Mathematical Programming (2024) 204:703–737
https://doi.org/10.1007/s10107-023-01982-0

FULL LENGTH PAPER

Series A

On the convex hull of convex quadratic optimization problems with indicators

Linchuan Wei[1] · Alper Atamtürk[2] · Andrés Gómez[3] · Simge Küçükyavuz[4]

Received: 6 January 2022 / Accepted: 20 April 2023 / Published online: 7 June 2023

Abstract

We consider the convex quadratic optimization problem in $\mathbb{R}^n$ with indicator variables and arbitrary constraints on the indicators. We show that a convex hull description of the associated mixed-integer set in an extended space with a quadratic number of additional variables consists of an $(n+1) \times (n+1)$ positive semidefinite constraint (explicitly stated) and linear constraints. In particular, convexification of this class of problems reduces to describing a polyhedral set in an extended formulation. While the vertex representation of this polyhedral set is exponential and an explicit linear inequality description may not be readily available in general, we derive a compact mixed-integer linear formulation whose solutions coincide with the vertices of the polyhedral set. We also give descriptions in the original space of variables: we provide a description based on an infinite number of conic-quadratic inequalities, which are "finitely generated." In particular, it is possible to characterize whether a given inequality is necessary to describe the convex hull. The new theory presented here unifies several previously

✉ Andrés Gómez
gomezand@usc.edu

Linchuan Wei
linchuanwei2022@u.northwestern.edu

Alper Atamtürk
atamturk@berkeley.edu

Simge Küçükyavuz
simge@northwestern.edu

[1] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, USA

[2] Department of Industrial Engineering and Operations Research, University of California Berkeley, Berkeley, USA

[3] Department of Industrial and System Engineering, University of Southern California, Los Angeles, USA

[4] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, USA

established results, and paves the way toward utilizing polyhedral methods to analyze the convex hull of mixed-integer nonlinear sets.

Mathematics Subject Classification 90C11 · 90C25

1 Introduction

Given a symmetric positive semidefinite matrix $Q \in \mathbb{R}^{n\times n}$, vectors $a, b \in \mathbb{R}^n$ and set $Z \subseteq \{0,1\}^n$, consider the mixed-integer quadratic optimization (MIQO) problem with indicator variables

$$
\text{(MIQO)} \quad
\begin{aligned}
\min\ & a^\top x + b^\top z + \tfrac{1}{2}t && \text{(1a)}\\
\text{s.t. } & x^\top Q x \le t && \text{(1b)}\\
& x_i(1-z_i)=0,\ \ i=1,\dots,n && \text{(1c)}\\
& x\in\mathbb{R}^n,\ \ z\in Z,\ \ t\in\mathbb{R}, && \text{(1d)}
\end{aligned}
$$

and the associated mixed-integer nonlinear set

$$X=\left\{(x,z,t)\in\mathbb{R}^n\times Z\times\mathbb{R} : t\ge x^\top Qx,\ \ x\circ(\boldsymbol{e}-z)=0\right\},$$

where $\boldsymbol{e}$ denotes a vector of ones, and $x\circ(\boldsymbol{e}-z)$ is the Hadamard product of vectors x and $\boldsymbol{e}-z$. There has recently been an increasing interest in problem (1) due to its statistical applications: the nonlinear term (1b) is used to model a quadratic loss function, as in regression, while Z represents logical conditions on the support of the variables x. For example, given model matrix $F\in\mathbb{R}^{m\times n}$ and responses $\beta\in\mathbb{R}^m$, setting $a=-\beta^\top F$, $Q=F^\top F$, $b=0$ and $Z=\left\{z\in\{0,1\}^n : \sum_{i=1}^n z_i\le r\right\}$ in (1) is equivalent to the best subset selection problem with a given cardinality r [10, 16]:

$$\min_{x,z}\ \|\beta-Fx\|_2^2 \quad \text{s.t.} \quad x\circ(\boldsymbol{e}-z)=0,\ \sum_{i=1}^n z_i\le r. \tag{2}$$

Other constraints defining Z that have been considered in statistical learning applications include multicollinearity [10], cycle prevention [28, 30], and hierarchy [12]. Set X arises as a substructure in many other applications, including portfolio optimization [13], optimal control [21], image segmentation [26], signal denoising [9].

A critical step toward solving MIQO effectively is to convexify the set X. Indeed, the mixed-integer optimization problem (1) is equivalent to the convex optimization problem

$$\min_{x,z,t}\left\{a^\top x+b^\top z+\tfrac{1}{2}t\ \ (x,z,t)\in \text{cl conv}(X)\right\},$$

where $\text{conv}(X)$ denotes the convex hull of X and $\text{cl conv}(X)$ is the closure of $\text{conv}(X)$. However, problem MIQO is $\mathcal{NP}$-hard even if $Z = \{0, 1\}^n$ [15]. Thus, a simple description of $\text{cl conv}(X)$ is, in general, not possible unless $\mathcal{NP}$= Co-$\mathcal{NP}$.

In practice, one aims to obtain a good convex relaxation of X, which can then be used either as a standalone method (as is pervasively done in the machine learning literature), to obtain high-quality solutions via rounding, or in a branch-and-bound framework. Nonetheless, it is unclear how to determine whether a given relaxation is good or not. In mixed-integer *linear* optimization, it is well-understood that facet-defining inequalities give strong relaxations. However, in MIQO (and, more generally, in mixed-integer nonlinear optimization problems), $\text{cl conv}(X)$ is not a polyhedron and there is no consensus on how to design good convex relaxations, or even what a good relaxation should be.

An important class of convex relaxations of X that has received attention in the literature is obtained by decomposing matrix $Q = \sum_{i=1}^{\ell} \Gamma_i + R$, where $\Gamma_i \succeq 0$, $i = 1, \ldots, \ell$, are assumed to be "simple" and $R \succeq 0$. Then

$$t \geq x^\top Qx \iff t \geq \sum_{i=1}^{\ell} \tau_i + x^\top Rx, \text{ and } \tau_i \geq x^\top \Gamma_i x, \ \forall i \in \{1, \ldots, \ell\}, \tag{3}$$

and each constraint $\tau_i \geq x^\top \Gamma_i x$ is replaced with a system of inequalities describing the convex hull of the associated "simple" mixed-integer set. This idea was originally used in [19], where $\ell = n$, $(\Gamma_i)_{ii} = d_i > 0$ and $(\Gamma_i)_{jk} = 0$ otherwise, and constraints $\tau_i \geq d_i x_i^2$ are strengthened using the perspective relaxation [1, 18, 22], i.e., reformulated as $z_i \tau_i \geq d_i x_i^2$. Similar relaxations based on separable quadratic terms were considered in [17, 35]. A generalization of the above approach is rank-one decomposition, which lets $\Gamma_i = h_i h_i^\top$ be a rank-one matrix [5, 6, 33, 34]; in this case, letting $S_i = \{i \in [n] : h_i \neq 0\}$, constraints $\left(\sum_{j \in S_i} z_j\right) \tau_i \geq (h_i^\top x)^2$ can be added to the formulation. Alternative generalizations of perspective relaxation that have been considered in the literature include exploiting substructures based on Γ_i where non-zeros are 2×2 matrices [4, 7, 8, 20, 24, 27] or tridiagonal [29].

Convexifications based on decomposition (3) have proven to be strong computationally, and are attractive from a theoretical perspective. The fact that a given formulation is ideal for the substructure $\tau_i \geq x^\top \Gamma_i x$ lends some theoretical weight to the strength of the convexification. However, approaches based on decomposition (3) have fundamental limitations as well. First, they require computing the convex hull description of a nonlinear mixed-integer set to establish (theoretically) the strength of the relaxation, a highly non-trivial task that restricts the classes of matrices Γ_i that can be used. Second, even if the ideal formulation for the substructure $\tau_i \geq x^\top \Gamma_i x$ is available, the convexification based on such decomposition can still be a poor relaxation of X—and there is currently no approach to establish the strength of the relaxation without numerical computations. Third, it is unclear whether the structure of the relaxations induced by (3) matches the structure of $\text{cl conv}(X)$, or if they are overly simple or complex.

Contributions and outline

In this paper, we close the aforementioned gaps in the literature by characterizing the structure of $\text{cl conv}(X)$. First, in Sect. 2, we review relevant background for the paper. In Sect. 3, we show that $\text{cl conv}(X)$ can be described in a compact extended formulation with $\mathcal{O}(n^2)$ additional variables with linear constraints and an $(n+1) \times (n+1)$ positive semidefiniteness constraint. In particular, convexification of X in this extended formulation reduces to describing a *base* polytope. We use the vertex description of this base polytope, which is exponential in general. However, we show that the set of vertices can be represented as the feasible points of a compact mixed-integer linear formulation (Sect. 5). In Sect. 4, we characterize $\text{cl conv}(X)$ in the original space of variables. While the resulting description has an infinite number of conic quadratic constraints, we show that $\text{cl conv}(X)$ is *finitely generated*, and thus we establish which inequalities are necessary to describe $\text{cl conv}(X)$—in precisely the same manner that facet-defining inequalities are required to describe a polyhedron. We also establish a relationship between $\text{cl conv}(X)$ and relaxations obtained from decompositions (3). In Sect. 5, we present a mixed-integer *linear* formulation of the MIQO problem using the theoretical results in Sect. 3. Finally, in Sect. 6 we conclude the paper with a few remarks.

We point out that, using standard disjunctive programming techniques [14, 20], it is possible to obtain a conic quadratic extended formulation of (1), although such representation typically requires adding $\mathcal{O}(|Z|n)$ number of variables and $\mathcal{O}(|Z|)$ *nonlinear* constraints. Since $|Z|$ is often exponential in n, these formulations are in general impractical, and therefore their use has been restricted to small instances with $n \leq 2$ [4, 7, 20, 22, 24] or problems with special structures that admit a compact representation [23]. We argue that the convexifications in this paper are significantly more tractable: regardless of Z, we require only $\mathcal{O}(n^2)$ variables instead of $\mathcal{O}(|Z|n)$, and only *one* nonlinear conic constraint instead of $\mathcal{O}(|Z|)$. The major complexity of the proposed formulations in this paper is the exponential number of *linear* inequalities, which can be generated, as needed, using mature mixed-integer linear optimization techniques.

2 Notation and preliminaries

In this section, we first review the relevant background and introduce the notation used in the paper.

Definition 1 ([31]) Given a matrix $W \in \mathbb{R}^{p\times q}$, its pseudoinverse $W^\dagger \in \mathbb{R}^{q\times p}$ is the unique matrix satisfying the four properties:

$$WW^\dagger W = W, \quad W^\dagger W W^\dagger = W^\dagger, \quad (WW^\dagger)^\top = (WW^\dagger), \quad (W^\dagger W)^\top = W^\dagger W.$$

Clearly, if W is invertible, then $W^{-1} = W^\dagger$. It also readily follows from the definition that $(W^\dagger)^\dagger = W$.

We recall the generalized Schur complement, relating pseudoinverses and positive semidefinite matrices.

Lemma 1 *([3]) Let* $W = \begin{pmatrix} W_{11} & W_{12} \\ W_{12}^\top & W_{22} \end{pmatrix}$, *with symmetric* $W_{11} \in \mathbb{R}^{p\times p}$, *symmetric* $W_{22} \in \mathbb{R}^{q\times q}$, *and* $W_{12} \in \mathbb{R}^{p\times q}$. *Then* $W \succeq 0$ *if and only if* $W_{11} \succeq 0$, $W_{11}W_{11}^\dagger W_{12} = W_{12}$ *and* $W_{22} - W_{12}^\top W_{11}^\dagger W_{12} \succeq 0$.

Note that if $W_{11} \succ 0$, then the second condition of Lemma 1 is automatically satisfied. Otherwise, this condition is equivalent to the system of equalities $W_{11}U = W_{12}$ having a solution $U \in \mathbb{R}^{p\times q}$.

Let $[n] = \{1, \ldots, n\}$. Throughout, we use the convention that $x_i^2/z_i = 0$ if $x_i = z_i = 0$ and $x_i^2/z_i = +\infty$ if $z_i = 0$ and $x_i \neq 0, i \in [n]$. For a vector $a \in \mathbb{R}^n$, $\|a\|_2$ and $\|a\|_\infty$ denote the vector ℓ_2-norm and the maximum absolute value among a_i's, respectively. Given two matrices V, W of matching dimensions, let $\langle V, W\rangle = \sum_i \sum_j V_{ij}W_{ij}$ denote the usual inner product. Given a matrix $W \in \mathbb{R}^{n\times n}$, let $\mathrm{Tr}(W) = \sum_{i=1}^n W_{ii}$ denote its trace, and let W^{-1} denote its inverse, if it exists. Let $\|W\|_F$ and $\|W\|_{\max}$denote the *Frobenius* norm and the maximum absolute value of entries of W respectively, and $\lambda_{\max}(W)$ means the maximum eigenvalue of W. We let $\mathrm{col}(W)$ denote the column space of matrix W. Given a matrix $W \in \mathbb{R}^{n\times n}$ and $S \subseteq [n]$, let $W_S \in \mathbb{R}^{S\times S}$ be the submatrix of W induced by S, and let $\hat{W}_S \in \mathbb{R}^{n\times n}$ be the $n \times n$ matrix obtained from W_S by filling the missing entries with zeros, i.e., matrices subscripted by S without "hat" refer to the lower-dimensional submatrices. For any two sets $S, T \subset [n]$, let $W_{S,T}$ denote the submatrix of W with rows in S and columns in T. Note that if matrix $W \succ 0$, then it can be easily be verified from Definition 1 that the submatrix of $\hat{W}_S^\dagger$ indexed by S coincides with W_S^{-1}, and $\hat{W}_S^\dagger$ is zero elsewhere; in this case, we abuse notation and write $\hat{W}_S^{-1}$ instead of $\hat{W}_S^\dagger$. Given $S \subseteq [n]$, let $\hat{\boldsymbol{e}}_S \in \{0, 1\}^n$ be the indicator vector of S. We define π_S as the projection onto the subspace indexed by S and $\pi_S^{-1}(x)$ as the preimage of x under π_S.

Example 1 Let $Q = \begin{pmatrix} d_1 & b \\ b & d_2 \end{pmatrix}$ with $d_1, d_2 > 0$ and $d_1 d_2 > b^2$. Then

$$\hat{Q}_\emptyset^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \ \hat{Q}_{\{1\}}^{-1} = \begin{pmatrix} 1/d_1 & 0 \\ 0 & 0 \end{pmatrix}, \ \hat{Q}_{\{2\}}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & 1/d_2 \end{pmatrix}, \text{ and}$$
$$Q_{\{1,2\}}^{-1} = \frac{1}{d_1 d_2 - b^2} \begin{pmatrix} d_2 & -b \\ -b & d_1 \end{pmatrix}.$$

3 Convexification in an extended space

In this section, we describe $\mathrm{cl\,conv}(X)$ in an extended space. In Sect. 3.1, we provide a "canonical" representation of $\mathrm{cl\,conv}(X)$ under the assumption that $Q \succ 0$. In Sect. 3.2, we provide alternative representations of $\mathrm{cl\,conv}(X)$, which can handle non-invertible matrices Q and may also lead to sparser formulations.

3.1 Canonical representation

Given $Q \succ 0$, define the polytope $P \subseteq \mathbb{R}^{n+n^2}$ as

$$P \overset{\text{def}}{=} \operatorname{conv}\left(\left\{(\hat{\boldsymbol{e}}_S, \hat{Q}_S^{-1})\right\}_{\hat{\boldsymbol{e}}_S \in Z}\right).$$

Proposition 1 below shows how to construct mixed-integer conic formulations of MIQO using polytope P.

Proposition 1 *If $Q \succ 0$, then the mixed-integer optimization model*

$$\min_{x,z,W,t} \; a^\top x + b^\top z + \tfrac{1}{2}t \tag{4a}$$

$$\text{s.t.} \begin{pmatrix} W & x \\ x^\top & t \end{pmatrix} \succeq 0 \tag{4b}$$

$$(z, W) \in P \tag{4c}$$

$$z \in \{0,1\}^n \tag{4d}$$

$$x \in \mathbb{R}^n, t \in \mathbb{R} \tag{4e}$$

is a valid formulation of problem (1).

Proof Consider a point (x, z, t, W) satisfying constraints (4b), (4c) with $z = \hat{\boldsymbol{e}}_S$ for some $\hat{\boldsymbol{e}}_S$. Constraint (4c) is satisfied if and only if $W = \hat{Q}_S^{-1}$. Therefore, constraint (4b) reduces to

$$\begin{pmatrix} Q_S^{-1} & \mathbf{0} & x_S \\ \mathbf{0} & \mathbf{0} & x_{[n]\setminus S} \\ x_S^\top & x_{[n]\setminus S}^\top & t \end{pmatrix} \succeq 0.$$

Since the pseudoinverse of matrix $W = \begin{pmatrix} Q_S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ is $W^\dagger = \begin{pmatrix} Q_S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, we find from Lemma 1 that constraint (4b) is satisfied if and only if:

- $W \succeq 0$, which is automatically satisfied.
- $WW^\dagger x = x \Leftrightarrow \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} x_S \\ x_{[n]\setminus S} \end{pmatrix} = \begin{pmatrix} x_S \\ x_{[n]\setminus S} \end{pmatrix} \Leftrightarrow x_{[n]\setminus S} = 0$. Thus, condition $WW^\dagger x = x$ simply enforces the complementarity constraints $x \circ (\boldsymbol{e} - z) = 0$.
- $t \geq x^\top W^\dagger x \Leftrightarrow t \geq x_S^\top Q_S x_S$, which is precisely the nonlinear constraint defining set X.

Now, it is clear that for any (x, z, t, W) satisfying constraints (4b), (4c), (4d), it holds $(x, z, t) \in X$. On the other hand, for any $(x, z, t) \in X$ with $z = \hat{\boldsymbol{e}}_S$ for some $S \subset [n]$, we can always let $W = \hat{Q}_S^{-1}$ and similarly, (x, z, W, t) satisfies constraints (4b), (4c), (4d). □

Note that condition $WW^\dagger x = x$ is used to enforce the complementarity constraints. We point out that a similar idea was recently used in the context of low-rank optimization [11].

Now consider the convex relaxation of (4), obtained by dropping the integrality constraints $z \in \{0, 1\}^n$:

$$\min_{x,z,W,t} \; a^\top x + b^\top z + \tfrac{1}{2}t \tag{5a}$$

$$\text{s.t. (4b), (4c), (4e).} \tag{5b}$$

Theorem 1 *Let Q be a positive definite matrix. Then*

$$cl\ conv(X) \;=\; \{(z, x, t) \in [0, 1]^n \times \mathbb{R}^{n+1} \mid \exists W \in \mathbb{R}^{n\times n} s.t.\ (4b), (4c)\}.$$

Consequently, that problem (5) has an optimal solution integral in z.

Proof First observe that constraints (4b),(4c) define a closed convex set. Projecting out variable t, we find that problem (5) reduces to

$$\min_{x,z,W} \; a^\top x + b^\top z + \tfrac{1}{2}x^\top W^\dagger x \tag{6a}$$

$$\text{s.t. } WW^\dagger x = x \tag{6b}$$

$$(z, W) \in P,\;\; x \in \mathbb{R}^n. \tag{6c}$$

Note that this formulation uses the pseudoinverse of a matrix of variables. Observe that we omit the constraint $W \succeq 0$. Since every extreme point $(\bar{z}, \bar{W})$ of P satisfies $\bar{W} \succeq 0$, it follows $(z, W) \in P$ already implies $W \succeq 0$.

We argue that for any fixed $(z, W) \in P$, setting $x = -Wa$ is optimal for (6). Using equality (6b), we replace the term $a^\top x$ in the objective with $a^\top WW^\dagger x$. Since the problem is convex in x, from KKT conditions we find that any point x satisfying

$$WW^\dagger x = x \tag{7a}$$

$$\exists \lambda \in \mathbb{R}^n \text{ s.t. } W^\dagger Wa + W^\dagger x + \lambda^\top (WW^\dagger - I) = 0 \tag{7b}$$

is optimal. In particular, setting $x = -Wa$, we find that (7b) is satisfied with $\lambda = 0$, and (7a) is satisfied since $WW^\dagger x = -WW^\dagger Wa = -Wa = x$.

Substituting $x = -Wa$ in the relaxed problem, we obtain

$$\min_{z,W} \; -\tfrac{1}{2}a^\top Wa + b^\top z \tag{8a}$$

$$\text{s.t. } (z, W) \in P. \tag{8b}$$

Since the objective $-\frac{1}{2}\langle aa^\top, W\rangle + b^\top z$ is linear in (z, W) and P is a polytope, there exists an optimal solution (z^*, W^*) that is an extreme point of P, and in particular there exists $\hat{\boldsymbol{e}}_S \in Z$ such that $z^* = \hat{\boldsymbol{e}}_S$ and $W^* = \hat{Q}_S^{-1}$. □

Remark 1 The convexification for the case where Q is tridiagonal [29] is precisely in the form given in Theorem 1, where the polyhedron P is described with a compact extended formulation. □

3.1.1 Bivariate quadratic functions

Consider set

$$X_{2\times 2}=\left\{(x,z,t)\in\mathbb{R}^2\times\{0,1\}^n\times\mathbb{R}: t\geq d_1x_1^2-2x_1x_2+d_2x_2^2,\ x\circ(\boldsymbol{e}-z)=0\right\},$$

where $d_1d_2>1, d_1, d_2>0$. Set $X_{2\times 2}$ corresponds (after scaling) to a generic strictly convex quadratic function of two variables. We now illustrate Theorem 1 by computing an extended formulation of $\operatorname{cl}\operatorname{conv}(X_{2\times 2})$, that is, for $Q=\begin{pmatrix} d_1 & -1\\ -1 & d_2\end{pmatrix}$. Let $\Delta:=d_1d_2-1>0$ be the determinant of Q.

Proposition 2 *The closure of the convex hull of $X_{2\times 2}$ is*

$$\begin{aligned}cl\ conv(X_{2\times 2})=&\left\{(x,z,t)\in\mathbb{R}^5:\exists W\in\mathbb{R}^{2\times 2}\ such\ that\ \begin{pmatrix} W_{11} & W_{12} & x_1\\ W_{12} & W_{22} & x_2\\ x_1 & x_2 & t\end{pmatrix}\succeq 0,\right.\\ &0\leq z_1\leq 1,\ 0\leq z_2\leq 1,\ d_1W_{11}=W_{12}+z_1,\ d_2W_{22}=z_2+W_{12},\\ &\left.W_{12}\geq 0,\ \Delta W_{12}\geq -1+z_1+z_2,\ \Delta W_{12}\leq z_1,\ \Delta W_{12}\leq z_2\right\}.\end{aligned}$$

Proof Polyhedron P is the convex hull of the four points given in Table 1.

Note that equalities $W_{11}=\frac{1}{d_1}(z_1+W_{12})$ and $W_{22}=\frac{1}{d_2}(z_2+W_{12})$ are valid. Letting $w=W_{12}$ and projecting out variables W_{11} and W_{22}, we find that

$$W=\begin{pmatrix}\frac{1}{d_1}z_1 & 0\\ 0 & \frac{1}{d_2}z_2\end{pmatrix}+\begin{pmatrix}1/d_1 & 1\\ 1 & 1/d_2\end{pmatrix}w. \tag{9}$$

Also note that $w=\frac{1}{\Delta}\min\{z_1,z_2\}$, and the convex hull of $\left\{(z_1,z_2,w)\in\{0,1\}^2\times\mathbb{R}\mid w=\frac{1}{\Delta}\min\{z_1,z_2\}\right\}$ is described by the following inequalities:

$$w\geq 0,\ w\geq\frac{1}{\Delta}(-1+z_1+z_2),\ w\leq\frac{1}{\Delta}z_2,\ w\leq\frac{1}{\Delta}z_1, 0\leq z_1,z_2\leq 1 \tag{10}$$

Then, (9) and (10) describe the polyhedron P. □

Conic quadratic disjunctive programming representations of $\operatorname{cl}\operatorname{conv}(X_{2\times 2})$ have been used in the literature [4]; explicit representations of $\operatorname{cl}\operatorname{conv}(X_{2\times 2}\cap\{(x,z,t):x\geq 0\})$ in the original space of variables have been given [8, 24], and descriptions of the rank-one case $d_1d_2=1$ were given in [5]. A description of $\operatorname{cl}\operatorname{conv}(X_{2\times 2}\cap\{(x,z,t):\ell\leq x\leq u\})$ in a conic quadratic extended formulation is given in [20] via disjunctive programming. This formulation can be easily adapted to

Table 1 Extreme points of P corresponding to set $X_{2\times 2}$

z_1	z_2	W
0	0	$\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$
1	0	$\begin{pmatrix}1/d_1 & 0\\ 0 & 0\end{pmatrix}$
0	1	$\begin{pmatrix}0 & 0\\ 0 & 1/d_2\end{pmatrix}$
1	1	$\frac{1}{\Delta}\begin{pmatrix}d_2 & 1\\ 1 & d_1\end{pmatrix}$

the case with no bounds (considered here), and requires three additional variables and three conic quadratic constraints. In Proposition 2, we give an alternative description of cl conv$(X_{2\times 2})$ using three additional variables, a compact 3×3 positive semidefinite constraint, and linear inequalities.

Remark 2 Since P is not full-dimensional, we require only one additional variable w (instead of three) for conic representation of cl conv$(X_{2\times 2})$ via the constraints $0 \le z \le 1$, (10), and

$$\begin{pmatrix}(1/d_1)(z_1+w) & w & x_1\\ w & (1/d_2)(z_2+w) & x_2\\ x_1 & x_2 & t\end{pmatrix} \succeq 0. \qquad \square$$

Remark 3 The matrix representation (9) suggests an interesting connection between cl conv$(X_{2\times 2})$ and McCormick envelopes. Indeed, from Table 1, we see that

$$W = \begin{pmatrix}1/d_1 & 0\\ 0 & 0\end{pmatrix} z_1 + \begin{pmatrix}0 & 0\\ 0 & 1/d_2\end{pmatrix} z_2 + \frac{1}{\Delta}\begin{pmatrix}1/d_1 & 1\\ 1 & 1/d_2\end{pmatrix} z_1 z_2.$$

Moreover, the usual McCormick envelopes of the bilinear term $z_1 z_2$, given by $\max\{0, -1+z_1+z_2\} \le z_1 z_2 \le \min\{z_1, z_2\}$, are sufficient to characterize the convex hull. □

3.1.2 Quadratic functions with “choose-one" constraints

Given $Q \succ 0$, consider set

$$X_{C1} = \left\{(x, z, t) \in \mathbb{R}^n \times \{0,1\}^n \times \mathbb{R} : t \ge x^\top Q x,\ x \circ (\boldsymbol{e} - z) = 0,\ \sum_{i=1}^{n} z_i \le 1\right\}.$$

Set X_{C1} arises, for example, in regression problems with multicollinearity constraints [10]: given a set J of features that are collinear, constraints $\sum_{i\in J} z_i \le 1$ are used to ensure that at most one such feature is chosen.

The closure of the convex hull of X_{C1} is [see, e.g.,20, 33]

$$\text{cl conv}(X_{C1}) = \left\{(x, z, t) \in \mathbb{R}^n \times \mathbb{R}^n_+ \times \mathbb{R} : t \geq \sum_{i=1}^n Q_{ii}x_i^2/z_i,\ \sum_{i=1}^n z_i \leq 1\right\}.$$

We now give an alternative derivation of this result using our technique. Polyhedron P is the convex hull of $n+1$ points: point $(0, 0)$ and points $\{(\hat{\boldsymbol{e}}_{\{i\}}, \hat{Q}^{-1}_{\{i\}})\}_{i=1}^n$. It can easily be seen that P is described by constraints $W_{ij} = 0$ whenever $i \neq j$, $W_{ii} = z_i/Q_{ii}$ for $i \in [n]$, and constraints $z \geq 0$, $\sum_{i=1}^n z_i \leq 1$. In particular, constraint (4b) reduces to

$$\begin{pmatrix} z_1/Q_{11} & 0 & \dots & 0 & x_1 \\ 0 & z_2/Q_{22} & \dots & 0 & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & z_n/Q_{nn} & x_n \\ x_1 & x_2 & \dots & x_n & t \end{pmatrix} \succeq 0,$$

which by Lemma 1 is equivalent to

$$t \geq \sum_{i=1}^n Q_{ii}x_i^2/z_i,\ z_i/Q_{ii} \geq 0,$$

and $x_i = 0$ if $z_i/Q_{ii} = 0$, $\forall i \in [n]$. Note that the second condition is the complementarity constraint, which is already included in the constraint $t \geq \sum_{i=1}^n Q_{ii}x_i^2/z_i$ (since $z_i = 0$ and $x_i > 0$ implies $\frac{x_i^2}{z_i} = +\infty$).

3.2 Factorable representation

A (possibly low-rank) matrix $Q \in \mathbb{R}^{n\times n}$ is positive semidefinite if and only if there exists some $F \in \mathbb{R}^{n\times k}$ such that $Q = FF^\top$. Then, letting $u = F^\top x$, one can rewrite $x^\top Qx$ as $x^\top FF^\top x = u^\top u$. Matrix F may be immediately available when formulating the problem, or may be obtained through a Cholesky decomposition or eigendecomposition of Q. Such a factorization is often employed by solvers, since it results in simpler (separable) nonlinear terms, and in many situations matrix F is sparse as well. In this section, we discuss representations of cl conv(X) amenable to such factorizations of Q. While the proofs of the propositions of this section are similar to those in Sect. 3.1, additional care is required to handle unbounded problems (1) arising from a rank-deficient Q.

Given $F \in \mathbb{R}^{n\times k}$, define $F_S \in \mathbb{R}^{S\times k}$ as the submatrix of F corresponding to the rows indexed by S, and let $\hat{F}_S \in \mathbb{R}^{n\times k}$ be the matrix obtained by filling the missing entries with zeros. Define the polytope $P_F \subseteq \mathbb{R}^{n+k^2}$ as

$$P_F = \text{conv}\left(\left\{(\hat{\boldsymbol{e}}_S, \hat{F}_S^\dagger \hat{F}_S)\right\}_{\hat{\boldsymbol{e}}_S \in Z}\right).$$

Remark 4 For any $S \subseteq [n]$, matrix $\hat{F}_S^\dagger \hat{F}_S$ is an orthogonal projection matrix (symmetric and idempotent), and in particular $(\hat{F}_S^\dagger \hat{F}_S)^\dagger = \hat{F}_S^\dagger \hat{F}_S$. These properties can be easily verified from Definition 1. Since all eigenvalues of an orthogonal projection matrix are either 0 or 1, it also follows that $\hat{F}_S^\dagger \hat{F}_S \succeq 0$. □

Proposition 3 *If $Q = FF^\top$, then the mixed-integer optimization model*

$$\min_{x,z,W,t} \; a^\top x + b^\top z + \tfrac{1}{2}t \tag{11a}$$

$$\text{s.t.} \; \begin{pmatrix} W & F^\top x \\ x^\top F & t \end{pmatrix} \succeq 0 \tag{11b}$$

$$(z, W) \in P_F \tag{11c}$$

$$z \in \{0,1\}^n, \;\; x \circ (\boldsymbol{e} - z) = 0 \tag{11d}$$

$$x \in \mathbb{R}^n, t \in \mathbb{R} \tag{11e}$$

is a valid formulation of problem (1).

Proof Consider a point $(x, z, t) \in X$ with $z = \hat{\boldsymbol{e}}_S$ for some $\hat{\boldsymbol{e}}_S \in Z$. Constraint (11d) is trivially satisfied. Constraint (11c) is satisfied if and only if $W = \hat{F}_S^\dagger \hat{F}_S$. Note that in any feasible solution, $x_i = 0$ whenever $i \notin S$, and in particular $F^\top x = \hat{F}_S^\top x$. From Lemma 1, we find that constraint (11b) is satisfied if and only if (recall properties in Remark 4):

- $\hat{F}_S^\dagger \hat{F}_S \succeq 0$, which is automatically satisfied.
- $\hat{F}_S^\dagger \hat{F}_S (\hat{F}_S^\dagger \hat{F}_S)^\dagger F^\top x = F^\top x$. We find that

  $$\hat{F}_S^\dagger \hat{F}_S (\hat{F}_S^\dagger \hat{F}_S)^\dagger \hat{F}_S^\top x = \hat{F}_S^\dagger \hat{F}_S \hat{F}_S^\dagger \hat{F}_S \hat{F}_S^\top x = \hat{F}_S^\dagger \hat{F}_S \hat{F}_S^\top x = \hat{F}_S^\top (\hat{F}_S^\dagger)^\top \hat{F}_S^\top x = \hat{F}_S^\top x,$$

  and, therefore, this condition is satisfied as well.
- $t \geq x^\top F W^\dagger F^\top x \;\Leftrightarrow\; t \geq x_S^\top \hat{F}_S (\hat{F}_S^\dagger \hat{F}_S)^\dagger \hat{F}_S^\top x_S = x_S^\top \hat{F}_S \hat{F}_S^\dagger \hat{F}_S \hat{F}_S^\top x_S = x_S^\top \hat{F}_S \hat{F}_S^\top x_S$, which is precisely the nonlinear constraint defining set X and is thus satisfied.

□

While the proofs of Propositions 1 and 3 are similar in spirit, we highlight a critical difference. In the proof of Proposition 1, with the assumption $Q \succ 0$, constraints $WW^\dagger x = x$ enforce the complementarity constraints $x \circ (\boldsymbol{e} - z) = 0$, and therefore, such constraints are excluded in (4). In contrast, in the proof of Proposition 3, with Q potentially of low-rank, constraints $WW^\dagger F^\top x = F^\top x$ alone are not sufficient to enforce $x \circ (\boldsymbol{e} - z) = 0$, and therefore, they are included in (11) and are used to prove the validity of the mixed-integer formulation. Indeed, if there exist $\hat{\boldsymbol{e}}_S \in Z$ and $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_S \neq 0$, $\bar{x}_{[n]\setminus S} = 0$ and $F^\top \bar{x} = 0$, then for any $(x, z, t) \in X$ we find that

$$\lim_{\lambda \to 0^+} (1 - \lambda)(x, z, t) + \lambda((1/\lambda)\bar{x}, \hat{\boldsymbol{e}}_S, 0) = (x + \bar{x}, z, t) \in \operatorname{cl} \operatorname{conv}(X).$$

In particular, the point $(x + \bar{x}, z, t)$, which may not satisfy the complementarity constraints, cannot be separated from cl conv(X), or any closed relaxation. On the other hand, if matrix Q is full-rank, then $F^\top \bar{x} = 0 \implies \bar{x} = 0$ (as shown in the proof of Proposition 1); therefore, the complementarity constraints are enforced by the conic constraint.

Recall that $\pi_S : \mathbb{R}^n \to \mathbb{R}^S$ is the projection onto the subspace indexed by S. Now we consider the natural convex relaxation of (11) by dropping constraint (11d), and show that it is ideal under certain technical conditions over F and the set Z, as stated in Theorem 2 below.

Theorem 2 *Let $Q = FF^\top$, where $F \in \mathbb{R}^{n\times k}$ is a full-column rank matrix satisfying* $\text{col}(F) = \bigcap_{\hat{e}_S \in Z} \pi_S^{-1}(\text{col}(F_S))$. *Then*

$$\textit{cl conv}(X) \;=\; \{(z, x, t) \in [0, 1]^n \times \mathbb{R}^{n+1} \mid \exists W \in \mathbb{R}^{k\times k} \text{ s.t. } (11b), (11c)\}.$$

Proof Clearly, constraints (11b),(11c) define a closed convex set. Consider the two optimization problems:

$$\begin{aligned} \min \quad & a^\top x + b^\top z + \tfrac{1}{2} t && (12a)\\ \text{s.t.} \quad & (x, z, t) \in \text{cl conv}(X), && (12b) \end{aligned}$$

and

$$\begin{aligned} \min \quad & a^\top x + b^\top z + \tfrac{1}{2} t && (13a)\\ \text{s.t.} \quad & \begin{pmatrix} W & F^\top x \\ x^\top F & t \end{pmatrix} \succeq \mathbf{0}, && (13b)\\ & (z, W) \in P_F,\ \ x \in \mathbb{R}^n, t \in \mathbb{R}. && (13c) \end{aligned}$$

It suffices to show that problem (12) and (13) always attain the same optimal value. Consider the following two cases:

- $FF^\dagger a \neq a$: In other words, a is not in the column space of F, i.e., $a \notin \text{col}(F)$. In this case, by the condition $\text{col}(F) = \bigcap_{\hat{e}_S \in Z} \pi_S^{-1}(\text{col}(F_S))$, there exists one $\hat{e}_S \in Z$ such that $a_S \notin \text{col}(F_S)$. Then, let z be such that $z_i = 1,\ \forall i \in S$. Since $a_S \notin \text{col}(F_S)$, there exists x such that $x_i = 0$ for all $i \in [n]\backslash S$, x_S is in the orthogonal complement of F_S and $a_S^\top x_S < 0$. Clearly, z and x satisfy the constraint $x_i(1-z_i) = 0$ for all $i = 1, \ldots, n$. Complementarity holds for λx for $\lambda > 0$ as well. Since, by construction, $x^\top FF^\top x = 0$, the objective $b^\top z + \lambda\langle a, x\rangle + \lambda^2(x^\top FF^\top x)$ tends to $-\infty$ for $(\lambda x, z)$ as $\lambda \to \infty$. Thus problem (12) is unbounded and since problem (13) is a convex relaxation of (12), problem (13) is unbounded as well.
- $FF^\dagger a = a$: For problem (13), we can project out t using the relation

$$\begin{pmatrix} W & F^\top x \\ x^\top F & t \end{pmatrix} \succeq 0 \quad \text{iff} \quad WW^\dagger F^\top x = F^\top x \quad \text{and} \quad t \geq x^\top FW^\dagger F^\top x.$$

Therefore, problem (13) is equivalent to

$$\begin{aligned} \min \quad & a^\top x + b^\top z + \tfrac{1}{2} x^\top F W^\dagger F^\top x && \text{(14a)}\\ \text{s.t.} \quad & W W^\dagger F^\top x = F^\top x && \text{(14b)}\\ & (z, W) \in P_F,\ x \in \mathbb{R}^n. && \text{(14c)} \end{aligned}$$

Since $FF^\dagger a = a$, we can write $a^\top x = (F^\dagger a)^\top F^\top x$. Define $\tilde{a} = F^\dagger a$, then $a^\top x = \tilde{a}^\top F^\top x$. Substituting $F^\top x$ with a new variable $u \in \mathbb{R}^k$ and since F has full column rank, problem (14) is equivalent to

$$\begin{aligned} \min \quad & b^\top z + \tilde{a}^\top u + \tfrac{1}{2} u^\top W^\dagger u && \text{(15a)}\\ \text{s.t.} \quad & W W^\dagger u = u && \text{(15b)}\\ & (z, W) \in P_F, u \in \mathbb{R}^k. && \text{(15c)} \end{aligned}$$

Using identical arguments as in the proof of Theorem 1, we find that there exists $\hat{\boldsymbol{e}}_S \in Z$ such that $(u^*, z^*, W^*) = (-\hat{F}_S^\dagger \hat{F}_S \tilde{a}, \hat{\boldsymbol{e}}_S, \hat{F}_S^\dagger \hat{F}_S)$ is optimal for (15). We now construct an optimal solution for (14). Let x^* be defined as $x_S^* = -(F_S^\dagger)^\top F_S^\dagger a_S$ and $x_{[n]\setminus S}^* = 0$, and observe that (x^*, z^*) is feasible for (12), with objective $\sum_{i \in S} b_i - \frac{1}{2} \| F_S^\dagger a_S \|_2^2$. Substituting $W^* = \hat{F}_S^\dagger \hat{F}_S$, the optimal value of problem (13) equals $\sum_{i \in S} b_i - \frac{1}{2} \| F_S^\dagger F_S F^\dagger a \|_2^2$. Note that both $\alpha_1 = F^\dagger a$ and $\alpha_2 = F_S^\dagger a_S$ satisfy the equation $F_S \alpha = a_S$ and thus $\alpha_1 - \alpha_2$ is orthogonal to the row space of F_S which means $F_S^\dagger F_S \alpha_1 = F_S^\dagger F_S \alpha_2 = \alpha_2$. Hence, we conclude that the optimal values of problem (12) and problem (13) coincide. □

Remark 5 From the first case analysis of the proof of Theorem 2, one sees that the technical condition $\operatorname{col}(F) = \bigcap_{\hat{\boldsymbol{e}}_S \in Z} \pi_S^{-1}(\operatorname{col}(F_S))$ is equivalent to stating that the mixed-integer optimization problem and the proposed convex relaxation are unbounded at the same time. The condition is automatically satisfied if $\boldsymbol{e} \in Z$. Moreover, if matrix Q is rank-one, then this condition is equivalent to the nondecomposability condition on Z given in [34]. If it fails to hold, the convexification presented is still valid but may be weak: the convex relaxation may be unbounded even if the mixed-integer optimization problem is bounded. We provide an example illustrating this phenomenon in Sect. 3.2.3. □

Remark 6 An immediate consequence of Theorem 2 is that if matrix Q is rank-deficient, i.e., $k < n$, then the extended formulation describing $\operatorname{cl}\operatorname{conv}(X)$ is simpler than the full rank case, i.e., it has fewer additional variables and lower-dimensional conic constraints. □

We now illustrate Theorem 2 by providing an alternative proof of the main result of [5] using our unifying framework.

3.2.1 Rank-one quadratic functions

Consider the rank-one set

$$X_{R1} = \left\{(x, z, t) \in \mathbb{R}^n \times \{0,1\}^n \times \mathbb{R} : t \geq \left(h^\top x\right)^2,\ x \circ (\boldsymbol{e} - z) = 0\right\},$$

where we assume $h_i \neq 0$ for all $i \in [n]$.

Proposition 4 *([5]) The closure of the convex hull of X_{R1} is*

$$cl\ conv(X_{R1}) = \left\{(x, z, t) \in \mathbb{R}^{2n+1} : \begin{pmatrix} \min\{1, \boldsymbol{e}^\top z\} & h^\top x \\ h^\top x & t \end{pmatrix} \succeq 0,\ 0 \leq z \leq \boldsymbol{e}\right\}.$$

Proof In the case of a rank-one function, we have $F = h$ and $W \in \mathbb{R}^1$. Note that the pseudoinverse of vector $\hat{h}_S$ is given by

$$\hat{h}_S^\dagger = \begin{cases} 0 & \text{if } \hat{h}_S = 0 \\ \hat{h}_S^\top / (\hat{h}_S^\top \hat{h}_S) & \text{otherwise,} \end{cases}$$

and, in particular, we find that $\hat{h}_S^\dagger \hat{h}_S = 1$ if $S \neq \emptyset$, and $\hat{h}_S^\dagger \hat{h}_S = 0$ otherwise. Thus, $\hat{h}_S^\dagger \hat{h}_S = \max\{z_1, \ldots, z_n\}$, and P_F is described by the linearization $0 \leq W \leq \min\{1, \boldsymbol{e}^\top z\}$. Projecting out variable W, we arrive at the result. □

We discuss generalizations of X_{R1} with arbitrary constraints on the indicator variables in Sect. 4.

3.2.2 An example with a rank-two quadratic function

In order to illustrate how convexification methods for polyhedra can be directly utilized to convexify the mixed-integer nonlinear set X, we consider a special rank-two quadratic function with three variables and the associated set

$$X_3 = \left\{(x, z, t) \in \mathbb{R}^3 \times \{0,1\}^3 \times \mathbb{R} : t \geq (x_1 + x_2 + x_3)^2 + x_3^2,\ x \circ (\boldsymbol{e} - z) = 0\right\}.$$

In this case, $Q = FF^\top$ with $F^\top = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. The extreme points of P_F are given in Table 2. Using PORTA [32] to switch from the extreme point representation of P_F to its facial description, we obtain the closure of the convex hull of X_3:

$$\text{cl conv}(X_3) = \Bigg\{(x, z, t) \in \mathbb{R}^7 : \exists W \in \mathbb{R}^{2\times 2} \text{ such that}$$

$$\begin{pmatrix} W_{11} & W_{12} & x_1 + x_2 + x_3 \\ W_{12} & W_{22} & x_3 \\ x_1 + x_2 + x_3 & x_3 & t \end{pmatrix} \succeq 0,$$

Table 2 Extreme points of P_F corresponding to set X_3

z	$\hat{F}_S^\top$	$\hat{F}_S^\dagger$	$\hat{F}_S^\dagger \hat{F}_S$
$(0, 0, 0)$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
$(0, 0, 1)$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}$	$\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$
$(0, 1, 0)$	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
$(0, 1, 1)$	$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$(1, 0, 0)$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
$(1, 0, 1)$	$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$(1, 1, 0)$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
$(1, 1, 1)$	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1/2 & 1/2 & 0 \\ -1/2 & -1/2 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$\begin{aligned} & z_3 = W_{12} + W_{22},\ 0 \le W_{12} \le W_{22} \le W_{11}, \\ & z_3 + \max\{z_1, z_2\} \le W_{11} + W_{22} \le z_1 + z_2 + z_3, \\ & \left. W_{11} + 2W_{12} + W_{22} \le 1 + z_3 \right\}. \end{aligned}$$

3.2.3 An example where the technical condition fails

Consider the set

$$X_{R1}^{C1} = \left\{ (x, z, t) \in \mathbb{R}^n \times \{0, 1\}^n \times \mathbb{R} : t \ge \left(h^\top x \right)^2, x \circ (\boldsymbol{e} - z) = 0, \sum_{i=1}^n z_i \le 1 \right\}$$

with $h_i \neq 0$ for $i \in [n]$. In this case, $F = h$ and $\text{col}(F_{\{i\}}) = \mathbb{R}$ and $\pi_S^{-1}(\text{col}(F_{\{i\}})) = \mathbb{R}^n$. Thus, $\bigcap_{\hat{\boldsymbol{e}}_S \in Z} \pi_S^{-1}(\text{col}(F_S)) = \mathbb{R}^n$, while $\text{col}(F) = \{x \in \mathbb{R}^n : x = \lambda h \text{ for some } \lambda \in \mathbb{R}\}$, and the technical assumption is not satisfied.

The relaxation induced by (11b), (11c), (11e), which is constructed as outlined in Proposition 4, results in the set induced by bound constraints $0 \le z \le 1$, $\boldsymbol{e}^\top z \le 1$ and $t \ge (h^\top x)^2/(\boldsymbol{e}^\top z)$. Moreover, the corresponding optimization problem

$$\min_{x,z}\ a^\top x + b^\top z + (h^\top x)^2/(\boldsymbol{e}^\top z) \text{ s.t. } \boldsymbol{e}^\top z \le 1,\ x \in \mathbb{R}^n,\ z \in [0, 1]^n$$

is unbounded unless $a \in \text{col}(F)$.

In contrast, $\text{cl conv}(X_{R1}^{C1})$ is described via constraint $t \ge \sum_{i=1}^n h_i^2 x_i^2 / z_i$ [33, 34] (similar to the result described in Sect. 3.1.2), and the corresponding optimization problem is always bounded.

4 Convexification in the original space

We now turn our attention to describing cl conv(X) in the original space of variables. The discussion of this section is based on projecting out the matrix variable W in the canonical description of cl conv(X) given in Theorem 1 for $Q \succ 0$. Identical arguments hold for the representation in Theorem 2 for low-rank matrices.

Suppose that a minimal description of polyhedron P is given by the facet-defining inequalities

$$\langle \Gamma_i, W\rangle - \gamma_i{}^\top z \le \beta_i, \quad i = 1, \ldots, m_1, \tag{16}$$

and equalities

$$\langle \Gamma_i, W\rangle - \gamma_i{}^\top z = \beta_i, \quad i = m_1 + 1, \ldots, m,$$

where $\Gamma_i \in \mathbb{R}^{n\times n}$, $\beta_i \in \mathbb{R}$ and $\gamma_i \in \mathbb{R}^n$. Theorem 3 describes cl conv(X) in the original space of variables. Note that, in practice, a complete description of P may not be explicitly available, in which case one can use a partial description to derive valid inequalities.

Before we give the description in the original space, we define a set of feasible coefficients used to derive the inequalities. Let

$$\mathcal{Y} \overset{\text{def}}{=} \left\{ y \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m-m_1} : \sum_{i=1}^{m} \Gamma_i y_i \succeq 0,\ \sum_{i=1}^{m} \mathrm{Tr}(\Gamma_i) y_i \le 1 \right\}.$$

Theorem 3 *If $Q \succ 0$, point $(x, z, t) \in$ cl conv(X) if and only if $z \in$ conv(Z), $t \ge 0$ and*

$$t \ge \frac{x^\top \left(\sum_{i=1}^m \Gamma_i y_i\right) x}{y^\top \beta + \left(\sum_{i=1}^m y_i \gamma_i\right)^\top z}, \qquad \forall y \in \mathcal{Y}, \tag{17}$$

or equivalently,

$$t \ge \max_{y\in\mathcal{Y}} \frac{x^\top \left(\sum_{i=1}^m \Gamma_i y_i\right) x}{y^\top \beta + \left(\sum_{i=1}^m y_i \gamma_i\right)^\top z}. \tag{18}$$

Proof A point $(x, z, t) \in$ cl conv(X) if and only if

$$\begin{aligned} 0 \ge \min_{W,\lambda}\ & \lambda \\ \text{s.t. } & \langle \Gamma_i, W\rangle \le \beta_i + \gamma_i{}^\top z, \quad i = 1, \ldots, m_1 \\ & \langle \Gamma_i, W\rangle = \beta_i + \gamma_i{}^\top z, \quad i = m_1 + 1, \ldots, m \\ & W - xx^\top/t + \lambda I \succeq 0,\ \lambda \ge 0. \end{aligned}$$

Strong duality holds since there exists $(z, W) \in P$ that satisfies the facet-defining inequalities strictly, and we can always increase λ to find a strictly feasible solution to the above minimization problem. Substituting $V = W - xx^\top/t + \lambda I$, the optimization problem simplifies to

$$\begin{aligned} 0 \geq \min_{V,\lambda}\ & \lambda \\ \text{s.t.}\ & -\langle \Gamma_i, V\rangle + \lambda \mathrm{Tr}(\Gamma_i) \geq -\beta_i - \gamma_i^\top z + \langle \Gamma_i, xx^\top/t\rangle,\ i = 1, \dots, m_1 \qquad (y_i) \\ & -\langle \Gamma_i, V\rangle + \lambda \mathrm{Tr}(\Gamma_i) = -\beta_i - \gamma_i^\top z + \langle \Gamma_i, xx^\top/t\rangle,\ i = m_1 + 1, \dots, m \\ & V \succeq 0,\ \lambda \geq 0. \qquad (y_i) \end{aligned}$$

Letting $y \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m-m_1}$ denote the dual variables, we find the equivalent representation

$$0 \geq \max_{y \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m-m_1}} \sum_{i=1}^m y_i \left(-\beta_i - \gamma_i^\top z + \langle \Gamma_i, xx^\top/t\rangle\right) \tag{20a}$$

$$\text{s.t.} - \sum_{i=1}^m y_i \Gamma_i \preceq 0,\ \sum_{i=1}^m \mathrm{Tr}(\Gamma_i) y_i \leq 1. \tag{20b}$$

In particular, inequality (20a) is valid for any fixed feasible y. Multiplying both sides of the inequality by t, we find the equivalent conic quadratic representation

$$t\left(y^\top \beta + \left(\sum_{i=1}^m y_i \gamma_i\right)^\top z\right) \geq \langle \sum_{i=1}^m y_i \Gamma_i, xx^\top\rangle. \tag{21}$$

Note that validity of inequalities (21) implies that $y^\top \beta + \left(\sum_{i=1}^m y_i \gamma_i\right)^\top z \geq 0$ for any primal feasible z and dual feasible y; dividing both sides of the inequality by $y^\top \beta + \left(\sum_{i=1}^m y_i \gamma_i\right)^\top z$, the theorem is proven. □

Note that even if inequalities (16) are not facet-defining or are insufficient to describe P, the corresponding inequalities (23) are still valid for cl conv(X).

We also state the analogous result for low-rank matrices, without proof, where $(\Gamma_i, \gamma_i, \beta_i), i \in [m]$ defines P_F.

Theorem 4 *Let $Q = FF^\top$, where $F \in \mathbb{R}^{n\times k}$ is a full-column rank matrix satisfying* $\mathrm{col}(F) = \bigcap_{\hat{e}_S \in Z} \pi_S^{-1}(\mathrm{col}(F_S))$. *Then point $(x, z, t) \in$ cl conv(X) if and only if $z \in$ conv(Z), $t \geq 0$ and*

$$t \geq \frac{x^\top F\left(\sum_{i=1}^m \Gamma_i y_i\right) F^\top x}{y^\top \beta + \left(\sum_{i=1}^m y_i \gamma_i\right)^\top z}, \qquad \forall y \in \mathcal{Y}, \tag{22}$$

or equivalently,

$$t \geq \max_{y \in \mathcal{Y}} \frac{x^\top F \left(\sum_{i=1}^m \Gamma_i y_i\right) F^\top x}{y^\top \beta + \left(\sum_{i=1}^m y_i \gamma_i\right)^\top z}. \tag{23}$$

We now illustrate Theorem 3 for the set $X_{2\times 2}$ discussed in Sect. 3.1.1.

Example 2 (Description of cl conv$(X_{2\times 2})$ in the original space) From Proposition 2, we find that for $X_{2\times 2}$, a minimal description of polyhedron P is given by the bound constraints $0 \leq z \leq 1$ and

$$\left\langle \begin{pmatrix} 1 & -1/(2d_1) \\ -1/(2d_1) & 0 \end{pmatrix}, W \right\rangle - (1/d_1)z_1 = 0 \tag{y_1}$$

$$\left\langle \begin{pmatrix} 0 & -1/(2d_2) \\ -1/(2d_2) & 1 \end{pmatrix}, W \right\rangle - (1/d_2)z_2 = 0 \tag{y_2}$$

$$\left\langle \begin{pmatrix} 0 & -1/2 \\ -1/2 & 0 \end{pmatrix}, W \right\rangle \leq 0 \tag{y_3}$$

$$\left\langle \begin{pmatrix} 0 & -1/2 \\ -1/2 & 0 \end{pmatrix}, W \right\rangle + (1/\Delta)z_1 + (1/\Delta)z_2 \leq 1/\Delta \tag{y_4}$$

$$\left\langle \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, W \right\rangle - (1/\Delta)z_1 \leq 0 \tag{y_5}$$

$$\left\langle \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, W \right\rangle - (1/\Delta)z_2 \leq 0. \tag{y_6}$$

Then, an application of Theorem 3 yields the inequality

$$t \geq \max_{y \in \mathbb{R}^6_+} \frac{y_1 x_1^2 + y_2 x_2^2 + (-y_1/d_1 - y_2/d_2 - y_3 - y_4 + y_5 + y_6)x_1 x_2}{(1/\Delta)y_4 + (y_1/d_1 - y_4/\Delta + y_5/\Delta)z_1 + (y_2/d_2 - y_4/\Delta + y_6/\Delta)z_2} \tag{24a}$$

$$\text{s.t. } 4y_1 y_2 \geq (-y_1/d_1 - y_2/d_2 - y_3 - y_4 + y_5 + y_6)^2, \ \ y_1 + y_2 \leq 1. \tag{24b}$$

Note that variables y_1, y_2 are originally free as dual variables for equality constraints, however, the nonnegativity constraints are imposed due to the positive definiteness constraint in $\mathcal{Y}$. In Appendix A we provide an independent verification that inequality

(24) is indeed valid, and reduces to the quadratic inequality $t \geq d_1 x_1^2 + d_2 x_2^2 - 2x_1 x_2$ at integral z. □

From Theorem 3, we see that $\mathrm{cl\,conv}(X)$ can be described by an infinite number of fractional quadratic/affine inequalities (23). More importantly, the convex hull is finitely generated: the infinite number of quadratic and affine functions are obtained from conic combinations of a *finite* number of base matrices Γ_i and vectors (γ_i, β_i), which correspond precisely to the minimal description of P. To solve the resulting semi-infinite problem in practice, one can employ a delayed cut generation scheme, where at each iteration, the problem with a subset of inequalities (22) is solved to obtain $(\bar{x}, \bar{z})$. Then, the separation problem to find a maximum violated inequality (i.e., y) at $(\bar{t}, \bar{x}, \bar{z})$, if it exists, is a convex optimization problem given by the inner maximization problem in (23).

Example 3 (Rank-one function with constraints) Given $Z \subseteq \{0, 1\}^n$, consider the set

$$X_{R1}^{Z} = \left\{ (x, z, t) \in \mathbb{R}^n \times Z \times \mathbb{R} : t \geq \left(h^\top x \right)^2, \ x \circ (\boldsymbol{e} - z) = 0 \right\},$$

that is, a rank-one function with arbitrary constraints on the indicator variables z defined by Z. As discussed in the proof of Proposition 4, $P_F \subseteq \mathbb{R}^{n+1}$ with one additional variable $W \in \mathbb{R}^1$ which, at integer points, is given by $W = \max\{z_1, \ldots, z_n\}$. For simplicity, assume that $0 \in Z$, and that both $\mathrm{conv}(Z)$ and $\mathrm{conv}(Z \backslash \{0\})$ are full-dimensional. Finally, consider all facet-defining inequalities of $\mathrm{conv}(Z \setminus \{0\})$ of the form ${\gamma_i}^\top z \geq 1$ (that is, inequalities that cut off point 0), for $i = 1, \ldots, m$. Now consider the inequalities

$$W \leq {\gamma_i}^\top z, \qquad \forall i \in [m]. \tag{25}$$

First, observe that inequalities (25) are valid for P_F: given $z \in Z$, if $z = 0$, then $W = 0$; otherwise, $z \in Z \backslash \{0\} \implies {\gamma_i}^\top z \geq 1 = W$. Second, note that inequalities (25) are facet-defining for P_F. Indeed, given $i \in [m]$, consider the face $Z_i = \{z \in \mathrm{conv}(Z \backslash \{0\}) : {\gamma_i}^\top z = 1\}$ of $\mathrm{conv}(Z \setminus \{0\})$: since $\mathrm{conv}(Z \setminus \{0\})$ is full-dimensional and ${\gamma_i}^\top z \geq 1$ is facet-defining, there are n affinely independent points $\{z^j\}_{j=1}^n$ such that $z^j \in Z_i$. Thus, we find that points $(z^j, 1)_{j=1}^n$ and $(0, 0)$ are $(n + 1)$-affinely independent points satisfying (25) at equality. Moreover, one can easily verify that inequality $W \leq 1$ is facet-defining as well. Thus, from (23) (adapted to the factorable representation discussed in Sect. 3.2), we conclude that the inequality

$$t \geq \max_{y \in \mathbb{R}_+^{m+1}} \left\{ \frac{\left(\sum_{i=0}^m y_i\right) (h^\top x)^2}{y_0 + \sum_{i=1}^m y_i ({\gamma_i}^\top z)} \ \text{ s.t. } \sum_{i=0}^m y_i \leq 1 \right\} \tag{26}$$

is valid for $\mathrm{cl\,conv}(X_{R1}^Z)$. Moreover, an optimal solution to optimization problem (26) corresponds to setting $y_i = 1$ for $i \in \arg\min_{i \in [m]}\{{\gamma_i}^\top z\}$, and we conclude that inequalities $t \geq (h^\top x)^2$ and $t \geq (h^\top x)^2 / ({\gamma_i}^\top z)$, $i \in [m]$ are valid for $\mathrm{cl\,conv}(X_{R1}^Z)$. Indeed, as shown in [34], these inequalities along with $z \in \mathrm{conv}(Z)$ fully describe $\mathrm{cl\,conv}(X_{R1}^Z)$ (when a nondecomposability condition holds). □

Connection with decomposition methods

From Theorem 3, we see that the convex hull, X, is obtained by adding conic quadratic inequalities $t \geq \frac{x^\top(\sum_{i=1}^m \Gamma_i y_i)x}{y^\top\beta+(\sum_{i=1}^m y_i\gamma_i)^\top z}$ with simpler quadratic structure $x^\top\Gamma_i x$ (corresponding to inequalities describing P). In particular, the intuition is similar to convexifications obtained from decompositions (3). We now show how the theory presented in this paper sheds light on the strength of the aforementioned decompositions.

Suppose inequalities (16), which we repeat for convenience:

$$\langle \Gamma_i, W\rangle - {\gamma_i}^\top z \leq \beta_i, \quad i = 1, \ldots, m, \tag{27}$$

are valid for P and, additionally, $\Gamma_i \succeq 0$ for all $i \in [m]$. Since P is not full-dimensional in general, positive semidefiniteness conditions may not be as restrictive as they initially seem.

Example 4 (Description of $\text{cl conv}(X_{2\times 2})$, continued) None of the matrices in the facets of P for $\text{cl conv}(X_{2\times 2})$ given in Example 2 are positive semidefinite. Nonetheless, the inequalities below also describe P (we abuse notation and encode using variables y how each inequality is obtained):

$$\left\langle \begin{pmatrix} 1 & -1/d_1 \\ -1/d_1 & d_2/d_1 \end{pmatrix}, W \right\rangle - (1/d_1)(z_1+z_2) = 0 \qquad (y_1 + (d_2/d_1)y_2)$$

$$\left\langle \begin{pmatrix} d_1/d_2 & -1/d_2 \\ -1/d_2 & 1 \end{pmatrix}, W \right\rangle - (1/d_2)(z_1+z_2) = 0 \qquad (y_2 + (d_1/d_2)y_1)$$

$$\left\langle \begin{pmatrix} d_1/2 & -1 \\ -1 & d_2/2 \end{pmatrix}, W \right\rangle - (1/2)(z_1+z_2) \leq 0 \qquad (y_3 + (d_1/2)y_1 + (d_2/2)y_2)$$

$$\left\langle \begin{pmatrix} d_1/2 & -1 \\ -1 & d_2/2 \end{pmatrix}, W \right\rangle + (1/\Delta - 1/2)z_1 + (1/\Delta - 1/2)z_2 \leq 1/\Delta$$
$$(y_4 + (d_1/2)y_1 + (d_2/2)y_2)$$

$$\left\langle \begin{pmatrix} d_1 & 0 \\ 0 & 0 \end{pmatrix}, W \right\rangle - (d_1d_2/\Delta)z_1 \leq 0 \qquad (y_5 + d_1y_1)$$

$$\left\langle \begin{pmatrix} 0 & 0 \\ 0 & d_2 \end{pmatrix}, W \right\rangle - (d_1d_2/\Delta)z_2 \leq 0 \qquad (y_6 + d_2y_2).$$

In particular, the last two inequalities satisfy positive semidefiniteness. Moreover, the relaxation of the first two equalities obtained by replacing them with inequalities also satisfies positive semidefiniteness. Finally, if Q is sufficiently diagonally dominant and $d_1 d_2 \geq 4$, then the third and fourth inequalities satisfy positive semidefiniteness as well. □

Now suppose that in (23), we fix $y_i = \lambda/(\beta_i + {\gamma_i}^\top z)$, where λ is small enough to ensure that constraint $\sum_{i=1}^m \mathrm{Tr}(\Gamma_i) y_i \leq 1$ is satisfied. Then inequality (23) reduces to

$$mt \geq \sum_{i=1}^m \frac{x^\top \Gamma_i x}{\beta_i + {\gamma_i}^\top z},$$

which is precisely the relaxations obtained from (3). We make the following two important observations.

Observation 1 Relaxations obtained by fixing a given decomposition (3) [19, 20] are, in general, *insufficient* to describe $\mathrm{cl\,conv}(X)$. Indeed, from Theorem 3, describing $\mathrm{cl\,conv}(X)$ requires one inequality per extreme point of the region $\mathcal{Y}$, whereas a given decomposition corresponds to a single point in this region.

Observation 2 On the other hand, the strong "optimal" or "dynamic" relaxations [5, 17, 35], where the decomposition is not fixed but instead is chosen dynamically, are *excessive* to describe $\mathrm{cl\,conv}(X)$. Indeed, they are of the form (23) for every possible (rank-one, 2×2, remainder) matrix, and are not finitely generated; whereas, our results imply that the necessary inequalities are finitely generated.

We conclude this section with an analysis of rank-one decompositions, where we assume for simplicity that $Q \succ 0$: given a subset $\mathcal{T} \subseteq 2^{[n]}$, rank-one relaxations are given by

$$t \geq \sum_{T \in \mathcal{T}} \frac{(\hat{h}_T^\top x)^2}{\hat{\boldsymbol{e}}_T^\top z} + x^\top R x, \tag{28}$$

where $R = Q - \sum_{T \in \mathcal{T}} \hat{h}_T \hat{h}_T^\top \succeq 0$, and $\hat{h}_T \in \mathbb{R}^n$ are given vectors that are zero in entries not indexed by T. Relaxation (28) can be interpreted as a decomposition obtained from valid inequalities for P of the form

$$\langle \hat{h}_T \hat{h}_T^\top, W \rangle \leq \gamma \hat{\boldsymbol{e}}_T^\top z, \tag{29}$$

where $\gamma \geq 0$. Note that inequality (29) is valid for P if

$$\gamma \geq \max_{\hat{\boldsymbol{e}}_S \in Z} \frac{1}{|S \bigcap T|} \langle \hat{h}_T \hat{h}_T^\top, \hat{Q}_S^{-1} \rangle. \tag{30}$$

Proposition 5 *If* $\gamma = \max_{\hat{\boldsymbol{e}}_S \in Z} \frac{1}{|S \bigcap T|} \langle \hat{h}_T \hat{h}_T^\top, \hat{Q}_S^{-1} \rangle$, *then inequality* (29) *defines a face of P of dimension at least* $\dim(P_0) + 1$, *where*

$$P_0 = \{(z, W) \in P : z_T = 0 \text{ and } W_T = 0\}.$$

Proof There are $\dim(P_0) + 1$ affinely independent points in P_0, and all satisfy (29) at equality. Letting $S^* \in \arg\max_{\hat{\boldsymbol{e}}_S \in Z} \frac{1}{|S \bigcap T|} \langle \hat{h}_T \hat{h}_T^\top, \hat{Q}_S^{-1} \rangle$, we find that $(\hat{\boldsymbol{e}}_{S^*}, \hat{Q}_{S^*}^{-1})$ is an additional affinely independent point satisfying (29) at equality. □

Note that if optimization problem (30) has multiple optimal solutions, then one can find additional affinely independent points. In particular, (29) is guaranteed to define a high dimensional face of P if $|T|$ is small. Indeed, inequalities (29) were found to be particularly effective computationally if $\mathcal{T} = \{T \subseteq [n] : |T| \leq \kappa\}$ for some small κ [5], although a theoretical justification of this observation has been missing until now.

Remark 7 (Description of cl conv$(X_{2\times 2})$, continued) Consider again the facet-defining inequalities given in Example 4. The last two inequalities correspond to a rank-one strengthening with $|T| = 1$, which leads to relaxations of $X_{2\times 2}$ similar to the perspective relaxation. Thus, we may argue that the perspective relaxation is required to describe cl conv$(X_{2\times 2})$. □

5 A mixed-integer linear formulation for *P*

The polyhedron P can (in theory) be studied using standard methods from mixed-integer linear optimization. However, the vertex representation of P is often not convenient, as most techniques require that the polyhedron be described explicitly via linear inequalities. Thus, in this section, we present such a mixed-integer *linear* formulation for the vertices of polytope P when the Hessian matrix Q is positive definite.

First, we describe the linear equalities necessary for P. Throughout this section, for ease of exposition, for a given $S \subseteq [n]$, we permute the rows and columns of Q such that indices in S appear first.

Proposition 6 *For any* $(z, W) \in P$,

$$\sum_k Q_{ik} W_{ki} = z_i, \quad \forall i \in [n]. \tag{31}$$

Proof For any $S \subseteq [n]$, $(\hat{\boldsymbol{e}}_S, \hat{Q}_S^{-1}) \in P$, we have

$$\hat{Q}_S^{-1} Q = \begin{pmatrix} Q_S^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_S & Q_{S,[n]\setminus S} \\ Q_{S,[n]\setminus S}^\top & Q_{[n]\setminus S} \end{pmatrix} = \begin{pmatrix} I_{|S|} & Q_S^{-1} Q_{S,[n]\setminus S} \\ 0 & 0 \end{pmatrix}. \tag{32}$$

Observe that the i^{th} diagonal entry of $\hat{Q}_S^{-1} Q$ is one if $i \in S$ and zero otherwise. Since at all extreme points of P we have $z = \hat{\boldsymbol{e}}_S$ and $W = \hat{Q}_S^{-1}$ for some $S \subseteq [n]$, it follows that $(WQ)_{ii} = (\hat{Q}_S^{-1} Q)_{ii} = z_i$. □

Since P satisfies n linearly independent equalities, we immediately get insights into the dimension of P.

Corollary 1 *The dimension of P is at most $n(n+1)/2$. If $Q_{ij} \neq 0$ for all $i, j \in [n]$ and $Z = \{0,1\}^n$, then this bound is tight.*

Proof Polyhedron P has $n + n^2$ variables, but symmetry constraints $W_{ij} = W_{ji}$ and equalities (31) imply the upper bound on the dimension. If $Q_{ij} \neq 0$ for all $i, j \in [n]$, the set of points $(\hat{\boldsymbol{e}}_{\{i,j\}}, Q^{-1}_{\{i,j\}})_{i \neq j}$ and $(\hat{\boldsymbol{e}}_i, Q^{-1}_{\{i\}})_{i \in [n]}$ are $n(n+1)/2$ affinely independent points of P, because each point is the unique one satisfying $W_{ij} \neq 0$. Together with point $(0, \mathbf{0})$, where $\mathbf{0}$ represents the null matrix, we find the required $n(n+1)/2 + 1$ affinely independent points in P. □

From Corollary 1, we see that (under mild conditions) there are no other equalities in the description of P. In order to construct a mixed-integer linear formulation for the vertices of P, we will use big-M constraints. Lemmas 2 and 3 are necessary to identify valid bounds for coefficients M.

Lemma 2 *For any $S \subseteq [n]$, $Q^{-1} \succeq \hat{Q}_S^{-1}$ and $\|\hat{Q}_S^{-1}\|_{\max} \leq \lambda_{\max}(Q^{-1})$.*

Proof To prove $Q^{-1} \succeq \hat{Q}_S^{-1}$ for $S \subseteq [n]$, it suffices to show $I \succeq Q^{1/2}\hat{Q}_S^{-1}Q^{1/2}$. Since switching the order of matrix multiplication does not change the set of nonzero eigenvalues, the nonzero eigenvalues of $Q^{1/2}\hat{Q}_S^{-1}Q^{1/2}$ coincide with those of $\hat{Q}_S^{-1}Q$. From (32) one sees that $\hat{Q}_S^{-1}Q = \begin{pmatrix} I_{|S|} & Q_S^{-1}Q_{S,[n]\setminus S} \\ 0 & 0 \end{pmatrix}$ is an upper triangular matrix, which has a maximum eigenvalue of one. Then we conclude that $I \succeq Q^{1/2}\hat{Q}_S^{-1}Q^{1/2}$ and thus $Q^{-1} \succeq \hat{Q}_S^{-1}$.

For the second part, it follows that for $i \in [n]$, $(\hat{Q}_S^{-1})_{ii} \leq Q^{-1}_{ii} \leq \lambda_{\max}(Q^{-1})$. Since $\hat{Q}_S^{-1} \succeq 0$, for any $i, j \in [n]$, $(\hat{Q}_S^{-1})^2_{ij} \leq (\hat{Q}_S^{-1})_{ii}(\hat{Q}_S^{-1})_{jj}$. As $\lambda_{\max}(Q^{-1})$ gives a uniform bound on the diagonal elements of $\hat{Q}_S^{-1}$, $\lambda_{\max}(Q^{-1})$ also bounds the absolute value of the off-diagonal elements of $\hat{Q}_S^{-1}$. □

Next, we define

$$M \overset{\text{def}}{=} \lambda_{\max}(Q^{-1}) \max_{i \in [n]} \left\{ \|Q_{[n],\{i\}}\|_2 \right\} \tag{33}$$

and prove that M provides a bound for the off-diagonal elements of $\hat{Q}_S^{-1}Q$ for any $S \subseteq [n]$ in the following lemma.

Lemma 3 *For any $S \subseteq [n]$, the off-diagonals of $\hat{Q}_S^{-1}Q$ are bounded by M.*

Proof Note that $\hat{Q}_S^{-1}Q = \begin{pmatrix} I_{|S|} & Q_S^{-1}Q_{S,[n]\setminus S} \\ 0 & 0 \end{pmatrix}$. For any $j \notin S$,

$$\begin{aligned} \|Q_S^{-1}Q_{S,\{j\}}\|_\infty \leq \|Q_S^{-1}Q_{S,\{j\}}\|_2 \leq \lambda_{\max}(Q_S^{-1})\|Q_{S,\{j\}}\|_2 \\ = \lambda_{\max}(\hat{Q}_S^{-1})\|Q_{S,\{j\}}\|_2 \leq \lambda_{\max}(Q^{-1})\|Q_{[n],\{j\}}\|_2, \end{aligned}$$

where the last inequality follows from Lemma 2. □

One can make a few observations about $P = \{(\hat{\boldsymbol{e}}_S, \hat{Q}_S^{-1})\}_{\hat{\boldsymbol{e}}_S \in Z}$. Note that at extreme points of P, $W = \hat{Q}_S^{-1}$ for some S. Thus, for any extreme point $(z, W) \in P$, W_{ij} is nonzero only if $z_i = z_j = 1$. Moreover, for any $S \subseteq [n]$, $(\hat{\boldsymbol{e}}_S, \hat{Q}_S^{-1}) \in P$, $Q\hat{Q}_S^{-1} = QW = \begin{pmatrix} I_{|S|} & 0 \\ Q_{S,[n]\setminus S}^{\top} Q_S^{-1} & 0 \end{pmatrix}$, and the off-diagonal entries in the i^{th} row of QW are all zeros if $i \in S$. These two observations lead to the formulation in the following proposition.

Proposition 7 *The extreme points of P are described as*

$$\left\{(\hat{e}_S, \hat{Q}_S^{-1})_{\hat{e}_S \in Z}\right\} = \left\{(z, W) \in Z \times \mathbb{R}^{n\times n} \mid \sum_{k=1}^n Q_{ik} W_{ki} = z_i,\ \forall i \in [n], \right.$$
$$- M(1 - z_i) \le \sum_{k=1}^n Q_{ik} W_{kj} \le M(1 - z_i),\ \forall i \ne j,$$
$$\left. |W_{ij}| \le \lambda_{\max}(Q^{-1}) \min\{z_i, z_j\},\ \forall i, j \in [n]\right\}.$$

Proof For any $z = \hat{e}_S \in Z$, the constraint

$$|W_{ij}| \le \lambda_{\max}(Q^{-1}) \min\{z_i, z_j\}, \quad \forall i, j \in [n],$$

implies that $W_{ij} = 0$ if either i or j is not in S. For $i \in S$, we have

$$\sum_{k=1}^n Q_{ik} W_{ki} = 1 \tag{34}$$
$$\sum_{k=1}^n Q_{ik} W_{kj} = 0, \quad \forall j \ne i. \tag{35}$$

Inequalities (34) and (35) imply that $\left(Q_S \ \ Q_{S,[n]\setminus S}\right) \begin{pmatrix} W_S \\ W_{S,[n]\setminus S}^{\top} \end{pmatrix} = I$. Since $W_{S,[n]\setminus S} = 0$, we have $Q_S W_S = I$ and $W = \hat{Q}_S^{-1}$. Therefore, $Q\hat{Q}_S^{-1} = \begin{pmatrix} I & 0 \\ Q_{S,[n]\setminus S}^{\top} Q_S^{-1} & 0 \end{pmatrix}$. It is clear that the off-diagonal elements in the i^{th} row are all zero if $i \in S$, otherwise (if $i \notin S$) they are bounded by M according to Lemma 3. In other words, constraints

$$- M(1 - z_i) \le \sum_{k=1}^n Q_{ik} W_{kj} \le M(1 - z_i), \quad \forall j \ne i$$

hold. Moreover, thanks to Lemma 2, the constraints

$$|W_{ij}| \le \lambda_{\max}(Q^{-1}) \min\{z_i, z_j\},\ \forall i, j \in [n] \tag{36}$$

hold at $W = \hat{Q}_S^{-1}$ and $z = \hat{\boldsymbol{e}}_S$ as well. □

Proposition 7 allows us to give a mixed-integer *linear* formulation for the MIQO problem (1). Substituting the mixed-integer linear representation of P in Proposition 7 in the equivalent MIQO formulation (8), we arrive at:

$$\min_{z,W} \ -\frac{1}{2}a^\top W a + b^\top z \tag{37a}$$

$$\text{s.t.} \ \sum_{k=1}^{n} Q_{ik} W_{ki} = z_i, \quad \forall i \in [n] \tag{37b}$$

$$\text{(MILO)} \qquad -M(1-z_i) \le \sum_{k=1}^{n} Q_{ik} W_{kj} \le M(1-z_i), \quad \forall i \neq j \tag{37c}$$

$$|W_{ij}| \le \lambda_{\max}(Q^{-1}) \min\{z_i, z_j\}, \quad \forall i, j \in [n] \tag{37d}$$

$$z \in Z, \tag{37e}$$

where M is defined in (33). MILO is the first polynomial-size *explicit* mixed-integer linear formulation given for (1).

We point out that the mixed-integer representation of P in Proposition 7 relies on big-M constraints and, therefore, it is not a strong formulation. Nonetheless, advanced mixed-integer linear optimization solvers have a plethora of built-in techniques to improve such formulations. Preliminary computations using Gurobi indicate the following findings:

(1) The natural relaxation of (37) is very weak and, therefore, (37) results in worse performance than alternative (nonlinear) formulations for problem (1) in most cases.

(2) In some cases, however, and notably when the matrix Q is sparse, Gurobi improves the relaxation in presolve to the point where the problems are solved at the root node, faster than existing formulations for (1). This situation illustrates that (in some cases) existing methods can improve even weak relaxations, whereas similar improvements are not currently available for nonlinear formulations.

Detailed computational results are presented in Appendix B. Overall, the results illustrate the potential benefits of reducing convexification to describing a polyhedral set, but also indicate that much work remains to be done for deriving better relaxations of P.

6 Conclusion

In this paper, we first describe the convex hull of the epigraph of a convex quadratic function with indicators in an extended space, which is given by one semi-definite constraint, and an exponential system of linear inequalities defining the convex hull of a polytope, P (or P_F). We then derive the convex hull description in the original space as a semi-infinite conic quadratic program. Furthermore, we give a *compact* mixed-integer linear representation of the vertices of the polytope P that results in the first compact mixed-integer *linear* formulation of MIQO problems. While this is a weak formulation, our preliminary computational experience indicates that for a class of sparse problems, off-the-shelf solvers are able to take advantage of the developments in MILO to improve the formulation substantially and it is competitive if not better than state-of-the-art approaches. To translate our theoretical developments into effective practical methods, it is crucial to exploit the structure of P. In our ongoing work, we explore the case when Q is a Stieltjes matrix for which P has a nice structure that allows us to use our results directly without resorting to the MILO formulation. Our results provide a unifying framework for several convex relaxations of MIQO problems in the literature and can also be used to evaluate their strength.

Acknowledgements We thank the AE and three reviewers for their suggestions that improved the presentation. Alper Atamtürk is supported, in part, by NSF AI Institute for Advances in Optimization Award 2112533, and DOD ONR Grant 12951270. Andrés Gómez is supported, in part, by NSF Grant 2006762 and AFOSR grant FA9550-22-1-0369. Simge Küçükyavuz and Linchuan Wei are supported, in part, by NSF Grant 2007814, and DOD ONR Grants N00014-19-1-2321 and N00014-22-1-2602.

Funding Open access funding provided by SCELC, Statewide California Electronic Library Consortium

Appendix A. Validity of inequalities (24)

Here we directly check the validity of the inequalities in Example 2, which are repeated for convenience.

$$t \geq \max_{y\in\mathbb{R}^6_+} \frac{y_1x_1^2 + y_2x_2^2 + (-y_1/d_1 - y_2/d_2 - y_3 - y_4 + y_5 + y_6)x_1x_2}{(1/\Delta)y_4 + (y_1/d_1 - y_4/\Delta + y_5/\Delta)z_1 + (y_2/d_2 - y_4/\Delta + y_6/\Delta)z_2}$$
$$\text{s.t. } 4y_1y_2 \geq (-y_1/d_1 - y_2/d_2 - y_3 - y_4 + y_5 + y_6)^2,\ y_1 + y_2 \leq 1.$$

If $z_1 = z_2 = x_1 = x_2 = 0$, then the inequality reduces to $t \geq 0$. If $z_1 = 1$ and $z_2 = x_2 = 0$, the inequality reduces to

$$t \geq \max_{y\in\mathbb{R}^2_+} \frac{y_1x_1^2}{y_1/d_1 + y_5/\Delta}.$$

The inequality can be maximized by setting $y_6 = y_1/d_1$ and $y_2 = y_3 = y_4 = y_5 = 0$, and reduces to $t \geq d_1x_1^2$. The case $z_2 = 1$, $z_1 = x_1 = 0$ is identical.

Finally, if $z_1 = z_2 = 1$, then the inequality reduces to

$$t \geq \max_{y \in \mathbb{R}^6_+} \frac{y_1 x_1^2 + y_2 x_2^2 + (-y_1/d_1 - y_2/d_2 - y_3 - y_4 + y_5 + y_6) x_1 x_2}{y_1/d_1 + y_2/d_2 - y_4/\Delta + y_5/\Delta + y_6/\Delta}. \tag{38}$$

Note that we can assume, without loss of generality, that $y_3 = 0$ (otherwise, if $y_3 > 0$, one can increase y_4 and reduce y_3 to obtain a feasible solution with better objective value). Let $\bar{y} = y_4 - y_5 - y_6$. With these simplifications, (38) reduces to

$$t \geq \max \frac{y_1 x_1^2 + y_2 x_2^2 + (-y_1/d_1 - y_2/d_2 - \bar{y}) x_1 x_2}{y_1/d_1 + y_2/d_2 - \bar{y}/\Delta} \tag{39a}$$

$$\text{s.t. } 4 y_1 y_2 \geq (-y_1/d_1 - y_2/d_2 - \bar{y})^2, \; y_1 + y_2 \leq 1 \tag{39b}$$

$$y_1, \; y_2 \geq 0, \; \bar{y} \text{ free.} \tag{39c}$$

By taking the derivative of the objective with respect to $\bar{y}$, we conclude that (for fixed values of y_1 and y_2) the objective is monotone, and thus $\bar{y}$ may be assumed to be set at a bound. In particular, the rotated cone constraint holds at equality, and $\bar{y} = -y_1/d_1 - y_2/d_2 \pm 2\sqrt{y_1 y_2}$. Thus, problem (39) further reduces to

$$t \geq \Delta \max \frac{y_1 x_1^2 + y_2 x_2^2 \pm 2\sqrt{y_1 y_2} x_1 x_2}{y_1 d_2 + y_2 d_1 \pm 2\sqrt{y_1 y_2}} \tag{40a}$$

$$\text{s.t. } y_1 + y_2 \leq 1 \tag{40b}$$

$$y_1, \; y_2 \geq 0 \tag{40c}$$

Substitute $\bar{y}_1 = \pm\sqrt{y_1}$ and $\bar{y}_2 = \pm\sqrt{y_2}$. By multiplying by $(\bar{y}_1^2 d_2 + \bar{y}_2^2 d_1 + 2\bar{y}_1\bar{y}_2)/(t\Delta) \geq 0$ on both sides of the inequality, we find that (40) is satisfied if and only if $\forall \bar{y}_1, \bar{y}_2 \in \mathbb{R}$ satisfying $\bar{y}_1 + \bar{y}_2 \leq 1$, it holds

$$\left\langle \begin{pmatrix} d_2/\Delta - x_1^2/t & 1/\Delta - x_1 x_2/t \\ 1/\Delta - x_1 x_2/t & d_1/\Delta - x_2^2/t \end{pmatrix}, \begin{pmatrix} \bar{y}_1^2 & \bar{y}_1 \bar{y}_2 \\ \bar{y}_1 \bar{y}_2 & \bar{y}_2^2 \end{pmatrix} \right\rangle \geq 0,$$

which in turn holds if and only if

$$\begin{pmatrix} d_2/\Delta - x_1^2/t & 1/\Delta - x_1 x_2/t \\ 1/\Delta - x_1 x_2/t & d_1/\Delta - x_2^2/t \end{pmatrix} \succeq 0 \iff \begin{pmatrix} t & x_1 & x_2 \\ x_1 & d_2/\Delta & 1/\Delta \\ x_2 & 1/\Delta & d_1/\Delta \end{pmatrix} \succeq 0$$
$$\iff t \geq d_1 x_1^2 - 2 x_1 x_2 + d_2 x_2^2.$$

Appendix B. Numerical Experiments

Formulation MILO provides one way of utilizing Theorem 1 for general problems for which an explicit linear description of P is not available. In this section, we discuss the practical effectiveness of MILO to solve problem (1). First, in Sect. B.1, we test

MILO on best subset selection problems (2). As MILO is a weak formulation due to big-M constraints, it is often outperformed by alternative formulations to solve MIQO problems in the literature. Then, in Sect. B.2, we test the formulations on a class of graphical models which result in MIQO problems where matrix Q is sparse. It turns out advanced optimization solvers are able to substantially improve the relaxation, and MILO is competitive against the usual alternatives for this class of problems.

We compare MILO with the following alternative formulations:

Natural:

The natural reformulation, where we replace the nonconvex constraint $x_i(1-z_i) = 0$ in (1) with $|x_i| \leq 5\|x^*\|_\infty z_i$, where x^* denotes the optimal solution of the problem without binary variables or cardinality constraints. Observe that $5\|x^*\|_\infty$ is not guaranteed to be a valid bound on $|x_i|$, thus this formulation may produce suboptimal solutions for (1).

PerspS: The perspective reformulation [2, 14, 18, 22] where we extract a diagonal term $\mathbf{diag}(\delta)$ from Q with $\delta \in \mathbb{R}^n_+$ and add the perspective constraints $s_i z_i \geq x_i^2,\ \forall i \in [n]$. For numerical stability, it is common to add bounds on the variables with the perspective reformulation [28, 30, 35]. Overall, the perspective reformulation we compare with is as follows:

$$\min_{z,x,s} \quad \frac{1}{2}x^\top (Q - \mathbf{diag}(\delta))\, x + a^\top x + \frac{1}{2}\sum_{i=1}^n \delta_i s_i + b^\top z \tag{41a}$$

$$\text{s.t.} \quad s_i z_i \geq x_i^2, \quad \forall i \in [n] \tag{41b}$$

$$|x_i| \leq \lambda_{\max}(Q^{-1})\|a\|_2 z_i, \quad \forall i \in [n] \tag{41c}$$

$$z \in Z. \tag{41d}$$

The validity of the bound $\lambda_{\max}(Q^{-1})\|a\|_2$ comes from the fact that for any subset $S \subset [n]$, the unconstrained optimal solution in the reduced space equals $Q_S^{-1}a_S$ whose infinity norm is bounded by $\lambda_{\max}(Q^{-1})\|a\|_2$ by a similar argument as Lemma 3. In [10], a similar bound on the maximum absolute value of the continuous variables is proposed, and the bound we use may be seen as a relaxation that is easy to compute and works for an arbitrary Z. The vector δ is chosen as the one which gives the highest lower bound of the perspective relaxation (41) by solving the SDP model in [35] with MOSEK 10. We also tested other methods for diagonal decomposition [18, 19, 35]; but we only present the results with PerspS.

In all experiments, Z is defined by a cardinality constraint, i.e., $Z = \{z \in \{0,1\}^n \mid \sum_{i=1}^n z_i \leq r\}$, where $r = kn$ for a given sparsity parameter $0 < k \leq 1$, and $b = 0$. The mixed-integer optimization problems are solved by Gurobi 9.0 on a laptop with Intel(R) Core(TM) i7-8750H 2.20 GHz and 32 GB RAM. We set the time limit to 30 min, and we use the default values of the Gurobi parameters. When tackling mixed-integer second-order conic problems (PerspS), Gurobi automatically decides between solving the continuous relaxation of (41) or using an outer linear approximation for the rotated cone constraints (41b).

Table 3 Benchmark datasets

dataset	m	n
Housing	506	13
Diabetes	442	11
Servo	167	19
AutoMPG	392	25

B.1. Best subset selection

In this section, we solve the best subset selection problem (2) with varying k on the benchmark datasets in Table 3, available from the UCI machine learning repository.[1] The performance measures considered are solution time in seconds (for PerspS, we include the time for solving the SDP problem in the total solution time), the number of nodes explored (denoted as #node), and the initial percentage optimality gap of the continuous relaxation (denoted as %gap). We also record the optimality gaps attained at the root node after presolve for MILO, under the %gap column (in parentheses). Denoting the optimal objective value of a continuous relaxation by **LB** and the exact optimal value (or the best upper bound) by **OPT**, the initial optimality gap is calculated as % gap = $100 \times \frac{\textbf{OPT}-\textbf{LB}}{\textbf{OPT}}$. For instances that hit the time limit, we report the average end gap in parentheses.

Table 4 shows the performance of the different formulations on these benchmark datasets. We observe that the relaxation quality of MILO is poor, with optimality gaps well above 100% (in the range of $10^3 - 10^7\%$). Indeed, even though the objective of (2) has a trivial lower bound of 0, the objective values produced by the continuous relaxation of MILO are in all cases negative. The bad relaxation quality leads to large numbers of branch-and-bound nodes and solution times. However, for the special case of $k = 0.1$ on the first three datasets, Gurobi is able to close *almost all* the gap at the root node and solve the problems with little or no branching. Thus, while the results clearly indicate that at the moment—in the context of a *general* MIQO—standard methods are better than the MILO formulation, in some cases, solvers might be able to exploit the polyhedrality of MILO. In the next section, we present experiments showcasing this phenomenon.

B.2. Inference with graphical models

Given a graph $G = (V, E)$, we consider the following MIQO problem

$$\min_{z,x} \quad \sum_{i\in V} \frac{1}{\sigma^2}(y_i - x_i)^2 + \sum_{(i,j)\in E} (x_i - x_j)^2 \tag{42a}$$

$$\text{s.t.} \quad x_i(1 - z_i) = 0 \quad \forall i \in [n] \tag{42b}$$

$$\sum_{i=1}^{n} z_i \le k|V|. \tag{42c}$$

[1] https://archive.ics.uci.edu/ml/datasets.php.

Table 4 Performance of MILO, Natural and PerspS on the datasets in Table 3

dataset	k	MILO			Natural			PerspS		
		% gap	#node	time(%endgap)	% gap	#node	time	% gap	#node	time
Housing	0.1	2.15E3(0)	1	0.02	43	1	0.04	3.8	3	0.04
	0.2	5.55E3(4.0E3)	91	0.16	28.2	53	0.04	1.3	6	0.05
	0.3	9.01E3(6.8E3)	302	0.35	19.3	19	0.04	0.5	4	0.05
	0.4	1.26E4(1.20E4)	1151	0.51	11.1	20	0.04	0.1	4	0.06
	0.5	1.43E4(1.39E4)	1667	0.58	8.7	18	0.04	0.3	18	0.06
Diabetes	0.1	4.24E3(0)	1	0.02	26.5	1	0.04	11.2	8	0.04
	0.2	1.03E4(9.3E3)	66	0.29	10.8	7	0.04	3.1	6	0.04
	0.3	1.61E4(1.46E4)	222	0.31	7.2	20	0.04	3.1	16	0.05
	0.4	2.17E4(1.46E4)	424	0.27	5.1	25	0.05	3.4	29	0.05
	0.5	2.37E4(2.33E4)	662	0.37	1.8	37	0.05	1.1	17	0.05
Servo	0.1	7.32E4(0)	0	0.05	40.8	1	0.03	40.1	10	0.05
	0.2	1.91E6(1.51E4)	1541	1.1	22.5	158	0.06	22.3	45	0.06
	0.3	3E6(2.72E6)	17556	8.88	15	1107	0.11	15	762	0.16
	0.4	3.88E6(3.71E6)	40491	47.27	8	2536	0.17	7.8	2490	0.29
	0.5	4.43E6(4.37E6)	1.2E5	144.74	1.8	2103	0.15	1.7	1965	0.24
AutoMPG	0.1	7.7E6(6.16E6)	549	2.71	51.3	177	0.06	50.1	162	0.13
	0.2	2.26E7(2.26E7)	18932	72.13	28.6	1320	0.16	28.5	667	0.24
	0.3	3.51E7(3.38E7)	3.4E5	844.37	20.3	1.27E4	0.47	20.2	1329	0.35
	0.4	4.39E7(4.39E7)	4.2E6	1800(3.28E5)	9.2	1.01E5	0.48	9.2	4086	0.54
	0.5	5.76E7(5.71E7)	9.46E6	1800(7.36E5)	3.9	9669	0.34	3.5	1964	0.32

A 1800 s solution time means Gurobi hits the time limit, and we report the best optimality gap in the following parenthesis

For MILO, the gap after Gurobi's presolve is reported in the following parenthesis

Problem (42) arises in the sparse inference problem of a two-dimensional Gaussian Markov random field (GMRF), see [29] for an in-depth discussion.

The graph G we consider in our experiment is a two-dimensional 10×10 grid.

The corresponding Hessian matrix Q in problem (42) is sparse: each row has at most five nonzero entries (including the diagonal element). We use the random instances from [25], available at https://sites.google.com/usc.edu/gomez/data, where $y_i = x_i + \mathcal{N}(0, \sigma)$ is a noisy observation of x, and there are three randomly sampled 3×3 blocks of x to be nonzero. Note that σ affects both the noise level of y and the diagonal dominance of Q in (42), with small noise values σ resulting in problems with larger diagonal dominance. We test on $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5$ and sparsity levels $k = 0.1, 0.2, 0.3, 0.4, 0.5$. For each σ, we use five randomly generated instances and report the average statistics.

Table 5 summarizes the results. Similar to the experiments reported in Sect. B.1, the continuous relaxation of MILO is the worst among the three formulations, with gaps well over 100%. However, in this case, Gurobi closes virtually all optimality gap in *all* the instances, and the problems are solved very fast with at most one branch-and-bound node. The overall performance is significantly better than using the natural MIQO formulation, and very close to and in some cases (e.g., $\sigma = 0.5$, $k > 0.1$) faster than perspective reformulation for these instances.

Table 5 Performance of MILO, Natural, and PerspS formulations on graphical models

σ	k	MILO			Natural			PerspS		
		% gap	#node	Time	% gap	#node	Time(%endgap)	% gap	#node	Time
0.1	0.1	1598.83(0)	0	1.32	21.5	77.8	0.36	0.12	1	0.33
	0.2	2257.57(0)	0.2	1.44	2.5	995.2	0.37	0.02	70.8	0.54
	0.3	2404.38(0)	0.8	1.66	0.67	7.21E4	2.19	0	717.4	0.65
	0.4	2473.30(0)	0.4	1.61	0.25	1.83E7	252.27	0	1579.8	0.74
	0.5	2500.19(0)	0.8	1.77	0.16	1.5E7	490.67	0	1597.8	0.73
0.2	0.1	983.97(0)	0	1.93	16.6	239	2.02	0.14	1.4	0.44
	0.2	1512.18(0)	0.4	2.12	4.6	1.34E6	46.87	0.04	267.6	0.6
	0.3	1783.68(0)	0.8	2.53	2.48	$3.79E6^2$	1281.78(0.2)	0.04	1798.8	1.80
	0.4	1958.07(0)	1	2.53	1.34	$3.75E7^3$	1487.74(0.2)	0	1544.4	1.84
	0.5	2045.22(0)	0.8	2.51	0.69	$3.74E7^3$	1341.41(1)	0	1761.4	1.79
0.3	0.1	704.64(0)	0.4	2.53	18.98	1.72E4	13.58	0.28	24.4	0.62
	0.2	1291.39(0)	0.4	2.57	9.12	$6.02E6^1$	440.89(0.6)	0.16	400	0.77
	0.3	1740.51(0)	1	3.8	5.5	$2.98E7^4$	1555.90(1.4)	0.2	1809.4	1.91
	0.4	2062.39(0)	1	3.16	3.18	$4.17E7^5$	1715.11(1.2)	0.1	4666	14.82
	0.5	2237.68(0)	1	3.72	1.8	$4.31E7^3$	1630.78(0.6)	0.06	3272.6	3.59
0.4	0.1	547.32(0.01)	0.4	2.64	23.4	1.10E6	179.06	0.32	164.4	0.9
	0.2	1188.23(0)	0.6	3.18	14.41	$1.84E7^4$	1414.86(2.6)	0.38	768	1.72
	0.3	1766(0)	1	4.97	9.3	$2.78E7^5$	1800.36(3.8)	0.42	4711.6	3.15
	0.4	2215.48(0)	0.8	7.24	5.77	$3.46E7^5$	1800.26(2.6)	0.32	5205.2	3.85
	0.5	2492.49(0)	1	8.9	3.4	$3.90E7^5$	1800.58(1.2)	0.2	4048.8	3.84

Table 5 continued

σ	k	MILO			Natural			PerspS		
		% gap	#node	Time	% gap	#node	Time(%endgap)	% gap	#node	Time
0.5	0.1	483.62(0)	0.2	2.42	26.27	$3.94E6^1$	662.67(1.2)	0.76	433.2	2.01
	0.2	1096.05(0)	1	5.86	18.17	$1.85E7^5$	1800.36(6.8)	1.06	4733.8	7.17
	0.3	1728.40(0)	0.4	5.8	12.07	$2.50E7^5$	1800.41(5.4)	0.88	30283.6	14.15
	0.4	2261.46(0)	0.8	8.17	7.75	$3.67E7^5$	1800.61(3.8)	0.72	112304.6	30.13
	0.5	2600.52(0)	0.4	9.55	4.58	$3.60E7^5$	1800.2(1.8)	0.42	31098.4	25.07

A super script ^{i} indicates that i out of five instances hit the time limit
For instances reaching the time limit, the average best optimality gap is reported in the parenthesis following the solution time
For MILO, the gap after Gurobi's presolve is recorded in parentheses

References

1. Aktürk, M.S., Atamtürk, A., Gürel, S.: A strong conic quadratic reformulation for machine-job assignment with controllable processing times. Oper. Res. Lett. **37**, 187–191 (2009)
2. Aktürk, M.S., Atamtürk, A., Gürel, S.: Parallel machine match-up scheduling with manufacturing cost considerations. J. Sched. **13**, 95–110 (2010)
3. Albert, A.: Conditions for positive and nonnegative definiteness in terms of pseudoinverses. SIAM J. Appl. Math. **17**(2), 434–440 (1969)
4. Anstreicher, K.M., Burer, S.: Quadratic optimization with switching variables: the convex hull for $n = 2$. Math. Program. **188**, 421–441 (2021)
5. Atamtürk, A., Gómez, A.: Rank-one convexification for sparse regression. arXiv preprint arXiv:1901.10334 (2019)
6. Atamtürk, A., Gómez, A.: Supermodularity and valid inequalities for quadratic optimization with indicators. Math. Program. (2022). https://doi.org/10.1007/s10107-022-01908-2
7. Atamtürk, A., Gómez, A.: Strong formulations for quadratic optimization with M-matrices and indicator variables. Math. Program. **170**, 141–176 (2018)
8. Atamtürk, A., Gómez, A., Han, S.: Sparse and smooth signal estimation: convexification of ℓ_0-formulations. J. Mach. Learn. Res. **22**(52), 1–43 (2021)
9. Bach, F.: Submodular functions: from discrete to continuous domains. Math. Program. **175**, 419–459 (2019)
10. Bertsimas, D., King, A.: OR forum–an algorithmic approach to linear regression. Oper. Res. **64**, 2–16 (2015)
11. Bertsimas, D., Cory-Wright, R., Pauphilet, J.: Mixed-projection conic optimization: a new paradigm for modeling rank constraints. Oper. Res. **70**(6), 3321–3344 (2022)
12. Bien, J., Taylor, J., Tibshirani, R.: A lasso for hierarchical interactions. Ann. Stat. **41**(3), 1111 (2013)
13. Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. Math. Program. **74**(2), 121–140 (1996)
14. Ceria, S., Soares, J.: Convex programming for disjunctive convex optimization. Math. Program. **86**, 595–614 (1999)
15. Chen, X., Ge, D., Wang, Z., Ye, Y.: Complexity of unconstrained $l_2 - l_p$ minimization. Math. Program. **143**(1), 371–383 (2014)
16. Cozad, A., Sahinidis, N.V., Miller, D.C.: Learning surrogate models for simulation-based optimization. AIChE J. **60**(6), 2211–2227 (2014)
17. Dong, H., Chen, K., Linderoth, J.: Regularization vs. relaxation: a conic optimization perspective of statistical variable selection. arXiv preprint arXiv:1510.06083 (2015)
18. Frangioni, A., Gentile, C.: Perspective cuts for a class of convex 0–1 mixed integer programs. Math. Program. **106**, 225–236 (2006)
19. Frangioni, A., Gentile, C.: SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. Oper. Res. Lett. **35**, 181–185 (2007)
20. Frangioni, A., Gentile, C., Hungerford, J.: Decompositions of semidefinite matrices and the perspective reformulation of nonseparable quadratic programs. Math. Oper. Res. **45**(1), 15–33 (2020)
21. Gao, J., Li, D.: Cardinality constrained linear-quadratic optimal control. IEEE Trans. Autom. Control **56**, 1936–1941 (2011)
22. Günlük, O., Linderoth, J.: Perspective reformulations of mixed integer nonlinear programs with indicator variables. Math. Program. **124**, 183–205 (2010)
23. Han, S., Gómez, A.: Compact extended formulations for low-rank functions with indicator variables. arXiv preprint arXiv:2110.14884 (2021)
24. Han, S., Gómez, A., Atamtürk, A.: 2x2 convexifications for convex quadratic optimization with indicator variables. Math. Program. (2023). https://doi.org/10.1007/s10107-023-01924-w. (**Online First Article**)
25. He, Z., Han, S., Gómez, A., Cui, Y., Pang, J.-S.: Comparing solution paths of sparse quadratic minimization with a stieltjes matrix. Math. Program. (2023). https://doi.org/10.1007/s10107-023-01966-0
26. Hochbaum, D.S.: An efficient algorithm for image segmentation, Markov random fields and related problems. J. ACM (JACM) **48**(4), 686–701 (2001)
27. Jeon, H., Linderoth, J., Miller, A.: Quadratic cone cutting surfaces for quadratic programs with on-off constraints. Discret. Optim. **24**, 32–50 (2017)

28. Küçükyavuz, S., Shojaie, A., Manzour, H., Wei, L. , Wu, H.-H.: Consistent second-order conic integer programming for learning Bayesian networks. arXiv preprint arXiv:2005.14346 (2020)
29. Liu, P., Fattahi, S., Gómez, A., Küçükyavuz, S.: A graph-based decomposition method for convex quadratic optimization with indicators. Math. Program. (2022). https://doi.org/10.1007/s10107-022-01845-0. (**Article in Advance**)
30. Manzour, H., Küçükyavuz, S., Wu, H.-H., Shojaie, A.: Integer programming for learning directed acyclic graphs from continuous data. INFORMS J. Optim. **3**(1), 46–73 (2021)
31. Penrose, R.: A generalized inverse for matrices. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol 51, pp. 406–413. Cambridge University Press (1955)
32. POlyhedron Representation Transformation Algorithm. https://porta.zib.de/#download. Accessed: 2021-11-20
33. Wei, L., Gómez, A., Küçükyavuz, S.: On the convexification of constrained quadratic optimization problems with indicator variables. In: International conference on integer programming and combinatorial optimization, pp. 433–447. Springer (2020)
34. Wei, L., Gómez, A., Küçükyavuz, S.: Ideal formulations for constrained convex optimization problems with indicator variables. Math. Program. **192**(1–2), 57–88 (2022)
35. Zheng, X., Sun, X., Li, D.: Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach. INFORMS J. Comput. **26**, 690–703 (2014)