Error Bounds and Singularity Degree
in Semidefinite Programming

Stefan Sremac*   Hugo J. Woerdeman†   Henry Wolkowicz‡

Abstract

In semidefinite programming a proposed optimal solution may be quite poor in spite of having sufficiently small residual in the optimality conditions. This issue may be framed in terms of the discrepancy between *forward error* (the unmeasurable 'true error') and *backward error* (the measurable violation of optimality conditions). In [29], Sturm provided an upper bound on forward error in terms of backward error and *singularity degree*. In this work we provide a method to bound the maximum rank over all optimal solutions and use this result to obtain a lower bound on forward error for a class of convergent sequences. This lower bound complements the upper bound of Sturm. The results of Sturm imply that semidefinite programs with slow convergence necessarily have large singularity degree. Here we show that large singularity degree is, in some sense, also a sufficient condition for slow convergence for a family of external-type 'central' paths. Our results are supported by numerical observations.

Keywords: Semidefinite programming, SDP, facial reduction, singularity degree, maximizing log det.

AMS subject classifications: 90C22, 90C25

1   Introduction

It is well known that for certain pathological instances of semidefinite programming, state-of-the-art algorithms, while theoretically guaranteed to converge to an optimal solution, do so very slowly or can fail to converge entirely. This issue is exacerbated in that it is generally undetectable. In this paper we propose a method to detect this type of slow convergence by lower bounding *forward error*, i.e., distance to the solution (optimal) set. This bound is obtained by analyzing a class of parametric curves that are proven to converge to a solution of maximum rank and then upper bounding that rank. In the second part of the paper we present a new analysis of the relation between forward error and *singularity degree*, a measure introduced by Sturm in [29] and shown to be a necessary condition for slow convergence. Our results indicate that large singularity degree is, in some sense, also a sufficient condition for slow convergence for a certain family of central paths.

To be more specific about the type of slow convergence we are concerned with, let $\mathcal{F} \subset \mathbb{S}^n$ be the solution set of a *semidefinite program (SDP)*. By *solution set* we mean the set of optimal solutions. For feasibility problems the solution set coincides with the feasible set. Throughout this paper we refer to $\mathcal{F}$ as a *spectrahedron*. Here $\mathbb{S}^n$ denotes the ambient space of $n \times n$ symmetric matrices. It is always possible to express $\mathcal{F}$ as the intersection of an affine subspace, $\mathcal{L}$, and the

set of positive semidefinite matrices, $\mathbb{S}_+^n$. Given a matrix $X \in \mathbb{S}^n$, the *forward error* is defined as,

$$\epsilon^f(X, \mathcal{F}) := \text{dist}(X, \mathcal{F}). \tag{1.1}$$

We cannot expect to measure forward error accurately without substantial knowledge of $\mathcal{F}$. For this reason forward error is generally unknown. What is readily available to users is *backward error*,

$$\epsilon^b(X, \mathcal{F}) := \text{dist}(X, \mathcal{L}) + \text{dist}(X, \mathbb{S}_+^n). \tag{1.2}$$

In backward error we recognize that $\mathcal{F}$ is the intersection of two sets with easily computable forward errors. Backward error serves as a proxy for the unknown forward error. The type of slow convergence we are concerned with is when *backward error is sufficiently small but forward error is much larger*. The problem with this scenario is not just the poor quality of the proposed solution. More than this, it is the lack of awareness of a poor solution.

To demonstrate the discrepancy between forward error and backward error, we consider an SDP, with $n = 5$, from the family introduced in [30] (see also [28, Example 4.2.6.]). The output of `cvx` using the solver `SDPT3` is,

$$X \approx \begin{bmatrix} 0.94 & 0 & 0.028 & 0.001 & 2.3 \times 10^{-6} \\ 0 & 0.057 & 0 & 0 & 0 \\ 0.028 & 0 & 0.028 & 4.1 \times 10^{-5} & 6.5 \times 10^{-8} \\ 0.001 & 0 & 4.1 \times 10^{-5} & 4.5 \times 10^{-6} & 3.1 \times 10^{-9} \\ 2.3 \times 10^{-6} & 0 & 6.5 \times 10^{-8} & 3.1 \times 10^{-9} & 0 \end{bmatrix}, \; \lambda(X) \approx \begin{pmatrix} 0.94 \\ 0.057 \\ 1.9 \times 10^{-3} \\ 2.4 \times 10^{-6} \\ -5.4 \times 10^{-12} \end{pmatrix},$$

where $\lambda(X)$ is the vector of eigenvalues of X. Similar results were obtained with the solvers `SeDuMi` and `MOSEK`. The backward error for X is quite small at 5.46×10^{-12} and `cvx` output states that the problem is *solved*. All indicators point to a 'good' solution. However, the solution set of the SDP is a singleton consisting of the matrix with 1 in the upper left entry and zeros everywhere else. Given this information, X does *not* look like a very good solution. Indeed, forward error is 9.15×10^{-2}, unacceptably large. Moreover, the eigenvalues of X hardly indicate that the solution is a rank one matrix. In the numerical case studies of Section 5, we show that our lower bound on forward error is significantly greater than backward error. Therefore, it serves as an alarm that the proposed solution is not as accurate as it appears to be.

In [29], Sturm defined singularity degree as the fewest number of iterations required in the *facial reduction algorithm*, a regularization scheme for conic optimization introduced in [3–5]. The singularity degree of a spectrahedron $\mathcal{F}$, denoted $\text{sd}(\mathcal{F})$, is an integer between 0 and $n-1$. Loosely speaking, Sturm showed that forward error is bounded in terms of backward error and singularity degree,

$$\epsilon^f(X, \mathcal{F}) = \mathcal{O}\left(\epsilon^b(X, \mathcal{F})^{2 - \text{sd}(\mathcal{F})}\right). \tag{1.3}$$

See Fact 2.2 below for a more precise presentation of the result. In particular, this bound implies that large singularity degree is a necessary condition for large forward error. Equivalently, small singularity degree implies small forward error. It is exactly this relation that has motivated our study of singularity degree.

The challenge with singularity degree is that, like forward error, it is unknown in most cases. In [6] it is shown that the facial reduction algorithm is stable when singularity degree is 0 or 1, but the authors were not able to show that stability holds for larger singularity degree. Moreover, the empirical evidence we have obtained indicates a lack of stability of the algorithm when singularity degree is greater than 1. For this reason, we view singularity degree as intractable for general

instances of SDP. Here we obtain a lower bound on singularity degree as a consequence of the upper bound on maximum rank.

Our final contribution in this work is showing that singularity degree is also a sufficient measure, in some sense, for large forward error. We prove that for a class of central paths the eigenvalues that vanish, do so at a *fast rate* if, and only if, singularity degree is at most 1. We also prove that among the elements of the dual path that converge to 0, there are at least $\mathrm{sd}(\mathcal{F})$ different rates of convergence.

The paper is organized as follows. In Section 2 we introduce our notation and basic concepts pertaining to facial reduction and singularity degree. The bounds on maximum rank, forward error, and singularity degree are presented in Section 3; and in Section 4 we present results that support the notion that singularity degree is a measure of hardness. The paper is concluded with numerical observations in Section 5.

2 Notation and Background

Throughout this paper the ambient space is the Euclidean space of $n \times n$ real symmetric matrices, denoted $\mathbb{S}^n$, with the standard trace inner product, $\langle X, Y \rangle := \mathrm{trace}(XY)$, and the induced Frobenius norm, $\|X\|_F := \sqrt{\langle X, X \rangle}$.

The eigenvalues of any $X \in \mathbb{S}^n$ are real and ordered so as to satisfy, $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$, and $\lambda(X) \in \mathbb{R}^n$ is the vector consisting of all the eigenvalues. In terms of this notation we have $\|X\|_F = \|\lambda(X)\|_2$, where $\|\cdot\|_2$ is the Euclidian norm when the argument is a vector in $\mathbb{R}^n$. When the argument to $\|\cdot\|_2$ is a symmetric matrix then we mean the operator 2-norm, defined as $\|X\|_2 := \max_i |\lambda_i(X)|$. In some of our discussion we use the notation $\lambda_{\max}(X) = \lambda_1(X)$ and $\lambda_{\min}(X) = \lambda_n(X)$, when we are not concerned with the dimensions of the matrix, or wish to stress the minimality and maximality of the values.

The set of positive semidefinite matrices, $\mathbb{S}^n_+$, is a closed convex cone in $\mathbb{S}^n$, with interior consisting of the positive definite matrices, $\mathbb{S}^n_{++}$. The cone $\mathbb{S}^n_+$ induces the *Löwner partial order* on $\mathbb{S}^n$. That is, for $X, Y \in \mathbb{S}^n$, we write $X \succeq Y$ when $X - Y \in \mathbb{S}^n_+$, and similarly $X \succ Y$ when $X - Y \in \mathbb{S}^n_{++}$.

2.1 Facial Reduction for SDPs

We begin with a brief discussion of the faces of $\mathbb{S}^n_+$. For further reading and proofs of some of our claims, we suggest [11, 25, 30, 34]. A *face* of $\mathbb{S}^n_+$, denoted $f \trianglelefteq \mathbb{S}^n_+$, is a convex subcone of $\mathbb{S}^n_+$ such that,

$$X, Y \in \mathbb{S}^n_+, \ X + Y \in f \implies X, Y \in f.$$

For a given $f \trianglelefteq \mathbb{S}^n_+$, there exists $r \in \{0, \ldots, n\}$, $W \in \mathbb{S}^n_+$, and $V \in \mathbb{R}^{n \times r}$ such that,

$$WV = 0, \ W + VV^T \succ 0, \ f = V\mathbb{S}^r_+ V^T = \mathbb{S}^n_+ \cap W^\perp. \tag{2.1}$$

When the matrix W in (2.1) is not the 0 matrix, it is referred to as an *exposing vector* for f.

An important notion regarding SDPs is that of *minimal face*. The minimal face of $\mathbb{S}^n_+$ containing a convex set C, denoted face(C), is the intersection of all faces of $\mathbb{S}^n_+$ that contain C. If the minimal face for an SDP is known, then the SDP may be transformed into an equivalent SDP for which the *Slater condition* – strict feasibility with respect to the positive semidefinite constraint – holds. See the survey [11] or [21–23,31] for further reading on regularization of SDPs via facial reduction.

For SDPs with special structure, the minimal face may be obtained through theoretical analysis. Alternatively, the minimal face may be obtained using the *facial reduction algorithm*, introduced in [3–5]. This algorithm generates a sequence of faces $f^1, \cdots, f^d$ satisfying,

$$f^1 \supsetneq \cdots \supsetneq f^d, \quad f^d = \text{face}(\mathcal{F}).$$

Equivalently, using the two characterizations of faces in (2.1), the algorithm generates a sequence of matrices $W^1, \ldots, W^d$, a sequence of decreasing positive integers $r_1, \ldots, r_d$, and a sequence of matrices $V^k \in \mathbb{R}^{n \times r_k}$ with $k \in \{1, \ldots, d\}$ such that,

$$f^k = V^k \mathbb{S}_+^{r_k} \left(V^k\right)^T = \mathbb{S}_+^n \cap \left(W^k\right)^\perp, \ k \in \{1, \ldots, d\}.$$

The facial reduction algorithm depends on the algebraic representation of the solution set of an SDP. As stated earlier, such a set is the intersection of an affine subspace and $\mathbb{S}_+^n$. We assume that the affine subspace is defined in terms of a linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, and a vector $b \in \mathbb{R}^m$, so that,

$$\mathcal{F} = \mathcal{F}(\mathcal{A}, b) = \{X \in \mathbb{S}_+^n : \mathcal{A}(X) = b\}. \tag{2.2}$$

The notation $\mathcal{F}(\mathcal{A}, b)$ stresses the dependence on the algebraic representation of the affine subspace. The facial reduction algorithm relies on the following theorem of the alternative.

Fact 2.1. *Let $\mathcal{F} = \mathcal{F}(\mathcal{A}, b) \neq \emptyset$ be defined as in (2.2). Then exactly one of the following holds:*

(i) $\mathcal{F} \cap \mathbb{S}_{++}^n \neq \emptyset$,

(ii) *there exists nonzero $W = \mathcal{A}^*(y) \succeq 0$ with $y^T b = 0$.*

A proof of this result may be found in [11], for instance.

Now let us briefly describe how this result may be used for facial reduction. Recall that the purpose of the facial reduction algorithm is to create an equivalent representation of $\mathcal{F}$ so that the Slater condition holds. Now if Fact 2.1 Item (i) holds, then $\mathcal{F}$ has a Slater point and we are done. On the other hand if Fact 2.1 Item (ii) holds, then we easily see that the matrix W is an exposing vector for a face containing face($\mathcal{F}$),

$$X \in \mathcal{F} \implies 0 = y^T b = y^T \mathcal{A}(X) = \langle \mathcal{A}^*(y), X \rangle = \langle W, X \rangle.$$

Letting r be the nullity of W and choosing V so as to satisfy the properties in (2.1), we see that for all $X \in \mathcal{F}$ we have $X \in V\mathbb{S}_+^r V^T$. It follows that,

$$\begin{aligned}
\mathcal{F} &= \{VRV^T \in \mathbb{S}_+^n : \mathcal{A}(VRV^T) = b\} \\
&= V\{R \in \mathbb{S}_+^r : \mathcal{A}(VRV^T) = b\}V^T \\
&= V\mathcal{F}\left(\mathcal{A}(V \cdot V^T), b\right)V^T.
\end{aligned}$$

In the next iteration of the algorithm we apply Fact 2.1 to the spectrahedron $\mathcal{F}\left(\mathcal{A}(V \cdot V^T), b\right)$. We continue in this way until eventually Fact 2.1 Item (i) – the Slater condition – holds. We have included Algorithm 2.1, page 5, as a more rigorous description of facial reduction.

At each iteration, the order of the semidefinite matrices is reduced by at least one, implying that the algorithm terminates in at most n iterations. In fact, the upper bound is actually $n - 1$ since in the case $\mathcal{F} = \{0\}$ it can be shown that we can choose W of full rank and the algorithm terminates in exactly 1 iteration. Moreover, it is shown in [5, Alg. B], [28, Thm 3.5.4], that at least one constraint becomes redundant at each iteration. Therefore the upper bound is $\min\{m, n - 1\}$.

4

Algorithm 2.1 Facial Reduction for: f, V, W, Rank r_d in (2.1), and Singularity Degree d

1: **INPUT:** $\mathcal{A}$, b.
2: **initialize:** $k = 0$, $\mathcal{A}^k = \mathcal{A}$, $V^k = I$, $W^k = 0$, $r_k = n$, $q_k = 0$
3: **define:** $S_k = \left\{ Z : 0 \neq Z = \left(\mathcal{A}^k\right)^* (y) \succeq 0, y^T b = 0 \right\}$
4: **while** $S_k \neq \emptyset$ **do**
5:      obtain (nonzero) $Z^{k+1} \in S_k$ and orthogonal $\begin{bmatrix} Q_1^{k+1} & Q_2^{k+1} \end{bmatrix}$ such that,

$$Z^{k+1} = \begin{bmatrix} Q_1^{k+1} & Q_2^{k+1} \end{bmatrix} \begin{bmatrix} \Lambda^{k+1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^{k+1} & Q_2^{k+1} \end{bmatrix},$$

   where $\Lambda^{k+1} \succ 0$ and $Q_1^{k+1} \in \mathbb{R}^{r_k \times q_{k+1}}$.
6:      **if** $Z^{k+1} \succ 0$ **then**
7:          $Q_2^{k+1} = 0 \in \mathbb{S}^{r_k}$, $V^{k+1} = 0 \in \mathbb{S}^n$, and $r_{k+1} = 0$
8:      **else**
9:          $Q_2^{k+1} \in \mathbb{R}^{r_k \times r_{k+1}}$ and $V^{k+1} = V^k Q_2^{k+1} \in \mathbb{R}^{n \times r_{k+1}}$
10:     **end if**
11:     $W^{k+1} = W^k + V^k Z^{k+1} \left(V^k\right)^T \in \mathbb{S}_+^n$
12:     $\mathcal{A}^{k+1} = \mathcal{A} \left(V^{k+1} \cdot \left(V^{k+1}\right)^T \right)$
13:     $k = k + 1$
14:     $S_k = \left\{ Z : 0 \neq Z = \left(\mathcal{A}^k\right)^* (y) \succeq 0, y^T b = 0 \right\}$
15: **end while**
16: **OUTPUT:** $d = k$, $V = V^k$, $W = W^k$, $r = r_d$, $f = V\mathbb{S}_+^r V^T (= \mathbb{S}_+^n \cap W^\perp)$.

2.2 Singularity Degree and the Bounds of Sturm

The number of iterations required by the facial reduction algorithm of the previous section is dependent on the choice of exposing vector, Z^{k+1}, obtained at each iteration. When the exposing vector is chosen to have maximum rank, the number of iterations generated by Algorithm 2.1 is defined as the *singularity degree*, and denoted $\mathrm{sd}(\mathcal{F})$. It can be shown that singularity degree is the least number of iterations required by the algorithm.

Two special cases deserve mention. The first case is that of $\mathrm{sd}(\mathcal{F}) = 0$. This case occurs if, and only if, the solution set satisfies the Slater condition. The second special case is when $\mathcal{F} = \{0\}$. Here our definition of singularity degree does not coincide with that of Sturm. By our definition we have $\mathrm{sd}(\mathcal{F}) = 1$, since exactly one iteration of the algorithm is required to obtain the Slater condition. On the other hand, Sturm defines $\mathrm{sd}(\mathcal{F}) = 0$ for this case, based on his error bounds. That is, the worst case error bounds when $\mathcal{F} = \{0\}$ are the same as when $\mathcal{F}$ satisfies the Slater condition.

We now state the error bounds of Sturm. For a proof see [29, Theorem 3.3].

Fact 2.2. *Let $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ be a nonempty spectrahedron, and let $\{X(\alpha) \in \mathbb{S}^n : \alpha > 0\}$ be a sequence where $\|X(\alpha)\|$ is bounded. Then,*

$$\epsilon^f(X(\alpha), \mathcal{F}) = \begin{cases} \mathcal{O}\left(\epsilon^b(X(\alpha), \mathcal{F})\right), & \text{if } \mathcal{F} = \{0\}, \\ \mathcal{O}\left(\epsilon^b(X(\alpha), \mathcal{F})^{2^{-\mathrm{sd}(\mathcal{F})}}\right), & \text{otherwise.} \end{cases}$$

In proving Fact 2.2, Sturm actually obtained the following more precise statement about the way in which $X(\alpha)$ approaches $\mathcal{F}$.

Fact 2.3. *Let $0 \neq \mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ be a nonempty spectrahedron with $\mathrm{sd}(\mathcal{F}) \geq 1$; and let $\{X(\alpha) \in \mathbb{S}^n : \alpha > 0\}$ be a sequence where $\epsilon^b(X(\alpha), \mathcal{F}) = \mathcal{O}(\alpha)$. For $i \in \{1, \ldots, \mathrm{sd}(\mathcal{F})\}$, let Z^i be a maximum rank exposing vector obtained as in Algorithm 2.1, with $q_i := \mathrm{rank}(Z^i)$. Let $\bar{\alpha} > 0$ be fixed. Then there exists an orthogonal matrix Q such that,*

$$\mathrm{face}(Q\mathcal{F}Q^T) = \begin{bmatrix} \mathbb{S}^r_+ & 0 \\ 0 & 0 \end{bmatrix},$$

and,

$$QX(\alpha)Q^T = \begin{bmatrix} X_0(\alpha) & * & \cdots & & * \\ * & X_1(\alpha) & & & \\ \vdots & & \ddots & & * \\ * & * & * & & X_{\mathrm{sd}(\mathcal{F})}(\alpha) \end{bmatrix},$$

where $X_0(\alpha) \in \mathbb{S}^r$ and for all $i \in \{1, \ldots, \mathrm{sd}(\mathcal{F})\}$ and $\alpha \in (0, \bar{\alpha})$ it holds that,

$$X_i(\alpha) \in \mathbb{S}^{q_i} \ \text{and} \ \|X_i(\alpha)\| = \mathcal{O}\left(\alpha^{\xi(i)}\right), \ \xi(i) := 2^{-(\mathrm{sd}(\mathcal{F})-i)}.$$

This result shows that under the correct orthogonal transformation, the diagonal blocks of $X(\alpha)$ that converge to 0 may do so at different rates.

3 Bounds on Maximum Rank, Forward Error, and Singularity Degree

We consider SDPs in the form,

$$\begin{aligned} p^* := \min \ & \langle C, X \rangle \\ \text{s.t.} \ & \widehat{\mathcal{A}}(X) = \widehat{b} \\ & X \succeq 0, \end{aligned} \tag{3.1}$$

where $\widehat{\mathcal{A}} : \mathbb{S}^n \to \mathbb{R}^{\widehat{m}}$ is a linear map, $\widehat{b} \in \mathbb{R}^{\widehat{m}}$, $C \in \mathbb{S}^n$, and p^* is finite. The solution set, optimality set of (3.1), is the spectrahedron

$$\mathcal{F} = \{X \in \mathbb{S}^n_+ : \widehat{\mathcal{A}}(X) = \widehat{b}, \ \langle C, X \rangle = p^*\}. \tag{3.2}$$

When (3.1) is a feasibility problem, then $C = 0$ and $p^* = 0$. In this case we let $m := \widehat{m}$, $\mathcal{A} := \widehat{\mathcal{A}}$, and $b := \widehat{b}$ so that $\mathcal{F} = \mathcal{F}(\mathcal{A}, b) = \mathcal{F}(\widehat{\mathcal{A}}, \widehat{b})$. On the other hand, when the objective of (3.1) is non-trivial, i.e., $C \neq 0$, then it may not be the case that $\mathcal{F} = \mathcal{F}(\widehat{\mathcal{A}}, \widehat{b})$. Thus we define $m := \widehat{m} + 1$ and define $\mathcal{A}$ and b as,

$$\mathcal{A}(X) := \begin{pmatrix} \widehat{\mathcal{A}}(X) \\ \langle C, X \rangle \end{pmatrix}, \ b := \begin{pmatrix} \widehat{b} \\ p^* \end{pmatrix}.$$

Then $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ is a spectrahedron in the notation we have already developed. Note that in the $C \neq 0$ case, strict feasibility fails in general as optimal solutions are singular in general. In the likely case that p^* is not known, the affine manifold cannot be constructed. However, in Remark 3.10 we show that the results of this section can still be applied.

Assumption 3.1. *Let $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ be the spectrahedron consisting of the solutions to the SDP in (3.1). We assume that $\mathcal{F}$ is nonempty, $\mathrm{sd}(\mathcal{F}) \geq 1$, and $\mathcal{F} \neq \{0\}$.*

The assumption that $\mathcal{F}$ is nonempty is for the purpose of defining singularity degree. The other two assumptions ensure that there is a possibility of discrepancy between forward error and backward error.

Our analysis in this section is based on *path-following* algorithms for SDP. The foundation of such algorithms is the *central path*, a smooth parametric curve, say $\{X(\alpha) : \alpha > 0\}$, that is known to converge to a solution of the SDP, granted that a solution exists. Specifically we mean that $X(\alpha) \to \bar{X}$ as $\alpha \searrow 0$ and $\bar{X} \in \mathcal{F}$. Then a path-following algorithm produces a sequence of positive numbers $\{\alpha_k\}$ and matrices $\{X^k\}$ such that α_k is successively closer to 0 and X^k is a successively better approximation of $X(\alpha_k)$. In other words, the iterates approach the solution set along a trajectory that approximates the central path.

Remark 3.2. *We have chosen to work with path-following methods. The approximations X^k for such methods are typically obtained by applying a second order approximation, a Newton type method, onto (perturbed) optimality conditions. For appropriate convergence, one requires full rank of the Jacobian in the limit, loss of which results in inaccurate search directions, slow convergence, and low precision solutions. For standard primal-dual path-following methods, it is known that loss of strict feasibility and/or strict complementarity results in degeneracy, loss of full rank, see e.g., [1, 9, 33]. First order methods require regularity as well, e.g. [10]. To avoid the slow convergence and/or loss of precision, one has to somehow remove the degeneracy. For loss of strict feasibility, that is what facial reduction is about. Removing the degeneracy arising from strict complementarity requires different approaches, e.g., recognizing and deflating appropriate eigenvectors of primal-dual solutions.*[1]

To obtain a method for bounding maximum rank and forward error, we study central paths that satisfy the following assumptions.

Assumption 3.3. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path and let $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ be the spectrahedron consisting of the solutions to the SDP in (3.1). We assume that,*

(i) there exists $\bar{X} \in \mathrm{relint}(\mathcal{F})$ such that $\lim_{\alpha \searrow 0} X(\alpha) = \bar{X}$,

(ii) $X(\alpha) \succ 0$ for all $\alpha > 0$.

Many of the well-known algorithms for SDP are based on central paths that satisfy this assumption, e.g., [18].

3.1 A Bound on Maximum Rank

We provide two ways to bound maximum rank. The first method is based on tracking the ratios,

$$\frac{\lambda_i(X(\alpha))}{\lambda_{i+1}(X(\alpha))}, \ i \in \{1, \ldots, n-1\}. \tag{3.3}$$

The ratios that diverge to $+\infty$ indicate one of two scenarios. The first scenario is that both eigenvalues converge to 0, but $\lambda_{i+1}(X(\alpha))$ does so much more quickly. The second is that $\lambda_i(X(\alpha))$ converges to a positive value and $\lambda_{i+1}(X(\alpha))$ vanishes. This only happens when i corresponds to the rank of the limit point $\bar{X}$. Thus the smallest index i for which the ratio tends to infinity as $\alpha \to 0$ corresponds exactly to the maximum rank. We state this observation formally in the following.

---

[1]It is conceivable that higher order methods of the recent type introduced in e.g., [13, 19] might outperform our second order approach. However, even these require some sort of regularity.

Proposition 3.4. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path satisfying Assumption 3.3 for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 3.1. Let i be the smallest index such that the ratio in (3.3) tends to infinity as $\alpha \to 0$. Then i is the maximum rank over $\mathcal{F}$.*

In practice, a ratio of the form in (3.3) may diverge to $+\infty$ slowly, so that it may appear to be bounded. In such cases an approximation of the maximum rank may be obtained by the smallest index i for which it is *clear that the ratio tends to infinity as $\alpha \to 0$.*

In addition to differentiating between eigenvalues that vanish and those that do not, the ratios in (3.3) also indicate the number of different rates of convergence among the eigenvalues that do converge to 0. We explore this further in Section 4.2.

The second method to bound maximum rank is more involved, but provides us with additional information on singularity degree. Moreover, this method appears to be more reliable in our experiments. The approach is to analyze the Q-convergence rates of the eigenvalues of $X(\alpha)$. We begin by translating Fact 2.3 into a statement about the eigenvalues of $X(\alpha)$.

Lemma 3.5. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path satisfying Assumption 3.3 for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ that satisfies Assumption 3.1. Assume further that $\epsilon^b(X(\alpha), \mathcal{F}) = \mathcal{O}(\alpha)$. For $i \in \{1, \dots, \mathrm{sd}(\mathcal{F})\}$, let Z^i be a maximum rank exposing vector obtained as in Algorithm 2.1 and let $q_i := \mathrm{rank}(Z^i)$. Let r denote the maximum rank over $\mathcal{F}$. Let $\mathcal{I}^0, \mathcal{I}^1, \dots, \mathcal{I}^{\mathrm{sd}(\mathcal{F})}$ form a partition of $\{1, \dots, n\}$ such that $\mathcal{I}^0 = \{1, \dots, r\}$ and,*

$$\mathcal{I}^1 = r + \{1, \dots, q^1\}, \ \mathcal{I}^2 = r + q^1 + \{1, \dots, q^2\},$$

and so on. Then, for $j \in \{1, \dots, n\}$ it holds that for sufficiently small $\alpha > 0$,

$$j \in \mathcal{I}^i \implies \lambda_j(X(\alpha)) = \begin{cases} \Theta(1) & \text{if } i = 0, \\ \mathcal{O}\left(\alpha^{\xi(i)}\right) & \text{otherwise,} \end{cases}$$

where $\xi(i) := 2^{-(\mathrm{sd}(\mathcal{F}) - i)}$.

Proof. By assumption, $X(\alpha) \to \bar{X} \in \mathrm{relint}(\mathcal{F})$ and $\mathrm{rank}(\bar{X}) = r$. Therefore for sufficiently small $\alpha > 0$, the r largest eigenvalues of $X(\alpha)$ are converging to positive numbers. It follows that for sufficiently small $\alpha > 0$,

$$j \in \mathcal{I}^0 \implies \lambda_j(X(\alpha)) = \Theta(1),$$

proving one part of the desired result.

Next, by Fact 2.3 there exists an orthogonal Q such that,

$$\mathrm{face}(Q\mathcal{F}Q^T) = \begin{bmatrix} \mathbb{S}_+^r & 0 \\ 0 & 0 \end{bmatrix} \text{ and } QX(\alpha)Q^T = \begin{bmatrix} X_0(\alpha) & * & \cdots & * \\ * & X_1(\alpha) & & \\ \vdots & & \ddots & * \\ * & * & * & X_{\mathrm{sd}(\mathcal{F})}(\alpha) \end{bmatrix} \ \forall \alpha > 0, \quad (3.4)$$

where $X_0(\alpha) \in \mathbb{S}^r$ and for all $i \in \{1, \dots, \mathrm{sd}(\mathcal{F})\}$ it holds that,

$$X_i(\alpha) \in \mathbb{S}^{q_i} \text{ and } \|X_i(\alpha)\| = \mathcal{O}\left(\alpha^{\xi(i)}\right). \quad (3.5)$$

Now let $i \in \{1, \dots, \mathrm{sd}(\mathcal{F})\}$ and let $j \in \mathcal{I}^i$. Consider the principal submatrix of $QX(\alpha)Q^T$,

$$S(\alpha) := \begin{bmatrix} X_i(\alpha) & & \\ & \ddots & \\ & & X_{\mathrm{sd}(\mathcal{F})}(\alpha) \end{bmatrix}. \quad (3.6)$$

8

By Assumption 3.3, $X(\alpha) \succ 0$ implying that $S(\alpha) \succ 0$. Thus by (3.5) we have,

$$\|S(\alpha)\| = \mathcal{O}\left(\max_{\ell \in \{i,\dots,\mathrm{sd}(\mathcal{F})\}} \|X_\ell(\alpha)\|\right) = \mathcal{O}\left(\alpha^{\xi(i)}\right).$$

It follows that $\lambda_1(S) = \mathcal{O}\left(\alpha^{\xi(i)}\right)$. Moreover, the interlacing eigenvalue theorem implies that,

$$\lambda_j(X(\alpha)) \leq \lambda_1(S).$$

Combining these inequalities yields the desired result, $\lambda_j(X(\alpha)) = \mathcal{O}\left(\alpha^{\xi(i)}\right)$. $\qquad\square$

Now that we have bounds on those eigenvalues of $X(\alpha)$ that converge to 0 we analyze their Q-convergence rates. First a lemma.

Lemma 3.6. *Let $\{a_k\}_{k\in\mathbb{N}}$ and $\{b_k\}_{k\in\mathbb{N}}$ be sequences of positive reals such that $a_k \to 0$ and $b_k \to 0$. If $a_k \leq b_k$ for all $k \in \mathbb{N}$ then,*

$$\liminf_{k\to\infty} \frac{a_{k+1}}{a_k} \leq \limsup_{k\to\infty} \frac{b_{k+1}}{b_k}. \tag{3.7}$$

Proof. Let L_a and L_b denote the limit inferior and limit superior of (3.7), respectively. For simplicity we assume that L_a and L_b are finite, but the arguments extend to the general case trivially. Suppose for the sake of contradiction that there exists $\tau > 0$ such that $L_a - \tau \geq L_b$. Then there exists $\bar{k} \in \mathbb{N}$ such that for all $k \geq \bar{k}$,

$$\frac{a_{k+1}}{a_k} \geq L_a - \frac{\tau}{3} \quad \text{and} \quad \frac{b_{k+1}}{b_k} \leq L_a - \frac{\tau}{2} \tag{3.8}$$

Rearranging the first equation in (3.8) gives us,

$$a_{k+1} \geq a_k \left(L_a - \frac{\tau}{3}\right), \quad \forall k \geq \bar{k}. \tag{3.9}$$

Replacing k with $k-1$ we get that,

$$a_k \geq a_{k-1} \left(L_a - \frac{\tau}{3}\right), \quad \forall k \geq \bar{k} + 1. \tag{3.10}$$

Combining (3.9) with (3.10) yields,

$$a_{k+1} \geq a_{k-1} \left(L_a - \frac{\tau}{3}\right)^2, \quad \forall k \geq \bar{k} + 1.$$

Continuing in this fashion we get,

$$a_k \geq a_{\bar{k}} \left(L_a - \frac{\tau}{3}\right)^{k-\bar{k}} = \frac{a_{\bar{k}}}{\left(L_a - \frac{\tau}{3}\right)^{\bar{k}}} \left(L_a - \frac{\tau}{3}\right)^k, \quad \forall k \geq \bar{k}. \tag{3.11}$$

Through an analogous approach applied to the second equation of (3.8) we get,

$$b_k \leq \frac{b_{\bar{k}}}{\left(L_a - \frac{\tau}{2}\right)^{\bar{k}}} \left(L_a - \frac{\tau}{2}\right)^k, \quad \forall k \geq \bar{k}. \tag{3.12}$$

Combining the hypothesis that b_k dominates a_k with (3.11) and (3.12) we get,

$$\frac{b_{\bar{k}}}{\left(L_a - \frac{\tau}{2}\right)^{\bar{k}}} \left(L_a - \frac{\tau}{2}\right)^k \geq \frac{a_{\bar{k}}}{\left(L_a - \frac{\tau}{3}\right)^{\bar{k}}} \left(L_a - \frac{\tau}{3}\right)^k, \quad \forall k \geq \bar{k}. \tag{3.13}$$

Observe that $L_b \geq 0$ since $b_k \geq 0$ for every $k \in \mathbb{N}$. Therefore, $L_a - \tau \geq 0$ and we have $L_a - \frac{\tau}{3} > L_a - \frac{\tau}{2} > 0$. It follows that for sufficiently large k, the inequality in (3.13) is violated, giving us the desired contradiction. $\qquad\square$

Our main result on Q-convergence rates of eigenvalues considers sequences of the form $\{\sigma_k\}$ for some $\sigma \in (0, 1)$ where the kth term, σ_k, is the kth power of σ, i.e. σ^k.

Theorem 3.7. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path satisfying Assumption 3.3 for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ that satisfies Assumption 3.1. Assume further that $\epsilon^b(X(\alpha), \mathcal{F}) = \mathcal{O}(\alpha)$. Let $\mathcal{I}^0, \mathcal{I}^1, \ldots, \mathcal{I}^{\mathrm{sd}(\mathcal{F})}$ be a partition of $\{1, \ldots, n\}$ as in Lemma 3.5. For $\sigma \in (0, 1)$ the following hold.*

(i) *If $j \in \mathcal{I}^0$ then,*
$$\lim_{k \to \infty} \frac{\lambda_j(X(\sigma^{k+1}))}{\lambda_j(X(\sigma^k))} = 1.$$

(ii) *If $j \in \mathcal{I}^i$ with $i \in \{1, \ldots, \mathrm{sd}(\mathcal{F})\}$ then,*
$$\liminf_{k \to \infty} \frac{\lambda_j(X(\sigma^{k+1}))}{\lambda_j(X(\sigma^k))} \leq \sigma^{\xi(i)} < 1,$$

where $\xi(i) = 2^{-(\mathrm{sd}(\mathcal{F})-i)}$.

Proof. By Assumption 3.3 and the definition of $\mathcal{I}^0$ we have that $\lambda_j(X(\sigma^k))$ converges, in k, to a positive number whenever $j \in \mathcal{I}^0$. We have proved Item (i).

Now let $j \in \mathcal{I}^i$ with $i \in \{1, \ldots, \mathrm{sd}(\mathcal{F})\}$. By Lemma 3.5 we have,
$$\lambda_j(X(\sigma^k)) = \mathcal{O}\left(\left(\sigma^k\right)^{\xi(i)}\right), \quad \forall k \in \mathbb{N}. \tag{3.14}$$

Thus there exists $M > 0$ such that $\lambda_j(X(\sigma^k)) \leq M\sigma^{k\xi(i)}$. Now the sequences $\{\lambda_j(X(\sigma^k))\}_{k \in \mathbb{N}}$ and $\{M\sigma^{k\xi(i)}\}_{k \in \mathbb{N}}$ satisfy the assumptions of Lemma 3.6. Therefore,
$$\liminf_{k \to \infty} \frac{\lambda_j(X(\sigma^{k+1}))}{\lambda_j(X(\sigma^k))} \leq \limsup_{k \to \infty} \frac{M\sigma^{(k+1)\xi(i)}}{M\sigma^{k\xi(i)}} = \sigma^{\xi(i)}.$$

Lastly $\sigma^{\xi(d)} < 1$ holds since, $\sigma \in (0, 1)$ and $\xi(i) > 0$. $\qquad\square$

The following corollary emphasizes how Theorem 3.7 may be used to upper bound the rank of $\bar{X}$.

Corollary 3.8. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path satisfying Assumption 3.3 for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ that satisfies Assumption 3.1. Assume further that $\epsilon^b(X(\alpha), \mathcal{F}) = \mathcal{O}(\alpha)$. Let r denote the maximum rank over $\mathcal{F}$ and let $\sigma \in (0, 1)$. Then,*
$$\liminf_{k \to \infty} \frac{\lambda_i(X(\sigma^{k+1}))}{\lambda_i(X(\sigma^k))} \leq \sigma^{\xi(1)} = \sigma^{2^{-(\mathrm{sd}(\mathcal{F})-1)}} < 1 \quad \Longleftrightarrow \quad i > r.$$

The number $\sigma^{2^{-(\mathrm{sd}(\mathcal{F})-1)}}$ serves as a threshold so that limit inferiors of the eigenvalue ratios lie below this number if, and only if, those eigenvalues converge to 0. For large singularity degree, it may be difficult to distinguish $\sigma^{2^{-(\mathrm{sd}(\mathcal{F})-1)}}$ from 1, numerically. However if we can identify another number, say $\tau \in (0, 1)$, that is numerically distinguishable from 1 and there exists a positive integer $\bar{r}$ such that,
$$\liminf_{k \to \infty} \frac{\lambda_i(X(\sigma^{k+1}))}{\lambda_i(X(\sigma^k))} \leq \tau \quad \Longleftrightarrow \quad i > \bar{r},$$

then $\bar{r}$ is an upper bound on the maximum rank, r, over $\mathcal{F}$. We state this result formally in the following.

Corollary 3.9. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path satisfying Assumption 3.3 for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ that satisfies Assumption 3.1. Assume further that $\epsilon^b(X(\alpha), \mathcal{F}) = \mathcal{O}(\alpha)$. Let r denote the maximum rank over $\mathcal{F}$ and let $\sigma \in (0, 1)$. Suppose there exists $\tau \in (0, 1)$ and $\bar{r} \in \{1, \ldots, n\}$ such that*

$$\liminf_{k \to \infty} \frac{\lambda_i(X(\sigma^{k+1}))}{\lambda_i(X(\sigma^k))} \leq \tau \iff i > \bar{r}.$$

Then $\bar{r} \geq r$.

Remark 3.10. *In the results of this section we require knowledge of the magnitude of backward error for $\mathcal{F}$. When the SDP in (3.1) has a non-trivial objective, we cannot expect to know the optimal value p^*. Consequently, backward error is intractable. However, taking a cue from Sturm [29], we use the duality gap to upper bound $(\langle C, X \rangle - p^*)$. Thus,*

$$\epsilon^b(X, \mathcal{F}) \leq \text{dist}\left(X, \{S : \widehat{\mathcal{A}}(S) = \widehat{b}\}\right) + \text{dist}(X, \mathbb{S}^n_+) + (\langle C, X \rangle - p^*).$$

In particular, for the central path $\{X(\alpha) : \alpha > 0\}$ it holds that,

$$\text{dist}\left(X(\alpha), \{S : \widehat{\mathcal{A}}(S) = \widehat{b}\}\right) + \text{dist}(X(\alpha), \mathbb{S}^n_+) + (\langle C, X(\alpha) \rangle - p^*) = \mathcal{O}(\alpha) \implies \epsilon^b(X(\alpha), \mathcal{F}) = \mathcal{O}(\alpha),$$

the assumed bound in the above results. This observation allows us to apply the main results of this section to SDPs with non-trivial objectives even when the optimal value is not known.

3.2 Bounds on Forward Error and Singularity Degree

Any bound on maximum rank, such as the one from the previous section, may be used to provide lower bounds on forward error and singularity degree.

Theorem 3.11. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path satisfying Assumption 3.3 for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ that satisfies Assumption 3.1. If $\bar{r}$ is an upper bound for the maximum rank over $\mathcal{F}$ then for all $\alpha > 0$ it holds that,*

$$\epsilon^f(X(\alpha), \mathcal{F}) \geq \|\left(\lambda_{\bar{r}+1}(X(\alpha)) \quad \cdots \quad \lambda_n(X(\alpha))\right)^T\|_2.$$

Proof. Let $\bar{r}$ be as in the hypothesis and let $\alpha > 0$. Since $\mathcal{F}$ is a closed convex set, there exists $X \in \mathcal{F}$ such that,
$$\epsilon^f(X(\alpha), \mathcal{F}) = \|X(\alpha) - X\|_F.$$
Then observing that $\|S\|_F = \|\lambda(S)\|_2$ for any $S \in \mathbb{S}^n$ we have,

$$\begin{aligned}
\epsilon^f(X(\alpha), \mathcal{F})^2 &= \|X(\alpha) - X\|_F^2 \\
&= \|X(\alpha)\|_F^2 + \|X\|_F^2 - 2\langle X(\alpha), X \rangle \\
&= \|\lambda(X(\alpha))\|_2^2 + \|\lambda(X)\|_2^2 - 2\langle X(\alpha), X \rangle.
\end{aligned}$$

Applying the classical bound $\langle X, Y \rangle \leq \lambda(X)^T \lambda(Y)$, e.g. [12, 14], we get,

$$\begin{aligned}
\epsilon^f(X(\alpha), \mathcal{F})^2 &\geq \|\lambda(X(\alpha))\|_2^2 + \|\lambda(X)\|_2^2 - 2\lambda(X(\alpha))^T \lambda(X) \\
&= \|\lambda(X(\alpha)) - \lambda(X)\|_2^2 \\
&\geq \|\left(\lambda_{\bar{r}+1}(X(\alpha)) \quad \cdots \quad \lambda_n(X(\alpha))\right)^T\|_2^2.
\end{aligned}$$

Taking the square root of both sides yields the desired result. $\qquad \square$

Theorem 3.12. *Let $\{X(\alpha) : \alpha > 0\}$ be a central path satisfying Assumption 3.3 for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ that satisfies Assumption 3.1. Assume further that $\epsilon^b(X(\alpha), \mathcal{F}) = \mathcal{O}(\alpha)$. Let $\bar{r}$ be an upper bound on the maximum rank over $\mathcal{F}$ and let $\sigma \in (0, 1)$. Suppose $\underline{d}$ is the smallest positive integer such that,*

$$\liminf_{k \to \infty} \frac{\lambda_i(X(\sigma^{k+1}))}{\lambda_i(X(\sigma^k))} \leq \sigma^{2^{-(\underline{d}-1)}} \quad \Longleftrightarrow \quad i > \bar{r}. \tag{3.15}$$

Then $\underline{d} \leq \mathrm{sd}(\mathcal{F})$.

Proof. Let r denote the maximum rank over $\mathcal{F}$. Suppose $\bar{r} > r$. Then by Corollary 3.8 and (3.15) we have,

$$\sigma^{2^{-(\underline{d}-1)}} < \liminf_{k \to \infty} \frac{\lambda_{r+1}(X(\sigma^{k+1}))}{\lambda_{r+1}(X(\sigma^k))} \leq \sigma^{2^{-(\mathrm{sd}(\mathcal{F})-1)}}.$$

It follows that $\underline{d} \leq \mathrm{sd}(\mathcal{F})$. Now suppose that $\bar{r} = r$. Then by Corollary 3.8 we have,

$$\liminf_{k \to \infty} \frac{\lambda_i(X(\sigma^{k+1}))}{\lambda_i(X(\sigma^k))} \leq \sigma^{2^{-(\mathrm{sd}(\mathcal{F})-1)}} \quad \Longleftrightarrow \quad i > r = \bar{r}. \tag{3.16}$$

Out of all positive integers that could replace $\mathrm{sd}(\mathcal{F})$ in (3.16), we chose $\underline{d}$ to be the smallest. Hence $\underline{d} \leq \mathrm{sd}(\mathcal{F})$, as desired. $\qquad\square$

4 Singularity Degree as a Measure of Hardness

It is certainly possible to construct a parametric curve $\{X(\alpha) : \alpha > 0\}$ with the properties of Assumption 3.3, for which singularity degree is large, but forward error is small. For instance, the path defined as $X(\alpha) := \bar{X} + \alpha I$, where $\bar{X} \in \mathrm{relint}(\mathcal{F})$, exhibits fast convergence and low forward error irrespective of the singularity degree. This demonstrates that singularity degree is not a sufficient condition for slow convergence for *all* parametric curves. However, for many of the central paths constructed by state of the art algorithms, empirical evidence indicates otherwise. In this section we present several results that give credence to the notion that singularity degree is a measure of hardness for a family of central paths. In Section 4.1 we introduce the family of central paths and in Section 4.2 we present the main results of the section.

4.1 Analysis of a Family of Central Paths

The classical *interior point* method for SDPs is based on a central path that is constructed by assuming that both the primal and the dual satisfy the Slater condition. As this assumption is quite restrictive, *infeasible* central paths assuming weaker conditions have been subsequently proposed. Among these are [7, 8, 16, 20, 24].

In [24], Potra and Sheng proposed a family of infeasible central paths that are based on perturbing the feasible region so as to satisfy the Slater condition, and then decreasing the perturbation. Our analysis requires a complete knowledge of the algebraic representation of the spectrahedron. Thus we assume that $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ for a known linear map $\mathcal{A}$ and vector b. In terms of the SDP of (3.1), this equates to feasibility problems, i.e., $C = 0$ and $\mathcal{A} := \widehat{\mathcal{A}}$, $b := \widehat{b}$. From the family of paths proposed by Potra and Sheng we choose the path $\{X(\alpha) : \alpha > 0\}$ defined by,

$$\begin{cases} X(\alpha) := \arg\max\{\alpha \log \det(X) : X \in \mathcal{F}(\alpha)\}, \\ \mathcal{F}(\alpha) := \{X \in \mathbb{S}_+^n : \mathcal{A}(X) = b(\alpha)\}, \\ b(\alpha) := b + \alpha \mathcal{A}(B), \end{cases} \tag{4.1}$$

where $B \succ 0$ is fixed. The matrix $X(\alpha)$ exists for each $\alpha > 0$ if, and only if, $\mathcal{F}(\alpha)$ is nonempty and bounded. Hence the following assumption.

Assumption 4.1. *Let $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ be a spectrahedron defined in terms of a map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ and a vector $b \in \mathbb{R}^m$. We assume that,*

  (i) $\mathcal{F}$ is nonempty, bounded, $\mathrm{sd}(\mathcal{F}) \geq 1$, and $\mathcal{F} \neq \{0\}$,

  (ii) $\mathcal{A}$ is surjective.

Assumption 4.1 differs from Assumption 3.1 in the additional requirements that $\mathcal{F}$ is bounded and that $\mathcal{A}$ is surjective. The need for a bounded $\mathcal{F}$ has already been discussed and the restriction on $\mathcal{A}$ ensures a unique $y \in \mathbb{R}^m$ for every $Z \in \mathrm{range}(\mathcal{A}^*)$, a property that will prove convenient in the subsequent discussion.

It is easy to see that if $\mathcal{F} \neq \emptyset$ then $\mathcal{F}(\alpha)$ has a Slater point for every $\alpha > 0$. For instance, the set $\mathcal{F} + \alpha B$ has positive definite elements and is contained in $\mathcal{F}(\alpha)$. Since $X(\alpha)$ is chosen to be the determinant maximizer over $\mathcal{F}(\alpha)$ it follows that $X(\alpha) \succ 0$ and $X(\alpha) \in \mathrm{relint}(\mathcal{F})$ for each $\alpha > 0$. We have thus shown that this central path satisfies Assumption 3.3 Item (ii). In the remainder of this section we show that it also possesses the other properties of Assumption 3.3. Namely, that $\{X(\alpha) : \alpha > 0\}$ is smooth and converges to a matrix in $\mathrm{relint}(\mathcal{F})$ as $\alpha \searrow 0$.

The optimality conditions for (4.1) yield the primal-dual central path,

$$\left\{ (X(\alpha), y(\alpha), Z(\alpha)) \in \mathbb{S}^n_{++} \times \mathbb{R}^m \times \mathbb{S}^n_{++} : \begin{bmatrix} \mathcal{A}^*(y(\alpha)) - Z(\alpha) \\ \mathcal{A}(X(\alpha)) - b(\alpha) \\ Z(\alpha) X(\alpha) - \alpha I \end{bmatrix} = 0, \ \alpha > 0 \right\}. \qquad (4.2)$$

This primal-dual central path is smooth around every $\alpha > 0$ and admits the following convergence result.

Theorem 4.2. *Let $\{(X(\alpha), y(\alpha), Z(\alpha)) : \alpha > 0\}$ be the primal-dual central path of (4.2) for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 4.1. Then,*

$$\lim_{\alpha \searrow 0} (X(\alpha), y(\alpha), Z(\alpha)) = (\bar{X}, \bar{y}, \bar{Z}) \in \mathbb{S}^n_+ \times \mathbb{R}^m \times \mathbb{S}^n_+,$$

with

$$\begin{cases} \bar{X} \in \mathrm{relint}(\mathcal{F}), \\ \bar{Z} = \mathcal{A}^*(\bar{y}), \\ \bar{Z} \in \mathrm{relint}\{ Z \in \mathbb{S}^n_+ \setminus \{0\} : Z = \mathcal{A}^*(y), \ y^T b = 0, \ y \in \mathbb{R}^m \}. \end{cases} \qquad (4.3)$$

Proof. For a proof see [28, Section 6.1]. $\qquad \square$

An immediate consequence of Theorem 4.2 is a statement about the convergence rates of eigenvalues of $X(\alpha)$ and $Z(\alpha)$.

Corollary 4.3. *Let $\{(X(\alpha), y(\alpha), Z(\alpha)) : \alpha > 0\}$ be the primal-dual central path of (4.2) for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 4.1. Let $(\bar{X}, \bar{y}, \bar{Z})$ be the limit point of the primal-dual central path and let r and q denote the rank of $\bar{X}$ and $\bar{Z}$, respectively. Let $\bar{\alpha} > 0$. Then for $\alpha \in (0, \bar{\alpha})$ it holds that $\epsilon^b(X(\alpha)) = \mathcal{O}(\alpha)$. Moreover,*

$$\lambda_i(X(\alpha)) = \begin{cases} \Theta(1) & i \leq r, \\ \Omega(\alpha) \ and \ \neq \mathcal{O}(\alpha) & i \in [r+1, n-q], \\ \Theta(\alpha) & i \geq n-q+1, \end{cases} \qquad (4.4)$$

13

and

$$\lambda_i(Z(\alpha)) = \begin{cases} \Theta(1) & i \leq q, \\ \mathcal{O}(1) \ and \ \neq \Omega(1) & i \in [q+1, n-r], \\ \Theta(\alpha) & i \geq n-r+1. \end{cases} \tag{4.5}$$

Proof. The claim $\epsilon^b(X(\alpha)) = \mathcal{O}(\alpha)$ is proved in Lemma 6.2.2 of [28]. The convergence result of Theorem 4.2 implies that the r largest eigenvalues of $X(\alpha)$ and the q largest eigenvalues of $Z(\alpha)$ converge to positive values, hence are $\Theta(1)$. Moreover, the relation $Z(\alpha) = \alpha X(\alpha)^{-1}$ gives that the q smallest eigenvalues of $X(\alpha)$ and the r smallest eigenvalues of $Z(\alpha)$ converge to 0 at a rate that is $\Theta(\alpha)$.

Now let $i \in \{r+1, \dots, n-q\}$. The lower bound $\Omega(\alpha)$ holds since,

$$\lambda_i(X(\alpha)) \geq \lambda_n(X(\alpha)) = \Theta(\alpha).$$

Now suppose, for the sake of contradiction, that $\lambda_i(X(\alpha)) = \mathcal{O}(\alpha)$. Combining with the lower bound, $\Omega(\alpha)$, we conclude that $\lambda_i(X(\alpha)) = \Theta(\alpha)$. Once again using the relation $Z(\alpha) = \alpha X(\alpha)^{-1}$ implies that there are $q+1$ eigenvalues of $Z(\alpha)$ that are bounded away from 0, contradicting the assumption that $\operatorname{rank}(\bar{Z}) = q$ and the statement of Theorem 4.2:

$$\bar{Z} \in \operatorname{relint}\{Z \in \mathbb{S}_+^n \setminus \{0\} : Z = \mathcal{A}^*(y), \ y^T b = 0, \ y \in \mathbb{R}^m \}.$$

The remaining bounds on $Z(\alpha)$ are obtained in similar fashion. $\qquad\square$

Since the time of our original submission, similar convergence bounds have been derived in [17] for a different central path.

4.2 On Singularity Degree and Slow Convergence

An immediate implication of Corollary 4.3 is that fast convergence of eigenvalues does not occur when singularity degree is greater than 1.

Theorem 4.4. *Let $\{(X(\alpha), y(\alpha), Z(\alpha)) : \alpha > 0\}$ be the primal-dual central path of* (4.2) *for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 4.1. If r denotes the maximum rank over $\mathcal{F}$ then,*

$$\lambda_i(X(\alpha)) = \mathcal{O}(\alpha), \ \forall i \in \{r+1, \dots, n\} \iff \operatorname{sd}(\mathcal{F}) = 1.$$

This result identifies two groups of spectrahedra, those with 'good' convergence of eigenvalues and those with 'bad' convergence of eigenvalues. There is an interesting connection here to another notion that has been used throughout the literature to guarantee good convergence: strict complementarity. In [28, Section 4.5] it is proved that strict complementarity holds for an SDP if, and only if, the optimal set of the primal has singularity degree 1. While strict complementarity separates SDPs into two classes, the framework of singularity degree creates a greater stratification of SDPs. As we will see later in this section, singularity degree gives a more comprehensive explanation of poor convergence than strict complementarity. In particular, the main result of this section, Theorem 4.7, states that larger singularity degree leads to greater irregularity in the way that components of $Z(\alpha)$ converge. To simplify the proof we introduce two lemmas. First we bound the projection of $y(\alpha)$ onto $\operatorname{span}(b)$.

Lemma 4.5. *Let $\{(X(\alpha), y(\alpha), Z(\alpha)) : \alpha > 0\}$ be the primal-dual central path of* (4.2) *for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 4.1. Suppose that $b \neq 0$ and let $v^1, \dots, v^{m-1} \in \mathbb{R}^m$*

form a basis for $b^\perp$. Let $\bar{\alpha} > 0$ and for all $\alpha \in (0, \bar{\alpha})$ let $\beta(\alpha)$ and $\nu_1(\alpha), \ldots, \nu_{m-1}(\alpha)$ be real coefficients such that,

$$y(\alpha) = \beta(\alpha)b + \sum_{i=1}^{m-1} \nu_i(\alpha)v^i.$$

Then for all $\alpha \in (0, \bar{\alpha})$,

$$|\beta(\alpha)| = \Theta(\alpha).$$

Proof. By definition of $b(\alpha)$ in (4.1) and by (4.2) we have,

$$\begin{aligned}
y(\alpha)^T b &= y(\alpha)^T (b(\alpha) - \alpha \mathcal{A}(B)) \\
&= y(\alpha)^T (\mathcal{A}(X(\alpha)) - \alpha \mathcal{A}(I)) \\
&= \langle \mathcal{A}^*(y(\alpha)), X(\alpha) \rangle - \alpha \langle B, \mathcal{A}^*(y(\alpha)) \rangle \\
&= \alpha \langle X(\alpha)^{-1}, X(\alpha) \rangle - \alpha \langle B, \mathcal{A}^*(y(\alpha)) \rangle \\
&= \alpha(n - \langle B, \mathcal{A}^*(y(\alpha)) \rangle).
\end{aligned}$$

Now $\langle B, \mathcal{A}^*(y(\alpha)) \rangle$ is bounded below and above on $\alpha \in (0, \bar{\alpha})$, implying that $|y(\alpha)^T b| = \Theta(\alpha)$. On the other hand, $\{v^1, v^2, \ldots, v^{m-1}\} \in b^\perp$ by construction. Therefore $y(\alpha)^T b = \beta(\alpha)\|b\|^2$, yielding,

$$|\beta(\alpha)| = \frac{|y(\alpha)^T b|}{\|b\|^2} = \Theta(\alpha),$$

as desired. $\qquad\qquad\square$

Lemma 4.6. *Let $\{(X(\alpha), y(\alpha), Z(\alpha)) : \alpha > 0\}$ be the primal-dual central path of (4.2) for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 4.1. Let r denote the maximum rank over $\mathcal{F}$ and let $\bar{\alpha} > 0$ be fixed. Let $\hat{Z}(\alpha) \in \mathbb{S}^{\hat{n}}$ be a principal submatrix of $Z(\alpha)$, for some $\hat{n} \leq n$. Then for all $\alpha \in (0, \bar{\alpha})$ it holds that,*

(i) $\|\hat{Z}(\alpha)\| = \Omega(\alpha)$,

(ii) if $\hat{n} \geq r + 1$ then $\|\hat{Z}(\alpha)\| = \Omega(\alpha^{1-\xi(1)})$.

Proof. By interlacing eigenvalues we have,

$$\|\hat{Z}(\alpha)\|_2 = \lambda_1(\hat{Z}(\alpha)) \geq \lambda_{n-r}(Z(\alpha)). \tag{4.6}$$

Now to see that item (i) holds, observe that $\lambda_{n-r}(Z(\alpha)) \geq \lambda_n(Z(\alpha))$ and that $\lambda_n(Z(\alpha)) = \Theta(\alpha)$ by 4.3. To see that item (ii) holds we recall that $Z(\alpha) = \alpha X(\alpha)^{-1}$. Substituting into (4.6) we get,

$$\|\hat{Z}(\alpha)\|_2 \geq \lambda_{n-r}(\alpha X(\alpha)^{-1}) = \alpha \lambda_{r+1}(X(\alpha))^{-1}. \tag{4.7}$$

From 4.3 we have that $\epsilon^b(X(\alpha)) = \mathcal{O}(\alpha)$, hence $\{X(\alpha) : \alpha > 0\}$ satisfies the assumptions of Fact 2.3. It follows by interlacing eigenvalues that $\lambda_{r+1}(X(\alpha)) = \mathcal{O}(\alpha^{\xi(1)})$. Combining this observation with (4.7) yields the desired result. Naturally, these bounds extend to any other norm on $\mathbb{S}^n$. $\qquad\qquad\square$

Now we show that after a suitable orthogonal transformation, $Z(\alpha)$ admits a block partition where at least $\mathrm{sd}(\mathcal{F})$ of these blocks converge to 0, each at a different rate.

Theorem 4.7. *Let $\{(X(\alpha), y(\alpha), Z(\alpha)) : \alpha > 0\}$ be the primal-dual central path of (4.2) for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 4.1. Let $\bar{\alpha} > 0$ be fixed. Then there exists an integer $d \in [\mathrm{sd}(\mathcal{F}), \bar{m}]$ where,*

$$\bar{m} = \begin{cases} m & \text{if } b = 0, \\ m - 1 & \text{otherwise,} \end{cases}$$

and a suitable orthogonal transformation of $\mathcal{F}$, such that,

$$Z(\alpha) = \begin{bmatrix} Z_{d+1}(\alpha) & * & \cdots & * \\ * & Z_d(\alpha) & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & Z_1(\alpha) \end{bmatrix},$$

where for all $\alpha \in (0, \bar{\alpha})$ it holds that,

(i) $Z_1(\alpha) \to S_1 \succ 0$ and $Z_i(\alpha) \to 0$ for all $i \in \{2, \dots, d+1\}$,

(ii) $\frac{\lambda_{\min}(Z_i(\alpha))}{\lambda_{\max}(Z_{i+1}(\alpha))} \to \infty$ for all $i \in \{1, \dots, d\}$,

(iii) $\lambda_{\min}(Z_i(\alpha)) = \Theta(\lambda_{\max}(Z_i(\alpha)))$ for all $i \in \{1, \dots, d+1\}$,

(iv) $\frac{\|Z_i(\alpha)\|}{\|Z_{i+1}(\alpha)\|} \to \infty$ for all $i \in \{1, \dots, d\}$,

(v) $\|Z_{d+1}(\alpha)\| = \Theta(\alpha)$.

Proof. We assume, without loss of generality, that,

$$\mathrm{face}(\mathcal{F}) = \begin{bmatrix} \mathbb{S}_+^r & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.8}$$

As above, let $(\bar{X}, \bar{y}, \bar{Z})$ be the limit point of the primal-dual central path. We know from Theorem 4.2 that $\bar{Z}$ is an exposing vector for a face containing $\mathrm{face}(\mathcal{F})$. Thus $\bar{y}^T b = 0$. Without loss of generality we may assume that,

$$\bar{Z} =: \begin{bmatrix} 0 & 0 \\ 0 & S_1 \end{bmatrix}, \quad S_1 \succ 0. \tag{4.9}$$

Now we define $y^1 := \bar{y}$ and choose $v^1, \dots, v^{m_v} \in \mathrm{span}\{b, y^1\}^\perp$ for some $m_v \leq m$ so that the collection of vectors $\{y^1, v^1, \dots, v^{m_v}\}$ is a basis for $b^\perp$. Note that when $b \neq 0$, then $\{b, y^1, v^1, \dots, v^{m_v}\}$ is also a basis for $\mathbb{R}^m$.

Now for each $\alpha > 0$ there exist coefficients $\beta(\alpha)$ and $\nu_1(\alpha), \dots, \nu_{m_v}(\alpha)$ and $\gamma_1(\alpha)$ such that,

$$y(\alpha) = \gamma_1(\alpha) y^1 + \beta(\alpha) b + \sum_{i=1}^{m_v} \nu_i(\alpha) v^i. \tag{4.10}$$

Since $y(\alpha) \to y^1$ we have $\gamma_1(\alpha) \to 1$ and $\gamma_1(\alpha)$ dominates the other coefficients. That is,

$$\lim_{\alpha \searrow 0} \frac{\sum_{i=1}^{m_v} |\nu_i(\alpha)|}{\gamma_1(\alpha)} = 0. \tag{4.11}$$

Now let us consider a block partition of $Z(\alpha)$ according to the block partition of $\bar{Z}$ in (4.9). We have,

$$Z(\alpha) = \begin{bmatrix} 0 & 0 \\ 0 & \gamma_1(\alpha) S_1 \end{bmatrix} + \mathcal{A}^* \left(\beta(\alpha) b + \sum_{i=1}^{m_v} \nu_i(\alpha) v^i \right). \tag{4.12}$$

16

Note that the two diagonal blocks of $Z(\alpha)$ in (4.12) possess the properties of item (i) and item (ii).

Let $Z_{11}(\alpha)$ denote the upper left block of $Z(\alpha)$. We consider two possibilities. First, suppose that $\|Z_{11}(\alpha)\| = \mathcal{O}(\alpha)$. In this case the lower bound of Lemma 4.6 (i) implies that $\|Z_{11}(\alpha)\| = \Theta(\alpha)$. Consequently the two blocks of $Z(\alpha)$ also satisfy the properties of items (iii)-(v) with $d = 1$. It is immediate that $d \leq \bar{m}$. Next, Lemma 4.6 (ii) implies that $Z_{11}(\alpha)$ has at most r rows. Moreover, by our assumption on the facial structure of face$(\mathcal{F})$ in (4.8) we conclude that $Z_{11}(\alpha)$ has exactly r rows, otherwise $\bar{Z}$ exposes a face that is strictly smaller than face$(\mathcal{F})$. Thus we have $\text{sd}(\mathcal{F}) = 1 \leq d$, as desired.

The second possibility is that $\|Z_{11}(\alpha)\| \neq \mathcal{O}(\alpha)$. In this case, at least one of the coefficients other than $\gamma_1(\alpha)$ converges to 0 at a rate not equal to $\mathcal{O}(\alpha)$. This coefficient is not $\beta(\alpha)$, since Lemma 4.5 implies that $\beta(\alpha) = \Theta(\alpha)$ when $b \neq 0$. When $b = 0$ we may set $\beta(\alpha) = 0$ as it is irrelevant. Thus we conclude that $|\nu_i(\alpha)| \neq \mathcal{O}(\alpha)$ for some $i \in \{1, \dots, m_v\}$.

Now let, y^2 be a limit point of $\beta(\alpha)b + \sum_{i=1}^{m_v} \nu_i(\alpha)v^i$, after normalizing. By the arguments above, $y^2 \in \text{span}\{v^1, \dots, v^{m_v}\}$ and thus $(y^2)^T b = 0$. Secondly, $(\mathcal{A}^*(\beta(\alpha)b + \sum_{i=1}^{m_v} \nu_i(\alpha)v^i))_{11}$ is positive definite for every $\alpha > 0$ by (4.12). This implies that $(\mathcal{A}^*(y^2))_{11} \succeq 0$. Thus if $(\mathcal{A}^*(y^2))_{11}$ is not the zero matrix it is an exposing vector in the second step of facial reduction.

Let us first address the case $(\mathcal{A}^*(y^2))_{11} = 0$. Here we let $w^1 := y^2$ and choose

$$v^1, \dots, v^{m_v} \in \text{span}\{b, y^1, w^1\}^\perp$$

for some m_v, different than the previously used m_v, so that $\{y^1, w^1, v^1, \dots, v^{m_v}\}$ is a basis for $b^\perp$. Now we repeat the above process until we obtain a new y^2 such that $(\mathcal{A}^*(y^2))_{11} \neq 0$, i.e., we are in the second case.

Now we may assume that we have obtained y^2 as above and $(\mathcal{A}^*(y^2))_{11} \neq 0$. We also assume, without loss of generality that,

$$(\mathcal{A}^*(y^2))_{11} = \begin{bmatrix} 0 & 0 \\ 0 & S_2 \end{bmatrix}, \ S_2 \succ 0. \tag{4.13}$$

Then the matrix,

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & S_2 & 0 \\ 0 & 0 & S_1 \end{bmatrix}, \tag{4.14}$$

exposes a face containing face$(\mathcal{F})$ and this face is smaller than the one exposed by $\bar{Z}$. In other, words, we have obtained a better exposing vector. Let us assume that we have accumulated m_w vectors of the type w^1 obtained in the case $(\mathcal{A}^*(y^2))_{11} = 0$. Then we choose,

$$v^1, \dots, v^{m_v} \in \text{span}\{b, y^1, y^2, w^1, \dots, w^{m_w}\}^\perp, \tag{4.15}$$

so that $\{y^1, y^2, w^1, \dots, w^{m_w}, v^1, \dots, v^{m_v}\}$ is a basis for $b^\perp$. As above, there exist coefficients,

$$\beta(\alpha), \ \gamma_1(\alpha), \gamma_2(\alpha), \ \omega_1(\alpha), \dots, \omega_{m_w}(\alpha), \ \nu_1(\alpha), \dots, \nu_{m_v}(\alpha), \tag{4.16}$$

such that,

$$y(\alpha) = \beta(\alpha)b + \sum_{i=1}^{2} \gamma_i(\alpha)y^i + \sum_{i=1}^{m_w} \omega_i(\alpha)w^i + \sum_{i=1}^{m_v} \nu_i(\alpha)v^i.$$

Then,

$$Z(\alpha) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \gamma_2(\alpha)S_2 & 0 \\ 0 & 0 & \gamma_1(\alpha)S_1 \end{bmatrix} + \mathcal{A}^*\left(\beta(\alpha)b + \sum_{i=1}^{m_v} \nu_i(\alpha)v^i + \sum_{i=1}^{m_w} \omega_i(\alpha)w^i\right), \tag{4.17}$$

where the upper left block is $\mathcal{A}_{11}^* \left(\beta(\alpha)b + \sum_{i=1}^{m_v} \nu_i(\alpha)v^i \right)$. By construction we have

$$\frac{\gamma_1(\alpha)}{\gamma_2(\alpha)} \to \infty \text{ and } \frac{\gamma_2(\alpha)}{b(\alpha) + \sum_{i=1}^{m_v}|\nu_i(\alpha)|} \to \infty. \tag{4.18}$$

Thus we conclude that the diagonal blocks of $Z(\alpha)$ satisfy the properties of item (i) and item (ii). In addition, the blocks containing $\gamma_1(S_1)$ and $\gamma_2(S_2)$ satisfy the properties of item (iii) and item (iv).

By Lemma 4.6 we may continue in this fashion until,

$$Z(\alpha) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \gamma_d(\alpha)S_d & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_1(\alpha)S_1 \end{bmatrix} + \mathcal{A}^* \left(\beta(\alpha)b + \sum_{i=1}^{m_v} \nu_i(\alpha)v^i + \sum_{i=1}^{m_w} \omega_i(\alpha)w^i \right), \tag{4.19}$$

for some positive integer d and the upper left block has norm that is $\mathcal{O}(\alpha)$. By reasoning analogous to that of the discussion following (4.12), we conclude that the blocks of $Z(\alpha)$, according to the block partition of (4.19), satisfy the properties of item (i)-item (v) and that $d \in [\text{sd}(\mathcal{F}), \bar{m}]$, as desired. $\qquad \square$

The issue with different rates of convergence among the blocks of $Z(\alpha)$ that vanish, is one of a practical nature. When singularity degree is high, it may be the case that $\|Z_{d+1}(\alpha)\| << \|Z_2(\alpha)\|$. By combining Corollary 4.3 and Theorem 4.7 we get the relation,

$$\epsilon^b(X(\alpha)) = \mathcal{O}(\|Z_{d+1}(\alpha)\|) << \|Z_2(\alpha)\| = \mathcal{O}(\|Z(\alpha) - \bar{Z}\|),$$

for spectrahedra with high singularity degree. Now suppose a path-following algorithm is applied and accurately follows the primal-dual central path. To obtain a good approximation of the dual limit point $\bar{Z}$ we may need a very small backward error. Due to numerical issues, such as ill-conditioning of the Jacobian of (4.2), a very small backward error is generally not attainable. For this reason, we cannot expect $Z(\alpha)$, and by extension $X(\alpha) = \alpha Z(\alpha)^{-1}$, to be accurate when singularity degree is high.

The integer d of Theorem 4.7 actually provides an upper bound on $\text{sd}(\mathcal{F})$ that complements the lower bound of Theorem 3.12. However, this upper bound is generally intractable due to its reliance on an unknown orthogonal transformation. If the statement of the theorem can be translated to a statement about convergence rates of blocks of eigenvalues, then we would have a tractable upper bound on $\text{sd}(\mathcal{F})$. However this may not be true as illustrated by the parametric sequence,

$$S(\alpha) := \begin{bmatrix} 3 & \alpha^{1/2} & 0 \\ \alpha^{1/2} & \frac{\alpha}{3-\alpha^2} & 0 \\ 0 & 0 & \alpha^3 \end{bmatrix}, \ \alpha > 0. \tag{4.20}$$

Here $S(\alpha)$ has different rates of convergence among the diagonal, but the two eigenvalues that vanish, do so at the same rate, $\Theta(\alpha^3)$. One way to guarantee that the diagonal blocks correspond to blocks of eigenvalues is the following.

Corollary 4.8. *Let $\{(X(\alpha), y(\alpha), Z(\alpha)) : \alpha > 0\}$ be the primal-dual central path of (4.2) for a spectrahedron $\mathcal{F} = \mathcal{F}(\mathcal{A}, b)$ satisfying Assumption 4.1. Let $\bar{\alpha} > 0$ be fixed and let d be as in Theorem 4.7. For every $\alpha \in (0, \bar{\alpha})$ we assume that $Z(\alpha)$ has the block structure of Theorem 4.7. If every principal submatrix of $Z(\alpha)$ of the form,*

$$S_i(\alpha) = \begin{bmatrix} Z_i(\alpha) & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & Z_1(\alpha) \end{bmatrix}, \ i \in \{2, \ldots, d\},$$

satisfies $\lambda_{\min}(S_i(\alpha)) = \Theta(\lambda_{\min}(Z_i(\alpha)))$, *then there are exactly d different rates of convergence among the eigenvalues of $X(\alpha)$ that vanish.*

Proof. Applying the interlacing eigenvalue theorem with the principal submatrices,

$$\begin{bmatrix} Z_{d+1}(\alpha) & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & Z_i(\alpha) \end{bmatrix} \text{ and } \begin{bmatrix} Z_i(\alpha) & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & Z_1(\alpha) \end{bmatrix},$$

yields the upper bound of $\lambda_{\max}(Z_i(\alpha))$ and the lower bound $\lambda_{\min}(Z_i(\alpha))$ on a block of $\lambda(Z(\alpha))$ having the same size as the number of rows $Z_i(\alpha)$. Since $\lambda_{\min}(Z_i(\alpha)) = \Theta(\lambda_{\max}(Z_i(\alpha)))$, these bounds are the same. Thus we conclude that for each $i \in \{2, \ldots, d+1\}$ there is a block of eigenvalues that converges to 0 at the same rate as $Z_i(\alpha)$ does. The desired result follows from the relation $X(\alpha) = \alpha Z(\alpha)^{-1}$. $\qquad\square$

The challenge with this result is that the hypothesis is unverifiable just as the block structure of $Z(\alpha)$ is unobservable without knowledge of the appropriate orthogonal transformation. However, our numerical observations indicate that the conclusion of the corollary holds for those test cases for which it can be verified.

5 Numerical Case Studies

In this section we apply the main results above to obtain bounds on forward error, singularity degree, and maximum rank for several problems from the literature. Our analysis is focused on problems with relatively large singularity degree; although we do study one instance with singularity degree 1, in order to demonstrate *good* convergence. For some of the instances, the exact singularity degree is known, allowing us to test the quality of our bounds. In other instances, the singularity degree is not known and we use our bounds to provide an estimate of it.

In order to study the notion that large singularity degree is sufficient, in some sense, for slow convergence, we consider bounded spectrahedra satisfying Assumption 4.1. We follow the primal-dual central path of (4.2) with a path-following algorithm based on the Gauss-Newton search direction, see [9, 15]. Like most interior point algorithms, our implementation generates a sequence of positive semidefinite matrices $\{\tilde{X}(\sigma^k)\}$ that approximates the true sequence $\{X(\sigma^k)\}$, contained on the central path. In our case $\sigma = 0.6$. These sequences are inherently well-suited to the results of this paper, in particular, those of Section 3.

For each of our test problems, we present two plots. The first is an approximation of the Q-convergence ratio of Theorem 3.7. To make the subsequent discussion less cumbersome, we introduce the notation,

$$R_Q(i, k) := \frac{\lambda_i(\tilde{X}(\sigma^{k+1}))}{\lambda_i(\tilde{X}(\sigma^k))}, \ i \in \{1, \ldots, n\}, \ k \geq 1. \tag{5.1}$$

We use this first ratio to bound maximum rank (Corollary 3.9), forward error (Theorem 3.11), and singularity degree (Theorem 3.12).

The second plot is an approximation of the ratio of adjacent eigenvalues from (3.3),

$$R_N(i, k) := \frac{\lambda_i(\tilde{X}(\sigma^k))}{\lambda_{i+1}(\tilde{X}(\sigma^k))}, \ i \in \{1, \ldots, n-1\}, \ k \geq 1. \tag{5.2}$$

19

This ratio is used to upper bound the maximum rank as in Proposition 3.4 and to determine the number of different rates of convergence among eigenvalues that vanish.

It is important to note that the only modification we made to our interior point implementation is a few lines of code to store the ratios $R_Q(i,k)$ and $R_N(i,k)$. In principle, just about any interior point implementation can be modified to track these ratios. Once the ratios have been obtained for k sufficiently large, around 60, we generate plots of the ratios against k for each i and obtain bounds on maximum rank, forward error, and singularity degree. The bounds as well as the true values (when available) are recorded in Table 5.1.
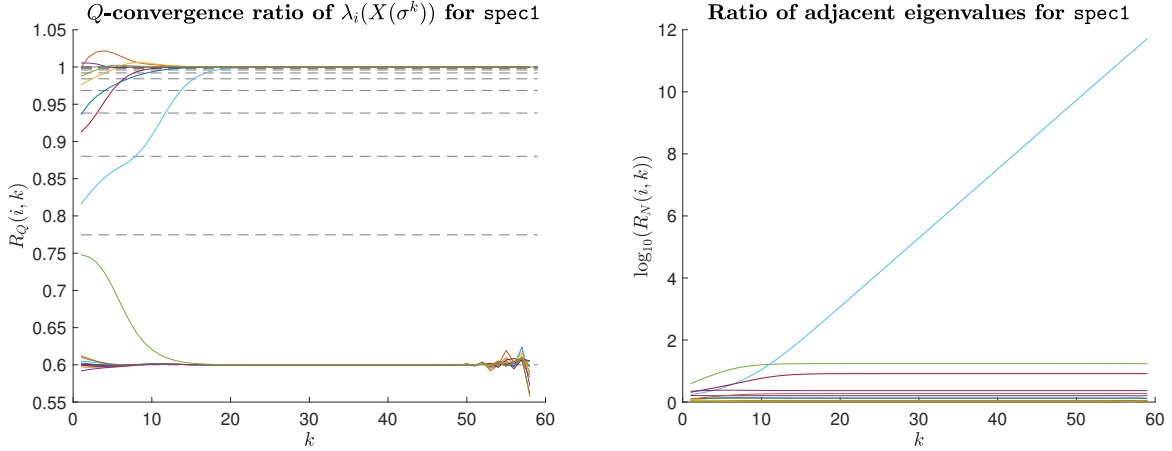

Figure 5.1: Based on these images we obtain the bounds $\bar{r} = 7$ and $\underline{d = 1}$ for `spec1`. The dashed lines coincide with the values $\sigma^{2^{-(d-1)}}$ for $d \in \{1, \ldots, n-1\}$.

We denote the first spectrahedron as `spec1` and it is taken from the class of SDPs introduced in [33]. There, the authors present a method to generate SDPs with specified *complementarity gap*. By considering the optimal set of one such SDP with strict complementarity, we obtain a spectrahedron with singularity degree 1. For details see [28, Section 4.5]. The problem we consider has size $n = 20$ and plots of the ratios $R_Q(i,k)$ and $R_N(i,k)$ are shown in Figure 5.1. In the left image, there is a clear distinction between curves that converge to 1 and curves that do not converge to 1. Moreover, if we disregard the irregularity in the last few values of the curves that do not converge to 1, we may conclude that those curves converge to the smallest dashed line located at 0.6. This observation, together with Theorem 4.4, correctly indicates that the spectrahedron has singularity degree 1. Exactly 13 of the curves converge to 0.6, yielding $\bar{r} = 7$, the correct approximation of the maximal rank r. The plot on the right side of the figure shows that exactly one curve tends to ∞ and it is the curve corresponding to $i = 7 = r$. This indicates, as expected, two groups of eigenvalues of X: those that converge to positive values and those that vanish.

The second spectrahedron, denoted `spec2`, is the classical 'worst case' problem as presented in [30, Section 2.6], with $n = 5$ and singularity degree $n - 1 = 4$. Plots of the two ratios are in Figure 5.2. The left image shows 5 distinct rates of Q-convergence, one for each eigenvalue. All but one of the curves converge to values that are clearly below 1. This indicates, correctly, that the maximum rank is at most 1. The largest of the curves appears to converge to the highest of the dashed lines. Thus we may infer that singularity degree is at least 4. Since $4 = n - 1$, the worst case upper bound, we may conclude that the singularity degree is exactly 4. The row corresponding to `spec2` in Table 5.1 shows a very large discrepancy between forward error and backward error. Our lower bound is actually quite close to the true forward error. Now let us consider the image on the right. It may be somewhat speculative to assert that the two
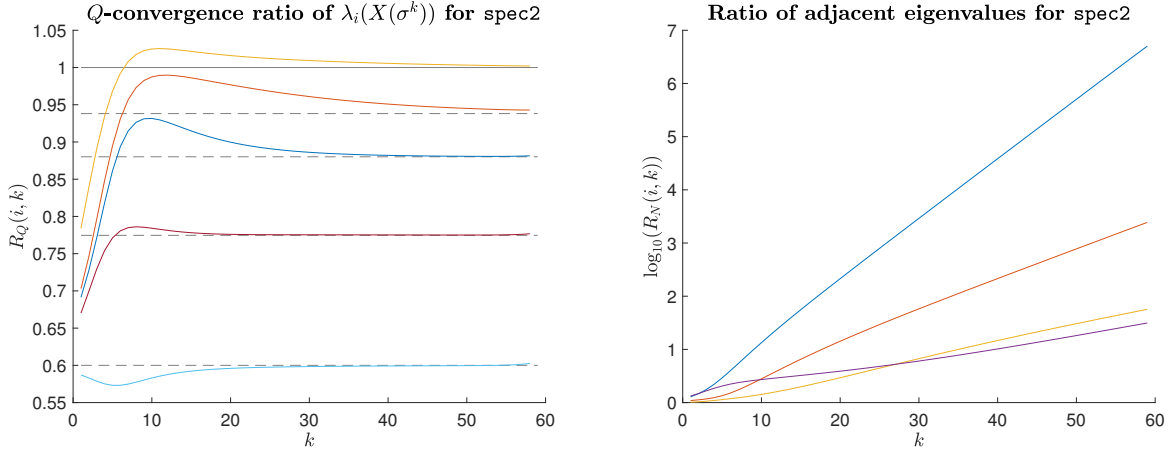
Figure 5.2: Based on these images we obtain the bounds $\bar{r} = 1$ and $\underline{d = 4}$ for `spec2`. The dashed lines coincide with the values $\sigma^{2^{-(d-1)}}$ for $d \in \{1, \ldots, n-1\}$.

lower curves tend to ∞. Thus, taking the more cautious approach we assume that only the two larger curves tend to ∞. Checking the indices of these curves yields an upper bound of 3 on the maximum rank. We choose the notably lower estimate of 1 based on the left plot. On the other hand if we are to apply Corollary 4.8 then we would want an overestimate of the number of different rates of convergence among the eigenvalues of $X(\alpha)$ that vanish. For this number we include the two lower curves, giving a bound of 4.
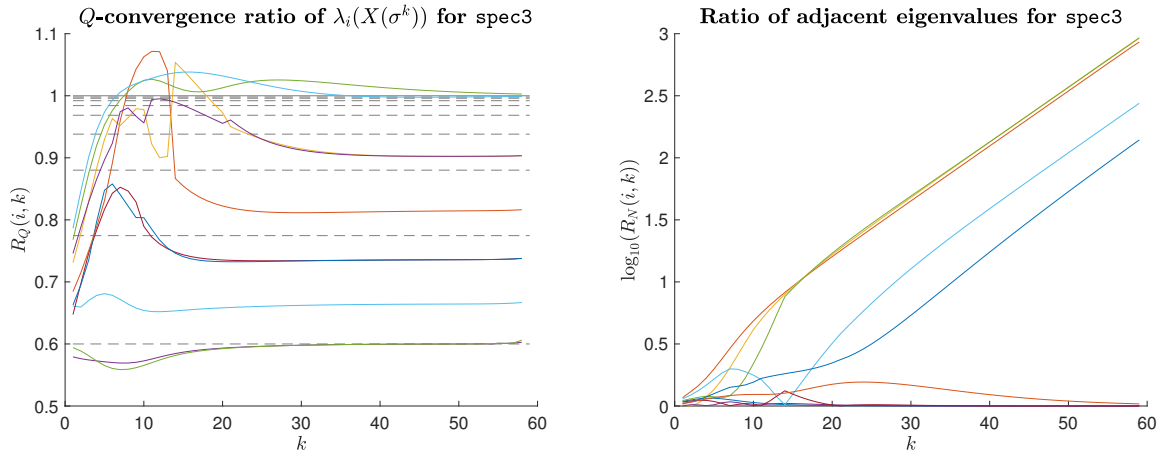

Figure 5.3: Based on these images we obtain the bounds $\bar{r} = 2$ and $\underline{d = 4}$ for `spec3`. The dashed lines coincide with the values $\sigma^{2^{-(d-1)}}$ for $d \in \{1, \ldots, n-1\}$.

The next spectrahedron, `spec3`, comes from an application in polynomial optimization. The authors of [32] observed *strange behaviour* when attempting to optimize over it with an interior point method. The dimension is $n = 10$ and the singularity degree is proven to be 5. In Table 5.1 we see a large discrepancy between forward error and backward error. The left image of Figure 5.3 shows six distinct groups of curves. It is clear, for all but two of the curves, that the limit point is different from 1. Thus we have an upper bound of 2 on the maximum rank. This upper bound gives quite an accurate lower bound on the forward error. The largest of the curves that does not converge to 1 appears to converge to a value that is below the fourth dashed line, indicating a lower bound of 4 on the singularity degree. Unlike the two previous spectrahedra, here the lower

bound on singularity degree is a strict one. The image on the right shows exactly five different rates of convergence among the eigenvalues of $X(\alpha)$ that converge to 0. Moreover the upper bound on maximum rank corresponds to the one obtained from the left image.
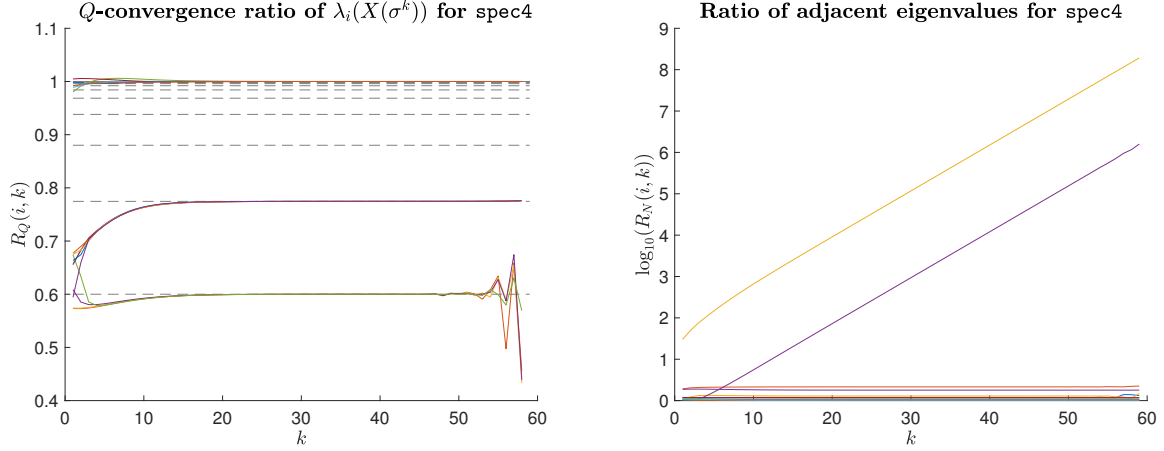

Figure 5.4: Based on these images we obtain the bounds $\bar{r} = 5$ and $\underline{d = 2}$ for `spec4`. The dashed lines coincide with the values $\sigma^{2^{-(d-1)}}$ for $d \in \{1, \ldots, n-1\}$.

The fourth spectrahedron, `spec4`, is generated by the algorithm of [33], just as `spec1` is. However, for this instance we require the existence of a complementarity gap, i.e., strict complementarity does not hold. While we do not know the exact value of the singularity degree, the lack of strict complementarity gives us a lower bound of 2. Plots of the ratios $R_Q(i, k)$ and $R_N(i, k)$ are shown in Figure 5.4. From the left image we can be quite sure that those curves that converge to the second dashed line or below do not converge to 1. A closer inspection reveals that there are 10 such curves, implying an upper bound of 5 on the maximum rank. The corresponding lower bound on forward error, as recorded in Table 5.1, is indeed a lower bound and more informative than the reported backward error. We also obtain a lower bound on the singularity degree that coincides with the theoretical lower bound of 2. The image on right shows exactly two rates of convergence among eigenvalues of $X(\alpha)$ that converge to 0 and, once again, provides the same upper bound on maximum rank as obtained from the image on left.
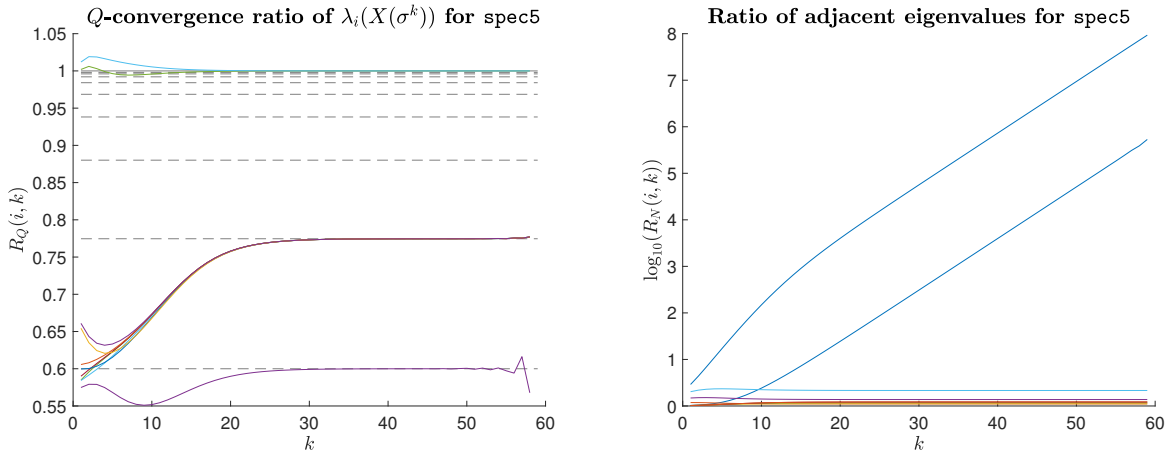

Figure 5.5: Based on these images we obtain the bounds $\bar{r} = 2$ and $\underline{d = 2}$ for `spec5`. The dashed lines coincide with the values $\sigma^{2^{-(d-1)}}$ for $d \in \{1, \ldots, n-1\}$.

The final spectrahedron we consider is taken from an application to *Toeplitz cycle completion problems* of the form introduced in [2, Corollary 6]. The specific instance we consider here is that of [27, Example 4.5] with $n = 10$. While the singularity degree for this problem is not known, a lower bound of 2 was proven in [27]. In Figure 5.5, we find images of plots of the ratios $R_Q(i,k)$ and $R_N(i,k)$. The left image indicates that all but two of the eigenvalues of $X(\alpha)$ converge to 0, yielding an exact approximation of maximum rank. Moreover, the corresponding eight curves appear to have limits below the second dashed line. Hence we have a lower bound of 2 on the singularity degree. The image on the right clearly indicates two curves tending to ∞, confirming the lower bound of 2 on singularity degree.[2]

$\mathcal{F}$	$\epsilon^b(\mathcal{F})$	r	$\bar{r}$	$\epsilon^f(\mathcal{F})$	$\underline{\epsilon}$	$\text{sd}(\mathcal{F})$	$\underline{d}$	N_λ
spec1	6.62×10^{-11}	7	7	4.36×10^{-11}	3.10×10^{-12}	1	1	1
spec2	4.44×10^{-13}	1	1	4.93×10^{-2}	3.19×10^{-2}	4	4	4
spec3	2.47×10^{-13}	2	2	3.39×10^{-2}	9.88×10^{-3}	5	4	5
spec4	1.88×10^{-11}	5	5	1.35×10^{-5}	4.16×10^{-7}	≥ 2	2	2
spec5	2.61×10^{-13}	2	2	1.22×10^{-6}	6.96×10^{-8}	≥ 2	2	2

Table 5.1: A record of relevant measures and their bounds for the spectrahedra considered in our analysis. Here $\bar{r}$, $\underline{\epsilon}$, and $\underline{d}$ denote the upper bound on rank, lower bound on forward error, and lower bound on singularity degree, respectively. The number of different rates of convergence among the eigenvalues of $X(\alpha)$ that vanish is denoted by N_λ.

In these case studies we have demonstrated the ability to upper bound maximum rank quite effectively. The resulting lower bound on forward error is of a much larger magnitude than backward error in all instances with the exception of spec1, where the singularity degree is 1. We see this feature as quite useful, as it alerts practitioners that the proposed solution is of substantially lower accuracy than backward error indicates. For spectrahedra with known singularity degree, we have demonstrated that the lower bound is quite accurate. In the other cases, the lower bound is in agreement with the theoretical lower bound. Lastly, for these test cases (as well as for others we have tested), the value N_λ seems to be an upper bound on singularity degree. Proving this, or demonstrating a counterexample, is an interesting topic for future research.

6 Conclusion

In this paper we studied the lower bounding of *forward error* for semidefinite programming and its relationship to the singularity degree. We showed that a large singularity degree is an indicator for slow convergence of the class of algorithms we looked at. Therefore, it provides an indication for practitioners that the obtained solution is of substantially lower accuracy than a measure of backward error indicates.

Currently we have not come across high singularity degree in practice in the literature, except in a theoretical context. We think that this will prove more and more important in solving **SDP** problems and cone optimization problems in the future. Note that though full rank for matrices is generic, singular and ill-conditioned linear systems of equations are clearly important and not ignored. And, in fact, low rank problems have become greatly desired for e.g., *data recovery problems*. Similarly, Rademacher's theorem guarantees that locally Lipschitz functions are differentiable almost everywhere. Yet, nonsmooth problems arise in optimization and are currently a very important area of study. Also, application of Saard's theorem implies that

---

[2]The link www.math.uwaterloo.ca/~hwolkowi/henry/reports/sdTestsStefan/ has the codes and examples.

optimality conditions are generic in a sense, e.g., [26]. However, in practice many problems suffer from degeneracies such as the ones discussed in our paper.

That singularity degree is at most one is in a sense *generic* is shown in [10]. A discussion for matrix completion where high singularity degree appears is given in [27]. Like the examples of rank, differentiability, optimality conditions, we think that high singularity degree will be an important issue in the future and one we are working on.

References

[1] F. Alizadeh, J-P.A. Haeberly, and M.L. Overton. Complementarity and nondegeneracy in semidefinite programming. *Math. Programming*, 77:111–128, 1997. 7

[2] W. Barrett, C.R. Johnson, and P. Tarazaga. The real positive definite completion problem for a simple cycle. *Linear Algebra Appl.*, 192:3–31, 1993. Computational linear algebra in algebraic and related problems (Essen, 1992). 23

[3] J.M. Borwein and H. Wolkowicz. Characterization of optimality for the abstract convex program with finite-dimensional range. *J. Austral. Math. Soc. Ser. A*, 30(4):390–411, 1980/81. 2, 4

[4] J.M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 30(3):369–380, 1980/81. 2, 4

[5] J.M. Borwein and H. Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. Appl.*, 83(2):495–530, 1981. 2, 4

[6] Y-L. Cheung, S. Schurr, and H. Wolkowicz. Preprocessing and regularization for degenerate semidefinite programs. In D.H. Bailey, H.H. Bauschke, P. Borwein, F. Garvan, M. Thera, J. Vanderwerff, and H. Wolkowicz, editors, *Computational and Analytical Mathematics, In Honor of Jonathan Borwein's 60th Birthday*, volume 50 of *Springer Proceedings in Mathematics & Statistics*, pages 225–276. Springer, 2013. 2

[7] E. de Klerk, C. Roos, and T. Terlaky. Initialization in semidefinite programming via a self-dual skew-symmetric embedding. *Oper. Res. Lett.*, 20(5):213–221, 1997. 12

[8] 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*, volume 18 of *The Fields Institute for Research in Mathematical Sciences, Communications Series*, pages 215–236. American Mathematical Society, 1998. 12

[9] X.V. Doan, S. Kruk, and H. Wolkowicz. A robust algorithm for semidefinite programming. *Optim. Methods Softw.*, 27(4-5):667–693, 2012. 7, 19

[10] D. Drusvyatskiy, G. Li, and H. Wolkowicz. A note on alternating projections for ill-posed semidefinite feasibility problems. *Math. Program.*, 162(1-2, Ser. A):537–548, 2017. 7, 24

[11] D. Drusvyatskiy and H. Wolkowicz. The many faces of degeneracy in conic optimization. *Foundations and Trends® in Optimization*, 3(2):77–170, 2017. 3, 4

[12] K. Fan. On a theorem of Weyl concerning eigenvalues of linear transformations ii. *Proc. Nat. Acad. Sci. U.S.A.*, 36:31–35, 1950. 11

[13] N.I.M. Gould, Daniel P. Robinson, and H. Sue Thorne. On solving trust-region and other regularised subproblems in optimization. *Math. Program. Comput.*, 2(1):21–57, 2010. 7

[14] A.J. Hoffman and H.W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematics*, 20:37–39, 1953. 11

[15] S. Kruk, M. Muramatsu, F. Rendl, R.J. Vanderbei, and H. Wolkowicz. The Gauss-Newton direction in semidefinite programming. *Optim. Methods Softw.*, 15(1):1–28, 2001. 19

[16] Z-Q. Luo, J.F. Sturm, and S. Zhang. Conic convex programming and self-dual embedding. *Optim. Methods Softw.*, 14(3):169–218, 2000. 12

[17] A. Mohammad-Nezhad and T. Terlaky. On the identification of the optimal partition for semidefinite optimisation. *INFOR: Information Systems and Operational Research*, pages 1–39, 2019. 14

[18] R.D.C. Monteiro and M.J. Todd. Path-following methods. In *Handbook of Semidefinite Programming*, pages 267–306. Kluwer Acad. Publ., Boston, MA, 2000. 7

[19] Y. Nesterov and B.T. Polyak. Cubic regularization of Newton method and its global performance. *Math. Program.*, 108(1, Ser. A):177–205, 2006. 7

[20] Y.E. Nesterov, M.J. Todd, and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Math. Program.*, 84(2, Ser. A):227–267, 1999. 12

[21] G. Pataki. Strong duality in conic linear programming: facial reduction and extended duals. In David Bailey, Heinz H. Bauschke, Frank Garvan, Michel Thera, Jon D. Vanderwerff, and Henry Wolkowicz, editors, *Computational and analytical mathematics*, volume 50 of *Springer Proc. Math. Stat.*, pages 613–634. Springer, New York, 2013. 3

[22] 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(3):1257–1282, 2017. 3

[23] F. Permenter and P. Parrilo. Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone. Technical Report Preprint arXiv:1408.4685, MIT, Boston, MA, 2014. 3

[24] F.A. Potra and R. Sheng. A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming. Technical Report Reports on Computational Mathematics, 78, University of Iowa, Iowa City, IA, 1995. 12

[25] R.T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970. 3

[26] J.E. Spingarn and R.T. Rockafellar. The generic nature of optimality conditions in nonlinear programming. *Math. Oper. Res.*, 4(4):425–430, 1979. 24

[27] S. Sremac, H.J. Woerdeman, and H. Wolkowicz. Maximum determinant positive definite Toeplitz completions. In *Operator Theory, Analysis and the State Space Approach: In Honor of Rien Kaashoek*, volume 271, pages 421–441. Birkhäuser/Springer, Cham, 2018. 23, 24

[28] Sremac, S. Error bounds and singularity degree in semidefinite programming, 2020. 2, 4, 13, 14, 20

[29] J.F. Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10(4):1228–1248 (electronic), 2000. 1, 2, 5, 11

[30] L. Tunçel. *Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization*, volume 27 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2010. 2, 3, 20

[31] H. Waki and M. Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013. 3

[32] H. Waki, M. Nakata, and M. Muramatsu. Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization. *Computational Optimization and Applications*, 53(3):823–844, 2012. 21

[33] H. Wei and H. Wolkowicz. Generating and measuring instances of hard semidefinite programs. *Math. Program.*, 125(1, Ser. A):31–45, 2010. 7, 20, 22

[34] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of semidefinite programming*. International Series in Operations Research & Management Science, 27. Kluwer Academic Publishers, Boston, MA, 2000. Theory, algorithms, and applications. 3