

Facial Reduction for Symmetry Reduced Semidefinite Programs

Hao Hu* Renata Sotirov† Henry Wolkowicz‡

December 21, 2019

Abstract

We consider both facial and symmetry reduction techniques for semidefinite programming, **SDP**. We show that the two together fit surprisingly well in an alternating direction method of multipliers, **ADMM**, approach. The combination of facial and symmetry reduction leads to a significant improvement in both numerical stability and running time for both the **ADMM** and interior point approaches.

We test our method on various doubly nonnegative, **DNN**, relaxations of hard combinatorial problems including quadratic assignment problems with sizes of more than $n = 500$.

Keywords: Semidefinite programming, **SDP**, group symmetry, facial reduction, quadratic assignment problem, vertex separator problem.

AMS subject classifications: 90C22, 90C25

Contents

1	Introduction	2
1.1	Outline	3
2	Background	3
2.1	Semidefinite programming	3
2.1.1	Strict feasibility and facial reduction	3
2.2	Group invariance and symmetry reduction	4
2.2.1	First symmetry reduction using $X = \mathcal{B}^*(x)$	6
2.2.2	Second symmetry reduction to block diagonal form using $X = Q\tilde{\mathcal{B}}^*(x)Q^T$	7
3	Facial reduction for the symmetry reduced program	7
3.1	Rank preserving	8
3.2	Facial reduction and exposing vectors	8
3.3	Simplifications	10
4	The alternating direction method of multipliers, ADMM	12
4.1	Augmented Lagrangian	12
4.2	On solving the \tilde{R} -subproblem	13
4.3	On solving the x -subproblem	13

*Department of Combinatorics and Optimization Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; h92hu@uwaterloo.ca.

†Department of Econometrics and Operations Research, Tilburg University, The Netherlands; r.sotirov@uvt.nl

‡Department of Combinatorics and Optimization Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; Research supported by The Natural Sciences and Engineering Research Council of Canada; www.math.uwaterloo.ca/~hwolkowi.

5	Numerical results	15
5.1	The quadratic assignment problem (QAP)	15
5.1.1	Background for the QAP	15
5.1.2	On solving QAP with ADMM	19
5.1.3	Numerical results for the QAP	22
5.2	The graph partition problem (GP)	25
5.2.1	The general GP	25
5.2.2	The vertex separator problem (VSP) and min-cut (MC)	27
6	Conclusion	29
	Index	33
	Bibliography	36

List of Tables

5.1	Lower and upper bounds for different QAP instances.	24
5.2	Esc instances	25
5.3	Graphs and partitions	27
5.4	Numerical results for the graph 3-partition.	27
5.5	The Queen graphs and partitions	28
5.6	The vertex separator problem on the Queen graphs	29
5.7	The min-cut problem on the Queen graphs	29

1 Introduction

We consider two reduction techniques, facial and symmetry reduction, for semidefinite programming, **SDP**. We see that the exposing vector approach for facial reduction, **FR**, moves naturally onto the symmetry reduction. We show that the combination of the two reductions fits surprisingly well within an alternating direction method of multipliers, **ADMM**. The combination of facial and symmetry reduction leads to a significant improvement in both numerical stability and running time for both the **ADMM** and interior point approaches. We test our method on various doubly nonnegative, **DNN**, relaxations of hard combinatorial problems including quadratic assignment problems with sizes of more than $n = 500$.

Semidefinite programming can be viewed as an extension of linear programming where the nonnegative orthant is replaced by the cone of positive semidefinite matrices. Although there are many algorithms for solving semidefinite programs, they currently do not scale well and often do not provide high accuracy solutions. Symmetry reduction is a methodology, pioneered by Schrijver [41], that exploits symmetries in the data matrices that allows for the problem size to be reduced, often significantly.

Without loss of generality, we consider the case where the primal problem has a finite optimal value. Then for linear programming, strong duality holds for both the primal and the dual problems. But, this is *not* the case for **SDP**, where the primal and/or the dual can be unattained, and one can even have a positive *duality gap* between the primal and dual optimal values. The usual constraint qualification to guarantee strong duality is Slater's condition, strict feasibility. Failure of Slater's condition may lead to theoretical and numerical problems when solving the **SDP**. Facial reduction, introduced by Borwein and Wolkowicz [5–7], addresses this issue by projecting the *minimal face* of the **SDP** into a lower dimensional space.

In this paper, we assume that we know how to do facial reduction and symmetry reduction separately for the input **SDP** instance. Under this assumption, we show that it is possible to implement facial reduction to the symmetry reduced **SDP**. The obtained **SDP** is both facially reduced and symmetry reduced, and it can be solved in a numerically stable manner by interior point methods. Finally, we also show that this approach allows us to solve the facially and symmetry reduced program using an alternating direction method of multipliers approach, **ADMM**. As a consequence, we are able to solve some very large *doubly nonnegative*, **DNN**, relaxations for highly symmetric instances of certain hard combinatorial problems, and do so in a reasonable amount of time.

1.1 Outline

In Section 2 we provide the background on using substitutions to first obtain facial reduction and then symmetry and block diagonal symmetry reduction. In Section 3 we show how to apply facial reduction to the symmetry reduced **SDP**, and we also provide conditions such that the obtained **SDP** is strictly feasible. In Section 4 we show that the facially and symmetry reduced **SDP** can be solved via an alternating direction method of multipliers, **ADMM**, approach in an efficient way. In Section 5 we apply our result to two classes of problems: the quadratic assignment and graph partition problems. Concluding comments are in Section 6.

2 Background

2.1 Semidefinite programming

The *semidefinite program*, **SDP**, in standard form is

$$p_{\text{SDP}}^* := \min\{\langle C, X \rangle \mid \mathcal{A}(X) = b, X \succeq 0\}, \quad (2.1)$$

where the linear transformation $\mathcal{A} : \mathbb{S}^n \rightarrow \mathbb{R}^m$ maps real $n \times n$ symmetric matrices to \mathbb{R}^m , and $X \in \mathbb{S}_+^n$ is positive semidefinite. In the case of a *doubly nonnegative*, **DNN**, relaxation, nonnegativity constraints, $X \succeq 0$, are added to (2.1). Without loss of generality, we assume that \mathcal{A} is onto. We let

$$\mathcal{F}_X = \{X \succeq 0 \mid \mathcal{A}(X) = b\}$$

denote the feasible set of (2.1). Note that the linear equality constraint is equivalent to

$$\mathcal{A}(X) = (\langle A_i, X \rangle) = (b_i) \in \mathbb{R}^m,$$

for some $A_i \in \mathbb{S}^n, i = 1, \dots, m$. The *dual program* utilizes the *adjoint* transformation $\mathcal{A}^* : \mathbb{R}^m \rightarrow \mathbb{S}^n$

$$d_{\text{SDP}}^* := \max\{\langle b, y \rangle \mid \mathcal{A}^*(y) \preceq C\}, \quad (2.2)$$

where the concrete expression for the adjoint is

$$\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i.$$

2.1.1 Strict feasibility and facial reduction

The standard constraint qualification to guarantee strong duality¹ for the primal **SDP** is the *Slater constraint qualification* (strict feasibility)

$$\exists \hat{X} : \mathcal{A}(\hat{X}) = b, \hat{X} \succ 0,$$

¹Strong duality for the primal means a zero duality gap, $p_{\text{SDP}}^* = d_{\text{SDP}}^*$, and dual attainment.

where $\hat{X} \succ 0$ denotes positive definiteness. ($\hat{X} \succeq 0$ denotes positive semidefiniteness.) For many problems where strict feasibility fails, one can exploit structure and facially reduce the problem to obtain strict feasibility, see e.g., [5,6] for the theory and [7] for the *facial reduction algorithm*. A survey with various views of facial reduction is given in [17]. Facial reduction means that there exists a full column rank matrix $V \in \mathbb{R}^{n \times r}$, $r < n$, and the corresponding adjoint of the linear transformation $\mathcal{V} : \mathbb{S}^n \rightarrow \mathbb{S}^r$ given in

$$\mathcal{V}^*(R) = VRV^T, R \in \mathbb{S}^r,$$

such that the *substitution* $X = \mathcal{V}^*(R)$ results in the *equivalent, regularized, smaller dimensional*, problem

$$p_{\mathbf{SDP}}^* = \min\{\langle V^T CV, R \rangle \mid \langle V^T A_i V, R \rangle = b_i, i = 1, \dots, m, R \in \mathbb{S}_+^r\}. \quad (2.3)$$

Strict feasibility holds for (2.3). The cone $V\mathbb{S}_+^r V^T$ is the *minimal face* of the **SDP**, i.e., the smallest face of \mathbb{S}_+^n that contains the feasible set, \mathcal{F}_X . And

$$\text{range}(V) = \text{range}(X), \forall X \in \text{relint}(\mathcal{F}_X).$$

If $U \in \mathbb{R}^{n \times n-r}$ with $\text{range}(U) = \text{null}(V^T)$, then $W := UU^T$ is an *exposing vector* for the minimal face, i.e.,

$$X \text{ feasible} \implies WX = 0.$$

Let \mathcal{F}_R denote the feasible set for (2.3). We emphasize the following constant rank result for the facial reduction substitution:

$$R \in \mathcal{F}_R, \text{rank}(R) = r \iff X = \mathcal{V}^*(R) \in \mathcal{F}_X, \text{rank}(X) = r.$$

Typical facial reduction algorithms for finding the minimal face require at most $n - 1$ steps that involve finding exposing vectors. Moreover, at each iteration the dimension is strictly reduced at least one redundant linear constraint can be discarded, [44].

2.2 Group invariance and symmetry reduction

We now find a *substitution* using the adjoint linear transformation \tilde{B}^* in (2.9) below, that obtains the symmetry reduction to block diagonal form. We first look at the procedure for simplifying an **SDP** that is invariant under the action of a symmetry group. This approach was introduced by Schrijver [41]; see also the survey [2]. The appropriate algebra isomorphism follows from the Artin-Wedderburn theory [49]. A more general framework is given in the thesis [35]. More details can be found in e.g., [12,19,21,47].

Let \mathcal{G} be a nontrivial group of permutation matrices of size n . The *commutant*, $A_{\mathcal{G}}$, (or *centralizer ring*) of \mathcal{G} is defined as the subspace

$$A_{\mathcal{G}} := \{X \in \mathbb{R}^{n \times n} \mid PX = XP, \forall P \in \mathcal{G}\}.$$

Thus, $A_{\mathcal{G}}$ is the set of matrices that are self-permutation-congruent for all $P \in \mathcal{G}$. An equivalent definition of the commutant is

$$A_{\mathcal{G}} = \{X \in \mathbb{R}^{n \times n} \mid \mathcal{R}_{\mathcal{G}}(X) = X\},$$

where

$$\mathcal{R}_{\mathcal{G}}(X) := \frac{1}{|\mathcal{G}|} \sum_{P \in \mathcal{G}} PXP^T, X \in \mathbb{R}^{n \times n},$$

is called the *Reynolds operator* (or *group average*) of \mathcal{G} . The operator $\mathcal{R}_{\mathcal{G}}$ is the orthogonal projection onto the commutant. The commutant $A_{\mathcal{G}}$ is a *matrix *-algebra*, i.e., it is a set of matrices that is closed under addition, scalar multiplication, matrix multiplication, and taking conjugate transposes. One may obtain a basis for $A_{\mathcal{G}}$ from the orbits of the action of \mathcal{G} on ordered pairs of vertices, where the *orbit* of $(u_i, u_j) \in \{0, 1\}^n \times \{0, 1\}^n$ under the action of \mathcal{G} is the set $\{(Pu_i, Pu_j) \mid P \in \mathcal{G}\}$, and $u_i \in \mathbb{R}^n$ is the i -th unit vector. In what follows, we denote

$$\{B_1, \dots, B_d\} \in \{0, 1\}^{n \times n}, \text{ basis for } A_{\mathcal{G}}. \quad (2.4)$$

Note that $\sum_{i=1}^d B_i = J_n$, where J_n is the matrix of all ones of order n . We will sometimes also use the notation J instead of J_n if the order of the matrix is clear from the context.

In what follows we obtain that the Reynolds operator maps the feasible set \mathcal{F}_X of (2.1) into itself and keeps the objective value the same, i.e.,

$$X \in \mathcal{F}_X \implies \mathcal{R}_{\mathcal{G}}(X) \in \mathcal{F}_X \text{ and } \langle C, \mathcal{R}_{\mathcal{G}}(X) \rangle = \langle C, X \rangle.$$

One can restrict optimization of an SDP problem to feasible points in a matrix *-algebra that contains the data matrices of that problem, see e.g., [13, 20]. In particular, the following result is known.

Theorem 2.1 ([13]). *Let $A_{\mathcal{G}}$ denote a matrix *-algebra that contains the data matrices of an SDP problem as well as the identity matrix. If the SDP problem has an optimal solution, then it has an optimal solution in $A_{\mathcal{G}}$.*

Therefore, we may restrict the feasible set of the optimization problem to its intersection with $A_{\mathcal{G}}$. In particular, we can use the basis matrices and assume that

$$X \in \mathcal{F}_X \cap A_{\mathcal{G}} \Leftrightarrow \left[X = \sum_{i=1}^d x_i B_i =: \mathcal{B}^*(x) \in \mathcal{F}_X, \text{ for some } x \in \mathbb{R}^d \right]. \quad (2.5)$$

Example 2.2 (Hamming Graphs). *We now present an example of an algebra that we use later in our numerics.*

The Hamming graph $H(d, q)$ is the Cartesian product of d copies of the complete graph K_q , with vertices represented by d -tuples of letters from an alphabet of size q . The Hamming distance between vertices u and v , denoted by $|(u, v)|$, is the number of positions in which d -tuples u and v differ.

The matrices

$$(B_i)_{u,v} := \begin{cases} 1 & \text{if } |(u, v)| = i \\ 0 & \text{otherwise} \end{cases}, \quad i = 0, \dots, d$$

form a basis of the Bose-Mesner algebra of the Hamming scheme, see [16]. In particular, $B_0 = I$ is the identity matrix and B_1 is the adjacency matrix of the Hamming graph $H(d, q)$ of size $q^d \times q^d$. In cases, like for the Bose-Mesner algebra, when one of the basis elements equals the identity matrix, it is common to set the index of the corresponding basis element to zero. The basis matrices B_i can be simultaneously diagonalized by the real, orthogonal matrix Q given by

$$Q_{u,v} = 2^{-\frac{d}{2}} (-1)^{u^T v}.$$

The distinct elements of the matrix $Q^T B_i Q$ equal $K_i(j)$ ($j = 0, \dots, d$) where

$$K_i(j) := \sum_{h=0}^i (-1)^h (q-1)^{i-h} \binom{j}{h} \binom{d-j}{i-h}, \quad j = 0, \dots, d,$$

are Krawtchouk polynomials. We denote by $\mu_j := \binom{d}{j}(q-1)^j$ the multiplicity of the j -th eigenvalue $K_i(j)$. The elements of the character table $P \in \mathbb{R}^{(d+1) \times (d+1)}$ of the Hamming scheme $H(d, q)$, given in terms of the Krawtchouk polynomials, are

$$p_{i,j} := K_i(j), \quad i, j = 0, \dots, d.$$

In the later sections, we use the following well-known orthogonality relations on the Krawtchouk polynomial

$$\sum_{j=0}^d K_r(j)K_s(j) \binom{d}{j} (q-1)^j = q^d \binom{d}{s} (q-1)^s \delta_{r,s}, \quad r, s = 0, \dots, d, \quad (2.6)$$

where $\delta_{r,s}$ is the Kronecker delta function.

2.2.1 First symmetry reduction using $X = \mathcal{B}^*(x)$

We now obtain our first reduced program using the substitution $X = \mathcal{B}^*(x)$. Note that the program is reduced in the sense that the feasible set can be smaller though the optimal value remains the same.

$$p_{\text{SDP}}^* = \min\{\langle \mathcal{B}(C), x \rangle \mid (\mathcal{A} \circ \mathcal{B}^*)(x) = b, \mathcal{B}^*(x) \succeq 0\}, \quad (\text{substitution } X = \mathcal{B}^*(x)). \quad (2.7)$$

Here, \mathcal{B} is the adjoint of \mathcal{B}^* . In the case of a doubly nonnegative relaxation, the structure of the basis in (2.4) allows us to equivalently set $x \geq 0$.

A matrix $*$ -algebra \mathcal{M} is called basic if $\mathcal{M} = \{\oplus_{i=1}^t M \mid M \in \mathbb{C}^{m \times m}\}$, where \oplus denotes the direct sum of matrices. A very important decomposition result for matrix $*$ -algebras is the following result due to Wedderburn.

Theorem 2.3 ([49]). *Let \mathcal{M} be a matrix $*$ -algebra containing identity matrix. Then there exists a unitary matrix Q such that $Q^* \mathcal{M} Q$ is a direct sum of basic algebras.*

To simplify our presentation, the matrix Q is assumed to be real orthogonal. The case when Q is complex can be derived analogously.

The matrices in the basis $B_j, j = 1 \dots, d$, can be mutually block-diagonalized by some orthogonal matrix Q . More precisely, there exists an orthogonal matrix Q such that we get the following block-diagonal transformation on B_j :

$$\tilde{B}_j := Q^T B_j Q =: \text{blkdiag}((\tilde{B}_j^k)_{k=1}^t), \quad \forall j = 1, \dots, d. \quad (2.8)$$

For $Q^T X Q = \sum_{j=1}^d x_j \tilde{B}_j$, we now define the linear transformation for obtaining the block matrix diagonal form:

$$\tilde{\mathcal{B}}^*(x) := \sum_{j=1}^d x_j \tilde{B}_j = \begin{bmatrix} \tilde{\mathcal{B}}_1^*(x) & & \\ & \ddots & \\ & & \tilde{\mathcal{B}}_t^*(x) \end{bmatrix} =: \text{blkdiag}((\tilde{\mathcal{B}}_k^*(x))_{k=1}^t), \quad (2.9)$$

where

$$\tilde{\mathcal{B}}_k^*(x) =: \sum_{j=1}^d x_j \tilde{B}_j^k \in \mathcal{S}_+^{n_i}$$

is the k -th diagonal block of $\tilde{\mathcal{B}}^*(x)$, and the sum of the t block sizes $n_1 + \dots + n_t = n$. Thus, for any feasible X we get

$$X = \mathcal{B}^*(x) = Q \tilde{\mathcal{B}}^*(x) Q^T \in \mathcal{F}_X.$$

2.2.2 Second symmetry reduction to block diagonal form using $X = Q\tilde{\mathcal{B}}^*(x)Q^T$

We now derive the second reduced program using the substitution $X = Q\tilde{\mathcal{B}}^*(x)Q^T$. The program is further reduced since we obtain the block diagonal problem

$$p_{\text{SDP}}^* = \min\{\langle \tilde{\mathcal{B}}(\tilde{C}), x \rangle \mid (\tilde{\mathcal{A}} \circ \tilde{\mathcal{B}}^*)(x) = b, \tilde{\mathcal{B}}^*(x) \succeq 0\}, \quad (2.10)$$

where $\tilde{C} = Q^T C Q$ and $\tilde{\mathcal{A}}$ is the linear transformation obtained from \mathcal{A} by using the rotation $\tilde{A}_j = Q^T A_j Q, \forall j$. We denote the corresponding blocks as $\tilde{A}_j^k, \forall j = 1, \dots, d, \forall k = 1, \dots, t$.

We see that the objective (2.10) satisfies

$$\tilde{c} := \tilde{\mathcal{B}}(\tilde{C}) = (\langle \tilde{B}_j, \tilde{C} \rangle) = (\langle B_j, C \rangle) \in \mathbb{R}^d.$$

The i -th row of the matrix constraint $\tilde{\mathcal{A}}x = b$ is

$$\begin{aligned} b_i &= (\tilde{\mathcal{A}}x)_i \\ &:= ((\tilde{\mathcal{A}} \circ \tilde{\mathcal{B}}^*)(x))_i \\ &= \langle \tilde{A}_i, \tilde{\mathcal{B}}^*(x) \rangle \\ &= \langle \tilde{\mathcal{B}}(\tilde{A}_i), x \rangle. \end{aligned}$$

Therefore

$$\tilde{A}_{ij} = (\tilde{\mathcal{B}}(\tilde{A}_i))_j = \langle \tilde{B}_j, \tilde{A}_i \rangle = \langle B_j, A_i \rangle, \quad i = 1, \dots, m, j = 1, \dots, d. \quad (2.11)$$

Without loss of generality, we can now define

$$c := \tilde{c}, \quad A := \tilde{\mathcal{A}}.$$

Moreover, just as for facial reduction, the symmetry reduction step can result in A not being full row rank (onto). We then have to choose a nice (well conditioned) submatrix that is full row rank and use the resulting subsystem of $Ax = b$. We see below how to do this and simultaneously obtain strict feasibility.

We can now rewrite the **SDP** (2.1) as

$$p_{\text{SDP}}^* = \min\{c^T x \mid Ax = b, \tilde{\mathcal{B}}_k^*(x) \succeq 0, k = 1, \dots, t\}. \quad (2.12)$$

For many applications, there are *repeated blocks*. We then take advantage of this to reduce the size of the problem and maintain stability.

The program (2.12) is a *symmetry reduced formulation* of (2.1). We denote its feasible set and feasible slacks as

$$\mathcal{F}_x := \{x \mid \tilde{\mathcal{B}}^*(x) \succeq 0, Ax = b, x \in \mathbb{R}^d\}, \quad \mathcal{S}_x := \{\tilde{\mathcal{B}}^*(x) \succeq 0 \mid Ax = b, x \in \mathbb{R}^d\}. \quad (2.13)$$

We bear in mind that $\tilde{\mathcal{B}}^*(x)$ is a block-diagonal matrix. But it is written as a single matrix for convenience in order to describe facial reduction for the symmetry reduced program below.

Since $\tilde{B}_1, \dots, \tilde{B}_d$ are block diagonal, symmetric matrices, the symmetry reduced formulation is typically much smaller than the original problem, i.e.,

$$x \in \mathbb{R}^d, \quad d \ll \sum_{i=1}^d t(n_i) \ll t(n),$$

where $t(k) = k(k+1)/2$ is the triangular number.

3 Facial reduction for the symmetry reduced program

In this section, we show how to apply facial reduction to the symmetry reduced **SDP** (2.12). The key is using the exposing vector view of facial reduction, [17]. Formally speaking, if an exposing vector of the *minimal face*² of the **SDP** (2.1) is given, then we are able to construct a

²The smallest face containing the feasible set.

corresponding exposing vector of the minimal face of the symmetry reduced program (2.12). In fact, we show that all the exposing vectors of the symmetry reduced program can be obtained from the exposing vectors of the original program.

3.1 Rank preserving

We begin with showing the maximum rank preserving properties of the symmetrization reduction. Note that

$$\begin{aligned} \max\{\text{rank}(X) : X \in \mathcal{F}_X\} &= \text{rank}(X), \forall X \in \text{relint}(\mathcal{F}_X) \\ &= \text{rank}(X), \forall X \in \text{relint}(\text{face}(\mathcal{F}_X)), \end{aligned}$$

where $\text{face}(\mathcal{F}_X)$ is the minimal face of \mathbb{S}_+^n containing the feasible set.

Theorem 3.1. *The feasible matrix $X \in \mathcal{F}_X \subseteq \mathbb{S}_+^n$ of maximum rank has $\text{rank}(X) = r$ if, and only if, the feasible slack matrix $\tilde{\mathcal{B}}^*(x) \in \mathcal{S}_x \subseteq \mathbb{S}_+^n$ of maximum rank has (the same) $\text{rank}(\tilde{\mathcal{B}}^*(x)) = r$.*

Proof. Let $X \in \mathcal{F}_X$ be the matrix with maximum rank r . Then X is in the relative interior of the minimal face $f \trianglelefteq \mathbb{S}_+^n$ containing \mathcal{F}_X , i.e.,

$$X \in \text{relint}(f) = [V \ U] \begin{bmatrix} \mathbb{S}_+^r & 0 \\ 0 & 0 \end{bmatrix} [V \ U]^T, \text{ for some orthogonal } [V \ U].$$

The nonsingular congruence $P^T X P$ is feasible for each $P \in \mathcal{G}$, and also has rank r . Note that

$$A, B \in \mathbb{S}_+^n \implies \text{rank}(A + B) \geq \max\{\text{rank}(A), \text{rank}(B)\}.$$

Therefore, applying the Reynolds operator, we have

$$X_0 = \frac{1}{|\mathcal{G}|} \sum_{P \in \mathcal{G}} P^T X P \in \text{relint}(f).$$

Since $X_0 \in \mathcal{A}_{\mathcal{G}}$, we have $Q^T X_0 Q \in \mathcal{S}_x (= Q^T(\mathcal{F}_X \cap \mathcal{A}_{\mathcal{G}})Q)$ and it has rank r , where Q is the orthogonal matrix given above in (2.8).

Conversely, if $\tilde{\mathcal{B}}^*(x) \in \mathcal{S}_x$ with rank r , then $X := Q\tilde{\mathcal{B}}^*(x)Q^T$ is in \mathcal{F}_X with rank r . \square

Thus, the Slater condition is preserved by symmetry reduction.

Corollary 3.2. *The program (2.1) is strictly feasible if, and only if, its symmetry reduced program (2.12) is strictly feasible.*

Remark 3.3. *From the proof of Theorem 3.1, if there is a linear transformation $X = \mathcal{L}(x)$ with a full rank feasible $\hat{X} \in \text{range}(\mathcal{L})$, $\hat{X} = \mathcal{L}(\hat{x})$, then in general we can conclude that the substitution $X = \mathcal{L}(x)$ results in a smaller **SDP** with strict feasibility holding at \hat{x} , i.e.,*

$$\hat{X} \succ 0, \mathcal{A}(\hat{X}) = b, \hat{X} = \mathcal{L}(\hat{x}) \implies \mathcal{L}(\hat{x}) \succ 0, (\mathcal{A} \circ \mathcal{L})(\hat{x}) = b.$$

3.2 Facial reduction and exposing vectors

Unfortunately, for many given combinatorial problems, the semidefinite relaxation is not strictly feasible and therefore degenerate and generally ill-conditioned. From the above, we see that this implies that the symmetry reduced problem is degenerate as well. Although both symmetry reduction and facial reduction can be performed independently, there has not been any study that implements these techniques simultaneously, i.e., to obtain a symmetry reduced problem which also guarantees strict feasibility.

In what follows, we show that the exposing vectors of the symmetry reduced program (2.12) can be obtained from the exposing vectors of the original program (2.1). This enables us to facially reduce the symmetry reduced program (2.12).

Let $W = UU^T$, with $U \in \mathbb{R}^{n \times (n-r)}$ full column rank; and let W be an exposing vector of the minimal face of \mathbb{S}_+^n containing the feasible region \mathcal{F}_X of (2.1). Let $V \in \mathbb{R}^{n \times r}$ be such that

$$\text{range}(V) = \text{null}(U^T).$$

Then facial reduction means that we can use the substitution $X = \mathcal{V}^*(R) = VRV^T$ and obtain the following equivalent, smaller, formulation of (2.1), with strict feasibility holding:

$$\min\{\langle V^T CV, R \rangle \mid \langle V^T A_i V, R \rangle = b_i, \quad i = 1, \dots, m, \quad R \in \mathbb{S}_+^r\}. \quad (3.1)$$

In fact, \hat{R} strictly feasible corresponds to $\hat{X} = \mathcal{V}^*(\hat{R}) \in \text{relint}(\mathcal{F}_X)$.

The following results show how to find an exposing vector that is in the commutant $\mathcal{A}_{\mathcal{G}}$.

Lemma 3.4. *Let W be an exposing vector of rank d of $\text{face}(\mathcal{F}_X)$, the minimal face of \mathbb{S}_+^n containing \mathcal{F}_X . Then there exists an exposing vector $W_{\mathcal{G}} \in \mathcal{A}_{\mathcal{G}}$ with $\text{rank}(W_{\mathcal{G}}) \geq d$.*

Proof. Let W be the exposing vector of rank d , i.e.,

$$X \in \mathcal{F}_X \implies \langle W, X \rangle = 0.$$

Since (2.1) is \mathcal{G} -invariant, $PXP^T \in \mathcal{F}_X$ for every $P \in \mathcal{G}$, we conclude that

$$\langle W, PXP^T \rangle = \langle P^T WP, X \rangle = 0.$$

Therefore, $P^T WP \succeq 0$ is an exposing vector of rank d . Let $W_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{P \in \mathcal{G}} P^T WP$. Then $W_{\mathcal{G}}$ is also an exposing vector.

It remains to show that the rank of $W_{\mathcal{G}}$ is at least d . Let v be a vector in the null space of $W_{\mathcal{G}}$, i.e., $W_{\mathcal{G}}v = 0$. Then $v^T W_{\mathcal{G}}v = \frac{1}{|\mathcal{G}|} \sum_{P \in \mathcal{G}} (Pv)^T W(Pv) = 0$. Since $W \succeq 0$, we have $(Pv)^T W(Pv) = 0$ and $W(Pv) = 0$ for every $P \in \mathcal{G}$. In particular, for the identity matrix $I \in \mathcal{G}$, we get $Wv = 0$. This shows that v is also in the null space of W . Therefore the null space of $W_{\mathcal{G}}$ is contained in the null space of W . We conclude that the rank of $W_{\mathcal{G}}$ is at least d . \square

Note that one can obtain an exposing vector $W_{\mathcal{G}} \in \mathcal{A}_{\mathcal{G}}$ from an exposing vector W by using the Reynolds operator. However, in some cases $W_{\mathcal{G}}$ can be more easily derived, as our examples in later numerical sections show. We now continue and show that $Q^T W_{\mathcal{G}} Q$ is also an exposing vector.

Lemma 3.5. *Let W be an exposing vector of the face of \mathbb{S}_+^n containing \mathcal{F}_X , and assume that $W \in \mathcal{A}_{\mathcal{G}}$. Let Q be the orthogonal matrix given above in (2.8). Then $\widetilde{W} = Q^T W Q$ exposes a face of \mathbb{S}_+^n containing \mathcal{S}_x .*

Proof. Let

$$Z = \sum_{i=1}^d x_i \tilde{B}_i = Q^T \left(\sum_{i=1}^d x_i B_i \right) Q \in \mathcal{S}_x.$$

Then, by construction Z is a block-diagonal matrix, say $Z = \text{blkdiag}(Z_1, \dots, Z_t)$. Now, since W is an exposing vector of the face of \mathbb{S}_+^n containing \mathcal{F}_X we have

$$\begin{aligned} WX = 0, \quad \forall X \in \mathcal{F}_X &\implies WX = 0, \quad \forall X = \sum_i x_i B_i \succeq 0, \quad \text{for some } x \text{ with } Ax = b \\ &\implies \widetilde{W}Z = 0, \quad \forall Z \in \mathcal{S}_x, \end{aligned}$$

where $\widetilde{W} = Q^T W Q \succeq 0$. Thus, \widetilde{W} is an exposing vector of a proper face of \mathbb{S}_+^n containing \mathcal{S}_x .

Since $Z = \text{blkdiag}(Z_1, \dots, Z_t)$ is a block-diagonal matrix and $W \in A_G$, we have that $\widetilde{W} = \text{blkdiag}(\widetilde{W}_1, \dots, \widetilde{W}_t)$ with \widetilde{W}_i the corresponding i -th diagonal block of $Q^T W Q$. \square

Since we may assume $W \in A_G$, the exposing vector $Q^T W Q$ is a block-diagonal matrix. Now, let us show that $Q^T W Q$ exposes the minimal face of \mathbb{S}_+^n containing \mathcal{S}_x , $\text{face}(\mathcal{S}_x)$. It suffices to show that the rank of $Q^T W Q$ is $n - r$, see Theorem 3.1.

Theorem 3.6. *Let $W \in A_G$ be an exposing vector of $\text{face}(\mathcal{F}_X)$, the minimal face of \mathbb{S}_+^n containing \mathcal{F}_X . Then the block-diagonal matrix $\widetilde{W} = Q^T W Q$ exposes the minimal face of \mathbb{S}_+^n containing \mathcal{S}_x , $\text{face}(\mathcal{S}_x)$.*

Proof. The minimality follows from Theorem 3.1, as $\text{rank}(\widetilde{W}) = \text{rank}(W) = n - r$. \square

Now let $\widetilde{W} = Q^T W Q$ expose the minimal face of \mathbb{S}_+^n containing \mathcal{S}_x , and let

$$\widetilde{W} = \text{blkdiag}(\widetilde{W}_1, \dots, \widetilde{W}_t), \quad \widetilde{W}_i = \widetilde{U}_i \widetilde{U}_i^T, \quad \widetilde{U}_i \text{ full rank}, \quad i = 1, \dots, t.$$

Let \widetilde{V}_i be a full rank matrix whose columns form a basis for the orthogonal complement to the columns of \widetilde{U}_i , $i = 1, \dots, t$. Take $\widetilde{V} = \text{blkdiag}(\widetilde{V}_1, \dots, \widetilde{V}_t)$. Then, the facially reduced formulation of (2.12) is

$$\begin{aligned} p_{FR}^* &= \min\{c^T x \mid Ax = b, \widetilde{\mathcal{B}}^*(x) = \widetilde{V} \widetilde{R} \widetilde{V}^T, \widetilde{R} \succeq 0\} \\ &= \min\{c^T x \mid Ax = b, \widetilde{\mathcal{B}}_k^*(x) = \widetilde{V}_k \widetilde{R}_k \widetilde{V}_k^T, \widetilde{R}_k \succeq 0, \forall k = 1, \dots, t\}, \end{aligned} \quad (3.2)$$

where $\widetilde{V}_k \widetilde{R}_k \widetilde{V}_k^T$ is the corresponding k -th block of $\widetilde{\mathcal{B}}^*(x)$, and $\widetilde{R} = \text{blkdiag}(\widetilde{R}_1, \dots, \widetilde{R}_t)$. Note that some of the blocks $\widetilde{\mathcal{B}}_k^*(x)$ are the same and thus can be removed in the computation, see Theorem 2.3.

3.3 Simplifications

After facial reduction, some of the constraints become redundant in the facially reduced program (3.1). The next result shows that these constraints are also redundant in the facially and symmetry reduced program (3.2).

Theorem 3.7. *Let $\mathcal{I} \subsetneq \{1, \dots, m\}$. Suppose that the constraints $\langle A_k, V R V^T \rangle = b_k, k \notin \mathcal{I}$, are redundant in (3.1), i.e., the facially reduced formulation (3.1) is equivalent to*

$$\min_{R \in \mathbb{S}_+^r} \{ \langle V^T C V, R \rangle \mid \langle V^T A_i V, R \rangle = b_i, \forall i \in \mathcal{I} \}. \quad (3.3)$$

Then the constraints

$$\sum_{j=1}^d A_{k,j} x_j = b_k, k \notin \mathcal{I},$$

are redundant in (3.2), i.e., the facially and symmetry reduced program (3.2) is equivalent to

$$\min_{x \in \mathbb{R}^d, \widetilde{R} \in \mathbb{S}_+^r} \{ c^T x \mid \sum_{j=1}^d A_{i,j} x_j = b_i, \forall i \in \mathcal{I}, \widetilde{\mathcal{B}}^*(x) = \widetilde{V} \widetilde{R} \widetilde{V}^T \}. \quad (3.4)$$

Proof. Note that $\text{range}(V) = \text{null}(W)$ and $\text{range}(\tilde{V}) = \text{null}(Q^T W Q)$. Let us show that $\text{range}(V) = \text{range}(Q\tilde{V})$. Let $v \in \text{range}(V)$. Since Q is orthogonal, there exists a unique vector u such that $v = Qu$. Thus we have $0 = v^T W v = u^T Q^T W Q u$. This means $u \in \text{range}(\tilde{V})$ and thus $u = \tilde{V}w$ for some vector w . This implies that $v = Q\tilde{V}w$ and thus $v \in \text{range}(Q\tilde{V})$. Conversely, let $v \in \text{range}(Q\tilde{V})$ and thus $v = Q\tilde{V}u$ for some vector u . Since $\tilde{V}^T Q^T W Q \tilde{V} = 0$, we have $v^T W v = 0$. As W is positive semidefinite, it holds that $Wv = 0$ and thus $v \in \text{null}(W) = \text{range}(V)$.

Since $\text{range}(V) = \text{range}(Q\tilde{V})$, we can assume without loss of generality, that the V in (3.3) satisfies $V = Q\tilde{V}$. Let (x, \tilde{R}) be feasible for (3.4). Define $X := \sum_{i=1}^d B_i x_i$. We will show the equivalence in the following order:

$$\begin{aligned} \boxed{(x, \tilde{R}) \text{ feasible for (3.4)}} &\implies \boxed{\tilde{R} \text{ feasible for (3.3)}} \\ &\implies \boxed{\tilde{R} \text{ feasible for (3.1)}} \\ &\implies \boxed{(x, \tilde{R}) \text{ feasible for (3.2)}}. \end{aligned}$$

Since

$$Q^T \left(\sum_{i=1}^d B_i x_i \right) Q = \tilde{V} \tilde{R} \tilde{V}^T = Q^T V \tilde{R} V^T Q,$$

we have $\sum_{i=1}^d B_i x_i = V \tilde{R} V^T$. Using the feasibility and (2.11), it holds that

$$\langle A_i, V \tilde{R} V^T \rangle = \langle A_i, \sum_{j=1}^d B_j x_j \rangle = \sum_{j=1}^d A_{i,j} x_j = b_i, \forall i \in \mathcal{I}.$$

Thus all the linear constraints in (3.3) are satisfied, and $\tilde{R} \succeq 0$ is feasible for (3.3). By assumption, \tilde{R} is also feasible for (3.1). Thus the constraints $\langle A_i, V \tilde{R} V^T \rangle = b_i, \forall i \notin \mathcal{I}$ are satisfied as well. This shows that (x, \tilde{R}) is feasible for (3.2). \square

To obtain a formulation in variable \tilde{R} only as in the facially reduced formulation (3.1), we can replace x in terms of \tilde{R} using the constraint $\tilde{\mathcal{B}}^*(x) = \tilde{V} \tilde{R} \tilde{V}^T$. For the facially and symmetry reduced program (3.2), this substitution is particularly important as the constraint $\tilde{\mathcal{B}}^*(x) = \tilde{V} \tilde{R} \tilde{V}^T$ can offset the benefits of symmetry reduction. This substitution can be done easily for (3.2) by rewriting the constraints as

$$b_i = \langle A_i, X \rangle = \langle Q^T A_i Q, Q^T X Q \rangle = \langle Q^T A_i Q, \tilde{\mathcal{B}}^*(x) \rangle = \langle Q^T A_i Q, \tilde{V} \tilde{R} \tilde{V}^T \rangle.$$

The objective can be similarly changed. This method, however, does not work for DNN relaxations that include non-negativity constraints. This difficulty could be resolved as follows.

Theorem 3.8. *Consider the facially and symmetry reduced DNN relaxation (3.2) with non-negativity constraints,*

$$\min\{c^T x \mid Ax = b, \tilde{\mathcal{B}}^*(x) = \tilde{V} \tilde{R} \tilde{V}^T, \tilde{R} \succeq 0, x \geq 0\}. \quad (3.5)$$

Equate x with

$$x \leftarrow f(\tilde{R}) = \text{Diag}(w^{-1}) \mathcal{B}(V \tilde{R} V^T),$$

with

$$w = (\langle B_i, B_i \rangle)_{i \in \mathbb{R}^d}, V = Q\tilde{V} \text{ and } Q \text{ from Theorem 2.3.}$$

Then (3.5) is equivalent to

$$\min\{c^T f(\tilde{R}) \mid Af(\tilde{R}) = b, \tilde{R} \succeq 0, f(\tilde{R}) \geq 0\}. \quad (3.6)$$

Proof. If (x, \tilde{R}) is feasible for (3.5), then $\mathcal{B}^*(x) = V\tilde{R}V^T$. As $w > 0$ and $\mathcal{B}\mathcal{B}^* = \text{Diag}(w)$, we have $x = f(\tilde{R})$ and thus \tilde{R} is feasible for (3.6).

Let \tilde{R} be feasible for (3.6). Assume that there are no repeating blocks in the decomposition. Since $\tilde{V}\tilde{R}\tilde{V}^T$ is a block-diagonal matrix in the basic algebra $Q^T A_G Q$, we have $V\tilde{R}V^T = Q\tilde{V}\tilde{R}\tilde{V}^T Q^T \in A_G$. It follows from Theorem 2.3 that there exists a unique x such that $V\tilde{R}V^T = \mathcal{B}^*(x)$. Then we must have $x = f(\tilde{R})$ and thus (x, \tilde{R}) is feasible for (3.5). If there are repeating blocks, then we can remove the repeating ones and the result follows with similar arguments. \square

For the Hamming scheme, we have an explicit expression for the orthogonal matrix Q used in Theorem 3.8, see Example 2.2 and Section 5.1. In general, we do not know the corresponding orthogonal matrix explicitly. In Section 5.2, we use the heuristics from [14] to compute a block diagonalization of A_G . In this case, the equivalence in Theorem 3.8 may not be true, and (3.6) may be weaker than (3.5). However our computational results indicate that all the bounds remain the same, see Table 5.7 below.

4 The alternating direction method of multipliers, ADMM

It is well known that interior-point methods do not scale well for **SDP**. Moreover, they have great difficulty with handling additional cutting planes such as nonnegativity constraints. In particular, solving the doubly nonnegative relaxation, **DNN**, using interior-point methods is extremely difficult. The alternating direction method of multipliers is a first-order method for convex problems developed in the 1970s, and rediscovered recently. This method decomposes an optimization problem into subproblems that may be easier to solve. In particular, it is extremely successful for splittings with two cones. This feature makes the **ADMM** well suited for large-scaled **DNN** problems. For state of the art in theory and applications of the **ADMM**, we refer the interested readers to [8].

Oliveira, Wolkowicz and Xu [34] propose a version of the **ADMM** for solving an **SDP** relaxation for the Quadratic Assignment Problem (**QAP**). Their computational experiments show that the proposed variant of the **ADMM** exhibits remarkable robustness, efficiency, and even provides improved bounds.

4.1 Augmented Lagrangian

We follow the approach from [34] for solving our facially and symmetry reduced **DNN** relaxation in (3.2).

Let $\tilde{V} = \text{blkdiag}(\tilde{V}_1, \dots, \tilde{V}_t)$ and $\tilde{R} = \text{blkdiag}(\tilde{R}_1, \dots, \tilde{R}_t)$. The augmented Lagrangian of (3.2) corresponding to the linear constraints $\tilde{\mathcal{B}}^*(x) = \tilde{V}\tilde{R}\tilde{V}^T$ is given by:

$$\mathcal{L}(x, \tilde{R}, \tilde{Z}) = \langle \tilde{C}, \tilde{\mathcal{B}}^*(x) \rangle + \langle \tilde{Z}, \tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T \rangle + \frac{\beta}{2} \|\tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T\|^2,$$

where, see (2.10), $\tilde{C} = Q^T C Q$ is a block-diagonal matrix as $C \in A_G$, \tilde{Z} is also in block-diagonal form, and $\beta > 0$ is the penalty parameter.

The alternating direction method of multipliers, **ADMM**, uses the *augmented Lagrangian*, $\mathcal{L}(x, \tilde{R}, \tilde{Z})$, and essentially solves the max-min problem

$$\max_{\tilde{Z}} \min_{x \in P, \tilde{R} \succeq 0} \mathcal{L}(x, \tilde{R}, \tilde{Z}),$$

where P is a simple polyhedral set of constraints on x , e.g., linear constraints $Ax = b$ and nonnegativity constraints, see (4.3) below. The advantage of the method is the simplifications

obtained for the constraints by taking advantage of the splitting in the variables. We then find the following updates $(x_+, \tilde{R}_+, \tilde{Z}_+)$:

$$\begin{aligned} x_+ &= \arg \min_{x \in P} \mathcal{L}(x, \tilde{R}, \tilde{Z}), \\ \tilde{R}_+ &= \arg \min_{\tilde{R} \succeq 0} \mathcal{L}(x_+, \tilde{R}, \tilde{Z}), \\ \tilde{Z}_+ &= \tilde{Z} + \gamma\beta(\tilde{\mathcal{B}}^*(x_+) - \tilde{V}\tilde{R}_+\tilde{V}^T). \end{aligned}$$

Here, $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is the step size for updating the dual variable \tilde{Z} . In the following sections we explain in details how to solve each subproblem.

4.2 On solving the \tilde{R} -subproblem

The \tilde{R} -subproblem can be explicitly solved. We complete the square and get the equivalent problem

$$\begin{aligned} \tilde{R}_+ &= \min_{\tilde{R} \succeq 0} \|\tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T + \frac{1}{\beta}\tilde{Z}\|^2 \\ &= \min_{\tilde{R} \succeq 0} \|\tilde{R} - \tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \frac{1}{\beta}\tilde{Z})\tilde{V}\|^2 \\ &= \sum_{k=1}^t \min_{\tilde{R}_k \succeq 0} \|\tilde{R}_k - (\tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \frac{1}{\beta}\tilde{Z})\tilde{V})_k\|^2. \end{aligned} \tag{4.1}$$

Here, we normalize each block \tilde{V}_k such that $\tilde{V}_k^T\tilde{V}_k = I$, and thus $(\tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \frac{1}{\beta}\tilde{Z})\tilde{V})_k$ is the k -th block of $\tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \frac{1}{\beta}\tilde{Z})\tilde{V}$ corresponding to \tilde{R}_k . So we only need to solve k small problems whose optimal solutions are

$$\tilde{R}_k = \mathcal{P}_{\mathbb{S}_+} \left(\tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \frac{1}{\beta}\tilde{Z})\tilde{V} \right)_k, \quad k = 1, \dots, t,$$

where $\mathcal{P}_{\mathbb{S}_+}(M)$ is the projection onto the cone of positive semidefinite matrices.

4.3 On solving the x -subproblem

For the x -subproblem, we have

$$x_+ = \arg \min_{x \in P} \left\| \tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T + \frac{\tilde{C} + \tilde{Z}}{\beta} \right\|^2. \tag{4.2}$$

For many combinatorial optimization problems, some of the constraints $Ax = b$ in (2.12) become redundant after facial reduction of their semidefinite programming relaxations, see Theorem 3.7. Thus, the set P often collapses to a simple set. This often leads to an analytic solution for the x -subproblem; e.g., this happens for the quadratic assignment, graph partitioning, vertex separator, and shortest path problems.

For some interesting applications, the x -subproblem is equivalent to the following special case of the weighted, relaxed, quadratic knapsack problem:

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|\mathcal{T}^*(x) - Y\|^2 \\ \text{s.t.} \quad & x \in P := \{x : w^T x = c, x \geq 0\}, \end{aligned} \tag{4.3}$$

where Y is a given matrix and $\mathcal{T}^*(x) = \sum_{i=1}^q x_i T_i$ for some given symmetric matrices T_i . The problem (4.3) is a projection onto the weighted simplex. We consider the following assumption on a linear transformation $\mathcal{T} : \mathbb{S}^n \rightarrow \mathbb{R}^q$ and its adjoint.

Assumption 4.1. The linear transformation $\mathcal{T} : \mathbb{S}^n \rightarrow \mathbb{R}^q$ in (4.3) satisfies

$$\mathcal{T}(\mathcal{T}^*(x)) = \text{Diag}(w)x, \forall x \in \mathbb{R}^q, \text{ for some } w > 0.$$

Lemma 4.2. Suppose that the linear transformation \mathcal{T} satisfies Assumption 4.1, and that (4.3) is feasible. Then the projection problem (4.3) can be solved efficiently (explicitly) using Algorithm 4.3.

Proof. The Lagrangian function of the problem is

$$\frac{1}{2} \|\mathcal{T}^*(x) - Y\|^2 - \tau(w^T x - c) - \lambda^T x,$$

where $\tau \in \mathbb{R}$ and $\lambda \in \mathbb{R}_+^q$ are the Lagrangian multipliers. The KKT optimality conditions for the problem are given by

$$\begin{aligned} \mathcal{T}(\mathcal{T}^*(x)) - \mathcal{T}(Y) - \tau w - \lambda &= 0, \\ x &\geq 0, \\ \lambda &\geq 0, \\ \lambda^T x &= 0, \\ w^T x &= c. \end{aligned}$$

Note that $\text{Diag}(w)$ is the matrix representation of $\mathcal{T} \circ \mathcal{T}^*$. This means that $\langle T_i, T_j \rangle = 0, \forall i \neq j$, and we can simplify the first condition.³ This yields

$$x_i = w_i^{-1}(\mathcal{T}(Y))_i + \tau + w_i^{-1}\lambda_i.$$

Define the data vector $y := \mathcal{T}(Y)$. The complementary slackness $\lambda^T x = 0$ implies that if $x_i > 0$, then $\lambda_i = 0$ and $x_i = w_i^{-1}y_i + \tau$. If $x_i = 0$, then $w_i^{-1}y_i + \tau = -w_i^{-1}\lambda_i \leq 0$. Thus the zero, respectively positive, entries of the optimal solution x correspond to the smaller, respectively larger, entries in $(w_i^{-1}y_i)_{i=1}^q$.

Let us assume, without loss of generality, that $(w_i^{-1}y_i)_{i=1}^q, x$ are sorted in non-increasing order:

$$\frac{y_1}{w_1} \geq \dots \geq \frac{y_k}{w_k} \geq \frac{y_{k+1}}{w_{k+1}} \geq \dots \geq \frac{y_q}{w_q}, \quad x_1 \geq \dots \geq x_k > x_{k+1} = \dots = x_q = 0.$$

The condition $w^T x = c$ implies that

$$w^T x = \sum_{i=1}^k w_i \left(\frac{y_i}{w_i} + \tau \right) = \sum_{i=1}^k y_i + \tau \sum_{i=1}^k w_i = c,$$

and thus

$$\tau = \frac{c - \sum_{i=1}^k y_i}{\sum_{i=1}^k w_i}.$$

Therefore, one can solve the problem by simple inspection once k is known. The following algorithm finds an optimal solution x to the problem (4.3). The correctness of the algorithm is then similar to the projection onto the (unweighted) simplex problem, see [10, 11]. \square

³Note that this is always satisfied for basis matrices from a coherent configuration.

Algorithm 4.3 (Finding an optimal solution for (4.3)).

Input: $w \in \mathbb{R}^q, y \in \mathbb{R}^q$

Sort $\{y_i/w_i\}$ such that $y_1/w_1 \geq \dots \geq y_q/w_q$

Set $k := \max_{1 \leq k \leq n} \{k \mid w_k^{-1}y_k + (\sum_{i=1}^k w_i)^{-1}(c - \sum_{i=1}^k y_i) > 0\}$

Set $\tau := (\sum_{i=1}^k w_i)^{-1}(c - \sum_{i=1}^k y_i)$

Set $x_i = \max\{w_i^{-1}y_i + \tau, 0\}$ for $i = 1, \dots, q$

Output: $x \in \mathbb{R}^q$

In our examples, see Sections 5.1 and 5.2, the x -subproblem (4.2) satisfies Assumption 4.1. Moreover, we have the following lemma.

Lemma 4.4. *The x -subproblem (4.2) satisfies Assumption 4.1, if*

$$P = \{x \in \mathbb{R}^q \mid \langle J, \mathcal{B}^*(x) \rangle = c, x \geq 0\}.$$

Proof. It holds that

$$(\tilde{B}(\tilde{B}^*(x)))_i = \langle \tilde{B}_i, \sum_{j=1}^q \tilde{B}_j x_j \rangle = \langle \tilde{B}_i, \tilde{B}_i x_i \rangle = \text{trace}(Q^T B_i^T Q Q^T B_i Q) x_i = w_i x_i, \quad (4.4)$$

where $w_i = \text{trace}(B_i^T B_i)$. Furthermore, $\langle J, \mathcal{B}^*(x) \rangle = w^T x$ with $w = (w_i) \in \mathbb{R}^q$. Thus we set $\mathcal{T} = \mathcal{B}$ and note that $\mathcal{T}(\mathcal{T}^*(x)) = \text{Diag}(w)x$. \square

5 Numerical results

We now demonstrate the efficiency of our new approach on two classes of problems: the quadratic assignment problem, **QAP**, and the graph partitioning problem, **GP**.

5.1 The quadratic assignment problem (QAP)

5.1.1 Background for the QAP

The Quadratic Assignment Problem was introduced in 1957 by Koopmans and Beckmann as a model for location problems that take into account the cost of placing a new facility on a certain site as well as the interaction with other facilities, i.e., a quadratic cost objective. The **QAP** contains the traveling salesman problem as a special case and is therefore NP-hard in the strong sense. It is generally considered to be one of the hardest of the NP-hard problems.

Let $A, B \in \mathbb{S}^n$ and let Π the set of $n \times n$ permutation matrices. The **QAP** can be stated as follows

$$\min_{X \in \Pi} \text{trace } AX^T BX.$$

The **QAP** is difficult to solve to optimality for large values of n , e.g., problems with $n \geq 30$ are still considered hard. It is well known that semidefinite programming relaxations provide strong bounds for the **QAP**, see [14, 51]. However even for sizes $n \geq 15$, it is difficult to solve the resulting **SDP** relaxation by interior point methods if one cannot exploit special structure such as symmetry. Solving the **DNN** relaxation is significantly more difficult.

Here, we first consider the **SDP** formulation for the **QAP** from Povh and Rendl [39], though we eventually solve the corresponding **DNN** relaxation:

$$\begin{aligned}
\min \quad & \text{trace}(A \otimes B)Y \\
\text{s.t.} \quad & \langle J_{n^2}, Y \rangle = n^2 \\
& \langle I_n \otimes (J_n - I_n) + (J_n - I_n) \otimes I_n, Y \rangle = 0 \\
& \langle I_n \otimes E_{ii}, Y \rangle = 1, \forall i = 1, \dots, n \\
& \langle E_{ii} \otimes I_n, Y \rangle = 1, \forall i = 1, \dots, n \\
& Y \succeq 0, Y \geq 0.
\end{aligned} \tag{5.1}$$

Here, I_n , or I when the meaning is clear, denotes the identity matrix of order n , and $E_{ii} = u_i u_i^T$, where $u_i \in \mathbb{R}^n$ is i -th unit vector. The **SDP** relaxation (5.1) is known to be equivalent to the **SDP** relaxation by Zhao et al. [51]. The authors in [14] show that

$$\mathcal{A}_{\mathcal{G}} = \mathcal{A}_{\text{aut}(A)} \otimes \mathcal{A}_{\text{aut}(B)},$$

where $\text{aut}(A)$ is the automorphism group of A . In the following Lemma 5.1 we derive the null space of the feasible solutions of (5.1), see also [46]. In what follows we use e_n or e when the meaning is clear, to denote the vector of all ones of order n .

Lemma 5.1. *Let $U = \frac{1}{\sqrt{n}}(nI - J) \in \mathbb{R}^{n \times n}$. Then the null space of the feasible solution Y of (5.1) is spanned by the columns of*

$$[U \otimes e_n \quad e_n \otimes U] \in \mathbb{R}^{n^2 \times 2n}.$$

Proof. Let $X \in \{0, 1\}^{n \times n}$ be a permutation matrix. Then $Xe_n = X^T e_n = e_n$, and thus

$$\begin{aligned}
(U \otimes e_n)^T \text{vec}(X) &= U^T e_n = 0, \\
(e_n \otimes U)^T \text{vec}(X) &= U^T e_n = 0.
\end{aligned}$$

Thus the columns of $[U \otimes e_n \quad e_n \otimes U]$ are in the null space of the barycenter of (5.1). The dimension of the column space of $[U \otimes e_n \quad e_n \otimes U]$ is $2(n-1)$, and the statement follows. \square

Let us now derive an exposing vector of the **SDP** relaxation (5.1).

Lemma 5.2. *The matrix*

$$W = I_n \otimes nJ_n + J_n \otimes (nI_n - 2J_n) \in \mathcal{S}_+^{n^2} \tag{5.2}$$

is an exposing vector of (5.1) of rank $2(n-1)$ in $\mathcal{A}_{\mathcal{G}}$.

Proof. Let U be defined as in Lemma 5.1. Using the properties of the Kronecker product, we have

$$\begin{aligned}
W &= [U \otimes e_n \quad e_n \otimes U] [U \otimes e_n \quad e_n \otimes U]^T \\
&= (UU^T) \otimes J + J \otimes (UU^T) \\
&= (nI - J) \otimes J + J \otimes (nI - J) \\
&= I \otimes nJ + J \otimes (nI - 2J)
\end{aligned} \tag{5.3}$$

as $UU^T = nI - J$. Clearly it holds that $W \succeq 0$, and from Lemma 5.1, we have W is an exposing vector of rank $2(n-1)$. Let P be any permutation matrix of order n . Then $P^T(UU^T)P = UU^T$ by construction. We have $(P_1 \otimes P_2)^T W (P_1 \otimes P_2) = W$ for any P_1, P_2 permutation matrices, and thus $W \in \mathcal{A}_{\mathcal{G}}$. \square

In the rest of this section, we first provide the facially reduced program and the symmetry reduced program of (5.1). Then we show how to do facial reduction for the symmetry reduced program based on our theory. The facially reduced formulation of (5.1) is also presented in [46]. We state it here for later use.

Lemma 5.3 ([46]). *The facially reduced program of the doubly nonnegative, **DNN** (5.1) is given by*

$$\begin{aligned}
\min \quad & \langle (V^T (A \otimes B) V), R \rangle \\
\text{s.t.} \quad & \langle V^T J V, R \rangle = n^2 \\
& \mathcal{G}(V R V^T) = 0 \\
& V R V^T \geq 0 \\
& R \in \mathcal{S}_+^{(n-1)^2+1},
\end{aligned} \tag{5.4}$$

where, by abuse of notation, \mathcal{G} is the transformation that imposes the zero structure,⁴ and the columns of $V \in \mathbb{R}^{n^2 \times (n-1)^2+1}$ form a basis of the null space of W , see Lemma 5.2.

Note that the constraints $\langle I \otimes E_{ii}, Y \rangle = 1$ and $\langle E_{ii} \otimes I, Y \rangle = 1$ become redundant after the facial reduction in (5.4).

We now discuss the symmetry reduced program. The symmetry reduced formulation of (5.1) is studied in [15]. We assume that the automorphism group of the matrix A is non-trivial. To simplify the presentation, we assume

$$A = \sum_{i=0}^d a_i A_i,$$

where $\{A_0, \dots, A_d\}$ is the basis of the commutant of the automorphism group of A . For instance the matrices A_i ($i = 0, 1, \dots, d$) may form a basis of the Bose-Mesner algebra of the Hamming scheme, see Example 2.2. Further, we assume from now on that A_0 is a diagonal matrix, which is the case for the Bose-Mesner algebra of the Hamming scheme. Here, we do not assume any structure in B . However the theory applies also when B has some symmetry structure and/or A_0 is not diagonal; see our numerical tests for the minimum cut problem in Section 5.2, below.

If the **SDP** (5.1) has an optimal solution $Y \in \mathcal{S}_+^{n^2}$, then it has an optimal solution of the form $Y = \sum_{i=0}^d A_i \otimes Y_i$ for some matrix variables $Y_0, \dots, Y_d \in \mathbb{R}^{n \times n}$, see Section 2.2. We write these matrix variables in a more compact way as $y = (\text{vec}(Y_0), \dots, \text{vec}(Y_d))$, if necessary. Denote by $\tilde{\mathcal{B}}_k^*(y) \in \mathcal{S}_+^{n^k}$ the k -th block of the block-diagonal matrix

$$\tilde{\mathcal{B}}^*(y) := (Q \otimes I)^T Y (Q \otimes I) = \sum_{i=0}^d (Q^T A_i Q) \otimes Y_i, \tag{5.5}$$

where Q is the orthogonal matrix block-diagonalizing A_i ($i = 0, \dots, d$).

⁴We use \mathcal{G} as the group and as a linear operator, usually referred to as the *gangster operator*, since the meaning is clear from the context.

Lemma 5.4. *The symmetry reduced program of the **DNN** relaxation (5.1) is given by*

$$\begin{aligned}
\min \quad & \sum_{i=0}^d a_i \text{trace}(A_i A_i) \text{trace}(B Y_i) \\
\text{s.t.} \quad & \sum_{i=0}^d \text{trace}(J A_i) \text{trace}(J Y_i) = n^2 \\
& \text{offDiag}(Y_0) = 0 \\
& \text{diag}(Y_i) = 0, \quad i = 1, \dots, d \\
& \text{diag}(Y_0) = \frac{1}{n} e_n \\
& Y_j \geq 0, \quad j = 0, \dots, d \\
& \tilde{\mathcal{B}}_k^*(y) \in \mathcal{S}_+^{n_k}, \quad k = 1, \dots, t,
\end{aligned} \tag{5.6}$$

where $\tilde{\mathcal{B}}_k^*(y)$ is the k -th block from (5.5) and offDiag is the linear transformation that zeros out the off-diagonal elements.

Proof. See e.g., [14, 46]. □

It remains to facially reduce the symmetry reduced program (5.6). Note that $W \in A_G$ can be written as $W = \sum_{i=0}^d A_i \otimes W_i$ for some matrices $W_0, \dots, W_d \in \mathbb{R}^{n \times n}$. Theorem 3.6 shows that the block-diagonal matrix

$$\widetilde{W} := (Q \otimes I)^T W (Q \otimes I) = \sum_{i=0}^d (Q^T A_i Q) \otimes W_i \tag{5.7}$$

is an exposing vector of the symmetry reduced program (5.6). Further, we denote by \widetilde{W}_k ($k = 1, \dots, t$) the k -th block of \widetilde{W} . Let us show how this applies to an example.

Example 5.5. *Let us consider Example 2.2 where A_i ($i = 0, \dots, d$) form a basis of the Bose-Mesner algebra of the Hamming scheme. Then, the exposing vector $W \in \mathcal{S}_+^{n^2}$ defined in Lemma 5.2 can be written as $W = \sum_{i=0}^d A_i \otimes W_i$ where*

$$W_0 = (n-2)J + nI \text{ and } W_i = nI_n - 2J \text{ for } i = 1, \dots, d. \tag{5.8}$$

Let $\widetilde{W}_k \in \mathcal{S}^n$ be the k -th block of \widetilde{W} , see (5.7). Then there are $d+1$ distinct blocks given by $\widetilde{W}_k = \sum_{i=0}^d p_{i,k} W_i \in \mathcal{S}^n$ for $k = 0, \dots, d$ where $p_{i,k}$ are elements in the character table P of the Hamming scheme, see Example 2.2. Using the fact that $Pe = (n, 0, \dots, 0)^T$ and $p_{1,k} = 1$ for every $k = 0, \dots, d$, we have

$$\widetilde{W}_0 = n^2 I - nJ \text{ and } \widetilde{W}_k = nJ \text{ for } k = 1, \dots, d, \tag{5.9}$$

and the matrices \tilde{V}_k , whose columns form a basis of the null space of $\widetilde{W}_k \in \mathcal{S}^n$, are given by

$$\tilde{V}_0 = e_n \text{ and } \tilde{V}_k = \begin{bmatrix} I_{n-1} \\ -e_{n-1}^T \end{bmatrix} \in \mathbb{R}^{n \times (n-1)} \text{ for } k = 1, \dots, d. \tag{5.10}$$

Similar results can be derived when one uses different groups. Now we are ready to present an **SDP** relaxation for the **QAP** which is both facially and symmetry reduced.

Proposition 5.6. *The facially reduced program of the symmetry reduced **DNN** relaxation (5.6) is given by*

$$\begin{aligned}
\min \quad & \sum_{i=1}^d a_i \text{trace}(A_i A_i) \text{trace}(B Y_i) \\
\text{s.t.} \quad & \sum_{i=0}^d \text{trace}(J A_i) \text{trace}(J Y_i) = n^2 \\
& \text{offDiag}(Y_0) = 0 \\
& \text{diag}(Y_i) = 0, i = 1, \dots, d \\
& Y_j \geq 0, j = 0, \dots, d \\
& \tilde{\mathcal{B}}_k^*(y) = \tilde{V}_k \tilde{R}_k \tilde{V}_k^T, k = 1, \dots, t \\
& \tilde{R}_k \in \mathcal{S}_+^{n'_k}, k = 1, \dots, t.
\end{aligned} \tag{5.11}$$

Here, the columns of $\tilde{V}_k \in \mathbb{R}^{n_k \times n'_k}$ form a basis of the null space of the $\tilde{W}_k \in \mathcal{S}^n$.

Proof. Applying Theorem 3.6 to the block-diagonal matrix $(Q \otimes I)^T W (Q \otimes I) = \sum_{i=0}^d (Q^T A_i Q) \otimes W_i$, the matrices \tilde{W}_k are the exposing vectors of the symmetry reduced program (5.6), and thus $\tilde{W}_k \tilde{\mathcal{B}}_k^*(y) = 0$ for every $k = 1, \dots, t$. This means that there exists a full column rank matrix $\tilde{V}_k \in \mathbb{R}^{n_k \times n'_k}$ such that $\tilde{\mathcal{B}}_k^*(y) = \tilde{V}_k \tilde{R}_k \tilde{V}_k^T$ where $\tilde{R}_k \in \mathcal{S}_+^{n'_k}$ for every $k = 1, \dots, t$. Finally, we apply Theorem 3.7 to remove redundant constraints, see also Lemma 5.3, and this yields the formulation (5.11). \square

Note that in the case that the basis elements A_i ($i = 0, \dots, d$) belong to the Hamming scheme, see Example 5.5, it follows that $t = d + 1$ in the above proposition.

5.1.2 On solving QAP with ADMM

Now we discuss how to use **ADMM** to solve the **DNN** relaxation (5.11), for the particular case when A_i ($i = 0, 1, \dots, d$) form a basis of the Bose-Mesner algebra of the Hamming scheme. We proceed similarly as in Section 4, and exploit properties of the known algebra, see Example 2.2. Clearly, for any other algebra we can proceed in the similar way. We assume without loss of generality that all the matrices \tilde{V}_j in this section have orthonormal columns.

First, we derive the equivalent reformulation of the **DNN** relaxation (5.11), by exploiting the following.

- (1) Since we remove the repeating blocks of positive semidefinite constraints, to use **ADMM** we have to reformulate the **DNN** in such a way that Assumption 4.1 is satisfied. Let us first derive an expression for the objective function as follows.

$$\begin{aligned}
\text{trace}((A \otimes B)Y) &= \text{trace} \left((Q \otimes I)^T \left(\sum_{i=0}^d a_i A_i \otimes B \right) (Q \otimes I) (Q \otimes I)^T \left(\sum_{j=0}^d A_j \otimes Y_j \right) (Q \otimes I) \right) \\
&= \text{trace} \left(\left(\sum_{i=0}^d (Q^T a_i A_i Q) \otimes B \right) \left(\sum_{j=0}^d (Q^T A_j Q) \otimes Y_j \right) \right) \\
&= \sum_{k=0}^d \mu_k \text{trace} \left(\left(\sum_{i=0}^d a_i p_{i,k} B \right) \left(\sum_{j=0}^d p_{j,k} Y_j \right) \right) \\
&= \sum_{k=0}^d \langle \tilde{C}_k, \sqrt{\mu_k} \sum_{i=0}^d p_{i,k} Y_i \rangle,
\end{aligned}$$

where $\tilde{C}_k := \sqrt{\mu_k} \left(\sum_{i=0}^d a_i p_{i,k} \right) B$. Recall that $\mu = (\mu_k) \in \mathbb{R}^{d+1}$, with $\mu_k := \binom{d}{k} (q-1)^k$. Then, we multiply the coupling constraints $\tilde{\mathcal{B}}_i^*(y) = \tilde{V}_i \tilde{R}_i \tilde{V}_i^T$ by the square root of its multiplicities. Thus, for the Bose-Mesner algebra, we end up with $\sqrt{\mu_j} \left(\sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T \right) = 0$.

- (2) In many applications, it is not necessary to compute high-precision solutions, and the **ADMM** can be terminated at any iteration. Then, one can use the dual variable \tilde{Z}_j from the current iteration to compute a valid lower bound, see Lemma 5.8. By adding redundant constraints, the so obtained lower bound is improved significantly when the **ADMM** is terminated with low-precision. Therefore we add the following redundant constraints

$$Y_0 = \frac{1}{n}I, \quad \text{trace}(\tilde{R}_j) = \sqrt{\mu_j}p_{0,j} \text{ for } j = 0, \dots, d. \quad (5.12)$$

To see the redundancy of the last $d+1$ constraints above, we use the fact that the columns of \tilde{V}_j are orthonormal, and that $\text{diag}(Y_i) = 0$ ($i = 1, \dots, d$), to derive

$$\text{trace}(\tilde{R}_j) = \text{trace}(\tilde{V}_j \tilde{R}_j \tilde{V}_j^T) = \text{trace} \sqrt{\mu_j} \left(\sum_{i=0}^d p_{i,j} Y_i \right) = \sqrt{\mu_j} p_{0,j}.$$

This technique can also be found in [23, 28, 34, 37].

We would like to emphasize that the techniques above are not restricted to the Bose-Mesner algebra of the Hamming scheme. Let us to present our reformulated **DNN** relaxation for **ADMM**. Define

$$\mathcal{P} := \{(Y_0, \dots, Y_d) \mid \sum_{i=0}^d \binom{d}{i} (q-1)^i q^d \text{trace}(JY_i) = n^2, Y_0 = \frac{1}{n}I, \text{diag}(Y_i) = 0, Y_j \geq 0, i = 1, \dots, d\} \quad (5.13)$$

and

$$\tilde{\mathcal{R}} := \{(\tilde{R}_0, \dots, \tilde{R}_d) \mid \text{trace}(\tilde{R}_j) = \sqrt{\mu_j}p_{0,j}, \tilde{R}_i \in \mathcal{S}_+^n, i = 0, \dots, d\}. \quad (5.14)$$

We obtain the following **DNN** relaxation for our **ADMM**.

$$\begin{aligned} p^* := \min \quad & \sum_{j=0}^d \langle \tilde{C}_j, \sqrt{\mu_j} \sum_{i=0}^d p_{i,j} Y_i \rangle \\ \text{s.t.} \quad & (Y_0, \dots, Y_d) \in \mathcal{P} \\ & (\tilde{R}_0, \dots, \tilde{R}_d) \in \tilde{\mathcal{R}} \\ & \sqrt{\mu_j} \left(\sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T \right) = 0, j = 0, \dots, d. \end{aligned} \quad (5.15)$$

The augmented Lagrangian is

$$\mathcal{L}(\tilde{Y}, \tilde{R}, \tilde{Z}) := \sum_{j=0}^d \left(\langle \tilde{C}_j, \sqrt{\mu_j} \sum_{i=0}^d p_{i,j} Y_i \rangle + \langle \tilde{Z}_j, \sqrt{\mu_j} \left(\sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T \right) \rangle + \frac{\beta}{2} \left\| \sqrt{\mu_j} \left(\sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T \right) \right\|^2 \right).$$

The Y -subproblem, the \tilde{R} -subproblem and the dual update are represented below.

1. The Y -subproblem:

$$\begin{aligned} \min \quad & \sum_{j=0}^d \left\| \sqrt{\mu_j} \sum_{i=0}^d p_{i,j} Y_i - \sqrt{\mu_j} \tilde{V}_j \tilde{R}_j \tilde{V}_j^T + \frac{\tilde{C}_j + \tilde{Z}_j}{\beta} \right\|^2 \\ \text{s.t.} \quad & Y_0 = \frac{1}{n}I \\ & \text{diag}(Y_i) = 0, i = 1, \dots, d \\ & \sum_{i=0}^d \binom{d}{i} (q-1)^i q^d \text{trace}(JY_i) = n^2 \\ & Y_i \geq 0, i = 0, \dots, d. \end{aligned} \quad (5.16)$$

2. The \tilde{R} -subproblems:

$$\begin{aligned} \min \quad & \|\tilde{R}_j - \tilde{V}_j^T (\sum_{i=0}^d p_{i,j} Y_i + \frac{\tilde{Z}_j}{\beta \sqrt{\mu_j}}) \tilde{V}_j\|^2 \\ \text{s.t.} \quad & \tilde{R}_j \in \mathcal{S}_+^{n'_j}, \end{aligned} \quad (5.17)$$

for $j = 0, \dots, d$.

3. Update the dual variable:

$$\tilde{Z}_j \leftarrow \tilde{Z}_j + \gamma \beta \sqrt{\mu_j} \left(\sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T \right), \quad j = 0, \dots, d. \quad (5.18)$$

Clearly, the \tilde{R} -subproblems can be solved in the same way as (4.1). To see that the Y -subproblem can also be solved efficiently, let us show that it is a problem of the form (4.3), and thus satisfies Assumption 4.1.

Let $\lambda_j = (p_{0,j}, \dots, p_{d,j})^T$,

$$y = \begin{bmatrix} \text{vec}(Y_0) \\ \vdots \\ \text{vec}(Y_d) \end{bmatrix} \quad \text{and} \quad \hat{y} = \begin{bmatrix} \text{vec}(\sqrt{\mu_0} \tilde{V}_0 \tilde{R}_0 \tilde{V}_0^T - \frac{\tilde{C}_0 + \tilde{Z}_0}{\beta}) \\ \vdots \\ \text{vec}(\sqrt{\mu_d} \tilde{V}_d \tilde{R}_d \tilde{V}_d^T - \frac{\tilde{C}_d + \tilde{Z}_d}{\beta}) \end{bmatrix}.$$

Define the linear transformation $\mathcal{T}^* : \mathbb{R}^{(d+1)n^2} \rightarrow \mathbb{R}^{(d+1)n^2}$ by

$$\mathcal{T}^*(y) = \begin{bmatrix} \sqrt{\mu_0} (\lambda_0^T \otimes I_{n^2}) \\ \vdots \\ \sqrt{\mu_d} (\lambda_d^T \otimes I_{n^2}) \end{bmatrix} y.$$

Lemma 5.7. *The Y -subproblem (5.16) is equivalent to the following projection to the weighted simplex problem*

$$\begin{aligned} \min \quad & \|\mathcal{T}^*(y) - \hat{y}\|^2 \\ \text{s.t.} \quad & y_i = 0, i \in \mathcal{I} \\ & w^T y = n^2 \\ & y \geq 0, \end{aligned} \quad (5.19)$$

where $w := q^d (\mu \otimes e_{n^2}) \in \mathbb{R}^{(d+1)n^2}$, and \mathcal{I} contains the indices of y associated to the off-diagonal entries of Y_0 . Furthermore, the problem (5.19) satisfies Assumption 4.1.

Proof. One can verify that (5.16) and (5.19) are equivalent. Furthermore, it holds that

$$\begin{aligned} \mathcal{T}(\mathcal{T}^*(y)) &= \begin{bmatrix} \sqrt{\mu_0} (\lambda_0^T \otimes I_{n^2}) \\ \vdots \\ \sqrt{\mu_d} (\lambda_d^T \otimes I_{n^2}) \end{bmatrix}^T \begin{bmatrix} \sqrt{\mu_0} (\lambda_0^T \otimes I_{n^2}) \\ \vdots \\ \sqrt{\mu_d} (\lambda_d^T \otimes I_{n^2}) \end{bmatrix} y \\ &= \left(\sum_{j=0}^d \mu_j (\lambda_j^T \otimes I_{n^2})^T (\lambda_j^T \otimes I_{n^2}) \right) y \\ &= \left(\sum_{j=0}^d \mu_j \lambda_j \lambda_j^T \right) \otimes I_{n^2} y. \end{aligned}$$

Applying the orthogonality relation of the Krawtchouk polynomial (2.6), the (r, s) -th entry of $\sum_{j=0}^d \mu_j \lambda_j \lambda_j^T$ is $\sum_{j=0}^d \mu_j p_{r,j} p_{s,j} = q^d \binom{d}{s} (q-1)^s \delta_{r,s} = q^d \mu_s \delta_{r,s}$ for $r, s = 0, \dots, d$. Thus $\mathcal{T}(\mathcal{T}^*(y)) = \text{Diag}(w)y$ and Assumption 4.1 is satisfied. \square

To efficiently solve the Y -subproblem for the **QAP**, we use Algorithm 4.3. Finally we describe how to obtain a valid lower bound when the **ADMM** model is solved approximately.

Lemma 5.8. *Let \mathcal{P} be the feasible set defined in (5.13), and consider the problem in (5.15). For any $\tilde{Z} = (\tilde{Z}_0, \dots, \tilde{Z}_d)$, the objective value*

$$g(\tilde{Z}) := \min_{(Y_0, \dots, Y_d) \in \mathcal{P}} \sum_{j=0}^d \langle \tilde{C}_j + \tilde{Z}_j, \sqrt{\mu_j} \sum_{i=0}^d p_{i,j} Y_i \rangle - \sum_{j=0}^d \mu_j p_{0,j} \lambda_{\max}(\tilde{V}_j^T \tilde{Z}_j \tilde{V}_j) \leq p^*, \quad (5.20)$$

i.e., it provides a lower bound to the optimal value p^ of (5.15).*

Proof. The dual of (5.15) with respect to the constraints $\sqrt{\mu_j}(\sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T) = 0$ is

$$d^* := \max_{(\tilde{Z}_0, \dots, \tilde{Z}_d)} \min_{\substack{(Y_0, \dots, Y_d) \in \mathcal{P} \\ (\tilde{R}_0, \dots, \tilde{R}_d) \in \mathcal{R}}} \sum_{j=0}^d \langle \tilde{C}_j, \sqrt{\mu_j} \sum_{i=0}^d p_{i,j} Y_i \rangle + \langle \tilde{Z}_j, \sqrt{\mu_j} (\sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T) \rangle. \quad (5.21)$$

The inner minimization problem can be written as

$$\min_{(Y_0, \dots, Y_d) \in \mathcal{P}} \sum_{j=0}^d \langle \tilde{C}_j + \tilde{Z}_j, \sqrt{\mu_j} \sum_{i=0}^d p_{i,j} Y_i \rangle + \min_{(\tilde{R}_0, \dots, \tilde{R}_d) \in \mathcal{R}} \sum_{j=0}^d \langle \tilde{Z}_j, \sqrt{\mu_j} (-\tilde{V}_j \tilde{R}_j \tilde{V}_j^T) \rangle. \quad (5.22)$$

It follows from the Rayleigh Principle, that the optimal value of the second minimization problem is $-\sum_{j=0}^d \mu_j p_{0,j} \lambda_{\max}(\tilde{V}_j^T \tilde{Z}_j \tilde{V}_j)$. Using strong duality, we have $g(\tilde{Z}) \leq d^* = p^*$. \square

5.1.3 Numerical results for the QAP

In this section we provide numerical results on solving the facially and symmetry reduced **DNN** relaxation (5.11). We used in this and remaining sections a computer with Two Intel Xeon E5-2637v3 4-core 3.5 GHz (Haswell) processors and 64GB of memory, unless specified differently. We use Mosek as the interior point solver, see [1]. We include *huge* problems of sizes up to 512.

Given a tolerance parameter ϵ , we terminate the **ADMM** when one of the following conditions is satisfied.

- The primal and dual residuals are smaller than ϵ , i.e.,

$$pres := \sum_{j=0}^d \left\| \sum_{i=0}^d p_{i,j} Y_i - \tilde{V}_j \tilde{R}_j \tilde{V}_j^T \right\| < \epsilon \text{ and } dres := \|\tilde{Z}^{\text{old}} - \tilde{Z}^{\text{new}}\| \leq \epsilon.$$

- Let p_k be the **ADMM** objective value, and $d_k := g(\tilde{Z})$ the dual objective value at some dual feasible point value at the k -th iteration, see (5.20). If the duality gap is not improving significantly, i.e.,

$$gap = \frac{p_{100k} - d_{100k}}{1 + p_{100k} + d_{100k}} < 10^{-4}$$

for 20 consecutive integers k , then we conclude that there is a stagnation in the objective value. (Note that we measure the gap only every 100 iterations, because computing the dual objective value d_k at every iteration is expensive.)

In our **QAP** experiments, we use $\epsilon = 10^{-12}$ if $n \leq 128$ and $\epsilon = 10^{-5}$ when $n = 256, 512$. The objective value from the **ADMM** is denoted by OBJ, and the valid lower bound obtained from the dual feasible solution is denoted by LB, see Lemma 5.8. The running times in all

tables are reported in seconds. We also list the maximum of the primal and dual residuals, i.e., $res := \max\{pres, dres\}$. If a result is not available, we put ‘-’ in the corresponding entry.

The first set of test instances are from Mittelman and Peng [33], where the authors compute **SDP** bounds for the **QAP** with A being the Hamming distance matrix. Choices of the matrix B^5 differ for different types of instances. In particular, in the Harper instance Harper- n where $n = 2^d$ we set $B_{ij} = |i - j|$ for all $i, j = 1, \dots, 2^d$. Further eng1- n and eng9- n with $n = 2^d$, $d = 4, \dots, 9$ refer to the engineering problems, and VQ- n instances have random matrices B . For details see [33]. In rand_256 and rand_512 instances, A is the Hamming distance matrix of appropriate size and B is a random matrix.

Table 5.1 reads as follows. In the first column we list the instance names where the sizes of the **QAP** matrices are indicated after the underscore. Upper bounds are given in the column two. For instances with up to 128 nodes we list the upper bounds computed in [33], and for the remaining instances we use our heuristics. Since data matrices for the Harper instances are integer, we round up lower bounds to the closest integer. In the column three (resp. four) we list **SDP**-based lower bounds (resp. computation times in seconds) from [33]. The bounds from [33] are obtained by solving an **SDP** relaxation having several matrix variables on order n . The bounds in [33] were computed on a 2.67GHz Intel Core 2 computer with 4GB memory. In the columns five to seven, we present the results obtained by using our **ADMM** algorithm.

Table 5.1 shows that we significantly improve bounds for all eng1- n and eng9- n instances. Moreover, we are able to compute bounds for huge QAP instances with $n = 256$ and $n = 512$ in a reasonable amount of time. Note that for each instance from Table 5.1 of size $n = 2^d$, the **DNN** relaxation boils down to $d + 1$ positive semidefinite blocks of order n . Clearly, there is no interior point algorithm that is able to solve so large problems.

⁵We thank Hans Mittelman for providing us generators for the mentioned instances.

		[33]		ADMM			
problem	UB	LB	time	OBJ	LB	time	res.
Harper_16	2752	2742	1	2743	2742	1.92	4.50e-05
Harper_32	27360	27328	3	27331	27327	9.70	1.67e-04
Harper_64	262260	262160	56	262196	261168	36.12	1.12e-05
Harper_128	2479944	2446944	1491	2446800	2437880	186.12	3.86e-05
Harper_256	22370940	-	-	22369996	22205236	432.10	9.58e-06
Harper_512	201329908	-	-	201327683	200198783	1903.66	9.49e-06
eng1_16	1.58049	1.5452	1	1.5741	1.5740	2.28	3.87e-05
eng1_32	1.58528	1.24196	4	1.5669	1.5637	14.63	5.32e-06
eng1_64	1.58297	0.926658	56	1.5444	1.5401	38.35	4.69e-06
eng1_128	1.56962	0.881738	1688	1.4983	1.4870	389.04	2.37e-06
eng1_256	1.57995	-	-	1.4820	1.3222	971.48	9.95e-06
eng1_512	1.53431	-	-	1.4553	1.3343	9220.13	9.66e-06
eng9_16	1.02017	0.930857	1	1.0014	1.0013	3.58	2.11e-06
eng9_32	1.40941	1.03724	3	1.3507	1.3490	12.67	3.80e-05
eng9_64	1.43201	0.887776	68	1.3534	1.3489	74.89	6.60e-05
eng9_128	1.43198	0.846574	2084	1.3331	1.3254	700.27	8.46e-06
eng9_256	1.45132	-	-	1.3152	1.2610	1752.72	9.74e-06
eng9_512	1.45914	-	-	1.3074	1.1168	23191.96	9.96e-06
VQ_32	297.29	294.49	3	296.3241	296.1351	11.82	1.27e-05
VQ_64	353.5	352.4	45	352.7621	351.4358	43.17	4.22e-04
VQ_128	399.09	393.29	2719	398.4269	396.2794	282.28	6.19e-04
rand_256	126630.6273	-	-	124589.4215	124469.2129	2054.61	3.78e-05
rand_512	577604.8759	-	-	570935.1468	569915.3034	9694.71	1.32e-04

Table 5.1: Lower and upper bounds for different **QAP** instances.

The second set of test instances are Eschermann and Wunderlich instances from the QAPLIB library [9]. In *esc_nx* instance, the distance matrix A is the Hamming distance matrix of order $n = 2^d$, whose automorphism group is the automorphism group of the Hamming graph $H(d, 2)$. In [14] the authors exploit symmetry in *esc* instances to solve the **DNN** relaxation (5.6) by the interior point method. That was the first time that one computes **SDP** bounds for large **QAP** instances by exploiting symmetry. In particular, the authors from [14] needed 98 seconds to compute the **SDP** bound for *esc64a*, and 52 seconds for computing *esc128* **SDP** bound. The bounds in [14] are computed by the interior point solver SeDuMi [45] using the Yalmip interface [31] and Matlab 6.5, implemented on a PC with Pentium IV 3.4 GHz dual-core processor and 3GB of memory.

In [34] the authors approximately solve the **DNN** relaxation (5.6) using the **ADMM** algorithm, but do not exploit symmetry. Here, we compare computational results from [34] with the approach we present in this paper. All the instances from [34] were tested on an Intel Xeon Gold 6130 2.10 Ghz PC with 32 cores and 64 GB of memory and running on 64-bit Ubuntu system.

In Table 5.2 we present numerical result for the *esc* instances, and we conclude that:

- There are notably large differences in computation times between the **ADMM** algorithm presented here and the one from [34], since the latter does not exploit symmetry.
- In [14], the authors use SeDuMi to solve an equivalent relaxation to the symmetry reduced program (5.6) and obtain 53.0844 for *esc128*. However, the bound for the same problem instance and for the facially and symmetry reduced program (5.11) computed by the interior point method solver of Mosek is 51.7516. Note that our **ADMM** algorithms reports 51.7518.

		ADMM from [34]		ADMM			
problem	opt	LB	time	OBJ	LB	time	res
esc16a	68	64	20.14	63.2856	63.2856	2.48	1.17e-11
esc16b	292	290	3.10	290.0000	290.0000	0.78	9.95e-13
esc16c	160	154	8.44	154.0000	153.9999	2.11	2.56e-09
esc16d	16	13	17.39	13.0000	13.0000	1.04	9.94e-13
esc16e	28	27	24.04	26.3368	26.3368	1.21	9.89e-13
esc16f	0	0	3.22e+02	0.0000	-0.0000	0.01	2.53e-14
esc16g	26	25	33.54	24.7403	24.7403	1.40	9.95e-13
esc16h	996	977	4.01	976.2293	976.2293	2.51	7.73e-13
esc16i	14	12	100.79	11.3749	11.3660	6.15	2.53e-06
esc16j	8	8	56.90	7.7942	7.7942	0.21	9.73e-13
esc32a	130	104	2.89e+03	103.3211	103.0465	12.36	3.62e-06
esc32b	168	132	2.52e+03	131.8843	131.8843	4.64	9.59e-13
esc32c	642	616	4.48e+02	615.1813	615.1813	8.04	2.05e-10
esc32d	200	191	8.68e+02	190.2271	190.2263	5.86	7.45e-08
esc32e	2	2	1.81e+03	1.9000	1.9000	0.70	4.49e-13
esc32f	2	2	1.80e+03	1.9000	1.9000	0.76	4.49e-13
esc32g	6	6	6.04e+02	5.8333	5.8333	3.50	9.97e-13
esc32h	438	425	3.02e+03	424.4027	424.3184	5.89	1.03e-06
esc64a	116	98	1.64e+04	97.7500	97.7500	5.33	8.95e-13
esc128	64	-	-	51.7518	51.7518	137.71	1.18e-12

Table 5.2: Esc instances

5.2 The graph partition problem (GP)

5.2.1 The general GP

The graph partition problem is the problem of partitioning the vertex set of a graph into a fixed number of sets of given sizes such that the sum of edges joining different sets is optimized. The problem is known to be NP-hard and it has many applications. The **GP** has many applications such as VLSI design, parallel computing, network partitioning, and floor planing. Graph partitioning also plays a role in machine learning (see e.g., [27]) and data analysis (see e.g., [36]). There exist several **SDP** relaxations for the **GP** of different complexity and strength, see e.g., [25, 42, 43, 48, 50].

Let $G = (V, E)$ be an undirected graph with vertex set V , $|V| = n$ and edge set E , and $k \geq 2$ be a given integer. We denote by A the adjacency matrix of G . The goal is to find a partition of the vertex set into k (disjoint) subsets S_1, \dots, S_k of specified sizes $m_1 \geq \dots \geq m_k$, where $\sum_{j=1}^k m_j = n$, such that the sum of weights of edges joining different sets S_j is minimized. Let

$$P_m := \left\{ S = (S_1, \dots, S_k) : S_i \subset V, |S_i| = m_i, \forall i, S_i \cap S_j = \emptyset, i \neq j, \cup_{i=1}^k S_i = V \right\} \quad (5.23)$$

denote the set of all partitions of V for a given $m = (m_1, \dots, m_k)$. In order to model the **GP** in binary variables we represent the partition $S \in P_m$ by the partition matrix $X \in \mathbb{R}^{n \times k}$ where the column j is the incidence vector for the set S_j .

The **GP** can be stated as follows

$$\min_{X \in \mathcal{M}_m} \frac{1}{2} \text{trace}(AX(J_k - I_k)X^T),$$

where

$$\mathcal{M}_m = \{X \in \{0, 1\}^{n \times k} : Xe_k = Xe_n, X^T e_n = m\} \quad (5.24)$$

is the set of partition matrices.

Here, we consider the following **SDP** relaxation that is equivalent to the **SDP** relaxation from [50]:

$$\begin{aligned}
\min \quad & \frac{1}{2} \text{trace}(A \otimes B)Y \\
\text{s.t.} \quad & \mathcal{G}(Y) = 0 \\
& \text{trace}(D_1 Y) - 2(e_n \otimes e_k)^T \text{diag}(Y) + n = 0 \\
& \text{trace}(D_2 Y) - 2(e_n \otimes m)^T \text{diag}(Y) + m^T m = 0 \\
& \mathcal{D}_O(Y) = \text{Diag}(m) \\
& \mathcal{D}_e(Y) = e \\
& \langle J, Y \rangle = n^2 \\
& Y \succeq 0, Y \succeq 0,
\end{aligned} \tag{5.25}$$

where $B = J_k - I_k$, and

$$Y = \begin{bmatrix} Y^{(11)} & \dots & Y^{(1n)} \\ \vdots & \ddots & \vdots \\ Y^{(n1)} & \dots & Y^{(nn)} \end{bmatrix} \in \mathcal{S}^{kn}$$

with each $Y^{(ij)}$ being a $k \times k$ matrix, and

$$\begin{aligned}
D_1 &= I_n \otimes J_k \\
D_2 &= J_n \otimes I_k \\
\mathcal{D}_O(Y) &= \sum_{i=1}^n Y^{ii} \in \mathcal{S}^k \\
\mathcal{D}_e(Y) &= (\text{trace } Y^{ii}) \in \mathbb{R}^n \\
\mathcal{G}(Y) &= \langle I_n \otimes (J_k - I_k), Y \rangle.
\end{aligned}$$

To compute **SDP** bounds for the **GP**, we apply facial reduction for symmetry reduced relaxation (5.25). The details are similar to the **QAP**, and thus omitted.

We present numerical results for different graphs from the literature. Matrix can161 is from the library Matrix Market [4], matrix grid3dt5 is 3D cubical mesh, and gridtxx matrices are 2D triangular meshes. Myciel7 is a graph based on the Mycielski transformation and 1_FullIns.4 graph is a generalization of the Mycielski graph. Both graphs are used in the COLOR02 symposium [24].

In Table 5.3 we provide information on the graphs and the considered 3-partition problems. In particular, the first column specifies graphs, the second column provides the number of vertices in a graph, the third column is the number of orbits after symmetrization, the fourth column lists the number of blocks in $Q^T A Q$. Here, the orthogonal matrix Q is computed by using the heuristic from [14]. The last column specifies sizes of partitions.

Table 5.4 lists lower bounds obtained by using Mosek and our **ADMM** algorithm. The table also presents computational times required to compute bounds by both methods as well as the number of interior point method iterations. The results show that the **ADMM** with precision $\epsilon = 10^{-3}$ provides competitive bounds in much shorter time than the interior point method solver. In Table 5.4, some instances are marked by *. This means that our 64GB machine did not have enough memory to solve these instances by the interior point method solver, and therefore they are solved on a machine with an Intel(R) Xeon(R) Gold 6126, 2.6 GHz quad-core processor and 192GB of memory. However, the **ADMM** algorithm has much lower memory requirements, and thus the **ADMM** is able to solve all instances from Table 5.4 on the smaller machine.

instance	$ V $	# orbits	blocks of A	m
1_FullIns.4	93	3629	(53,27,9,3,1)	(30,31,32)
can161	161	921	(20,20,20,20,20,20,20,11,10)	(52,53,56)
grid3dt5	125	4069	(39,36,26,24)	(40,41,44)
gridt15	120	2432	(80,24,16)	(39,40,41)
gridt17	153	3942	(102,30,21)	(50,51,52)
myciel7	191	6017	(64,64,63)	(62,63,66)

Table 5.3: Graphs and partitions

instance	IPM (Symmetry &Facially reduced)			ADMM ($\epsilon = 10^{-3}$)			
	LB	time	iter.	OBJ	LB	time	res
1_FullIns.4	194.2825	311.95	26	194.2686	194.0523	141.29	1.50e-01
can161	33.0151	124.32	19	33.0392	30.4470	19.74	2.58e-01
grid3dt5	68.3175	245.65	17	68.3029	68.0436	200.35	2.02e-01
gridt15	12.1153	1302.10	41	12.1116	11.3654	97.17	1.91e-01
gridt17*	12.2482	1865.67	21	12.2532	11.1459	357.53	1.80e-01
myciel7*	1126.0309	2579.65	17	1126.0385	1123.8526	553.67	9.50e-02

Table 5.4: Numerical results for the graph 3-partition.

5.2.2 The vertex separator problem (VSP) and min-cut (MC)

We consider the problem of partitioning the node set of a graph into k subsets of given sizes in order to minimize the cut obtained removing the k -th set. If the resulting cut has value zero, then one has obtained a vertex separator. The described problem is known as the vertex separator problem. This problem is NP-hard and it is related to the general graph partitioning problem. The vertex separator problem was studied by Helmberg, Mohar, Poljak and Rendl [22], Povh and Rendl [38], Rendl and Sotirov [40], Pong, Sun, Wang, Wolkowicz [37].

The **VSP** appears in many different fields such as VLSI design [3] and bioinformatics [18]. Finding vertex separators of minimal size is an important problem in communications network [26] and finite element methods [32]. The VSP also appears in divide-and-conquer algorithms for minimizing the work involved in solving system of equations, see e.g., [29, 30].

The **VSP** is closely related to the following graph partitioning problem. Let $\delta(S_i, S_j)$ denote the set of edges between S_i and S_j , where S_i and S_j are defined as in (5.23). We denote the set of edges with endpoints in distinct partition sets S_1, \dots, S_{k-1} by

$$\delta(S) = \cup_{i < j < k} \delta(S_i, S_j).$$

The min-cut (**MC**) problem is

$$\text{cut}(m) = \min\{|\delta(S)| : S \in P_m\}.$$

The graph has a vertex separator if there exists $S \in P_m$ such that after the removal of S_k the induced subgraph has no edges across S_i and S_j for $1 \leq i < j < k$. Thus, if $\text{cut}(m) = 0$ or equivalently $\delta(S) = \emptyset$, there exists a vertex separator. On the other hand $\text{cut}(m) > 0$ shows that no separator S_k for the cardinalities specified in m exists.

Clearly, $|\delta(S)|$ can be represented in terms of a quadratic function of the partition matrix X , i.e., as $\frac{1}{2} \text{trace}(AXBX^T)$ where

$$B := \begin{bmatrix} J_{k-1} - I_{k-1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{S}^k. \quad (5.26)$$

Therefore,

$$\text{cut}(m) = \min_{X \in \mathcal{M}_m} \frac{1}{2} \text{trace}(AXBX^T),$$

where \mathcal{M}_m is given in (5.24). To compute **SDP** bounds for the **MC** problem and provide bounds for the vertex separator problem, we use the **SDP** relaxation (5.25) with B defined as in (5.26).

We present numerical results for the Queen graphs, where the $n \times n$ Queen graph has the squares of $n \times n$ chessboard for its vertices and two such vertices are adjacent if the corresponding squares are in the same row, column, or diagonal. The instances in this class come from the DIMACS challenge on graph coloring. In Table 5.5 we provide information on the Queen graphs. The table is arranged in the same way as Table 5.3.

instance	$ V $	# orbits	blocks of A	m
queen5_5	25	91	(12,6,3,3,1)	(4,5,16)
queen6_6	36	171	(18,6,6,3,3)	(6,7,23)
queen7_7	49	325	(24,10,6,6,3)	(9,9,31)
queen8_8	64	528	(32,10,10,6,6)	(11,12,41)
queen9_9	81	861	(40,15,10,10,6)	(14,15,52)
queen10_10	100	1275	(50,15,15,10,10)	(18,18,64)
queen11_11	121	1891	(60,21,15,15,10)	(21,22,78)
queen12_12	144	2628	(72,21,21,15,15)	(25,26,93)
queen13_13	169	3655	(84,28,21,21,15)	(30,30,109)

Table 5.5: The Queen graphs and partitions

In Table 5.6 we provide the numerical results for the vertex separator problem. More specifically, we are computing the largest integer m_3 such that the solution value of the **DNN** relaxation (5.25) is positive with partition

$$m = (\lfloor \frac{n - m_3}{2} \rfloor, \lceil \frac{n - m_3}{2} \rceil, m_3). \quad (5.27)$$

Then $m_3 + 1$ is a lower bound for the vertex separator problem with respect to the choice of m . One may tend to solve (5.25) for all possible m_3 between $0, 1, \dots, |V| - 1$ to find the largest m_3 for which the **DNN** bound is positive. However, the optimal value of (5.25) is monotone in m_3 , and thus we find the appropriate m_3 using binary search starting with $m_3 = \lceil \frac{n}{2} \rceil$. We present the lower bound on the vertex separator, i.e., $m_3 + 1$ in the third column of Table 5.6. The total number of problems solved is listed in the fourth column of the same table. The running time given in the last two columns is the total amount of time used to find a positive lower bound for (5.25) for some m_3 by using Mosek and our **ADMM** algorithm, respectively. This task is particularly suitable for the **ADMM**, as we can terminate the **ADMM** once the lower bound in an iterate is positive. For example, it takes 786 seconds to solve the min-cut relaxation on queen12_12 by Mosek, see Table 5.7. However, though not shown in the table, it takes **ADMM** only 120 seconds to conclude that the optimal value is positive.

instance	$ V $	$m_3 + 1$	#problems	IPM (Symmetry&Facially reduced) time	ADMM ($\epsilon = 10^{-12}$) time
queen 5_5	25	17	4	7.49	2.69
queen 6_6	36	24	5	9.62	2.91
queen 7_7	49	32	5	25.34	5.95
queen 8_8	64	42	6	85.72	34.35
queen 9_9	81	53	6	304.44	64.10
queen 10_10	100	65	7	1309.85	131.66
queen 11_11	121	79	7	3416.01	387.38
queen 12_12	144	94	7	6147.20	671.02
queen 13_13	169	110	8	-	1352.17

Table 5.6: The vertex separator problem on the Queen graphs

In Table 5.7 we compare bounds and computational times required to solve, for fixed m , symmetry reduced **DNN** relaxation (5.25) by the interior point algorithm, as well as symmetry and facially reduced relaxation (5.25) by using Mosek our **ADMM** algorithm.

instance	IPM (Symmetry reduced)			IPM (Symmetry &Facially reduced)			ADMM ($\epsilon = 10^{-12}$)			
	LB	time	iter.	LB	time	iter.	OBJ	LB	time	res
queen5_5	0.0908	1.04	38	0.1658	0.27	10	0.1658	0.1658	6.88	7.36e-11
queen6_6	0.0962	3.43	31	0.1411	0.91	11	0.1411	0.1411	11.37	1.83e-10
queen7_7	0.5424	15.42	32	0.6196	1.92	10	0.6196	0.6196	17.97	5.53e-11
queen8_8	0.1967	127.60	39	0.3087	7.38	13	0.3087	0.3087	61.50	1.15e-10
queen9_9	0.0698	377.77	32	0.2175	19.98	12	0.2175	0.2175	204.39	1.16e-06
queen10_10	0.8159	1664.09	42	1.0211	85.42	14	1.0211	1.0211	239.75	1.09e-09
queen11_11	-	-	-	0.2131	275.20	16	0.2131	0.2131	877.85	1.82e-05
queen12_12	-	-	-	0.3248	786.12	25	0.3248	0.3248	1474.45	1.20e-06
queen13_13	-	-	-	-	-	-	0.9261	0.9261	1864.30	5.71e-09

Table 5.7: The min-cut problem on the Queen graphs

We conclude from Tables 5.6 and 5.7 that

- For small instances, the interior point algorithm is faster than the **ADMM** as shown in Table 5.7. For larger instances, the interior point algorithm has memory issues. However, the **ADMM** algorithm can still handle large instances due to its low memory demand.
- To obtain bounds on the the vertex separator of a graph, one does not need to solve the **DNN** relaxation to high-precision. The **ADMM** is able to exploit this fact, and find a lower bound on the size of the vertex separator in significantly less amount of time than the interior point algorithm, see Table 5.6.
- The symmetry reduced program is heavily ill-conditioned and the interior point method is not able to solve it correctly for any of the instances. The running time is also significantly longer than the symmetry and facially reduced program, see Table 5.7.

6 Conclusion

In this paper, we propose a method to implement facial reduction to the symmetry reduced **SDP**. More specifically, if an exposing vector of the minimal face for the input **SDP** is given, then we are able to construct an exposing vector of the minimal face for the symmetry reduced

SDP. The obtained **SDP** is symmetry reduced and satisfies Slater's condition, and thus can be solved with improved numerical stability. For many combinatorial problems, a facially reduced **SDP** relaxation can be solved extremely fast using alternating direction method of multipliers. We also show that a symmetry and facially reduced **SDP** can be solved more efficiently by interior point methods than only symmetry reduced **SDP**. As a result, we are able to compute improved lower bounds for some **QAP** instances in a significantly less amount of time.

Index

- A_G , commutant, 4
- $Af(\tilde{R}) = b$, 11
- $Ax = b$, 7, 10, 11
- $E_{ii} = u_i u_i^T$, 16
- I_n , 16
- J , the matrix of all ones, 5
- P , polyhedral constraints on x , 12
- \mathcal{F}_X , feasible set, 3
- \mathcal{F}_x , feasible set with x , 7
- \mathcal{G} , group of permutation matrices, 4
- \mathcal{G} , gangster operator, 17
- $\text{null}(U^T)$, 9
- Π , permutation matrices order n , 15
- $\mathcal{R}_G(X)$, Reynolds operator, 4
- $\text{range}(V)$, 9
- \mathcal{S}_x , feasible set with $\tilde{B}^*(x)$, 7
- \mathbb{S}^n , symmetric matrices, 3
- \mathbb{S}_+^n , positive semidefinite cone, 3
- $\delta_{r,s}$, Kronecker delta, 6
- $\text{face}(S)$, minimal face of \mathbb{S}_+^n containing S , 8
- $\hat{X} \succ 0$, 4
- $\hat{X} \succeq 0$, 4
- $\tilde{A}x = b$, 7
- \tilde{A}_j^k , 7
- $\tilde{B}_j := Q^T B_j Q =: \text{blkdiag}((\tilde{B}_j^k)_{k=1}^t)$, 6
- \tilde{B}_j^k , 6
- $\tilde{c} := \tilde{B}(\tilde{C})$, 7
- $\tilde{B}^*(x)$, 6
- \tilde{B}_k^* , 6
- $\mathcal{P}_{\mathbb{S}_+}(\cdot)$, projection onto positive semidefinite matrices, 13
- $\succeq 0$, positive semidefinite, 3
- $\{B_1, \dots, B_d\} \in \{0, 1\}^{n \times n}$, basis for A_G , 5
- d , dimension of basis for A_G , 5
- d_{SDP}^* , dual optimal value, 3
- e_n or e , 16
- p_{SDP}^* , primal optimal value, 3
- t , number of blocks, 6
- $t(k) = k(k+1)/2$, triangular number, 7
- $\mathcal{L}(x, R, Z)$, augmented Lagrangian, 13
- $\text{aut}(A)$, automorphism group of A , 16
- $\mathcal{B}^*(x)$, 5
- $\mathcal{A}(X)$, 3
- $\mathcal{A}^*(y)$, 3
- \mathcal{F}_R , 4
- $\mathcal{V}^*(R) = VRV^T$, 4
- DNN**, doubly nonnegative, 3, 17
- QAP**, quadratic assignment problem, 15
- SDP**, semidefinite program, 3
- adjoint, 3
- augmented Lagrangian, $\mathcal{L}(x, \tilde{R}, \tilde{Z})$, 12
- automorphism group of A , $\text{aut}(A)$, 16
- basis for A_G , $\{B_1, \dots, B_d\}$, 5
- centralizer ring, 4
- commutant, A_G , 4
- dimension of basis for A_G , d , 5
- doubly nonnegative, **DNN**, 3, 17
- dual optimal value, d_{SDP}^* , 3
- dual program, 3
- exposing vector, 4
- facial reduction, 4
- feasible set with $\tilde{B}^*(x)$, \mathcal{S}_x , 7
- feasible set with x , \mathcal{F}_x , 7
- feasible set, \mathcal{F}_X , 3
- gangster operator, \mathcal{G} , 17
- group average, 5
- group of permutation matrices, \mathcal{G} , 4
- Hamming graph, 5
- Kronecker delta, $\delta_{r,s}$, 6
- matrix *-algebra, 5
- minimal face, 2, 4, 7
- minimal face of \mathbb{S}_+^n containing S , $\text{face}(S)$, 8
- number of blocks, t , 6
- orbit, 5
- permutation matrices order n , Π , 15
- polyhedral constraints on x , P , 12
- positive semidefinite cone, \mathbb{S}_+^n , 3
- positive semidefinite, $\succeq 0$, 3
- primal optimal value, p_{SDP}^* , 3
- projection onto positive semidefinite matrices, $\mathcal{P}_{\mathbb{S}_+}(\cdot)$, 13

quadratic assignment problem, **QAP**, 15

repeated blocks, 7

Reynolds operator, $\mathcal{R}_G(X)$, 4

semidefinite program, **SDP**, 3

Slater constraint qualification, 3

substitution

 block diagonal symmetry reduction, 7

 facial reduction, 4

 symmetry reduction, 6

symmetric matrices, \mathbb{S}^n , 3

symmetry reduced formulation, 7

triangular number, $t(k) = k(k + 1)/2$, 7

References

- [1] Erling D Andersen and Knud D Andersen. The mosek interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In *High performance optimization*, pages 197–232. Springer, 2000. 22
- [2] C. Bachoc, D.C. Gijswijt, A. Schrijver, and F. Vallentin. Invariant semidefinite programs. In *Handbook on semidefinite, conic and polynomial optimization*, volume 166 of *Internat. Ser. Oper. Res. Management Sci.*, pages 219–269. Springer, New York, 2012. 4
- [3] S.N. Bhatt and F.T. Leighton. A framework for solving vlsi graph layout problems. *Journal of Computer and System Sciences*, 28(2):300–343, 1984. 27
- [4] R. F. Boisvert, R. Pozo, K. Remington, R. F. Barrett, and J. J. Dongarra. Matrix market: a web resource for test matrix collections. In *Quality of Numerical Software*, pages 125–137. Springer, 1997. 26
- [5] J.M. Borwein and H. Wolkowicz. Characterization of optimality for the abstract convex program with finite-dimensional range. *J. Austral. Math. Soc. Ser. A*, 30(4):390–411, 1980/81. 2, 4
- [6] J.M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 30(3):369–380, 1980/81. 2, 4
- [7] J.M. Borwein and H. Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. Appl.*, 83(2):495–530, 1981. 2, 4
- [8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Machine Learning*, 3(1):1–122, 2011. 12
- [9] R.E. Burkard, S.E. Karisch, and F. Rendl. QAPLIB—a quadratic assignment problem library. *J. Global Optim.*, 10(4):391–403, 1997. 24
- [10] Y. Chen and X. Ye. Projection onto a simplex. 2011. 14
- [11] L. Condat. Fast projection onto the simplex and the l_1 ball. *Mathematical Programming*, 158(1-2):575–585, 2016. 14
- [12] E. de Klerk. Exploiting special structure in semidefinite programming: A survey of theory and applications. *European J. Oper. Res.*, 201(1):1–10, 2010. 4
- [13] E. de Klerk, C. Dobre, and D.V. Pasechnik. Numerical block diagonalization of matrix *-algebras with application to semidefinite programming. *Math. Program.*, 129(1, Ser. B):91–111, 2011. 5
- [14] E. de Klerk and R. Sotirov. Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem. *Math. Program.*, 122(2, Ser. A):225–246, 2010. 12, 15, 16, 18, 24, 26
- [15] E. de Klerk and R. Sotirov. Improved semidefinite programming bounds for quadratic assignment problems with suitable symmetry. *Mathematical programming*, 133(1-2):75–91, 2012. 17

- [16] P. Delsarte. *An Algebraic Approach to the Association Schemes of Coding Theory*. PhD thesis, Universite Catholique de Louvain, Louvain, France, 1973. 5
- [17] D. Drusvyatskiy and H. Wolkowicz. The many faces of degeneracy in conic optimization. *Foundations and Trends[®] in Optimization*, 3(2):77–170, 2017. 4, 7
- [18] B. Fu, S.A. Oprisan, and L. Xu. Multi-directional width-bounded geometric separator and protein folding. In *International Symposium on Algorithms and Computation*, pages 995–1006. Springer, 2005. 27
- [19] K. Gatermann and P.A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *Journal of Pure and Applied Algebra*, 192(1-3):95–128, 2004. 4
- [20] K. Gatermann and P.A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *J. Pure Appl. Algebra*, 192(1-3):95–128, 2004. 5
- [21] D. Gijswijt. Matrix algebras and semidefinite programming techniques for codes. *PhD Thesis*, 2010. 4
- [22] C. Helmberg, F. Rendl, B. Mohar, and S. Poljak. A spectral approach to bandwidth and separator problems in graphs. *Linear and Multilinear Algebra*, 39(1-2):73–90, 1995. 27
- [23] H. Hu and R. Sotirov. On solving the quadratic shortest path problem. *INFORMS Journal on Computing*, 2019. 20
- [24] D. S. Johnson, A. Mehrotra, and M. A. Trick. COLOR02/03/04: Graph coloring and its generalizations, 2002. 26
- [25] S. E. Karisch and F. Rendl. Semidefinite programming and graph equipartition. *Topics in Semidefinite and Interior-Point Methods*, 18:77–95, 1998. 25
- [26] F.T. Leighton. Complexity issues in vlsi: Optimal layouts for the shuffle exchange and other networks, 1983. 27
- [27] M. Li, D. G. Andersen, and A. J. Smola. Graph partitioning via parallel submodular approximation to accelerate distributed machine learning. *arXiv preprint arXiv:1505.04636*, 2015. 25
- [28] X. Li, T.K. Pong, H. Sun, and H. Wolkowicz. A strictly contractive Peaceman-Rachford splitting method for the doubly nonnegative relaxation of the minimum cut problem. Technical report, University of Waterloo, Waterloo, Ontario, 2019. 40 pages, research report. 20
- [29] R.J. Lipton, D.J. Rose, and R.E. Tarjan. Generalized nested dissection. *SIAM journal on numerical analysis*, 16(2):346–358, 1979. 27
- [30] R.J. Lipton and R.E. Tarjan. Applications of a planar separator theorem. *SIAM journal on computing*, 9(3):615–627, 1980. 27
- [31] J. Löfberg. Yalmip: A toolbox for modeling and optimization in matlab. In *Proceedings of the CACSD Conference*, volume 3. Taipei, Taiwan, 2004. 24
- [32] G.L. Miller, S.-H. Teng, W. Thurston, and S.A. Vavasis. Geometric separators for finite-element meshes. *SIAM Journal on Scientific Computing*, 19(2):364–386, 1998. 27

- [33] H. Mittelmann and J. Peng. Estimating bounds for quadratic assignment problems associated with hamming and manhattan distance matrices based on semidefinite programming. *SIAM Journal on Optimization*, 20(6):3408–3426, 2010. 23, 24
- [34] D.E. Oliveira, H. Wolkowicz, and Y. Xu. ADMM for the SDP relaxation of the QAP. *Math. Program. Comput.*, 10(4):631–658, 2018. 12, 20, 24, 25
- [35] F.N. Permenter. *Reduction methods in semidefinite and conic optimization*. PhD thesis, Massachusetts Institute of Technology, 2017. 4
- [36] H. Pirim, B. Ekşioğlu, A. D. Perkins, and Ç. Yüceer. Clustering of high throughput gene expression data. *Computers & operations research*, 39(12):3046–3061, 2012. 25
- [37] T.K. Pong, H. Sun, N. Wang, and H. Wolkowicz. Eigenvalue, quadratic programming, and semidefinite programming relaxations for a cut minimization problem. *Comput. Optim. Appl.*, 63(2):333–364, 2016. 20, 27
- [38] J. Povh and F. Rendl. A copositive programming approach to graph partitioning. *SIAM Journal on Optimization*, 18(1):223–241, 2007. 27
- [39] J. Povh and F. Rendl. Copositive and semidefinite relaxations of the quadratic assignment problem. *Discrete Optim.*, 6(3):231–241, 2009. 16
- [40] F. Rendl and R. Sotirov. The min-cut and vertex separator problem. *Computational Optimization and Applications*, 69(1):159–187, 2018. 27
- [41] A. Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Transactions on Information Theory*, 25(4):425–429, 1979. 2, 4
- [42] R. Sotirov. An efficient semidefinite programming relaxation for the graph partition problem. *INFORMS Journal on Computing*, 26(1):16–30, 2013. 25
- [43] R. Sotirov. Graph bisection revisited. *Ann. Oper. Res.*, 265(1):143–154, 2018. 25
- [44] S. Sremac, H.J. Woerdeman, and H. Wolkowicz. Error bounds and singularity degree in semidefinite programming. Technical report, University of Waterloo, Waterloo, Ontario, 2019. 24 pages, research report revision. 4
- [45] J.F Sturm. Using sedumi 1.02. *A MATLAB toolbox for optimization over symmetric cones*, (5):10, 2001. 24
- [46] U. Truetsch. *A semidefinite programming based branch-and-bound framework for the quadratic assignment problem*. Tilburg University, 2014. 16, 17, 18
- [47] F. Vallentin. Symmetry in semidefinite programs. *Linear Algebra Appl.*, 430(1):360–369, 2009. 4
- [48] E.R van Dam and R. Sotirov. Semidefinite programming and eigenvalue bounds for the graph partition problem. *Mathematical Programming*, 151(2):379–404, 2015. 25
- [49] J.H.M. Wedderburn. On Hypercomplex Numbers. *Proc. London Math. Soc. (2)*, 6:77–118, 1908. 4, 6
- [50] H. Wolkowicz and Q. Zhao. Semidefinite programming relaxations for the graph partitioning problem. *Discrete Appl. Math.*, 96/97:461–479, 1999. Selected for the special Editors’ Choice, Edition 1999. 25, 26

- [51] Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. volume 2, pages 71–109. 1998. Semidefinite programming and interior-point approaches for combinatorial optimization problems (Toronto, ON, 1996). 15, 16