SIAM J. OPTIM.
Vol. 7, No. 3, pp. 641–662, August 1997

004

STRONG DUALITY FOR SEMIDEFINITE PROGRAMMING*

MOTAKURI V. RAMANA[†], LEVENT TUNÇEL[‡], AND HENRY WOLKOWICZ[‡]

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 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 subject classifications. Primary, 65K10; Secondary, 90C25, 90M45, 15A45, 47D20

PII. S1052623495288350

1. Introduction.

1.1. Semidefinite programming (SDP).

We study strong duality theorems for the semidefinite linear programming problem

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

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 $\lambda K \subset K$ for all $\lambda \geq 0$. We consider the space of symmetric matrices, $\mathcal{S}_n$, as a vector space with the *trace inner product* $\langle U, X\rangle := \operatorname{trace} UX$. (Over the space of $n \times n$ matrices, $\langle U, X\rangle := \operatorname{trace}(U^t X)$.) 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. (This implies that the feasible set $F \neq \emptyset$.)

1.2. Background.

1.2.1. Cone of semidefinite matrices.

The cone of positive semidefinite matrices has been studied extensively for both its importance and geometric 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 on $\mathcal{S}_n$ called the Löwner partial order. Various monotonicity results were studied

* Received by the editors June 22, 1995; accepted for publication (in revised form) April 23, 1996. http://www.siam.org/journals/siopt/7-3/28835.html

† Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611 (ramana@iseipc.ise.ufl.edu).

‡ Department of Combinatorics and Optimization, Faculty of Mathematics, Waterloo, Ontario N2L 3G1, Canada (ltuncel@math.uwaterloo.ca, henry@orion.uwaterloo.ca).

with respect to this partial order [28, 29]. An early paper in this area is the one by Bohnenblust [9]. Optimization problems over cones of matrices are also discussed in the monograph by Berman [8].

More recently, we have seen a strong renewed interest in semidefinite programming. This is due to new applications in engineering (e.g., Ben-Tal and Nemirovskii [7], Boyd et al. [14], and Vandenberghe and Boyd [36]) and combinatorial optimization (e.g., Alizadeh [1], Goemans and Williamson [19], Lovász and Schrijver [27], Nesterov and Nemirovskii [30], Delorme and Poljak [15], and Helmberg et al. [22]). Other applications of SDP arise from the study of correlation matrices in statistics, e.g., Pukelsheim [31]; matrix completion problems, see [20, 5, 24]; and multiquadratic programs, e.g., [32].

Nesterov and Nemirovskii's book provides a unifying framework for polynomial-time interior-point algorithms in convex programming (which includes SDP). Currently, interior-point algorithms seem to be the best algorithms (from both theoretical and practical viewpoints) for solving SDP problems, e.g., [36]. An infeasible-start interior-point algorithm was presented in Freund [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 given tolerances. 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). This illustrates some of the difficulties encountered with nonzero duality gaps. 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.

1.2.2. Early duality results. Extensions of finite linear programming duality to infinite dimensions and/or to optimization problems over cones have been studied in the literature. We do not give a comprehensive survey, 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 in [16] is the notion of optimization with respect to a partial order induced by a cone. Duality theory is also central in the related notion of continuous programming, e.g., [25, 26, 34], which is closely tied in with infinite programming. 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 paper is motivated by the recent paper of Ramana [33]. A dual program, called an *extended Lagrange–Slater dual program* and denoted (ELSD), is presented therein. Strong duality holds for this dual and, in addition, it can be written down in polynomial time. Previous work on general (convex) cone constrained programs [13, 38, 11, 10] also presented dual programs for which strong duality holds. The results were based on regularization and on finding the so-called minimal cone of the program (P). We denote these duals 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 paper is 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 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 such duals are uniquely identified 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 extended Lagrange–Slater dual of the Lagrangian dual of (P). 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.

2. Geometry of the SDP cone.

We now outline several known and some new results on the geometry of the cone $\mathcal{P}$. More details can be found in [3, 4]. For an introduction to the geometry of convex sets, see Rockafellar [35].

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

$$x, y \in T,\ x + y \in K \Rightarrow x, y \in K. \tag{2.1}$$

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,

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

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,

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

Equivalent characterizations for K and K^c are given in (2.6) and (2.7).

Two additional facts about the faces of the cone $\mathcal{P}$ are as follows:

(i) 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

$X \in \operatorname{relint}(K^c)$ (respectively, $\operatorname{relint}(K)$). Also, complementary faces are orthogonal and satisfy $XY = 0$ for all $X \in K, Y \in K^c$.

(ii) The cone $\mathcal{P}$ is *projectionally exposed* (see [11]); 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 $\operatorname{relint}(K)$, 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 shows that we can express the orthogonal complement of a face completely in terms of a system of semidefinite inequalities. The semidefinite inequalities are based on the data of the original problem. The description is made possible by using a semidefinite completion problem.

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

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

Proof. Suppose that $W + W^t \in K$, i.e., $U \succeq WW^t$ for some $U \in C$. Since $x^t(U - WW^t)x \geq 0$ for all x, we get $\mathcal{N}(U) \subset \mathcal{N}(W^t)$. Equivalently, $\mathcal{R}(U) \supset \mathcal{R}(W)$. Since $UU^\dagger$ is the orthogonal projection 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

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

(See, e.g., [33].) 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 $Q = [Q_1, Q_2]$:

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

with $Q_1, n \times r,\ d_1 > 0$. Therefore, the minimal face can be written using block matrices as follows:

$$\begin{aligned} \mathcal{F}(C) &= \{Q_1BQ_1^t : B \succeq 0,\ B \in \mathcal{S}_r\} \\ &= \left\{Q \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} Q^t : B \succeq 0,\ B \in \mathcal{S}_r\right\} \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} (\mathcal{F}(C))^c &= \{Q_2BQ_2^t : B \succeq 0,\ B \in \mathcal{S}_{n-r}\} \\ &= \left\{Q \begin{bmatrix} 0 & 0 \\ 0 & B \end{bmatrix} Q^t : B \succeq 0,\ B \in \mathcal{S}_{n-r}\right\}. \end{aligned} \tag{2.7}$$

This implies that V in $((\mathcal{F}(C))^c)^\perp$ can be written in terms of blocks as

$$V = Q\left(\begin{bmatrix} .5T & C \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} .5T & 0 \\ C^t & 0 \end{bmatrix}\right)Q^t.$$

We then have

$$X = Q\left(\begin{bmatrix} T^2 + CC^t & 0 \\ 0 & 0 \end{bmatrix}\right)Q^t \preceq \alpha U$$

for sufficiently large α, i.e., $V + V^t \in K$.

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

Now, we 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 face K satisfies*

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

Then

$$\mathcal{P} + K^\perp = \overline{\mathcal{P} + \operatorname{span} K^c}; \tag{2.8}$$

$$\mathcal{P} + \operatorname{span} K \quad \textit{is not closed.} \tag{2.9}$$

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 k, 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 k unit vectors. Therefore, $X \in K$ has a $k \times k$ zero block, i.e.,

$$X = \begin{bmatrix} 0_k & 0 \\ 0 & \bar{X} \end{bmatrix}.$$

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

$$K^\perp = \left\{Y : Y = \begin{bmatrix} C & D \\ D^t & 0 \end{bmatrix},\ C \in \mathcal{S}_k,\ D \in \mathcal{M}_{k,n-k}\right\}.$$

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

$$T^n + P^n \to L = \begin{bmatrix} L_1 & L_2 \\ L_2^t & L_3 \end{bmatrix}.$$

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

$$L = \begin{bmatrix} L_1 & L_2 \\ L_2^t & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & L_3 \end{bmatrix},$$

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

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

To prove the converse inclusion, suppose that

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

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

$$\text{(2.10)} \qquad \operatorname{trace} \Phi W < 0 \le \operatorname{trace} \Phi(P + w) \; \forall P \in \mathcal{P}, \; w \in \operatorname{span} K^c.$$

This implies that $\Phi \succeq 0$ and $\Phi \in (K^c)^\perp$. But then $\operatorname{trace} \Phi W = \operatorname{trace} \Phi P + \operatorname{trace} \Phi w$, 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 $U \in K^c$, so $\operatorname{trace} \Phi w = 0$. This implies that $\operatorname{trace} \Phi W = \operatorname{trace} \Phi P \ge 0$, which contradicts (2.10). This completes the proof of (2.8).

Now suppose that $X \in \operatorname{relint}(K)$ 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 $\operatorname{span} K = \{Q_2 B Q_2^t : B \in \mathcal{S}\}$. Now let $B \succ 0, T \succ 0$, and $n = 1, 2, \ldots$. Choose T, L so that

$$[Q_1 \; Q_2] \begin{bmatrix} \frac{1}{n} T & L \\ L^t & nB \end{bmatrix} \begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix} \in \mathcal{P}.$$

But

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

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

$$[Q_1 \; Q_2] \begin{bmatrix} 0 & L \\ L^t & 0 \end{bmatrix} \begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix},$$

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

COROLLARY 2.1.

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

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

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

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. (See, for instance, the arguments in Duffin [16] and Alizadeh [1].)

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

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

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

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

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:

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

3.1.1. Linear programming special case. We note that the SDP pair (P) and (D) look exactly like 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 (since we assumed that the primal feasible set is nonempty), 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.,

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

where $A_{i:}$ denotes the ith row of A. 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 smallest 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 $(b - A\hat{x}) \in \operatorname{int} \mathcal{P}$. 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 primal program*; see [11, 38]:

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

Moreover, by the definition of faces, there exists $\hat{x}$ such that $(b - A\hat{x}) \in \operatorname{relint}(\mathcal{P}^f)$. 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:

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

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

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

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

$$\text{(RD)} \qquad \begin{array}{rcc} d^* = & \min & \operatorname{trace} bU \\ & \text{subject to} & A^*U = c \\ & & U \succeq_{\mathcal{P}^f_D} 0. \end{array}$$

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

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

The above presents two pairs of symmetric dual programs: (RP) and (DRP); (RD) and (DRD). The following theorem states that these dual pairs have all the nice properties of dual pairs in ordinary linear programming, i.e., [38, Theorem 4.1]. (Part 3 of Theorem 3.1 modifies and corrects the statement in [3.8].) 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.*

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*

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

3. *If both (RP) and (DRP) are consistent, then their optimal values are equal and (DRP) has an optimal solution.*

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*

$$\operatorname{trace} U^0(b - Ax^0) = 0$$

and if and only if

$$U^0(b - Ax^0) = 0.$$

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 $\Re^m \times (\mathcal{P}^f)^+$. Then,*

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

3.3. Extended duals. The above dual program (DRP) uses the minimal cone explicitly. In [33], the *extended Lagrange–Slater dual* program, $(ELSD)$, is proposed. First define the following sets:

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

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 [33] it is shown that strong duality holds for the following (ELSD) dual of (P):

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

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 bounded by a polynomial function of the size of the input problem (P).

At a first glance, the duals (DRP) and (ELSD) appear very different. This is especially true 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 the 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 the same, 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 problem (P). We include a proof for completeness.

LEMMA 4.1. *Suppose* $\mathcal{P}^f \lhd K \lhd \mathcal{P}$. *For every solution* U *of the system*

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

we have

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

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, the fact that $\{U\}^\perp \cap K$ is a face of K follows from $U \succeq_{K^+} 0$. ☐

The result in [11, Lemma 7.1] is for more general convex, vector valued functions. However, the linearity of (P) means that it is equivalent to our statement above.

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 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),

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

Choose $\hat{U}_1 \in \operatorname{relint}(\mathcal{U}_1)$. (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.

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

$$(\mathrm{DRP}_1) \qquad \begin{array}{rll} d_1^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & A^*U = c \\ & & U \succeq_{(\mathcal{P}_1)^+} 0. \end{array}$$

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$_1$).

$$(\text{ELSD}_1)\qquad \begin{array}{rll} d_1^* = & \min & \operatorname{trace} b(U + (W + W^t)) \\ & \text{s.t.} & A^*(U + (W + W^t)) = c \\ & & A^*U_1 = 0, \operatorname{trace} U_1 b = 0 \\ & & U \succeq 0, \begin{bmatrix} I & W^t \\ W & U_1 \end{bmatrix} \succeq 0. \end{array}$$

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.

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

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

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$_2$).

$$(\text{ELSD}_2)\qquad \begin{array}{rll} d_2^* = & \min & \operatorname{trace} b(U + (W + W^t)) \\ & \text{s.t.} & A^*(U + (W + W^t)) = c \\ & & A^*U_1 = 0, \operatorname{trace} U_1 b = 0 \\ & & A^*(U_2 + (W_1 + W_1^t)) = 0, \\ & & \operatorname{trace}(U_2 + (W_1 + W_1^t))b = 0 \\ & & U \succeq 0, \begin{bmatrix} I & W_1^t \\ W_1 & U_1 \end{bmatrix} \succeq 0 \\ & & \begin{bmatrix} I & W^t \\ W & U_2 \end{bmatrix} \succeq 0. \end{array}$$

... **Step k** ...

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$_\mathrm{k}$), and (ELSD) with (ELSD$_\mathrm{k}$). The following theorem clarifies some of the relationships between the various sets.

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

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

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

Proof. The nesting is clear from the definitions and is discussed in [33, Lemma 3] (for $\mathcal{W}_k$). Moreover, in [33, Lemma 2] it is shown that for $k \in \{1, 2, \ldots, m\}$,

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

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]. □

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

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

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

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

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

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

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

The dual (D) is equivalent to

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

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

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

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. □

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

LEMMA 4.3.

1. *If $D \subset S^+$, then $\mathcal{F}(D) = 0 \times K$, where $K \lhd \mathcal{P}$.*
2. *If $D \subset S$, then $\mathcal{F}(D) = \Re^m \times K$, where $K \lhd \mathcal{P}$.*

Proof. The statements follow from the definitions. □

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*

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

Then

$$K = ((\mathcal{F}(D))^c)^\perp$$
$$= \left\{ \begin{pmatrix} x \\ W + W^t \end{pmatrix} : \begin{bmatrix} I & W^t \\ W & U \end{bmatrix} \succeq 0 \text{ for some } \begin{pmatrix} y \\ U \end{pmatrix} \in D \right\}.$$

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 $0^m \times \mathcal{P}$. 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

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

Now

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

implies that

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

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 VV^t, this implies that $\alpha U \succeq VV^t$ for some $\alpha > 0$ large enough; i.e., $V \in K$. □

Now define the following sets:

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

The extended Lagrange–Slater dual of the dual (D) can now be stated.

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

Step 1

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

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

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

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

Therefore,

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

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.

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

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

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).

$$(\text{ELSDD}_1) \qquad \begin{array}{rll} p_1^* = & \max & c^t x \\ & \text{s.t.} & Ax + (Z + Z^t) \preceq b \\ & & Ay \succeq 0, c^t y = 0 \\ & & \begin{bmatrix} I & Z^t \\ Z & Ay \end{bmatrix} \succeq 0. \end{array}$$

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

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

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

$$(\text{RED}_2) \qquad \begin{array}{rll} d^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & A^*U = c \\ & & U \succeq_{\mathcal{P}_2} 0 \\ & \text{or} & G^*U \succeq_{T_2} \begin{pmatrix} c \\ 0 \end{pmatrix}. \end{array}$$

$$(\text{DRED}_2) \qquad \begin{array}{rll} p_2^* = & \max & c^t x \\ & \text{subject to} & Ax \preceq_{(\mathcal{P}_2)^+} b \\ & & x \in \Re^m \\ & \text{or} & G\phi =_{T_2^+} b,\ \phi \succeq_{T_2^+} 0. \end{array}$$

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$_2$).

$$(\text{ELSDD}_2) \qquad \begin{array}{rll} p_2^* = & \max & c^t x \\ & \text{s.t.} & Ax + (Z + Z^t) \preceq b \\ & & Ay + (Z + Z^t) \succeq 0, c^t y = 0 \\ & & \begin{bmatrix} I & Z^t \\ Z & Ay \end{bmatrix} \succeq 0 \\ & & Ay_1 + (Z_1 + Z_1^t) \succeq 0, c^t y = 0 \\ & & \begin{bmatrix} I & Z_1^t \\ Z_1 & Ay_1 \end{bmatrix} \succeq 0. \end{array}$$

. . . **Step k** . . .

5. Homogenization. In section 3.1.1, we have shown that an 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 [33, Example 1]. Moreover, the duality gap may be 0, but the dual may not be attained, see [33, 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 ($\Re^n_+$ arises as the direct sum of n 1×1 semidefinite cones), 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 by using the homogenized semidefinite program (assume the optimal objective function value p^* is known):

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

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 $B(\alpha w) = 0$ for all $\alpha \in \Re$, and

$$w = \begin{pmatrix} x \\ 0 \\ Z \end{pmatrix}.$$

Therefore, $c^t x > 0$ implies that $p^* = \infty$ (since there exists x such that $Ax \preceq 0$, $c^t x > 0$ implies (P) is unbounded). 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 the *polar* of a set C:

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

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 that

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

This yields the asymptotic optimality conditions (up to closure):

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

where the adjoint operator

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

and the polar cone

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

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; i.e., $\mathcal{P} = \mathcal{P}^+$. Note that if the closure in (5.3) is not needed, then these optimality conditions, along with weak duality for (P) and (D), $p \leq \text{trace}\, bU$, yield optimality conditions for (P); i.e., (5.3) with closure is equivalent to

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

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

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

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

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 way of generating examples where strong duality fails; i.e., we need to find examples where the right-hand side of (5.5) is not closed, and then we can 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 $x + C \subset C$.) 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 find sets, T, to add to attain the closure. Equivalently, find sets, C, $C^+ = T$, to intersect with K to attain the closure so that

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

On the other hand, note that the following is always true:

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

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. Some SDPs can be decomposed 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. (For a similar result, see, for instance, Boyd et al. [14, pp. 31–32].)

Suppose both problems (P) and (D) have feasible solutions (so that if there is a duality gap then it is finite). Consider the set

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

If $\mathcal{Z} \cap int(\mathcal{P}) \neq \emptyset$ then we have an interior point and strong duality holds for the Lagrangian dual. 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: 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: Block i is not an LP, but (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: Block i is not an LP, but conditions (5.3) and (5.5) both 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 the case where the objective function is separable with respect to this partition, 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). The reader may find it useful to generate examples by taking direct sums of examples from Freund [17] and Ramana [32].

Finally, we make some remarks about the ramifications of these results. We assumed throughout that (P) is feasible. Under this assumption, (ELSD) is feasible if and only if $p^* < +\infty$. If we also assume that $p^* < +\infty$, then we have $d^* = p^*$ (here, d^* is the optimal value of (ELSD)) and d^* is attained. We showed that in the dual problem (ELSD), the set W_m is precisely the subspace $(\mathcal{P}^f)^\perp$. Let us consider the following family of problems parameterized by $M > 0$:

$$(\tilde{\mathrm{P}}_{\mathrm{M}}) \qquad \begin{array}{cc} \sup & c^t x \\ \text{subject to} & Ax \preceq b \\ & A^*(I)^t x \leq M - \operatorname{trace}(b) \\ & x \in \Re^m, \end{array}$$

$$(\mathrm{ELS\tilde{D}_M}) \qquad \begin{array}{cc} \min & \operatorname{trace} b(U+W) + (M - \operatorname{trace}(b))z \\ \text{subject to} & A^*(U+W) - zA^*(I) = c \\ & W \in \mathcal{W}_m = (\mathcal{P}^f)^\perp \\ & U \succeq 0,\ z \geq 0. \end{array}$$

PROPOSITION 5.1. *Suppose* (P) *is feasible and* $p^* < +\infty$. *Then there exists a feasible solution* $(\tilde{U}, \tilde{W}, \tilde{z})$ *of* $(\mathrm{ELS\tilde{D}_M})$ *such that* $\tilde{U} \succ 0,\ \tilde{z} > 0$. *Moreover, for a given* M*, there exist optimal solutions of* (P) *with* $\operatorname{trace}(b - Ax) \leq M$ *if and only if there exist optimal solutions of* $(\tilde{\mathrm{P}}_{\mathrm{M}})$ *and every optimal solution of* $(\tilde{\mathrm{P}}_{\mathrm{M}})$ *is an optimal solution of* (P).

Proof. We apply the strong duality theorem to the pair (P) and (ELSD) to establish the existence of $(\bar{U}, \bar{W})$ such that $A^*(\bar{U} + \bar{W}) = c$, $\bar{W} \in \mathcal{W}_m$, and $\bar{U} \succeq 0$. Now, defining $\tilde{U} := \bar{U} + I$, $\tilde{W} := \bar{W}$, $\tilde{z} := 1$, we see that the first part of the proposition is proved. The second part of the proposition easily follows from the definition of $(\tilde{\mathrm{P}}_{\mathrm{M}})$. ◻

6. Conclusion. In this paper we have studied dual programs that guarantee strong duality for SDP. In particular, we have seen the relationships that exist between (DRP) (the dual of the regularized primal program (RP)) and (ELSD) (the extended

Lagrange–Slater dual). (DRP) uses the minimal cone $\mathcal{P}^f$ which, in general, cannot be computed exactly. (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., [39]. These problems arise from relaxations of 0,1 combinatorial optimization problems with linear constraints. In fact, for these problems, the Slater 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, one can question whether (D) is still 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 when 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 (i.e., not if we replace $Z \succeq 0$ by a nonnegativity constraint on the smallest eigenvalue $\lambda_{\min}(Z) \geq 0$).

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. For 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). In [33] it was shown, by using (ELSD), that SDFP is not NP-complete unless NP=Co-NP.

It may be interesting to try to interpret the significance of (ELSD) in terms of the computational complexity of solving SDPs which do not satisfy the Slater CQ. We do have a dual program, (ELSD), that can be written down in polynomial time. However, we still do not know how to solve (P) and (ELSD) in polynomial time by a symmetric, primal–dual interior-point algorithm.

Acknowledgments. The second and third authors thank the National Science and Engineering Research Council Canada for their support.

REFERENCES

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM J. Optim., 5 (1995), pp. 13–51.

[2] E. Anderson and P. Nash, *Linear Programming in Infinite Dimensional Spaces*, John Wiley and Sons, New York, 1987.

[3] G. P. Barker, *The lattice of faces of a finite dimensional cone*, Linear Algebra Appl., 7 (1973), pp. 71–82.

[4] G. P. Barker and D. Carlson, *Cones of diagonally dominant matrices*, Pacific J. Math., 57 (1975), pp. 15–32.

[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, VA, 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 (1969), pp. 367–389.

[7] A. Ben-Tal and A. Nemirovskii, *Potential reduction polynomial time method for truss topology design*, SIAM J. Optim., 4 (1994), pp. 596–612.

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

[9] F. Bohnenblust, *Joint Positiveness of Matrices*, California Institute of Technology, Pasadena, CA, 1948, manuscript.

[10] J. M. Borwein and H. Wolkowicz, *Characterizations of optimality for the abstract convex program with finite dimensional range*, J. Austral. Math. Soc. Ser. A, 30 (1981), pp. 390–411.

[11] J. M. Borwein and H. Wolkowicz, *Regularizing the abstract convex program*, J. Math. Anal. Appl., 83 (1981), pp. 495–530.

[12] J. M. Borwein and H. Wolkowicz, *Characterizations of optimality without constraint qualification for the abstract convex program*, Math. Programming Study, 19 (1982), pp. 77–100.

[13] J. M. Borwein and H. Wolkowicz, *A simple constraint qualification in infinite dimensional programming*, Math. Programming, 35 (1986), pp. 83–96.

[14] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*, in Studies in Applied Mathematics 15, SIAM, Philadelphia, PA, 1994.

[15] C. Delorme and S. Poljak, *The performance of an eigenvalue bound on the max-cut problem in some classes of graphs*, Discrete Math., 111 (1993), pp. 145–156.

[16] R. J. Duffin, *Infinite programs*, in Linear Equalities and Related Systems, A. W. Tucker, ed., Princeton University Press, Princeton, NJ, 1956, pp. 157–170.

[17] R. M. Freund, *Complexity of an Algorithm for Finding an Approximate Solution of a Semidefinite Program with No Regularity Assumption*, Technical report OR 302-94, MIT, Cambridge, MA, 1994.

[18] K. Glashoff and S. Gustafson, *Linear optimization and approximation*, in Applied Mathematical Sciences 45, 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. Wolkowicz, *Positive definite completions of partial Hermitian matrices*, Linear Algebra Appl., 58 (1984), pp. 109–124.

[21] M. Guignard, *Generalized Kuhn–Tucker conditions for mathematical programming problems in a banach space*, SIAM J. Control, 7 (1969), pp. 232–241.

[22] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz, *An interior point method for semidefinite programming*, SIAM J. Optim., 6 (1996), pp. 342–361.

[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*, Comput. Optim. Appl., to appear.

[25] K. Kretschmer, *Programming in paired spaces*, Canad. J. Math., 13 (1961), pp. 221–238.

[26] N. Levinson, *A class of continuous linear programming problems*, J. Math. Anal. Appl., 16 (1966), pp. 73–83.

[27] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and* 0–1 *optimization*, SIAM J. Optim., 1 (1991), pp. 166–190.

[28] K. Löwner, *Uber monotone matrixfunctionen*, Math. Z., 49 (1934), pp. 375–392.

[29] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, NY, 1979.

[30] Y. E. Nesterov and A. S. Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming: Theory and Algorithms*, SIAM, Philadelphia, PA, 1994.

[31] F. Pukelsheim *Optimal Design of Experiments*, Wiley, New York, 1993.

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

[33] M. V. Ramana, *An Exact Duality Theory for Semidefinite Programming and its Complexity Implications*, DIMACS Technical report 95-02R, RUTCOR, Rutgers University, New

Brunswick, NJ, 1995.
[34] T. W. Reiland, *Optimality conditions and duality in continuous programming.* ii. *The linear problem revisited*, J. Math. Anal. Appl., 77 (1980), pp.329–343.
[35] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[36] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.
[37] H. Wolkowicz, *Geometry of optimality conditions and constraint qualifications: The convex case*, Math. Programming, 19 (1980), pp. 32–60.
[38] H. Wolkowicz, *Some applications of optimization in matrix theory*, Linear Algebra Appl., 40 (1981), pp. 101–118.
[39] Q. Zhao, S. Karisch, F. Rendl, and H. Wolkowicz, *Semidefinite Programming Relaxations for the Quadratic Assignment Problem*, Research report, University of Waterloo, Waterloo, Ontario, Canada, 1995.