A Note on Lack of Strong Duality for Quadratic Problems with Orthogonal Constraints*

Henry Wolkowicz†

December 5, 2001

Contents

1 Introduction **2**
1.1 Notation 3

2 Lagrangian Duals **3**
2.1 Lagrangian Duals; the Homogeneous Case 4
2.2 Lagrangian Dual; the General Case 7
2.2.1 Schur Complement 9
2.3 The Linear Case 10

3 Examples with Duality Gaps **11**

4 Conclusion **12**

Abstract

The general quadratically constrained quadratic program (QQP) is an important modelling tool for many diverse problems. The QQP is in general NP hard, and numerically intractable. Lagrangian relaxations often provide good approximate solutions to these hard problems. Such relaxations are equivalent to semidefinite programming (SDP) relaxations and can be solved efficiently.

For several special cases of QQP, the Lagrangian relaxation provides the exact optimal value. This means that there is a zero duality gap and the problem is tractable. It is important to know for which cases this is true, since they can then be used as subproblems to improve Lagrangian relaxation for intractable QQPs.

*Department of Combinatorics and Optimization, Waterloo, Ontario N2L 3G1, Canada, Research Report CORR 99-59.

†Research partially supported by The Natural Sciences Engineering Research Council Canada.

In this paper we study the special QQP with orthogonal (matrix) constraints $XX^T = I$. If $C = 0$, the zero duality gap result holds if the redundant orthogonal constraints $X^TX = I$ are added. We show that this is not true in the general case. However, we show how to close the duality gap in the pure linear case by adding variables in addition to constraints.

Keywords Quadratic Objective, Orthogonal Constraints, Semidefinite Programming, Lagrangian Relaxation, Redundant Constraints, Strong Duality, Procrustes Problem.

1 Introduction

We study the quadratic (matrix) program with orthogonal constraints

$$\text{QQP}_{\mathcal{O}} \qquad \mu^{\mathcal{O}} := \begin{array}{ll} \min & \text{Trace}\, AXBX^T - 2CX^T \\ \text{s.t.} & XX^T = I, \end{array} \tag{1.1}$$

where A, B are $n \times n$ symmetric matrices and C is $n \times n$. This constraint set is often called the *Stiefel manifold*, [13]. If the objective function is written as $||AY - XB||^2$, with both X, Y orthogonal, then this is the *orthogonal Procrustes Problem.* (See e.g. [11, 5, 13] for references, theory, and applications.)

The special case that $C = 0$ in (1.1), the homogeneous case, is studied in [2]. These problems arise as orthogonal relaxations of the quadratic assignment and graph partitioning problems, e.g. [6, 1, 15]. It is shown that the resulting, well-known, eigenvalue bounds for these problems can be obtained from the Lagrangian dual of the orthogonally constrained relaxations, but only if the *redundant constraint* $X^TX = I$ is explicitly added to the orthogonality constraint $XX^T = I$.

In this paper we show that this strong duality result does *not* hold if $C \neq 0$. Nor does it hold for the pure linear case. We do this with a simple counterexample in the purely linear case using the property that the dual is independent of the signs of the individual components of C, see Lemma 2.4; while the optimal value of the primal is based on the sum of the singular values of C, see Example 3.2. We then show how to close the duality gap in the pure linear case by adding variables in addition to adding constraints. (See Theorem 2.6.)

We still leave open the question: what modifications are required to the constraints and/or variables to close the duality gap for the general case. One purpose of the paper is to present an approach that might lead to closing this duality gap. This approach is outlined in the second proof of Theorem 2.2.

The paper is organized as follows. We complete this section with notation in §1.1. In §2, we derive optimality conditions as well as the dual of (1.1). We do this in stages starting with the homogeneous case in §2.1 and proceeding to the general case in §2.2. We specialize this to the linear case in §2.3, where we also show how to close the duality gap. The main results are: in §2.1 we present a new proof of strong duality for the homogeneous case; in §2.3 we also show how to close the duality gap in the pure linear case; and in §3 we present the examples with the duality gaps. We summarize our results in §4.

1.1 Notation

We work in the space of real $n \times n$ matrices, $\mathcal{M}_n$, with the trace inner product, $\langle M, N \rangle = \text{Trace } M^T N$. The subspace of symmetric matrices is denoted $\mathcal{S}_n$. This space is equipped with the Löwner partial order, i.e. $A \succeq B$ denotes $A - B$ is positive semidefinite.

We will use several linear operators, e.g. vec (X) denotes the vector formed (columnwise) from the matrix X. The adjoint of a linear operator $\mathcal{A}$ is denoted $\mathcal{A}^*$, i.e. the adjoint satisfies

$$\langle \mathcal{A}x, y \rangle = \langle x, \mathcal{A}^* y \rangle, \quad \forall x, y.$$

2 Lagrangian Duals

The simplest example of a quadratic constrained quadratic problem is the eigenvalue problem. Let A be an $n \times n$ symmetric matrix. Then the Rayleigh Principle yields the following formulation of the smallest eigenvalue.

$$\lambda_{\min}(A) = \min_{x^T x = 1} q(x) \quad (= x^T A x).$$

This result can be proved easily using Lagrange multipliers, i.e. the optimum x must be a stationary point of the Lagrangian $q(x) + \lambda \left(1 - x^T x\right)$. We can get an equivalent semidefinite programming (SDP) problem using Lagrangian duality and relaxation. Note that

$$\begin{aligned}
\lambda_{\min}(A) &= \min_x \max_\lambda x^T A x + \lambda \left(1 - x^T x\right) && (2.1)\\
&\geq \max_\lambda \min_x x^T A x + \lambda \left(1 - x^T x\right) && (2.2)\\
&= \max_{A - \lambda I \succeq 0} \min_x x^T (A - \lambda I) x + \lambda && (2.3)\\
&= \max_{A - \lambda I \succeq 0} \lambda && (2.4)\\
&= \lambda_{\min}(A). && (2.5)
\end{aligned}$$

This follows from the hidden constraints, i.e. the inner problems have hidden constraints. For example, $A - \lambda I \succeq 0$ arises since a homogeneous quadratic function in bounded below if and only if the Hessian is positive semidefinite.

Note that the above strong duality result still holds if the quadratic objective function $q(x)$ has a linear term. This case is called *the Trust Region Subproblem, TRS*. (See [12, Theorem 5.1] for the strong duality theorem.) This problem is also equivalent to a max-min eigenvalue problem, see [10], which is another way to see that the problem is tractable. However, strong duality can fail if there are two constraints, i.e. the so-called CDT problem [4]. Thus we see that going from one to two constraints, even if both constraints are convex, can result in a duality gap. However, we will see below (Theorem 2.2) that strong duality will hold for a nonconvex problem with $\binom{n+1}{2} = n(n+1)/2$ constraints.

2.1 Lagrangian Duals; the Homogeneous Case

We now consider our more general problem (1.1), but with $C = 0$. Because of the similarity of the orthogonality constraint to the norm constraint $x^T x = 1$, the result of this section can be viewed as a matrix generalization of the strong duality result for the Rayleigh Principle given above.

$$\text{QQPH}_{\mathcal{O}} \qquad \mu_H^{\mathcal{O}} := \begin{array}{ll} \min & \text{Trace}\, AXBX^T \\ \text{s.t.} & XX^T = I. \end{array} \tag{2.6}$$

Though this is a nonconvex problem with many nonconvex constraints, this problem can be solved efficiently using Lagrange multipliers and eigenvalues, see e.g. [7], or using the classical Hoffman-Wielandt inequality, e.g. [3]. The optimal value is the so-called minimal scalar product of the eigenvalues of A and B. We include a simple proof for completeness using Lagrange multipliers. As was done for the ordinary eigenvalue problem above, we note that Lagrange multipliers can be used in two ways. First, one can use them in the necessary conditions (Karush-Kuhn-Tucker) for optimality, i.e. in the stationarity of the Lagrangian. This is how we apply them now. (The other use is in Lagrangian duality or Lagrangian relaxation where the Lagrangian is positive semidefinite. This is done below.) Also, the Lagrange multipliers here are symmetric matrices since the image of the constraint $XX^T - I$ is a symmetric matrix.

Proposition 2.1 *Suppose that the orthogonal diagonalizations of A, B are $A = V\Sigma V^T$ and $B = U\Lambda U^T$, respectively, where the eigenvalues in Σ are in nonincreasing order, and the eigenvalues in Λ are in nondecreasing order. Then the optimal value of* $\text{QQPH}_{\mathcal{O}}$ *is $\mu_H^{\mathcal{O}} =$* $\text{Trace}\, \Sigma\Lambda$*, and the optimal solution is obtained using the orthogonal matrices that yield the diagonalizations, i.e. $X^* = VU^T$.*

Proof. The constraint $G(X) := XX^T - I$ maps $\mathcal{M}_n$ to $\mathcal{S}_n$. The Jacobian of the constraint at X acting on the direction h is $J(X)(h) = Xh^T + hX^T$. (This can be found by simple expansion and neglecting the second order term.) The adjoint of the Jacobian acting on $S \in \mathcal{S}_n$ is $J^*(X)(S) = 2SX$, since

$$\text{Trace}\, SJ(X)(h) = \text{Trace}\, h^T J^*(X)(S).$$

But $J^*(X)(S) = 0$ implies $S = 0$, i.e. J^* is one-one for all X orthogonal. Therefore J is onto, i.e. the standard constraint qualification holds at the optimum. It follows that the necessary conditions for optimality are that the gradient of the Lagrangian

$$L(X, S) = \text{Trace}\, AXBX^T - \text{Trace}\, S(XX^T - I), \tag{2.7}$$

is 0, i.e.

$$AXB - SXI = 0.$$

Therefore,

$$AXBX^T = S = S^T,$$

i.e. $AXBX^T$ is symmetric, which means that A and XBX^T commute and so are mutually diagonalizable by the orthogonal matrix U. Therefore, we can assume that both A and B are

diagonal and we choose X to be a product of permutations that gives the correct ordering of the eigenvalues. ■

The second use of Lagrange multipliers is in forming the Lagrangian dual. The Lagrangian dual of $\mathrm{QQPH}_{\mathcal{O}}$ is

$$\max_{S=S^T} \min_{X} \operatorname{Trace} AXBX^T - \operatorname{Trace} S(XX^T - I). \tag{2.8}$$

However, there can be a nonzero duality gap for the Lagrangian dual, see [16] and Example 3.1 below. The inner minimization in the dual problem (2.8) is an unconstrained quadratic minimization in the variables $x = \operatorname{vec}(X)$, with Hessian

$$B \otimes A - I \otimes S.$$

We apply the hidden semidefinite constraint again. This minimization is unbounded only if the Hessian is not positive semidefinite. To close the duality gap, we need a larger class of quadratic functions. Note that in $\mathrm{QQPH}_{\mathcal{O}}$ the constraints $XX^T = I$ and $X^TX = I$ are equivalent. We add the redundant constraints $X^TX = I$ and arrive at

$$\mathrm{QQPH}_{\mathcal{OO}} \qquad \begin{array}{rll} \mu_H^{\mathcal{O}} := & \min & \operatorname{Trace} AXBX^T \\ & \text{s.t.} & XX^T = I,\ X^TX = I. \end{array} \tag{2.9}$$

We can use symmetric matrices S and T to relax the constraints $XX^T = I$ and $X^TX = I$, respectively. We obtain a dual problem

$$\mathrm{DQQPH}_{\mathcal{OO}} \qquad \begin{array}{rll} \mu_H^{\mathcal{O}} \geq \nu_H^{\mathcal{O}} := & \max & \operatorname{Trace} S + \operatorname{Trace} T \\ & \text{s.t.} & (I \otimes S) + (T \otimes I) \preceq (B \otimes A) \\ & & S = S^T,\ T = T^T. \end{array}$$

We now prove the strong duality presented in [2] for the case $C = 0$. We include two proofs. The first proof is from [2]. It uses the well known strong duality for LAP, the linear assignment problem, and the known optimal value from Proposition 2.1. The second proof exploits the LAP duality results from the first proof. We include this second proof because it illustrates where convexity and complementary slackness arise without using Proposition 2.1. If we hope to obtain a strong duality result for the general case, then these are the sufficient optimality conditions that we need to satisfy, i.e. we have the curious statement: these are the necessary *sufficient optimality conditions*. We now present the strong duality theorem.

Theorem 2.2 *Strong duality holds for* $\mathrm{QQPH}_{\mathcal{OO}}$ *and* $\mathrm{DQQPH}_{\mathcal{OO}}$, *i.e.* $\mu_H^{\mathcal{O}} = \nu_H^{\mathcal{O}}$ *and both primal and dual are attained.*

Proof I. Let $A = V\Sigma V^T$, $B = U\Lambda U^T$, where V and U are orthonormal matrices whose columns are the eigenvectors of A and B, respectively, σ and λ are the corresponding vectors of eigenvalues, and $\Sigma = \operatorname{diag}(\sigma)$, $\Lambda = \operatorname{diag}(\lambda)$. Then for any S and T,

$$(B \otimes A) - (I \otimes S) - (T \otimes I) = (U \otimes V)\left[(\Lambda \otimes \Sigma) - (I \otimes \bar{S}) - (\bar{T} \otimes I)\right](U^T \otimes V^T),$$

where $\bar{S} = V^T S V$, $\bar{T} = U^T T U$. Since $U \otimes V$ is nonsingular, $\operatorname{Trace} S = \operatorname{Trace} \bar{S}$ and $\operatorname{Trace} T = \operatorname{Trace} \bar{T}$, the dual problem DQQP$_{\mathcal{OO}}$ is equivalent to

$$\begin{array}{rll} \nu_H^{\mathcal{O}} = & \max & \operatorname{Trace} S + \operatorname{Trace} T \\ & \text{s.t.} & (\Lambda \otimes \Sigma) - (I \otimes S) - (T \otimes I) \succeq 0 \\ & & S = S^T,\ T = T^T. \end{array} \tag{2.10}$$

However, since Λ and Σ are diagonal matrices, (2.10) is equivalent to the ordinary linear program:

$$\begin{array}{lll} \text{LD} & \max & e^T s + e^T t \\ & \text{s.t.} & \lambda_i \sigma_j - s_j - t_i \geq 0, \quad i, j = 1, \ldots, n. \end{array}$$

But LD is the dual of the linear assignment problem:

$$\begin{array}{lll} \text{LP} & \min & \displaystyle\sum_{i,j} \lambda_i \sigma_j y_{ij} \\ & \text{s.t.} & \displaystyle\sum_{j=1}^{n} y_{ij} = 1, \quad i = 1, \ldots, n \\ & & \displaystyle\sum_{i=1}^{n} y_{ij} = 1, \quad j = 1, \ldots, n \\ & & y_{ij} \geq 0, \quad i, j = 1, \ldots, n. \end{array}$$

Assume without loss of generality that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. Then LP can be interpreted as the problem of finding a permutation $\pi(\cdot)$ of $\{1, \ldots, n\}$ so that $\sum_{i=1}^{n} \lambda_i \sigma_{\pi(i)}$ is minimized. But the minimizing permutation is then $\pi(i) = i$, $i = 1, \ldots, n$, and from Proposition 2.1 the solution value $\nu_H^{\mathcal{O}}$ is exactly $\mu_H^{\mathcal{O}}$.

Proof II. Using the above notation in Proof I, we diagonalize A and B. We can write (2.9) with diagonal matrices, i.e.

$$\text{QQP}_{\mathcal{OO}} \qquad \begin{array}{rll} \mu^{\mathcal{O}} := & \min & \operatorname{Trace} V \Sigma V^T X U \Lambda U^T X^T \\ & \text{s.t.} & X X^T = I,\ X^T X = I. \end{array}$$

With

$$Y = V^T X U, \tag{2.11}$$

we get the equivalent problem

$$\text{QAP}_{\mathcal{OO}} \qquad \begin{array}{rll} \mu^{\mathcal{O}} := & \min & \operatorname{Trace} \Sigma Y \Lambda Y^T \\ & \text{s.t.} & Y Y^T = I,\ Y^T Y = I. \end{array} \tag{2.12}$$

The Lagrangian for this problem is

$$L(Y, S, T) = \operatorname{Trace} \Sigma Y \Lambda Y^T - \operatorname{Trace} S(Y Y^T - I) - \operatorname{Trace} (Y T Y^T - T).$$

Stationarity for the Lagrangian is

$$0 = \nabla L(Y, S, T) = \Sigma Y \Lambda - SYI - IYT. \tag{2.13}$$

Similar to the Proof I, the dual program is equivalent to the ordinary linear program LD, which is the dual of the LAP, LP above. Let Y be the optimal permutation of LP above and let S, T be the optimal solutions of LD above. Then the constraints of LD guarantee that the Hessian of the Lagrangian $L(Y, S, T)$ is positive semidefinite, i.e. the Lagrangian is convex in Y. In addition, complementary slackness between LD and LP can be written as:

$$y_{ij} \left(\Lambda_{ii} \Sigma_{jj} - S_{jj} - T_{ii} \right) = 0, \tag{2.14}$$

while stationarity can be rewritten as

$$0 = Y^T \Sigma Y \Lambda - Y^T S Y - T. \tag{2.15}$$

Since Y is a permutation, these are equivalent statements. Therefore, we have the three sufficient conditions for optimality:

$$\begin{array}{l} \text{primal feasibility (and so complementary slackness);} \\ \text{stationarity of the Lagrangian;} \\ \text{convexity of the Lagrangian.} \end{array} \tag{2.16}$$

Therefore Y is optimal for (2.12). After using the transformation (2.11), we get the optimal X for the original problem. ■

Remark 2.3 *We first observe that the optimal values* $\mu_H^{\mathcal{O}} = \nu_H^{\mathcal{O}} = \text{Trace}\, S = \text{Trace}\, T$, *i.e. the sum of traces of the Lagrange multipliers. This type of relationship appears to be common in these types of problems, e.g. see the eigenvalue problem above and also the trust region subproblem.*

The addition of the redundant constraints closes the duality gap because the resulting equivalent linear program LD has a basic feasible solution, which yields an optimal solution for the original problem. The addition of the redundant constraints $X^T X = I$ *results in the extra linear equality constraints needed in LD to ensure that the extreme points are permutation matrices.*

2.2 Lagrangian Dual; the General Case

We now look at the general case where $C \neq 0$. Note that, as we saw above, we can assume that A, B are both diagonal if desired, i.e. once we solve the problem in the diagonalized case then we can recover the original solution using (2.11). We now derive the Lagrangian dual for the general nonhomogeneous case.

We begin with the homogenized version of (2.9) above, i.e. we homogenize the linear part.

$$\begin{array}{rrll} & \mu^{\mathcal{O}} := & \min & \text{Trace}\, AXBX^T - 2x_0 CX^T \\ \text{QAP}_{\mathcal{OO}} & & \text{s.t.} & XX^T = I,\ X^T X = I \\ & & & x_0^2 = 1. \end{array} \tag{2.17}$$

This does not change the problem or its Lagrangian dual. The Lagrangian is

$$L(X,S,T,w) = \text{Trace}\, AXBX^T - 2x_0CX^T - S(XX^T - I) - T(X^TX - I) - w(x_0^2 - 1),$$

where we have introduced a Lagrange multiplier w for the constraint on x_0 and Lagrange multipliers S for $XX^T = I$ and T for $X^TX = I$. Note that the gradient of the Lagrangian set to zero is equivalent to

$$0 = AXB - Cx_0 - SX - XT, \quad w = \text{Trace}\, CX^T. \tag{2.18}$$

We get the Lagrangian dual lower bound $\nu^{\mathcal{O}}$.

$$\begin{array}{l} \mu^{\mathcal{O}} \geq \nu^{\mathcal{O}} := \\ \quad \max\limits_{S,T,w} \min\limits_{X,x_0} \quad \left\{ \text{Trace}\, \left[AXBX^T - SXX^T - TX^TX - wx_0^2 \right] \right. \\ \qquad\qquad\qquad - \text{Trace}\, x_0 2CX^T \\ \qquad\qquad\qquad \left. + \text{Trace}\, S + \text{Trace}\, T + w \right\}. \end{array} \tag{2.19}$$

With $x = \text{vec}\,(X),\ y = \begin{pmatrix} x_0 \\ x \end{pmatrix}$ we get

$$\begin{array}{l} \nu^{\mathcal{O}} = \\ \max\limits_{S,T,w} \min\limits_{y} \left\{ y^T \left[L_Q - \text{B}^{\text{O}}\text{Diag}\,(S) - \text{O}^{\text{O}}\text{Diag}\,(T) - wE_{00} \right] y \right. \\ \qquad\qquad \left. + \text{Trace}\, S + \text{Trace}\, T + w \right\}, \end{array} \tag{2.20}$$

where we define the $(n^2+1) \times (n^2+1)$ matrices

$$L_Q := \begin{bmatrix} 0 & -\text{vec}\,(C)^T \\ -\text{vec}\,(C) & B \otimes A \end{bmatrix}, \quad E_{00} := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.21}$$

and the linear operators

$$\text{B}^{\text{O}}\text{Diag}\,(S) := \begin{bmatrix} 0 & 0 \\ 0 & I \otimes S \end{bmatrix}, \tag{2.22}$$

$$\text{O}^{\text{O}}\text{Diag}\,(T) := \begin{bmatrix} 0 & 0 \\ 0 & T \otimes I \end{bmatrix}. \tag{2.23}$$

There is a hidden semidefinite constraint in (2.20), i.e. the inner minimization problem is bounded below only if the Hessian of the quadratic form is positive semidefinite. In this case the quadratic form has minimum value 0. This yields the following SDP.

$$\begin{array}{lll} (D_{\mathcal{O}}) & \max & \text{Trace}\, S + \text{Trace}\, T + w \\ & \text{s.t.} & L_Q - \text{B}^{\text{O}}\text{Diag}\,(S) - \text{O}^{\text{O}}\text{Diag}\,(T) - wE_{00} \succeq 0. \end{array}$$

Equivalently, we get

$$\begin{array}{lll} (D_{\mathcal{O}}) & \max & \text{Trace}\, S + \text{Trace}\, T + w \\ & \text{s.t.} & \text{B}^{\text{O}}\text{Diag}\,(S) + \text{O}^{\text{O}}\text{Diag}\,(T) + wE_{00} \preceq L_Q, \end{array}$$

where we recall that we have assumed, without loss of generality, that L_Q is an *arrow matrix*, i.e. $B \otimes A$ is diagonal. With $c = \mathrm{vec}\,(C)$, we can write the constraint in matrix form as

$$\begin{bmatrix} -w & -c^T \\ -c & B \otimes A - I \otimes S - T \otimes I \end{bmatrix} \succeq 0. \tag{2.24}$$

We now obtain our desired SDP relaxation of $(QQP_{\mathcal{OO}})$ as the Lagrangian dual of $(D_{\mathcal{O}})$. We introduce the $(n^2+1) \times (n^2+1)$ dual matrix variable $Y \succeq 0$ and derive the dual program to the SDP $(D_{\mathcal{O}})$.

$$(SDP_{\mathcal{O}}) \qquad \begin{array}{ll} \min & \mathrm{Trace}\, L_Q Y \\ \text{s.t.} & \mathrm{b^O diag}\,(Y) = I, \quad \mathrm{o^O diag}\,(Y) = I \\ & Y_{00} = 1, \\ & Y \succeq 0, \end{array} \tag{2.25}$$

The *block-0-diagonal operator* and *off-0-diagonal operator* acting on Y are defined by

$$\mathrm{b^O diag}\,(Y) := \sum_{k=1}^{n} Y_{(k,\cdot),(k,\cdot)} \tag{2.26}$$

and

$$\mathrm{o^O diag}\,(Y) := \sum_{k=1}^{n} Y_{(\cdot,k),(\cdot,k)}. \tag{2.27}$$

These are the adjoint operators of $\mathrm{B^O Diag}\,(\cdot)$ and $\mathrm{O^O Diag}\,(\cdot)$, respectively. The block-0-diagonal operator guarantees that the sum of the diagonal blocks equals the identity. The off-0-diagonal operator guarantees that the trace of each diagonal block is 1, while the trace of the off-diagonal blocks is 0. These constraints come from the orthogonality constraints, $XX^T = I$ and $X^T X = I$, respectively.

2.2.1 Schur Complement

The constraints in (2.24) can be rewritten using the Shur complement. First, we assume that the optimal w is known and fixed. From (2.18), we know that $w = \mathrm{Trace}\, CX^T$. That $-w > 0$ holds by the semidefiniteness constraint in (2.24) and since $c \neq 0$. Therefore,

$$(2.24) \text{ holds} \quad \Leftrightarrow \quad (B \otimes A - I \otimes S - T \otimes I) + \frac{1}{w} cc^T \succeq 0 \tag{2.28}$$

$$\Leftrightarrow \quad I \otimes S + T \otimes I \preceq B \otimes A + \frac{1}{w} cc^T \preceq B \otimes A, \tag{2.29}$$

since $w < 0$. We immediately have that the diagonal elements satisfy

$$S_{ii} + T_{jj} \leq A_{ii} B_{jj} - \frac{1}{|w|} |c_{(j-1)n+i}|^2, \quad \forall i, j. \tag{2.30}$$

Moreover, since the objective function of the dual involves only the traces of S, T, we can restrict ourselves to diagonal matrices S, T. What is remarkable about these equations is that they are *independent of the sign of the individual components of* C. This is a strong hint on how to obtain an example with a duality gap, see Example 3.2.

Lemma 2.4 *Suppose that $A, B \in \mathcal{S}_n$ are diagonal and $c \in \Re^{n^2}$. Then the optimal value of the dual program $D_{\mathcal{O}}$ is independent of the signs of the elements c_i of c.*

Proof. Let

$$\mathcal{F} := \left\{ S, T \in \mathcal{S}_n : (B \otimes A - I \otimes S - T \otimes I) + \frac{1}{w} cc^T \succeq 0 \right\}.$$

Suppose that $(S, T) \in \mathcal{F}$. Then (2.30) holds, i.e. the diagonal values are independent of the signs of the elements of c. Since the objective function depends only on the diagonals of S, T, we can assume that both these matrices are diagonal. ■

2.3 The Linear Case

We now assume that $A = B = 0$.

$$\begin{array}{lrll} \text{QQPL}_{\mathcal{O}} & \mu_L^{\mathcal{O}} := & \min & \text{Trace} - 2CX^T \\ & & \text{s.t.} & XX^T = I. \end{array} \tag{2.31}$$

Just as in the homogeneous (quadratic) case, we can characterize the optimal solution, except that the solution uses singular values rather than eigenvalues.

Proposition 2.5 *Suppose we have the singular value decomposition $C = U\Sigma V^T$, where the singular values, σ_i, are in the diagonal matrix Σ, and U, V are orthogonal matrices. Then the optimal value of* QQPL$_{\mathcal{O}}$ *is $\mu_L^{\mathcal{O}} = -2\text{Trace}\,\Sigma = -2\sum_{i=1}^{n} \sigma_i$. The optimal solution is obtained using the orthogonal matrices that yield the decomposition, i.e. $X^* = UV^T$.*

Proof. The decomposition implies

$$\text{Trace}\, CX^T = \text{Trace}\, \Sigma V^T X^T U.$$

Since X, U, V are orthogonal, the diagonal (in fact, all) elements of $V^T X^T U$ are ≤ 1. Therefore the minimum is attained with $V^T X^T U = I$. ■

From (2.28), the dual $D_{\mathcal{O}}$ in this purely linear case can be written as follows. Recall that we can assume S, T are diagonal.

$$\begin{array}{lrll} \text{LD} & \max & e^T s + e^T t + w & \\ & \text{s.t.} & s_j + t_i \leq -\dfrac{1}{|w|} |c_{(j-1)n+i}|^2, & i, j = 1, \ldots, n. \end{array}$$

Again we notice that the dual is independent of the sign of the individual elements of C. This results in a duality gap, unlike the homogeneous case, see Example 3.2.

However, we can solve this pure linear case efficiently using singular values, i.e. it is a tractable problem. In [14, 9], it was conjectured that quadratic problems that are tractable can be solved with Lagrangian relaxation if appropriate redundant constraints are chosen. We now see that this holds here if we add variables as well as constraints.

Theorem 2.6 *Strong duality holds between the pure linear case and its Lagrangian dual if the following equivalent problem for the sum of the singular values is used.*

$$\begin{array}{rcl} \sum_{i=1}^{n} \sigma_i(C) = \max & & 2\text{Trace}\, YCX^T \\ \text{(SVD)} & \text{s.t.} & WW^T = I,\ W^TW = I, \\ & & W = \begin{bmatrix} X & Y \\ V & Z \end{bmatrix}. \end{array} \tag{2.32}$$

Proof. The singular values of C are the largest n eigenvalues of the symmetric matrix $\begin{pmatrix} 0 & C^T \\ C & 0 \end{pmatrix}$. (This can be seen from the variational characterization of the singular values, e.g. [8].) Therefore, using our results in Proposition 2.1 above, we get

$$\begin{array}{rcl} \sum_{i=1}^{n} \sigma_i(C) = \max & \text{Trace} & \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X & Y \\ V & Z \end{bmatrix} \begin{bmatrix} 0 & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} X^T & V^T \\ Y^T & Z^T \end{bmatrix} \\ \text{s.t.} & & WW^T = I,\ W^TW = I, \end{array} \tag{2.33}$$

where

$$W = \begin{bmatrix} X & Y \\ V & Z \end{bmatrix}. \tag{2.34}$$

We know that there is no duality gap for this program and its Lagrangian dual. And, the objective function for this program can be simplified to yield the objective function of the theorem. ■

3 Examples with Duality Gaps

We now present two examples of problems with duality gaps. First we present the duality gap for the homogeneous case, $C = 0$, before adding the redundant orthogonal constraints. (See [16].)

Example 3.1 *Consider the the pure quadratic, orthogonally constrained problem*

$$\begin{array}{rcl} \mu^* := & \min & \text{Trace}\, AXBX^T \\ & s.t. & XX^T = I, \end{array} \tag{3.1}$$

with 2×2 *matrices*

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}.$$

The dual problem is

$$\begin{array}{rcl} \mu^D := & \max & -\text{Trace}\, S \\ & s.t. & (B \otimes A + I \otimes S) \succeq 0 \\ & & S = S^T. \end{array} \tag{3.2}$$

Then $\mu^* = 10$. But the dual optimal value $\mu^D < 10$, i.e. we have

$$B \otimes A = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix},$$

and for dual feasibility, we must have $S_{11} \geq -3$ and $S_{22} \geq -6$. To maximize the dual, equality must hold. Therefore $-\text{Trace}\, S = 9$.

The next example is for the pure linear case after adding the redundant orthogonal constraints.

Example 3.2 *This example uses $A = B = 0$.*

$$\begin{array}{rll} \mu^* := & \min & \text{Trace} - 2CX^T \\ & s.t. & XX^T = I,\ X^TX = I, \end{array} \tag{3.3}$$

with 2×2 matrix

$$C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

We then solve this example with the sign changed on -1, i.e.

$$C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The dual problem in the second case is

$$\begin{array}{rll} (D_{\mathcal{O}}) & \max & \text{Trace}\, S + \text{Trace}\, T + w \\ & s.t. & \text{B}^{\text{O}}\text{Diag}\,(S) + \text{O}^{\text{O}}\text{Diag}\,(T) + wE_{00} \preceq L_Q, \end{array}$$

where

$$L_Q := \begin{bmatrix} 0 & -1 & -1 & -1 & -1 \\ -1 & & & & \\ -1 & & \mathit{0} & & \\ -1 & & & & \\ -1 & & & & \end{bmatrix}. \tag{3.4}$$

We saw that the dual optimal value does not change, see Lemma 2.4. But here, the primal does, since the sum of the singular values of the two matrices are: $2\sqrt{2}$ in the first instance and just 2 for the symmetric C, i.e. the optimal values are $-4(\sqrt{2})$ and -4, respectively. Therefore, we have a duality gap in the second case.

4 Conclusion

In this note we have studied Lagrangian (and so SDP) duality gaps for problems with matrix orthogonality constraints $XX^T = I$. We saw that in the homogeneous case we can have a

duality gap, which is closed if we double the number of constraints by adding the redundant constraint $X^T X = I$ to the primal problem before taking the Lagrangian dual.

We then presented a counterexample to show that one can still have a nonzero duality gap for the general inhomogeneous problem. The duality gap can occur even for the pure linear problem, even though the pure linear problem can be solved efficiently using singular values. The duality gap can be seen to occur because the sign of an individual element of C does not change the dual problem, see Lemma 2.4.

We then saw that we can close the duality gap in the pure linear case if we replace the objective with the homogenized form YCX^T, and add both variables and constraints, see Theorem 2.6. Effectively, this doubles the number of variables but changes the linear case to a quadratic case so we can apply our previous results on the quadratic homogeneous case.

We still have the open question of whether we can find redundant constraints and/or relaxations to close the duality gap in the general case; or show that it is not possible.

References

[1] K.M. ANSTREICHER and N.W. BRIXIUS. A new bound for the quadratic assignment problem based on convex quadratic programming. Technical report, University of Iowa, Iowa City, IA, 1999.

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

[3] R. BHATIA. *Perturbation Bounds for Matrix Eigenvalues : Pitman Research Notes in Mathematics Series 162.* Longman, New York, 1987.

[4] 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.

[5] A. EDELMAN, T. ARIAS, and S.T. SMITH. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353 (electronic), 1999.

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

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

[8] R.A. HORN and C.R. JOHNSON. *Matrix analysis.* Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[9] Y.E. NESTEROV, H. WOLKOWICZ, and Y. YE. Semidefinite programming relaxations of nonconvex quadratic optimization. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *HANDBOOK OF SEMIDEFINITE PROGRAMMING: Theory, Al-*

gorithms, and Applications, pages 361–420. Kluwer Academic Publishers, Boston, MA, 2000.

[10] F. RENDL and H. WOLKOWICZ. A semidefinite framework for trust region subproblems with applications to large scale minimization. *Math. Programming*, 77(2, Ser. B):273–299, 1997.

[11] A. SHAPIRO and J.D. BOTHA. Dual algorithms for orthogonal Procrustes rotations. *SIAM J. Matrix Anal. Appl.*, 9(3):378–383, 1988.

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

[13] E. STIEFEL. Richtungsfelder und fernparallelismus in n-dimensionalem mannig faltigkeiten. *Commentarii Math. Helvetici*, 8:305–353, 1935-36.

[14] H. WOLKOWICZ. Semidefinite and Lagrangian relaxations for hard combinatorial problems. In M.J.D. Powell, editor, *Proceedings of 19th IFIP TC7 Conference on System Modelling and Optimization, July, 1999, Cambridge*, pages 269–309. Kluwer Academic Publishers, Boston, MA, 2000.

[15] H. WOLKOWICZ and Q. ZHAO. Semidefinite programming relaxations for the graph partitioning problem. *Discrete Appl. Math.*, 96/97:461–479, 1999. Selected for the special Editors' Choice, Edition 1999.

[16] 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 (Toronto, ON, 1996).