On Efficient Semidefinite Relaxations for Quadratically Constrained Quadratic Programming

by

Yichuan Ding

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

Waterloo, Ontario, Canada, 2007

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

Two important topics in the study of Quadratically Constrained Quadratic Programming (QCQP) are how to exactly solve a QCQP with few constraints in polynomial time and how to find an inexpensive and strong relaxation bound for a QCQP with many constraints. In this thesis, we first review some important results on QCQP, like the S-Procedure, and the strength of Lagrangian Relaxation and the semidefinite relaxation. Then we focus on two special classes of QCQP, whose objective and constraint functions take the form $\operatorname{trace}(X^TQX + 2C^TX) + \beta$, and $\operatorname{trace}(X^TQX + XPX^T + 2C^TX) + \beta$ respectively, where X is an n by r real matrix. For each class of problems, we proposed different semidefinite relaxation formulations and compared their strength. The theoretical results obtained in this thesis have found interesting applications, e.g., solving the Quadratic Assignment Problem.

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Dr. Henry Wolkowicz for his excellent supervision and infinite help throughout my graduate study and research.

I am deeply thankful for Dr. Chek Beng Chua, Dr. Levent Tuncel and Dr. Romy Shioda. I learned a lot of knowledge, and also gained much help from them. I also thank Dr. Tuncel and Dr. Vavasis for their effort of reviewing this thesis.

I would like to thank all my friends, who make my life in Canada full of pleasure. Especially, I would like to thank Ms. Haiyan Xu and her family for their especially considerate help.

At last, I would thank my parents, who bring me up to this stage of life and let me enjoy the great interests of academic research!

Contents

1 Introduction **1**
1.1 Problem and Motivation 1
1.2 Outline 3
1.3 Symbols and Notations 3
1.4 Convex Relaxations for QCQP 5

2 QCQP with Few Constraints **11**
2.1 S-Procedure 11
2.2 Strength of Semidefinite Relaxations 17

3 Theory of SDP Relaxations for QMP **22**
3.1 Quadratic Matrix Programming 23
3.2 QMP_1 24
3.3 QMP_2 31
3.3.1 VSDR-2 V.S. MSDR-2 33
3.3.2 Improving the Bound of MSDR-2 44

4 Applications: Quadratic Assignment Problem **46**
4.1 Matrix-Lifting Semidefinite Relaxation for QAP 47
4.2 The Convex Hull of the Orthogonal Similarity Set of B 48
4.3 Projected Bound MSR 51
4.4 Comparing Bounds for QAPLIB Problems 53

5 Conclusions and Future Work **57**

List of Tables

1.1 The Optimization Problems Corresponding to μ_j^* 6

3.1 Comparison of Costs for Different Semidefinite Relaxations 32

4.1 Comparison of Bounds for QAPLIB Instances 55
4.2 Complexity of Relaxations 56
4.3 CPU Time and Iterations for Computing $\boldsymbol{MSR}$ on the Nugent Problems . 56

Chapter 1

Introduction

1.1 Problem and Motivation

Consider the minimization problem with a finite number of constraints

$$\begin{array}{ll} \min & q_0(x) \\ \text{s.t.} & q_j(x) \le 0, \quad j = 1, 2, \ldots, m \\ & x \in \mathbb{R}^n. \end{array} \tag{1.1.1}$$

If for $j = 0, 1, \ldots, m$, $q_j(x)$ are all affine functions (i.e., $q_j(x) = c_j^T x + \beta_j$ for some $c_j \in \mathbb{R}^n$, $\beta_j \in \mathbb{R}$), then (1.1.1) is the well-known linear programming problem (***LP***); if q_0 is a quadratic function ($q_0(x) = x^T Q_0 x + 2c_0^T x + \beta_0$ for some n by n real symmetric matrix Q_0, $c_0 \in \mathbb{R}^n$, and $\beta_0 \in \mathbb{R}$), and for $j = 1, 2, \ldots, m$, $q_j(x)$ are affine functions, then (1.1.1) is a Quadratic Programming problem (***QP***). In this paper, we are interested in situations when for each $j = 0, 1, \ldots, m$, q_j is a quadratic function. Usually, such an optimization problem is called a Quadratically Constrained Quadratic Programming problem (***QCQP***), and has the form

$$(\boldsymbol{QCQP}) \qquad \begin{array}{rll} \mu_0^* := & \min & q_0(x) := x^T Q_0 x + 2c_0^T x \\ & \text{s.t.} & q_j(x) := x^T Q_j x + 2c_j^T x + \beta_j \le 0, \ j = 1, 2, \ldots, m \\ & & x \in \mathbb{R}^n. \end{array}$$

In our definition, we allow the Q_j to be zero matrices. Since the set of linear functions is a subset of quadratic functions, ***QCQP*** is a more general category than ***QP*** or ***LP***.

We can write any integer in a binary system with digits 0 and 1, and the binary constraint $x \in \{0, 1\}$ is equivalent with the quadratic equality constraint $x^2 - x = 0$.

Therefore, any integer programming problem can be reformulated as a ***QCQP***. For example, consider the classical Max-Cut problem

$$(\boldsymbol{MAXCUT}) \qquad \begin{array}{rl} \max & x^T L x \\ \text{s.t.} & x \in \{-1, 1\}^n, \end{array}$$

where L is the n by n Laplacian Matrix [11]. This problem can be formulated as a ***QCQP*** like

$$\begin{array}{rl} \max & x^T L x \\ \text{s.t.} & x^T E_{ii} x = 1, \; i = 1, 2, \ldots, n \\ & x \in \mathbb{R}^n, \end{array} \tag{1.1.2}$$

where E_{ii} is the zero matrix with only the i-th diagonal entry equaling 1.

Some interesting subproblems in nonlinear optimization can also be formulated as a ***QCQP*** with few constraints (typically one or two). For example, to minimize a nonlinear smooth objective $f(x)$, we often use a quadratic function to approximate f in the region around the current iterate point x^k, and find the next search direction d^k. This is the well-known trust-region method [10], and includes solving the trust-region subproblem

$$(\boldsymbol{TRS}) \qquad \begin{array}{rl} \max & \frac{1}{2}(d^k)^T \nabla^2 f(x^k) d^k + (\nabla f(x^k))^T d^k \\ \text{s.t.} & \|d^k\|^2 \leq \delta^2 \\ & d^k \in \mathbb{R}^n. \end{array}$$

The trust-region subproblem is actually a ***QCQP*** which has a single convex quadratic constraint.

Also, consider applying the sequential quadratic programming method, e.g., [6], to minimize a nonlinear objective $f(x)$ subject to m nonlinear constraints $c_j(x) = 0$, $j = 1, 2, \ldots, m$. At the $k-$th iteration point x^k, let $A(x^k) = \begin{pmatrix} \nabla c_1(x^k) & \nabla c_2(x^k) & \ldots & \nabla c_m(x^k) \end{pmatrix}^T$, and $c^k = \begin{pmatrix} c_1(x^k) & c_2(x^k) & \ldots & c_m(x^k) \end{pmatrix}^T$. We need to solve a quadratic program, e.g.,[37],

$$\begin{array}{rl} \max & \frac{1}{2}(d^k)^T \nabla^2 f(x^k) d^k + (\nabla f(x^k))^T d^k \\ \text{s.t.} & A(x^k) d^k + c^k = 0, \end{array} \tag{1.1.3}$$

and also maintain the search step length $\|d^k\| \leq \delta$, for some $\delta > 0$. However, such a d^k may not exist. For this reason, Celis, Dennis and Tapia proposed a new subproblem in 1985 [9] (now called the CDT trust-region subproblem). They replace the linear equality constraint with a quadratic inequality constraint and reformulate (1.1.3) as

$$\begin{array}{rl} \max & \frac{1}{2}(d^k)^T \nabla^2 f(x^k) d^k + (\nabla f(x^k))^T d^k \\ \text{s.t.} & \|A(x^k) d^k + c^k\|^2 \leq \theta^2 \\ & \|d^k\|^2 \leq \delta^2, \end{array} \tag{1.1.4}$$

where δ, θ are some positive real numbers. The CDT trust-region subproblem is actually a $\boldsymbol{QCQP}$ which has two convex quadratic constraints.

Theory on solving such kinds of $\boldsymbol{QCQP}$ will contribute to the development of efficient nonlinear optimization algorithms. We will come to details in Chapter 2.

A $\boldsymbol{QCQP}$ with a large number of constraints is generally too difficult to be solved. But for many application problems, we do not need to solve the $\boldsymbol{QCQP}$ exactly. Instead, we may just want to obtain a bound which involves less expensive computations. This will be another topic we will carefully study in this thesis.

1.2 Outline

We continue in this chapter to provide some preliminary background about semidefinite relaxation and Lagrangian relaxation, and to introduce some notations which will be extensively used in the rest of the thesis. In Chapter 2, we will review some important theoretical results for solving a $\boldsymbol{QCQP}$ with one or two constraints. Our main contributions lie in Chapter 3 and Chapter 4. In Chapter 3, we will discuss a special class of $\boldsymbol{QCQP}$, namely Quadratic Matrix Programming ($\boldsymbol{QMP}$). We divide the $\boldsymbol{QMP}$ problems into two classes, namely $\boldsymbol{QMP_1}$ and $\boldsymbol{QMP_2}$. For each class of problems, we propose two different semidefinite relaxation formulations, i.e., vector-lifting semidefinite relaxation and matrix-lifting semidefinite relaxation. The latter is cheaper to compute, but it generates a bound never stronger than the first. For the $\boldsymbol{QMP_1}$ case, we proved in Theorem 3.2.3 that the two relaxations always generate the same bound; while for the $\boldsymbol{QMP_2}$ case, we proved in Theorem 3.3.1 that the two relaxations generate the same bound under a certain condition. The two theorems (Theorems 3.2.3 and 3.3.1) are our main theoretical contributions. In Chapter 4, we apply the matrix-lifting semidefinite relaxation to generate relaxation bounds for the Quadratic Assignment Problem ($\boldsymbol{QAP}$). The numerical results show that the matrix-lifting semidefinite relaxation is much faster in practical computations compared with the current vector-lifting semidefinite relaxation methods, while the bounds from matrix-lifting semidefinite relaxation are also competitive compared with other relaxation bounds for $\boldsymbol{QAP}$.

1.3 Symbols and Notations

Denote $A \in \mathcal{M}^{nr}$ if A is an n by r real matrix. For two matrices $A,\ B \in \mathcal{M}^{nr}$, the trace inner-product over $\mathcal{M}^{nr}$ is defined as $A \cdot B = \operatorname{trace} A^T B$. We use $A \otimes B$ to denote the

Kronecker product of A and B, i.e.,

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1r}B \\ a_{21}B & \cdots & a_{2r}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{nr}B \end{pmatrix}.$$

Let $\mathcal{R}(A)$ denote the column space of A, and let $\mathcal{N}(A)$ denote the nullspace of A. $\mathcal{S}^n$ denotes the space of $n \times n$ real symmetric matrices; $\mathcal{H}^n$ denotes the space of $n \times n$ Hermitian matrices defined on the complex field; while $\mathcal{S}^n_+$, $\mathcal{H}^n_+$ denote the cones of positive semidefinite matrices in $\mathcal{S}^n$ and $\mathcal{H}^n$ respectively. We let $A \succeq B$ denote the Löwner partial order when $A - B \in \mathcal{S}^n_+$. The linear transformation $\operatorname{diag} X$ denotes the vector formed from the diagonal of the matrix X; the adjoint linear transformation is $\operatorname{diag}^* v = \operatorname{Diag} v$, i.e. the diagonal matrix formed from the vector v; $\operatorname{vec}(X)$ denotes the vector in $\mathbb{R}^{nr}$ obtained from the columns of an $n \times r$ matrix X. For a block-wise n^2 by n^2 matrix $A = (B_{ij})_{i,j=1,2,\dots,n}$, where each B_{ij} is an $n \times n$ submatrix, we denote

$$\begin{array}{rcl} \mathrm{b}^0\mathrm{diag}\,(A) & = & \sum_{i=1}^n B_{ii} \\ \mathrm{o}^0\mathrm{diag}\,(A) & = & (b_{ij})_{i,j=1,2,\dots,n}, \end{array} \tag{1.3.5}$$

where $b_{ij} = \operatorname{trace}(B_{ij})$.

Let e denote the all-ones vector, and E denote the all-ones matrix, both with dimensions consistent with the context. For square matrices in $\mathcal{M}^{nn}$, $\mathcal{O}$ denotes the set of orthogonal matrices, $XX^T = X^TX = I$; while, for a given $B \in \mathcal{S}^n$, $\mathcal{O}(B)$ denotes the set of symmetric matrices orthogonally similar to B, i.e., $\mathcal{O}(B) := \{Y \in \mathcal{S}^n : Y = XBX^T, X \in \mathcal{O}\}$. Π denotes the set of permutation matrices; $\mathcal{N}$ denotes the nonnegative, element-wise, matrices, i.e., $X \geq 0$; $\mathcal{E}$ denotes matrices with row and column sums one. We denote the vector of eigenvalues of a matrix A by $\lambda(A)$. We define the *minimal product* of two vectors $\langle x, y\rangle_- := \min\limits_{\pi} \sum_{i=1}^n x_i y_{\pi(i)}$, where the minimum is over all permutations of the indices $\{1, 2, \dots, n\}$. Clearly, the minimum is attained if the vectors (x_i) and $(y_{\pi(i)})$ are sorted in reverse order by magnitude, i.e., $x_1 \leq x_2 \leq \dots \leq x_n$ and $y_{\pi(1)} \geq y_{\pi(2)} \geq \dots \geq y_{\pi(n)}$. Similarly, we define the maximal product of x, y, $\langle x, y\rangle_+ := \max\limits_{\pi} \sum_{i=1}^n x_i y_{\pi(i)}$.

For a square singular matrix A with singular value decomposition [19] $A = U\Sigma_A V^T$, where $\Sigma_A = \operatorname{Diag}(\sigma(A))$, define its Moore-Penrose inverse [49]

$$A^\dagger = V\Sigma_A^\dagger U^T,$$

where the diagonal matrix $\Sigma_A^\dagger$ has its diagonal entries

$$(\Sigma_A^\dagger)_{ii} = \begin{cases} \frac{1}{\sigma_i(A)} & \text{if } \sigma_i(A) \neq 0 \\ 0 & \text{if } \sigma_i(A) = 0. \end{cases}$$

The Moore-Penrose inverse is characterized by the following four properties:

1. $AA^\dagger A = A$;
2. $A^\dagger AA^\dagger = A^\dagger$;
3. $AA^\dagger = \mathcal{P}_{\mathcal{R}(A)}$;
4. $A^\dagger A = \mathcal{P}_{\mathcal{R}(A^T)}$,

where $\mathcal{P}_{\mathcal{R}(A)}$, $\mathcal{P}_{\mathcal{R}(A^T)}$ denotes the orthogonal projection onto $\mathcal{R}(A)$, $\mathcal{R}(A^T)$ respectively.

For simplicity, instead of inf/sup, we use min/max; however it does not mean the optimum is always finite or attained. When we say an optimal solution for a ***SDP***, we allow it has an error bounded by ϵ. We also use the following simplified notations throughout the thesis:

$$\begin{array}{rcl} Q_\lambda & := & Q_0 + \sum_{j=1}^m \lambda_j Q_j \\ P_\lambda & := & P_0 + \sum_{j=1}^m \lambda_j P_j \\ c_\lambda & := & c_0 + \sum_{j=1}^m \lambda_j c_j \\ \beta_\lambda & := & \sum_{j=1}^m \lambda_j \beta_j, \end{array} \tag{1.3.6}$$

for any $\lambda = \{\lambda_j\}_{j=1,2,\ldots,m}$.

Throughout this thesis, we use μ_j^* to denote optimal values of various mathematical programming problems. We listed them in Table 1.1 for the convenience of the readers.

1.4 Convex Relaxations for QCQP

Consider a minimization problem with the form

$$\begin{array}{ll} \min & C \cdot X \\ \text{s.t.} & A_i \cdot X = b_i, \ i = 1, 2, \ldots, m_1 \\ & B_j \cdot X \leq d_j, \ j = 1, 2, \ldots, m_2 \\ & X \succeq 0. \end{array} \tag{1.4.7}$$

μ_j^*	Corresponding Optimization Problems
μ_0^*	$\boldsymbol{QCQP}$
μ_1^*	$\boldsymbol{SDPP}$
μ_2^*	(2.1.12)
μ_3^*	$\boldsymbol{VSDR-1}$
μ_4^*	$\boldsymbol{MSDR-1}$
μ_5^*	$\boldsymbol{VSDR-1'}$
μ_6^*	$\boldsymbol{VSDR-2}$
μ_7^*	$\boldsymbol{MSDR-2}$
μ_8^*	$\boldsymbol{DV-2}$
μ_9^*	$\boldsymbol{DM-2}$
μ_{10}^*	(3.3.2)
μ_{11}^*	(3.3.71)
μ_{12}^*	(3.3.89)
μ_{13}^*	(4.1.97)
μ_{14}^*	(4.2.100)
μ_{15}^*	$\boldsymbol{MSR}$
μ_{LR}^*	Bounds from Lagrangian Relaxations

Table 1.1: The Optimization Problems Corresponding to μ_j^*

We refer to (1.4.7) as a semidefinite program (***SDP***). It has been studied as far back as the 1940s, and became popular in the 1990s due to its important applications and the development of interior point algorithms. ***SDP*** is much more general than ***LP***, because the positive semidefinite constraint is a nonlinear constraint. With a primal-dual interior point method, ***SDP*** can be solved to optimality with an error bounded by a given ϵ in polynomial time [36]. In this section, we will discuss the semidefinite relaxations for ***QCQP***, i.e., generating a lower bound for a ***QCQP*** with a ***SDP***. If the generated bound equals the optimal value of the original program, the relaxation is called to be *tight*.

We can rewrite a quadratic function as a trace inner-product between two matrices: $q_j(x) := x^T Q_j x + 2c_j^T x + \beta_j = M(q_j(\bullet)) \cdot \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix}$, where $M(q_j(\bullet)) := \begin{pmatrix} \beta_j & c_j^T \\ c_j & Q_j \end{pmatrix}$. So the feasible set in ***QCQP*** admits a lifted representation

$$\begin{aligned} &\{x \in \mathbb{R}^n | x^T Q_j x + 2c_j^T x + \beta_j \leq 0,\ j = 1, 2, \ldots, m\} \\ &\quad = \left\{ x \in \mathbb{R}^n | M(q_j(\bullet)) \cdot Z \leq 0,\ j = 1, 2, \ldots, m, \text{ for some } Z = \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \right\}. \end{aligned}$$

The equality constraint $Z = \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix}$ is still nonconvex. A usual approach is to relax the nonconvex equality constraint to the convex constraint: $Z = \begin{pmatrix} 1 & x^T \\ x & Y \end{pmatrix} \succeq 0$ for some n by n real symmetric matrix Y. Then we get a semidefinite relaxation (named ***SDPP*** in this thesis)

$$(\boldsymbol{SDPP}) \qquad \begin{array}{rll} \mu_1^* := & \min & M(q_0(\bullet)) \cdot Z \\ & \text{s.t.} & M(q_j(\bullet)) \cdot Z \leq 0,\ j = 1, 2, \ldots, m \\ & & Z_{11} = 1 \\ & & Z \succeq 0. \end{array}$$

Now let $P :=$ cone $\{M(q_j(\bullet)) | j = 1, 2, \ldots, m\}$, i.e., the closed convex cone generated by $M(q_j(\bullet))$ $(j = 1, 2, \ldots, m)$, and let $P^* := \{Y | X \cdot Y \geq 0$ for all $X \in P\}$, i.e., the dual cone of P. We also denote the feasible set of ***SDPP*** with $\mathcal{F}_{\boldsymbol{SDPP}}$, then we have the following theorem, which is from Theorem 4.2 in [29].

Theorem 1.4.1 *The semidefinite relaxation* ***SDPP*** *is as strong as a* ***QCQP*** *which includes all the convex quadratic constraints* $q(x) \leq 0$ *with* $M(q(\bullet))$ *in the cone generated by*

$M(q_j(\bullet))$, or equivalently,

$$\left\{x \in \mathbb{R}^n | Z = \begin{pmatrix} 1 & x^T \\ x & Y \end{pmatrix} \in \mathcal{F}_{\boldsymbol{SDPP}} \text{ for some } Y \in \mathcal{S}^n \right\}$$
$$= \left\{x \in \mathbb{R}^n | x^T Q x + 2c^T x + \beta \leq 0,\ \forall \begin{pmatrix} \beta & c^T \\ c & Q \end{pmatrix} \in P \text{ with } Q \succeq 0 \right\}. \tag{1.4.8}$$

Proof. See the proof to Theorem 4.2 in [29]. ■

A quadratic function $q(x) = x^T Q x + 2c^T x + \beta$ is convex if and only if $Q \succeq 0$. A ***QCQP*** is called convex if all the objective and constraint functions are convex. A convex ***QCQP*** can be solved in polynomial time by nonlinear optimization methods [37] or a Second Order Cone Programming (***SOCP***), or more trivially, a ***SDP*** [5].

Proposition 1.4.1 *If a **QCQP** is convex and its minimum is attained, then **SDPP** solves **QCQP** to optimality, i.e. $\mu_1^* = \mu_0^*$.*

Proof. Suppose ***SDPP*** is minimized at $Z^* = \begin{pmatrix} 1 & (x^*)^T \\ x^* & Y^* \end{pmatrix}$, then $Z^* \succeq 0$ implies $Y^* \succeq x^*(x^*)^T$. Denote $\bar{Y} = x^*(x^*)^T$. Then $Y^* \succeq \bar{Y}$. Because $Y^* \succeq \bar{Y}$ and $Q_j \succeq 0$, we have $M(q_j(\bullet)) \cdot \begin{pmatrix} 1 & (x^*)^T \\ x^* & \bar{Y} \end{pmatrix} \leq M(q_j(\bullet)) \cdot \begin{pmatrix} 1 & (x^*)^T \\ x^* & Y^* \end{pmatrix}$, for $j = 1, \ldots, m$. Therefore, $\bar{Z} := \begin{pmatrix} 1 & (x^*)^T \\ x^* & \bar{Y} \end{pmatrix}$ is also feasible for ***SDPP*** . Furthermore, $Y^* \succeq \bar{Y}$ indicates the objective $M(q_0(\bullet)) \cdot \begin{pmatrix} 1 & (x^*)^T \\ x^* & \bar{Y} \end{pmatrix} \leq M(q_0(\bullet)) \cdot \begin{pmatrix} 1 & (x^*)^T \\ x^* & Y^* \end{pmatrix}$. Together with the fact that $Z^* = \begin{pmatrix} 1 & (x^*)^T \\ x^* & Y^* \end{pmatrix}$ minimizes ***SDPP*** , we conclude that $\bar{Z}$ also minimizes ***SDPP*** . Because $\bar{Y} = x^*(x^*)^T$, and x^* is feasible to the original ***QCQP*** , we know $\mu_0^* \leq q_0(x^*) = \mu_1^*$, which implies the tightness of ***SDPP*** for ***QCQP*** . ■

Remark 1.4.1 *In fact, minimum attainment is not the necessary condition to establish Proposition 1.4.1. It has been shown in [18] that semidefinite relaxation, or Lagrangian relaxation (see succeeding paragraph), always admits zero gap for a convex **QCQP** .*

■

If not all $Q_j \succeq 0$ for $j = 0, 1, \ldots, m$, then we get a nonconvex ***QCQP***. In this case, ***SDPP*** can fail to solve the ***QCQP*** exactly but provides a lower bound, i.e., $\mu_1^* \leq \mu_0^*$.

Another form of convex relaxation for ***QCQP*** is the so-called Lagrangian dual relaxation (***LR***)

$$\begin{aligned}
\mu_0^* &= \min_{x\in\mathbb{R}^n} \max_{\lambda\in\mathbb{R}_+^m} x^T(Q_0 + \textstyle\sum_{j=1}^m \lambda_j Q_j)x + 2(c_0 + \sum_{j=1}^m \lambda_j c_j)^T x + \sum_{j=1}^m \lambda_j \beta_j \\
&\geq \max_{\lambda\in\mathbb{R}_+^m} \min_{x\in\mathbb{R}^n} x^T(Q_0 + \textstyle\sum_{j=1}^m \lambda_j Q_j)x + 2(c_0 + \sum_{j=1}^m \lambda_j c_j)^T x + \sum_{j=1}^m \lambda_j \beta_j \\
&=: \mu_{LR}^*.
\end{aligned}$$

Here μ_{LR}^* denotes the Lagrangian relaxation bound for ***QCQP***. The inequality here is due to the weak duality $\min_x \max_\lambda L(x,\lambda) \geq \max_\lambda \min_x L(x,\lambda)$ for any real valued function $L(x,\lambda)$ [7].

In fact, the Lagrangian dual program can be formulated as a ***SDP***. To see that, we first homogenize $L(x,\lambda)$ as

$$\begin{aligned}
\mu_{LR}^* &= \max_{\lambda\in\mathbb{R}_+^m} \min_{x\in\mathbb{R}^n} x^T Q_\lambda x + 2c_\lambda^T x + \beta_\lambda \\
&= \max_{\lambda\in\mathbb{R}_+^m} \min_{x\in\mathbb{R}^n, x_0^2=1} x^T Q_\lambda x + 2x_0 c_\lambda^T x + \beta_\lambda x_0^2
\end{aligned} \tag{1.4.9a}$$

$$\begin{aligned}
&= \max_{\lambda\in\mathbb{R}_+^m, \alpha\in\mathbb{R}} \min_{x\in\mathbb{R}^n, x_0\in\mathbb{R}} x^T Q_\lambda x + 2x_0 c_\lambda^T x + \beta_\lambda + \alpha(x_0^2 - 1) \\
&= \max_{\lambda\in\mathbb{R}_+^m, \alpha\in\mathbb{R}} \min_{x\in\mathbb{R}^n, x_0\in\mathbb{R}} \beta_\lambda - \alpha + \begin{pmatrix} x_0 & x^T \end{pmatrix} \begin{pmatrix} \alpha & c_\lambda^T \\ c_\lambda & Q_\lambda \end{pmatrix} \begin{pmatrix} x_0 \\ x \end{pmatrix},
\end{aligned} \tag{1.4.9b}$$

where $Q_\lambda,\ c_\lambda,\ \beta_\lambda$ follow from the notations introduced in (1.3.6). The equality (1.4.9b) follows from the losslessness of S-Procedure [15], which will be discussed in later sections.

Note that the value $\min_{x\in\mathbb{R}^n, x_0\in\mathbb{R}} \begin{pmatrix} x_0 & x^T \end{pmatrix} \begin{pmatrix} \alpha & c_\lambda^T \\ c_\lambda & Q_\lambda \end{pmatrix} \begin{pmatrix} x_0 \\ x \end{pmatrix}$ is finite only when $\begin{pmatrix} \alpha & c_\lambda^T \\ c_\lambda & Q_\lambda \end{pmatrix} \succeq 0$, and then the minimal value always equals zero (when $x_0 = 0$, $x = 0$). So the right hand side of (1.4.9b) can be further formulated as [42]

$$(\boldsymbol{SDPD}) \qquad \begin{aligned}
\mu_{LR}^* = \ \max \ & \textstyle\sum_{j=1}^m \lambda_j \beta_j - \alpha \\
\text{s.t.} \ & \begin{pmatrix} \alpha & (c_0 + \sum_{j=1}^m \lambda_j c_j)^T \\ c_0 + \sum_{j=1}^m \lambda_j c_j & Q_0 + \sum_{j=1}^m \lambda_j Q_j \end{pmatrix} \succeq 0 \\
& \lambda_j \geq 0,\ j = 1, 2, \ldots m.
\end{aligned}$$

It is easy to check that $\boldsymbol{SDPD}$ is just the conic dual program of $\boldsymbol{SDPP}$. According to [25], the minimal value of $\boldsymbol{SDPP}$ (μ_1^*) and the maximal value of $\boldsymbol{SDPD}$ (μ_{LR}^*) coincide if the so-called generalized Slater condition holds, i.e.,

$$\begin{aligned}
&\exists \alpha \in \mathbb{R}, \quad \lambda_j \geq 0, \quad j = 1, 2 \ldots, m, \\
&\qquad\qquad \text{s.t.} \begin{pmatrix} \alpha & (c_0 + \sum_{j=1}^m \lambda_j c_j)^T \\ c_0 + \sum_{j=1}^m \lambda_j c_j & Q_0 + \sum_{j=1}^m \lambda_j Q_j \end{pmatrix} \succ 0.
\end{aligned} \tag{1.4.10}$$

Because α can be an arbitrarily large number, condition (1.4.10) is equivalent with

$$\exists \lambda_j \geq 0, \; j = 1, 2, \ldots m, \qquad \text{s.t. } Q_0 + \sum_{j=1}^{m} \lambda_j Q_j \succ 0. \tag{1.4.11}$$

So under condition (1.4.11), $\boldsymbol{SDPP}$ and $\boldsymbol{SDPD}$ generate the same bounds. Because $\boldsymbol{SDPD}$ is just the Lagrangian relaxation, we have the following theorem.

Theorem 1.4.2 *For a* $\boldsymbol{QCQP}$*, if condition (1.4.11) holds, then the semidefinite relaxation* $\boldsymbol{SDPP}$ *and Lagrangian relaxation* $\boldsymbol{LR}$ *generate the same bounds .*

■

Chapter 2

QCQP with Few Constraints

In this chapter, we will move to the theory behind solving $\boldsymbol{QCQP}$ with few constraints. Let m denote the number of constraints in a $\boldsymbol{QCQP}$. It can be shown that the difficulty of solving a $\boldsymbol{QCQP}$ grows dramatically with the increase of m. For example, a $\boldsymbol{QCQP}$ with $m = 1$ is solvable in polynomial time; a $\boldsymbol{QCQP}$ with $m = 2$ is solvable in polynomial time when one of the constraints is not active at optimality; whereas efficiently solving a $\boldsymbol{QCQP}$ with $m > 2$ to optimality is still an open problem. A $\boldsymbol{QCQP}$ can be solved in polynomial time if the convex relaxation (semidefinite relaxation or Lagrangian relaxation) provides a tight bound. However, it is quite challenging to recognize special nonconvex $\boldsymbol{QCQP}$ problems that admit tight convex relaxations. Different approaches have been employed to study this issue, including methods based on separation theory, Lagrangian saddle function, and the rank-reduction procedure for a positive semidefinite matrix in an affine space. And several remarkable theories have been established, e.g., the S-Procedure [16].

2.1 S-Procedure

One important result on the strength of the Lagrangian relaxation for $\boldsymbol{QCQP}$ is the so-called S-Procedure [16]. Suppose $q_j : V \to \mathbb{R}$, for $j = 0, 1, \ldots, m$, are m real valued functionals defined on a vector space V. Consider a minimization problem

$$\mu_2^* := \min\{q_0(x)|\ q_j(x) \leq 0,\ j = 1, 2, \ldots, m\}. \tag{2.1.12}$$

We define the Lagrangian function

$$L(x, \lambda) = q_0(x) + \sum_{j=1}^{m} \lambda_j q_j(x), \tag{2.1.13}$$

where $\lambda_1, \ldots, \lambda_m$ are real numbers. Then consider the following two conditions:

$$q_0(x) \geq 0 \quad \text{for} \quad \forall x \in V, \text{ s.t. } q_j \leq 0, \; j = 1, 2, \ldots, m, \tag{2.1.14}$$

$$\exists \lambda_j \geq 0, \; j = 1, 2, \ldots, m, \quad \text{s.t. } \forall x \in V, \; L(x, \lambda) \geq 0. \tag{2.1.15}$$

It is straightforward to check that (2.1.15) implies (2.1.14); however, the converse may not be true. People call the S-procedure lossless for (2.1.12) if (2.1.14) also leads to (2.1.15) [16].

In fact, the losslessness of the S-Procedure is equivalent with the tightness of Lagrangian relaxation for (2.1.12). To see this, put $\hat{q}_0(x) := q_0(x) - \mu_2^*$, then it is straightforward to check that the S-Procedure is lossless for $\hat{q}_0$, q_j, $j = 1, 2, \ldots, m$, if and only if $\max_{\lambda} \min_{x} \; q_0(x) + \sum_{j=1}^{m} \lambda_j q_j(x) = \mu_2^*$, i.e., the Lagrangian relaxation is tight.

(2.1.12) is said to satisfy the *regularity* condition ([16]) if

$$\exists x \in V, \qquad \text{s.t.} \quad q_j(x) < 0 \text{ for each } j = 1, 2, \ldots, m. \tag{2.1.16}$$

A sufficient condition for the losslessness of the S-Procedure is the convexity of the joint image and the regularity condition (2.1.16).

Lemma 2.1.1 ([52]) *For* (2.1.12)*, define a mapping from V to $\mathbb{R}^{m+1}$,*

$$\varphi(x) = \begin{pmatrix} q_0(x) \\ q_1(x) \\ \vdots \\ q_m(x) \end{pmatrix}. \tag{2.1.17}$$

If its image $\Gamma := \{\varphi(x) | x \in X\}$ is convex, and condition (2.1.16) *holds, then the S-procedure is lossless for* (2.1.12)*.*

Proof. Let

$$S := \left\{ t := \begin{pmatrix} t_0 \\ t_1 \\ \vdots \\ t_m \end{pmatrix} \in \mathbb{R}^{m+1} | t_0 < 0, \; t_j \leq 0, \; j = 1, 2, \ldots, m \right\}. \tag{2.1.18}$$

Suppose (2.1.14) holds, then $\Gamma \cap S = \emptyset$. Both Γ and S are convex sets. By the separation theorem [27], and the fact $0 \in \text{cl}(S)$, we deduce that there exist $\hat{\lambda}_j,\ j = 0, 1, \ldots, m$, such that

$$\hat{\lambda}_0 q_0(x) + \sum_{j=1}^{m} \hat{\lambda}_j q_j(x) \geq 0, \qquad \text{for all } x \in V, \tag{2.1.19a}$$

$$\hat{\lambda}_0 t_0 + \sum_{j=1}^{m} \hat{\lambda}_j t_j \leq 0, \qquad \text{for all } t \in S. \tag{2.1.19b}$$

(2.1.19b) implies $\hat{\lambda}_j \geq 0$, $j = 0, 1, \ldots, m$; while (2.1.19a) together with condition (2.1.16) implies $\hat{\lambda}_0 > 0$. By multiplying $1/\hat{\lambda}_0$ through (2.1.19a), we get

$$q_0(x) + \sum_{j=1}^{m} \frac{\hat{\lambda}_0}{\hat{\lambda}_j} q_j(x) \geq 0, \qquad \text{for all } x \in V. \tag{2.1.20}$$

Let $\lambda_j := \hat{\lambda}_j / \hat{\lambda}_0$, $j = 1, 2, \ldots, m$. Then $\lambda_j \geq 0$, $j = 1, 2, \ldots, m$. Therefore, (2.1.20) leads to (2.1.15), and the S-Procedure is lossless for (2.1.12). ■

In [16], based on Lemma 2.1.1 and the convexity of the image $\left\{ \begin{pmatrix} x^T Q_0 x \\ x^T Q_1 x \end{pmatrix} | x \in \mathbb{R}^n \right\}$, the S-procedure is proved lossless for (2.1.12) when $m = 1$ and $q_0,\ q_1$ are both quadratic functions. This leads to the following result.

Theorem 2.1.1 *The Lagrangian relaxation (**LR**), or the semidefinite relaxation (**SDPP**) is tight for a **QCQP** with $m = 1$.*

Proof. Equivalent results have been proved in [16, 52, 48] with different methods. ■

Now we will extend this result to the complex ***QCQP*** case.

Theorem 2.1.2 *Consider a **QCQP** over the complex space*

$$\begin{array}{ll} \min & x^* Q_0 x + 2Re(c_0^* x) + \beta_0 \\ s.t. & x^* Q_j x + 2Re(c_j^* x) + \beta_j,\ j = 1, 2, \ldots, m \\ & \quad x \in \mathcal{C}^n, \end{array} \tag{2.1.21}$$

*where Q_j are real symmetric matrices, x^*, c_j^* denote the conjugate transposition of the vectors $x,\ c \in \mathcal{C}^2$. If $m \leq 2$, then the Lagrangian relaxation (**LR**), or equivalently the semidefinite relaxation (**SDPP**) is tight for* (2.1.21).

Proof. The proof to Theorem 2.1.1 [16] can be extended to prove Theorem 2.1.2 word-by-word, except for the need to show the convexity of the joint image of three quadratic mappings over the complex space (Lemma 2.1.2), instead of two quadratic mappings over the real space. ■

Lemma 2.1.2 *Let* $\varphi(x) = \begin{pmatrix} x^*Q_1x \\ x^*Q_2x \\ x^*Q_3x \end{pmatrix}$, $x \in \mathcal{C}^n$. *Then its image:* $\Gamma := \{\varphi(x) | x \in \mathcal{C}^n\} \subseteq \mathbb{R}^3$ *is convex.*

Proof. We first prove the case of $n = 2$, i.e., when the variable vector is in the two dimensional complex space $\mathcal{C}^2$. For given $x_1, x_2 \in \mathcal{C}^2$, $\lambda \in [0, 1]$, we want to prove

$$\lambda\varphi(x_1) + (1 - \lambda)\varphi(x_2) \in \Gamma,$$

that is,

$$\{x \in \mathcal{C}^2 | \varphi(x) = \lambda\varphi(x_1) + (1 - \lambda)\varphi(x_2)\} \neq \emptyset. \tag{2.1.22}$$

We define a mapping from $\mathcal{H}^2$ to $\mathcal{C}^3$,

$$\psi(X) := \begin{pmatrix} \mathrm{trace}\,(Q_1X) \\ \mathrm{trace}\,(Q_2X) \\ \mathrm{trace}\,(Q_3X) \end{pmatrix}.$$

If $X \succeq 0$ and $\mathrm{rank}\,(X) = 1$, then $X = xx^*$ and $\psi(X) = \varphi(x)$. Therefore, to prove (2.1.22), we only need to prove

$$\{X \in \mathcal{H}^2_+ | \mathrm{rank}\,(X) = 1,\ \psi(X) = \lambda\psi(X_1) + (1 - \lambda)\psi(X_2)\} \neq \emptyset. \tag{2.1.23}$$

Since ψ is a linear mapping, for any $\lambda \in [0, 1]$, by defining $X_\lambda = \lambda X_1 + (1 - \lambda)X_2$, we have $\psi(X_\lambda) = \lambda\psi(X_1) + (1 - \lambda)\psi(X_2)$ and $X_\lambda \succeq 0$. So if X_λ is a singular 2 by 2 matrix, then it should belong to the left hand side of (2.1.23). Otherwise, we have $X_\lambda \succ 0$. Then we consider the linear equation system

$$\begin{array}{rcl} \mathrm{trace}\,(Q_1X) &=& 0 \\ \mathrm{trace}\,(Q_2X) &=& 0 \\ \mathrm{trace}\,(Q_3X) &=& 0 \\ X &\in& \mathcal{H}^2, \end{array} \tag{2.1.24}$$

which admits a nontrivial solution $\bar{X} := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \in \mathcal{H}^2$. Thus, $\forall c \in \mathbb{R}$, $\psi(X_\lambda + c\bar{X}) = \psi(X_\lambda) = \lambda\psi(X_1) + (1-\lambda)\psi(X_2)$. Because $\bar{X} \neq 0$ and $X_\lambda \succ 0$, we know the matrix pencil $X_\lambda + c\bar{X}$ will intersect the positive semidefinite cone for some $\bar{c} \in \mathbb{R}$. Therefore, $\exists \bar{c} \in \mathbb{R}$, such that rank $(X_\lambda + \bar{c}\bar{X}) < n = 2$, and

$$X_\lambda + \bar{c}\bar{X} \in \{X \in \mathcal{H}_+^2 | \text{rank}\,(X) = 1,\ \psi(X) = \lambda\psi(X_1) + (1-\lambda)\psi(X_2)\},$$

i.e., (2.1.23) holds. Hence, Lemma 2.1.2 is proved in the case $n = 2$.

Now we move to the case $n > 2$. For any $x_1, x_2 \in C^n$ that are linearly dependent, and for any given $\lambda \in (0,1)$, we can find a scalar α^* such that $\alpha^* x_1$ belongs to the set in (2.1.22), which makes (2.1.22) satisfied. If x_1 and x_2 are linearly independent, then we will prove there exists an $\bar{x}$ in the two dimensional subspace spanned by x_1 and x_2, i.e., $\bar{x} = (\,x_1 \quad x_2\,)\, v$ for some $v \in \mathcal{C}^2$, and $\bar{x}$ belongs to the set in (2.1.22). The proof is as following. Let $\bar{Q}_i = (\,x_1 \quad x_2\,)^T Q_i\, (\,x_1 \quad x_2\,)$ for $i = 1, 2, 3$, define $\bar{\varphi}:\ \mathcal{C}^2 \to \mathbb{R}^3$ as

$$\bar{\varphi}(v) := \begin{pmatrix} v^*\bar{Q}_1 v \\ v^*\bar{Q}_2 v \\ v^* Q_3 v \end{pmatrix}. \tag{2.1.25}$$

Thus, if $x = (\,x_1 \quad x_2\,)\, v$, we have

$$\varphi(x) = \begin{pmatrix} x^* Q_1 x \\ x^* Q_2 x \\ x^* Q_3 x \end{pmatrix} = \begin{pmatrix} v^*\bar{Q}_1 v \\ v^*\bar{Q}_2 v \\ v^*\bar{Q}_3 v \end{pmatrix} = \bar{\varphi}(v). \tag{2.1.26}$$

Therefore, to prove (2.1.22), we just need to prove

$$\{v \in \mathcal{C}^2 | \bar{\varphi}(v) = \lambda\bar{\varphi}(v_1) + (1-\lambda)\bar{\varphi}(v_2)\} \neq \emptyset, \tag{2.1.27}$$

where $v_1 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $v_2 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, so that $\bar{\varphi}(v_1) = \varphi(x_1)$, $\bar{\varphi}(v_2) = \varphi(x_2)$. Note that (2.1.27) is just the $n = 2$ case we have proved. So Lemma 2.1.2 holds true for any $n \geq 2$. ■

Theorem 2.1.3 ([41]) *Consider a **QCQP** with m constraints, and each q_j is a homogeneous quadratic function ($c_j = 0$) for $j = 0, 1, \ldots m$ with $m \leq n$. If (2.1.14) holds, then there exists $\lambda_j \geq 0$, such that $Q_0 + \Sigma_{i=1}^m \lambda_i Q_i$ has at most $m - 1$ negative eigenvalues.*

Theorem 2.1.3 is a more general result about the S-Procedure for quadratic functions. It is interesting that if $m = 2$ and $q_j = x^*Q_jx$ $(j = 0, 1, 2)$, then the losslessness of the S-Procedure for (2.1.12) can be easily proved by Theorem 2.1.3. Note that any Hermitian form z^*Qz could be represented with

$$\begin{pmatrix} \Re(z) & \Im(z) \end{pmatrix} \begin{pmatrix} \Re(Q) & \Im(Q) \\ -\Im(Q) & \Re(Q) \end{pmatrix} \begin{pmatrix} \Re(z) \\ \Im(z) \end{pmatrix},$$

where $\Re(Q)$ and $\Im(Q)$ denote the real and imaginary part of Q, respectively. By assuming Q is real, the Hermitian matrix will take the form of $\begin{pmatrix} \Re(Q) & 0 \\ 0 & \Re(Q) \end{pmatrix}$, which indicates all the eigenvalues of the Hessian $Q_0 + \Sigma_{j=1}^m \lambda_j Q_j$ will appear in pairs. So the number of negative eigenvalues will always be even. According to Theorem 2.1.3, if (2.1.14) holds, then there exists $\lambda \in \mathbb{R}^m$ such that $Q_0 + \Sigma_{j=1}^m \lambda_j Q_j$ has at most one negative eigenvalue (but the number should be even!), which actually implies $Q_0 + \Sigma_{j=1}^m \lambda_j Q_j \succeq 0$. So (2.1.15) holds.

Finally, we give a result derived in [32] that generalizes the S-Lemma to the matrix case.

Lemma 2.1.3 ([32]) *Suppose* $Q_0,\ B_0,\ Q_1 \in \mathcal{S}^n$, $C_0 \in \mathcal{M}^{nn}$, *then*

$$X^TQ_0X + C_0^TX + X^TC_0 + B_0 \succeq 0, \text{ for all } X \in \mathcal{M}^{nn} \text{ with } I - X^TQ_1X \succeq 0, \quad (2.1.28)$$

if and only if

$$\begin{pmatrix} B_0 & C_0^T \\ C_0 & Q_0 \end{pmatrix} - t \begin{pmatrix} I & 0 \\ 0 & -Q_1 \end{pmatrix} \succeq 0, \quad \text{for some } t \geq 0. \quad (2.1.29)$$

Proof. First establish the equivalence between (2.1.28) and

$$x_1^TB_0x_1 + 2x_2^TC_0x_1 + x_2^TQ_0x_2 \geq 0, \text{ for all } x_1, x_2 \text{ with } x_1^Tx_1 - x_2^TQ_1x_2 \geq 0. \quad (2.1.30)$$

Then deduce the equivalence between (2.1.30) and (2.1.29) by the losslessness of S-Procedure for the one quadratic constraint case. For details, see [32]. ■

2.2 Strength of Semidefinite Relaxations

Another approach to study the tightness of Lagrangian relaxation is to check the existence of a rank-one optimal solution to the semidefinite relaxation program $\boldsymbol{SDPP}$. Suppose there is a rank-one solution $Z^* = \begin{pmatrix} 1 & (x^*)^T \\ x^* & Y^* \end{pmatrix} \succeq 0$ that minimizes $\boldsymbol{SDPP}$, then it is not difficult to see $Y^* = x^*(x^*)^T$ by the Schur complement [33], which implies the feasibility of x^* and the tightness of $\boldsymbol{SDPP}$.

There is a well-known theorem saying that a low rank positive semidefinite matrix always exists in an affine space that has dimensionality high enough.

Theorem 2.2.1 ([39, 40, 3]) *Consider the semidefinite program* (1.4.7)*. If* $m_1 + m_2 \leq \begin{pmatrix} r+2 \\ 2 \end{pmatrix} - 1$*, then* (1.4.7) *has an optimal solution* X^* *with* $\operatorname{rank}(X^*) \leq r$.

■

With Theorem 2.2.1, the tightness of semidefinite relaxation for a $\boldsymbol{QCQP}$ with $m = 1$ can be easily proved. To see this, we notice that the relaxation $\boldsymbol{SDPP}$ for a $\boldsymbol{QCQP}$ with $m = 1$ has two constraints (with an additional constraint $Z_{11} = 1$). Because the inequality $2 \leq \begin{pmatrix} r+2 \\ 2 \end{pmatrix} - 1$ holds for $r = 1$, Theorem 2.2.1 implies the existence of a rank-one optimal solution to $\boldsymbol{SDPP}$. So the $\boldsymbol{SDPP}$ is tight for a $\boldsymbol{QCQP}$ with $m = 1$.

However, Theorem 2.2.1 is too general, so we can not expect to get very surprising results by directly applying Theorem 2.2.1. To further explore the low rank attributes, a new rank-one decomposition technique is proposed in [50].

Proposition 2.2.1 ([50]) *Given a symmetric matrix* G*, and a positive semidefinite symmetric matrix* X *with* $\operatorname{rank}(X) = r$*, then* $G \cdot X = 0$ *if and only if there is a rank-one decomposition* $X = \sum_{m=1}^{r} p_i p_i^T$*, such that*

$$p_i^T G p_i = 0 \text{ for } i = 1, 2, \ldots, r.$$

Proof. The sufficiency is clear. We can prove the necessity by showing that the following rank-one decomposition procedure is correct.

Rank-one Decomposition Procedure [50]: *Input*: $X, G \in \mathcal{S}^n$, such that $X \succeq 0$, $\operatorname{rank}(X) = r$ and $G \cdot X = 0$. *Output*: Vector $y \in \mathbb{R}^n$, such that $y^T G y = 0$, $X - yy^T \succeq 0$ and with rank $r - 1$.
Step 0: Compute r linear independent vectors $p_1, p_2, \ldots, p_r$ such that $X = \sum_{i=1}^{r} p_i p_i^T$.

Step 1: If $p_1^T G p_1 = 0$, then return $y = p_1$, and the procedure completes; otherwise, suppose $p_1^T G p_1 > 0$, and choose j such that $p_j^T G p_j < 0$.
Step 2: Determine α such that $(p_1 + \alpha p_j)^T G (p_1 + \alpha p_j) = 0$. Return $y = \frac{1}{\sqrt{1+\alpha^2}}(p_1 + \alpha p_j)$.

By repeating this procedure, we can decompose $X = \sum_{i=1}^r p_i p_i^T$ and for each i, $p_i^T G p_i = 0$.

We now show the correctness of this procedure. Firstly, if $p_1^T X p_1 > 0$, then $\sum_{i=1}^r p_i^T G p_i = 0$ implies the existence of a j such that $p_j^T X p_j < 0$. Then, consider the quadratic function

$$(p_1 + \alpha p_j)^T G (p_1 + \alpha p_j) = (p_j^T G p_j)\alpha^2 + 2(p_1^T G p_j)\alpha + p_1^T G p_1. \tag{2.2.31}$$

Since $(p_1^T G p_1)(p_j^T G p_j) \leq 0$, the discriminant of (2.2.31) as a polynomial in α is nonnegative. Therefore, there exists α such that $(p_1 + \alpha p_j)^T G (p_1 + \alpha p_j) = 0$. Let $y = \frac{1}{\sqrt{1+\alpha^2}}(p_1 + \alpha p_j)$, define $u = \frac{1}{\sqrt{1+\alpha^2}}(p_j - \alpha p_1)$, we will have $X - yy^T = uu^T + \sum_{i=2,3,\ldots,r,\ i\neq j} p_i p_i^T$ with its rank equaling $r-1$. So the procedure can successfully return a y that satisfies the requirements.
■

The rank-one decomposition procedure can be trivially implemented to get a similar result for $G \cdot X \geq 0$ case.

Proposition 2.2.2 ([50]) *Given a symmetric matrix G, and a positive semidefinite symmetric matrix X with rank r, then $G \cdot X \geq 0$ if and only if there is a rank-one decomposition $X = \sum_{m=1}^r p_i p_i^T$, such that*

$$p_i^T G p_i \geq 0 \text{ for } i = 1, 2, \ldots, r.$$

■

The rank-one decomposition procedure is a powerful tool to study the strength of semidefinite relaxation. Below are several interesting results obtained by applying the rank-one decomposition procedure.

Lemma 2.2.1 ([50]) *Consider the problem of minimizing a quadratic function subject to a convex quadratic constraint together with a linear inequality constraint:*

$$\begin{array}{ll} \min & q_0(x) := x^T Q_0 x + 2c_0^T x \\ s.t. & q_1(x) := x^T Q_1 x + 2c_1^T x + \beta_1 \leq 0 \\ & q_2(x) := 2c_2^T x + \beta_2 \leq 0. \end{array} \tag{2.2.32}$$

(2.2.32) *could be solved by the following semidefinite formulation:*

$$\begin{array}{ll} \min & M(q_0(\bullet)) \cdot Z \\ s.t. & M(q_1(\bullet)) \cdot Z \leq 0 \\ & Z \begin{pmatrix} \beta_2 \\ c_2 \end{pmatrix} \in S \\ & Z_{11} = 1 \\ & Z \succeq 0, \end{array} \tag{2.2.33}$$

where S is a convex set defined as

$$S := \{ \begin{pmatrix} t_0 \\ t \end{pmatrix} \in \mathbb{R}^{n+1} | t_0 \geq 0,\ t_0^2 \beta_1 + 2 t_0 c_1^T t + t^T Q_1 t \leq 0,\ t_0 \beta_1 + 2 c_1^T t \leq 0 \}.$$

■

Remark 2.2.1 *Due to the difficulty to formulate S,* (2.2.33) *turns out no easier than the original problem, but it is an interesting theoretical result. Because a direct semidefinite relaxation*

$$\begin{array}{ll} \min & M(q_0(\bullet)) \cdot Z \\ s.t. & M(q_1(\bullet)) \cdot Z \leq 0 \\ & c_2^T x + \beta_2 \leq 0 \\ & Z_{11} = 1 \\ & Z (:= \begin{pmatrix} 1 & x^T \\ x & Y \end{pmatrix}) \succeq 0 \end{array} \tag{2.2.34}$$

may not be tight. See the following example.

■

Example 2.2.1 *In* (2.2.32), *take $n = 1$, $q_0(x) = -x^2 + 2x$, $q_1(x) = x^2 - 1$ and $q_2(x) = -x$, then a direct semidefinite relaxation* (2.2.34) *returns an optimal value -1; whereas for this example the minimum is attained at $x = 0$ with an optimal value equaling 0. So a direct semidefinite relaxation* (2.2.34) *is not tight. In fact, this example is a trust region problem with an additional linear inequality constraint. In this example, the linear constraint has cut off the global minimizer of the trust region problem, and left a local minimizer at which the Hessian of the Lagrangian function is not positive semidefinite [35]. As a result, the Lagrangian relaxation, or equivalently, the semidefinite relaxation can never be tight.*

Lemma 2.2.2 ([53]) *Consider the problem of minimizing a quadratic function subject to two quadratic constraints*

$$\begin{array}{rl} \min & q_0(x) := x^T Q_0 x + 2c_0^T x \\ s.t. & q_1(x) := x^T Q_1 x + 2c_1^T x + \beta_1 \leq 0 \\ & q_2(x) := x^T Q_2 x + 2c_2^T x + \beta_2 \leq 0, \end{array} \tag{2.2.35}$$

and its semidefinite relaxation

$$\begin{array}{rl} \min & M(q_0(\bullet)) \cdot Z \\ s.t. & M(q_1(\bullet)) \cdot Z \leq 0 \\ & M(q_2(\bullet)) \cdot Z \leq 0 \\ & Z_{11} = 1 \\ & Z \succeq 0. \end{array} \tag{2.2.36}$$

Suppose (2.2.36) *attains its optimal at* Z^*, *and* $M(q_i(\bullet)) \cdot Z^* < 0$ *for at lease one of* $i = 1, 2$. *Then the relaxation* (2.2.36) *is tight for* (2.2.35). *Moreover, an optimal solution to* (2.2.35) *can be constructed in polynomial time.*

■

Lemma 2.2.3 ([53]) *Consider a homogeneous* ***QCQP*** *with* $m = 2$

$$\begin{array}{rl} \min & q_0(x) := x^T Q_0 x \\ s.t. & q_1(x) := x^T Q_1 x \leq 1 \\ & q_2(x) := x^T Q_2 x \leq 1. \end{array} \tag{2.2.37}$$

Then its semidefinite relaxation

$$\begin{array}{rl} \min & Q_0 \cdot Z \\ s.t. & Q_1 \cdot Z \leq 1 \\ & Q_2 \cdot Z \leq 1 \\ & Z_{11} = 1 \\ & Z \succeq 0 \end{array} \tag{2.2.38}$$

is tight. Moreover, an optimal solution to (2.2.37) *can be constructed in polynomial time.*

■

We have seen that the rank reduction procedure often help us to recognize tight semidefinite relaxations. However, it is still an open problem to determine the minimal rank optimal solution to a given $\boldsymbol{SDP}$. To be less ambitious, the following theorem shows that it is likely to find a sufficiently low rank solution in an approximate affine space.

Theorem 2.2.2 ([46]) *Let $A_1, A_2, \ldots, A_m \in \mathcal{S}^n_+$, and let $b_1, b_2, \ldots, b_m \geq 0$. Suppose that $\exists Z \succeq 0$ such that $A_i \cdot Z = b_i$ for $i = 1, 2, \ldots, m$. Let $r = \min\{\sqrt{2m}, n\}$. Then, for any $d \geq 1$, $\exists Z_0 \succeq 0$ with $rank(Z_0) \leq d$, such that*

$$\beta(m,n,d)b_i \leq A_i \cdot Z_0 \leq \alpha(m,n,d)b_i \qquad \textit{for} \ \ i = 1, 2, \ldots, m,$$

where:

$$\alpha(m,n,d) = \begin{cases} 1 + \frac{12\ln(4mr)}{d} & \textit{for} 1 \leq d \leq 12\ln(4mr) \\ 1 + \sqrt{\frac{12\ln(4mr)}{d}} & \textit{otherwise} \end{cases}, \tag{2.2.39}$$

$$\beta(m,n,d) = \begin{cases} \frac{1}{5em^{2/d}} & \textit{for } 1 \leq d \leq \frac{2\ln m}{\ln\ln(2m)} \\ \frac{1}{4e\ln^{f(m)/d}(2m)} & \textit{for } \frac{2\ln m}{\ln\ln(2m)} < d \leq 4\ln(4mr) \\ 1 - \sqrt{\frac{4\ln(4mr)}{d}} & \textit{for } d > 4\ln(4mr) \end{cases}, \tag{2.2.40}$$

and $f(m) = \frac{3\ln m}{\ln\ln(2m)}$. Moreover, such an Z_0 can be found in randomized polynomial time.

Proof. Proof to Theorem 2.2.2 is based on probabilistic methods, see [46]. ■

Chapter 3

Theory of SDP Relaxations for QMP

A nonconvex $\boldsymbol{QCQP}$ with a large m is generally considered intractable. To be less ambitious, we are more interested in generating inexpensive and strong relaxation bounds. Either $\boldsymbol{SDPP}$ or $\boldsymbol{SDPD}$ is polynomially solvable via a primal-dual interior point method [36], and will provide a lower bound in polynomial time. However, both are difficult to compute when the problem scale is large. A possible way to improve the computational efficiency is to explore the sparsity of the matrices and use the concepts of *positive semidefinite completion*, as illustrated below.

Definition 3.0.1 ([2, 51, 47]) *For an undirected graph G, a chord is an edge joining two non-consecutive vertices of a cycle. G is said to be chordal if every cycle with length greater than three has a chord.*

Definition 3.0.2 ([47]) *A partial symmetric matrix is a symmetric matrix in which not all of its entries are specified.*

Definition 3.0.3 ([51]) *A partial positive semidefinite matrix is a partial symmetric matrix with each of its fully specified principal submatrices positive semidefinite.*

Definition 3.0.4 ([26, 47]) *A positive semidefinite completion for a partial symmetric matrix X is a fully specified symmetric matrix $\bar{X}$ with $\bar{X} \succeq 0$ and $\bar{X}_{ij} = X_{ij}$ for any specified entries X_{ij} in X.*

For a sparse matrix X, its *sparsity pattern*, which is defined as the set of row/column indices of nonzero entries of X, can be represented with an undirected graph $G = (V, E)$, where $V = \{1, 2, \ldots, n\}$ and $E = \{\{i, j\} : X_{ij} \neq 0, i \neq j\}$ [47]. A partial symmetric matrix can also be treated as a sparse matrix by regarding its unspecified entries as zeros, and hence its sparsity pattern can be represented by a graph G [47]. For a simple undirected graph G, let $S(G)$ denote the set of partial symmetric matrices with sparsity pattern represented by G. Then there is a well-known theorem that establishes the connections between chordal graphs and positive semidefinite completions.

Theorem 3.0.3 ([20, 26]) *Every partial positive semidefinite matrix $X \in S(G)$ has a positive semidefinite completion if and only if G is a chordal graph.*

■

Therefore, for the dual program $\boldsymbol{SDPD}$, if $\begin{pmatrix} \alpha & (c_0 + \sum_{j=1}^m \lambda_j c_j)^T \\ c_0 + \sum_{j=1}^m \lambda_j c_j & Q_0 + \sum_{j=1}^m \lambda_j Q_j \end{pmatrix}$ is sparse and its nonzero entries correspond to a chordal graph, we can simplify the positive semidefinite constraint according to Theorem 3.0.3. In this chapter, we will introduce a particular type of $\boldsymbol{QCQP}$, called *Quadratic Matrix Programming*, whose dual program can be reduced to smaller $\boldsymbol{SDP}$ with this trick.

3.1 Quadratic Matrix Programming

We now study a special class of $\boldsymbol{QCQP}$ that allows for smaller $\boldsymbol{SDP}$ relaxations. In [4], the term *Quadratic Matrix Programming* ($\boldsymbol{QMP}$) is used to define the following type of $\boldsymbol{QCQP}$, which we refer to as $\boldsymbol{QMP_1}$

$$(\boldsymbol{QMP_1}) \qquad \begin{array}{ll} \min & \operatorname{trace}(X^T Q_0 X + 2C_0^T X) \\ \text{s.t.} & \operatorname{trace}(X^T Q_j X + 2C_j^T X) + \beta_j \leq 0, \ j = 1, 2, \ldots, m \\ & X \in \mathcal{M}^{nr}, \end{array}$$

where $Q_j \in \mathcal{S}^n$, $C \in \mathcal{M}^{nr}$.

In this thesis, we will extend the class of problems studied in [4]. We will study another quadratic minimization problem for which all the objective and constraint functions are in the form of $f(X) := \operatorname{trace}(X^T Q X + X P X^T + 2C^T X) + \beta$. We name it as $\boldsymbol{QMP_2}$, which has the general form

$$(\boldsymbol{QMP_2}) \qquad \begin{array}{ll} \min & \operatorname{trace}(X^T Q_0 X + X P_0 X^T + 2C_0^T X) \\ \text{s.t.} & \operatorname{trace}(X^T Q_j X + X P_j X^T + 2C_j X) + \beta_j \leq 0, \ j = 1, 2, \ldots, m \\ & X \in \mathcal{M}^{nr}, \end{array}$$

where $Q_j \in \mathcal{S}^n$, $P_j \in \mathcal{S}^r$, $C_j \in \mathcal{M}^{nr}$.

Both $\boldsymbol{QMP_1}$ and $\boldsymbol{QMP_2}$ are named as Quadratic Matrix Programming ($\boldsymbol{QMP}$) in this thesis.

Note that the trace of the matrix products can be expressed using Kronecker products as follows:

$$\begin{aligned}\text{trace}\,(X^TQX) &= \text{vec}\,(X)^T(I_r \otimes Q)\text{vec}\,(X);\\ \text{trace}\,(XPX^T) &= \text{vec}\,(X)^T(P \otimes I_n)\text{vec}\,(X).\end{aligned}$$

Therefore, $\boldsymbol{QMP}$ can be reformulated in vectorized forms, which implies that $\boldsymbol{QMP}$ is a special class of $\boldsymbol{QCQP}$. However, by exploiting the special structure of $\boldsymbol{QMP}$ problems, we can formulate a semidefinite relaxation for $\boldsymbol{QMP}$ which maintains the matrices rather than vectorizing them. We call the semidefinite relaxation based on the vectorized form of $\boldsymbol{QMP}$ the *Vector-Lifting Semidefinite Relaxation* ($\boldsymbol{VSDR}$); while the new type of semidefinite relaxation is called *Matrix-Lifting Semidefinite Relaxation* ($\boldsymbol{MSDR}$). Later, we will prove that the bounds from $\boldsymbol{MSDR}$ are equal or weaker than those from $\boldsymbol{VSDR}$. And we will focus on the comparison between $\boldsymbol{VSDR}$ and $\boldsymbol{MSDR}$, and provide sufficient conditions under which $\boldsymbol{VSDR}$ and $\boldsymbol{MSDR}$ generate the same bounds.

By the fact $\text{trace}\,(X^TQX + XPX^T) = \text{vec}\,(X)^T(I_r \otimes Q + P \otimes I_n)\text{vec}\,(X)$, $\boldsymbol{QMP_2}$ is no more general than $\boldsymbol{QMP_1}$. However, reformulating with matrix $I_r \otimes Q + P \otimes I_n$ can destroy the sparsity patterns, which will be discussed in later sections. Therefore, $\boldsymbol{QMP_2}$ is interesting to be studied independently.

3.2 QMP_1

We will focus on $\boldsymbol{QMP_1}$ in this section. To distinguish between the vector and matrix formulations, we adopt the notation

$$\begin{aligned}q_j^V(X) &:= \text{vec}\,(X)^T(I_r \otimes Q_j)\text{vec}\,(X) + 2\text{vec}\,(C_j)^T\text{vec}\,(X) + \beta_j,\\ q_j^M(X) &:= \text{trace}\,(X^TQ_jX + 2C_j^TX) + \beta_j.\end{aligned}$$

Now denote

$$\begin{aligned}M(q_j^V(\bullet)) &:= \begin{pmatrix} \beta_j & \text{vec}\,(C_j)^T \\ \text{vec}\,(C_j) & I_r \otimes Q_j \end{pmatrix},\\ M(q_j^M(\bullet)) &:= \begin{pmatrix} \frac{\beta_j}{r}I_r & C_j^T \\ C_j & Q_j \end{pmatrix}.\end{aligned}$$

We now propose two different $\boldsymbol{SDP}$ relaxations for $\boldsymbol{QMP_1}$, i.e. $\boldsymbol{VSDR-1}$ and $\boldsymbol{MSDR-1}$:

$$(\boldsymbol{VSDR-1}) \quad \begin{array}{rll} \mu_3^* := & \min & M(q_0^V(\bullet)) \cdot Z_V \\ & \text{s.t.} & M(q_j^V(\bullet)) \cdot Z_V \leq 0, \ j = 1, 2, \ldots, m \\ & & (Z_V)_{1,1} = 1 \\ & & Z_V \left(:= \begin{pmatrix} 1 & \text{vec}\,(X)^T \\ \text{vec}\,(X) & Y_V \end{pmatrix} \right) \succeq 0; \end{array}$$

$$(\boldsymbol{MSDR-1}) \quad \begin{array}{rll} \mu_4^* := & \min & M(q_0^M(\bullet)) \cdot Z_M \\ & \text{s.t.} & M(q_j^M(\bullet)) \cdot Z_M \leq 0, \ j = 1, 2, \ldots, m \\ & & (Z_M)_{1:r,1:r} = I_r \\ & & Z_M \left(:= \begin{pmatrix} I_r & X^T \\ X & Y_M \end{pmatrix} \right) \succeq 0. \end{array}$$

$\boldsymbol{VSDR-1}$ is relaxing the quadratic equality constraint $Y_V = \text{vec}\,(X)\text{vec}\,(X)^T$ to $Y_V \succeq \text{vec}\,(X)\text{vec}\,(X)^T$. This can be formulated as a conic constraint $Z_V = \begin{pmatrix} 1 & \text{vec}\,(X)^T \\ \text{vec}\,(X) & Y_V \end{pmatrix} \succeq 0$; while $\boldsymbol{MSDR-1}$ is relaxing $Y_M = XX^T$ to the conic constraint $Z_M = \begin{pmatrix} I_r & X^T \\ X & Y_M \end{pmatrix} \succeq 0$.

Both $\boldsymbol{VSDR-1}$ and $\boldsymbol{MSDR-1}$ can be interpreted from the perspective of Lagrange relaxation. By the equivalence between the Lagrangian relaxation and semidefinite relaxation for $\boldsymbol{QCQP}$ (Theorem 1.4.2), we may deduce that $\boldsymbol{VSDR-1}$ generates a bound as strong as Lagrangian relaxation under condition (1.4.11), which can be illustrated as

$$\begin{array}{rcl} \mu_{LR}^* & = & \max\limits_{\lambda \in \mathbb{R}_+^m} \min\limits_{X \in \mathcal{M}^{nr}} \text{vec}\,(X)^T (I_r \otimes Q_\lambda) \text{vec}\,(X) + 2x_0 (\text{vec}\,(C_\lambda))^T \text{vec}\,(X) + \beta_\lambda \\ & = & \max\limits_{\lambda \in \mathbb{R}_+^m} \min\limits_{X \in \mathcal{M}^{nr}, x_0^2 = 1} \text{vec}\,(X)^T (I_r \otimes Q_\lambda) \text{vec}\,(X) + 2x_0 (\text{vec}\,(C_\lambda))^T \text{vec}\,(X) + \beta_\lambda \\ & = & \max\limits_{\lambda \in \mathbb{R}_+^m, \alpha \in \mathbb{R}} \min\limits_{X \in \mathcal{M}^{nr}, x_0 \in \mathbb{R}} \text{vec}\,(X)^T (I_r \otimes Q_\lambda) \text{vec}\,(X) + 2x_0 (\text{vec}\,(C_\lambda))^T \text{vec}\,(X) \\ & & \qquad\qquad + \beta_\lambda + \alpha(x_0^2 - 1) \\ & = & \max \ -\alpha + \sum_{j=1}^m \lambda_j \beta_j \\ & & \text{s.t.} \begin{pmatrix} \alpha & (C_0 + \sum_{j=1}^m \lambda_j C_j)^T \\ \text{vec}\,(C_0 + \sum_{j=1}^m \lambda_j C_j) & I_r \otimes (Q_0 + \sum_{j=1}^m \lambda_j Q_j) \end{pmatrix} \succeq 0 \\ & & \qquad \lambda_j \geq 0, \ j = 1, 2, \ldots, m, \end{array}$$

where μ_{LR}^* denotes the Lagrangian relaxation bound, $Q_\lambda, C_\lambda, \beta_\lambda$ follows the definition introduced in (1.3.6). It is easy to see the right hand side of the last equality is the dual of $\boldsymbol{VSDR-1}$. Therefore, under condition (1.4.11), $\mu_{LR}^* = \mu_3^*$.

We could also explain $\boldsymbol{MSDR-1}$ from the perspective of Lagrangian relaxation

$$
\begin{array}{rcl}
\mu^*_{LR} & = & \max\limits_{\lambda\in\mathbb{R}^M_+} \min\limits_{X\in\mathcal{M}^{nr}} \operatorname{trace}(X^TQ_\lambda X + 2C_\lambda^T X) + \beta_\lambda \\
& = & \max\limits_{\lambda\in\mathbb{R}^M_+} \min\limits_{X\in\mathcal{M}^{nr}, X_0X_0^T = I_r} \operatorname{trace}(X^TQ_\lambda X + 2X_0^T C_\lambda^T X) + \beta_\lambda \\
& \geq & \max\limits_{\lambda\in\mathbb{R}^M_+, T\in\mathcal{S}^r} \min\limits_{X, X_0\in\mathcal{M}^{nr}} \operatorname{trace}(X^TQ_\lambda X + 2X_0^T C_\lambda^T X + T(X_0X_0^T - I_r)) + \beta_\lambda \\
& = & \max \quad -trT + \sum_{j=1}^m \lambda_j\beta_j \\
& & \text{s.t.} \begin{pmatrix} T & (C_0 + \sum_{j=1}^m \lambda_j C_j)^T \\ C_0 + \sum_{j=1}^m \lambda_j C_j & Q_0 + \sum_{j=1}^m \lambda_j Q_j \end{pmatrix} \succeq 0 \\
& & \qquad \lambda_j \geq 0,\ j = 1, 2, \ldots, m \\
& & \qquad T \in \mathcal{S}^n.
\end{array}
\tag{3.2.41}
$$

The right-hand side of the last equality is the conic dual of $\boldsymbol{MSDR-1}$. Therefore, $\boldsymbol{MSDR-1}$ can never be stronger than the Lagrangian relaxation or $\boldsymbol{VSDR-1}$, i.e., $\mu_4^* \leq \mu_{LR}^* = \mu_3^*$. However, the following theorem indicates that under condition (1.4.11), the inequality in (3.2.41) can be replaced with an equality.

Theorem 3.2.1 ([4]) *For $\boldsymbol{QMP_1}$, if the equivalent generalized Slater condition (1.4.11) holds, then the optimal values of $\boldsymbol{MSDR-1}$ and $\boldsymbol{VSDR-1}$ coincide, i.e., $\mu_3^* = \mu_4^*$.*

■

The proof to Theorem 3.2.1 in [4] needs the following theorem.

Theorem 3.2.2 *If the minimum of a $\boldsymbol{QMP_1}$ is attainable and its total number of constraints (equality or inequality) does not exceed r, then $\boldsymbol{MSDR-1}$ solves $\boldsymbol{QMP_1}$ exactly.*

Proof. Apply Theorem 2.2.1 to deduce that $\boldsymbol{MSDR-1}$ has an optimal solution Z_M^* with $\operatorname{rank}(Z_M^*) \leq r$. For details, see [4]. ■

We will provide a similar but even stronger result than Theorem 3.2.1 by using Theorem 3.0.3. To be different from the proof in [4], our proof is completely based on the analysis to the primal program and hence does not need the equivalent generalized Slater condition (1.4.11). And we will also prove that the feasible sets of the two relaxations have the same projection on the X part.

Lemma 3.2.1 *For $\boldsymbol{QMP_1}$, denote the columns of X with x_j, $j = 1, 2, \ldots, r$, and let $\boldsymbol{VSDR{-}1'}$ denote the semidefinite relaxation*

$$(\boldsymbol{VSDR{-}1'}) \qquad \begin{array}{rl} \mu_5^* := \min & Q_0 \cdot \sum_{j=1}^r Y_{jj} + 2C_0 \cdot X \\ s.t. & Q_j \cdot \sum_{j=1}^r Y_{jj} + 2C_j \cdot X + \beta_j \le 0, \; j = 1, 2, \ldots, m \\ & Z_{jj} = \begin{pmatrix} 1 & x_j^T \\ x_j & Y_{jj} \end{pmatrix} \succeq 0, \; j = 1, 2, \ldots, m. \end{array}$$

If both $\boldsymbol{VSDR{-}1}$ and $\boldsymbol{VSDR{-}1'}$ attain their minima, then their optimal values coincide, i.e., $\mu_3^* = \mu_5^*$.

Proof.

1. $\mu_3^* \le \mu_5^*$:

   Suppose $\boldsymbol{VSDR{-}1'}$ has an optimal solution $Z_{jj}^* = \begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix}$ $(j = 1, 2, \ldots, r)$, we may construct a partial symmetric matrix

$$Z_V^* = \begin{pmatrix} 1 & (x_1^*)^T & (x_2^*)^T & \ldots & (x_r^*)^T \\ x_1^* & Y_{11}^* & ? & ? & ? \\ x_2^* & ? & Y_{22}^* & ? & ? \\ \vdots & ? & ? & \ddots & ? \\ x_r^* & ? & ? & ? & Y_{rr}^* \end{pmatrix}.$$

   By observation, the unspecified entries of Z_V^* (marked by ?) are not involved in any of the constraint and objective functions of $\boldsymbol{VSDR{-}1}$. In other words, we can assign any values to those positions without changing the feasibility and the objective value. Because Z_{jj}^* $(j = 1, 2, \ldots, r)$ is feasible to $\boldsymbol{VSDR{-}1'}$, we have $\begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix} \succeq 0$ for each $j = 1, 2, \ldots, r$. So all the principal submatrices of Z_V^* are positive semidefinite, and hence Z_V^* is a *partial positive semidefinite matrix* according to Definition 3.0.3. It is clear that the *arrow* sparsity pattern of Z_V^* corresponds to a chordal graph. By Theorem 3.0.3, Z_V^* has a semidefinite completion $\bar{Z}_V^*$ which is feasible to $\boldsymbol{VSDR{-}1}$, so

$$\mu_3^* \le M(q_0^V(\bullet)) \cdot \bar{Z}_V^*. \qquad (3.2.42)$$

Let $X^* = (x_1^* \quad x_2^* \quad \cdots \quad x_r^*)$, then

$$M(q_0^V(\bullet)) \cdot \bar{Z}_V^* = Q_0 \cdot \sum_{j=1}^{r} Y_{jj}^* + 2C_0 \cdot X^*. \tag{3.2.43}$$

Because $Z_{jj}^* = \begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix}$ $(j = 1, 2, \ldots, r)$ minimizes $\boldsymbol{VSDR{-}1'}$, we got $Q_0 \cdot \sum_{j=1}^{r} Y_{jj}^* + 2C_0 \cdot X^* = \mu_5^*$. Together with (3.2.42) and (3.2.43), we conclude $\mu_3^* \leq \mu_5^*$.

2. $\mu_3^* \geq \mu_5^*$:

   Suppose

$$Z_V^* = \begin{pmatrix} 1 & (x_1^*)^T & (x_2^*)^T & \ldots & (x_r^*)^T \\ x_1^* & Y_{11}^* & Y_{12}^* & \ldots & Y_{1r}^* \\ x_2^* & Y_{21}^* & Y_{22}^* & \ldots & Y_{2r}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_r^* & Y_{r1}^* & Y_{r2}^* & \ldots & Y_{rr}^* \end{pmatrix} \succeq 0$$

is an optimal solution to $\boldsymbol{VSDR{-}1}$. Now we construct $Z_{jj}^* := \begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix}$, $j = 1, 2, \ldots, r$. Because each Z_{jj}^* is a principal submatrix of the positive semidefinite matrix Z_V^*, we have $Z_{jj}^* \succeq 0$ for each $j = 1, 2, \ldots, r$. Because Z_V^* is feasible to $\boldsymbol{VSDR{-}1}$, we have

$$M(q_i^V(\bullet)) \cdot Z_V^* \leq 0, \qquad i = 1, 2, \ldots, m. \tag{3.2.44}$$

If $X^* = (x_1^* \quad x_2^* \quad \cdots \quad x_r^*)$, it is easy to check

$$Q_i \cdot \sum_{j=1}^{r} Y_{jj}^* + 2C_i \cdot X^* + \beta_i = M(q_i^V(\bullet)) \cdot Z_V^*, \qquad i = 1, 2, \ldots, m. \tag{3.2.45}$$

So the m constraints of $\boldsymbol{VSDR{-}1'}$

$$Q_i \cdot \sum_{j=1}^{r} Y_{jj}^* + 2C_i \cdot X^* + \beta_i \leq 0, \qquad i = 1, 2, \ldots, m \tag{3.2.46}$$

are satisfied by $Z_{jj}^* = \begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix}$ $(j = 1, 2, \ldots, r)$. Because μ_5^* denotes the optimal value of $\boldsymbol{VSDR{-}1'}$, we have

$$\mu_5^* \leq Q_0 \cdot \sum_{i=1}^{r} Y_{ii}^* + 2C_0 \cdot X^*. \tag{3.2.47}$$

Because Z_V^* minimizes $\boldsymbol{VSDR{-}1}$, we have

$$\mu_3^* = M(q_0^V(\bullet)) \cdot Z_V^*. \tag{3.2.48}$$

Also, by inspection,

$$M(q_0^V(\bullet)) \cdot Z_V^* = \cdot Q_0 \cdot \sum_{i=1}^{r} Y_{ii}^* + 2C_0 \cdot X^*. \tag{3.2.49}$$

Therefore, (3.2.47), (3.2.48) and (3.2.49) together lead to $\mu_3^* \geq \mu_5^*$.

■

Lemma 3.2.1 shows that $\boldsymbol{VSDR{-}1'}$ is as strong as $\boldsymbol{VSDR{-}1}$, although it includes fewer variables. The following lemma shows that the $\boldsymbol{VSDR{-}1'}$ can be further reduced to $\boldsymbol{MSDR{-}1}$ without weakening the bounds.

Lemma 3.2.2 *If both $\boldsymbol{MSDR{-}1}$ and $\boldsymbol{VSDR{-}1'}$ attain their minima, then their optimal values coincide, i.e., $\mu_4^* = \mu_5^*$.*

Proof.

1. $\mu_4^* \leq \mu_5^*$:

   Suppose $\boldsymbol{VSDR{-}1'}$ has an optimal solution $Z_{jj}^* = \begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix}$ $(j = 1, 2, \cdots, m)$. Let $X^* = (\, x_1^* \quad x_2^* \quad \cdots \quad x_r^* \,)$, and $Y_M^* = \sum_{j=1}^{r} Y_{jj}^*$. Now we construct $Z_M^* := \begin{pmatrix} I_r & (X^*)^T \\ X^* & Y_M^* \end{pmatrix}$. Then by $Z_{jj}^* \succeq 0$ for each $j = 1, 2, \ldots, r$, we have $Z_M^* \succeq 0$, and

$$M(q_0^M(\bullet)) \cdot Z_M^* = Q_0 \cdot \sum_{i=1}^{r} Y_{ii}^* + 2C_0 \cdot X^*, \tag{3.2.50}$$

$$M(q_j^M(\bullet)) \cdot Z_M^* = Q_j \cdot \sum_{i=1}^{r} Y_{ii}^* + 2C_j \cdot X^* + \beta_j, \qquad j = 1, \ldots, m. \tag{3.2.51}$$

Because $Z_{jj}^* = \begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix}$ $(j = 1, 2, \ldots, m)$ minimize $\boldsymbol{VSDR-1'}$, we have

$$Q_0 \cdot \sum_{i=1}^{r} Y_{ii}^* + 2C_0 \cdot X^* = \mu_5^*, \tag{3.2.52}$$

and

$$Q_j \cdot \sum_{i=1}^{r} Y_{ii}^* + 2C_j \cdot X^* + \beta_j \leq 0, \qquad j = 1, \ldots, m. \tag{3.2.53}$$

By (3.2.51) and (3.2.53), we deduce

$$M(q_j^M(\bullet)) \cdot Z_M^* \leq 0, \qquad j = 1, \ldots, m. \tag{3.2.54}$$

(3.2.54) and $Z_M^* \succeq 0$ imply that Z_M^* is a feasible solution to $\boldsymbol{MSDR-1}$, whose minimal objective value equals μ_4^*. So

$$\mu_4^* \leq M(q_0^M(\bullet)) \cdot Z_M^*. \tag{3.2.55}$$

Then by (3.2.50), (3.2.52) and (3.2.55), we conclude $\mu_4^* \leq \mu_5^*$.

2. $\mu_5^* \leq \mu_4^*$:

   Suppose $Z_M^* = \begin{pmatrix} I_r & (X^*)^T \\ X^* & Y_M^* \end{pmatrix} \succeq 0$ minimizes $\boldsymbol{MSDR-1}$, and $X^* = (\, x_1^* \quad x_2^* \quad \cdots \quad x_r^* \,)$. Then let $Y_{ii}^* = x_i^*(x_i^*)^T$ for $i = 1, 2, \ldots, r-1$, and let $Y_{rr} = x_r^*(x_r^*)^T + (Y_M^* - X^*(X^*)^T)$. As a result, $Y_{ii}^* \succeq x_i^*(x_i^*)^T$ for $i = 1, 2, \ldots, r$, and $\sum_{i=1}^{r} Y_{ii}^* = Y_M^*$. So by constructing $Z_{jj}^* = \begin{pmatrix} 1 & (x_j^*)^T \\ x_j^* & Y_{jj}^* \end{pmatrix}$ $(j = 1, 2, \cdots, r)$, (3.2.50) and (3.2.51) hold.

   Because Z_M^* minimizes $\boldsymbol{MSDR-1}$, we deduce (3.2.54) holds and

$$\mu_4^* = M(q_0^M(\bullet)) \cdot Z_M^*. \tag{3.2.56}$$

By (3.2.51) and (3.2.54), we can deduce (3.2.53), i.e., Z_{jj}^* ($j = 1, 2, \ldots, r$) feasible to $\boldsymbol{VSDR\!-\!1'}$. Because the minimal objective value of $\boldsymbol{VSDR\!-\!1'}$ equals μ_5^*, we have

$$\mu_5^* \leq Q_0 \cdot \sum_{i=1}^{r} Y_{ii}^* + 2C_0 \cdot X^*. \tag{3.2.57}$$

Now (3.2.50), (3.2.56) and (3.2.57) together imply $\mu_5^* \leq \mu_4^*$.

■

As a conclusion for this section, we propose the following theorem, which is a stronger result than Theorem 3.2.1.

Theorem 3.2.3 *The two programs,* $\boldsymbol{VSDR\!-\!1}$ *and* $\boldsymbol{MSDR\!-\!1}$*, either both reach the same optimal value, i.e.,* $\mu_4^* = \mu_5^*$*, or neither of them attains its minima.*

Proof. According to Lemma 3.2.1 and Lemma 3.2.2, given any $Z_V^* = \begin{pmatrix} 1 & \mathrm{vec}\,(X)^T \\ \mathrm{vec}\,(X) & Y_V \end{pmatrix}$ feasible to $\boldsymbol{VSDR\!-\!1}$, we can construct a $Z_M^* = \begin{pmatrix} I_r & X^T \\ X & Y_M \end{pmatrix}$ that is feasible to $\boldsymbol{MSDR\!-\!1}$ and reaches the same objective value. And the converse is also true. Therefore, the primal optimum of $\boldsymbol{VSDR\!-\!1}$ is attained if and only if the primal optimum of $\boldsymbol{MSDR\!-\!1}$ is attained. Furthermore, if they both attain their minima, then their objective values coincide according to Lemma 3.2.1 and Lemma 3.2.2. ■

3.3 QMP_2

In this section, we will move to the second type of Quadratic Matrix Programming, i.e., $\boldsymbol{QMP_2}$. We propose the Vector-Lifting Semidefinite Relaxation ($\boldsymbol{VSDR\!-\!2}$) for

$\boldsymbol{QMP_2}$ as

$$(\boldsymbol{VSDR-2}) \qquad \begin{array}{rll} \mu_6^* := & \min & \begin{pmatrix} 0 & \operatorname{vec}(C_0)^T \\ \operatorname{vec}(C_0) & I_r \otimes Q_0 + P_0 \otimes I_n \end{pmatrix} \cdot Z_V \\ & \text{s.t.} & \begin{pmatrix} \beta_j & \operatorname{vec}(C_j)^T \\ \operatorname{vec}(C_j) & I_r \otimes Q_j + P_j \otimes I_n \end{pmatrix} \cdot Z_V \le 0, \ j = 1, 2, \ldots, m \\ & & Z_V (= \begin{pmatrix} 1 & \operatorname{vec}(X)^T \\ \operatorname{vec}(X) & Y_V \end{pmatrix}) \succeq 0. \end{array}$$

We also propose the Matrix-Lifting Semidefinite Relaxation ($\boldsymbol{MSDR-2}$) for $\boldsymbol{QMP_2}$ as

$$(\boldsymbol{MSDR-2}) \qquad \begin{array}{rll} \mu_7^* := & \min & Y_1 \cdot Q_0 + Y_2 \cdot P_0 + 2C_0 \cdot X \\ & \text{s.t.} & Y_1 \cdot Q_j + Y_2 \cdot P_j + 2C_j \cdot X + \beta_j \le 0, \ j = 1, 2, \ldots, m \\ & & Z_1 (= \begin{pmatrix} I_r & X^T \\ X & Y_1 \end{pmatrix}) \succeq 0 \\ & & Z_2 (= \begin{pmatrix} I_n & X \\ X^T & Y_2 \end{pmatrix}) \succeq 0 \\ & & \operatorname{trace} Y_1 = \operatorname{trace} Y_2. \end{array}$$

Now we conclude the comparison for the numbers of variables and constraints of $\boldsymbol{VSDR-1}$ and $\boldsymbol{MSDR-1}$, $\boldsymbol{VSDR-2}$ and $\boldsymbol{MSDR-2}$ in Table 3.1.

Methods	$\boldsymbol{VSDR-1}$	$\boldsymbol{MSDR-1}$	$\boldsymbol{VSDR-2}$	$\boldsymbol{MSDR-2}$
Variables	$(nr+1)^2$	$(n+r)^2$	$(nr+1)^2$	$2(n+r)^2$
Constraints	$m+1$	$m+0.5r^2$	$m+1$	$m+1+0.5(n+r)^2$

Table 3.1: Comparison of Costs for Different Semidefinite Relaxations

Because most problems can be reformulated as a $\boldsymbol{QMP}$ with m as the same order of nr, it is often cheaper to compute an $\boldsymbol{MSDR-2}$ rather than to compute a $\boldsymbol{VSDR-2}$. However, the bound from $\boldsymbol{MSDR-2}$ turns out never stronger than that from $\boldsymbol{VSDR-2}$, because for any $Z_V = \begin{pmatrix} 1 & \operatorname{vec}(X)^T \\ \operatorname{vec}(X) & Y_V \end{pmatrix}$ that minimizes $\boldsymbol{VSDR-2}$, by constructing $Y_1 = \mathrm{o}^0\mathrm{diag}(Y_V)$, $Y_2 = \mathrm{b}^0\mathrm{diag}(Y_V)$ ($\mathrm{o}^0\mathrm{diag}$, $\mathrm{b}^0\mathrm{diag}$ defined in (1.3.5)), it is easy to check that $Z_1 = \begin{pmatrix} I_r & X^T \\ X & Y_1 \end{pmatrix}$, $Z_2 = \begin{pmatrix} I_n & X \\ X^T & Y_2 \end{pmatrix}$ are feasible to $\boldsymbol{MSDR-2}$ and reaches the same objective value μ_6^*. So $\mu_6^* \ge \mu_7^*$ holds consistently. We can also interpret the fact $\mu_6^* \ge \mu_7^*$ in another way. By the fact that Z_1, Z_2 can be constructed with the entries of

Z_V (using $\mathrm{o}^0\mathrm{diag}$, $\mathrm{b}^0\mathrm{diag}$), $\boldsymbol{MSDR-2}$ actually only restricts the sum of some principal submatrices of Z_V to be positive semidefinite; while in $\boldsymbol{VSDR-2}$, the whole matrix Z_V is restricted to stay positive semidefinite. So the constraints of $\boldsymbol{MSDR-2}$ are not as strong as those of $\boldsymbol{VSDR-2}$.

Now the question arises. Will $\boldsymbol{VSDR-2}$ be strictly stronger than $\boldsymbol{MSDR-2}$ for some instances? For the $\boldsymbol{QMP_1}$ case, the answer is NO, so it encourages us to use the cheaper $\boldsymbol{MSDR-1}$ formulation instead of $\boldsymbol{VSDR-1}$ whenever we encounter a $\boldsymbol{QMP_1}$. However, the situation is different for the $\boldsymbol{QMP_2}$. By a close inspection, the entries of Y_V involved in the operators $\mathrm{o}^0\mathrm{diag}(\bullet)$, $\mathrm{b}^0\mathrm{diag}(\bullet)$ will form a partial semi-definite matrix whose sparsity pattern is not chordal and hence they do not necessarily admit a semidefinite completion. Therefore, for given $Y_1,\ Y_2$ feasible to $\boldsymbol{MSDR-2}$, there is no guarantee to find a $Z_V = \begin{pmatrix} 1 & \mathrm{vec}(X)^T \\ \mathrm{vec}(X) & Y_V \end{pmatrix} \succeq 0$ with $\mathrm{o}^0\mathrm{diag}(Y_V) = Y_1$ and $\mathrm{b}^0\mathrm{diag}(Y_V) = Y_2$. In later sections, we will see that $\boldsymbol{MSDR-2}$ is as strong as $\boldsymbol{VSDR-2}$ only under certain conditions.

3.3.1 VSDR-2 V.S. MSDR-2

Due to the computational advantage of $\boldsymbol{MSDR-2}$, we are motivated to study the conditions under which $\boldsymbol{MSDR-2}$ is as strong as $\boldsymbol{VSDR-2}$. We come to study the dual programs of $\boldsymbol{VSDR-2}$ and $\boldsymbol{MSDR-2}$, i.e., $\boldsymbol{DM-2}$ and $\boldsymbol{DV-2}$, because the structures of the dual programs are easier to explore. The dual of the $\boldsymbol{VSDR-2}$ is

$$(\boldsymbol{DV-2}) \qquad \begin{array}{rll} \mu_8^* := & \max & \beta_\lambda - \alpha \\ & \text{s.t.} & \begin{pmatrix} \alpha & \mathrm{vec}(C_\lambda)^T \\ \mathrm{vec}(C_\lambda) & I_r \otimes Q_\lambda + P_\lambda \otimes I_n \end{pmatrix} \succeq 0 \\ & & \lambda_j \geq 0,\ j = 1, 2, \ldots, m. \end{array}$$

And the dual of $\boldsymbol{MSDR-2}$ is

$$(\boldsymbol{DM-2}) \qquad \begin{array}{rll} \mu_9^* := & \max & \beta_\lambda - \mathrm{trace}\, S_1 - \mathrm{trace}\, S_2 \\ & \text{s.t.} & \begin{pmatrix} S_1 & R_1^T \\ R_1 & Q_\lambda - tI_n \end{pmatrix} \succeq 0 \\ & & \begin{pmatrix} S_2 & R_2 \\ R_2^T & P_\lambda + tI_r \end{pmatrix} \succeq 0 \\ & & R_1 + R_2 = C_\lambda \\ & & \lambda_j \geq 0,\ j = 1, 2, \ldots, m \\ & & S_1 \in \mathcal{S}^r,\ S_2 \in \mathcal{S}^n,\ R_1, R_2 \in \mathcal{M}^{nr},\ t \in R, \end{array}$$

where $\beta_\lambda,\ C_\lambda,\ Q_\lambda\ , P_\lambda$ follow the definitions in (1.3.6). The following lemma provides a key observation for the connections between the two dual programs.

Lemma 3.3.1 *Let $P \in \mathcal{S}^r,\ Q \in \mathcal{S}^n$, and I_r, I_n denote the identity matrices in $\mathcal{S}^r, \mathcal{S}^n$, respectively. Then*

$$P \otimes I_n + I_r \otimes Q \succeq (\succ) 0,$$

if and only if

$$P + tI_r \succeq (\succ) 0,\ Q - tI_n \succeq 0, \qquad \text{for some } t \in R.$$

Proof.

1. *Necessity*

   Let the symmetric matrix P, Q have spectral decomposition

   $$P = U\Sigma_P U^T, Q = V\Sigma_Q V^T,$$

   where

   $$\Sigma_P = \mathrm{Diag}\,(\lambda_1(P), \lambda_2(P), \ldots, \lambda_r(P)),$$

   $$\Sigma_Q = \mathrm{Diag}\,(\lambda_1(Q),\ \lambda_2(Q), \ldots, \lambda_n(Q)).$$

   So

   $$P \otimes I_n + I_r \otimes Q = (U \otimes V)(\Sigma_P \otimes I_n + I_r \otimes \Sigma_Q)(U \otimes V)^T \succeq (\succ) 0 \tag{3.3.58}$$

   implies the diagonal matrix, $\Sigma_P \otimes I_n + I_r \otimes \Sigma_Q \succeq (\succ) 0$. In other words,

   $$\forall i = 1, 2, \ldots r,\ \ \forall j = 1, 2, \ldots n, \qquad \lambda_i(P) + \lambda_j(Q) \geq (>) 0. \tag{3.3.59}$$

   Therefore,

   $$\min_i \lambda_i(P) + \min_j \lambda_j(Q) \geq (>) 0. \tag{3.3.60}$$

   Now with $t = \min_j \lambda_j(Q)$, we have $Q - tI \succeq 0$ and $P + tI \succeq (\succ) 0$.

2. *Sufficiency*

   We still use the spectral decomposition and deduce $P \otimes I_n + I_m \otimes Q \succeq (\succ) 0$ if (3.3.60) holds. Then by $P + tI_r \succeq (\succ) 0$, we know $\min_i \lambda_i(P) \geq (>) -t$; also by $Q - tI_n \succeq 0$, we get $\min_j \lambda_j(Q) \geq t$. So $\min_i \lambda_i(P) + \min_j \lambda_j(P) \geq (>) -t + t = 0$, which leads to $P \otimes I_n + I_m \otimes Q \succeq (\succ) 0$.

■

Remark 3.3.1 *By the equivalence between* (1.4.11) *and* (1.4.10)*, we know the generalized Slater condition [25] holds for* $\boldsymbol{DV-2}$ *if and only if*

$$\exists \lambda \in \mathbb{R}^m_+, \qquad s.t. \ (\sum_{j=1}^m \lambda_j P_j + P_0) \otimes I_n + I_r \otimes (\sum_{j=1}^m \lambda_j Q_j + Q_0) \succ 0. \tag{3.3.61}$$

As a result of Lemma 3.3.1, (3.3.61) *holds if and only if*

$$\exists t \in \mathbb{R}, \ \lambda \in \mathbb{R}^m_+, \quad s.t. \ \sum_{j=1}^m \lambda_j P_j + P_0 + tI_n \succ 0, \ \sum_{j=1}^m \lambda_j Q_j + Q_0 - tI_r \succ 0, \tag{3.3.62}$$

which is equivalent with the generalized Slater condition for $\boldsymbol{DM-2}$ *. So by classical convex analysis theory [45], under condition* (3.3.62)*, both* $\boldsymbol{VSDR-2}$ *and* $\boldsymbol{MSDR-2}$ *attain their minima and their optimal values coincide with their dual programs respectively, i.e.,*

$$\mu_6^* = \mu_8^*, \qquad \mu_7^* = \mu_9^*.$$

Therefore, under condition (3.3.62)*, to prove that* $\boldsymbol{VSDR-2}$ *and* $\boldsymbol{MSDR-2}$ *generate the same bound (i.e.,* $\mu_6^* = \mu_7^*$*), it is enough to show that the optimal values of* $\boldsymbol{DV-2}$ *and* $\boldsymbol{DM-2}$ *coincide (i.e.,* $\mu_8^* = \mu_9^*$*).*

■

Now we come to prove the main results, i.e., Theorem 3.3.1. Please recall the definition of the Moore-Penrose inverse (†) of a singular symmetric matrix in section 1.3. We have to prove several useful propositions and lemmas before proving Theorem 3.3.1.

Proposition 3.3.1 ([4]) *A quadratic mapping* $f(x) = x^T Qx + 2c^T x + \beta$ *is nonnegative for any* $x \in R^n$ *if and only if the matrix* $\begin{pmatrix} \beta & c^T \\ c & Q \end{pmatrix} \succeq 0$.

■

Proposition 3.3.2 *Suppose* $\boldsymbol{DV-2}$ *is feasible, and its maximum is attained at* $\lambda^* \in \mathcal{R}_+^m$, *then*

$$\mu_8^* = -\mathrm{vec}\,(C_{\lambda^*})^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*}. \tag{3.3.63}$$

Proof. Due to Proposition 3.3.1,

$$\begin{pmatrix} \alpha & \mathrm{vec}\,(C_{\lambda^*})^T \\ \mathrm{vec}\,(C_{\lambda^*}) & I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n \end{pmatrix} \succeq 0 \tag{3.3.64}$$

if and only if

$$\forall X \in \mathcal{M}^{nr},\ \mathrm{vec}\,(X)^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)\mathrm{vec}\,(X) + 2\mathrm{vec}\,(C_{\lambda^*})^T \mathrm{vec}\,(X) + \alpha \geq 0.$$

Therefore, α satisfies (3.3.64) if and only if $-\alpha \leq \min_{X \in \mathcal{M}^{nr}} \mathrm{vec}\,(X)^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)\mathrm{vec}\,(X) + 2\mathrm{vec}\,(C_{\lambda^*})^T \mathrm{vec}\,(X)$. So we can deduce

$$\begin{array}{rll} & \min\limits_{X \in \mathcal{M}^{nr}} & \mathrm{vec}\,(X)^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)\mathrm{vec}\,(X) + 2\mathrm{vec}\,(C_{\lambda^*})^T \mathrm{vec}\,(X) \\ = & \max\limits_{\alpha \in \mathbb{R}} & -\alpha \\ & \text{s.t.} & \begin{pmatrix} \alpha & \mathrm{vec}\,(C_{\lambda^*})^T \\ \mathrm{vec}\,(C_{\lambda^*}) & I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n \end{pmatrix} \succeq 0. \end{array} \tag{3.3.65}$$

The feasibility of λ^* to $\boldsymbol{DV-2}$ implies $\mathrm{vec}\,(C_{\lambda^*}) \in \mathcal{R}(I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)$, so

$$\begin{array}{l} \min\limits_{X \in \mathcal{M}^{nr}} \mathrm{vec}\,(X)^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)\mathrm{vec}\,(X) + 2\mathrm{vec}\,(C_{\lambda})^T \mathrm{vec}\,(X) \\ \qquad = -\mathrm{vec}\,(C_{\lambda^*})^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}). \end{array} \tag{3.3.66}$$

Since $\boldsymbol{DV-2}$ is maximized at $\lambda^* \in \mathbb{R}_+^m$, by fixing $\lambda = \lambda^*$ in $\boldsymbol{DV-2}$, we have

$$\begin{array}{rll} \mu_8^* & = & \max\limits_{\alpha \in \mathbb{R}} -\alpha + \beta_{\lambda^*} \\ & & \text{s.t.} \begin{pmatrix} \alpha & \mathrm{vec}\,(C_{\lambda^*})^T \\ \mathrm{vec}\,(C_{\lambda^*}) & I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n \end{pmatrix} \succeq 0. \end{array} \tag{3.3.67}$$

Then (3.3.63) follows from (3.3.65), (3.3.66) and (3.3.67). ■

Proposition 3.3.3 *If $A, B \in \mathcal{S}_+^n$ and $A \preceq B$, then $\mathcal{N}(B) \subseteq \mathcal{N}(A)$, and $\mathcal{R}(A) \subseteq \mathcal{R}(B)$.*

Proof. Suppose $v \in \mathcal{N}(B)$, then $B \succeq A \succeq 0$ leads to $0 = v^T B v \geq v^T A v \geq 0$, so $v^T A v = 0$. Since $A \in \mathcal{S}_+^n$, A has a Cholesky decomposition $A = R^T R$. Then $v^T R^T R v = 0$ implies $Rv = 0$, which leads to $Av = R^T(Rv) = 0$. Therefore, $v \in \mathcal{N}(A)$, and hence $\mathcal{N}(B) \subseteq \mathcal{N}(A)$. Because A, B are symmetric matrices, by [23], we have $\mathcal{R}(A) = (\mathcal{N}(A^T))^\perp = (\mathcal{N}(A))^\perp$, and $\mathcal{R}(B) = (\mathcal{N}(B^T))^\perp = (\mathcal{N}(B))^\perp$, so $\mathcal{N}(B) \subseteq \mathcal{N}(A)$ implies $\mathcal{R}(B) = (\mathcal{N}(B))^\perp \supseteq (\mathcal{N}(A))^\perp = \mathcal{R}(A)$. ■

The following Lemma is a key step to establish the connections between $\boldsymbol{DV-2}$ and $\boldsymbol{DM-2}$.

Lemma 3.3.2 *Suppose $\boldsymbol{DV-2}$ has its maximum attained at $\lambda^* \in \mathcal{R}_+^m$, and condition (3.3.62) holds, then $\exists t$ such that $Q_{\lambda^*,t} := \sum_{j=1}^m \lambda_j Q_j + Q_0 - tI \succeq 0$, $P_{\lambda^*,t} := \sum_{j=1}^m \lambda_j P_j + P_0 + tI \succeq 0$. If we denote*

$$\begin{array}{rll} \mu_{10}^* := & max & -\mathrm{vec}\,(R_1)^T (I_r \otimes Q_{\lambda^*,t})^\dagger \mathrm{vec}\,(R_1) - \mathrm{vec}\,(R_2)^T (P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(R_2) + \beta_{\lambda^*} \\ & s.t. & R_1 + R_2 = C_{\lambda^*} \\ & & R_1, R_2 \ \in \mathcal{M}^{nr}, \end{array}$$

Then $\mu_8^* = \mu_{10}^*$.

We first introduce some notations before proving the lemma. Let $P_{\lambda^*,t} = U_p \Sigma_p U_p^T$, $Q_{\lambda^*,t} = U_q \Sigma_q U_q^T$ be the spectral decompositions for $P_{\lambda^*,t}, Q_{\lambda^*,t}$ (so U_p, U_q are orthogonal square matrices and Σ_p, Σ_q are diagonal matrices with eigenvalues of $P_{\lambda^*,t},\ Q_{\lambda^*,t}$). Now we begin to prove Lemma 3.3.2.

Proof. Condition 3.3.62 and remark 3.3.1 implies the existence of such a t (by taking $t = \min_{j=1,2,\ldots,n} \lambda_j(Q_{\lambda^*})$). In fact, if $(I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)$ is positive semidefinite but not positive definite, then such a t is unique. By Proposition 3.3.2,

$$\mu_8^* = -\mathrm{vec}\,(C_{\lambda^*})^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*}.$$

Thus, to prove $\mu_8^* = \mu_{10}^*$, it suffice to prove

$$\mu_{10}^* = -\mathrm{vec}\,(C_{\lambda^*})^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*}. \tag{3.3.68}$$

By denoting

$$\begin{array}{rcl} \phi(R_1) & := & \beta_{\lambda^*} - \mathrm{vec}\,(R_1)^T ((I_r \otimes Q_{\lambda^*,t})^\dagger + (P_{\lambda^*,t} \otimes I_n)^\dagger) \mathrm{vec}\,(R_1) \\ & & +2\mathrm{vec}\,(C_{\lambda^*})^T (P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(R_1) - \mathrm{vec}\,(C_{\lambda^*})^T (P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}), \end{array}$$

according to the definition of μ_{10}^*, we have

$$\mu_{10}^* = \max_{R_1 \in \mathcal{M}^{nr}} \phi(R_1). \tag{3.3.69}$$

Since $Q_{\lambda^*,t} \succeq 0$, $P_{\lambda^*,t} \succeq 0$, we know $(I_r \otimes Q_{\lambda^*,t})^\dagger + (P_{\lambda^*,t} \otimes I_n)^\dagger \succeq 0$ and hence ϕ is concave. Because $(P_{\lambda^*,t} \otimes I_n)^\dagger \preceq (I_r \otimes Q_{\lambda^*,t})^\dagger + (P_{\lambda^*,t} \otimes I_n)^\dagger$ and Proposition 3.3.3, we can further deduce $(P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) \in \mathcal{R}((P_{\lambda^*,t} \otimes I_n)^\dagger) \subseteq \mathcal{R}((I_r \otimes Q_{\lambda^*,t})^\dagger + (P_{\lambda^*,t} \otimes I_n)^\dagger)$. Therefore, the maximum of the quadratic concave function $\phi(R_1)$ is finite and attained at

$$\begin{aligned} \mathrm{vec}\,(R_1^*) &= ((I_r \otimes Q_{\lambda^*,t})^\dagger + (P_{\lambda^*,t} \otimes I_n)^\dagger)^\dagger (P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) \\ &= (P_{\lambda^*,t} \otimes Q_{\lambda^*,t}^\dagger + P_{\lambda^*,t} P_{\lambda^*,t}^\dagger \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}). \end{aligned} \tag{3.3.70}$$

Now put $R_1 = R_1^*$ into (3.3.69),

$$\begin{aligned} \mu_{10}^* &= \phi(R_1^*) \\ &= \mathrm{vec}\,(C_{\lambda^*})^T (P_{\lambda^*,t} \otimes Q_{\lambda^*,t}^\dagger + P_{\lambda^*,t} P_{\lambda^*,t}^\dagger \otimes I_n)^\dagger (P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) \\ &\qquad - \mathrm{vec}\,(C_{\lambda^*})^T (P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*} \\ &= \mathrm{vec}\,(C_{\lambda^*})^T ((P_{\lambda^*,t}^2 \otimes Q_{\lambda^*,t}^\dagger + P_{\lambda^*,t} \otimes I_n)^\dagger - (P_{\lambda^*,t} \otimes I_n)^\dagger) \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*} \\ &= \mathrm{vec}\,(C_{\lambda^*})^T (U_p \otimes U_q) \Sigma (U_p \otimes U_q)^T \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*}, \end{aligned}$$

where

$$\Sigma := (\Sigma_p^2 \otimes \Sigma_q^\dagger + \Sigma_p \otimes I_n)^\dagger - \Sigma_p^\dagger \otimes I_n.$$

For $\Sigma_p = \mathrm{Diag}\,(\lambda_{p,i})$, $i = 1, 2, \ldots, r$, and $\Sigma_q = \mathrm{Diag}\,(\lambda_{q,j})$, $j = 1, 2, \ldots, n$, the diagonal entries of the diagonal matrix Σ can be calculated as

$$\Sigma_{ni+j} = \begin{cases} -\frac{1}{\lambda_{p,i} + \lambda_{q,j}} & \text{if } \lambda_{p,i} \neq 0,\ \lambda_{q,j} \neq 0 \\ 0 & \text{if } \lambda_{p,i} = 0 \text{ or } \lambda_{q,j} = 0. \end{cases}$$

So we may deduce $\Sigma = -(\Sigma_p \otimes I_n + I_r \otimes \Sigma_q)^\dagger$. As a consequence,

$$\begin{aligned} \mu_{10}^* &= \mathrm{vec}\,(C_{\lambda^*})^T (U_p \otimes U_q) \Sigma (U_p \otimes U_q)^T \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*} \\ &= -\mathrm{vec}\,(C_{\lambda^*})^T (I_r \otimes Q_{\lambda^*,t} + P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*} \\ &= -\mathrm{vec}\,(C_{\lambda^*})^T (I_r \otimes Q_{\lambda^*} + P_{\lambda^*} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\lambda^*}) + \beta_{\lambda^*}. \end{aligned}$$

So (3.3.68) is proved and Lemma 3.3.2 has been established. ■

Then we come to establish the connections between μ_9^* and μ_{10}^*.

Proposition 3.3.4 **([4])** *$f(X) = \mathrm{trace}\,(X^TQX + 2C^TX) + \beta \geq 0$ for any $X \in \mathcal{M}^{nr}$ if and only if $\exists S \in S_r$, such that $\mathrm{trace}\, S \leq 0$ and $\begin{pmatrix} S + \frac{\beta}{r} I_r & C^T \\ C & Q \end{pmatrix} \succeq 0$.*

■

Remark 3.3.2 *An immediate result following Proposition 3.3.4 is, if $\mathcal{R}(C) \subseteq \mathcal{R}(Q)$, then*

$$\begin{array}{lll} & \min\limits_{X\in\mathcal{M}^{nr}} & \mathrm{trace}\,(X^TQX + 2C^TX) \\ = & \max\limits_{\psi\in\mathcal{S}^r} & -\mathrm{trace}\,\psi \\ & s.t. & \begin{pmatrix} \psi & C^T \\ C & Q \end{pmatrix} \succeq 0. \end{array}$$

■

Suppose $\boldsymbol{DV-2}$ is maximized at $\lambda^* \in \mathcal{R}^m_+$. Then consider the following subproblem of $\boldsymbol{DM-2}$

$$\begin{array}{lll} \mu_{11}^* := & \max & \beta_{\lambda^*} - \mathrm{trace}\, S_1 - \mathrm{trace}\, S_2 \\ & \text{s.t.} & R_1 + R_2 = C_{\lambda^*} \\ & & \begin{pmatrix} S_1 & R_1^T \\ R_1 & Q_{\lambda^*,t} \end{pmatrix} \succeq 0 \\ & & \begin{pmatrix} S_2 & R_2 \\ R_2^T & P_{\lambda^*,t} \end{pmatrix} \succeq 0 \\ & & R_1, R_2 \in \mathcal{M}^{nr},\ S_1 \in S_r\ S_2 \in S_n, \end{array} \tag{3.3.71}$$

where we have fixed $\lambda = \lambda^*$ and fixed t as the real number satisfying $Q_{\lambda^*,t} := Q_{\lambda^*} - tI_n \succeq 0,\ P_{\lambda^*,t} := P_{\lambda^*} + tI_r \succeq 0$. Because we have fixed λ to be particular values, the maximal value of (3.3.71) can never exceed the maximum of the original problem $\boldsymbol{DM-2}$, i.e., $\mu_{11}^* \leq \mu_9^*$. The following Proposition further reveals $\mu_{11}^* = \mu_{10}^*$ under a certain condition.

Lemma 3.3.3 *Suppose $\boldsymbol{DV-2}$ is maximized at $\lambda^* \in \mathcal{R}^m_+$, and t is the real number satisfying $Q_{\lambda^*,t} := Q_{\lambda^*} - tI_n \succeq 0,\ P_{\lambda^*,t} := P_{\lambda^*} + tI_r \succeq 0$. μ_{11}^* follows from the definition in* (3.3.71)*. Then*

$$\mu_{10}^* = \mu_{11}^*,$$

if and only if

$$\exists M \in \mathcal{M}^{nr},\ C_{\lambda^*} = Q_{\lambda^*,t} M P_{\lambda^*,t}. \tag{3.3.72}$$

We firstly introduce some notations to simplify the proof to this proposition. We rearrange $\Sigma_p = \begin{pmatrix} \Sigma_{p1} & 0 \\ 0 & 0 \end{pmatrix}$, $\Sigma_q = \begin{pmatrix} \Sigma_{q1} & 0 \\ 0 & 0 \end{pmatrix}$, where Σ_{p1}, Σ_{q1} are diagonal matrix with positive diagonal entries, and we then rewrite $U_p = \begin{pmatrix} U_{p1} & U_{p2} \end{pmatrix}$, $U_q = \begin{pmatrix} U_{q1} & U_{q2} \end{pmatrix}$, where U_{p1}, U_{q1} are eigenvectors corresponding to nonzero eigenvalues of $P_{\lambda^*,t}, Q_{\lambda^*,t}$ respectively.

Proof. By Proposition 3.3.4 and Remark 3.3.2, we can deduce

$$\begin{aligned} \mu_{11}^* &= \max_{R_1+R_2=C_{\lambda^*}} (\beta_{\lambda^*} + \min_{X\in\mathcal{M}^{nr}} \operatorname{trace}(X^T Q_{\lambda^*,t} X + 2R_1^T X) \\ &\qquad\qquad + \min_{Y\in\mathcal{M}^{nr}} (\operatorname{trace} Y P_{\lambda^*,t} Y^T + 2R_2 Y^T)) \\ &= \begin{array}{ll} \max & \beta_{\lambda^*} - \operatorname{vec}(R_1)^T (I_r \otimes Q_{\lambda^*,t})^\dagger \operatorname{vec}(R_1) - \operatorname{vec}(R_2)^T (P_{\lambda^*,t} \otimes I_n)^\dagger \operatorname{vec}(R_2) \\ \text{s.t.} & R_1 + R_2 = C_{\lambda^*} \\ & \operatorname{vec}(R_1) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t}) \\ & \operatorname{vec}(R_2) \in \mathcal{R}(P_{\lambda^*,t} \otimes I_n). \end{array} \end{aligned}$$

Recall the definition of $\phi(R_1)$ in Lemma 3.3.2, we can deduce

$$\mu_{11}^* = \begin{array}{ll} \max & \phi(R_1) \\ \text{s.t.} & \operatorname{vec}(R_1) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t}) \\ & \operatorname{vec}(C_{\lambda^*} - R_1) \in \mathcal{R}(P_{\lambda^*,t} \otimes I_n). \end{array} \tag{3.3.73}$$

We want to find the sufficient and necessary condition for $\mu_{11}^* = \mu_{10}^*$. By comparing (3.3.69) and (3.3.73), we find that (3.3.73) has two more constraints than (3.3.69). Therefore, $\mu_{11}^* \le \mu_{10}^*$, and the equality holds if and only if (3.3.69) has a maximizer which also satisfies the two constraints in (3.3.73), i.e.,

$$\exists R_1 \in \operatorname{argmax}(\phi),\ \operatorname{vec}(R_1) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t}),\ \operatorname{vec}(C_{\lambda^*} - R_1) \in \mathcal{R}(P_{\lambda^*,t} \otimes I_n), \tag{3.3.74}$$

where $\operatorname{argmax}(\phi)$ denotes the set of the maximizers of ϕ.

So (3.3.74) is the sufficient and necessary condition for $\mu_{11}^* = \mu_{10}^*$. Now our aim is to find a simpler condition equivalent with (3.3.74). We first study the conditions under which there is an $R_1 \in \operatorname{argmax}(\phi)$, such that $\operatorname{vec}(R_1) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t})$.

By (3.3.70), $\operatorname{vec}(R_1^*) = (P_{\lambda^*,t} \otimes Q_{\lambda^*,t}^\dagger + P_{\lambda^*,t} P_{\lambda^*,t}^\dagger \otimes I_n)^\dagger \operatorname{vec}(C_{\lambda^*}) \in \operatorname{argmax}(\phi)$. So the set $\operatorname{argmax}(\phi)$ can be characterized as

$$\operatorname{argmax}(\phi) = \{R_1^* + V |\ \operatorname{vec}(V) \in \mathcal{N}(I_r \otimes Q_{\lambda^*,t} + P_{\lambda^*,t} \otimes I_n)\}. \tag{3.3.75}$$

If $\operatorname{vec}(R_1^*) \notin \mathcal{R}(I_r \otimes Q_{\lambda^*,t})$ (but always $\operatorname{vec}(R_1^*) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t} + P_{\lambda^*,t} \otimes I_n)$ by (3.3.70)), then the projection of R_1^* on $\mathcal{R}(P_{\lambda^*,t} \otimes I_n)$ is nonzero. So for any $V \in \mathcal{N}(I_r \otimes Q_{\lambda^*,t} + P_{\lambda^*,t} \otimes I_n)$, the projection of $R_1^* + V$ on $\mathcal{R}(P_{\lambda^*,t} \otimes I_n)$ is still nonzero. So if $\text{vec}\,(R_1^*) \notin \mathcal{R}(I_r \otimes Q_{\lambda^*,t})$, we get $\text{argmax}(\phi) \cap \{R_1 | \text{vec}\,(R_1) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t})\} = \emptyset$. Therefore, we only need to concern under which conditions R_1^* satisfies the constraint $\text{vec}\,(R_1^*) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t})$, instead of checking the feasibility of each member of $\text{argmax}(\phi)$.

We know $\text{vec}\,(R_1^*) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t})$ if and only if there exists a vector $v \in \mathcal{R}^{nr}$, such that

$$\text{vec}\,(R_1^*) = (I_r \otimes Q_{\lambda^*,t})v = (U_p \otimes U_q)(I_r \otimes \Sigma_q)(U_p \otimes U_q)^T v. \tag{3.3.76}$$

By (3.3.70),

$$\begin{aligned} \text{vec}\,(R_1^*) &= (P_{\lambda^*,t} \otimes Q_{\lambda^*,t}^\dagger + P_{\lambda^*,t} P_{\lambda^*,t}^\dagger \otimes I_n)^\dagger \text{vec}\,(C_{\lambda^*}) \\ &= (U_p \otimes U_q)(\Sigma_p \otimes \Sigma_q^\dagger + \Sigma_p \Sigma_p^\dagger \otimes I_n)^\dagger (U_p \otimes U_q)^T \text{vec}\,(C_{\lambda^*}) \\ &= (U_p \otimes U_q) D (U_p \otimes U_q)^T \text{vec}\,(C_{\lambda^*}), \end{aligned} \tag{3.3.77}$$

where the diagonal matrix $D := (\Sigma_p \otimes \Sigma_q^\dagger + \Sigma_p \Sigma_p^\dagger \otimes I_n)^\dagger$.

So by (3.3.76) and (3.3.77), $\text{vec}\,(R_1^*) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t})$ if and only if $\exists v \in \mathbb{R}^{nr}$, such that

$$(U_p \otimes U_q)(I_r \otimes \Sigma_q)(U_p \otimes U_q)^T v = (U_p \otimes U_q) D (U_p \otimes U_q)^T \text{vec}\,(C_{\lambda^*}). \tag{3.3.78}$$

Furthermore, (3.3.78) holds if and only if

$$v = (U_p \otimes U_q)(I_r \otimes \Sigma_q^\dagger) D (U_p \otimes U_q)^T \text{vec}\,(C_{\lambda^*}), \tag{3.3.79}$$

and

$$(I_r \otimes \Sigma_q \Sigma_q^\dagger) D (U_p \otimes U_q)^T \text{vec}\,(C_{\lambda^*}) = D (U_p \otimes U_q)^T \text{vec}\,(C_{\lambda^*}). \tag{3.3.80}$$

By a close observation for (3.3.80), the diagonal entry of the diagonal matrix $I_r \otimes \Sigma_q \Sigma_q^\dagger$ equals zero if and only if it finally corresponds to a row of $(U_p \otimes U_{q,2})^T$ in the production of (3.3.80) (rows of $(U_p \otimes U_{q,2})^T$ are a subset of rows of $(U_p \otimes U_q)^T$). And the diagonal entries of D can be calculated as

$$D_{ni+j} \begin{cases} > 0 & \text{if } (\Sigma_p)_{ii} \neq 0 \\ = 0 & \text{if } (\Sigma_p)_{ii} = 0 \end{cases} \quad i = 1, 2, \ldots, r,\ j = 1, 2, \ldots, n.$$

So the diagonal entry of the diagonal matrix D equals zero if and only if it corresponds to a row of $(U_{p,2} \otimes U_q)^T$ in the production of (3.3.80). Therefore, $(I_r \otimes \Sigma_q \Sigma_q^\dagger) D$ actually diminishes the positive diagonal entries of D that correspond to the rows of $(U_{p,1} \otimes U_{q,2})^T$ in the production of (3.3.80). Therefore, (3.3.80) holds if and only if $\text{vec}\,(C_{\lambda^*})$ originally has no components on the subspace spanned by $U_{p,1} \otimes U_{q,2}$. So we conclude

$$(U_{p,1} \otimes U_{q,2})^T \text{vec}\,(C_{\lambda^*}) = 0 \tag{3.3.81}$$

is the sufficient and necessary condition for (3.3.78), and hence for $\mathrm{vec}\,(R_1^*) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t})$. For the other constraint $\mathrm{vec}\,(C_{\lambda^*} - R_1) \in \mathcal{R}(P_{\lambda^*,t} \otimes I_n)$ in (3.3.74), note

$$\mathrm{vec}\,(C_{\lambda^*} - R_1^*) = (P_{\lambda^*,t}^\dagger \otimes Q_{\lambda^*,t} + I_r \otimes Q_{\lambda^*,t} Q_{\lambda^*,t}^\dagger)^\dagger \mathrm{vec}\,(C_{\lambda^*}),$$

which has an identical structure with the formulation (3.3.70) for R_1^*, except swapping the positions of $P_{\lambda^*,t}$ and $Q_{\lambda^*,t}$. So by analogy, the sufficient and necessary condition for $\mathrm{vec}\,(C_{\lambda^*} - R_1^*) \in \mathcal{R}(P_{\lambda^*,t} \otimes I_n)$ is

$$(U_{p,2} \otimes U_{q,1})^T \mathrm{vec}\,(C_{\lambda^*}) = 0. \tag{3.3.82}$$

Because λ^* is feasible to $\boldsymbol{DV-2}$, we have $\mathrm{vec}\,(C_{\lambda^*}) \in \mathcal{R}(I_r \otimes Q_{\lambda^*,t} + P_{\lambda^*,t} \otimes I_n)$, which implies the projection of $\mathrm{vec}\,(C_{\lambda^*})$ on $\mathcal{N}(I_r \otimes Q_{\lambda^*,t} + P_{\lambda^*,t} \otimes I_n)$ is zero, i.e.,

$$(U_{p,2} \otimes U_{q,2})^T \mathrm{vec}\,(C_{\lambda^*}) = 0. \tag{3.3.83}$$

We know, $U_{p,1} \otimes U_{q,1}$, $U_{p,2} \otimes U_{q,1}$, $U_{p,1} \otimes U_{q,2}$ and $U_{p,2} \otimes U_{q,2}$ span the whole space, so condition (3.3.81), (3.3.82), (3.3.83) together imply that

$$\mathrm{vec}\,(C_{\lambda^*}) \in \mathcal{R}(U_{p,1} \otimes U_{q,1}) \tag{3.3.84}$$

is the sufficient and necessary condition for (3.3.74), hence also the sufficient and necessary condition for $\mu_{11}^* = \mu_{10}^*$. Furthermore, condition (3.3.84) can be equivalently formulated as (3.3.72), which establishes Lemma 3.3.3. ■

Remark 3.3.3 *The proof to Lemma 3.3.3 shows that the difference between* $\boldsymbol{DV-2}$ *and* $\boldsymbol{DM-2}$ *originates from the different feasible sets for* (3.3.73) *and* (3.3.69). *One can verify that the maximizers for* (3.3.69) *are*

$$\begin{aligned} \mathrm{vec}\,(R_1^*) &= (I_r \otimes U_{q,2} U_{q,2}^T)\mathrm{vec}\,(C_{\lambda^*}) + \mathrm{vec}\,(\bar{R}_1^*), \\ \mathrm{vec}\,(R_2^*) &= (U_{p,2} U_{p,2}^T \otimes I_n)\mathrm{vec}\,(C_{\lambda^*}) + \mathrm{vec}\,(\bar{R}_2^*) \end{aligned} \tag{3.3.85}$$

for some $\mathrm{vec}\,(\bar{R}_1^*)$, $\mathrm{vec}\,(\bar{R}_2^*) \in \mathcal{R}(P_{\lambda^*,t} \otimes Q_{\lambda^*,t})$. *One can also verify that the maximizers for* (3.3.73) *are*

$$\begin{aligned} \mathrm{vec}\,(\hat{R}_1^*) &= (U_{p,2} U_{p,2}^T \otimes I_n)\mathrm{vec}\,(C_{\lambda^*}) + \mathrm{vec}\,(\bar{R}_1^*), \\ \mathrm{vec}\,(\hat{R}_2^*) &= (I_r \otimes U_{q,2} U_{q,2}^T)\mathrm{vec}\,(C_{\lambda^*}) + \mathrm{vec}\,(\bar{R}_2^*), \end{aligned} \tag{3.3.86}$$

where $\bar{R}_1^*,\ \bar{R}_2^*$ are the same as in (3.3.85).
Therefore, the difference between μ_{11}^* and μ_{10}^* can be calculated as

$$\begin{aligned} \mu_{10}^* - \mu_{11}^* &= \phi(R_1^*) - \phi(\hat{R}_1^*) \\ &= \text{vec}\,(C_{\lambda^*})^T (U_{p,2}U_{p,2}^T \otimes I_n)^T (I_r \otimes Q_{\lambda^*,t})^\dagger (U_{p,2}U_{p,2}^T \otimes I_n)\text{vec}\,(C_{\lambda^*}) \\ &\quad + \text{vec}\,(C_{\lambda^*})^T (I_r \otimes U_{q,2}U_{q,2}^T)^T (P_{\lambda^*,t} \otimes I_n)^\dagger (I_r \otimes U_{q,2}U_{q,2}^T)\text{vec}\,(C_{\lambda^*}). \end{aligned} \tag{3.3.87}$$

Because $\mu_9^* \geq \mu_{11}^*$, $\mu_8^* = \mu_{10}^*$, (3.3.87) *provides a bound for the difference between* $\boldsymbol{DM-2}$ *and* $\boldsymbol{DV-2}$ *, i.e.,*

$$\begin{aligned} \mu_8^* - \mu_9^* \leq\ & \text{vec}\,(C_{\lambda^*})^T (U_{p,2}U_{p,2}^T \otimes I_n)^T (I_r \otimes Q_{\lambda^*,t})^\dagger (U_{p,2}U_{p,2}^T \otimes I_n)\text{vec}\,(C_{\lambda^*}) \\ & + \text{vec}\,(C_{\lambda^*})^T (I_r \otimes U_{q,2}U_{q,2}^T)^T (P_{\lambda^*,t} \otimes I_n)^\dagger (I_r \otimes U_{q,2}U_{q,2}^T)\text{vec}\,(C_{\lambda^*}). \end{aligned} \tag{3.3.88}$$

■

Now we give the main theorem.

Theorem 3.3.1 *If* $\boldsymbol{DV-2}$ *satisfies condition* (3.3.62) *and attains its maximum at λ^*, and condition* (3.3.72) *holds, then* $\boldsymbol{VSDR-2}$ *and* $\boldsymbol{MSDR-2}$ *generate the same bound for* $\boldsymbol{QMP_2}$ *, i.e., $\mu_6^* = \mu_7^*$. Furthermore, if λ^* is the unique maximizer for* $\boldsymbol{DV-2}$ *, then the condition* (3.3.72) *is also necessary to establish $\mu_6^* = \mu_7^*$.*

Proof. Through previous arguments, we have established $\mu_8^* = \mu_{10}^* = \mu_{11}^* \leq \mu_9^*$ under condition (3.3.72). We also have shown $\boldsymbol{DM-2}$ is never stronger than $\boldsymbol{DV-2}$, i.e. $\mu_8^* \geq \mu_9^*$. Thus under condition (3.3.72), the two dual programs, $\boldsymbol{DV-2}$ and $\boldsymbol{DM-2}$ generate the same bounds, i.e., $\mu_8^* = \mu_9^*$. By Remark 3.3.1, under condition (3.3.62), both dual programs have their optimal values equaling their primal programs, so the optimal values of $\boldsymbol{VSDR-2}$ and $\boldsymbol{MSDR-2}$ coincide, i.e., $\mu_6^* = \mu_7^*$.
Furthermore, the condition (3.3.72) is also necessary to establish $\mu_6^* = \mu_7^*$ when λ^* is the unique maximizer for $\boldsymbol{DV-2}$. Because if (3.3.72) fails, then by Lemma 3.3.3, we have $\mu_{11}^* < \mu_{10}^* = \mu_8^*$. If $\boldsymbol{DM-2}$ is also maximized at $\lambda = \lambda^*$, then $\mu_9^* = \mu_{11}^*$, so we have $\mu_9^* < \mu_8^*$; if $\boldsymbol{DM-2}$ is maximized at $\bar{\lambda} \neq \lambda^*$, then its optimal value μ_9^* will never exceed the subproblem of $\boldsymbol{DV-2}$ by fixing $\lambda = \bar{\lambda}$. Because λ^* is the unique maximizer to $\boldsymbol{DV-2}$, we conclude $\mu_9^* < \mu_8^*$ in this case. Then by condition (3.3.62) and Remark 3.3.1, the objective values of their primal programs also have the strict inequality $\mu_7^* < \mu_6^*$. ■

There is an immediate result following Theorem 3.3.1, which is useful and nontrivial.

Lemma 3.3.4 *For a homogeneous* $\boldsymbol{QMP_2}$ *(*$C_j = 0,$ *for all* $j = 0, 1, \ldots, m$*) satisfying condition* (3.3.62)*,* $\boldsymbol{VSDR-2}$ *and* $\boldsymbol{MSDR-2}$ *always generate the same bound.*

■

3.3.2 Improving the Bound of MSDR-2

If a $\boldsymbol{MSDR-2}$ is solved and returns an optimal value μ_9^* and the dual maximizers $\bar{\lambda},\ R_1^*$, then we can further improve the bound of μ_9^* without much computational effort.

Suppose $\bar{\lambda}, t, \hat{R}_1, \hat{R}_2$ maximizes $\boldsymbol{DM-2}$, then the feasibility implies $P_{\bar{\lambda},t} \succeq 0$, $Q_{\bar{\lambda},t} \succeq 0$, and $\mathrm{vec}\,(\hat{R}_1) \in \mathcal{R}(I_r \otimes Q_{\bar{\lambda},t})$, $\mathrm{vec}\,(\hat{R}_2) \in \mathcal{R}(P_{\bar{\lambda},t} \otimes I_n)$. Therefore, $I_r \otimes Q_{\bar{\lambda},t} + P_{\bar{\lambda},t} \otimes I_n \succeq 0$ and $\mathrm{vec}\,(C_{\bar{\lambda}}) = \mathrm{vec}\,(\hat{R}_1 + \hat{R}_2) \in \mathcal{R}(I_r \otimes Q_{\bar{\lambda},t} + P_{\bar{\lambda},t} \otimes I_n)$. So there is an $\alpha \in \mathcal{R}$ such that $(\bar{\lambda}, \alpha)$ is feasible to $\boldsymbol{DV-2}$, and let μ_{12}^* denotes the objective value of the subproblem of $\boldsymbol{DV-2}$ by fixing $\lambda = \bar{\lambda}$, i.e.,

$$\begin{array}{rll} \mu_{12}^* = & \max\limits_{\alpha} & \beta_{\bar{\lambda}} - \alpha \\ & \text{s.t.} & \begin{pmatrix} \alpha & \mathrm{vec}\,(C_{\bar{\lambda}})^T \\ \mathrm{vec}\,(C_{\bar{\lambda}}) & I_r \otimes Q_{\bar{\lambda},t} + P_{\bar{\lambda},t} \otimes I_n \end{pmatrix} \succeq 0. \end{array} \tag{3.3.89}$$

Then $\mu_{12}^* \le \mu_8^*$, so μ_{12}^* is also a *valid* lower bound for $\boldsymbol{QMP_2}$, which is no stronger than the bound of $\boldsymbol{VSDR-2}$.

According to the discussion in Lemma 3.3.2, if we denote

$$\begin{array}{rcl} \hat{\phi}(R_1) & := & -\mathrm{vec}\,(R_1)^T((I_r \otimes Q_{\bar{\lambda},t})^\dagger + (P_{\bar{\lambda},t} \otimes I_n)^\dagger)\mathrm{vec}\,(R_1) \\ & & +2\mathrm{vec}\,(C_{\bar{\lambda}})^T(P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(R_1) - \mathrm{vec}\,(C_{\bar{\lambda}})^T(P_{\lambda^*,t} \otimes I_n)^\dagger \mathrm{vec}\,(C_{\bar{\lambda}}) + \beta_{\bar{\lambda}}, \end{array}$$

then

$$\mu_{12}^* = \max \hat{\phi}(R_1), \tag{3.3.90}$$

and its maximizer R_1^* satisfies

$$\begin{array}{rcl} \mathrm{vec}\,(R_1^*) & = & (I_r \otimes \hat{U}_{q,2}\hat{U}_{q,2}^T)\mathrm{vec}\,(C_{\bar{\lambda}}) + \mathrm{vec}\,(\bar{R}_1^*), \\ \mathrm{vec}\,(R_2^*) & = & (\hat{U}_{p,2}\hat{U}_{p,2}^T \otimes I_n)\mathrm{vec}\,(C_{\bar{\lambda}}) + \mathrm{vec}\,(\bar{R}_2^*), \end{array} \tag{3.3.91}$$

where $\hat{U}_{p,2},\ \hat{U}_{q,2}$ consist of eigenvectors corresponding to the zero eigenvalues of $Q_{\bar{\lambda},t}, P_{\bar{\lambda},t}$, and $\mathrm{vec}\,(\bar{R}_1^*),\ \mathrm{vec}\,(\bar{R}_2^*) \in \mathcal{R}(Q_{\bar{\lambda},t} \otimes P_{\bar{\lambda},t})$.

At the other side, because $\bar{\lambda}$ maximizes $\boldsymbol{DM-2}$, by Proposition 3.3.4,

$$\begin{array}{rll} \mu_9^* = & \max & \hat{\phi}(R_1) \\ = & \text{s.t.} & \operatorname{vec}(R_1) \in \mathcal{R}(I_r \otimes Q_{\bar{\lambda},t}) \\ & & \operatorname{vec}(C_{\bar{\lambda}} - R_1) \in \mathcal{R}(P_{\bar{\lambda},t} \otimes I_n). \end{array} \tag{3.3.92}$$

The above program is maximized at $\hat{R}_1$, $\hat{R}_2 = C_{\bar{\lambda}} - \hat{R}_1$ (the maximizers for $\boldsymbol{DM-2}$), which satisfy

$$\begin{array}{rcl} \operatorname{vec}(\hat{R}_1) & = & (\hat{U}_{p,2}\hat{U}_{p,2}^T \otimes I_n)\operatorname{vec}(C_{\bar{\lambda}}) + \operatorname{vec}(\bar{R}_1^*), \\ \operatorname{vec}(\hat{R}_2) & = & (I_r \otimes \hat{U}_{q,2}\hat{U}_{q,2}^T)\operatorname{vec}(C_{\bar{\lambda}}) + \operatorname{vec}(\bar{R}_2^*). \end{array} \tag{3.3.93}$$

Therefore, by (3.3.90), (3.3.91),(3.3.92) and (3.3.93), we can establish the relations between μ_9^* and μ_{12}^* as

$$\begin{array}{rcl} \mu_{12}^* - \mu_9^* & = & \hat{\phi}(R_1^*) - \hat{\phi}(\hat{R}_1) \\ & = & \operatorname{vec}(C_{\bar{\lambda}})^T(\hat{U}_{p,2}\hat{U}_{p,2}^T \otimes I_n)^T(I_r \otimes Q_{\bar{\lambda},t})^\dagger(\hat{U}_{p,2}\hat{U}_{p,2}^T \otimes I_n)\operatorname{vec}(C_{\bar{\lambda}}) \\ & & \quad +\operatorname{vec}(C_{\bar{\lambda}})^T(I_r \otimes \hat{U}_{q,2}\hat{U}_{q,2}^T)^T(P_{\bar{\lambda},t} \otimes I_n)^\dagger(I_r \otimes \hat{U}_{q,2}\hat{U}_{q,2}^T)\operatorname{vec}(C_{\bar{\lambda}}) \\ & \geq & 0. \end{array} \tag{3.3.94}$$

Thus, we can always improve the $\boldsymbol{DM-2}$ bound μ_9^* to μ_{12}^*, which is also a valid lower bound. By (3.3.88) and (3.3.94), the difference between the improved bound μ_{12}^* and the bound from $\boldsymbol{DV-2}$ can be bounded as

$$\begin{array}{rcl} \mu_8^* - \mu_{12}^* & \leq & \operatorname{vec}(C_{\lambda^*})^T(U_{p,2}U_{p,2}^T \otimes I_n)^T(I_r \otimes Q_{\lambda^*,t})^\dagger(U_{p,2}U_{p,2}^T \otimes I_n)\operatorname{vec}(C_{\lambda^*}) \\ & & + \operatorname{vec}(C_{\lambda^*})^T(I_r \otimes U_{q,2}U_{q,2}^T)^T(P_{\lambda^*,t} \otimes I_n)^\dagger(I_r \otimes U_{q,2}U_{q,2}^T)\operatorname{vec}(C_{\lambda^*}) \\ & & - \operatorname{vec}(C_{\bar{\lambda}})^T(\hat{U}_{p,2}\hat{U}_{p,2}^T \otimes I_n)^T(I_r \otimes Q_{\bar{\lambda},t})^\dagger(\hat{U}_{p,2}\hat{U}_{p,2}^T \otimes I_n)\operatorname{vec}(C_{\bar{\lambda}}) \\ & & - \operatorname{vec}(C_{\bar{\lambda}})^T(I_r \otimes \hat{U}_{q,2}\hat{U}_{q,2}^T)^T(P_{\bar{\lambda},t} \otimes I_n)^\dagger(I_r \otimes \hat{U}_{q,2}\hat{U}_{q,2}^T)\operatorname{vec}(C_{\bar{\lambda}}), \end{array}$$

where λ^* is the maximizer for $\boldsymbol{DV-2}$, and $U_{p,2}$, $U_{q,2}$ are eigenvectors corresponding to zero eigenvalues of $P_{\lambda^*,t},\ Q_{\lambda^*,t}$.

Chapter 4

Applications: Quadratic Assignment Problem

As we introduced in the first section, any integer programming can be formulated as a ***QCQP***. Furthermore, most of them admit a ***QMP*** formulation. For example, the quadratic assignment problem (***QAP***) proposed by Koopmans and Beckman in 1957 [30] is well-known to be NP-hard, and many famous problems like traveling-salesman-problem (TSP) can be reformulated as a ***QAP***. The original ***QAP*** problem is a minimization problem over all the possible permutations

$$\min_{\pi \in \Pi(n)} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} b_{\pi(i),\pi(j)},$$

where $\Pi(n)$ denotes the set of permutations of n entries, $a_{i,j}, b_{i,j}$ $i,j = 1,2,\ldots,n$ are weights corresponding to the distance or cost in real life. Because permutations can also be represented with a permutation matrix, the ***QAP*** has a nice Koopmans-Beckman formulation [30]

$$\min_{X \in \Pi(n)} \quad \operatorname{trace} AXBX^T.$$

In practice, the objective function may include a linear term $2C \cdot X$. By the fact $\Pi(n) = \{X \in \mathcal{M}^{nn} | XX^T = I,\ Xe = X^T e = e,\ X \geq 0\}$, the ***QAP*** can be formulated as

$$\begin{array}{ll} \min & \operatorname{trace}\left(AXBX^T + 2C^T X\right) \\ \text{s.t.} & XX^T = X^T X = I \\ & Xe = X^T e = e \\ & X \geq 0. \end{array} \tag{4.0.95}$$

Some constraints in (4.0.95) are redundant, but they help improve the bound when we do a relaxation.

4.1 Matrix-Lifting Semidefinite Relaxation for QAP

Due to its severe nonconvexity, QAP has been a challenging problem to optimizers and engineers for many decades. Solving a QAP to optimality normally requires to run a branch and bound algorithm, in which getting strong, inexpensive relaxation bound is critical to expedite the computation. One of the earliest and least expensive relaxations for ***QAP*** is based on a Linear Programming (***LP***) formulation, e.g. Gilmore-Lawler (GLB) [17, 12], and related dual-based ***LP*** bounds [28, 38, 12]. These inexpensive formulations are able to handle problems with n up to several tens, see [17, 31]. However, the bounds are too weak to handle large problems. Improved bounds include classes of: eigenvalue and parametric eigenvalue bounds EB [14, 44], projected eigenvalue bounds PB [21, 13], convex quadratic programming bounds QPB [1], and ***SDP*** bounds [43, 54]. Classical ***SDP*** bounds vectorize the matrix X and lift $(1, \operatorname{vec}(X)) \in R^{n^2+1}$ into the semidefinite cone $\mathcal{S}_+^{n^2+1}$ and provide a rich amount of cuts for the convex hull of the feasible set. However, this vector-lifting SDP formulation, even the cheapest SDP1, requires $O(n^4)$ variables and hence turns out to be too expensive in practical computation. Problems with $n > 25$ become impractical to solve by such methods.

Here we are interested in applying the theory in Chapter 3 and reformulate (4.0.95) as a ***QMP*** by replacing $R = XB$

$$
(QAP-QMP) \qquad
\begin{array}{ll}
\min & \operatorname{trace}\begin{pmatrix} X & R \end{pmatrix}\begin{pmatrix} 0 & \frac{1}{2}A \\ \frac{1}{2}A & 0 \end{pmatrix}\begin{pmatrix} X \\ R \end{pmatrix} + \operatorname{trace} CX \\
s.t. & R = XB \\
 & Xe = X^T e = e \\
 & X \geq 0 \\
 & X^T E^{(ii)} X = 1,\ i = 1, 2, \ldots n \\
 & X^T E^{(ij)} X = 0,\ i, j = 1, 2, \ldots n,
\end{array}
$$

where $E^{(ij)},\ i, j = 1, 2, \ldots, n$ denotes the sparse matrix with a unique 1 at the $i-$th row and $j-$th column. The constraints $X^T E^{(ii)} X = 1,\ X^T E^{(ij)} X = 0$ $(i, j = 1, 2, \ldots n)$ stands for $XX^T = I$.

So we have formulated a ***QAP*** as a $\boldsymbol{QMP_1}$. By Theorem 3.2.3, the relaxations ***VSDR−1*** and ***MSDR−1*** always generate the same bound. Therefore, we can use the inexpensive $\boldsymbol{MSDR-1}$ formulation to get a relaxation bound

$$\begin{array}{ll} \min & \operatorname{trace} AY + \operatorname{trace} CX \\ s.t. & R = XB \\ & Xe = X^T e = e \\ & X \geq 0 \\ & \begin{pmatrix} I & X^T & R^T \\ X & I & Y \\ R & Y & Z \end{pmatrix} \succeq 0 \\ & X, R \in \mathbb{R}^{n,n}, \; Y, Z \in \mathcal{S}^n. \end{array} \tag{4.1.96}$$

Also, given that $Xe = X^T e = e$ and $Y = XBX^T, \; Z = XBX^T$ for all $X, \; Y, \; Z$ feasible to original QAP, we have $Ye = XBe, Ze = XB^2 e$. These constraints can be added into (4.1.96) to improve the bound, and we get a new formulation

$$\begin{array}{lll} \mu_{13}^* := & \min & \operatorname{trace} AY + \operatorname{trace} CX \\ & s.t. & R = XB \\ & & Xe = X^T e = e \\ & & X \geq 0 \\ & & \operatorname{diag}(Y) = X \operatorname{diag}(B) \\ & & \operatorname{diag}(Z) = X \operatorname{diag}(B^2) \\ & & Ye = XBe \\ & & Ze = XB^2 e \\ & & \begin{pmatrix} I & X^T & R^T \\ X & I & Y \\ R & Y & Z \end{pmatrix} \succeq 0 \\ & & X, R \in \mathbb{R}^{n,n}, \; Y, Z \in \mathcal{S}^n. \end{array} \tag{4.1.97}$$

4.2 The Convex Hull of the Orthogonal Similarity Set of B

Just as in any $\boldsymbol{LP}$ relaxation, the relaxation (4.1.97) uses only $O(n^2)$ variables. So the $\boldsymbol{MSDR-1}$ formulation is indeed very cheap to compute. However, our numerical tests show that the bound generated by (4.1.97) is often weaker than that from QPB. Therefore, we are motivated to add more constraints in order to strengthen (4.1.97) . Here we will first introduce the concept of *majorization*, which is an equivalence relation on $\mathbb{R}^n$. By abuse of notation, we denote x *majorizes* y or y *is majorized by* x with $x \succeq y$ or $y \preceq x$. The precise definition is as follows, see e.g. [34]: Let $x, y \in \mathbb{R}^n$ and, without loss of generality, let the components of both vectors be sorted in non-increasing order. Then $x \succeq y$ if and only if

$$\begin{array}{rcll} x_1 + x_2 + \ldots + x_p & \geq & y_1 + y_2 + \ldots + y_p, & \forall p = 1, 2, \ldots, n-1, \\ x_1 + x_2 + \ldots + x_n & = & y_1 + y_2 + \ldots + y_n. & \end{array}$$

In [34], it is shown that $x \succeq y$ if and only if $\exists S \in \mathcal{E} \cap \mathcal{N}$ with $Sx = y$. Note that for a fixed y, the constraint $x \succeq y$ is not a convex constraint; while $x \preceq y$ is a convex constraint and it has an equivalent LP formulation [22].

Now in (4.1.97), Y actually stands for XBX^T with X an orthogonal matrix. So $\lambda(Y) = \lambda(B)$. Suppose $A = U_A^T \mathrm{Diag}\,(\lambda(A)) U_A$ is the spectral decomposition of A, and the elements of $\lambda(A)$ are in nondecreasing order. We have $\lambda(U_A^T Y U_A) = \lambda(B)$, and by the Schur-Horn Theorem [24], we have

$$\mathrm{diag}\,(U_A^T Y U_A) \preceq \lambda(B). \tag{4.2.98}$$

For each $p = 1, 2, \ldots, n-1$, let Γ_p denote the index set $\{n-p+1, n-p+2, \ldots, n\}$, i.e., the last p indices of $\{1, 2, \ldots, n\}$, and define the vector δ^p as

$$\delta_i^p = \begin{cases} 1 & \text{if } i \in \Gamma_p \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{array}{l} \mathrm{diag}\,(U_A^T Y U_A) \preceq \lambda(B) \Rightarrow \\ \qquad \langle \delta^p, \mathrm{diag}\,(U_A^T Y U_A) \rangle \geq \langle \delta^p, \lambda(B) \rangle_-, \qquad \text{for } \; p = 1, 2, \ldots, n-1. \end{array} \tag{4.2.99}$$

Now we take $\langle \delta^p, \mathrm{diag}\,(U_A^T Y U_A) \rangle \geq \langle \delta^p, \lambda(B) \rangle_-$, $p = 1, 2, \ldots, n-1$, as an approximation to the constraint $\mathrm{diag}\,(U_A^T Y U_A) \preceq \lambda(B)$, and get an even stronger relaxation than (4.1.97) as

$$\begin{array}{lll} \mu_{14}^* := & \min & \langle A, Y \rangle + \langle C, X \rangle \\ & \text{s.t.} & Xe = X^T e = e \\ & & X \geq 0 \\ & & \mathrm{diag}\,(Y) = X \mathrm{diag}\,(B) \\ & & \mathrm{diag}\,(Z) = X \mathrm{diag}\,(B^2) \\ & & Ye = XBe \\ & & Ze = XB^2 e \\ & & \langle \delta^p, \mathrm{diag}\,(U_A^T Y U_A) \rangle \geq \langle \delta^p, \lambda(B) \rangle_-, \; p = 1, 2, \ldots, n-1 \\ & & \begin{pmatrix} I & X^T & B^T X^T \\ X & I & Y \\ XB & Y & Z \end{pmatrix} \succeq 0 \\ & & X \in \mathcal{M}^{nn}, \; Y, Z \in \mathcal{S}^n. \end{array} \tag{4.2.100}$$

Note that, after taking the inner-product with e, the constraint $\text{diag}\,(Y) = X\text{diag}\,(B)$ implies $\text{trace}\,(Y) = \text{trace}\,(B)$. So there is no need to include the constraint $\text{trace}\,(Y) = \text{trace}\,(B)$.

Lemma 4.2.1 $\quad u_{14}^* \geq \langle \lambda(A), \lambda(B)\rangle_- \;+ \min_{Xe=X^Te=e, X\geq 0} \langle C, X\rangle,$

i.e. the bound of (4.2.100) *is no weaker than the eigenvalue bound, EB, proposed in [14, 44].*

Proof. It is enough to show that the first term of the objective of (4.2.100) satisfies

$$\langle A, Y\rangle \;\geq\; \langle \lambda(A), \lambda(B)\rangle_-,$$

for any Y feasible in (4.2.100) . Note that

$$\langle A, Y\rangle = \langle U_A \text{Diag}\,(\lambda(A)) U_A^T, Y\rangle = \langle \lambda(A), \text{diag}\,(U_A^T Y U_A)\rangle,$$

and

$$\lambda(A) = \sum_{p=1}^{n-1} (\lambda_{n-p+1}(A) - \lambda_{n-p}(A))\delta^p + \lambda_1(A)e.$$

Therefore we have

$$\langle A, Y\rangle = \sum_{p=1}^{n-1} (\lambda_{n-p+1}(A) - \lambda_{n-p}(A))\langle \delta^p, \text{diag}\,(U_A Y U_A^T)\rangle + \lambda_1(A)\langle e, \lambda(B)\rangle.$$

Since $\langle \delta^p, \text{diag}\,(U_A^T Y U_A)\rangle \geq \langle \delta^p, \lambda(B)\rangle_-$, $p = 1, 2, \ldots, n-1$, holds for any Y feasible to (4.2.100), we deduce that

$$\langle A, Y\rangle \geq \sum_{p=1}^{n-1} (\lambda_{n-p+1}(A) - \lambda_{n-p}(A))\langle \delta^p, \lambda(B)\rangle_- + \lambda_1(A)\langle e, \lambda(B)\rangle_-.$$

By ordering the entries of $\lambda(B)$ into nondecreasing order, $\lambda_{j(1)}(B) \leq \lambda_{j(2)}(B) \leq \ldots \leq \lambda_{j(n)}(B)$, we get $\langle \delta^p, \lambda(B)\rangle_- = \sum_{i=1}^p \lambda_{j(i)}(B)$, i.e. $\langle \delta^p, \lambda(B)\rangle$ is equal to the sum of the p minimal entries of $\lambda(B)$. Therefore, for any Y feasible to (4.2.100), we have

$$\begin{array}{rcl}\langle A, Y\rangle & \geq & \sum_{p=1}^{n-1} (\lambda_{n-p+1}(A) - \lambda_{n-p}(A))\langle \delta^p, \lambda(B)\rangle_- + \lambda_1(A)\langle e, \lambda(B)\rangle_- \\ & = & \sum_{p=1}^{n-1} (\lambda_{n-p+1}(A) - \lambda_{n-p}(A)) \sum_{i=1}^p \lambda_{j(i)}(B) + \lambda_1(A) \sum_{i=1}^n \lambda_{j(i)}(B) \\ & = & \sum_{i=1}^n \lambda_{j(i)}(B)(\sum_{p=i}^{n-1} \lambda_{n-p+1}(A) - \lambda_{n-p}(A)) + \lambda_1(A)) \\ & = & \langle \lambda(A), \lambda(B)\rangle_-.\end{array}$$

■

4.3 Projected Bound MSR

The row and column sum equality constraints of $\boldsymbol{QAP}$, $\mathcal{E} = \{X \in \mathcal{M}^{nn} : Xe = X^T e = e\}$, can be eliminated using a null-space method.

Proposition 4.3.1 ([21]) *Let $V \in \mathcal{M}^{n,n-1}$ be full column rank and satisfy $V^T e = 0$. Then $X \in \mathcal{E} \cap \mathcal{O}$ if and only if*

$$X = \frac{1}{n}E + V\hat{X}V^T, \text{ for some } \hat{X} \in \mathcal{O}. \tag{4.3.101}$$

■

After substituting X with (4.3.101), and denoting $\hat{A} = V^T AV$, $\hat{B} = V^T BV$, the $\boldsymbol{QAP}$ can now be reformulated as the projected version ($\boldsymbol{PQAP}$)

$$(PQAB) \quad \begin{array}{ll} \min & \operatorname{trace}(\hat{A}\hat{X}\hat{B}\hat{X}^T + \frac{1}{n}\hat{A}\hat{X}\hat{B}E + \frac{1}{n}\hat{A}E\hat{B}\hat{X}^T + \frac{1}{n^2}\hat{A}E\hat{B}E) \\ \text{s.t.} & \hat{X}\hat{X}^T = \hat{X}^T\hat{X} = I \\ & X(\hat{X}) = \frac{1}{n}E + V\hat{X}V^T \geq 0. \end{array}$$

We now define $\hat{Y} = \hat{X}\hat{B}\hat{X}^T$, and $\hat{Z} = \hat{Y}\hat{Y} = \hat{X}\hat{B}\hat{B}\hat{X}^T$; and we replace X with $\frac{1}{n}E + V\hat{X}V^T$. Then the terms XBX, $XBVV^TBX^T$ admit linear representations as

$$XBX^T = V\hat{X}\hat{B}\hat{X}^T V^T + \frac{1}{n}EBV\hat{X}^T V^T + \frac{1}{n}V\hat{X}V^T BE + \frac{1}{n^2}E\hat{B}E,$$

$$XBVV^TBX^T = V\hat{Z}V^T + \frac{1}{n}EBVV^TBVX^TV^T + \frac{1}{n}VXV^TBVV^TBE + \frac{1}{n^2}EBVV^TBE.$$

In (4.2.100) , we use Y to represent/approximate XBX^T, and use Z to represent/approximate $XBBX^T$. However, $XBBX^T$ cannot be linearly represented with $\hat{X}, \hat{Y}$. Therefore, in the projected version, we have to let Z represent $XBVV^TBX^T$ instead of $XBBX^T$, and replace the corresponding diagonal constraint with $\operatorname{diag}(Z) = X\operatorname{diag}(BVV^TB)$.

Based on these definitions, ***PQAP*** can be formulated as

$$\begin{array}{ll}
\min & \operatorname{trace}(AY + CX) \\
\text{s.t.} & \operatorname{diag} Y = X \operatorname{diag}(B) \\
& \operatorname{diag} Z = X \operatorname{diag}(BVV^T B) \\
& X(\hat{X}) = V\hat{X}V^T + \frac{1}{n}E \\
& Y(\hat{X}, \hat{Y}) = V\hat{Y}V^T + \frac{1}{n}EBV\hat{X}^T V^T + \frac{1}{n}V\hat{X}V^T BE + \frac{1}{n^2}E\hat{B}E \\
& Z(\hat{X}, \hat{Z}) = V\hat{Z}V^T + \frac{1}{n}EBVV^T BVX^T V^T + \frac{1}{n}VXV^T BVV^T BE + \frac{1}{n^2}EBVV^T BE \\
& \hat{R} = \hat{X}\hat{B} \\
& \begin{pmatrix} I & \hat{Y} \\ \hat{Y} & \hat{Z} \end{pmatrix} = \begin{pmatrix} \hat{X}\hat{X}^T & \hat{X}\hat{R}^T \\ \hat{R}\hat{X}^T & \hat{R}\hat{R}^T \end{pmatrix} \\
& X(\hat{X}) \geq 0 \\
& \hat{X}, \hat{R} \in \mathcal{M}^{n-1}, \; \hat{Y}, \hat{Z} \in \mathcal{S}^{n-1}.
\end{array}$$

We still use idea of the matrix-lifting semidefinite relaxation, relaxing the nonconvex quadratic equality constraint

$$\begin{pmatrix} I & \hat{Y} \\ \hat{Y} & \hat{Z} \end{pmatrix} = \begin{pmatrix} \hat{X}\hat{X}^T & \hat{X}\hat{R}^T \\ \hat{R}\hat{X}^T & \hat{R}\hat{R}^T \end{pmatrix}$$

into the semidefinite constraint

$$\begin{pmatrix} I & \hat{X}^T & \hat{R}^T \\ \hat{X} & I & \hat{Y} \\ \hat{R} & \hat{Y} & \hat{Z} \end{pmatrix} \succeq 0.$$

And as in (4.2.100) , we add cuts

$$\langle \delta^p, \operatorname{diag}(U_{\hat{A}}^T \hat{Y} U_{\hat{A}}\rangle \geq \langle \delta^p, \lambda(\hat{B})\rangle, \quad p = 1, 2, \ldots, n-2$$

for $\hat{Y} \in \operatorname{conv} \mathcal{O}(\hat{X})$, where $\hat{A} = U_{\hat{A}} \operatorname{Diag}(\lambda(\hat{A})) U_{\hat{A}}^T$ is the spectral decomposition of $\hat{A}$, and $\lambda_1(\hat{A}) \leq \lambda_2(\hat{A}) \leq \ldots \leq \lambda_n(\hat{A})$. Let δ^p follow the definition in section 4.2, i.e., $\delta^p \in \mathbb{R}^{n-1}$, $\delta^p = \{0, 0, \ldots, 0, 1, \ldots, 1\}^T$. Then we reach the final relaxation formulation named $\boldsymbol{MSR}$,

$$
(\boldsymbol{MSR}) \quad
\begin{array}{rll}
\mu_{15}^* := & \min & \langle A, Y(\hat{X},\hat{Y})\rangle + \langle C, X(\hat{X})\rangle \\
& \text{s.t.} & \operatorname{diag}(Y(\hat{X},\hat{Y})) = X(\hat{X})\operatorname{diag}(B) \\
& & \operatorname{diag}(Z(\hat{X},\hat{Z})) = X(\hat{X})\operatorname{diag}(BVV^TB) \\
& & \langle \delta^p, \operatorname{diag}(U_{\hat{A}}^T \hat{Y} U_{\hat{A}}\rangle \geq \langle \delta^p, \lambda(\hat{B})\rangle, \; p = 1,2,\ldots,n-2 \\
& & X(\hat{X}) \geq 0 \\
& & \begin{pmatrix} I & \hat{X}^T & \hat{B}^T\hat{X}^T \\ \hat{X} & I & \hat{Y} \\ \hat{X}\hat{B} & \hat{Y} & \hat{Z} \end{pmatrix} \succeq 0 \\
& & \hat{X} \in \mathcal{M}^{n-1}, \; \hat{Y}, \hat{Z} \in \mathcal{S}^{n-1},
\end{array}
$$

where

$$
\begin{array}{rcl}
X(\hat{X}) & = & \frac{1}{n}E + V\hat{X}V^T, \\
Y(\hat{X},\hat{Y}) & = & V\hat{Y}V^T + \frac{1}{n}EBV\hat{X}^TV^T + \frac{1}{n}V\hat{X}V^TBE + \frac{1}{n^2}E\hat{B}E, \\
Z(\hat{X},\hat{Z}) & = & V\hat{Z}V^T + \frac{1}{n}EBVV^TBV\hat{X}^TV^T + \frac{1}{n}V\hat{X}V^TBVV^TBE + \frac{1}{n^2}EBVV^TBE.
\end{array}
$$

Notice $Ye = XBe$, $Ze = XB^2e$ are no longer included in $\boldsymbol{MSR}$, because any X, Y, Z as the projection of $\hat{X}$, $\hat{Y}$, $\hat{Z}$ automatically satisfy these linear equalities.

In $\boldsymbol{MSR}$, all the constraints act on the lower dimensional space. Such a strategy, i.e., first a projection followed by a cut, has also been used in the projected eigenvalue bound PB and the quadratic programming bound QPB. It is numerically superior to directly adding cuts to the high-dimensional image space of the projection, e.g., the projected eigenvalue bound PB is much stronger than the eigenvalue bound EB. For this reason, we propose using $\boldsymbol{MSR}$ instead of (4.2.100) .

4.4 Comparing Bounds for QAPLIB Problems

Table 4.1 is a comparison of bounds obtained from $\boldsymbol{MSR}$ and other relaxation methods applied to the instances from QAPLIB [8]. The first column OPT denotes the exact optimal value of the problem instance, while the following columns contain the lower bounds from the relaxation methods: GLB, the Gilmore-Lawler bound [17]; KCCEB, the dual linear programming bound [28]; PB, the projected eigenvalue bound [21]; QPB, the convex quadratic programming bound [1]; SDR1, SDR2, SDR3, the three vector-lifting semidefinite relaxation bounds [54] computed by the bundle method [43]; the last column is our $\boldsymbol{MSR}$. All output values are rounded up to the nearest integer.

To solve $\boldsymbol{QAP}$, the minimization of trace $AXBX^T$ and trace $BXAX^T$ are equivalent. But for the relaxation $\boldsymbol{MSR}$, exchanging the roles of A and B results in two different formulations and bounds. In our tests we use both versions and take the larger output as the bound of $\boldsymbol{MSR}$. We then keep the better formulation throughout the branch and bound process, so that we do not double the computational work.

From Table 4.1, we see that the relative performances between the $\boldsymbol{LP}$-based bounds GLB, KCCEB are unpredictable. At some instances, both are weaker than even the least expensive PB bounds. For the other bounds, the *average* performance can be ranked as follows: $PB < QPB < \boldsymbol{MSR} \approx SDR1 < SDR2 < SDR3$.

By comparing the number of variables and constraints of QPB and different $\boldsymbol{SDP}$ methods in Table 4.2, we see $\boldsymbol{MSR}$ uses $O(n^2)$ variables and $O(n^2)$ constraints, which is the same with QPB and strictly less than other vector-lifting $\boldsymbol{SDP}$ methods. If we solve $\boldsymbol{MSR}$ or QPB with an interior point method, the complexity of computing the Newton direction in each iteration are both $O(n^6)$. And, the number of iterations of an interior point method is bounded by $O(\sqrt{n^2}\ln\frac{1}{\epsilon}) = O(n\ln\frac{1}{\epsilon})$ [36]. So the complexity of computing $\boldsymbol{MSR}$ or QPB with an interior point methods to accuracy ϵ are both $O(n^7\ln\frac{1}{\epsilon})$. Note that for the most expensive $\boldsymbol{SDP}$ formulation, SDP3, the computation complexity is $O(n^{14}\ln\frac{1}{\epsilon})$ for ϵ accuracy. So $\boldsymbol{MSR}$ is significantly faster than SDR3. Compared with QPB, computing $\boldsymbol{MSR}$ is still slower, but their complexity with respect to the order of n is the same.

Table 4.3 listed the CPU time (in seconds) for $\boldsymbol{MSR}$ to compute some of the *Nugent* instances on a sun4c UNIX machine. Here we use the well known $\boldsymbol{SDP}$ package SeDuMi[1].

We believe the speed can be much improved by using more specific software. So $\boldsymbol{MSR}$ is an excellent relaxation algorithm and can be used under the Branch and Bound frame and solve the $\boldsymbol{QAP}$ to optimality.

As a conclusion of this chapter, the idea of matrix-lifting semidefinite relaxation proposed in Chapter 3 has found an important application in efficiently solving the $\boldsymbol{QAP}$ problem. We believe that there are more examples where the matrix-lifting semidefinite relaxation can contribute, and are still studying those examples.

[1] sedumi.mcmaster.ca

Problem	OPT	GLB	KCCEB	PB	QPB	SDR1	SDR2	SDR3	MSR
esc16a	68	38	41	47	55	47	49	59	50
esc16b	292	220	274	250	250	250	275	288	276
esc16c	160	83	91	95	95	95	111	142	123
esc16d	16	3	4	-19	-19	-19	-13	8	1
esc16e	28	12	12	6	6	6	11	23	14
esc16g	26	12	12	9	9	9	10	20	13
esc16h	996	625	704	708	708	708	905	970	906
esc16i	14	0	0	-25	-25	-25	-22	9	0
esc16j	8	1	2	-6	-6	-6	-5	7	0
had12	1652	1536	1619	1573	1592	1604	1639	1643	1595
had14	2724	2492	2661	2609	2630	2651	2707	2715	2634
had16	3720	3358	3553	3560	3594	3612	3675	3699	3587
had18	5358	4776	5078	5104	5141	5174	5282	5317	5153
had20	692	6166	6567	6625	6674	6713	6843	6885	6681
kra30a	88900	68360	75566	63717	68257	69736	68526	77647	72480
kra30b	91420	69065	76235	63818	68400	70324	71429	81156	73155
Nug12a	578	493	521	472	482	486	528	557	502
Nug14	1014	852	n.a.	871	891	903	958	992	918
Nug15	1150	963	1033	973	994	1009	1069	1122	1016
Nug16a	1610	1314	1419	1403	1441	1461	1526	1570	1460
Nug16b	1240	1022	1082	1046	1070	1082	1136	1188	1082
Nug17	1732	1388	1498	1487	1523	1548	1619	1669	1549
Nug18	1930	1554	1656	1663	1700	1723	1798	1852	1726
Nug20	2570	2057	2173	2196	2252	2281	2380	2451	2291
Nug21	2438	1833	2008	1979	2046	2090	2244	2323	2099
Nug22	3596	2483	2834	2966	3049	3140	3372	3440	3137
Nug24	3488	2676	2857	2960	3025	3068	3217	3310	3061
Nug25	3744	2869	3064	3190	3268	3305	3438	3535	3300
Nug27	5234	3701	n.a.	4493	n.a.	n.a.	4887	4965	4621
Nug30	6124	4539	4785	5266	5362	5413	5651	5803	5446
rou12	235528	202272	223543	200024	205461	208685	219018	223680	207445
rou15	354210	298548	323589	296705	303487	306833	320567	333287	303456
rou20	725522	599948	641425	597045	607362	615549	641577	663833	609102
scr12	31410	27858	29538	4727	8223	11117	23844	29321	18803
scr15	51140	44737	48547	10355	12401	17046	41881	48836	39399
scr20	110030	86766	94489	16113	23480	28535	82106	94998	50548
tai12a	224416	195918	220804	193124	199378	203595	215241	222784	202134
tai15a	388214	327501	351938	325019	330205	333437	349179	364761	331956
tai17a	491812	412722	441501	408910	415576	419619	440333	451317	418356
tai20a	703482	580674	616644	575831	584938	591994	617630	637300	587266
tai25a	1167256	962417	1005978	956657	981870	974004	908248	1041337	970788
tai30a	1818146	1504688	1565313	1500407	1517829	1529135	1573580	1652186	1521368
tho30	149936	90578	99855	119254	124286	125972	134368	136059	122778

Table 4.1: Comparison of Bounds for QAPLIB Instances

Methods	QPB	SDR1	SDR2	SDR3	MSR
Variables	$O(n^2)$	$O(n^4)$	$O(n^4)$	$O(n^4)$	$O(n^2)$
Constraints	$O(n^2)$	$O(n^2)$	$O(n^3)$	$O(n^4)$	$O(n^2)$

Table 4.2: Complexity of Relaxations

Instances	Nug12	Nug15	Nug18	Nug20	Nug25	Nug27	Nug30
CPU time(s)	15.1	57.6	203.9	534.9	3236.4	5211.3	12206.0
Number of iterations	18	19	22	26	27	25	29

Table 4.3: CPU Time and Iterations for Computing $\boldsymbol{MSR}$ on the Nugent Problems

Chapter 5

Conclusions and Future Work

In this thesis, we discussed several important issues in solving the $\boldsymbol{QCQP}$. Solving a $\boldsymbol{QCQP}$ with few constraints to optimality and deriving inexpensive and strong relaxation bounds for large scale $\boldsymbol{QCQP}$ are of primal interests. Especially, we recognized a special class of $\boldsymbol{QCQP}$ that admits an alternative efficient semidefinite relaxation, i.e., the $\boldsymbol{QMP}$. We then divided the $\boldsymbol{QMP}$ problems into two classes, $\boldsymbol{QMP_1}$ and $\boldsymbol{QMP_2}$, and proved that vector-lifting semidefinite relaxation and matrix-lifting semidefinite relaxation are equally strong for a $\boldsymbol{QMP_1}$ or a homogeneous $\boldsymbol{QMP_2}$. At last, we studied the efficient semidefinite relaxations for the quadratic assignment problem. Based on the first class of Matrix-lifting Semidefinite Relaxation, our formulation generates competitive bounds, while spending much less time than the usual vector-lifting $\boldsymbol{SDP}$.

Future work includes further exploring the condition in Theorem 3.3.1 and making it more applicable. Also, we are interested in quadratic matrix programming including quadratic matrix inequality constraints such like $X^TQX + C^TX + X^TC + D \succeq 0$. Such optimization models arise extensively in robust optimization fields and deserve serious study.

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] W.W. BARRETT, C.R. JOHNSON, and M. LUNDQUIST. Determinantal formulae for matrix completions associated with chordal graphs. *Linear Algebra Appl.*, 121:265–289, 1989.

[3] A. BARVINOK. Problems of distance geometry and convex properties of quadratic maps. *Discrete Comput. Geom.*, 13(2):189–202, 1995.

[4] A. BECK. Quadratic matrix programming. *SIAM J. Optim.*, to appear:– (electronic), 2006.

[5] A. BEN-TAL and A.S. NEMIROVSKI. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Analysis, algorithms, and engineering applications.

[6] P.T. BOGGS and J.W. TOLLE. Sequential quadratic programming. *Acta Numerica*, 4:1–50, 1995.

[7] S. BOYD and L. VANDENBERGHE. *Convex Optimization.* Cambridge University Press, Cambridge, 2004.

[8] R.E. BURKARD, S. KARISCH, and F. RENDL. QAPLIB – a quadratic assignment problem library. *European J. Oper. Res.*, 55:115–119, 1991. www.opt.math.tu-graz.ac.at/qaplib/.

[9] M.R. CELIS, J.E. DENNIS Jr., and R.A. TAPIA. A trust region strategy for nonlinear equality constrained optimization. In *Proceedings of the SIAM Conference on Numerical Optimization, Boulder, CO*, 1984. Also available as Technical Report TR84-1, Department of Mathematical Sciences, Rice University, Houston, TX.

[10] A.R. CONN, N.I.M. GOULD, and P.L. TOINT. *Trust-Region Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[11] M.M. DEZA and M. LAURENT. *Geometry of cuts and metrics*. Springer-Verlag, Berlin, 1997.

[12] Z. DREZNER. Lower bounds based on linear programming for the quadratic assignment problem. *Comput. Optim. Appl.*, 4(2):159–165, 1995.

[13] J. FALKNER, F. RENDL, and H. WOLKOWICZ. A computational study of graph partitioning. *Math. Programming*, 66(2, Ser. A):211–239, 1994.

[14] G. FINKE, R.E. BURKARD, and F. RENDL. Quadratic assignment problems. *Ann. Discrete Math.*, 31:61–82, 1987.

[15] A.L. FRADKOV and V.A. JAKUBOVIČ. The S-procedure and the duality relation in convex quadratic programming problems. *Vestnik Leningrad. Univ.*, 1(1 Mat. Meh. Astronom. Vyp. 1):81–87, 155, 1973.

[16] A.L. FRADKOV and V.A. JAKUBOVIČ. The S-procedure and a duality relations in convex problems of quadratic programming. *Vestnik Leningrad. Univ.*, 6(1 Mat. Meh. Astronom. Vyp. 1):101–109, 1979. English, translated by K. Makowski.

[17] P.C. GILMORE. Optimal and suboptimal algorithms for the quadratic assignment problem. *SIAM Journal on Applied Mathematics*, 10:305–313, 1962.

[18] F. GLINEUR and T. TERLAKY. A conic formulation for l_p-norm optimization. Technical report, Faculte Polytechnique de Mons, Mons, Belgium, 2001. URL: http://www.optimization-online.org/DB_HTML/2001/03/292.html.

[19] G.H. GOLUB and C.F. VAN LOAN. *Matrix Computations*. The Johns Hopkins University Press, 1983.

[20] B. GRONE, C.R. JOHNSON, E. MARQUES de SA, and H. WOLKOWICZ. Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.*, 58:109–124, 1984.

[21] 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.

[22] G.H. HARDY, J.E. LITTLEWOOD, and G. POLYA. *Inequalities.* Cambridge Univesity Press, London and New York, 1934. 2nd edition 1952.

[23] K. HOFFMAN and R. KUNZE. *Linear algebra, Second edition.* Prentice-Hall, Inc, Englewood Cliffs, N.J., 1971.

[24] R. HORN and C. JOHNSON. *Matrix Analysis.* Cambridge University Press, New York, 1985.

[25] V. JEYAKUMAR and H. WOLKOWICZ. Generalizations of Slater's constraint qualification for infinite convex programs. *Math. Programming*, 57(1, Ser. B):85–101, 1992.

[26] C.R. JOHNSON and R.L. SMITH. The positive semidefinite completion problem relative to a subspace. *Linear Algebra Appl.*, 307:1–14, 2000.

[27] L.V. KANTOROVICH and G.P. AKILOV. *Functional analysis.* Pergamon Press, Oxford, second edition, 1982. Translated from the Russian by Howard L. Silcock.

[28] S.E. KARISCH, E. ÇELA, J. CLAUSEN, and T. ESPERSEN. A dual framework for lower bounds of the quadratic assignment problem based on linearization. *Computing*, 63(4):351–403, 1999.

[29] M. KOJIMA and L. TUNÇEL. Cones of matrices and successive convex relaxations of nonconvex sets. *SIAM J. Optim.*, 10(3):750–778, 2000.

[30] T.C. KOOPMANS and M.J. BECKMANN. Assignment problems and the location of economic activities. *Econometrica*, 25:53–76, 1957.

[31] E. LAWLER. The quadratic assignment problem. *Management Sci.*, 9:586–599, 1963.

[32] Z-Q. LUO, J. F. STURM, and S. ZHANG. Multivariate nonnegative quadratic mappings. *SIAM J. Optim.*, 14(4):1140–1162, 2004.

[33] H. LÜTKEPOHL. *Handbook of matrices.* John Wiley & Sons Ltd., Chichester, 1996.

[34] A.W. MARSHALL and I. OLKIN. *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York, NY, 1979.

[35] J.M. MARTINEZ. Local minimizers of quadratic functions on Euclidean balls and spheres. *SIAM J. Optim.*, 4(1):159–176, 1994.

[36] R.D.C. MONTEIRO and M.J. TODD. Path-following methods. In *Handbook of Semidefinite Programming*, pages 267–306. Kluwer Acad. Publ., Boston, MA, 2000.

[37] J. NOCEDAL and S.J. WRIGHT. *Numerical optimization*. Springer-Verlag, New York, 1999.

[38] P.M. PARDALOS, K.G. RAMAKRISHNAN, M.G.C. RESENDE, and Y. LI. Implementation of a variance reduction-based lower bound in a branch-and-bound algorithm for the quadratic assignment problem. *SIAM J. Optim.*, 7(1):280–294, 1997.

[39] G. PATAKI. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.*, 23(2):339–358, 1998.

[40] G. PATAKI. Geometry of Semidefinite Programming. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *HANDBOOK OF SEMIDEFINITE PROGRAMMING: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Boston, MA, 2000.

[41] I. POLIK and T. TERLAKY. A comprehensive study of the s-lemma. Technical report, Advanced Optimization Laboratory, McMaster University, Hamilton, Ontario, 2004.

[42] S. POLJAK, F. RENDL, and H. WOLKOWICZ. A recipe for semidefinite relaxation for $(0,1)$-quadratic programming. *J. Global Optim.*, 7(1):51–73, 1995.

[43] F. RENDL and R. SOTIROV. Bounds for the quadratic assignment problem using the bundle method. *Math. Programming*, to appear.

[44] F. RENDL and H. WOLKOWICZ. Applications of parametric programming and eigenvalue maximization to the quadratic assignment problem. *Math. Programming*, 53(1, Ser. A):63–78, 1992.

[45] R.T. ROCKAFELLAR. *Convex Analysis*. Princeton University Press, Princeton, NJ, USA, 1970.

[46] A. SO, Y. YE, and J. ZHANG. A unified theorem on SDP rank reduction. Technical report, Operation Research, Stanford University, Stanford, CA, 2006.

[47] G. SRIJUNTONGSIRI and S. VAVASIS. A fully sparse implementation of a primal-dual interior-point potential reduction method for semidefinite programming. Technical report, Cornell University, Ithaca, NY, 2004. Available at URL: `http:www.optimization-online.orgDB_HTML2004121015.html`.

[48] R. STERN and H. WOLKOWICZ. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM J. Optim.*, 5(2):286–313, 1995.

[49] G.W. STEWART and J-G. SUN. *Matrix Perturbation Theory.* Academic Press, 1990.

[50] J.F. STURM and S. ZHANG. On cones of nonnegative quadratic functions. Technical report, Tilburg University, The Netherlands, 2001.

[51] Z. WANG, S. ZHENG, S. BOYD, and Y. YE. Further relaxations of the SDP approach to sensor network localization. Technical report, Stanford University, Stanford, CA, 2006.

[52] V.A. YAKUBOVICH. The S-procedure and duality theorems for nonconvex problems of quadratic programming. *Vestnik Leningrad. Univ.*, 1973(1):81–87, 1973.

[53] Y. YE and S. ZHANG. New results on quadratic minimization. Technical Report SEEM2001-03, Department of Systems Engineering & Engineering Management The Chinese University of Hong Kong, Hong Kong, 2001.

[54] 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).