STRONG DUALITY FOR SEMIDEFINITE PROGRAMMING *

Motakuri Ramana [†]   Levent Tunçel and Henry Wolkowicz [‡]

University of Waterloo
Department of Combinatorics and Optimization
Faculty of Mathematics
Waterloo, Ontario N2L 3G1, Canada

June 1, 1995

Abstract

It is well known that the duality theory for linear programming (LP) is powerful and elegant and lies behind algorithms such as simplex and interior-point methods. However, the standard Lagrangian for nonlinear programs requires constraint qualifications to avoid duality gaps.

Semidefinite linear programming (SDP) is a generalization of LP where the nonnegativity constraints are replaced by a semidefiniteness constraint on the matrix variables. There are many applications, e.g. in systems and control theory and in combinatorial optimization. However, the Lagrangian dual for SDP can have a duality gap.

We discuss the relationships among various duals and give a unified treatment for strong duality in semidefinite programming. These duals guarantee strong duality, i.e. a zero duality gap and dual attainment. This paper is motivated by the recent paper by Ramana where one of these duals is introduced.

Key words: Semidefinite linear programming, strong duality, Löwner partial order, symmetric positive semidefinite matrices.

AMS 1991 Subject Classification: Primary 65K10, Secondary 90C25, 90M45, 15A45, 47D20.

Contents

1 INTRODUCTION

1.1 Semidefinite Programming

We study the semidefinite linear programming problem

$$
\begin{array}{rl}
p^* = & \sup \quad c^t x \\
(\mathbf{P}) & \text{subject to} \quad Ax \preceq b \\
& x \in \Re^m,
\end{array}
$$

where: $c, x \in \Re^m$; $b = Q_0 \in \mathcal{S}_n$, the *space of symmetric $n \times n$ matrices*; the linear operator $Ax = \sum_{i=1}^m x_i Q_i$, for $Q_i \in \mathcal{S}_n$, $i = 1, \ldots, m$; and $\preceq$ denotes the Löwner partial order, i.e. $X \preceq (\prec) Y$ means $Y - X$ is positive semidefinite (positive definite). We let $\mathcal{P}$ denote the cone of semidefinite matrices. By a *cone* we mean a convex cone, i.e. a set K satisfying $K + K \subset K$, and $\lambda K \subset K$, $\forall \lambda \geq 0$. We consider the space of symmetric matrices, $\mathcal{S}_n$, as a vector space with the *trace inner product* $\langle U, X \rangle = \operatorname{trace} UX$. The corresponding norm is the *Frobenius matrix norm* $\|X\| = \sqrt{\operatorname{trace} X^2}$.

We let F denote the feasible set of P and we assume that the optimal value p is finite; so that the feasible set $F \neq \emptyset$.

1.2 Background

1.2.1 Cone of Semidefinite Matrices

The cone of semidefinite matrices has been studied extensively both for its importance and elegance. Positive definite matrices arise naturally in many areas, including differential equations, statistics, and systems and control theory. The cone $\mathcal{P}$ induces a partial order, called the Löwner partial order. Various monotonicity results were studied with respect to this partial order [29, 30]. An early paper is that by Bohnenblust [9]. Optimization problems over cones of matrices are also discussed in the monograph by Berman [8].

More recently, there has been renewed interest in semidefinite programming mostly due to applications in engineering (see, for instance, Ben-Tal and Nemirovskii [7], Boyd et al. [14], Vandenberghe and Boyd [36]) and combinatorial optimization (see, for instance, Alizadeh [1], Goemans and Williamson [19], Lovász and Schrijver [28], Nesterov and Nemirovskii [31], Poljak [15], Helmberg et al. [22]). Nesterov and Nemirovskii's book provides a unifying framework for polynomial-time interior-point algorithms in convex programming. Up to this date, interior-point algorithms seem to be the best algorithms (from both theoretical and practical viewpoints) for

solving semidefinite programming problems. Freund presents an infeasible-interior-point algorithm [17]. Complexity of the algorithm depends on the distances (in a norm induced by the initial solution) of the initial solution to the sets of approximately feasible and approximately optimal solutions, where approximate feasibility and optimality are defined in terms of tolerances which are given. The algorithm does not assume that the zero duality gap (or even feasibility) is attainable. Indeed, for the case when the given problem exhibits a finite nonzero duality gap, we can ask for a tolerance in the duality gap that is not attainable (for such a tolerance, the distance from the set of approximately optimal solutions would be infinite for any starting point). Our goal here is to study and unify the ways in which a dual problem can be modified to ensure a zero duality gap at optimality.

Other related issues arise from the study of correlation matrices in statistics, e.g. Pukelsheim [32], and matrix completion problems, see [20, 5, 24]. Results for multiquadratic programs is studied in [33].

1.2.2 Early Duality

Extensions of finite linear programming duality to infinite dimensions and/or to optimization problems over cones has been studied in the literature. We do not give a comprehensive survey, since that would probably be impossible. But we mention several early results.

In [16], Duffin studies infinite linear programs, i.e. programs for which there are an infinite number of constraints and/or an infinite number of variables. Also studied is the notion of optimization with respect to a partial order induced by a cone. Infinite linear programming is closely related to the notion of continuous programming, e.g. [26, 27, 35]. A major question is the formulation of duals that close the duality gap. Infinite dimensional linear programming is also studied in the books by Glashoff and Gustafson [18] and Anderson and Nash [2].

More recently, duals that guarantee strong duality for general abstract convex programs have been given in [13, 12, 11, 10]. The special case of a linear program with cone constraints is treated in [38].

1.3 Outline

This note is motivated by the recent paper of Ramana [34]. Therein a dual program, called an extended Lagrange-Slater dual program and denoted ELSD, is presented for which strong duality holds and more importantly it

can be written down in polynomial time. Previous work on general (convex) cone constrained programs, see [13, 38, 11, 10], presented dual programs for which strong duality holds and was based on regularization and on finding the so-called minimal cone of $\mathcal{P}$. We denote this dual by DRP. A procedure for defining the minimal cone was presented in [11]. This procedure started with an initial feasible point and reduced the program, in a finite number of steps, to a regularized program.

The main result in this note is, first, to show that the extended Lagrange dual program ELSD is equivalent to the regularized dual DRP. This equivalence is in the sense that the constraints and the set of Lagrange multipliers are the same. The difference in the duals is in the fact that the feasible set of Lagrange multipliers, denoted $(\mathcal{P}^f)^+$, is expressed implicitly in ELSD as the solution of m systems of constraints included in the dual, whereas it is defined explicitly in DRP as the output of the separate procedure mentioned above. This separate procedure finds the minimal cone by solving a system of constraints equivalent to that in ELSD. Also presented, is an extended dual of the dual, i.e. this closes the duality gap from the dual side.

The fact that the two duals ELSD and DRP are found using different techniques and then result in being equivalent is more than a coincidence. In fact, we show that these duals are unique in a certain sense.

In Section 2 we discuss the geometry of the cone of semidefinite matrices. In particular, we present old and new results on the faces of this cone. Lemmas 2.1 and 2.2 provide a description of the faces and characterization of the cases in which the sum of the positive semidefinite cone and a subspace is closed. The two strong duality schemes are outlined in Section 3. The relationships between the duals is presented in Section 4. We include the results on the dual of the dual. In Section 5, we present a homogenized program which is equivalent to SDP and provides a different view of optimality conditions. We conclude with some remarks on perturbations of SDP and computational complexity issues.

1.4   Notation

relint   the relative interior

∂ the boundary

$\mathcal{M}_{k,l}$ the space of $k \times l$ matrices

$\mathcal{M}_n$ the space of $n \times n$ matrices

$\mathcal{S}_n$ the space of symmetric $n \times n$ matrices

$\mathcal{P}_n$ **or** $\mathcal{P}$ the cone of positive semidefinite matrices in $\mathcal{S}_n$

$Q \preceq R$ $R - Q$ is positive semidefinite

$Q \preceq_{\mathcal{P}} R$ $R - Q$ is positive semidefinite

$Q \preceq_K R$ $R - Q \in K$, where K is a closed convex cone

$\mathcal{N}(A)$ the null space of the linear operator A

$\mathcal{R}(A)$ the range space of the linear operator A

A^* the adjoint of the linear operator A, (3.1)

$A^\dagger$ the Moore-Penrose generalized inverse of the linear operator A

$K \lhd T$ K is a face of T, (2.1)

K^c the complementary (or conjugate) face of K, (2.2)

K^+ the polar cone of K, $K^+ = \{\phi : \langle \phi, k \rangle \geq 0, \; \forall k \in K\}$

$K^\perp$ the orthogonal complement of K, $K^\perp = \{\phi : \langle \phi, k \rangle = 0, \; \forall k \in K\}$

$\mathcal{F}(C)$ the minimal face of $\mathcal{P}$ containing the subset $C \subset \mathcal{P}$

$\mathcal{P}^f$ the minimal cone of P, i.e. $\mathcal{F}(b - A(F))$, (2.3)

S_D^f the minimal cone in S, i.e. $\mathcal{F}(G^*(F_D) - \begin{pmatrix} c \\ 0 \end{pmatrix})$

SDP   a semidefinite program

LP   a linear program

P   the primal SDP

RP   the regularization of the primal SDP

p^* the optimal value of the primal SDP

F   the feasible set of P

$L(x, U)$ the Lagrangian of P, $L(x, U) = c^t x + \operatorname{trace} U(b - Ax)$

4

CQ	constraint qualification
D	the dual SDP
ELSD	the extended Lagrange-Slater dual SDP to P
ELSDD	the extended Lagrange-Slater dual SDP to D
ED	an equivalent program to the dual SDP
RD	the regularization of the dual SDP
d^*	the optimal value of the dual SDP
F_D	the feasible set of D
$\mathcal{P}_D^f$	the minimal cone of D, i.e. $\mathcal{F}(F_D)$
$\mathcal{C}_k$	$\left\{ (U_i, W_i)_{i=1}^k : A^*(U_i + W_{i\perp 1}) = 0, \ \text{trace}\, b(U_i + W_{i\perp 1}) = 0, \right.$ $\left. U_i \succeq W_i W_i^t, \ \forall i = 1, \ldots, k, W_0 = 0 \right\}$
$\mathcal{U}_k$	$\{ U_k : (U_i, W_i)_{i=1}^k \in \mathcal{C}_k \}$
$\mathcal{W}_k$	$\{ W_k : (U_i, W_i)_{i=1}^k \in \mathcal{C}_k \}$
$\mathcal{W}_k^S$	$\{ W + W^t : W \in \mathcal{W}_k \}$
$\mathcal{W}$	$\mathcal{W}_m^S$

2 GEOMETRY of the PSD CONE

We now outline several known and some new results on the geometry of the cone $\mathcal{P}$. More details can be found in e.g. [3, 4].

The cone $K \subset T$ is a *face* of the cone T, denoted $K \lhd T$, if

$$x, y \in T, \ x + y \in K \Rightarrow x, y \in K. \qquad (2.1)$$

The faces of $\mathcal{P}$ have a very special structure. Each face, $K \lhd \mathcal{P}$, is characterized by a unique subspace, $S \subset \Re^n$:

$$K = \{ X \in \mathcal{P} : \mathcal{N}(X) \supset S \}.$$

Moreover,

$$\text{relint}\, K = \{ X \in \mathcal{P} : \mathcal{N}(X) = S \}.$$

5

The *complementary* (or conjugate) face of K is $K^c = K^\perp \cap \mathcal{P}$ and

$$K^c = \{X \in \mathcal{P} : \mathcal{N}(X) \supset S^\perp\}. \tag{2.2}$$

Moreover,

$$\text{relint } K^c = \{X \in \mathcal{P} : \mathcal{N}(X) = S^\perp\}.$$

Note: Each face K (respectively, K^c) is *exposed*, i.e. it is equal to the intersection of $\mathcal{P}$ with a supporting hyperplane; the supporting hyperplane corresponds to any $X \in \text{relint } K^c$ (respectively, relint K). Also, complementary faces are orthogonal and satisfy $XY = 0, \forall X \in K, Y \in K^c$.

Another property of the cone $\mathcal{P}$, see [11], is that it is *projectionally exposed*, i.e. every face of $\mathcal{P}$ is the image of $\mathcal{P}$ under some projection. In fact, if $Q \in \mathcal{S}_n$ is the projection onto the subspace S, the null space of matrices in relint K, then the face K satisfies

$$K = (I - Q)\mathcal{P}(I - Q).$$

The *minimal cone* of P is defined as

$$\mathcal{P}^f = \cap\{K \lhd \mathcal{P} : K \supset (b - A(F))\}, \tag{2.3}$$

i.e. the minimal cone is the intersection of all faces of $\mathcal{P}$ containing the feasible slacks.

The following lemma describes orthogonal complements to faces in terms of the data. The connection is in terms of a semidefinite completion problem. (For more on completion problems see e.g. [20].) The lemma shows that we can express the orthogonal complement of a face in terms of a system of semidefinite inequalities.

Lemma 2.1 *Suppose that C is a convex cone and $C \subset \mathcal{P}$. Let*

$$K := \{W + W^t : U \succeq WW^t, \text{ for some } U \in C\}.$$

Then

$$
\begin{aligned}
((\mathcal{F}(C))^c)^\perp \;&=\; K \\
&=\; \left\{ W + W^t : \begin{bmatrix} I & W^t \\ W & U \end{bmatrix} \succeq 0, \text{ for some } U \in C \right\}. \tag{2.4}
\end{aligned}
$$

6

Proof. Suppose that $W + W^t \in K$, i.e. $U \succeq WW^t$, for some $U \in C$. Since $x^t(U - WW^t)x \geq 0 \ \forall x$, we get $\mathcal{N}(U) \subset \mathcal{N}(W^t)$. Equivalently, $\mathcal{R}(U) \supset \mathcal{R}(W)$. Since $UU^\dagger$ is the projection (orthogonal) onto the range of U, where $U^\dagger$ denotes the Moore-Penrose generalized inverse of U, we conclude that $W = UU^\dagger W$. We have shown that

$$U \succeq WW^t \Rightarrow W = UH, \text{ for some H.} \tag{2.5}$$

Therefore, $\operatorname{trace} WV = 0$, for all $V \in (\mathcal{F}(C))^c$, i.e. $W + W^t \in ((\mathcal{F}(C))^c)^\perp$.

To prove the converse inclusion, suppose that $V \in ((\mathcal{F}(C))^c)^\perp$ and $U \in C \cap \operatorname{relint} \mathcal{F}(C)$. Let U be orthogonally diagonalized by $Q = [Q_1 Q_2]$

$$U = Q^t \operatorname{Diag}(d_1 \ 0)Q, \ Q^t Q = I,$$

with $Q_1, n \times r, \ d_1 > 0$. Therefore,

$$\mathcal{F}(C) = \{Q_1 B Q_1^t : B \succeq 0, \ B \in \mathcal{S}_r\}$$

and

$$(\mathcal{F}(C))^c = \{Q_2 B Q_2^t : B \succeq 0, \ B \in \mathcal{S}_{n \perp r}\}.$$

Now $V \in ((\mathcal{F}(C))^c)^\perp$ implies that

$$0 = \operatorname{trace} VQ_2 BQ_2^t = \operatorname{trace} Q_2^t V Q_2 B, \ \forall B \succeq 0,$$

i.e.

$$Q_2^t V Q_2 = 0.$$

This implies that $Q_2 Q_2^t V Q_2 Q_2^t = 0$ as well. Note that $Q_2 Q_2^t$ is the orthogonal projection onto $\mathcal{N}(U)$. Therefore the nonzero eigenvalues of V correspond to the eigenvectors in the eigenspace formed from the column space of Q_1. Since the same must be true for VV^t, this implies that $\alpha U \succeq VV^t$, for some $\alpha > 0$ large enough, i.e. $V \in K$.

The alternative expression for K in (2.4) follows from the Schur complement.

$\square$

Now, note the following interesting and surprising closure property of the faces of $\mathcal{P}$. This is surprising because it is not true in general that the sum of a cone and a subspace is closed.

Lemma 2.2 *Suppose that the proper face K satisfies*

$$\{0\} \neq K \triangleleft \mathcal{P}, \ K \neq \mathcal{P}.$$

7

Then

$$\mathcal{P} + K^{\perp} = \overline{\mathcal{P} + \operatorname{span} K^c}; \tag{2.6}$$

$$\mathcal{P} + \operatorname{span} K \quad \textit{is not closed.} \tag{2.7}$$

Proof. Since $\operatorname{span} K^c \subset K^{\perp}$, we get

$$\mathcal{P} + K^{\perp} \supset \mathcal{P} + \operatorname{span} K^c.$$

From the characterization of faces in [3, 4], there exists a subspace $S \subset \Re^n$, with dimension t, such that

$$K = \{X \succeq 0 : \mathcal{N}(X) \supset S\}.$$

After applying an orthogonal transformation to $\Re^n$, we can assume that S is the span of the first t unit vectors. Therefore, $X \in K$ has a $t \times t$ 0 block, i.e.

$$X = \left[\begin{array}{cc} 0_t & 0 \\ 0 & \bar{X} \end{array} \right].$$

Moreover, for X in the relative interior of K, we have $\bar{X} \succ 0$. This implies that

$$K^{\perp} = \left\{ Y : Y = \left[\begin{array}{cc} C & D \\ D^t & 0 \end{array} \right], \ C \in \mathcal{S}_t, \ D \in \mathcal{M}_{t, n \perp t} \right\}.$$

Now suppose that we are given $T^n \in K^{\perp}$, $P^n \in \mathcal{P}$, $n = 1, 2, \ldots$ and the sequence

$$T^n + P^n \to L = \left[\begin{array}{cc} L_1 & L_2 \\ L_2^t & L_3 \end{array} \right].$$

Comparing the corresponding bottom right blocks, we see that necessarily $L_3 \succeq 0$. Therefore

$$L = \left[\begin{array}{cc} L_1 & L_2 \\ L_2^t & 0 \end{array} \right] + \left[\begin{array}{cc} 0 & 0 \\ 0 & L_3 \end{array} \right],$$

i.e. $L \in K^{\perp} + \mathcal{P}$. This proves $\mathcal{P} + K^{\perp}$ is closed, i.e.

$$\mathcal{P} + K^{\perp} \supset \overline{\mathcal{P} + \operatorname{span} K^c}.$$

To prove the converse inclusion, suppose that

$$W \in (\mathcal{P} + K^{\perp}) \setminus (\overline{\mathcal{P} + \operatorname{span} K^c}).$$

8

Then there exists a separating hyperplane, i.e. there exists Φ such that

$$\text{trace } \Phi W < 0 \leq \text{trace } \Phi(P + w), \ \forall P \in \mathcal{P}, \ w \in \text{span } K^c. \qquad (2.8)$$

This implies that $\Phi \succeq 0$ and $\Phi \in (K^c)^\perp$. But then $\text{trace } \Phi W = \text{trace } \Phi P + \text{trace } \Phi w$, with $P \in \mathcal{P}, w \in K^\perp$. From Lemma 2.1 and (2.5) we get that $w = UH + H^t U$, for some $U \in K^c$ and so $\text{trace } \Phi w = 0$. This implies that $\text{trace } \Phi W = \text{trace } \Phi P \geq 0$, which contradicts (2.8). This completes the proof of (2.6).

Now suppose that $X \in \text{relint } K$ and $X = QDQ^t$, $Q = [Q_1 \ Q_2]$, $QQ^t = I$, is an orthogonal diagonalization of X with the columns of Q_1 spanning $\mathcal{N}(X)$ and the columns of Q_2 spanning $\mathcal{R}(X)$. Then $K = \{Q_2 B Q_2^t : B \succeq 0\}$, and $\text{span } K = \{Q_2 B Q_2^t : B \in \mathcal{S}\}$. Now let $B \succ 0$ and $n = 1, 2, \ldots$. We see that

$$[Q_1 \ Q_2] \begin{bmatrix} \frac{1}{n} B^{\perp 1} & I \\ I & nB \end{bmatrix} \begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix} \in \mathcal{P}$$

while

$$[Q_1 \ Q_2] \begin{bmatrix} 0 & 0 \\ 0 & -nB \end{bmatrix} \begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix} \in \text{span } K.$$

However, the limit of the sum of the two sequences is

$$[Q_1 \ Q_2] \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix}$$

which is not in the sum $(\mathcal{P} + \text{span } K)$.

$\square$

Corollary 2.1

$$(\mathcal{P}^f)^+ = \mathcal{P}^+ + (\mathcal{P}^f)^\perp = \mathcal{P} + (\mathcal{P}^f)^\perp = \overline{\mathcal{P} + \text{span} \, (\mathcal{P}^f)^c}.$$

Proof. From the definition of a face and the closure condition above, we get

$$\begin{aligned} (\mathcal{P}^f)^+ &= (\mathcal{P} \cap \mathcal{P}^f)^+ \\ &= (\mathcal{P} \cap \text{span} \, (\mathcal{P}^f))^+ \\ &= \mathcal{P}^+ + (\mathcal{P}^f)^\perp. \end{aligned}$$

$\square$

3 DUALITY SCHEMES

3.1 Lagrangian Duality

The Lagrangian for P is

$$L(x, U) = c^t x + \operatorname{trace} U(b - Ax).$$

Consider the max-min problem

$$p^* = \max_x \min_{U \succeq 0} L(x, U).$$

The inner minimization problem has the hidden constraint $Ax \preceq b$, i.e. the minimization problem is unbounded otherwise. Once this hidden constraint is added to the outer maximization problem, the minimization problem has optimum $U = 0$. Therefore we see that this max-min problem is equivalent to the primal P. This illustrates that we have the correct constraint on the dual variable U.

   The Lagrangian dual to P is obtained by reversing the max-min to a min-max and rewriting the Lagrangian, i.e.

$$p^* \leq d^* = \min_{U \succeq 0} \max_x L(x, U) = \operatorname{trace} bU + x^t(c - A^*U).$$

Here A^* denotes the *adjoint* of the linear operator A, i.e.

$$(A^*U)_i = \operatorname{trace} Q_i U. \qquad (3.1)$$

The inner maximization now has the hidden constraint $c - A^*U = 0$. Once this hidden constraint is added to the outer minimization problem, the inner maximization has optimum $x = 0$. Therefore we see that this min-max problem is equivalent to the following dual program.

$$
\begin{aligned}
d^* = \quad &\min \quad &&\operatorname{trace} bU \\
(\mathbf{D}) \quad &\text{subject to} \quad &&A^*U = c \\
& &&U \succeq 0.
\end{aligned}
$$

3.1.1 Linear Programming Special Case

We note that the SDP pair P and D look exactly like linear programming (LP) duals but with $\geq$ replaced by $\succeq$. In fact, if the adjoint operator A^* includes constraints that force U to be diagonal, then we see that LP is a special case of SDP.

Now suppose that we consider P and D as LPs, i.e. suppose that we replace $\succeq$ with $\geq$. Then the operator A is an $n \times m$ matrix, and $U \in \Re^n$. In this special case, we always have strong duality, i.e. $p^* = d^*$ and d^* is attained. Moreover, we can have more than one dual of P. Let $P^=$ denote the set of indices of the rows of A corresponding to the implicit equality constraints, i.e.

$$P^= := \{i : x \in F \text{ implies } A_{i:}x = b_i\}.$$

Then we can consider the equality constraints $A_{i:}x = b_i$, for any subset of $P^=$, without changing P. This is equivalent to allowing the dual variables U_i, $i \in P^=$ to be free rather than nonnegative. Thus we see that we can have different duals for P while maintaining strong duality. In fact, there are an infinite number of duals, since the space of dual variables can be any set which includes the nonnegative orthant and restricts $U_i \geq 0$, $i \notin P^=$.

It is clearly better to have a smaller set of dual variables. In fact, in the case of LP discussed above, if some of the inactive constraints at the optimum can be identified, then we can restrict the corresponding dual variables to be 0. This is equivalent to ignoring the inactive constraints. Of course, we do not, in general, know which constraints will be active at the optimum.

Having more than one dual program occurs because there is no strictly feasible solution for P. We see below that a similar phenomenon occurs for P in the SDP case, but with the additional complication of possible loss of strong duality. In addition, the semidefinite constraint is not as simple as the nonnegativity constraint in LP. The question arises whether or not we get the same dual if we treat the semidefinite constraint $U \succeq 0$ as a functional constraint using the minimal eigenvalue of U.

3.2 Strong Duality and Regularization

If a *constraint qualification*, denoted CQ, see Section 5, holds for P, then we have strong duality for the Lagrange dual program, i.e. $p^* = d^*$ and d^* is attained. The usual CQ is *Slater's condition,* there exists $\hat{x}$ such that $(b - A\hat{x}) \in \text{int}\,\mathcal{P}$. Examples where $p^* < d^*$, and/or one of d^*, p^* is not attained, have appeared in the literature, see e.g. [17]. One can close the duality gap by using the minimal cone of $\mathcal{P}$. Therefore, an equivalent program is the *regularized P program*, see [11, 38],

$$
\begin{array}{rl}
p^* = & \max \quad c^t x \\
(\mathbf{RP}) & \text{subject to} \quad Ax \preceq_{\mathcal{P}^f} b \\
& \qquad\qquad\quad x \in \Re^m.
\end{array}
$$

Moreover, by the definition of faces, there exists $\hat{x}$ such that $(b - A\hat{x}) \in$ relint $\mathcal{P}^f$. Therefore, the generalized Slater's constraint qualification holds, i.e. strong duality holds for this program. (This is proved in detail in [11, 38].) Thus, the following is a dual program for P for which strong duality holds

$$
\textbf{(DRP)} \quad \begin{array}{rcl} p^* = & \min & \operatorname{trace} bU \\ & \text{subject to} & A^*U = c \\ & & U \succeq_{(\mathcal{P}^f)^+} 0, \end{array}
$$

where the polar cone

$$
(\mathcal{P}^f)^+ := \{U : \operatorname{trace} UP \geq 0, \ \forall P \in \mathcal{P}^f\}.
$$

One can also close the duality gap from the dual side. Let F_D denote the feasible set of D. The *minimal cone* of D is defined as

$$
\mathcal{P}_D^f = \cap\{K : K \lhd P, K \supset F_D\}. \tag{3.2}
$$

Therefore, an equivalent program is the *regularized D program*,

$$
\textbf{(RD)} \quad \begin{array}{rcl} d^* = & \min & \operatorname{trace} bU \\ & \text{subject to} & A^*U = c \\ & & U \succeq_{\mathcal{P}_D^f} 0. \end{array}
$$

Strong duality holds for this program. We therefore get the following strong dual of the dual.

$$
\textbf{(DRD)} \quad \begin{array}{rcl} d^* = & \max & c^t x \\ & \text{subject to} & Ax \preceq_{(\mathcal{P}_D^f)^+} b \\ & & x \in \Re^m. \end{array}
$$

The above presents two pairs of symmetric dual programs: RP and DRP; and RD and DRD. These dual pairs have all the nice properties of dual pairs in ordinary linear programming, i.e. [38, Theorem 4.1] yields the following for our dual pairs. This extends the duality results over polyhedral cones presented in [6].

Theorem 3.1 *Consider the paired regularized programs RP and DRP.*

  1. *If one of the problems is inconsistent, then the other is inconsistent or unbounded.*

12

2. *Let the two problems be consistent, and let x^0 be a feasible solution for P and U^0 be a feasible solution for DRP. Then*

$$c^t x^0 \leq \operatorname{trace} b U^0.$$

3. *If both RP and DRP are consistent, then they have optimal solutions and their optimal values are equal.*

4. *Let x^0 and U^0 be feasible solutions of RP and DRP, respectively. Then x^0 and U^0 are optimal if and only if*

$$\operatorname{trace} U^0 (b - A x^0) = 0,$$

if and only if

$$U^0 (b - A x^0) = 0.$$

5. *The vector $x^0 \in \Re^m$ and matrix $U \in \mathcal{S}_n$ are optimal solutions of RP and DRP, respectively, if and only if (x^0, U^0) is a saddle point of the Lagrangian $L(x, U)$ for all (x, U) in $\Re^m \otimes (\mathcal{P}^f)^+$, and then*

$$L(x^0, U^0) = c^t x^0 = \operatorname{trace} b U^0.$$

3.3   Extended Duals

The above dual program DRP uses the minimal cone explicitly. In [34], the *Extended Lagrange-Slater Dual* program is proposed. First define the following sets:

$$
\begin{aligned}
\mathcal{C}_k &= \{(U_i, W_i)_{i=1}^k : A^*(U_i + W_{i\perp 1}) = 0, \ \operatorname{trace} b(U_i + W_{i\perp 1}) = 0, \\
&\qquad U_i \succeq W_i W_i^t, \ \forall i = 1, \dots, k, W_0 = 0\}, \\
\mathcal{U}_k &= \{U_k : (U_i, W_i)_{i=1}^k \in \mathcal{C}_k\}, \qquad\qquad\qquad\qquad (3.3) \\
\mathcal{W}_k &= \{W_k : (U_i, W_i)_{i=1}^k \in \mathcal{C}_k\}.
\end{aligned}
$$

Note that Schur complements imply that

$$U_i \succeq W_i W_i^t \iff \begin{bmatrix} I & W_i^t \\ W_i & U_i \end{bmatrix} \succeq 0.$$

In [34] it is shown that strong duality holds for the following dual of P

$$
\begin{aligned}
p^* = \quad &\min \quad &&\operatorname{trace} b(U + W) \\
\textbf{(ELSD)} \qquad &\text{subject to} \quad && A^*(U + W) = c \\
& && W \in \mathcal{W}_m \\
& && U \succeq 0.
\end{aligned}
$$

13

The advantage for this dual is that it is stated completely in terms of the data of the original program, whereas DRP uses the minimal cone explicitly. Moreover, the size of ELSD is polynomially bounded in the size of the input problem P.

At a first glance the duals DRP and ELSD appear very different, especially in light of the fact that the matrices W do not have to be symmetric. However, the adjoint operator A^* involves traces which are unchanged by taking the symmetric part of matrices. Therefore, we can replace W by $W + W^t$, or equivalently, replace $\mathcal{W}_m$ by $\mathcal{W}_m^s$. We show below that, after this change, the two duals are actually equal, i.e. $\mathcal{P} + \mathcal{W} = (\mathcal{P}^f)^+$, where

$$\mathcal{W} = \mathcal{W}_m^S = \{W + W^t : W \in \mathcal{W}_m\}.$$

4   RELATIONSHIP BETWEEN DUALS

4.1   Duals of P

We now show the relationships between the above two strong dual programs.

The algorithm to find the minimal cone is based on [11, Lemma 7.1] which we now phrase for our specific P problem; we include a proof for completeness.

Lemma 4.1 *Suppose $\mathcal{P}^f \lhd K \lhd \mathcal{P}$. The system*

$$A^*U = 0, U \succeq_{K^+} 0, \operatorname{trace} Ub = 0 \tag{4.1}$$

is consistent only if

$$\text{the minimal cone } \mathcal{P}^f \subset \{U\}^\perp \cap K \lhd K. \tag{4.2}$$

Proof.   Since $\operatorname{trace} U(Ax - b) = 0$, for all x, we get $(A(F) - b) \subset U^\perp$, i.e. $\mathcal{P}^f \subset \{U\}^\perp$. Also, that $\{U\}^\perp \cap K$ is a face of K follows from $U \succeq_{K^+} 0$. $\qquad\square$

The result in [11, Lemma 7.1] is for more general convex, cone valued functions. However, the linearity of P means that it is equivalent to our statement above, i.e. the subgradient removes the need for the initial feasible point and then, complementary slackness is equivalent to $\operatorname{trace} Ub = 0$ in the presence of the stationarity condition $A^*U = 0$.

We now use the algorithm for finding $\mathcal{P}^f$ presented in [11] to show the relation between the two duals of P. We see that each step of the algorithm

14

finds a smaller dimensional face $\mathcal{P}_k$ which contains the minimal cone $\mathcal{P}^f$. We show that

$$\mathcal{P}_k^+ = \mathcal{P} + \mathcal{W}_k^s, \; \mathcal{W}_k^s = (\mathcal{P}_k)^\perp.$$

There is one difference with the algorithm discussed here and the one from [11]; here we find the points in the relative interior of the complementary faces, rather than an arbitrary point (which may be on the boundary). This guarantees the immediate correspondence with the dual ELSD.

Step 1
Define $\mathcal{P}_0 := \mathcal{P}$ and note that, since $W_0 = 0$ in (3.3),

$$\mathcal{U}_1 := \{U \succeq 0 : A^*U = 0, \text{trace } Ub = 0\}.$$

Choose $\hat{U}_1 \in \text{relint}\,\mathcal{U}_1$. (If $\hat{U}_1 = 0$, then Slater's condition holds for P and we STOP.) Further, let

$$\mathcal{P}_1 := (\mathcal{F}(\mathcal{U}_1))^c \; (= \{\hat{U}_1\}^\perp \cap \mathcal{P}_0 \lhd \mathcal{P}_0).$$

We can now define the following equivalent program to P and its Lagrangian dual.

$$\begin{array}{rlc}
& p^* = \max & c^t x \\
(\mathbf{RP_1}) & \text{s.t.} & Ax \preceq_{\mathcal{P}_1} b \\
& & x \in \Re^m.
\end{array}$$

$$\begin{array}{rlc}
& d_1^* = \min & \text{trace } bU \\
(\mathbf{DRP_1}) & \text{s.t.} & A^*U = c \\
& & U \succeq_{(\mathcal{P}_1)^+} 0.
\end{array}$$

Note that $p^* \le d_1^* \le d^*$. From Corollary 2.1 and Lemma 2.1 we conclude that

$$(\mathcal{P}_1)^+ = (\mathcal{P} \cap \mathcal{P}_1)^+ = \mathcal{P} + (\mathcal{P}_1)^\perp$$

so that

$$(\mathcal{P}_1)^+ = \mathcal{P} + ((\mathcal{F}(\mathcal{U}_1))^c)^\perp, \; (\mathcal{P}_1)^\perp = \mathcal{W}_1^S.$$

Therefore we get the following equivalent program to DRP_1.

$$\begin{array}{rlc}
& d_1^* = \min & \text{trace } b(U + (W + W^t)) \\
& \text{s.t.} & A^*(U + (W + W^t)) = c \\
(\mathbf{ELSD_1}) & & A^*U_1 = 0, \text{trace } U_1 b = 0 \\
& & U \succeq 0, \; \begin{bmatrix} I & W^t \\ W & U_1 \end{bmatrix} \succeq 0.
\end{array}$$

Step 2

We can now apply the same procedure to the program RP_1. Since $\mathcal{W}_1^S = (\mathcal{P}_1)^\perp$, we get

$$\mathcal{U}_2 := \{U \succeq 0 : (U + V) \succeq_{(\mathcal{P}_1)^+} 0, A^*(U + V) = 0, \operatorname{trace}(U + V)b = 0\}.$$

Choose $\hat{U}_2 \in \operatorname{relint}\mathcal{U}_2$. (If $\hat{U}_2 = 0$, then the generalized Slater's condition holds for RP_1 and we STOP.)

$$\mathcal{P}_2 := (\mathcal{F}(\mathcal{U}_2))^c \ (= \{\hat{U}_2\}^\perp \cap \mathcal{P}_1 \lhd \mathcal{P}_1).$$

We get a new equivalent program to P and its Lagrangian dual.

$$
\begin{array}{rll}
p^* = & \max & c^t x \\
(\textbf{RP}_2) & \text{s.t.} & Ax \preceq_{\mathcal{P}_2} b \\
& & x \in \Re^m.
\end{array}
$$

$$
\begin{array}{rll}
d_2^* = & \min & \operatorname{trace} bU \\
(\textbf{DRP}_2) & \text{s.t.} & A^*U = c \\
& & U \succeq_{(\mathcal{P}_2)^+} 0.
\end{array}
$$

We now have $p^* \leq d_2^* \leq d_1^* \leq d^*$. From Corollary 2.1 and Lemma 2.1 we conclude that

$$(\mathcal{P}_2)^+ = (\mathcal{P} \cap \mathcal{P}_2)^+ = \mathcal{P} + (\mathcal{P}_2)^\perp$$

and

$$(\mathcal{P}_2)^+ = \mathcal{P} + ((\mathcal{F}(\mathcal{U}_2))^c)^\perp, \ (\mathcal{P}_2)^\perp = \mathcal{W}_2^S.$$

Therefore we get the following equivalent program to DRP_2.

$$
\begin{array}{rl}
d_1^* = & \min \quad \operatorname{trace} b(U + (W + W^t)) \\
& \text{s.t.} \quad A^*(U + (W + W^t)) = c \\
& \qquad A^*U_1 = 0, \operatorname{trace} U_1 b = 0 \\
& \qquad A^*(U_2 + (W_1 + W_1^t)) = 0, \\
& \qquad \operatorname{trace}(U_2 + (W_1 + W_1^t))b = 0 \\
(\textbf{ELSD}_2) & \qquad U \succeq 0, \ \begin{bmatrix} I & W_1^t \\ W_1 & U_1 \end{bmatrix} \succeq 0 \\
& \qquad \begin{bmatrix} I & W^t \\ W & U_2 \end{bmatrix} \succeq 0.
\end{array}
$$

... Step k ...

16

The remaining steps of the algorithm and the regularization are similar and we see that after $k \leq \min\{m, n\}$ steps we obtain the equivalence of RP with RP_k, and ELSD with $ELSD_k$. The following theorem clarifies some of the relationships between the various sets.

Theorem 4.1 *For some $k \leq \min\{m, n\}$, we have*

$$\mathcal{F}(\mathcal{U}_k) = (\mathcal{P}_k)^c, \ and \ \mathcal{U}_1 \subset \mathcal{U}_2 \subset \ldots \subset \mathcal{U}_k = \ldots = \mathcal{U}_m = (\mathcal{P}^f)^c. \qquad (4.3)$$

$$\mathcal{W}_k^s = (\mathcal{P}_k)^\perp = ((\mathcal{F}(\mathcal{U}_k))^c)^\perp, \ \ \mathcal{W}_1^S \subset \ldots \subset \mathcal{W}_k^S = \ldots = \mathcal{W}_m^S = (\mathcal{P}^f)^\perp. \tag{4.4}$$

Proof.   The nesting is clear from the definitions and is discussed in [34, Lemma 3] (for $\mathcal{W}_k$). Moreover, in [34, Lemma 2] it is shown that for $k = 1, 2, \ldots, m,$

$$(b - Ax)U = 0, \ and \ (b - Ax)W = 0, \ \forall x \in F, U \in \mathcal{U}_k, W \in \mathcal{W}_k.$$

Therefore, the inclusions in $(\mathcal{P}^f)^c, (\mathcal{P}^f)^\perp$ follow. Equality follows from the dimension of the feasible set, $F \subset \Re^m$, and a partial converse of Lemma 4.1, i.e. if $\mathcal{U}_k^c \neq \mathcal{P}^f$, then the system (4.1), with $U \neq 0$, is consistent, see [11, Corollary 7.1].

$\square$

4.2   Duals of D

Similar results can be obtained for the dual of D, i.e. we can use the minimal cone to close the duality gap and we can get an explicit representation for the minimal cone. The extended Lagrange-Slater dual of the dual D is

$$\begin{array}{lll} & d^* = & \max & \text{trace } c^t x \\ (\textbf{ELSDD}) & & \text{subject to} & A(x + (Z + Z^t)) \preceq b \\ & & & Z \in \mathcal{Z}_m, \end{array}$$

for $\mathcal{Z}_m$ to be derived below.

We can reformulate the dual D to the form of P, i.e. define the cone

$$S = \Re^m \otimes \mathcal{P}, \ (S^+ = \{0\}^m \otimes \mathcal{P})$$

and the constraint operator $G : \Re^m \otimes \mathcal{S}_n \to \mathcal{S}_n$

$$G \begin{pmatrix} x \\ V \end{pmatrix} := Ax + V, \ \ G^* U = \begin{pmatrix} A^* U \\ U \end{pmatrix}.$$

The dual D is equivalent to

$$d^* = \quad \min \quad \operatorname{trace} bU$$
$$(\mathbf{ED}) \qquad \text{subject to} \quad G^*U \succeq_{S^+} \begin{pmatrix} c \\ 0 \end{pmatrix}.$$

We have the following equivalence to Lemma 4.1.

Lemma 4.2 *Suppose $S_D^f \lhd K \lhd S^+$. The system*

$$\phi = \begin{pmatrix} x \\ Ax \end{pmatrix} \succeq_{K^+} 0, \ \operatorname{trace} x^t c = 0 \tag{4.5}$$

is consistent only if

$$\text{the minimal cone } S_D^f \subset (\{\phi\}^\perp \cap K) \lhd K. \tag{4.6}$$

Proof. Suppose that ϕ is found from (4.5) and $U \in F_D$. Now

$$\left\langle \phi, G^*U - \begin{pmatrix} c \\ 0 \end{pmatrix} \right\rangle = x^t(A^*U - c) + \operatorname{trace} U(Ax)$$
$$= -x^t c + \operatorname{trace} U(Ax - Ax) = 0,$$

since $x^t c = 0$. We get $G(F_D) - \begin{pmatrix} c \\ 0 \end{pmatrix} \subset \phi^\perp$, i.e. the minimal cone $S_D^f \subset$ $\{\phi\}^\perp$. Finally, the fact that $\{\phi\}^\perp \cap K$ is a face of K follows from $\phi \in K^+$, i.e. $\{\phi\}^\perp$ is a supporting hyperplane containing S^f. $\qquad\qquad\square$

The faces of S and S^+ directly correspond to faces of $\mathcal{P}$.

Lemma 4.3 *1. If $D \subset S^+$, then $\mathcal{F}(D) = 0 \otimes K$, where $K \lhd \mathcal{P}$.*

2. If $D \subset S$, then $\mathcal{F}(D) = \Re^m \otimes K$, where $K \lhd \mathcal{P}$.

Proof. The statements follow from the definitions. $\qquad\qquad\square$

We also need a result similar to Lemma 2.1.

Lemma 4.4 *Suppose that D is a convex cone and $D \subset S$. Let*

$$K := \left\{ \begin{pmatrix} x \\ W + W^t \end{pmatrix} : x \in \Re^m, \ U \succeq WW^t, \text{ for some } \begin{pmatrix} y \\ U \end{pmatrix} \in D \right\}.$$

18

Then

$$K = ((\mathcal{F}(D))^c)^\perp$$

$$= \left\{ \begin{pmatrix} x \\ W + W^t \end{pmatrix} : \begin{bmatrix} I & W^t \\ W & U \end{bmatrix} \succeq 0, \; \textit{for some } \begin{pmatrix} y \\ U \end{pmatrix} \in D \right\}.$$

Proof.   The proof is very similar to the proof of Lemma 2.1. The difference is that we have to account for the cone S^+ being the direct sum $0^m \otimes \mathcal{P}$. We include the details for completeness.

Suppose that $\begin{pmatrix} x \\ W + W^t \end{pmatrix} \in K$, i.e. $U \succeq WW^t$, for some $\begin{pmatrix} y \\ U \end{pmatrix} \in D$. Then there exists a matrix H such that $W = UH$, see (2.5). Therefore, $\operatorname{trace} WV = 0$, for all $\begin{pmatrix} 0 \\ V \end{pmatrix} \in (\mathcal{F}(D))^c \subset S^+$, i.e.

$$\begin{pmatrix} x \\ W + W^t \end{pmatrix} \in ((\mathcal{F}(D))^c)^\perp.$$

To prove the converse, suppose that $\begin{pmatrix} x \\ V \end{pmatrix} \in ((\mathcal{F}(D))^c)^\perp$ and $\begin{pmatrix} y \\ U \end{pmatrix} \in D \cap \operatorname{relint} \mathcal{F}(D)$. Let U be orthogonally diagonalized by $Q = [Q_1 Q_2]$

$$U = Q^t \operatorname{Diag}(d_1 \; 0)Q, \; Q^t Q = I,$$

with $Q_1, n \times r, \; d_1 > 0$. Therefore,

$$\mathcal{F}(D) = \left\{ \begin{pmatrix} x \\ Q_1 B Q_1^t \end{pmatrix} : B \succeq 0, \; B \in \mathcal{S}_r, \; x \in \Re^m \right\}$$

and

$$(\mathcal{F}(D))^c = \left\{ \begin{pmatrix} 0 \\ Q_2 B Q_2^t \end{pmatrix} : B \succeq 0, \; B \in \mathcal{S}_{n \perp r}, 0 \in \{0\}^m \right\}.$$

Now

$$\begin{pmatrix} x \\ V \end{pmatrix} \in ((\mathcal{F}(D))^c)^\perp$$

implies that

$$0 = \operatorname{trace} V Q_2 B Q_2^t = \operatorname{trace} Q_2^t V Q_2 B, \; \forall B \succeq 0,$$

19

i.e.

$$Q_2^t V Q_2 = 0.$$

This implies that $Q_2 Q_2^t V Q_2 Q_2^t = 0$ as well. Note that $Q_2 Q_2^t$ is the orthogonal projection onto $\mathcal{N}(U)$. Therefore the nonzero eigenvalues of V correspond to eigenvectors in the eigenspace formed from the column space of Q_1. Since the same must be true for $V V^t$, this implies that $\alpha U \succeq V V^t$, for some $\alpha > 0$ large enough, i.e. $V \in K$.

$\square$

Now define the following sets:

$$
\begin{aligned}
\mathcal{D}_k &= \{ (V_i, Z_i)_{i=1}^k : A x_i + (Z_{i \perp 1} + Z_{i \perp 1}^t) \succeq 0,\ x_i^t c = 0, \\
&\qquad V_i = A x_i,\ V_i \succeq Z_i Z_i^t,\ \forall i = 1, \ldots, k, Z_0 = 0 \} \\
\mathcal{V}_k &= \{ V_k : (V_i, Z_i)_{i=1}^k \in \mathcal{D}_k \} \\
\mathcal{Z}_k &= \{ Z_k : (V_i, Z_i)_{i=1}^k \in \mathcal{D}_k \}
\end{aligned}
$$

The extended Lagrange-Slater dual of the dual D can now be stated.

$$
\textbf{(ELSDD)} \qquad
\begin{aligned}
d^* = \quad &\max &&\operatorname{trace} c^t x \\
&\text{subject to} \quad &&A\big(x + (Z + Z^t)\big) \preceq b \\
& &&Z \in \mathcal{Z}_m
\end{aligned}
$$

Step 1
Define $T_0 := S^+$ and $\mathcal{P}_0 := \mathcal{P}$ and note that, since $Z_0 = 0$,

$$
\begin{aligned}
\mathcal{V}_1 &:= \left\{ A x : \phi = \begin{pmatrix} x \\ A x \end{pmatrix}, \phi \succeq_{T_0^+} 0,\ x^t c = 0 \right\} \\
&= \{ V : V = A x \succeq 0,\ x^t c = 0 \}.
\end{aligned}
$$

Choose $\hat{V}_1 \in \operatorname{relint} \mathcal{V}_1$. (If $\hat{V}_1 = 0$, then the generalized Slater's condition holds for ED and we STOP.) Further, let

$$T_1 := (\mathcal{V}_1^f)^c \ (= \{\hat{V}_1\}^\perp \cap T_0 \lhd T_0).$$

Therefore,

$$T_1 = \{0\}^m \otimes \mathcal{P}_1,$$

thus defining the face $\mathcal{P}_1 \lhd \mathcal{P}_0$.

We can now define the following equivalent program to ED and its Lagrangian dual.

$$
\begin{array}{rll}
d^* = & \min & \operatorname{trace} bU \\
& \text{s.t.} & A^*U = c \\
& & U \succeq_{\mathcal{P}_1} 0 \\
& \text{or} & G^*U \succeq_{T_1} \begin{pmatrix} c \\ 0 \end{pmatrix}.
\end{array}
$$

$(\mathbf{RED_1})$

$$
\begin{array}{rll}
p_1^* = & \max & c^t x \\
& \text{subject to} & Ax \preceq_{(\mathcal{P}_1)^+} b \\
& & x \in \Re^m \\
& \text{or} & G\phi =_{T_1^+} b, \ \phi \succeq_{T_1^+} 0.
\end{array}
$$

$(\mathbf{DRED_1})$

Note that $p^* \leq p_1^* \leq d^*$. From Corollary 2.1 we conclude that

$$
(\mathcal{P}_1)^+ = (\mathcal{P} \cap \mathcal{P}_1)^+ = \mathcal{P} + (\mathcal{P}_1)^\perp
$$

so that

$$
(T_1)^+ = S + ((\mathcal{F}(\mathcal{V}_1))^c)^\perp .
$$

Therefore Lemma 4.4 yields the following equivalent SDP to $DRED_1$.

$$
\begin{array}{rll}
p_1^* = & \max & c^t x \\
& \text{s.t.} & Ax + (Z + Z^t) \preceq b \\
& & Ay \succeq 0, c^t y = 0 \\
& & \begin{bmatrix} I & Z^t \\ Z & Ay \end{bmatrix} \succeq 0.
\end{array}
$$

$(\mathbf{ELSDD_1})$

Step 2

We can now apply the same procedure to the program RED_1.

$$
\begin{aligned}
\mathcal{V}_2 & := \left\{ Ax : \phi = \begin{pmatrix} x \\ Ax \end{pmatrix}, \phi \succeq_{T_1^+} 0, \ x^t c = 0 \right\} \\
& = \{ V : V = Ax \succeq_{\mathcal{P}_1} 0, \ x^t c = 0 \}.
\end{aligned}
$$

Choose $\hat{V}_2 \in \operatorname{relint} \mathcal{V}_2$. (If $\hat{V}_2 = 0$, then the generalized Slater's condition holds for DRP_1 and we STOP.) Let

$$
T_2 := (\mathcal{F}(\mathcal{V}_2))^c \ (= \{\hat{V}_2\}^\perp \cap T_1 \vartriangleleft T_1).
$$

21

We get a new equivalent program to D and its Lagrangian dual.

$$
\begin{aligned}
d^* = \quad &\min & \operatorname{trace} bU \\
&\text{s.t.} & A^*U = c \\
&& U \succeq_{\mathcal{P}_2} 0
\end{aligned}
$$

$(\mathbf{RED_2})$

$$
\text{or} \quad G^*U \succeq_{T_2} \begin{pmatrix} c \\ 0 \end{pmatrix}.
$$

$$
\begin{aligned}
p_2^* = \quad &\max & c^t x \\
&\text{subject to} & Ax \preceq_{(\mathcal{P}_2)^+} b \\
&& x \in \Re^m
\end{aligned}
$$

$(\mathbf{DRED_2})$

$$
\text{or} \quad G\phi =_{T_2^+} b, \ \phi \succeq_{T_2^+} 0.
$$

We now have $p^* \le p_1^* \le p_2^* \le d^*$. From Corollary 2.1 we get

$$
(\mathcal{P}_2)^+ = (\mathcal{P} \cap \mathcal{P}_2)^+ = \mathcal{P} + (\mathcal{P}_2)^\perp
$$

so that

$$
(T_2)^+ = S + ((\mathcal{F}(\mathcal{V}_2))^c)^\perp.
$$

Therefore Lemma 4.4 yields the following equivalent SDP to DRP_2.

$$
\begin{aligned}
p_2^* = \quad &\max & c^t x \\
&\text{s.t.} & Ax + (Z + Z^t) \preceq b \\
&& A(y + (Z + Z^t)) \succeq 0, c^t y = 0 \\
&& \begin{bmatrix} I & Z^t \\ Z & Ay \end{bmatrix} \succeq 0 \\
&& A(y_1 + (Z_1 + Z_1^t)) \succeq 0, c^t y = 0 \\
&& \begin{bmatrix} I & Z_1^t \\ Z_1 & Ay_1 \end{bmatrix} \succeq 0.
\end{aligned}
$$

$(\mathbf{ELSDD_2})$

... Step k ...

5   HOMOGENIZATION

In Section 3.1.1, it is illustrated that the ordinary linear programming problem can have an infinite number of dual programs for which strong duality holds. This includes the standard Lagrangian dual. However, this is not the case for SDP. First, the standard Lagrangian dual can result in a duality

gap, see [34, Example 1]. Moreover, the duality gap may be 0, but the dual may not be attained, see [34, Example 5].

However, we have seen that the two equivalent duals DRP and ELSD both provide a zero duality gap and dual attainment, i.e. strong duality. Since LP is a special case of SDP, we conclude that there are examples of SDP where there are many duals for which strong duality holds. A natural question to ask is whether there is any type of uniqueness for the strong duals? And, among the strong duals, what is the "strongest", i.e. which is the "closest" to the standard Lagrangian dual?

Therefore, we now look at general optimality conditions for P. We do this using the homogenized semidefinite program

$$
\begin{array}{rlll}
0 = & \max & c^t x + t(-p^*) & (= \langle a, w \rangle) \\
& \text{subject to} & Ax + t(-b) + Z = 0 & (Bw = 0) \\
(\textbf{HP}) & & & \\
& & w \in K = \Re^m \otimes \Re_+ \otimes \mathcal{P} & \left(w = \begin{pmatrix} x \\ t \\ Z \end{pmatrix} \right)
\end{array}
$$

The above defines the vector a, the linear operator B, and the convex cone K. Let F_H denote the feasible set, i.e.

$$F_H = \mathcal{N}(B) \cap K,$$

where $\mathcal{N}$ denotes null space.

Note that if $t = 0$ in a feasible solution of HP, then $A(\alpha w) = 0$, $\forall \alpha \in \Re$, and $w = \begin{pmatrix} x \\ 0 \\ Z \end{pmatrix}$. Therefore $c^t x > 0$ implies that $p^* = \infty$. While if $t > 0$ in a feasible solution of HP, then $w = \begin{pmatrix} \frac{1}{t} x \\ 1 \\ \frac{1}{t} Z \end{pmatrix}$ is feasible which implies that $c^t x + t(-p^*) \leq 0$. Therefore

$$Bw = 0, w \in K \text{ implies } \langle a, w \rangle \leq 0. \tag{5.1}$$

This shows that 0 is in fact the optimal value of HP and HP is an equivalent problem to P.

One advantage of HP is that we know a feasible solution, namely the origin. Recall that the *polar* of a set C

$$C^+ = \{ \phi : \langle \phi, c \rangle \geq 0, \forall c \in C \}.$$

23

With this definition, the optimality conditions for HP are simply that the negative of the gradient of the objective function is in the polar of the feasible set, i.e. from (5.1) we conclude

$$a = \begin{pmatrix} c \\ -p \\ 0 \end{pmatrix} \in -(\mathcal{N}(B) \cap K)^+. \quad \begin{pmatrix} \text{optimality} \\ \text{conditions} \\ \text{for HP} \end{pmatrix} \quad (5.2)$$

This yields the asymptotic optimality conditions (up to closure)

$$\begin{pmatrix} c \\ -p \\ 0 \end{pmatrix} \in -(\overline{\mathcal{R}(B^*) + K^+}), \quad (5.3)$$

where the adjoint operator

$$B^* U = \begin{pmatrix} A^* U \\ -\operatorname{trace} bU \\ U \end{pmatrix},$$

and the polar cone

$$K^+ = \{0\} \otimes \Re_+ \otimes \mathcal{P}.$$

We have used the fact that the polar of the intersection of sets is the closure of the sum of the polars of the sets and that $\mathcal{P}$ is self-polar, $\mathcal{P} = \mathcal{P}^+$. Note that if the closure in (5.3) is not needed, then these optimality conditions, along with weak duality for P and D, $p \leq \operatorname{trace} bU$, yield optimality conditions for P, i.e. (5.3) with closure is equivalent to

$$\begin{pmatrix} c \\ -p \\ 0 \end{pmatrix} = \begin{pmatrix} A^* U \\ -\operatorname{trace} bU \\ U \end{pmatrix} - \begin{pmatrix} 0 \\ \alpha \\ V \end{pmatrix} \begin{pmatrix} \text{dual feasibility} \\ \text{strong duality} \\ \text{dual feasibility} \end{pmatrix}, \quad (5.4)$$

for some $\alpha \geq 0$, $V \succeq 0$. This yields the optimality conditions for P:

$$A^* U = c, U \succeq 0 \quad \text{(dual feasibility)}$$
$$p = \operatorname{trace} bU \quad \text{(strong duality)}$$

(Note that strong duality is equivalent to complementary slackness.) We have proved the following.

24

Theorem 5.1 *$p \in \Re$ is the optimal value of P if and only if (5.3) holds. Moreover, suppose that (5.3) holds but*

$$\begin{pmatrix} c \\ -p \\ 0 \end{pmatrix} \notin \mathcal{R}(B^*) - K^+. \tag{5.5}$$

Then p is still the optimal value of P but either there is a duality gap or the dual D is unattained, i.e. strong duality fails for P and D.

□

The above theorem provides a means of generating examples where strong duality fails, i.e. we need to find examples where the closure does not hold and then pick a vector that is in the closure but not the preclosure.

There are many conditions, called *constraint qualifications*, that guarantee the closure condition in (5.3). In fact, this closure has been referred to as a *weakest constraint qualification*, [21, 37]. As an example of a closure condition, see e.g. [23, pp. 104-105], if C, D are closed convex sets and the intersection of their recession cones is $\{0\}$, then $D - C$ is closed. (Here the recession cone of a convex set C is the set of all points x such that $x + D \subset D$.) Therefore, for a subspace $\mathcal{V}$ and a convex cone K,

$$\mathcal{V} \cap K = \{0\} \text{ implies } \mathcal{V} + K \text{ is closed.}$$

In our case, several conditions for the closure (constraint qualifications) are given in [13, Theorem 3.1]. For example, the cone generated by the set $F_H - K$ is the whole space; or Slater's condition

$$\exists \hat{x} \in F \text{ such that } A\hat{x} \prec b.$$

One approach to guarantee the closure condition is to try and find sets, T, to add to attain the closure. Equivalently, find sets, C, $C^+ = T$, to intersect with K to attain the closure since

$$(\mathcal{N}(B) \cap K)^+ = (\mathcal{N}(B) \cap (K \cap C))^+ = \overline{\mathcal{R}(B^*) + K^+ + C^+}. \tag{5.6}$$

There are some trivial choices for the set, e.g. $C = \mathcal{N}(B) \cap K$. Another choice would be $(\mathcal{N}(B) \cap K)^f$.

The above translates into choosing sets that contain the minimal cone $\mathcal{P}^f$. Since we want a small set of dual multipliers, we would like to find large sets that contain $\mathcal{P}^f$ but for which the above closure conditions hold.

25

It is conjectured that every SDP can be divided into parts, a linear part and a nonlinear part. Multipliers for the linear part correspond to linear programming, i.e. we choose the standard set of multipliers. However, we cannot choose a smaller set than $(\mathcal{P}^f)^+$ for the nonlinear part.

The following portion of the conjecture is easily established: Suppose both problems P and D have feasible solutions (so that if there is a duality gap then it is finite). Consider the cone

$$\mathcal{Z} = \{Z \in \mathcal{P} : \ Z = b - Ax, \text{ for some } x \in \Re^m\}.$$

If $\mathcal{Z} \cap int(\mathcal{P}) \neq \emptyset$ then we have an interior point and Lagrangian dual is good (strong duality holds). Otherwise, $\mathcal{Z} \subset \partial\mathcal{P}$. In particular, there exists a permutation matrix P and a block diagonal matrix structure in $\mathcal{S}_n$ such that $Z \in \mathcal{Z}$ implies that $P Z P^T$ is a block diagonal matrix which lies in the subspace defined by the block diagonal structure. We pick P such that each of the blocks has one of the following properties:

Type I blocks
Semidefinite programming problem arising from block i is an LP (that is the block matrix is a diagonal matrix). In this case strong duality holds for many duals including the Lagrangian dual.

Type II blocks
Semidefinite programming problem arising from block i is not an LP and condition (5.3) holds and (5.5) does not hold. In this case strong duality holds for many duals including the Lagrangian dual.

Type III blocks
Semidefinite programming problem arising from block i is not an LP and conditions (5.3) and (5.5) do hold. In this case, we can find linear objective functions for which D is feasible but strong duality does not hold for the Lagrangian dual.

In case the objective function is separable with respect to this partition, then the duality for Type I and Type II blocks is well understood. For Type III blocks we showed that as long as (5.3) and (5.5) hold, there will be objective functions for which D is feasible, yet strong duality does not hold for P and D.

6 CONCLUSION

In this paper we have studied dual programs that guarantee strong duality for SDP. In particular, we have seen the relationship between the dual, DRP, of the regularized program, RP, and the extended Lagrange-Slater dual ELSD. DRP uses the minimal cone $\mathcal{P}^f$ which, in general, cannot be computed exactly (numerically). ELSD shows that a regularized dual can be written down explicitly.

The pair P and D are the usual pair of dual programs used in SDP. This yields primal-dual interior-point methods when both programs satisfy the Slater CQ, i.e. strict feasibility. However, there are classes of problems where the CQ fails, see e.g. [25]. In fact, for these problems, which arise from relaxations of 0,1 combinatorial optimization problems with linear constraints, CQ fails for the primal while it is satisfied for the dual. Therefore, in theory, there is no duality gap between P and D. However, is D the true dual of P in this case? It is true that perturbations in b will yield the dual value d^* as the perturbations go to 0, if we can guarantee that we maintain the semidefinite constraint exactly. If we could do this, then we could solve any SDP independent of any regularity condition, i.e. we would only have to solve a perturbed dual to get the optimum value of the primal. However, the key here is that we cannot maintain the semidefinite constraint exactly, i.e. D is not a true dual of P in this case. It is the dual with respect to perturbations in the equality constraint $Ax + Z = b$ but not if we allow perturbations in the constraint $Z \succeq 0$ as well.

Unlike LP, the solutions and optimal values of SDP may be doubly exponential rational numbers or even irrational. Note that the optimal value being doubly exponential means that the size (the number of bits required to express the value in binary) is an exponential function of the size of the input problem P. However, in some cases it may be possible to find, a priori, upper bounds on the sizes of some primal and dual optimal solutions. Alizadeh [1] suggests that it may even be possible to bound the feasible solution sets of P and D a priori. Nevertheless this is impossible (even for an LP), if the feasible region of P is bounded then the feasible region of D is unbounded and vice versa. Hence one cannot hope to solve an SDP to exact optimality, or for that matter find feasible solutions of semidefinite inequality systems in polynomial time. However, a challenging open problem is to determine if a given rational semidefinite system has a solution. This problem is called the Semidefinite Feasibility Problem (SDFP), and in [34] it was shown by using ELSD that SDFP is not NP-Complete unless NP=Co-NP.

27

It may be interesting to try to interpret the significance of ELSD in terms of computational complexity of solving SDPs which do not satisfy the Slater CQ. A dual program, ELSD, can be written down in polynomial time; however, we do not know how to solve P or ELSD in polynomial time or we could get an explicit representation for $\mathcal{P}^f$, the minimal cone.

References

[1] F. ALIZADEH. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5:13–51, 1995.

[2] E. ANDERSON and P. NASH. *Linear Programming in Infinite Dimensional Spaces*. John Wiley and Sons, 1987.

[3] G.P. BARKER. The lattice of faces of a finite dimensional cone. *Linear Algebra and its Appl.*, 7:71–82, 1973.

[4] G.P. BARKER and D. CARLSON. Cones of diagonally dominant matrices. *Pacific J. of Math.*, 57:15–32, 1975.

[5] W.W. BARRETT, C.R. JOHNSON, and R. LOEWY. The real positive definite completion problem: cycle completability. Technical report, Department of Mathematics, College of William and Mary, Williamsburg, Virginia, 1993.

[6] A. BEN-ISRAEL. Linear equations and inequalities on finite dimensional, real or complex, vector spaces: a unified theory. *J. Math. Anal. Appl.*, 27:367–389, 1969.

[7] A. BEN-TAL and A NEMIROVSKI. Potential reduction polynomial time method for truss topology design. *SIAM Journal on Optimization*, 4:596–612, 1994.

[8] A. BERMAN. *Cones, Matrices and Mathematical Programming*. Springer-Verlag, Berlin, New York, 1973.

[9] F. BOHNENBLUST. Joint positiveness of matrices. Technical report, 1948. unpublished manuscript.

[10] J.M. BORWEIN and H. WOLKOWICZ. Characterizations of optimality for the abstract convex program with finite dimensional range. *J. Austra. Math. Soc. Series A*, 30:390–411, 1981.

[11] J.M. BORWEIN and H. WOLKOWICZ. Regularizing the abstract convex program. *Journal of Mathematical Analysis and its Applications*, 83:495–530, 1981.

[12] J.M. BORWEIN and H. WOLKOWICZ. Characterizations of optimality without constraint qualification for the abstract convex program. *Mathematical Programming Study*, 19:77–100, 1982.

[13] J.M. BORWEIN and H. WOLKOWICZ. A simple constraint qualification in infinite dimensional programming. *Mathematical Programming*, 35:83–96, 1986.

[14] S. BOYD, L. El GHAOUI, E. FERON, and V. BALAKRISHNAN. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.

[15] C. DELORME and S. POLJAK. The performance of an eigenvalue bound on the max-cut problem in some classes of graphs. *Discrete Mathematics*, 111:145–156, 1993.

[16] R.J. DUFFIN. Infinite programs. In A.W. Tucker, editor, *Linear Equalities and Related Systems*, pages 157–170. Princeton University Press, Princeton, NJ, 1956.

[17] R.M. FREUND. Complexity of an algorithm for finding an approximate solution of a semi-definite program with no regularity assumption. Technical Report OR 302-94, MIT, Cambridge, MA, 1994.

[18] K. GLASHOFF and S. GUSTAFSON. *Linear Optimzation and Approximation*, volume 45 of *Applied Mathematical Sciences*. Springer-Verlag, Verlag Basel, 1978.

[19] M.X. GOEMANS and D.P. WILLIAMSON. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. Technical report, Department of Mathematics, MIT, 1994.

[20] B. GRONE, C. JOHNSON, E. MARQUES DE SA, and H. WOLKOW-ICZ. Positive definite completions of partial Hermitian matrices. *Linear Algebra and its Applications*, 58:109–124, 1984.

[21] M. GUIGNARD. Generalized Kuhn-Tucker conditions for mathematical programming problems in a banach space. *SIAM J. of Control*, 7:232–241, 1969.

[22] C. HELMBERG, F. RENDL, R. J. VANDERBEI, and H. WOLKOW-ICZ. An interior point method for semidefinite programming. *SIAM Journal on Optimization*, To appear. Accepted Aug/94.

[23] R.B. HOLMES. *Geometric Functional Analysis and its Applications*. Springer-Verlag, Berlin, 1975.

[24] C. JOHNSON, B. KROSCHEL, and H. WOLKOWICZ. An interior-point method for approximate positive semidefinite completions. Technical Report CORR Report 95-11, University of Waterloo, Waterloo, Canada, 1995. Submitted.

[25] S. KARISCH, F. RENDL, H. WOLKOWICZ, and Q. ZHAO. Quadratic Lagrangian relaxation for the quadratic assignment problem. Research report, University of Waterloo, Waterloo, Ontario, In progress.

[26] K. KRETSCHMER. Programming in paired spaces. *Canad. J. Math.*, 13:221–238, 1961.

[27] N. LEVINSON. A class of continuous linear programming problems. *J. Math. Anal. Appl.*, 16:73–73, 1966.

[28] L. LOVÁSZ and A. SCHRIJVER. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

[29] K. LÖWNER. Uber monotone matrixfunctionen. *Math. Z.*, 49:375–392, 1934.

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

[31] Y. E. NESTEROV and A. S. NEMIROVSKY. *Interior Point Polynomial Algorithms in Convex Programming : Theory and Algorithms*. SIAM Publications. SIAM, Philadelphia, USA, 1994.

[32] F. PUKELSHEIM. *Optimal Design of Experiments*. Wiley, New York, 1993.

[33] M.V. RAMANA. *An algorithmic analysis of multiquadratic and semidefinite programming problems*. PhD thesis, Johns Hopkins University, Baltimore, Md, 1993.

[34] M.V. RAMANA. An exact duality theory for semidefinite programming and its complexity implications. Technical Report DIMACS Technical Report, 95-02R, RUTCOR, Rutgers University, New Brunswick, NJ, 1995. DIMACS URL: http://dimacs.rutgers.edu.

[35] T.W. REILAND. Optimality conditions and duality in continuous programming. ii. the linear problem revisited. *J. Math. Anal. Appl.*, 77:329–343, 1980.

[36] L. VANDENBERGHE and S. BOYD. Positive definite programming. Technical report, Electrical Engineering Department, Stanford, CA, 1994.

[37] H. WOLKOWICZ. Geometry of optimality conditions and constraint qualifications: The convex case. *Mathematical Programming*, 19:32–60, 1980.

[38] H. WOLKOWICZ. Some applications of optimization in matrix theory. *Linear Algebra and its Applications*, 40:101–118, 1981.