On Integrality in Semidefinite Programming for Discrete Optimization

Meijer, Frank de; Sotirov, Renata

DOI
10.1137/23M1580905

Publication date
2023

Document Version
Final published version

Published in
SIAM Journal on Optimization

SIAM J. OPTIM.
Vol. 34, No. 1, pp. 1071–1096

ON INTEGRALITY IN SEMIDEFINITE PROGRAMMING FOR DISCRETE OPTIMIZATION*

FRANK DE MEIJER† AND RENATA SOTIROV‡

Abstract. It is well known that by adding integrality constraints to the semidefinite programming (SDP) relaxation of the max-cut problem, the resulting integer semidefinite program is an exact formulation of the problem. In this paper we show similar results for a wide variety of discrete optimization problems for which SDP relaxations have been derived. Based on a comprehensive study on discrete positive semidefinite matrices, we introduce a generic approach to derive mixed-integer SDP (MISDP) formulations of binary quadratically constrained quadratic programs and binary quadratic matrix programs. Applying a problem-specific approach, we derive more compact MISDP formulations of several problems, such as the quadratic assignment problem, the graph partition problem, and the integer matrix completion problem. We also show that several structured problems allow for novel compact MISDP formulations through the notion of association schemes. Complementary to the recent advances on algorithmic aspects related to MISDP, this work opens new perspectives on solution approaches for the here considered problems.

Key words. mixed-integer semidefinite programming, discrete positive semidefinite matrices, binary quadratic programming, quadratic matrix programming, association schemes

MSC codes. 90C11, 90C22, 90C27

DOI. 10.1137/23M1580905

1. Introduction.

Semidefinite programming deals with the optimization of a linear function over the cone of positive semidefinite matrices under the presence of affine constraints. Over the last decades, semidefinite programs (SDPs) have proven themselves particularly useful in providing tight relaxations of discrete optimization problems [39, 52, 55]. Following the extension from linear programming to integer linear programming initiated in the 1960s, a recent interest in incorporating integer variables into SDPs has arisen. When the variables in an SDP are required to be integer, we refer to the problem as an integer SDP (ISDP). When an SDP contains both integer and continuous variables, we refer to the program as a mixed-integer SDP (MISDP).

The combination of positive semidefiniteness and integrality induces a lot of structure in matrices. Exploiting this fact, it has been shown that several classical discrete optimization problems allow for a formulation as an (M)ISDP [2, 13, 19, 38, 44, 45]. Recently, programs including positive semidefinite matrix variables and integrality constraints have also been used to model more applied problems; see [10, 17, 21, 24, 36, 42, 50, 56, 57, 58]. Despite the literature on these particular problems, a generic approach for deriving MISDPs has not been applied.

In this paper we advocate that MISDPs are suitable as a general modeling technique for many optimization problems. We particularly focus on binary quadratic

*Received by the editors June 20, 2023; accepted for publication (in revised form) November 15, 2023; published electronically March 15, 2024.
https://doi.org/10.1137/23M1580905
†Delft Institute of Applied Mathematics, Delft University of Technology, Delft, 2628 CD, The Netherlands (f.j.j.demeijer@tudelft.nl).
‡CentER, Department of Econometrics and OR, Tilburg University, Tilburg, 5037 AB, The Netherlands (r.sotirov@uvt.nl).

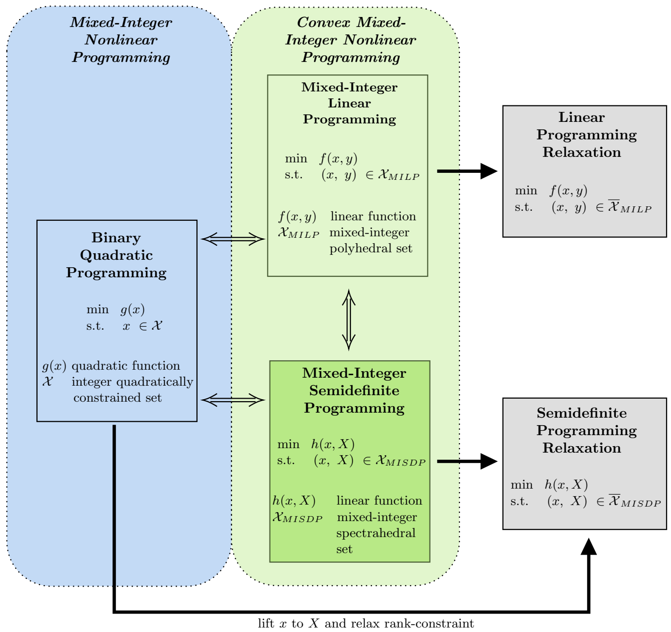

FIG. 1. *Overview of various exact formulations of BQPs and their relaxations. A double arrow ($\Longleftrightarrow$) denotes equivalence between the formulations, while a solid arrow ($\longrightarrow$) denotes that the formulation is relaxed from the former to the latter. The sets $\mathcal{X}$, $\mathcal{X}_{MILP}$, and $\mathcal{X}_{MISDP}$ are defined by nonconvex integer constraints, while $\overline{\mathcal{X}}_{MILP}$ and $\overline{\mathcal{X}}_{MISDP}$ are convex relaxations. Note: color appears only in the online article.*

programs (BQPs); see Figure 1. A common approach to solve a BQP is by exploiting linearization techniques to model it as an MILP, which is solved in a branch-and-bound framework. This research line is depicted in the top stream of Figure 1. An alternative approach is to apply a lifting of the variables to model the problem as an SDP with a nonconvex rank constraint. After relaxing the rank constraint, one obtains an SDP relaxation of the problem. This approach corresponds to the bottom arrow in Figure 1. It is generally disregarded that this relaxation can also be obtained via relaxing integrality in an MISDP model that is equivalent to the BQP. Realizing this fact provides a systematic way of approaching BQPs via MISDPs. The MISDP formulation has the advantages to have both a linear objective and a convex relaxation that is often stronger than linear programming relaxations.

The focus of this paper is primarily on the modeling aspect of discrete optimization problems as (M)ISDPs. There exist several general-purpose solution approaches for (M)ISDPs such as branch-and-bound methods [22, 43, 31] and branch-and-cut methods [29, 34, 44]. The computational ingredients of the above-mentioned approaches combined with the theoretical framework of modeling problems as (M)ISDPs that we derive in this paper, provide a complementary foundation of mixed-integer semidefinite programming in discrete optimization.

Main results and outline. This paper studies the theoretical role of MISDP in discrete optimization. We show that many problems can be modeled as an (M)ISDP, either by a generic approach for certain large problem classes, or by a more problem-specific approach. Our approach is accompanied with a large number of examples of various discrete optimization problems.

In section 2, we provide various theoretical results on positive semidefinite (PSD) $\{0,1\}$-, $\{\pm 1\}$-, and $\{0,\pm 1\}$-matrices. For instance we derive an integer semidefinite characterization of PSD $\{0,1\}$-matrices of rank (at most) r. We also present a combinatorial, polyhedral, set-completely positive, and integer hull description of the set of PSD $\{0,1\}$-matrices bounded by a certain rank, among other results.

These matrix theoretical results are exploited in section 3, when proving that many BQPs allow for a formulation as a binary SDP (BSDP). We establish this result for binary quadratically constrained quadratic programs and, in particular, for binary quadratic matrix programs. Problems that allow for a formulation as a binary quadratic matrix program, e.g., quadratic clustering or packing problems, can be modeled as a compact BSDP with a PSD matrix variable of relatively low order.

In section 4, we treat several specific problem classes for which we obtain MISDP formulations that do not follow from the above-mentioned framework or for which it is possible to obtain a more compact formulation. Among these are the quadratic assignment problem, several graph partition problems, and graph problems that can be modeled based on association schemes. Also, as most formulations that we discuss include binary variables, we present two problems that have an MISDP formulation where the variables are nonbinary.

Notation. We denote by $\mathbf{0}_n \in \mathbb{R}^n$ the vector of all zeros, and by $\mathbf{1}_n \in \mathbb{R}^n$ the vector of all ones. The identity matrix and the matrix of ones of order n are denoted by $\mathbf{I}_n$ and $\mathbf{J}_n$, respectively. We omit the subscripts of these matrices when there is no confusion about the order. The ith unit vector is denoted by $\mathbf{e}_i$ and we define $\mathbf{E}_{ij} := \mathbf{e}_i \mathbf{e}_j^\top$. The set of $n \times n$ permutation matrices is denoted by Π_n. For $n \in \mathbb{Z}_+$, we define the set $[n] := \{1, \ldots, n\}$. For any $S \subseteq [n]$, we let $\mathbb{1}_\mathbf{S}$ be the binary indicator vector of S. The support of $x \in \mathbb{R}^n$ is denoted by $\operatorname{supp}(x)$.

We denote the set of all $n \times n$ real symmetric matrices by $\mathcal{S}^n$. The cone of symmetric PSD matrices is defined as $\mathcal{S}^n_+ := \{X \in \mathcal{S}^n : X \succeq \mathbf{0}\}$, where $X \succeq \mathbf{0}$ denotes that X is PSD. The trace of a square matrix $X = (X_{ij})$ is given by $\operatorname{tr}(X) = \sum_i X_{ii}$. For any $X, Y \in \mathbb{R}^{n \times m}$ the trace inner product is defined as $\langle X, Y \rangle := \operatorname{tr}(X^\top Y) = \sum_{i=1}^n \sum_{j=1}^m X_{ij} Y_{ij}$.

The operator $\operatorname{diag} : \mathbb{R}^{n \times n} \to \mathbb{R}^n$ maps a square matrix to a vector consisting of its diagonal elements. We denote by $\operatorname{Diag} : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ its adjoint operator. The direct sum of matrices X and Y is defined as $X \oplus Y = \left(\begin{smallmatrix} X & \mathbf{0} \\ \mathbf{0} & Y \end{smallmatrix}\right)$. The Kronecker product $X \otimes Y$ of matrices $X \in \mathbb{R}^{p \times q}$ and $Y \in \mathbb{R}^{r \times s}$ is defined as the $pr \times qs$ matrix composed of pq blocks of size $r \times s$ with block ij given by $x_{ij} Y$, $i \in [p]$, $j \in [q]$.

2. Theory on discrete PSD matrices. Most discrete optimization problems that we consider in this paper are defined using binary variables, i.e., variables taking values in $\{0,1\}$ or $\{\pm 1\}$, or ternary variables, i.e., variables whose values are in $\{0, \pm 1\}$. In this section we derive several useful results on these matrix sets with respect to positive semidefiniteness. We start by considering the PSD $\{0,1\}$-matrices, after which we extend these results to PSD $\{\pm 1\}$- and $\{0, \pm 1\}$-matrices.

2.1. Theory on PSD $\{0,1\}$-matrices. In this section we consider the set of PSD $\{0,1\}$-matrices. We derive and recall several formulations of this matrix set, including a combinatorial, polyhedral and a set-completely positive description. We start this section with two known results, i.e., Theorem 1 and Proposition 1, after which we present a series of new results.

PSD $\{0,1\}$-matrices are studied explicitly by Letchford and Sørensen [40]. They derive the following decomposition result on PSD $\{0,1\}$-matrices.

THEOREM 1 (see [40]). *Let $X \in \{0,1\}^{n\times n}$ be a symmetric matrix. Then $X \succeq \mathbf{0}$ if and only if $X = \sum_{j=1}^{r} x_j x_j^\top$ for some $x_j \in \{0,1\}^n$, $j \in [r]$.*

Given $S \subseteq \mathbb{R}_+$, a matrix X is called S-completely positive if $X = PP^\top$ for some $P \in S^{n\times k}$. In case $S = \mathbb{R}_+$, we call X completely positive. It follows from Theorem 1 that any PSD $\{0,1\}$-matrix is $\{0,1\}$-completely positive.

The decomposition of PSD $\{0,1\}$-matrices gives rise to a useful combinatorial interpretation on the complete graph K_n. Viewing each vector $x_j \in \{0,1\}^n$ as an indicator vector on the vertices of K_n, the matrix $x_j x_j^\top$ can be seen as the characteristic matrix of a clique in K_n. Given a decomposition $X = \sum_{j=1}^{k} x_j x_j^\top$, the cliques indexed by $j \in [k]$ are pairwise disjoint, since the diagonal of X is at most one. Therefore, each PSD $\{0,1\}$-matrix is the characteristic matrix of a set of pairwise disjoint cliques in K_n. This combinatorial structure is in the literature also known as a clique packing.

As we will see in the next section, many SDP formulations arise from a lifting $PP^\top$, where P is an appropriate $n\times r$ $\{0,1\}$–matrix. Consequently, the resulting PSD $\{0,1\}$-matrix has rank at most r. From that perspective, it makes sense to consider the set of PSD $\{0,1\}$-matrices that have an upper bound on the rank. For positive integers r, n with $r \le n$, let us define the discrete set

$$\mathcal{D}_r^n := \left\{ X \in \{0,1\}^{n\times n} \,:\, X \succeq \mathbf{0},\, \operatorname{rank}(X) \le r \right\}. \tag{1}$$

Theorem 1 induces the following $\{0,1\}$-completely positive description of $\mathcal{D}_r^n$:

$$\mathcal{D}_r^n = \left\{ PP^\top \,:\, P \in \{0,1\}^{n\times r},\, P\mathbf{1}_r \le \mathbf{1}_n \right\}. \tag{2}$$

Next, we will derive another formulation of $\mathcal{D}_r^n$, where the constraint $\operatorname{rank}(X) \le r$ is established by an appropriate linear matrix inequality. To that end, we exploit the following result that is implicitly proved in many sources and explicitly by Dukanovic and Rendl [18].

PROPOSITION 1 (see [18]). *Let $X \in \{0,1\}^{n\times n}$ be symmetric. Then, the following are equivalent:*

(i) $\operatorname{diag}(X) = \mathbf{1}_n$, $\operatorname{rank}(X) = r$, $X \succeq \mathbf{0}$.
(ii) *There exists a permutation matrix Q such that $QXQ^\top = \mathbf{J}_{n_1} \oplus \cdots \oplus \mathbf{J}_{n_r}$ with $n = n_1 + \cdots + n_r$.*
(iii) $\operatorname{diag}(X) = \mathbf{1}_n$, $\operatorname{rank}(X) = r$, *and X satisfies the triangle inequalities $X_{ij} + X_{ik} - X_{jk} \le 1$ for all $(i,j,k) \in [n]\times[n]\times[n]$.*
(iv) $\operatorname{diag}(X) = \mathbf{1}_n$ *and* $(tX - \mathbf{J} \succeq \mathbf{0} \Longleftrightarrow t \ge r)$.

Proposition 1 establishes the equivalence between useful characterizations of rank-r PSD $\{0,1\}$-matrices that have ones on the diagonal. In the following corollary we generalize this result by relaxing the condition $\operatorname{diag}(X) = \mathbf{1}_n$. The proof is similar to the proof of Proposition 1.

COROLLARY 1. *Let $X \in \{0,1\}^{n\times n}$ be symmetric. Then, the following statements are equivalent:*

(i) $\operatorname{rank}(X) = r$, $X \succeq \mathbf{0}$.
(ii) *There exists a permutation matrix Q such that $QXQ^\top = \mathbf{J}_{n_1} \oplus \cdots \oplus \mathbf{J}_{n_r} \oplus \mathbf{0}_{n_z\times n_z}$ with $n = n_1 + \cdots + n_r + n_z$.*
(iii) $\operatorname{rank}(X) = r$ *and X satisfies the triangle inequalities $X_{ij} \le X_{ii}$ for all $i \ne j$ and $X_{ij} + X_{ik} - X_{jk} \le X_{ii}$ for all $j < k, i \ne j, k$.*
(iv) $tX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$ *if and only if $t \ge r$.*

Proof. Throughout the proof, let $N_1 := \{i \in [n] : X_{ii} = 1\}$ and let $N_0 := [n] \setminus N_1$. Moreover, let Q' denote a permutation matrix corresponding to a permutation of $[n]$ that maps the ordered set $(1, 2, \ldots, n)$ to an ordered set where the elements in N_1 occupy the first N_1 positions.

(i) $\iff$ (ii) : Let X be PSD with $\operatorname{rank}(X) = r$. Then, the rows and columns indexed by N_0 only contain zeros. As a consequence, $Q'X(Q')^\top$ is of the form $Y \oplus \mathbf{0}_{|N_0|\times|N_0|}$ with $\operatorname{diag}(Y) = \mathbf{1}_{|N_1|}$ and $Y \succeq \mathbf{0}$. By Proposition 1 there exists a permutation matrix $\bar{Q}$ such that $\bar{Q}Y\bar{Q}^\top = \mathbf{J}_{n_1} \oplus \cdots \oplus \mathbf{J}_{n_r}$ with $|N_1| = n_1 + \cdots + n_r$. Let $Q := (\bar{Q} \oplus \mathbf{I}_{|N_0|})Q'$, then $QXQ^\top = \mathbf{J}_{n_1} \oplus \cdots \oplus \mathbf{J}_{n_r} \oplus \mathbf{0}_{|N_0|\times|N_0|}$.

Conversely, suppose that $QXQ^\top = \mathbf{J}_{n_1} \oplus \cdots \oplus \mathbf{J}_{n_r} \oplus \mathbf{0}_{n_z\times n_z}$ with $n = n_1 + \cdots + n_r + n_z$ for some permutation matrix Q. Then, obviously, $QXQ^\top$ is PSD with $\operatorname{rank}(QXQ^\top) = r$, from which it follows that $X \succeq \mathbf{0}$ with $\operatorname{rank}(X) = r$.

(i) $\iff$ (iii) : For $n = 2$, the inequalities $X_{ij} \leq X_{ii}$, $i \neq j$, are trivially necessary and sufficient for $X \succeq \mathbf{0}$. For $n \geq 3$, the result follows from Letchford and Sørensen [40, Proposition 3].

(i) $\iff$ (iv) : Suppose X is PSD with $\operatorname{rank}(X) = r$. Then, $Q'X(Q')^\top = Y \oplus \mathbf{0}_{|N_0|\times|N_0|}$ with $\operatorname{diag}(Y) = \mathbf{1}_{|N_1|}$ and $Y \succeq \mathbf{0}$. This leads to the following sequence of equivalences:

$$
\begin{aligned}
tX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0} &\iff tQ'X(Q')^\top - Q'\operatorname{diag}(X)\operatorname{diag}(X)^\top(Q')^\top \succeq \mathbf{0} \\
&\iff t\left(Y \oplus \mathbf{0}_{|N_0|\times|N_0|}\right) - \begin{pmatrix} \mathbf{1}_{|N_1|} \\ \mathbf{0}_{|N_0|} \end{pmatrix} \begin{pmatrix} \mathbf{1}_{|N_1|} \\ \mathbf{0}_{|N_0|} \end{pmatrix}^\top \succeq \mathbf{0}
\end{aligned}
$$

which holds if and only if $t \geq r$, by statement (iv) of Proposition 1.

Conversely, suppose that $tX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$ if and only if $t \geq r$. If $r = 0$, then $t = 0$ induces $\operatorname{diag}(X) = \mathbf{0}_n$, while $t = 1$ implies $X - \operatorname{diag}(X)\operatorname{diag}(X)^\top = X \succeq \mathbf{0}$, since $\operatorname{diag}(X) = \mathbf{0}_n$. Hence, X must be the zero matrix, which is PSD with rank zero. Now, assume that $r \geq 1$. Then, $rX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$ can be written as $X \succeq \frac{1}{r}\operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$. Let $r^* := \operatorname{rank}(X)$. It follows from the previously proven implication, (i) $\implies$ (iv), that $r^* = \min\{t : tX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}\}$. By assumption, this value equals r, so $r^* = r$. We conclude that X is PSD with $\operatorname{rank}(X) = r$. ◻

Corollary 1 can be exploited to prove the following result.

COROLLARY 2. *Let $X \in \{0,1\}^{n\times n}$ be symmetric. If $Y = \begin{pmatrix} r & \operatorname{diag}(X)^\top \\ \operatorname{diag}(X) & X \end{pmatrix} \succeq \mathbf{0}$, then $X \succeq \mathbf{0}$ with $\operatorname{rank}(X) \leq r$.*

Proof. The assertion $X \succeq \mathbf{0}$ is trivial, so it suffices to show that $Y \succeq \mathbf{0}$ implies $\operatorname{rank}(X) \leq r$. If $r = 0$, then $\operatorname{diag}(X) = \mathbf{0}_n$. Since $X \succeq \mathbf{0}$, X must be the zero matrix and, thus, $\operatorname{rank}(X) = 0$.

Now, let $r \geq 1$. The Schur complement lemma implies that $rX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$. Let $r^* := \min\left\{t : tX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}\right\} \leq r$. Since $r^*X - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$ and $X \succeq \mathbf{0}$, it follows that $tX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$ for all $t \geq r^*$. Therefore, $tX - \operatorname{diag}(X)\operatorname{diag}(X)^\top \succeq \mathbf{0}$ if and only if $t \geq r^*$. Corollary 1 then implies $\operatorname{rank}(X) = r^* \leq r$. ◻

Corollary 2 implies the following characterization of $\mathcal{D}_r^n$, where the rank constraint is merged into a lifted linear matrix inequality:

$$
\mathcal{D}_r^n = \left\{ X \in \{0,1\}^{n\times n} : \begin{pmatrix} r & \operatorname{diag}(X)^\top \\ \operatorname{diag}(X) & X \end{pmatrix} \succeq \mathbf{0} \right\}. \tag{3}
$$

For some optimization problems, the upper bound constraint on the rank of X is not sufficient, as we require that X is exactly of rank r. The max k-cut problem, for instance, requires partitioning the vertex set of a graph into exactly k nonempty and pairwise disjoint subsets. The following two results show that the description given in (3) can be extended to also include a lower bound on the rank of X.

THEOREM 2. *Let $X \in \{0,1\}^{n\times n}$ be symmetric. If there exists a matrix $P \in \{0,1\}^{n\times r}$ with $P^\top \mathbf{1} \geq \mathbf{1}$ such that $Y = \begin{pmatrix} \mathbf{I}_r & P^\top \\ P & X \end{pmatrix} \succeq \mathbf{0}$, then $X \succeq \mathbf{0}$ with* $\operatorname{rank}(X) \geq r$.

Proof. The assertion $X \succeq \mathbf{0}$ is trivial. It suffices to show that $\operatorname{rank}(X) \geq r$. As $Y \succeq \mathbf{0}$ and Y has binary entries, it follows from Theorem 1 that $Y = \sum_{j=1}^k \binom{u_j}{x_j}\binom{u_j}{x_j}^\top$ for some $u_j \in \{0,1\}^r$ and $x_j \in \{0,1\}^n$, $j \in [k]$. Since $\sum_{j=1}^k u_j u_j^\top = \mathbf{I}_r$, we must have $k \geq r$. Moreover, the set $\{u_j : j \in [k]\}$ must contain $\mathbf{e}_1, \ldots, \mathbf{e}_r$ and $k - r$ copies of $\mathbf{0}_r$. Without loss of generality, let us assume that the first r vectors in $\{u_j : j \in [k]\}$ correspond to the elementary vectors. Then, it follows that $P = \sum_{j=1}^k x_j u_j^\top = \sum_{j=1}^r x_j \mathbf{e}_j^\top = [x_1 \ldots x_r]$. Since $P^\top \mathbf{1} \geq \mathbf{1}$, it follows that the vectors x_j, $j \in [r]$, cannot be the zero vector. Since these are moreover linearly independent, we have $\operatorname{rank}(X) \geq \operatorname{rank}(\sum_{j=1}^r x_j x_j^\top) = r$. ∎

Theorem 2 and Corollary 2 together impose the following integer semidefinite characterization of PSD $\{0,1\}$-matrices of rank r.

COROLLARY 3. *Let $X \in \{0,1\}^{n\times n}$ be symmetric. If there exists a matrix $P \in \{0,1\}^{n\times r}$ with $P^\top \mathbf{1} \geq \mathbf{1}$, $P\mathbf{1} = \operatorname{diag}(X)$, such that $Y = \begin{pmatrix} \mathbf{I}_r & P^\top \\ P & X \end{pmatrix} \succeq \mathbf{0}$, then $X \succeq \mathbf{0}$ with* $\operatorname{rank}(X) = r$.

Proof. It immediately follows from Theorem 2 that $X \succeq \mathbf{0}$ with $\operatorname{rank}(X) \geq r$. Moreover, since $Y \succeq \mathbf{0}$, we also know that

$$\left(\mathbf{1}_r^\top \oplus \mathbf{I}_n\right) Y \left(\mathbf{1}_r^\top \oplus \mathbf{I}_n\right)^\top = \begin{pmatrix} \mathbf{1}_r^\top \mathbf{I}_r \mathbf{1}_r & \mathbf{1}_r^\top P^\top \mathbf{I}_n \\ \mathbf{I}_n P \mathbf{1}_r & \mathbf{I}_n X \mathbf{I}_n \end{pmatrix} = \begin{pmatrix} r & \operatorname{diag}(X)^\top \\ \operatorname{diag}(X) & X \end{pmatrix} \succeq \mathbf{0}.$$

It then follows from Corollary 2 that $\operatorname{rank}(X) \leq r$. ∎

The integer semidefinite characterization of $\mathcal{D}_r^n$ given in (3) shows that if a $\{0,1\}$-matrix satisfies a certain linear matrix inequality, then a rank condition is implied. For the case of rank-one matrices, we can show that the converse implication does also hold, i.e., if a rank-one matrix satisfies a certain linear matrix inequality, then its entries must be in $\{0,1\}$.

THEOREM 3. *Let $Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq \mathbf{0}$ with $\operatorname{diag}(X) = x$. Then, $\operatorname{rank}(Y) = 1$ if and only if $X \in \{0,1\}^{n\times n}$.*

Proof. $(\Longrightarrow)$: If $\operatorname{rank}(Y) = 1$, then $Y = \bar{x}\bar{x}^\top$ with $\bar{x} = [1 \ x^\top]^\top \in \mathbb{R}^{n+1}$ and $X = xx^\top$. From the positive semidefiniteness of order two principal submatrices of Y we obtain $\mathbf{0}_n \leq x \leq \mathbf{1}_n$. Since $\operatorname{diag}(xx^\top) = x$, we have $x_i^2 = x_i$ for all $i \in [n]$, so $x \in \{0,1\}^n$. We conclude that $X = xx^\top \in \{0,1\}^{n\times n}$.

$(\Longleftarrow)$: Since $X \in \{0,1\}^{n\times n}$ and $x = \operatorname{diag}(X)$, it follows that $Y \in \{0,1\}^{(n+1)\times(n+1)}$. From Theorem 1 it follows that $Y = \sum_{j=1}^k x_j x_j^\top$ for some $x_j \in \{0,1\}^{n+1}$, $j \in [k]$, i.e., Y can be decomposed in terms of cliques. Since $Y_{11} = 1$ and $\operatorname{diag}(Y) = (1, x^\top)^\top$, all indices $i \in [n+1]$ for which $Y_{ii} = 1$ must be in the same clique as the first index. Hence, the decomposition consists of only one clique and $\operatorname{rank}(Y) = 1$. ∎

Theorem 3 plays a central role in deriving integer SDP formulations of BQPS defined over vectors of variables in section 3.1. However, Theorem 3 cannot be extended to matrices with a rank larger than one. That is, if Y is a PSD matrix satisfying $\mathrm{diag}(Y) = Y\mathbf{e}_1$ and $Y_{11} = r$, then integrality of Y is not equivalent to $\mathrm{rank}(Y) = r$. For example, the matrix $Y = \frac{1}{2}(\mathbf{J}_3 + 3\mathbf{E}_{11})$ satisfies $Y \succeq \mathbf{0}$, $\mathrm{diag}(Y) = Y\mathbf{e}_1$ and $\mathrm{rank}(Y) = 2$, but Y is not integer.

The characterizations given in (2) and (3) rely on conditions involving discreteness. Let us now move on to continuous descriptions. Of course, since $\mathcal{D}_r^n$ is itself a discrete set, a continuous description does not aim at describing $\mathcal{D}_r^n$, but rather its convex hull, i.e.,

$$\mathcal{P}_r^n := \mathrm{conv}(\mathcal{D}_r^n). \tag{4}$$

Observe that although the matrices in $\mathcal{D}_r^n$ have an upper bound on the rank, the polytopes $\mathcal{P}_r^n$ are full dimensional, since $\frac{1}{n}\mathbf{I}_n \in \mathcal{P}_r^n$ for all $1 \leq r \leq n$. In order to gain more insight into the structure of $\mathcal{P}_r^n$, we introduce the notion of a so-called packing family.

Definition 1. *Let T be a finite set of elements. A collection $\mathcal{F}$ of nonempty subsets of T is called a packing of T if the subsets in $\mathcal{F}$ are pairwise disjoint. The family of all packings of T is called the packing family of T, denoted by $\mathbf{F}(T)$.*

Observe that $\mathcal{F} = \emptyset$ also belongs to $\mathbf{F}(T)$. Next, we define the notion of an r-packing of T.

Definition 2. *Let T be a finite set of elements. A packing $\mathcal{F}$ of T is called an r-packing of T if $|\mathcal{F}| \leq r$. The family of all r-packings of T is called the r-packing family of T, denoted by $\mathbf{F}_r(T)$.*

The r-packing family of $[n]$ can be exploited to describe $\mathcal{P}_r^n$. Let $X \in \mathcal{D}_r^n$. By Theorem 1 we know that X is the sum of at most r rank-one PSD $\{0,1\}$-matrices. From a combinatorial point of view, this implies that X corresponds to an r-packing of $[n]$. In fact, there is a bijection between the matrices in $\mathcal{D}_r^n$ and the r-packings in $\mathbf{F}_r([n])$. For any r-packing $\mathcal{F}$, let $\mathbf{E}_{\mathcal{F}} := \sum_{S \in \mathcal{F}} \mathbb{1}_S \mathbb{1}_S^\top$. Then, we obtain the following polyhedral description of $\mathcal{P}_r^n$ for all positive integers $r \leq n$:

(5)
$$\mathcal{P}_r^n = \left\{ X \in \mathcal{S}^n : X = \sum_{\mathcal{F} \in \mathbf{F}_r([n])} \lambda_{\mathcal{F}} \mathbf{E}_{\mathcal{F}}, \sum_{\mathcal{F} \in \mathbf{F}_r([n])} \lambda_{\mathcal{F}} = 1, \, \lambda_{\mathcal{F}} \geq 0 \text{ for all } \mathcal{F} \in \mathbf{F}_r([n]) \right\}.$$

We call the description above the *packing description* of $\mathcal{P}_r^n$. Let us now consider the cardinality of the vertices of $\mathcal{P}_r^n$.

In the vein of Definition 2, we call $\mathcal{F} \subseteq \mathbb{P}([n])$ an r-partition of $[n]$ if it is an r-packing with $\bigcup_{S \in \mathcal{F}} = [n]$. Here, $\mathbb{P}([n])$ denotes the power set of $[n]$. The number of partitions of the set $[n]$ into k nonempty subsets is in the literature known as the Stirling number of the second kind, denoted by $\left\{ {n \atop k} \right\}$. The total number of partitions of $[n]$ equals the Bell number B_n [5], for which we have $B_n = \sum_{k=0}^n \left\{ {n \atop k} \right\}$. We can now show the following result regarding the cardinality of $\mathcal{D}_r^n$.

Theorem 4. *For $n \geq 1$ and $0 \leq r \leq n$, we have $|\mathcal{D}_r^n| = \sum_{k=1}^{r+1} \left\{ {n+1 \atop k} \right\}$. In particular, $|\mathcal{D}_1^n| = 2^n$ and $|\mathcal{D}_n^n| = B_{n+1}$.*

Proof. It follows from the discussion above that $|\mathcal{D}_r^n|$ equals the number of r-packings in $\mathbf{F}_r([n])$. In order to count these, we count the number of packings that consist of exactly k subsets, while k ranges from 0 to r. Any packing of $[n]$ into k subsets corresponds to a partition of $[n+1]$ into $k+1$ subsets. To see this, observe that to each packing $\mathcal{F}$ of $[n]$ into k subsets one can add a $(k+1)$th set containing the element $n+1$ and the elements not covered by $\mathcal{F}$. Conversely, given a partition of $[n+1]$ into $k+1$ subsets, dropping the set containing the element $n+1$ yields a packing of $[n]$ consisting of exactly k subsets. Hence, the number of packings of $[n]$ consisting of exactly k subsets equals $\left\{ {n+1 \atop k+1} \right\}$ and $|\mathcal{D}_r^n| = \sum_{k=0}^{r} \left\{ {n+1 \atop k+1} \right\} = \sum_{k=1}^{r+1} \left\{ {n+1 \atop k} \right\}$. For the special case $r=1$, we obtain $|\mathcal{D}_1^n| = \left\{ {n+1 \atop 1} \right\} + \left\{ {n+1 \atop 2} \right\} = 1 + \frac{2^{n+1}-2}{2} = 2^n$. When $r=n$, we exploit $\left\{ {n+1 \atop 0} \right\} = 0$ to conclude that $|\mathcal{D}_n^n| = \sum_{k=1}^{n+1} \left\{ {n+1 \atop k} \right\} = \sum_{k=0}^{n+1} \left\{ {n+1 \atop k} \right\} = B_{n+1}$. □

The polytope $\mathcal{P}_r^n$ has several relationships with other well-known polytopes from the literature. Letchford and Sørensen [40] study the polytope $\mathcal{P}_n^n$, albeit in a different embedding, and refer to it as the binary PSD polytope of order n. They emphasize its relationship with the clique partitioning polytope that was introduced by Grötschel and Wakabayashi [27] and later studied in [47]. Given the complete graph $G=(V,E)$, a clique partition is a subset $A \subseteq E$ such that there is a partition of V into nonempty disjoint sets $V_1, \dots, V_k$ such that each V_j, $j \in [k]$, induces a clique in G and $A = \bigcup_{j \in [k]} \{\{i,\ell\} \,:\, i,\ell \in V_j, i \neq \ell\}$. The incidence vectors of clique partitions are only defined on the edge set, and therefore the clique partition polytope can be seen as a projection of $\mathcal{P}_n^n$.

Among one of the first graph partition problems is the one considered by Chopra and Rao [12]. Given an undirected graph G, the vertices need to be partitioned into at most k subsets so as to minimize the total cost of edges cut by the partition. If G is the complete graph, the partition polytope $P1(r)$ considered in [12] coincides with $\mathcal{P}_r^n$ (apart from the embedding).

The polytope $\mathcal{P}_n^n$ can also be related to the stable set polytope. Let $G_{\mathbb{P}} = (V_{\mathbb{P}}, E_{\mathbb{P}})$ be the power set graph, i.e., each vertex in $V_{\mathbb{P}}$ corresponds to a nonempty subset of $[n]$ and the edge set is defined as $E_{\mathbb{P}} := \{\{S,T\} \in V_{\mathbb{P}}^{(2)} \,:\, S \cap T \neq \emptyset\}$. A set of vertices is stable in $G_{\mathbb{P}}$ if and only if its corresponding collection of subsets is a packing of $[n]$. Hence, the packing family $\mathbf{F}_n([n])$ is the collection of all stable sets in $G_{\mathbb{P}}$. It follows that there is a bijection between the elements in $\mathcal{P}_n^n$ and the stable set polytope of $G_{\mathbb{P}}$.

Finally, for $r=1$, the r-packings of $[n]$ are subsets of $[n]$, so the polytope $\mathcal{P}_1^n$ simplifies to

$$\mathcal{R}_1^n := \left\{ X \in \mathcal{S}^n \,:\, X = \sum_{S \subseteq [n]} \theta_S \mathbb{1}_S \mathbb{1}_S^\top, \sum_{S \subseteq [n]} \theta_S = 1, \, \theta_S \geq 0 \text{ for all } S \subseteq [n] \right\}. \tag{6}$$

The polytope $\mathcal{R}_1^n$ relates to the convex hull of the characteristic vectors of all cliques in K_n, i.e., the clique polytope of K_n. This polytope in the literature is also known as the complete set packing polytope; see [7]. Finally, apart from the embedding, the polytope $\mathcal{R}_1^n$ also coincides with the Boolean quadric polytope [48].

Another continuous formulation of the convex hull of PSD $\{0,1\}$-matrices is given by a conic description. The cone of completely positive matrices is defined as $\mathcal{CP}^n := \operatorname{conv}\left(\{xx^\top \,:\, x \in \mathbb{R}_+^n\}\right)$. An extension of the completely positive matrices are the so-called set-completely positive matrices (see, e.g., [41]), where the membership condition $x \in \mathbb{R}_+^n$ is replaced by $x \in \mathcal{K}$ for a general convex cone $\mathcal{K}$. Lieder, Rad, and Jarre [41] considered the following set-completely positive matrix cone

$$\mathcal{SCP}^n := \operatorname{conv}\left(\{xx^\top \,:\, x \in \mathbb{R}_+^n, \, x_1 \geq x_i \text{ for all } i \in \{2, \dots, n\}\}\right). \tag{7}$$

Since the membership condition given in (7) is more restricted than $x \in \mathbb{R}_+^n$, we have $\mathcal{SCP}^n \subsetneq \mathcal{CP}^n$. Let us now consider the following set-completely positive matrix set

$$\mathcal{C}_1^n := \left\{ X \in \mathcal{S}^n : \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathcal{SCP}^{n+1},\ \operatorname{diag}(X) = x \right\}. \tag{8}$$

The following result follows from Lieder, Rad, and Jarre [41].

THEOREM 5 (see [41]). *We have* $\mathcal{P}_1^n = \mathcal{C}_1^n$.

A natural question is whether the descriptions $\mathcal{R}_1^n$ and $\mathcal{C}_1^n$ for $\mathcal{P}_1^n$ given in (6) and (8), respectively, can be extended to higher ranks. The extensions of these sets are as follows:

(9)

$$\mathcal{R}_r^n := \left\{ X \in \mathcal{S}^n : X = \sum_{S \subseteq [n]} \theta_S \mathbb{1}_S \mathbb{1}_S^\top,\ \sum_{S \subseteq [n]} \theta_S = r,\ \sum_{S : i \in S} \theta_S \leq 1\ \forall i \in [n],\ \theta_S \geq 0\ \forall S \subseteq [n] \right\},$$

(10)

$$\mathcal{C}_r^n := \left\{ X \in \mathcal{S}^n : \begin{pmatrix} r & \operatorname{diag}(X)^\top \\ \operatorname{diag}(X) & X \end{pmatrix} \in \mathcal{SCP}^{n+1},\ \operatorname{diag}(X) \leq \mathbf{1}_n \right\}.$$

The extension from $\mathcal{C}_1^n$ to $\mathcal{C}_r^n$ follows from the intersection of the Minkowski sum of r copies of $\mathcal{C}_1^n$ with the upper bound constraint $X \leq \mathbf{J}_n$. Since $X_{ii} \geq X_{ij}$ for all $i, j \in [n]$ if $X \in \mathcal{C}_1^n$, it suffices to add $\operatorname{diag}(X) \leq \mathbf{1}_n$. The extension from $\mathcal{R}_1^n$ to $\mathcal{R}_r^n$ is derived as follows. If $X \in \mathcal{P}_r^n$, then $X = \sum_{\mathcal{F} \in \mathbf{F}_r([n])} \lambda_{\mathcal{F}} \mathbf{E}_{\mathcal{F}}$ for some nonnegative weights $\lambda_{\mathcal{F}}$. By splitting each r-packing into its separate subsets, we obtain

$$X = \sum_{\mathcal{F} \in \mathbf{F}_r([n])} \lambda_{\mathcal{F}} \mathbf{E}_{\mathcal{F}} = \sum_{\mathcal{F} \in \mathbf{F}_r([n])} \lambda_{\mathcal{F}} \sum_{S \in \mathcal{F}} \mathbb{1}_S \mathbb{1}_S^\top = \sum_{S \subseteq [n]} \sum_{\substack{\mathcal{F} \in \mathbf{F}_r([n]): \\ S \in \mathcal{F}}} \lambda_{\mathcal{F}} \mathbb{1}_S \mathbb{1}_S^\top = \sum_{S \subseteq [n]} \theta_S \mathbb{1}_S \mathbb{1}_S^\top,$$

where $\theta_S := \sum_{\mathcal{F} \in \mathbf{F}_r([n]) : S \in \mathcal{F}} \lambda_{\mathcal{F}}$. Moreover, $\sum_{S \subseteq [n]} \theta_S = \sum_{\mathcal{F} \in \mathbf{F}_r([n])} \lambda_F |\mathcal{F}| \leq r \sum_{\mathcal{F} \in \mathbf{F}_r([n])} \lambda_F = r$. By increasing $\theta_\emptyset$, we obtain $\sum_{S \subseteq [n]} \theta_S = r$. Finally, since $X_{ii} \leq 1$ for $i \in [n]$, we have $\sum_{S : i \in S} \theta_S \leq 1$. We conclude that $\mathcal{P}_r^n \subseteq \mathcal{R}_r^n$.

Unfortunately, for $r \geq 2$, the sets $\mathcal{R}_r^n$ and $\mathcal{C}_r^n$ no longer exactly describe $\mathcal{P}_r^n$. Namely, consider the matrix $X = \frac{1}{2} P P^\top$, where $P = \mathbf{E}_{11} + \mathbf{E}_{22} + \mathbf{E}_{23} + \mathbf{E}_{32} + \mathbf{E}_{31} + \mathbf{E}_{41} + \mathbf{E}_{43}$. For this matrix one can verify that $X \in \mathcal{R}_2^4$ and $X \in \mathcal{C}_2^4$, while $X \notin \mathcal{P}_2^4$. For $r \geq 2$, the following relationship between $\mathcal{P}_r^n$, $\mathcal{C}_r^n$, $\mathcal{R}_r^n$ holds.

THEOREM 6. *We have* $\mathcal{P}_r^n \subseteq \mathcal{C}_r^n = \mathcal{R}_r^n$*, while for* $r = 1$ *the three sets are equal.*

Proof. Since $\mathcal{P}_r^n = \operatorname{conv}(\mathcal{D}_r^n)$, it suffices to consider membership of the elements in $\mathcal{D}_r^n$ in $\mathcal{C}_r^n$. Let $X \in \mathcal{D}_r^n$, then $X = \sum_{j=1}^r x_j x_j^\top$ for some $x_j \in \{0,1\}^n$, $j \in [r]$. Let $Y^j := x_j x_j^\top$ for all $j \in [r]$. We clearly have $\begin{pmatrix} 1 & \operatorname{diag}(Y^j) \\ \operatorname{diag}(Y^j) & Y^j \end{pmatrix} \in \mathcal{SCP}^{n+1}$ for all $j \in [r]$, from which it follows that

$$\sum_{j=1}^r \begin{pmatrix} 1 & \operatorname{diag}(Y^j) \\ \operatorname{diag}(Y^j) & Y^j \end{pmatrix} = \begin{pmatrix} r & \operatorname{diag}(X) \\ \operatorname{diag}(X) & X \end{pmatrix} \in \mathcal{SCP}^{n+1}.$$

Moreover, $X \in \{0,1\}^{n \times n}$, so $\operatorname{diag}(X) \leq \mathbf{1}_n$. We conclude that $X \in \mathcal{C}_r^n$.

To prove $\mathcal{C}_r^n = \mathcal{R}_r^n$, let $X \in \mathcal{C}_r^n$. We define the matrix Y as

$$Y := \frac{1}{r} \begin{pmatrix} r & \operatorname{diag}(X)^\top \\ \operatorname{diag}(X) & X \end{pmatrix} = \begin{pmatrix} 1 & \operatorname{diag}(\frac{1}{r}X)^\top \\ \operatorname{diag}(\frac{1}{r}X) & \frac{1}{r}X \end{pmatrix}. \tag{11}$$

From the fact that $X \in \mathcal{C}_r^n$, it follows that $Y \in \mathcal{SCP}^{n+1}$. Applying Theorem 5 to the matrix Y implies that $\frac{1}{r}X$ is a convex combination of rank-one binary PSD matrices, i.e., there exists $\theta'_S \geq 0$ for all $S \subseteq [n]$ with $\sum_{S\subseteq[n]} \theta'_S = 1$, such that $\frac{1}{r}X = \sum_{S\subseteq[n]} \theta'_S \mathbb{1}_S \mathbb{1}_S^\top$ or, equivalently, $X = \sum_{S\subseteq[n]} r\theta'_S \mathbb{1}_S \mathbb{1}_S^\top$. Now, let $\theta_S := r\theta'_S$ for all $S \subseteq [n]$, from which it follows that $\sum_{S\subseteq[n]} \theta_S = r$. Since $\operatorname{diag}(X) \leq \mathbf{1}_n$, it follows that $X_{ii} = \sum_{S:i\in S} \theta_S \leq 1$. We conclude that $X \in \mathcal{R}_r^n$.

Finally, observe that the argument above can also be followed in the converse direction. That is, given $X \in \mathcal{R}_r^n$ with corresponding weights θ_S for all $S \subseteq [n]$, we define $\theta'_S := \frac{1}{r}\theta_S$, $S \subseteq [n]$, which implies that $\frac{1}{r}X \in \mathcal{P}_1^n$. By Theorem 5, we know that Y (see (11)), is contained in $\mathcal{SCP}^{n+1}$, implying $X \in \mathcal{C}_r^n$. ∎

2.2. Theory on PSD {±1}-matrices. In this section we present several results for PSD matrices that have entries in $\{\pm 1\}$. Let us first state the following well-known result; see, e.g., [2].

PROPOSITION 2. *Let X be a symmetric matrix. Then, $X \succeq \mathbf{0}$, $X \in \{\pm 1\}^{n\times n}$ if and only if $X = xx^\top$ for some $x \in \{\pm 1\}^n$.*

A simple necessary condition for $X \in \{\pm 1\}^{n\times n}$ to be PSD is that $\operatorname{diag}(X) = \mathbf{1}$. The next result establishes the equivalence between $\{0,1\}$- and $\{\pm 1\}$-PSD matrices by exploiting their rank.

PROPOSITION 3. *Let $X \in \{\pm 1\}^{n\times n}$ be a symmetric matrix and $Y := \frac{1}{2}(X + J) \in \{0,1\}^{n\times n}$. Then, $X \succeq \mathbf{0}$ if and only if $\operatorname{diag}(Y) = \mathbf{1}$, $Y \succeq \mathbf{0}$, and $\operatorname{rank}(Y) \leq 2$.*

Proof. ($\Longrightarrow$): Let $X \succeq \mathbf{0}$. Since $\mathbf{J} \succeq \mathbf{0}$, it follows that $Y \succeq \mathbf{0}$. Moreover, $\operatorname{diag}(X) = \operatorname{diag}(\mathbf{J}) = \mathbf{1}$ implies that $\operatorname{diag}(Y) = \mathbf{1}$. Finally, by Proposition 2 we know that $X = xx^\top$ for some $x \in \{\pm 1\}^n$. Therefore, Y is the weighted sum of two rank-one matrices, so $\operatorname{rank}(Y) \leq 2$.

($\Longleftarrow$): Let $Y = \frac{1}{2}(X + \mathbf{J}) \succeq \mathbf{0}$, $\operatorname{diag}(Y) = \mathbf{1}$, and $\operatorname{rank}(Y) \leq 2$ for some symmetric matrix $X \in \{\pm 1\}^{n\times n}$. Since Y is binary PSD with rank at most two, it follows from Theorem 1 that $Y = x_1x_1^\top + x_2x_2^\top$ for some $x_1, x_2 \in \{0,1\}^n$. Then, $X = 2Y - \mathbf{J} = (x_1 - x_2)(x_1 - x_2)^\top$, which implies that $X \succeq \mathbf{0}$. ∎

Note that the matrix Y from the previous theorem has rank one if and only if $Y = X = \mathbf{J}$. Similarly to (1), we define the discrete set of all $\{\pm 1\}$-matrices as

$$\widehat{\mathcal{D}}^n := \left\{ X \in \{\pm 1\}^{n\times n} \,:\, X \succeq \mathbf{0} \right\}, \tag{12}$$

where the subscript r is not present anymore, as all matrices in $\widehat{\mathcal{D}}^n$ have rank one. Based on Proposition 2, we can easily establish that $|\widehat{\mathcal{D}}^n| = 2^{n-1}$. Next, we summarize known results on sets related to $\{\pm 1\}$-matrices. The convex hull of all PSD $\{\pm 1\}$-matrices is known as the cut polytope:

$$\widehat{\mathcal{P}}^n := \operatorname{conv}(\widehat{\mathcal{D}}^n); \tag{13}$$

see, e.g., [38]. Also, we define the following set-completely positive matrix cone

$$\mathcal{SC}\widehat{\mathcal{P}}^n := \operatorname{conv}\left(\left\{ xx^\top \,:\, x \in \mathbb{R}^n,\, x_1 + x_i \geq 0,\, x_1 - x_i \geq 0 \text{ for all } i \in \{2,\ldots,n\} \right\}\right). \tag{14}$$

The cone $\mathcal{SC}\widehat{\mathcal{P}}^n$ is considered in [41], where the authors show that $\mathcal{SC}\widehat{\mathcal{P}}^n$ and $\mathcal{SCP}^n$ (see (7)), are related as follows, $\mathcal{T}(\mathcal{SCP}^n) = \mathcal{SC}\widehat{\mathcal{P}}^n$ and $\mathcal{T}^{-1}(\mathcal{SC}\widehat{\mathcal{P}}^n) = \mathcal{SCP}^n$, where $\mathcal{T}$ is an appropriate linear mapping. Lieder, Rad, and Jarre [41] consider the following set-completely positive matrix set,

$$\widehat{\mathcal{C}}^n := \left\{ X \in \mathcal{S}^n \,:\, \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathcal{SC}\widehat{\mathcal{P}}^{n+1},\, \operatorname{diag}(X) = \mathbf{1}_n \right\}, \tag{15}$$

which is the analogue of the set $\mathcal{C}_1^n$ for $\{0,1\}$-matrices; see (8).

THEOREM 7 (see [41]). *We have* $\widehat{\mathcal{P}}^n = \widehat{\mathcal{C}}^n$.

This theorem is the analogue of Theorem 5 that provides a result for $\{0,1\}$-matrices. For the equivalence transformation between $\{\pm 1\}$- and $\{0,1\}$-representations of SDP relaxations of binary quadratic optimization problems, we refer the interested reader to Helmberg [28].

2.3. Theory on PSD $\{0,\pm 1\}$-matrices. In the following we generalize several results from the previous sections to PSD $\{0,\pm 1\}$-matrices. The following result shows that a PSD $\{0,\pm 1\}$-matrix is block-diagonalizable, which is the analogue of Proposition 1 for $\{0,1\}$-matrices.

PROPOSITION 4. *Let $X \in \{0,\pm 1\}^{n\times n}$ be symmetric. Then, the following statements are equivalent:*

(i) $\mathrm{diag}(X) = \mathbf{1}_n$, $\mathrm{rank}(X) = r$, $X \succeq \mathbf{0}$.

(ii) *There exists a permutation matrix Q such that $QXQ^\top = \boldsymbol{B}_{n_1} \oplus \cdots \oplus \boldsymbol{B}_{n_r}$, where $\boldsymbol{B}_{n_i} = b_i b_i^\top$, $b_i \in \{\pm 1\}^{n_i}$ for $i \in [r]$, $n = n_1 + \cdots + n_r$.*

Proof. Suppose that $QXQ^\top$ is in the block form given in (ii), then it trivially satisfies the conditions given in (i). Conversely, let $X \in \{0,\pm 1\}^{n\times n}$ satisfy (i). Let us consider the ith row in X. Suppose j and k are two distinct indices not equal to i in the support of this row, i.e., $X_{ij}, X_{ik} \neq 0$. For the sake of contradiction, suppose that $X_{jk} = 0$. Then, the submatrix of X induced by i, j, and k is one of the following matrices:

$$\begin{array}{c} \\ i \\ j \\ k \end{array}\begin{array}{c} \begin{array}{ccc} i & j & k \end{array} \\ \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \end{array}, \quad \begin{array}{c} \\ i \\ j \\ k \end{array}\begin{array}{c} \begin{array}{ccc} i & j & k \end{array} \\ \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \end{array}, \quad \begin{array}{c} \\ i \\ j \\ k \end{array}\begin{array}{c} \begin{array}{ccc} i & j & k \end{array} \\ \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \end{array}, \quad \text{or} \quad \begin{array}{c} \\ i \\ j \\ k \end{array}\begin{array}{c} \begin{array}{ccc} i & j & k \end{array} \\ \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \end{array}.$$

One easily checks that the determinants of these matrices are all negative, contradicting that $X \succeq \mathbf{0}$. Hence, $X_{jk} \neq 0$. This argument can be repeated to conclude that the submatrix of X indexed by the support of row i has entries in $\{\pm 1\}$. Since the submatrix of X is also PSD, it follows from Proposition 2 that the submatrix is of the form $bb^\top$ with $b \in \{\pm 1\}^{n_i}$ for some positive integer n_i.

By the same argument, it follows that the other indices in the submatrix induced by row i have the same support as row i. Indeed, if this would not be the case, one of the four matrices above should be a submatrix of X. We conclude that X can be fully constructed from nonoverlapping submatrices of the form $bb^\top$ with $b \in \{\pm 1\}^{n_i}$ for some positive integer n_i. Since its rank equals r, there must be r of those submatrices. From here the claim follows. ∎

Proposition 4 extends easily to the following result.

COROLLARY 4. *Let $X \in \{0,\pm 1\}^{n\times n}$ be symmetric. Then, the following statements are equivalent:*

(i) $\mathrm{rank}(X) = r$, $X \succeq \mathbf{0}$.

(ii) *There exists a permutation matrix Q such that $QXQ^\top = \boldsymbol{B}_{n_1} \oplus \cdots \oplus \boldsymbol{B}_{n_r} \oplus \mathbf{0}_{n_z\times n_z}$, where $\boldsymbol{B}_{n_i} = b_i b_i^\top$, $b_i \in \{\pm 1\}^{n_i}$ for $i \in [r]$, $n = n_1 + \cdots + n_r + n_z$.*

Proof. The proof is similar to the proof of Corollary 1. ∎

Let $X \in \{0,\pm 1\}^{n\times n}$ be given as in Corollary 4, then

$$QXQ^\top = \mathbf{B}_{n_1} \oplus \cdots \oplus \mathbf{B}_{n_r} \oplus \mathbf{0}_{n_z\times n_z} = b_1 b_1^\top \oplus \cdots \oplus b_r b_r^\top \oplus \mathbf{0}_{n_z\times n_z} = \sum_{i=1}^r \bar{x}_i \bar{x}_i^\top,$$

where $\bar{x}_1^\top = [b_1^\top \ \mathbf{0}_{n-n_1}^\top]$, $\bar{x}_2^\top = [\mathbf{0}_{n_1}^\top \ b_2^\top \ \mathbf{0}_{n-n_1-n_2}^\top]$, and so on. Let $x_i := Q\bar{x}_i$ for $i \in [r]$, then $X = \sum_{i=1}^r x_i x_i^\top$, where $x_i \in \{0, \pm 1\}^n$. This construction yields the following decomposition of PSD $\{0, \pm 1\}$-matrices.

THEOREM 8. *Let $X \in \{0, \pm 1\}^{n \times n}$ be symmetric. Then, $X \succeq \mathbf{0}$ if and only if $X = \sum_{j=1}^r x_j x_j^\top$ for some $x_j \in \{0, \pm 1\}^n$, $j \in [r]$.*

The previous result is an extension of Theorem 1 to $\{0, \pm 1\}$matrices. We now consider an equivalence between a PSD $\{0, \pm 1\}$matrix of rank one and an extended linear matrix inequality, i.e., the analogue of Theorem 3.

PROPOSITION 5. *Let $Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathcal{S}^{n+1}$ with supp$(\operatorname{diag}(X)) =$ supp(x). Then, $Y \in \{0, \pm 1\}^{(n+1) \times (n+1)}$, $Y \succeq \mathbf{0}$ if and only if $X = xx^\top$.*

Proof. Let $Y_{ij} \in \{0, \pm 1\}$ for all $i, j \in [n+1]$ and $Y \succeq \mathbf{0}$. Then $x \in \{0, \pm 1\}^n$. The Schur complement lemma implies $X - xx^\top \succeq \mathbf{0}$. If $X_{ii} = 0$ then $x_i = 0$, and if $X_{ii} = 1$ then $x_i = 1$ or $x_i = -1$. Thus $\operatorname{diag}(X - xx^\top) = \mathbf{0}$, from which it follows that $X = xx^\top$. The converse direction is trivial. ◻

Clearly, the condition $\operatorname{supp}(\operatorname{diag}(X)) = \operatorname{supp}(x)$ can be replaced by $\operatorname{diag}(X)_{ii} = |x_i|$ for all $i \in [n]$, where $|\cdot|$ denotes the absolute value.

3. Binary quadratic optimization problems.

In this section we exploit the theoretical results on discrete PSD matrices from the previous section to derive exact reformulations of BQPs as BSDPs. In section 3.1 we consider the general class of binary quadratically constrained quadratic programs. In section 3.2 we consider a subclass of these programs that allow for a formulation as a binary quadratic matrix program.

3.1. Binary quadratically constrained quadratic programs.

A quadratically constrained quadratic program (QCQP) is an optimization problem with a quadratic objective function under the presence of quadratic constraints. Many discrete optimization problems can be formulated as QCQPs.

Let $Q_0, Q_i \in \mathcal{S}^n$, $c_0, c_i \in \mathbb{R}^n$ for all $i \in [m]$, and $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ for all $i \in [p]$, where $m, p \in \mathbb{N}$. We consider binary programs of the following form:

$$(QCQP) \qquad \begin{aligned} \min \quad & x^\top Q_0 x + c_0^\top x \\ \text{s.t.} \quad & x^\top Q_i x + c_i^\top x \le d_i \quad \forall i \in [m], \\ & {a_i}^\top x = b_i \quad \forall i \in [p], \\ & x \in \{0, 1\}^n. \end{aligned}$$

The quadratic terms in $(QCQP)$ can be written as $\langle Q_i, X \rangle + c_i^\top x$ for all i, where we substitute X for $xx^\top$. This yields the following exact reformulation of $(QCQP)$:

$$\begin{aligned} \min \quad & \langle Q_0, X \rangle + c_0^\top x \\ \text{s.t.} \quad & \langle Q_i, X \rangle + c_i^\top x \le d_i \quad \forall i \in [m], \\ & {a_i}^\top x = b_i \quad \forall i \in [p], \\ & Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq \mathbf{0}, \ \operatorname{diag}(X) = x, \ \operatorname{rank}(Y) = 1. \end{aligned}$$

Here we used the conventional notion of exactness, i.e., the nonconvex constraint $\operatorname{rank}(Y) = 1$. We also exploit here Theorem 3 in order to not explicitly require that

x is binary. However, one can utilize an alternative notion of exactness in terms of integrality, namely, by exploiting Theorem 3. This leads to the following BSDP:

$$(BSDP_{QCQP}) \qquad \begin{aligned} \min \quad & \langle Q_0, X\rangle + c_0^\top x \\ \text{s.t.} \quad & \langle Q_i, X\rangle + c_i^\top x \le d_i \quad \forall i \in [m], \\ & {a_i}^\top x = b_i \quad \forall i \in [p], \\ & Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq \mathbf{0},\ \mathrm{diag}(X) = x,\ x \in \{0,1\}^n. \end{aligned}$$

Observe that it is sufficient to impose integrality on the diagonal of X. Namely, it follows from the determinants of the 3×3 principal submatrices of the matrix Y that $X_{ij} \in \{0,1\}$ whenever $X_{ii}, X_{jj} \in \{0,1\}$ for all i and j. Note that a binary matrix X that satisfies the linear matrix inequality from $(BSDP_{QCQP})$ with $x = \mathrm{diag}(X)$, is an element of $\mathcal{D}_1^n$; see (3). The next result follows directly from the previous discussion and Theorem 3.

Theorem 9. *$(BSDP_{QCQP})$ is equivalent to $(QCQP)$.*

To provide a more compact BSDP formulation of $(QCQP)$, we prove the following result.

Lemma 1. *Let $S = \sum_{i=1}^p \binom{-b_i}{a_i}\binom{-b_i}{a_i}^\top$ and $Y = \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq \mathbf{0}$, where $\mathrm{diag}(X) = x$ and $X \in \{0,1\}^{n\times n}$. Then, $a_i^\top x = b_i$ for all $i \in [p]$ if and only if $\langle S, Y\rangle = 0$.*

Proof. It follows from Theorem 3 that $Y = \binom{1}{x}\binom{1}{x}^\top$. If $a_i^\top x = b_i$ for all $i \in [p]$, it is not difficult to verify that $\langle S, Y\rangle = 0$. Conversely, let $\langle S, Y\rangle = 0$. Then, $0 = \sum_{i=1}^p \left\langle \binom{-b_i}{a_i}\binom{-b_i}{a_i}^\top, \binom{1}{x}\binom{1}{x}^\top \right\rangle = \sum_{i=1}^p (b_i - a_i^\top x)^2$, from which it follows that $a_i^\top x = b_i$ for all $i \in [p]$. ∎

Lemma 1 induces the following compact BSDP that is equivalent to $(QCQP)$:

$$\begin{aligned} \min \quad & \langle Q_0, X\rangle + c^\top x \\ \text{s.t.} \quad & \langle Q_i, X\rangle + c_i^\top x \le d_i \quad \forall i \in [m], \\ & \sum_{i=1}^p \left\langle \begin{pmatrix} -b_i \\ a_i \end{pmatrix} \begin{pmatrix} -b_i \\ a_i \end{pmatrix}^\top, \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \right\rangle = 0, \\ & \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq \mathbf{0},\ \mathrm{diag}(X) = x,\ x \in \{0,1\}^n. \end{aligned}$$

There are various equivalent formulations of the BQP $(QCQP)$ in the literature. We finalize this subsection by mentioning below only those that are closely related to our approach.

Assume that $Q_i = \mathbf{0}$, $c_i = \mathbf{0}$, and $d_i = 0$ for all $i \in [m]$ in $(QCQP)$. Burer [8] proved that the resulting optimization problem with a quadratic objective and linear constraints is equivalent to the following completely positive program,

$$\begin{aligned} \min \quad & \langle Q_0, X\rangle + c^\top x \\ \text{s.t.} \quad & {a_i}^\top x = b_i \quad \forall i \in [p],\ \langle a_i {a_i}^\top, X\rangle = b_i^2 \quad \forall i \in [p], \\ & \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathcal{CP}^{n+1},\ \mathrm{diag}(X) = x, \end{aligned}$$

provided that the inequalities $0 \le x_i \le 1$ for $i \in [n]$ are implied by the constraints of the original problem. Here $\mathcal{CP}^{n+1}$ is the cone of completely positive matrices.

On the other hand, Lieder, Rad, and Jarre [41] proved the following equivalent formulation of the BQP with quadratic objective and linear constraints,

$$\begin{aligned} \min \quad & \langle Q_0, X\rangle + c^\top x \\ \text{s.t.} \quad & {a_i}^\top x = b_i \quad \forall i \in [p], \\ & \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathcal{SCP}^{n+1},\ \operatorname{diag}(X) = x, \end{aligned}$$

where the cone $\mathcal{SCP}^{n+1}$ is defined in (7). The authors of [41] also proved that, under mild assumptions, the BQP $(QCQP)$ also with quadratic constraints can be equivalently reformulated as an optimization problem over the set-completely positive matrix cone $\mathcal{SCP}^{n+1}$.

We end this section by providing an example of a problem that can be modeled as $(BSDP_{QCQP})$.

Example 1 (the stable set problem). Let $G = (V, E)$ be a simple graph on n vertices. A stable set in G is a subset $S \subseteq V$ such that no two vertices in S are adjacent in G. The stable set problem (SSP) asks for the largest size of a stable set in G. To model this problem, let $x \in \{0,1\}^n$ be such that $x_i = 1$ if and only if $i \in S$. Then, x is the characteristic vector of a stable set if $x^\top(\mathbf{E}_{ij} + \mathbf{E}_{ij}^\top)x = 0$ for all $\{i,j\} \in E$, hence the SSP is of the form $(QCQP)$. Applying Theorem 9, the following BSDP models the SSP:

$$\begin{aligned} \alpha(G) := \max \quad & \langle \mathbf{I}_n, X\rangle \\ \text{s.t.} \quad & X_{ij} = 0 \quad \forall \{i,j\} \in E,\ \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succeq \mathbf{0},\ \operatorname{diag}(X) = x,\ x \in \{0,1\}^n. \end{aligned} \tag{16}$$

The doubly nonnegative relaxation of (16) obtained after replacing $x \in \{0,1\}^n$ by $\mathbf{0} \le x \le \mathbf{1}$, is well studied in the literature [26], and is equivalent to a strengthened version of the Lovász theta number [53].

3.2. Binary quadratic matrix programs. A quadratic matrix program (QMP) [4] is a programming formulation where the objective and constraints are given by

$$\operatorname{tr}(P^\top Q_i P) + 2\operatorname{tr}(B_i^\top P) + d_i \tag{17}$$

for some $Q_i \in \mathcal{S}^n$, $B_i \in \mathbb{R}^{n\times k}$, and $c_i \in \mathbb{R}$, where P is an $n \times k$ matrix variable. QMPs are a special case of QCQPs and are particularly useful to model optimization problems where the matrix P has entries in $\{0,1\}$ and represents a classification of n objects over k classes, i.e., $P_{ij} = 1$ if and only if object i is assigned to class j. For example, if each object needs to be assigned in exactly (resp., at most) one class, we call P a partition (resp., packing) matrix.

In this section we consider two different binary QMPs of increasing generality and show how to reformulate them as BSDPs. For both QMPs, we consider some problems that fit into the framework.

Our first QMP incorporates a specific objective and constraint structure, while optimizing over the packing or partition matrices. Let $Q_0, Q_i \in \mathcal{S}^n$, $d_i \in \mathbb{R}$ for all $i \in [m]$, $a_i \in \mathbb{R}^n$, and $b_i \in \mathbb{R}_+$ for all $i \in [p]$. We consider the binary quadratic matrix program

$$(QMP_1) \qquad \begin{aligned} \min \quad & \operatorname{tr}(P^\top Q_0 P) \\ \text{s.t.} \quad & \operatorname{tr}(P^\top Q_i P) + d_i \le 0 \quad \forall i \in [m],\ P^\top a_i \le b_i \mathbf{1}_k \quad \forall i \in [p], \\ & P\mathbf{1}_k \le \mathbf{1}_n,\ P \in \{0,1\}^{n\times k}. \end{aligned}$$

Observe that $P\mathbf{1}_k \leq \mathbf{1}_n$ implies that P is a packing matrix. This constraint is replaced by $P\mathbf{1}_k = \mathbf{1}_n$ in case we deal with partition matrices. The constraints $\operatorname{tr}(P^\top Q_i P) + d_i \leq 0$ and $P^\top a_i \leq b_i \mathbf{1}_k$ might follow from the structure of the problem under consideration. Observe that these constraints differ from the general form (17) in the sense that the linear part $\operatorname{tr}(B_i^\top P)$ is only included in a very specific form.

A possible way to deal with the quadratic terms in (QMP_1) is by lifting the variables in a higher-dimensional space. By vectorizing the matrix P, the problem (QMP_1) can be written in the form $(QCQP)$, after which we can follow the approach of section 3.1. This results in a BSDP where the matrix variable is of order $nk+1$. Since the resulting program is obtained from a lifting of the vectorization of P, we say that we applied a vector-lifting approach. To obtain a more compact problem formulation where the matrix variable is of lower order, we here consider a matrix-lifting approach. In particular, the objective function can be written as $\operatorname{tr}(P^\top Q_0 P) = \operatorname{tr}(Q_0 PP^\top) = \operatorname{tr}(Q_0 X)$, where $X = PP^\top$. By doing so, we obtain the following BSDP:

$$(BSDP_{QMP1}) \qquad \begin{aligned} \min \quad & \langle Q_0, X\rangle \\ \text{s.t.} \quad & \langle Q_i, X\rangle + d_i \leq 0 \quad \forall i \in [m],\ Xa_i \leq b_i x \quad \forall i \in [p], \\ & \begin{pmatrix} k & x^\top \\ x & X \end{pmatrix} \succeq \mathbf{0},\ \operatorname{diag}(X) = x,\ X \in \{0,1\}^{n\times n}. \end{aligned}$$

If a QMP is defined over the partition matrices, then $P\mathbf{1}_k \leq \mathbf{1}_n$ is replaced by $P\mathbf{1}_k = \mathbf{1}_n$ in (QMP_1), and consequently $\operatorname{diag}(X) = x$ is replaced by $\operatorname{diag}(X) = \mathbf{1}_n$ in $(BSDP_{QMP1})$. By exploiting theory from section 2.1, we show the following equivalence.

THEOREM 10. *$(BSDP_{QMP1})$ is equivalent to (QMP_1).*

Proof. Let P be feasible for (QMP_1) and define $X = PP^\top$ and $x = P\mathbf{1}_k$. Since P represents a packing matrix, we have $X \in \{0,1\}^{n\times n}$, where x is a $\{0,1\}$-vector indicating whether object i is packed in one of the classes or not. Then, $\langle Q_i, X\rangle + d_i = \langle Q_i, PP^\top\rangle + d_i = \operatorname{tr}(P^\top Q_i P) + d_i \leq 0$ for all $i \in [m]$. Moreover, we have $Xa_i = PP^\top a_i \leq b_i P\mathbf{1}_k = b_i x$. To show that $\operatorname{diag}(X) = x$, observe that $X_{ii} = \sum_{j=1}^k P_{ij}^2 = \sum_{j=1}^k P_{ij} = \mathbf{e}_i^\top P\mathbf{1}_k = x_i$. Finally, we can decompose the matrix $\left(\begin{smallmatrix} k & x^\top \\ x & X\end{smallmatrix}\right)$ into $\left(\begin{smallmatrix} k & x^\top \\ x & X\end{smallmatrix}\right) = \left(\begin{smallmatrix} \mathbf{1}_k^\top \\ P\end{smallmatrix}\right)\left(\begin{smallmatrix} \mathbf{1}_k^\top \\ P\end{smallmatrix}\right)^\top$, showing that it is PSD. We conclude that X and x are feasible for $(BSDP_{QMP1})$.

To show the converse inclusion, let X and $x = \operatorname{diag}(X)$ be feasible for $(BSDP_{QMP1})$. It follows from Corollary 2 that X can be decomposed as the sum of at most k rank-one symmetric $\{0,1\}$-matrices. By adding copies of the zero matrix in the case $\operatorname{rank}(X) < k$, we may assume that there exist $x_1, \dots, x_k \in \{0,1\}^n$ such that $X = \sum_{j=1}^k x_j x_j^\top$. Now, let $P = [x_1 \ \dots \ x_k]$. Then, $P \in \{0,1\}^{n\times k}$ with $P\mathbf{1}_k = \sum_{j=1}^k x_j = \operatorname{diag}(X) \leq \mathbf{1}_n$. To prove that $P^\top a_i \leq b_i\mathbf{1}_k$, consider column j^* of P. If all entries in $P\mathbf{e}_{j^*}$ $(= x_{j^*})$ are zero, this implies that $\mathbf{e}_{j^*}^\top P^\top a_i = 0 \leq b_i$, since $b_i \in \mathbb{R}_+$. Otherwise, there exists a row i^* such that $P_{i^* j^*} = 1$. For the i^*th row of X, we know $\mathbf{e}_{i^*}^\top X = \sum_{j=1}^k (x_j)_{i^*} x_j^\top = x_{j^*}^\top$. The i^*th row of the system $Xa_i \leq b_i x$ then reads $x_{j^*}^\top a_i \leq b_i x_{i^*} = b_i$. Hence, $P^\top a_i \leq b_i \mathbf{1}_k$. Finally, the constraint $\operatorname{tr}(P^\top Q_i P) + d_i \leq 0$ follows immediately from $\langle Q_i, X\rangle + d_i \leq 0$ for all $i \in [m]$. Thus, P is feasible for (QMP_1).

As the objective functions of (QMP_1) and $(BSDP_{QMP1})$ clearly coincide with respect to the given mapping between P and X, we conclude that the two programs are equivalent. ∎

The matrix P no longer appears explicitly in $(BSDP_{QMP1})$, and therefore we will not be able to write all quadratic problems over the packing or partition matrices in this form. The typical problems that can be modeled as $(BSDP_{QMP1})$, are the ones that are symmetric over the classes $[k]$, i.e., we do not add constraints for one specific class. Below we discuss two examples from the literature that fit in the framework of (QMP_1).

Example 2 (the maximum k-colorable subgraph problem). Let $G=(V,E)$ be a simple graph on n vertices. Given a positive integer k, a graph is called k-colorable if it is possible to assign to each vertex a color of $[k]$ such that any two adjacent vertices get assigned a different color. The maximum k-colorable subgraph (MkCS) problem [37, 46] asks to find an induced subgraph $G'=(V',E')$ of G that is k-colorable such that $|V'|$ is maximized. The MkCS problem can be modeled as (QMP_1), where $P \in \{0,1\}^{n\times k}$ is such that $P_{ij}=1$ if and only if vertex $i \in [V]$ is in color class $j \in [k]$. In order to model that P induces a coloring in G, we include the constraints $\operatorname{tr}(P^\top(\mathbf{E}_{ij}+\mathbf{E}_{ji})P)=0$ for all $\{i,j\}\in E$.

Now, it follows from Theorem 10 that the MkCS problem can be modeled as the following BSDP:

$$\begin{aligned} &\max \quad \langle \mathbf{I}_n, X\rangle \\ &\text{s.t.} \quad X_{ij}=0 \quad \forall \{i,j\}\in E,\ \begin{pmatrix} k & x^\top \\ x & X\end{pmatrix} \succeq \mathbf{0},\ \operatorname{diag}(X)=x,\ X\in\{0,1\}^{n\times n}. \end{aligned} \tag{18}$$

After replacing $X\in\{0,1\}^{n\times n}$ by $\mathbf{0}\le X\le \mathbf{J}$ in (18), we obtain the formulation $\theta_k^3(G)$ derived in [37].

The next example shows that the parameter k in $(BSDP_{QMP1})$ can also be used as a variable in order to quantify the number of classes in the solution.

Example 3 (the quadratic bin packing problem). Consider a set of n items, each with a weight $w_i \in \mathbb{R}_+$, and an unbounded number of bins, each with capacity $W\in\mathbb{R}_+$ and cost $c\in\mathbb{R}_+$. Let $D=(d_{ij})\in\mathcal{S}^n$ denote a dissimilarity matrix, where d_{ij} is the cost of packing item i and j in the same bin. The goal of the quadratic bin packing problem (QBPP) [11] is to assign each item to exactly one bin, such that the sum of the dissimilarity and the cost of the used bins is minimized, while not violating the capacity constraints.

Suppose the number of available bins is k. Then, the problem is of the form (QMP_1), where $P\in\{0,1\}^{n\times k}$ is a matrix with $P_{ij}=1$ if and only if item i is contained in bin j, where we require that $P\mathbf{1}_k=\mathbf{1}_n$ and $P^\top w\le W\mathbf{1}_k$. Theorem 10 shows that this problem can be modeled as a BSDP, where k appears as a parameter. If we replace k by a variable z, we obtain the following formulation of the QBPP:

$$\begin{aligned} &\min \quad \left\langle \begin{pmatrix} z & \mathbf{1}_n^\top \\ \mathbf{1}_n & X\end{pmatrix}, c\oplus D\right\rangle \\ &\text{s.t.} \quad Xw\le W\mathbf{1}_n,\ \operatorname{diag}(X)=\mathbf{1}_n,\ \begin{pmatrix} z & \mathbf{1}_n^\top \\ \mathbf{1}_n & X\end{pmatrix}\succeq \mathbf{0},\ X\in\{0,1\}^{n\times n},\ z\in\mathbb{R}. \end{aligned}$$

The variable z is not restricted to be integer, since at an optimal solution it always equals $\operatorname{rank}(X)$.

The quadratic matrix program (QMP_1) only includes specific types of constraints of the form (17). We now consider a generalization of (QMP_1). Let $Q_0, Q_i\in\mathcal{S}^n$, $B_0, B_i\in\mathbb{R}^{n\times k}$, and $d_0, d_i\in\mathbb{R}$ for all $i\in[m]$ and consider the quadratic matrix program

$$(QMP_2) \qquad \begin{aligned} \min \quad & \operatorname{tr}(P^\top Q_0 P) + 2\operatorname{tr}(B_0^\top P) + d_0 \\ \text{s.t.} \quad & \operatorname{tr}(P^\top Q_i P) + 2\operatorname{tr}(B_i^\top P) + d_i \le 0 \quad \forall i \in [m], \\ & P\mathbf{1}_k \le \mathbf{1}_n,\ P \in \{0,1\}^{n\times k}. \end{aligned}$$

Again, the constraint $P\mathbf{1}_k \le \mathbf{1}_n$ can be replaced by $P\mathbf{1}_k = \mathbf{1}_n$ when optimizing over partition matrices. Now, let us consider the BSDP

$(BSDP_{QMP2})$

$$\begin{aligned} \min \quad & \left\langle \begin{pmatrix} \frac{d_0}{k}\mathbf{I}_k & B_0^\top \\ B_0^\top & Q_0 \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix} \right\rangle \\ \text{s.t.} \quad & \left\langle \begin{pmatrix} \frac{d_i}{k}\mathbf{I}_k & B_i^\top \\ B_i^\top & Q_i \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix} \right\rangle \le 0 \quad \forall i \in [m], \\ & \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix} \succeq \mathbf{0},\ \operatorname{diag}(X) = P\mathbf{1}_k,\ X \in \{0,1\}^{n\times n},\ P \in \{0,1\}^{n\times k}, \end{aligned}$$

which is equivalent to (QMP_2), as shown below.

THEOREM 11. *$(BSDP_{QMP2})$ is equivalent to (QMP_2).*

Proof. Let P be feasible for (QMP_2) and define $Y \in \{0,1\}^{(n+k)\times(n+k)}$ as $Y = \binom{\mathbf{I}_k}{P}\binom{\mathbf{I}_k}{P}^\top = \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix}$, where $X := PP^\top$. Clearly, we have $Y \succeq \mathbf{0}$ and $X_{ii} = \sum_{j=1}^k P_{ij}^2 = \sum_{j=1}^k P_{ij} = \mathbf{e}_i^\top P\mathbf{1}_k$ for all $i \in [n]$, showing that $\operatorname{diag}(X) = P\mathbf{1}_k$. Moreover, we have

$$\begin{aligned} \operatorname{tr}(P^\top Q_i P) + 2\operatorname{tr}(B_i^\top P) + d_i &= \operatorname{tr}(Q_i X) + 2\operatorname{tr}(B_i^\top P) + d_i \\ &= \left\langle \begin{pmatrix} \frac{d_i}{k}\mathbf{I}_k & B_i^\top \\ B_i^\top & Q_i \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix} \right\rangle \end{aligned}$$

for all $i \in [m]$ and $i = 0$. Hence, X and P are feasible for $(BSDP_{QMP2})$ and the objective functions coincide.

Conversely, let $P \in \{0,1\}^{n\times k}$ and $X \in \{0,1\}^{n\times n}$ be feasible for $(BSDP_{QMP2})$. Following the proof of Theorem 2, it follows that there exist $x_1, \ldots, x_{k'} \in \{0,1\}^n$ with $k' \ge k$ such that $P = [x_1 \ \ldots \ x_k]$ and $X = \sum_{j=1}^{k'} x_j x_j^\top$. Since $\operatorname{diag}(X) = P\mathbf{1}_k$, it follows that for all $i \in [n]$ we have $X_{ii} = \mathbf{e}_i^\top P\mathbf{1}_k$ implying that $\sum_{j=1}^{k'} (x_j)_i^2 = \sum_{j=1}^k (x_j)_i$. Since $(x_j)_i \in \{0,1\}$, the equality above only holds if $(x_j)_i = 0$ for all $j = k+1, \ldots, k'$. As this is true for all $i \in [n]$, we have $x_j = \mathbf{0}_n$ for all $j = k+1, \ldots, k'$, implying that $X = \sum_{j=1}^k x_j x_j^\top = PP^\top$. We can now follow the derivation of the first part of the proof in the converse order to conclude that P is feasible for (QMP_2). ∎

Typical problems that fit in the framework of (QMP_2) and $(BSDP_{QMP2})$ are QMPs over the packing or partition matrices that require constraints for specific classes; see, e.g., Example 4. Another important feature of $(BSDP_{QMP2})$ is that it is possible to impose a condition on the rank of X. Corollary 3 implies that if we add the constraint $P^\top\mathbf{1}_n \ge \mathbf{1}_k$ to $(BSDP_{QMP2})$, the resulting matrix X has rank exactly k. This makes this formulation suitable for quadratic classification problems that require an exact number of classes, e.g., the (capacitated) max-k-cut problem [23].

Example 4 (the quadratic multiple knapsack problem). Consider a set of n items, each with a weight $w_i \in \mathbb{R}_+$ and a profit $p_i \in \mathbb{R}_+$, and a set of k knapsacks, each with a capacity $c_j \in \mathbb{R}_+$. Let $R = (r_{i\ell})$ denote a revenue matrix, where $r_{i\ell}$ denotes the revenue of including items i and ℓ in the same knapsack. The quadratic multiple knapsack problem (QMKP) aims at allocating each item to at most one knapsack such that we maximize the profits of the included items and their interaction revenues; see [30].

Let $P \in \{0,1\}^{n\times k}$ be a packing matrix, where $P_{ij} = 1$ if and only if item i is allocated to knapsack j, which should satisfy $P^\top w \le c$, where $w \in \mathbb{R}^n_+$ and $c \in \mathbb{R}^k_+$ denote the vector of weights and capacities, respectively. The total profit equals $\langle R, PP^\top\rangle + p^\top P\mathbf{1}_k$, where $p \in \mathbb{R}^n$ denotes the vector of item profits. Theorem 11 implies that we can model the QMKP as the following binary SDP:

$$\begin{aligned} \max \quad & \left\langle \begin{pmatrix} \mathbf{0} & \frac{1}{2}\mathbf{1}_k p^\top \\ \frac{1}{2}p\mathbf{1}_k^\top & R \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix} \right\rangle \\ \text{s.t.} \quad & P^\top w \le c,\, \operatorname{diag}(X) = P\mathbf{1}_k,\, \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix} \succeq \mathbf{0},\, X \in \{0,1\}^{n\times n},\, P \in \{0,1\}^{n\times k}. \end{aligned}$$

4. Problem-specific formulations. In this section we consider MISDP formulations of problems that do not belong to the BQPs or for which the reformulation technique differs from the ones in section 3.

4.1. The QAP as an MISDP. We present a MISDP formulation of the quadratic assignment problem (QAP) that is derived by a matrix-lifting approach. To the best of our knowledge, our QAP formulation provides the most compact convex mixed-integer formulation of the problem in the literature. The formulation is motivated by the matrix-lifting SDP relaxations of the QAP derived in [16].

The QAP is an optimization problem of the following form,

$$\min_{X\in\Pi_n} \operatorname{tr}(AXBX^\top) + \operatorname{tr}(CX^\top), \tag{19}$$

where $A, B \in \mathcal{S}^n$, $C \in \mathbb{R}^{n\times n}$, and Π_n is the set of $n\times n$ permutation matrices. The QAP is among the most difficult $\mathcal{NP}$-hard combinatorial optimization problems to solve in practice. The QAP was introduced in 1957 by Koopmans and Beckmann [35] as a model for location problems. Nowadays, the QAP is known as a generic model for various (real-life) problems.

By exploiting properties of the Kronecker product and Theorem 3, one can lift the QAP into the space of $(n^2+1)\times(n^2+1)$ $\{0,1\}$-matrix variables and obtain a BSDP formulation of the QAP; see section 3.1. Since this vector-lifting approach results in a problem formulation with a large matrix variable, we consider here a matrix-lifting approach for the QAP. Ding and Wolkowicz [16] introduce several matrix-lifting SDP relaxations of the QAP with matrix variables of order $3n$. By imposing integrality on the matrix variable X in one of these SDP relaxations, i.e., the relaxation $MSDR_0$ in [16], we obtain the following MISDP:

$$\begin{aligned} \min \quad & \langle A, Y\rangle + \langle C, X\rangle \\ \text{s.t.} \quad & \begin{pmatrix} \mathbf{I}_n & X^\top & R^\top \\ X & \mathbf{I}_n & Y \\ R & Y & Z \end{pmatrix} \succeq \mathbf{0},\, R = XB, \\ & X \in \Pi_n,\, R \in \mathbb{R}^{n\times n},\, Y, Z \in \mathcal{S}^n. \end{aligned} \tag{20}$$

Note that if B is an integer matrix, then R is also an integer matrix. However, we do not have to impose integrality on R explicitly.

The Schur complement lemma implies that the linear matrix inequality in (20) is equivalent to

$$\begin{pmatrix} \mathbf{I}_n & Y \\ Y & Z \end{pmatrix} - \begin{pmatrix} XX^\top & XR^\top \\ RX^\top & RR^\top \end{pmatrix} \succeq \mathbf{0}. \tag{21}$$

Now, we are ready to prove the following result.

PROPOSITION 6. *The MISDP* (20) *is equivalent to* (19).

Proof. Let (X,Y,Z,R) be feasible for (20). Then $XX^\top = \mathbf{I}_n$ and $\begin{pmatrix} \mathbf{I}_n - XX^\top & Y - XR^\top \\ Y - RX^\top & Z - RR^\top \end{pmatrix} \succeq \mathbf{0}$ imply that $Y = XR^\top$. Thus, $Y = XB^\top X^\top = XBX^\top$, meaning that the two objectives coincide.

Conversely, let X be feasible for (19). Define $R := XB$, $Y := XR^\top$, and $Z := RR^\top$. It trivially follows that the constraints in (20) are satisfied and that the two objective functions coincide. ∎

Many combinatorial optimization problems can be formulated as the QAP; see, e.g., [9]. We provide an example below.

Example 5 (the traveling salesman problem). Given is a complete undirected graph $K_n = (V,E)$ on n vertices and a matrix $D = (d_{ij}) \in \mathcal{S}^n$, where d_{ij} is the cost of edge $\{i,j\} \in E$. The goal of the traveling salesman problem (TSP) is to find a Hamiltonian cycle of minimum cost in K_n.

Let B be the adjacency matrix of the tour on n vertices, i.e., B is a symmetric Toeplitz matrix whose first row is $[0\ 1\ \mathbf{0}_{n-3}^\top\ 1]$. It is well known (see, e.g., [33]), that (19) with this matrix B and $A = D$ is a formulation of the TSP. Thus, a MISDP formulation of the TSP is the optimization problem (20) where the objective is replaced by $\frac{1}{2}\langle D, Y\rangle$. Another MISDP formulation of the TSP is given in subsection 4.3.

4.2. MISDP formulations of the graph partition problem. We present here various MISDP formulations of the graph partition problem (GPP). Several of the here derived formulations cannot be obtained by using results from subsections 3.1 and 3.2.

The GPP is the problem of partitioning the vertex set of a graph into a fixed number of sets, say k, of given sizes such that the sum of weights of edges joining different sets is optimized. If all sets are of equal size, then the corresponding problem is known as the k-equipartition problem (k-EP). The case of the GPP with $k = 2$ is known as the graph bisection problem (GBP). To formalize, let $G = (V,E)$ be an undirected graph on $n := |V|$ vertices and let $W := (w_{ij}) \in \mathcal{S}^n$ denote a weight matrix with $w_{ij} = 0$ if $\{i,j\} \notin E$. The graph partition problem aims to partition the vertices of G into k $(2 \le k \le n-1)$ disjoint sets $S_1,\ldots,S_k$ of specified sizes $m_1 \ge \cdots \ge m_k \ge 1$, $\sum_{j=1}^k m_j = n$ such that the total weight of edges joining different sets S_j is minimized.

For a given partition of V into k subsets, let $P = (P_{ij}) \in \{0,1\}^{n\times k}$ be the partition matrix, where $P_{ij} = 1$ if and only if $i \in S_j$ for $i \in [n]$ and $j \in [k]$. The total weight of the partition equals

$$\frac{1}{2}\mathrm{tr}\left(W(\mathbf{J}_n - PP^\top)\right) = \frac{1}{2}\mathrm{tr}(LPP^\top), \tag{22}$$

where $L := \mathrm{Diag}(W\mathbf{1}_n) - W$ is the weighted Laplacian matrix of G. The GPP can be formulated as the following QMP,

$$\min \quad \frac{1}{2}\langle L, PP^\top\rangle \quad \text{s.t.} \quad P\mathbf{1}_k = \mathbf{1}_n,\ P^\top\mathbf{1}_n = \mathbf{m},\ P \in \{0,1\}^{n\times k}, \tag{23}$$

where $\mathbf{m} = [m_1\ \ldots\ m_k]^\top$. The formulation (23) is a special case of the quadratic matrix program (QMP_2). Therefore, applying Theorem 11, the GPP can be modeled as follows:

$$
\begin{aligned}
&\min && \frac{1}{2}\langle L, X\rangle \\
(24)\qquad &\text{s.t.} && P\mathbf{1}_k = \mathbf{1}_n,\ P^\top \mathbf{1}_n = \mathbf{m},\ \operatorname{diag}(X) = \mathbf{1}_n, \\
& && \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X \end{pmatrix} \succeq \mathbf{0},\ X \in \{0,1\}^{n\times n},\ P \in \{0,1\}^{n \times k}.
\end{aligned}
$$

The doubly nonnegative relaxation of (24) is similar to the relaxation for the k-partition problem from [20]. For the k-EP and the GBP, we can derive simpler formulations by removing P from the model.

In the case of the k-EP, the QMP (23) is a special case of (QMP_1), and therefore the k-EP can be modeled as follows:

$$
\begin{aligned}
&\min && \frac{1}{2}\langle L, X\rangle \\
(25)\qquad &\text{s.t.} && \operatorname{diag}(X) = \mathbf{1}_n,\ X\mathbf{1}_n = \frac{n}{k}\mathbf{1}_n, \\
& && kX - \mathbf{J}_n \succeq \mathbf{0},\ X \in \mathcal{S}^n,\ X \in \{0,1\}^{n\times n}.
\end{aligned}
$$

This result follows from Theorem 10. An alternative proof is provided below.

Proposition 7. *Let $\mathbf{m} = \frac{n}{k}\mathbf{1}_k$. Then, the QMP (23) for the k-EP is equivalent to the BSDP (25).*

Proof. Let P be feasible for (23), where $\mathbf{m} = \frac{n}{k}\mathbf{1}_k$. We define $X := PP^\top$. The first and second constraint in (25), as well as $X \in \{0,1\}^{n\times n}$ follow by direct verification. Let p_i be the ith column of P for $i \in [k]$, then

$$
kX - \mathbf{J}_n = kPP^\top - \mathbf{1}_n\mathbf{1}_n^\top = k\sum_{i=1}^k p_i p_i^\top - \left(\sum_{i=1}^k p_i\right)\left(\sum_{i=1}^k p_i\right)^\top = \sum_{i<j}(p_i - p_j)(p_i - p_j)^\top \succeq \mathbf{0}.
$$

Conversely, let X be feasible for (25). Then, it follows from Theorem 1 and Proposition 1 that there exist $x_i \in \{0,1\}^n$, $i \in [r]$, $k \geq r$ such that $X = \sum_{i=1}^r x_i x_i^\top$, where $\sum_{i=1}^r x_i = \mathbf{1}_n$. Since the constraint $X\mathbf{1}_n = \frac{n}{k}\mathbf{1}_n$ is invariant under permutation of rows and columns of X, we have that the sum of the elements in each row and column of the block matrix $\mathbf{J}_{n_1} \oplus \cdots \oplus \mathbf{J}_{n_r}$ equals n/k. From this it follows that $r = k$ and $\mathbf{1}_n^\top x_i = n/k$ for $i \in [k]$. It is easy to verify that $P := [x_1\ \dots\ x_k] \in \{0,1\}^{n\times k}$ is feasible for (23). Since the two objectives coincide, the result follows. ∎

The next result shows that the MISDP (24) also simplifies for the GBP. It has to be noted, however, that the GBP is not a special case of (QMP_1).

Proposition 8. *Let $\mathbf{m} = [m_1\ n - m_1]^\top$, $1 \leq m_1 \leq n/2$. Then, the QMP (23) for the GBP is equivalent to the following BSDP:*

$$
\begin{aligned}
&\min && \frac{1}{2}\langle L, X\rangle \\
(26)\qquad &\text{s.t.} && \operatorname{diag}(X) = \mathbf{1}_n,\ \langle \boldsymbol{J}_n, X\rangle = m_1^2 + (n-m_1)^2, \\
& && 2X - \boldsymbol{J}_n \succeq \mathbf{0},\ X \in \mathcal{S}^n,\ X \in \{0,1\}^{n\times n}.
\end{aligned}
$$

Proof. Let P be feasible for (23). We define $X := PP^\top$. The first and second constraints in (26) follow by direct verification. Let p_i be the ith column of P for $i \in [2]$, then

$$
2X - \mathbf{J}_n = 2PP^\top - \mathbf{1}_n\mathbf{1}_n^\top = 2\sum_{i=1}^2 p_i p_i^\top - \left(\sum_{i=1}^2 p_i\right)\left(\sum_{i=1}^2 p_i\right)^\top = (p_1 - p_2)(p_1 - p_2)^\top \succeq \mathbf{0}.
$$

Conversely, let X be feasible for (26). Then, it follows from Theorem 1 and Corollary 2 that there exist $x_1, x_2 \in \{0,1\}^n$ such that $X = x_1x_1^\top + x_2x_2^\top$, where $x_1 + x_2 = \mathbf{1}_n$. Note that X cannot have rank one or zero for $1 \leq m_1 < n$. From $\langle \mathbf{J}_n, X\rangle = m_1^2 + (n-m_1)^2$, it follows that $\mathbf{1}_n^\top x_1 = m_1$ or $\mathbf{1}_n^\top x_1 = n - m_1$. Without loss of generality, we assume that $\mathbf{1}^\top x_1 = m_1$. Clearly, $P := [x_1 \; x_2]$ is feasible for (23). Moreover, the two objective functions coincide. ∎

In the remainder of this section, we derive yet another alternative MISDP formulation of the GPP, different from (24). For that purpose we notice that the GPP can also be formulated as a QMP of the following form:

$$(QMP_3) \qquad \begin{array}{rl} \min & \operatorname{tr}(P^\top Q_0 P) + \operatorname{tr}(PC_0P^\top) + 2\operatorname{tr}(B_0^\top P) + d_0 \\ \text{s.t.} & \operatorname{tr}(P^\top Q_i P) + \operatorname{tr}(PC_iP^\top) + 2\operatorname{tr}(B_i^\top P) + d_i \leq 0 \quad \forall i \in [m], \\ & P \in \mathbb{R}^{n\times k}, \end{array}$$

where $Q_i \in \mathcal{S}^n$, $C_i \in \mathcal{S}^k$, $B_i \in \mathbb{R}^{n\times k}$, $d_i \in \mathbb{R}$ for $i = 0,1,\ldots,m$. Note that (QMP_2) is a special case of (QMP_3). Examples of problems that are of this form are quadratic problems with orthogonality constraints; see, e.g., [3]. The GPP can be formulated as follows (see e.g., [15]):

$$(27) \qquad \begin{array}{rl} \min & \frac{1}{2}\langle L, PP^\top\rangle \\ \text{s.t.} & P^\top \mathbf{1}_n = \mathbf{m},\; P^\top P = \operatorname{Diag}(\mathbf{m}), \\ & \operatorname{diag}(PP^\top) = \mathbf{1}_n,\; P \geq \mathbf{0},\; P \in \mathbb{R}^{n\times k}. \end{array}$$

To reformulate (27) as an MISDP we introduce matrices $X_1 \in \mathcal{S}^n$ and $X_2 \in \mathcal{S}^k$ such that $X_1 = PP^\top$ and $X_2 = P^\top P$ and relax these matrix equalities to the linear matrix inequalities (LMIs) $X_1 - PP^\top \succeq \mathbf{0}$ and $X_2 - P^\top P \succeq \mathbf{0}$, respectively. These can be rewritten as

$$\begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X_1 \end{pmatrix} \succeq \mathbf{0} \quad \text{and} \quad \begin{pmatrix} \mathbf{I}_n & P \\ P^\top & X_2 \end{pmatrix} \succeq \mathbf{0}.$$

After introducing the constraints $\operatorname{diag}(X_1) = \mathbf{1}_n$ and $X_2 = \operatorname{Diag}(\mathbf{m})$, we obtain the following MISDP:

$$(28) \qquad \begin{array}{rl} \min & \frac{1}{2}\langle L, X_1\rangle \\ \text{s.t.} & P\mathbf{1}_k = \mathbf{1}_n,\; \operatorname{diag}(X_1) = \mathbf{1}_n,\; X_2 = \operatorname{Diag}(\mathbf{m}), \\ & \begin{pmatrix} \mathbf{I}_k & P^\top \\ P & X_1 \end{pmatrix} \succeq \mathbf{0},\; \begin{pmatrix} \mathbf{I}_n & P \\ P^\top & X_2 \end{pmatrix} \succeq \mathbf{0},\; X_1 \in \mathcal{S}^n,\; X_2 \in \mathcal{S}^k,\; P \in \{0,1\}^{n\times k}. \end{array}$$

We prove below that (28) is an exact formulation of the GPP.

PROPOSITION 9. *The MISDP* (28) *is an exact formulation of the GPP.*

Proof. We prove the result by showing the equivalence between (27) and (28).

Let $P \in \mathbb{R}^{n\times k}$ be feasible for (27). Then, it follows from $\operatorname{diag}(PP^\top) = \mathbf{1}_n$ that $(PP^\top)_{ii} = \sum_{j=1}^k P_{ij}^2 = 1$ for $i \in [n]$. From this and $P \geq \mathbf{0}$, we obtain $0 \leq P_{ij} \leq 1$ for all $i \in [n]$, $j \in [k]$. From $P^\top \mathbf{1}_n = \mathbf{m}$ it follows that $\sum_{i,j} P_{ij} = n$ and from $P^\top P = \operatorname{Diag}(\mathbf{m})$ that $\operatorname{tr}(P^\top P) = n$, and thus $\sum_{i,j} P_{ij}^2 = n$. Therefore, $P_{ij} \in \{0,1\}$ for all $i \in [n]$, $j \in [k]$. The equality $\operatorname{diag}(PP^\top) = \mathbf{1}_n$ then implies that $P\mathbf{1}_k = \mathbf{1}_n$. It follows from the discussion prior to the proposition that $X_1 := PP^\top$ and $X_2 := P^\top P$ are feasible for (28).

Conversely, let X_1, X_2, and P be feasible for (28). From $P \in \{0,1\}^{n\times k}$ and $P\mathbf{1}_k = \mathbf{1}_n$ it follows that $\mathrm{diag}(PP^\top) = \mathbf{1}_n$. From $X_2 - P^\top P \succeq \mathbf{0}$ and $\mathbf{1}_k^\top(X_2 - P^\top P)\mathbf{1}_k = 0$ it follows that $(X_2 - P^\top P)\mathbf{1}_k = \mathbf{0}$ and thus $P^\top \mathbf{1}_n = \mathbf{m}$. Moreover, we have $(P^\top P)_{ii} = \sum_{j=1}^n P_{ji}^2 = \sum_{j=1}^n P_{ji} = m_i$ for $i \in [k]$, implying that $\mathrm{diag}(P^\top P) = \mathbf{m}$. Finally, we have $X_2 - P^\top P = \mathrm{Diag}(\mathbf{m}) - P^\top P \succeq \mathbf{0}$, where it follows from above that the latter matrix has a diagonal of zeros. Thus, we must have $P^\top P = \mathrm{Diag}(\mathbf{m})$, which concludes the proof. ∎

The MISDP (28) has two LMIs and requires integrality constraints only on a matrix of size $n \times k$, while (24) has only one LMI and asks for integrality on matrices of size $n \times n$ and $n \times k$.

4.3. MISDP formulations via association schemes. Association schemes provide a unifying framework for the treatment of problems in several branches of mathematics, including algebraic graph theory, coding theory, and optimization; see, e.g., [14, 54]. For the background on association schemes, we refer to [6, 25]. De Klerk, Filho, and Pasechnik [32] introduce a framework for deriving SDP relaxations of optimization problems on graphs by using association schemes. By exploiting a similar approach, one can obtain exact formulations of discrete problems via association schemes. We provide two examples below.

Example 6 (the TSP). Let us reconsider the TSP; see Example 5. The following MISDP is an exact model of the TSP,

$$
\begin{aligned}
\min \quad & \frac{1}{2}\langle D, X_1\rangle \\
\text{s.t.} \quad & \mathbf{I}_n + \sum_{i=1}^r X_i = \mathbf{J}_n,\ \mathbf{I}_n + \sum_{i=1}^r \cos\left(\frac{2ij\pi}{n}\right) X_i \succeq \mathbf{0}\ \forall j \in [r], \\
& X_1 = X_1^\top \in \{0,1\}^{n\times n},\ X_i \geq \mathbf{0},\ X_i \in \mathcal{S}^n\ \forall i = 2,\dots,r,
\end{aligned} \tag{29}
$$

where $r = \lfloor n/2 \rfloor$ and n is odd. A similar model can be derived for n even. One can show that (29) is an exact formulation of the TSP by exploiting the ISDP formulation of the TSP by Cvetković, Čangalović, and Kovačević-Vujčić and [33, Theorem 4.1].

Example 7 (the k-equipartition problem). Let $n, k, m \in \mathbb{Z}_+$, and D be a nonnegative symmetric matrix of order n, where $n = mk$. The k-EP can be formulated as finding a complete regular k-partite subgraph on n vertices in K_{mk} of minimum weight; see also subsection 4.2. It is not difficult to show that the following MISDP is an exact model of the k-EP:

$$
\begin{aligned}
\min \quad & \langle D, X_1\rangle \\
\text{s.t.} \quad & (m-1)\mathbf{I}_n - X_2 \succeq \mathbf{0},\ (k-1)\mathbf{I}_n - X_1 + (k-1)X_2 \succeq \mathbf{0}, \\
& \mathbf{I}_n + X_1 + X_2 = \mathbf{J}_n,\ X_1, X_2 \in \mathcal{S}^n,\ X_1 \geq \mathbf{0},\ X_2 \in \{0,1\}^{n\times n}.
\end{aligned} \tag{30}
$$

Note that the MISDP (30) has two LMIs and the BSDP (25) only one. Moreover, observe that one may replace in (30) the constraint $(m-1)\mathbf{I}_n - X_2 \succeq \mathbf{0}$ by $X_2\mathbf{1}_n = (m-1)\mathbf{1}_n$ and obtain an MISDP for the k-EP with only one PSD constraint.

The examples show that one can impose integrality conditions to *only one* of the matrices in the models and obtain an exact problem formulation.

4.4. MISDP formulations beyond binarity. Almost all problem formulations discussed before involve variables restricted to $\{0,1\}$. We now consider several problems possessing a MISDP formulation where the integer variables are not necessarily binary.

Example 8 (the integer matrix completion problem). A well-known problem in data analysis is the problem of low-rank matrix completion. Suppose a partially observed data matrix is given, i.e., let $\Omega \subseteq [n] \times [m]$ denote the set of observed entries and let $D \in \mathbb{R}^{n\times m}$ denote a given data matrix that has its support on Ω. The goal of the low-rank matrix completion problem is to find a minimum rank matrix $X \in \mathbb{R}^{n\times m}$ such that X coincides with D on the set Ω; e.g., see [1].

Since minimizing $\operatorname{rank}(X)$ leads to a nonconvex problem, a related but tractable alternative is given by

$$\min_{X\in\mathbb{R}^{n\times m}} \quad ||X||_* \quad \text{s.t.} \quad X_{ij} = D_{ij} \quad \text{for all } (i,j) \in \Omega,$$

where $||X||_*$ denotes the nuclear norm of X, i.e., the sum of its singular values. Recht, Fazel, and Parrilo [51] have shown that this problem can be modeled as an SDP. A possible generalization would be to require the entries in X to be integer; see, e.g., [1]. Given a discrete set $S \subseteq \mathbb{Z}$, this leads to the following ISDP:

$$\begin{aligned} \min \quad & \langle \mathbf{I}_n, Z_1\rangle + \langle \mathbf{I}_m, Z_2\rangle \\ \text{s.t.} \quad & \begin{pmatrix} Z_1 & X \\ X^\top & Z_2 \end{pmatrix} \succeq \mathbf{0},\ X_{ij} = D_{ij} \text{ for all } (i,j)\in\Omega,\ X_{ij}\in S \text{ for all } (i,j)\notin\Omega, \end{aligned}$$

which models the integer matrix completion problem [1].

Example 9 (the sparse integer least squares problem). In the integer least squares problem we are given a matrix $M \in \mathbb{R}^{n\times k}$ and a column $b \in \mathbb{R}^n$ and we seek the closest point to b in the lattice spanned by the columns of M. Pia and Zhou [49] consider the related sparse integer least squares (SILS) problem:

$$\min \quad \frac{1}{n}||Mx-b||_2^2 \quad \text{s.t.} \quad x \in \{0,\pm1\}^k,\ ||x||_0 \le K. \tag{31}$$

The SILS problem has applications in, among others, multiuser detection and sensor networks; see [49] and the references therein. Now, consider the following ternary SDP:

$$\begin{aligned} \min \quad & \frac{1}{n}\left\langle \begin{pmatrix} 1 & x^\top \\ x & X\end{pmatrix}, \begin{pmatrix} b^\top b & -b^\top M \\ -M^\top b & M^\top M\end{pmatrix}\right\rangle \\ \text{s.t.} \quad & \operatorname{tr}(X) \le K,\ \operatorname{diag}(X) = y_1 + y_2,\ x = y_1 - y_2, \\ & \begin{pmatrix} 1 & x^\top \\ x & X\end{pmatrix} \succeq \mathbf{0},\ \begin{pmatrix} 1 & x^\top \\ x & X\end{pmatrix} \in \{0,\pm1\}^{(k+1)\times(k+1)},\ y_1, y_2 \in \mathbb{R}^n_+. \end{aligned} \tag{32}$$

It is not difficult to verify that if x is a solution to (31), then x, $X = xx^\top$, $y_1 = \max(x, \mathbf{0})$, and $y_2 = \max(-x, \mathbf{0})$ is feasible for (32) with the same objective value. Conversely, if (x, X, y_1, y_2) is feasible for (32), it follows from Proposition 5 that $x = y_1 - y_2$ is a solution to (31) with the same objective value.

5. Conclusions. In this paper we showed that the class of MISDPs embodies a rich structure, allowing for compact formulations of many well-known discrete optimization problems. Due to the recent progress in computational methods for solving

MISDPs [22, 29, 31, 34, 43, 44], these formulations can be exploited to obtain alternative methods for solving the problems to optimality.

As most problems are naturally encoded using binary or ternary variables, we started our research with a study on the general theory related to PSD $\{0,1\}$-, $\{\pm 1\}$-, and $\{0,\pm 1\}$-matrices. Section 2 provides a comprehensive overview on this matter, including known and new results. In particular, we presented a combinatorial, polyhedral, set-completely positive and integer hull description of the set of PSD $\{0,1\}$-matrices bounded by a certain rank; see section 2.1. Several of these results are extended to matrices having entries in $\{\pm 1\}$ and $\{0,\pm 1\}$.

Based on these matrix results, in particular, Theorems 1–3 and Corollary 3, we derived a generic approach to model BQPs as BSDPs. We derived a BSDP for the class of binary QCQPs (see $(BSDP_{QCQP})$), and for two types of binary QMPs, see $(BSDP_{QMP1})$ and $(BSDP_{QMP2})$. These results are widely applicable to a large number of discrete optimization problems; see also the examples in section 3.

We moreover considered problem-specific MISDP formulations that are derived in a different way than through this generic approach. We provided compact MISDP formulations of the QAP (see (20)), and various variants of the GPP; see (24), (25), and (26). We derived several MISDP formulations of discrete optimization problems that can be modeled using association schemes; see section 4.3. We also considered problems that have discrete but nonbinary variables; see Examples 8 and 9.

Given the wide range of discrete optimization problems for which we derived new formulations based on MISDP, we expect more problems to allow for such representations. It is also interesting to study the behavior of MISDP solvers on the presented formulations to see whether this leads to competitive solution approaches for the considered problems.

Acknowledgment. We thank two anonymous reviewers for providing us with insightful feedback on an earlier version of this manuscript.

REFERENCES

[1] Proceedings of KDD Cup and Workshop, 2007, ACM, New York, https://www.cs.uic.edu/∼liub/KDD-cup-2007/proceedings.html.

[2] M. Anjos and H. Wolkowicz, *Strengthened semidefinite relaxations via a second lifting for the max-cut problem*, Discrete Appl. Math., 119 (2002), pp. 79–106.

[3] K. Anstreicher and H. Wolkowicz, *On Lagrangian relaxation of quadratic matrix constraints*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 41–55, https://doi.org/10.1137/S0895479898340299.

[4] A. Beck, *Quadratic matrix programming*, SIAM J. Optim., 17 (2007), pp. 1224–1238.

[5] E. Bell, *Exponential numbers*, Amer. Math. Monthly, 41 (1934), pp. 411–419.

[6] A. Brouwer and W. Haemers, *Association schemes*, in Handbook of Combinatorics, Elsevier Science, Amsterdam, 1995, pp. 747–771.

[7] T. Bulhões, A. Pessoa, F. Protti, and E. Uchoa, *On the complete set packing and set partitioning polytopes: Properties and rank* 1 *facet*, Oper. Res. Lett., 46 (2018), pp. 389–392.

[8] S. Burer, *On the copositive representation of binary and continuous nonconvex quadratic programs*, Math. Program., 120 (2009), pp. 479–495, https://doi.org/10.1007/s10107-008-0223-z.

[9] R. E. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*, SIAM, Philadelphia, 2009.

[10] A. Cerveira, A. Agra, F. Bastos, and J. Gromicho, *A new branch and bound method for a discrete truss topology design problem*, Comput. Optim. Appl., 54 (2013), pp. 163–187.

[11] V. Chagas, *Exact Algorithms for the Quadratic Bin Packing Problem*, Ph.D. thesis, Universidade Estadual de Campinas, Campinas, Brazil, 2021.

[12] S. Chopra and M. Rao, *The partition problem*, Math. Program., 59 (1993), pp. 87–115.

[13] D. Cvetković, M. Čangalović, and V. Kovačević-Vujčić, *Semidefinite programming methods for the symmetric traveling salesman problem*, in Integer Programming and Combinatorial Optimization (IPCO 1999), Lecture Notes in Comput. Sci. 1610, G. Cornujóls, R. Burkard, and G. Woeginger, eds., Springer, Berlin, 1999.
[14] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Res. Rep. Suppl., 10 (1973).
[15] Y. Ding, D. Ge, and H. Wolkowicz, *On equivalence of semidefinite relaxations for quadratic matrix programming*, Math. Oper. Res., 36 (2011), pp. 88–104, http://www.jstor.org/stable/23011949 (accessed 2023-01-21).
[16] Y. Ding and H. Wolkowicz, *A low-dimensional semidefinite relaxation for the quadratic assignment problem*, Math. Oper. Res., 34 (2009), pp. 1008–1022.
[17] B. Duarte, *Exact optimal designs of experiments for factorial models via mixed-integer semidefinite programming*, Mathematics, 11 (2023), 854.
[18] I. Dukanovic and F. Rendl, *Semidefinite programming relaxations for graph coloring and maximal clique problems*, Math. Program., 109 (2007), pp. 345–365.
[19] A. Eisenblätter, *Frequency Assignment in GSM Networks: Models, Heuristics, and Lower Bounds*, Ph.D. thesis, Technische Universität Berlin, Berlin, 2001.
[20] J. Fairbrother and A. N. Letchford, *Projection results for the k-partition problem*, Discrete Optim., 26 (2017), pp. 97–111.
[21] T. Gally and M. Pfetsch, *Computing restricted isometry constants via mixed-integer semidefinite programming*, Optimization Online, (2016), 13907, http://www.optimization-online.org/DB_FILE/2016/04/5395.pdf.
[22] T. Gally, M. Pfetsch, and S. Ulbrich, *A framework for solving mixed-integer semidefinite programs*, Optim. Methods Softw., 33 (2018), pp. 594–632.
[23] D. Gaur, R. Krishnamurti, and R. Kohli, *The capacitated max k-cut problem*, Math. Program., 115 (2008), pp. 65–72.
[24] W. Gil-González, A. Molina-Cabrera, O. Montoya, and L. Grisales-Noreña, *An MI-SDP model for optimal location and sizing of distributed generators in DC grids that guarantees the global optimum*, Appl. Sci., 10 (2020), 7681.
[25] C. Godsil, *Algebraic Combinatorics*, Chapman & Hall, New York, 1993.
[26] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Vol. 2, Springer, Berlin, 1988.
[27] M. Grötschel and Y. Wakabayashi, *Facets of the clique partitioning polytope*, Math. Program., 47 (1990), pp. 367–387.
[28] C. Helmberg, *Fixing variables in semidefinite relaxations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 952–969.
[29] H. Hijazi, G. Wang, and C. Coffrin, *coin-or/Gravity,* https://github.com/coin-or/Gravity/tree/GravitySDP (2018).
[30] A. Hiley and B. Julstrom, *The quadratic multiple knapsack problem and three heuristic approaches to it*, in Genetic and Evolutionary Computation Conference (GECCO), P. Heggernes, ed., ACM, New York, 2006, pp. 547–552.
[31] C. Hojny and M. Pfetsch, *Handling symmetries in mixed-integer semidefinite programs*, in Integration of Constraint Programming, Artificial Intelligence, and Operations Research, Springer, Cham, Switzerland, 2023, pp. 69–78.
[32] E. de Klerk, F. O. Filho, and D. Pasechnik, *Relaxations of combinatorial problems via association schemes*, in Handbook on Semidefinite, Conic and Polynomial Optimization, Internat. Ser. Oper. Res. Management Sci., M. F. Anjos and J. B. Lasserre, eds., Springer, New York, 2012, pp. 171–199, https://ideas.repec.org/h/spr/isochp/978-1-4614-0769-0_7.html.
[33] E. de Klerk, D. V. Pasechnik, and R. Sotirov, *On semidefinite programming relaxations of the traveling salesman problem*, SIAM J. Optim., 19 (2008), pp. 1559–1573, https://doi.org/10.1137/070711141.
[34] K. Kobayashi and Y. Takano, *A branch-and-cut algorithm for solving mixed-integer semidefinite optimization problems*, Comput. Optim. Appl., 75 (2020), pp. 493–513.
[35] T. C. Koopmans and M. Beckmann, *Assignment problems and the location of economic activities*, Econometrica, 25 (1957), pp. 53–76.
[36] M. Kočvara, *Truss topology design with integer variables made easy*, Optimization Online, (2010), 11135, https://optimization-online.org/2010/05/2614/.
[37] O. Kuryatnikova, R. Sotirov, and J. Vera, *The maximum k-colorable subgraph problem and related problems*, INFORMS J. Comput., 34 (2021), pp. 656–669.
[38] M. Laurent, S. Poljak, and F. Rendl, *Connections between semidefinite relaxations of the max-cut and stable set problems*, Math. Program., 77 (1997), pp. 225–246.
[39] M. Laurent and F. Rendl, *Semidefinite programming and integer programming*, Handb. Oper. Res. Manag. Sci., 12 (2005), pp. 393–514.

[40] A. Letchford and M. Sørensen, *Binary positive semidefinite matrices and associated integer polytopes*, Math. Program. Ser. A, 131 (2012), pp. 253–271.
[41] F. Lieder, F. Rad, and F. Jarre, *Unifying semidefinite and set-copositive relaxations of binary problems and randomization techniques*, Comput. Optim. Appl., 61 (2015), pp. 669–688.
[42] S. Mars, *Mixed-integer Semidefinite Programming with an Application to Truss Topology Design*, Ph.D. thesis, FAU Erlangen-Nürnberg, Erlangen, Germany, 2013.
[43] F. Matter and M. Pfetsch, *Presolving for mixed-integer semidefinite optimization*, INFORMS J. Optim., 5 (2022), pp. 131–154.
[44] F. de Meijer and R. Sotirov, *The Chvátal-Gomory Procedure for Integer SDPs with Applications in Combinatorial Optimization*, preprint, arXiv:2201.10224v2, 2023.
[45] P. Meurdesoif, *Strengthening the Lovász bound for graph coloring*, Math. Program., 102 (2005), pp. 577–588, https://doi.org/10.1007/s101070100246.
[46] G. Narasimhan, *The Maximum k-Colorable Subgraph Problem*, Ph.D. thesis, University of Wisconsin-Madison, Madison, WI, 1989.
[47] M. Oosten, J. Rutten, and F. Spieksma, *The clique partitioning problem: Facets and patching facets*, Networks, 38 (2001), pp. 209–226.
[48] M. Padberg, *The Boolean quadric polytope: Some characteristics, facets and relatives*, Math. Program., 45 (1989), pp. 139–172.
[49] A. D. Pia and D. Zhou, *An SDP Relaxation for the Sparse Integer Least Square Problem*, preprint, arXiv:2203.02607, 2023.
[50] M. Pilanci, M. Wainwright, and L. El Ghaoui, *Sparse learning via Boolean relaxations*, Math. Program., 151 (2015), pp. 63–87.
[51] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501, https://doi.org/10.1137/070697835.
[52] F. Rendl, *Semidefinite relaxations for integer programming*, in 50 Years of Integer Programming 1958-2008, Springer, Berlin, 2010, pp. 687–726.
[53] A. Schrijver, *A comparison of the Delsarte and Lovász bounds*, IEEE Trans. Inform. Theory, 25 (1979), pp. 425–429.
[54] A. Schrijver, *New code upper bounds from the Terwilliger algebra and semidefinite programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 2859–2866.
[55] N. Shor, *Quadratic optimization problems*, Sov. J. Comput. Syst. Sci., 25 (1987), pp. 1–11.
[56] L. Wei, A. Atamtürk, A. Gómez, and S. Küçükyavuz, *On the convex hull of convex quadratic optimization problems with indicators*, Math. Program., to appear.
[57] K. Yonekura and Y. Kanno, *Global optimization of robust truss topology via mixed integer semidefinite programming*, Optim. Eng., 11 (2010), pp. 355–379.
[58] X. Zheng, H. Chen, Y. Xu, Z. Li, Z. Lin, and Z. Liang, *A mixed-integer SDP solution to distributionally robust unit commitment with second order moment constraints*, CSEE J. Power Energy Syst., 6 (2020), pp. 374–383.