Numerical algebraic geometry and semidefinite programming

Jonathan D. Hauenstein [a] Alan C. Liddell, Jr.[b] Sanesha McPherson[c] and Yi Zhang[d]

[a] Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN USA; [b] Vidrio Technologies, Ashburn, VA USA; [c] Department of Mathematics and Physics, North Carolina Central University, Durham, NC USA; [d] Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC USA

ARTICLE HISTORY
Compiled June 23, 2021

ABSTRACT
Standard interior point methods in semidefinite programming can be viewed as tracking a solution path for a homotopy defined by a system of bilinear equations. By considering this in the context of numerical algebraic geometry, we employ numerical algebraic geometric techniques such as adaptive precision path tracking, endgames, and projective space to accurately solve semidefinite programs. We develop feasibility tests for both primal and dual problems which can distinguish between the four feasibility types of semidefinite programs. Finally, we couple our feasibility tests with facial reduction to develop a solving approach that can handle every scenario arising in semidefinite programming, including problems with nonzero duality gap. Various examples are used to demonstrate the new methods with comparisons to commonly used semidefinite programming software.

1. Introduction

Semidefinite programs are nonlinear convex optimization problems arising in many applications in engineering, control, and combinatorial optimization, e.g., see [1, 9, 13, 15, 45, 48, 52] and the references contained therein. The primal form of a semidefinite program is

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \qquad i = 1,\ldots,m, \\ & X \succeq 0, \end{aligned} \tag{SDP-P}$$

where $b = (b_1,\ldots,b_m) \in \mathbb{R}^m$ and $A_1,\ldots,A_m, C, X \in \mathbb{R}^{n\times n}$ are all symmetric matrices with $\langle R, S\rangle = \text{trace}\,(R^T S) = \text{trace}\,(RS)$. The inequality $X \succeq 0$ means that X is a symmetric positive semidefinite matrix, i.e., every eigenvalue of X is nonnegative. Since the set of symmetric positive semidefinite matrices is a convex set, (SDP-P) is a convex program that minimizes a linear function over a spectahedron [9] which is a linear slice of the cone of symmetric positive semidefinite matrices. The corresponding

CONTACT J.D.H. Author. Email: hauenstein@nd.edu

dual form of (SDP-P) is

$$\begin{aligned} \underset{S,y}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & C - \sum_{i=1}^{m} y_i A_i = S, \\ & S \succeq 0. \end{aligned} \tag{SDP-D}$$

A standard approach in optimization is to consider the first-order Karush-Kuhn-Tucker (KKT) optimality conditions. If both (SDP-P) and (SDP-D) have feasibility sets with a nonempty interior, i.e., *strictly feasible*, then the KKT conditions are both necessary and sufficient for solving (SDP-P) and (SDP-D), namely:

$$\begin{aligned} \langle A_i, X \rangle &= b_i, \qquad i = 1, , \ldots, m, \\ C - \sum_{i=1}^{m} y_i A_i &= S, \\ SX &= 0, \\ X, S &\succeq 0. \end{aligned} \tag{KKT}$$

Interior point methods, e.g., see [1, 2, 23, 34–36, 54] and the references contained therein, based on the KKT conditions using the barrier function $\mu \log \det X$ yield

$$\begin{aligned} \langle A_i, X \rangle &= b_i, \qquad i = 1, \ldots, m, \\ C - \sum_{i=1}^{m} y_i A_i &= S, \\ SX &= \mu I, \\ X, S &\succ 0 \end{aligned} \tag{1}$$

where I is the $n \times n$ identity matrix. When $\mu > 0$, the matrices X and S arising in the unique solution of (1) are required to be positive definite, i.e., every eigenvalue is positive, denoted $X, S \succ 0$. Hence, upon removing the positive definite condition from (1), the remaining system of equations is simply a bilinear homotopy parameterized by μ which defines the *central path*. This permits techniques from numerical algebraic geometry, e.g., see [6, 43], to be employed to solve semidefinite optimization problems.

When both (SDP-P) and (SDP-D) are strictly feasible, the central path converges to optimizers of both (SDP-P) and (SDP-D), say X^* and (S^*, y^*), respectively. That is, the optimal values are achieved with $\langle C, X^* \rangle = b^T y^*$, i.e., a *zero duality gap*. However, when both (SDP-P) and (SDP-D) are not strictly feasible, the optimal values need not be achieved and the duality gap may be nonzero. The papers [26, 37, 50, 51] demonstrate the difficulty of current semidefinite programming software to identify cases such as nonzero duality gaps and so-called weakly infeasible programs. By utilizing techniques from numerical algebraic geometry, we develop a solving approach which handles every possible scenario in semidefinite programming.

The rest of the paper is organized as follows. Section 2 summarizes the three key aspects from numerical algebraic geometry which will be utilized in homotopy-based approach for solving (SDP-P) and (SDP-D): adaptive precision path tracking, endgames, and projective space. Section 3 applies this methodology to create feasibility tests for (SDP-P) and (SDP-D). In particular, we show that our approach can distinguish between the four feasibility types of semidefinite programs: strictly feasible, feasible but not strictly feasible, weakly infeasible, and strongly infeasible. Section 4 provides a complete solving approach that can handle every scenario arising in semidefinite programs. In addition to illustrative examples throughout, Section 5 includes various examples to demonstrate our numerical algebraic geometric approach with comparisons to other commonly used software for solving semidefinite programs. A short conclusion is provided in Section 6.

2. Primal-dual solvers and numerical algebraic geometry

As mentioned in the Introduction, we can view interior point methods as path tracking for a bilinear homotopy. In this section, we will consider homotopies for solving (SDP-P) and (SDP-D) and three computational techniques from numerical algebraic geometry, namely adaptive precision path tracking, endgames, and projective spaces. For more details on numerical algebraic geometry, see the books [6, 43]. Some of these ideas have already been utilized in optimization, e.g., [16–18, 32, 33, 38, 53].

2.1. Interior point homotopy

When both (SDP-P) and (SDP-D) are strictly feasible, that is, their feasibility sets have a nonempty interior, a standard solving approach is to utilize an interior point method. That is, one first computes an interior point that (approximately) lies on the central path for some $\mu > 0$ defined by the *interior point homotopy*

$$H(X,S,y;\mu) = \begin{bmatrix} \langle A_i, X\rangle - b_i & i=1,\ldots,m \\ C - \sum_{i=1}^{m} y_i A_i - S & \\ SX - \mu I & \end{bmatrix} = 0. \tag{H}$$

This is typically called the *Phase I* stage. The *Phase II* stage is to track along the central path as $\mu \to 0^+$ which limits to a primal-dual solution. Since only the constant terms are changing, (H) is a so-called *Newton homotopy*, e.g., see [20].

We can simplify (H) to a well-constrained homotopy, i.e., one which has the same number of variables and equations, as follows. The total number of variables is $n^2 + n + m$ since S and X are symmetric $n \times n$ matrices and y is a vector of length m. In terms of equations, the first collection in (H) consists of m linear equations. The next equation is a linear matrix equation of symmetric matrices so this yields $(n^2+n)/2$ linear equations. Thus, to have a well-constrained system, we need the final matrix equation, arising from the complimentary slackness condition, to also yield $(n^2+n)/2$ equations. There are three common approaches for this: use the upper triangular part of this matrix equation, take the upper triangular part of the symmetrized version of this matrix equation [2], namely

$$\frac{1}{2}(SX + XS) = \mu I,$$

or apply techniques from [21] to adaptively select $(n^2+n)/2$ linear combinations of the n^2 bilinear equations based on local conditioning.

Remark 1. Since $S = C - \sum_{i=1}^m y_i A_i$, one can trivially eliminate S to reduce from a well-constrained homotopy involving $n^2 + n + m$ equations and variables as written in (H) to one involving $(n^2+n)/2+m$ equations and variables. Similar reductions are possible for all of the homotopies presented below.

With interior point methods and homotopy continuation, there is a start point, say $z^* := (X^*, S^*, y^*)$ corresponding with $\mu^* > 0$, i.e., $H(X^*, S^*, y^*; \mu^*) = 0$. Hence, by taking $\mu = t \cdot \mu^*$, $z = (X, S, y)$, and $N = n^2 + n + m$, we can view (H) as a well-constrained homotopy $H(z;t) : \mathbb{R}^N \times [0,1] \to \mathbb{R}^N$ and we wish to track the solution path, i.e., the central path, $z(t) : (0,1] \to \mathbb{R}^N$ defined by $z(1) = z^*$ and $H(z(t);t) \equiv 0$ to compute $z(0)$. Even though $z(t)$ is a smooth path on $(0,1]$ for cases of interest, ill-conditioning along the path and particularly near $t = 0$ can cause numerical challenges, which is addressed in Section 2.2 using adaptive precision path tracking and endgames. To ameliorate scaling issues, we consider compactifying using projective space in Section 2.3 which will be essential in subsequent sections. These methods are implemented in the software package `Bertini` [5].

2.2. Adaptive precision path tracking and endgames

A solution path $z(t)$ for a homotopy $H(z;t)$ satisfies $H(z(t);t) \equiv 0$. Differentiating with respect to t, one sees that the path $z(t)$ satisfies the Davidenko differential equation

$$\dot{z}(t) = -J_z H(z;t)^{-1} \cdot J_t H(z;t) \tag{2}$$

where $J_z H(z;t)$ and $J_t H(z;t)$ are the Jacobian matrix and vector, respectively, with respect to z and t. Hence, one can track along $z(t)$ with the aim of computing $z(0)$ using a predictor-corrector strategy with adaptive stepsize and adaptive precision [4, 7, 8]. The predictor, e.g., Euler or Runge-Kutta predictor, is based on solving (2). The corrector, e.g., Newton's method, is based on solving $H = 0$. The stepsize and precision used in the computations are based on local conditioning along the path as well as the history of success or failure of previous steps as described in [4, 7, 8]. In fact, even though the path is smooth on $(0,1]$, it may become ill-conditioned, e.g., near $t = 0$ as one approaches an optimizer for a problem which has infinitely many optimizers. The advantage of using such adaptive approaches is to avoid unnecessary computational costs resulting from employing too high precision or too small steps when not needed.

Since the goal is to numerically approximate $z(0)$ and ill-conditioning can naturally arise near $t = 0$, one switches from path tracking to using endgames to compute accurate approximations of $z(0)$ using points along the path $z(t)$ for selected values of t near 0. See, e.g., [6, Chap. 3] for more details. The essential aspect of endgames in the context of interior point methods is that $z(t)$ can be written as a Puiseux series in a neighborhood of $t = 0$, called the *endgame operating zone*. That is, there exists $c \in \mathbb{N}$, called the *cycle number* of the path, and coefficients $a_i \in \mathbb{R}^N$ such that

$$z(t) = \sum_{i=0}^{\infty} a_i \cdot t^{i/c}$$

for all t in a neighborhood of 0. Since the goal is to approximate $z(0) = a_0$, the power series endgame [30] uses values along the path $z(t)$ near $t = 0$ to first determine the cycle number c and then uses interpolation to approximate the coefficients up to a given order based on the number of sample points selected. The Cauchy endgame [29] first determines c as the minimum number of loops encircling $t = 0$ over $\mathbb{C}$ needed to return to the starting value, i.e., the winding number. Since $s \mapsto z(s^c)$ is analytic in a neighborhood of $s = 0$, the Cauchy integral theorem yields $z(0)$ by integrating along the loops encircling 0. Numerical integration using the trapezoid rule is exponentially convergent due to periodicity, e.g., see [47]. The endpoint can be computed to arbitrary accuracy by rerunning the endgame, e.g., see [6, § 7.2.2], and using deflation techniques to restore local quadratic convergence of Newton's method [22, 25].

Example 2.1. To illustrate, consider solving the semidefinite program

$$\begin{array}{ll} \text{minimize} & 2x_{11} - 1 \\ \text{subject to} & \begin{bmatrix} x_{11} & 0 & -x_{22}/2 \\ 0 & x_{22} & 0 \\ -x_{22}/2 & 0 & x_{11} - 1 \end{bmatrix} \succeq 0 \end{array}$$

which corresponds with (SDP-P) where $n = 3$, $m = 4$, $b = (0, 0, 0, 1)$,

$$C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \; A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$
$$A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \; A_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

For $\mu(t) = t$ and $\alpha(t) = (-3t + \sqrt{9t^2 + 4})/2$, the corresponding solution path of (H) is

$$X(t) = \begin{bmatrix} (3t + \alpha(t) + 1)/2 & 0 & -\sqrt{t\alpha(t) + 3t^2}/2 \\ 0 & \sqrt{t\alpha(t) + 3t^2} & 0 \\ -\sqrt{t\alpha(t) + 3t^2}/2 & 0 & (3t + \alpha(t) - 1)/2 \end{bmatrix},$$
$$S(t) = \begin{bmatrix} 1 - \alpha(t) & 0 & \sqrt{t\alpha(t)} \\ 0 & \sqrt{t\alpha(t)} & 0 \\ \sqrt{t\alpha(t)} & 0 & 1 + \alpha(t) \end{bmatrix}, \; y(t) = \begin{bmatrix} 1 \\ 1 - \sqrt{t\alpha(t)} \\ 1 \\ \alpha(t) \end{bmatrix}$$

which is illustrated in Figure 1(a).

Ill-conditioning for this path near $t = 0$ arises from two places. First, the cycle number is $c = 2$ so that the solution path has an infinite derivative at $t = 0$. Second, the endpoint

$$X(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; S(0) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \; y(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

is actually an embedded point on the following linear solution component of (H) parameterized by y_3:

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; S = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1-y_3 \\ 0 & 1-y_3 & 2 \end{bmatrix}, \; y = \begin{bmatrix} 1 \\ 1 \\ y_3 \\ 1 \end{bmatrix}.$$

Figure 1(b) plots the condition number of the Jacobian matrix of (H) with respect to the variables along the path. By reparameterizing with $t = s^2$, the solution path is analytic on $s \in [0, 1]$, i.e., the cycle number with respect to s is 1. Figure 1(c) plots the original path parameterized by t, which is not analytic at $t = 0$ since it has a vertical tangent at $t = 0$, and its reparameterization by s with $t = s^2$, which is analytic at $s = 0$.

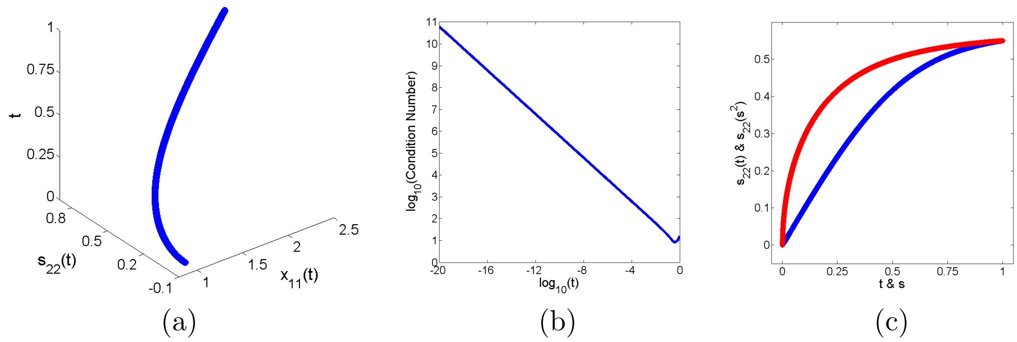

Figure 1. Plot of (a) two coordinates of the path, (b) condition number, and (c) comparison of $s_{22}(t)$ (red) which is not analytic at $t = 0$ due to a vertical tangent and $s_{22}(s^2)$ (blue) which is analytic at $s = 0$.

Remark 2. When the cycle number is not known *a priori* but is expected to be larger than 1, i.e., the path is not analytic at $t = 0$, reparameterizing by $t = s^r$ by some integer $r > 1$ can help to improve the performance of the endgame for accurately approximating the endpoint. For example, if $\delta \in (0, 1)$ such that $\{|t| < \delta\}$ is contained in the endgame operating zone with respect to t, then $\{|s| < \sqrt[r]{\delta}\}$ is contained in the endgame operating zone with respect to s where $t = s^r$. Increasing the endgame operating zone is counterbalanced by a potential increase in ill-conditioning requiring higher precision computations. Thus, when the cycle number is not known *a priori*, reasonable choices are $r = 2$, 3, or 4 (For example, we use $r = 4$ in Section 5.2.).

2.3. Projective space

Since (H) is bilinear, i.e., linear in X and linear in $\{S, y\}$, it is natural to view the variables as lying in the product space $\mathbb{R}^{(n^2+n)/2} \times \mathbb{R}^{(n^2+n)/2+m}$ corresponding to the primal and dual variables, respectively. To ameliorate scaling issues and to ensure that every solution path has finite length, we compactify this product space using projective space. We provide a brief summary with more details in, e.g., [43, Chap. 3].

Projective space $\mathbb{P}^a$, by definition, is the space of all lines in $\mathbb{C}^{a+1}$ passing through the origin. The unique line passing through a nonzero point $(x_0, \ldots, x_a) \in \mathbb{C}^{a+1}$ and the origin is denoted $[x_0, \ldots, x_a] \in \mathbb{P}^a$. Since, for every $\lambda \neq 0$, the points $(x_0, \ldots, x_a)$

and $(\lambda x_0, \ldots, \lambda x_a)$ lie on the same line passing through the origin,

$$[x_0, \ldots, x_a] = [\lambda x_0, \ldots, \lambda x_a].$$

There is a natural embedding of $\mathbb{C}^a$ into $\mathbb{P}^a$ via $(x_1, \ldots, x_a) \mapsto [1, x_1, \ldots, x_a]$. With this embedding, the *hyperplane at infinity* is $\mathcal{H} = \{[0, x_1, \ldots, x_a] \in \mathbb{P}^a\}$. In particular, $\mathbb{P}^a \setminus \mathcal{H} \cong \mathbb{C}^a$ with

$$[x_0, x_1, \ldots, x_a] \in \mathbb{P}^a \setminus \mathcal{H} \mapsto (x_1/x_0, \ldots, x_a/x_0) \in \mathbb{C}^a.$$

For a polynomial f defined on $\mathbb{C}^a$ of degree d, the homogenization of f is the homogeneous polynomial

$$f^h(x_0, \ldots, x_a) = x_0^d \cdot f(x_1/x_0, \ldots, x_a/x_0).$$

Hence, $\{f = 0\} \cong \{f^h = 0\} \setminus \mathcal{H}$ where $\{f^h = 0\} \subset \mathbb{P}^a$ is compact.

Example 2.2. The two circles defined by $f_1(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0$ and $f_2(x_1, x_2) = (x_1 - 1)^2 + x_2^2 - 1 = 0$ intersect in two points in $\mathbb{R}^2$, namely $(1/2, \pm\sqrt{3}/2)$ and two points at infinity. To see this, we have

$$f_1^h(y_0, y_1, y_2) = y_1^2 + y_2^2 - y_0^2 \quad \text{and} \quad f_2^h(y_0, y_1, y_2) = (y_1 - y_0)^2 + y_2^2 - y_0^2 = y_1^2 + y_2^2 - 2y_0y_1.$$

Hence, $\{f_1^h = f_2^h = 0\} = \{[2, 1, \pm\sqrt{3}], [0, 1, \pm\sqrt{-1}]\}$ with the first two being finite and the last two are contained in the hyperplane at infinity.

For (H), the embedding $\mathbb{R}^{(n^2+n)/2} \times \mathbb{R}^{(n^2+n)/2+m} \hookrightarrow \mathbb{P}^{(n^2+n)/2} \times \mathbb{P}^{(n^2+n)/2+m}$ yields

$$H^h([x_0, X], [s_0, S, y]; \mu) = \begin{bmatrix} \langle A_i, X\rangle - b_i x_0 \qquad i = 1, \ldots, m \\ s_0 C - \sum_{i=1}^{m} y_i A_i - S \\ SX - \mu x_0 s_0 I \end{bmatrix} = 0. \qquad (3)$$

Hence, this permits the independent rescaling of the homogenization of the primal and dual variables.

2.4. Primal-dual homotopy

The interior point homotopy (H) requires one to start with an interior point on the central path. One can modify the homotopy to create another solution path that still ends at a primal-dual solution when both (SDP-P) and (SDP-D) are strictly feasible. For this approach to work, the following assumption is needed:

(A1) The matrices $A_1, \ldots, A_m$ are linearly independent.

By using (numerical) linear algebra computations, we can always replace (SDP-P) with a problem where Assumption (A1) is satisfied and update (SDP-D) accordingly.

This modified approach builds the start point and homotopy simultaneously. To that end, arbitrary select $\widehat{y} \in \mathbb{R}^m$ and then compute $\sigma \in \mathbb{R}$ such that

$$\widehat{S} := C - \sum_{i=1}^{m} \widehat{y}_i A_i - \sigma I \succ 0. \tag{4}$$

For example, after selecting $\widehat{y}$, the choice of σ can be done either by directly computing the eigenvalues of the constant matrix $C - \sum_{i=1}^{m} \widehat{y}_i A_i$ or by bounding them, e.g., using Gershgorin's theorem.

Since $\widehat{S} \succ 0$, we next compute its inverse $\widehat{X} := \widehat{S}^{-1} \succ 0$ and evaluate the linear functions $\widehat{b}_i := \langle A_i, \widehat{X} \rangle$ for $i = 1, \dots, m$. Let $\widehat{b} = (\widehat{b}_1, \dots, \widehat{b}_m)$, and define $C_\mu := C - \mu\sigma I$ and $b_\mu := (1-\mu)b + \mu\widehat{b}$. Hence, it is straightforward to check that $\widehat{X}$ is strictly feasible for the following perturbation of (SDP-P) when $\mu = 1$:

$$\begin{array}{lll} \underset{X}{\text{minimize}} & \langle C_\mu, X \rangle & \\ \text{subject to} & \langle A_i, X \rangle = b_{\mu_i} & i = 1, \dots, m \\ & X \succeq 0. & \end{array} \tag{SDP-Pμ}$$

Also, $(\widehat{S}, \widehat{y})$ is strictly feasible for the following perturbation of (SDP-D) when $\mu = 1$:

$$\begin{array}{ll} \underset{S,y}{\text{maximize}} & b_\mu^T y \\ \text{subject to} & C_\mu - \sum_{i=1}^{m} y_i A_i = S \\ & S \succeq 0. \end{array} \tag{SDP-Dμ}$$

Note that (SDP-P) and (SDP-D) correspond to (SDP-Pμ) and (SDP-Dμ), respectively, when $\mu = 0$. Thus, the following system corresponds with (KKT) when $\mu = 0$:

$$\begin{aligned} \langle A_i, X \rangle &= b_{\mu_i}, \qquad i = 1, \dots, m, \\ C_\mu - \sum_{i=1}^{m} y_i A_i &= S, \\ SX &= \mu I, \\ X, S &\succ 0, \end{aligned} \tag{5}$$

and homotopy

$$H(X, S, y; \mu) = \begin{bmatrix} \langle A_i, X \rangle - b_{\mu_i} \qquad i = 1, \dots, m \\ C_\mu - \sum_{i=1}^{m} y_i A_i - S \\ SX - \mu I \end{bmatrix} = 0. \tag{H$_\mu$}$$

The following shows that we can use (H$_\mu$) to solve (SDP-P) and (SDP-D) when both are strictly feasible.

Theorem 2.3. *With the setup described above, if* (SDP-P) *and* (SDP-D) *are both strictly feasible and Assumption (A1) holds, then the solution path of* (H$_\mu$) *starting* *at* $(\widehat{X},\widehat{S},\widehat{y})$ *with* $\mu = 1$ *is smooth for* $\mu \in (0,1]$ *and ends at* $\mu = 0$ *at a solution of* (SDP-P) *and* (SDP-D).

Proof. Select X^0 and (S^0, y^0) which are strictly feasible for (SDP-P) and (SDP-D), respectively. Consider defining $X_\mu := (1-\mu)X^0 + \mu\widehat{X}$, $S_\mu := (1-\mu)S^0 + \mu\widehat{S}$, and $y_\mu := (1-\mu)y^0 + \mu\widehat{y}$. Hence,

$$C_\mu - \sum_{i=1}^{m} y_{\mu_i} A_i = S_\mu \qquad \text{and} \qquad \langle A_i, X_\mu \rangle = b_{\mu_i} \ \text{ for } \ i = 1,\dots,m.$$

Moreover, since $X^0, \widehat{X}, S^0, \widehat{S} \succ 0$, convexity yields $X_\mu, S_\mu \succ 0$ for $\mu \in [0,1]$. Thus, we have shown (SDP-Pμ) and (SDP-Dμ) are strictly feasible for $\mu \in [0,1]$ so that standard theory [19, 23, 35] shows that (5) has a unique solution for $\mu \in (0,1]$ producing a smooth path for $\mu \in (0,1]$. Since (KKT), which are both necessary and sufficient conditions in this case, corresponds with (5) when $\mu = 0$, the endpoint solves (SDP-P) and (SDP-D). The same result holds for the solution path defined by (H_μ) starting at $(\widehat{X},\widehat{S},\widehat{y})$ since the inequality conditions of (5) are trivially satisfied along the path for $\mu \in (0,1]$. □

Example 2.4. We illustrate on a simple example in order to plot the corresponding path and feasible sets for the original problem and perturbed problem. To that end, consider solving

$$\begin{array}{ll} \text{minimize} & 2x_{11} + 2x_{12} \\ \text{subject to} & \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & (1-2x_{11})/3 \end{bmatrix} \succeq 0 \end{array}$$

which corresponds with (SDP-P) with $n = 2$, $m = 1$, $b_1 = 1$,

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad A_1 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

It is easy to verify that both (SDP-P) and (SDP-D) are strictly feasible where the constraints in (SDP-P) are satisfied if and only if, for $r = 1$,

$$(4x_{11} - r)^2 + 24x_{12}^2 \le r^2. \tag{6}$$

We now employ the homotopy approach described above by arbitrarily selecting $\widehat{y} = 0.1$ and taking $\sigma = -1.7$. Rounding to 4 decimal places for presentation, this yields

$$\widehat{S} = \begin{bmatrix} 3.5000 & 1.0000 \\ 1.0000 & 1.4000 \end{bmatrix}, \widehat{X} = \widehat{S}^{-1} = \begin{bmatrix} 0.3590 & -0.2564 \\ -0.2564 & 0.8974 \end{bmatrix}, \quad \text{and} \quad \widehat{b}_1 = \langle A_1, \widehat{X} \rangle = 3.4103.$$

The corresponding feasible set for (SDP-Dμ) is the ellipse defined by (6) with $r = b_\mu = (1-\mu)b_1 + \mu\widehat{b}_1$. Tracking the solution path defined by (H_μ) starting at $\mu = 1$ with $(\widehat{X},\widehat{S},\widehat{y})$ yields

$$X = \begin{bmatrix} 0.0564 & -0.1291 \\ -0.1291 & 0.2958 \end{bmatrix}, \quad S = \begin{bmatrix} 2.2910 & 1.0000 \\ 1.0000 & 0.4365 \end{bmatrix}, \quad \text{and} \quad y = -0.1455,$$

rounded to 4 decimal places. Figure 2 plots the path along with the feasiblity sets for the original problem (smaller ellipse) and the perturbed problem (larger ellipse) in the (x_{11}, x_{12})-plane.

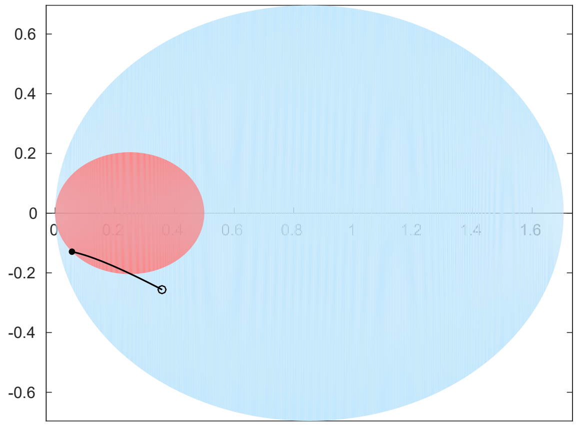

Figure 2. Homotopy path with start point (∘), endpoint (•), and feasible sets.

By utilizing projective space, one can extend Theorem 2.3 to be used for solving (SDP-P) and (SDP-D) when both are feasible and the so-called *duality gap* is zero, i.e., the optimal values of (SDP-P) and (SDP-D) are both finite and equal, but may not be attained. For example, if both (SDP-P) and (SDP-D) are feasible and one of the problems is strictly feasible, then the duality gap is zero but the optimal value for the problem which is strictly feasible may not be attained as shown in the following.

Example 2.5. Consider the following problems:

$$\begin{array}{ll} \text{minimize} & x_{11} \\ \text{subject to} & \begin{bmatrix} x_{11} & 1 \\ 1 & x_{22} \end{bmatrix} \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & y_1 \\ \text{subject to} & \begin{bmatrix} 1 & -y_1/2 \\ -y_1/2 & 0 \end{bmatrix} \succeq 0 \end{array}$$

which correspond to (SDP-P) and (SDP-D), respectively, with $n = 2$, $m = 1$, $b_1 = 1$,

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad A_1 = \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}.$$

It is easy to see that the primal problem is strictly feasible while the only feasible value for the dual problem is $y_1 = 0$. Moreover, the optimal value for both is 0 which is a maximum for the dual problem but an infimum that is not achieved for the primal problem. Due to this, `SDPT3` [46] truncates with $x_{11} \approx 10^{-5}$ and $x_{22} \approx 10^{4}$ while `MOSEK` [3] truncates with $x_{11} \approx 10^{-7}$ and $x_{22} \approx 10^{6}$. We return to this problem in Ex. 2.8.

The following family results from a pathological example (cf. [9, 40]).

Example 2.6. For $\alpha \geq 0$, consider the following problems:

$$\begin{array}{llcll} \text{minimize} & x_{11} & & \text{maximize} & \alpha y_2 \\ \text{subject to} & \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & 0 & (\alpha - x_{11})/2 \\ x_{13} & (\alpha - x_{11})/2 & x_{33} \end{bmatrix} \succeq 0 & & \text{subject to} & \begin{bmatrix} 1-y_2 & 0 & 0 \\ 0 & -y_1 & -y_2 \\ 0 & -y_2 & 0 \end{bmatrix} \succeq 0 \end{array}$$

which correspond to (SDP-P) and (SDP-D), respectively, with $n = 3$, $m = 2$,

$$b = (0, \alpha), \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

A direct calculation shows that both the primal and dual problems are feasible but not strictly feasible. In particular, $x_{11} = \alpha$ and $y_2 = 0$ at every feasible point for the primal and dual problems, respectively. Hence, the duality gap is α. We return to this problem in Ex. 2.9.

By simply tracking the solution path of (H_μ) on the product of projective spaces corresponding to the primal and dual variables as described in Section 2.3, one has the following result.

Theorem 2.7. *With the setup described above, if* (SDP-P) *and* (SDP-D) *are both feasible with duality gap zero and Assumption (A1) holds, then the solution path of* (H_μ) *starting at* $(\widehat{X}, \widehat{S}, \widehat{y})$ *with* $\mu = 1$ *is smooth for* $\mu \in (0, 1]$ *and converges in* $\mathbb{P}^{(n^2+n)/2} \times \mathbb{P}^{(n^2+n)/2+m}$ *corresponding to a solution of* (SDP-P) *and* (SDP-D).

Proof. Similar to the proof of Theorem 2.3, let X^0 and (S^0, y^0) both be feasible for (SDP-P) and (SDP-D). As above, $X_\mu := (1-\mu)X^0 + \mu\widehat{X}$, $S_\mu := (1-\mu)S^0 + \mu\widehat{S}$, and $y_\mu := (1-\mu)y^0 + \mu\widehat{y}$ yield feasible points for (SDP-Pμ) and (SDP-Dμ) for $\mu \in [0, 1]$. Moreover, since $\widehat{X}, \widehat{S} \succ 0$, convexity yields that $X_\mu, S_\mu \succ 0$ for $\mu \in (0, 1]$ which shows that (SDP-Pμ) and (SDP-Dμ) are strictly feasible for $\mu \in (0, 1]$. Standard theory [19, 23, 35] again yields that (5) has a unique solution for $\mu \in (0, 1]$ producing a smooth path $\mu \in (0, 1]$ yielding the corresponding solution path defined by (H_μ) starting at $(\widehat{X}, \widehat{S}, \widehat{y})$. By compactifying

$$\mathbb{R}^{(n^2+n)/2} \times \mathbb{R}^{(n^2+n)/2+m} \hookrightarrow \mathbb{P}^{(n^2+n)/2} \times \mathbb{P}^{(n^2+n)/2+m},$$

this solution path must converge in $\mathbb{P}^{(n^2+n)/2} \times \mathbb{P}^{(n^2+n)/2+m}$. Since the corresponding optimal values are finite (both feasible) and equal (duality gap is zero), and the limit of (5) is (KKT) as μ converges to 0 via a parameter homotopy [28], the limit corresponds with a solution of (SDP-P) and (SDP-D). □

We demonstrate this method on the examples from above.

Example 2.8. For the primal and dual problems from Ex. 2.5, we set up our homotopy method using the arbitrary choice of $\widehat{y} = 0.1$ and taking $\sigma = -0.5$. Rounding to 4 decimal places for presentation, this yields

$$\widehat{S} = \begin{bmatrix} 1.5000 & -0.0500 \\ -0.0500 & 0.5000 \end{bmatrix}, \quad \widehat{X} = \widehat{S}^{-1} = \begin{bmatrix} 0.6689 & 0.0669 \\ 0.0669 & 2.0067 \end{bmatrix}, \quad \text{and} \quad \widehat{b}_1 = \langle A_1, \widehat{X} \rangle = 0.0669.$$

Figure 3 plots the affine coordinates of x_{11}, y_1, and x_{22} as well as the projection of this path onto the (x_{11}, y_1)-plane showing that x_{11} and y_1 converge to 0 while x_{22} diverges. In fact, tracking on $\mathbb{P}^3 \times \mathbb{P}^4$ yields

$$[x_0, x_{11}, x_{12}, x_{22}] = [0, 0, 0, 1] \quad \text{and} \quad [s_0, s_{11}, s_{12}, s_{22}, y_1] = [1, 1, 0, 0, 0] \tag{7}$$

which shows that the solution to the primal problem is "at infinity" since $x_0 = 0$. We can confirm that, in affine space, the coordinate x_{22} diverges since, for the projective endpoint, $x_{22} \neq 0$. The solution to the dual problem is finite since $s_0 \neq 0$ which confirms that the minimum for the dual problem, which is the affine value of y_1 at the endpoint, is attained at 0.

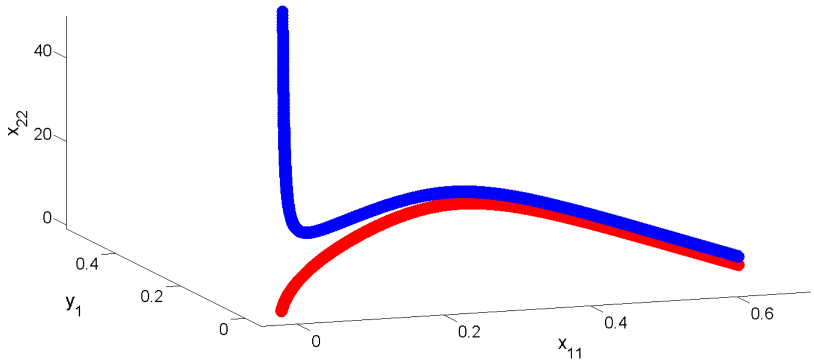

Figure 3. Solution path in x_{11}, y_1, and x_{22} affine coordinates with its projection onto the (x_{11}, y_1)-plane.

Example 2.9. For the primal and dual problems from Ex. 2.6, we compare our homotopy method when $\alpha = 0$ and $\alpha = 1$. Since the duality gap is α, Theorem 2.7 guarantees that our method will solve the problem when $\alpha = 0$. In either case, we can use the same start point by using the arbitrary choice of $\widehat{y} = (0.3, 0.7)$ and taking $\sigma = -1$. Rounding to 4 decimal places for presentation, we have

$$\widehat{S} = \begin{bmatrix} 1.3000 & 0 & 0 \\ 0 & 0.7000 & -0.7000 \\ 0 & -0.7000 & 1.0000 \end{bmatrix}, \widehat{X} = \begin{bmatrix} 0.7692 & 0 & 0 \\ 0 & 4.7619 & 3.3333 \\ 0 & 3.3333 & 3.3333 \end{bmatrix}, \begin{matrix} \widehat{b}_1 = \langle A_1, \widehat{X} \rangle = 4.7619, \\ \widehat{b}_2 = \langle A_2, \widehat{X} \rangle = 7.4359. \end{matrix} \tag{8}$$

When $\alpha = 0$, the path converges in $\mathbb{R}^6 \times \mathbb{R}^8$ with endpoint

$$\begin{aligned} (x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}) &= (0, 0, 0, 0, 0, 1), \\ (s_{11}, s_{12}, s_{13}, s_{22}, s_{23}, s_{33}, y_1, y_2) &= (1, 0, 0, 0.21, 0, 0, -0.21, 0), \end{aligned}$$

showing that the minimum for both primal and dual problems is indeed 0.

When $\alpha = 1$, the duality gap is $1 > 0$ so that the only information that we can gather from the proof of Theorem 2.7 is that the path will converge in $\mathbb{P}^6 \times \mathbb{P}^8$, but it need not correspond with a solution. In this case, the endpoint in $\mathbb{P}^6 \times \mathbb{P}^8$ is

$$\begin{aligned} [x_0, x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}] &= [0, 0, 0, 0, 0, 0, 1], \\ [s_0, s_{11}, s_{12}, s_{13}, s_{22}, s_{23}, s_{33}, y_1, y_2] &= [0, 0, 0, 0, 1, 0, 0, -1, 0], \end{aligned}$$

showing that both are "at infinity" since $x_0 = s_0 = 0$. The limit of x_{11} and y_2 in affine coordinates is 0 and 0.105, respectively, which suggests that the duality gap is nonzero and further computations are needed. We consider the feasibility of the primal and dual problems in Examples 3.3 and 3.7, respectively, and solve the primal and dual problems independently using facial reduction in Section 4.6.

We conclude this section with an infeasible example.

Example 2.10. Consider the following problems:

$$\begin{array}{llll} \text{minimize} & -1 & \text{maximize} & -y_1 \\ \text{subject to} & -1 \geq 0 & \text{subject to} & 1 - y_1 \geq 0 \end{array}$$

which correspond to (SDP-P) and (SDP-D), respectively, with $n = 1$, $m = 1$, $b_1 = -1$, and $C = A_1 = [1]$. The primal problem is clearly infeasible while the dual problem is unbounded, i.e., maximum is ∞. We demonstrate that a solution path to our homotopy method need not exist in this case by arbitrarily selecting $\widehat{y} = 5$ and taking $\sigma = -6$. Hence, $\widehat{S} = [2]$, $\widehat{X} = \widehat{S}^{-1} = [1/2]$, and $\widehat{b}_1 = \langle A_1, \widehat{X} \rangle = 1/2$. This yields

$$H(x_{11}, s_{11}, y_{11}; \mu) = \begin{bmatrix} x_{11} + (1-\mu) - \mu/2 \\ 1 + 6\mu - y_1 - s_{11} \\ s_{11}x_{11} - \mu \end{bmatrix} = 0.$$

For the path starting at $(1/2, 2, 5)$ at $\mu = 1$, one can compute the solution path is

$$(x_{11}(\mu), s_{11}(\mu), y_1(\mu)) = \left(\frac{3\mu - 2}{2}, \frac{2\mu}{3\mu - 2}, \frac{18\mu^2 - 11\mu - 2}{3\mu - 2} \right)$$

which is not defined at $\mu = 2/3$ and hence not smooth for $\mu \in (0, 1]$.

In the next section, we extend our homotopy method to decide the feasiblity of a given problem.

3. Testing for feasibility

The homotopy techniques described in Section 2 can also be applied to test the feasibility of (SDP-P) and (SDP-D). This is motivated by providing an alternative approach to the certificates of infeasibility developed by [26]. For simplicity, in addition to Assumption (A1), we assume that the linear equations are consistent for symmetric matrices which can be easily tested using (numerical) linear algebra computations:

(A2) There exists a symmetric matrix X such that $\langle A_i, X \rangle = b_i$ for $i = 1, \ldots, m$.

Clearly, if Assumption (A2) does not hold, then (SDP-P) is trivially infeasible. When Assumption (A2) holds, we aim to classify (SDP-P) and (SDP-D) as belonging to one of the four "feasibility" types [52]: strictly feasible, feasible but not strictly feasible, weakly infeasible, and strongly infeasible. The partitioning of the infeasible problems into two types (strongly infeasible and weakly infeasible) was shown in [27].

We recall definitions for completeness starting with the primal problem (SDP-P) with feasibility set

$$\mathcal{F}_P = \{X \succeq 0 \mid \langle A_i, X \rangle = b_i, i = 1, \ldots, m\} :$$

- *strictly feasible* if there exists $X \in \mathcal{F}_P$ with $X \succ 0$;
- *feasible but not strictly feasible* if $\mathcal{F}_P \neq \emptyset$ and $\det X = 0$ for every $X \in \mathcal{F}_P$;
- *weakly infeasible* if $\mathcal{F}_P = \emptyset$ and, for every $\epsilon > 0$, there exists an $X \succeq 0$ such that

$$|\langle A_i, X\rangle - b_i| \leq \epsilon \qquad \text{for } i = 1, \dots, m;$$

- *strongly infeasible* if $\mathcal{F}_P = \emptyset$ and there is an improving ray $y \in \mathbb{R}^m$ with

$$-\sum_{i=1}^{m} y_i A_i \succeq 0 \qquad \text{and} \qquad b^T y > 0.$$

A similar classification exists for the dual problem (SDP-D) with feasibility set

$$\mathcal{F}_D = \left\{ (S, y) \;\middle|\; C - \sum_{i=1}^{m} y_i A_i = S,\ S \succeq 0 \right\}:$$

- *strictly feasible* if there exists $(S, y) \in \mathcal{F}_D$ with $S \succ 0$;
- *feasible but not strictly feasible* if $\mathcal{F}_D \neq \emptyset$ and $\det S = 0$ for every $(S, y) \in \mathcal{F}_D$;
- *weakly infeasible* if $\mathcal{F}_D = \emptyset$ and, for every $\epsilon > 0$, there exists a (S, y) with $S \succeq 0$ such that

$$\left\| C - \sum_{i=1}^{m} y_i A_i - S \right\| \leq \epsilon;$$

- *strongly infeasible* if $\mathcal{F}_D = \emptyset$ and there is an improving ray $X \succeq 0$ such that

$$\langle A_i, X\rangle = 0 \quad \text{for } i = 1, \dots, m \qquad \text{and} \qquad \langle C, X\rangle < 0.$$

3.1. Primal feasibility

In order to test the feasibility of (SDP-P), we consider the following convex optimization problem:

$$\begin{aligned} \underset{X,\lambda}{\text{minimize}} \quad & \lambda \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i \qquad i = 1, \dots, m \\ & X + \lambda I \succeq 0 \\ & \lambda + M \geq 0, \end{aligned} \tag{9}$$

where $M > 0$ is a given constant.

Remark 3. One may replace λI in (9) by λD for any $D \succ 0$. We utilize I to simplify the presentation.

The Lagrange dual of (9) is

$$
\begin{array}{ll}
\underset{S,y,\gamma}{\text{maximize}} & b^T y - \gamma M \\
\text{subject to} & \sum_{i=1}^{m} y_i A_i + S = 0 \\
& \text{trace}(S) + \gamma - 1 = 0 \\
& S \succeq 0, \ \gamma \geq 0,
\end{array} \tag{10}
$$

and the corresponding KKT conditions are

$$
\begin{aligned}
\langle A_i, X\rangle &= b_i \qquad i = 1, \ldots, m \\
-\sum_{i=1}^{m} y_i A_i - S &= 0 \\
\text{trace}(S) + \gamma - 1 &= 0 \\
S(X + \lambda I) &= 0 \\
\gamma(\lambda + M) &= 0 \\
X + \lambda I \succeq 0, \ S \succeq 0,& \ \lambda + M \geq 0, \ \gamma \geq 0.
\end{aligned} \tag{11}
$$

Let $\overline{p}$ and $\overline{d}$ denote the optimal value for (9) and (10), respectively.

Theorem 3.1. *Under Assumptions (A1) and (A2), for any $M > 0$, $\overline{p} = \overline{d}$ is a finite value where* (10) *always attains the optimal value such that:*

(1) $\overline{p} < 0$ *if and only if* (SDP-P) *is strictly feasible;*
(2) $\overline{p} = 0$ *and the minimum is attained in* (9) *if and only if* (SDP-P) *is feasible but not strictly feasible;*
(3) $\overline{p} = 0$ *but the minimum is not attained in* (9) *if and only if* (SDP-P) *is weakly infeasible;*
(4) $\overline{p} > 0$ *if and only if* (SDP-P) *is strongly infeasible.*

Proof. Clearly, Assumption (A2) implies that (9) is feasible so that $\overline{p} < \infty$. The constraint $\lambda + M \geq 0$ provides a lower bound on $\overline{p}$, i.e., $\overline{p} \geq -M > -\infty$, so that $\overline{p}$ is a finite value. Since the feasible set for (9) clearly has a nonempty interior, $\overline{p} = \overline{d}$ and the optimal value for (10) is always attained.

If $\overline{p} < 0$, then there exists (X, λ) with $\overline{p} \leq \lambda < 0$ which is feasible for (9). Hence, X is strictly feasible for (SDP-P). Conversely, if X is strictly feasible for (SDP-P), then (X, λ) is feasible for (9) where

$$\lambda = -\min\{M, \lambda_{\min}(X)\} < 0$$

such that $\lambda_{\min}(X)$ is the minimum eigenvalue of X. Hence, $\overline{p} < 0$.

If $\overline{p} = 0$ and the minimum is attained, then select $(X, 0)$ which is feasible for (9). Hence, $X \succeq 0$ is feasible for (SDP-P). Moreover, if (SDP-P) was strictly feasible, then $\overline{p} < 0$ so that (SDP-P) is feasible but not strictly feasible. Conversely, if (SDP-P) is feasible but not strictly feasible, then, for every X that is feasible for (SDP-P), $(X, 0)$ is feasible for (9) showing that $\overline{p} \leq 0$. Since (SDP-P) is not strictly feasible, $\overline{p} \geq 0$ showing that $\overline{p} = 0$ for which the minimum is attained.

If $\overline{p} > 0$, then there exists (S, y, γ) which is feasible for (10) with $b^T y - \gamma M = \overline{d} = \overline{p} > 0$. Since $\gamma M \geq 0$, this shows that (SDP-P) is strongly infeasible due to the improving ray y since $S = -\sum_{i=1}^m y_i A_i \succeq 0$ and $b^T y = \overline{d} + \gamma M > 0$. Conversely, if (SDP-P) is strongly infeasible with improving ray y, then take $S = -\sum_{i=1}^m y_i A_i \succeq 0$ with $b^T y > 0$ implying $y \neq 0$. Since $S \succeq 0$, trace$(S) = 0$ if and only if $S = 0$. By Assumption (A1), $y \neq 0$ implies $S \neq 0$ so that trace$(S) > 0$. Hence, we can find $\delta > 0$ such that trace$(\delta S) = 1$. Thus, since $(\delta S, \delta y, 0)$ is easily observed to be feasible for (10), we know that $\overline{p} = \overline{d} \geq \delta \cdot (b^T y) > 0$.

If $\overline{p} = 0$ and the minimum is not attained, then (SDP-P) is clearly infeasible and not strongly infeasible. Hence, (SDP-P) must be weakly infeasible. Conversely, if (SDP-P) is weakly infeasible, we know that $\overline{p} \leq 0$ since it is not strongly infeasible. Since it is not strictly feasible, $\overline{p} \geq 0$ showing that $\overline{p} = 0$. Since attaining the minimum of 0 yields a feasible point for (SDP-P), the minimum of 0 is not attained. □

Given $M > 0$, we construct our homotopy approach as follows. As before, we arbitrarily select $\widehat{y} \in \mathbb{R}^m$, take $\widehat{\lambda} = 0$ and $\widehat{\gamma} = M^{-1}$, and compute $\sigma \in \mathbb{R}$ such that

$$\widehat{S} := -\sum_{i=1}^m \widehat{y}_i A_i - \sigma I \succ 0.$$

Let $\widehat{X} := \widehat{S}^{-1} \succ 0$ and $\widehat{b}_i := \langle A_i, \widehat{X} \rangle$ for $i = 1, \ldots, m$ with $\widehat{b} = (\widehat{b}_1, \ldots, \widehat{b}_m)$. For $b_\mu := (1-\mu)b + \mu\widehat{b}$, $C_\mu := -\mu\sigma I$, and $t_\mu := \mu(\text{trace}(\widehat{S}) + \widehat{\gamma} - 1)$, consider

$$H(X, \lambda, S, y, \gamma; \mu) = \begin{bmatrix} \langle A_i, X \rangle - b_{\mu_i} \qquad i = 1, \ldots, m \\ C_\mu - \sum_{i=1}^m y_i A_i - S \\ \text{trace}(S) + \gamma - 1 - t_\mu \\ S(X + \lambda I) - \mu I \\ \gamma(\lambda + M) - \mu \end{bmatrix} = 0 \qquad (\mathrm{H}_\mu^P)$$

with start point $(\widehat{X}, \widehat{\lambda}, \widehat{S}, \widehat{y}, \widehat{\gamma})$ at $\mu = 1$ with $\widehat{X} + \widehat{\lambda} I \succ 0$, $\widehat{S} \succ 0$, $\widehat{\lambda} + M > 0$, and $\widehat{\gamma} > 0$.

Theorem 3.2. *With the setup described above, if Assumptions (A1) and (A2) hold, the solution path of* (H_μ^P) *starting at* $(\widehat{X}, \widehat{\lambda}, \widehat{S}, \widehat{y}, \widehat{\gamma})$ *with* $\mu = 1$ *is smooth for* $\mu \in (0, 1]$ *and converges in* $\mathbb{P}^{(n^2+n)/2+1} \times \mathbb{R}^{(n^2+n)/2+m+1}$ *corresponding to a solution of* (9) *and* (10).

Proof. The result follows using a similar proof to that of Theorem 2.7 since both (9) and (10) are feasible, the former strictly feasible, the optimal for the latter is attained, the duality gap is zero, and the inequalities are strictly satisfied at the start point. □

Example 3.3. For $\alpha = 1$, consider the primal problem in Ex. 2.6 which is feasible but not strictly feasible. To test our feasibility homotopy approach, we take $M = 1$ and select $\widehat{y} = (0.3, 0.7)$ and $\sigma = -1$ as in Ex. 2.9 which yields the same $\widehat{S}$, $\widehat{X}$, and $\widehat{b}$ in (8) with $\widehat{\lambda} = 0$ and $\widehat{\gamma} = 1$. The path defined by (H_μ^P) converges in $\mathbb{R}^7 \times \mathbb{R}^9$ with the following endpoint:

$$\begin{array}{rcl}(x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, \lambda) & = & (1, 0, 0, 0, 0, 1, 0),\\ (s_{11}, s_{12}, s_{13}, s_{22}, s_{23}, s_{33}, y_1, y_2, \gamma) & = & (0, 0, 0, 1, 0, 0, -1, 0, 0).\end{array}$$

Hence, $\overline{p} = \overline{d} = 0$ are both attained with Theorem 3.1 confirming the primal problem in Ex. 2.6 is feasible but not strictly feasible.

Example 3.4. A classic primal problem that is weakly infeasible is

$$\begin{array}{ll}\text{minimize} & x_{11}\\ \text{subject to} & \begin{bmatrix} x_{11} & 1\\ 1 & 0\end{bmatrix} \succeq 0\end{array} \tag{12}$$

which corresponds to (SDP-P) with $n = 2$, $m = 2$, $b_1 = 1$, $b_2 = 0$,

$$C = \begin{bmatrix} 1 & 0\\ 0 & 0\end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 1/2\\ 1/2 & 0\end{bmatrix}, \quad \text{and} \quad A_2 = \begin{bmatrix} 0 & 0\\ 0 & 1\end{bmatrix}.$$

Selecting $M = 1$ and using the arbitrary choice of $\widehat{y} = (-0.4, 0.7)$, we take $\sigma = -1$ with $\widehat{\lambda} = 0$, $\widehat{\gamma} = 1$,

$$\widehat{S} = \begin{bmatrix} 1.0000 & 0.2000\\ 0.2000 & 0.3000\end{bmatrix}, \quad \widehat{X} = \begin{bmatrix} 1.1538 & -0.7692\\ -0.7692 & 3.8462\end{bmatrix}, \quad \begin{array}{l}\widehat{b}_1 = \langle A_1, \widehat{X}\rangle = -0.7692,\\ \widehat{b}_2 = \langle A_2, \widehat{X}\rangle = 3.8462,\end{array}$$

rounded to four digits for presentation. The endpoint of the corresponding path defined by (H_μ^P) in $\mathbb{P}^4 \times \mathbb{R}^6$ is

$$[x_0, x_{11}, x_{12}, x_{22}, \lambda] = [0, 1, 0, 0, 0], \qquad (s_{11}, s_{12}, s_{22}, y_1, y_2, \gamma) = (0, 0, 1, 0, -1, 0).$$

By Theorem 3.1, $\overline{p} = \overline{d} = b_1 y_1 + b_2 y_2 = 0$ which is not attained confirming that (12) is weakly infeasible.

3.2. Dual feasibility

Similar to Section 3.1, we test the feasibility of (SDP-D) using the following convex optimization problem:

$$\begin{array}{ll}\underset{S, y, \lambda}{\text{maximize}} & \lambda\\ \text{subject to} & \displaystyle\sum_{i=1}^m y_i A_i + \lambda I + S = C\\ & S \succeq 0\\ & M - \lambda \geq 0,\end{array} \tag{13}$$

where $M > 0$ is a constant.

Remark 4. Similar to Remark 3, one may replace λI in (13) by λD for any $D \succ 0$.

The Lagrange dual of (13) is

$$\begin{array}{lll} \underset{X,\beta}{\text{minimize}} & \langle C, X\rangle + \beta M & \\ \text{subject to} & \langle A_i, X\rangle = 0 \qquad i = 1, \ldots, m \\ & \operatorname{trace}(X) + \beta - 1 = 0 \\ & X \succeq 0, \;\; \beta \geq 0, \end{array} \tag{14}$$

and the corresponding KKT conditions are

$$\begin{aligned} \langle A_i, X\rangle &= 0 \qquad i = 1, \ldots, m \\ \operatorname{trace}(X) + \beta - 1 &= 0 \\ C - \sum_{i=1}^{m} y_i A_i - \lambda I - S &= 0 \\ SX &= 0 \\ \beta(M - \lambda) &= 0 \\ X \succeq 0, \;\; S \succeq 0, \;\; M - \lambda \geq 0, \;\; \beta &\geq 0. \end{aligned} \tag{15}$$

Let $\overline{d}$ and $\overline{p}$ denote the optimal value for (13) and (14), respectively.

Theorem 3.5. *Under Assumption (A1), for any $M > 0$, $\overline{p} = \overline{d}$ is a finite value where* (14) *always attains the optimal value such that:*

(1) $\overline{d} > 0$ *if and only if* (SDP-D) *is strictly feasible;*
(2) $\overline{d} = 0$ *and the maximum is attained in* (13) *if and only if* (SDP-D) *is feasible but not strictly feasible;*
(3) $\overline{d} = 0$ *but the maximum is not attained in* (13) *if and only if* (SDP-D) *is weakly infeasible;*
(4) $\overline{d} < 0$ *if and only if* (SDP-D) *is strongly infeasible.*

Proof. Clearly, (13) is always feasible so that $\overline{d} > -\infty$. The constraint $M - \lambda \geq 0$ provides an upper bound on $\overline{d}$, i.e., $\overline{d} \leq M < \infty$, so that $\overline{d}$ is a finite value. Since the feasible set for (13) clearly has a nonempty interior, $\overline{p} = \overline{d}$ and the optimal value for (14) is always attained.

If $\overline{d} > 0$, then there exists (S, y, λ) with $\overline{d} \geq \lambda > 0$ which is feasible for (13). Hence, (S, y) is strictly feasible for (SDP-D). Conversely, if (S, y) is strictly feasible for (SDP-D), then (S, y, λ) is feasible for (13) where

$$\lambda = \min\{M, \lambda_{\min}(S)\} > 0$$

such that $\lambda_{\min}(S)$ is the minimum eigenvalue of S. Hence, $\overline{d} > 0$.

If $\overline{d} = 0$ and the minimum is attained, then select $(S, y, 0)$ which is feasible for (13). Hence, (S, y) is feasible for (SDP-D). Moreover, if (SDP-D) was strictly feasible, then $\overline{d} > 0$ so that (SDP-D) is feasible but not strictly feasible. Conversely, if (SDP-D) is feasible but not strictly feasible, then, for every (S, y) that is feasible for (SDP-D), $(S, y, 0)$ is feasible for (13) showing that $\overline{d} \geq 0$. Since (SDP-D) is not strictly feasible, $\overline{d} \leq 0$ showing that $\overline{d} = 0$ for which the minimum is attained.

If $\overline{d} < 0$, then there exists (X, β) which is feasible for (14) with $\langle C, X\rangle + \beta M = \overline{p} = \overline{d} < 0$. Since $\beta M \geq 0$, this shows that (SDP-D) is strongly infeasible due to the improving ray $X \succeq 0$ since $\langle A_i, X\rangle = 0$ and $\langle C, X\rangle = \overline{p} - \beta M < 0$. Conversely, if (SDP-D) is strongly infeasible with improving ray $X \succeq 0$, then $\langle C, X\rangle < 0$ implies $X \neq 0$. Hence, $\text{trace}(X) > 0$ so that we can find $\delta > 0$ such that $\text{trace}(\delta X) = 1$. Since $(\delta X, 0)$ is easily observed to be feasible for (14), we know that $\overline{d} = \overline{p} \leq \delta \cdot \langle C, X\rangle < 0$.

If $\overline{d} = 0$ and the minimum is not attained, then (SDP-D) is clearly infeasible and not strongly infeasible. Hence, (SDP-D) must be weakly infeasible. Conversely, if (SDP-D) is weakly infeasible, we know that $\overline{d} \geq 0$ since it is not strongly infeasible. Since it is not strictly feasible, $\overline{d} \leq 0$ showing that $\overline{d} = 0$. Since attaining the minimum of 0 yields a feasible point for (SDP-D), the minimum of 0 is not attained. □

Given $M > 0$, we construct our homotopy approach as follows. As before, we arbitrarily select $\widehat{y} \in \mathbb{R}^m$, take $\widehat{\lambda} = 0$ and $\widehat{\beta} = M^{-1}$, and compute $\sigma \in \mathbb{R}$ such that

$$\widehat{S} := C - \sum_{i=1}^{m} \widehat{y}_i A_i - \widehat{\lambda} I - \sigma I \succ 0.$$

Let $\widehat{X} := \widehat{S}^{-1} \succ 0$ and $\widehat{b}_i := \langle A_i, \widehat{X}\rangle$ for $i = 1, \ldots, m$ with $\widehat{b} = (\widehat{b}_1, \ldots, \widehat{b}_m)$. For $b_\mu := \mu\widehat{b}$, $C_\mu := C - \mu\sigma I$, and $t_\mu := \mu(\text{trace}(\widehat{X}) + \widehat{\beta} - 1)$, we consider the homotopy

$$H(X, \beta, S, y, \lambda; \mu) = \begin{bmatrix} \langle A_i, X\rangle - b_{\mu_i} \quad i = 1, \ldots, m \\ \text{trace}(X) + \beta - 1 - t_\mu \\ C_\mu - \sum_{i=1}^{m} y_i A_i - \lambda I - S \\ SX - \mu I \\ \beta(M - \lambda) - \mu \end{bmatrix} = 0 \qquad (\mathrm{H}_\mu^D)$$

with start point $(\widehat{X}, \widehat{\beta}, \widehat{S}, \widehat{y}, \widehat{\lambda})$ at $\mu = 1$ with $\widehat{X} \succ 0$, $\widehat{S} \succ 0$, $M - \widehat{\lambda} > 0$, and $\widehat{\beta} > 0$.

Theorem 3.6. *With the setup described above, if Assumption (A1) holds, then the solution path of* (H_μ^D) *starting at* $(\widehat{X}, \widehat{\beta}, \widehat{S}, \widehat{y}, \widehat{\lambda})$ *with* $\mu = 1$ *is smooth for* $\mu \in (0, 1]$ *and converges in* $\mathbb{R}^{(n^2+n)/2+1} \times \mathbb{P}^{(n^2+n)/2+m+1}$ *corresponding to a solution of* (14) *and* (13).

Proof. The result follows using a similar proof to that of Theorem 2.7 since both (14) and (13) are feasible, the latter strictly feasible, the optimal for the former is attained, the duality gap is zero, and the inequalities are strictly satisfied at the start point. □

Example 3.7. For $\alpha = 1$, consider the dual problem in Ex. 2.6 which is feasible but not strictly feasible. To test our feasibility homotopy approach, we take $M = 1$ yielding $\widehat{\lambda} = 0$ and $\widehat{\beta} = 1$. We select $\widehat{y} = (0.3, 0.7)$ and $\sigma = -1$ as in Ex. 2.9 which yields the same $\widehat{S}$, $\widehat{X}$, and $\widehat{b}$ in (8). The path defined by (H_μ^D) converges in $\mathbb{R}^7 \times \mathbb{R}^9$ with the following endpoint:

$$\begin{aligned} (x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, \beta) &= (0, 0, 0, 0, 0, 1, 0), \\ (s_{11}, s_{12}, s_{13}, s_{22}, s_{23}, s_{33}, y_1, y_2, \lambda) &= (1, 0, 0, 0.21, 0, 0, -0.21, 0, 0). \end{aligned}$$

Hence, $\overline{p} = \overline{d} = 0$ are both attained with Theorem 3.5 confirming the dual problem in Ex. 2.6 is feasible but not strictly feasible.

Example 3.8. A classic dual problem that is weakly infeasible is

$$\begin{array}{ll} \text{maximize} & y_1 \\ \text{subject to} & \begin{bmatrix} -y_1 & 1 \\ 1 & 0 \end{bmatrix} \succeq 0 \end{array} \tag{16}$$

which corresponds to (SDP-D) with $n = 2$, $m = 1$, $b_1 = 1$,

$$C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Selecting $M = 1$ and using the arbitrary choice of $\widehat{y} = 0.8$, we take $\sigma = -2$ with $\widehat{\lambda} = 0$, $\widehat{\beta} = 1$,

$$\widehat{S} = \begin{bmatrix} 1.2000 & 1.0000 \\ 1.0000 & 2.0000 \end{bmatrix}, \quad \widehat{X} = \begin{bmatrix} 1.4286 & -0.7143 \\ -0.7143 & 0.8571 \end{bmatrix}, \quad \widehat{b}_1 = \langle A_1, \widehat{X} \rangle = 1.4286,$$

rounded to four digits for presentation. The endpoint of the corresponding path defined by (H_μ^P) in $\mathbb{R}^4 \times \mathbb{P}^5$ is

$$(x_{11}, x_{12}, x_{22}, \beta) = (0, 0, 1, 0), \qquad [s_0, s_{11}, s_{12}, s_{22}, y_1, \lambda] = [0, 1, 0, 0, 1, 0].$$

By Theorem 3.5, $\overline{d} = \overline{p} = \langle C, X \rangle + \beta M = 0$ which is not attained confirming that (16) is weakly infeasible.

4. Solving all semidefinite programs

Motivated by [37] showing that commonly used semidefinite program software packages have difficulty solving problems when the duality gap is nonzero, the following describes a three-step process (feasibility testing, facial reduction, and primal-dual solving) for solving all semidefinite programs. The outcome of the feasibility tests (Section 3.1 for (SDP-P) and Section 3.2 for (SDP-D)) dictate how one proceeds in the subsequent steps with Table 1 providing a summary of how this section proceeds.

		Dual		
		SF	F	IF
Primal	SF	§4.1	§4.1	§4.2
	F	§4.1	§4.5	§4.3
	IF	§4.2	§4.4	§4.2

Table 1. Summary of solving based on the feasibility testing (SF: strictly feasible, F: feasible but not strictly feasible, IF: infeasible): §4.1: apply primal-dual solver; §4.2: no additional computations needed; §4.3–4.5: apply facial reduction and solve individually as needed.

In the following, $\overline{p}$ and $\overline{d}$ are the optimal value of (SDP-P) and (SDP-D), respectively. After describing how to proceed in each of the cases, we conclude with some examples in Section 4.6

4.1. Both feasible and at least one strictly feasible

When both (SDP-P) and (SDP-D) are feasible and at least one is strictly feasible, $\overline{p} = \overline{d}$ and this value can be computed using a primal-dual solver, e.g., via Theorem 2.7.

4.2. At most one strictly feasible and otherwise infeasible

If (SDP-P) is infeasible and (SDP-D) is strictly feasible, $\overline{p} = \overline{d} = \infty$. If (SDP-P) is strictly feasible and (SDP-D) is infeasible, $\overline{p} = \overline{d} = -\infty$. If both (SDP-P) and (SDP-D) are infeasible, $\overline{p} = \infty$ and $\overline{d} = -\infty$.

4.3. Primal feasible and dual infeasible

When (SDP-P) is feasible but not strictly feasible and (SDP-D) is infeasible, then $\overline{d} = -\infty$ and one can use facial reduction, e.g., see [11, 12, 39, 41, 42, 49, 55], to reduce to an equivalent problem which has a zero duality gap to compute $\overline{p}$. Since the outcome of the feasibility test for (SDP-P) in Theorem 3.2 is a relative interior point X^*, the feasibility set for (SDP-P) is simply

$$\{X \succeq 0 \mid \text{null } X^* \subset \text{null } X, \langle A_i, X\rangle = b_i, i = 1, \ldots, m\}. \tag{17}$$

Hence, one simply uses (numerical) linear algebra to replace the linear constraints in (17) with a linearly independent set of linear equations $\langle A_i', X\rangle = b_i'$ for $i = 1, \ldots, m'$. Updating (SDP-P) with these new linear constraints yields a problem with zero duality gap which can be solved via the aforementioned cases.

4.4. Dual feasible and primal infeasible

When (SDP-D) is feasible but not strictly feasible and (SDP-P) is infeasible, then $\overline{p} = \infty$ and one can use facial reduction to compute $\overline{d}$. As in Section 4.3, since the outcome of the feasibility test for (SDP-D) in Theorem 3.6 is a relative interior point (S^*, y^*), the feasibility set for (SDP-D) is simply

$$\left\{(S, y) \;\middle|\; \text{null } S^* \subset \text{null } S, S = C - \sum_{i=1}^m y_i A_i, S \succeq 0\right\}. \tag{18}$$

Let N^* be a basis for null S^* and W^* be a basis for the linear space

$$\left\{y \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^m y_i A_i N^* = 0\right\}.$$

Then, if W^* has m' columns, (18) is equal to

$$\left\{(S, y) \;\middle|\; y = W^* y' + y^*, S = C - \sum_{i=1}^m y_i A_i, S \succeq 0, y' \in \mathbb{R}^{m'}\right\}.$$

Next, simple linear algebra computations yield matrices C' and $A'_1, \ldots, A'_{m'}$ such that (18) is equal to

$$\left\{ (S, y) \;\middle|\; y = W^* y' + y^*, S = C' - \sum_{i=1}^{m'} y'_i A'_i, S \succeq 0, y' \in \mathbb{R}^{m'} \right\}.$$

Therefore, $\overline{d} = b^T y^* + d'$ where d' is the optimal value of

$$\begin{aligned} \underset{S,y'}{\text{maximize}} \quad & b'^T y' \\ \text{subject to} \quad & C' - \sum_{i=1}^{m'} y'_i A'_i = S, \\ & S \succeq 0 \end{aligned}$$

in which $b' = (W^*)^T b$. This problem has a zero duality gap and can be solved via the aforementioned cases.

4.5. Both feasible but not strictly feasible

When (SDP-P) and (SDP-D) are feasible but not strictly feasible, $\infty > \overline{p} \geq \overline{d} > -\infty$. Using facial reduction on (SDP-P) as in Section 4.3 yields a relative strictly feasible problem in which the corresponding dual problem is feasible. Applying a primal-dual solver, e.g., via Theorem 2.7, yields $\overline{p}$. Similarly, applying facial reduction on (SDP-D) as in Section 4.4 yields a relative strictly feasible problem in which the corresponding primal problem is feasible. Applying a primal-dual solver, e.g., via Theorem 2.7, yields $\overline{d}$.

4.6. Illustrative example

To demonstrate, we consider Example 2.6 with $\alpha = 1$. It was shown in Examples 3.3 and 3.7 that the primal and dual problems are feasible but not strictly feasible, respectively. Thus, as described in Section 4.5, we separately apply facial reduction to the primal and dual problems, and then solve each independently to obtain $\overline{p}$ and $\overline{d}$.

For the primal, Example 3.3 yields

$$X^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since null X^* is spanned by $[0,1,0]^T$, just considering the condition null $X^* \subset$ null X yields three linear constraints, namely $\langle A_i, X \rangle = 0$ for $i = 3, 4, 5$ where

$$A_3 = \begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad A_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix}.$$

Since $A_1, \ldots, A_5$ are not linearly independent, a linear independent collection of $m' = 4$

constraints is $\langle A'_i, X\rangle = b'_i$ for $i = 1, \ldots, m'$ where

$$A'_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; A'_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; A'_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and } A'_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

with $b'_1 = 1$ and $b'_2 = b'_3 = b'_4 = 0$. Thus, the feasibility set of this updated problem is

$$X = \begin{bmatrix} 1 & 0 & x_{13} \\ 0 & 0 & 0 \\ x_{13} & 0 & x_{33} \end{bmatrix} \succeq 0$$

which clearly yields an optimal value of $\overline{p} = 1$.

For the dual, Example 3.7 yields

$$S^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.21 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad y^* = \begin{bmatrix} -0.21 \\ 0 \end{bmatrix}.$$

Following the notation of Section 4.4, $N^* = [0, 0, 1]^T$ is a basis for null S^* and $W^* = [1, 0]^T$ is a basis for

$$\{y \in \mathbb{R}^2 \mid y_1 A_1 N^* + y_2 A_2 N^* = 0\}.$$

Therefore, we have $m' = 1$ with

$$C' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.21 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A'_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad b' = [0]$$

where $\overline{d} = 0 + d'$. Since d' is clearly equal to 0, $\overline{d} = 0$.

Therefore, this procedure correctly identifies that there is a nonzero duality gap with $\overline{p} = 1 > 0 = \overline{d}$.

5. Examples

The following examples utilized `Bertini` [5] to perform the path tracking. Files associated with these examples are located in a repository at `http://dx.doi.org/10.7274/R0D798G4`. All timings reported are based on using a single core of a 2.4 GHz AMD Opteron Processor 6378 with 128 GB RAM.

5.1. Sums of squares

An application of semidefinite programming is to decide if a given polynomial is a sum of squares and thus providing a certificate of nonnegativity, e.g., see [9, Chaps. 3-4] and [24]. In particular, a polynomial $p(x) = p(x_1, \ldots, x_n)$ with real coefficients of degree $2d$ is a *sum of squares* if there exists polynomials $p_1(x), \ldots, p_k(x)$ with real coefficients of degree d such that $p(x) = p_1(x)^2 + \cdots + p_k(x)^2$. Letting $v(x)$ be the vector of length $N = \binom{n+d}{d}$ consisting of all monomials in x of degree at most d, it is easy to verify that $p(x)$ is a sum of squares if and only there exists an $N \times N$ matrix $Q \succeq 0$ with

$$p(x) = v(x)^T \cdot Q \cdot v(x).$$

We consider applying our homotopy approach to several applications involving sums of squares.

5.1.1. *Verifying real solution set*

The polynomial $f(x) = x^3 - 2$ has a unique real root, namely $\alpha = \sqrt[3]{2}$. One way to verify this is through the real Nullstellensatz (see, e.g., [10, Chap. 4]) together with sums of squares decomposition as described in [14]. In particular, if there exists $c \in \mathbb{R}^2$ such that

$$g(x; c) := -(x - \alpha)^4 + (c_1 x + c_2)(x^3 - 2) \text{ is a sum of squares,}$$

then, due to the nonnegativity of sums of squares, it is clear that x is a real root of f if and only if $x = \alpha$. Since g has degree 4 in x, g is a sum of squares if and only if there exists a 3×3 matrix $Q \succeq 0$ such that

$$g(x; c) = v(x)^T \cdot Q \cdot v(x) \quad \text{where} \quad v(x) = \begin{bmatrix} 1 & x & x^2 \end{bmatrix}^T.$$

Since $x^2 = x \cdot x$, $g(x; c)$ is a sum of squares if and only if there exists $c_3 \in \mathbb{R}$ such that

$$Q = C - c_1 A_1 - c_2 A_2 - c_3 A_3 \succeq 0 \tag{19}$$

where

$$C = \begin{bmatrix} -2\alpha & 4 & 0 \\ 4 & -6\alpha^2 & 2\alpha \\ 0 & 2\alpha & -1 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$
$$A_2 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & -1/2 \\ 0 & -1/2 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

Hence, we aim to decide the feasiblity of

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & Q = C - c_1 A_1 - c_2 A_2 - c_3 A_3 \succeq 0. \end{array} \tag{20}$$

To accomplish this, we consider the following strictly feasible problem:

$$\begin{array}{ll} \text{maximize} & \lambda \\ \text{subject to} & C - c_1 A_1 - c_2 A_2 - c_3 A_3 - \lambda I \succeq 0. \end{array} \tag{21}$$

For $b = (0, 0, 0, 1)$, $y = (c_1, c_2, c_3, \lambda)$, and $A_4 = I$, we utilized our homotopy approach upon selecting $\widehat{y} = (0, 0, 0, -12)$ and $\sigma = 0$ yielding the solution (to 4 decimal places):

$$c_1 = 4.0681, \quad c_2 = -5.1255, \quad c_3 = -4.8163, \quad \lambda = 0. \tag{22}$$

Tracking this homotopy path in all 16 variables using `Bertini` took 13 steps in 0.030 seconds while the homotopy path in 10 variables using the trivial reduction described in Remark 1 took 9 steps in 0.027 seconds. Since the optimal $\lambda = 0$ is attained, (20) is feasible but not strictly feasible. Nonetheless, this is enough to show that $x = \alpha$ is the unique real root of f by writing $g(x; c)$ as a sum of two squares. Moreover, using [22] starting with (22), this point is a smooth point on the following line parameterized by c_1

$$(c_1, c_2, c_3) = (c_1, -\alpha \cdot c_1, -\alpha^2 \cdot (c_1/2 + 1)) \tag{23}$$

allowing one to easily perform additional computations on this line. For example, on this line, one needs $c_1 \geq 4$ to satisfy (19). In particular, we can write $g(x; c)$ using one square when $c_1 = 4$ so that $c_2 = -4\alpha$ yielding the following decomposition:

$$-(x - \alpha)^4 + 4(x - \alpha)(x^3 - 2) = 3(x^2 - \alpha^2)^2.$$

For comparison, we also solved (21) using `MOSEK` [3] which took 0.02 seconds in 8 iterations to compute

$$c_1 = 4.4785, \quad c_2 = -5.6426, \quad c_3 = -5.1421, \quad \lambda = 0$$

lying on the line in (23). The software `SDPT3` [46] also took 8 iterations in 0.33 seconds to compute

$$c_1 = 8.1075, \quad c_2 = -10.2149, \quad c_3 = -8.0244, \quad \lambda = 0$$

which also lies on the line in (23).

5.1.2. Motzkin polynomial

The Motzkin polynomial $p(x, y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ is a nonnegative polynomial that is not a sum of squares. If it was a sum of squares, then there exists a 10×10 matrix $Q \succeq 0$ with $p(x, y) = v(x, y)^T \cdot Q \cdot v(x, y)$ where

$$v(x, y) = \begin{bmatrix} 1 & x & y & x^2 & xy & y^2 & x^3 & x^2y & xy^2 & y^3 \end{bmatrix}^T$$

which describes an $m = 27$ dimensional linear space on the 55 variables in Q. Applying our homotopy-based dual feasibility test from Section 3.2, e.g., with $M = 1$, to

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & Q \succeq 0 \text{ such that } p(x, y) = v(x, y)^T \cdot Q \cdot v(x, y) \end{array} \tag{24}$$

yields the optimal value of the problem corresponding to (13) is $\overline{d} \approx -0.006989 < 0$. Tracking in `Bertini` with a total of 139 variables took 18 steps in approximately 1.05 seconds while the homotopy path in 84 variables using the trivial reduction described in Remark 1 took 14 steps in 0.27 seconds. By Theorem 3.5, this shows that (24) is strongly infeasible confirming that $p(x, y)$ is not a sum of squares.

Both `MOSEK` [3] and `SDPT3` [46] confirm the infeasibility of (24). After 0.35 seconds with 12 iterations, `SDPT3` stops with the message "suspected of being infeasible" while `MOSEK` detects the infeasibility certificate in 0.03 seconds using 4 iterations.

We also solved the following using MOSEK and SDPT3:

$$\begin{array}{ll} \text{maximize} & \lambda \\ \text{subject to} & Q - \lambda I \succeq 0 \text{ such that } p(x,y) = v(x,y)^T \cdot Q \cdot v(x,y). \end{array} \tag{25}$$

Both MOSEK and SDPT3 compute the optimal value is approximately $-0.006989 < 0$ confirming strong infeasibility. The software SDPT3 used 10 iterations in 0.34 seconds while MOSEK used 9 iterations in 0.03 seconds. However, the output $Q - \lambda I$ matrix from MOSEK was not positive semidefinite.

5.2. Identify weakly and strongly infeasible

A test suite of infeasible dual problems was created in [26] which have integer entries and thus can be verified as infeasible using exact arithmetic. Since the structure for why these are infeasible can be easily observed, they call these instances *clean*. To hide this structure, they perform a "messing operation" on each clean instance via row operations and a rotation yielding *messy* instances. In total, this test suite consists of 800 instances with $n = 10$: 400 instances have $m = 20$ linear constraints and the other 400 instances have $m = 10$ linear constraints. The results from [26, Tables 1 & 2] are presented in Tables 2 and 3 for the four options they tested, namely MOSEK [3], SDPT3 [46], SeDuMi [44], and SeDuMi with preprocessing algorithm of [39].

We employed our homotopy-based dual feasibility test described in Section 3.2 to these 800 instances using Bertini [5]. To improve the performance of the endgame without *a priori* knowledge of the cycle number as described in Remark 2, we tracked each homotopy with respect to s where $\mu = s^4$. As summarized in Tables 2 and 3, Theorem 3.5 successfully permitted the distinction between strongly infeasible and weakly feasible by accurately computing the endpoint in projective space as described in Theorem 3.6.

5.3. Nonzero duality gaps

Similar to Section 5.2, a test suite of problems with nonzero duality gaps was created in [37] which have integer entries with *clean* and *messy* instances. In half of these problems, both the primal and dual are feasible but not strictly feasible with $\overline{p} = 10 > 0 = \overline{d}$, i.e., a finite duality gap. The others have the primal infeasible and the dual feasible but not strictly feasible with $\overline{p} = \infty > 0 = \overline{d}$, i.e., an infinite duality gap.

We employed our three-step process described in Section 4 to 20 instances, 4 each in $\mathbb{R}^{n\times n}$ for $n = 3, \ldots, 7$, using Bertini [5]. The results are presented in Table 4, together with the four options that are tested in [37], namely MOSEK [3], SDPA-GMP [31], MOSEK with preprocessing algorithm of [39], and MOSEK with preprocessing algorithm of [55]. As summarized in Table 4, the solving approach described in Section 4 successfully used the output of the feasibility tests to employ the proper facial reduction needed to create and solve related problems with zero duality gap.

The challenge with this test suite is the feasibility testing for the primal problems in $\mathbb{R}^{n\times n}$. For the problems tested, the cycle number in primal feasibility testing was 2^{n-3} for both the instances with finite duality gap and 2^{n-2} for both the instances with infinite duality gap. Thus, the exponential increase in the cycle number caused extreme ill-conditioning near $t = 0$ forcing Bertini to utilize increasingly higher precision computations as the size of the problem increased even after utilizing the analytic

	Strongly Infeasible		Weakly Infeasible	
	Clean	Messy	Clean	Messy
SeDuMi [44]	100	100	1	0
SDPT3 [46]	100	96	0	0
MOSEK [3]	100	100	11	0
Preprocess [39] + SeDuMi [44]	100	100	100	0
Bertini [5]	100	100	100	100

Table 2. Results for a test suite of dual problems [26] with integer entries: 400 instances with $m = 20$ linear constraints. Clean: infeasible instances that can be easily observed. Messy: infeasible instances that are constructed from "messing operations" on clean instances.

	Strongly Infeasible		Weakly Infeasible	
	Clean	Messy	Clean	Messy
SeDuMi [44]	87	27	0	0
SDPT3 [46]	10	5	0	0
MOSEK [3]	63	17	0	0
Preprocess [39] + SeDuMi [44]	100	27	100	0
Bertini [5]	100	100	100	100

Table 3. Results for a test suite of dual problems [26] with integer entries: 400 instances with $m = 10$ linear constraints. Clean: infeasible instances that can be easily observed. Messy: infeasible instances that are constructed from "messing operations" on clean instances.

	Finite Duality Gap		Infinite Duality Gap	
	Clean	Messy	Clean	Messy
MOSEK [3]	1	1	0	0
SDPA-GMP [31]	1	1	0	0
Preprocess [39] + MOSEK [3]	5	1	5	0
Preprocess [55] + MOSEK [3]	5	1	5	0
Bertini [5]	5	5	5	5

Table 4. Results for a test suite of problems with nonzero duality gaps [37] with integer entries: 20 instances. Clean: infeasible instances that can be easily observed. Messy: infeasible instances that are constructed from "messing operations" on clean instances.

reparameterization discussed in Section 2.2. The cycle number in the feasibility testing of the dual problems as well as the primal-dual solver after facial reduction was 1 resulting in easy-to-track paths.

6. Conclusion

By viewing interior point methods as a homotopy defined by a system of bilinear equations, techniques from numerical algebraic geometry can be used to handle various cases that arise, such as adaptive precision path tracking to handle ill-conditioned areas, endgames to accurately approximate an optimizer, and projective space when the optimal value is not achieved. We applied our homotopy-based approach to develop a feasibility test for both primal and dual problems. In particular, Theorems 3.1 and 3.5 show that the four feasibility types of semidefinite programs can be distinguished with our homotopy approach. Section 5.2 demonstrates the success of our approach on the 800 instances of the test suite of [26]. Finally, to solve any semidefinite program, we developed a three-step approach consisting of feasibility testing, facial reduction, and primal-dual solving with Section 5.3 demonstrating the success of our approach on 20 instances from the test suite of [37]. In the future, we aim to understand the scalability of this numerical algebraic geometric approach by analyzing the test suites in [50, 51].

Acknowledgments

All authors were partially supported in part by Office of Naval Research grant N00014-16-1-2722, NSF grants ACI-1460032 and CCF-1812746, and Sloan Research Fellowship BR2014-110 TR14.

References

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

[2] F. Alizadeh, J. Haeberly, and M. Overton, *Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results*, SIAM J. Optim. 8 (1998), pp. 746–768. Available at `http://dx.doi.org/10.1137/S1052623496304700`.

[3] M. ApS, *The MOSEK Python optimizer API manual, Version 7.1 (Revision 62)*, `http://docs.mosek.com/7.1/pythonapi/index.html` (2015). Available at `http://docs.mosek.com/7.1/pythonapi/index.html`.

[4] D. Bates, J. Hauenstein, and A. Sommese, *Efficient path tracking methods*, Numer. Algorithms 58 (2011), pp. 451–459.

[5] D. Bates, J. Hauenstein, A. Sommese, and C. Wampler, *Bertini: Software for numerical algebraic geometry.* Available at `bertini.nd.edu`.

[6] D. Bates, J. Hauenstein, A. Sommese, and C. Wampler, *Numerically Solving Polynomial Systems with Bertini*, Vol. 25, SIAM, 2013.

[7] D. Bates, J. Hauenstein, A. Sommese, and C. Wampler II, *Adaptive multiprecision path tracking*, SIAM J. Numer. Anal. 46 (2008), pp. 722–746.

[8] D. Bates, J. Hauenstein, A. Sommese, and C. Wampler II, *Stepsize control for path tracking*, in *Interactions of classical and numerical algebraic geometry*, Contemp. Math. Vol. 496, Amer. Math. Soc., Providence, RI, 2009, pp. 21–31.

[9] G. Blekherman, P. Parrilo, and R. Thomas, *Semidefinite Optimization and Convex Algebraic Geometry*, SIAM, Philadelphia, PA, USA, 2012.

[10] J. Bochnak, M. Coste, and M.F. Roy, *Real algebraic geometry; 3rd ed.*, Ergebnisse der Mathematik und ihrer Grenzgebiete A Series of Modern Surveys in Mathematics, Springer, Berlin, 1998.

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

[12] J. Borwein and H. Wolkowicz, *Facial reduction for a cone-convex programming problem*, J. Austral. Math. Soc. Ser. A 30 (1980/81), pp. 369–380.

[13] S. Boyd, L. El Ghoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM Studies in Applied Mathematics Vol. 15, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994, Available at `http://dx.doi.org/10.1137/1.9781611970777`. MR 1284712

[14] D. Brake, J. Hauenstein, and A. Liddell Jr., *Validating the Completeness of the Real Solution Set of a System of Polynomial Equations*, in *Proceedings of the ACM on International Symposium on Symbolic and Algebraic Computation*, New York, NY, USA. ACM, ISSAC '16, 2016, pp. 143–150.

[15] J. Dattorro, *Convex Optimization & Euclidean Distance Geometry*, Meboo Publishing USA, 2011.

[16] E. de Klerk, C. Roos, and T. Terlaky, *A short survey on semidefinite programming*, in *Ten*

years LNMB, CWI Tract Vol. 122, Math. Centrum, Centrum Wisk. Inform., Amsterdam, 1997, pp. 323–339. MR 1465717

[17] E. de Klerk, C. Roos, and T. Terlaky, *Infeasible-start semidefinite programming algorithms via self-dual embeddings*, in *Topics in semidefinite and interior-point methods (Toronto, ON, 1996)*, Fields Inst. Commun. Vol. 18, Amer. Math. Soc., Providence, RI, 1998, pp. 215–236. MR 1607349

[18] E. de Klerk, T. Terlaky, and K. Roos, *Self-dual embeddings*, in *Handbook of semidefinite programming*, Internat. Ser. Oper. Res. Management Sci. Vol. 27, Kluwer Acad. Publ., Boston, MA, 2000, pp. 111–138. Available at `https://doi.org/10.1007/978-1-4615-4381-7_5`. MR 1778227

[19] B. Gärtner and J. Matoušek, *Approximation Algorithms and Semidefinite Programming*, Springer, Heidelberg, 2012, Available at `http://dx.doi.org/10.1007/978-3-642-22015-9`. MR 3015090

[20] J. Hauenstein and A. Liddell, *Certified predictor-corrector tracking for newton homotopies*, J. Symbolic Comput. 74 (2016), pp. 239–254.

[21] J. Hauenstein and M. Regan, *Adaptive strategies for solving parameterized systems using homotopy continuation*, Appl. Math. Comput. 332 (2018), pp. 19–34.

[22] J. Hauenstein and C. Wampler, *Isosingular sets and deflation*, Found. Comput. Math. 13 (2013), pp. 371–403.

[23] M. Kojima, S. Shindoh, and S. Hara, *Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices*, SIAM J. Optim. 7 (1997), pp. 86–125. Available at `http://dx.doi.org/10.1137/S1052623494269035`. MR 1430559

[24] M. Laurent, *Sums of squares, moment matrices and optimization over polynomials*, in *Emerging applications of algebraic geometry*, IMA Vol. Math. Appl. Vol. 149, Springer, New York, 2009, pp. 157–270. Available at `https://doi.org/10.1007/978-0-387-09686-5_7`. MR 2500468

[25] A. Leykin, J. Verschelde, and A. Zhao, *Newton's method with deflation for isolated singularities of polynomial systems*, Theoret. Comput. Sci. 359 (2006), pp. 111–122.

[26] M. Liu and G. Pataki, *Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming*, Math. Program. 167 (2018), pp. 435–480.

[27] Z.Q. Luo, J. Sturm, and S. Zhang, *Duality results for conic convex programming*, Econometric Institute Research Papers EI 9719/A, Erasmus University Rotterdam, Erasmus School of Economics (ESE), Econometric Institute, 1997. Available at `http://EconPapers.repec.org/RePEc:ems:eureir:1412`.

[28] A. Morgan and A. Sommese, *Coefficient-parameter polynomial continuation*, Appl. Math. Comput. 29 (1989), pp. 123–160.

[29] A. Morgan, A. Sommese, and C. Wampler, *Computing singular solutions to nonlinear analytic systems*, Numer. Math. 58 (1991), pp. 669–684.

[30] A. Morgan, A. Sommese, and C. Wampler, *A power series method for computing singular solutions to nonlinear analytic systems*, Numer. Math. 63 (1992), pp. 391–409.

[31] M. Nakata, *A numerical evaluation of highly accurate multiple-precision arithmetic version of semidefinite programming solver: SDPA-GMP, -QD and -DD*, in *2010 IEEE International Symposium on Computer-Aided Control System Design*, Sep. 2010, pp. 29–34.

[32] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, SIAM Studies in Applied Mathematics Vol. 13, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[33] Y. Nesterov, M. Todd, and Y. Ye, *Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems*, Math. Program. 84 (1999), pp. 227–267.

[34] Y. Nesterov and A. Nemirovskii, *Self-concordant Functions and Polynomial-time Methods in Convex Programming*, USSR Academy of Sciences, Central Economic & Mathematic Institute, 1989.

[35] Y. Nesterov and A. Nemirovskii, *Interior Point Methods in Convex Programming: Theory and Applications*, SIAM, Philadelphia, 1994.

[36] Y. Nesterov and M. Todd, *Primal-dual interior-point methods for self-scaled cones*,

SIAM J. Optim. 8 (1998), pp. 324–364. Available at http://dx.doi.org/10.1137/S1052623495290209. MR 1618802

[37] G. Pataki, *On positive duality gaps in semidefinite programming*, arXiv:1812.11796 (2018).

[38] F. Permenter, H. Friberg, and E. Andersen, *Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach*, SIAM J. Optim. 27 (2017), pp. 1257–1282. Available at https://doi.org/10.1137/15M1049415. MR 3668602

[39] F. Permenter and P. Parrilo, *Partial facial reduction: simplified, equivalent sdps via approximations of the psd cone*, Math. Program. (2017).

[40] M. Ramana, *An exact duality theory for semidefinite programming and its complexity implications*, Math. Program. 77 (1997), pp. 129–162. Available at http://dx.doi.org/10.1007/BF02614433.

[41] M. Ramana, *An exact duality theory for semidefinite programming and its complexity implications*, Math. Programming 77 (1997), pp. 129–162. Semidefinite programming.

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

[43] A. Sommese and C. Wampler, *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*, World Scientific, 2005.

[44] J. Sturm, *Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones*, Optim. Methods Softw. 11 (1999), pp. 625–653.

[45] M. Todd, *Semidefinite optimization*, Acta Numer. 10 (2001), pp. 515–560.

[46] K. Toh, M. Todd, and R. Tütüncü, *SDPT3 – a MATLAB software package for semidefinite programming*, Optim. Methods Softw. 11 (1999), pp. 545–581.

[47] L. Trefethen and J. Weideman, *The exponentially convergent trapezoidal rule*, SIAM Rev. 56 (2014), pp. 385–458.

[48] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev. 38 (1996), pp. 49–95. Available at http://dx.doi.org/10.1137/1038003.

[49] H. Waki and M. Muramatsu, *Facial reduction algorithms for conic optimization problems*, J. Optim. Theory Appl. 158 (2013), pp. 188–215.

[50] H. Waki, *How to generate weakly infeasible semidefinite programs via Lasserre's relaxations for polynomial optimization*, Optim. Lett. 6 (2012), pp. 1883–1896.

[51] H. Waki, M. Nakata, and M. Muramatsu, *Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization*, Comput. Optim. Appl. 53 (2012), pp. 823–844.

[52] H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.), *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, International Series in Operations Research & Management Science Vol. 27, Kluwer Academic Publishers, Boston, MA, 2000, Available at http://dx.doi.org/10.1007/978-1-4615-4381-7. MR 1778223

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

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

[55] Y. Zhu, G. Pataki, and Q. Tran-Dinh, *Sieve-sdp: a simple facial reduction algorithm to preprocess semidefinite programs*, Math. Program. Comput. (2019). Available at https://doi.org/10.1007/s12532-019-00164-4.