Semidefinite Optimization *

M. J. Todd †

August 22, 2001

Abstract

Optimization problems in which the variable is not a vector but a symmetric matrix which is required to be positive semidefinite have been intensely studied in the last ten years. Part of the reason for the interest stems from the applicability of such problems to such diverse areas as designing the strongest column, checking the stability of a differential inclusion, and obtaining tight bounds for hard combinatorial optimization problems. Part also derives from great advances in our ability to solve such problems efficiently in theory and in practice (perhaps "or" would be more appropriate: the most effective computational methods are not always provably efficient in theory, and vice versa). Here we describe this class of optimization problems, give a number of examples demonstrating its significance, outline its duality theory, and discuss algorithms for solving such problems.

†School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853, USA (miketodd@cs.cornell.edu). Research supported in part by NSF through grant DMS-9805602 and ONR through grant N00014-96-1-0050.

1 Introduction

Semidefinite optimization is concerned with choosing a symmetric matrix to optimize a linear function subject to linear constraints and a further crucial constraint that the matrix be positive semidefinite. It thus arises from the well-known linear programming problem by replacing the vector of variables with a symmetric matrix and replacing the nonnegativity constraints with a positive semidefinite constraint. (An alternative way to write such a problem is in terms of a vector of variables, with a linear objective function and a constraint that some symmetric matrix that depends affinely on the variables be positive semidefinite.) This generalization nevertheless inherits several important properties from its vector counterpart: it is convex, has a rich duality theory (although not as strong as linear programming's), and admits theoretically efficient solution procedures based on iterating interior points to either follow the central path or decrease a potential function. Here we will investigate this class of problems and survey the recent results and methods obtained.

While linear programming (LP) as a subject grew very fast during the '50s and '60s, due to the availability of the very efficient simplex method of G.B. Dantzig, semidefinite optimization (also known as semidefinite programming or SDP, the term we shall use) was slower to attract as much attention. Partly this was because, since the feasible region is no longer polyhedral, the simplex method was not applicable, although related methods do exist. As soon as theoretically efficient (as well as practically useful) algorithms became available in the late '80s and '90s, research in the area exploded. The recent Handbook of Semidefinite Programming [67] lists 877 references, while the online bibliography on semidefinite programming collected by Wolkowicz [66] lists 722, almost all since 1990.

The development of efficient algorithms was only one trigger of this explosive growth: another key motivation was the power of SDP to model problems arising in a very wide range of areas. We will describe some of these applications in Section 3, but these only cover part of the domain. The handbook [67] has chapters on applications in combinatorial optimization, on nonconvex quadratic programming, on eigenvalue and nonconvex optimization, on systems and control theory, on structural design, on matrix completion problems, and on problems in statistics.

Bellman and Fan seem to have been the first to formulate a semidefinite programming problem, in 1963. Instead of considering a linear programming problem in vector form and replacing the vector variable with a matrix variable, they started with a scalar LP formulation and replaced each scalar variable with a matrix. The resulting problem (although equivalent to the general formulation) was somewhat cumbersome, but they derived a dual problem and established several key duality theorems, showing that additional regularity is needed in the SDP case to prove strong duality. However, the importance of constraints requiring that a certain matrix be positive (semi)definite had been recognised much earlier in control theory: Lyapunov's characterization of the stability of the solution of a linear differential equation in 1890 involved just such a constraint (called a linear matrix inequality, or LMI), and subsequent work of Luré, Postnikov, and Yakubovich in the Soviet Union in the '40s, '50s, and '60s established the importance of LMIs in control theory (see Boyd at al. [9]). In the early '70s, Donath and Hoffman [13] and then Cullum, Donath, and Wolfe [12] showed that some hard graph-partitioning problems could be attacked by considering a related eigenvalue optimization problem – as we shall see, these are closely connected with SDP. Then in 1979, Lovász [35] formulated an SDP problem that provided a bound on the Shannon capacity of a graph and thereby found the capacity of the pentagon, solving a long-open conjecture. At that time, the most efficient method known for SDP problems was the ellipsoid method, and Grötschel, Lovász, and Schrijver [24] investigated in detail its application to combinatorial optimization problems by using it to approximate the solution of both LP and SDP relaxations. Lovász and Schrijver [36] later showed how SDP problems can provide tighter relaxations of $(0,1)$-programming problems than can LP.

Fletcher [17, 18] revived interest in SDP among nonlinear programmers in the '80s, and this led to a series of papers by Overton and Overton and Womersley; see [50] and the references therein. The key contributions of Nesterov and Nemirovski [44, 45] and Alizadeh [1] showed that the new generation of interior-point methods pioneered by Karmarkar [30] for LP could be extended to SDP. In particular, Nesterov and Nemirovski established a general framework for solving nonlinear convex optimization problems in a theoretically efficient way using interior-point methods, by developing the powerful theory of self-concordant barrier functions. These works led to the huge recent interest in semidefinite programming, which was further increased by the result of Goemans and Williamson [22] which showed that an SDP relaxation could provide a provably good approximation to the max-cut problem in combinatorial optimization.

Our coverage will necessarily be incomplete and biased. Let us therefore refer the reader to a survey paper by Vandenberghe and Boyd [63] which discusses in particular a number of applications, especially in control theory; the book of Boyd et al. which describes the latter in much further detail and gives the history of SDP in control theory; the excellent paper of Lewis and Overton [34] in this journal on the very closely related topic of eigenvalue optimization; and the aforementioned handbook edited by Wolkowicz et al. [67]. We also mention that SDP is both an extension of LP and a special case of more general conic optimization problems. Nesterov and Nemirovski [44, 45] consider general convex cones, with the sole proviso that a self-concordant barrier is known for the cone. Nesterov and Todd [46, 47] consider the subclass of self-scaled cones, which admit symmetric primal-dual algorithms (these cones turn out to coincide with symmetric (homogeneous self-dual) cones). Another viewpoint is that of Euclidean Jordan Algebras, developed by Faybusovich [15, 16] and now investigated by a number of authors: see Alizadeh and Schmieta [5]. Since the area is receiving so much attention, it is hard to keep abreast of recent developments, but this is immeasurably assisted by three web sites, those of Helmberg [25] and Alizadeh [2] on semidefinite programming, and that of Wright [68] on interior-point methods. The latter also allows one to sign up for the interior-point methods mailing list, where almost all papers addressing interior-point methods for SDP are announced.

The rest of the paper is organised as follows. In the next section, we define the SDP problem in both primal and dual form and introduce some useful notation for expressing it. We also establish weak duality. Then Section 3 gives nine examples of the application of SDP to diverse areas; along the way, we list a number of useful facts about symmetric matrices that allow this development. The following section is devoted to duality, and presents some examples demonstrating the anomalies that can occur; then conditions sufficient for strong duality to hold are established. Section 5 introduces the very important logarithmic barrier function for the cone of positive semidefinite matrices and uses it to define, and then prove some important properties of, the central path. Then in Section 6 we consider path-following and potential-reduction algorithms and also methods based on nonlinear programming reformulations of the SDP problem. Section 7 contains some concluding remarks.

Notation. Most matrices occurring in this paper will be real symmetric matrices of order n: we let $S\mathbb{R}^{n\times n}$ denote the space of such matrices. $U \bullet V$ denotes the inner product between two such matrices, defined by trace (U^TV) (the transpose makes this valid for nonsymmetric and even nonsquare matrices also). The associated norm is the Frobenius norm, written $\|U\|_F := (U \bullet U)^{\frac{1}{2}}$ or just $\|U\|$, while $\|P\|_2$ denotes the L_2-operator norm of a matrix. Norms on vectors will always be Euclidean unless otherwise noted.

We write $U \succeq 0$ to mean that U is positive semidefinite. Similarly, $U \succ 0$ indicates that U is positive definite, and these terms always refer to symmetric matrices unless there is an explicit statement otherwise. We write $S\mathbb{R}^{n\times n}_+$ ($S\mathbb{R}^{n\times n}_{++}$) to denote the set of positive semidefinite (positive definite) symmetric matrices of order n. We use $U \preceq V$ or $V \succeq U$ to mean $V - U \succeq 0$, and $U \prec V$ and $V \succ U$ similarly mean $V - U \succ 0$. If $U \succeq 0$, we write $U^{\frac{1}{2}}$ for the (symmetric) positive semidefinite square root of U.

We write $\mathrm{diag}(U)$ for the vector of diagonal entries of $U \in S\mathbb{R}^{n\times n}$, and $\mathrm{Diag}(u)$ for the diagonal matrix with the vector $u \in \mathbb{R}^n$ on its diagonal. We extend this to general block diagonal matrices: if U_1, U_2, ..., U_k are symmetric matrices, then $\mathrm{Diag}(U_1, U_2, \ldots, U_k)$ denotes the block diagonal matrix with the U_i's down its diagonal.

As is customary, lower-case Roman letters usually denote vectors and upper-case letters $n \times n$ matrices; we reserve K, L, P, and Q (Q will usually be orthogonal) for not necessarily symmetric matrices, with all other letters denoting members of $S\mathbb{R}^{n\times n}$. We use lower-case Greek letters for scalars, and script letters for linear operators on (usually symmetric) matrices. We introduce the useful notation $P \odot Q$ for $n\times n$ matrices P and Q (usually P and Q are symmetric). This is an operator from $S\mathbb{R}^{n\times n}$ to itself defined by

$$(P \odot Q)U := \frac{1}{2}(PUQ^T + QUP^T). \tag{1}$$

2 Problems

The SDP problem in *primal standard form* is

$$\begin{array}{llrcl} (P) & \min_X & C \bullet X & & \\ & & A_i \bullet X & = & b_i, \quad i = 1, \ldots, m \\ & & X & \succeq & 0, \end{array}$$

where all $A_i \in S\mathbb{R}^{n\times n}$, $b \in \mathbb{R}^m$, $C \in S\mathbb{R}^{n\times n}$ are given, and $X \in S\mathbb{R}^{n\times n}$ is the variable. We also consider SDP problems in *dual standard form*:

$$\begin{array}{lllrcl} (D) & \max_{y,S} & b^T y & & & \\ & & \sum_{i=1}^m y_i A_i & + \; S & = & C \\ & & & S & \succeq & 0, \end{array}$$

where $y \in \mathbb{R}^m$ and $S \in S\mathbb{R}^{n\times n}$ are the variables. This can also be written as

$$\max_y b^T y, \quad \sum_{i=1}^m y_i A_i \preceq C,$$

or

$$\max_y b^T y, \quad C - \sum_{i=1}^m y_i A_i \succeq 0,$$

but we shall see the benefit of having the "slack matrix" S available when we discuss algorithms.

We should strictly write "inf" and "sup" instead of "min" and "max" above, not just because the problems might be unbounded, but also because even if the optimal values are finite they might not be attained. We stick to "min" and "max" both to highlight the fact that we are interested in optimal solutions, not just values, and because we shall often impose conditions that ensure that the optimal values are in fact attained where finite.

The last form of the problem in dual standard form shows that we are trying to optimize a linear function of several variables, subject to the constraint that a symmetric matrix that depends affinely on the variables is restricted to be positive semidefinite. (Henceforth, as is common in mathematical programming, we use "linear" to mean "affine" in most cases: however, linear operators will always be linear, not affine.) We will encounter several examples of such problems, and will not see the need to express them explicitly in the form above, but it is straightforward to do so.

We have been somewhat coy in referring to the problems above as SDP problems in primal and dual form respectively. If they are defined by the same data A_i, $i = 1, \ldots, m$, b, and C, they are in fact dual problems, and have a beautiful theory that will be studied in Section 4. However, we find it useful to discuss some examples before we investigate duality in detail. Here we just note the following trivial but key fact:

Proposition 2.1 *(Weak Duality) If X is feasible in (P) and (y, S) in (D), then*

$$C \bullet X - b^T y = X \bullet S \geq 0. \tag{2}$$

Proof:
We find

$$C \bullet X - b^T y = (\sum_{i=1}^m y_i A_i + S) \bullet X - b^T y = \sum_{i=1}^m (A_i \bullet X) y_i + S \bullet X - b^T y = S \bullet X = X \bullet S.$$

Moreover, since X is positive semidefinite, it has a square root $X^{\frac{1}{2}}$, and so $X \bullet S = \mathrm{trace}\,(XS) = \mathrm{trace}\,(X^{\frac{1}{2}} X^{\frac{1}{2}} S) = \mathrm{trace}\,(X^{\frac{1}{2}} S X^{\frac{1}{2}}) \geq 0$. Here we used the facts that trace (PQ) = trace (QP), that $X^{\frac{1}{2}}SX^{\frac{1}{2}}$ is positive semidefinite since S is (from the definition), and that the trace of a positive semidefinite matrix is nonnegative (as the sum of its nonnegative diagonal elements or the sum of its nonnegative eigenvalues). $\square$

It is convenient to introduce some notation to make the problems above easier to state. We define the linear operator $\mathcal{A} : S\mathbb{R}^{n\times n} \to \mathbb{R}^m$ by

$$\mathcal{A}X := (A_i \bullet X)_{i=1}^m \in \mathbb{R}^m.$$

Note that, for any $X \in S\mathbb{R}^{n\times n}$ and $v \in \mathbb{R}^m$, $(\mathcal{A}X)^T v = \sum_{i=1}^m (A_i \bullet X) v_i = (\sum_{i=1}^m v_i A_i) \bullet X$, so the adjoint of $\mathcal{A}$ is given by

$$\mathcal{A}^* v = \sum_{i=1}^m v_i A_i,$$

a mapping from $\mathbb{R}^m$ to $S\mathbb{R}^{n\times n}$. Using this notation, we can rewrite our problems as

$$(P) \qquad \min \quad C \bullet X, \qquad \mathcal{A}X = b, \quad X \succeq 0,$$

and

$$(D) \qquad \max \quad b^T y, \qquad \mathcal{A}^* y + S = C, \quad S \succeq 0.$$

The weak duality chain of equations can then be written as

$$C \bullet X - b^T y = (\mathcal{A}^* y + S) \bullet X - b^T y = (\mathcal{A}X)^T y + S \bullet X - b^T y = X \bullet S.$$

We call the difference between the optimal value of (P) and that of (D), which is always nonnegative by the result above, the *duality gap*. Strong duality is the assertion that the duality gap is zero and both problems attain their optima whenever both problems are feasible, but it does not always hold for SDP problems. We investigate this in detail in Section 4.

3 Examples

In this section we present a number of examples of SDP problems. In order to do so, we also introduce some simple facts about symmetric matrices. Here is our first example:

Example 1: minimizing the maximum eigenvalue. This problem arises in stabilizing a differential equation, for instance. Suppose we have a symmetric matrix, say $M(z)$, depending linearly (affinely) on a vector z. We wish to choose z to minimize the maximum eigenvalue of $M(z)$. Note that $\lambda_{\max}(M(z)) \le \eta$ iff $\lambda_{\max}(M(z) - \eta I) \le 0$, or equivalently iff $\lambda_{\min}(\eta I - M(z)) \ge 0$. This holds iff $\eta I - M(z) \succeq 0$. So we get the SDP problem in dual form:

$$\max \ -\eta, \quad \eta I - M(z) \succeq 0, \tag{3}$$

where the variable is $y := (\eta; z)$.

To introduce other examples, we need to use a collection of very handy tools concerning symmetric matrices. We list these below (usually) without proof, but most are not hard to show.

Fact 1. If $P \in I\!R^{m\times n}$ and $Q \in I\!R^{n\times m}$, then $\mathrm{trace}\,(PQ) = \mathrm{trace}\,(QP)$.

Fact 2. $\mathcal{A}$ and $\mathcal{A}^*$ above are adjoints.

Fact 3. If $U, V \in SI\!R^{n\times n}$, and Q is orthogonal, then $U \bullet V = (Q^T UQ) \bullet (Q^T V Q)$. More generally, if P is nonsingular, $U \bullet V = (PUP^T) \bullet (P^{-T} V P^{-1})$.

Fact 4. Every $U \in SI\!R^{n\times n}$ can be written as $U = Q\Lambda Q^T$, where Q is orthogonal and Λ is diagonal. Then $UQ = Q\Lambda$, so the columns of Q are the eigenvectors, and the diagonal entries of Λ the corresponding eigenvalues of U. We write $Q(U) := Q$, $\Lambda(U) := \Lambda$ and $\lambda(U) := \mathrm{diag}(\Lambda)$. (Together with Fact 3, this means that we can often assume that one symmetric matrix under study is diagonal, which can simplify some proofs.)

Fact 5. The following are norms on $SI\!R^{n\times n}$: $\|\lambda(U)\|_2 = \|U\|_F$, $\|\lambda(U)\|_\infty = \|U\|_2$, and $\|\lambda(U)\|_1$. If $U \succeq 0$, $\|\lambda(U)\|_1 = \sum_j |\lambda_j(U)| = \sum_j \lambda_j(U) = I \bullet \Lambda(U) = \mathrm{trace}\,(U) = I \bullet U$.

Fact 6. For $U \in SI\!R^{n\times n}$, the following are equivalent:
a) $U \succeq 0$ ($U \succ 0$);
b) $v^T U v \geq 0$ for all $v \in I\!R^n$ ($v^T U v > 0$ for nonzero $v \in I\!R^n$);
c) $\lambda(U) \geq 0$ ($\lambda(U) > 0$); and
d) $U = P^T P$ for some matrix P ($U = P^T P$ for some square nonsingular matrix P).

Immediate corollaries are that $uu^T \succeq 0$ for all $u \in I\!R^n$, that every $U \succeq 0$ has a positive semidefinite square root $U^{\frac{1}{2}}$ (take $U^{\frac{1}{2}} = Q(U)\Lambda^{\frac{1}{2}}(U)Q^T(U)$, where $\Lambda^{\frac{1}{2}}(U)$ is the diagonal matrix whose diagonal contains the (nonnegative) square roots of the eigenvalues of U), and that if $U \succ 0$, then U is nonsingular, with $U^{-1} = Q(U)\Lambda^{-1}(U)Q^T(U)$. It also follows that $SI\!R^{n\times n}_+$ is a closed convex cone, pointed (i.e., $(SI\!R^{n\times n}_+) \cap (-SI\!R^{n\times n}_+) = \{0\}$) and with nonempty interior $SI\!R^{n\times n}_{++}$, an open convex cone. Finally, hence we get $\{(\eta; z) : \eta \geq \lambda_{\max}(M(z))\} = \{(\eta; z) : \eta I - M(z) \succeq 0\}$, as used above, and since this is a convex set, $\lambda_{\max}(M(\cdot))$ is a convex function.

Fact 7. If $U \succeq 0$, then each $u_{jj} \geq 0$, and if $u_{jj} = 0$, $u_{jk} = u_{kj} = 0$ for all k. Similarly, if $U \succ 0$, then each $u_{jj} > 0$.

Fact 8. If $U \succeq 0$, then $PUP^T \succeq 0$ for any P of appropriate column dimension. If P is square and nonsingular, then $U \succ 0$ iff $PUP^T \succ 0$.

Fact 9. If $U = \begin{pmatrix} U_{11} & U_{12} \\ U_{12}^T & U_{22} \end{pmatrix} \succeq 0$ ($\succ 0$), then $U_{11} \succeq 0$ ($\succ 0$). Using Fact 8 with P a permutation matrix, we see that every principal submatrix of a positive semidefinite (definite) matrix is also positive semidefinite (definite).

Fact 10. $U \succeq 0$ ($\succ 0$) iff every principal minor is nonnegative (positive). In fact, $U \succ 0$ iff every leading principal minor is positive. Also, $U \succ 0$ iff $U = LL^T$ for some nonsingular lower triangular matrix L (the Cholesky factorization).

We can prove Fact 10 using the preceding facts, induction, and the following very useful property:

Fact 11. Suppose $U = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ with A and C symmetric and $A \succ 0$. Then

$$U \succeq 0 \quad (\succ 0) \qquad \text{iff} \qquad C - B^T A^{-1} B \succeq 0 \quad (\succ 0).$$

The matrix $C - B^T A^{-1} B$ is called the *Schur complement* of A in U. This is easily proved using the factorization

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^T A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}.$$

Fact 12. (Representing quadratics) If $U \in S\mathbb{R}^{n\times n}$, then $x^T U x = U \bullet xx^T$.

Fact 13. (Self-duality) $S\mathbb{R}^{n\times n}_+ = (S\mathbb{R}^{n\times n}_+)^* := \{V : U \bullet V \geq 0 \text{ for all } U \in S\mathbb{R}^{n\times n}_+\}$.

Proof:
(i) $S\mathbb{R}^{n\times n}_+ \subseteq (S\mathbb{R}^{n\times n}_+)^*$: We want to show that $U \bullet V \geq 0$ for all positive semidefinite U and V. We can show this directly using Facts 3 and 4 to assume that one is diagonal, or use Fact 6 to obtain a square root of U, and then note that $U \bullet V = \operatorname{trace} UV = \operatorname{trace} U^{\frac{1}{2}} V U^{\frac{1}{2}} \geq 0$ since $U^{\frac{1}{2}} V U^{\frac{1}{2}}$ is positive semidefinite.
(ii) $S\mathbb{R}^{n\times n}_+ \subseteq (S\mathbb{R}^{n\times n}_+)^*$: We show that if $U \notin S\mathbb{R}^{n\times n}_+$, then $U \notin (S\mathbb{R}^{n\times n}_+)^*$. Indeed, in this case we have $v^T U v < 0$ for some $v \in \mathbb{R}^n$, and then $U \bullet vv^T < 0$ shows that U is not in $(S\mathbb{R}^{n\times n}_+)^*$. ∎

Fact 14. If $U \succ 0$, then $U \bullet V > 0$ for every nonzero $V \succeq 0$, and $\{V \succeq 0 : U \bullet V \leq \beta\}$ is bounded for every positive β. Indeed, if $\lambda := \lambda_{\min}(U) > 0$, then $U \bullet V = (U - \lambda I) \bullet V + \lambda I \bullet V \geq \lambda I \bullet V = \lambda I \bullet \Lambda(V) = \lambda \|\lambda(V)\|_1 \geq \lambda \|V\|_F$ for $V \succeq 0$. This shows the first part directly, and the second since then any V in the set has Frobenius norm at most β/λ.

Fact 15. If $U, V \succeq 0$, then $U \bullet V = 0$ iff $UV = 0$. This is easy to show using the eigenvalue decomposition of U, and considering separately its positive and zero eigenvalues.

Fact 16. If $U, V \in S\mathbb{R}^{n\times n}$, then U and V commute iff UV is symmetric, iff U and V can be simultaneously diagonalised (i.e., they have eigenvalue decompositions with the same Q).

We can now return to considering other examples of semidefinite programming problems.

Example 2: minimizing the L_2-operator norm of a matrix. By considering the two cases where $P \in \mathbb{R}^{m\times n}$ is zero and nonzero, and using Fact 11, we can easily see that $\eta \geq \|P\|_2$ iff $\begin{pmatrix} \eta I & P \\ P^T & \eta I \end{pmatrix} \succeq 0$. Hence we can solve the problem of minimizing $\|P(z)\|_2$, where $P(z)$ depends affinely on z, by solving the SDP problem

$$\max \quad -\eta, \qquad \begin{pmatrix} \eta I & P(z) \\ P(z)^T & \eta I \end{pmatrix} \succeq 0, \tag{4}$$

where the variable is $y := (\eta; z)$.

Example 3: LP. The linear programming $\max\{b^T y : A^T y \le c\}$, where $A \in I\!R^{m\times n}$ and the vectors have appropriate dimensions, can be written as the SDP problem in dual form

$$\max \quad b^T y, \qquad \mathrm{Diag}(c - A^T y) \succeq 0.$$

Here, in our standard notation, $C = \mathrm{Diag}(c)$ and $A_i = \mathrm{Diag}(a_i)$, with a_i the ith column of A^T. Note that its semidefinite dual problem involves a symmetric $n \times n$ matrix X, and hence seems to differ from the usual linear programming dual. We will discuss this further very shortly.

Example 4: a quasi-convex nonlinear programming problem. Consider now the problem

$$\min \quad \frac{(b^T y)^2}{d^T y}, \qquad A^T y \le c,$$

where we assume that $d^T y > 0$ for all feasible y. If we note that the objective function can be written as $(b^T y)(d^T y)^{-1}(b^T y)$ we see the resemblance to the Schur complement, and then it is easy to check that (for feasible y),

$$\eta \ge \frac{(b^T y)^2}{d^T y} \qquad \text{iff} \qquad \begin{pmatrix} \eta & b^T y \\ b^T y & d^T y \end{pmatrix} \succeq 0.$$

It follows that our nonlinear programming problem can be written as

$$\max_{\eta, y} \quad -\eta, \quad \mathrm{Diag}(c - A^T y) \succeq 0, \quad \begin{pmatrix} \eta & b^T y \\ b^T y & d^T y \end{pmatrix} \succeq 0.$$

This has two semidefinite constraints, but of course they can be combined into a single constraint:

$$\mathrm{Diag}\left(\mathrm{Diag}(c - A^T y), \begin{pmatrix} \eta & b^T y \\ b^T y & d^T y \end{pmatrix}\right) \succeq 0.$$

Here C and the A_i's are all block diagonal, with m 1×1 blocks and one 2×2 block.

In the last two examples we have seen cases where the data matrices C and the A_i's share the same block diagonal structure. Indeed, as in the last example, this arises whenever several semidefinite constraints are combined into a single constraint. Let $\mathcal{S}$ denote the space of block diagonal symmetric matrices of the form

$$M = \begin{pmatrix} M_{11} & 0 & \cdots & 0 \\ 0 & M_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{kk} \end{pmatrix},$$

where $M_{jj} \in I\!R^{n_j \times n_j}$ for $j = 1, \ldots, k$. Let us suppose C and all A_i's lie in $\mathcal{S}$. Then any feasible S in the dual problem, with $\mathcal{A}^* y + S = C$, also lies in $\mathcal{S}$. So (D) can alternatively be written as

$$\max_{(y,S) \in \boldsymbol{R}^m \times \mathcal{S}} \quad b^T y, \qquad \mathcal{A}^* y + S = C, \quad S \succeq 0.$$

Its dual is

$$\min_{X \in S\mathbb{R}^{n\times n}} \quad C \bullet X, \qquad \mathcal{A}X = b, \quad X \succeq 0;$$

can we restrict X also to $\mathcal{S}$? If so, then in the LP case, X will be block diagonal with 1×1 blocks, and thus we regain the usual LP dual. Consider any $X \in S\mathbb{R}^{n\times n}$, and partition it as M above:

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & \cdots & X_{kk} \end{pmatrix}.$$

Then, with obvious notation,

$$\begin{aligned} A_i \bullet X &= A_{i11} \bullet X_{11} + \cdots + A_{ikk} \bullet X_{kk} \quad \text{for each } i, \\ C \bullet X &= C_{11} \bullet X_{11} + \cdots + C_{kk} \bullet X_{kk}. \end{aligned}$$

Also, if $X \succeq 0$, then $X_{jj} \succeq 0$ for $j = 1, \ldots, k$, and then

$$\tilde{X} := \begin{pmatrix} X_{11} & 0 & \cdots & 0 \\ 0 & X_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{kk} \end{pmatrix} \succeq 0.$$

Hence, if X is feasible in (P), then so is $\tilde{X} \in \mathcal{S}$, and with the same objective value. It follows that we can restrict X to $\mathcal{S}$ without loss of generality.

It is important to realize that, if $\mathcal{S}$ denotes instead the set of symmetric matrices with a given sparsity structure and all A_i's and C lie in $\mathcal{S}$, then any feasible S also lies in $\mathcal{S}$ but it is no longer the case that we can restrict feasible X's to $\mathcal{S}$.

Let us mention two ways in which block diagonal structure arises in SDP problems in primal form. First, if we have inequality constraints like $A_i \bullet X \le b_i, i = 1, \ldots, m, X \succeq 0$, then we can add slack variables to reach $A_i \bullet X + \xi_i = b_i, i = 1, \ldots, m, X \succeq 0, \xi \ge 0$. But these can be written as equality constraints in the positive semidefinite variable $\tilde{X} := \mathrm{Diag}(X, \mathrm{Diag}(\xi))$, and then all matrices have the same block diagonal structure with one $n \times n$ block followed by m 1×1 blocks.

Similarly, if we have several matrix variables and our problem is

$$\begin{array}{lccccccl} \min & C_{11} \bullet X_{11} & + & \cdots & + & C_{kk} \bullet X_{kk}, & & \\ & A_{i11} \bullet X_{11} & + & \cdots & + & A_{ikk} \bullet X_{kk} & = & b_i, \quad i = 1, \ldots, m, \\ & X_{11} & \succeq & 0, & \ldots & X_{kk} & \succeq & 0, \end{array}$$

then we can express this as an SDP problem involving just one positive semidefinite variable

$$X := \begin{pmatrix} X_{11} & 0 & \cdots & 0 \\ 0 & X_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{kk} \end{pmatrix} \succeq 0,$$

and again we have common block diagonal structure in all the matrices.

All of our results below and all algorithms can exploit this block diagonal structure (and obviously must to be efficient), but for simplicity we write $S\mathbb{R}^{n\times n}$ as the matrix space henceforth.

Now we return to our examples of SDP problems with

Example 5: convex quadratically constrained programming. Here we consider optimizing a linear function subject to convex quadratic constraints (we can easily convert the minimization of a convex quadratic constraint subject to similar constraints to this form). So we address

$$\max \quad b^T y, \qquad f_i(y) \le 0, \quad i = 1, \ldots, l.$$

where $f_i(y) := y^T C_i y - d_i^T y - \epsilon_i, C_i \succeq 0, i = 1, \ldots, l$. Let $C_i = G_i^T G_i$. Then $f_i(y) \le 0$ can be written as

$$\begin{pmatrix} I & G_i y \\ (G_i y)^T & d_i^T y + \epsilon_i \end{pmatrix} \succeq 0$$

using Schur complements, or alternatively as

$$\begin{pmatrix} (1 + d_i^T y + \epsilon_i) I & \begin{pmatrix} 1 - d_i^T y - \epsilon_i \\ 2G_i y \end{pmatrix} \\ \begin{pmatrix} 1 - d_i^T y - \epsilon_i \\ 2G_i y \end{pmatrix}^T & 1 + d_i^T y + \epsilon_i \end{pmatrix} \succeq 0.$$

The advantage of the second formulation is that the semidefinite constraint $\begin{pmatrix} \alpha I & v \\ v^T & \alpha \end{pmatrix} \succeq 0$ can be expressed as $\begin{pmatrix} \alpha \\ v \end{pmatrix} \in K_2 := \left\{ \begin{pmatrix} \beta \\ w \end{pmatrix} : \beta \ge \|w\|_2 \right\}$, the second-order or Lorentz cone. This is a more efficient way to solve the problem — second-order cones are to be preferred to semidefinite cones in general: see Nesterov and Nemirovski [44, 45].

Example 6: robust mathematical programming. This is a way to model uncertainty in the data of an optimization problem (or in the implementation of a solution) by requiring that the solution be feasible whatever the realization of the data (see Ben-Tal and Nemirovski [7]). Without loss of generality we can assume that the objective function is deterministic. Let us consider robust LP with ellipsoidal uncertainty. The problem

$$\max \quad b^T y, \qquad a_j^T y \le c_j \text{ for all } (a_j; c_j) \in \mathcal{E}_j, \quad j = 1, \ldots, n,$$

can be rewritten, after introducing an extra variable and changing notation, as

$$\begin{array}{ll} \max & b^T y \\ & a_j^T y \le 0, \quad \text{for all } a_j \in \mathcal{E}_j \text{ and } j = 1, \ldots, k, \\ & a_j^T y \le 0, \quad \text{for } j = k+1, \ldots, n. \end{array}$$

Suppose $\mathcal{E}_j = \{\bar{a}_j + G_j u_j : \|u_j\|_2 \le 1\}$. Then, for a given vector y, we have $a_j^T y \le 0$ for all $a_j \in \mathcal{E}_j$ iff $\bar{a}_j^T y + (G_j u_j)^T y \le 0$ for all u_j of norm at most one, iff $\|G_j^T y\|_2 \le -\bar{a}_j^T y$

or $(-\bar{a}_j^T y; G_j^T y) \in K_2$, the second-order cone. So we can model the robust LP above using second-order cones, or if we wish as an SDP problem. Ben-Tal and Nemirovski discuss a number of other robust mathematical programming problems; for instance the robust version of the convex quadratically constrained programming problem above can be formulated as an SDP problem, see [7].

Example 7: control theory. There are many applications of semidefinite programming (or the feasibility version, called a linear matrix inequality in the field) in control systems. We will describe a very simple case, leaving the discussion of more general and realistic situations to Vandenberghe and Boyd [63] and Boyd et al. [9].

Suppose $x = x(t)$ satisfies the differential inclusion

$$\dot{x} \in \operatorname{conv}\{A_1, \ldots, A_m\}x, \quad x(0) = x_0,$$

where $A_1, \ldots, A_m$ are given matrices in $\mathbb{R}^{n \times n}$. We want to determine whether $x(t)$ necessarily remains bounded.

This holds iff there is some $P \succ 0$ so that $v(x) := x^T P x$ remains uniformly bounded, and this certainly follows if v is nonincreasing. Such a function is called a Lyapunov function. Hence a sufficient condition for uniform boundedness is that

$$\frac{d}{dt}(x^T P x) = \dot{x}^T P x + x^T P \dot{x} \leq 0.$$

If x_0 is arbitrary, and $\dot{x}(0)$ can be anywhere in the appropriate convex set, then we need

$$A_i^T P + P A_i \preceq 0, \quad \text{for all } i = 1, \ldots, m.$$

We also want $P \succ 0$, and since the constraints above are homogeneous, we may require $P \succeq I$. If we seek a matrix P with say minimum condition number, we are then led to the SDP problem

$$\begin{array}{ll} \max & -\eta \\ & A_i^T P + P A_i \preceq 0, \quad \text{for all } i = 1, \ldots, m \\ & \eta I \succeq P \succeq I, \end{array}$$

where the variables are η and the entries of the symmetric matrix P. Note that again we have block diagonal structure in this SDP problem.

Now we turn to applications of SDP in obtaining good relaxations of combinatorial optimization problems. Here relaxations mean optimization problems where the feasible region of the problem of interest is enlarged to obtain a tractable problem whose optimal value provides a bound for that of the problem of interest (in some cases, the optimal solution of the relaxed problem is also of great use). We discuss two: Lovász's theta function and the max-cut problem.

Example 8: Lovász's theta function [35]. Here we seek a bound on the Shannon capacity, or on the stability number, of an undirected graph $G = (N, E)$ with node set N and edge set E; we write ij instead of $\{i, j\}$ for an edge linking nodes i and j. We will assume that $N = \{1, \ldots, n\}$. A *stable* or *independent* set is a set of mutually nonadjacent nodes, and $\alpha(G)$ is the maximum size of a stable set: it is NP-hard to compute. A *clique* of G is a set of mutually adjacent nodes, and $\bar{\chi}(G)$ is the minimum cardinality of a collection of cliques that together include all the nodes of G (a clique cover): this is also NP-hard to compute. Note that $\bar{\chi}(G) = \chi(\bar{G})$, the chromatic number of the complement $\bar{G}$ of G in which two nodes are adjacent iff they are nonadjacent in G. Clearly, since each node in a stable set must be in a different clique in a clique cover,

$$\alpha(G) \leq \bar{\chi}(G).$$

Our aim is to find approximations to these numbers: in particular, we'll define $\theta(G)$, which lies between $\alpha(G)$ and $\bar{\chi}(G)$ and is the optimal value of an SDP problem. If G is a so-called *perfect* graph, then $\alpha(G) = \bar{\chi}(G)$ and we can calculate these invariants of the graph exactly by computing $\theta(G)$.

We define

$$\theta(G) := \max\{ee^T \bullet X : I \bullet X = 1, x_{ij} = 0 \text{ if } ij \in E, X \succeq 0\}, \tag{5}$$

where $e \in I\!R^n$ denotes the vector of ones. Clearly this is an SDP problem in primal form, but in maximization form. Its dual can be written as $\min\{\eta : \eta I + \sum_{ij \in E} y_{ij} M_{ij} \succeq ee^T\}$, where M_{ij} is the symmetric matrix that is all zero except for ones in the ijth and jith positions. The constraint on η can also be written as $\eta I \succeq V + ee^T$, where V is a symmetric matrix that is zero on the diagonal and in positions $ij \notin E$. As we shall see in the next section, strong duality holds for this pair of SDP problems, so we can also define

$$\theta(G) := \min\{\lambda_{\max}(V + ee^T) : v_{ii} = 0 \text{ for all } i, v_{ij} = 0 \text{ for all } ij \notin E, V \in SI\!R^{n \times n}\}.$$

It is also instructive to give another definition of $\theta(G)$. An *orthonormal representation of* G is a set $\{u_i : i \in N\}$ of unit vectors in $I\!R^n$ with u_i and u_j orthogonal if $ij \notin E$. Then $\theta(G)$ can also be defined as the minimum, over all orthonormal representations $\{u_i : i \in N\}$ of G and all unit vectors c, of

$$\max_{i \in N} \frac{1}{(c^T u_i)^2},$$

where $1/0$ is taken to be $+\infty$. To illustrate these three definitions, consider the square viewed as a graph on four nodes, with edges 12, 23, 34, and 14. Then $\alpha(G) = \bar{\chi}(G) = 2$, so $\theta(G) = 2$ also. For the first definition, an optimal X has $1/4$ in positions 11, 13, 22, 24, 31, 33, 42, and 44, with zeroes elsewhere. An optimal $V + ee^T$ for the second definition is 4 times this matrix. And to get an optimal orthonormal representation, consider an umbrella with just four ribs, and imagine opening it up until nonadjacent ribs are orthogonal. Then the u_i's are unit vectors along the ribs ($(\pm 1; 0; -1)/\sqrt{2}$ and $(0; \pm 1; -1)/\sqrt{2}$) and the unit vector c is a unit vector along the handle: $(0; 0; -1)$. (A similar example for the pentagon uses a five-ribbed umbrella, and gives $\theta = \sqrt{5}$, while $\alpha = 2$ and $\bar{\chi} = 3$; $\sqrt{5}$ is also the Shannon capacity of the pentagon.) It is not immediately apparent that this last definition gives the same value as the previous ones: we refer to Grötschel, Lovász, and Schrijver [24] for a proof of this (and several other definitions of $\theta = \theta(G)$). See also the survey article of Goemans [21]. We just note here the relationship between positive semidefinite matrices and sets of vectors. If V gives an optimal solution to the problem defining θ as a maximum eigenvalue, then $\theta I - V - ee^T$ is positive semidefinite and hence can be factored as W^TW, and clearly we then know something about the inner products of the columns w_i of W. We obtain an orthonormal system for G by manipulating these vectors w_i.

We conclude our discussion of this application by showing that all three definitions give an upper bound on $\alpha(G)$. Let $K \subseteq N$ be a maximum-cardinality stable set of G, with cardinality $k = \alpha(G)$. For the first definition, choose for X the symmetric matrix that is all zeroes, except that $x_{ij} = 1/k$ for all $i, j \in K$. It is clear that this is feasible, and it gives an objective value of k. The maximum value is thus at least as large. For the second definition, consider any feasible V, and note that the (K, K) principal submatrix of $V + ee^T$ consists of all ones, and hence has maximum eigenvalue equal to its order, k. Since the largest eigenvalue of any matrix is at least that of any principal submatrix (e.g., from considering Rayleigh quotients), we conclude that the optimal value of the eigenvalue problem is at least k. Finally, let $\{u_i : i \in N\}$ be an orthonormal representation of G and c a unit vector. Then $\{u_i : i \in K\}$ is a set of orthonormal vectors (which can be completed to an orthonormal basis), and so

$$1 = \|c\|^2 \geq \sum_{j \in K} (c^T u_j)^2.$$

It follows that one of the summands is at most $1/k$, and hence

$$\max_{i \in N} \frac{1}{(c^T u_i)^2} \geq \max_{j \in K} \frac{1}{(c^T u_j)^2} \geq k.$$

Example 9: the max-cut problem. Once again we have an undirected graph $G = (N, E)$, and a nonnegative vector $w = (w_{ij}) \in I\!R^E_+$. For $K \subseteq N$, $\delta(K)$ denotes $\{ij \in E : i \in K, j \notin K\}$, the *cut* determined by K, with *weight* equal to $w(\delta(K)) := \sum_{ij \in \delta(K)} w_{ij}$. We want to find the cut of maximum weight. (This problem arises in VLSI and in finding the ground state of a spin glass; see Poljak and Tuza [52].) We can assume that the graph is complete (each node is adjacent to all others) by setting $w_{ij} = 0$ for all non-edges ij; we also set $w_{ii} = 0$ for all i.

We start with two (nonconvex) quadratic programming formulations. We use $x \in I\!R^n$, with each $x_i = \pm 1$, to represent the cut $\delta(K)$, where $x_i = 1$ iff $i \in K$. Then clearly $x_i x_j$ is -1 if $ij \in \delta(K)$, $+1$ otherwise. Let us define $C \in SI\!R^{n \times n}$ by setting $c_{ij} = -w_{ij}/4$ for $i \neq j$ and $c_{ii} = \sum_j w_{ij}/4$ for all i. Then for the x above, we have

$$w(\delta(K)) = \frac{1}{2} \sum_{i<j} w_{ij}(1 - x_i x_j) = \frac{1}{4} \sum_i \sum_j w_{ij}(1 - x_i x_j) = x^T C x.$$

Since every $(+1, -1)$-vector corresponds to a cut, the max-cut problem can be written as the integer quadratic programming problem

$$(IQP): \quad \max x^T C x, \quad x_i \in \{+1, -1\}, \, i \in N,$$

or as the nonconvex quadratically constrained quadratic problem

$$(NQCQP): \quad \max x^T Cx, \quad x_i^2 = 1,\, i \in N.$$

We now discuss three different ways to arrive at an SDP relaxation of this problem. First, we note that (NQCQP) is linear in the products $x_i x_j$, and these are the entries of the rank one matrix $X = xx^T$. Note that $X \in S\mathbb{R}^{n\times n}$, with $X_{ii} = 1$ for all i and $X \succeq 0$. (We write the entries of X as X_{ij} to avoid confusion with the vector x.) Conversely, it is easy to see that any such matrix that also has rank one is of the form xx^T for some $(+1,-1)$-vector x. Since $x^T Cx = C \bullet (xx^T)$, we see that (IQP) is equivalent to

$$\max C \bullet X, \quad X_{ii} = 1,\, i \in N,\; X \succeq 0,\; X \text{ rank one}.$$

If we relax the last constraint, we get the SDP problem

$$\max C \bullet X, \quad X_{ii} = 1,\, i \in N,\; X \succeq 0. \tag{6}$$

Secondly, note that in (IQP), we associate a 1-dimensional unit vector x_i (± 1) with each node. As in the previous example, we now associate an n-dimensional unit vector p_i with each node, and let P be the matrix whose rows are these vectors. (P corresponds to the vector x, whose rows correspond to the 1-dimensional vectors.) We then replace the objective $C \bullet (xx^T)$ with $C \bullet (PP^T)$, and the constraints $x_i \in \{+1,-1\}$ by $\operatorname{diag}(PP^T) = e$. Since PP^T is positive semidefinite, and every such matrix can be factored as PP^T, we see that our problem has become the SDP problem above. It is clearly a relaxation, since if we restrict each row of P to a multiple of a fixed unit vector (± 1), then we recover (IQP).

The third way to derive the SDP relaxation is by taking the dual twice. (This approach was apparently first considered by Shor [57]; see also Poljak et al. [51].) Given any optimization problem $\max\{f(x) : g(x) = b, x \in \Xi\}$, where we have distinguished a certain set of m equality constraints and left the rest as an abstract set restriction, the Lagrangian dual obtained by dualizing the $g(x) = b$ constraints is defined to be

$$\min_{y \in \mathbf{R}^m} h(y), \quad \text{where } h(y) := \max_{x \in \Xi} [f(x) - y^T(g(x) - b)].$$

Note that h, as the pointwise maximum of a set of linear functions, is always convex. It is easy to see that the optimal value of this dual problem always provides an upper bound on that of the original.

We now apply this scheme to (NQCQP), dualizing the constraints $x_i^2 = 1$, $i = 1, \ldots, n$. The dual problem is to minimize over all $y \in \mathbb{R}^n$

$$\begin{aligned} h(y) &:= \max_{x \in \mathbf{R}^n} (x^T Cx - \sum_i y_i (x_i^2 - 1)) \\ &= e^T y - \min_{x \in \mathbf{R}^n} (x^T (\operatorname{Diag}(y) - C)x). \end{aligned}$$

The minimum here is 0 if $\operatorname{Diag}(y) - C$ is positive semidefinite, and $-\infty$ otherwise. Hence there is an implicit semidefinite constraint, and the dual problem becomes

$$\min e^T y, \quad \operatorname{Diag}(y) - C \succeq 0.$$

This is an SDP problem in dual form, and its dual is precisely the SDP problem above. Again, these dual problems satisfy the conditions of the next section guaranteeing strong duality, so either provides a relaxation of the original max-cut problem. These bounds on the value of a maximum weight cut were obtained by Delorme and Poljak [14].

Since we have a relaxation, the optimal *value* of the SDP problem provides an upper bound on the value of the max cut. But in this case, we can also use the *solution* of the primal problem to generate a provably good cut, as was shown in a beautiful contribution of Goemans and Williamson [22] (see also the survey article of Goemans [21]). This uses the second derivation of the SDP problem above. So let us suppose the optimal solution is $X \succeq 0$, and then factor $X = PP^T$. Then the rows of P, p_i for each i, give unit vectors for each node. If these vectors were all collinear, then we could obtain a maximum weight cut by choosing the nodes whose vectors were equal to p_1 as K, and those with vectors equal to $-p_1$ as $N \setminus K$. In general, we proceed as follows. Choose a random vector v uniformly on the unit sphere, and set $K := \{i \in N : v^T p_i \geq 0\}$. Then we get a random cut, and it is not hard to show that its expected weight, $Ew(\delta(K))$, is at least .878 of the optimal value of the SDP problem, which is at least the value of a maximum weight cut. Hence we achieve at least this fraction of the best cut (on average) in this way. In fact, it is possible to derandomize this procedure, to achieve a deterministic cut that is provably close to maximum weight. For the pentagon (again!) with all weights equal to one, the ratio of the optimal values of the max-cut problem and its SDP relaxation is about .884, so the bound above is about the best one could hope for.

4 Duality

Now it is time to discuss the relation between (P) and (D). We have already shown weak duality, and here we will give conditions for strong duality to hold. But first, since we have discussed Lagrangian duality in Example 9, we show that each of these problems is the Lagrangian dual of the other, dualizing the equality constraints in each case. It is easy to see that (D) is the Lagrangian dual of (P) (of course, we have to switch max and min in our derivation of the dual). Let us show that (P) is the Lagrangian dual of (D), when we dualise the constraints $\mathcal{A}^*y + S = C$. Since this is an equation between symmetric matrices, our dual variable will also be a symmetric matrix, and we shall denote it by X. Hence our dual problem is

$$\min_{X \in \mathcal{S}\mathbb{R}^{n\times n}} h(X), \quad h(X) := \max_{y \in \mathbb{R}^m, S \succeq 0} [b^T y - (\mathcal{A}^* y + S - C) \bullet X].$$

The maximum can be written as

$$C \bullet X - \min_{y \in \mathbb{R}^m} [(\mathcal{A}X - b)^T y] - \min_{S \succeq 0} [S \bullet X].$$

Since y ranges over all of $\mathbb{R}^m$, this is $+\infty$ unless $\mathcal{A}X - b = 0$. Also, by Fact 13 (self-duality of the cone of positive semidefinite matrices), it is $+\infty$ unless $X \succeq 0$. If X

satisfies both these conditions, the maximum reduces to just $C \bullet X$, and we retrieve (P).

Next we present a number of examples, from Vandenberghe and Boyd [63] and Luo, Sturm, and Zhang [38], showing how strong duality can fail. Further examples can be found in the latter reference.

Consider first

$$\max -y_1, \quad \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Equivalently, we require that $\begin{pmatrix} y_1 & 1 \\ 1 & y_2 \end{pmatrix} \succeq 0$. It is easy to see that the feasible region is $\{(y_1; y_2) : y_1 > 0, y_2 > 0, y_1 y_2 \geq 1\}$. So the optimal value is 0, but it is not attained. (We can get arbitrarily close with solutions of the form $(\epsilon; 1/\epsilon)$ for arbitrarily small positive ϵ.) The dual of this problem is

$$\min \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \bullet X, \quad \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \bullet X = -1, \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \bullet X = 0, \; X \succeq 0,$$

for which the only feasible (and hence optimal) solution is $X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ with optimal value 0. Here there is no duality gap, but one of the values is not attained.

Our next example is

$$\begin{array}{rrcl} \min & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \bullet X & & \\ & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \bullet X & = & 0, \\ & \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \bullet X & = & 2, \\ & X & \succeq & 0. \end{array}$$

Any feasible X is of the form $\begin{pmatrix} 0 & \xi_1 & \xi_2 \\ \xi_1 & \xi_3 & \xi_4 \\ \xi_2 & \xi_4 & 1 - \xi_1 \end{pmatrix}$, and, since it must be positive semidefinite, in fact $\begin{pmatrix} 0 & 0 & 0 \\ 0 & \xi_3 & \xi_4 \\ 0 & \xi_4 & 1 \end{pmatrix}$ for suitable ξ_i's. It follows that an optimal X is

$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, with optimal value 1. The dual problem is

$$\max\ 2y_2,\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Equivalently, we require

$$S = \begin{pmatrix} -y_1 & -y_2 & 0 \\ -y_2 & 0 & 0 \\ 0 & 0 & 1-2y_2 \end{pmatrix} \succeq 0,$$

so y_2 must equal 0 and y_1 be nonpositive. Thus $y = (0;0)$ is optimal, with optimal value 0. Here both problems attain their optimal values, but there is a gap between them.

Note that in both primal and dual, a matrix that is required to be positive semidefinite has a zero on the diagonal, and this forces the off-diagonal entries in that row and column to be zero also. It is instructive to see what happens when this implicit constraint is removed by perturbing the diagonal entry. The reader may wish to check that if b_1 (using the usual notation) is changed to $\epsilon > 0$, then both optimal values become 0, while if c_{22} is changed to $\epsilon > 0$, then both optimal values become 1. (If both changes are made, the optimal values again become equal, but now both are 3/4.)

Our last example is

$$\begin{array}{rrcl} \min & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \bullet X & & \\ & \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \bullet X & = & 0, \\ & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \bullet X & = & 2, \\ & X & \succeq & 0. \end{array}$$

Any feasible X must have (1,1) entry 0 and (1,2) entry 1, and such a matrix cannot be positive semidefinite, so the optimal value (using the usual convention) is $+\infty$. The dual problem is

$$\max\ 2y_2,\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Equivalently, we require

$$S = \begin{pmatrix} -y_1 & -y_2 \\ -y_2 & 0 \end{pmatrix} \succeq 0,$$

so y_2 must equal 0 and y_1 be nonpositive. Thus $y = (0;0)$ is optimal, with optimal value 0. Here there is an infinite gap between the optimal values.

Having given examples of strong duality failing, we now turn to conditions ensuring that it holds. It turns out that a Slater (strict feasibility) condition suffices. Let us define

$$\begin{aligned} F(P) &:= \{X \in S\mathbb{R}^{n\times n} : \mathcal{A}X = b, X \succeq 0\}, \\ F^0(P) &:= \{X \in F(P) : X \succ 0\}, \\ F(D) &:= \{(y,S) \in \mathbb{R}^m \times S\mathbb{R}^{n\times n} : \mathcal{A}^* y + S = C, S \succeq 0\}, \\ F^0(D) &:= \{(y,S) \in F(D) : S \succ 0\}. \end{aligned}$$

Also, we say the *linear independence* condition holds if $A_1, \ldots, A_m$ are linearly independent in $S\mathbb{R}^{n\times n}$.

Theorem 4.1 *(Strong Duality) Suppose that $F(P)$ and $F^0(D)$ are nonempty. Then (P) has a nonempty compact set of optimal solutions, and the optimal values of (P) and (D) are equal.*

Proof:
Let $\hat{X} \in F(P)$ and $(\hat{y}, \hat{S}) \in F^0(D)$. Then we can add the constraint $C \bullet X \le C \bullet \hat{X}$ to (P) without changing its optimal value or the set of its optimal solutions. But, using (2), this inequality is equivalent to $\hat{S} \bullet X = C \bullet X - b^T\hat{y} \le C \bullet \hat{X} - b^T\hat{y} = \hat{S} \bullet \hat{X}$. So (P) has the same optimal value and set of optimal solutions as

$$(P'): \qquad \min C \bullet X, \quad \mathcal{A}X = b,\ \hat{S} \bullet X \le \hat{S} \bullet \hat{X},\ X \succeq 0.$$

But by Fact 14, this problem has a compact feasible region since $\hat{S} \succ 0$. The objective function being continuous, this implies the first assertion of the theorem.

Now let ζ_* denote the optimal value of (P) and ϵ be positive. We want to show that there is a feasible solution of (D) with objective value at least $\zeta_* - \epsilon$. Consider the two sets $\mathcal{G}_1 := S\mathbb{R}^{n\times n}_+$ and $\mathcal{G}_2 := \{X \in S\mathbb{R}^{n\times n} : \mathcal{A}X = b, C \bullet X \le \zeta_* - \epsilon\}$. These two sets are closed convex, and disjoint, and have no common direction of recession (any such would be a nonzero $X \succeq 0$ satisfying $C \bullet X = 0$, $\mathcal{A}X = 0$, showing that the set of optimal solutions of (P) is unbounded, a contradiction). Hence, by a separating hyperplane theorem (Rockafellar [56], Corollary 11.4.1), there exist $S \in S\mathbb{R}^{n\times n}$ and $\sigma \in \mathbb{R}$ with

$$S \bullet X > \sigma \quad \text{for any } X \in \mathcal{G}_1, \qquad S \bullet X < \sigma \quad \text{for any } X \in \mathcal{G}_2.$$

Since $0 \in \mathcal{G}_1$, σ is negative. Since $\lambda u u^T \in \mathcal{G}_1$ for any positive λ and any $u \in \mathbb{R}^n$, it follows that $S \succeq 0$.

Next we have that $\mathcal{A}X = b$, $C \bullet X \le \zeta_* - \epsilon$ imply $S \bullet X \le \sigma$. By a theorem of the alternative (for linear inequalities — there are no semidefinite constraints here), there exist $y \in \mathbb{R}^m$ and $\eta \ge 0$ with

$$C\eta - \mathcal{A}^* y = S, \quad (\zeta_* - \epsilon)\eta - b^T y \le \sigma.$$

Suppose $\eta = 0$. Then $-b^T y \le \sigma < 0$ and also $-b^T y = -(\mathcal{A}\hat{X})^T y = \hat{X} \bullet (-\mathcal{A}^* y) = \hat{X} \bullet S \ge 0$, a contradiction. Hence η is positive, and by scaling y, S, and σ we can assume that $\eta = 1$. But then $C - \mathcal{A}^* y = S \succeq 0$ and $b^T y \geq \zeta_* - \epsilon - \sigma \geq \zeta_* - \epsilon$, and we have a feasible solution to (D) with value at least $\zeta_* - \epsilon$. Since ϵ was arbitrary, we have shown that there is no duality gap. ∎

The result above is asymmetric between (P) and (D). We now make a few remarks concerning these two problems and their presentation. The gist is that each can be rewritten in the format of the other. For this, we assume the linear independence condition. (Our assumption is basically without loss of generality. If the A_i's are linearly dependent, and $\mathcal{A}^* y = 0$ implies $b^T y = 0$, then we can choose a basis, say A_i, $i = 1, \ldots, k$, for them and remove the last $m - k$ primal constraints and dual variables to get equivalent problems where the assumption holds. If we have $\mathcal{A}^* \hat{y} = 0$ and $b^T \hat{y} > 0$, then (P) is infeasible and (D) unbounded if it is feasible.)

Given the assumption, we choose $D \in S\mathbb{R}^{n \times n}$ satisfying $\mathcal{A} D = b$ and let $G_1, \ldots, G_k$ be a basis for the orthogonal complement of the span of the A_i's in $S\mathbb{R}^{n \times n}$. Finally, let $h_j := C \bullet G_j$, $j = 1, \ldots, k$. Then it is not hard to see that (P) is equivalent to

$$C \bullet D - \max_{w \in \mathbf{R}^k, X \in S\mathbf{R}^{n \times n}} \{h^T w : \sum_j w_j G_j + X = D,\, X \succeq 0\},$$

an SDP problem in dual form. Similarly, (D) is equivalent to

$$C \bullet D - \min_{S \in S\mathbf{R}^{n \times n}} \{D \bullet S : G_j \bullet S = h_j,\, j = 1, \ldots, k,\, S \succeq 0\},$$

an SDP problem in primal form. We can use this construction for moving between the two forms. Let us note that, given the linear independence condition, for any S there is at most one y with (y, S) feasible for (D), which allows us to extend boundedness results from just S to the pair (y, S).

Applying this procedure to the previous result, we obtain

Corollary 4.1 *Suppose the linear independence condition holds and that $F^0(P)$ and $F(D)$ are nonempty. Then (D) has a nonempty compact set of optimal solutions, and there is no duality gap.*

We also find

Corollary 4.2 *Suppose the linear independence condition holds and that both (P) and (D) have strictly feasible solutions. Then each has a nonempty compact set of optimal solutions, and there is no duality gap.*

We will give an alternative proof of this corollary in the next section. We note the historical fact that Corollary 4.2 was proved in 1963 by Bellman and Fan [6] for the following pair of SDP problems:

$$\begin{array}{llcl} \min & \sum_j C_j \bullet X_j & & \\ & \sum_j (A_{ij} X_j + X_j A_{ij}^T) & = & B_i, \text{ for all } i \\ & X_j & \succeq & 0, \text{ for all } j, \end{array}$$

$$\begin{array}{llcl} \max & \sum_i B_i \bullet Y_i & & \\ & \sum_i (Y_i A_{ij} + A_{ij}^T Y_i) & \preceq & C_j, \text{ for all } j. \end{array}$$

Here all B_i's and C_j's, as well as the variables X_j and Y_i, are symmetric matrices of order n, while the A_{ij}s are not necessarily symmetric matrices of the same order; also j runs from 1 to k and i from 1 to m throughout. Clearly this form of the problems was inspired by systems arising in control theory, but no connections were made. It is possible to show that (P) and (D) can be formulated as above, and that these problems can be formulated as (P) and (D).

Since the "standard" dual (D) of (P) may lead to a positive duality gap, we can ask whether there is a perhaps more complicated dual problem for which strong duality always holds, without any additional regularity assumptions. The answer is in the affirmative: see Ramana [54] and Ramana, Tunçel, and Wolkowicz [55].

Finally, if we assume that strong duality holds, then we have as necessary and sufficient optimality conditions the following:

$$\begin{array}{rcccll} & \mathcal{A}^*y & + & S & = & C, \quad S \succeq 0, \\ \mathcal{A}X & & & & = & b, \quad X \succeq 0, \\ & XS & & & = & 0. \end{array}$$

(Here the natural last condition stating that the duality gap is zero, $X \bullet S = 0$, has been replaced by the seemingly stronger condition that the matrix product is zero using Fact 15, since both matrices are positive semidefinite.)

5 The Logarithmic Barrier Function and the Central Path

We define f on $S\mathbb{R}^{n\times n}$ by

$$f(X) := -\ln\det X \text{ if } X \succ 0, \quad f(X) := +\infty \text{ otherwise.}$$

and call it the *logarithmic barrier function* for the cone $S\mathbb{R}^{n\times n}_+$ of positive semidefinite matrices. For $n = 1$, we get the smooth function $-\ln x$, which is defined on the positive axis and tends to $+\infty$ as x approaches 0 from above. In the same way, f is defined on the positive definite matrices and tends to $+\infty$ as X approaches a matrix $\bar{X} \in \partial S\mathbb{R}^{n\times n}_+$ through positive definite values. We say that f *has the barrier property* for $S\mathbb{R}^{n\times n}_+$.

The idea, which we shall investigate in detail below, is to replace the problem (P) with the somewhat awkward constraint that X be positive semidefinite by the sequence of problems (barrier problems parametrised by $\nu > 0$):

$$BP(\nu): \quad \min C \bullet X + \nu f(X), \quad \mathcal{A}X = b \quad (X \succ 0),$$

where there is only a linear constraint, the implicit positive definite constraint being enforced by the barrier property of f.

Clearly f is smooth on the interior $S\mathbb{R}^{n\times n}_{++}$ of $S\mathbb{R}^{n\times n}_+$: we evaluate its first few derivatives. Let $X \succ 0$, $H \in S\mathbb{R}^{n\times n}$. Then

$$\begin{array}{rcl} f(X+\alpha H) & = & -\ln\det[X(I+\alpha X^{-1}H)] \\ & = & -\ln\det X - \ln(1+\alpha \operatorname{trace} X^{-1}H + O(\alpha^2)) \\ & = & f(X) - \alpha X^{-1} \bullet H + O(\alpha^2), \end{array}$$

so that $f'(X) = -X^{-1}$ and $Df(X)[H] = -X^{-1} \bullet H$.

Similarly we find

$$\begin{aligned} f'(X+\alpha H) &= -[X(I+\alpha X^{-1}H)]^{-1} \\ &= -[I-\alpha X^{-1}H) + O(\alpha^2)]X^{-1} \\ &= f'(X) + \alpha X^{-1}HX^{-1} + O(\alpha^2). \end{aligned}$$

Hence $f''(X)[H] = X^{-1}HX^{-1}$ and $D^2f(X)[H,J] = X^{-1}HX^{-1} \bullet J$. In the notation introduced in (1), $f''(X) = X^{-1} \odot X^{-1}$. It is easy to see that the adjoint of $P \odot Q$ is $P^T \odot Q^T$, so this operator is self-adjoint (i.e., $[(P\odot Q)U] \bullet V = [(P \odot Q)V] \bullet U$) if P and Q are symmetric; further, it is positive definite (i.e., $[(P \odot Q)U] \bullet U > 0$ if $U \in S\mathbb{R}^{n\times n}$ is nonzero) if $P \succ 0$, $Q \succ 0$. Hence $f''(X)$ is self-adjoint and positive definite. In the same way we find $f'''(X)[H,J] = -X^{-1}HX^{-1}JX^{-1} - X^{-1}JX^{-1}HX^{-1}$.

We now introduce the important notion of *self-concordance*, defined and developed in great detail by Nesterov and Nemirovski [45]. We know that f is convex if, for every $X \succ 0$ and every $H \in S\mathbb{R}^{n\times n}$, $\phi(\alpha) := f(X+\alpha H)$ is convex in α. We say that f is *self-concordant* if it is convex and 3-times differentiable and if, for every such X and H, ϕ defined as above satisfies

$$|\phi'''(0)| \le 2[\phi'(0)]^{3/2}.$$

Finally, f is a *θ-normal barrier for $S\mathbb{R}^{n\times n}_{++}$* (or for $S\mathbb{R}^{n\times n}_{+}$) if it is convex, self-concordant, has the barrier property, and is logarithmically homogeneous of degree θ:

$$f(\alpha X) = f(X) - \theta \ln \alpha, \quad \text{for all } X \succ 0,\ \alpha > 0.$$

We now have

Theorem 5.1 *$f(X) := -\ln\det X$ is an n-normal barrier for $S\mathbb{R}^{n\times n}_{+}$.*

Proof:
Define ϕ as above. Then it is finite on the convex set of α such that $X + \alpha H \succ 0$, and on this set

$$\phi''(\alpha) = D^2 f(\bar{X})[H,H] = (\bar{X}^{-1}H\bar{X}^{-1}) \bullet H,$$

where $\bar{X} := X + \alpha H$. Since this matrix is positive definite, so is $V := \bar{X}^{-\frac{1}{2}}$, and then $\phi''(\alpha) = V^2HV^2 \bullet H = \operatorname{trace}(V^2HV^2H) = \operatorname{trace}([VHV][VHV]) = \|VHV\|_F^2 \ge 0$. So ϕ is convex. Indeed, the quantity above is positive if H is nonzero, so in fact then ϕ and hence f is strictly convex. We have also shown that $f''(X)$ is a positive definite and hence nonsingular operator.

Let ϕ be as above and now let $V := X^{-\frac{1}{2}} \succ 0$. Then $\phi''(0) = \operatorname{trace}([VHV][VHV])$. If $\lambda := \lambda(VHV)$, then $\phi''(0) = \operatorname{trace}(\operatorname{Diag}(\lambda)\operatorname{Diag}(\lambda)) = \|\lambda\|_2^2$. Next,

$$\begin{aligned} \phi'''(0) &= -2(X^{-1}HX^{-1}HX^{-1}) \bullet H \\ &= -2\operatorname{trace}(V^2HV^2HV^2H) = -2\operatorname{trace}([VHV][VHV][VHV]) \\ &= -2\sum \lambda_i^3. \end{aligned}$$

So we conclude that

$$|\phi'''(0)| = 2|\sum \lambda_i^3| \leq 2\sum |\lambda_i^3| = 2\|\lambda\|_3^3 \leq 2\|\lambda\|_2^3 = 2[\phi''(0)]^{3/2}.$$

Finally, we already checked the barrier property, and

$$f(\alpha X) = -\ln\det(\alpha X) = -\ln(\alpha^n \det X) = f(X) - n\ln\alpha,$$

so the proof is complete. ∎

Having the positive definite operator $f''(X)$, we can define the X-*norm* of a symmetric matrix by

$$\|H\|_X := (f''(X)H \bullet H)^{frac12} = \|\lambda(X^{-\frac{1}{2}}HX^{-\frac{1}{2}}\|_2 = \|X^{-\frac{1}{2}}HX^{-\frac{1}{2}}\|_F$$

and the *dual* X-*norm* of a symmetric matrix by

$$\|J\|_X^* := ([f''(X)]^{-1}J \bullet J)^{frac12} = \|XqJXq\|_F.$$

Note that $|J \bullet H| \leq \|J\|_X^*\|H\|_X$ as in the Cauchy-Schwarz inequality.

The following properties follow from our formulae, but can also be obtained directly by differentiating the equation for logarithmic homogeneity:

Proposition 5.1 *For $\alpha > 0$, $X \succ 0$ of order n,*

$$\begin{array}{cc} f'(\alpha X) = \alpha^{-1}f'(X), & f''(\alpha X) = \alpha^{-2}f''(X); \\ f'(X) \bullet X = -n, & f''(X)X = -f'(X); \\ \|X\|_X = \sqrt{n}, & \|f'(X)\|_X^* = \sqrt{n}. \end{array}$$

∎

The last line also states that the X-norm of the Newton step for minimizing f from X, $-[f''(X)]^{-1}f'(X)$, is exactly $\sqrt{n}$. This shows that f satisfies also the original definition of Nesterov and Nemirovski (which applies also to functions that are not logarithmically homogeneous) to be an n-self-concordant barrier function.

We now return to the barrier problem mentioned at the beginning of this section, defining the primal and dual barrier problems (parametrised by $\nu > 0$) to be

$$BP(\nu): \quad \min C \bullet X + \nu f(X), \quad \mathcal{A}X = b \quad (X \succ 0),$$

and

$$BD(\nu): \quad \max b^T y - \nu f(S), \quad \mathcal{A}^* y + S = C \quad (S \succ 0).$$

It is not hard to check that each is in fact the Lagrangian dual of the other up to an additive constant.

Suppose BP(ν) has an optimal solution X. Then $X \in F^0(P)$ and, by Lagrange's theorem, for some $y \in I\!R^m$ we have

$$C - \nu X^{-1} - \mathcal{A}^* y = C + \nu f'(X) - \mathcal{A}^* y = 0.$$

Let us set $S := \nu X^{-1} \succ 0$. Then we see that $(y, S) \in F^0(D)$, and we have a solution to the set of equations

$$CPE(\nu): \quad \begin{array}{rcrcll} & & \mathcal{A}^* y & + & S & = & C, & S \succ 0, \\ \mathcal{A}X & & & & & = & b, & X \succ 0, \\ & & XS & & & = & \nu I. & \end{array} \tag{7}$$

We call these the *central path equations* for reasons that will become clearer shortly. Note that, except for the final right-hand side, these equations coincide with the optimality conditions stated at the end of the previous section.

If BD(ν) has an optimal solution (y, S), a similar derivation shows that, for some X, the above equations again hold.

Theorem 5.2 *Suppose $F^0(P)$ and $F^0(D)$ are nonempty and the linear independence assumption holds. Then for every positive ν, there is a unique solution $(X(\nu), y(\nu), S(\nu))$ to CPE(ν). Further, $X(\nu)$ is the unique solution to BP(ν) and $(y(\nu), S(\nu))$ to BD(ν). Finally, if the assumption of strict feasibility fails, then CPE(ν), BP(ν), and BD(ν) have no solution.*

Proof:
First we establish existence. Choose $\hat{X} \in F^0(P)$ and $(\hat{y}, \hat{S}) \in F^0(D)$, and consider BP($\nu$). Suppose $\sigma := \lambda_{\min}(\hat{S}) > 0$. Now $\hat{X}$ is feasible for BP(ν), and for feasible X, $C \bullet X$ differs by a constant from $\hat{S} \bullet X$ (2). Hence BP(ν) has the same set of optimal solutions as

$$BP'(\nu): \quad \min \hat{S} \bullet X + \nu f(X), \quad \mathcal{A}X = b, \, \hat{S} \bullet X + \nu f(X) \leq \hat{S} \bullet \hat{X} + \nu f(\hat{X}) \quad (X \succ 0).$$

Our aim is to show that this amounts to the minimization of a continuous function on a compact set, yielding existence.

Suppose X is feasible in BP$'(\nu)$, and let $\lambda := \lambda(X)$ and $e \in I\!R^n$ be again a vector of ones. Then we have $\lambda > 0$ and $\sigma e^T \lambda - \nu \sum \ln \lambda_j = \sigma I \bullet X + \nu f(X) \leq \hat{S} \bullet X + \nu f(X) \leq \hat{S} \bullet \hat{X} + \nu f(\hat{X}) =: \alpha$, so

$$\sum_j (\sigma \lambda_j - \nu \ln \lambda_j) \leq \alpha.$$

Now the function $\sigma\tau - \nu \ln \tau$ has a unique minimizer at $\tau_* = \nu/\sigma$ and goes to $+\infty$ as τ goes to either 0 or $+\infty$. Let the minimum value be β and suppose that $\sigma\tau - \nu \ln \tau > \alpha - (n-1)\beta$ for $\tau \in (0, \underline{\tau}]$ or $\tau \in [\bar{\tau}, +\infty)$. Then the inequality above implies that $\lambda_j \in [\underline{\tau}, \bar{\tau}]$ for all j, so $\|X\|_F = \|\lambda\|_2 \leq \sqrt{n}\bar{\tau}$. Hence we have a bounded feasible set. Moreover, $\lambda_j \geq \underline{\tau} > 0$ for all j implies that $\hat{S} \bullet X + \nu f(X)$ is continuous on this set, so it is also closed and hence compact. We have just seen that the objective function of BP$'(\nu)$ is continuous on the feasible set, and hence existence of a minimizer for BP(ν) follows. Now such a minimizer must satisfy the necessary conditions, and hence we see as above that we have a solution to CPE(ν).

Since the barrier problem is convex, these conditions are also sufficient for optimality. So any solution to CPE(ν) yields a minimizer for BP(ν). Moreover, the objective here is strictly convex, so the minimizer is unique. The equations $XS = \nu I$ show that

S is also unique, and then the equations $\mathcal{A}^*y + S = C$ and the linear independence assumption imply that y is also unique. The equations CPE(ν) also provide necessary and sufficient conditions for the dual barrier problem. Finally, if strict feasibility fails for (P), there is no solution yielding a finite value for the objective function of BP(ν); there is no solution satisfying the necessary conditions for optimality in BD(ν); and there is no solution to CPE(ν), since the X-part would give a strictly feasible solution. A similar argument applies to the dual, and the proof is complete. $\square$

So far we have established the existence of a unique solution to CPE(ν) for each positive ν, but not that these solutions form a smooth path. This will follow from the implicit function theorem if we show that the equations defining it are differentiable, with a derivative (with respect to (X, y, S)) that is square and nonsingular at points on the path. Unfortunately, while the equations of (7) are certainly differentiable, the derivative is not even square since the left-hand side maps $(X, y, S) \in S\mathbb{R}^{n\times n} \times \mathbb{R}^m \times S\mathbb{R}^{n\times n}$ to a point in $S\mathbb{R}^{n\times n} \times \mathbb{R}^m \times \mathbb{R}^{n\times n}$; XS is usually not symmetric even if X and S are. We therefore need to change the equations defining the central path. There are many possible approaches, which as we shall see lead to different search directions for our algorithms, but for now we choose a simple one: we replace $XS = \nu I$ by $-\nu X^{-1} + S = 0$. As in our discussion of the barrier function f, the function $X \to -\nu X^{-1}$ is differentiable at nonsingular symmetric matrices, with derivative $\nu(X^{-1} \odot X^{-1})$. So the central path is defined by the equations

$$\Phi_P(X, y, S; \nu) := \begin{pmatrix} & \mathcal{A}^*y & + & S \\ \mathcal{A}X & & & \\ -\nu X^{-1} & & + & S \end{pmatrix} = \begin{pmatrix} C \\ b \\ 0 \end{pmatrix}, \tag{8}$$

whose derivative (with respect to (X, y, S)) is

$$\Phi_P'(X, y, S; \nu) := \begin{pmatrix} 0 & \mathcal{A}^* & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ \nu(X^{-1} \odot X^{-1}) & 0 & \mathcal{I} \end{pmatrix}, \tag{9}$$

where $\mathcal{I}$ denotes the identity operator on $S\mathbb{R}^{n\times n}$. We have been rather loose in writing this in matrix form, since the blocks are operators rather than matrices, but the meaning is clear. We want to show that this derivative is nonsingular, and for this it suffices to prove that its null-space is trivial. Since similar equations will occur frequently, let us derive this from a more general result.

Theorem 5.3 *Suppose the operators $\mathcal{E}$ and $\mathcal{F}$ map $S\mathbb{R}^{n\times n}$ to itself, and that $\mathcal{E}$ is nonsingular and $\mathcal{E}^{-1}\mathcal{F}$ is positive definite (but not necessarily self-adjoint). Assume that the linear independence condition holds. Then, for any P, $R \in S\mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^m$, the solution to*

$$\begin{array}{rcccl} & \mathcal{A}^*v & + & W & = & P, \\ \mathcal{A}U & & & & = & q, \\ \mathcal{E}U & & + & \mathcal{F}W & = & R \end{array} \tag{10}$$

is uniquely given by

$$\begin{array}{rcl} v & = & (\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*)^{-1}(q - \mathcal{A}\mathcal{E}^{-1}(R - \mathcal{F}P)), \\ W & = & P - \mathcal{A}^* v, \\ U & = & \mathcal{E}^{-1}(R - \mathcal{F}W). \end{array} \tag{11}$$

Proof:
The formulae for W and U follow directly from the first and third equations. Now substituting for W in the formula for U, and inserting this in the second equation, we obtain after some manipulation

$$(\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*)v = q - \mathcal{A}\mathcal{E}^{-1}(R - \mathcal{F}P).$$

Since $\mathcal{E}^{-1}\mathcal{F}$ is positive definite and the A_i's are linearly independent, the $m \times m$ matrix on the left is positive definite (but not necessarily symmetric) and hence nonsingular. This verifies that v is uniquely determined as given, and then so are W and U. Moreover, these values solve the equations. □

In our case, $\mathcal{F}$ is the identity, while $\mathcal{E}$ is $\nu(X^{-1} \odot X^{-1})$ with inverse $\nu^{-1}(X \odot X)$. This is easily seen to be positive definite, just as $f''(X)$ is. Hence the theorem applies, and so the derivative of the function Φ_P is nonsingular on the central path (and throughout $S\mathbb{R}^{n\times n}_{++} \times \mathbb{R}^m \times S\mathbb{R}^{n\times n}_{++}$); thus the central path is indeed a differentiable path.

By taking the trace of the last equation of (7), we obtain the last part of the following theorem, which summarises what we have observed:

Theorem 5.4 *Assume that both (P) and (D) have strictly feasible solutions and the linear independence condition holds. Then the set of solutions to (7) for all positive ν forms a nonempty differentiable path, called the central path. If $(X(\nu), y(\nu), S(\nu))$ solve these equations for a particular positive ν, then $X(\nu)$ is a strictly feasible solution to (P) and $(y(\nu), S(\nu))$ a strictly feasible solution to (D), with duality gap*

$$C \bullet X(\nu) - b^T y(\nu) = X(\nu) \bullet S(\nu) = n\nu. \tag{12}$$

□

We claimed above that we could use the central path to prove strong duality. Indeed, we have:

Theorem 5.5 *The existence of strictly feasible solutions to (P) and (D) and the linear independence condition imply that both have bounded nonempty optimal solution sets, with zero duality gap.*

Proof:
The last part follows from the existence of the central path, since by (12) the duality gap associated to $X(\nu)$ and $(y(\nu), S(\nu))$ is $n\nu$, and this approaches zero as ν tends to zero. (In fact, the central path approaches optimal solutions to the primal and dual problems as ν decreases to zero [37, 23], but we shall not prove this here.)

To show that (P) has a bounded nonempty set of optimal solutions, we proceed as in the proof of Theorem 5.2, again choosing $(\hat{y}, \hat{S}) \in F^0(D)$. Clearly, the set of optimal solutions is unchanged if we change the objective function of (P) to $\hat{S} \bullet X$ and add the constraint $\hat{S} \bullet X \leq \hat{S} \bullet \hat{X}$. But this latter constraint (for $X \in \mathcal{P}$) implies that all the eigenvalues of X are bounded by $(\hat{S} \bullet \hat{X})/\sigma$, where again $\sigma > 0$ denotes the smallest eigenvalue of $\hat{S}$. This shows that all optimal solutions of (P) (if any) lie in a compact set of feasible solutions; but the minimum of the continuous function $\hat{S} \bullet X$ over this compact set (containing $\hat{X}$) is attained, and so the set of optimal solutions is nonempty and bounded. The proof that the set of optimal dual solutions is bounded and nonempty is similar: we start by noting that the objective of maximizing $b^T y$ can be replaced by that of minimizing $\hat{X} \bullet S$ using (2). $\square$

6 Algorithms

In this section we will discuss three classes of algorithms for solving SDP problems: path-following methods, potential-reduction methods and algorithms based on smooth or nonsmooth nonlinear programming formulations. The first two classes consist of interior-point methods, while the last contains both interior-point and non-interior-point approaches. Interior-point methods for SDP were first introduced by Nesterov and Nemirovski (see [45]) and independently by Alizadeh [1]. In all cases we shall concentrate on feasible methods, in which all iterates are (strictly in the first two cases) feasible; if we are using Newton steps, this implies that P and q in the system (10) will be zero, while R will depend on the method. One easy way to allow infeasible iterates (satisfying positive definiteness, but not the equality constraints) is to just let P and q be the negatives of the residuals in the dual and primal equality constraints, but some theoretical results then do not hold. Alternatively, the problems (P) and (D) can be embedded in a larger self-dual system that always has strictly feasible solutions at hand and whose solution gives the required information about the original problems: see [31, 38, 53], based on the work of Ye, Todd, and Mizuno [69] for linear programming.

6.1 Path-following methods

These methods are motivated by Theorem 5.4, and attempt to track points on the central path as the parameter ν is decreased to zero. We mention first primal and dual versions, and then discuss primal-dual methods.

Primal and dual path-following methods conform to the general scheme of Nesterov and Nemirovski [44, 45], where they were first introduced and analysed. The basic strategy of the primal method is to take some Newton steps towards the minimizer of BP(ν) for some parameter $\nu > 0$, and then decrease ν and repeat. It is easy to see that Newton steps for minimizers of BP(ν) are just the X-part of Newton steps for the zeroes of $\Phi_P(\cdot;\nu)$ in (8), and Theorem 5.3 shows how these may be computed. It is not necessary to maintain the S iterates, but the y iterates are useful to give a test for when the Newton steps can be terminated and ν reduced. We want the gradient of BP(ν), modified by a Lagrangian term, to be sufficiently small, and since gradients "live in dual space", we measure this using the dual X-norm. Hence our proximity criterion is

$$\|C - \nu X^{-1} - \mathcal{A}^* y\|_X^* \le \tau\nu,$$

where $\tau \in (0,1)$. This has two nice consequences. Suppose we set $S := C - \mathcal{A}^* y$. Then we have $\|S - \nu X^{-1}\|_X^* \le \tau\nu$, so that $\|\nu^{-1}S - X^{-1}\|_{X^{-1}} \le \tau$, and using the eigenvalue characterization of this norm we see that $\nu^{-1}X^{\frac{1}{2}}SX^{\frac{1}{2}}$ and hence S is positive definite, and so (y, S) strictly feasible for (D). Secondly, the duality gap is

$$X \bullet S = X \bullet (\nu X^{-1} + [S - \nu X^{-1}]) \le \nu n + \|X\|_X \|S - \nu X^{-1}\|_X^* \le \nu(n + \tau\sqrt{n}).$$

so that we are provably close to optimality when ν is small. The algorithm then becomes:

Choose a strictly feasible X for (P), $y \in \mathbb{R}^m$, and $\nu > 0$. Perform damped Newton steps, maintaining X positive definite, until the proximity criterion is satisfied. Stop if ν is sufficiently small. Otherwise, replace ν by $\theta\nu$ for some $\theta \in (0,1)$ and continue.

Here by damped Newton steps we mean that (X, y) is replaced by $(X_+, y_+) := (X + \alpha\Delta X, y + \alpha\Delta y)$ for some $\alpha \in (0,1]$, where $(\Delta X, \Delta y)$ is the usual (full) Newton step obtained by setting the linearization of $\Phi_P(\cdot,\nu)$ to zero, which will now be called the Newton direction. Using Theorem 5.3, it is not hard to see that this direction can be found by first computing the $m \times m$ matrix M with entries $m_{ij} := \nu^{-1}A_i \bullet (XA_jX)$, then solving

$$M\Delta y = -\mathcal{A}(X - \nu^{-1}X[C - \mathcal{A}^* y]X),$$

and finally setting $\Delta X = X - \nu^{-1}X[C - \mathcal{A}^*(y + \Delta y)]X$. Note that the proximity criterion is satisfied (for X and $y + \Delta y$) iff the Newton step for X is small: $\|\Delta X\|_X \le \tau$.

The beautiful theory of self-concordant functions developed by Nesterov and Nemirovski enables them to establish a polynomial convergence result for this method. Suppose the initial (X, y, ν) are such that the proximity criterion is satisfied for $\tau = .1$ (so that the first action of the algorithm will be to reduce ν). Suppose also that ν is reduced each time by the factor $\theta = 1 - .1/\sqrt{n}$. Then at each iteration we can choose $\alpha = 1$ (we do not need to damp the Newton steps), the proximity criterion will be satisfied after a single Newton step, and in $O(\sqrt{n}\ln(1/\epsilon))$ steps, the duality gap will be reduced to ϵ times its original value. (The occurrence of $\sqrt{n}$ in these results arises since f is an n-normal barrier for the positive semidefinite cone, and more particularly from the size of $f'(X)$ established in Proposition 5.1. This shows that ν can be reduced by the factor θ above while not losing too much proximity, so that one Newton step restores it.)

Next we discuss the dual method. This can be viewed as taking Newton steps for the minimizer of BD(ν), or equivalently for the zero of $\Phi_D(\cdot;\nu)$, defined as $\Phi_P(\cdot;\nu)$ but with $X - \nu S^{-1}$ replacing $-\nu X^{-1} + S$ as its last part. Here it is not necessary to maintain the X iterates. It is not hard to see that the Newton direction is computed as follows. First find the $m \times m$ matrix M with entries $m_{ij} := \nu A_i \bullet (S^{-1}A_jS^{-1})$, then solve

$$M\Delta y = b - \nu\mathcal{A}S^{-1}, \qquad (13)$$

and finally set $\Delta S = -\mathcal{A}^*\Delta y$. Continue taking damped Newton steps until the following proximity criterion is satisfied:

$$\|\Delta S\|_S \le \tau.$$

Then reduce ν and continue. Here, $X := \nu[S^{-1} + S^{-1}(\mathcal{A}^*\Delta y)S^{-1}]$ is strictly feasible in (P) when this criterion holds. The same theoretical results hold as in the primal case. One advantage of the dual method arises when C and the A_i's share a sparsity pattern. Then S will have the same sparsity, while X may well be dense. Of course, S^{-1} is likely to be dense, but we may be able to perform operations cheaply with this matrix using a sparse Cholesky factorization of S. Recently, Fukuda et al. [20] have investigated ways in which the primal-dual methods discussed next can exploit this form of sparsity.

Now we turn to primal-dual path-following methods. Here we maintain (X, y, S), and our steps are determined by both the current primal and the current dual iterates. Apart from the sparsity issue above, this seems to be worthwhile computationally, and leads to fewer difficulties if an iterate gets close to the boundary of the positive semidefinite cone. In addition, the Newton step is based on a system more like $XS - \nu I = 0$, which is certainly smoother than one involving inverses, especially for near-singular iterates. The Newton step is then regarded as a search direction, and damped steps are taken (possibly with different damping in the primal and dual spaces) to get the next iterates. As discussed in the previous section, we cannot take Newton steps for the function whose last part is defined by $XS - \nu I$, so we have to symmetrize this somehow, but now we do this without using the inverse function. The first idea is to replace this condition with $(XS + SX)/2 - \nu I$, and this was proposed by Alizadeh, Haeberly, and Overton [4]. Linearizing this system gives the equation (in addition to the feasibility equations)

$$\frac{1}{2}(\Delta X S + S\Delta X + X\Delta S + \Delta S X) = \nu I - \frac{1}{2}(XS + SX).$$

Thus the resulting Newton direction (called the AHO search direction) is defined by a system as in (10) with

$$\mathcal{E} = S \odot I, \quad \mathcal{F} = X \odot I.$$

One difficulty with this system is that we do not have an explicit form for the inverse of $\mathcal{E}$; instead, to find $\mathcal{E}^{-1}U$ we need to solve a Lyapunov system. Also, the sufficient conditions of Theorem 5.3 do not hold for this choice, and Todd, Toh, and Tütüncü [60] give an example where the Newton direction is not well-defined at a pair of strictly feasible solutions. (This does not seem to cause difficulties in practice.)

A more general approach is to apply a similarity to XS before symmetrizing it. This was discussed for a specific pair of similarities by Monteiro [39], and then in general by Zhang [70]. So let P be nonsingular, and let us replace the last part of Φ_P by

$$\frac{1}{2}(PXSP^{-1} + P^{-T}SXP^T) - \nu I. \tag{14}$$

(Zhang showed that this is zero exactly when $XS = \nu I$ as long as X and S are symmetric.) An alternative way to view this is to scale (P) so that the variable X is replaced by $\hat{X} := PXP^T$ and (D) so that S is replaced by $\hat{S} := P^{-T}SP^{-1}$; then apply the Alizadeh-Haeberly-Overton approach in this scaled space. The resulting search directions form the *Monteiro-Zhang* family. Of course, with $P = I$, we retrieve the AHO direction.

Since the need for symmetrization occurs because X and S do not commute, it seems reasonable to choose P so that the scaled matrices do commute. Three ways to do this are: choose $P = S^{\frac{1}{2}}$ so that $\hat{S} = I$; choose $P = X^{-\frac{1}{2}}$ so that $\hat{X} = I$; and choose $P = W^{-\frac{1}{2}}$, where

$$W = X^{\frac{1}{2}}(X^{\frac{1}{2}}SX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}} \tag{15}$$

is the unique positive definite matrix with $WSW = X$, so that $\hat{X} = \hat{S}$. The resulting search directions are known as the HRVW/KSH/M, dual HRVW/KSH/M, and NT directions. The first was introduced by Helmberg, Rendl, Vanderbei, and Wolkowicz [28], and independently Kojima, Shindoh, and Hara [33], using different motivations, and then rediscovered from the perspective above by Monteiro [39]. The second was also introduced by Kojima, Shindoh, and Hara [33] and rediscovered by Monteiro; since it arises by switching the roles of X and S, it is called the dual of the first direction. The last was introduced by Nesterov and Todd [46, 47], from yet another motivation, and shown to be derivable in this form by Todd, Toh, and Tütüncü [60]. These and several other search directions are discussed in Kojima et al. [32] and Todd [59].

In the first case, the Newton direction can be obtained from the solution of a linear system as in (10) with

$$\mathcal{E} = \mathcal{I}, \quad \mathcal{F} = X \odot S^{-1};$$

in the second case with

$$\mathcal{E} = S \odot X^{-1}, \quad \mathcal{F} = \mathcal{I};$$

and in the third case with

$$\mathcal{E} = \mathcal{I}, \quad \mathcal{F} = W \odot W$$

(it is not immediate that this last corresponds to the Newton system for (14) with $P = W^{\frac{1}{2}}$; see [60] for the analysis). In all cases, it is easy to see that $\mathcal{E}^{-1}\mathcal{F}$ is positive definite (and in fact also self-adjoint), so that the Newton direction is well-defined. However, in the second, a Lyapunov system must again be solved to apply $\mathcal{E}^{-1}$ to a matrix. For the first case, we define M by setting $m_{ij} = A_i \bullet (XA_jS^{-1})$, while for the last, $m_{ij} = A_i \bullet (WA_jW)$. We then solve (13) for Δy, set $\Delta S = -\mathcal{A}^*\Delta y$, and then set

$$\Delta X = -X + \nu S^{-1} + \frac{1}{2}[X(\mathcal{A}^*\Delta y)S^{-1} + S^{-1}(\mathcal{A}^*\Delta y)X]$$

for the first case, and

$$\Delta X = -X + \nu S^{-1} + W(\mathcal{A}^*\Delta y)W$$

for the last. Once again, damped steps are taken to preserve positive definiteness. We still need a proximity criterion, and here two possibilities have been considered. In both, we let $\mu := \mu(X,S) := (X \bullet S)/n$. Then the narrow neighbourhood (parametrised by $\tau \in (0,1)$) is

$$\mathcal{N}_F(\tau) := \{(X,y,S) \in F^0(P) \times F^0(D) : \|X^{\frac{1}{2}}SX^{\frac{1}{2}} - \mu I\|_F = \|\lambda(XS - \mu I)\|_2 \leq \tau\mu\},$$

while the wide neighbourhood is

$$\mathcal{N}_{-\infty}(\tau) := \{(X, y, S) \in F^0(P) \times F^0(D) : \lambda_{\min}(XS) \geq (1-\tau)\mu\}.$$

Algorithms that maintain all iterates in a narrow neighbourhood are called short-step methods, while those that keep the iterates in a wide neighbourhood are termed long-step methods. In practice, algorithms frequently ignore such criteria and just take steps a proportion α (say .99) of the way to the boundary; different steps can be taken for the primal and dual iterates.

Here is a typical short-step primal-dual path-following algorithm. Assume given an initial strictly feasible point $(X, y, S) \in \mathcal{N}_F(\tau)$. Choose $\nu = \sigma\mu$ for some $\sigma \in (0, 1)$, compute the search direction chosen from the AHO, HRVW/KSH/M, dual HRVW/KSH/M, and NT search directions, and take a full Newton step. Repeat.

Monteiro [40] showed that such an algorithm, with $\tau = .1$ and $\sigma = 1 - .1/\sqrt{n}$, generates a sequence of iterates all in the narrow neighbourhood, and produces a strictly feasible point with duality gap at most ϵ times that of the original point in $O(\sqrt{n}\ln(1/\epsilon))$ steps. (Included in this is the result of Monteiro and Zanjacomo [42] that the AHO search direction is well-defined within such a narrow neighbourhood.) Predictor-corrector methods, which alternate taking $\sigma = 1$ (with a line search) and $\sigma = 0$, and use two sizes of narrow neighbourhood, also have the same complexity. Also see Monteiro and Todd [41].

A typical long-step primal-dual path-following algorithm assumes given an initial strictly feasible point $(X, y, S) \in \mathcal{N}_{-\infty}(\tau)$. Choose $\nu = \sigma\mu$ for some $\sigma \in (0, 1)$, compute the search direction chosen from the AHO, HRVW/KSH/M, dual HRVW/KSH/M, and NT search directions, and take the longest step that keeps the iterate in $\mathcal{N}_{-\infty}(\tau)$. Here it is not certain that the AHO search direction will be well-defined, so our theoretical results are for the other cases.

Monteiro and Zhang [43] showed that such an algorithm, with any τ and σ in $(0, 1)$ and independent of n, and using the NT search direction, generates a strictly feasible point with duality gap at most ϵ times that of the original point in $O(n\ln(1/\epsilon))$ steps; using the HRVW/KSH/M or dual HRVW/KSH/M search direction increases the bound to $O(n^{3/2}\ln(1/\epsilon))$ steps. Again, another reference for these results is [41].

6.2 Potential-reduction methods

The methods of the previous subsection were based on approximately solving the barrier problems BP(ν) and BD(ν), and the parameter ν had to be explicitly adjusted towards zero. Here we combine the objective function and the barrier function in a different way, and avoid the need to adjust a parameter. Such *potential functions* were first introduced by Karmarkar in his seminal work on interior-point methods for linear programming [30].

Consider the Tanabe-Todd-Ye [58, 61] primal-dual potential function

$$\Psi_\rho(X, y, S) := (n+\rho)\ln X \bullet S - \ln\det X - \ln\det S - n\ln n,$$

defined for strictly feasible points (X, y, S). If $\lambda := \lambda(X^{\frac{1}{2}}SX^{\frac{1}{2}})$, then it is easy to see that $\Psi_0(X, y, S) = n\ln(e^T\lambda/n) - \ln(\Pi_j\lambda_j)$, so the arithmetic-geometric mean inequality shows that this is always nonnegative. In fact, it is zero iff all eigenvalues of $X^{\frac{1}{2}}SX^{\frac{1}{2}}$ are equal, or equivalently iff the point is on the central path. $\Psi_\rho(X, y, S)$ increases the weight on the logarithm of the duality gap, and therefore pushes points towards the optimum. Our aim is to decrease this function by a constant at each iteration:

Theorem 6.1 *Suppose* $(X_0, y_0, S_0) \in F^0(P) \times F^0(D)$ *satisfies*

$$\Psi_0(X_0, y_0, S_0) \le \rho \ln \frac{1}{\epsilon}$$

for some $\epsilon > 0$. *Then, if we generate a sequence of strictly feasible points* (X_k, y_k, S_k) *with*

$$\Psi_\rho(X_k, y_k, S_k) \le \Psi_\rho(X_{k-1}, y_{k-1}, S_{k-1}) - \delta$$

for some constant $\delta > 0$ *and all* $k \ge 1$, *then in* $O(\rho \ln(1/\epsilon))$ *steps, we will have a strictly feasible point* (X_K, y_K, S_K) *with duality gap at most* ϵ *times that of* (X_0, y_0, S_0).

Proof:
Let $K := 2\rho \ln(1/\epsilon)/\delta$. Then, using the fact above, we have

$$\begin{aligned}
\rho \ln X_K \bullet S_K &\le \rho \ln X_K \bullet S_K + \Psi_0(X_K, y_K, S_K) \\
&= \Psi_\rho(X_K, y_K, S_K) \\
&\le \Psi_\rho(X_0, y_0, S_0) - K\delta \\
&= \rho \ln X_0 \bullet S_0 + \Psi_0(X_0, y_0, S_0) - K\delta \\
&\le \rho \ln X_0 \bullet S_0 - \rho \ln \frac{1}{\epsilon}.
\end{aligned}$$

$\square$

Notice that there is no need to control the proximity of the iterates to the central path, as long as the requisite decrease in the potential function can be obtained. It turns out that this is possible as long as $\rho \ge \sqrt{n}$. A reasonable way to try to effect such a decrease is to move in the direction of steepest descent with respect to some norm.

Let us consider first a dual method. Suppose our current dual strictly feasible iterate is (y, S), and that we have available a primal strictly feasible solution X (in fact, initially it is only necessary to have an upper bound on the dual optimal value). Let $\nabla_S \Psi$ denote the derivative of Ψ_ρ with respect to S,

$$\nabla_S \Psi = \frac{n + \rho}{X \bullet S} X - S^{-1},$$

let U be positive definite, and consider

$$\min \nabla_S \Psi \bullet \Delta S + \frac{1}{2} \|\Delta S\|_U^2, \quad \mathcal{A}^* \Delta y + \Delta S = 0. \tag{16}$$

Of course, it is natural to take $U = S$, but we shall soon see the value of the generality we have allowed. For now, let us choose $U = S$ and see what the resulting direction is. If we let P denote the Lagrange multiplier for the constraint, we need to solve

$$\mathcal{A}P = 0, \quad \mathcal{A}^* \Delta y + \Delta S = 0, \quad P + S^{-1} \Delta S S^{-1} = -\nabla_S \Psi. \tag{17}$$

Let us set $\nu := (X \bullet S)/(n+\rho)$. Then the last equation above, multiplied by ν, becomes $(\nu P) + \nu S^{-1}\Delta S S^{-1} = -X + \nu S^{-1}$. It follows that $(\Delta y, \Delta S)$ is exactly the same as the search direction in the dual path-following algorithm — see the paragraph including (13) — for this value for ν. If the resulting $\|\Delta S\|_S$ is sufficiently large, then a suitable step is taken in the direction $(\Delta y, \Delta S)$ and one can show that the potential function is thus decreased by a constant (X is unchanged). If not, then the solution of the problem above suffices to generate an improved X, exactly as we found below (13), and then updating X while holding (y, S) unchanged also can be shown to give a constant decrease in the potential function. It follows that we can attain the iteration complexity bound given in Theorem 6.1. Details can be found in, for example, Benson et al. [8], which describes why this method is attractive for SDP problems arising in combinatorial optimization problems and gives some excellent computational results.

Now let us consider a symmetric primal-dual method. Suppose we have a strictly feasible point (X, y, S). In addition to the dual direction-finding problem (16) above, we need a primal problem to determine ΔX. Let $\nabla_X \Psi$ denote the derivative of Ψ_ρ with respect to X,

$$\nabla_X \Psi = \frac{n+\rho}{X \bullet S} S - X^{-1},$$

let V be positive definite, and consider

$$\min \nabla_X \Psi \bullet \Delta X + \frac{1}{2}\|\Delta X\|_V^2, \quad \mathcal{A}\Delta X = 0. \tag{18}$$

Here it is natural to choose $V = X$, and this would lead to a primal potential-reduction method with the same iteration complexity. But we would like to get search directions for both primal and dual problems without solving two optimization subproblems. This can be achieved by using $V = W$ in (18) and $U = W^{-1}$ in (16), where W is the scaling matrix of (15). The dual direction then comes from equations like (17), with W replacing S^{-1} on the left-hand side of the last equation. The primal direction, if we use a Lagrange multiplier q for the constraint, comes from the solution to

$$\mathcal{A}\Delta X = 0, \quad W^{-1}\Delta X W^{-1} - \mathcal{A}^* q = -\nabla_X \Psi. \tag{19}$$

If we write R for $-\mathcal{A}^* q$ and pre- and postmultiply the last equation by W (noting that $WSW = X$ and $WX^{-1}W = S^{-1}$), we get

$$\mathcal{A}\Delta X = 0, \quad \mathcal{A}^* q + R = 0, \quad \Delta X + WRW = -\nabla_S \Psi.$$

Comparing these two systems, we see that they are *identical* if we identify ΔX with P and $(\Delta y, \Delta S)$ with (q, R). It thus turns out that both search directions can be obtained simultaneously by solving one system of the form (10). In fact, the search directions are exactly (up to a scalar factor) those of the NT path-following method of the previous subsection, and we have already discussed how those can be computed. (Again, we need to take $\nu = X \bullet S/(n+\rho)$.) It turns out that, by taking a suitable step in these directions, we can again achieve a constant decrease in the potential function. The analysis is somewhat complicated, and the reader is referred to the original article of Nesterov and Todd [46], the subsequent paper [47] which gives a simplified proof for the key Theorem 5.2 in the first paper, and the paper of Tunçel [62] which provides an easier analysis for the SDP case.

The important point again is that a constant decrease leads easily (via Theorem 6.1) to the best known complexity bound for the number of iterations, and that this is achieved without any concern for the iterates staying close to the central path, yielding great flexibility for the algorithms.

6.3 Nonlinear programming approaches

Finally we turn to methods that are based on nonsmooth or smooth optimization techniques for nonlinear programming formulations of (P) or (D). Some of these place restrictions on the SDP problems that can be handled.

First we discuss nonsmooth methods for minimizing the maximum eigenvalue of a matrix which depends affinely on some parameters. This was our first example in Section 3, but it is remarkably general. Suppose X is bounded for feasible solutions to (P). Then we can add an inequality on the trace of X, and by adding a slack variable and making a block diagonal matrix, we can assume that the trace of X is fixed at some positive value; by scaling, we suppose this is 1. So we assume that $\operatorname{trace} X = 1$ for all feasible X. Note that this holds for Examples 8 and (after scaling) 9. If we add this constraint explicitly, the dual problem then becomes to minimize $\lambda_{\max}(\mathcal{A}^*y - C) - b^T y$ over $y \in I\!R^m$ (we switched to a minimization problem by changing the sign of the objective). We can also assume that the linear objective $b^T y$ does not appear by incorporating it into the first term (each A_i is replaced by $A_i - b_i I$). Hence any such constant trace problem has a dual that is a maximum eigenvalue minimization problem.

We now have a convex but nonsmooth optimization problem, to which standard methods of nonlinear programming can be applied. One such is the bundle method, which builds up a cutting-plane model of the objective function by computing subgradients of the maximum eigenvalue function. Let us set $g(y) := \lambda_{\max}(\mathcal{A}^*y - C)$. A *subgradient* of g at y is a vector z with $g(y') \geq g(y) + z^T(y' - y)$ for all y'; and in our case, one can be found as $\mathcal{A}(vv^T)$, where v is an eigenvector of $\mathcal{A}^*y - C$ associated with its maximum eigenvalue. It is useful also to consider so-called ϵ-*subgradients* for $\epsilon > 0$: z is one such if

$$g(y') \geq g(y) + z^T(y' - y) - \epsilon$$

for all y', and the set of them all is called the ϵ-*subdifferential* $\partial g_\epsilon(y)$. In our case this turns out to be

$$\partial g_\epsilon(y) = \{\mathcal{A}W : (\mathcal{A}^*y - C) \bullet W \geq \lambda_{\max}(\mathcal{A}^*y - C) - \epsilon, \operatorname{trace} W = 1, W \succeq 0\}.$$

Helmberg and Rendl [27] develop a very efficient algorithm, the *spectral bundle method*, by modifying the classical bundle method to exploit this structure. From the result above, it is easy to see that

$$g(y') \geq (\mathcal{A}^*y - C) \bullet W + (\mathcal{A}W)^T(y' - y) = (\mathcal{A}^*y' - C) \bullet W$$

for any $W \succeq 0$ with trace 1 and any y'. Hence if we choose any subset $\mathcal{W}$ of such matrices,

$$g(y') \geq \hat{g}_{\mathcal{W}}(y') := \max\{(\mathcal{A}^*y' - C) \bullet W : W \in \mathcal{W}\}.$$

At every iteration, Helmberg and Rendl generate a search direction d for the current iterate y by minimizing $\hat{g}_{\mathcal{W}}(y+d) + (u/2)d^T d$ for some regularizing parameter u and some $\mathcal{W}$. Let $P \in \mathbb{R}^{n\times k}$ have orthonormal columns (think of them as approximate eigenvectors corresponding to almost maximal eigenvalues of $\mathcal{A}^*y - C$), and let $\bar{W} \succeq 0$ have trace 1 (think of this as a matrix containing useful past information). Then the spectral bundle method chooses

$$\mathcal{W} := \{\alpha\bar{W} + PVP^T : \alpha + \operatorname{trace} V = 1, \alpha \geq 0, V \succeq 0\}.$$

The dual of the direction-finding subproblem turns out to be an SDP problem with a quadratic objective function in a lower dimensional space (V is of order k). This problem is solved to yield the search direction d and a new y trial vector is computed. If there is a suitable improvement in the objective function, this new point replaces the old; otherwise we stay at the old point. In either case, an approximate eigenvector corresponding to the maximum eigenvalue of the trial $\mathcal{A}^*y - C$ is computed, and this is added as a column to the P matrix. If there are too many columns, old information is incorporated into the aggregate matrix $\bar{W}$, and the process continues. Many details have been omitted, but the rough idea of the method is as above; it can be thought of as providing an approximation by considering only a subset of feasible X matrices, using this to improve the dual solution y, and using this in turn to improve the subset of feasible solutions in the primal.

As a version of the bundle method, the algorithm above has good global convergence properties, but no iteration bounds as for the interior-point methods of the previous subsections are known. Nevertheless, excellent computational results have been obtained for problems that are inaccessible to the latter methods due to their size; see [27].

It is known that for smooth optimization problems, second-order methods are much more attractive than first-order techniques such as the spectral bundle method, but it is not clear how second-order information can be incorporated in nonsmooth optimization. However, for the maximum eigenvalue problem, this is possible: Oustry [48, 49] devises the so-called *$\mathcal{U}$-Lagrangian* of the maximum eigenvalue function, uses this to get a quadratic approximation to the latter along a manifold where the maximum eigenvalue has a fixed multiplicity, and then develops a second-order bundle method using these ideas. This method retains the global convergence of the first-order method, but also attains asymptotic quadratic convergence under suitable regularity conditions. These bundle methods are further discussed, and improved computational results given, in Helmberg and Oustry [26].

Fukuda and Kojima [19] have recently proposed an interior-point method for the same class of problems, working just in the space of y to avoid difficulties for large-scale problems. This paper also has an excellent discussion of recent attempts to solve such problems efficiently. Note that Vavasis [65] has developed an efficient way to compute the barrier and its gradient for this dual formulation.

Now we turn to methods that generate nonconvex nonlinear programming problems in a lower dimension, and apply interior-point or other techniques for their solution. Suppose first that (P) includes constraints specifying the diagonal entries of X:

$$(P): \quad \min C \bullet X, \quad \operatorname{diag}(X) = d,\ \mathcal{A}X = b,\ X \succeq 0,$$

with dual problem

$$(D): \quad \max d^T z + b^T y, \quad \mathrm{Diag}(z) + \mathcal{A}^* y + S = C,\ S \succeq 0.$$

Burer, Monteiro, and Zhang [10] suggest solving (D) by an equivalent nonlinear programming problem obtained by eliminating variables. In fact, they only consider strictly feasible solutions of (D). Their procedure is based on a theorem stating that, for each $(w, y) \in \mathbb{R}^n_{++} \times \mathbb{R}^m$, there is a unique strictly lower triangular matrix $\bar{L} = \bar{L}(w, y)$ and a unique $z = z(w, y) \in \mathbb{R}^n$ satisfying

$$C - \mathrm{Diag}(z) - \mathcal{A}^* y = (\mathrm{Diag}(w) + \bar{L})(\mathrm{Diag}(w) + \bar{L})^T,$$

and that $\bar{L}(w, y)$ and $z(w, y)$ are infinitely differentiable. This takes care of the constraint that S be positive definite implicitly by requiring it to have a nonsingular Cholesky factorization. (D) is then replaced by the smooth but nonconvex problem

$$(D'): \quad \max_{w,y} d^T z(w, y) + b^T y,\ w > 0.$$

The authors then suggest algorithms to solve this problem: a log-barrier method and a potential-reduction method. A subsequent paper relaxes the requirement that the diagonal of X be fixed. Instead, they require in [11] that the diagonal be bounded below, so the first constraint becomes $\mathrm{diag}(X) \geq d$. This constraint can be without loss of generality, since it holds for any positive semidefinite matrix if we choose the vector d to be zero. The corresponding change to (D) is that now z must be nonnegative, and so the constraint $z(w, y) > 0$ is added to (D′) (as we noted, Burer et al. only consider strictly feasible solutions to (D)). Once again, they consider log-barrier and potential-reduction methods to solve (D′). Although the problem (D′) is nonconvex, Burer, Monteiro, and Zhang prove global convergence of their methods, and have obtained some excellent computational results on large-scale problems.

Finally, we mention the approach of Vanderbei and Yurttan Benson [64]: the primal variable X is factored as $L(X)\,\mathrm{Diag}(d(X))L(X)^T$, where $L(X)$ is unit lower triangular and $d(X) \in \mathbb{R}^n$, and the constraint that X be positive semidefinite is replaced with the requirement that $d(X)$ be a nonnegative vector. The authors show that d is a concave function, and give some computational results for this reformulation.

We should mention that research is very active in new methods to solve large sparse SDP problems. The reader is urged to consult the web pages of Helmberg [25] and Wright [68] to see the latest developments.

7 Concluding remarks

We have investigated semidefinite programming from several viewpoints, examining its applications, duality theory, and several algorithms for solving SDP problems. The area has a rich history, drawing from several fields, and recently powerful methods for solving small- and medium-scale problems have been developed. The interior-point methods we have discussed can solve most problems with up to about a thousand linear constraints and matrices of order up to a thousand or so. However, as problems get larger, it is not clear that this class of methods can successfully compete with special-purpose algorithms that better exploit sparsity, and we have also considered a number of these. The limitations of such methods are being reduced, and they have successfully solved problems with matrices of order 10,000 and more. One limitation is that these more efficient methods usually solve the dual problem, and if a primal near-optimal solution is required (as in the max-cut problem using the technique of Goemans and Williamson to generate a cut), they may not be as appropriate. The topic remains exciting and vibrant, and significant developments can be expected over the next several years.

Acknowledgements. I would like to thank Michael Overton, Jos Sturm, and Henry Wolkowicz for helpful comments on this paper.

References

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

[2] F. Alizadeh. Semidefinite programming home page. `http://karush.rutgers.edu/~alizadeh/Sdppage/index.html`

[3] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. A new primal-dual interior-point method for semidefinite programming. In: J. G. Lewis, ed., *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra*, SIAM, Philadelphia, USA, 1994, pp. 113–117.

[4] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM J. Optim.*, 8:746–768, 1998.

[5] F. Alizadeh and S. Schmieta. Symmetric cones, potential reduction methods and word-by-word extensions. In: H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., *Handbook of Semidefinite Programming*, Kluwer Academic Publishers, Boston-Dordrecht-London, 2000, pp. 195 – 233.

[6] R. Bellman and K. Fan. On systems of linear inequalities in Hermitian matrix variables. In: *Convexity*, Proceedings of Symposia in Pure Mathematics, Vol. 7, American Mathematical Society, Providence, RI, 1963, pp. 1–11.

[7] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23:769–805, 1998.

[8] S. J. Benson, Y. Ye, and X. Zhang, Solving large-scale sparse semidefinite programs for combinatorial optimization. *SIAM J. Optim.*, 10:443–461, 2000.

[9] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* SIAM Studies in Applied Mathematics. SIAM, Philadelphia, USA, 1994.

[10] S. Burer, R. D. C. Monteiro, and Y. Zhang. Solving semidefinite programs via nonlinear programming II: Interior point methods for a subclass of SDPs. Tech-

nical Report TR99-23, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1999.

[11] S. Burer, R. D. C. Monteiro, and Y. Zhang. Interior point algorithms for semidefinite programming based on a nonlinear programming formulation. Technical Report TR99-27, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1999.

[12] J. Cullum, W. E. Donath, and P. Wolfe. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. *Math. Programming Study*, 3:35–55, 1975.

[13] W. E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM J. of Research and Development*, 17:420–425, 1973.

[14] C. Delorme and S. Poljak. Laplacian eigenvalues and the maximum cut problem. *Math. Programming*, 62:557–574, 1993.

[15] L. Faybusovich. Euclidean Jordan algebras and interior-point algorithms. *Positivity*, 1:331–357, 1997.

[16] L. Faybusovich. Linear systems in Jordan algebras and primal-dual interior-point algorithms. *Journal of Computational and Applied Mathematics*, 86:149–175, 1997.

[17] R. Fletcher. A nonlinear programming problem in statistics (educational testing). *SIAM J. Sci. Stat. Comput.*, 2:257–267, 1981.

[18] R. Fletcher. Semi-definite matrix constraints in optimization. *SIAM J. Control Optim.*, 23:493–513, 1985.

[19] M. Fukuda and M. Kojima. Interior-point methods for Lagrangian duals of semidefinite programs. Technical Report B-365, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, 2000.

[20] M. Fukuda, M. Kojima, K. Murota, and K. Nakata. Exploiting sparsity in semidefinite programming via matrix completions I: General framework. *SIAM J. Optim.*, 11:647–674, 2000.

[21] M. X. Goemans, Semidefinite programming in combinatorial optimization. *Math. Programming*, 79:143–161, 1997.

[22] M. X. Goemans and D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[23] D. Goldfarb and K. Scheinberg. Interior point trajectories in semidefinite programming. *SIAM J. Optim.*, 8:871–886, 1998.

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

[25] C. Helmberg. Semidefinite programming home page. `http://www.zib.de/helmberg/semidef.html`

[26] C. Helmberg and F. Oustry. Bundle methods to minimize the maximum eigenvalue function. In: H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., *Handbook of Semidefinite Programming*, Kluwer Academic Publishers, Boston-Dordrecht-London, 2000, pp. 307 – 337.

[27] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM J. Optim.*, 10:673–696, 2000.

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

[29] F. Jarre. A QQP-minimization method for semidefinite and smooth nonconvex programs. Technical Report, Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Germany, 1999.

[30] N. K. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[31] E. de Klerk, C. Roos, and T. Terlaky. Initialization in semidefinite programming via a self-dual skew-symmetric embedding. *Operations Research Letters*, 20:213–221, 1997.

[32] M. Kojima, M. Shida, and S. Shindoh. Search directions in the SDP and the monotone SDLCP: generalization and inexact computation. *Math. Programming*, 85:51–80, 1999.

[33] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7:86–125, 1997.

[34] A. S. Lewis and M. L. Overton. Eigenvalue optimization. *Acta Numerica*, 5:149–190, 1996.

[35] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:1–7, 1979.

[36] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optim.*, 1:166-190, 1991.

[37] Z.-Q. Luo, J. F. Sturm, and S. Zhang. Superlinear convergence of a symmetric primal-dual path-following algorithm for semidefinite programming. *SIAM J. Optim.*, 8:59–81, 1998.

[38] Z.-Q. Luo, J. F. Sturm, and S. Zhang. Conic convex programming and self-dual embedding. *Optim. Methods Softw.*, 14:169–218, 2000.

[39] R. D. C. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optim.*, 7:663–678, 1997.

[40] R. D. C. Monteiro. Polynomial convergence of primal-dual algorithms for semidefinite programming based on the Monteiro and Zhang family of directions. *SIAM J. Optim.*, 8:797–812, 1998.

[41] R. D. C. Monteiro and M. J. Todd. Path-following methods. In: H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., *Handbook of Semidefinite Programming*, Kluwer Academic Publishers, Boston-Dordrecht-London, 2000, pp. 267 – 306.

[42] R. D. C. Monteiro and P. R. Zanjacomo. A note on the existence of the Alizadeh-Haeberly-Overton direction for semidefinite programming. *Math. Programming*, 78 (1997), 393–396.

[43] R.D.C. Monteiro and Y. Zhang. A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Math. Programming*, 81:281–299, 1998.

[44] Yu. E. Nesterov and A. S. Nemirovski. Conic formulation of a convex programming problem and duality. *Optim. Methods Softw.*, 1:95–115, 1992.

[45] Yu. E. Nesterov and A. S. Nemirovski. *Interior Point Polynomial Algorithms in Convex Programming.* SIAM Publications. SIAM, Philadelphia, USA, 1994.

[46] Yu. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22:1–42, 1997.

[47] Yu. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8:324–364, 1998.

[48] F. Oustry. The $\mathcal{U}$-Lagrangian of the maximum eigenvalue function. *SIAM J. Optim.*, 9:526–549, 1999.

[49] F. Oustry. A second-order bundle method to minimize the maximum eigenvalue function. *Math. Programming*, 89:1–33, 2000.

[50] M. L. Overton and R. S. Womersley. Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. *Math. Programming*, 62:321–357, 1993.

[51] S. Poljak, F. Rendl, and H. Wolkowicz. A recipe for semidefinite relaxation for $(0,1)$-quadratic programming. *Journal of Global Optimization*, 7:51–73, 1995.

[52] S. Poljak and Z. Tuza. Maximum cuts and largest bipartite subgraphs. In: W. Cook, L. Lovász, and P. Seymour, eds., *Combinatorial Optimization*, American Mathematical Society, Providence, RI, 1995, pp. 181–244.

[53] F. Potra and R. Sheng. On homogeneous interior-point algorithms for semidefinite programming. *Optim. Methods Softw.*, 9:161–184, 1998.

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

[55] M. Ramana, L. Tunçel, and H. Wolkowicz. Strong duality for semidefinite programming. *SIAM J. Optim.*, 7:641–662, 1997.

[56] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[57] N. Z. Shor. Dual quadratic estimates in polynomial and Boolean programming. *Annals of Operations Research*, 25:163–168, 1990.

[58] K. Tanabe. Centered Newton method for mathematical programming. In: *System Modeling and Optimization*, Springer-Verlag, NY, 1988, pp. 197–206.

[59] M. J. Todd. A study of search directions in interior-point methods for semidefinite programming. *Optim. Methods Softw.*, 11&12:1–46, 1999.

[60] M. J. Todd, K.-C. Toh, and R. H. Tütüncü. On the Nesterov-Todd direction in semidefinite programming. *SIAM J. Optim.*, 8:769–796, 1998.

[61] M. J. Todd and Y. Ye, A centered projective algorithm for linear programming. *Math. Oper. Res.*, 15:508–529, 1990.

[62] L. Tunçel. Potential reduction and primal-dual methods. In: H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., *Handbook of Semidefinite Programming*, Kluwer Academic Publishers, Boston-Dordrecht-London, 2000, pp. 235 – 265.

[63] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38:49–95, 1996.

[64] R. J. Vanderbei and H. Yurttan Benson. On formulating semidefinite programming problems as smooth convex nonlinear optimization problems. Report ORFE 99-01, Operations Research and Financial Engineering, Princeton, NJ, 1999.

[65] S. A. Vavasis. A note on efficient computation of the gradient in semidefinite programming. Technical Report, Department of Computer Science, Cornell University, Ithaca, NY, 1999.

[66] H. Wolkowicz. Bibliography on semidefinite programming. `http:liinwww.ira.uka.de/bibliography/Math/psd.html`

[67] H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.). *Handbook of Semidefinite Programming.* Kluwer Academic Publishers, Boston-Dordrecht-London, 2000.

[68] S. J. Wright. Interior-point methods online home page. `http://www-unix.mcs.anl.gov/otc/InteriorPoint/`

[69] Y. Ye, M. J. Todd, and S. Mizuno, An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Math. Oper. Res.*, 19:53–67, 1994.

[70] Y. Zhang. On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming. *SIAM J. Optim.*, 8:365–386, 1998.