DIMACS Technical Report 95-02
January 1995

An Exact duality Theory for Semidefinite Programming and its Complexity Implications

by

Motakuri Ramana[1]
RUTCOR, Rutgers University
New Brunswick, New Jersey 08903

[1] I would like to acknowledge the support from RUTCOR and DIMACS.

DIMACS is a cooperative project of Rutgers University, Princeton University, AT&T Bell Laboratories and Bellcore.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

In this paper, we present a new and more complete duality for Semidefinite Programming (SDP), with the following features:

- This dual is an explicit semidefinite program, whose number of variables and the coefficient bitlengths are polynomial in those of the primal.
- If the Primal is feasible, then it is bounded if and only if the dual is feasible.
- The duality gap, i.e. the difference between the primal and the dual objective function values, is zero whenever the primal is feasible and bounded. Also, in this case, the dual attains its optimum.
- It yields a precise Farkas Lemma for semidefinite feasibility systems, i.e. characterization of the *infeasibility* of a semidefinite inequality in terms of the *feasibility* of another polynomial size semidefinite inequality.

Note that the standard duality for Linear Programming satisfies all of the above features, but no such duality theory was previously known for SDP, without Slater-like conditions being assumed.

Then we apply the dual to derive certain complexity results for Semidefinite Programming Problems. The decision problem of Semidefinite Feasibility (SDFP), i.e. that of determining if a given semidefinite inequality system is feasible, is the central problem of interest. The complexity of SDFP is unknown, but we show the following: 1) In the Turing machine model, SDFP is not NP-Complete unless NP=Co-NP; 2) In the real number model of Blum, Shub and Smale[BSS89], SDFP is in NP$\cap$Co-NP. We then give polynomial reductions from the following problems to SDFP: 1) Checking whether an SDP is bounded; 2) Checking whether a feasible and bounded SDP attains the optimum; 3) Checking the optimality of a feasible solution.

Part 1: MOTIVATION AND BACKGROUND

1 Introduction and Motivation

1.1 The Main Problem of Interest

We consider the following (Primal) Semidefinite Program (SDP):

$$\begin{array}{ll} \sup: & c^T x \\ \text{s.t.} & \sum_{i=1}^m x_i Q_i \preceq Q_0, \end{array} \tag{P}$$

(Throughout this paper, (P) will refer to this semidefinite program.)

Formally, we define,

Definition 1 : The Semidefinite Programming Problem (SDPP)
Determine and/or compute the following:

1. *Is the feasible region nonempty?*
2. *If so, is the objective function bounded?*
3. *If so, is the optimum attained?*
4. *If so, compute an optimum solution.*

Note that the first three parts are decision problems and the last is of numerical computation nature. Of the three decision problems, the first one is probably the most central:

Definition 2 : Semidefinite Feasibility Problem (SDFP)
Determine whether there exists an x such that $Q(x) \succeq 0$, where $Q(x) = Q_0 - \sum_i x_i Q_i$ for given symmetric matrices $Q_i, i = 0, .., m$ with integral entries.

We will show that there are polynomial reductions[GJ79] from parts 2. and 3. of SDPP to SDFP. At the moment, the complexity of SDFP is OPEN in both the Turing machine as well as the Real number models. Asking for the exact complexity of SDPP is not a well stated problem in the Turing model, as it may turn out that the optimal solutions of a rational semidefinite programs are irrational (in contrast with Linear Programming, where a rational LP always has rational optimal solution of polynomial bitlength); see §1.6. The existing algorithms for SDPP can be characterized as those that find *approximate* optimal solutions to SDP, and for this very reason are incapable of solving the decision problems 1 to 3 of SDPP in polynomial time.

The purpose of this paper is to present a new, explicit polynomial size dual program for SDP (§3) and employ it derive characterizations for several properties for semidefinite systems (§7,§8) and apply them to obtain certain complexity relations for Semidefinite Programming (§9).

1.2 On Semidefinite Programming

Semidefinite Programming (name given by Alizadeh[Ali91]) is a generalization of Linear Programming, obtained by replacing the nonnegativity constraints of Linear Programming by the semidefiniteness of some matrix variables and maps.

Historically, semidefinite programming has been studied in more general contexts such as convex and cone programming (see [Ali94] for some references).

However, the more recent surge of interest in SDP was primarily inspired by the work of [GLS84] (see [GLS88], Chapter 9). In that work, the authors associate with every graph G, a convex set denoted by TH(G), and show that when G is perfect, this set equals the stable set polytope. Then they demonstrate that one can optimize over TH(G) in polynomial time, and hence the stable set problem (along with many other related problems) can be solved in polynomial time for perfect graphs.

The algorithms of [GLS84] employ the ellipsoid method, and are not considered to be efficient in practice. In his Ph.D. work[Ali91], Farid Alizadeh showed that one can extend, in an almost mechanical fashion, many of the known interior point methods for LP into polynomial time algorithms for solving SDP *approximately* (see §1.5). Independently of Alizadeh's work, Nesterov and Nemirovskii[NN94] developed efficient interior point methods for a wider class of convex programs, by employing self-concordant barrier functions.

Standard Linear Programming has also been extended to Semidefinite Programming ([Ali94], [NN94]), but this extended duality theory is not complete, for instance, there are examples where the duality gap is a finite nonzero number (Example 1, §1.3.1). As already mentioned, in this paper, we remedy this situation, by deriving an explicit dual for SDP, that always has zero duality gap, among other properties.

A recent result of Goemans and Williamson[GW94] showing that one can use the solution obtained from a semidefinite relaxation to obtain a .878-approximation algorithm[1] for the Max-Cut problem, gave further impetus to Semidefinite Programming. Their result employs an ingenious randomized rounding scheme. This result has inspired other results on the application of SDP to combinatorial optimization problems.

1.3 Existing Duality Theories for SDP: A Critique

The dual to be proposed (see §3) in this paper satisfies the following criteria:

1. The dual is an explicit semidefinite program, which can be written down mechanically (i.e. without performing any computations), and its number of variables and bitlengths of its coefficients are polynomial in those of the primal.

2. The duality gap, i.e. the difference between the primal and the dual optimal objective function values, is zero whenever the primal is feasible and bounded. Also, in this case, the dual attains its optimum.

3. It yields a precise Farkas Lemma for semidefinite feasibility systems, i.e. characterization of the *infeasibility* of a semidefinite inequality in terms of the feasibility of another polynomial size semidefinite inequality.

We will now examine two previously known duality theories for SDP.

- The dual proposed in ([Ali94],[NN94]) satisfies 1 above, but needs to assume certain Slater-like constraint qualifications for 2 and 3 to hold.

- On the other hand, the duality theory proposed in [BW81] (when specialized to SDP) while theoretically satisfying 2 and 3 it fails to meet the specifications of 1: their dual is not an explicit semidefinite program, and it requires certain computations to be performed to extract the dual. Moreover, these computations are not shown to be and are unlikely to be polynomial time.

Let us now take a closer look at these two approaches.

1.3.1 The Lagrange-Slater Dual for SDP

Here is the dual proposed in [Ali94] and [NN94] (note that we are using Primal and Dual in reversed sense here as compared to [Ali94]). We will call this dual the *Lagrange-Slater Dual (LSD)* of (P).

$$\begin{array}{llcl} \inf & U \bullet Q_0 & & \\ \text{s.t.} & U \bullet Q_i & = & c \;\forall\; i = 1, .., m \\ & U & \succeq & 0 \end{array} \tag{LSD}$$

[1] The arithmetic complexity of their overall algorithm is polynomial. Unfortunately, neither the authors establish an explicit polynomial bound for the bitlengths of the intermediate numbers, nor does such a bound follow trivially from other results in the literature.

When a slater-like condition holds, for instance when there exists an x such that $\sum_i x_i Q_i \succ Q_0$, then the optimum values of the primal (P) and the dual (LSD) can be shown to be equal ([NN94]). While it is not clear whether these Slater conditions are restrictive in practical applications, it certainly makes the duality theory rather incomplete, as their absence can result in duality gaps. Here is a very simple SDP for which there is a nonzero finite duality gap (slight modification of an example in [VB94]).

Example 1: *SDP with nonzero duality gap*
Let the primal is

$$\begin{array}{ll} \sup: & x_2 \\ \text{s.t.} & \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \preceq \begin{bmatrix} \alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

where α is a positive constant. Then the LSD becomes:

$$\begin{array}{llll} \inf: & \alpha U_{11} & & \\ \text{s.t.} & U_{22} & = & 0 \\ & U_{11} + 2U_{23} & = & 1 \\ & U & \succeq & 0 \end{array}$$

Any primal feasible solution has $x_2 = 0$, and hence the primal optimum value is 0. Where as, in the dual, $U_{22} = 0$ forces $U_{23} = 0$, so that $U_{11} = 1$ for every dual feasible solution. Therefore the dual optimum value is α, resulting in a duality gap of α.

An example of failure of Farkas lemma that one obtains from LSD is given in §1.4 (Example 2). These deficiencies of LSD strongly motivated our investigations.

1.3.2 The Minimal Cone Based Approach

This approach was proposed by Borwein and Wolkowicz[BW81]. In simplified terms, they consider a *cone programming* problem

$$\sup\{c^T x | b - Ax \in S\},$$

where S is a convex cone, which in the case of SDP will be the cone of PSD matrices. For a convex subset T of S, they define the *minimal cone* $S^f(T)$ (slightly different notation being used here) of T to be the intersection of all the faces of S containing T. Then the following is an equivalent problem to the above cone-program:

$$\sup\{c^T x | b - Ax \in S^f(T)\},$$

where $T = \{b - Ax | x \in \Re^m\} \cap S$. The authors have shown that the "standard lagrange multiplier theorem" as applied to this modified problem always has zero duality gap, among other things. The authors then develop a theoretical algorithm for computing the minimal cone $S^f(T)$. While the results are mathematically very interesting, their approach suffers from the following drawbacks:

- Since their algorithm for computing the minimal cone has not been shown to be polynomial time, their approach doesn't yield a dual that can written in polynomial time; in contrast, the dual to be proposed ((ELSD);§3) has polynomial size (both the number of variables and the bitlengths of the coefficients) and it can be written down explicitly, without performing any computations whatsoever.

- It appears that, when specialized to SDP, their algorithm involves solving several semidefinite inequality systems; this is dubious for the following reason: semidefinite inequalities may not always have polynomial size (bitlengths) solutions (see examples 3-5 in §1.6) and hence their algorithm will most likely result in duals of exponential bitlength.

- Even if we put aside the polynomial time computability issues, it is quite far-fetched to have to solve several semidefinite programs just in order to write down the dual of a semidefinite program; duality theory for a mathematical program is an analytical tool and should provide a starting point in an algorithmic treatment of the problem and not vice versa.
- In particular, the complexity results established in the current paper seem to be impossible to derive based on their approach.

1.4 Farkas Lemma and Certificates

A related issue is that of Farkas Lemma and other *Theorems of Alternative.* In the case of Linear Programming, Farkas Lemma and its relatives characterize the solvability of one linear inequality system in terms of the unsolvability of another linear inequality system: The system $Ax \leq b$ is infeasible iff the system $A^T y = 0, b^T y = -1, y \geq 0$ is feasible. Such pairs of systems are called *Systems of Alternative.*

Systems of Alternatives essentially give "certificates" of infeasibility. Prior to the publication of the Ellipsoid Algorithm in 1979, which showed that Linear Programming was polynomial time solvable, the LP Farkas Lemma of the above para enabled one to conclude that the feasibility problem for linear inequalities is in NP$\cap$Co-NP (see Chapter 7 of [GJ79]). The significance of the membership of a problem in NP$\cap$Co-NP derives from the widely held belief that P=NP$\cap$Co-NP ([GJ79]).

Let us now turn to semidefinite inequality systems: Suppose that we want to characterize the feasibility/infeasibility of the system

$$Q(x) \succeq 0. \tag{1}$$

Then a direct extension of the LP Farkas Lemma ([Ali94]) gives the following as a system of alternative for (1).

$$Q_i \bullet U = 0 \;\forall\; i = 1, .., m, Q_0 \bullet U = -1, U \succeq 0. \tag{2}$$

However, this extension if inadequate as witnessed in the following example, where (1) and (2) are simultaneously infeasible.

Example 2: *Failure of "Farkas Lemma"*
Consider the inequality:

$$Q(x) = \begin{bmatrix} x & 1 & 0 \\ 1 & y & 0 \\ 0 & 0 & -x \end{bmatrix} \succeq 0,$$

and observe that it is infeasible. The system (2) for this map is given by:

$$U_{33} - U_{11} = 0, -U_{22} = 0, 2U_{12} = -1, U \succeq 0,$$

which is also infeasible since a positive semidefinite U with $U_{22} = 0$ will force U_{12} to be zero.

We will derive a polynomial size explicit semidefinite system of alternative (aka Farkas Lemma) for (1), and employ it to derive the following complexity results:

1. If SDFP$\in$NP, then SDFP$\in$Co-NP, and vice versa.
2. SDFP is not NP-Complete unless NP=Co-NP, in the Turing Machine model[GJ79].
3. SDFP is in NP$\cap$Co-NP in the real number model of Blum, Shub and Smale[BSS89].
4. There are polynomial time reductions from the following problems to SDFP:
   (a) Checking whether an SDP is bounded.
   (b) Checking whether a feasible and bounded SDP attains the optimum.
   (c) Checking the optimality of a feasible solution to an SDP.

1.5 An Overview of Known Complexity Results

By applying ellipsoid and interior point methods, one can deduce the following complexity results concerning the Semidefinite Program (P) of §1.1. First let

$$Q(x) = \sum_{i=1}^{m} x_i Q_i - Q_0,$$

where $Q_i, i = 0, .., m$ are given rational symmetric matrices, c is a rational vector, and let

$$G = \{x | Q(x) \succeq 0\}$$

be the feasible region of (P).

We let L denote the maximum of the bitlengths of the entries of the Q_i and the components of c, and define for $\epsilon > 0$,

$$S(G, \epsilon) = G + B(0, \epsilon) \text{ and } S(G, -\epsilon) = \{x | B(x, \epsilon) \subset G\}.$$

- If a positive integer R is known a priori such that either $G = \emptyset$ or $G \cap B(0, R) \neq \emptyset$, then there is an algorithm that solves the "weak optimization" problem, i.e. for any rational $\epsilon > 0$, the algorithm either finds a point $y \in S(G, \epsilon)$ that satisfies $c^T x \leq c^T y + \epsilon \ \forall \ x \in S(G, -\epsilon)$, or asserts that $S(G, -\epsilon)$ is empty ([GLS88]). The complexity of the algorithm is polynomial in n, m, L, and $\log(1/\epsilon)$.

- There is an algorithm which, given any rational $\epsilon > 0$ and x_0 such that $Q(x_0) \succ 0$, computes a rational vector $\overline{x}$ such that $Q(\overline{x}) \succ 0$, and $c^T \overline{x}$ is with in an additive factor ϵ of the optimum value of SDP. The arithmetic complexity of the algorithm is polynomial in $n, m, L, \log(1/\epsilon), \log(R)$ and the bitlength of x_0, where R is an integer such that the feasible region of the SDP lies the ball of radius R around the origin ([Ali94],[NN94]). However, it should be mentioned that a polynomial bound has not been established for the bitlengths of the intermediate numbers occurring in these algorithms.

- For any fixed m, there is a polynomial time algorithm (in n, L), that checks whether there exists an x such that $Q(x) \succ 0$, and if so, computes such a vector ([Ram93]).

In a recent work[Fre95], Bob Freund discusses interior-point algorithms for SDPs in which no regularity (Slater-like) conditions are assumed. Progression from "weak optimization" to "strong optimization" is made difficult as there exist SDPs with certain undesirable features as described in the following.

1.6 "Ill-Conditioned" SDPs

1. *SDPs with no rational optimal solutions*: For a symmetric matrix A, consider the SDP

   $$\max\{t | A - tI \succeq 0\},$$

   and note that its optimal value equals the smallest eigenvalue of A, and this is usually irrational even for rational matrices A.

2. *Feasible regions with small volume*:

   Example 3: *Khachian's Example:*
   Consider the map:

   $$Q(x) = Q_1(x) \oplus Q_2(x) \oplus \ldots Q_n(x) \oplus [1 - x_n]$$

   where

   $$Q_i(x) = \begin{bmatrix} x_0 & 2x_{i-1} \\ 2x_{i-1} & x_i \end{bmatrix},$$

   and $\oplus$ denotes the direct sum operation for matrices (§2.1).

It is then easily shown that for any x with $Q(x) \succeq 0$, $x_i \leq 1, 4^{2^i-1}x_0 \leq x_i \ \forall \ i = 1,..,n$. Hence, the volume of G is doubly exponentially small in n. Further, there exists no rational point satisfying $Q(x) \succ 0$ whose bitlength is single exponential in n. Hence, any polynomial algorithm that attempts to check whether $\{x|Q(x) \succeq 0\} \neq \emptyset$ can not actually output such an x. This example was found by L.G. Khachian.

3. *SDPs with doubly exponential optimal solutions*: Rational optimal solutions exist, but they require exponentially many bits: In the above example $\max\{x_0|Q(x) \succeq 0\} = 4^{1-2^n}$.

4. *Feasible region too far from the origin* The distance of G from the origin is doubly exponential. In the above example, if we replace $x_n \leq 1$ by $x_0 \geq 1$, then the feasible region is at a doubly exponential distance from the origin. Another (somewhat simpler) example of this type is the following.

   Example 4: Consider the map:

$$Q(x) = [x_1 - 2] \oplus Q_1(x) \oplus \ldots Q_{n-1}(x)$$

   where

$$Q_i(x) = \begin{bmatrix} 1 & x_i \\ x_i & x_{i+1} \end{bmatrix}, i = 1,..,n-1.$$

   Then

$$G = \{Q(x) \succeq 0\} = \{x|x_1 \geq 2, x_2 \geq x_1^2,..,x_n \geq x_{n-1}^2\},$$

   and hence for any $x \in G$,

$$x_i \geq 2^{2^{i-1}} \ \forall \ i,$$

   which is doubly exponential and therefore any rational solution in G has exponential bitlength and is at an exponential distance from the origin. Clearly, like the Khachian's example, even this semidefinite system has no rational solutions of polynomial bitlength.

5. *SDPs with unattained optimum:*

   Example 5: Consider the map

$$Q(x) = \begin{bmatrix} x_1 & 1 \\ 1 & x_2 \end{bmatrix} \succeq 0,$$

   and note that $\inf\{x_1|Q(x) \succeq 0\} = 0$ (this is the infimum of x_1 over the top right lobe of the hyperbola $x_1x_2 \geq 1$), but it is not attained.

The success of ellipsoid and interior point methods in establishing the polynomial complexity for the strong versions of linear and convex quadratic programming relies on the fact that the above idiosyncrasies do not arise for these problems. For this reason, it seems reasonable to surmise that a completely new approach might indeed be required to establish polynomial time solvability of SDFP, if at all it is possible. We elaborate this point as follows.

As we have seen, the optimal as well as feasible solutions of SDPs could potentially be exclusively irrational, double exponential in bitlength, or sometimes the optimum may not even be attained, so that one can not hope to write down explicitly an optimal solution or for that matter an arbitrary feasible solutions of general SDPs in polynomial time. But, there might be alternate implicit polynomial size representations for these solutions. For instance, the optimal solution of the SDP in 3. above is the *unique* solution to the system

$$x_n = 1, x_0x_i = 4x_{i-1}^2, i = 1,..,n,$$

and it takes polynomially many bits to write down this system. Similarly, the optimum value of $\max\{t|A - tI \succeq 0\}$ is the smallest root of the characteristic polynomial of A, and clearly it is the *unique* solution to the polynomial system

$$S_k(A - tI) \geq 0 \ \forall \ k = 1,..,n-1, S_n(A - tI) = \det(A - tI) = 0,$$

where $S_k(\bullet)$ denotes the kth elementary symmetric function of a matrix[HJ85]. Now, if A is an integral matrix whose entries have bitlength at most L, then the above polynomial system has size polynomial in n, L (as each of the $S_k(A-tI)$ is the sum of single exponentially many subdeterminants of $A-tI$, which are polynomials whose coefficients are of polynomial bitlength).

2 Preliminaries

2.1 Notation

Most of the notation is taken from [HJ85] and [Roc70]. The main matrix spaces of interest here are:

- $\mathcal{M}_n$: The space of $n \times n$ real matrices.
- $\mathcal{S}_n$: The subspace of symmetric matrices in $\mathcal{M}_n$.

For $A, B \in \mathcal{S}_n$, we write $A \succeq B$ (resp. $\succ$), if $A - B$ is positive semidefinite (resp. positive definite), i.e. all the eigenvalues of $A - B$ are nonnegative (resp. positive). We will sometimes abbreviate "positive semidefinite" by PSD and "positive definite" by PD.

The inner product on $\mathcal{M}_n$ (and $\mathcal{S}_n$) is given by:

$$A \bullet B = \sum_{i,j} A_{ij} B_{ij}.$$

Given $A \in \mathcal{M}_m, B \in \mathcal{M}_k$, their *direct sum* is the block partitioned $(m+k) \times (m+k)$ matrix:

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

The *spectral decomposition* of a symmetric matrix A is $A = V^T D V$, where V is an orthogonal matrix (i.e. $V^T V = I$) whose columns are eigenvectors of A, and D is the diagonal matrix of the eigenvalues of A. The *spectral radius* of a symmetric matrix A, denoted by $\rho(A)$, is the largest of the magnitudes of the eigenvalues of A.

For $A, B \subset \Re^n$, $A + B$ denotes the *Minkowski Sum* (also called the *set sum*). For a scalar $\alpha \in \Re$, and a set A in $\Re^n$, $\alpha A = \{\alpha x | x \in A\}$. For $A \subset \Re^n$, $\mathrm{Conv}(A)$ (resp. $\mathrm{Aff}(A)$) denotes the smallest convex set (resp. affine subspace) containing A. The dimension of A is $\dim(\mathrm{Aff}(A))$, and $B(x, r)$ is the ball of radius r around x.

For a convex set $G \subset \Re^n$, the interior and the relative interior of G are denoted by $\mathrm{Int}(G)$ and $\mathrm{ri}(G)$ respectively. The *recession cone* of G is defined by:

$$0^+(G) = \{v \in \Re^n \mid \forall\ x \in G, t \geq 0, x + tv \in G\},$$

and the *lineal hull* of G (contrast with linear hull) is the subspace $0^+(G) \cap 0^+(-G)$.

We consider a *Semidefinite Program* to be the optimization problem:

$$\begin{array}{ll} \sup: & c^T x \\ \text{s.t.} & \sum_{i=1}^m x_i Q_i \preceq Q_0. \end{array} \tag{P}$$

Note that any semidefinite program as considered in [Ali94] can be cast in the above form. We say that (P) is *homogeneous*, if $Q_0 = 0$.

Definition 3 *A* **Spectrahedron** *is defined to be a closed convex set of the following type:*

$$G = \{x | Q(x) \succeq 0\},$$

where $Q(x)$ is an affine symmetric matrix map.

It is clear that spectrahedra are precisely the feasible regions of semidefinite programs.

2.2 Some Useful Facts About PSD Matrices

We now collect as a proposition, some useful and well known properties of semidefinite matrices, most of whose proofs can be found in [HJ85].

Proposition 1 *The following hold.*

1. *If $U \succeq 0$, and $U_{ii} = 0$, then $U_{ij} = 0 \; \forall \; j$.*
2. *If $A \in \mathcal{S}_n$ and $S \in \mathcal{M}_n$ and nonsingular, then $A \succeq 0$ iff $S^T AS \succeq 0$. (Special case of the Sylvester's law of inertia.)*
3. *If $A \succeq 0$ and $u \in \Re^n$, then $u^T Au = 0$ iff $Au = 0$.*
4. *Given a block partitioned matrix*
$$U = \begin{bmatrix} A & B^T \\ B & C \end{bmatrix},$$
where A and C are square, and A is nonsingular, then the Schur Complement of A in U is the matrix $S = C - BA^{-1}B^T$. We have that, if $A \succ 0$, then $U \succeq 0 \Leftrightarrow S \succeq 0$ and $U \succ 0 \Leftrightarrow S \succ 0$.
5. *$A \in \mathcal{S}_n$ is PSD iff $A \bullet B \geq 0$ for all $B \succeq 0$.*
6. *If $A, B \succeq 0$, then $A \bullet B = 0$ if and only if $AB = 0$.*
7. *If $U - WW^T \succeq 0$, then there exists a matrix $H \in \mathcal{M}_n$ such that $W = UH$.*
8. *If $U_1, U_2 \succeq 0$, then $Null(U_1 + U_2) = Null(U_1) \cap Null(U_2)$.*

2.3 Polarity and Duality

Our approach employs polars of convex sets, and in this subsection we describe some results concerning polars and gauges(see [Roc70] and [Hol75]).

Let G be a closed convex set containing the origin. Then the *polar* of G is

$$G^\circ = \{y | y^T x \leq 1 \; \forall \; x \in G\},$$

and its *value function* (also known as support function) is

$$v_G(y) = \sup\{y^T x | x \in G\},$$

and let the *gauge* function of G be defined as

$$\rho_G(y) = \inf\{t > 0 | y \in tG\}.$$

(By convention, infimum is ∞ if for no $t > 0$, $y \in tG$.)

Proposition 2 *Let $G \subset \Re^m$ be a closed convex set and $0 \in G$, and let $y \in \Re^m$. Then the following hold.*

1. *G° is a closed convex set containing the origin.*
2. *$G = G^{\circ\circ}$.*

The proofs are quite standard, and may be found in [Roc70]. The following lemma, linking polars, support functions and gauges, serves as a prototype for the proof of our duality theorem (§6).

Lemma 1 *For closed convex set $0 \in G \subset \Re^m$, and a vector $c \in \Re^m$, the following hold.*

1. *$c^T x$ is bounded on G iff there exists $t > 0$ such that $tc \in G^\circ$.*

2. *The origin maximizes $c^T x$ iff $ty \in G^\circ$ for every $t > 0$.*

3. *When $c^T x$ is bounded on nonempty G, then $v_G(y) = \rho_{G^\circ}(y)\ \forall\ y \in \Re^m$.*

Proof: Let $y \in \Re^m$.

$$\begin{array}{rcl} \rho_{G^\circ}(y) & = & \inf\{t > 0 | y \in tG^\circ\} \\ & = & \inf\{t > 0 | x^T y \le t\ \forall\ x \in G\} \\ & = & \sup\{y^T x | x \in G\} \\ & = & v_G(y) \Box \end{array}$$

The lemma trivially follows. □ We will use the following folklore result. A quick proof is included.

Proposition 3 *If $K \subset \Re^m$ is a convex set such that $Cl(K) = \Re^m$, then $K = \Re^m$.*

Proof: From Theorem 6.3 of [Roc70],

$$\mathrm{ri}(K) = \mathrm{ri}(\mathrm{Cl}(K)) = \Re^m,$$

and hence $K = \Re^m$ □.

2.4 A Brief Sketch of the Approach

The technique used here emerged from the results of [RG94a], where certain geometric properties of spectrahedra (G) were investigated. In particular, in order to understand the polars (G°) of spectrahedra, an object called the algebraic polar(G^*) was introduced(see §5). This is an "algebraic approximation" of polar, which is a geometric object. The relations established there were: 1) $G^\circ = \mathrm{Cl}(G^*)$ 2) $G^\circ = G^* + \mathrm{Aff}(G)^\perp$ (Lemmas 5 and 6 here). A closer examination of the proof of 2) revealed that one can extract very useful information concerning $\mathrm{Aff}(G)$ whenever G^* turns out not to be closed (and hence not equal to G°). This observation is exploited here in §5.

In the next section, we introduce the dual, called *Expanded Lagrange-Slated Dual(ELSD)* for SDP. Then we establish weak duality for ELSD in §4. In section 5, we establish an exact algebraic description of G° (§5). The polar then readily yields the strong duality properties of the dual ELSD(Theorem 1; proof in §6). A Farkas Lemma for semidefinite systems is derived in §7, and further properties of SDP are then characterized in §8. Finally, in §9, we establish the stated complexity results for SDP.

Part 2: DERIVATION OF THE DUAL

3 The Expanded Lagrange-Slater Dual for SDP

The following is our primal semidefinite program:

$$\begin{array}{lll} \sup: & c^T x & \\ \text{s.t.} & \sum_{i=1}^m x_i Q_i & \preceq \quad Q_0, \end{array} \tag{P}$$

Let us define the following.

- $G := \{x | \sum_i x_i Q_i \preceq Q_0\}$ is feasible region of (P).
- $\hat{Q}(x) = \sum_{i=1}^m x_i Q_i$.
- $Q(x) = Q_0 - \hat{Q}(x)$ (so that $G = \{x | Q(x) \succeq 0\}$).
- $Q^* : \mathcal{M}_n \to \Re^m$ is defined by $Q^*(U) = U \bullet Q_i, i = 1, .., m$.
- $Q^\# : \mathcal{M}_n \to \Re^{m+1}$, defined by

$$Q^\#(U) = \begin{pmatrix} Q_0 \bullet U \\ Q^*(U) \end{pmatrix}.$$

Note that the restriction of Q^* to $\mathcal{S}_n$ is the adjoint of $\hat{Q}$.

We now introduce the new dual semidefinite program, named **Extended Lagrange-Slater Dual (ELSD)** for SDP.

$$\begin{array}{llll} \inf: & (U + W_m) \bullet Q_0 & & \\ \text{s.t.} & Q^*(U + W_m) & = & c \\ & Q^\#(U_i + W_{i-1}) & = & 0, i = 1, .., m \\ & U_i & \succeq & W_i W_i^T, i = 1, .., m \\ & W_0 & = & 0 \end{array} \tag{ELSD}$$

Note that the constraint $U \succeq WW^T$ can alternately be written as (by Propositions 1(4))

$$\begin{bmatrix} I & W^T \\ W & U \end{bmatrix} \succeq 0,$$

and consequently (ELSD) is indeed a semidefinite program. A glossary of the domains of the variables is: $U \in \mathcal{S}_n, U_i \in \mathcal{S}_n \ \forall\ i = 1, .., m$ and $W_i \in \mathcal{M}_n \ \forall\ i = 1, .., m$ (and we use an auxiliary matrix variable $W_0 = 0$ for notational brevity).

Size of ELSD: Note that the ELSD has $O(mn^2)$ number of variables, and the coefficients are essentially those of the primal. Hence ELSD is a polynomial size semidefinite program.

Let us define for a arbitrary positive integer k,

$$\begin{array}{lll} \mathcal{C}_k & = & \{(U_i, W_i)_{i=1}^k | Q^\#(U_i + W_{i-1}) = 0, U_i \succeq W_i W_i^T, \ \forall\ i = 1, .., k, W_0 = 0\} \\ \mathcal{U}_k & = & \{U_k | (U_i, W_i)_{i=1}^k \in \mathcal{C}_k\} \\ \mathcal{W}_k & = & \{W_k | (U_i, W_i)_{i=1}^k \in \mathcal{C}_k\} \end{array}$$

In terms of the above notation, we can write ELSD compactly and in a form that closely resembles LSD os §1.3.1. Future references to (ELSD) will be to this form of the dual.

$$\begin{array}{llll} \inf: & (U + W) \bullet Q_0 & & \\ \text{s.t.} & Q^*(U + W) & = & c \\ & W & \in & \mathcal{W}_m \\ & U & \succeq & 0, \end{array} \tag{ELSD}$$

A pair (U, W) is said to be *dual feasible*, if it satisfies the above constraints.

In the next few sections, we will derive the properties of ELSD and its relation to the primal SDP. Specifically, we will prove the following strong duality theorem, which closely resembles LP Duality.

Theorem 1 (Duality Theorem) *The following hold for the Primal-Dual pair (P)-(ELSD).*

1. *(***Weak Duality***) If x is primal feasible and (U, W) is dual feasible, then $c^T x \le (U + W) \bullet Q_0$.*

2. *(***Primal Boundedness***) If the primal is feasible, then the primal optimum value is finite if and only if the dual is feasible. In fact, this result holds for the weaker version of (ELSD), obtained by replacing $W \in \mathcal{W}_m$ by $W \in \mathcal{W}_{m-1}$.*

3. *(***Zero Gap***) If both the primal and the dual are feasible, then their optimum values are equal.*

4. *(***Dual Attainment***) Whenever the common optimum value is finite, the dual attains this value.*

The only difference between a standard LP Duality theorem and the above it borne by the fact that some semidefinite programs (i.e. primal) may not achieve their optimum values (example 5 in §1.6). The proof of the theorem is given in §6.

We would like to note that, a somewhat more stringent argument can be used to show that one may replace $W \in \mathcal{W}_m$ in (ELSD) by $W \in \mathcal{W}_K$, where $K = \min\{m-1, n\}$, while preserving the correctness of the duality theorem. This improvement reduces the numbers of variables in the dual. We will not delve into these details here.

4 Weak Duality

In this section, we will establish weak duality for (ELSD). We need the following lemma.

Lemma 2 *Let $k \le m$. Then for any $x \in G$ and $U \in \mathcal{U}_k$ or $W \in \mathcal{W}_k$, we have*

$$Q(x)U = 0, Q(x)W = 0,$$

and hence $Q(x) \bullet U = 0 = Q(x) \bullet W$. In particular, if $0 \in G$, then $Q_0 \bullet U = 0 = Q_0 \bullet W$.

Proof: Let $x \in G$ and

$$(U_i, W_i)_{i=1}^k \in \mathcal{C}_k.$$

To start with, note that as $W_0 = 0$, and $Q^{\#}(U_1) = 0$, we have

$$Q(x) \bullet U_1 = 0,$$

Since $Q(x), U_1 \succeq 0$, by Proposition 1(6),

$$Q(x)U_1 = 0.$$

Since $U_1 \succeq W_1 W_1^T$, by Proposition 1(7), we infer that there exists a matrix H such that

$$W_1 = U_1 H,$$

and consequently,

$$Q(x)W_1 = 0 \text{ and } Q(x) \bullet W_1 = 0.$$

Since $Q^{\#}(U_1 + W_1) = 0$, the above gives

$$Q(x) \bullet U_2 = 0,$$

which as above gives $Q(x)U_2 = 0$. It is clear that by continuing in this fashion, one can conclude the truth of the lemma. □

That weak duality holds for the Primal-Dual pair (P) and (ELSD) is shown below.

Proof of Weak Duality (Theorem 1(1)): Let x be primal feasible, $k \leq m$, and (U, W) satisfy:

$$\begin{array}{rcl} Q^*(U+W) & = & c \\ W & \in & \mathcal{W}_k \\ U & \succeq & 0. \end{array}$$

We then have

$$\begin{array}{rcl} c^T x & = & x^T Q^*(U+W) \\ & = & -Q(x) \bullet (U+W) + Q_0 \bullet (U+W) \\ & = & -Q(x) \bullet U - Q(x) \bullet W + Q_0 \bullet (U+W) \end{array}$$

Since $Q(x), U \succeq 0$, $-Q(x) \bullet U \leq 0$. From the above lemma, $Q(x) \bullet W = 0$, and we get

$$c^T x \leq Q_0 \bullet (U+W).$$

Applying the above with $k = m$ yields that weak duality holds for the pair (P)-(ELSD). □

We will establish the following lemma for use in future sections. It sheds some light on the incremental structure of the dual (ELSD).

Lemma 3 *The following hold.*

1. *For any $k \leq m$, $\mathcal{W}_k$ and $Q^*(\mathcal{W}_k)$ are linear subspaces.*
2. *$\mathcal{W}_1 \subset \ldots \subset \mathcal{W}_m$.*
3. *$Q^*(\mathcal{W}_1) \subset \ldots \subset Q^*(\mathcal{W}_m)$.*
4. *If $0 \in G$, $G \subset (Q^*(\mathcal{W}_k))^\perp$ for every $k = 1, .., m$.*

Proof: To prove 1, let $W \in \mathcal{W}_k$ and let

$$(U_i, W_i)_{i=1}^k \in \mathcal{C}_k$$

such that $W = W_k$. We will first show that $\mu W \in \mathcal{W}_k$ for every $\mu \in \Re$. Define matrices

$$\begin{array}{rcl} W_i' & = & \mu^{2^{k-i}} W_i \\ U_i' & = & \mu^{2^{k-i+1}} U_i \end{array}$$

for every $i = 1, .., k$. Then,

$$Q^\#(U_i' + W_{i-1}') = \mu^{2^{k-i+1}} Q^\#(U_i + W_{i-1}) = 0,$$

and

$$U_i' - W_i' W_i'^T = \mu^{2^{k-i+1}} (U_i - W_i W_i^T) \succeq 0,$$

and it follows that

$$(U_i', W_i')_{i=1}^k \in \mathcal{C}_k,$$

hence

$$\mu W_k \in \mathcal{W}_k.$$

We can conclude that $\mathcal{W}_k$ is subspace, once we show that $\mathcal{W}_k$ is convex, for which it suffices to show that $\mathcal{C}_k$ is convex, which follows trivially using the fact that one can write the condition $U - WW^T \succeq 0$ as

$$\begin{bmatrix} I & W^T \\ W & U \end{bmatrix} \succeq 0.$$

The proof of 2 is as follows: If

$$(U_i, W_i)_{i=1}^k \in \mathcal{C}_k,$$

then set

$$U_i' = U_{i-1}, W_i' = W_{i-1} \ \forall \ i = 2, .., k+1, \text{ and } U_1' = 0 = W_1',$$

and observe that

$$(U_i', W_i')_{i=1}^{k+1} \in \mathcal{U}_{k+1}.$$

Clearly 2 implies 3, and the proof is complete, as 4 is an easy consequence of Lemma 3.□

5 A Description of Polars of Spectrahedra

In this section, we derive an expression for the polar of an arbitrary spectrahedron defined by

$$G = \{x|Q(x) \succeq 0\},$$

and containing the origin, i.e. $Q_0 \succeq 0$.

The polar is defined to be

$$G^\circ = \{y|x^T y \leq 1, \ \forall \ x \in G\}.$$

Our main result of the section is

Theorem 2 (Polar Description) *If G contains the origin, then its polar is given by*

$$G^\circ = \{Q^*(U+W)|W \in \mathcal{W}_k, U \succeq 0, U \bullet Q_0 \leq 1\} \ \forall \ k \geq m-1.$$

The proof will be given shortly, but first consider following object, called the *Algebraic Polar* of G (with respect to the representation $Q(x) \succeq 0$). It serves as an "initial approximation" for G°:

$$G^* = \{Q^*(U)|U \bullet Q_0 \leq 1, U \succeq 0\}.$$

We reproduce here the proof that G^* indeed approximates G° except on a set of measure zero.

Lemma 4 *Assuming that $0 \in G$, we have $G^\circ = Cl(G^*)$.*

Proof: Let us first prove that $G^\circ \supset G^*$:

$$\begin{array}{rcl} G^\circ & = & \{y|\sup\{y^T x|Q(x) \succeq 0\} \leq 1\} \\ & \supset & \{Q^*(U)|U \bullet Q_0 \leq 1, U \succeq 0\} \\ & = & G^* \end{array}$$

Since G° is always closed, $\mathrm{Cl}(G^*) \subset G^\circ$. To show the reverse inclusion, put $H = \mathrm{Cl}(G^*)$ and consider any $w \in H^\circ$. Then, we have

$$w^T(Q^*(U)) \leq 1 \text{ whenever } U \succeq 0, U \bullet Q_0 \leq 1. \qquad (*)$$

We claim that $Q(w) \succeq 0$. First, if $Q_0 = 0$, then (*) implies

$$Q(w) \bullet U \geq -1 \ \forall \ U \succeq 0,$$

which happens if and only if $Q(w) \succeq 0$.

Suppose that $Q_0 \neq 0$. Let $V \succ 0$, and choose $\lambda > 0$ such that $\lambda V \bullet Q_0 = 1$, which is possible since $Q_0 \succeq 0$ and nonzero. Then

$$(\lambda V) \bullet Q(w) = (\lambda V) \bullet Q_0 - w^T Q^*(\lambda V) = 1 - w^T Q^*(\lambda V) \geq 0 \text{ by } (*).$$

Therefore, $Q(w) \bullet V \geq 0 \ \forall \ V \succ 0$, implying that $w \in G$, and hence $H^\circ \subset G$. But then,

$$H = H^{\circ\circ} \supset G^\circ,$$

and the proof is complete. □.

The following fact was established in [RG94a] and its proof is omitted as we will not use the result directly here.

Lemma 5 *Let $G = \{x|Q(x) \succeq 0\}$ with $Q_0 \succeq 0$. Then*

$$G^\circ = Cl(G^*) = G^* + Aff(G)^\perp$$

The essence of the lemma is that, the difference of the polar G° and its algebraic approximation G^* is offset by the orthogonal complement of the affine hull of G. However, the disadvantage of the above lemma is that, one needs to first have at their disposal, a description of the affine hull of G, say as $\{x|Ax = 0\}$ for some A. However, such a representation is not readily available. We attack the problem as follows.

As seen, if G^* turns out to be closed, we already have a description of G°. So suppose that this is not the case. Interstingly enough, from the nonclosedness, we can obtain information on the annihilators of G (as will be witnessed in Claim 1 and 2 below).

Proof of Theorem 2:

Let us first consider the algebraic polar G^* of G:

$$G^* = \{Q^*(U)|U \bullet Q_0 \leq 1, U \succeq 0\},$$

and the subspaces (see Lemma 4) of $\Re^m$

$$S_0 = \{0\}, S_k = Q^*(\mathcal{W}_k) \ \forall\ k = 1, .., m,$$

which satisfy

$$S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_m.$$

Let us invoke the assumption that $0 \in G$, and apply Lemma 2 with $x = 0$ to conclude that for any $W \in \mathcal{W}_k$,

$$Q_0 W = 0,$$

and hence by another application of Lemma 2, we get that

$$x^T Q^*(W) = 0 \ \forall\ x \in G.$$

This implies easily that

$$G^* + S_k \subset G^\circ.$$

Therefore, we have

$$G^* \subset G^* + S_1 \subset G^* + S_2 \subset \ldots \subset G^\circ = \mathrm{Cl}(G^*).$$

If for some $k \leq m$, $G^* + S_k$ is closed, then clearly

$$G^* + S_j = G^\circ \ \forall\ j \geq k.$$

We will then show (Claim 1 below) that whenever $G^* + S_k$ (with $0 \leq k \leq m-1$) is not closed, then

$$S_k \subsetneq S_{k+1},$$

and therefore

$$\dim(S_k) < \dim(S_{k+1}).$$

We now assert that the above implies

$$G^\circ = G^* + S_{m-1}.$$

For the sake of a contradiction, suppose that this is not the case.

If $G^* + S_{m-1}$ is not closed, then $G + S_k$ is not closed for every $0 \leq k \leq m-1$, which by Claim 1 will result in

$$0 = \dim(S_0) < \dim(S_1) < \dim(S_2) < \ldots \dim(S_{m-1}) < \dim(S_m),$$

implying that $S_m = \Re^m$. But by Lemma 4(4), this implies that $G = \{0\}$, and hence $G^\circ = \Re^m$. But then $\mathrm{Cl}(G^*) = \Re^m$, and by Proposition 3, the only convex set whose closure is the whole space is the whole space itself and hence $G^* = \Re^m$, a contradiction. $\Box$.

Let us now prove the claim made in the above proof.

Claim 1 *If $G^* + S_k$ is not closed, then*

$$S_k \subsetneq S_{k+1} \text{ and } \dim(S_k) < \dim(S_{k+1}).$$

Proof: Let A be a $r \times m$ full row rank matrix such that

$$T := S_k^{\perp} = \{x | Ax = 0\},$$

and suppose that $G^* + S_k = G^* + T^{\perp}$ is not closed.

Our main tool is the result depicted in claim 2 below, according to which, there exists a matrix U and a vector $\lambda \in \Re^r$ such that

$$\begin{aligned} Q^*(U) + A^T\lambda &= 0 \\ Q_0 \bullet U &= 0 \end{aligned}$$

and $\{x | Ax = 0, Q(x)U = 0\}$ is a subspace with a strictly smaller dimension than T. We will use the quantities U, λ as follows. First, note that

$$T^{\perp} = S_k = Q^*(\mathcal{W}_k),$$

and hence we may choose $W \in \mathcal{W}_k$ such that

$$A^T\lambda = Q^*(W).$$

Since $W \in \mathcal{W}_k$, $Q_0 W = 0$ from an application of Lemma 3, and hence

$$\begin{aligned} Q^*(U + W) &= 0 \\ Q_0 \bullet (U + W) &= 0. \end{aligned}$$

From the claim, the subspace

$$\{x \in T | Q(x)U = 0\} = \{x \in T | \sum_{i=1}^{m} x_i Q_i U = 0\}$$

(the equality follows from $Q_0 U = 0$) is strictly smaller than $T = S_k^{\perp}$, and hence there exists a matrix $V \in \mathcal{M}_n$ (multiplier for the equations $\sum_{i=1}^m x_i Q_i U = 0$) such that the vector

$$v = Q^*(UV) \notin S_k.$$

The proof of the claim will be completed by showing that $v \in S_{k+1} = Q^*(\mathcal{W}_{k+1})$. Let

$$(U_i, W_i)_{i=1}^k \in \mathcal{C}_k,$$

such that $W_k = W$. Obviously neither U nor V is zero, for otherwise, $v = 0$ will be in the subspace S_k. Now let us define

$$\mu = 1/(\rho(V)\sqrt{\rho(U)}),$$

and set $U_{k+1} = U$ and $W_{k+1} = \mu UV$. Let us show that $U_{k+1} \succeq W_{k+1} W_{k+1}^T$:

$$\begin{aligned} W_{k+1} W_{k+1}^T &= (1/\rho(U)\rho(V)^2) UVV^TU \\ &\preceq (1/\rho(U))U^2 \\ &\preceq U. \end{aligned}$$

It is now clear that

$$(U_i, W_i)_{i=1}^{k+1} \in \mathcal{C}_{k+1},$$

and hence

$$\mu v = Q^*(W_{k+1}) \in S_{k+1},$$

and since the latter is a subspace, it contains v. The proof of Claim 1 is complete. □

The proof of the following Claim 2 was inspired and closely parallels that of Lemma 5 given in [RG94a].

Claim 2 *Let $T = \{x|Ax = 0\}$, where A is an $r \times m$ full row rank matrix, and let $G = \{x|Q(x) \succeq 0\}$ be any spectrahedron containing the origin. If $G^* + T^\perp$ is not closed, then there exist $U \succeq 0$ and $\lambda \in \Re^r$ such that*

$$\begin{aligned} Q^*(U) + A^T\lambda &= 0 \\ Q_0 \bullet U &= 0 \end{aligned}$$

and $\{x|Ax = 0, Q(x)U = 0\}$ is a subspace whose dimension is strictly smaller than T.

Proof: The proof is by induction on n (the size of the matrices Q_i), with the base case being $n = 1$, which is trivial to verify. Put $Z = G^* + T^\perp$ and suppose that $w \in \mathrm{Cl}(Z)\backslash Z$, and pick a sequence $[U(i) \succeq 0, \lambda(i) \in \Re^r], i = 1, .., \infty$ such that

$$U(i) \bullet Q_0 \leq 1 \;\forall\; i \text{ and } \lim_{i\to\infty} Q^*(U(i)) + A^T\lambda(i) = w \notin Z$$

If the sequence $[U(i), \lambda(i)]$ has an accumulation point, say, $[\tilde{U}, \tilde{\lambda}]$, we will then have

$$w = Q^*(\tilde{U}) + A^T\tilde{\lambda} \in Z,$$

a contradiction. Hence either the sequence $\{Tr(U(i)) = U(i) \bullet I\}$ or the sequence $\{||\lambda(i)||\}$ is unbounded.

Now define the normalized sequence

$$[\hat{U}(i), \hat{\lambda}(i)] = [U(i), \lambda(i)]/(U(i) \bullet I + ||\lambda(i)||) \;\forall\; i,$$

and assume, by passing to a subsequence if necessary, that the sequence converges to $(\hat{U}, \hat{\lambda})$. Note that the following hold:

- $\hat{U} \succeq 0$.
- $Q^*(\hat{U}) + A^T\hat{\lambda} = \lim w/(U(i) \bullet I + ||\lambda(i)||) = 0$.
- At least one of $\hat{U}$ or $\hat{\lambda}$ is nonzero.
- In fact, $\hat{U} \neq 0$. This can be seen as follows: if $\hat{U} = 0$, then $A^T\hat{\lambda} = 0$, which contradicts the fact that A has full row rank.
- Also, since $0 \leq U(i) \bullet Q_0 \leq 1 \;\forall\; i$ (as $Q_0 \succeq 0$ from $0 \in G$), it follows that $\hat{U} \bullet Q_0 = 0$.

Now consider the affine subspace

$$T' = \{x|Ax = 0, Q(x)\hat{U} = 0\}.$$

Since $Q_0 \bullet \hat{U} = 0$ (and $Q_0 \succeq 0$), by Propositions 1(6), $Q_0\hat{U} = 0$, and hence

$$T' = \{x|Ax = 0, \hat{Q}(x)\hat{U} = 0\},$$

and consequently T' is a linear subspace.

Now, if $\dim(T') < \dim(T)$, we are done. Suppose that this is not the case, so that $T' = T$. We will first consider the case when

$$\hat{U} = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix},$$

where the identity block is $l \times l$, for some $l > 0$ (since $\hat{U} \neq 0$). Also, block partition $Q(x)$ accordingly

$$Q(x) = \begin{bmatrix} P(x) & R(x)^T \\ R(x) & S(x) \end{bmatrix},$$

giving

$$Q(x)\hat{U} = \begin{bmatrix} 0 & R(x)^T \\ 0 & S(x) \end{bmatrix}, \text{ and}$$

$$T' = \{x|Ax = 0, R(x) = 0, S(x) = 0\}.$$

Since $T' = T$, we have

$$Ax = 0 \Rightarrow R(x) = 0 \text{ and } S(x) = 0,$$

implying that

$$G \cap T = \{x|P(x) \succeq 0, Ax = 0\} = H \cap T,$$

where H is the spectrahedron defined by

$$H = \{x|P(x) \succeq 0\}.$$

Evidently, $H^* \subset G^*$, and so

$$H^* + T^\perp \subset G^* + T^\perp.$$

We have,

$$\begin{array}{rcl} \mathrm{Cl}(H^* + T^\perp) & = & \mathrm{Cl}(H^\circ + T^\perp) \\ & = & (H \cap T)^\circ \\ & = & (G \cap T)^\circ \\ & \supset & G^* + T^\perp \\ & \supset & H^* + T^\perp, \end{array}$$

in which we made use of the fact that the closure of the algebraic polar of a spectrahedron is its polar. If $H^* + T^\perp$ is closed, then equality holds throughout, and in particular, $G^* + T^\perp$ is closed. Hence we may assume that $H^* + T^\perp$ is not closed. But since the dimension of the matrix map $P(x)$ is $n - l < n$, by induction there exists ($(n-l) \times (n-l)$ matrix) $V \succeq 0$ and $\tilde{\lambda} \in \Re^r$ such that

$$P^*(V) + A^T\tilde{\lambda} = 0, V_0 \bullet P_0 = 0,$$

and

$$\{x|Ax = 0, P(x)V = 0\} \subsetneq T.$$

To obtain the U, λ as required by the claim, we just set

$$U = \begin{bmatrix} V & 0 \\ 0 & I \end{bmatrix},$$

and $\lambda = \hat{\lambda} + \tilde{\lambda}$. The result follows when $\hat{U}$ has the special partial identity form. In the general situation, we simply pick a nonsingular matrix X such that

$$\hat{U} = X^T \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} X.$$

The proof above will apply after minor modifications. The proof of claim 2 is now complete. □.

6 Proof of Strong Duality for ELSD

With a complete description of the polar of a spectrahedron at hand, it is quite easy to derive the dual for SDP along the lines of Lemma 1. We will first prove (2) and (3) of Theorem 1, and then apply the weak version of (2) to prove (4), i.e. dual attainment.

First, let us consider the following simplification that allows us to assume that $0 \in G$. In the special case of $Q_0 = 0$, 0 is in G, and in the general case, since we may assume for the purposes of Theorem 1 that the primal is feasible, let $\overline{x} \in G$. Let us perform the simple translation of G that sends $\overline{x}$ to 0: $z = x - \overline{x}$, so that we get

$$G' = G - \overline{x} = \{z | Q'(z) \succeq 0\},$$

where

$$Q'(z) := (Q(\overline{x})) - \hat{Q}(z).$$

So the primal SDP (P) transforms to:

$$\begin{array}{lll} \sup: & c^T z + c^T \overline{x} & \\ \text{s.t.} & \sum_i z_i Q_i & \preceq \quad Q(\overline{x}). \end{array} \tag{P'}$$

Note that $\hat{Q}$ has not changed, and it is not difficult to show that the $\mathcal{U}_k$ remain unchanged, as the equations $Q^{\#}(X) = 0$ are equivalent to $Q^*(X) = 0, Q(\overline{x}) \bullet X = 0$. And therefore, the (ELSD) corresponding to (P') is given by

$$\begin{array}{llll} \inf: & (U + W) \bullet Q(\overline{x}) + c^T \overline{x} & & \\ \text{s.t.} & Q^*(U + W) & = & c \\ & W & \in & \mathcal{W}_m \\ & U & \succeq & 0. \end{array} \tag{ELSD'}$$

Since $Q^*(U + W) = c$

$$(U + W) \bullet Q(\overline{x}) + c^T \overline{x} = (U + W) \bullet Q_0 - \overline{x}^T Q^*(U + W) - c^T \overline{x} = (U + W) \bullet Q_0,$$

and thus there is an exact correspondence between (ELSD) and (ELSD') and hence we may WOLOG assume that 0 is feasible for (P). Note that this implies

$$\sup\{c^T x | x \in G\} \geq 0.$$

Let us now define the following weaker version of ELSD.

$$\begin{array}{llll} \inf: & (U + W) \bullet Q_0 & & \\ \text{s.t.} & Q^*(U + W) & = & c \\ & W & \in & \mathcal{W}_{m-1} \\ & U & \succeq & 0, \end{array} \tag{W-ELSD}$$

Proof of Theorem 1(2)-(4): We have the following for any $k \geq m - 1$.

$$\begin{array}{rcl} \sup\{c^T x | x \in G\} & = & \inf\{t > 0 | c^T x \leq t \ \forall \ x \in G\} \\ & = & \inf\{t > 0 | c^T x / t \leq 1 \ \forall \ x \in G\} \\ & = & \inf\{t > 0 | c/t \in G^\circ\} \\ & = & \inf\{t > 0 | Q^*(U' + W') = c/t, W' \in \mathcal{W}_k, U' \succeq 0, U' \bullet Q_0 \leq 1\} \\ & = & \inf\{t > 0 | Q^*(U + W) = c, W \in \mathcal{W}_k, U \succeq 0, U \bullet Q_0 \leq t\} \\ & = & \inf\{U \bullet Q_0 | Q^*(U + W) = c, W \in \mathcal{W}_k, U \succeq 0\} \\ & = & \inf\{(U + W) \bullet Q_0 | Q^*(U + W) = c, W \in \mathcal{W}_k, U \succeq 0\}. \end{array}$$

Explanation: The first equality is from the fact that the primal optimal value is nonnegative. The second is obvious, and the third follow from the definition of polar. The fourth equality is from Theorem 2, and the fifth is obtained by a change of variables $U = tU', W = tW'$, and (since $\mathcal{W}_k$ is a subspace, $W \in \mathcal{W}_k$). The sixth equality is quite obvious and the last one comes from the fact that $W \bullet Q_0 = 0$ whenever $W \in \mathcal{W}_k$ (of course, under the assumption that $0 \in G$).

Thus, we deduce that

- Whenever the primal (P) is bounded, the duals (ELSD) and (W-ELSD) are feasible.
- In such a case, the optimal values of (P), (ELSD) and (W-ELSD) are equal.

The parts 2 and 3 of Theorem 1 follow. Note that the duality gap is zero for the weak version of the dual (W-ELSD) itself. Let us now show that the dual (ELSD) attains its optimum value whenever the latter is finite (it doesn't from the proof below that (W-ELSD) attains its optimum, though a stringent argument will imply it). Let $0 \leq \alpha < \infty$ be the common optimal value of the two programs (P) and (ELSD). We distinguish between two cases.

Case 1: $\alpha > 0$. Clearly $c \neq 0$. Consider

$$0 < \alpha = \inf\{t > 0 | c^T x/t \in G^\circ\}$$

and hence

$$1/\alpha = \sup\{s \geq 0 | sc \in G^\circ\},$$

and since G° is a closed convex set and $c \neq 0$, the above supremum is attained, say for $s = s^*$. Then clearly,

$$\alpha = 1/s^* = \min\{t > 0 | c^T x/t \in G^\circ\},$$

and it easily follows that (ELSD) attains the optimal value α in this case.

Case 2: $\alpha = 0$.

This means that 0 is an optimal solution to the primal[1]. If $c = 0$, then $(U, W) = 0$ will be a dual feasible solution with an optimal value of 0. So assume that $c \neq 0$.

For the remaining case, we want to show that there exists a dual feasible solution (to (ELSD)) with optimum value of 0, i.e. we want to show that the following is feasible:

$$(U + W) \bullet Q_0 = 0, Q^*(U + W) = c, U \succeq 0, W \in \mathcal{W}_m. \qquad (\ddagger)$$

It is easy to that the above system is the set of feasible solutions of the (W-ELSD) of the following homogeneous semidefinite program[2] in $m + 1$ variables:

$$\begin{array}{lll} \sup: & c^T x & \\ \text{s.t.} & \sum_{i=1}^m x_i Q_i & \preceq x_0 Q_0. \end{array} \qquad (\dagger\dagger)$$

Since the above primal problem is feasible (with $x = 0, x_0 = 0$), the infeasibility of the (W-ELSD) system ($\ddagger$) implies that the ($\dagger\dagger$) is unbounded (this is precisely the motivation behind introducing the weak version of the dual. The improvement obtained in having $W \in \mathcal{W}_{m-1}$ is now absorbed by the above increase in the number of variables from m to $m + 1$) From the homogeneity of ($\dagger\dagger$), we conclude that there exists (x, x_0) such that

$$c^T x > 0, \sum_{i=1}^m x_i Q_i \preceq x_0 Q_0.$$

We will pick an $r \geq 0$ such that $r + x_0 > 0$, set

$$z = x/(r + x_0)$$

and claim that $Q(z) \succeq 0$:

$$\begin{array}{rcl} (r + x_0)Q(z) & = & (r + x_0)(Q_0 - \sum_{i=1}^m z_i Q_i) \\ & = & (r + x_0)Q_0 - \sum_{i=1}^m x_i Q_i \\ & = & rQ_0 + (x_0 Q_0 - \sum_{i=1}^m x_i Q_i) \\ & \succeq & 0. \end{array}$$

Therefore z is feasible for (P) and quite clearly $c^T z > 0$, which is a contradiction since $\alpha = 0$. Thus, ($\ddagger$) is feasible, and the dual (ELSD) attains its optimum value of 0.

This concludes the proof of our strong duality result. □

[1] Therefore, the primal attains its optimum, and it is quite surprising that showing dual attainment is harder in this case

[2] There is a cause for confusion here as the matrix map has changed slightly after homogenization. We are using $\mathcal{W}_m$ (and implicitly Q^* and $Q^\#$) as they are defined for the original $Q(x)$.

Part 3: APPLICATION OF THE DUAL

7 An Exact Farkas Lemma for SDP

By applying Theorem 1(2) (the weak version), we can now trivially derive an exact Farkas Lemma for semidefinite systems.

Let $Q(x) = Q_0 - \sum_{i=1}^m x_i Q_i$ be a given matrix map, and suppose we would like characterize the nonemptyness of $G = \{x | Q(x) \succeq 0\}$ (or the feasibility of $Q(x) \succeq 0$). Consider the following homogeneous SDP in $m+1$ variables:

$$\begin{array}{ll} \sup: & x_0 \\ \text{s.t.} & \sum_{i=1}^m x_i Q_i - x_0 Q_0 \preceq 0 \end{array}$$

It is quite clear that G is nonempty if and only if the above supremum is ∞, i.e. unbounded. Therefore, we can apply the Duality Theorem (Theorem 1(2); for the weaker version of ELSD) to conclude that G is empty iff there exist U, W such that

$$(U + W) \bullet Q_0 = -1, Q^*(U + W) = 0, W \in \mathcal{W}_m, U \succeq 0.$$

We now state this fact as a system of alternative.

Theorem 3 (Farkas Lemma for SDP) *Let $Q(x) = Q_0 - \sum_{i=1}^m x_i Q_i$ be a given affine matrix map. Then exactly one of the following semidefinite systems is consistent:*

1. $Q(x) \succeq 0$
2. $$\begin{array}{rcl} (U + W_m) \bullet Q_0 & = & -1 \\ Q^*(U + W_m) & = & 0 \\ Q^{\#}(U_i + W_{i-1}) & = & 0 \ \forall\ i = 1, .., m \\ W_0 & = & 0 \\ U & \succeq & 0 \\ \begin{bmatrix} I & W_i^T \\ W_i & U_i \end{bmatrix} & \succeq & 0 \ \forall\ i = 1, .., m \end{array}.$$

8 Optimality and Related Conditions

Given a semidefinite program (P) and a feasible vector $\overline{x}$, suppose we wish to test the optimality of this vector. The the strong duality theorem, in particular the dual attainment part, gives us an answer.

Theorem 4 (Optimality Condition) *A given feasible vector $\overline{x}$ is optimal to SDP (P) if and only there exist U, W such that*

$$(U + W) \bullet Q_0 = c^T \overline{x}, Q^*(U + W) = 0, W \in \mathcal{W}_m, U \succeq 0.$$

Thus the *Optimality Detection Problem* for SDP reduces to semidefinite feasibility. The proof is an obvious application of Theorem 1. On a related note, let us now characterize the condition that a feasible and bounded semidefinite program (P) attains its optimum value (as seen in Example 5 of §1.6, not all SDPs attain their optima).

Theorem 5 (Primal Attainment) *Suppose that (P) as well (ELSD) are feasible. Then the primal attains the optimal value iff the following system if feasible:*

$$Q(x) \succeq 0, c^T x = (U + W) \bullet Q_0, Q^*(U + W) = 0, W \in \mathcal{W}_m, U \succeq 0.$$

9 Complexity Results for SDP

In this section, we will address certain complexity issues concerning Semidefinite Programming. The two computational models of our interest are

- **The Turing Machine model**[GJ79] (abbreviated as *TM Model* here). We will mainly borrow their notation in Chapters 2 and 7.
- **The real number model** of Blum, shub and Smale[BSS89] (abbreviated as :*BSS Model*).

The primary decision problem of interest is:

Semidefinite Feasibility Problem(SDFP):
INSTANCE: Positive integers m and n, and $n \times n$ symmetric matrices $Q_0, \ldots, Q_m$ with integer entries
QUESTION: Are there real numbers such that $x_1, \ldots, x_m$ such that $Q(x) = Q_0 - \sum_{i=1}^{m} x_i Q_i$ is positive semidefinite?

Sizes of instances: In the TM model, the size of an instance of SDFP is $m+n$ plus the total number of bits required to write down the matrices. In the BSS model, the size if $m+n$ (bitlengths are not allowed as a part of the size here).

At present, it is not known whether SDFP is in NP (and of course, it is open whether SDFP is P) for the TM Model. In the BSS model, it follows from some previous results that SDFP is in NP, and it is shown here that it is also in Co-NP.

A language[GJ79] (or a decision problem) is defined to be **Semidefinite Reducible**, if there exists a polynomial reduction from that language to SDFP.

For any language A, Co-A denotes the complementary language associated with A (for instance, Co-SDFP is the set of all instances of SDFP for which the answer to SDFP is "No").

Our main result of the section is the following.

Theorem 6 *The following hold concerning SDFP:*

1. *If SDFP$\in$ NP, then SDFP$\in$Co-NP and vice-versa.*
2. *In TM model, unless NP=Co-NP, SDFP is neither NP-Complete nor Co-NP-Complete.*
3. *In BSS model, SDFP is in NP$\cap$Co-NP.*

Proof: The result is essentially an application of Farkas Lemma (Theorem 3).

First note that, the system of alternative given for $Q(x) \succeq 0$, written here in short form:

$$(U+W) \bullet Q_0 = -1, Q^*(U+W) = 0, W \in \mathcal{W}_m, U \succeq 0,$$

can itself be cast in the form "$Q(x) \succeq 0$": Specifically, we can replace linear constraints $Ax = b$

$$\mathrm{Diag}(Ax - b) \oplus \mathrm{Diag}(b - Ax) \succeq 0,$$

and one can combine two semidefinite inequalities $P(x) \succeq 0, Q(x) \succeq 0$ in to one as:

$$P(x) \oplus Q(x) \succeq 0.$$

These two "tricks" are sufficient to show that the system of alternative 2 of the Farkas Lemma (Theorem 3) can be written as a semidefinite inequality whose size is *polynomial* in the size of the input instance of the SDFP. So, given any instance of SDFP, we write the above alternative system and deduce from the Farkas Lemma that the former has a "No" answer iff the latter has a "Yes" answer. This directly implies 1.

To prove 2, suppose that SDFP is NP-Complete (in the TM model). Then, in particular, it is in NP, and then 1 implies that SDFP is in NP$\cap$Co-NP. By Theorem 7.2 of [GJ79], the existence of an NP-Complete problem in NP=Co-NP implies that NP=Co-NP.

The proof of 3 will follow if we show that SDFP is in NP for the BSS model. If we can check in time polynomial (in m, n), given a "guess vector" x, whether $Q(x) \succeq 0$, then it follows that SDFP is in NP. We can perform this task in $O(\max\{n^3, mn^2\})$ arithmetic computations as follows: first compute the combination matrix $Q(x)$ in $O(mn^2$ steps and then apply partial Cholesky decomposition (see [GLS88] or [Ram93]) to check if this matrix is positive semidefinite in $O(n^3)$ steps. □

The reason one cannot extend the above proof that SDFP∈ NP∩Co-NP for the BSS model to the TM model is because feasible semidefinite inequalities need not always have rational solutions of polynomial size (see examples in §1.6). However, part 2 of the theorem is quite interesting as it results in: either SDFP∈ NP∩ Co-NP or SDFP∉ NP∪ Co-NP, which may be interpreted as, "either SDFP is too easy or it is too hard".

By an application of the results of last section, we can conclude that there exist polynomial time reductions (in both TM and BSS models) from the following problems to SDFP:

1. **The Primal Boundedness Problem:** Checking whether a feasible SDP is bounded; By the strong duality theorem, primal boundedness is equivalent to the feasibility of the ELSD.

2. **The Primal attainment Problem:** Verifying whether a feasible and bounded SDP attains its optimum value; This follows from Theorem 5 of last section

3. **The SDP Optimality Problem:** Given an SDP and a feasible, determine whether this solution is optimal; The reduction is evident from Theorem 4.

It is also clear that each of these problems naturally inherit all the all the properties of SDFP given by Theorem 6.

10 Concluding Remarks

While ELSD has polynomially many variables ($O(mn^2)$), from a practical computation point of view, this is too large. Hence it would be nice to find refinements of the dual that have a lot fewer variables. It should however be mentioned that, despite the large number of varaibles, the dual has a great deal of structure, which might lend itself to efficient matrix algebraic manipulations, and reduce the effective algebraic complexity of algorithms involving ELSD.

Perhaps the most outstanding open problem in the theory of Semidefinite Programming is whether SDFP (see §9) is polynomial time solvable. While Theorem 6(2) shows that SDFP is not NP-Complete (unless NP=Co-NP), the examples of §1.6 are reasonably strong reason to suspect that SDFP is not NP to start with; one has to verify feasibility without actually computing a feasible solution. But one can not rule out the possibility that there may be a representation theory for solutions of semidefinite systems, which lends itself to a polynomial time treatment (see the remarks at the end of §1.6).

Also, there is good enough reason to suspect that Graph/Matrix Isomorphism is *Semidefinite Reducible*(§9), i.e. one can reduce it to SDFP. This comes from certain results in [RSU94]. Let A, B be the adjacency matrices of two graphs. Then these graphs are isomorphic iff there exists a permutation matrix P such that $A = PBP^T$, i.e. the following system is feasible

$$A = PBP^T, Pe = e, P^Te = e, P_{ij} \in \{0, 1\}, \tag{GI1}$$

where e denotes the vector of all ones. In [RSU94], it was shown that one may relax the last integrality condition, i.e. let P be doubly stochastic instead, without losing exactness, i.e. GI1 is feasible iff the following is feasible

$$A = PBP^T, Pe = e, P^Te = e, 0 \leq P \leq J, \tag{GI2}$$

where J is the matrix of all ones. It is probably not far-fetched to conjecture that some semidefinite relaxation of GI1 or GI2 is exact. Another seemingly good candidate for reduction to SDFP is the recognition problem for perfect graphs.

Acknowledgements I thank Farid Alizadeh, Adi Ben-Israel and Lóránt Porkoláb for many exciting discussions, and Bob Freund for supplying me an example of an SDP for which the Lagrange-Slater dual(LSD) has a finite nonzero duality gap.

References

[Ali91] F. Alizadeh, *Combinatorial Optimization with Interior Point Methods and Semi-Definite Matrices*, Ph.D. Thesis, Computer Science Department, University of Minnesota, Minneapolis, Minnesota, 1991.

[Ali94] F. Alizadeh, Interior Point Methods in Semidefinite Programming with Applications to Combinatorial Optimization, *SIAM J. Opt.*, Vol. 5(1), 1995.

[BW81] J. Borwein, H. Wolkowicz, Regularizing the Abstract Convex Program, *J. Math. Anal. Appl.*, 83(1981).

[BSS89] L. Blum, M. Shub and S. Smale, On a Theory of Computation and Complexity over the Real Numbers: NP-Completeness, Recursive Functions and Universal Machines, *Bull. (New Series) of the AMS*, Vol. 21(1), pp 1-46, 1989.

[Bro83] A. Brøndsted, *An Introduction to Convex Polytopes*, Springer-Verlag, New York, 1983.

[Chv83] V. Chvátal, *Linear Programming*, W.H. Freeman and Co., New York, 1983.

[Fre95] R. Freund, Complexity of an Algorithm for Finding an Approximate Solution of a Semi-Definite Program, with no Regularity Condition, *manusript*, 1995.

[GJ79] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, New York, 1979.

[GW94] M.X. Goemans and D.P. Williamson, Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming, *Submitted to J. ACM.* (contact goemans@math.mit.edu for copies)

[GLS84] M. Grötschel, L. Lovász and A. Schrijver, Polynomial Algorithms for Perfect Graphs, *Annals of Discrete Mathematics* 21, C. Berge and V. Chvátal, eds., North Holland, 1984.

[GLS88] M. Grötschel, L. Lovász and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, Berlin, 1988.

[Hol75] R.B. Holmes, *Geometric Functional Analysis and its Applications*, Springer-Verlag, New York, 1975.

[HJ85] R. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[JPY88] D.S. Johnson, C.H. Papadimitriou and M. Yannakakis, How Easy is Local Search?, *Jnl. of Comp. Sys. Sci.*, Vol. 37, pp. 79-100, 1988.

[Koi94] P. Koiran, A Weak Version of the Blum, Shub & Smale Model, *DIMACS Technical Report* # 94-10, 1994. (URL: http://dimacs/TechnicalReports/1994.html)

[NN94] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Methods for Convex Programming: Theory and applications*, SIAM, 1994.

[Ram93] M.V. Ramana, *An Algorithmic Analysis of Multiquadratic and Semidefinite Programming Problems*, Ph.D. Thesis, The Johns Hopkins University, Baltimore, 1993.

[RG94a] M.V. Ramana and A.J. Goldman, Some Geometric Results in Semidefinite Programming, *To appear in Jnl. Glob. Opt.*, 1995.

[RSU94] M.V. Ramana, E.R. Scheinerman and D. Ullman, Fractional Isomorphism of Graphs, *Disc. Math.* 132, pp 247-265, 1994.

[Roc70] T.R. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, 1970.

[VB94] L. Vandenberghe, S. Boyd, Positive-Definite Programming, *Mathematical Programming: State of the Art 1994*, J.R. Birge and K.G. Murty ed.s, U. of Michigan, 1994.