Semidefinite Programming

Henry Wolkowicz *

October 14, 2002

University of Waterloo
Department of Combinatorics & Optimization
Waterloo, Ontario N2L 3G1, Canada
Research Report CORR 2002-04

Key words: Semidefinite Programming, Cone Programming, Lagrangian Relaxation.

Contents

*Research supported by The Natural Sciences and Engineering Research Council of Canada. E-mail hwolkowicz@uwaterloo.ca

[0] Invited article for the Encyclopedia of Statistical Sciences, 2nd Edition. URL for paper: http://orion.math.uwaterloo.ca/~hwolkowi/henry/reports/ABSTRACTS.html

1 Introduction

Semidefinite Programming, denoted SDP, has been studied (under various names) as far back as the 1940s. The interest has grown tremendously since the early 1990s and it is currently considered to be the hottest area in Optimization. The research activity was motivated by the discovery of new applications in several areas, combined with the development of efficient new algorithms. This article serves as an introduction to the basics of SDP.

2 What is SDP?

Primal and dual SDPs look like Linear Programs, LPs, i.e. the primal SDP is

$$\text{(PSDP)} \qquad \begin{array}{rcl} p^* := & \min & C \bullet X := \operatorname{trace} CX \\ & \text{subject to} & \mathcal{A}\, X = b \\ & & X \succeq 0 \end{array}$$

and its dual is

$$\text{(DSDP)} \qquad \begin{array}{rcl} d^* := & \max & b^T y \\ & \text{subject to} & \mathcal{A}^* y + Z = C \\ & & Z \succeq 0. \end{array}$$

Here: $\mathcal{S}^n$ is the vector space of $n \times n$ real symmetric matrices and $X, Z, C \in \mathcal{S}^n$ equipped with the trace inner product; the nonnegativity symbol $\succ 0$ (resp. $\succeq 0$) denotes positive (semi) definiteness, often referred to as the *Löwner partial order*. $\mathcal{A} : \mathcal{S}^n \to \Re^m$ is a linear operator and $\mathcal{A}^*$ is the adjoint operator (reduces to transpose for LP), i.e. the adjoint satisfies

$$\mathcal{A}(X) \bullet y = X \bullet \mathcal{A}^*(y), \quad \forall X \in \mathcal{S}^n, \forall y \in \Re^m.$$

The action of the linear operator can be expressed as the vector

$$\mathcal{A}\, X = (\operatorname{trace} A_i X) \in \Re^m,$$

where $A_i \in \mathcal{S}^n$, $i = 1, \ldots, m$. The adjoint operator is then the matrix

$$\mathcal{A}^* y = \sum_{i=1}^{m} y_i A_i \in \mathcal{S}^n .$$

For LP, $x \geq 0$ denotes nonnegativity elementwise, i.e. $x \in \Re^n$ the nonnegative orthant, a polyhedral cone. For SDP, $X \succeq 0$ denotes positive semidefiniteness, i.e. X is in the nonlinear cone (closed convex) of positive semidefinite matrices.

2.1 Duality

We now use the concept of a *hidden constraint* and derive the dual and the principle of weak duality. (The notion of hidden constraint appears again in the derivation of SDP relaxations below.)

Let $\mathcal{P}$ denote the cone of positive semidefinite matrices in $\mathcal{S}^n$. The *polar cone* of $\mathcal{P}$ is

$$\mathcal{P}^+ = \{S \in \mathcal{S}^n : S \bullet T \geq 0, \ \forall T \in \mathcal{P}\}.$$

Lemma 2.1

$$\mathcal{P}^+ = \mathcal{P},$$

i.e. $\mathcal{P}$ *is self-polar.*

Proof. Note that trace is commutative, i.e. $\operatorname{trace} MN = \operatorname{trace} NM$. Suppose that $S, T \in \mathcal{S}^n$, $T \succeq 0$ and $T^{\frac{1}{2}} \succeq 0$ is its square root. Then

$$\operatorname{trace} ST \geq 0, \ \forall T \in \mathcal{P} \quad \Leftrightarrow \quad \operatorname{trace} T^{\frac{1}{2}} S T^{\frac{1}{2}} \geq 0, \ \forall T \in \mathcal{P}.$$

The result now follows from Sylvester's Theorem of Inertia. ■

Theorem 2.2 *(Weak Duality) The primal-dual pair of SDPs satisfy*

$$p^* \geq d^*.$$

Proof. We show the following

$$\begin{aligned} p^* &= \min_{X \succeq 0} \max_{y} C \bullet X + y^T(b - \mathcal{A}X) \\ &\geq \max_{y} \min_{X \succeq 0} y^T b + (C - \mathcal{A}^* y) \bullet X \\ &= \max_{C - \mathcal{A}^* y \succeq 0} y^T b = d^* \quad (DSDP). \end{aligned} \tag{2.1}$$

In the first equality, the inner maximization is unconstrained in y. Therefore, there is a hidden constraint for the minimization problem that $b - \mathcal{A}X = 0$. Once we add this constraint under the minimization, the maximization problem disappears and we are left with PSDP, the primal problem.

The inequality follows by interchanging the minimization and maximization and using the adjoint equation. This second line now has a hidden constraint for the inner minimization, i.e. $C - \mathcal{A}^* y \succeq 0$. For if $T := C - \mathcal{A}^* y$ is not positive semidefinite, let $T = QDQ^T$ be its orthogonal diagonalization. Let $X = QD^a Q^T$, where the diagonal matrix

$$D^a_{ij} = \begin{cases} D_{ij} & \text{if } D_{ij} < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then $\lim_{\alpha \to \infty} \alpha X T = -\infty$. Once we add this constraint under the maximization, the equivalence with DSDP results. ■

If X, y, Z are primal-dual feasible, then the relations in (2.1) show that $p^* = d^*$ if and only if *complementary slackness* $Z \cdot X = (C - \mathcal{A}^* y) \cdot X = 0$. However, using the same argument given in Lemma 2.1, we get

$$Z \cdot X = 0 \quad \Leftrightarrow \quad ZX = 0.$$

This and the duality theory yields the elegant characterization of optimality which drives interior- point methods.

Theorem 2.3 *The variables X, y, Z are a primal-dual optimal pair for the SDPs if and only if the following hold.*

$$\begin{array}{ccll} \mathcal{A}^* y + Z - C & = & 0 & \textit{dual feasibility} \\ b - \mathcal{A}(X) & = & 0 & \textit{primal feasibility} \\ ZX & = & 0 & \textit{complementary slackness} \\ Z, X \succeq 0. & & & \end{array}$$

■

Primal-dual interior-point (p-d i-p) methods perturb the complementarity equation to

$$ZX = \mu I, \quad \mu > 0.$$

They then apply Newton's method and stay *interior*, $X, Z \in \operatorname{int} \mathcal{P}$ equivalently $X, Z \succ 0$, while reducing $\mu \downarrow 0$.

2.1.1 Comparisons with LP

The above arguments are similar to those for LP, where Z, X usually represent diagonal matrices formed from the nonnegative elementwise vectors $x, z \geq 0$. The product of diagonal matrices ZX is a diagonal matrix. However, for $Z, X \in \mathcal{S}^n$ the matrix ZX is not necessarily symmetric! Therefore, the system of equations in Theorem 2.3 is *overdetermined* and Newton's method cannot be directly applied.

There are other interesting differences between LP and SDP. These are both cone programs, i.e. the minimization of a linear function subject to linear and cone constraints. In the LP case, we use the nonnegative orthant. The partial order $x \geq y$ means that $x - y \geq 0$ elementwise, or $x - y$ is in the nonnegative orthant. For SDP, the nonnegative orthant is replaced by the nonpolyhedral cone $\mathcal{P}$. The geometry of this cone is well understood, see e.g. [21]. For 2×2 matrices we can visualize $\mathcal{P}$ as an *ice-cream cone* in $\Re^3$. However, the nonpolyhedral nature of the cone introduces several nonlinear complications which differ from LP.

1. Just as in LP, a *zero duality gap* holds $p^* = d^*$ if and only if complementary slackness $Z \cdot X = 0$ if and only if $ZX = 0$. However, in LP the zero duality gap always holds (unless both problems are infeasible). Nonzero duality gaps can occur for SDP. Constraint qualifications (CQ) are needed to guarantee a zero duality gap and also attainment. The standard CQ is *Slater's Condition*: strict feasibility. However, a regularization process is possible which closes the duality gaps, [4, 25].

2. *Strict complementarity* for SDP is equivalent to $Z + X \succ 0$. However, the Theorem of Goldman and Tucker [7], that guarantees the existence of a strictly complementary pair, can fail for SDP even if the Slater constraint qualification holds for both primal-dual SDPs. The strict complementarity conditions (and nondegeneracy) hold generically, see e.g. [28, 22, 1].

3. The existence of *polynomial time algorithms* for LP was shown in [11]. Polynomial time algorithms for more general convex programs, including SDP, was shown in [17, 18]. The development of p-d i-p methods for SDP followed those for LP, i.e. as mentioned above a Newton type method is applied to the perturbed optimality conditions. The solution to the perturbed conditions for each $\mu > 0$ is called the *central path*. For LP the central path converges to the analytic center of the set of optimal solutions. However, this does not hold true for SDP if strict complementarity fails, see e.g. [29, 9].

3 Why Use SDP?

For many computationally hard problems, quadratic programs provide stronger models than linear programs. These quadratic programs (quadratic objective and quadratic constraint) are, in general, intractable. However, the Lagrangian relaxations can be solved efficiently using SDP.

We now look at two applications. We start with perhaps the simplest and most successful SDP relaxation, the Max-Cut problem, (MC). We then look at the Quadratic Assignment Problem, (QAP). Both of these are hard combinatorial problems. There are many other such applications, e.g. Max-Clique, Graph-Partitioning, Graph-Colouring, Max-Satisfiability, closest correlation matrix, Ricatti equations, min-max eigenvalue problems, matrix norm minimization, eigenvalue localization, etc..., see [32].

3.1 Tractable Relaxations of Max-Cut

The Max-Cut problem consists in finding a partition of the set of vertices of a given undirected graph with weights on the edges so that the sum of the weights of the edges cut by the partition is maximized. This NP-hard discrete optimization problem can be formulated as the following (quadratic) program (e.g. Q is a multiple of the Laplacian matrix of the graph).

$$\text{(MC0)} \qquad \begin{array}{rll} \text{mc}^* := & \max & v^T Q v \\ & \text{s.t.} & v_i^2 = 1, \quad i = 1, \ldots, n. \end{array}$$

More generally, it is a special case of quadratic boolean programming.

$$(MCQ) \qquad \mu^* := \max_{x \in \mathcal{F}} q_0(x) \quad (:= x^T Q x - 2c^T x).$$

where $\mathcal{F} = \{\pm 1\}^n$. Perturbing the diagonal of Q on $\mathcal{F}$ yields an equivalent problem:

$$\begin{array}{rcl} q_u(x) & := & x^T (Q + \mathrm{Diag}\,(u)) x - 2c^T x - u^T e \\ & = & q_0(x), \quad \forall x \in \mathcal{F}, \end{array}$$

where e is the vector of ones, $\mathrm{Diag}\,(u)$ denotes the diagonal matrix formed from u.

MCQ is an NP-hard problem. Therefore, one usually attempts to solve relaxations. Exact solutions often use branch and bound methods. We now look at several different bounds obtained by relaxing the feasible set $\mathcal{F}$.

3.1.1 Simple Relaxation

A trivial bound formed from the diagonal perturbations is

$$\mu^* \leq f_0(u) := \max_x q_u(x).$$

Define the set of perturbations

$$S := \left\{u : u^T e = 0, Q + \mathrm{Diag}\,(u) \preceq 0\right\}.$$

Then we get

$$\boxed{\begin{array}{c}\mu^* \leq B_0 := \min_u f_0(u) \\ \left(= \min_{u^T e=0} f_0(u), \text{ if } S \neq \emptyset\right).\end{array}}$$

We can use the hidden semidefinite constraint that: $q_u(x)$ bounded above implies that the Hessian $\nabla^2 q_u \preceq 0$. This yields our first bound B_0:

$$\mu^* \leq B_0 = \min_{Q+\mathrm{Diag}\,(u)\preceq 0} f_0(u).$$

3.1.2 Trust Region Relaxation

We now relax the feasible set to the sphere of radius $\sqrt{n}$. This uses the tractable trust region subproblem, TRS,

$$\mu^* \leq f_1(u) := \max_{||x||^2=n} q_u(x),$$

i.e. TRS is a hidden convex problem in that strong duality holds and the Lagrangian dual is the maximization of a concave function over an interval, [30], or an unconstrained concave maximization, [27]. This yields our next bound

$$\boxed{\mu^* \leq B_1 := \min_u f_1(u).}$$

3.1.3 Box Constraint Relaxation

Another relaxation uses the box constraint

$$\mu^* \leq f_2(u) := \max_{|x_i|\leq 1} q_u(x).$$

This relaxation can still be NP-hard unless $q_u(x)$ is concave, [20]. Therefore, we add the hidden semidefinite constraint to make the bound tractable.

$$\mu^* \leq \min_u f_2(u)$$

and

$$\boxed{\mu^* \leq B_2 := \min_{Q+\mathrm{Diag}\,(u) \preceq 0} f_2(u).}$$

3.1.4 Eigenvalue Bound

We can lift the problem to a higher dimension and homogenize. We use the matrix

$$Q^c := \begin{bmatrix} 0 & -c^T \\ -c & Q \end{bmatrix}$$

and the homogenized function

$$q_u^c(y) := y^T(Q^c + \mathrm{diag}\,(u))y - u^T e.$$

Then

$$\mu^* \leq f_1^c(u) := \max_{||y||^2=n+1} q_u^c(y)$$

where the maximum is an eigenvalue problem

$$\max_{||y||^2=n+1} q_u^c(y) = (n+1)\lambda_{\max}(Q^c + \mathrm{diag}\,(u)) - u^T e.$$

Therefore, our min-max eigenvalue bound is

$$\boxed{\mu^* \leq B_1^c := \min_u f_1^c(u).}$$

Similarly, we can get equivalent other homogenized bounds from the previous listed bounds.

3.1.5 SDP Bound

After homogenization if needed (i.e. assume $c = 0$), we use

$$x^T Q x = \mathrm{trace}\, x^T Q x = \mathrm{trace}\, Q x x^T$$

and, for $x \in \mathcal{F}$, $y_{ij} = x_i x_j$ defines a symmetric, rank one, positive semidefinite matrix Y with diagonal elements 1. We relax the (hard) rank one condition to get

$$\boxed{\begin{array}{lcc} B_3 := & \max & \mathrm{trace}\, QY \\ & \text{subject to} & \mathrm{diag}\,(Y) = e \\ & & Y \succeq 0. \end{array}}$$

3.1.6 Summary and Lagrangian Relaxation

Suppose that we restrict the perturbations with $u^Te = 0$. Then the bounds are

$$
\boxed{
\begin{array}{rcl}
B_0 & = & \min_u \max_x q_u(x) \\
B_1 & = & \min_u \max_{x^Tx=n} q_u(x) \\
B_2 & = & \min_u \max_{-1\leq x_i \leq 1} q_u(x) \\
B_3 & = & \max\{\operatorname{trace} Q^cY : \operatorname{diag}(Y) = e,\ Y \succeq 0.\} \\
B_1^c & = & \min_u \max_{y^Ty=n+1} q_u^c(y)
\end{array}
}
$$

We now do something that seems of no value; we replace the ± 1 constraints with $x_i^2 = 1, \forall i$. This yields the following quadratic, quadratically constrained, equivalent program to MCQ.

$$
\begin{array}{lcl}
(P_E) & \max & q_0(x) = x^TQx - 2c^Tx \\
 & \text{subject to} & x_i^2 = 1, \quad i = 1, \cdots, n.
\end{array}
$$

Then the Lagrangian relaxation bound is

$$
B_L = \min_\lambda \max_x q_0(x) + \sum_{i=1}^n \lambda_i (1 - x_i^2).
$$

The following theorem is proved in [24, 23].

Theorem B_L equals all the above bounds.

Thus, we see that the Lagrangian relaxation is as strong as all the other relaxations. Moreover, it can be calculated efficiently using the SDP relaxation.

3.2 Recipe for SDP relaxations

The homogenization and relaxation techniques yield the following recipe, [23].

1. add redundant constraints
2. take Lagrangian dual
3. homogenize
4. use hidden semidefinite constraint to obtain equivalent SDP *(check Slater's constraint qualification - strict feasibility)*
5. take Lagrangian dual again
6. check Slater's CQ again - project if it fails
7. delete redundant constraints

3.3 SDP Relaxation for the Quadratic Assignment Problem, QAP

We now apply the recipe in Section 3.2 to QAP. (See [34] for more details.) The QAP is one of the most difficult combinatorial problems, e.g. $n = 30$ instances have only recently been solved using new bounding techniques and high performance parallel computing.

The QAP in the trace formulation is

$$(QAP) \quad \mu^* := \min_{X \in \Pi} \ \operatorname{trace} AXBX^T - 2CX^T,$$

where A, B are real symmetric $n \times n$ matrices, C is a real $n \times n$ matrix, and Π is the set of permutation matrices. (We assume $n \geq 4$ to avoid trivialities.) One of the many applications of QAP is the modelling of the allocation of a set of n facilities to a set of n locations while minimizing the quadratic objective arising from the distance between the locations in combination with the flow between the facilities. See e.g. [19, 5].

There are several interesting numerical and theoretical difficulties that arise, e.g. what to do with the loss of a constraint qualification and loss of sparsity in the optimality conditions. Can the new bound compete with other bounding techniques in speed and quality? Can we incorporate the new bound in a branch and bound algorithm? We now apply our recipe to QAP.

We need the following notation. $\mathcal{E} := \{X : Xe = X^T e = e\}$ is the set of matrices satisfying the *assignment constraints*, i.e. row and column sums are all equal to one. $\mathcal{Z} := \{X : X_{ij} \in \{0,1\}\}$ is the set of $(0,1)$-matrices. $\mathcal{N} := \{X : X_{ij} \geq 0\}$ is the set of *nonnegative matrices*. $\mathcal{O} := \{X : XX^T = X^T X = I\}$ is the set of orthogonal matrices.

Permutation matrices are $0, 1$ matrices with exactly one element equal to 1 in each column and each row. It is well known that

$$\Pi = \mathcal{E} \cap \mathcal{Z} = \mathcal{O} \cap \mathcal{Z}.$$

This gives us a group of redundant constraints to add to get the following equivalent program to QAP.

$$\begin{array}{ll} \min & \operatorname{trace} AXBX^T - 2CX^T \\ \text{s.t.} & XX^T = X^T X = I \quad \text{(orthog.; redundancy important)} \\ & X_{ij}^2 - X_{ij} = 0, \quad \forall i, j. \quad (0,1 \text{ constraints}) \\ & \left.\begin{array}{l} \|Xe - e\|^2 = 0 \\ \|X^T e - e\|^2 = 0 \end{array}\right\} \quad \text{(row \& col. sums 1)} \\ & \left.\begin{array}{ll} \operatorname{diag}(X_{:i} X_{:j}^T) = 0, & \text{if } i \neq j \\ X_{:i} X_{:j}^T - \operatorname{Diag}\left(\operatorname{diag}(X_{:i} X_{:j}^T)\right) = 0, & \text{if } i = j \end{array}\right\} \quad \text{(gangster)} \end{array}$$

We use both $XX^T = I$ and $X^T X = I$. These constraints are equivalent but are not redundant in the relaxation. They provide a significant strengthening, see [2]. We change the linear constraints in the set $\mathcal{E}$ into quadratic constraints, as linear constraints are ignored in the Lagrangian relaxation. The so-called *gangster constraints* come from the property that the columns of a permutation matrix are elementwise orthogonal while the elementwise product of a column with itself is equal to itself. The term gangster comes from the fact that the operator *shoots holes* in a matrix, as we see below.

We can now take the Lagrangian relaxation

$$
\begin{array}{rl}
\mu_{\mathcal{O}} \geq \mu_{\mathcal{L}} & \\
:= \max\limits_{W,u_0,v_0\cdots} \min\limits_{XX^T=X^TX=I} & \{\operatorname{trace} AXBX^T - 2CX^T \\
& + \sum_{ij} W_{ij}(X_{ij}^2 - X_{ij}) \\
& + u_0\|Xe - e\|^2 \\
& + v_0\|X^Te - e\|^2 \\
& + \ldots\}.
\end{array}
$$

and homogenize the Lagrangian using a scalar x_0 and constraint $x_0^2 = 1$. We get the lower bound (separating quadratic, linear, and constant terms in X)

$$
\begin{array}{rl}
\max\limits_{W,S_b,S_o,u_0v_0,,w_0} \min\limits_{X,\ x_0} & \{\operatorname{trace}[AXBX^T \\
& + u_0\|Xe\|^2 + v_0\|X^Te\|^2 \\
& + W(X \circ X)^T + w_0x_0^2 \\
& + S_bXX^T + S_oX^TX] \\
& - \operatorname{trace} x_0(2C + W)X^T \\
& - 2x_0u_0e^T(X + X^T)e \\
& + \ldots \\
& - w_0 - \operatorname{trace} S_b - \operatorname{trace} S_o + 2nu_0x_0^2\}.
\end{array}
$$

Applying the hidden semidefinite constraint that the Hessian of a quadratic bounded below is positive semidefinite, leads to the SDP:

$$
\begin{array}{ll}
\max & -w_0 - \operatorname{trace} S_b - \operatorname{trace} S_o + \ldots \\
\text{s.t.} & L_Q + \operatorname{Arrow}(w) + \mathrm{B^0Diag}(S_b) \\
& \quad + \mathrm{O^0Diag}(S_o) + u_0D + \ldots \succeq 0.
\end{array}
\tag{3.1}
$$

where the matrix L_Q is formed using the Kronecker product

$$
L_Q := \begin{bmatrix} 0 & -\operatorname{vec}(C)^T \\ -\operatorname{vec}(C) & B \otimes A \end{bmatrix},
$$

the matrix

$$
D := \begin{bmatrix} n & -e^T \otimes e^T \\ -e \otimes e & I \otimes E \end{bmatrix} + \begin{bmatrix} n & -e^T \otimes e^T \\ -e \otimes e & E \otimes I \end{bmatrix}
$$

and the linear operators

$$
\operatorname{Arrow}(w) := \begin{bmatrix} w_0 & -\frac{1}{2}w_{1:n^2}^T \\ -\frac{1}{2}w_{1:n^2} & \operatorname{Diag}(w_{1:n^2}) \end{bmatrix},
\tag{3.2}
$$

$$
\mathrm{B^0Diag}(S) := \begin{bmatrix} 0 & 0 \\ 0 & I \otimes S_b \end{bmatrix}, \quad \text{block diagonal}
\tag{3.3}
$$

$$
\mathrm{O^0Diag}(S) := \begin{bmatrix} 0 & 0 \\ 0 & S_o \otimes I \end{bmatrix}, \quad \text{block off-diagonal.}
\tag{3.4}
$$

Slater's constraint qualification (strict feasibility) holds for (3.1). Therefore, we can take the Lagrangian dual again which yields an SDP relaxation

$$\begin{array}{ll} \min & \operatorname{trace} L_Q Y \\ \text{s.t.} & \mathrm{b}^0\mathrm{diag}\,(Y) = I, \quad \mathrm{o}^0\mathrm{diag}\,(Y) = I \\ & \operatorname{arrow}(Y) = e_0, \quad \operatorname{trace} DY = 0 \\ & Y \succeq 0, \end{array}$$

where the *arrow operator*, acting on the $(n^2+1)\times(n^2+1)$ matrix Y, is the adjoint operator to $\operatorname{Arrow}(\cdot)$ and is defined by

$$\operatorname{arrow}(Y) := \operatorname{diag}(Y) - \left(0, (Y_{0,1:n^2})^T\right), \tag{3.5}$$

i.e. the arrow constraint guarantees that the diagonal and 0-th row (or column) are identical. The *block-0-diagonal operator* and *off-0-diagonal operator* acting on Y are defined by

$$\mathrm{b}^0\mathrm{diag}\,(Y) := \sum_{k=1}^{n} Y_{(k,\cdot),(k,\cdot)} \tag{3.6}$$

and

$$\mathrm{o}^0\mathrm{diag}\,(Y) := \sum_{k=1}^{n} Y_{(\cdot,k),(\cdot,k)}. \tag{3.7}$$

These are the adjoint operators of $\mathrm{B}^0\mathrm{Diag}\,(\cdot)$ and $\mathrm{O}^0\mathrm{Diag}\,(\cdot)$, 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 $X^TX = I$, respectively.

We now check Slater's CQ again. But $0 \neq D \succeq 0$, $\operatorname{trace} YD = 0$, implies that Y is singular. However, we can project onto the minimal face of the semidefinite cone that contains the feasible set. Define the following $(n^2+1)\times((n-1)^2+1)$ matrix

$$\hat{V} := \left[\begin{array}{c|c} 1 & 0 \\ \hline \frac{1}{n}(e\otimes e) & V\otimes V \end{array}\right], \tag{3.8}$$

where V is an $n\times(n-1)$ matrix containing a basis of the orthogonal complement of e, i.e. $V^Te = 0$, e.g.

$$V := \left[\begin{array}{c} I_{n-1} \\ \hline -e_{n-1}^T \end{array}\right].$$

After removing redundant constraints, we get the following simplified projected relaxation with $n^3 - 2n^2 + 1$ constraints.

$$\begin{array}{rll} \mu_{R2} := & \min & \operatorname{trace}(\hat{V}^T L_Q \hat{V})R \\ & \text{s.t.} & \mathcal{G}_{\bar{J}}(\hat{V}R\hat{V}^T) = E_{00} \\ & & R \succeq 0. \end{array}$$

	Sol.	GLB	ELI	EVB3	*SDP*	rel. error SDP
Nug20	2570	2057	2196	2290	2386	0.0771
Nug21	2438	1833	1979	2116	2253	0.0821
Nug22	3596	2483	2966	3174	3396	0.0589
Nug24	3488	2676	2960	3074	3235	0.0782
Nug25	3744	2869	3190	3287	3454	0.0840
Nug30	6124	4539	5266	5448	5695	0.0753

QAPLIB (Nugent) instances
NEOS Server time for n=30; 1400 hours on SUN E6500

The above description uses the so-called *gangster operator*. Let $J \subset \{(i,j) : 1 \leq i,j \leq n^2+1\}$. The operator $\mathcal{G}_J : \mathcal{S}_{n^2+1} \to \mathcal{S}_{n^2+1}$ is called the *Gangster* operator. For matrix Y, and $i,j = 1,\ldots,n^2+1$, the ij component of the image of the gangster operator is defined as

$$(\mathcal{G}_J(Y))_{ij} := \begin{cases} Y_{ij} & \text{if } (i,j) \in J \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

(The indices for $\bar{J}$ are given in [34].)

The dual problem is (the gangster operator is self-adjoint)

$$\begin{array}{rll} \mu_{R2} = & \max & -Y_{00} \\ & \text{s.t.} & \hat{V}^T(L_Q + \mathcal{G}^*_{\bar{J}}(Y))\hat{V} \succeq 0. \end{array}$$

Table 3.3 illustrates the strength of the SDP relaxation on several Nugent problems from QAPLIB compared to other bounds in the literature: Gilmore-Lawler bound (GLB) [6, 14], the projection or elimination bound ELI of [8], and the improved eigenvalue bound EVB3 from [26]. We note the high cost of the SDP bound for n=30 in the table and the low relative error for the bounds. A relaxed form of the SDP bound played a major role in the solution to optimality of several hard QAPs, see [3].

4 How to Solve SDP?

The similarity of SDP with Linear Programming, LP, motivated researchers to apply techniques that proved successful for LP, in particular primal-dual interior-point (p-d i-p) methods, see e.g. [32]. Newton type methods are applied on a perturbation of the characterization of optimality for the primal-dual pair, i.e. suppose that

$$X_c \succ 0, Z_c \succ 0$$

are the current strictly positive estimates. And $\mu > 0$ is the barrier parameter. Then we would like to solve the following system of nonlinear equations

$$\begin{array}{rcll} R_D & := & \mathcal{A}^*y - Z - C = 0 & \text{(dual feasibility)} \\ R_P & := & AX - b = 0 & \text{(primal feasibility)} \\ R_C & := & ZX - \mu I = 0 & \text{(perturbed complementary slackness)} \end{array}$$

Linearization leads to the following system for the search direction $\Delta s = \begin{pmatrix} \Delta X \\ \Delta y \\ \Delta Z \end{pmatrix}$

$$\begin{array}{rcl} \mathcal{A}^* \Delta y - \Delta Z &=& -R_D \\ A \Delta X &=& -R_P \\ Z_c \Delta X + \Delta Z X_c &=& -R_C. \end{array} \tag{4.1}$$

However, the product ZX is not necessarily symmetric, though Z, X are. Therefore, the above is an *overdetermined* linear system. This has led to symmetrization schemes that apply Newton's method, see e.g. [15]. Alternatively, a Gauss-Newton approach is used in [13]. See also the recent books [31, 33].

The HKM search direction [10, 12, 16] is, arguably, the most popular and efficient among the primal-dual interior-point (p-d i-p) directions for SDP. It is based on applying Newton's method to a symmetrized form of the optimality conditions for PSDP. Therefore, in theory, we get fast asymptotic and polynomial time algorithms. We now derive the HKM search direction $\Delta s = \begin{pmatrix} \Delta X \\ \Delta y \\ \Delta Z \end{pmatrix}$ from the above linearization (4.1) of the perturbed optimality conditions. We get

$$\Delta Z = \mathcal{A}^*(\Delta y) + R_D \tag{4.2}$$

and

$$\Delta X = -Z^{-1}(\Delta Z)X - Z^{-1}R_C = -Z^{-1}(\mathcal{A}^*(\Delta y) + R_D)X + \mu Z^{-1} - X. \tag{4.3}$$

We substitute this into the second equation and solve for Δy using

$$\mathcal{A}\left(Z^{-1}\mathcal{A}^*(\Delta y)X\right) = \mathcal{A}\left(\mu Z^{-1} - X - Z^{-1}R_D X\right) + R_P = \mathcal{A}\left(\mu Z^{-1} - Z^{-1}R_D X\right) - b. \tag{4.4}$$

We can now backsubstitute to get the symmetric matrix ΔZ using (4.2). However, ΔX in (4.3) need not be symmetric. Therefore we *cheat* and symmetrize ΔX after backsubstition in (4.3), i.e. we solve for the system by assuming ΔX is a general matrix and then symmetrize by projecting the solution back into $\mathcal{S}^n$.

The p-d i-p algorithms have the following simple framework. The reduction in μ is adaptive. In addition, an adaptive *centering parameter* is used.

Given $(X^0, y^0, Z^0) \in \mathcal{F}^0$ (strictly feasible)
for $k = 0, 1, 2 \ldots$
solve the linearization for the search direction

$$F'_\mu(X^k, y^k, Z^k) \begin{pmatrix} \Delta X^k \\ \Delta y^k \\ \Delta Z^k \end{pmatrix} = \begin{pmatrix} -R_D \\ -R_P \\ -X^k Z^k + \sigma_k \mu_k I \end{pmatrix}$$

where σ_k centering, $\mu_k = \operatorname{trace} X^k Z^k / n$

$$(X^{k+1}, y^{k+1}, Z^{k+1}) = \\ (X^k, y^k, Z^k) + \alpha_k (\Delta X^k, \Delta y^k, \Delta Z^k)$$

so that $(X^{k+1}, Z^{k+1}) \succ 0$
end (for).

The above need to symmetrize illustrates one of the subtle differences between SDP and LP. Other differences include: possible duality gaps for SDP in the absence of strictly feasible solutions (Slater's constraint qualification, CQ); strict complementarity can fail at the optimum. Two illustrations follow.

Example 4.1 (Duality gap)

$$(\mathbf{P}) \quad \begin{array}{rll} p^* = & \max & x_2 \\ & s.\ t. & \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$(\mathbf{D}) \quad \begin{array}{rlc} d^* = & \min & \operatorname{trace} U_{11} \\ & s.\ t. & U_{22} = 0 \\ & & U_{11} + 2U_{23} = 1 \\ & & U \succeq 0. \end{array}$$

Then $p^* = 0 < d^* = 1$.

Example 4.2 (Strict complementarity)

$$(\mathbf{P}) \quad \begin{array}{rll} p^* = & \max & x_1 \\ & s.\ t. & \begin{bmatrix} x_1 & x_3 & x_2 \\ x_3 & x_2 & 0 \\ x_2 & 0 & x_3 \end{bmatrix} \preceq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$(\mathbf{D}) \quad \begin{array}{rlc} d^* = & \min & \operatorname{trace} U_{33} \\ & s.\ t. & U_{11} = 1 \\ & & U_{22} + 2U_{13} = 0 \\ & & U_{33} + 2U_{12} = 0 \\ & & U \succeq 0. \end{array}$$

Then the (unique) optimum pair is: $U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $x = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ *with slack* $Z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

5 Conclusion

We have presented an introduction to SDP. We started with the basic properties and optimality conditions and emphasized the similarities/differences with LP.

We then motivated the many applications by illustrating its use on quadratic models. These quadratic models are generally stronger relaxations of NP-hard problems than linear models, though they are often themselves NP-hard problems. However, we can solve the Lagrangian relaxation of these quadratic models efficiently using SDP.

We included a discussion on numerical approaches and software for solving SDP.

References

[1] F. ALIZADEH, J-P.A. HAEBERLY, and M.L. OVERTON. A new primal-dual interior-point method for semidefinite programming. In J.G. Lewis, editor, *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra*, pages 113–117. SIAM, 1994.

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

[4] J.M. BORWEIN and H. WOLKOWICZ. Characterization of optimality for the abstract convex program with finite-dimensional range. *J. Austral. Math. Soc. Ser. A*, 30(4):390–411, 1980/81.

[5] F. CELA. *The Quadratic Assignment Problem: Theory and Algorithms*. Kluwer, Massachessets, USA, 1998.

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

[7] A. J. GOLDMAN and A. W. TUCKER. Theory of linear programming. In *Linear inequalities and related systems*, pages 53–97. Princeton University Press, Princeton, N.J., 1956. Annals of Mathematics Studies, no. 38.

[8] 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.

[9] M. HALICKA, E. de KLERK, and C. ROOS. On the convergence of the central path in semidefinite optimization. Technical Report June, Faculty ITS, Delft University of Technology, Delft, The Netherlands, 2001.

[10] C. HELMBERG, F. RENDL, R. J. VANDERBEI, and H. WOLKOWICZ. An interior-point method for semidefinite programming. *SIAM J. Optim.*, 6(2):342–361, 1996.

[11] L.G. KHACHIAN. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, 244:1093–1096, 1979.

[12] M. KOJIMA, S. SHINDOH, and S. HARA. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7(1):86–125, 1997.

[13] S. KRUK, M. MURAMATSU, F. RENDL, R.J. VANDERBEI, and H. WOLKOWICZ. The Gauss-Newton direction in linear and semidefinite programming. *Optimization Methods and Software*, 15(1):1–27, 2001.

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

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

[16] R.D.C. MONTEIRO. Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optim.*, 7(3):663–678, 1997.

[17] Y.E. NESTEROV and A.S. NEMIROVSKI. Polynomial barrier methods in convex programming. *Èkonom. i Mat. Metody*, 24(6):1084–1091, 1988.

[18] Y.E. NESTEROV and A.S. NEMIROVSKI. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM Publications. SIAM, Philadelphia, USA, 1994.

[19] P. PARDALOS and H. WOLKOWICZ, editors. *Quadratic assignment and related problems*. American Mathematical Society, Providence, RI, 1994. Papers from the workshop held at Rutgers University, New Brunswick, New Jersey, May 20–21, 1993.

[20] P.M. PARDALOS and S.A. VAVASIS. Quadratic programming with one negative eigenvalue is NP-hard. *J. Global Optim.*, 1(1):15–22, 1991.

[21] 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.

[22] G. PATAKI and L. TUNÇEL. On the generic properties of convex optimization problems in conic form. *Math. Programming*, to appear.

[23] 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.

[24] S. POLJAK and H. WOLKOWICZ. Convex relaxations of $(0,1)$-quadratic programming. *Math. Oper. Res.*, 20(3):550–561, 1995.

[25] M.V. RAMANA. An exact duality theory for semidefinite programming and its complexity implications. *Math. Programming*, 77:129–162, 1997.

[26] 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.

[27] 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.

[28] A. SHAPIRO. Duality and optimality conditions. In *HANDBOOK OF SEMIDEFINITE PROGRAMMING: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Boston, MA, 2000.

[29] G. SPORRE and A. FORSGREN. Characterization of the limit point of the central path in semidefinite programming. Technical Report TRITA-MAT-02-OS12, Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden, 2002.

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

[31] R.J. VANDERBEI. *Linear Programming: Foundations and Extensions*. Kluwer Acad. Publ., Dordrecht, 1998.

[32] H. WOLKOWICZ, R. SAIGAL, and L. VANDENBERGHE, editors. *HANDBOOK OF SEMIDEFINITE PROGRAMMING: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Boston, MA, 2000. xxvi+654 pages.

[33] S. WRIGHT. *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, 1996.

[34] 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).