	0.32	600.05	480	0.62	600.00	20871	1411	0.18	600.00	96071	3649	0.19	600.03	99064	3455
pard300-2	nc	0.82	600.15	442	0.00	182.56	120650	288	0.00	59.91	96237	272	0.00	56.76	99555	285
pard300-3	6	0.61	600.04	727	0.57	600.25	12023	1511	0.00	69.45	17349	1963	0.00	72.77	15152	1921
pard300-3	8	0.46	600.07	704	0.76	600.01	12290	1495	0.00	190.47	38768	3198	0.00	199.00	38904	3178
pard300-3	10	0.27	600.16	712	0.64	600.29	15660	1460	0.00	362.65	53598	3738	0.00	437.73	64013	3740
pard300-3	nc	8.16	600.27	577	0.00	135.29	77614	275	0.00	35.35	77204	276	0.00	34.21	74277	282
pard300-4	6	0.00	220.47	537	0.00	166.65	3011	423	0.00	11.48	3074	447	0.00	11.66	2827	429
pard300-4	8	0.00	150.93	504	0.00	201.27	4923	504	0.00	14.78	4841	494	0.00	14.93	5097	490
pard300-4	10	0.00	313.88	563	0.00	310.42	18959	712	0.00	36.47	12623	748	0.00	42.76	15618	759
pard300-4	nc	0.64	600.15	593	0.00	50.10	46204	83	0.00	13.68	39924	78	0.00	11.22	33926	86
pard300-5	6	0.00	134.41	521	0.00	173.29	2965	437	0.00	11.49	3050	428	0.00	12.16	2934	418
pard300-5	8	0.00	265.28	574	0.00	301.53	6647	757	0.00	23.44	7598	745	0.00	24.93	7533	756
pard300-5	10	0.39	600.06	718	0.77	600.26	19129	1437	0.69	600.01	63240	5211	0.66	600.01	69370	4972
pard300-5	nc	0.97	600.04	606	0.00	183.41	211273	266	0.00	70.42	213307	270	0.00	87.83	205647	258
pard300-6	6	0.13	600.04	469	0.00	133.93	1813	335	0.00	8.67	1721	330	0.00	10.68	1561	331
pard300-6	8	0.26	600.10	459	0.00	129.34	2246	318	0.00	9.00	1904	289	0.00	9.88	2193	292
pard300-6	10	0.00	307.04	365	0.00	91.63	2168	216	0.00	10.60	2750	261	0.00	9.80	3004	221
pard300-6	nc	2.43	600.27	396	0.00	21.63	3455	41	0.00	3.65	3985	45	0.00	3.32	2400	42
pard300-7	6	0.00	270.95	576	0.00	244.85	3842	620	0.00	16.59	3469	629	0.00	19.60	3899	613
pard300-7	8	0.00	462.82	653	0.00	418.45	8075	1049	0.00	32.32	7947	1046	0.00	37.14	11121	1035
pard300-7	10	0.00	317.12	579	0.00	430.79	16710	1004	0.00	71.12	25051	1041	0.00	65.78	20430	1006
pard300-7	nc	1.26	600.14	613	0.00	243.62	115781	298	0.00	85.81	73289	308	0.00	117.56	98849	315
pard300-8	6	0.00	500.10	568	0.00	404.61	7295	1025	0.00	31.58	7667	1046	0.00	38.90	7913	1032
pard300-8	8	0.30	600.02	599	0.50	600.36	12524	1497	0.00	112.81	27362	2333	0.00	127.49	23651	2333
pard300-8	10	0.00	568.35	572	0.44	600.31	24616	1364	0.00	442.80	85912	2473	0.00	534.75	98009	2552
pard300-8	nc	1.66	600.13	486	0.00	194.54	164052	316	0.00	80.78	165395	315	0.00	68.98	144705	332
pard300-9	6	0.00	251.57	490	0.00	219.89	3555	561	0.00	14.66	3240	560	0.00	18.45	3632	568
pard300-9	8	0.00	228.12	482	0.00	318.92	5977	800	0.00	25.26	6042	835	0.00	26.72	5659	785
pard300-9	10	0.00	271.77	538	0.00	203.33	4540	501	0.00	16.89	4721	496	0.00	17.75	3904	484
pard300-9	nc	1.49	600.00	551	0.56	600.01	1025175	316	0.00	589.91	1251427	356	0.00	430.20	926608	344

TABLE 3. Instance-wise comparison results for portfolio selection problems when $n = 400$.

Problem	k	MISOCP			OA-BC				OA-persp				OA-persp+ro			
		Gap	Time	Node	Gap	Time	Node	Cut	Gap	Time	Node	Cut	Gap	Time	Node	Cut
pard400-0	6	0.67	600.23	625	0.85	600.33	5162	1126	0.00	114.78	16102	2789	0.00	129.38	15852	2776
pard400-0	8	0.53	600.13	691	0.69	600.35	7677	1109	0.00	390.52	47432	4873	000	411.78	51595	4781
pard400-0	10	0.09	600.20	587	0.00	575.93	34205	930	0.00	68.89	20167	893	0.00	106.13	32837	943
pard400-0	nc	2.79	600.15	601	1.23	600.01	820944	202	1.14	600.00	1066345	221	1.19	600.01	899022	250
pard400-1	6	0.00	28.46	15	0.00	14.59	51	21	0.00	0.88	49	21	0.00	0.81	69	19
pard400-1	8	0.00	67.75	31	0.00	15.71	26	20	0.00	0.86	89	20	0.00	1.08	167	19
pard400-1	10	0.00	91.25	45	0.00	10.36	97	12	0.00	0.92	126	15	0.00	0.84	103	15
pard400-1	nc	0.00	5.21	1	0.00	18.19	449	23	0.00	0.97	378	17	0.00	1.02	364	16
pard400-2	6	1.09	600.01	330	0.00	87.69	1038	157	0.00	4.56	896	163	0.00	4.85	931	167
pard400-2	8	0.58	600.19	327	0.00	161.98	2080	278	0.00	9.68	2268	263	0.00	9.84	1947	279
pard400-2	10	0.14	600.24	332	0.00	208.95	10134	336	0.00	21.00	7510	314	0.00	19.44	7014	304
pard400-2	nc	0.31	600.31	307	0.00	26.04	1616	32	0.00	2.97	1793	22	0.00	2.62	1071	24
pard400-3	6	0.37	600.37	559	0.52	600.30	8111	1117	0.00	63.02	13294	1723	0.00	64.94	12359	1724
pard400-3	8	0.33	600.32	521	0.57	600.02	7347	1087	0.00	251.99	38879	3672	0.00	250.50	36804	3631
pard400-3	10	0.43	600.27	549	0.60	600.47	10242	1015	0.49	600.08	55783	3727	0.50	600.02	52198	3530
pard400-3	nc	1.16	600.24	530	0.00	101.38	34868	115	0.00	22.26	22862	118	0.00	38.40	39031	121
pard400-4	6	0.39	600.25	715	0.00	335.62	5146	627	0.00	21.17	4993	646	0.00	25.37	5081	648
pard400-4	8	0.32	600.22	722	0.00	526.71	10625	951	0.00	40.43	11851	973	0.00	42.31	9458	965
pard400-4	10	0.29	600.00	713	0.44	600.33	18136	1009	0.00	304.55	65101	2084	0.00	320.14	70781	2111
pard400-4	nc	3.44	600.46	673	1.09	600.00	773944	233	1.09	600.03	836874	247	1.17	600.02	902718	228
pard400-5	6	0.00	361.13	317	0.00	16.54	39	25	0.00	0.92	65	24	0.00	2.01	56	24
pard400-5	8	0.00	213.83	295	0.00	12.74	55	14	0.00	0.77	31	14	0.00	1.42	34	14
pard400-5	10	0.00	150.25	297	0.00	30.91	243	48	0.00	1.69	328	41	0.00	3.90	499	61
pard400-5	nc	1.33	600.02	394	0.00	43.53	2240	56	0.00	7.34	5186	68	0.00	7.87	3325	66
pard400-6	6	0.66	600.22	719	0.73	600.16	7635	1122	0.00	104.90	14965	2559	0.00	168.34	12845	2532
pard400-6	8	0.50	600.19	716	0.65	600.56	6986	1113	0.00	219.46	34701	3380	0.00	296.05	31447	3350
pard400-6	10	0.12	600.17	738	0.22	600.15	15369	1051	0.00	74.9	20279	1264	0.00	109.27	22383	1182
pard400-6	nc	1.92	600.37	715	0.82	600.02	1304288	78	0.54	600.01	1421540	82	0.73	600.01	1566134	76
pard400-7	6	1.07	600.25	488	1.00	600.13	6699	1098	0.00	147.22	21138	3189	0.00	157.47	19730	3234
pard400-7	8	0.66	600.32	432	0.86	600.47	7015	1077	0.00	457.43	52436	5581	0.00	510.10	55535	5492
pard400-7	10	0.36	600.00	460	0.49	600.59	20688	993	0.00	266.85	60777	2109	0.00	284.53	70818	2060
pard400-7	nc	1.22	600.52	423	0.00	119.40	44938	105	0.00	64.92	98069	94	0.00	46.19	36017	101
pard400-8	6	0.44	600.34	746	0.71	600.42	13874	1116	0.00	94.72	19385	2380	0.00	129.77	17848	2396
pard400-8	8	0.48	600.03	741	0.62	600.03	9496	1103	0.00	386.84	62479	4231	0.00	428.76	61514	4246
pard400-8	10	0.33	600.01	733	0.45	600.25	16468	1040	0.17	600.02	80269	3012	0.10	600.01	75529	2998
pard400-8	nc	0.45	600.68	699	0.00	52.82	1867	74	0.00	4.54	1825	67	0.00	5.23	1744	67
pard400-9	6	0.63	600.21	721	0.85	600.39	6717	1095	0.00	170.26	26773	3505	0.00	182.38	27287	3482
pard400-9	8	0.51	600.00	740	0.68	600.29	7388	1051	0.00	495.54	53731	5548	0.00	524.17	53215	5515
pard400-9	10	0.21	600.35	733	0.38	600.02	11406	996	0.11	600.03	72666	2755	0.17	600.02	80224	2675
pard400-9	nc	2.31	600.34	694	0.99	600.02	1227572	128	0.92	600.01	1276260	127	0.95	600.02	1334464	129