2×2-CONVEXIFICATIONS FOR CONVEX QUADRATIC OPTIMIZATION WITH INDICATOR VARIABLES

SHAONING HAN, ANDRÉS GÓMEZ AND ALPER ATAMTÜRK

Abstract. In this paper, we study the convex quadratic optimization problem with indicator variables. For the 2×2 case, we describe the convex hull of the epigraph in the original space of variables, and also give a conic quadratic extended formulation. Then, using the convex hull description for the 2×2 case as a building block, we derive an extended SDP relaxation for the general case. This new formulation is stronger than other SDP relaxations proposed in the literature for the problem, including the optimal perspective relaxation and the optimal rank-one relaxation. Computational experiments indicate that the proposed formulations are quite effective in reducing the integrality gap of the optimization problems.

Keywords. Mixed-integer quadratic optimization, semidefinite programming, perspective formulation, indicator variables, convexification.

April 2020; December 2021

1. Introduction

We consider the convex quadratic optimization with indicators:

$$\text{(QI)} \quad \min \left\{ a'x + b'y + y'Qy : (x,y) \in \mathcal{I}_n \right\}, \tag{1}$$

where the indicator set is defined as

$$\mathcal{I}_n = \left\{ (x,y) \in \{0,1\}^n \times \mathbb{R}^n_+ : y_i(1-x_i) = 0,\ \forall i \in [n] \right\},$$

where a and b are n-dimensional vectors, $Q \in \mathbb{R}^{n \times n}$ is a positive semidefinite (PSD) matrix and $[n] := \{1, 2, \ldots, n\}$. For each $i \in [n]$, the complementarity constraint $y_i(1-x_i) = 0$, along with the indicator variable $x_i \in \{0,1\}$, is used to state that $y_i = 0$ whenever $x_i = 0$. Numerous applications, including

S. Han, A. Gómez : Daniel J. Epstein Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California, CA 90089. gomezand@usc.edu, shaoning@usc.edu

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

portfolio optimization [12], optimal control [25], image segmentation [33], signal denoising [9] are either formulated as (QI) or can be relaxed to (QI).

Building strong convex relaxations of (QI) is instrumental in solving it effectively. A number of approaches for developing linear and nonlinear valid inequalities for (QI) are considered in literature. Dong and Linderoth [21] describe lifted linear inequalities from its continuous quadratic optimization counterpart with bounded variables. Bienstock and Michalka [13] derive valid linear inequalities for optimization of a convex objective function over a non-convex set based on gradients of the objective function. Valid linear inequalities for (QI) can also be obtained using the epigraph of bilinear terms in the objective [e.g. 14, 19, 29, 38]. In addition, several specialized results concerning optimization problems with indicator variables exist in the literature [7, 10, 11, 16, 18, 26, 27, 36, 39].

There is a substantial body of research on the perspective formulation of convex univariate functions with indicators [1, 20, 21, 22, 28, 32, 43]. When Q is diagonal, $y'Qy$ is separable and the perspective formulation provides the convex hull of the epigraph of $y'Qy$ with indicator variables by strengthening each term $Q_{ii}y_i^2$ with its perspective counterpart $Q_{ii}y_i^2/x_i$, individually. For the general case, however, convex relaxations based on the perspective reformulation may not be strong. The computational experiments in [24] demonstrate that as Q deviates from a diagonal matrix, the performance of the perspective formulation deteriorates.

Beyond the perspective reformulation, which is based on the convex hull of the epigraph of a univariate convex quadratic function with one indicator variable, the convexification for the 2×2 case has received attention recently. Convex hulls of univariate and 2×2 cases can be used as building blocks to strengthen (QI) by decomposing $y'Qy$ into a sequence of low-dimensional terms. Castro et al. [17] study convexification of a special class of two-term quadratic function controlled by a single indicator variable. Jeon et al. [35] give conic quadratic valid inequalities for the 2×2 case. Frangioni et al. [24] combine perspective reformulation and disjunctive programming and apply them to the 2×2 case. Atamtürk and Gómez [5] study the convex hull of the mixed-integer epigraph of $(y_1 - y_2)^2$ with indicators. Atamtürk et al. [4] give the convex hull of the more general set

$$\mathcal{Z}_- := \left\{(x, y, t) \in \mathcal{I}_2 \times \mathbb{R}_+ : t \geq d_1 y_1^2 - 2y_1y_2 + d_2 y_2^2 \right\},$$

with coefficients $d \in \mathcal{D} := \{d \in \mathbb{R}^2 : d_1 \geq 0, d_2 \geq 0, d_1 d_2 \geq 1\}$. The conditions on the coefficients d_1, d_2 imply convexity of the quadratic. Atamtürk and Gómez [6] study the case where the continuous variables are free and the rank of the coefficient matrix is one in the context of sparse linear regression. Anstreicher and Burer [3] give an extended SDP formulation for the convex hull of the 2×2 bounded set $\left\{(y, yy', xx') : 0 \le y \le x \in \{0,1\}^2\right\}$. Their formulation does not assume convexity of the quadratic function and contain PSD matrix variables X and Y as proxies for xx' and yy' as additional variables. Anstreicher and Burer [2] study computable representations of convex hulls of low dimensional quadratic forms without indicator variables. More general convexifications for low-rank quadratic functions [8, 31] or quadratic functions with tridiagonal matrices [37] have also been proposed.

To design convex relaxations for (QI) based on convexifications for simpler substructures, a standard approach is to decompose the matrix Q as $Q = R + \sum_{j\in J} Q_j$, for some index set J, where $R, Q_j \succeq 0$, $j \in J$. After writing problem (1) as

$$\min\ a'x + b'y + y'Ry + \sum_{j\in J} t_j \tag{2a}$$

$$\text{s.t. } t_j \ge y'Q_jy \qquad \forall j \in J \tag{2b}$$

$$(x,y) \in \mathcal{I}_n,\ t \in \mathbb{R}^J_+, \tag{2c}$$

formulation (2) can then be strengthened based on convexifications of the simpler structures induced by constraints (2b) (e.g., matrices Q_j are diagonal or 2×2). There are two main approaches to implement convexifications based on (2). On the one hand, one may choose fixed R, Q_j, $j \in J$, *a priori* and treat them as parameters, as done in [8, 23, 24, 37, 44], resulting in simpler formulations (e.g., conic quadratic representable) that may be amenable to use with off-the-shelf solvers for mixed-integer optimization. On the other hand, one may treat matrices R,Q_j, $j \in J$, as decision variables that are chosen with the goal of obtaining the optimal relaxation bound after strengthening, as done in [4, 6, 20]. The resulting formulations with the second approach are stronger but more complex (e.g., SDP representable). In general, neither approach is preferable to the other.

Contributions. The contributions of this paper is two-fold.

1. 2×2 case: We describe the convex hull of the epigraph of a convex bivariate quadratic with a positive cross product and indicators. Consider

$$\mathcal{Z}_+ := \left\{(x, y, t) \in \mathcal{I}_2 \times \mathbb{R}_+ : t \ge d_1y_1^2 + 2y_1y_2 + d_2y_2^2 \right\},$$

where $d \in \mathcal{D}$. Observe that any bivariate convex quadratic with positive off-diagonals can be written as $d_1y_1^2 + 2y_1y_2 + d_2y_2^2$, by scaling appropriately. Therefore, $\mathcal{Z}_+$ is the *complementary* set to $\mathcal{Z}_-$ and, together, $\mathcal{Z}_+$ and $\mathcal{Z}_-$ model epigraphs of all bivariate convex quadratics with indicators and nonnegative continuous variables.

In this paper, we propose *conic quadratic* extended formulations to describe $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$ and $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$. These extended formulations are more compact than alternatives previously proposed in the literature. More importantly, a distinguishing contribution of this paper is that we also give the explicit description of $\mathrm{cl\,conv}(\mathcal{Z}_+)$ in the original space of the variables. The corresponding convex envelope of the bivariate function is a four-piece function. While convexifications in the original space of variables are more difficult to implement using current off-the-shelf mixed-integer optimization solvers, they offer deeper insights on the structure of the convex hulls. Whereas the ideal formulations of $\mathcal{Z}_-$ can be conveniently described with two simpler valid "extremal" inequalities [4], a similar result does not hold for $\mathcal{Z}_+$ (see Example 1 in §3). The derivation of ideal formulations for the more involved set $\mathcal{Z}_+$ differs significantly from the methods in [4]. The complementary results of this paper and [4] for $\mathcal{Z}_-$ complete the convex hull descriptions of bivariate convex functions with indicators and nonnegative continuous variables.

2. General case: We develop an optimal SDP relaxation based on 2×2 *convexifications for (QI).* In order to construct a strong convex formulation for (QI), we extract a sequence of 2×2 PSD matrices from Q such that the residual term is a PSD matrix as well, and convexify each bivariate quadratic term utilizing the descriptions of $\mathrm{cl\,conv}(\mathcal{Z}_+)$ and $\mathrm{cl\,conv}(\mathcal{Z}_-)$. This approach works very well when Q is 2×2 PSD decomposable, i.e., when Q is scaled-diagonally dominant [15]. Otherwise, a natural question is how to optimally decompose $y'Qy$ into bivariable convex quadratics and a residual convex quadratic term so as to achieve the best strengthening.

We address this question by deriving an optimal convex formulation using SDP duality. The new SDP formulation dominates any formulation obtained through a 2×2-decomposition scheme. This formulation is also stronger than other SDP formulations in the literature, including the optimal perspective formulation [20] and the optimal rank-one convexification [6]. In addition, the proposed formulation is solved many orders of magnitude faster than the 2×2-decomposition approaches based on disjunctive programming [24], and delivers higher quality bounds than standard mixed-integer optimization approaches in difficult portfolio index tracking problems.

Outline. The rest of the paper is organized as follows. In §2 we review the convex hull results on $\mathcal{Z}_-$ and illustrate the structural difference between $\mathcal{Z}_+$ and $\mathcal{Z}_-$. In §3 we provide a conic quadratic formulation of $\mathrm{cl\,conv}(\mathcal{Z}_+)$ and $\mathrm{cl\,conv}(\mathcal{Z}_-)$ in an extended space and derive the explicit form of $\mathrm{cl\,conv}(\mathcal{Z}_+)$ in the original space. In §4, employing the results in §3, we give a strong convex relaxation for (QI) using SDP techniques. In §5, we compare the strength of the proposed SDP relaxation with others in literature. In §6, we present computational results demonstrating the effectiveness of the proposed convex relaxations. Finally, in §7, we conclude with a few final remarks.

Notation. To simplify the notation throughout, we adopt the following convention for division by 0: given $x \geq 0$, $x^2/0 = \infty$ if $x \neq 0$ and $x^2/0 = 0$ if $x = 0$. Thus, x^2/z, the closure of the perspective of x^2, is a closed convex function [see 40, pages 67-68]. For a set $\mathcal{X} \subseteq \mathbb{R}^n$, $\operatorname{cl}\operatorname{conv}(\mathcal{X})$ denotes the closure of the convex hull of $\mathcal{X}$. For a vector v, $\operatorname{diag}(v)$ denotes the diagonal matrix V with $V_{ii} = v_i$ for each i. Finally, $\mathbb{S}^n_+$ refers to the cone of $n \times n$ real symmetric PSD matrices.

2. Preliminaries

In this section, we review the existing results on convex hulls of sets $\mathcal{Z}_-$, $\mathcal{Z}_+$, and their relaxation $\mathcal{Z}_f$ with free continuous variables:

$$\mathcal{Z}_f := \left\{(x,y,t) \in \{0,1\}^2 \times \mathbb{R}^3 : t \geq d_1 y_1^2 \pm 2y_1y_2 + d_2 y_2^2,\ y_i(1-x_i) = 0, i \in [2]\right\}.$$

Note that when the continuous variables are free, the sign associated with the cross term $2y_1y_2$ is irrelevant, since one can state it equivalently with the opposite sign by substituting $\bar{y}_i = -y_i$. In contrast, if $y \geq 0$, such a substitution is not possible; hence, the need for separate analyses for sets $\mathcal{Z}_+$ and $\mathcal{Z}_-$.

We first point out that all three sets can be naturally seen as disjunctions of four convex sets corresponding to the four possible values for $x \in \{0,1\}^2$. Thus, a direct application of disjunctive programming yields similar (conic quadratic) representations of the three sets [24] but such representations require several additional variables. While the disjunctive approach might suggest that $\mathcal{Z}_f$, $\mathcal{Z}_+$, $\mathcal{Z}_-$ may be similar, we now argue that the sign of the cross terms materially affect the complexity of the optimization problems as well as the structure of the convex hulls.

2.1. Optimization.

The sign of the off-diagonals of matrix Q critically affect the complexity of the optimization problem (QI). We first state a result concerning optimization with Stieltjes matrices Q, first proven in [5].

Proposition 1 (Atamtürk and Gómez [5])**.** *Problem* (1) *can be solved in polynomial time if* $Q \succ 0$ *and* $Q_{ij} \leq 0$ *for all* $i \neq j$ *and* $b \leq 0$*.*

In contrast, an analogous result does not hold if the off-diagonal terms of matrix Q are nonnegative.

Proposition 2. *Problem* (1) *is* $\mathcal{NP}$*-hard if* $Q \succ 0$ *and* $Q_{ij} \geq 0$ *for all* $i \neq j$ *and* $b \leq 0$*.*

Proof. We show that (QI) includes the $\mathcal{NP}-$hard subset sum problem as a special case under the assumptions of the proposition: given $w \in \mathbb{Z}_+^n, K \in \mathbb{Z}_+$, solve the equation

$$w'x = K,\ x \in \{0,1\}^n. \tag{3}$$

Set $Q = (I + qq')/2 \succ 0$ where $q \in \mathbb{R}_{++}^n$ is a parameter to be specified later. Let $p_i = q_i^2,\ i \in [n]$, $b = -q$ and $a = \gamma p$ for some $\gamma > 0$ to be specified later as well. For a vector $z \in \mathbb{R}^n$ and matrix $M \in \mathbb{S}_+^n$, let z_S and M_S denote the subvector and principle submatrix defined by $S \subseteq [n]$, respectively. Then (QI) reduces to

$$\begin{aligned}
&\min_{(x,y)\in\mathcal{I}_n} \frac{1}{2}y'(I+qq')y - q'y + \gamma p'x \\
=&\min_{S\subseteq[n], y_S\ge 0} \frac{1}{2}y_S'(I_S + q_Sq_S')y_S - q_S'y_S + \gamma\sum_{i\in S}p_i && (S := \{i : x_i = 1\}) \\
=&\min_{S\subseteq[n]} -\frac{1}{2}q_S'\left(I_S + q_Sq_S'\right)^{-1}q_S + \gamma\sum_{i\in S}p_i && \text{(4a)} \\
=&\min_{S\subseteq[n]} -\frac{1}{2}q_S'\left(I_S - \frac{q_Sq_S'}{1+\|q_S\|_2^2}\right)q_S + \gamma\sum_{i\in S}p_i && \text{(Woodbury matrix identity)} \\
=&\min_{S\subseteq[n]} -\frac{1}{2}\frac{\|q_S\|_2^2}{1+\|q_S\|_2^2} + \gamma\|q_S\|_2^2 && (p_i = q_i^2) \\
=&\min_{S\subseteq[n]} \left[\frac{1}{2(1+\|q_S\|_2^2)} + \gamma(1+\|q_S\|_2^2)\right] - \gamma - \frac{1}{2}. && \text{(4b)}
\end{aligned}$$

Note that the nonnegativity constraints are dropped in (4a) because they are trivially satisfied by the optimal solution as

$$y_S = \left(I_S - \frac{q_Sq_S'}{1+\|q_S\|_2^2}\right)q_S = \frac{1}{1+\|q_S\|_2^2}q_S \ge 0.$$

Now, let $q_i = \sqrt{w_i}$, $i \in [n]$ and $\gamma = \frac{1}{2(1+K)^2}$. Then (4b) simplifies to (after dropping the constant term $-\gamma - 1/2$ and multiplying by 2)

$$\begin{aligned}
&= \min_S \frac{1}{1+w(S)} + \frac{1+w(S)}{(1+K)^2} \\
&\ge \frac{2}{1+K},
\end{aligned}$$

where the lower bound is attained if and only if $w(S) = K$. Hence, the subset sum problem (3) has a solution if and only if the optimal value of (QI) as constructed above equals $2/(1+K)$. $\square$

Propositions 1 and 2 suggest that convex hulls of sets with negative cross terms are substantially simpler than those with positive terms.

2.2. **Rank-one results.** It is convenient to formulate convex hulls of sets via conic quadratic constraints as they are readily supported by modern mixed-integer optimization software. While such representations are easy to obtain via disjunctive programming, the resulting formulations generally have a prohibitive number of variables and constraints, which hamper the performance of solvers. Therefore, it is of interest to find *the most compact* conic quadratic formulations. In this regard, as well, $\mathcal{Z}_+$ is significantly more complex than $\mathcal{Z}_f$ and $\mathcal{Z}_-$. Consider the existing results for the simpler sets in the rank-one case, i.e., $d_1 = d_2 = 1$.

Proposition 3. *Atamtürk and Gómez [6] If $d_1 = d_2 = 1$, then*

$$\operatorname{cl}\operatorname{conv}(\mathcal{Z}_f) = \left\{(x, y, t) \in [0,1]^2 \times \mathbb{R}^3 : t \geq (y_1 \pm y_2)^2,\ t(z_1 + z_2) \geq (y_1 \pm y_2)^2\right\}.$$

In particular, for the rank-one case with free continuous variables, the $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_f)$ is conic quadratic representable in the original space of variables, without the need for additional variables.

Proposition 4. *Atamtürk and Gómez [5] If $d_1 = d_2 = 1$, then*

$$\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-) = \left\{(x, y, t) \in [0,1]^2 \times \mathbb{R}+^3 : t \geq \phi(x_1, x_2, y_1, y_2)\right\},$$

where

$$\phi(x_1, x_2, y_1, y_2) = \begin{cases} (y_1 - y_2)^2/x_1 & \text{if } y_1 \geq y_2 \\ (y_1 - y_2)^2/x_2 & \text{if } y_1 \leq y_2. \end{cases}$$

Since constraints $t \geq (y_1 - y_2)^2/x_i$, $i \in [2]$, are not valid for $\mathcal{Z}_-$, $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$ is not conic quadratic representable in the original space of variables. A conic quadratic representation with two additional variables is given in [4].

In Section 3, Corollary 2, we describe $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ in the original space for the rank-one case. This description is more complex than $\mathcal{Z}_-$ as it requires four pieces instead of two and it is not conic-quadratic representable. We also provide a compact extended formulation with three additional variables.

2.3. **Full-rank results.** A description of $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$ in the original space of variables is given in [4]. Interestingly, it can be expressed as two valid inequalities involving function ϕ introduced in Proposition 4.

Proposition 5 (Atamtürk et al. [4])**.** *Set* $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$ *is described by bound constraints $y \geq 0$, $0 \leq x \leq 1$, and the two valid inequalities*

$$t \geq d_1 \phi(x_1, x_2, y_1, y_2/d_1) + \frac{y_2^2}{x_2}\left(d_2 - \frac{1}{d_1}\right),$$

$$t \geq d_2 \phi(x_1, x_2, y_1/d_2, y_2) + \frac{y_1^2}{x_1}\left(d_1 - \frac{1}{d_2}\right).$$

Proposition 5 reveals that $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$ requires only homogeneous functions that are sums of rank-one and perspective convexifications. In Section 3, Proposition 7, we give $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ in the original space of variables, and show that the resulting function does not have either of these properties. The discrepancy between the results highlights that $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ is fundamentally different from $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$, and helps explain why optimization with positive matrices Q (Proposition 2) is substantially more difficult than optimization with Stieltjes matrices (Proposition 1).

3. Convex hull description of $\mathcal{Z}_+$

In this section, we give ideal convex formulations for

$$\mathcal{Z}_+ = \left\{(x,y,t) \in \mathcal{I}_2 \times \mathbb{R}_+ : t \geq d_1 y_1^2 + 2y_1y_2 + d_2 y_2^2 \right\}.$$

When $d_1 = d_2 = 1$, $\mathcal{Z}_+$ reduces to the simpler rank-one set

$$\mathcal{X}_+ = \left\{(x,y,t) \in \mathcal{I}_2 \times \mathbb{R}_+ : t \geq (y_1 + y_2)^2 \right\}.$$

Set $\mathcal{X}_+$ is of special interest as it arises naturally in (QI) when Q is a diagonally dominant matrix, see computations in §6.1 for details. As we shall see, the convex hulls of $\mathcal{Z}_+$ and $\mathcal{X}_+$ are significantly more complicated than their complementary sets $\mathcal{Z}_-$ and $\mathcal{X}_-$ studied earlier. In §3.1, we develop an SOCP-representable extended formulation of $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$. Then, in §3.2, we derive the explicit form of $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ in the original space of variables.

3.1. **Conic quadratic-representable extended formulation.** We start by writing $\mathcal{Z}_+$ as the disjunction of four convex sets defined by all values of the indicator variables; that is,

$$\mathcal{Z}_+ = \mathcal{Z}_+^1 \cup \mathcal{Z}_+^2 \cup \mathcal{Z}_+^3 \cup \mathcal{Z}_+^4,$$

where $\mathcal{Z}_+^i, i = 1,2,3,4$ are convex sets defined as:

$$\begin{aligned}
\mathcal{Z}_+^1 &= \{(1,0,u,0,t_1) : t_1 \geq d_1 u^2, u \geq 0\},\\
\mathcal{Z}_+^2 &= \{(0,1,0,v,t_2) : t_2 \geq d_2 v^2, v \geq 0\},\\
\mathcal{Z}_+^3 &= \{(1,1,w_1,w_2,t_3) : t_3 \geq d_1 w_1^2 + 2w_1w_2 + d_2 w_2^2, w_1 \geq 0, w_2 \geq 0\},\\
\mathcal{Z}_+^4 &= \{(0,0,0,0,t_4) : t_4 \geq 0\}.
\end{aligned}$$

By the definition, a point $(x_1, x_2, y_1, y_2, t) \in \operatorname{conv}(\mathcal{Z}_+)$ if and only if it can be written as a convex combination of four points belonging in $\mathcal{Z}_+^i, i = 1,2,3,4$. Using $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ as the corresponding weights, $(x_1, x_2, y_1, y_2, t) \in \operatorname{conv}(\mathcal{Z}_+)$ if and only if the following inequality system has a feasible solution

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1 \tag{5a}$$
$$x_1 = \lambda_1 + \lambda_3,\ x_2 = \lambda_2 + \lambda_3 \tag{5b}$$
$$y_1 = \lambda_1 u + \lambda_3 w_1,\ y_2 = \lambda_2 v + \lambda_3 w_2 \tag{5c}$$
$$t = \lambda_1 t_1 + \lambda_2 t_2 + \lambda_3 t_3 + \lambda_4 t_4 \tag{5d}$$
$$t_1 \geq d_1 u^2,\ t_2 \geq d_2 v^2,\ t_3 \geq d_1 w_1^2 + 2w_1 w_2 + d_2 w_2^2,\ t_4 \geq 0 \tag{5e}$$
$$u, v, w_1, w_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0. \tag{5f}$$

We will now simplify (5). First, by Fourier–Motzkin elimination, one can substitute t_1, t_2, t_3, t_4 with their lower bounds in (5e) and reduce (5d) to $t \geq \lambda_1 d_1 u^2 + \lambda_2 d_2 v^2 + \lambda_3(d_1 w_1^2 + 2w_1 w_2 + d_2 w_2^2)$. Similarly, since $\lambda_4 \geq 0$, one can eliminate λ_4 and reduce (5a) to $\sum_{i=1}^3 \lambda_i \leq 1$. Next, using (5c), one can substitute $u = (y_1 - \lambda_3 w_1)/\lambda_1$ and $v = (y_2 - \lambda_3 w_2)/\lambda_2$. Finally, using (5b), one can substitute $\lambda_1 = x_1 - \lambda_3$ and $\lambda_2 = x_2 - \lambda_3$ to arrive at

$$\max\{0, x_1 + x_2 - 1\} \leq \lambda_3 \leq \min\{x_1, x_2\} \tag{6a}$$
$$\lambda_3 w_i \leq y_i, \quad i = 1, 2 \tag{6b}$$
$$w_i \geq 0, \quad i = 1, 2 \tag{6c}$$
$$t \geq \frac{d_1(y_1 - \lambda_3 w_1)^2}{x_1 - \lambda_3} + \frac{d_2(y_2 - \lambda_3 w_2)^2}{x_2 - \lambda_3} + \lambda_3(d_1 w_1^2 + 2w_1 w_2 + d_2 w_2^2), \tag{6d}$$

where (6a) results from the nonnegativity of $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, (6b) from the nonnegativity of u and v. Finally, observe that (6b) is redundant for (6): indeed, if there is a solution (λ, w, t) satisfying (6a), (6c) and (6d) but violating (6b), one can decrease w_1 and w_2 such that (6b) is satisfied without violating (6d).

Redefining variables in (6), we arrive at the following conic quadratic-representable extended formulation for $\mathrm{cl\,conv}(\mathcal{Z}_+)$ and its rank-one special case $\mathrm{cl\,conv}(\mathcal{X}_+)$.

Proposition 6. *The set* $\mathrm{cl\,conv}(\mathcal{Z}_+)$ *can be represented as*

$$\mathrm{cl\,conv}(\mathcal{Z}_+) = \Big\{(x, y, t) \in [0,1]^2 \times \mathbb{R}_+^3 : \exists \lambda \in \mathbb{R}_+, z \in \mathbb{R}_+^2 \ s.t.\ x_1 + x_2 - 1 \leq \lambda \leq \min\{x_1, x_2\},\ t \geq \tfrac{d_1(y_1 - z_1)^2}{x_1 - \lambda} + \tfrac{d_2(y_2 - z_2)^2}{x_2 - \lambda} + \tfrac{d_1 z_1^2 + 2z_1 z_2 + d_2 z_2^2}{\lambda}\Big\}.$$

Corollary 1. *The set* $\operatorname{cl}\operatorname{conv}(\mathcal{X}_+)$ *can be represented as*

$$\operatorname{cl}\operatorname{conv}(\mathcal{X}_+) = \Bigg\{(x,y,t) \in [0,1]^2 \times \mathbb{R}^3_+ : \exists \lambda \in \mathbb{R}_+, z \in \mathbb{R}^2_+ \ s.t.$$
$$x_1 + x_2 - 1 \le \lambda \le \min\{x_1, x_2\},$$
$$t \ge \frac{(y_1 - z_1)^2}{x_1 - \lambda} + \frac{(y_2 - z_2)^2}{x_2 - \lambda} + \frac{(z_1 + z_2)^2}{\lambda} \Bigg\}.$$

Remark 1. One can apply similar arguments to the complementary set $\mathcal{Z}_-$ to derive an SOCP representable formulation of its convex hull as

$$\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-) = \Bigg\{(x,y,t) \in [0,1]^2 \times \mathbb{R}^3_+ : \exists \lambda \in \mathbb{R}_+, z \in \mathbb{R}^2 \text{ s.t.}$$
$$x_1 + x_2 - 1 \le \lambda \le \min\{x_1, x_2\},$$
$$z_1 \le y_1,\ z_2 \le y_2,$$
$$t \ge \tfrac{d_1(y_1 - z_1)^2}{x_1 - \lambda} + \tfrac{d_2(y_2 - z_2)^2}{x_2 - \lambda} + \tfrac{d_1 z_1^2 - 2z_1 z_2 + d_2 z_2^2}{\lambda} \Bigg\}.$$

This extended formulation is smaller than the one given Atamtürk et al. [4] for $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$.

3.2. **Description in the original space of variables** x, y, t**.** The purpose of this section is to express $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ and $\operatorname{cl}\operatorname{conv}(\mathcal{X}_+)$ in the original space.

Let $\Lambda_x := \{\lambda \in \mathbb{R} : \max\{0, x_1 + x_2 - 1\} \le \lambda \le \min\{x_1, x_2\}\}$, i.e., the set of feasible λ implied by constraint (6a). Define

$$G(\lambda, w) := \frac{d_1(y_1 - \lambda w_1)^2}{x_1 - \lambda} + \frac{d_2(y_2 - \lambda w_2)^2}{x_2 - \lambda} + \lambda(d_1 w_1^2 + 2w_1 w_2 + d_2 w_2^2)$$

and $g(\lambda) : \Lambda_x \to \mathbb{R}$ as

$$g(\lambda) := \min_{w \in \mathbb{R}^2_+} G(\lambda, w).$$

Note that as G is SOCP-representable, it is convex. We first prove an auxiliary lemma that will be used in the derivation.

Lemma 1. *Function* $g(\lambda)$ *is non-decreasing over* Λ_x.

Proof. Note that for any fixed w and $\lambda < \min\{x_1, x_2\}$, we have

$$\begin{aligned}
\frac{\partial G(\lambda, w)}{\partial \lambda} &= \frac{d_1[2(\lambda w_1 - y_1)w_1(x_1-\lambda) + (y_1 - \lambda w_1)^2]}{(x_1-\lambda)^2} \\
&\quad + \frac{d_2[2(\lambda w_2 - y_2)w_2(x_2-\lambda) + (y_2 - \lambda w_2)^2]}{(x_2-\lambda)^2} \\
&\quad + (d_1 w_1^2 + 2w_1w_2 + d_2 w_2^2) \\
&= \frac{d_1[w_1^2(x_1-\lambda)^2 + 2(\lambda w_1 - y_1)w_1(x_1-\lambda) + (y_1 - \lambda w_1)^2]}{(x_1-\lambda)^2} \\
&\quad + \frac{d_2[w_2^2(x_2-\lambda_2)^2 + 2(\lambda w_2 - y_2)w_2(x_2-\lambda) + (y_2 - \lambda w_2)^2]}{(x_2-\lambda)^2} \\
&\quad + 2w_1w_2 \\
&= \frac{d_1(w_1x_1 - y_1)^2}{(x_1-\lambda)^2} + \frac{d_2(w_2x_2 - y_2)^2}{(x_2-\lambda)^2} + 2w_1w_2 \\
&\geq 0.
\end{aligned}$$

Therefore, for fixed w, $G(\cdot, w)$ is nondecreasing. Now for $\tilde{\lambda} \leq \hat{\lambda}$, let $\tilde{w}$ and $\hat{w}$ be optimal solutions defining $g(\tilde{\lambda})$ and $g(\hat{\lambda})$. Then,

$$g(\tilde{\lambda}) = G(\tilde{\lambda}, \tilde{w}) \leq G(\tilde{\lambda}, \hat{w}) \leq G(\hat{\lambda}, \hat{w}) = g(\hat{\lambda}),$$

proving the claim. □

We now state and prove the main result in this subsection.

Proposition 7. *Define*

$$f(x,y,\lambda;d) := \frac{(d_1d_2-1)(d_1x_2y_1^2 + d_2x_1y_2^2) + 2\lambda d_1d_2y_1y_2 + \lambda(d_1y_1^2 + d_2y_2^2)}{(d_1d_2-1)x_1x_2 - \lambda^2 + \lambda(x_1+x_2)},$$

and

$$f_+^*(x,y;d) := \begin{cases} \frac{d_1y_1^2}{x_1} + \frac{d_2y_2^2}{x_2} & \text{if } x_1 + x_2 \leq 1 \\ \frac{d_1y_1^2}{1-x_2} + \frac{d_2y_2^2}{x_2} & \text{if } 0 \leq x_1 + x_2 - 1 \leq (x_1y_2 - d_1x_2y_1)/y_2 \\ \frac{d_1y_1^2}{x_1} + \frac{d_2y_2^2}{1-x_1} & \text{if } 0 \leq x_1 + x_2 - 1 \leq (x_2y_1 - d_2x_1y_2)/y_1 \\ f(x,y,x_1+x_2-1) & \text{o.w.} \end{cases}$$

Then, the set $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ *can be expressed as*

$$\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+) = \{(x,y,t) \in [0,1]^2 \times \mathbb{R}_+^3 : t \geq f_+^*(x_1,x_2,y_1,y_2;d_1,d_2)\}.$$

Proof. First, observe that we may assume $x_1, x_2 > 0$, as otherwise $x_1 + x_2 \leq 1$ and f_+^* reduces to the perspective function for the univariate case. To find the representation in the original space of variables, we first project out variables z in Proposition 6. Specifically, notice that $g(\lambda)$ can be rewritten in the following form by letting $z_i = \lambda w_i, i = 1, 2$:

$$g(\lambda) = \min \ \frac{d_1(y_1 - z_1)^2}{x_1 - \lambda} + \frac{d_2(y_2 - z_2)^2}{x_2 - \lambda} + \frac{d_1 z_1^2 + 2z_1 z_2 + d_2 z_2^2}{\lambda} \quad \text{s.t. } z_i \geq 0, \ \ i = 1, 2. \ \ (s_i) \tag{7}$$

By Proposition 6, a point $(x, y, t) \in [0, 1]^2 \times \mathbb{R}^3_+$ belongs to $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ if and only if $t \geq \min_{\lambda \in \Lambda_x} \ g(\lambda)$. We first assume $x_1 + x_2 - 1 > 0$, which implies $\lambda > 0, \forall \lambda \in \Lambda_x$. For given $\lambda \in \Lambda_x$, optimization problem (7) is convex with affine constraints, thus Slater condition holds. Hence, the following KKT conditions are necessary and sufficient for the minimizer:

$$\frac{2d_1}{x_1 - \lambda}(z_1 - y_1) + \frac{2(d_1 z_1 + z_2)}{\lambda} - s_1 = 0 \tag{8a}$$

$$\frac{2d_2}{x_2 - \lambda}(z_2 - y_2) + \frac{2(d_2 z_2 + z_1)}{\lambda} - s_2 = 0 \tag{8b}$$

$$z_1 s_1 = 0 \tag{8c}$$

$$z_2 s_2 = 0 \tag{8d}$$

$$s_i, z_i \geq 0, \ \ i = 1, 2. \tag{8e}$$

Let us analyze the KKT system considering the positiveness of s_1 and s_2.

- *Case* $s_1 > 0$. By (8c), $z_1 = 0$ and by (8a), $z_2 > 0$, which implies $s_2 = 0$ from (8d). Hence, (8a) and (8b) reduce to

$$\frac{2z_2}{\lambda} = \frac{2d_1}{x_1 - \lambda} y_1 + s_1$$

$$\frac{2d_2}{x_2 - \lambda}(z_2 - y_2) + \frac{2d_2 z_2}{\lambda} = 0.$$

  Solving these two linear equations, we get $z_2 = \frac{y_2}{x_2}\lambda$ and $s_1 = 2(\frac{y_2}{x_2} - \frac{d_1 y_1}{x_1 - \lambda})$. This also indicates $s_1 \geq 0$ iff $\lambda \leq (x_1 y_2 - d_1 x_2 y_1)/y_2$. By replacing the variables with their optimal values in the objective function (7), we find that

$$g(\lambda) = \frac{d_1 y_1^2}{x_1 - \lambda} + \frac{d_2}{x_2 - \lambda}\left(y_2 - \frac{y_2}{x_2}\lambda\right)^2 + \frac{d_2}{\lambda}\left(\frac{y_2}{x_2}\lambda\right)^2 \tag{9a}$$

$$= \frac{d_1 y_1^2}{x_1 - \lambda} + \frac{d_2 y_2^2}{x_2} \tag{9b}$$

  when $\lambda \in [0, (x_1 y_2 - d_1 x_2 y_1)/y_2] \cap \Lambda_x$.
- *Case* $s_2 > 0$. Similarly, we find that

$$g(\lambda) = \frac{d_1 y_1^2}{x_1} + \frac{d_2 y_2^2}{x_2 - \lambda} \tag{10}$$

  when $\lambda \in [0, (x_2 y_1 - d_2 x_1 y_2)/y_1] \cap \Lambda_x$.

- *Case* $s_1 = s_2 = 0$. In this case, (8a) and (8b) reduce to

$$\begin{pmatrix} d_1x_1 & x_1-\lambda \\ x_2-\lambda & d_2x_2 \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \lambda \begin{pmatrix} d_1y_1 \\ d_2y_2 \end{pmatrix}.$$

If $\lambda > 0$, the determinant of the matrix is $(d_1d_2-1)x_1x_2 + \lambda(x_1 + x_2 - \lambda) > 0$ and the system has a unique solution. It follows that

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \lambda \begin{pmatrix} d_1x_1 & x_1-\lambda \\ x_2-\lambda & d_2x_2 \end{pmatrix}^{-1}\begin{pmatrix} d_1y_1 \\ d_2y_2 \end{pmatrix},$$

i.e.,

$$z_1 = \frac{\lambda(d_1d_2x_2y_1 + (\lambda - x_1)d_2y_2)}{(d_1d_2-1)x_1x_2 - \lambda^2 + \lambda(x_1+x_2)},$$

$$z_2 = \frac{\lambda(d_1d_2x_1y_2 + (\lambda - x_2)d_1y_1)}{(d_1d_2-1)x_1x_2 - \lambda^2 + \lambda(x_1+x_2)}.$$

Therefore, the bounds $z_1, z_2 \geq 0$ imply lower bounds

$$\lambda \geq (x_1y_2 - d_1x_2y_1)/y_2, \quad \lambda \geq (x_2y_1 - d_2x_1y_2)/y_1$$

on λ. Moreover, from (8a) and (8b), we have

$$\frac{d_1(y_1-z_1)}{x_1-\lambda} = \frac{d_1z_1+z_2}{\lambda} \text{ and } \frac{d_2(y_2-z_2)}{x_2-\lambda} = \frac{d_2z_2+z_1}{\lambda}.$$

By substituting the two equalities in (7), we find that

$$\begin{aligned} g(\lambda) =& \big(d_1y_1z_1 + y_1z_2 + d_2y_2z_2 + y_2z_1\big)/\lambda \\ =& \frac{(d_1d_2-1)(d_1x_2y_1^2 + d_2x_1y_2^2) + 2\lambda d_1d_2y_1y_2 + \lambda(d_1y_1^2 + d_2y_2^2)}{(d_1d_2-1)x_1x_2 - \lambda^2 + \lambda(x_1+x_2)}. \end{aligned}$$

Therefore,

$$g(\lambda) = f(x, y, \lambda; d) \tag{11}$$

when $\lambda \in [\max\{(x_1y_2 - d_1x_2y_1)/y_2, (x_2y_1 - d_2x_1y_2)/y_1\}, +\infty) \cap \Lambda_x$.

To see that the three pieces of $g(\lambda)$ considered above are, indeed, mutually exclusive, observe that when $\lambda \leq (x_1y_2 - d_1x_2y_1)/y_2$, this is, $\frac{y_2(x_1-\lambda)}{x_2y_1} \geq d_1$, we have $\frac{d_2y_2}{y_1}\frac{x_1-\lambda}{x_2} \geq d_1d_2 \geq 1$. Since $\frac{x_1-\lambda}{x_2}\frac{x_2-\lambda}{x_1} \leq \frac{x_1}{x_2}\frac{x_2}{x_1} = 1$, it holds $\frac{d_2y_2}{y_1}\frac{x_1-\lambda}{x_2} \geq \frac{x_1-\lambda}{x_2}\frac{x_2-\lambda}{x_1}$, that is, $\lambda \geq (x_2y_1 - d_2x_1y_2)/y_1$.

Finally, notice when $x_1 + x_2 - 1 \leq 0$, λ may take the value 0. In this case, (7) reduces to

$$g(0) = \frac{d_1y_1^2}{x_1} + \frac{d_2y_2^2}{x_2}.$$

By Lemma 1, $\min_{\lambda\in\Lambda_x} g(\lambda) = g(\max\{0, x_1 + x_2 - 1\})$. Combining this fact with the above discussion, Proposition 7 holds. □

Remark 2. For further intuition, we now comment on the validity of each piece of $t \geq f_+^*(x,y;d)$ over $[0,1]^2 \times \mathbb{R}_+^3$ for $\mathcal{Z}_+$. Because the first piece can be obtained by dropping the nonnegative cross product term y_1y_2 and then strengthening $t \geq y_1^2 + y_2^2$ using perspective reformulation, it is valid everywhere. When $x_1 + x_2 < 1$ and $y_1, y_2 > 0$, $t \geq y_i^2/x_i + y_j^2/(1-x_i) > f_+^*(x,y;1,1)$ for $i \neq j$. Therefore, the second and the third pieces are not valid on the domain $[0,1]^2 \times \mathbb{R}_+^3$.

If $d_1d_2 > 1$, the last piece $t \geq f(x,y,x_1+x_2-1;d)$ is not valid for $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ everywhere, as seen by exhibiting a point $(x,y,t) \in \operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ violating $t \geq f(x,y,x_1+x_2-1;d)$. To do so, let

$$(x_1,x_2,y_1,y_2,t) = (0.5, \frac{1}{d_1d_2+1} + \epsilon, \frac{1}{\sqrt{d_1}}, 2\sqrt{d_1}, f_+^*(x,y)),$$

where $\epsilon > 0$ is small enough so that $x_1 + x_2 < 1$, i.e., $x_2 < 0.5$. With this choice, $f_+^*(x,y) = d_1y_1^2/x_1 + d_2y_2^2/x_2$. Let $\tilde{\lambda} = x_1 + x_2 - 1$, then $\tilde{\lambda}(x_1+x_2) - \tilde{\lambda}^2 = \tilde{\lambda}$. Hence, for point (x,y,t), we have

$$\begin{aligned}
f(x,y,\tilde{\lambda};d) =& \frac{(d_1d_2-1)(d_1x_2y_1^2 + d_2x_1y_2^2) + 2\tilde{\lambda}d_1d_2y_1y_2 + \tilde{\lambda}(d_1y_1^2 + d_2y_2^2)}{(d_1d_2-1)x_1x_2 + \tilde{\lambda}} \\
=& \frac{(d_1d_2-1)x_1x_2(d_1y_1^2/x_1 + d_2y_2^2/x_2) + \tilde{\lambda}(d_1y_1^2 + 2d_1d_2y_1y_2 + d_2y_2^2)}{(d_1d_2-1)x_1x_2 + \tilde{\lambda}} \\
=& (1-\alpha)f_+^*(x,y) + \alpha(d_1y_1^2 + 2d_1d_2y_1y_2 + d_2y_2^2),
\end{aligned}$$

where $\alpha = \tilde{\lambda}/((d_1d_2-1)x_1x_2 + \tilde{\lambda})$. Since $\tilde{\lambda} < 0$, $\alpha < 0$ if and only if

$$\begin{aligned}
& (d_1d_2-1)x_1x_2 + x_1 + x_2 - 1 > 0 \\
\iff & d_1d_2 > \frac{(1-x_1)(1-x_2)}{x_1x_2} = \frac{1}{x_2} - 1 \quad (\text{by } x_1 = 0.5) \\
\iff & x_2 > \frac{1}{d_1d_2+1},
\end{aligned}$$

which is true by the choice of x_2. Moreover,

$$\begin{aligned}
f_+^*(x,y) = & d_1y_1^2/x_1 + d_2y_2^2/x_2 = 2 + 8d_1d_2 \\
> & d_1y_1^2 + 2d_1d_2y_1y_2 + d_2y_2^2 = 1 + 8d_1d_2.
\end{aligned}$$

This indicates $f(x,y,\tilde{\lambda};d) > (1-\alpha)f_+^*(x,y) + \alpha f_+^*(x,y) = f_+^*(x,y) = t$, that is, $t \geq f(x,y,x_1+x_2-1;d)$ is violated.

Observe that if $d_1d_2 = 1$, then $f(x,y,x_1+x_2-1;d)$ reduces to the original quadratic $d_1y_1^2 + 2y_1y_2 + d_2y_2$. Otherwise, although $t \geq f(x,y,x_1+x_2-1;d)$ appears complicated, the next proposition implies that it is convex over its restricted domain and can, in fact, be stated as an SDP constraint. This results strongly indicates that SOCP-representable relaxations of (QI) may be inadequate to describe the convex hull of the relevant mixed-integer sets, unless a large number of additional variables are added. The proof of Proposition 8 can be found in the Appendix.

Proposition 8. *If* $d_1 d_2 > 1$ *and* $x_1 + x_2 - 1 > 0$, *then* $t \geq f(x, y, x_1 + x_2 - 1; d)$ *can be rewritten as the SDP constraint*

$$\begin{pmatrix} t/(d_1 d_2 - 1) & y_1 & y_2 \\ y_1 & d_2 x_1 + x_2/d_1 - 1/d_1 & -x_1 - x_2 + 1 \\ y_2 & -x_1 - x_2 + 1 & x_1/d_2 + d_1 x_2 - 1/d_2 \end{pmatrix} \succeq 0.$$

From Proposition 7, we get the convex hull of rank-one case $\mathcal{X}_+$ by setting $d_1 = d_2 = 1$.

Corollary 2.

$$\operatorname{cl}\operatorname{conv}(\mathcal{X}_+) = \left\{(x, y, t) \in [0, 1]^2 \times \mathbb{R}^3_+ : t \geq f_{1+}(x, y)\right\},$$

where

$$f_{1+}(x, y) = \begin{cases} \frac{y_1^2}{x_1} + \frac{y_2^2}{x_2} & \text{if } x_1 + x_2 \leq 1 \\ \frac{y_2^2}{x_2} + \frac{y_1^2}{1 - x_2} & \text{if } 0 \leq x_1 + x_2 - 1 \leq (x_1 y_2 - x_2 y_1)/y_2 \\ \frac{y_1^2}{x_1} + \frac{y_2^2}{1 - x_1} & \text{if } 0 \leq x_1 + x_2 - 1 \leq (x_2 y_1 - x_1 y_2)/y_1 \\ (y_1 + y_2)^2 & \text{o.w.} \end{cases}$$

3.3. **Rank-one approximations of $\mathcal{Z}_+$.** We now consider valid inequalities analogous to the ones given in Proposition 5 for $\mathcal{Z}_-$. Consider the two decompositions of the bivariate quadratic function given by

$$\begin{aligned} d_1 y_1^2 + 2 y_1 y_2 + d_2 y_2^2 &= d_1 (y_1 + \frac{y_2}{d_1})^2 + (d_2 - \frac{1}{d_1}) y_2^2 \\ &= d_2 (\frac{y_1}{d_2} + y_2)^2 + (d_1 - \frac{1}{d_2}) y_2^2. \end{aligned}$$

Applying perspective reformulation and Corollary 2 to the separable and pairwise quadratic terms, respectively, one can obtain two simple valid inequalities for $\mathcal{Z}_+$:

$$t \geq d_1 f_{1+}(x_1, x_2, y_1, \frac{y_2}{d_1}) + (d_2 - \frac{1}{d_1}) \frac{y_2^2}{x_2} \tag{12a}$$

$$t \geq d_2 f_{1+}(x_1, x_2, \frac{y_1}{d_2}, y_2) + (d_1 - \frac{1}{d_2}) \frac{y_1^2}{x_1}. \tag{12b}$$

The following example shows that the inequalities above do not describe $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$, highlighting the more complicated structure of $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ compared to its complementary set $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$.

Example 1. Consider $\mathcal{Z}_+$ with $d_1 = d_2 = d = 2$, and let $x_1 = x_2 = x = 2/3$, $y_1 = y_2 = y > 0$ and $t = f_+^*(x, y)$. Then $(x, y, t) \in \operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$. On the one hand, $x_1 + x_2 > 1$ implies

$$t = f_+^*(x, y) = f(2/3, 2/3, y, y, 1/3) = \frac{133}{11} y^2.$$

On the other hand, $f_{1+}(x, x, y, y/d) = (y + y/d)^2 = 9/2y^2$ indicates that (12) reduces to

$$t \geq \frac{27}{4} y^2.$$

Since $\frac{133}{11} y^2 > \frac{27}{4} y^2$, (12) holds strictly at this point. □

4. An SDP relaxation for (QI)

In this section, we will give an extended SDP relaxation for (QI) utilizing the convex hull results obtained in the previous section. Introducing a symmetric matrix variable Y, let us write (QI) as

$$\min \left\{ a'x + b'y + \langle Q, Y \rangle : Y \succeq yy', (x, y) \in \mathcal{I}_n \right\}. \tag{13}$$

Suppose for a class of PSD matrices $\Pi \subseteq \mathbb{S}_+^n$ we have an underestimator $f_P(x, y)$ for $y'Py$ for any $P \in \Pi$. Then, since $\langle P, Y \rangle \geq y'Py$, we obtain a valid inequality

$$f_P(x, y) - \langle P, Y \rangle \leq 0, \;\; P \in \Pi \tag{14}$$

for (13). For example, if Π is the set of diagonal PSD matrices and $f_P(x, y) = \sum_i P_{ii} y_i^2 / x_i$, for $P \in \Pi$, then inequality (14) is the perspective inequality.

Furthermore, since (14) holds for any $P \in \Pi$, one can take the supremum over all $P \in \Pi$ to get an optimal valid inequality of the type (14)

$$\sup_{P \in \Pi} f_P(x, y) - \langle P, Y \rangle \leq 0. \tag{15}$$

In the example of perspective reformulation, inequality (15) becomes

$$\sup_{P \succeq 0 \text{ diagonal}} \left\{ \sum_i P_{ii} \left(y_i^2 / x_i - Y_{ii} \right) \right\} \leq 0,$$

which can be further reduced to the closed form $y_i^2 \leq Y_{ii} x_i, \forall i \in [n]$. This leads to the the optimal perspective formulation [20]

$$\begin{aligned}
(\mathsf{OptPersp}) \qquad & \min\; a'x + b'y + \langle Q, Y \rangle && && \text{(16a)} \\
& \text{s.t. } Y - yy' \succeq 0 && && \text{(16b)} \\
& \quad y_i^2 \leq Y_{ii} x_i && \forall i \in [n] && \text{(16c)} \\
& \quad 0 \leq x \leq 1, \; y \geq 0. && && \text{(16d)}
\end{aligned}$$

Han et al. [30] show that OptPersp is equivalent to the Shor's SDP relaxation [41] for problem (1).

Letting Π be the class of 2×2 PSD matrices and $f_P(\cdot)$ as the function describing the convex hull of the mixed-integer epigraph of $y'Py$, one can derive new valid inequalities for (QI). Specifically, using the extended formulations for $f^*_+(x,y;d)$ and $f^*_-(x,y;d)$ describing $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ and $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$, we have

$$f^*_+(x,y;d) = \min_{z,\lambda} \frac{d_1(y_1-z_1)^2}{x_1-\lambda} + \frac{d_2(y_2-z_2)^2}{x_2-\lambda} + \frac{d_1z_1^2+2z_1z_2+d_2z_2^2}{\lambda} \tag{17a}$$

$$\text{s.t. } z_1 \geq 0, z_2 \geq 0 \tag{17b}$$

$$\max\{0, x_1+x_2-1\} \leq \lambda \leq \min\{x_1,x_2\}, \tag{17c}$$

and

$$f^*_-(x,y;d) = \min_{z,\lambda} \frac{d_1(y_1-z_1)^2}{x_1-\lambda} + \frac{d_2(y_2-z_2)^2}{x_2-\lambda} + \frac{d_1z_1^2-2z_1z_2+d_2z_2^2}{\lambda} \tag{18a}$$

$$\text{s.t. } z_1 \leq y_1, z_2 \leq y_2 \tag{18b}$$

$$\max\{0, x_1+x_2-1\} \leq \lambda \leq \min\{x_1,x_2\}. \tag{18c}$$

Since any 2×2 symmetric PSD matrix P can be rewritten in the form of $P = p\begin{pmatrix} d_1 & 1 \\ 1 & d_2 \end{pmatrix}$ or $P = p\begin{pmatrix} d_1 & -1 \\ -1 & d_2 \end{pmatrix}$, we can take $f_P(x,y) = pf^*_+(x,y;d)$ or $f_P(x,y) = pf^*_-(x,y;d)$, correspondingly. Since we have the explicit form of $f^*_+(\cdot)$ and $f^*_-(\cdot)$, for any fixed d, (14) gives a nonlinear valid inequality which can be added to (13). Alternatively, (17) and (18) can be used to reformulate these inequalities as conic quadratic inequalities in an extended space. Moreover, maximizing the inequalities gives the optimal valid inequalities among the class of of 2×2 PSD matrices stated below. Recall that $\mathcal{D} := \{d \in \mathbb{R}^2 : d_1 \geq 0, d_2 \geq 0, d_1d_2 \geq 1\}$.

Proposition 9. *For any pair of indices $i < j$, the following inequalities are valid for (QI):*

$$\max_{d\in\mathcal{D}} \left\{ f^*_+(x_i,x_j,y_i,y_j;d_1,d_2) - d_1Y_{ii} - d_2Y_{jj} - 2Y_{ij} \right\} \leq 0, \tag{19a}$$

$$\max_{d\in\mathcal{D}} \left\{ f^*_-(x_i,x_j,y_i,y_j;d_1,d_2) - d_1Y_{ii} - d_2Y_{jj} + 2Y_{ij} \right\} \leq 0. \tag{19b}$$

Optimal inequalities (19) may be employed effectively if they can be expressed explicitly. We will now show how to write inequalities (19) explicitly using an auxiliary 3×3 matrix variable W.

Lemma 2. *A point $(x_1,x_2,y_1,y_2,Y_{11},Y_{12},Y_{22})$ satisfies inequality (19a) if and only if there exists $W^+ \in \mathbb{S}^3_+$ such that the inequality system*

$$W^+_{12} \leq Y_{12} \tag{20a}$$

$$(Y_{11}-W^+_{11})(x_1-W^+_{33}) \geq (y_1-W^+_{31})^2, W^+_{11} \leq Y_{11}, W^+_{33} \leq x_1 \tag{20b}$$

$$(Y_{22}-W^+_{22})(x_2-W^+_{33}) \geq (y_2-W^+_{32})^2, W^+_{22} \leq Y_{22}, W^+_{33} \leq x_2 \tag{20c}$$

$$W^+_{31} \geq 0, W^+_{32} \geq 0 \tag{20d}$$

$$W^+_{33} \geq x_1+x_2-1 \tag{20e}$$

is feasible.

Lemma 3. *A point* $(x_1, x_2, y_1, y_2, Y_{11}, Y_{12}, Y_{22})$ *satisfies inequality* (19b) *if and only if there exists* $W^- \in \mathbb{S}^3_+$ *such that the inequality system*

$$Y_{12} \leq W^-_{12} \tag{21a}$$
$$(Y_{11} - W^-_{11})(x_1 - W^-_{33}) \geq (y_1 - W^-_{31})^2, W^-_{11} \leq Y_{11}, W^-_{33} \leq x_1 \tag{21b}$$
$$(Y_{22} - W^-_{22})(x_2 - W^-_{33}) \geq (y_2 - W^-_{32})^2, W^-_{22} \leq Y_{22}, W^-_{33} \leq x_2 \tag{21c}$$
$$W^-_{31} \leq y_1, W^-_{32} \leq y_2 \tag{21d}$$
$$W^-_{33} \geq x_1 + x_2 - 1 \tag{21e}$$

is feasible.

Proof of Lemma 2. The Lemma is proved by means of conic duality. For brevity, dual variables associated with each constraint are introduced in the formulation below. Writing f^*_+ as a conic quadratic minimization problem as in (17), we first express inequality (19a) as

$$\begin{aligned}
0 \geq \max_{d\in\mathcal{D}} \min_{t,\lambda,z} \quad & d_1t_1 + d_2t_2 + t_3 - d_1Y_{11} - d_2Y_{22} - 2Y_{12} \\
\text{s.t.} \quad & t_1(x_1 - \lambda) \geq (y_1 - z_1)^2, t_1 \geq 0, x_1 - \lambda \geq 0 && (d_1, s_1, \eta_1) \\
& t_2(x_2 - \lambda) \geq (y_2 - z_2)^2, t_2 \geq 0, x_2 - \lambda \geq 0 && (d_2, s_2, \eta_2) \\
& \lambda t_3 \geq \|B_+ z\|_2^2, \lambda \geq 0, t_3 \geq 0 && (1, s_3, \gamma) \\
& \lambda \geq x_1 + x_2 - 1 && (\alpha) \\
& z_1, z_2 \geq 0, && (r_1, r_2)
\end{aligned}$$

where $B^2_+ = \begin{pmatrix} d_1 & 1 \\ 1 & d_2 \end{pmatrix}$. Taking the dual of the inner minimization, the inequality can be written as

$$\begin{aligned}
0 \geq \max_{d\in\mathcal{D}} \max_{\alpha,\eta,\gamma,s,r} \quad & -\sum_{i=1,2}(x_is_i + 2y_i\eta_i) + (x_1 + x_2 - 1)\alpha - d_1Y_{11} - d_2Y_{22} - 2Y_{12} \\
\text{s.t.} \quad & d_is_i \geq \eta_i^2, \ i = 1, 2 \\
& s_3 \geq \|\gamma\|_2^2 \\
& r_1, r_2, \alpha \geq 0 \\
& \alpha + s_3 = s_1 + s_2 && (\lambda) \\
& \begin{pmatrix} r_1 - 2\eta_1 \\ r_2 - 2\eta_2 \end{pmatrix} = 2B_+ \begin{pmatrix} \gamma_1 \\ \gamma_2, \end{pmatrix}, && (z)
\end{aligned}$$

Note that one can obtain a strictly dual feasible solution by taking $s_i, i \in [3]$ sufficiently large. Due to Slater condition, we deduce that strong duality holds. We first assume $d_1d_2 > 1$. Then, the last equation implies $\gamma =$

$B_+^{-1}(r/2-\eta)$. Substituting out γ and s_3, and letting $u_i = \eta_i - r_i/2, i = 1, 2$, the maximization problem is further reduced to

$$\begin{aligned} 0 \geq \max_{d\in\mathcal{D}} \max_{\alpha,\eta,s,u} \; & -\sum_{i=1,2}(x_i s_i + 2y_i\eta_i) + (x_1+x_2-1)\alpha - d_1Y_{11} - d_2Y_{22} - 2Y_{12} \\ \text{s.t.} \; & d_i s_i \geq \eta_i^2, \; i = 1,2 \\ & \eta_i \geq u_i, \; i = 1,2 \\ & \alpha \geq 0 \\ & s_1 + s_2 - \alpha \geq u' \begin{bmatrix} d_1 & 1 \\ 1 & d_2 \end{bmatrix}^{-1} u. \end{aligned}$$

Applying Schur Complement Lemma to the last inequality, we reach

$$\begin{aligned} 2Y_{12} \geq \max_{\eta,s,u,r,d} \; & -\sum_{i=1,2}(x_i s_i + 2y_i\eta_i) + (x_1+x_2-1)\alpha - d_1Y_{11} - d_2Y_{22} \\ \text{s.t.} \; & d_i s_i \geq \eta_i^2, \; i = 1,2 && (p_i, q_i, w_i) \\ & \eta_i \geq u_i, \; i = 1,2 && (v_i) \\ & \alpha \geq 0 && (\beta) \\ & \begin{pmatrix} s_1+s_2-\alpha & u_1 & u_2 \\ u_1 & d_1 & 1 \\ u_2 & 1 & d_2 \end{pmatrix} \succeq 0. && (W^+) \end{aligned}$$

Note the SDP constraint implies $d \in D$. If $d_1d_2 = 1$, then B_+ is singular. In this case, one can apply the same argument to the Moore-Penrose pseudo inverse of B_+ (see p108, Ch12 and Corollary 15.3.2 in [40]) and use the generalized Schur Complement Lemma (see 7.3.P8 in [34]) to deduce the last SDP constraint. Finally, taking the SDP dual of the maximization problem we arrive at

$$\begin{aligned} 2Y_{12} \geq \min_{p,q,w,v,W^+} \; & 2W_{12}^+ \\ \text{s.t.} \; & p_i q_i \geq w_i^2, \; p_i, q_i \geq 0, \; i = 1,2 \\ & v_i \geq 0, \; i = 1,2 \\ & q_i + W_{33}^+ = x_i, \; i = 1,2 && (s_i) \\ & p_i + W_{ii}^+ = Y_{ii}, \; i = 1,2 && (d_i) \\ & 2w_i + v_i = 2y_i, \; i = 1,2 && (\eta_i) \\ & 2W_{3i}^+ = v_i, \; i = 1,2 && (u_i) \\ & \beta - W_{33}^+ = 1 - x_1 - x_2 && (\alpha) \\ & \beta \geq 0, \; W^+ \succeq 0. \end{aligned}$$

One can obtain a strictly primal feasible solution by taking $d_i, s_i, i = 1, 2$ sufficiently large, which implies strong SDP duality holds due to Slater condtion. Substituting out p, q, w, v, β, we arrive at (20). □

The proof of Lemma 3 is similar and is omitted for brevity. Since both (19a) and (19b) are valid, using (20) and (21) together, one can obtain an SDP relaxation of (QI). While inequalities in (20) and (21) are quite similar, in general, W^+ and W^- do not have to coincide. However, we show below that choosing $W^+ = W^-$, the resulting SDP formulation is still valid and it is at least as strong as the strengthening obtained by valid inequalities (19).

Let $\mathcal{W}$ be the set of points $(x_1, x_2, y_1, y_2, Y_{11}, Y_{12}, Y_{22})$ such that there exists a 3×3 matrix W satisfying

$$W_{12} = Y_{12} \tag{25a}$$
$$(Y_{11} - W_{11})(x_1 - W_{33}) \geq (y_1 - W_{31})^2, W_{11} \leq Y_{11}, W_{33} \leq x_1 \tag{25b}$$
$$(Y_{22} - W_{22})(x_2 - W_{33}) \geq (y_2 - W_{32})^2, W_{22} \leq Y_{22}, W_{33} \leq x_2 \tag{25c}$$
$$0 \leq W_{31} \leq y_1,\ 0 \leq W_{32} \leq y_2 \tag{25d}$$
$$W_{33} \geq x_1 + x_2 - 1 \tag{25e}$$
$$W \succeq 0 \tag{25f}$$

Then, using $\mathcal{W}$ for every pair of indices, we can define the strengthened SDP formulation

$$\begin{aligned} (\mathsf{OptPairs}) \quad & \min a'x + b'y + \langle Q, Y \rangle && (26a)\\ & \text{s.t. } Y - yy' \succeq 0 && (26b)\\ & (x_i, x_j, y_i, y_j, Y_{ii}, Y_{ij}, Y_{jj}) \in \mathcal{W} \qquad \forall i < j && (26c)\\ & 0 \leq x \leq 1,\ y \geq 0. && (26d) \end{aligned}$$

Proposition 10. *OptPairs is a valid convex relaxation of (QI) and every feasible solution to it satisfies all valid inequalities* (19).

Proof. To see that OptPairs is a valid relaxation, consider a feasible solution (x, y) of (QI) and let $Y = yy'$. For $i < j$, if $x_i = x_j = 1$, constraint (26c) is satisfied with $W = \begin{pmatrix} Y_{ii} & Y_{ij} & y_i \\ Y_{ij} & Y_{jj} & y_j \\ y_i & y_j & 1 \end{pmatrix}$. Otherwise, without loss of generality, one may assume $x_i = 0$. It follows that $Y_{ii} = y_i^2 = Y_{ij} = y_i y_j = 0$. Then, constraint (26c) is satisfied with $W = 0$. Moreover, if W satisfies (26c), then W satisfies (20) and (21) simultaneously. □

5. Comparison of convex relaxations

In this section, we compare the strength of OptPairs with other convex relaxations of (QI). The perspective relaxation and the optimal perspective relaxation OptPersp for (QI) are well-known.

Proposition 11. *OptPairs is at least as strong as OptPersp.*

Proof. Note that (26c) includes constraints

$$\begin{pmatrix} Y_{ii} & y_i \\ y_i & x_i \end{pmatrix} \succeq \begin{pmatrix} W_{11} & W_{31} \\ W_{31} & W_{33} \end{pmatrix} \succeq 0,$$

corresponding to (25b)-(25c). Thus, the perspective constraints $Y_{ii}x_i \geq y_i^2$ are implied. □

In the context of linear regression, Atamtürk and Gómez [6] study the convex hull of the epigraph of rank-one quadratic with indicators

$$\mathcal{X}_f = \left\{ (x, y, t) \in \{0,1\}^n \times \mathbb{R}^{n+1} : t \geq \left(\sum_{i=1}^n y_i \right)^2, \; y_i(1 - x_i) = 0, i \in [n] \right\},$$

where the continuous variables are unrestricted in sign. Their extended SDP formulation based on $\operatorname{cl}\operatorname{conv}(\mathcal{X}_f)$, leads to the following relaxation for (QI)

$$\begin{aligned}
&\min a'x + b'y + \langle Q, Y \rangle && && \text{(27a)} \\
&\text{s.t. } Y - yy' \succeq 0 && && \text{(27b)} \\
&\quad y_i^2 \leq Y_{ii}x_i && \forall i && \text{(27c)} \\
\text{(OptRankOne)} \quad &\begin{pmatrix} x_i + x_j & y_i & y_j \\ y_i & Y_{ii} & Y_{ij} \\ y_j & Y_{ij} & Y_{jj} \end{pmatrix} \succeq 0, && \forall i < j && \text{(27d)} \\
&\quad y \geq 0, \; 0 \leq x \leq 1. && && \text{(27e)}
\end{aligned}$$

With the additional constraints (27d), it is immediate that OptRankOne is stronger than OptPersp. The following proposition compares OptRankOne and OptPairs.

Proposition 12. *OptPairs is at least as strong as OptRankOne.*

Proof. It suffices to show that for each pair $i < j$, constraint (26c) of OptPairs implies (27d) of OptRankOne. Rewriting (25b)–(25c), we get

$$W_{11} \leq Y_{11} - \frac{(y_1 - W_{31})^2}{x_1 - W_{33}}, \; W_{22} \leq Y_{22} - \frac{(y_2 - W_{32})^2}{x_2 - W_{33}}.$$

Combining the above and (25a) to substitute out W_{11}, W_{22} and W_{12} in $W \succeq 0$, we arrive at

$$\begin{pmatrix} Y_{11} - \frac{(y_1 - W_{31})^2}{x_1 - W_{33}} & Y_{12} & W_{31} \\ Y_{12} & Y_{22} - \frac{(y_2 - W_{32})^2}{x_2 - W_{33}} & W_{32} \\ W_{31} & W_{32} & W_{33} \end{pmatrix} \succeq 0, \; W_{33} \le x_1, \; W_{33} \le x_2,$$

which is equivalent to the following matrix inequality by Shur Complement Lemma

$$\begin{pmatrix} Y_{11} & Y_{12} & W_{31} & y_1 - W_{31} & 0 \\ Y_{12} & Y_{22} & W_{32} & 0 & y_2 - W_{32} \\ W_{31} & W_{32} & W_{33} & 0 & 0 \\ y_1 - W_{31} & 0 & 0 & x_1 - W_{33} & 0 \\ 0 & y_2 - W_{32} & 0 & 0 & x_2 - W_{33} \end{pmatrix} \succeq 0.$$

By adding the third row/column to the forth row/column and then adding the forth row/column to the fifth row/column, the large matrix inequality can be rewritten as

$$\begin{pmatrix} Y_{11} & Y_{12} & W_{31} & y_1 & y_1 \\ Y_{12} & Y_{22} & W_{32} & W_{32} & y_2 \\ W_{31} & W_{32} & W_{33} & W_{33} & W_{33} \\ y_1 & W_{32} & W_{33} & x_1 & x_1 \\ y_1 & y_2 & W_{33} & x_1 & x_1 + x_2 - W_{33} \end{pmatrix} \succeq 0.$$

Because $W_{33} \ge 0$, it follows that

$$\begin{pmatrix} Y_{11} & Y_{12} & y_1 \\ Y_{12} & Y_{22} & y_2 \\ y_1 & y_2 & x_1 + x_2 \end{pmatrix} \succeq \begin{pmatrix} Y_{11} & Y_{12} & y_1 \\ Y_{12} & Y_{22} & y_2 \\ y_1 & y_2 & x_1 + x_2 - W_{33} \end{pmatrix} \succeq 0.$$

Therefore, constraints (27d) are implied by (26c), proving the claim. □

The example below illustrates that OptPairs is indeed strictly stronger than OptPersp and OptRankOne.

Example 2. For $n = 2$, OptPairs is the ideal (convex) formulation of (QI). For the instance of (QI) with

$$a = \begin{pmatrix} 1 \\ 5 \end{pmatrix}, b = \begin{pmatrix} -8 \\ -5 \end{pmatrix}, Q = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

each of the other convex relaxations has a fractional optimal solution as demonstrated in Table 1.

Notably, the fractional x values for OptPersp and OptRankOne are far from their optimal integer values. A common approach to quickly obtain feasible solutions to NP-hard problems is to round a solution obtained from a suitable convex relaxation. This example indicates that feasible solutions

Table 1. Comparison of convex relaxations of (QI).

	obj val	x_1	x_2	y_1	y_2
OptPersp	-2.866	0.049	0.268	0.208	1.369
OptRankOne	-2.222	0.551	0.449	0.0	2.007
OptPairs	-2.200	1.0	0.0	0.800	0.0

obtained in this way from formulation OptPairs may be of higher quality than those obtained from weaker relaxations – our computations in §6.2 further corroborates this intuition. □

An alternative way of constructing strong relaxations for (QI) is to decompose the quadratic function $y'Qy$ into a sum of univariate and bivariate convex quadratic functions and utilize the convex hull results of 2×2 quadratics

$$\alpha_{ij} q_{ij}(y_i, y_j) = \beta_{ij} x_i^2 \pm 2 y_i y_j + \gamma_{ij} y_j^2,$$

where $\alpha_{ij} > 0$, in Section 3 for each term, see [24] for such an approach. Specifically, let

$$y'Qy = y'Dy + \sum_{(i,j)\in\mathcal{P}} \alpha_{ij} q_{ij}(y_i, y_j) + \sum_{(i,j)\in\mathcal{N}} \alpha_{ij} q_{ij}(y_i, y_j) + y'Ry$$

where D is a diagonal PSD matrix, $\mathcal{P}/\mathcal{N}$ is the set of quadratics $q_{ij}(\cdot)$ with positive/negative off-diagonals and R is PSD remainder matrix. Applying the convex hull description for each univariate and bivariate term we obtain the following convex relaxation for (QI):

$$\begin{aligned} \min\ & a'x + b'y + \sum_{i=1}^{n} D_{ii} y_i^2 / x_i + \sum_{(i,j)\in\mathcal{P}} \alpha_{ij} f_+^*(x_i, x_j, y_i, y_j; \beta_{ij}, \gamma_{ij}) \\ \text{(Decomp)} \quad & + \sum_{(i,j)\in\mathcal{N}} \alpha_{ij} f_-^*(x_i, x_j, y_i, y_j; \beta_{ij}, \gamma_{ij}) + y'Ry \\ \text{s.t.}\ & 0 \le x \le 1,\ y \ge 0. \end{aligned}$$

The next proposition shows that OptPairs dominates Decomp. Similar duality arguments were used in [20, 24, 44].

Proposition 13. *OptPairs is at least as strong as Decomp. Moreover, there exists a decomposition for which Decomp is equivalent to OptPairs.*

Proof. We prove the result via the minimax theory of concave-convex programs and show that Decomp can be viewed as a dual formulation of OptPairs. To make the dual relationship more transparent, we define $z_i^{ij} = W_{31}^{ij}$,

$z_j^{ij} = W_{32}^{ij}$, $\lambda_{ij} = W_{33}^{ij}$ and

$$\Lambda = \Big\{(x,y,z,\lambda) \in [0,1]^n \times \mathbb{R}_+^n \times \mathbb{R}^{n(n-1)} \times \mathbb{R}^{n(n-1)/2} : 0 \le z_i^{ij} \le y_i,\\ 0 \le z_j^{ij} \le y_j,\ \min\{0, x_i + x_j - 1\} \le \lambda_{ij} \le \max\{x_i, x_j\},\ \forall i<j\Big\}.$$

Then, OptPairs can be rewritten as

$$\min_{x,y,z,\lambda} \min_{Y,W} a'x + b'y + \langle Q, Y\rangle \tag{28a}$$

$$\text{s.t. } Y \succeq yy' \tag{R}$$

$$Y_{ii} - \frac{\left(y_i - z_i^{ij}\right)^2}{x_i - \lambda_{ij}} \ge W_{11}^{ij} \ge \frac{\left(z_i^{ij}\right)^2}{\lambda_{ij}} \quad \forall i<j \tag{ℓ_i^{ij}, u_i^{ij}}$$

$$Y_{jj} - \frac{\left(y_j - z_j^{ij}\right)^2}{x_j - \lambda_{ij}} \ge W_{22}^{ij} \ge \frac{\left(z_j^{ij}\right)^2}{\lambda_{ij}} \quad \forall i<j \tag{ℓ_j^{ij}, u_j^{ij}}$$

$$\begin{pmatrix} W_{11}^{ij} & Y_{ij} \\ Y_{ij} & W_{22}^{ij}\end{pmatrix} \succeq \frac{1}{\lambda_{ij}} \begin{pmatrix} z_i^{ij} \\ z_j^{ij}\end{pmatrix} \begin{pmatrix} z_i^{ij} & z_j^{ij}\end{pmatrix} \quad \forall i<j \tag{Q^{ij}}$$

$$(x,y,z,\lambda) \in \Lambda. \tag{28b}$$

Taking the SDP dual with respect to the inner minimization problem, one arrives at

$$\min_{x,y,z,\lambda} \max_{R,Q^{ij},\ell,u} a'x + b'y + \sum_{i<j}\Bigg[\ell_i^{ij}\frac{\left(z_i^{ij}\right)^2}{\lambda_{ij}} + \ell_j^{ij}\frac{\left(z_j^{ij}\right)^2}{\lambda_{ij}} + u_i^{ij}\frac{\left(y_i - z_i^{ij}\right)^2}{x_i - \lambda_{ij}} + \\ u_j^{ij}\frac{\left(y_j - z_j^{ij}\right)^2}{x_j - \lambda_{ij}} + \frac{1}{\lambda_{ij}}\left(z^{ij}\right)' Q^{ij} z^{ij}\Bigg] + \langle R, yy'\rangle \tag{29a}$$

$$\text{s.t. } Q_{ii} = R_{ii} + \sum_{i<j} u_i^{ij} + \sum_{i>j} u_i^{ji} \quad \forall i \tag{Y_{ii}}$$

$$Q_{ij} = R_{ij} + Q_{12}^{ij} \quad \forall i<j \tag{Y_{ij}}$$

$$0 = \ell_i^{ij} - u_i^{ij} + Q_{11}^{ij} \quad \forall i<j \tag{W_{11}^{ij}}$$

$$0 = \ell_j^{ij} - u_j^{ij} + Q_{22}^{ij} \quad \forall i<j \tag{W_{22}^{ij}}$$

$$R \succeq 0, Q^{ij} \succeq 0, \ell^{ij} \ge 0, u^{ij} \ge 0 \quad \forall i<j \tag{29b}$$

$$(x,y,z,\lambda) \in \Lambda. \tag{29c}$$

Since one can take the diagonal elements of Y and W^{ij} large enough, there exists a strictly feasible solution to the inner minimization of (28), which implies strong duality holds and, thus, (28) is equivalent to (29). Next, substituting out u^{ij} in (29a), one gets

$$\begin{aligned}
&\ell_i^{ij}\frac{\left(z_i^{ij}\right)^2}{\lambda_{ij}}+\ell_j^{ij}\frac{\left(z_j^{ij}\right)^2}{\lambda_{ij}}+u_i^{ij}\frac{\left(y_i-z_i^{ij}\right)^2}{x_i-\lambda_{ij}}+u_j^{ij}\frac{\left(y_j-z_j^{ij}\right)^2}{x_j-\lambda_{ij}}+\frac{1}{\lambda_{ij}}\left(z^{ij}\right)'Q^{ij}z^{ij}\\
=&\tilde{Q}_{11}^{ij}\frac{\left(y_i-z_i^{ij}\right)^2}{x_i-\lambda_{ij}}+\tilde{Q}_{22}^{ij}\frac{\left(y_j-z_j^{ij}\right)^2}{x_j-\lambda_{ij}}+\frac{\left(z^{ij}\right)'\tilde{Q}^{ij}z^{ij}}{\lambda_{ij}},
\end{aligned}$$

where $\tilde{Q}^{ij}=\begin{bmatrix}\ell_i^{ij}+Q_{11}^{ij} & Q_{12}^{ij}\\ Q_{12}^{ij} & \ell_j^{ij}+Q_{22}^{ij}\end{bmatrix}$. By changing variables $Q^{ij}\leftarrow\tilde{Q}^{ij}$, one arrives at

$$\begin{aligned}
\min_{(x,y,z,\lambda)\in\Lambda}\max_{R,Q^{ij}}\ &\sum_{i<j}\left[Q_{11}^{ij}\frac{\left(y_i-z_i^{ij}\right)^2}{x_i-\lambda_{ij}}+Q_{22}^{ij}\frac{\left(y_j-z_j^{ij}\right)^2}{x_j-\lambda_{ij}}+\frac{\left(z^{ij}\right)'Q^{ij}z^{ij}}{\lambda_{ij}}\right]\\
&+\langle R,yy'\rangle+a'x+b'y && \text{(30a)}\\
\text{s.t.}\ \ &Q_{ii}\geq R_{ii}+\sum_{i<j}Q_{11}^{ij}+\sum_{i>j}Q_{22}^{ji} \qquad \forall i && \text{(30b)}\\
&Q_{ij}=R_{ij}+Q_{12}^{ij} \qquad \forall i<j && \text{(30c)}\\
&R\succeq 0,Q^{ij}\succeq 0, \qquad \forall i<j && \text{(30d)}
\end{aligned}$$

which is equivalent to (29). Notice that (30b) is, in fact, tight. Thus, (30b), (30c),and (30d) define a valid decomposition of Q. Moreover, $\|Q^{ij}\|_2,\|R\|_2\leq$ Trace(Q) by (30b), which implies the feasible region of the inner maximization problem is compact. Therefore, according to Von Neumann's Minimax Theorem [42], one can interchange max and min without loss of equivalence and arrive at

$$\begin{aligned}
\max_{R,Q^{ij}:\text{(30b)-(30d)}}\ \min_{z^{ij},\lambda_{ij}}\ &\sum_{i<j}\left[Q_{11}^{ij}\frac{\left(y_i-z_i^{ij}\right)^2}{x_i-\lambda_{ij}}+Q_{22}^{ij}\frac{\left(y_j-z_j^{ij}\right)^2}{x_j-\lambda_{ij}}+\frac{\left(z^{ij}\right)'Q^{ij}z^{ij}}{\lambda_{ij}}\right]\\
&+\langle R,yy'\rangle+a'x+b'y\\
\text{s.t.}\ \ &(x,y,z,\lambda)\in\Lambda
\end{aligned}$$

where the inner minimization problem is in the form Decomp from Proposition 6. □

6. Computations

In this section, we report on computational experiments performed to test the effectiveness the formulations derived in the paper. Section 6.1 is devoted to synthetic portfolio optimization instances, where matrix Q is diagonally dominant and the conic quadratic-representable extended formulations developed in Section 3 can be readily used in a branch-and-bound algorithm without the need for an SDP constraint. The instances here are generated similarly to [5], and serve to check the incremental value of convexifications based on $\mathcal{Z}_+$ compared to those based on only $\mathcal{Z}_-$. In Section 6.2, we use real instances derived from stock market returns and test the SDP relaxation OptPairs derived in Section 4, as well as mixed-integer optimization approaches based on decompositions of the quadratic matrices.

6.1. **Synthetic instances – the diagonally dominant case.** We consider a standard cardinality-constrained mean-variance portfolio optimization problem of the form

$$\min_{x,y} \left\{ y'Qy : \begin{array}{l} b'y \geq r,\ 1'x \leq k \\ 0 \leq y \leq x,\ x \in \{0,1\}^n \end{array} \right\} \tag{31}$$

where Q is the covariance matrix of returns, $b \in \mathbb{R}^n$ is the vector of the expected returns, r is the target return and k is the maximum number of securities in the portfolio. All experiments are conducted using Mosek 9.1 solver on a laptop with a 2.30GHz Intel® Core™ i9-9880H CPU and 64 GB main memory. The time limit is set to one hour and all other settings are default by Mosek.

6.1.1. *Instance generation.* We adopt the method used in [5] to generate the instances. The instances are designed to control the integrality gap of the instances and the effectiveness of the perspective formulation. Let $\rho \geq 0$ be a parameter controlling the ratio of the magnitude positive off-diagonal entries of Q to the magnitude of the negative off-diagonal entries of Q. Lower values of ρ lead to higher integrality gaps. Let $\delta \geq 0$ be the parameter controlling the diagonal dominance of Q. The perspective formulation is more effective in closing the integrality gap for higher values of δ. The following steps are followed to generate the instances:

- Construct an auxiliary matrix $\bar{Q}$ by drawing a factor covariance matrix $G_{20\times 20}$ uniformly from $[-1,1]$, and generating an exposure matrix $H_{n\times 20}$ such that $H_{ij} = 0$ with probability 0.75, and H_{ij} drawn uniformly from $[0,1]$, otherwise. Let $\bar{Q} = HGG'H'$.
- Construct off-diagonal entries of Q: For $i \neq j$, set $Q_{ij} = \bar{Q}_{ij}$, if $\bar{Q}_{ij} < 0$ and set $Q_{ij} = \rho\bar{Q}_{ij}$ otherwise. Positive off-diagonal elements of $\bar{Q}$ are scaled by a factor of ρ.
- Construct diagonal entries of Q: Pick μ_i uniformly from $[0, \delta\bar{\sigma}]$, where $\bar{\sigma} = \frac{1}{n}\sum_{i\neq j}|Q_{ij}|$. Let $Q_{ii} = \sum_{i\neq j}|Q_{ij}| + \mu_i$. Note that if $\delta = \mu_i = 0$, then matrix Q is already diagonally dominant.

- Construct b, r, k: b_i is drawn uniformly from $[0.5Q_{ii}, 1.5Q_{ii}]$, $r = 0.25\sum_{i=1}^{n} b_i$, and $k = \lfloor n/5 \rfloor$.

Matrices Q generated in this way have only 20.1% of the off-diagonal entries negative on average.

6.1.2. *Formulations.* With above setting, the portfolio optimization problem can be rewritten as

$$\begin{aligned} \min \; & \sum_{i\in[n]} \mu_i z_i + \sum_{Q_{ij}<0} |Q_{ij}| t_{ij} + \sum_{Q_{ij}>0} |Q_{ij}| t_{ij} \\ \text{s.t.} \; & (x_i, y_i, z_i) \in \mathcal{X}_0, \; \forall i \in N, \\ & (x_i, x_j, y_i, y_j, t_{ij}) \in \mathcal{Z}_-, \; \forall i > j : Q_{ij} < 0, \\ & (x_i, x_j, y_i, y_j, t_{ij}) \in \mathcal{Z}_+, \; \forall i > j : Q_{ij} > 0, \\ & b'y \geq r, \; 1'x \leq k, \end{aligned} \tag{32}$$

where $\mathcal{Z}_+$ and $\mathcal{Z}_-$ are defined as before with $d_1 = d_2 = 1$. Four strong formulations are tested by replacing the mixed-integer sets with their convex hulls: ConicQuadPersp by replacing $\mathcal{X}_0$ with $\operatorname{cl}\operatorname{conv}(\mathcal{X}_0)$ using the perspective reformulation (2) ConicQuadN by replacing $\mathcal{X}_0$ and $\mathcal{Z}_-$ with $\operatorname{cl}\operatorname{conv}(\mathcal{X}_0)$ and $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$ using the corresponding extended formulation, (3) ConicQuadP by replacing $\mathcal{X}_0$ and $\mathcal{Z}_+$ with $\operatorname{cl}\operatorname{conv}(\mathcal{X}_0)$ and $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$ respectively, and (4) ConicQuadP+N by replacing $\mathcal{X}_0$, $\mathcal{Z}_-$, and $\mathcal{Z}_+$ with $\operatorname{cl}\operatorname{conv}(\mathcal{X}_0)$, $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_-)$ and $\operatorname{cl}\operatorname{conv}(\mathcal{Z}_+)$, correspondingly.

6.1.3. *Results.* Table 2 shows the results for matrices with varying diagonal dominance δ for $\rho = 0.3$. Each row in the table represents the average for five instances generated with the same parameters. Table 2 displays the dimension of the problem n, the initial gap (`igap`), the root gap improvement (`rimp`), the number of branch and bound nodes (`nodes`), the elapsed time in secons (`time`), and the end gap provided by the solver at termination (`egap`). In addition, in brackets, we report the number of instances solved to optimality within the time limit. The initial gap is computed as $\texttt{igap} = \frac{\texttt{obj}_{\texttt{best}} - \texttt{obj}_{\texttt{cont}}}{|\texttt{obj}_{\texttt{best}}|} \times 100$, where $\texttt{obj}_{\texttt{best}}$ is the objective value of the best feasible solution found and $\texttt{obj}_{\texttt{cont}}$ is the objective value of the natural continuous relaxation of (31), i.e. obtained by dropping the integral constraints; `rimp` is computed as $\texttt{rimp} = \frac{\texttt{obj}_{\texttt{relax}} - \texttt{obj}_{\texttt{cont}}}{\texttt{obj}_{\texttt{best}} - \texttt{obj}_{\texttt{cont}}} \times 100$, where $\texttt{obj}_{\texttt{relax}}$ is the objective value of the continuous relaxation of the corresponding formulation.

In Table 2, as expected, ConicQuadPersp has the worst performance in terms of both root gap and end gap as well as the solution time. It can only solve instances with dimension $n = 40$ and some instances with dimension $n = 60$ to optimality. The `rimp` of ConicQuadPersp is less than 10% when the diagonal dominance is small. This reflects the fact that

ConicQuadPersp provides strengthening only for diagonal terms. ConicQuadN performs better than ConicQuadPersp with rimp about 10%–25%, and it can solve all low-dimensional instances and most instances of dimension $n = 60$. However, ConicQuadN is still unable to solve high-dimensional instances effectively. ConicQuadP performs much better than ConicQuadN for the instances considered: The rimp results in significantly stronger root improvements (between 70–80% on average). Moreover, ConicQuadP can solve almost all instances to near-optimality for $n = 80$. For the instances that ConicQuadP is unable to solve to optimality, the average end gap is less than 5%. By strengthening both the negative and positive off-diagonal terms, ConicQuadP+N provides the best performance with rimp above 90%. ConicQuadP+N can solve all instances and most of them are solved within 10 minutes. Finally, observe that as the diagonal dominance increases, the performance of all formulations improves. Specifically, larger diagonal dominance results in more instances solved to optimality, smaller egap and shorter solving time for all formulations. For these instances, on average, the gap improvement is raised from 50.69% to 92.90% by incorporating strengthening from off-diagonal coefficients.

Table 3 displays the computational results for different values of ρ with fixed $\delta = 0.1$. The relative comparison of formulations is similar as discussed before, with ConicQuadP+N resulting in the best performance. As ρ increases, the performance of ConicQuadN deteriorates in terms of Rimp while the performance of ConicQuadP improves, as expected. The performance of ConicQuadP+N also improves for high values of ρ, and always results in significant improvement compared to other formulations for all instances. For these instances, on average, the gap improvement is raised from 9.77% to 85.38% by incorporating strengthening from off-diagonal coefficients.

In summary, we conclude that utilizing convexification for $\mathcal{Z}_+$ complement those previously obtained for $\mathcal{Z}_-$, and together result in significantly higher root gap improvement over the simpler perspective relaxation. For the experiments in this section, we use the results of Section 3 to convexify pairwise quadratic terms, but do not utilize the more sophisticated SDP formulations in Section 4. For the instances in this section, the optimal perspective formulation [20, 44] achieves close to 100% root improvement, and all the mixed-integer optimization problems are solved in a few seconds. Moreover, the new convex formulation OptPairs produces integer (thus optimal) solutions in all instances. In the next section, we consider these stronger conic relaxations for the more realistic and challenging instances.

TABLE 2. Experiments with varying diagonal dominance, $\rho = 0.3$.

n	δ	igap	ConicQuadPersp				ConicQuadN				ConicQuadP				ConicQuadP+N			
			Rimp	Nodes	Time	Egap	Rimp	Nodes	Time	Egap	Rimp	Nodes	Time	Egap	Rimp	Nodes	Time	Egap
40	0.1	53.37	9.74	9,537	46	0.00[5]	23.44	3,439	44	0.00[5]	66.07	526	18	0.00[5]	86.93	65	13	0.00[5]
	0.5	51.10	33.17	3,896	26	0.00[5]	47.86	1,335	18	0.00[5]	79.48	198	9	0.00[5]	95.01	24	9	0.00[5]
	1.0	52.73	60.86	1,463	9	0.00[5]	74.62	375	7	0.00[5]	86.83	146	7	0.00[5]	97.46	23	8	0.00[5]
Avg		**52.40**	**34.59**	**4,965**	**27**	**0.00[15]**	**48.64**	**1,717**	**23**	**0.00[15]**	**77.46**	**290**	**11**	**0.00[15]**	**93.13**	**37**	**10**	**0.00[15]**
60	0.1	46.90	9.05	316,000	3,363	5.53[1]	19.07	135,052	3,261	3.83[2]	76.78	4,898	498	0.00[5]	89.72	445	140	0.00[5]
	0.5	50.97	38.46	134,542	1,888	2.65[3]	49.94	55,434	1,321	0.98[4]	82.72	1,652	267	0.00[5]	95.13	203	75	0.00[5]
	1.0	47.22	60.04	21,440	317	0.00[5]	66.52	8,579	209	0.00[5]	94.69	86	35	0.00[5]	98.69	17	22	0.00[5]
Avg		**48.36**	**35.85**	**157,328**	**1,856**	**2.73[9]**	**45.18**	**66,355**	**1,597**	**1.60[11]**	**84.73**	**2,212**	**267**	**0.00[15]**	**94.51**	**222**	**79**	**0.00[15]**
80	0.1	49.91	4.76	155,000	3,600	20.25[0]	21.96	69,609	3,600	14.38[0]	65.11	8,017	2,742	4.69[2]	83.33	2,142	1,416	0.00[5]
	0.5	50.53	37.33	136,638	3,600	12.06[0]	49.16	63,897	3,600	7.49[0]	81.57	6,525	2,473	1.70[2]	94.21	341	261	0.00[5]
	1.0	53.78	56.96	152,704	3,600	7.41[0]	69.41	45,388	3,068	2.95[2]	84.42	5,870	2,116	1.27[3]	95.67	365	275	0.00[5]
Avg		**51.41**	**33.02**	**148,114**	**3,600**	**13.24[0]**	**46.84**	**59,632**	**3,423**	**8.27[2]**	**77.03**	**6,804**	**2,443**	**2.55[7]**	**91.07**	**950**	**651**	**0.00[15]**

Table 3. Experiments with varying positive off-diagonal entries, $\delta = 0.1$.

n	ρ	igap	ConicQuadPersp				ConicQuadN				ConicQuadP				ConicQuadP+N			
			Rimp	Nodes	Time	Egap	Rimp	Nodes	Time	Egap	Rimp	Nodes	Time	Egap	Rimp	Nodes	Time	Egap
40	0.1	60.67	9.88	6,928	36	0.00[5]	23.11	1,869	22	0.00[5]	50.68	1,134	34	0.00[5]	72.23	158	19	0.00[5]
	0.5	47.67	8.7	8,572	46	0.00[5]	21.67	3,181	41	0.00[5]	75.82	272	12	0.00[5]	91.78	53	11	0.00[5]
	1.0	43.23	10.05	8529	44	0.00[5]	18.08	4,903	60	0.00[5]	82.33	149	8	0.00[5]	92.56	51	11	0.00[5]
Avg		**50.52**	**9.54**	**8,010**	**42**	**0.00[15]**	**20.95**	**3,317**	**41**	**0.00[15]**	**69.61**	**519**	**18**	**0.00[15]**	**85.53**	**87**	**14**	**0.00[15]**
60	0.1	60.26	10.7	256,480	2,585	4.03[2]	27.85	34,563	847	0.00[5]	53.37	16,190	1,983	3.10[3]	78.5	1,016	264	0.00[5]
	0.5	45.98	9.38	319,534	3,230	5.57[1]	19.21	103,869	3,043	4.24[2]	78.22	2,715	315	0.00[5]	91.09	259	107	0.00[5]
	1.0	40.87	10.09	197,140	3,258	4.52[2]	15.93	98,289	2,982	4.34[2]	85.23	564	100	0.00[5]	91.66	135	72	0.00[5]
Avg		**49.03**	**10.06**	**257,718**	**3,024**	**4.71[5]**	**21**	**78,907**	**2,291**	**2.86[9]**	**72.27**	**6,490**	**799**	**1.03[13]**	**87.08**	**470**	**148**	**0.00[15]**
80	0.1	64.85	9.88	142,299	3,600	24.78[0]	26.42	60,081	3,600	14.81[0]	46.63	11,367	3,172	17.40[1]	69.6	4,948	2,920	6.22[1]
	0.5	47.97	9.27	148,252	3,600	18.46[0]	20.75	48,887	3,600	15.70[0]	73.3	7,245	3,019	2.72[2]	89.11	1,131	827	0.00[5]
	1.0	41.69	10	149,563	3,600	14.79[0]	16.61	52,485	3,600	14.34[0]	84.7	3,769	1,444	0.88[4]	91.93	1,068	716	0.00[5]
Avg		**51.51**	**9.72**	**146,705**	**3,600**	**19.34[0]**	**21.26**	**53,818**	**3,600**	**14.95[0]**	**68.21**	**7,460**	**2,545**	**7.00[7]**	**83.54**	**2,382**	**1,487**	**2.07[11]**

6.2. **Real instances – the general case.** Now using real stock market data, we consider portfolio index tracking problem of the form

$$
\text{(IT)} \quad \begin{aligned} \min\ & (y - y_B)'Q(y - y_B) \\ \text{s.t.}\ & 1'y = 1,\ 1'x \le k \\ & 0 \le y \le x,\ x \in \{0,1\}^n, \end{aligned}
$$

where $y_B \in \mathbb{R}^n$ is a benchmark index portfolio, Q is the covariance matrix of security returns and k is the maximum number of securities in the portfolio.

6.2.1. *Instance generation.* We use the daily stock return data provided by Boris Marjanovic in Kaggle[1] to compute the covariance matrix Q. Specifically, given a desired start date (either 1/1/2010 or 1/1/2015 in our computations), we compute the sample covariance matrix based on the stocks with available data in at least 99% of the days since the start (returns for missing data are set to 0). The resulting covariance matrices are available at `https://sites.google.com/usc.edu/gomez/data`. We then generate instances as follows:

- we randomly sample an $n \times n$ covariance matrix Q corresponding to n stocks, and
- we draw each element of y_B from uniform [0,1], and then scale y_B so that $1'y_B = 1$.

6.2.2. *Convex relaxations.* The natural convex relaxation of IT always yields a trivial lower bound of 0, as it is possible to set $z = y = y_B$. Thus, we do not report results concerning the natural relaxation. Instead, we consider the optimal perspective relaxation OptPersp of [20]:

$$
\text{(OptPersp)} \quad \begin{aligned} \min_{x,y,Y}\ & y_B'Qy_B - 2y_B'Qy + \langle Q, Y \rangle && \text{(34a)} \\ \text{s.t.}\ & Y - yy' \succeq 0 && \text{(34b)} \\ & y_i^2 \le Y_{ii}x_i \quad \forall i \in [n] && \text{(34c)} \\ & 0 \le x \le 1,\ y \ge 0 && \text{(34d)} \\ & 1'y = 1,\ 1'x \le k && \text{(34e)} \end{aligned}
$$

[1]https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs

and the proposed OptPairs exploiting off-diagonal elements of Q:

$$\text{(OptPairs)} \quad \begin{aligned} \min_{x,y,Y,W}\ & y_B'Qy_B - 2y_B'Qy + \langle Q, Y\rangle \\ \text{s.t. } & Y - yy' \succeq 0 \\ & W^{ij} \succeq 0 && \forall i<j \\ & (Y_{ii} - W^{ij}_{11})(x_i - W^{ij}_{33}) \geq (y_i - W^{ij}_{31})^2,\ W^{ij}_{11} \leq Y_{ii} && \forall i<j \\ & (Y_{jj} - W^{ij}_{22})(x_j - W^{ij}_{33}) \geq (y_j - W^{ij}_{32})^2,\ W^{ij}_{22} \leq Y_{jj} && \forall i<j \\ & W^{ij}_{33} \leq x_i + x_j - 1,\ W^{ij}_{33} \leq x_i,\ W^{ij}_{33} \leq x_j && \forall i<j \\ & 0 \leq W^{ij}_{31} \leq y_i,\ 0 \leq W^{ij}_{32} \leq y_j,\ W^{ij}_{12} = Y_{ij} && \forall i<j \\ & 0 \leq x \leq 1,\ y \geq 0 \\ & 1'y = 1,\ 1'x \leq k, \end{aligned}$$

For each relaxation, we consider a simple rounding heuristic to obtain feasible solutions to (IT): given an optimal solution $(\bar{x}, \bar{y})$ to the continuous relaxation, we fix $x_i = 1$ for the k-largest values of $\bar{x}$ and the remaining $x_i = 0$, and resolve the continuous relaxation to compute y.

6.2.3. *Exact mixed-integer optimization approaches.* We also consider three mixed-integer optimization approaches, each associated with a different convex relaxation. The first one is the Natural relaxation corresponding to the mixed-integer quadratic formulation (IT).

The second one is the corresponding OptPersp formulation

$$\min\ y_B'Qy_B - 2y_B'Qy + y'Ry + \sum_{i=1}^{n} D_{ii}t_i \tag{35a}$$

$$\text{s.t. } t_i x_i \geq y_i^2,\ i \in [n] \tag{35b}$$

$$1'y = 1,\ 1'x \leq k \tag{35c}$$

$$0 \leq y \leq x,\ x \in \{0,1\}^n, \tag{35d}$$

where $D + R = Q$ and R are the dual variables associated with constraint (34b). The third one is the OptPairs formulation based on the decomposition

$$\min\ y_B'Qy_B - 2y_B'Qy + y'Ry + \sum_{i<j} t_{ij} \tag{36a}$$

$$\text{s.t. } t_{ij} \geq Q^{ij}_{ii}y_i^2 + Q^{ij}_{jj}y_j^2 + 2Q^{ij}_{ij}y_iy_j \qquad \forall i<j \tag{36b}$$

$$1'y = 1,\ 1'x \leq k \tag{36c}$$

$$0 \leq y \leq x,\ x \in \{0,1\}^n, \tag{36d}$$

where matrix R is the dual variable associated with constraint $Y - yy' \succeq 0$, and matrices Q^{ij} are the dual variables associated with constraints $W^{ij} \succeq 0$.

The formulation is then obtained from the SOCP-representable convexification of constraints (36b) using Proposition 6 (if $Q_{ij}^{ij} \geq 0$) or Remark 1 (if $Q_{ij}^{ij} < 0$).

6.2.4. *Results.* In these experiments, the solution time limit is set to 20 minutes, which includes the time required to solve the SDP relaxations to find suitable decompositions. Tables 4 and 5 present the results using historical data since 2010 and 2015, respectively. They show, for different values of n and k, and for each conic relaxation: the time required to solve the convex relaxations in seconds, the lower bound (LB) corresponding to the optimal objective value of the continuous relaxation, the upper bound (UB) corresponding to the objective value of the heuristic, the gap between these two values, computed as $\texttt{Gap} = \frac{\texttt{UB}-\texttt{LB}}{\texttt{UB}}$; they also show the best objective found at termination, and the associated gap, number of nodes explored, time spent in branch-and-bound in seconds, and number of instances that could be solved to optimality within the time limit (#). The lower bounds, upper bounds from the convex relaxations, and objective from branch-and-bound, are scaled so that the best upper bound found for a given instance is 100. Each row represents an average of five instances generated with the same parameters.

We first summarize our conclusions, then discuss in depth the relative performance of the mixed-integer optimization formulations, and finally discuss the performance of the conic formulations (which, we argue, perform best for this class of problems).

• **Summary**. The perspective reformulation (35) remains the best approach to solve the problems to optimality with the current off-the-shelf MISOCP solvers, as MIP solvers struggle with more the sophisticated formulations. However, the stronger formulations are very effective in producing comparable or better solutions (especially in challenging instances with poor natural convex relaxations) via rounding the convex relaxation solutions in a fraction of the computational time.

• **Comparison of mixed-integer optimization approaches**. For instances with $n = 50$, we see that, among the mixed-integer optimization approaches, the one based on OptPersp is arguably the best, solving to optimality 22/30 instances (compared with Natural: 15/22, and OptPairs: 8/22). The Natural mixed-integer optimization formulation is able to explore more nodes, but the relaxations are weaker, ultimately leading to inferior performance. In contrast, the stronger mixed-integer formulation based on OptPairs needs more time to process each node (by orders-of-magnitude) due to the increased complexity of the relaxations, resulting in poor performance overall. Nonetheless, for instances where it can prove optimality

Table 4. Results with stock return data since 2010.

n	k	Method	Convex				Branch & Bound				
			LB	UB	Gap	Time(s)	Obj	Gap	Nodes	Time(s)	#
50	5	Natural	-	-	-	-	100.0	0.0%	1,248,360	140.3	5
		OptPersp	94.2	106.0	10.9%	0.9	100.0	0.0%	147,146	142.8	5
		OptPairs	99.3	100.7	1.4%	2.9	100.0	1.4%	1,844	843.6	3
	7	Natural	-	-	-	-	100.0	23.1%	10,188,792	1,155.2	1
		OptPersp	93.3	106.1	12.0%	0.8	100.0	7.1%	984,569	900.7	2
		OptPairs	98.9	101.5	2.6%	2.6	100.0	5.3%	2,186	1197.5	0
	10	Natural	-	-	-	-	100.0	47.8%	11,397,070	1,200.0	0
		OptPersp	92.7	106.8	13.2%	0.8	101.2	20.5%	1,252,352	1,112.7	1
		OptPairs	98.0	105.5	6.8%	2.5	101.7	9.6%	1,678	1,197.5	0
100	10	Natural	-	-	-	-	100.0	87.2%	4,121,436	1,200.0	0
		OptPersp	91.4	107.9	14.6%	25.3	107.6	40.5%	415,954	1,174.7	0
		OptPairs	99.3	100.9	1.6%	55.9	176.7	228.3%	239	1,144.1	0
	15	Natural	-	-	-	-	100.0	91.0%	4,681,628	1,200.0	0
		OptPersp	91.3	108.7	15.4%	21.7	107.2	46.2%	514,894	1,178.4	0
		OptPairs	99.8	100.5	0.8%	43.9	193.7	311.0%	259	1,156.1	0
	20	Natural	-	-	-	-	100.0	92.6%	4,852,043	1,200.0	0
		OptPersp	90.6	107.4	15.5%	21.5	109.7	52.7%	566,917	1,178.5	0
		OptPairs	99.3	100.4	1.2%	43.4	202.3	269.7%	189	1,156.7	0

(e.g., $n = 50$, $k = 5$), it does so with substantially fewer nodes, illustrating the power of the stronger relaxations. Interestingly, in the more challenging instances with data from 2010, $n = 50$ and $k = 10$, OptPairs is able to prove the best optimality gap of 9.6% (compared with OptPersp: 20.5%, and Natural: 4.8%).

For larger instances with $n = 100$, all mixed-integer optimization formulations struggle. Formulations based on OptPairs result in gaps well-above 100%, that is, the best lower bound achieved by branch-and-bound is negative; for instances with data since 2015 and $k = 20$, the root node relaxations cannot be fully processed in 20 minutes, and the branch-and-bound solver terminates without an incumbent solution. Indeed, MISOCP solvers based on outer approximations struggle to solve highly nonlinear instances with a large number of variables and exhibit pathological behavior, e.g., see [5, 8, 31] for similar documented results. Formulations based on Natural produce the best incumbent solutions, due to the large number of nodes explored, but terminate with optimality gaps close to 100% in all cases. Formulations

TABLE 5. Results with stock return data since 2015.

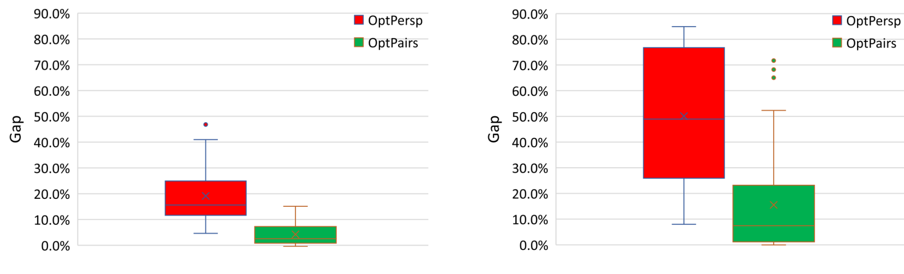
n	k	Method	Convex LB	UB	Gap	Time(s)	Branch & Bound Obj	Gap	Nodes	Time(s)	#
		Natural	-	-	-	-	100.0	0.0%	404,98	45.9	5
	5	OptPersp	90.6	116.3	20.9%	0.8	100.0	0.0%	49,455	55.6	5
		OptPairs	99.3	100.0	0.7%	2.9	100.0	0.8%	2,436	652.9	4
		Natural	-	-	-	-	100.0	4.5%	3,396,633	461.8	4
50	7	OptPersp	91.4	116.9	20.9%	0.9	100.1	0.0%	230,606	365.9	5
		OptPairs	99.5	100.0	0.5%	3.2	100.8	8.1%	2,387	1,129.7	1
		Natural	-	-	-	-	100.0	25.8%	7,582,274	1,200.0	0
	10	OptPersp	90.7	114.5	20.2%	1.1	100.3	4.6%	594,498	815.2	4
		OptPairs	99.3	100.4	1.1%	3.7	105.6	17.5%	1,477	1,196.3	0
		Natural	-	-	-	-	100.0	77.1%	4,345,144	1,200.0	0
	10	OptPersp	80.3	246.4	56.5%	20.9	102.1	41.8%	329,855	1,179.1	0
		OptPairs	96.6	102.7	6.0%	47.2	348.4	126.5%	195	1,152.9	0
		Natural	-	-	-	-	100.0	84.8%	4,388,528	1,200.0	0
100	15	OptPersp	80.1	278.9	66.2%	20.8	109.8	49.2%	378,832	1,179.3	0
		OptPairs	96.3	105.4	8.5%	45.7	223.6	143.1%	189	1,154.4	0
		Natural	-	-	-	-	100.0	87.6%	4,326,808	1,200	0
	20	OptPersp	80.2	204.3	56.5%	22.0	100.9	53.1%	409,343	1,178.3	0
		OptPairs	96.4	102.6	6.0%	51.3	†	†	†	†	0

†: Unable to fully process the root node in the time limit

based on OptPersp achieve a middle ground of producing reasonably good solutions with moderate gaps, although the optimality gaps of 50% are still quite high.

• **Discussion of conic formulations**. First, note that the continuous conic formulation OptPairs produces better lower bounds and upper bounds (via the rounding heuristic) than the continuous OptPersp: in particular, gaps are on average reduced by 66%, see Figure 1 for a summary of the gaps across all instances. The better performance comes at the expense of increased computational times by a factor of three, which does not depend on the dimension of the problem. For the instances considered, the additional computation time is at most 30 seconds, which is negligible compared with the cost of solving the mixed-integer optimization problem.

We now compare rounding OptPairs solution with the mixed-integer optimization based on OptPersp, henceforth referred to as MIO, which produced the best results among branch-and-bound approaches. For instances



(A) Data since 2010. (B) Data since 2015.

FIGURE 1. Distribution of gaps for OptPersp and OptPairs.

MIO solves to optimality (typically requiring between one and ten minutes), OptPairs produces optimality gaps under 2% in less than four seconds, indicating the effectiveness of rounding the strong OptPairs solutions. More importantly, in all other instances, OptPairs invariably produces much better gaps than MIO in a fraction of the time. For example, in Table 4 with $n = 100$, OptPairs provides optimality gaps under 2% in one minute, whereas MIO terminates with gaps above 40% after 20 minutes of branch-and-bound. While the improved gaps are mostly caused by considerably better lower bounds, in many cases the rounding heuristic based on OptPairs delivers better primal bounds than MIO: for example, in Table 4, $n = 100$ and $k = 20$, OptPairs produces feasible solutions with an average objective value of 100.4, whereas MIO results in incumbents with average value of 109.7.

7. Conclusions

In this paper, we describe the convex hull of the mixed-integer epigraph of the bivariate convex quadratic functions with nonnegative variables and off-diagonals with an SOCP-representable extended formulation as well as in the original space of variables. Furthermore, we develop a new technique for constructing an optimal convex relaxation from elementary valid inequalities. Using this technique, we develop a new strong SDP relaxation for (QI), based on the convex hull descriptions of the bivariate cases as building blocks. Moreover, the computational results with synthetic and real portfolio optimization instances indicate that the proposed formulations provide substantial improvement over existing alternatives in the literature.

Acknowledgments

The authors are grateful to two anonymous referees whose feedback have been very valuable in expanding the analyses. Andrés Gómez is supported, in part, by NSF grants 1930582 and 1818700. Alper Atamtürk is supported, in part, by NSF AI Institute for Advances in Optimization Award 211253, NSF grant 1807260, and DOD ONR grant 12951270.

References

[1] Aktürk, M. S., Atamtürk, A., and Gürel, S. (2009). A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Operations Research Letters*, 37(3):187–191.

[2] Anstreicher, K. and Burer, S. (2010). Computable representations for convex hulls of low-dimensional quadratic forms. *Mathematical Programming*, 124(1):33–43.

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

[4] Atamtürk, A., Gómez, A., and Han, S. (2021). Sparse and smooth signal estimation: Convexification of ℓ_0-formulations. *Journal of Machine Learning Research*, 22(52):1–43.

[5] Atamtürk, A. and Gómez, A. (2018). Strong formulations for quadratic optimization with M-matrices and indicator variables. *Mathematical Programming*, 170(1):141–176.

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

[7] Atamtürk, A. and Gómez, A. (2020). Safe screening rules for ℓ_0-regression from perspective relaxations. In *International Conference on Machine Learning*, pages 421–430. PMLR.

[8] Atamtürk, A. and Gómez, A. (2020). Supermodularity and valid inequalities for quadratic optimization with indicators. *arXiv preprint arXiv:2012.14633*.

[9] Bach, F. (2019). Submodular functions: from discrete to continuous domains. *Mathematical Programming*, 175(1-2):419–459.

[10] Belotti, P., Bonami, P., Fischetti, M., Lodi, A., Monaci, M., Nogales-Gómez, A., and Salvagnin, D. (2016). On handling indicator constraints in mixed integer programming. *Computational Optimization and Applications*, 65(3):545–566.

[11] Bertsimas, D., Cory-Wright, R., and Pauphilet, J. (2019). A unified approach to mixed-integer optimization: Nonlinear formulations and scalable algorithms. *arXiv preprint arXiv:1907.02109*.

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

[13] Bienstock, D. and Michalka, A. (2014). Cutting-planes for optimization of convex functions over nonconvex sets. *SIAM Journal on Optimization*, 24(2):643–677.

[14] Boland, N., Dey, S. S., Kalinowski, T., Molinaro, M., and Rigterink, F. (2017). Bounding the gap between the McCormick relaxation and the convex hull for bilinear functions. *Mathematical Programming*, 162(1):523–535.

[15] Boman, E. G., Chen, D., Parekh, O., and Toledo, S. (2005). On factor width and symmetric H-matrices. *Linear Algebra and Its Applications*, 405:239–248.

[16] Bonami, P., Lodi, A., Tramontani, A., and Wiese, S. (2015). On mathematical programming with indicator constraints. *Mathematical Programming*, 151(1):191–223.

[17] Castro, J., Frangioni, A., and Gentile, C. (2014). Perspective reformulations of the CTA problem with l_2 distances. *Operations Research*, 62(4):891–909.

[18] Dedieu, A., Hazimeh, H., and Mazumder, R. (2021). Learning sparse classifiers: Continuous and mixed integer optimization perspectives. *Journal of Machine Learning Research*, 22(135):1–47.

[19] Dey, S. S., Santana, A., and Wang, Y. (2019). New SOCP relaxation and branching rule for bipartite bilinear programs. *Optimization and Engineering*, 20(2):307–336.

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

[21] Dong, H. and Linderoth, J. (2013). On valid inequalities for quadratic programming with continuous variables and binary indicators. In Goemans, M. and Correa, J., editors, *Proceedings of IPCO 2013*, page 169–180, Berlin. Springer.

[22] Frangioni, A. and Gentile, C. (2006). Perspective cuts for a class of convex 0–1 mixed integer programs. *Mathematical Programming*, 106(2):225–236.

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

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

[25] Gao, J. and Li, D. (2011). Cardinality constrained linear-quadratic optimal control. *IEEE Transactions on Automatic Control*, 56(8):1936–1941.

[26] Gómez, A. (2021a). Outlier detection in time series via mixed-integer conic quadratic optimization. *SIAM Journal on Optimization*, 31(3):1897–1925.

[27] Gómez, A. (2021b). Strong formulations for conic quadratic optimization with indicator variables. *Mathematical Programming*, 188(1):193–226.

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

[29] Gupte, A., Kalinowski, T., Rigterink, F., and Waterer, H. (2020). Extended formulations for convex hulls of some bilinear functions. *Discrete Optimization*, 36:100569.

[30] Han, S., Gómez, A., and Atamtürk, A. (2021). The equivalence of optimal perspective formulation and Shor's SDP for quadratic programs with indicator variables. *https://arxiv.org/abs/2112.04618*.

[31] Han, S. and Gómez, A. (2021). Compact extended formulations for low-rank functions with indicator variables. *arXiv preprint arXiv:2110.14884*.

[32] Hijazi, H., Bonami, P., Cornuéjols, G., and Ouorou, A. (2012). Mixed-integer nonlinear programs featuring "on/off" constraints. *Computational Optimization and Applications*, 52:537–558.

[33] Hochbaum, D. S. (2001). An efficient algorithm for image segmentation, Markov random fields and related problems. *Journal of the ACM*, 48:686–701.

[34] Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge university press.

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

[36] Lim, C. H., Linderoth, J., and Luedtke, J. (2018). Valid inequalities for separable concave constraints with indicator variables. *Mathematical Programming*, 172(1-2):415–442.

[37] Liu, P., Fattahi, S., Gómez, A., and Küçükyavuz, S. (2021). A graph-based decomposition method for convex quadratic optimization with indicators. *arXiv preprint arXiv:2110.12547*.

[38] Locatelli, M. and Schoen, F. (2014). On convex envelopes for bivariate functions over polytopes. *Mathematical Programming*, 144(1):56–91.

[39] Mahajan, A., Leyffer, S., Linderoth, J., Luedtke, J., and Munson, T. (2017). Minotaur: A mixed-integer nonlinear optimization toolkit. Technical report, ANL/MCS-P8010-0817, Argonne National Lab.

[40] Rockafellar, R. T. (1970). *Convex Analysis.*

[41] Shor, N. Z. (1987). Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25:1–11.

[42] Sion, M. (1958). On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176.

[43] Wu, B., Sun, X., Li, D., and Zheng, X. (2017). Quadratic convex reformulations for semicontinuous quadratic programming. *SIAM Journal on Optimization*, 27:1531–1553.

[44] Zheng, X., Sun, X., and Li, D. (2014). 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.

Appendix A. Proof of Proposition 8

Proof of Proposition 8. Notice that for $\lambda = x_1 + x_2 - 1 > 0$, $f(x, y, \lambda; d)$ can be rewritten in the form

$$f(x, y, \lambda; d) = \frac{1}{D}\hat{y}' A^* \hat{y},$$

where $D = (d_1 d_2 - 1)x_1 x_2 + x_1 + x_2 - 1 > 0, \hat{y}' = (\sqrt{d_1} y_1, \sqrt{d_2} y_2)$ and

$$A^* = \begin{pmatrix} (d_1 d_2 - 1)x_2 + \lambda & \sqrt{d_1 d_2}\lambda \\ \sqrt{d_1 d_2}\lambda & (d_1 d_2 - 1)x_1 + \lambda \end{pmatrix}.$$

Observe $\det(A^*) = (d_1 d_2 - 1)D$. Hence,

$$f(x, y, \lambda; d) = \frac{(d_1 d_2 - 1)}{\det(A^*)}\hat{y}^T A^* \hat{y} = (d_1 d_2 - 1)\hat{y}^T A^{-1} \hat{y},$$

where A is the adjugate of A^*, i.e.,

$$A = \begin{pmatrix} (d_1 d_2 - 1)x_1 + \lambda & -\sqrt{d_1 d_2}\lambda \\ -\sqrt{d_1 d_2}\lambda & (d_1 d_2 - 1)x_2 + \lambda \end{pmatrix}.$$

Note that $A \succ 0$. By Schur Complement Lemma, $t/(d_1 d_2 - 1) \geq \hat{y}' A^{-1} \hat{y}$ if and only if

$$\begin{pmatrix} t/(d_1 d_2 - 1) & \hat{y}^T \\ \hat{y} & A \end{pmatrix} \succeq 0,$$

i.e.,

$$\begin{pmatrix} t/(d_1 d_2 - 1) & \sqrt{d_1} y_1 & \sqrt{d_2} y_2 \\ \sqrt{d_1} y_1 & (d_1 d_2 - 1)x_1 + \lambda & -\sqrt{d_1 d_2}\lambda \\ \sqrt{d_2} y_2 & -\sqrt{d_1 d_2}\lambda & (d_1 d_2 - 1)x_2 + \lambda \end{pmatrix} \succeq 0,$$

which is further equivalent to

$$\begin{pmatrix} t/(d_1 d_2 - 1) & y_1 & y_2 \\ y_1 & (d_1 d_2 - 1)x_1/d_1 + \lambda/d_1 & -\lambda \\ y_2 & -\lambda & (d_1 d_2 - 1)x_2/d_2 + \lambda/d_2 \end{pmatrix} \succeq 0.$$

The conclusion follows by taking $\lambda = x_1 + x_2 - 1$. □