Eigenvalue, Quadratic Programming and Semidefinite Programming Bounds for Graph Partitioning Problems

by

Ningchuan Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2014

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The Graph Partitioning problems are hard combinatorial optimization problems. We are interested in both lower bounds and upper bounds. We introduce several methods including basic eigenvalue and projected eigenvalue techniques, convex quadratic programming techniques, and semidefinite programming (SDP). In particular, we show that the SDP relaxation is equivalent to and arises from the Lagrangian relaxation for a particular quadratically constrained quadratic model. Moreover, the bounds obtained by the eigenvalue techniques are good and cheap.

Acknowledgements

I would like to especially thank my advisor Henry Wolkowicz, who gave me the valuable opportunity of Master degree and two years' financial support. He patiently taught me many optimization knowledge and guided me all the time.

I am from mainland of China. I have been in Canada since July 2008. I deeply thank my parents for both mental and financial support during my studies. There were a lot of difficulties in my life and study during these years. I cannot success without their encouragement. Before I came to Canada, I barely read them. My father is a chief engineer in a ship. He is on the sea most time for making money to support our family, even in the Chinese new year eve. Now I more and more understand my parents. Thanks to my parents for understanding. I would like to thank my uncle, Hongwei Wang, who immigrated to Canada in 2003 and picked me up from the Pearson airport. Thanks to my uncle for providing me a comfortable living environment at the beginning of my daily life in Canada.

I was brought up by my grandmother who is the most important person in my whole life. I never left her untill 2008. From 2012, her momery became worse and worse because of Alzheimer disease. Every time I thought of her, I lapsed into apologies and regrets. I hope she can maintain a happy attitude.

I would like to thank Professors Chris Godsil and Thomas Coleman for their valuable time and comments.

I would like to thank my girl friend, Lei Tong. Owing to her, I have the feeling of family in Canada.

I would like to thank my friends, Zhihan Gao, Jiaxin Liu, Junbo Huang, and Longji Guo, who helped me.

Table of Contents

List of Tables

Chapter 1

Introduction

We consider *graph partitioning, GP* problems where we partition the node set of a graph into k sets of given sizes in order to *minimize the sum of the weights of the cut edges*. This problem contains the cut minimization problem as a special case. In both problems, we can use a model with a quadratic objective function over the set of partition matrices. A common problem in circuit board and microchip design, computer program segmentation, floor planning and other layout problems can be modelled as GP problems. More applications of GP problem can be found in [11].

The GP problems have a history since 1869, due to Jordan who had results for trees. Three new algorithms, as well as three earlier algorithms, are summarized in [11]. We briefly introduce some of these heuristic algorithms.

The Kernighan-Lin algorithm is one of the earliest algorithms proposed for partitioning graphs. The algorithm starts with an initial partition into two sets, A and B whose sizes are specified. At each iteration, we choose subsets, A' of A and B' of B with $|A'| = |B'|$ such that $\delta(A, B) > \delta((A \backslash A') \cup B', (B \backslash B') \cup A')$. This algorithm is possible generated to partition the graph into arbitrary number of sets. However, the running time and storage costs of the algorithm will increase rapidly with the number of parts. In fact, most variants of graph partitioning problem are $\mathcal{NP}$-hard, i.e., it is unlikely to find a polynomial time algorithm. Worse, Bui and Jones have shown that it is $\mathcal{NP}$-hard to find approximately optimal vertex and edge separators, even in graphs with maximum degree three.

Level-structure partitioning is another early algorithm. We first find a approximately longest path in the graph, say uv-path. Then we apply breadth-first search from u to label the vertices. We label u as level 0, and neighbors of u as level 1. We label the neighbors of ith level vertices as $(i + 1)$th level. Then the algorithm chooses the vertices in the median level as the vertex separator.

A spectral partitioning algorithm is introduced in [10]. We need to solve an eigenvector corresponding to the second smallest eigenvalue of L, the Laplacian matrix of the graph. Then we use the median of the components of the eigenvector to partition the vertices. We say the median is x_l. Let A contain all vertices whose components are less than x_l. Let B contain all vertices whose components are greater than x_l. We put the vertices whose components are equal to x_l into either A or B such that the size of A differs the size of B at most one. Let $A' \subseteq A$ be the set of vertices having neighbors in B. Let $B' \subseteq B$ be the set of vertices having neighbors in A. Let E' be the set of edges joining A' and B'. Obviously, removing E' disconnect the graph. $H := (A', B', E')$ is a bipartite graph. König's Theorem says that the size of maximum matching equals the size of minimum cover. Finding the minimum cover of H gives us a vertex separator.

In this thesis, we don't talk about algorithms. We focus on lower bounds for graph partitioning. The lower bounds as well as upper bounds are also important because of the following reasons. First, we can improve our bounds to get close to the optimal value. Second, the good quality bounds are very helpful for a branch-and-bound scenario. We will study both existing and new bounds and provide both theoretical properties and empirical results.

In 1953, Hoffman and Wielandt proved Theorem (3.1.2) in [9]. In the early 70s, Donath and Hoffman provided an eigenvalue-based bound in [5] using the Hoffman-Wielandt result. The projection technique is studied and applied in [7, 14, 6] to eliminate two linear constriants. These are based on a parametrization of the affine span of the linear equality constraints. In [14], it shows that we can separate the objective function into three parts, and further we can perturbate the diagonal of A to improve the bounds. Computational results on variaty of randomly generated graphs are provided in [6].

Furthermore, we extend the approach in [1, 3, 2] from the quadratic assignment problem, QAP, to our GP case. This allows for a convex quadratic programming (QP) bound that is based on semidefinite programming (SDP) duality and that can be solved efficiently.

Finally, the SDP bounds are studied in [17, 18]. In [18], it shows that SDP relaxation can be obtained from the dual of the homogenized Lagrangian dual of the quadratically constrained quadratic problem. In [15], authors showed that the two constraints $X^TX = \mathrm{Diag}(m)$ and $\mathrm{diag}(XX^T) = u$ are redundent.

Many of the results in this thesis are taken from the recent research report [15].

1.1 Outline

This thesis is organized as follows. We continue in Chapter 2 with preliminary discriptions of graph partitioning problem and its formulation. We give a brief introduction to semidefinite programming, which is applied in Chapter 5. In Chapter 3, we introduce the definition of minimal scaler product and the projection technique. We get lower bounds by Theorem (3.1.2) and Theorem (3.2.3). The quadratic programming (QP) bound is introduced in Chapter 4. The semidefinite programming (SDP) bound is described in Chapter 5.

The Cut Minimization Problem (CM) is introduced in Chapter 6, including lower bounds.

Our empirical numerical tests are presented in Chapter 7.

We close with Chapter 8 presenting some conclusions.

Chapter 2

Preliminaries

2.1 Graphs and Partition Matrices

Definition 2.1.1. *A graph* G *is a finite nonempty set,* $\mathsf{N}(\mathsf{G})$*, of objects, called vertices, together with a set,* $\mathsf{E}(\mathsf{G})$*, of unordered pairs of distinct vertices. The elements of* $\mathsf{E}(\mathsf{G})$ *are called edges.*

Let $\mathsf{G} = (\mathsf{N}, \mathsf{E})$ be an edge-weighted undirected graph with node set $\mathsf{N} = \{1, \dots, n\}$ and edge weights $w_{ij} > 0$. In addition, we have an ordered positive integer vector of set sizes $m = (m_1, \dots, m_k)^T \in \mathbb{N}^k, m_1, \geq \dots, \geq m_k, k > 2$, such that the sum of the components $\sum_{i=1}^k m_i = {u_k}^T m = n$. Here u_k is the vector of ones and k indicates its size. We define

$$\mathcal{P}_m := \left\{(S_1, \dots, S_k) : S_i, S_j \subset \mathsf{N}, S_i \cap S_j = \emptyset, \text{ for } i \neq j, \cup_{i=1}^k S_i = \mathsf{N}, |S_i| = m_i, \forall i\right\}$$

to be the set of all partitions of N with the appropriate sizes specified by m. The partitioning is encoded using an $n \times k$ *partition matrix* $X \in \{0,1\}^{n \times k}$ where the column $X_{:j}$ is the incidence vector for the set S_j

$$X_{ij} = \begin{cases} 1 & \text{if } i \in S_j, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the set cardinality constraints are given by $X^T u_n = m$; while the constraints that each vertex appears in exactly one set is given by $X u_k = u_n$. We collect these matrices in the set $\mathcal{M}_m$,

$$\mathcal{M}_m := \{X \in \{0,1\}^{n \times k} : X u_k = u_n, X^T u_n = m\}.$$

Remark 2.1.2. *There is a one-to-one corresponding relation between* $\mathcal{P}_m$ *and* $\mathcal{M}_m$*.*

Definition 2.1.3. *We denote the set of zero-one, nonnegative, linear equalities, doubly stochastic type, m-diagonal orthogonality type, e-diagonal orthogonality type, and gangster* *constraints as, respectively,*

$$
\begin{array}{rcl}
\mathcal{Z} & := & \{X \in \mathbb{R}^{n\times k} : X_{ij} \in \{0,1\}, \forall ij\} = \{X \in \mathbb{R}^{n\times k} : (X_{ij})^2 = X_{ij}, \forall ij\}. \\
\mathcal{N} & := & \{X \in \mathbb{R}^{n\times k} : X_{ij} \geq 0, \forall ij\}. \\
\mathcal{E} & := & \{X \in \mathbb{R}^{n\times k} : Xu_k = u_n, X^T u_n = m\} = \{X \in \mathbb{R}^{n\times k} : \|Xu_k - u_n\|^2 + \|X^T u_n - m\|^2 = 0\}. \\
\mathcal{D} & := & \mathcal{N} \cap \mathcal{E}. \\
\mathcal{D}_O & := & \{X \in \mathbb{R}^{n\times k} : X^T X = \mathrm{Diag}(m)\}. \\
\mathcal{D}_e & := & \{X \in \mathbb{R}^{n\times k} : \mathrm{diag}(XX^T) = u_n\}. \\
\mathcal{G} & := & \{X \in \mathbb{R}^{n\times k} : X_{:i} \circ X_{:j} = 0, \forall i \neq j\}.
\end{array}
$$

Here $\mathrm{Diag}(v)$ denotes the diagonal matrix formed using the vector v; the adjoint $\mathrm{diag}(Y) = \mathrm{Diag}^*(Y)$ is the vector formed from the main diagonal of Y. We will indroduce the concept of adjoint later. $A \circ B$ denotes the *Hadamard product*.

A nonnegative matrix X is called *doubly stochastic* if every row sum and column sum are both equal to 1. $\mathcal{D}$ looks like the set of doubly stochastic matrices but not quite since the column sums of elements in $\mathcal{D}$ are not 1. So we call $\mathcal{D}$ the set of *doubly stochastic type.*

There are many equivalent ways of representing the set of all partition matrices. Following are a few.

Proposition 2.1.4.

$$
\begin{array}{rcl}
\mathcal{M}_m & = & \mathcal{E} \cap \mathcal{Z} \\
& = & \mathcal{E} \cap \mathcal{D}_O \cap \mathcal{N} \\
& = & \mathcal{E} \cap \mathcal{D}_O \cap \mathcal{D}_e \cap \mathcal{N} \\
& = & \mathcal{E} \cap \mathcal{Z} \cap \mathcal{D}_O \cap \mathcal{G} \cap \mathcal{N}
\end{array}
\tag{2.1.1}
$$

Proof. The first equality follows immediately from the definitions.

The second equality is shown in [13, Prop. 1]. Here we include the proof for completeness: $X \in \mathcal{M}_m \Longrightarrow X \in \mathcal{E} \cap \mathcal{D}_O \cap \mathcal{N}$ is trivial.

Conversely, let $X \in \mathcal{E} \cap \mathcal{D}_O \cap \mathcal{N}$. $X \in \mathcal{E} \cap \mathcal{N}$ implies $0 \leq X_{ij} \leq 1$, hence $(X_{ij})^2 \leq X_{ij}$. $X \in \mathcal{D}_O$ implies $\mathrm{tr}(X^T X) = \sum_i m_i = n$.

So we have $\mathrm{tr}(X^T X) = \sum_{ij}(X_{ij})^2 = n$ and $s(X) = {u_n}^T X u_k = \sum_{ij} X_{ij} = n$.

Therefore $n = \sum_{ij}(X_{ij})^2 \leq \sum_{ij} X_{ij} = n$. Getting equality throughout gives $(X_{ij})^2 = X_{ij}$. So $X_{ij} \in \{0,1\}$.

The third and fourth equivalences contain redundant sets of constraints. □

2.2 Formulation of GP using a Quadratic Program

Now we are going to formulate/model the GP Problem.

Let A be the matrix such that

$$
A_{ij} = \begin{cases} w_{ij} & \text{if } ij \in \mathsf{E}, \\ 0 & \text{otherwise.} \end{cases}
$$

The matrix A is called the *weighted adjacency matrix* of the graph. Since G is a undirected graph, $A = A^T$. Now we use $\mathcal{S}^n$ to denote the set of all $n \times n$ symmetric matrices, i.e.,

$$\mathcal{S}^n = \{H \in \mathbb{R}^{n\times n} : H = H^T\}.$$

Symmetric matrices are orthogonally diagonalizable, $S = PDP^T$, where $P \in \mathbb{R}^{n\times n}$ is the *orthogonal*, $P^TP = PP^T = I_n$, matrix of eigenvectors, and D is the diagonal matrix of (real) eigenvalues. Denote the matrices with orthogonal columns as

$$\mathcal{O}_{n\times m} := \{Q \in \mathbb{R}^{n\times m} : Q^TQ = I_m\}.$$

We use $\mathcal{O}_n$ to simply denote $\mathcal{O}_{n\times n}$.

Throughout this thesis, we use the vector notation $\lambda(H) = (\lambda_1(H), \lambda_2(H), \cdots, \lambda_n(H))^T \in \mathbb{R}^n$ to denote the eigenvalues of an n-by-n symmetric matrix H in non-increasing order, where $\lambda_1(H) \geq \lambda_2(H) \geq \cdots, \geq \lambda_n(H)$

For each partition matrix X, we can verify

$$(XX^T)_{ij} = \begin{cases} 1 & \text{if node } i \text{ and node } j \text{ are in the same set,} \\ 0 & \text{otherwise.} \end{cases}$$

So

$$w_{\text{uncut}}(X) := \frac{1}{2}\operatorname{tr}(AXX^T) = \frac{1}{2}\operatorname{tr}(X^TAX)$$

is the total weight of the uncut edges induced by the partition matrix X. Here, tr denotes the trace of the matrix. Note that $\operatorname{tr}(\cdot)$ is commutative, $\operatorname{tr}(XY) = \operatorname{tr}(YX)$ and $\operatorname{tr}(X) = \sum_{i=1}^n \lambda_i(X)$.

Let $r(A)$ denote the row sums of a $n \times n$ matrix A which indicates the degree of every vertex, i.e.,

$$r(A) = Au_n.$$

Let $s(A)$ denote the sum of all entries of A, i.e.,

$$s(A) = {u_n}^T A u_n.$$

Notice that $\frac{1}{2}s(A) = |\mathsf{E}(\mathsf{G})|$. Then the total weight for the cut edges induced by the partition matrix X is

$$w_{\text{cut}}(X) := \frac{1}{2}s(A) - \frac{1}{2}\operatorname{tr}(X^TAX) = \frac{1}{2}({u_n}^TAu_n - \operatorname{tr}(X^TAX)),$$

which is our *objective function*.

For each partition matrix X, we have

$$\operatorname{diag}(XX^T) = u_n.$$

So

$$\begin{aligned} {u_n}^TAu_n &= (\operatorname{diag}(XX^T))^TAu_n \\ &= \operatorname{tr}(\operatorname{Diag}(Au_n)XX^T) \\ &= \operatorname{tr}(X^T \operatorname{Diag}(Au_n)X). \end{aligned}$$

The second last equality above is due to that Diag is the adjoint of diag. We will introduce the adjoint of linear mapping later. So our objective function can be also written as:

$$w_{\text{cut}}(X) = \frac{1}{2}\operatorname{tr}(X^T L X) \tag{2.2.1}$$

where the matrix

$$L := \operatorname{Diag}(Au_n) - A$$

is called the *Laplacian matrix of the graph* G.

So the minimum weight of cut edges can be solved as:

$$\text{(GP)} \quad w^*_{\text{cut}} := \begin{array}{ll} \min & \frac{1}{2}\operatorname{tr}(X^T L X) \\ \text{s.t.} & X \in \mathcal{M}_m. \end{array} \tag{2.2.2}$$

Notice that

$$Lu_n = r(L) = r(\operatorname{Diag}(Au_n) - A) = r(A) - r(A) = 0.$$

So u_n is an eigenvector of L with the eigenvalue 0. Oberse that $\operatorname{rank}(L) = n - \kappa$, where κ is the number of components of G. By rank-and-nullity theorem, the multiplicity of 0 eigenvalue is κ.

Also,

$$s(L) = {u_n}^T L u_n = 0.$$

2.3 Semidefinite Programming

A semidefinite programming (SDP) problem is a problem of minimizing or maximizing a linear function of finitely many symmetric matrix variables with real entries subject to finitely many linear equations and linear inequalities on these variables and subject to positive semidefiniteness constraints on some of them. In this section, we will introduce some background of SDP, which is needed in chapter 5. We also conclude some theorems and proofs, as well as some related important results for SDP. Most of contents are taken from [16], which is the textbook of the course CO 671, Semidefinite Optimization.

2.3.1 Positive Semidefinite Matrices

Definition 2.3.1. *Let $A \in \mathcal{S}^n$. A is called* positive semidefinite (PSD) *if $\forall x \in \mathbb{R}^n$, we have $x^T A x \geq 0$. The set of all positive semidefinite matrices is denoted by $\mathcal{S}^n_+$. Similarly, A is called* positive definite (PD) *if $\forall x \in \mathbb{R}^n$ and $x \neq 0$, we have $x^T A x > 0$. The set of all positive definite matrices is denoted by $\mathcal{S}^n_{++}$.*

Proposition 2.3.2. *[16, Proposition 1.10]*(Characterization of PSD matrices) *Let $A \in \mathcal{S}^n$. Then the following are equivalent:*

1. *A is positive semidefinite;*

2. $\lambda_j(A) \geq 0, \forall j \in \{1, 2, \ldots, n\}$;

3. *there exist* $\mu \in \mathbb{R}^n_+$ *and* $h^{(i)} \in \mathbb{R}^n$, $\forall i \in \{1, 2, \ldots, n\}$ *such that*

$$A = \sum_{i=1}^{n} \mu_i h^{(i)} h^{(i)^T};$$

4. *there exists* $B \in \mathbb{R}^{n \times n}$ *such that* $A = BB^T$ *(here, B can be chosen as a lower triangular matrix-the Cholesky decomposition of A);*

5. *for every nonempty* $J \subseteq \{1, 2, \ldots, n\}$, $\det(X_J) \geq 0$, *where* $X_J := \{[X_{ij}] : i, j \in J\}$;

6. $\forall S \in \mathcal{S}^n_+, \langle X, S \rangle \geq 0$.

Remark 2.3.3. *Note that the number of nonzero eigenvalues of $A \in \mathcal{S}^n_+$ is equal to the rank of A. In Item 4, if* $\operatorname{rank}(A) = r$, *then we can choose $B \in \mathbb{R}^{n \times r}$. In Item 5, we call X_J the* symmetric minors *of A.*

Proposition 2.3.4. *[16, Proposition 1.11]*(Characterization of PD matrices) *Let $A \in \mathcal{S}^n$. Then the following are equivalent:*

1. *A is positive definite;*

2. $\lambda_j(A) > 0, \forall j \in \{1, 2, \ldots, n\}$;

3. *there exist* $\mu \in \mathbb{R}^n_{++}$ *and* $h^{(i)} \in \mathbb{R}^n$, $\forall i \in \{1, 2, \ldots, n\}$ *linearly independent such that*

$$A = \sum_{i=1}^{n} \mu_i h^{(i)} h^{(i)^T};$$

4. *there exists* $B \in \mathbb{R}^{n \times n}$ *nonsingular such that* $A = BB^T$ *(here, B can be chosen as a lower triangular matrix-the Cholesky decomposition of A);*

5. *for every* $J_k := \{1, 2, \ldots, k\}, k \in \{1, 2, \ldots n\}, \det(A_{J_k}) > 0$;

6. $\forall S \in \mathcal{S}^n_+ \backslash \{0\}, \langle X, S \rangle > 0$;

7. $A \succeq 0$ *and* $\operatorname{rank}(A) = n$.

Definition 2.3.5. *Let $A \in \mathcal{S}^n$. A is called* diagonally dominant *if* $A_{ii} \geq \sum_{j \neq i} |A_{ij}|$, *for every* $1 \leq i \leq n$. *Similarly, A is called* strictly diagonally dominant *if* $A_{ii} > \sum_{j \neq i} |A_{ij}|$, *for every* $1 \leq i \leq n$.

Remark 2.3.6. *The Laplacian matrix L is obviously diagonally dominant.*

Lemma 2.3.7. *[16, Lemma 1.22]*(Schur Complement) *Let $X \in \mathcal{S}^n$ and $T \in \mathcal{S}^m_{++}$. Then*

$$M := \begin{pmatrix} T & U^T \\ U & X \end{pmatrix} \succeq 0 \text{ if and only if } X - UT^{-1}U^T \succeq 0.$$

Moreover, $M \succ 0$ if and only if $X - UT^{-1}U^T \succ 0$.

Proof. Consider the following decomposition of M:

$$\underbrace{\begin{pmatrix} I & 0 \\ UT^{-1} & I \end{pmatrix}}_{R} \begin{pmatrix} T & 0 \\ 0 & X - UT^{-1}U^T \end{pmatrix} \underbrace{\begin{pmatrix} I & T^{-1}U^T \\ 0 & I \end{pmatrix}}_{R^T} = \begin{pmatrix} T & U^T \\ U & X \end{pmatrix}.$$

Since R is lower triangular and $\det(R) = 1$, R is nonsingular. Therefore,

$$M \succeq 0 \Longleftrightarrow X - UT^{-1}U^T \succeq 0.$$

Also,

$$M \succ 0 \Longleftrightarrow X - UT^{-1}U^T \succ 0.$$

□

Theorem 2.3.8. *If A is diagonally dominant, then $A \succeq 0$.*

Proof. We prove it by induction on n.

Base case: n=2. $A_{11} \geq |A_{12}| \geq 0$ and $A_{22} \geq |A_{12}| \geq 0$. $\det(A) = A_{11}A_{22} - {A_{12}}^2 \geq 0$. So $A \succeq 0$.

Suppose it is true for $2 \leq n \leq k$, where k is some natrual number. We need to prove it is also true for $n = k+1$.

Let $A = \begin{bmatrix} A_{11} & a^T \\ a & \bar{A} \end{bmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$ be a diagonally dominant matrix, where $A_{11} \in \mathbb{R}, a \in \mathbb{R}^k$, and $\bar{A} \in \mathbb{R}^{k\times k}$.

If $A_{11} = 0$, then $a = 0$. By inductive hypothesis, $\bar{A} \succeq 0$, which implies $A \succeq 0$.

If $A_{11} > 0$, by the Schur Complement Lemma,

$$A = \begin{bmatrix} A_{11} & a^T \\ a & \bar{A} \end{bmatrix} \succeq 0 \Longleftrightarrow \bar{A} - \frac{1}{A_{11}} aa^T \succeq 0.$$

If we can show $\bar{A} - \frac{1}{A_{11}}aa^T$ is a diagonal dominant matrix, we are done by the inductive hypothesis.

$(\bar{A} - \frac{1}{A_{11}}aa^T)_{ij} = \bar{A}_{ij} - \frac{a_i a_j}{A_{11}}$.

$$\begin{array}{rcl} \sum_{j\neq i} \left|(\bar{A} - \frac{1}{A_{11}}aa^T)_{ij}\right| & = & \sum_{j\neq i} \left|\bar{A}_{ij} - \frac{a_i a_j}{A_{11}}\right| \\ & \leq & \sum_{j\neq i} \left|\bar{A}_{ij}\right| + \frac{|a_i|}{A_{11}} \sum_{j\neq i} |a_j| \\ & \leq & \sum_{j\neq i} \left|\bar{A}_{ij}\right| + \frac{|a_i|}{A_{11}}(A_{11} - |a_i|) \\ & = & \sum_{j\neq i} \left|\bar{A}_{ij}\right| + |a_i| - \frac{{a_i}^2}{A_{11}} \\ & = & \sum_{j\neq i} |A_{ij}| - \frac{{a_i}^2}{A_{11}} \\ & \leq & |A_{ii}| - \frac{{a_i}^2}{A_{11}} \\ & = & (\bar{A} - \frac{1}{A_{11}}aa^T)_{ii} \end{array}$$

So $A \succeq 0$, as desired. □

Remark 2.3.9. *Note that if A is strictly diagonally dominant, then $A \succ 0$.*

Corollary 2.3.10. *The Laplacian matrix $L \succeq 0$ and $\lambda(L) \geq 0$.*

2.3.2 Inner Product and Norms

An inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^{n \times m}$ is defined as

$$\langle X, Y \rangle = \operatorname{tr}(X^T Y),$$

where $X, Y \in \mathbb{R}^{n \times m}$.

Remark 2.3.11. *$\langle \cdot, \cdot \rangle$ is indeed an inner product on $\mathbb{R}^{n \times m}$, i.e. $\forall \alpha \in \mathbb{R}, \forall X, Y, Z \in \mathbb{R}^{n \times m}$, we can easily verify:*

1. *positive semidefinite:*

$$\langle X, X \rangle \geq 0 \text{ and } \langle X, X \rangle = 0 \text{ if and only if } X = 0.$$

2. *linearity:*

$$\langle \alpha X, Y \rangle = \alpha \langle X, Y \rangle \quad \text{and} \quad \langle X + Z, Y \rangle = \langle X, Y \rangle + \langle Z, Y \rangle .$$

3. *Symmetry:*

$$\langle X, Y \rangle = \langle Y, X \rangle .$$

Theorem 2.3.12. *[16, Proposition 1.19] Let $X, Y \succeq 0$. Then $\langle X, Y \rangle = 0$ if and only if $XY = 0$.*

Proof. Suppose $XY = 0$. Then $\langle X, Y \rangle = \operatorname{tr}(XY) = \operatorname{tr}(0) = 0$.

Now suppose $X, Y \succeq 0$ and $\langle X, Y \rangle = 0$. Then $\langle X, Y \rangle = \operatorname{tr}(XY) = \operatorname{tr}(X^{1/2} Y X^{1/2}) = 0$. Since $Y \succeq 0$ and $X^{1/2}$ is symmetric matrix, we have $X^{1/2} Y X^{1/2} \succeq 0$. So $\lambda(X^{1/2} Y X^{1/2}) \geq 0$. Since $\operatorname{tr}(X^{1/2} Y X^{1/2}) = 0$, we have $\lambda(X^{1/2} Y X^{1/2}) = 0$. It implies that

$$0 = X^{1/2} Y X^{1/2} = X^{1/2} Y^{1/2} (X^{1/2} Y^{1/2})^T.$$

So $X^{1/2} Y^{1/2} = 0$. Then

$$XY = X^{1/2} (X^{1/2} Y^{1/2}) Y^{1/2} = 0.$$

□

Now we talk about norms on $\mathcal{S}^n$.

Definition 2.3.13. *Let $\alpha \in \mathbb{R}$ and $X, Y \in \mathcal{S}^n$. The* norm, *$\|\cdot\|$, on $\mathcal{S}^n$ satisfies the following three axioms:*

1. *$\|X\| > 0, \forall X \neq 0$ and $\|X\| = 0$ if and only if $X = 0$.*
2. *$\|\alpha X\| = |\alpha| \, \|X\|$.*
3. *$\|X + Y\| \leq \|X\| + \|Y\|$ (triangle inequality).*

Recall that for $h \in \mathbb{R}^n$, $\|h\|_p := (\sum_{j=1}^n |h_j|^p)^{\frac{1}{p}}$. We introduce the *Frobenius norm*:

$$\|H\|_F := \sqrt{\sum_{i,j} (H_{ij})^2},$$

and the *operator p-norm*:

$$\|H\|_p := \max \left\{ \|Hh\|_p : h \in \mathbb{R}^n, \|h\|_p = 1 \right\}.$$

Lemma 2.3.14. *Let $H \in \mathcal{S}^n$. Then $\|H\|_F = \langle H, H \rangle^{1/2} = \|\lambda(H)\|_2$.*

Proof. Since $H \in \mathcal{S}^n$, there exists an orthogonal matrix P such that

$$P^T H P = \operatorname{Diag}(\lambda(H)).$$

Then

$$\begin{aligned} \|H\|_F^2 &= \textstyle\sum_{ij} (H_{ij})^2 = \operatorname{tr}(HH) = \operatorname{tr}(HPP^T HPP^T) = \operatorname{tr}\left((P^T HP)(P^T HP)\right) \\ &= \operatorname{tr}\left(\operatorname{Diag}(\lambda(H)) \operatorname{Diag}(\lambda(H))\right) = \textstyle\sum_{j=1}^n \left(\lambda_j(H)\right)^2 = \|\lambda(H)\|_2^2 . \end{aligned}$$

□

Next, we talk about adjoints of linear operators.

Definition 2.3.15. *Let $\mathcal{A} : \mathcal{S}^n \longrightarrow \mathbb{R}^m$ be a linear operator. We define the* adjoint *of $\mathcal{A}$ as a linear operator*

$$\mathcal{A}^* : \mathbb{R}^m \longrightarrow \mathcal{S}^n$$

such that

$$\langle \mathcal{A}^*(y), X \rangle_{\mathcal{S}^n} := y^T \mathcal{A}(X) = \langle y, \mathcal{A}(X) \rangle_{\mathbb{R}^m}, \forall X \in \mathcal{S}^n, \forall y \in \mathbb{R}^m.$$

Notice that if we choose $y = e_i$, then $\langle \mathcal{A}^*(e_i), X \rangle = e_i^T \mathcal{A}(X) = [\mathcal{A}(X)]_i$. So we can write the explicit form of the linear operator $\mathcal{A}$ as

$$[\mathcal{A}(X)]_i := \langle A_i, X \rangle, \forall i \in \{1, 2, \cdots, m\},$$

where $A_i = \mathcal{A}^*(e_i) \in \mathcal{S}^n, \forall i \in \{1, 2, \cdots, m\}$. The adjoint $\mathcal{A}^*$ gives

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

Example 2.3.16. *The adjoint of* Diag *is* diag, *i.e.,*

$$\langle \operatorname{diag}(X), y \rangle = \langle X, \operatorname{Diag}(y) \rangle, \forall X \in \mathcal{S}^n, \forall y \in \mathbb{R}^n.$$

2.3.3 Kronecker Product

Definition 2.3.17. *Let $X \in \mathbb{R}^{n\times k}$.* vec$(X)$*, the vector formed from the column of X, is defined as*

$$\operatorname{vec}(X) = [X_{11}, X_{21}, \cdots, X_{n1}, X_{12}, \cdots, X_{n2}, \cdots, X_{nk}]^T \in \mathbb{R}^{nk}.$$

vec is a linear mapping. The adjoint, as well as the inverse mapping of vec is Mat, which maps nk-dimensional vectors to $n \times k$ matrices. Let $x \in \mathbb{R}^{nk}$. $[\operatorname{Mat}(x)]_{:i} = x_{n(i-1)+1:ni}$.

Definition 2.3.18. *Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$. We define the* Kronecker product *to be*

$$A \otimes B := \begin{bmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \cdots & A_{mn}B \end{bmatrix} \in \mathbb{R}^{mp\times nq}.$$

Note that, for compatible matrices, there are four obvious identities we use offen throughout this paper:

1. $(A \otimes B)^T = A^T \otimes B^T$.
2. $(A \otimes B)(C \otimes D) = AC \otimes BD$.
3. $\operatorname{vec}(AXB) = (B^T \otimes A)\operatorname{vec}(X)$.
4. $\operatorname{tr}(AXBX^T) = \operatorname{vec}(X)^T(B \otimes A)\operatorname{vec}(X)$.

2.3.4 Duality Theory

We define the semidefinite programming problem in standard form and its dual. Suppose $C \in \mathcal{S}^n, b \in \mathbb{R}^m$ and a linear transformation $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ are given. Then we define

$$\begin{array}{lllll} \text{(P)} & \inf & \langle C, X\rangle & & \\ & \text{s.t.} & \mathcal{A}(X) & = & b, \\ & & X & \succeq & 0. \end{array}$$

$$\begin{array}{lllll} \text{(D)} & \sup & b^T y & & \\ & \text{s.t.} & \mathcal{A}^*(y) + S & = & C, \\ & & S & \succeq & 0. \end{array}$$

As we noted, $\mathcal{A}$ can be represented in a more explicit form. So let $A_1, A_2, \cdots, A_m \in \mathcal{S}^n$ such that $[\mathcal{A}(X)]_i = \langle A_i, X\rangle, \forall X \in \mathcal{S}^n$. Then we can write (P) and (D) as

$$\begin{array}{llllll} \text{(P)} & \inf & \langle C, X\rangle & & & \\ & \text{s.t.} & \langle A_i, X\rangle & = & b_i, & \forall i \in \{1, 2, \cdots, m\} \\ & & X & \succeq & 0. & \end{array}$$

$$\begin{array}{lllll} \text{(D)} & \sup & b^T y & & \\ & \text{s.t.} & \sum_{i=1}^m y_i A_i + S & = & C, \\ & & S & \succeq & 0. \end{array}$$

This primal-dual pair is useful even in more general settings. We can replace $\mathcal{S}^n_+$ by an arbitrary convex cone K. Given $c \in \mathbb{R}^n, b \in \mathbb{R}^m, \mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ a linear mapping. Then the primal-dual pair is defined as

$$\begin{array}{lll} \text{(CP)} & \inf & \langle c, x \rangle \\ & \text{s.t.} & \mathcal{A}(x) = b \\ & & x \in K. \end{array}$$

$$\begin{array}{lll} \text{(CD)} & \sup & b^T y \\ & \text{s.t.} & \mathcal{A}^*(y) + s = c \\ & & s \in K^*, \end{array}$$

where $K^* = \{s \in \mathbb{R}^n : \langle s, x \rangle \geq 0, \forall x \in K\}$ is called the dual cone of K.

Definition 2.3.19. *$\bar{X}$ is a* Slater point *for (P) if it is feasible for (P) and $\bar{X} \succ 0$. $(\bar{y}, \bar{S})$ is a* Slater point *for (D) if it is feasible and $\bar{S} \succ 0$.*

Theorem 2.3.20. *[16, Theorem 1.17]* (Weak Duality Theorem for SDP) *If $(\bar{X}, (\bar{y}, \bar{S}))$ are feasible to (P) and (D),respectively, then*

$$\left\langle C, \bar{X} \right\rangle - b^T \bar{y} = \left\langle \bar{X}, \bar{S} \right\rangle \geq 0.$$

Proof.

$$\begin{aligned} \left\langle C, \bar{X} \right\rangle - b^T \bar{y} &= \left\langle C, \bar{X} \right\rangle - \mathcal{A}(\bar{X})^T \bar{y} \\ &= \left\langle C, \bar{X} \right\rangle - \left\langle \mathcal{A}(\bar{X}), \bar{y} \right\rangle \\ &= \left\langle C, \bar{X} \right\rangle - \left\langle \mathcal{A}^*(\bar{y}), \bar{X} \right\rangle \\ &= \left\langle C - \mathcal{A}^*(\bar{y}), \bar{X} \right\rangle \\ &= \left\langle \bar{S}, \bar{X} \right\rangle \geq 0, \end{aligned}$$

since $\bar{X}, \bar{S} \in \mathcal{S}^n_+$. □

The $\left\langle \bar{S}, \bar{X} \right\rangle$ is called the *duality gap* of $(\bar{X}, (\bar{y}, \bar{S}))$.

Theorem 2.3.21. *[16, Theorem 2.14]* (Strong Duality Theorem for SDP) *Suppose (D) has a Slater point. If the objective value of (D) is bounded from above then (P) attains its optimum value and the optimum values of (P) and (D) coincide.*

Corollary 2.3.22. *[16, Corollary 2.17] If both (P) and (D) have Slater points, then both optima are attained and they agree.*

2.3.5 Facial Structures

Definition 2.3.23. *A set $C \subseteq \mathbb{R}^n$ is* convex*, if for every $x, y \in C$ and every $\lambda \in [0, 1]$, we have $\lambda x + (1 - \lambda) y \in C$.*

In above definition, the set $\{\lambda x + (1 - \lambda) y : \lambda \in [0, 1]\}$ is called the *line segment* of x and y. So C is convex, if the line segment of every two points of C also lies in C.

Definition 2.3.24. *A set $K \subseteq \mathbb{R}^n$ is a* cone*, if for every $x \in K$ and every $\lambda \in \mathbb{R}_+$, we have $\lambda x \in K$.*

If a cone is nonempty and closed, it must contain the 0-element by definition.

Definition 2.3.25. *Let $S_1, S_2 \subseteq \mathbb{R}^n$. We define the* Minkowski Sum *of S_1 and S_2 as*

$$S_1 + S_2 = \{s_1 + s_2 : s_1 \in S_1, s_2 \in S_2\}.$$

It is not hard to verify that a set $K \in \mathbb{R}^n$ is a *convex cone*, i.e., K is a cone and a convex set, if $K + K \subseteq K$ and $\lambda K \subseteq K, \forall \lambda \in \mathbb{R}_+$, where $\lambda K = \{\lambda x : x \in K\}$.

Next, we are going to introduce the notion of an *extreme ray.* Let K be a cone. The set $\{\lambda x : \lambda \in \mathbb{R}_+\}$ for $x \in K \backslash \{0\}$ defines a *ray* inside K. A ray $R \subseteq K$ is called an *extreme ray* of K if for every pair of rays $R_1, R_2 \subseteq K$, such that $R \subseteq R_1 + R_2$ implies either $R_1 = R$ or $R_2 = R$ possibly both. The union of all extreme rays is denoted by $\text{Ext}(K)$.

Let R be an extreme ray of a cone K. We use a single nonzero normalized element of R to represent R. The set of all representatives of extreme rays is denoted by $\text{ext}(K)$.

Theorem 2.3.26.

$$\text{ext}(\mathcal{S}_+^n) = \{xx^T : x \in \mathbb{R}^n, \|x\| = 1\}.$$

Proof. Let R be an extreme ray of $\mathcal{S}_+^n$. We need to show that

$$R = \{\lambda x x^T : \lambda \in \mathbb{R}_+\}, \quad \text{for some} \quad x \in \mathbb{R}^n.$$

Since R is a ray of $\mathcal{S}_+^n$, it can be written as

$$R = \{\lambda X : \lambda \in \mathbb{R}_+\}, \quad \text{for some} \quad X \in \mathcal{S}_+^n.$$

By spectral decomposition of X,

$$X = \sum_{i=1}^{n} \alpha_i x_i x_i^T,$$

where x_i is the normalized eigenvector of X corresponding to the i-th largest eigenvalue α_i.

We prove by contradiction. Suppose $\text{rank}(X) > 1$. Since $\text{rank}(X)$ equals the number of nonzero eigenvalues of X, then at least two of eigenvalues of X are greater than 0. Let $R_1 = \{\lambda x_1 x_1^T : \lambda \geq 0\}$ and $R_2 = \left\{\lambda(\sum_{i=2}^{n} \alpha_i x_i x_i^T) : \lambda \geq 0\right\}$. Both R_1 and R_2 are rays, and $R \subseteq R_1 + R_2$. But neither $R = R_1$ nor $R = R_2$, which is a contradiction. □

Remark 2.3.27. *The above theorem tells us $\text{Ext}(\mathcal{S}_+^n)$ is the set of all rank one positive semidefinite matrices.*

Definition 2.3.28. *Let $C \subseteq \mathbb{R}^d$ be a closed convex cone. A convex cone $K \subseteq C$ is a* face *of C, if*

$$x, y \in C, x + y \in K \Rightarrow x, y \in K.$$

We denote $K \lhd C$.

A face K of C is exposed *if there exists $a \in \mathbb{R}^d$ such that*

$$K = \{x \in C : \langle a, x \rangle = 0\} \text{ and } C \subseteq \{x \in \mathbb{R}^d : \langle a, x \rangle \leq 0\},$$

i.e., K is the intersection of C with one of its supporting hyperplanes.

A face K of C is a proper face *of C if*

$$\{0\} \subset K \subset C.$$

The notation $K \subset C$ means that K is a proper subset of C. $K \subseteq C$ means that K is a subset of C and K may equal C.

Theorem 2.3.29. *[16, Theorem 2.25]*

1. *Every nonempty face F of $\mathcal{S}_+^n$ is characterized by a unique subspace $S \subseteq \mathbb{R}^n$ such that*

$$F = \{X \in \mathcal{S}_+^n : S \subseteq \mathcal{N}(X)\}$$

and

$$\operatorname{relint}(F) = \{X \in \mathcal{S}_+^n : S = \mathcal{N}(X)\}.$$

2. *Every proper face F of $\mathcal{S}_+^n$ is exposed.*

3. *Every nonempty face F of $\mathcal{S}_+^n$ can be expressed as*

$$F = (I - Q)\mathcal{S}_+^n(I - Q),$$

where $Q \in \mathcal{S}^n$ is the projection onto the unique subspace S defining F.

Remark 2.3.30. *The above theorem implies that every proper face of $\mathcal{S}_+^n$ is isomorphic to $\mathcal{S}_+^p$ for some $p < n$.*

Chapter 3

Eigenvalue Based Bounds

We first present bounds on w^*_{cut} based on $X \in \mathcal{D}_O$, the m-diagonal orthogonality type constraint $X^T X = M$, where we let $M := \operatorname{Diag}(m)$; and $\tilde{m} =: \left(\sqrt{m_1}, \dots, \sqrt{m_k}\right)^T, \tilde{M} := \operatorname{Diag}(\tilde{m})$ for notational simplicity. Note that $\tilde{M} = M^{1/2}$.

Notice that $\operatorname{tr}(X^T L X) = s(A) - \operatorname{tr}(X^T A X)$, under the condition $X \in \mathcal{M}_m$. But we don't guarantee that the equality holds if we only subject to $X \in \mathcal{D}_O$. That is

$$\begin{array}{lllll} \min & \frac{1}{2}\operatorname{tr}(X^T L X) & \neq & \min & \frac{s(A)}{2} - \frac{1}{2}\operatorname{tr}(X^T A X) \\ \text{s.t.} & X \in \mathcal{D}_O. & & \text{s.t.} & X \in \mathcal{D}_O. \end{array} \tag{3.0.1}$$

It gives us two options of objective functions when we do the relaxation over $X \in \mathcal{D}_O$. But we can perturb the diagonal of L by adding a parameter $d \in \mathbb{R}^n$ to combine these two cases together.

Lemma 3.0.31. $\operatorname{tr}(X^T L X) = \operatorname{tr}\left(X^T (L + \operatorname{Diag}(d)) X\right) - s(d), \forall X \in \mathcal{M}_m, \forall d \in \mathbb{R}^n$.

Proof. Since $X \in \mathcal{M}_m$, we have $\operatorname{diag}(XX^T) = u_n$. So $\operatorname{tr}(X^T \operatorname{Diag}(d) X) = \operatorname{tr}(\operatorname{Diag}(d) X X^T) = \left\langle \operatorname{Diag}(d), XX^T \right\rangle = \left\langle d, \operatorname{diag}(XX^T) \right\rangle = d^T u_n = s(d)$. □

We denote $L(d) := L + \operatorname{Diag}(d) = \operatorname{Diag}(Au_n + d) - A$. So our objective function can be written as $\frac{1}{2}\operatorname{tr}(X^T L(d) X) - \frac{s(d)}{2}$. Notice that

$$\frac{1}{2}\operatorname{tr}(X^T L(d) X) - \frac{s(d)}{2} = \begin{cases} \frac{1}{2}\operatorname{tr}(X^T L X) & \text{if } d = 0, \\ \frac{s(A)}{2} - \frac{1}{2}\operatorname{tr}(X^T A X) & \text{if } d = -Au_n. \end{cases}$$

So our Graph Partitioning Problem is equivalent to

$$\begin{array}{rll} w^*_{\text{cut}} = & \min & \frac{1}{2}\operatorname{tr}(X^T L(d) X) - \frac{s(d)}{2} \\ & \text{s.t.} & X \in \mathcal{M}_m. \end{array} \tag{3.0.2}$$

Because we are allowed to choose any $d \in \mathbb{R}^n$, the objective function in (3.0.2) performs better than the previous two when doing relaxation on $X \in \mathcal{D}_O$ if we choose appropriate $d \in \mathbb{R}^n$.

3.1 Basic Eigenvalue Bound

The Donath-Hoffman [9] bound can be applied to get a simple eigenvalue bound, i.e., we solve the relaxed problem

$$p^*_{eig}(d) := \begin{array}{ll} \min & \frac{1}{2}\operatorname{tr}(X^T L(d) X) - \frac{s(d)}{2} \\ \text{s.t.} & X \in \mathcal{D}_O, \end{array} \tag{3.1.1}$$

where $d \in \mathbb{R}^n$.

Here, by setting $X = YM^{1/2}$ gives

$$\operatorname{tr}(X^T L(d) X) = \operatorname{tr}\left(M^{1/2} Y^T L(d) Y M^{1/2}\right) = \operatorname{tr}\left(L(d) Y M Y^T\right).$$

Since $X \in \mathcal{D}_O$ if and only if $Y^T Y = I_k$, our relaxation (3.1.1) is equivalent to

$$\text{(BE)} \quad p^*_{eig}(d) = \begin{array}{ll} \min & \frac{1}{2}\operatorname{tr}\left(L(d) Y M Y^T\right) - \frac{s(d)}{2} \\ \text{s.t.} & Y^T Y = I_k. \end{array} \tag{3.1.2}$$

We first introduce the following definition.

Definition 3.1.1. *For two vectors $x, y \in \mathbb{R}^n$, the minimal scalar product of x and y is defined by*

$$\langle x, y \rangle_- := \min \left\{ \sum_{i=1}^n x_i y_{\phi(i)} : \phi \text{ is a permutation on } N \right\}.$$

We need the following theorem to get the optimal value of (BE) hence a lower bound for (GP).

Theorem 3.1.2 ([9])**.** *Let A, B be symmetric matrices of order n, k, respectively, with $k \leq n$. Then*

$$\min\left\{\operatorname{tr}(AXBX^T) : X^T X = I_k\right\} = \left\langle \lambda(A), \begin{pmatrix} \lambda(B) \\ 0 \end{pmatrix} \right\rangle_- . \tag{3.1.3}$$

The minimum is attained for $X = \left(p_{\phi(1)}, \ldots, p_{\phi(k)}\right) Q^T$, where $p_{\phi(i)}$ is a normalized eigenvector to $\lambda_{\phi(i)}(A)$ and the columns of $Q = \begin{bmatrix} q_1 & \ldots & q_k \end{bmatrix}$ contains the normalized eigenvectors q_i of $\lambda_i(B)$, and ϕ is the permutation of N attaining the minimum in the minimal scalar product.

Proof. For completeness we include an optimization based proof.

Let $G(X) := X^T X - I_k = 0$ denote the orthogonality constraint. Then the derivative acting on $H \in \mathbb{R}^{n \times k}$ is $\nabla G(X)(H) := X^T H + H^T X$. We note that $\operatorname{vec}(X^T H) = (I \otimes X^T)\operatorname{vec}(H)$ and that the Kronecker product $(I \otimes X^T)$ has full row rank at the minimizer X since X^T has full row rank. Therefore the standard linear independence constraint qualification (LICQ) that $\nabla G(X)$ is onto holds. We can now apply the Lagrange multiplier approach to the minimization problem in (3.1.3). Recall that the Lagrangian is defined as

$$\mathcal{L}(X, S) = \operatorname{tr}(AXBX^T) - \operatorname{tr}(S(X^T X - I)),$$

where $S \in \mathcal{S}^k$. From the LICQ, there exists a Lagrangian multiplier $S \in \mathcal{S}^k$ so that the minimizer X satisfies the stationary condition

$$0 = \nabla_X \mathcal{L}(X, S) = 2AXB - 2IXS.$$

Therefore, $X^T AXB = S = S^T$ which implies that at the optimum X, the matrices $X^T CX$ and D commute and hence are mutually orthogonally diagonalizable by a $k \times k$ orthogonal matrix Q. We can then write the optimal value at the optimum X to be

$$\begin{aligned} \operatorname{tr}(AXBX^T) &= \operatorname{tr}(Q^T X^T AXQQ^T BQ) \\ &= \operatorname{tr}((XQ)^T A(XQ))(\operatorname{Diag}(\lambda(B))) \\ &\geq \lambda_\phi(A)^T \begin{pmatrix} \lambda(B) \\ 0 \end{pmatrix}. \end{aligned}$$

where the last inequality follows from the interlacing of eigenvalues and the fact that $(XQ)^T(XQ) = I_k$. We have padded the vector of eigenvalues of B with zeros and $\lambda_\phi(A)$ is a suitable permutation of the eigenvalues of A. The conclusion now follows from this last observation, the definition of minimal scalar product, and the attainment for the choices of X, Q stated in the hypothesis. □

Remark 3.1.3. *By Theorem* (3.1.2)*, the optimal value* $p^*_{eig}(d)$ *of* (3.1.2) *is attained at*

$$Y = \begin{bmatrix} p_n & p_{n-1} & \cdots & p_{n-k+1} \end{bmatrix},$$

where p_i *is the normalized eigenvector corresponding to* $\lambda_i(L(d))$*. Then we can recover an approximate solution* $X = YM^{1/2}$*.*

Theorem 3.1.4. *Let* $d \in \mathbb{R}^n$*. Then:*

$$\begin{aligned} w^*_{\text{cut}} \geq p^*_{eig}(d) &= \frac{1}{2} \left\langle \lambda(L(d)), \begin{pmatrix} m \\ 0 \end{pmatrix} \right\rangle_- - \frac{s(d)}{2} \\ &= \frac{1}{2} \left(\sum_{i=1}^k m_i \cdot \lambda_{n-i+1}(L(d)) \right) - \frac{s(d)}{2}. \end{aligned} \tag{3.1.4}$$

Proof. This result is a direct application of the Donath-Hoffman bound. We provide a proof for completeness.

We now solve the equivalent problem (BE) in (3.1.2):

$$\begin{array}{ll} \min & \frac{1}{2} \operatorname{tr}\left(L(d) YMY^T\right) - \frac{s(d)}{2} \\ \text{s.t.} & Y^T Y = I_k. \end{array}$$

The optimal value of quadratic part is obtained using the minimal scalar product of eigenvalues as done in the Hoffman-Wielandt result, Theorem 3.1.2. □

Observe that $w^*_{\text{cut}} \geq 0$. So 0 is a natural lower bound. If $p^*_{eig}(d) < 0$, then the lower bound is useless. However we can simply use eigenvalues of L and A to get a useful lower bound as below.

Corollary 3.1.5.

$$
\begin{aligned}
w^*_{\text{cut}} &\geq \max\{p^*_{eig}(0), p^*_{eig}(-Au_n)\} \\
&= \max\left\{ \frac{1}{2}\left\langle \lambda(L), \begin{pmatrix} m \\ 0 \end{pmatrix} \right\rangle_-, \frac{1}{2}\left\langle \lambda(-A), \begin{pmatrix} m \\ 0 \end{pmatrix} \right\rangle_- + \frac{s(A)}{2} \right\} \\
&= \max\left\{ \frac{1}{2}\sum_{i=2}^k m_i \cdot \lambda_{n-i+1}(L), -\frac{1}{2}\sum_{i=1}^k m_i \cdot \lambda_i(A) + \frac{s(A)}{2} \right\} \\
&\geq 0.
\end{aligned}
\tag{3.1.5}
$$

Proof. If we choose $d = 0$, the objective function is purely quadratic. $L(0) = L \succeq 0 \Longrightarrow \lambda(L) \geq 0 \Longrightarrow \sum_{i=2}^k m_i \cdot \lambda_{n-i+1}(L) \geq 0 \Longrightarrow p^*_{eig}(0) \geq 0$. □

Remark 3.1.6. *There is no relation between $p^*_{eig}(0)$ and $p^*_{eig}(-Au_n)$ so far. We don't know which one is greater in general. However, we can always obtain nonnegative bounds by using $p^*_{eig}(0)$.*

Here we provide an alternative envidence of $p^*_{eig}(0) \geq 0$ by using the fact

$$\operatorname{tr}(LYMY^T) = \operatorname{vec}(Y)^T(M \otimes L)\operatorname{vec}(Y).$$

First, we need the following Lemma:

Lemma 3.1.7. *If $A \succeq 0$ and $B \succeq 0$, then $A \otimes B \succeq 0$. Moreover, if u_i is an eigenvector of A with $\lambda_i(A)$ and v_j is an eigenvector of B with $\lambda_j(B)$, then $u_i \otimes v_j$ is an eigenvector of $A \otimes B$ with eigenvalue $\lambda_i(A)\lambda_j(B)$.*

By the above Lemma, we have $M \otimes L \succeq 0$. $Y^TY = I_k$ implies $\|\operatorname{vec}(Y)\|_2 = \sqrt{k}$. Consider the problem:

$$
\begin{aligned}
p^* = \quad & \min \quad x^T(M \otimes L)x \\
& \text{s.t.} \quad \|x\|_2 = \sqrt{k}
\end{aligned}
$$

which is a relaxation of

$$
\begin{aligned}
p^*_{eig}(0) = \quad & \min \quad \operatorname{tr}(LYMY^T) \\
& \text{s.t.} \quad Y^TY = I_k.
\end{aligned}
$$

So we have $0 \leq k \cdot \lambda_{nk}(M \otimes L) = p^* \leq p^*_{eig}(0)$.

3.2 Projected Eigenvalue Bound

3.2.1 The projection technique

Consider the minimization problem

$$
\begin{aligned}
\min \quad & \frac{1}{2}\operatorname{tr}\left(X^TL(d)X\right) - \frac{s(d)}{2} \\
\text{s.t.} \quad & X \in \mathcal{D}_O \cap \mathcal{E}.
\end{aligned}
\tag{3.2.1}
$$

We now project $X \in \mathbb{R}^{n \times k}$ onto the $\mathcal{E}$-space.

Let the $n \times (n-1)$ matrix V be such that

$$V^T u_n = 0; \qquad V^T V = I_{n-1}.$$

The columns of V form an orthonormal basis to the orthogonal complement of u_n.
Similarly, let the $k \times (k-1)$ matrix W be such that

$$W^T \tilde{m} = 0; \qquad W^T W = I_{k-1}.$$

The columns of W represent the orthogonal complement to $\tilde{m}$.
Clearly, both V and W are not uniquely determined.
We define the $n \times n$ and $k \times k$ orthogonal matrices P, Q with

$$P = \begin{bmatrix} \frac{1}{\sqrt{n}} u_n & V \end{bmatrix} \in \mathcal{O}_n, \qquad Q = \begin{bmatrix} \frac{1}{\sqrt{n}} \tilde{m} & W \end{bmatrix} \in \mathcal{O}_k. \tag{3.2.2}$$

Note that

$$PP^T = I_n \Rightarrow VV^T = I_n - \frac{1}{n} u_n u_n{}^T$$

and

$$QQ^T = I_k \Rightarrow WW^T = I_k - \frac{1}{n} \tilde{m}\tilde{m}^T.$$

Lemma 3.2.1. *[14, Lemma 3.1] Let P, Q be defined in (3.2.2). Suppose that $X \in \mathbb{R}^{n\times k}$ and $Z \in \mathbb{R}^{(n-1)\times(k-1)}$ are related by*

$$X = P \begin{bmatrix} 1 & 0 \\ 0 & Z \end{bmatrix} Q^T \tilde{M}. \tag{3.2.3}$$

Then the following holds:

1. $X \in \mathcal{E}$.
2. $X \in \mathcal{N} \iff VZW^T \geq -\frac{1}{n} u_n \tilde{m}^T$.
3. $X \in \mathcal{D}_O \iff Z \in \mathcal{O}_{(n-1)\times(k-1)}$.

Conversely, if $X \in \mathcal{E}$, then there exists Z such that the representation (3.2.3) holds.

Proof. Define $\widehat{X} := \frac{1}{n} u_n m^T$. We expand (3.2.3) by substituting (3.2.2) yields

$$\begin{aligned} X &= P \begin{bmatrix} 1 & 0 \\ 0 & Z \end{bmatrix} Q^T \tilde{M} \\ &= \begin{bmatrix} \frac{u_n}{\sqrt{n}} & V \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Z \end{bmatrix} \begin{bmatrix} \frac{\tilde{m}^T}{\sqrt{n}} \\ W^T \end{bmatrix} \tilde{M} \\ &= \frac{1}{n} u_n m^T + VZW^T \tilde{M} \\ &= \widehat{X} + VZW^T \tilde{M}. \end{aligned} \tag{3.2.4}$$

Since $V^T u_n = 0$, we have

$$X^T u_n = \frac{1}{n} M u_k {u_n}^T u_n + \tilde{M} W Z^T V^T u_n = M u_k = m.$$

Similarly, since $W^T \tilde{M} u_k = W^T \tilde{m} = 0$, we have

$$X u_k = \frac{1}{n} u_n {u_k}^T M u_k + V Z W^T \tilde{M} u_k = u_n.$$

So

$$X \in \mathcal{E}.$$

By (3.2.4), we can write

$$X = \frac{1}{n} u_n \tilde{m}^T \tilde{M} + V Z W^T \tilde{M} = (\frac{1}{n} u_n \tilde{m}^T + V Z W^T) \tilde{M}.$$

Thus

$$X \in \mathcal{N} \Longleftrightarrow (\frac{1}{n} u_n \tilde{m}^T + V Z W^T) \tilde{M} \geq 0 \Longleftrightarrow V Z W^T \geq -\frac{1}{n} u_n \tilde{m}^T,$$

because multiplying with the positive diagonal matirx $\tilde{M}^{-1}$ does not change the inequality.

Finally, since $P \in \mathcal{O}_n$ and $Q \in \mathcal{O}_k$, we have

$$X^T X = M \Longleftrightarrow Q \begin{bmatrix} 1 & 0 \\ 0 & Z^T \end{bmatrix} P^T P \begin{bmatrix} 1 & 0 \\ 0 & Z \end{bmatrix} Q^T = I \Longleftrightarrow Z \in \mathcal{O}_{(n-1)\times(k-1)}.$$

Suppose $X \in \mathcal{E}$. Then

$$P^T X \tilde{M}^{-1} Q = \begin{bmatrix} \frac{{u_n}^T}{\sqrt{n}} \\ V^T \end{bmatrix} X \tilde{M}^{-1} \begin{bmatrix} \frac{\tilde{m}}{\sqrt{n}} & W \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & V^T X \tilde{M}^{-1} W \end{bmatrix}.$$

□

Please note that the XZ relation (3.2.4) in Lemma (3.2.1) will be used frequently in Chapter 4. So we emphasize it here.

If we substitute (3.2.4) into our objective function $\frac{1}{2}\operatorname{tr}(X^T L(d) X) - \frac{s(d)}{2}$, we will obtain an equivalent formulation of the graph partitioning problem in the lower dimensional Z-space.

$$\begin{aligned} \frac{1}{2}\operatorname{tr}(X^T L(d) X) - \frac{s(d)}{2} &= \frac{1}{2}\operatorname{tr}\left((\widehat{X} + V Z W^T \tilde{M})^T L(d) (\widehat{X} + V Z W^T \tilde{M})\right) - \frac{s(d)}{2} \\ &= \frac{1}{2}\operatorname{tr}(\widehat{X}^T L(d) \widehat{X}) + \operatorname{tr}\left(V^T L(d) \widehat{X} \tilde{M} W Z^T\right) \\ &\quad + \frac{1}{2}\operatorname{tr}\left((W^T M W) Z^T (V^T L(d) V) Z\right) - \frac{s(d)}{2} \\ &= \left(\alpha(d) - \frac{s(d)}{2}\right) + \operatorname{tr}\left(C(d) Z^T\right) + \frac{1}{2}\operatorname{tr}\left(\hat{L}(d) Z \hat{M} Z^T\right), \end{aligned} \tag{3.2.5}$$

where

$$\alpha(d)=\frac{1}{2}\operatorname{tr}(\widehat{X}^T L(d)\widehat{X})=\frac{s(M^2)s(d)}{2n^2},\;\; C(d)=\frac{1}{n}V^T dm^T\tilde{M}W,\;\; \hat{L}(d)=V^TL(d)V \text{ and } \hat{M}=W^TMW. \tag{3.2.6}$$

Notice that the linear term can be further written as

$$\begin{aligned}\operatorname{tr}(C(d)Z^T) &= \tfrac{1}{n}\operatorname{tr}\left(V^Tdm^T\tilde{M}WZ^T\right)\\ &= \tfrac{1}{n}\operatorname{tr}\left(dm^T(VZW^T\tilde{M})^T\right)\\ &= \tfrac{1}{n}\operatorname{tr}\left(dm^T(X-\widehat{X})^T\right)\\ &= \tfrac{1}{n}\operatorname{tr}\left(dm^TX^T\right)-\tfrac{1}{n}\operatorname{tr}\left(dm^T\widehat{X}^T\right)\\ &= \tfrac{1}{n}(Xm)^Td-2\alpha(d).\end{aligned} \tag{3.2.7}$$

So

$$\frac{1}{2}\operatorname{tr}(X^TL(d)X)-\frac{s(d)}{2}=\left(-\alpha(d)-\frac{s(d)}{2}\right)+\frac{1}{n}(Xm)^Td+\frac{1}{2}\operatorname{tr}\left(\hat{L}(d)Z\hat{M}Z^T\right). \tag{3.2.8}$$

3.2.2 The PE Bound

By Lemma (3.2.1), we can obtain the new formulation (PE) by using the variable $Z\in\mathbb{R}^{(n-1)\times(k-1)}$, which is equivalent to our original graph partitioning problem (GP) in (2.2.2):

$$\text{(PE)}\quad w^*_{\text{cut}}=\begin{array}{ll}\min & \left(\alpha(d)-\frac{s(d)}{2}\right)+\operatorname{tr}\left(C(d)Z^T\right)+\frac{1}{2}\operatorname{tr}\left(\hat{L}(d)Z\hat{M}Z^T\right)\\ \text{s.t.} & Z^TZ=I_{k-1},\\ & VZW^T\geq -\frac{1}{n}u_n\tilde{m}^T.\end{array} \tag{3.2.9}$$

Theorem 3.2.2. *[14, Theorem 3.1] Suppose X and Z are related by* (3.2.3). *Then X solves* (2.2.2) *if and only if Z solves* (3.2.9).

Theorem 3.2.3. *Let $d\in\mathbb{R}^n$. Then:*

$$\begin{aligned}w^*_{\text{cut}}\geq p^*_{peig}(d):&=\left(\alpha(d)-\tfrac{s(d)}{2}\right)+\min_{0\leq\frac{1}{n}u_n\tilde{m}^T+VZW^T}\operatorname{tr}\left(C(d)Z^T\right)+\tfrac{1}{2}\left\langle\lambda(\hat{L}(d)),\begin{pmatrix}\lambda(\hat{M})\\0\end{pmatrix}\right\rangle_-\\ &=\left(-\alpha(d)-\tfrac{s(d)}{2}\right)+\tfrac{1}{n}\min_{X\in\mathcal{D}}(Xm)^Td+\tfrac{1}{2}\left\langle\lambda(\hat{L}(d)),\begin{pmatrix}\lambda(\hat{M})\\0\end{pmatrix}\right\rangle_-.\end{aligned} \tag{3.2.10}$$

Proof. Notice that the objective function in (3.2.9) has three parts: the constant, the linear part, and the quadratic part. We separate the objective function into three parts and subject to different partial constraints. For the notational simplicity, we denote

$$S_Z:=\{Z\in\mathbb{R}^{(n-1)\times(k-1)}: Z^TZ=I_{k-1},\;\; \frac{1}{n}u_n\tilde{m}^T+VZW^T\geq 0\}.$$

Then

$$
\begin{aligned}
w^*_{\text{cut}} &= \min\left\{\left(\alpha(d)-\tfrac{s(d)}{2}\right)+\operatorname{tr}(C(d)Z^T)+\tfrac{1}{2}\operatorname{tr}(\hat{L}(d)Z\hat{M}Z^T) : Z\in S_Z\right\}\\
&\geq \left(\alpha(d)-\tfrac{s(d)}{2}\right)+\min\left\{\operatorname{tr}(C(d)Z^T) : Z\in S_Z\right\}+\tfrac{1}{2}\min\left\{\operatorname{tr}(\hat{L}(d)Z\hat{M}Z^T) : Z\in S_Z\right\}\\
&\geq \left(\alpha(d)-\tfrac{s(d)}{2}\right)+\min_{\frac{1}{n}u_n\tilde{m}^T+VZW^T\geq 0}\operatorname{tr}(C(d)Z^T)+\tfrac{1}{2}\min_{Z^TZ=I_{k-1}}\operatorname{tr}(\hat{L}(d)Z\hat{M}Z^T)\\
&= \left(\alpha(d)-\tfrac{s(d)}{2}\right)+\min_{0\leq\frac{1}{n}u_n\tilde{m}^T+VZW^T}\operatorname{tr}(C(d)Z^T)+\tfrac{1}{2}\left\langle\lambda(\hat{L}(d)),\begin{pmatrix}\lambda(\hat{M})\\0\end{pmatrix}\right\rangle_- =: p^*_{peig}(d)\\
&= \left(-\alpha(d)-\tfrac{s(d)}{2}\right)+\tfrac{1}{n}\min_{X\in\mathcal{D}}(Xm)^Td+\tfrac{1}{2}\left\langle\lambda(\hat{L}(d)),\begin{pmatrix}\lambda(\hat{M})\\0\end{pmatrix}\right\rangle_-.
\end{aligned}
\tag{3.2.11}
$$

The first equality is due to theorem (3.2.2). The second equality follows from

$$\min_{Z^TZ=I_{k-1}}\operatorname{tr}(\hat{L}(d)Z\hat{M}Z^T)=\left\langle\lambda(\hat{L}(d)),\begin{pmatrix}\lambda(\hat{M})\\0\end{pmatrix}\right\rangle_-$$

by theorem (3.1.2). The last equality follows from the relation in (3.2.7). □

Notice that $p^*_{peig}(0)=\frac{1}{2}\left\langle\lambda(\hat{L}),\begin{pmatrix}\lambda(\hat{M})\\0\end{pmatrix}\right\rangle_-$. Next we explore the information of eigenvalues of $\hat{L}$ and $\hat{M}$.

Lemma 3.2.4. *[13, Lemma 7] Let $v_1,v_2,\cdots,v_n=u_n$ be n eigenvectors of L, pairwise orthogonal, with eigenvalues $\lambda_1(L),\lambda_2(L),\cdots,\lambda_n(L)=0$. Then the eigenvalues of $\hat{L}$ are $\lambda_1(L),\lambda_2(L),\cdots,\lambda_{n-1}(L)$ with eigenvectors V^Tv_i for $i=1,2,\cdots,n-1$.*

Proof. Since $VV^T=I_n-\frac{1}{n}u_n{u_n}^T$ and $Lu_n=0$, we have $\hat{L}(V^Tv_i)=(V^TLV)(V^Tv_i)=V^TL(I_n-\frac{1}{n}u_n{u_n}^T)v_i=V^TLv_i=\lambda_iV^Tv_i$. □

Lemma 3.2.5. *Let $U\in\mathbb{R}^{p\times q}$ where $p\leq q$. Then* $\operatorname{rank}(U^TU)=\operatorname{rank}(U)=\operatorname{rank}(U^T)=\operatorname{rank}(UU^T)$.

Here, the notation $\mathcal{N}(U)=\{x\in\mathbb{R}^q : Ux=0\}$ is called the *null space* of U.

Proof. We start by showing $\mathcal{N}(U^TU)=\mathcal{N}(U)$. $\mathcal{N}(U)\subseteq\mathcal{N}(U^TU)$ is trivial.

Now we prove $\mathcal{N}(U^TU)\subseteq\mathcal{N}(U)$. Suppose $x\in\mathcal{N}(U^TU)$. Then $U^TUx=0$. So $(Ux)^T(Ux)=x^TU^TUx=0$, which implies $Ux=0$. So $x\in\mathcal{N}(U)$. So we have $\mathcal{N}(U^TU)\subseteq\mathcal{N}(U)$.

By the Rank-and-Nullity Theorem, we have $\operatorname{rank}(U^TU)=q-\dim\mathcal{N}(U^TU)=q-\dim\mathcal{N}(U)=\operatorname{rank}(U)$. □

Lemma 3.2.6. *$\hat{M}=W^TMW\succ 0$.*

Proof. $\hat{M}=W^T\tilde{M}\tilde{M}W\succeq 0$. Since $\tilde{M}$ is nonsingular, $\operatorname{rank}(\tilde{M}W)=\operatorname{rank}(W)=k-1$. By the above lemma (3.2.5), we have $\operatorname{rank}(\hat{M})=\operatorname{rank}(\tilde{M}W)=k-1$. Hence $\hat{M}\succ 0$. □

Theorem 3.2.7.

$$p^*_{peig}(0) = \frac{1}{2}\left\langle \lambda(\hat{L}), \begin{pmatrix}\lambda(\hat{M})\\ 0\end{pmatrix}\right\rangle_- = \frac{1}{2}\sum_{i=1}^{k-1}\lambda_i(\hat{M})\cdot\lambda_{n-i}(L) \geq 0.$$

Proof. The first equality follows from the definition of $p^*_{peig}(0)$. The second equality follows from the definition of minimal scalar product. We proved $\hat{M} \in \mathcal{S}^{k-1}_{++}$ in lemma (3.2.6) and $L \in \mathcal{S}^n_+$. So $\lambda_i(\hat{M}) > 0$ and $\lambda_i(L) \geq 0$. □

As we did in the basic eigenvalue bound method, we can simply use the eigenvalues of $\hat{L}$ and $\hat{M}$ to obtain a useful lower bound by above theorem.

Remark 3.2.8. *Since $p^*_{peig}(d)$ has both linear term and quadratic term and we optimize the two terms separately, the optimal solution of quadratic term will not be an optimal solution of linear term in general. Here we obtain an approximate solution by solving the optimal solution of the minimal scalar product. Let $Q \in \mathcal{O}_{k-1}$ with columns consisting of the eigenvectors of $\hat{M}$, defined in (3.2.6), corresponding to the eigenvalues of $\hat{M}$ in nondecreasing order. Let $P \in \mathbb{R}^{(n-1)\times(k-1)}$ be the matrix with orthonormal columns consisting of the $k-1$ eigenvectors of $\hat{L}(d)$, corresponding to the smallest $k-1$ eigenvalues of $\hat{L}(d)$ in nonincreasing order. From the theorem (3.1.2), the minimal scalar product term in (3.2.10) are attained at*

$$Z = PQ^T, \tag{3.2.12}$$

and the corresponding point in $\mathcal{E} \cap \mathcal{D}_O$ is

$$X = \widehat{X} + VZW^T\tilde{M}. \tag{3.2.13}$$

Corollary 3.2.9. *If the problem is graph equipartitioning, i.e. $m_1 = m_2 = \cdots = m_k = \frac{n}{k}$, then*

$$p^*_{peig}(0) = \frac{n}{2k}\sum_{i=1}^{k-1}\lambda_{n-i}(L).$$

Proof. If the problem is equipartitioning, then

$$\hat{M} = W^TMW = W^T(\frac{n}{k}I_k)W = \frac{n}{k}W^TW = \frac{n}{k}I_{k-1}.$$

So $\lambda_i(\hat{M}) = \frac{n}{k}, \forall i = 1, \ldots, k-1$. □

3.2.3 Explicit Solution for Linear Term

The constant term $-\alpha(d) - \frac{s(d)}{2} = \frac{s(d)}{2}\left(-\frac{s(M^2)}{n^2} - 1\right)$ in (3.2.10) can be computed easily. The minimal scalar product $\left\langle \lambda(\hat{L}(d)), \begin{pmatrix}\lambda(\hat{M})\\ 0\end{pmatrix}\right\rangle_- = \sum_{i=1}^{k-1}\lambda_i(\hat{M})\cdot\lambda_{n-i}(\hat{L}(d))$ in (3.2.10) can be also computed efficiently. We now going to show the linear term can be computed efficiently by deriving a explicit solution. In theorem (7.1.1), we have shown that $\mathcal{D}$ is the convex hull of $\mathcal{M}_m$. So $\min_{X\in\mathcal{D}}(Xm)^T d$ is equivalent to $\min_{X\in\mathcal{M}_m} \operatorname{tr}(Xm)^T d$. We define

$$x_0 := (\underbrace{m_1, \cdots, m_1}_{m_1}, \cdots, \underbrace{m_k, \cdots, m_k}_{m_k})^T \in \mathbb{R}^n,$$

and

$$X_0 := \begin{bmatrix} u_{m_1} & 0 & \cdots & 0 \\ 0 & u_{m_2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{m_k} \end{bmatrix} \in \mathcal{M}_m. \tag{3.2.14}$$

Notice that $X_0 m = x_0$.

Lemma 3.2.10. *Let $d \in \mathbb{R}^n$. Then*

$$\min_{X\in\mathcal{M}_m} (Xm)^T d = \langle d, x_0 \rangle_- \, .$$

Proof. Oberve that $X \in \mathcal{M}_m$ if and only if there exists a permutation matrix P on N such that $X = PX_0$. Let Π denote the set of all permutation matrices on N. Then

$$\begin{aligned} \min_{X\in\mathcal{M}_m} (Xm)^T d &= \min_{P\in\Pi}(PX_0 m)^T d \\ &= \min_{P\in\Pi}(Px_0)^T d \\ &= \langle d, x_0 \rangle_- \, . \end{aligned}$$

where the last equality follows from the definition of minimal scalar product. □

3.2.4 PE Bound via QAP

In this subsection, we first show GP problem can be convert into *Quadratic Assignment Problem QAP*, hence a special case of QAP. Formally, QAP consists of minimization

$$f(Y) = \operatorname{tr}((AYB^T + C)Y^T)$$

over the set of permutation matrices. A, B, and C are given real matrices defining the QAP.

We now convert the problem (3.0.2) into QAP.

Recall that $X \in \mathcal{M}_m$ if and only if there exists a permutation matrix Y on N such that $X = YX_0$, where X_0 is defined in (3.2.14).

So

$$\operatorname{tr}(X^T L(d)X) = \operatorname{tr}(L(d)XX^T) = \operatorname{tr}(L(d)YX_0{X_0}^T Y^T).$$

Define

$$T_0 := X_0{X_0}^T = \begin{bmatrix} E_{m_1} & 0 & \cdots & 0 \\ 0 & E_{m_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{m_k} \end{bmatrix} \in \mathbb{R}^{n\times n}$$

where X_0 is defined in (3.2.14) and E is a square matrix of all ones. Notice that $T_0 \succeq 0, \operatorname{rank}(T_0) = k$ and $\lambda_i(T_0) = m_i$ for $i = 1, \ldots, k$. Also, $r(T_0) = T_0 u_n = x_0$.

Let Π denote the set of permutation matrices. So the problem (3.0.2) is equivalent to the QAP

$$\begin{array}{rll} w^*_{\text{cut}} = & \min & \frac{1}{2} \operatorname{tr}(L(d) Y T_0 Y^T) - \frac{s(d)}{2} \\ & \text{s.t.} & Y \in \Pi. \end{array} \tag{3.2.15}$$

By considering the relaxation $Y \in \mathcal{O}_n$, we have the lower bound for w^*_{cut} as same as the basic eigenvalue bound.

Now define

$$\begin{array}{rcl} \widehat{\mathcal{N}} & := & \{Y \in \mathbb{R}^{n \times n} : Y_{ij} \geq 0, \forall ij\}. \\ \widehat{\mathcal{E}} & := & \{Y \in \mathbb{R}^{n \times n} : Y u_n = u_n, X^T u_n = u_n\}. \end{array}$$

Notice that

$$\Pi = \mathcal{O}_n \cap \widehat{\mathcal{N}} \cap \widehat{\mathcal{E}}.$$

We now project Π onto $\widehat{\mathcal{E}}$ by the following lemma, see [7].

Lemma 3.2.11. *[7, Lemma 3.1] Let Y be $n \times n$ and Z be $(n-1) \times (n-1)$. Suppose Y and Z satisfy*

$$Y = P \begin{bmatrix} 1 & 0 \\ 0 & Z \end{bmatrix} P^T. \tag{3.2.16}$$

Then

$$\begin{array}{c} Y \in \widehat{\mathcal{E}}, \\ Y \in \widehat{\mathcal{N}} \Leftrightarrow VZV^T \geq \frac{1}{n} u_n {u_n}^T, \\ Y \in \mathcal{O}_n \Leftrightarrow Z \in \mathcal{O}_{n-1}. \end{array}$$

Conversely, if $Y \in \widehat{\mathcal{E}}$, then there is a Z such that (3.2.16) *houlds.*

After substituting $Y = \frac{1}{n} u_n {u_n}^T + VZV^T$ into the problem (3.2.15), we have the following bound which is equal to our projected eigenvalue bound $p^*_{peig}(d)$:

$$\begin{aligned} w^*_{\text{cut}} & \geq \left(\alpha(d) - \tfrac{s(d)}{2}\right) + \tfrac{1}{n} \min_{VZV^T \geq \frac{1}{n} u_n {u_n}^T} \operatorname{tr}\left((V^T d {x_0}^T V) Z^T\right) + \tfrac{1}{2} \left\langle \lambda(\hat{L}(d)), \lambda(\hat{T}_0) \right\rangle_- \\ & = \left(-\alpha(d) - \tfrac{s(d)}{2}\right) + \tfrac{1}{n} \min_{Y \in \widehat{\mathcal{E}} \cap \widehat{\mathcal{N}}} (Y x_0)^T d + \tfrac{1}{2} \left\langle \lambda(\hat{L}(d)), \lambda(\hat{T}_0) \right\rangle_- \\ & = \left(-\alpha(d) - \tfrac{s(d)}{2}\right) + \tfrac{1}{n} \left\langle d, x_0 \right\rangle_- + \tfrac{1}{2} \left\langle \lambda(\hat{L}(d)), \lambda(\hat{T}_0) \right\rangle_- \\ & = p^*_{peig}(d), \end{aligned} \tag{3.2.17}$$

where $d \in \mathbb{R}^n, \hat{T}_0 = V^T T_0 V$. The last equality follows from the fact

$$\left\langle \lambda(\hat{L}(d)), \lambda(\hat{T}_0) \right\rangle_- = \left\langle \lambda(\hat{L}(d)), \lambda(\hat{M}) \right\rangle_- .$$

Chapter 4

Convex Quadratic Programming Bound

4.1 Introduction and Related Work

A new successful and efficient bound used for the quadratic assignment problem (QAP) is given in [1, 3]. In this chapter, we adapt the idea described there to obtain a lower bound for w^*_{cut}, which is stronger than the projected eigenvalue bound. This bound is obtained from a relaxation that is a Convex Quadratic Programming, i.e., the minimization of a quadratic function that is convex on the linear manifold defined by linear constraints. Approaches based on nonconvex QPs are given in e.g., [8].

The main idea in [1, 3] is to use the zero duality gap result for a homogeneous QAP [2, Theorem 3.2] on an objective obtained via a suitable reparametrization of the original problem. Following this idea, we consider the parametrization in (3.2.5) where our objective function in (3.0.2) is rewritten as:

$$\frac{1}{2}\operatorname{tr}(X^TLX) = \left(\alpha(d) - \frac{s(d)}{2}\right) + \operatorname{tr}\left(C(d)Z^T\right) + \frac{1}{2}\operatorname{tr}\left(\hat{L}(d)Z\hat{M}Z^T\right) \tag{4.1.1}$$

with X and Z related according to (3.2.3). Now we look at the homogeneous part:

$$\begin{array}{rl} v_p^* := \min & \frac{1}{2}\operatorname{tr}\left(\hat{L}(d)Z\hat{M}Z^T\right) \\ \text{s.t.} & Z^TZ = I_{k-1}, \\ & (ZZ^T \preceq I_{n-1}.) \end{array} \tag{4.1.2}$$

Lemma 4.1.1. *If $H \in \mathcal{O}_{n,k}$, then $HH^T \preceq I_n$.*

Proof. Since $H \in \mathcal{O}_{n,k}$, there exists $H_0 \in \mathcal{O}_{n,n-k}$, such that $[H_0 \;\; H] \in \mathcal{O}_n$. Then $I_n = [H_0 \;\; H][H_0 \;\; H]^T = H_0H_0^T + HH^T$. Since $0 \preceq H_0H_0^T = I_n - HH^T$, we have $HH^T \preceq I_n$. □

By lemma (4.1.1), $ZZ^T \preceq I_{n-1}$ is a redundant constraint in (4.1.2). But it does not mean that it is redundant in our Lagrangian dual problem because we may close the duality gap if we add the redundant constraint.

The Lagrangian dual problem of (4.1.2) is the following, with variables in Lagrangian multipliers S and T:

$$\begin{array}{rl} v_d^* := \max & \frac{1}{2}\operatorname{tr}(S) + \frac{1}{2}\operatorname{tr}(T) \\ \text{s.t.} & I_{k-1} \otimes S + T \otimes I_{n-1} \preceq \hat{M} \otimes \hat{L}(d), \\ & S \preceq 0, \\ & S \in \mathcal{S}^{n-1},\ T \in \mathcal{S}^{k-1}. \end{array} \tag{4.1.3}$$

Claim 4.1.2. *[12, Theorem 2]* $v_p^* = v_d^*$.

Proof. It is clear that $v_p^* \geq v_d^*$. Next we are going to show $v_p^* \leq v_d^*$.

Write $\hat{M} = U_1 \operatorname{Diag}(\lambda) U_1^T$ and $\hat{L}(d) = U_2 \operatorname{Diag}(\sigma) U_2^T$ in eigenvalue orthogonal decomposition forms. We substitute $\hat{S} = U_2^T S U_2$ and $\hat{T} = U_1^T T U_1$ and rename of $\hat{S}$ and $\hat{T}$. Then we have

$$\begin{array}{rl} 2v_d^* = \max & \operatorname{tr}(S) + \operatorname{tr}(T) \\ \text{s.t.} & I_{k-1} \otimes S + T \otimes I_{n-1} \preceq \operatorname{Diag}(\lambda) \otimes \operatorname{Diag}(\sigma), \\ & S \preceq 0, \\ & S \in \mathcal{S}^{n-1},\ T \in \mathcal{S}^{k-1}. \end{array} \tag{4.1.4}$$

Suppose (S_0, T_0) is a pair of optimal solution to (4.1.4). It is easy to verify that

$$(\operatorname{Diag}(\operatorname{diag}(S_0)), \operatorname{Diag}(\operatorname{diag}(T_0))$$

is also a optimal solution to (4.1.4). So it suffices to consider the variables S and T to be diagnal matrices. So we can reduce (4.1.4) to solving the following LP with variables $s \in \mathbb{R}^{n-1}$ and $t \in \mathbb{R}^{k-1}$,

$$\begin{array}{rl} 2v_d^* = \max & {u_{n-1}}^T s + {u_{k-1}}^T t \\ \text{s.t.} & t_i + s_j \leq \lambda_i \sigma_j, \quad i = 1, \ldots, k-1, \quad j = 1, \ldots, n-1, \\ & s_j \leq 0, \quad j = 1, \ldots, n-1. \end{array} \tag{4.1.5}$$

The dual problem of (4.1.5) is

$$\begin{array}{rl} 2v_d^* = \min & \sum_{i=1}^{k-1} \sum_{j=1}^{n-1} \lambda_i \sigma_j z_{ij} \\ \text{s.t.} & \sum_{j=1}^{n-1} z_{ij} = 1, \quad i = 1, \ldots, k-1, \\ & \sum_{i=1}^{k-1} z_{ij} + y_j = 1, \quad j = 1, \ldots, n-1, \\ & z_{ij} \geq 0, \quad i = 1, \ldots, k-1, j = 1, \ldots, n-1, \\ & y_j \geq 0, \quad j = 1, \ldots, n-1. \end{array} \tag{4.1.6}$$

Notice that (4.1.6) is totally unimodular. So there is an optimal solution (z^*, y^*) which is integral. z^* defines an injection $\phi^* : \{1, \ldots, k-1\} \to \{1, \ldots, n-1\}$ with $\phi^*(i) = j$, if $z_{ij} = 1$. Hence we have

$$2v_d^* = \sum_{i=1}^{k-1} \sum_{j=1}^{n-1} \lambda_i \sigma_j z_{ij}^* = \sum_{i=1}^{k-1} \lambda_i \sigma_{\phi^*(i)} \geq \left\langle \lambda(\hat{L}(d)), \begin{pmatrix} \lambda(\hat{M}) \\ 0 \end{pmatrix} \right\rangle_- = 2v_p^*.$$

□

Remark 4.1.3. *The* SDP *in* (4.1.3) *can be efficiently solved as the* LP *in* (4.1.5). *If we have an optimal solution* (s^*, t^*) *of* (4.1.5), *we can recover an optimal solution of* (4.1.3) *as*

$$S^* = U_2 \operatorname{Diag}(s^*) U_2^T \quad \text{and} \quad T^* = U_1 \operatorname{Diag}(t^*) U_1^T. \tag{4.1.7}$$

4.2 QP Bound

Next, suppose that the optimal value of the dual problem (4.1.3) is attained at (S^*, T^*). Let Z be such that the X defined according to (3.2.3) is a partition matrix. Then we have

$$\begin{array}{rcl} \frac{1}{2}\operatorname{tr}(\hat{L}(d) Z \hat{M} Z^T) & = & \frac{1}{2}\operatorname{vec}(Z)^T(\hat{M} \otimes \hat{L}(d))\operatorname{vec}(Z) \\ & = & \frac{1}{2}\operatorname{vec}(Z)^T \underbrace{(\hat{M} \otimes \hat{L}(d) - I_{k-1} \otimes S^* - T^* \otimes I_{n-1})}_{\widehat{Q}} \operatorname{vec}(Z) \\ & & +\frac{1}{2}\operatorname{tr}(S^* Z I_{k-1} Z^T) + \frac{1}{2}\operatorname{tr}(I_{n-1} Z T^* Z^T) \\ & = & \frac{1}{2}\operatorname{vec}(Z)^T \widehat{Q} \operatorname{vec}(Z) + \frac{1}{2}\operatorname{tr}(Z Z^T S^*) + \frac{1}{2}\operatorname{tr}(T^*) \\ & = & \frac{1}{2}\operatorname{vec}(Z)^T \widehat{Q} \operatorname{vec}(Z) + \frac{1}{2}\operatorname{tr}([Z Z^T - I_{n-1}] S^*) + \frac{1}{2}\operatorname{tr}(S^*) + \frac{1}{2}\operatorname{tr}(T^*) \\ & \geq & \frac{1}{2}\operatorname{vec}(Z)^T \widehat{Q} \operatorname{vec}(Z) + \frac{1}{2}\operatorname{tr}(S^*) + \frac{1}{2}\operatorname{tr}(T^*), \end{array}$$

where the last inequality uses $S^* \preceq 0$ and $ZZ^T \preceq I_{n-1}$. Notice that our $\widehat{Q} \succeq 0$ since $I_{k-1} \otimes S^* + T^* \otimes I_{n-1} \preceq \hat{M} \otimes \hat{L}(d)$ since (S^*, T^*) is a feasible solution to (4.1.3).

Recall that the original nonconvex problem (3.0.2) is equivalent to minimizing the right hand side of (4.1.1) over the set of all Z so that the X defined in (3.2.3) corresponds to a partition matrix. From the above relations, the third equality in (2.1.1) and Lemma 3.2.1, we see that

$$\begin{array}{rl} w^*_{\text{cut}} \geq \min & \left(\left(\alpha(d) - \frac{s(d)}{2}\right) + \operatorname{tr}\left(C(d) Z^T\right) + \frac{1}{2}\operatorname{vec}(Z)^T \widehat{Q} \operatorname{vec}(Z) \right) + \frac{1}{2}\operatorname{tr}(S^*) + \frac{1}{2}\operatorname{tr}(T^*) \\ \text{s.t.} & Z^T Z = I_{k-1}, \\ & \widehat{X} + V Z W^T \tilde{M} \geq 0. \end{array} \tag{4.2.1}$$

We also recall from (4.1.3) that $\frac{1}{2}\operatorname{tr}(S^*) + \frac{1}{2}\operatorname{tr}(T^*) = v_d^* = v_p^*$, which further equals

$$\frac{1}{2}\left\langle \lambda(\hat{L}(d)), \begin{pmatrix} \lambda(\hat{M}) \\ 0 \end{pmatrix} \right\rangle_-$$

according to (4.1.2) and Theorem (3.1.2).

A lower bound can now be obtained by relaxing the constraints in (4.2.1). For example, by dropping the orthogonality constraints, we obtain the following lower bound on w^*_{cut}:

$$\begin{array}{rcl} p^*_{QP}(d) & := & \min \quad R(Z) := \left(\alpha(d) - \frac{s(d)}{2}\right) + \operatorname{tr}\left(C(d) Z^T\right) + \frac{1}{2}\operatorname{vec}(Z)^T \widehat{Q} \operatorname{vec}(Z) \\ & & \qquad\qquad + \frac{1}{2}\left\langle \lambda(\hat{L}(d)), \begin{pmatrix} \lambda(\hat{M}) \\ 0 \end{pmatrix} \right\rangle_- \\ & & \text{s.t.} \quad \widehat{X} + V Z W^T \tilde{M} \geq 0. \end{array} \tag{4.2.2}$$

Notice that this is a QP with $(n-1)(k-1)$ variables and nk constraints.

As in [1, Page 346], it is possible to reformulate (4.2.2) into a QP in variables $X \in \mathcal{D}$. Note that $\tilde{Q}$ defined in (4.2.4) is not positive semidefinite in general. Nevertheless, the QP is implicitly convex. Also notice that

$$p^*_{QP}(d) \geq p^*_{peig}(d) + \min_{\widehat{X}+VZW^T\tilde{M}\geq 0} \frac{1}{2}\mathrm{vec}(Z)^T\widehat{Q}\mathrm{vec}(Z).$$

Since $\widehat{Q} \succeq 0$, we have $p^*_{PQ}(d) \geq p^*_{peig}(d)$. The euquality can hold. If $d=0$, then $\alpha(d) - \frac{s(d)}{2} = 0$ and $C(d) = 0$. $Z = 0$ is in our feasible region $\widehat{X} + VZW^T\tilde{M} \geq 0$. So we have $p^*_{QP}(0) = p^*_{peig}(0) \geq 0$.

Theorem 4.2.1. *Let* (S^*, T^*) *be optimal solutions of* (4.1.3) *as defined in* (4.1.7)*. A lower bound on* w^*_{cut} *is obtained from the following QP:*

$$w^*_{\mathrm{cut}} \geq p^*_{QP}(d) = \min_{X\in\mathcal{D}} \frac{1}{2}\,\mathrm{vec}(X)^T\tilde{Q}\,\mathrm{vec}(X) + \frac{1}{2}\left\langle \lambda(\hat{L}(d)), \begin{pmatrix}\lambda(\hat{M})\\ 0\end{pmatrix}\right\rangle_- - \frac{s(d)}{2} \qquad (4.2.3)$$

where

$$\tilde{Q} := I_k \otimes L(d) - M^{-1} \otimes (VS^*V^T) - (\tilde{M}^{-1}WT^*W^T\tilde{M}^{-1}) \otimes I_n. \qquad (4.2.4)$$

The QP in (4.2.3) *is implicitly convex since* $\tilde{Q}$ *is positive semidefinite on the tangent space of* $\mathcal{E}$*.*

Proof. We start by rewriting the quadratic term of $R(Z)$ in (4.2.2) using the relation (3.2.3).

Since $V^TV = I_{n-1}$ and $W^TW = I_{k-1}$, we have from the definitions of $\hat{M}$ and $\hat{L}(d)$ that

$$\begin{aligned}\widehat{Q} &= \hat{M} \otimes \hat{L}(d) - I_{k-1} \otimes S^* - T^* \otimes I_{n-1}\\ &= (W^T\tilde{M}I_k\tilde{M}W) \otimes (V^TL(d)V) - I_{k-1} \otimes S^* - T^* \otimes I_{n-1}\\ &= \Big((\tilde{M}W)\otimes V\Big)^T \underbrace{[I_k \otimes L(d) - M^{-1} \otimes (VS^*V^T) - (\tilde{M}^{-1}WT^*W^T\tilde{M}^{-1}) \otimes I_n]}_{\tilde{Q}} \Big((\tilde{M}W)\otimes V\Big).\end{aligned} \qquad (4.2.5)$$

On the other hand, from (3.2.4), we have

$$\mathrm{vec}(X - \widehat{X}) = \mathrm{vec}(VZW^T\tilde{M}) = \Big((\tilde{M}W)\otimes V\Big)\,\mathrm{vec}(Z).$$

Hence, the quadratic term in $R(Z)$ can be rewritten as

$$\mathrm{vec}(Z)^T\widehat{Q}\,\mathrm{vec}(Z) = \mathrm{vec}(X-\widehat{X})^T\tilde{Q}\,\mathrm{vec}(X-\widehat{X}), \qquad (4.2.6)$$

where $\tilde{Q}$ is defined in (4.2.4). Next, we see from $\mathrm{vec}(\widehat{X}) = m \otimes u_n$ and $V^Tu_n = 0$ that

$$\big(M^{-1}\otimes(VS^*V^T)\big)\,\mathrm{vec}(\widehat{X}) = \frac{1}{n}\big(M^{-1}\otimes(VS^*V^T)\big)(m\otimes u_n) = \frac{1}{n}u_k \otimes (VS^*V^Tu_n) = 0.$$

Similarly, since $W^T\tilde{m} = 0$, we also have

$$\begin{aligned}\left((\tilde{M}^{-1}WT^*W^T\tilde{M}^{-1}) \otimes I_n\right)\text{vec}(\widehat{X}) &= \frac{1}{n}\left((\tilde{M}^{-1}WT^*W^T\tilde{M}^{-1}) \otimes I_n\right)(m \otimes u_n)\\ &= \frac{1}{n}(\tilde{M}^{-1}WT^*W^T\tilde{m}) \otimes u_n = 0.\end{aligned}$$

Combining the above two relations with (4.2.6), we obtain further that

$$\begin{aligned}&\frac{1}{2}\text{vec}(Z)^T\widehat{Q}\text{vec}(Z)\\ =&\frac{1}{2}\text{vec}(X)^T\tilde{Q}\text{vec}(X) - \text{vec}(\widehat{X})^T[I_k \otimes L(d)]\text{vec}(X) + \frac{1}{2}\text{vec}(\widehat{X})^T[I_k \otimes L(d)]\text{vec}(\widehat{X})\\ =&\frac{1}{2}\text{vec}(X)^T\tilde{Q}\text{vec}(X) - \text{tr}\left(L(d)\widehat{X}X^T\right) + \frac{1}{2}\text{tr}(L(d)\widehat{X}\widehat{X}^T)\\ =&\frac{1}{2}\text{vec}(X)^T\tilde{Q}\text{vec}(X) - \frac{1}{n}\text{tr}(dm^TX^T) + \alpha(d).\end{aligned}$$

For the first two terms of $R(Z)$, we have

$$\left(\alpha(d) - \frac{s(d)}{2}\right) + \text{tr}\left(C(d)Z^T\right) = \left(-\alpha(d) - \frac{s(d)}{2}\right) + \frac{1}{n}\text{tr}(dm^TX^T).$$

Furthermore, recall from Lemma (3.2.1) that with X and Z related by (3.2.3), $X \in \mathcal{D}$ if, and only if, $VZW^T\tilde{M} \geq -\widehat{X}$.

The conclusion in (4.2.3) now follows by substituting the above expressions into (4.2.2).

Finally, from (4.2.5) we see that $\tilde{Q}$ is positive semidefinite when restricted to the range of $\tilde{M}W \otimes V$. This is precisely the tangent space of $\mathcal{E}$. □

Notice that (4.2.2) and (4.2.3) are equivalent. (4.2.3) has nk variables and $nk + n + k$ constraints. But the constraints of (4.2.3) look simpler than that of (4.2.2).

Chapter 5

Semidefinite Programming Relaxation Bound

In this chapter, we are going to apply semidefinite programming (SDP) method to compute lower bounds for GP problem. We are going to present two methods to obtain the *same* SDP relaxation. One way is doing *lifting process* through quadratic formulation. This method is called the direct approach. The other method is using Lagrangian relaxation.

5.1 The Direct Approach to SDP Relaxation

We now show that SDP relaxation can be obtained from lifting process, i.e., we lift the vector $x = \text{vec}(X)$ into the matrix space $\mathcal{S}^{n^2+1}$.

We starts with the equivalent quadratically constrained quadratic formulation:

$$\begin{array}{rll} w^*_{cut} = & \min & \frac{1}{2}\,\text{tr}(X^T L X) \\ & \text{s.t.} & X \circ X = X, \\ & & \|Xu_k - u_n\|^2 = 0, \\ & & \left\|X^T u_n - m\right\|^2 = 0, \\ & & X_{:i} \circ X_{:j} = 0 \quad \forall i \neq j. \end{array} \tag{5.1.1}$$

Here: $\circ$ is the Hadamard (elementwise) product. The last constraint is redundant. But it may not be redundant in our SDP relaxation.

Now we do the lifting process.

First, we define

$$Y_X := \begin{pmatrix} 1 \\ \text{vec}(X) \end{pmatrix} \begin{pmatrix} 1 & \text{vec}(X)^T \end{pmatrix} = \begin{bmatrix} 1 & \text{vec}(X)^T \\ \text{vec}(X) & \text{vec}(X)\text{vec}(X)^T \end{bmatrix} \in \mathbb{R}^{nk+1}.$$

Then $Y_X \succeq 0$ and rank(Y_X)=1.

Remark 5.1.1. *$Y \in \mathcal{S}^p_+$ and $\text{rank}(Y) = 1$ if and only if there exists $x \in \mathbb{R}^p$ such that $Y = xx^T$ by Item 4 of Theorem* (2.3.2).

Define

$$L_0 := \begin{bmatrix} 0 & 0 \\ 0 & I_k \otimes L \end{bmatrix}.$$

Then our objective function becomes

$$\frac{1}{2}\operatorname{tr}(X^T L X) = \frac{1}{2}\operatorname{tr}(L X I_k X^T) = \frac{1}{2}\operatorname{vec}(X)^T (I_k \otimes L)\operatorname{vec}(X) = \frac{1}{2}\operatorname{tr}(L_0 Y_X).$$

The first constraint $X \circ X = X$ is equivalent to $X_{ij} \in \{0,1\}, \forall ij$ which is further equivalent to

$$\operatorname{vec}(X) = \operatorname{diag}\left(\operatorname{vec}(X)\operatorname{vec}(X)^T\right).$$

We now define the linear mapping arrow : $\mathbb{R}^{(nk+1)\times(nk+1)} \longrightarrow \mathbb{R}^{nk+1}$ to be

$$\operatorname{arrow}(Y) := \operatorname{diag}(Y) - (0, Y_{0,1:nk})^T.$$

Therefore, $X \circ X = X$ is equivalent to

$$\operatorname{arrow}(Y_X) = e_0.$$

Observe that

$$\begin{aligned} \|Xu_k - u_n\|^2 &= (Xu_k - u_n)^T (Xu_k - u_n) \\ &= {u_k}^T X^T X u_k - 2{u_n}^T X u_k + u_n^T u_n \\ &= \operatorname{tr}(I_n X u_k {u_k}^T X^T) - 2{u_n}^T X u_k + n \\ &= \operatorname{vec}^T(X)[(u_k {u_k}^T) \otimes I_n]\operatorname{vec}(X) - 2\operatorname{vec}^T(X)(u_k \otimes u_n) + n \end{aligned}$$

and

$$\begin{aligned} \left\|X^T u_n - m\right\|^2 &= (X^T u_n - m)^T (X^T u_n - m) \\ &= u_n^T X X^T u_n - 2{u_n}^T X m + m^T m \\ &= \operatorname{tr}(u_n {u_n}^T X I_k X^T) - 2{u_n}^T X m + m^T m \\ &= \operatorname{vec}^T(X)[I_k \otimes (u_n {u_n}^T)]\operatorname{vec}(X) - 2\operatorname{vec}^T(X)(m \otimes u_n) + m^T m. \end{aligned}$$

We define $D_1, D_2 \in \mathcal{S}^{nk+1}$ to be

$$D_1 := \begin{bmatrix} n & -{u_k}^T \otimes {u_n}^T \\ -u_k \otimes u_n & (u_k {u_k}^T) \otimes I_n \end{bmatrix} \quad \text{and} \quad D_2 := \begin{bmatrix} m^T m & -m^T \otimes {u_n}^T \\ -m \otimes u_n & I_k \otimes (u_n {u_n}^T) \end{bmatrix}.$$

Then $\|Xu_k - u_n\|^2 = 0$ and $\left\|X^T u_n - m\right\|^2 = 0$ is equivalent to

$$\operatorname{tr}(D_1 Y_X) = 0 \text{ and } \operatorname{tr}(D_2 Y_X) = 0.$$

We now define the gangster operator $\mathcal{G}_J : \mathcal{S}^{nk+1} \to \mathcal{S}^{nk+1}$ to be

$$(\mathcal{G}_J(Y))_{ij} := \begin{cases} Y_{ij} & \text{if } (i,j) \text{ or } (j,i) \in J, \\ 0 & \text{otherwise,} \end{cases}$$

where the set

$$J := \left\{ (i,j) : \begin{array}{l} i = (p-1)n + q, \\ j = (r-1)n + q, \end{array} \quad \text{for} \quad \begin{array}{c} p < r, \\ p, r \in \{1, \dots, k\}, q \in \{1, \dots, n\} \end{array} \right\}.$$

The Hadamard constraint $X_{:i} \circ X_{:j} = 0, \forall i \neq j$ is equivalent to

$$\mathcal{G}_J(Y_X) = 0.$$

We can see that the gangster operator $\mathcal{G}_J$ shoots many "holes" in the matrix Y_X.

If we ignore the last rank-one hard constraint and use a general symmetric matrix Y rather than Y_X, we obtain the following SDP relaxation:

$$\begin{array}{rrll} w^*_{\text{cut}} \geq & p^*_{SDP} := & \min & \frac{1}{2}\operatorname{tr}(L_0 Y) \\ & & \text{s.t.} & \operatorname{arrow}(Y) = e_0, \\ & & & \operatorname{tr}(D_1 Y) = 0, \\ & & & \operatorname{tr}(D_2 Y) = 0, \\ & & & \mathcal{G}_J(Y) = 0, \\ & & & Y_{00} = 1, \\ & & & Y \succeq 0. \end{array} \tag{5.1.2}$$

5.2 Lagrangian Relaxation

In this section, we develop the SDP relaxation constructed from the various equality constriants in the representation in (2.1.3) and the objective function in (2.2.1). We follow the approach in [18].

We start with the following equivalent quadratically constrained quadratic problem to (GP) in (2.2.2):

$$\begin{array}{rlllll} w^*_{cut} = & \min & \frac{1}{2}\operatorname{tr}(AXBX^T) = & & \min & \frac{1}{2}\operatorname{tr}(AXBX^T) \\ & \text{s.t.} & X \circ X = X, & & \text{s.t.} & X \circ X = x_0 X, \\ & & \|Xu_k - u_n\|^2 = 0, & & & \|Xu_k - x_0 u_n\|^2 = 0, \\ & & \left\|X^T u_n - m\right\|^2 = 0, & & & \left\|X^T u_n - x_0 m\right\|^2 = 0, \\ & & X_{:i} \circ X_{:j} = 0, \forall i \neq j, & & & X_{:i} \circ X_{:j} = 0, \forall i \neq j, \\ & & X^T X - M = 0, & & & X^T X - M = 0, \\ & & \operatorname{diag}(XX^T) - u_n = 0, & & & \operatorname{diag}(XX^T) - u_n = 0, \\ & & & & & {x_0}^2 = 1. \end{array} \tag{5.2.1}$$

Here we use a trick of adding a new variable x_0 and a new constraint ${x_0}^2 = 1$ to the second optimization problem in (5.2.1). The reason is that we can kill all linear terms in our Lagrangian $\mathcal{L}$, while not changing the optimal value. x_0 can only take values 1 or -1 in the second problem. $(X, 1)$ is an optimal solution of second problem if and only if X is an optimal solution of first problem. (X,-1) is an optimal solution of second problem if and only if $-X$ is an optimal solution of the first problem. In both cases, the two problems are equivalent and their optimal values agree. Again, the last two constraints in the first problem of (5.2.1) are redundant. They may be not redundant in the Lagrangian dual.

The Lagrangian of second optimization problem in (5.2.1) is the sum of the objective function along with inner-products of the Lagrangian multipliers and the corresponding constraints.

$$\begin{aligned}\mathcal{L}(X, x_0, \Gamma, \beta, \mathcal{G}, \Psi, \phi, t) = \; & \tfrac{1}{2}\operatorname{tr}(AXBX^T) \\ & + \langle \Gamma, X \circ X \rangle + \beta({u_k}^T X^T X u_k + {u_n}^T X X^T u_n) + \textstyle\sum_{i \neq j} \mathcal{G}_{ij}(X_{:i} \circ X_{:j}) \\ & + \left\langle \Psi, X^T X \right\rangle + \left\langle \phi, \operatorname{diag}(XX^T) \right\rangle + t{x_0}^2 \\ & - \langle \Gamma, x_0 X \rangle - 2\beta(x_0 {u_n}^T X u_k + x_0 m^T X^T u_n) \\ & + \beta(n + m^T m){x_0}^2 - \langle \Psi, M \rangle - \langle \phi, u_n \rangle - t.\end{aligned}$$

Then we use the *implicit constraint* that the Hessian of the Lagrangian must be positive semidefinite in the Lagrangian relaxation

$$\max_{\Gamma, \beta, \mathcal{G}, \Psi, \phi, t} \left(\min_{X, x_0} \mathcal{L}(X, x_0, \Gamma, \beta, \mathcal{G}, \Psi, \phi, t) \right).$$

Moreover, there is a hidden constraint that we want the inner minimization problem to be bounded below. So the inner minimization is attained at $x_0 = 0$ and $X = 0$. Plugging these in, we obtain a maximization SDP in the Lagrangian multipliers.

$$\begin{array}{rcll} w^*_{\text{cut}} & \geq & \max\limits_{\Gamma, \beta, \mathcal{G}, \Psi, \phi, t} & - \langle \Psi, M \rangle - \langle \phi, u_n \rangle - t \\ & & \text{s.t.} & \nabla^2_{(X, x_0)} \mathcal{L}(X, x_0, \Gamma, \beta, \mathcal{G}, \Psi, \phi, t) \succeq 0. \end{array} \tag{5.2.2}$$

Finally, we take the dual of (5.2.2), using the adjoints of the linear transformations in the constraints in (5.2.2) and obtain an SDP relaxation of (5.2.1):

$$\begin{array}{rll} w^*_{\text{cut}} \geq p^*_{LSDP} := & \min & \tfrac{1}{2}\operatorname{tr}(L_0 Y) \\ & \text{s.t.} & \operatorname{arrow}(Y) = e_0, \\ & & \operatorname{tr}(D_1 Y) = 0, \\ & & \operatorname{tr}(D_2 Y) = 0, \\ & & \mathcal{G}_J(Y) = 0, \\ & & \mathcal{D}_O(Y) = M, \\ & & \mathcal{D}_e(Y) = u_n, \\ & & Y_{00} = 1, \\ & & Y \succeq 0. \end{array} \tag{5.2.3}$$

Now we denote the set of feasible solution of (5.2.3) by $\mathcal{F}$.

By abuse of notation, we use the symbols for the sets of constraints $\mathcal{D}_O, \mathcal{D}_e$ to represent the linear transformations in the SDP relaxation in (5.2.3). Note that

$$\left\langle \Psi, X^T X \right\rangle = \operatorname{tr}(I_n X \Psi X^T) = \operatorname{vec}(X)^T (\Psi \otimes I_n) \operatorname{vec}(X).$$

Therefore, ${\mathcal{D}_O}^*$, the adjoint of $\mathcal{D}_O$, is made up of a zero row and column and k^2 blocks that are multiples of the identity matrix:

$$
{\mathcal{D}_O}^*(\Psi) = \begin{bmatrix} 0 & 0 \\ 0 & \Psi \otimes I_n \end{bmatrix}.
$$

If we block Y appropriately as:

$$
Y = \begin{bmatrix} Y_{00} & Y_{0,:} \\ Y_{:,0} & \bar{Y} \end{bmatrix}, \bar{Y} = \begin{bmatrix} \bar{Y}_{(11)} & \bar{Y}_{(12)} & \cdots & \bar{Y}_{(1k)} \\ \bar{Y}_{(21)} & \bar{Y}_{(22)} & \cdots & \bar{Y}_{(2k)} \\ \vdots & \ddots & \ddots & \vdots \\ \bar{Y}_{(k1)} & \bar{Y}_{(k2)} & \cdots & \bar{Y}_{(kk)} \end{bmatrix},
$$

with each $\bar{Y}_{(ij)}$ being a $n \times n$ matrix, then

$$
\mathcal{D}_O(Y) = [\operatorname{tr}(\bar{Y}_{(ij)})] \in \mathcal{S}^k.
$$

Similarly,

$$
\left\langle \phi, \operatorname{diag}(XX^T) \right\rangle = \left\langle \operatorname{Diag}(\phi), XX^T \right\rangle = \operatorname{vec}(X)^T \left(I_k \otimes \operatorname{Diag}(\phi) \right) \operatorname{vec}(X).
$$

So we get the ${\mathcal{D}_e}^*$, the adjoint of $\mathcal{D}_e$:

$$
{\mathcal{D}_e}^*(\phi) = \begin{bmatrix} 0 & 0 \\ 0 & I_k \otimes \operatorname{Diag}(\phi) \end{bmatrix}.
$$

Therefore we get the sum of the diagonal parts

$$
\mathcal{D}_e(Y) = \sum_{i=1}^{k} \operatorname{diag}(\bar{Y}_{(ii)}) \in \mathbb{R}^n.
$$

Claim 5.2.1. *$\mathcal{D}_O(Y) = M$ and $\mathcal{D}_e(Y) = u_n$ are redundant in* (5.2.3).

Proof. Write $v := Y_{0:kn,0}, v_1 := Y_{1:kn,0}$, and $X = \operatorname{Mat}(v_1)$. So we have

$$
\begin{aligned}
\|Xu_k - u_n\|^2 &= u_k^T X^T X u_k - 2u_k^T X^T u_n + u_n^T u_n \\
&= \operatorname{vec}^T(X)[(u_k u_k^T) \otimes I_n]\operatorname{vec}(X) - 2\operatorname{vec}^T(X)(u_k \otimes u_n) + n \\
&= v_1^T[(u_k u_k^T) \otimes I_n]v_1 - 2v_1^T(u_k \otimes u_n) + n \\
&= \operatorname{tr}(D_1 v v^T)
\end{aligned}
$$

and

$$
\begin{aligned}
\left\|X^T u_n - m\right\|^2 &= u_n^T X X^T u_n - 2m^T X^T u_n + m^T m \\
&= \operatorname{vec}^T(X)[I_k \otimes (u_n u_n^T)]\operatorname{vec}(X) - 2\operatorname{vec}^T(X)(m \otimes u_n) + m^T m \\
&= v_1^T[I_k \otimes (u_n u_n^T)]v_1 - 2v_1^T(m \otimes u_n) + m^T m \\
&= \operatorname{tr}(D_2 v v^T).
\end{aligned}
$$

By Schur complement of $Y_{00} = 1$, $Y \succeq 0$ if and only if $\bar{Y} \succeq v_1 v_1^T$ if and only if $Y \succeq \begin{bmatrix} 1 & v_1^T \\ v_1 & v_1 v_1^T \end{bmatrix} = vv^T$. We see further that

$$0 = \operatorname{tr}(D_i Y) \geq \operatorname{tr}(D_i vv^T) = \begin{cases} \|Xu_k - u_n\|^2 & \text{if} \quad i = 1, \\ \|X^T u_n - m\|^2 & \text{if} \quad i = 2. \end{cases}$$

$X^T u_n = m$ together with the arrow constraint imply that $\operatorname{tr}(\bar{Y}_{ii}) = \sum_{j=(i-1)n+1} Y_{j0} = m_i$. Thus $\mathcal{D}_O(Y) = M$ holds. Similarly, $Xu_k = u_n$ together with the arrow constraint imply that $\mathcal{D}_e(Y) = u_n$. □

5.3 The Final Semidefinite Relaxation Through Facial Reduction

Claim 5.3.1. *$D_1 \succeq 0$ and $D_2 \succeq 0$.*

Proof. Observe that $\operatorname{rank}(D_1) = n$ and $\operatorname{rank}(D_2) = k$.

Let $B_1 = \begin{pmatrix} -u_n^T \\ I_n \\ I_n \\ \vdots \\ I_n \end{pmatrix} \in \mathbb{R}^{(nk+1)\times n}$. Then $D_1 = B_1 B_1^T$. So $D_1 \succeq 0$.

Let $B_2 = \begin{pmatrix} -m^T \\ J_1 \\ J_2 \\ \vdots \\ J_k \end{pmatrix} \in \mathbb{R}^{(nk+1)\times k}$, where $J_l \in \mathbb{R}^{n\times k}$ with everywhere 0 but l-th column all ones. Then $D_2 = B_2 B_2^T$. So $D_2 \succeq 0$. □

Since $D_1, D_2, Y \succeq 0$, by Theorem (2.3.12), we have $D_1 Y = 0$ and $D_2 Y = 0$. So we cannot find a feasible $Y \succ 0$ such that $\operatorname{tr}(D_1 Y) = \operatorname{tr}(D_2 Y) = 0$. So we encounter numerical difficulties if we apply the Interior Point Method. Actually, for many problems in the reality, the Slater's condition fails. But by Theorem (2.3.29), every nonempty face of $\mathcal{S}_+^n$ is uniquely characterized. So we can find the minimal face of $\mathcal{S}_+^n$ which contains $\mathcal{F}$, the feasible set of (5.2.3), by finding the barycenter point in the relative interior of the minimal face. Because the minimal face we found is isomorphic to a smaller dimensional space $\mathcal{S}_+^q$, where $q < n$, we can project $\mathcal{F}$ onto $\mathcal{S}_+^q$. This procedure is called the *facial reduction*. We now explain the procedure in detail.

Let $X \in \mathcal{M}_m$ and $x = \operatorname{vec}(X)$. $Y_X = \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 & x^T \end{pmatrix} = \begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix} \in \mathcal{F}$. Observe that

$|\mathcal{M}_m| = \frac{m_1!\cdots m_k!}{n!}$. We define the *barycenter* point

$$\hat{Y} := \frac{m_1!\cdots m_k!}{n!}\sum_{X\in\mathcal{M}_m}\begin{bmatrix}1 & x^T\\ x & xx^T\end{bmatrix}.$$

For each $X \in \mathcal{M}_m$, $\operatorname{rank}(Y_X) = 1$. By Theorem (2.3.26), Y_X is on an extreme ray of $\mathcal{S}^n$. We need only consider the intersection of faces of $\mathcal{S}^n_+$ which contain all Y_X. To achieve this, we need to find a matrix $\hat{V}$ with range equal to the intersection of the nullspaces of D_1 and D_2.

Let $V_j \in \mathbb{R}^{j\times(j-1)}, r(V_j^T) = V_j^T u_j = 0$, e.g.,

$$V_j := \begin{bmatrix}1 & 0 & \cdots & \cdots & 0\\ 0 & 1 & \cdots & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & 1\\ -1 & \cdots & \cdots & -1 & -1\end{bmatrix}.$$

Let

$$\hat{V} := \begin{bmatrix}1 & 0\\ \frac{1}{n}m\otimes u_n & V_k\otimes V_n\end{bmatrix}.$$

Theorem 5.3.2. *[17, Theorem 3.1]*

1. *The barycenter*

$$\hat{Y} = \begin{bmatrix}1 & \frac{m_1}{n}u_n^T & \cdots & \frac{m_k}{n}u_n^T\\ \frac{m_1}{n}u_n & (\frac{m_1}{n}I_n + \frac{m_1(m_1-1)}{n(n-1)}(E_n - I_n)) & \cdots & \frac{m_1 m_k}{n(n-1)}(E_n - I_n)\\ \vdots & \vdots & \ddots & \vdots\\ \frac{m_k}{n} & \frac{m_1 m_k}{n(n-1)}(E_n - I_n) & \cdots & (\frac{m_1}{n}I_n + \frac{m_1(m_1-1)}{n(n-1)}(E_n - I_n))\end{bmatrix},$$

where E_n is the $n\times n$ matrix with entries all 1's.

2. *The rank of the barycenter*

$$\operatorname{rank}(\hat{Y}) = (k-1)(n-1)+1.$$

3. *The rows of*

$$T := \begin{bmatrix}-m_1 & u_n^T & 0 & \cdots & \cdots & 0\\ -m_2 & 0 & u_n^T & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ -m_k & 0 & \cdots & \cdots & 0 & u_n^T\\ -u_n & I_n & I_n & \cdots & \cdots & I_n\end{bmatrix}$$

form a basis of the null space of $\hat{Y}$.

4. The columns of $\hat{V}$ form a basis of the range of $\hat{Y}$.

By the above theorem, we conclude that $Y \succeq 0$ is in the minimal face if and only if $Y = \hat{V} Z \hat{V}^T$, for some $Z \succeq 0$. By substituting

$$Y = \hat{V} Z \hat{V}^T \in \mathcal{S}^{kn+1}, Z \in \mathcal{S}^{(k-1)(n-1)+1}$$

into (5.2.3), we get the reduced SDP

$$\begin{array}{rll} w^*_{\text{cut}} \geq p^*_{SDP} = & \min & \frac{1}{2}\operatorname{tr}(\hat{V}^T L_0 \hat{V} Z) \\ & \text{s.t.} & \operatorname{arrow}(\hat{V} Z \hat{V}^T) = e_0 \\ & & \mathcal{G}_J(\hat{V} Z \hat{V}^T) = 0 \\ & & (\hat{V} Z \hat{V}^T)_{00} = 1 \\ & & Z \succeq 0, Z \in \mathcal{S}^{(k-1)(n-1)+1}. \end{array} \tag{5.3.1}$$

Lemma 5.3.3. *[17, Lemma 4.1] Let Z be an arbitrary symmetric matrix of order $(n-1)(k-1)+1$ with*

$$Z = \begin{bmatrix} Z_{00} & Z_{01} & \cdots & Z_{0(k-1)} \\ Z_{10} & Z_{11} & \cdots & Z_{1(k-1)} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{(k-1)0} & Z_{(k-1)1} & \cdots & Z_{(k-1)(k-1)} \end{bmatrix},$$

where Z_{00} is a scalar, $Z_{i0} \in \mathbb{R}^{n-1}$ for $i = 1, \ldots, k-1$ and $Z_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ for $i, j = 1, \ldots, k-1$ are blocks of Z. Let $Y = \hat{V} Z \hat{V}^T$ and block Y as

$$Y = \begin{bmatrix} Y_{00} & Y_{01} & \cdots & Y_{0k} \\ Y_{10} & Y_{11} & \cdots & Y_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{k0} & Y_{k1} & \cdots & Y_{kk} \end{bmatrix},$$

where Y_{00} is a scalar, $Y_{i0} \in \mathbb{R}^{n-1}$ for $i = 1, \ldots, k$ and $Y_{ij} \in \mathbb{R}^{n\times n}$ are blocks of Y. Then

1. $Y_{00} = Z_{00}$,

   $Y_{0i} u_n = Z_{00} m_i \quad$ *for* $\quad i = 1, \ldots, k$

   and

   $\sum_{i=1}^{k} Y_{0i} = Z_{00} u_n^T$.

2. $m_i Y_{0j} = u_n^T Y_{ij} \quad$ *for* $\quad i, j = 1, \ldots, k$.

3. $\sum_{i=1}^{k} Y_{ij} = u_n Z_{0j} \quad$ *for* $\quad j = 1, \ldots, k$.

   and

   $\sum_{i=1}^{k} \operatorname{diag}(Y_{ij}) = Z_{0j} \quad$ *for* $\quad j = 1, \ldots, k$.

By Lemma (5.3.3), the arrow operator is redundant if both the gangster constraint holds and $(\hat{V}Z\hat{V}^T)_{00} = 1$.

Lemma 5.3.4. *[17, Lemma 4.2] Suppose that $W \in \mathcal{S}^{nk+1}$. Then*

$$\hat{V}^T \mathcal{G}_J(W)\hat{V} = 0 \quad \Rightarrow \quad \mathcal{G}_J(W) = 0.$$

Lemma (5.3.4) tells us there are no other redundant constraints.

Theorem 5.3.5.

$$(\text{SDP}_{final}) \qquad \begin{array}{rll} w^*_{\text{cut}} \geq p^*_{SDP} = & \min & \frac{1}{2}\operatorname{tr}\left((\hat{V}^T L_0 \hat{V})Z)\right) \\ & s.t. & \mathcal{G}_{\bar{J}}(\hat{V}Z\hat{V}^T) = \mathcal{G}_{\bar{J}}(e_0 e_0^T) \\ & & Z \succeq 0, Z \in \mathcal{S}^{(k-1)(n-1)+1}, \end{array} \tag{5.3.2}$$

where $\bar{J} := J \cup (0,0)$.

The dual problem is

$$\begin{array}{ll} \max & \frac{1}{2}W_{00} \\ s.t. & \hat{V}^T \mathcal{G}_{\bar{J}}(W)\hat{V} \preceq \hat{V}^T L_0 \hat{V}. \end{array} \tag{5.3.3}$$

Theorem 5.3.6. *[17, Theorem 4.1]*

$$\hat{Z} = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \frac{1}{n^2(n-1)}(n\operatorname{Diag}(\bar{m}_{k-1}) - \bar{m}_{k-1}\bar{m}_{k-1}^T) \otimes (nI_{n-1} - E_{n-1}) \end{array}\right] \in \mathcal{S}_+^{(k-1)(n-1)+1},$$

where $\bar{m}_{k-1} = (m_1, \ldots, m_{k-1})^T$ is a Slater point for (5.3.2).

Theorem 5.3.7. *[17, Theorem 4.2]*

$$\hat{W} = \begin{bmatrix} \alpha & 0 \\ 0 & (E_k - I_k) \otimes I_n \end{bmatrix}$$

is a Slater point for (5.3.3), if α is a sufficiently negative real scalar.

We next present two properties for recoving approximate solutions X from a solution Z of (SDP_{final}).

Proposition 5.3.8. *[15, Proposition 5.2] Suppose that Z is feasible for (SDP_{final}) and $Y = \hat{V}Z\hat{V}^T$. Let $v_1 = Y_{1:kn,0}$. Then $X_1 := \operatorname{Mat}(v_1) \in \mathcal{E} \cap \mathcal{N}$. Let $(v_0 \quad v_2^T)^T$ denote a unit eigenvector of Y corresponding to the largest eigenvalue. If $v_0 \neq 0$, then $X_2 := \operatorname{Mat}(\frac{1}{v_0}v_2) \in \mathcal{E}$. Moreover, if $Y \geq 0$, then $v_0 \neq 0$ and $X_2 \in \mathcal{N}$.*

Proof. The fact that $X_1 \in \mathcal{E}$ was shown in the proof of Lemma (5.2.1). That $X_1 \in \mathcal{N}$ follows from the arrow constraint saying that the first column of Y equals the diagonal of Y which is nonnegative since $Y \succeq 0$. We now prove the result for X_2. Suppose $v_0 \neq 0$. Then

$$Y \succeq \lambda_1(Y) \begin{pmatrix} v_0 \\ v_2 \end{pmatrix} \begin{pmatrix} v_0 & v_2^T \end{pmatrix}.$$

Using this and the definition of D_i and X_2, we see further that

$$0 = \mathrm{tr}(D_i Y) \geq \begin{cases} \lambda_1(Y) v_0^2 \left\| X_2 u_k - u_n \right\|^2, & \text{if } \ i = 1, \\ \lambda_1(Y) v_0^2 \left\| X_2^T u_n - m \right\|^2, & \text{if } \ i = 2. \end{cases} \tag{5.3.4}$$

Since $\lambda_1(Y) \neq 0$ and $v_0 \neq 0$, we must have $\|X_2 u_k - u_n\|^2 = 0$ and $\left\|X_2^T u_n - m\right\|^2 = 0$. So $X_2 \in \mathcal{E}$.

Finally, we suppose $Y \geq 0$. We claim that for any eigenvector $(v_0 \;\; v_2^T)^T$ corresponding to the largest eigenvalue must satisfy:

1. $v_0 \neq 0$.
2. all entries if the eigenvector have the same sign, i.e., $v_0 v_2 \geq 0$.

From this claim, we have $X_2 = \mathrm{Mat}(\frac{1}{v_0} v_2) \in \mathcal{N}$.

To prove the cliam, we note from the classical Perron-Fröbenius theory, e.g. [4], that the vector $(|v_0| \;\; |v_2|^T)^T$ is also an eigenvector to the largest eigenvalue. Let $\chi := \mathrm{Mat}(v_2)$. We do the same procedure as in (5.3.4), we conclude that

$$\|\chi u_m - v_0 u_n\|^2 = 0 \quad \text{and} \quad \| |\chi| \, u_m - |v_0| \, u_n \|^2 = 0. \tag{5.3.5}$$

Suppose by contradiction that $v_0 = 0$. Then the second equality implies $\chi = 0$. Then $v_2 = \mathrm{vec}(\chi) = 0$. It is a contradiction since eigenvectors cannot be 0. So we conclude that $v_0 \neq 0$.

Now suppose $v_0 > 0$. Then the two equalities give us

$$\sum_{j=1}^{k} \chi_{ij} = v_0 = \sum_{j=1}^{k} |\chi_{ij}| \, ,$$

for all $i = 1, \ldots, n$. So we have $\chi_{ij} \geq 0$ for all i, j, which implies $v_2 \geq 0$. One can show similarly for the case $v_0 < 0$. Hence we proved $v_0 v_2 \geq 0$. $\square$

Chapter 6

Cut Minimization Problem

6.1 Introduction

There is another type of GP Problem. We partition the node set of a graph into k sets of given sizes. The goal is to minimize the size of cut edges obtained by removing the k-th set. This problem is called *Cut Minimization, CM Problem*, which is contained as a special case of GP Problem because the two minimization problems share the same constraints. Most of the contents in this chapter is taken from [15].

We let $\delta(S_i, S_j)$ denote the set of cut edges between the sets S_i and S_j, i.e.,

$$\delta(S_i, S_j) = \{uv \in E(G) : u \in S_i, v \in S_j\}.$$

We denote the set of edges with endpoints in distinct partition sets $S_1, \ldots, S_{k-1}$ by

$$\delta(S) = \cup_{i<j<k} \delta(S_i, S_j). \tag{6.1.1}$$

The minimum of the cardinality $|\delta(S)|$ is denoted

$$\operatorname{cut}(m) = \min \{|\delta(S)| : S \in \mathcal{P}_m\}. \tag{6.1.2}$$

The graph G has a *vertex separator* if there exists an $S \in \mathcal{P}_m$ such that $\delta(S) = \emptyset$, i.e., $\operatorname{cut}(m) = 0$. Otherwise, $\operatorname{cut}(m) > 0$. We call the later problem the *Vertex Separator Problem*.

We define the k ordered matrix

$$B := \begin{bmatrix} uu^T - I_{k-1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{S}^k.$$

Proposition 6.1.1. *[15, Proposition 2.3] For $S \in \mathcal{P}_m$ let $X \in \mathcal{M}_m$ be the associated partition matrix. Then*

$$|\delta(S)| = \frac{1}{2} \operatorname{tr}\left((A - \operatorname{Diag}(d))XBX^T\right), \forall d \in \mathbb{R}^n. \tag{6.1.3}$$

Proof. We include the proof for completeness. Let $X \in \mathcal{M}_m$ be the partition matrix associated to $S \in \mathcal{P}_m$. Write $X = \begin{pmatrix} x_1 & x_2 & \cdots & x_k \end{pmatrix}$, where x_i is the i-th column of X. Then $XBX^T = \sum_{0<i<j<k} x_i x_j^T + x_j x_i^T$. So XBX^T is the adjacency matrix of the complete $(k-1)$-partite graph with the parttition $S \in \mathcal{P}_m$. In particular,

$$(XBX^T)_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the distinct sets } S_1, \ldots, S_{k-1} \ , \\ 0 & \text{otherwise.} \end{cases}$$

So

$$A_{ij}(XBX^T)_{ij} = \begin{cases} 1 & \text{if } ij \in \delta(S), \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $\operatorname{tr}(AXBX^T) = \sum_{ij} A_{ij}(XBX^T)_{ij} = 2\,|\delta(S)|$.

Since $\operatorname{diag}(XBX^T) = 0$ and $A_{ij} = (A - \operatorname{Diag}(d))_{ij}$ for $i \neq j$, we have $\operatorname{tr}(AXBX^T) = \operatorname{tr}\left((A - \operatorname{Diag}(d))XBX^T\right), \forall d \in \mathbb{R}^n$. □

Remark 6.1.2. *$\forall S \in \mathcal{P}_m, |\delta(S)| = \operatorname{tr}(AXBX^T)$ if $d = 0$, while $|\delta(S)| = \operatorname{tr}(-LXBX^T)$ if $d = Au_n$. So the above proposition is a general version of [13, Prop.2].*

So the Vertex Separator Problem can be solved as

$$\begin{array}{ll} \min & \frac{1}{2} \operatorname{tr}\left(G(d)XBX^T\right) \\ \text{s.t.} & X \in \mathcal{M}_m, \end{array} \tag{6.1.4}$$

where $G(d) = A - \operatorname{Diag}(d), d \in \mathbb{R}^n$.

The format of objective function in (6.1.4) is same as that of (3.0.2), i.e., they are both quadratic. So we can apply the same strategies to derive the eigenvalue bounds, convex quadratic programming bounds, and semidefinite programming bounds for the Vertex Seperator Problem. We are going to go over some main details in the following sections.

6.2 Lower Bounds for the CM Problem

We briefly conclude the three types of lower bound for CM problem. Numerical tests for CM problem can be found in [15].

6.2.1 Eigenvalue Bounds

The same idea in Chapter 2 can be applied to derive the basic eigenvalue bound for the CM problem. Consider the relaxed problem

$$\operatorname{cut}(m) \geq \begin{array}{ll} \min & \frac{1}{2} \operatorname{tr}(GXBX^T) \\ \text{s.t.} & X \in \mathcal{D}_O. \end{array} \tag{6.2.1}$$

Lemma 6.2.1. *[13, Lemma 4] The $k-$ordered eigenvalues of the matrix $\tilde{B} := \tilde{M}B\tilde{M}$ satisfy*

$$\lambda_1(\tilde{B}) > 0 = \lambda_2(\tilde{B}) > \lambda_3(\tilde{B}) \geq \cdots \geq \lambda_k(\tilde{B}).$$

Proof. We include the proof for completeness. The matrix $u_{k-1}u_{k-1}^T$ has rank one and $\mathrm{tr}(uu^T) = k-1$ so it has eigenvalues $(k-1, 0, \cdots, 0)^T \in \mathbb{R}^{k-1}$. So the matrix $uu^T - I_{k-1}$ has eigenvalues $(k-2, -1, \cdots, -1)^T \in \mathbb{R}^{k-1}$. Because B has a row of 0, then 0 is an eigenvalue of B. So B has eigenvalues $(k-2), 0, -1, \cdots, -1)^T \in \mathbb{R}^k$.

The conclusion for $\tilde{B}$ follows from the Sylvester Law of Inertia for nonsingular congruences. □

Remark 6.2.2. *The Sylvester Law of Inertia states that if $H \in \mathcal{S}^p$ and $Q \in \mathbb{R}^{p\times p}$ is nonsingular, then H and QHQ^T have the same number of positive, negative and zero eigenvalues (same inertia). QHQ^T is said to be congruent to H.*

Theorem 6.2.3. *[15, Theorem 3.4] Let $d \in \mathbb{R}^n, G = A - \mathrm{Diag}(d)$. Then*

$$\mathrm{cut}(m) \geq 0 > q^*_{eig}(G) := \frac{1}{2}\left\langle \lambda(G), \begin{pmatrix} \lambda(\tilde{B}) \\ 0 \end{pmatrix} \right\rangle_- = \frac{1}{2}\left(\sum_{i=1}^{k-2} \lambda_{k-i+1}(\tilde{B})\lambda_i(G) + \lambda_1(\tilde{B})\lambda_n(G) \right).$$

Moreover, the function $q^*_{eig}(G(d))$ is a concave function of $d \in \mathbb{R}^n$.

Proof. Let $Y = X\tilde{M}^{-1}$. We have $X^TX = M$ if and only if $Y^TY = I_k$. We substitude $X = Y\tilde{M}$ into (6.2.1) to get the equivalent problem to (6.2.1):

$$\begin{array}{ll} \min & \frac{1}{2}\mathrm{tr}(GY\tilde{B}Y^T) \\ \text{s.t.} & Y^TY = I_k, \end{array} \tag{6.2.2}$$

where $\tilde{B} = \tilde{M}B\tilde{M}$.

Since (6.2.2) is a relaxation to the CM Problem, we have $\mathrm{cut}(m) \geq q^*_{eig}(G)$. The explicit formula for the minimal scalar product follows immediately from Lemma (6.2.1).

We are going to show $0 > q^*_{eig}(G)$. Let $\hat{\phi}$ be a permutation of $N = \{1, 2, \ldots, n\}$ that attains the minimal value $\min\left\{\sum_{i=1}^k \lambda_{\phi(i)}(G)\lambda_i(\tilde{B}) : \phi \text{ is a permutation}\right\}$. Then for any permutation ψ, we have

$$\sum_{i=1}^k \lambda_{\psi(i)}(G)\lambda_i(\tilde{B}) \geq \sum_{i=1}^k \lambda_{\hat{\phi}(i)}(G)\lambda_i(\tilde{B}). \tag{6.2.3}$$

Note that $\sum_{i=1}^k \lambda_i(\tilde{B}) = \mathrm{tr}(\tilde{B}) = \mathrm{tr}(\tilde{M}B\tilde{M}) = \mathrm{tr}(MB) = 0$, since $\mathrm{diag}(B) = 0$ and M is a diagonal matrix.

Let $\mathcal{T}$ be the set of all permutations of N, then we have

$$\sum_{\psi\in\mathcal{T}}\left(\sum_{i=1}^k \lambda_{\psi(i)}(G)\lambda_i(\tilde{B})\right) = \sum_{i=1}^k\left(\sum_{\psi\in\mathcal{T}}\lambda_{\psi(i)}(G)\right)\lambda_i(\tilde{B}) = \left(\sum_{\psi\in\mathcal{T}}\lambda_{\psi(1)}(G)\right)\left(\sum_{i=1}^k \lambda_i(\tilde{B})\right) = 0, \tag{6.2.4}$$

since $\sum_{\psi\in\mathcal{T}}\lambda_{\psi(i)}(G)$ is independent of i. It implies that $\sum_{i=1}^k \lambda_{\hat{\psi}(i)}(G)\lambda_i(\tilde{B}) \leq 0$.

Now we prove by contradiction. Suppose $\sum_{i=1}^{k} \lambda_{\hat{\psi}(i)}(G)\lambda_i(\tilde{B}) = 0$ which implies that $\sum_{i=1}^{k} \lambda_{\psi(i)}(G)\lambda_i(\tilde{B}) = 0, \forall \psi \in \mathcal{T}$. Recall from the Lemma (6.2.1) that $\lambda_1(\tilde{B}) > \lambda_k(\tilde{B})$. It implies that all eigenvalues of G are equal. Moreover, if all eigenvalues of G were equal, then necessarily $G = \beta I_n$ for some $\beta \in \mathbb{R}$ and A must be diagonal matrix. This implies that A=0 since $\operatorname{diag}(A) = 0$, which is a contradiction. Therefore $q^*_{eig}(G) < 0$.

Finally, the concavity follows by observing from (6.2.2) that

$$q^*_{eig}(G(d)) = \min_{Y^TY=I_k} \frac{1}{2} \operatorname{tr}\left(G(d)Y\tilde{B}Y^T\right),$$

is a function obtained as a minimum of a set of affine functions in d, and recall that the minimum of affine functions is concave. □

Since $q^*_{eig}(G(d)) \le 0, \forall d \in \mathbb{R}^n$, the basic eigenvalue bound for the CM problem is not a useful bound. Next theorem provides the projective eigenvalue bounds.

Theorem 6.2.4. *[15, Theorem 3.7] Let $d \in \mathbb{R}^n, G = A - \operatorname{Diag}(d)$. Let V and W be defined in (3.2.2) and $\widehat{X} = \frac{1}{n}u_n m^T \in \mathbb{R}^{n\times k}$. Then:*

1. *For any $X \in \mathcal{E}$ and $Z \in \mathbb{R}^{(n-1)\times(k-1)}$ related by (3.2.3), we have*

$$\begin{aligned} \operatorname{tr}(GXBX^T) &= \alpha + \operatorname{tr}(\hat{G}Z\hat{B}Z^T) + \operatorname{tr}(FZ^T) \\ &= -\alpha + \operatorname{tr}(\hat{G}Z\hat{B}Z^T) + 2\operatorname{tr}(G\widehat{X}BX^T), \end{aligned} \tag{6.2.5}$$

and

$$\operatorname{tr}(-LXBX^T) = \operatorname{tr}(-\hat{L}Z\hat{B}Z^T), \tag{6.2.6}$$

where

$$\hat{G} = V^TGV, \hat{L} = V^TLV, \hat{B} = W^T\tilde{M}B\tilde{M}W, \alpha = \frac{1}{n^2}(u^TGu)(m^TBm), F = 2V^TG\widehat{X}B\tilde{M}W. \tag{6.2.7}$$

2. *We have the following lower bounds:*

(a)

$$\begin{aligned} \operatorname{cut}(m) &\ge q^*_{peig}(G) := \tfrac{1}{2}\left\{-\alpha + \left\langle \lambda(\hat{G}), \begin{pmatrix}\lambda(\hat{B})\\ 0\end{pmatrix}\right\rangle_- + 2\min_{X\in\mathcal{D}} \operatorname{tr}(G\widehat{X}BX^T)\right\} \\ &= \tfrac{1}{2}\left\{\alpha + \left\langle \lambda(\hat{G}), \begin{pmatrix}\lambda(\hat{B})\\ 0\end{pmatrix}\right\rangle_- + \min_{0\le \widehat{X}+VZW^T\tilde{M}} \operatorname{tr}(FZ^T)\right\} \\ &= \tfrac{1}{2}\left\{-\alpha + \textstyle\sum_{i=1}^{k-2} \lambda_{k-i}(\hat{B})\lambda_i(\hat{G}) + \lambda_1(\hat{B})\lambda_{n-1}(\hat{G}) + 2\min_{X\in\mathcal{D}} \operatorname{tr}(G\widehat{X}BX^T)\right\}. \end{aligned} \tag{6.2.8}$$

(b)

$$\operatorname{cut}(m) \ge q^*_{peig}(-L) := \frac{1}{2}\left\langle \lambda(-\hat{L}), \begin{pmatrix}\lambda(\hat{B})\\ 0\end{pmatrix}\right\rangle_- \ge q^*_{eig}(-L). \tag{6.2.9}$$

Proof. We substitute the parametrization (3.2.3) into our objective function $\mathrm{tr}(GXBX^T)$. Then we get a constant, quadratic, and linear term:

$$\begin{aligned}\mathrm{tr}(GXBX^T) &= \mathrm{tr}\left(G(\widehat{X}+VZW^T\tilde{M})B(\widehat{X}+VZW^T\hat{M})^T\right)\\ &= \mathrm{tr}(G\widehat{X}B\widehat{X}^T)+\mathrm{tr}\left((V^TGV)Z(W^T\tilde{M}B\tilde{M}W)Z^T\right)+\mathrm{tr}(2V^TG\widehat{X}B\tilde{M}WZ^T)\end{aligned}$$

and

$$\begin{aligned}\mathrm{tr}(GXBX^T) &= \mathrm{tr}(G\widehat{X}B\widehat{X}^T)+\mathrm{tr}\left((V^TGV)Z(W^T\tilde{M}B\tilde{M}W)Z^T\right)+2\,\mathrm{tr}\left(G\widehat{X}B(VZW^T\tilde{M})^T\right)\\ &= \mathrm{tr}(G\widehat{X}B\widehat{X}^T)+\mathrm{tr}\left((V^TGV)Z(W^T\tilde{M}B\tilde{M}W)Z^T\right)+2\,\mathrm{tr}\left(G\widehat{X}B(X-\hat{X})^T\right)\\ &= -\,\mathrm{tr}(G\widehat{X}B\widehat{X}^T)+\mathrm{tr}\left((V^TGV)Z(W^T\tilde{M}B\tilde{M}W)Z^T\right)+2\,\mathrm{tr}(G\widehat{X}BX^T).\end{aligned}$$

These together with (6.2.7) yield the two equations in (6.2.5). Since $Lu=0$ and hence $L\widehat{X}=0$, we obtain (6.2.6) by replacing G with $-L$. We proved Item 1.

Now we are going to prove (6.2.8), i.e., Item (2a). Recall from (6.1.4) and (2.1.1) that

$$\mathrm{cut}(m)=\min\left\{\frac{1}{2}\,\mathrm{tr}(GXBX^T):X\in\mathcal{D}\cap\mathcal{D}_O\right\}.$$

Combining this with (6.2.5), we see further that

$$\begin{aligned}\mathrm{cut}(m) &= \tfrac{1}{2}\left(-\alpha+\min_{X\in\mathcal{D}\cap\mathcal{D}_O}\left\{\mathrm{tr}(\hat{G}Z\hat{B}Z^T)+2\,\mathrm{tr}(G\widehat{X}BX^T)\right\}\right)\\ &\geq \tfrac{1}{2}\left(-\alpha+\min_{X\in\mathcal{D}\cap\mathcal{D}_O}\mathrm{tr}(\hat{G}Z\hat{B}Z^T)+2\min_{X\in\mathcal{D}\cap\mathcal{D}_O}\mathrm{tr}(G\widehat{X}BX^T)\right)\\ &\geq \tfrac{1}{2}\left(-\alpha+\min_{Z^TZ=I_{k-1}}\mathrm{tr}(\hat{G}Z\hat{B}Z^T)+2\min_{X\in\mathcal{D}}\mathrm{tr}(G\widehat{X}BX^T)\right)\\ &= \tfrac{1}{2}\left(-\alpha+\left\langle\lambda(\hat{G}),\begin{pmatrix}\lambda(\hat{B})\\0\end{pmatrix}\right\rangle_-+2\min_{X\in\mathcal{D}}\mathrm{tr}(G\widehat{X}BX^T)\right)=q^*_{peig}(G),\end{aligned}\qquad(6.2.10)$$

where X and Z are related via (3.2.4), and the last equality follows from Lemma (3.2.1) and Theorem (3.1.2).

Furthermore, notice that

$$\begin{aligned}-\alpha+2\min_{X\in\mathcal{D}}\mathrm{tr}(G\widehat{X}BX^T) &= \alpha+2\min_{X\in\mathcal{D}}\mathrm{tr}\left(G\widehat{X}B(X-\widehat{X})^T\right)\\ &= \alpha+2\min_{0\leq\widehat{X}+VZW^T\tilde{M}}\mathrm{tr}\left(G\widehat{X}B(VZW^T\tilde{M})^T\right)\\ &= \alpha+\min_{0\leq\widehat{X}+VZW^T\tilde{M}}\mathrm{tr}(FZ^T),\end{aligned}\qquad(6.2.11)$$

where the second equality follows from Lemma (3.2.1), and the last equality follows from the definition of F in (6.2.7). Combining this last relation with (6.2.10) proves the first two equalities in (6.2.8).

The last equality in (6.2.8) follows from the fact that

$$\lambda_k(\tilde{B}) \le \lambda_{k-1}(\hat{B}) \le \lambda_{k-1}(\tilde{B}) \le \cdots \le \lambda_2(\tilde{B}) = 0 \le \lambda_1(\hat{B}) \le \lambda_1(\tilde{B}), \qquad (6.2.12)$$

which is a consequence of the eigenvalue interlacing theorem, the definition of $\hat{B}$ and Lemma(6.2.1).

Next, we prove (6.2.9), i.e., Item 2b. Recall from (6.1.4) and (2.1.1) that

$$\operatorname{cut}(m) = \min\left\{\frac{1}{2}\operatorname{tr}(-LXBX^T) : X \in \mathcal{D} \cap \mathcal{D}_O\right\}.$$

Using (6.2.6), we see further that

$$\begin{array}{rl} \operatorname{cut}(m) & \ge \min\left\{\frac{1}{2}\operatorname{tr}(-LXBX^T) : X \in \mathcal{E} \cap \mathcal{D}_O\right\} \\ & = \min\left\{\frac{1}{2}\operatorname{tr}(-\hat{L}Z\hat{B}Z^T) : Z^TZ = I_{k-1}\right\} \\ & = \frac{1}{2}\left\langle \lambda(-\hat{L}), \begin{pmatrix}\lambda(\hat{B})\\ 0\end{pmatrix}\right\rangle_- \ (= q^*_{peig}(-L)) \\ & \ge \min\left\{\frac{1}{2}\operatorname{tr}(-LXBX^T) : X \in \mathcal{D}_O\right\} (= q^*_{eig}(-L)), \end{array}$$

where X and Z are related via (3.2.4). The last inequality follows from dropping the constraint $X \in \mathcal{E}$. □

6.2.2 Convex Quadratic Programming Bounds

We follow the approach we used in chapter 4. Let (S^{**}, T^{**}) be an optimal solution to the following problem

$$\begin{array}{rl} \max & \frac{1}{2}\operatorname{tr}(S) + \frac{1}{2}\operatorname{tr}(T) \\ \text{s.t.} & I_{k-1} \otimes S + T \otimes I_{n-1} \preceq \hat{B} \otimes \hat{G}, \\ & S \preceq 0, \\ & S \in \mathcal{S}^{n-1}, T \in \mathcal{S}^{k-1}, \end{array}$$

where $\hat{B}$ and $\hat{G}$ are defined in (6.2.7).

We define

$$\widehat{\widehat{Q}} := \hat{B} \otimes \hat{G} - I_{k-1} \otimes S^{**} - T^{**} \otimes I_{n-1} \succeq 0.$$

Then,

$$\begin{array}{rl} \operatorname{cut}(m) \ge q^*_{QP}(G) := \min & \frac{1}{2}\left(\alpha + \operatorname{tr}(FZ^T) + \operatorname{vec}(Z)^T\widehat{\widehat{Q}}\operatorname{vec}(Z) + \left\langle \lambda(\hat{G}), \begin{pmatrix}\lambda(\hat{B})\\ 0\end{pmatrix}\right\rangle_-\right) \\ \text{s.t.} & \widehat{X} + VZW^T\tilde{M} \ge 0, \end{array} \qquad (6.2.13)$$

where F is define in (6.2.7).

Notice that (6.2.13) is a convex QP with $(n-1)(k-1)$ variables and nk constraints.

Theorem 6.2.5. *[15, Theorem 4.1]*

$$\operatorname{cut}(m) \geq q^*_{QP}(G) = \min_{X\in\mathcal{D}} \frac{1}{2}\operatorname{vec}(X)^T \tilde{\tilde{Q}} \operatorname{vec}(X) + \frac{1}{2}\left\langle \lambda(\hat{G}), \begin{pmatrix} \lambda(\hat{B}) \\ 0 \end{pmatrix} \right\rangle_- \tag{6.2.14}$$

where

$$\tilde{\tilde{Q}} := B \otimes G - M^{-1} \otimes (VS^{**}V^T) - (\tilde{M}^{-1}WT^{**}W^T\tilde{M}^{-1}) \otimes I_n.$$

The QP in (6.2.14) *is implicitly convex.*

6.2.3 SDP Bounds

Like what we did in Chapter 5, we can obtain a lower bound for $\operatorname{cut}(m)$ through the semidefinite programming. Let

$$L_G := \begin{bmatrix} 0 & 0 \\ 0 & B \otimes G \end{bmatrix}.$$

So we have the semidefinite programming bounds for $\operatorname{cut}(m)$ as following.

Theorem 6.2.6. *[15, Theorem 5.1]*

$$\begin{array}{rclll} \operatorname{cut}(m) \geq p^*_{VSDP}(G) & := & \min & \frac{1}{2}\operatorname{tr}(\widehat{V}^T L_G \widehat{V} Z) & \\ & & s.t. & \mathcal{G}_{\bar{J}}(\widehat{V} Z \widehat{V}^T) = \mathcal{G}_{\bar{J}}(e_0 e_0^T), & (\text{VSDP}_{final}) \\ & & & Z \succeq 0, Z \in \mathcal{S}^{(k-1)(n-1)+1}. & \end{array}$$

The dual problem is

$$\begin{array}{ll} \max & \frac{1}{2} W_{00} \\ s.t. & \widehat{V}^T \mathcal{G}_{\bar{J}}(W) \widehat{V} \preceq \widehat{V}^T L_G \widehat{V}. \end{array}$$

Both primal and dual satisfy the Slater constraint qualification and the objective function is independent of the $d \in \mathbb{R}^n$ chosen to form G.

Proof. The only thing we need to prove is the independence of choice of d. Let $Y = \widehat{V} Z \widehat{V}^T$ with Z feasible for (VSDP_{final}). Then Y satisfies the gangster constriants, i.e., $\operatorname{diag}(\bar{Y}_{(ij)}) = 0, \forall i \neq j$. On the other hand, notice that $\operatorname{tr}(L_G Y) = \operatorname{tr}(L_A Y) - \operatorname{tr}(L_{\operatorname{Diag}(d)} Y)$. From the structure of $B \otimes \operatorname{Diag}(d)$, $L_{\operatorname{Diag}(d)}$ has nonzero elements only in the diagonal positions of the off-diagonal blocks. So we have $\operatorname{tr}(L_{\operatorname{Diag}(d)} Y) = 0$. As a result,

$$\operatorname{tr}\left((\widehat{V}^T L_G \widehat{V}) Z\right) = \operatorname{tr}(L_G \widehat{V} Z \widehat{V}^T) = \operatorname{tr}(L_G Y) = \operatorname{tr}(L_A Y) = \operatorname{tr}(\widehat{V}^T L_A \widehat{V} Z),$$

for all $d \in \mathbb{R}^n$. □

Chapter 7

Numerical Tests

In this chapter, we provide some empirical comparisons for the lower and upper bounds obtained from above methods. All the numerical tests are performed in MATLAB version 2013b on a single node of the *COPS* cluster at University of Waterloo.

7.1 Feasible Solutions Upper Bounds

As an extension of the well-known Birkhoff-von Neumann Theorem relating the extreme points of the doubly stochastic matrices to the permutation matrices, we have the following. (We include a proof for completeness.)

Theorem 7.1.1. *[15, Theorem 6.1] The set of extreme points of the doubly stochastic type matrices $\mathcal{D}$ equals the set of partition matrices $\mathcal{M}_m$, i.e.,*

$$\operatorname{ext}(\mathcal{D}) = \mathcal{M}_m.$$

Proof. It is clear that $\mathcal{M}_m \subseteq \operatorname{ext}(\mathcal{D})$. Next we prove $\operatorname{ext}(\mathcal{D}) \subseteq \mathcal{M}_m$ by showing that all entries of elements in $\operatorname{ext}(\mathcal{D})$ are integral. Let $\tilde{X} \in \mathcal{D}$ have a non-integral entry. We are going to show that $\widetilde{X}$ is a nontrivial convex combination of partition matrices, and hence is not extremal. We define

$$X_0 = \begin{bmatrix} u_{m_1} & 0 & \cdots & 0 \\ 0 & u_{m_2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{m_k} \end{bmatrix} \in \mathcal{M}_m.$$

Consider the linear map $h : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times k}$ with $h(D) = DX_0$. Then for any doubly stochastic matrix D of order n, we have $h(D) \in \mathcal{D}$. Moreover, for any permutation matrix P of order n, we have $h(P) \in \mathcal{M}_m$. Next we define

$$\bar{X} := [\underbrace{\frac{1}{m_1}\widetilde{X}_{:1} \quad \cdots \quad \frac{1}{m_1}\widetilde{X}_{:1}}_{m_1} \quad \cdots \quad \underbrace{\frac{1}{m_k}\widetilde{X}_{:k} \quad \cdots \quad \frac{1}{m_k}\widetilde{X}_{:k}}_{m_k}] \in \mathbb{R}^{n\times n}.$$

It is easy to verify $\bar{X}$ is doubly stochastic and $h(\bar{X}) = \widetilde{X}$. Since $\bar{X}$ is doubly stochastic matrix, by the Birkhoff-von Neumann Theorem, there exist $\lambda_i > 0$ and permutation matrices P_i, $i = 1, \dots, l$, such that $\bar{X} = \sum_{i=1}^{l} \lambda_i P_i$ and $\sum_{i=1}^{l} \lambda_i = 1$. Applying the linear map h on both sides gives

$$\widetilde{X} = h(\bar{X}) = h\left(\sum_{i=1}^{l} \lambda_i P_i\right) = \sum_{i=1}^{l} \lambda_i h(P_i).$$

Since there is a entry of $\widetilde{X}$ that is fractional, there is at least one λ_i that is fractional. Consequently, since $\sum_{i=1}^{l} \lambda_i = 1$, there are two $\lambda_i' s$ that are fractional. So $\tilde{X}$ is a nontrivial convex combination of partition matrices. □

Theorem 7.1.2. *[15, Theorem 6.2] Let $\bar{X} \in \mathbb{R}^{n \times k}$. Then an optimal solution of $\min_{X \in \mathcal{M}_m} \left\|X - \bar{X}\right\|_F$ can be found by using simplex method to solve the LP problem*

$$\begin{array}{ll} \min & -\operatorname{tr}(\bar{X}^T X) \\ s.t. & X u_k = u_n, \\ & X^T u_n = m, \\ & X \geq 0. \end{array} \tag{7.1.1}$$

Proof. If $X \in \mathcal{M}_m$, then $\operatorname{Diag}(X^T X) = m$. So $\operatorname{tr}(X^T X) = n$. Hence we have

$$\begin{aligned} \min_{X \in \mathcal{M}_m} \left\|X - \bar{X}\right\|_F^2 &= \min_{X \in \mathcal{M}_m} \left\langle \bar{X} - X, \bar{X} - X \right\rangle \\ &= \operatorname{tr}(\bar{X}^T \bar{X}) + \min_{X \in \mathcal{M}_m} \operatorname{tr}(X^T X - 2\bar{X}^T X) \\ &= \operatorname{tr}(\bar{X}^T \bar{X}) + n + 2 \min_{X \in \mathcal{M}_m} \operatorname{tr}(-\bar{X}^T X). \end{aligned}$$

The simplex algorithm can give us an extreme point of $\mathcal{D}$ which is optimal. So this optimal solution is in $\mathcal{M}_m$ by the theorem (7.1.1). □

7.2 Random Tests with Various Size

We first fix a positive integer $k \geq 4$ and generate k integers $m_1, \dots, m_k$ each chosen randomly from $\{1, \dots, \mathsf{imax}\}$. If any of m_i happened to be 1, then we increase all the m_i by 1. Next we construct our graphs in two ways:

1. Structured graphs: We construct k disjoint cliques. The i-th clique has m_i nodes. Then we add u_0 edges between the k cliques, chosen uniformly at random from the complement graph. In our tests, we set $u_0 = \lfloor e_c p \rfloor$, where e_c is the number of edges in the complement graph and $0 \leq p < 1$. By our construction, u_0 is very likely to be the optimal value, i.e., $u_0 = w^*_{\text{cut}}$.

2. Random graphs: We generate a graph with $n = u_k^T m$ nodes. The adjacency matrix is generated by

$$\mathsf{A = round(rand(n)); A = round((A + A')/2); A = A - diag(diag(A));}$$

As a consequence, an edge is chosen with probability 0.75.

we define the relative gap (Rel. gap) as

$$\text{Rel. gap} = \frac{\text{best upper bound} - \text{best lower bound}}{\text{best upper bound} + \text{best lower bound}}.$$

In Tables 7.1 and 7.2, we consider small instances where $k = 4, 5$, $p = 20\%$ and $\mathsf{imax} = 10$. The tables include BE_L with value $p^*_{eig}(0)$, BE_A with value $p^*_{eig}(-Au)$, PE_L with value $p^*_{peig}(0)$, PE_A with value $p^*_{peig}(-Au)$, QP with value $p^*_{QP}(-Au)$, SDP bounds and the doubly nonnegative programming (DNN) bounds[1]. For each approach, we present the lower bounds (rounded up to the nearest interger) in the first line and the corresponding upper bounds (rounded down to the nearest integer) obtained via the linear programming technique described in Section (7.1)[2] in the second line.

In terms of lower bounds, the DNN approach usually gives the best lower bound. While the SDP bounds are better than the QP bounds for random graphs, they are comparable for structured graphs.

Data				$\frac{\text{Lower}}{\text{Upper}}$ bounds							Rel. gap
n	k	$\lvert E \rvert$	u_0	BE_L	BE_A	PE_L	PE_A	QP	SDP	DNN	
21	4	93	29	16	23	20	26	27	28	29	0.0000
				51	57	56	39	29	36	29	
27	4	139	52	33	45	34	46	47	46	52	0.0877
				77	83	68	79	62	62	62	
25	5	115	46	28	36	33	40	42	40	46	0.0000
				66	71	71	55	57	55	46	
31	5	173	73	43	61	48	65	66	65	73	0.0000
				117	119	117	80	73	92	73	

Table 7.1: Results for small structured graphs

We consider medium-sized instances in Tables 7.3 and 7.4, where $k = 8, 10, 12, p = 20\%$ and $\mathsf{imax} = 20$. We do not consider DNN bounds due to computational complexity. It is very interesting that QP bounds are better than SDP bounds in the medium-sized structured instances while SDP bounds are better than QP bounds in the medium-sized random instances. BE bounds and PE bounds are comparable.

Finally, in Tables 7.5 and 7.6, we consider larger instances with $k = 35, 45, 55, p = 20\%$ and $\mathsf{imax} = 100$. We do not consider QP, SDP and DNN bounds due to computational complexity.

[1] The doubly nonnegative programming relaxation is obtained by imposing the constraint $\widehat{V} Z \widehat{V}^T \geq 0$ onto (SDP $_{final}$).

[2] The SDP and DNN problems are solved via SDPT3 (version 4.0) with tolerence **gaptol** set to be 1e-6 and 1e-3 respectively. The problem (4.1.5) and (4.2.2) are solved via SDPT3 (version 4.0) called by CVX (version 1.22), using the default settings. The problem (7.1.1) is solved using simplex method in MATLAB using the default settings.

Data			$\frac{\text{Lower}}{\text{Upper}}$ bounds							Rel. gap
n	k	$\lvert\mathsf{E}\rvert$	BE_L	BE_A	PE_L	PE_A	QP	SDP	DNN	
16	4	93	36	40	46	50	52	52	58	0.0085
			60	60	60	59	59	62	59	
27	4	271	149	165	156	172	173	176	188	0.0105
			196	195	195	192	195	199	208	
31	5	360	196	224	212	238	241	244	262	0.0113
			268	270	275	269	274	275	286	
35	5	446	242	276	261	294	299	302	324	0.0167
			344	335	340	337	338	347	347	

Table 7.2: Results for small random graphs

In all tables, we have $\text{PE}_A \geq \text{PE}_L \geq \text{BE}_L$ and $\text{PE}_A \geq \text{BE}_A \geq \text{BE}_L$, while PE_L and BE_A are comparable.

Before ending this section, we briefly talk about the computational time measured by MATLAB tic-toc function. For lower bounds, the eigenvalue bounds are fastest to compute. The computational time for small, medium and large problems are usually less than 0.01 seconds, 0.1 seconds and 0.5 seconds, respectively. The QP bounds are more expensive to compute, taking around 0.5 to 2 seconds for small problems and 0.5 to 15 minutes for medium problems. The SDP bounds are even more expensive to compute, taking 0.5 to 3 seconds for small problems and 2 minutes to 2 hours for medium problems. The DNN bounds are the most expensive to compute. Even for small problems, it can take 20 seconds to 40 minutes to compute a bound. For upper bounds, using the MATLAB simplex method, the time for solving (7.1.1) is usually less than 1 second for small and medium problems; while for the large problems in Table (7.5) and Table (7.6), it takes 1 to 5 minutes.

Data				$\frac{\text{Lower}}{\text{Upper}}$ bounds						Rel. gap
n	k	$\lvert \mathsf{E} \rvert$	u_0	BE_L	BE_A	PE_L	PE_A	QP	SDP	
58	8	501	287	202	235	220	251	256	232	0.1065
				380	382	364	357	317	394	
62	8	594	324	225	275	248	294	300	291	0.1071
				422	438	442	426	372	394	
95	8	1481	745	583	665	619	691	701	678	0.0377
				1045	972	1047	988	756	1079	
117	10	1979	1201	994	1084	1033	1115	1123	1097	0.0369
				1565	1474	1464	1521	1209	1472	
94	10	1387	745	590	652	633	682	692	675	0.0559
				992	1003	971	774	878	911	
123	10	2147	1338	1061	1190	1120	1241	1253	1197	0.0504
				1645	1641	1662	1597	1386	1757	
132	12	2346	1575	1259	1408	1316	1460	1469	1438	0.0498
				1952	1876	1940	1623	1717	1983	
132	12	2368	1569	1242	1402	1285	1438	1449	1380	0.1003
				1907	1851	1966	1772	1790	2012	
115	12	1845	1177	912	1025	972	1073	1085	1047	0.0683
				1513	1440	1501	1244	1326	1477	

Table 7.3: Results for medium-sized structured graphs

Data			$\frac{\text{Lower}}{\text{Upper}}$ bounds						Rel. gap
n	k	$\|\mathsf{E}\|$	BE_L	BE_A	PE_L	PE_A	QP	SDP	
65	8	1571	980	1055	1089	1162	1170	1175	0.0435
			1292	1290	1297	1282	1299	1316	
67	8	1681	1050	1126	1182	1261	1270	1272	0.0450
			1407	1392	1408	1396	1398	1401	
73	8	1987	1245	1323	1420	1497	1511	1516	0.0408
			1656	1662	1663	1645	1653	1669	
92	10	3130	2138	2289	2322	2474	2486	2492	0.0384
			2716	2708	2720	2691	2708	2724	
122	10	5457	3898	4156	4089	4343	4358	4362	0.0377
			4732	4704	4735	4711	4718	4783	
108	10	4362	3121	3250	3323	3444	3457	3467	0.0355
			3722	3732	3740	3722	3732	3775	
130	12	6296	4701	4897	4936	5120	5135	5146	0.0331
			5549	5517	5546	5498	5549	5580	
144	12	7728	5674	6089	5983	6402	6419	6429	0.0340
			6926	6881	6911	6891	6929	6976	
137	12	6982	5127	5443	5414	5726	5743	5747	0.0348
			6181	6173	6192	6162	6184	6271	

Table 7.4: Results for medium-sized random graphs

Data				$\frac{\text{Lower}}{\text{Upper}}$ bounds				Rel. gap
n	k	$\lvert\mathsf{E}\rvert$	u_0	BE_L	BE_A	PE_L	PE_A	
2004	35	461865	386285	361900	376315	366025	380173	0.0170
				431210	418626	433334	393350	
1763	35	359293	298477	271810	288338	276742	292997	0.0273
				335801	327283	340000	309453	
1631	35	305901	255840	235396	247030	239323	250704	0.0269
				285787	277977	287345	264577	
2238	45	557743	486365	452193	469991	459627	477125	0.0352
				528404	519478	533745	511978	
2429	45	655197	573402	533846	556720	540877	563493	0.0168
				624029	607715	627066	582761	
2363	45	620375	542582	506838	526087	514241	533235	0.0222
				588947	569792	589190	557422	
2834	55	878964	783849	733871	761522	742441	769788	0.0233
				844913	826628	848004	806582	
3195	55	1113466	997237	941234	973997	948921	981431	0.0194
				1069036	1049898	1074344	1020181	
2863	55	892381	801142	750958	777796	760249	786835	0.0239
				860769	844305	862399	825361	

Table 7.5: Results for larger structured graphs

Data			$\frac{\text{Lower}}{\text{Upper}}$ bounds				Rel. gap
n	k	$\|\mathsf{E}\|$	BE_L	BE_A	PE_L	PE_A	
1863	35	1300266	1179478 1252578	1199702 1250970	1202727 1252942	1222947 1251010	0.0113
1952	35	1428503	1300322 1372084	1322867 1370239	1317912 1372118	1340382 1370104	0.0110
2089	35	1635134	1495514 1576666	1520705 1574543	1515972 1576794	1541096 1574286	0.0107
2383	45	2128866	1966436 2064791	1994338 2062844	1993357 2065059	2021177 2062515	0.0101
2262	45	1918577	1765491 1861114	1792213 1859275	1794208 1861269	1820953 1859358	0.0104
2429	45	2211867	2044687 2144815	2076008 2142597	2068729 2144933	2099995 2142573	0.0100
2764	55	2863291	2664605 2787232	2701727 2785471	2695612 2787807	2732598 2785103	0.0095
2744	55	2822053	2625297 2746913	2662913 2744450	2655062 2747459	2692538 2744324	0.0095
2936	55	3231089	3018448 3150582	3057504 3148539	3051330 3150742	3090316 3147990	0.0092

Table 7.6: Results for larger random graphs

Chapter 8

Conclusion

In this thesis, we have introduced eigenvalue bounds, QP bounds, and SDP bounds for the GP problem. We have used the Hoffman-Wielandt result together with projection techniques to find the eigenvalue bounds. In particular, we break the projected eigenvalue bound into three parts and find their optimal values separately. We have used a zero duality gap result and implicit convexity to find the QP bound and have shown that the QP bound is stronger than the projected eigenvalue bound. We have used the lifting process, or equivalently Lagrangian duality, to derive the SDP relaxation for the GP problem. We then obtain a faciallly reduced SDP relaxation and show that the so-called gangster constraint is very strong so that many other constraints are redundant. We have shown how to recover a feasible solution from an approximate solution by solving an LP . We have also summarized the eigenvalue, quadratic programming and semidefinite programming bounds for the CM problem, a special case of the GP problem.

Our eigenvalue bounds and QP bound can be found efficiently. The computational expense for the basic eigenvalue bound is less than for the projected eigenvalue bound which is less than for the QP bound which is less than for the SDP bound. In our numerical tests, we conclude that the quality of the eigenvalue bounds is comparable to the QP and SDP bounds, but the computational expense of the eigenvalue bounds is much cheaper. Surprisingly, we found that the bounds found by setting the parameter $d = -Au$ are stronger than the bounds using $d = 0$.

Index

Bibliography

[1] K.M. Anstreicher and N.W. Brixius. A new bound for the quadratic assignment problem based on convex quadratic programming. *Math. Program.*, 89(3, Ser. A):341–357, 2001.

[2] K.M. Anstreicher and H. Wolkowicz. On Lagrangian relaxation of quadratic matrix constraints. *SIAM J. Matrix Anal. Appl.*, 22(1):41–55, 2000.

[3] N.W. Brixius and K.M. Anstreicher. Solving quadratic assignment problems using convex quadratic programming relaxations. *Optim. Methods Softw.*, 16(1-4):49–68, 2001. Dedicated to Professor Laurence C. W. Dixon on the occasion of his 65th birthday.

[4] R. A. Brualdi and H. J. Ryser. *Combinatorial Matrix Theory.* Cambridge University Press, New York, 1991.

[5] W.E. Donath and A.J. Hoffman. Lower bounds for the partitioning of graphs. *IBM J. Res. Develop.*, 17:420–425, 1973.

[6] J. Falkner, F. Rendl, and H. Wolkowicz. A computational study of graph partitioning. *Math. Programming*, 66(2, Ser. A):211–239, 1994.

[7] S.W. Hadley, F. Rendl, and H. Wolkowicz. A new lower bound via projection for the quadratic assignment problem. *Math. Oper. Res.*, 17(3):727–739, 1992.

[8] W.W. Hager and J.T. Hungerford. A continuous quadratic programming formulation of the vertex separator problem. Report, University of Florida, Gainesville, 2013.

[9] A.J. Hoffman and H.W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematics*, 20:37–39, 1953.

[10] A. Pothen, H.D. Simon, and K.-P. Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. Matrix Anal. Appl.*, 11(3):430–452, 1990. Sparse matrices (Gleneden Beach, OR, 1989).

[11] Alex Pothen. Graph partitioning algorithms with applications to scientific computing. In *Parallel numerical algorithms (Hampton, VA, 1994)*, volume 4 of *ICASE/LaRC Interdiscip. Ser. Sci. Eng.*, pages 323–368. Kluwer Acad. Publ., Dordrecht, 1997.

[12] Janez Povh and Franz Rendl. Approximating non-convex quadratic programs by semidefinite and copositive programming. In *KOI 2006—11th International Conference on Operational Research*, pages 35–45. Croatian Oper. Res. Soc., Zagreb, 2008.

[13] F. Rendl, A. Lisser, and M. Piacentini. Bandwidth, vertex separators and eigenvalue optimization. In *Discrete Geometry and Optimization*, volume 69 of *The Fields Institute for Research in Mathematical Sciences, Communications Series*, pages 249–263. Springer, 2013.

[14] F. Rendl and H. Wolkowicz. A projection technique for partitioning the nodes of a graph. *Ann. Oper. Res.*, 58:155–179, 1995. Applied mathematical programming and modeling, II (APMOD 93) (Budapest, 1993).

[15] H. Sun, N. Wang, T.K. Pong, and H. Wolkowicz. Eigenvalue, quadratic programming, and semidefinite programming bounds for vertex separators. Technical report, University of Waterloo, Waterloo, Ontario, 2014. 32 pages, submitted.

[16] Levent Tunçel. *Polyhedral and semidefinite programming methods in combinatorial optimization*, volume 27 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2010.

[17] 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.

[18] Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *J. Comb. Optim.*, 2(1):71–109, 1998. Semidefinite programming and interior-point approaches for combinatorial optimization problems (Fields Institute, Toronto, ON, 1996).