Bad semidefinite programs with short proofs, and the closedness of the linear image of the semidefinite cone

Gábor Pataki*

June 11, 2018

Abstract

Semidefinite programs (SDPs) – some of the most useful and pervasive optimization problems of the last few decades – often behave pathologically: the optimal values of the primal and dual problems may differ and may not be attained. Such SDPs are theoretically interesting and often difficult or impossible to solve. Yet, the pathological SDPs in the literature look strikingly similar, and to explain why, our recent paper [37] characterized pathological semidefinite systems by certain *excluded matrices*, which are easy to spot in all published examples.

Here we give short and elementary proofs of the results of [37] using techniques mostly from elementary linear algebra. Our main tool is a standard (canonical) form of semidefinite systems, from which their pathological behavior is easy to verify. The standard form is constructed in a simple manner, using mostly elementary row operations inherited from Gaussian elimination.

As a byproduct, we show how to transform any linear map acting on symmetric matrices into a standard form; the standard form allows us to quickly check whether the image of the semidefinite cone under the map is closed. We thus provide a stepping stone to understand a fundamental issue in convex analysis: the linear image of a closed convex set may not be closed, and often simple conditions are available to verify the closedness, or lack of it.

Key words: semidefinite programming; duality; duality gap; pathological semidefinite programs; closedness of the linear image of the semidefinite cone

MSC 2010 subject classification: Primary: 90C46, 49N15; secondary: 52A40

OR/MS subject classification: Primary: convexity; secondary: programming-nonlinear-theory

1 Introduction. Main results

Semidefinite programs (SDPs) – optimization problems with semidefinite matrix variables, a linear objective, and linear constraints – are some of the most practical, widespread, and interesting optimization problems of the last three decades. They naturally generalize linear programs, and find applications in diverse areas such as combinatorial optimization, engineering, and economics. They are covered in many surveys, see e.g. [44] and textbooks, see e.g. [11, 3, 41, 10, 15, 5, 25, 45].

They are also a subject of intensive research: in the last 30 years several thousand papers have been published on SDPs.

*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill

To ground our discussion, let us write an SDP in the form

$$\begin{array}{ll} \sup & \sum_{i=1}^m c_i x_i \\ s.t. & \sum_{i=1}^m x_i A_i \preceq B \end{array} \tag{SDP_c}$$

where $A_1, \ldots, A_m,$ and B are $n \times n$ symmetric matrices, $c_1, \ldots, c_m$ are scalars, and for symmetric matrices S and T, we write $S \preceq T$ to say that $T - S$ is positive semidefinite (psd).

To solve (SDP_c) we rely on a natural dual, namely

$$\begin{array}{ll} \inf & B \bullet Y \\ s.t. & A_i \bullet Y = c_i \\ & Y \succeq 0, \end{array} \tag{SDD_c}$$

where the inner product of symmetric matrices S and T is $S \bullet T := \text{trace}(ST)$. Since the weak duality inequality

$$\sum_{i=1}^m c_i x_i \leq B \bullet Y \tag{1.1}$$

always holds between a pair of feasible solutions, if we find a pair (x^*, Y^*) that satisfies (1.1) with equality, we know that they are both optimal. Indeed, SDP solvers seek to find such an x^* and Y^*.

However, SDPs often behave pathologically: the optimal values of (SDP_c) and (SDD_c) may differ and may not be attained.

The duality theory of SDPs – together with their pathological behaviors – is covered in several references on optimization theory and in textbooks written for broader audiences. For example, [11] gives an extensive, yet concise account of Fenchel duality; [44] and [41] provide very succinct treatments; [3] treats SDP duality as special case of duality theory in infinite dimensional spaces; [10] covers stability and sensitivity analysis; [5] and [15] contain many engineering applications; and [25] and [45] are accessible to an audience with combinatorics background.

Why are the pathological behaviors interesting? First, they do not appear in linear programs, which makes it apparent that SDPs are a much less innocent generalization of linear programs, than one may think at first. [1] Note that the pathologies can come in "batches": in extreme cases (SDP_c) and (SDD_c) *both* can have unattained, and different, optimal values! The variety of thought-provoking pathological SDPs makes the teaching of SDP duality to undergraduate and graduate students (who are mostly used to clean and pathology-free linear programming) a quite a rewarding experience.

Second, these pathologies also appear in other convex optimization problems, thus SDPs make excellent "model problems" to study.

Last but not least: pathological SDPs are often difficult or impossible to solve.

Our recent paper [37] was motivated by the curious similarity of pathological SDPs in the literature. To build intuition, we recall two examples; they or their variants appear in a number of papers and surveys.

Example 1. *In the problem*

$$\begin{array}{ll} \sup & 2x_1 \\ s.t. & x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \end{array} \tag{1.2}$$

[1] Suppose the A_i and B are diagonal: then (SDP_c) is a linear program, since a diagonal matrix is psd if and only if it is nonnegative. In this case we can also restrict Y in (SDD_c) to be diagonal.

Of course, in linear programming the only possible pathology is when both the primal and the dual are infeasible.

the only feasible solution is $x_1 = 0$. The dual, with a variable matrix $Y = (y_{ij})$, is equivalent to

$$\begin{array}{ll} \inf & y_{11} \\ s.t. & \begin{pmatrix} y_{11} & 1 \\ 1 & y_{22} \end{pmatrix} \succeq 0 \end{array} \tag{1.3}$$

so it has an unattained 0 infimum.

Note the interesting connection to conic sections. The primal (1.2) seeks x_1 such that $-x_1^2 \geq 0$, i.e., a nonnegative point on a downward parabola: this point is "degenerate," i.e., unique.

The dual (1.3) seeks the smallest nonnegative y_{11} such that $y_{11}y_{22} \geq 1$, i.e., the lowest point on a hyperbola, which of course does not exist. See Figure 1.

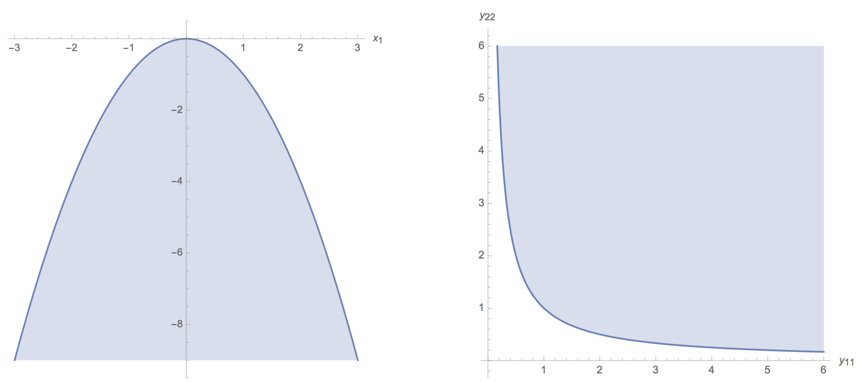

Figure 1: Parabola for the primal SDP, vs. hyperbola for the dual SDP in Example 1

Example 2. *The problem*

$$\begin{array}{lll} \sup & x_2 & \\ s.t. & x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \tag{1.4}$$

again has an attained 0 supremum. It readily follows [2] that the dual has optimal value 1 (and it is attained), so there is a finite, positive duality gap.

Curiously, while their pathologies differ, Examples 1 and 2 still look similar. First, in both examples there is a certain "antidiagonal" structure in a matrix on the left. Second, if we delete the second row and second column in all matrices in Example 2, and remove the first matrix, we get back Example 1! This raises the following questions: Do all pathological semidefinite systems "look the same"? Does the system of Example 1 appear in all of them as a "minor"?

The paper [37] made these questions precise and gave a "yes" answer to both.

We assume throughout that (P_{SD}) is feasible, and now recap some terminology from [37]. We say that the semidefinite system

$$\sum_{i=1}^{m} x_i A_i \preceq B \tag{P_{SD}}$$

[2]If the dual variable matrix is $Y = (y_{ij})$, then by the first dual constraint $y_{11} = 0$ hence by $Y \succeq 0$ we deduce $y_{12} = y_{13} = 0$, so $y_{22} = 1$.

is *badly behaved* if there is $c \in \mathbb{R}^m$ for which the value of (SDP_c) is finite but its dual has no solution with the same value. [3]

We say that (P_{SD}) is *well behaved*, if not badly behaved. A *slack matrix* or *slack* in (P_{SD}) is a positive semidefinite matrix of the form $Z = B - \sum_{i=1}^m x_i A_i$. It is not difficult to see that (P_{SD}) has a slack matrix whose rank is maximum, and our characterizations will rely on such a slack.

The following assumption [4] makes our characterizations nicer:

Assumption 1. *The maximum rank slack in (P_{SD}) is*

$$Z = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \text{ for some } 0 \leq r \leq n. \tag{1.5}$$

For the rest of the paper we fix this r. A slightly strengthened version of the main result of [37] follows.

Theorem 1. *The system (P_{SD}) is badly behaved if and only if the "Bad condition" below holds:*

Bad condition: *There is a V matrix, which is a linear combination of the A_i of the form*

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}^T & V_{22} \end{pmatrix}, \text{ where } V_{11} \text{ is } r \times r,\ V_{22} \succeq 0,\ \mathcal{R}(V_{12}^T) \not\subseteq \mathcal{R}(V_{22}), \tag{1.6}$$

where $\mathcal{R}()$ stands for rangespace. □

The Z and V matrices are *certificates* of the bad behavior. They can be chosen as

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ in Example 1, and}$$

$$Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, V = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ in Example 2.}$$

Theorem 1 is appealing: it is simple, and the excluded matrices Z and V are easy to spot in essentially all badly behaved semidefinite systems in the literature. For instance, we invite the reader to spot Z and V (after ensuring Assumption 1) in the SDP

$$\sup\ x_2 \ s.t. \begin{pmatrix} x_2 - \alpha & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{pmatrix} \preceq 0,$$

which is Example 5.79 in [10]. Here $\alpha > 0$ is a parameter, and the gap between this SDP and its dual is α. More examples can be found in [40, 22, 47, 46, 30, 45]; e.g., in an example [45, page 43] *any* matrix on the left hand side can serve as a V certificate matrix.

[3] In full detail this means that for this c the optimal value of (SDD_c)

- is unattained, and the same as that of (SDP_c) ; or
- it is finite, it may be attained or unattained, and it differs from the optimal value of (SDP_c); or
- it is $+\infty$, meaning (SDD_c) is infeasible.

[4] This assumption is easy to satisfy (at least in theory): if Z is a maximum rank slack in (P_{SD}), and Q is a matrix of suitably scaled eigenvectors of Z, then replacing all A_i by $Q^T A_i Q$ and B by $Q^T B Q$ puts Z into the required form.

Theorem 1 has an interesting geometric interpretation. Let $\operatorname{dir}(Z, \mathcal{S}_+^n)$ be the set of *feasible directions* at Z in $\mathcal{S}_+^n$, i.e.,

$$\operatorname{dir}(Z, \mathcal{S}_+^n) \;=\; \{\, Y \,|\, Z + \epsilon Y \succeq 0 \,\text{for some}\, \epsilon > 0 \,\}. \tag{1.7}$$

Then V is in the *closure* of $\operatorname{dir}(Z, \mathcal{S}_+^n)$, but it is not a feasible direction (see [37, Lemma 3]). That is, for small $\epsilon > 0$ the matrix $Z + \epsilon V$ is "almost" psd, but not quite.

We illustrate this point with the Z and V of Example 1. The shaded region of Figure 2 is the set of 2×2 psd matrices with trace equal to 1 : this set is an ellipse, so conic sections make a third appearance! The figure shows Z and $Z + \epsilon V$ for a small $\epsilon > 0$.

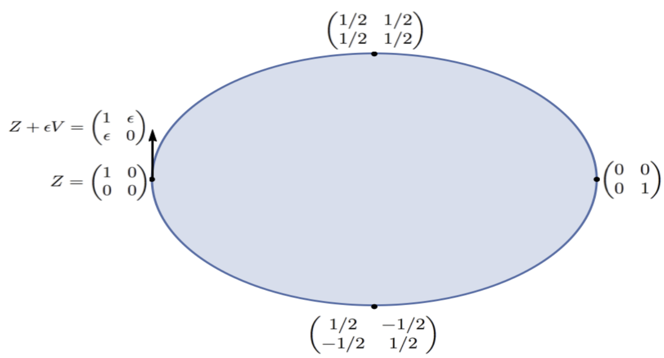

Figure 2: The matrix $Z + \epsilon V$ is "almost" psd, but not quite

In [37] we derived Theorem 1 from a much more general result ([37, Theorem 1]), which characterizes badly (and well) behaved conic linear systems. That paper does more: it also shows how to transform (P_{SD}) into a standard form, so its bad or good behavior becomes trivial to verify. The standard form is inspired by the row echelon form of a linear system of equations, [5] and indeed, the operations that we use to construct it come mostly from Gaussian elimination.

As a byproduct, [37] obtained a result of independent interest in convex geometry/convex analysis: a standard form of linear maps

$$\mathcal{M} : \; n \times n \,\text{symmetric matrices}\; \to \mathbb{R}^m,$$

to easily verify whether the image of the cone of semidefinite matrices under $\mathcal{M}$ is closed. We can thus introduce students to a fundamental issue in convex analysis: the linear image of a closed convex set is not always closed, and to verify its (non)closedness it is desirable to have simple conditions. For recent literature on closedness criteria we refer to [4, 1, 6, 35]; for perturbation results, see [12, 13]; for connections to duality theory, see e.g. [3, Theorem 7.2], [19, Theorem 2], [37, Lemma 2].

[5] The system $Ax = b$ has no solution if and only if its row echelon form contains the trivially infeasible equation $\langle 0, x \rangle = 1$.

A closely related subject is that of *asymptotes* of convex sets: an asymptote of a convex set C is an affine set which has zero distance to C without actually intersecting it. For classical and more recent studies on asymptotes, see [21, 31].

The paper [37] proved more general results than Theorem 1. However, it took a technical route: it combined a chain of lemmas and theorems, including a result on the closedness of the linear image of a closed convex cone in [35] and classical results on conic linear programs dating back to Duffin [17].

Here we give a short and self-contained proof of Theorem 1, of Theorem 2, which characterizes well behaved semidefinite systems, and show how to construct the canonical forms. We use the building blocks from [37], but we take shortcuts. Thus we mostly need elementary linear algebra, and use a basic separation argument (from convex analysis) only once.

We next state our characterization of well behaved semidefinite systems.

Theorem 2. *The system (P_{SD}) is well behaved if and only if both "Good conditions" below hold.*

Good condition 1: *There is $U \succ 0$ such that*

$$A_i \bullet \begin{pmatrix} 0 & 0 \\ 0 & U \end{pmatrix} = 0 \, \text{for all} \, i. \tag{1.8}$$

Good condition 2: *If V is a linear combination of the A_i of the form*

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}^T & 0 \end{pmatrix}, \text{ then } V_{12} = 0.$$

□

In Theorem 2 and the rest of the paper, $U \succ 0$ means that U is symmetric and positive definite, and we use the following convention:

Convention 1. *If a matrix is partitioned as in Theorems 1 or 2, then we understand that the upper left block is $r \times r$.*

Example 3. At first glance, the system

$$x_1 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.9}$$

looks very similar to the system in Example 1. However, (1.9) is well behaved, and Theorem 2 verifies this by $U = I_2$ in "Good condition 1" ("Good condition 2" trivially holds).

We next describe how to *reformulate* (P_{SD}).

Definition 1. *A semidefinite system is an* elementary reformulation, *or* reformulation *of (P_{SD}) if it is obtained from (P_{SD}) by a sequence of the following operations:*

(1) Choose an invertible matrix

$$T = \begin{pmatrix} I_r & 0 \\ 0 & M \end{pmatrix},$$

and replace A_i by $T^T A_i T$ for all i and B by $T^T B T$.

(2) Choose $\mu \in \mathbb{R}^m$ and replace B by $B + \sum_{j=1}^m \mu_j A_j$.

(3) Choose indices $i \neq j$ and exchange A_i and A_j.

(4) Choose $\lambda \in \mathbb{R}^m$ and an index i such that $\lambda_i \neq 0$, and replace A_i by $\sum_{j=1}^m \lambda_j A_j$.

Where do these operations come from? Mostly from Gaussian elimination: the last three can be viewed as elementary row operations done on (SDD_c) with some $c \in \mathbb{R}^m$. For example, operation (3) exchanges the constraints

$$A_i \bullet Y = c_i \text{ and } A_j \bullet Y = c_j.$$

Reformulating (P_{SD}) keeps the maximum rank slack the same (cf. Assumption 1). Of course, (P_{SD}) is badly behaved if and only if its reformulations are.

We organize the rest of the paper as follows. In Section 2 we give a brief review of SDP duality and complexity, and mention some applications. In Section 3 we prove Theorems 1 and 2 and show how to construct the standard reformulations. We prove the chain of implications

$$\begin{aligned} (P_{SD}) \text{ satisfies the "Bad condition"} \quad &\Longrightarrow \quad \text{it has a "Bad reformulation"} \\ &\Longrightarrow \quad \text{it is badly behaved,} \end{aligned} \tag{1.10}$$

and the "good" counterpart

$$\begin{aligned} (P_{SD}) \text{ satisfies the "Good conditions"} \quad &\Longrightarrow \quad \text{it has a "Good reformulation"} \\ &\Longrightarrow \quad \text{it is well behaved.} \end{aligned} \tag{1.11}$$

In these proofs we only use elementary linear algebra.

Of course, if (P_{SD}) is badly behaved, then it is not well behaved. Thus the implication

$$\text{Any of the "Good conditions" fail} \quad \Longrightarrow \quad \text{the "Bad condition" holds,} \tag{1.12}$$

ties everything together and shows that in (1.10) and (1.11) equivalence holds.

In Subsection 3.4 we discuss the complexity of finding the certificate matrices in Theorems 1 and 2, and a short version of [37, Theorem 1].

In Section 4 we look at linear maps that act on symmetric matrices. As promised, we show how to bring them into a canonical form, to easily check whether the image of the cone of semidefinite matrices under a given map is closed. We also point out connections to asymptotes of convex sets, and weak infeasibility in SDPs.

In Section 5 we close with a discussion: in particular we highlight how studying bad behavior of (P_{SD}) sheds light on *all* SDP pathologies.

Notation and reader's guide We write $\mathcal{S}^n$ for the set of $n \times n$ symmetric matrices, and $\mathcal{S}^n_+$ for the set of $n \times n$ symmetric positive semidefinite matrices. We need minimal additional notation in Subsection 3.3 and Section 4, which we introduce there.

This work has connections to several areas of optimization, for instance, to the complexity theory of SDPs, and to asymptotes of convex sets. We explore these connections in remarks, which can be safely skipped at first reading. The reader familiar with the duality theory, complexity, and applications of SDPs can also safely skip Section 2.

2 SDP duality, complexity, and some applications

In this section we give a brief account of SDP duality and of applications. Our review is necessarily far from complete, but we hope it is sufficient to motivate the results of the paper and help the reader appreciate the versatility of SDP.

2.1 Background on duality and complexity

We first prove the weak duality inequality (1.1). Let x be feasible in (SDP_c) and Y be feasible in (SDD_c). Then

$$B \bullet Y - \sum_{i=1}^{m} c_i x_i \;=\; B \bullet Y - \sum_{i=1}^{m} (A_i \bullet Y) x_i \;=\; (B - \sum_{i=1}^{m} x_i A_i) \bullet Y \;\geq\; 0,$$

where the last inequality follows, since the $\bullet$ product of two psd matrices is nonnegative. Clearly, x and Y are both optimal iff the last inequality holds at equality.

Two well known conditions ensure that (P_{SD}) is well behaved:

- The first is Slater's condition, i.e., when there is a positive definite slack in (P_{SD}).
- The second requires the A_i and B to be diagonal; in that case (P_{SD}) is a polyhedron.

The sufficiency of these conditions is immediate from Theorem 1. If Slater's condition holds, then Z in Theorem 1 is just I_n, so the V certificate matrix cannot exist; if the A_i and B are diagonal, then so are their linear combinations, so V again cannot exist.

Thus Theorem 1 unifies these two (seemingly unrelated) conditions, and we invite the reader to check that so does Theorem 2.

We next comment on the complexity of SDP. We can efficiently compute solutions of (SDP_c) and (SDD_c) within ϵ of being optimal, for any $\epsilon > 0$ tolerance, if we initialize a suitable algorithm with a positive definite slack in (SDP_c) and a $Y \succ 0$ feasible in (SDD_c): see e.g. [44].

However, we don't know how to solve SDPs exactly in polynomial time either in the Turing model, or in the real number model of computing. Some of the issues that arise in the Turing model are irrationality or exponential size of solutions; see [40] for a nice discussion. In fact, we do not even know how to decide in polynomial time, whether an SDP is feasible in either of these models.

Next we discuss some applications of SDP.

2.2 SDP relaxations of quadratic optimization problems

Our first example is the *maximum cut (or max cut) problem:* given an undirected graph $G = (V, E)$ with nonnegative edge weights denoted by w_{ij}, we seek to partition V into sets S and T to maximize the total weight of the edges between S and T. We identify such a partition by $x \in \{\pm 1\}^V$ as

$$x_i \;=\; \begin{cases} \;\;1 & \text{if } i \in S, \\ -1 & \text{if } i \in T. \end{cases}$$

We thus formulate the max cut problem as the quadratic optimization problem

$$\begin{array}{rl} \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - x_i x_j) \\ s.t. & x \in \{\pm 1\}^V. \end{array} \tag{2.13}$$

For a feasible solution x of (2.13) let us call the matrix xx^T a *cut matrix*. Note that $xx^T \succeq 0$ and it has rank 1. Thus the SDP

$$\begin{array}{rl} \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - y_{ij}) \\ s.t. & y_{ii} = 1 \, \forall i \\ & Y \succeq 0 \end{array} \tag{2.14}$$

is a relaxation of (2.13), since the rank 1 solutions in (2.14) are exactly the cut matrices (but of course (2.14) has other feasible solutions, which have higher rank).

Most importantly, the SDP (2.14) is a strong relaxation: the seminal work of [18] showed how to round an optimal solution of (2.14) to a cut whose weight is at least 0.8785 times the optimal. We also refer to [18] for the origin of this relaxation ([39], [16]).

We mention that the feasible set of (2.14), the set of *correlation matrices*, which has been recently termed *the elliptope,* has a fascinating geometry: see, e.g. [43, 24] When $n = 3$ we can visualize the elliptope by plotting the set of (y_{12}, y_{13}, y_{23}) such that the matrix Y with all ones on the diagonal is positive semidefinite: see Figure 3.

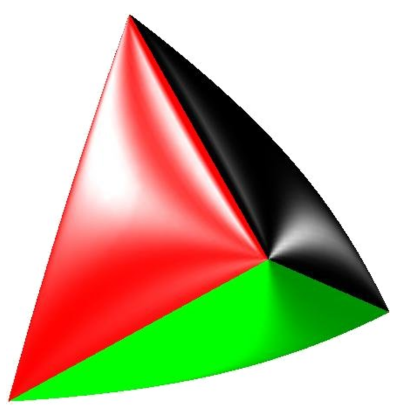

Figure 3: The elliptope in $\mathcal{S}^3_+$

Moving on from max cut, we can similarly construct relaxations of other quadratic (and typically nonconvex, thus difficult) optimization problems. To do so, we first express the constraints and objective in terms of xx^T, where x is a feasible solution, then replace xx^T by $Y \succeq 0$ to obtain a natural SDP relaxation. See [29] and [25] for surveys of such SDPs in combinatorial optimization and integer programming.

2.3 Polynomial optimization

As a first example, suppose we want to prove that the polynomial

$$f(x, y) := 5x^4 + 4x^3y + 10x^2y^2 - 12xy^3 + 23y^4 - 6y^2 + 2 \tag{2.15}$$

is nonnegative for all x and y. This is a nontrivial task, but SDP can help. Let us define the vector of monomials

$$z := [1,\, x^2,\, y^2,\, xy]^T.$$

If we find a matrix $Q \succeq 0$ such that

$$z^T Q z \;=\; f(x,y), \tag{2.16}$$

then writing $Q = LL^T$ (using Cholesky decomposition), we deduce $f(x,y) = (L^T z)^2$, which is a certificate of nonnegativity.

To find $Q = (q_{ij})_{i,j=1}^4$ we solve an SDP, whose constraints ensure that the coefficients of each monomial in (2.16) match. Thus the constraints include, for example,

$$q_{13} = -3 \text{ and } 2q_{23} + q_{44} = 10$$

(to capture the terms $-6y^2$ and $10x^2y^2$, respectively).

A particular Q and L that work for the above polynomial is

$$Q = \begin{pmatrix} 2 & 0 & -3 & 0 \\ 0 & 5 & 3 & 2 \\ -3 & 3 & 23 & -6 \\ 0 & 2 & -6 & 4 \end{pmatrix},\; L = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 2 \\ -2 & -3 & 1 & 3 \\ 0 & 2 & 0 & 0 \end{pmatrix}$$

giving the decomposition

$$f(x,y) = (1-2y^2)^2 + (x^2 - 3y^2 + 2xy)^2 + (1-y^2)^2 + (2x^2 - 3y^2)^2.$$

More generally, we say that a multivariate polynomial $f(x)$ is a sum of squares (SOS) if $f(x) = \sum_{i=1}^{\ell}(g_i(x))^2$, for some ℓ and for some g_i polynomials. Such a decomposition (if one exists) can be found by solving an SDP in a similar fashion.

SOS decompositions also help solve constrained polynomial optimization problems (POPs): see [23, 33], and the books [9, 23] that summarize the last several years of research in this area.

The SDPs that arise in this fashion are sometimes pathological and difficult; interestingly, this may happen even when the underlying POP is trivial. For in-depth studies, see e.g. [50] and [49]. Some remedies are available: e.g. [20] shows how to formulate these SDPs to ensure no duality gap and primal attainment – however, this technique does not guarantee attainment in the dual.

3 Proofs and some connections

In this section we prove the implications (1.10), (1.11), and (1.12), then point out some connections in Subsection 3.4.

3.1 The Bad

3.1.1 "Bad condition" gives a "Bad reformulation"

We assume the "Bad condition" holds in (P_{SD}) and show how to reformulate it as

$$\sum_{i=1}^{k} x_i \begin{pmatrix} F_i & 0 \\ 0 & 0 \end{pmatrix} + \sum_{i=k+1}^{m} x_i \begin{pmatrix} F_i & G_i \\ G_i^T & H_i \end{pmatrix} \preceq \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = Z, \tag{$P_{\mathrm{SD,bad}}$}$$

where

(1) matrix Z is the maximum rank slack,

(2) matrices
$$\begin{pmatrix} G_i \\ H_i \end{pmatrix} \; (i = k+1, \dots, m)$$
are linearly independent, and

(3) $H_m \succeq 0$.

We denote the constraint matrices on the left hand side by A_i throughout the process.

We first replace B by Z in (P_{SD}). We then choose $V = \sum_i \lambda_i A_i$ to satisfy the "Bad condition," We pick an i such that $\lambda_i \neq 0$, then switch A_i and A_m. Next we choose a maximal subset of the A_i matrices whose blocks comprising the last $n-r$ columns are linearly independent. We let A_m be one of these matrices (we can do this since A_m is now the V certificate matrix), and permute the A_i so this special subset becomes $A_{k+1}, \dots, A_m$ for some $k \geq 0$.

By taking linear combinations, we zero out the last $n-r$ columns of $A_1, \dots, A_k$, and arrive at the required reformulation.

Note that the systems in Examples 1 and 2 are already in the standard form.

Example 4. (Large bad example) *The system*

$$x_1 \begin{pmatrix} 9 & 7 & 7 & 1 \\ 7 & 12 & 8 & -3 \\ 7 & 8 & 2 & 4 \\ 1 & -3 & 4 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 17 & 7 & 8 & -1 \\ 7 & 8 & 7 & -3 \\ 8 & 7 & 4 & 2 \\ -1 & -3 & 2 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 & 2 & 2 & 1 \\ 2 & 6 & 3 & -1 \\ 2 & 3 & 0 & 2 \\ 1 & -1 & 2 & 0 \end{pmatrix} + x_4 \begin{pmatrix} 9 & 6 & 7 & 1 \\ 6 & 13 & 8 & -3 \\ 7 & 8 & 2 & 4 \\ 1 & -3 & 4 & 0 \end{pmatrix} \preceq \begin{pmatrix} 45 & 26 & 29 & 2 \\ 26 & 47 & 31 & -12 \\ 29 & 31 & 10 & 14 \\ 2 & -12 & 14 & 0 \end{pmatrix} \tag{3.17}$$

is badly behaved, but this would be difficult to verify by any ad hoc method.

System (3.17) satisfies the "Bad condition" with Z and V certificate matrices

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; V = \begin{pmatrix} 7 & 2 & 3 & -1 \\ 2 & 1 & 2 & -1 \\ 3 & 2 & 2 & 0 \\ -1 & -1 & 0 & 0 \end{pmatrix} \tag{3.18}$$

Indeed, $Z = B - A_1 - A_2 - 2A_4$, $V = A_4 - 2A_3$ (where we write A_i for the matrices on the left hand side, and B for the right hand side), and we explain shortly why Z is a maximum rank slack.

After the operations

$$\begin{aligned} B &:= B - A_1 - A_2 - 2A_4, \\ A_4 &= A_4 - 2A_3, \\ A_2 &= A_2 - A_3 - 2A_4, \\ A_1 &= A_1 - 2A_3 - A_4 \end{aligned} \tag{3.19}$$

system (3.17) becomes

$$x_1 \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 & 2 & 2 & 1 \\ 2 & 6 & 3 & -1 \\ 2 & 3 & 0 & 2 \\ 1 & -1 & 2 & 0 \end{pmatrix}$$
$$+ x_4 \begin{pmatrix} 7 & 2 & 3 & -1 \\ 2 & 1 & 2 & -1 \\ 3 & 2 & 2 & 0 \\ -1 & -1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3.20)$$

Besides looking simpler than (3.17), the bad behavior of (3.20) is much easier to verify, as we will see soon.

How do we convince a "user" that Z in formula (3.18) is indeed a maximum rank slack? Matrices

$$Y_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } Y_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \quad (3.21)$$

have zero $\bullet$ product with all constraint matrices, and hence also with any slack. Let S be an arbitrary slack. Since $S \bullet Y_1 = 0$, the $(4,4)$ element of S is zero, and so is its last row and column (since $S \succeq 0$). Similarly, $S \bullet Y_2 = 0$ shows the 3rd row and column of S is zero, thus Z indeed has maximum rank.

In fact, Lemma 5 in [37] proves that (P_{SD}) can always be reformulated, so that a similar sequence of matrices certifies the maximality of Z. To do so, we need to use operation (1) in Definition 1.

3.1.2 If (P_{SD}) has a "Bad reformulation", then it is badly behaved

Let x be feasible in ($P_{\text{SD,bad}}$) with a corresponding slack S. The last $n-r$ rows and columns of S must be zero, otherwise $\frac{1}{2}(S+Z)$ would be a slack with larger rank than Z. Hence, by condition (2) (after the statement of ($P_{\text{SD,bad}}$)), we deduce $x_{k+1} = \ldots = x_m = 0$, so the optimal value of the SDP

$$\sup \{ -x_m \mid x \text{ is feasible in } (P_{\text{SD,bad}}) \} \quad (3.22)$$

is 0. We prove that its dual cannot have a feasible solution with value 0, so suppose that

$$Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^T & Y_{22} \end{pmatrix} \succeq 0$$

is such a solution. By $Y \bullet Z = 0$ we get $Y_{11} = 0$, and since $Y \succeq 0$ we deduce $Y_{12} = 0$. Thus

$$\begin{pmatrix} F_m & G_m \\ G_m^T & H_m \end{pmatrix} \bullet Y = H_m \bullet Y_{22} \geq 0,$$

so Y cannot be feasible in the dual of (3.22), a contradiction. □

Example 4 (Large bad example) continued *Revisiting this example, the bad behavior of (3.17) is quite difficult to prove, whereas that of (3.20) is easy: the objective function* $\sup -x_4$ *gives a* 0 *optimal value over it, while there is no dual solution with the same value.*

3.2 The Good

3.2.1 "Good conditions" give a "Good reformulation"

Let us assume that both "Good conditions" hold. We show how to reformulate (P_{SD}) as

$$\sum_{i=1}^{k} x_i \begin{pmatrix} F_i & 0 \\ 0 & 0 \end{pmatrix} + \sum_{i=k+1}^{m} x_i \begin{pmatrix} F_i & G_i \\ G_i^T & H_i \end{pmatrix} \preceq \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = Z, \qquad (P_{\text{SD,good}})$$

with the following properties:

(1) matrix Z is the maximum rank slack.

(2) matrices H_i $(i = k+1, \dots, m)$ are linearly independent.

(3) $H_{k+1} \bullet U = \dots = H_m \bullet U = 0$ for some $U \succ 0$.

We construct the system ($P_{\text{SD,good}}$) similarly to how we constructed ($P_{\text{SD,bad}}$). Again we denote the matrices on the left hand side by A_i throughout the process.

We first replace B by Z in (P_{SD}), then choose a maximal subset of the A_i whose lower principal $(n-r) \times (n-r)$ blocks are linearly independent. We permute the A_i if needed, to make this subset $A_{k+1}, \dots, A_m$ for some $k \geq 0$.

Finally we take linear combinations to zero out the lower principal $(n-r) \times (n-r)$ block of $A_1, \dots, A_k$. By "Good condition 2" the upper right $r \times (n-r)$ block of $A_1, \dots, A_k$ (and the symmetric counterpart) also become zero. Item (3) holds by "Good condition 1", and the proof is complete. □

Example 5. (Large good example) *The system*

$$x_1 \begin{pmatrix} 9 & 7 & 7 & 1 \\ 7 & 12 & 8 & -3 \\ 7 & 8 & 2 & 4 \\ 1 & -3 & 4 & -2 \end{pmatrix} + x_2 \begin{pmatrix} 17 & 7 & 8 & -1 \\ 7 & 8 & 7 & -3 \\ 8 & 7 & 4 & 2 \\ -1 & -3 & 2 & -4 \end{pmatrix} + x_3 \begin{pmatrix} 1 & 2 & 2 & 1 \\ 2 & 6 & 3 & -1 \\ 2 & 3 & 0 & 2 \\ 1 & -1 & 2 & 0 \end{pmatrix}$$
$$+ x_4 \begin{pmatrix} 9 & 6 & 7 & 1 \\ 6 & 13 & 8 & -3 \\ 7 & 8 & 2 & 4 \\ 1 & -3 & 4 & -2 \end{pmatrix} \preceq \begin{pmatrix} 45 & 26 & 29 & 2 \\ 26 & 47 & 31 & -12 \\ 29 & 31 & 10 & 14 \\ 2 & -12 & 14 & -10 \end{pmatrix} \qquad (3.23)$$

is well behaved, but it would be difficult to improvise a method to verify this.

Instead, let us check that the "Good conditions" hold: to do so, we write A_i for the matrices on the left, and B for the right hand side.

First let us verify "Good condition 1." The Z matrix of equation (3.18) is also a maximum rank slack in (3.23), since $Z = B - A_1 - A_2 - 2A_4$, and

$$Y = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

has zero inner product with all A_i and B, hence also with any slack. "Good condition 2" can be verified directly.

The operations listed in (3.19) turn system (3.23) into

$$
\begin{aligned}
x_1 \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 & 2 & 2 & 1 \\ 2 & 6 & 3 & -1 \\ 2 & 3 & 0 & 2 \\ 1 & -1 & 2 & 0 \end{pmatrix} \\
+ x_4 \begin{pmatrix} 7 & 2 & 3 & -1 \\ 2 & 1 & 2 & -1 \\ 3 & 2 & 2 & 0 \\ -1 & -1 & 0 & -2 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},
\end{aligned}
\tag{3.24}
$$

and as we shall see shortly, the good behavior of (3.24) is much easier to verify.

3.2.2 If (P_{SD}) has a "Good reformulation," then it is well behaved

Let c be such that

$$v := \sup \{ \sum_{i=1}^{m} c_i x_i \mid x \text{ is feasible in } (P_{\mathrm{SD,good}}) \} \tag{3.25}$$

is finite. An argument like the one in Subsubsection 3.1.2 proves $x_{k+1} = \cdots = x_m = 0$ holds for any x feasible in (3.25), so

$$v = \sup \{ \sum_{i=1}^{k} c_i x_i \mid \sum_{i=1}^{k} x_i F_i \preceq I_r \}. \tag{3.26}$$

Since (3.26) satisfies Slater's condition, there is Y_{11} feasible in its dual with $Y_{11} \bullet I_r = v$.

We next choose a Y_{22} symmetric matrix (which may not be not positive semidefinite), such that

$$Y := \begin{pmatrix} Y_{11} & 0 \\ 0 & Y_{22} \end{pmatrix}$$

satisfies the equality constraints of the dual of (3.25) (this can be done, by condition (2)). We then replace Y_{22} by $Y_{22} + \lambda U$ for some $\lambda > 0$ to make it psd. After this, Y is still feasible in the dual of (3.25) (by condition (3)), and clearly $Y \bullet Z = v$ holds. The proof is now complete. □

Example 5 (Large good example) continued. *As we noted, the good behavior of (3.23) is quite difficult to prove; however that of (3.24) is easy, following the argument above.*

We remark that the procedure of constructing Y from Y_{11} was recently generalized in [38] to the case when (P_{SD}) satisfies only "Good condition 2."

3.3 Tying everything together

Now we tie everything together: we show that if any of the "Good conditions" fail, then the "Bad condition" holds.

This is the only part of the paper which needs some convex analysis, so we now introduce some terminology. We say that a set K is a *cone* if it contains all nonnegative multiples of its elements. Given a convex cone K, its dual cone is

$$K^* \;=\; \{\, y \,|\, \langle x, y \rangle \geq 0 \,\text{for all}\, x \in K \,\}.$$

We denote by $\operatorname{ri} K$ the relative interior of K and by $K^\perp$ the orthogonal complement of the linear span of K.

It is well known that $\mathcal{S}^n_+$ is full dimensional, its interior is the set of symmetric $n \times n$ positive definite matrices, and its dual cone is itself (using the $\bullet$ inner product).

In the proof we will rely on the Gordan-Stiemke theorem from convex analysis, which we state below:

Theorem 3. *Let L be a linear subspace and K a closed convex cone. Then*

$$\operatorname{ri} K \cap L^\perp = \emptyset \;\Leftrightarrow\; (K^* \setminus K^\perp) \cap L \neq \emptyset. \tag{3.27}$$

□

This result directly follows from a separation theorem in convex analysis e.g., from [11, Theorem 1.1.1].

Clearly, if "Good condition 2" fails, then the "Bad condition" holds, so assume that "Good condition 1" fails.

We invoke equivalence (3.27) with

$$L = \operatorname{lin}\{\, A_1, \ldots, A_m \,\},\; K = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & U \end{pmatrix} \,|\, U \succeq 0 \right\} \tag{3.28}$$

(i.e., L is the set of all linear combinations of the A_i). "Good condition 1" failing means $L^\perp \cap \operatorname{ri} K = \emptyset$, so there is $V \in (K^* \setminus K^\perp) \cap L$.

Since $\mathcal{S}^{n-r}_+$ is self-dual and full dimensional, we deduce

$$V \;=\; \begin{pmatrix} V_{11} & V_{12} \\ V_{12}^T & V_{22} \end{pmatrix} \text{ with } V_{22} \succeq 0,\, V_{22} \neq 0. \tag{3.29}$$

We are done if we show $\mathcal{R}(V_{12}^T) \not\subseteq \mathcal{R}(V_{22}),$ so assume otherwise, i.e., assume $V_{12}^T = V_{22}D$ for some $D \in \mathbb{R}^{(n-r)\times r}$. Define

$$M = \begin{pmatrix} I & 0 \\ -D & I \end{pmatrix},$$

and replace all A_i by M^TA_iM and B by M^TBM. After this, the maximum rank slack in (P_{SD}) remains the same (see equation (1.5)) and V is transformed into

$$M^TVM = \begin{pmatrix} V_{11} - D^TV_{12}^T & 0 \\ 0 & V_{22} \end{pmatrix}.$$

Since $V_{22} \neq 0,$ we deduce $Z + \epsilon V$ has larger rank than Z for a small $\epsilon > 0,$ which is a contradiction. The proof is complete. □

We thus proved the following corollary:

Corollary 1. *The system (P_{SD}) is badly behaved if and only if it has a reformulation of the form ($P_{SD,bad}$).*

It is well behaved if and only if it has a reformulation of the form ($P_{SD,good}$).

3.4 Complexity issues and generalization to conic linear programming

We now discuss some issues related to Theorems 1 and 2.

Remark 1. Can we actually compute the Z and V matrices of Theorem 1, or the U of Theorem 2? Regrettably, we don't know how to do this in polynomial time either in the Turing model, or in the real number model of computing. However, we will argue below that we can reduce this task to solving SDPs.

To start with the theoretical aspect of the reduction, we can find Z by invoking a facial reduction algorithm [14, 50, 36]. These algorithms must solve a sequence of SDPs in exact arithmetic.

Given Z, we can then verify whether "Good condition 1" holds by choosing L and K as in (3.28) and checking which of the alternatives in (3.27) is true. This again can be done by solving SDPs. In particular, it is immediate that $L^{\perp} \cap \operatorname{ri} K \neq \emptyset$ iff the optimal value of the SDP

$$\sup \left\{ t \,:\, A_i \bullet \begin{pmatrix} 0 & 0 \\ 0 & U \end{pmatrix} = 0 \ \forall i, \begin{pmatrix} 0 & 0 \\ 0 & U \end{pmatrix} \succeq t \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \right\} \tag{3.30}$$

is positive, and in this case this SDP provides a U. Also, $(K^* \setminus K^{\perp}) \cap L \neq \emptyset$ iff the dual of (3.30) is feasible, and in that case the dual provides a V in the form of (3.29).

In practice, heuristic and reasonably effective implementations of facial reduction algorithms exist [38, 51], and we may solve (3.30) and its dual approximately, to deduce that (P_{SD}) is nearly badly or well behaved.

We mention here that the complexity of checking attainment and the existence of a positive gap in SDPs is unknown.

Remark 2. We now briefly discuss Theorem 1 in [37] (with mildly simplified notation) from which we extracted Theorems 1 and 2 of this paper.

Let us consider the primal-dual pair of conic linear programs

$$(P_{\text{conic}}) \quad \begin{array}{ll} \sup & \sum_{i=1}^m c_i x_i \\ s.t. & \sum_{i=1}^m x_i a_i \leq_K b \end{array} \qquad\qquad \begin{array}{ll} \inf & \langle b, y \rangle \\ s.t. & \langle a_i, y \rangle = c_i \ (i = 1, \dots, m) \\ & y \geq_{K^*} 0 \end{array} \quad (D_{\text{conic}})$$

where the a_i and b are elements of a Euclidean space, K is a closed convex cone, K^* is its dual cone, and $s \geq_K t$ means $s - t \in K$.

Conic linear programs generalize SDPs, [6] and can be just as pathological. On one hand the weak duality inequality $\sum_{i=1}^m c_i x_i \leq \langle b, y \rangle$ is immediate if x and y are feasible. On the other hand, the optimal values of (P_{conic}) and (D_{conic}) are not always attained, and there may be a positive duality gap.

[6] We recover SDPs by choosing the underlying space as $\mathcal{S}^n$, $K = K^* = \mathcal{S}^n_+$. and the $\bullet$ product as inner product $\langle \rangle$.

To capture most of these pathologies, we say that the conic linear system

$$\sum_{i=1}^{m} x_i a_i \leq_K b \tag{3.31}$$

is *badly behaved* if there is $c \in \mathbb{R}^m$ for which the optimal value of (P_{conic}) is finite but (D_{conic}) has no solution with the same value.

Naturally, we say that $s \in K$ is a *slack* in (P_{conic}) if $s = b - \sum_{i=1}^m x_i a_i$ for some x; and that $z \in K$ is a *maximum slack* if it is in the relative interior of the set of all slacks.

An excerpt of Theorem 1 in [37] follows:

Theorem 4. *Suppose that K is polyhedral; or $K = \mathcal{S}_+^n$; or K is the p-cone*

$$\{\, x \in \mathbb{R}^k \,:\, x_1 \geq (|x_2|^p + \cdots + |x_k|^p)^{1/p} \,\}, \text{ where } p > 0;$$

or K is the direct product of such cones. Then (3.31) is badly behaved iff the "Conic bad condition" below holds:

Conic bad condition

There is v, which is a linear combination of the a_i such that

$$v \in \text{closure}(\text{dir}(z, K)) \setminus \text{dir}(z, K).$$

Here z is a maximum slack in (P_{conic}), and $\text{dir}(z, K) = \{\, y \,|\, z + \epsilon y \in K \text{ for some } \epsilon > 0 \,\}$ is the set of feasible directions at z in K.

□

Although Theorem 4 is a bit more technical than previous results of this paper, we next give an intuitive explanation:

- First, the "Conic bad condition" is quite natural, because it evidently fails, if K is polyhedral: in that case $\text{dir}(z, K)$ is a closed set.
- Second, we readily recover Theorem 1 of this paper as follows. If the pair of conic linear programs are a pair of SDPs, then z is a maximum *rank* slack. Invoking the characterization of feasible directions in $\mathcal{S}_+^n$ in [37, Lemma 3], the "Conic bad condition" translates into the "Bad condition".

We remark that Theorem 1 of [37] is more general: it states a conic version of the two "Good conditions" as well, and proves that the "Conic bad condition" is necessary and sufficient for a large class of cones, called *nice cones*.

4 When is the linear image of the semidefinite cone closed?

We now address a question of independent interest in convex analysis/convex geometry:

Given a linear map, is the image of $\mathcal{S}_+^n$ under the map closed?

This question fits in a much broader context. More generally, we can ask: when is the linear image of a closed convex set, say C, closed? Such closedness criteria are fundamental in convex analysis, and Chapter 9 in Rockafellar's classic text [42] is entirely dedicated to them. See also Chapter 2.3 in [1]. For more recent work on this subject we refer to [4, 6]; and [12, 13]. The latter paper shows that the set of linear maps under which the image of a closed convex cone is *not* closed is small both in measure and in category.

The closedness of the linear image of a closed convex cone ensures that a conic linear system is well-behaved (as defined in Remark 2); see e.g., [3, Theorem 7.2], [19, Theorem 2], [37, Lemma 2]. We studied criteria for the closedness of the linear image of a closed convex cone in [35], and the results therein led to [37], and to this paper.

The special case $C = \mathcal{S}^n_+$ is interesting, since the semidefinite cone is one of the simplest nonpolyhedral sets [7] whose geometry is well understood, see, e.g. [2, 34]. It turns out that the (non)closedness of the image of $\mathcal{S}^n_+$ admits simple combinatorial characterizations.

We need some basic notation: for a set S we define its *frontier* $\text{front}(S)$ as the difference between its closure and the set itself,

$$\text{front}(S) := \text{closure}(S) \setminus S.$$

Example 6. *Define the map*

$$\mathcal{S}^2 \ni Y \to (y_{11}, 2y_{12}). \tag{4.32}$$

The image set – shown on Figure 4 in blue, and its frontier in red – is

$$\{(0,0)\} \cup \{\, (\alpha, \beta) : \alpha > 0 \,\}, \tag{4.33}$$

so it is not closed. [8]

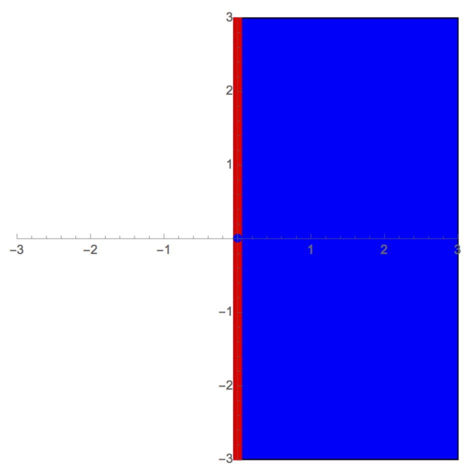

Figure 4: The image set is in blue, and its frontier is in red

In more involved examples, however, the (non)closedness of the image is much harder to check.

[7] The linear image of a polyhedron is also a polyhedron, so it is trivially closed.

[8] For example, $(0, 2)$ is in the frontier since $(\epsilon, 2)$ is the image of the psd matrix $\begin{pmatrix} \epsilon & 1 \\ 1 & 1/\epsilon \end{pmatrix}$ for all $\epsilon > 0$, but no psd matrix is mapped to $(0, 2)$.

Example 7. *This example is based on Example 6 in [27]. Define the linear map*

$$\mathcal{S}^3 \ni Y \to (5y_{11} + 4y_{22} + 4y_{13}, 3y_{11} + 3y_{22} + 2y_{13}, 2y_{11} + y_{22} + 2y_{13}). \tag{4.34}$$

As we will see, the image of $\mathcal{S}^3_+$ is not closed, but verifying this by any ad hoc method seems very difficult.

For convenience, we will represent linear maps from $\mathcal{S}^n$ to $\mathbb{R}^m$ by matrices $A_1, \dots, A_m \in \mathcal{S}^n$ and write

$$\mathcal{A}(x) = \sum_{i=1}^m x_i A_i, \text{ and } \mathcal{A}^*(Y) \; = \; (A_1 \bullet Y, \dots, A_m \bullet Y). \tag{4.35}$$

That is, we consider a linear map from $\mathcal{S}^n$ to $\mathbb{R}^m$ as the *adjoint* of a suitable linear map in the opposite direction, to better fit the framework of [35, 37].

The next proposition connects the closedness of the linear image of $\mathcal{S}^n_+$ and the bad (or good) behavior of a homogeneous semidefinite system. A simple proof follows, e.g., from the classic separation theorem [11, Theorem 1.1.1].

Proposition 1. *Given a linear map $\mathcal{A}$ and its adjoint $\mathcal{A}^*$ as in (4.35), the set $\mathcal{A}^*(\mathcal{S}^n_+)$ is not closed if and only if the system*

$$\sum_{i=1}^m x_i A_i \preceq 0 \tag{P_{SDH}}$$

is badly behaved. In particular, $c \in \operatorname{front}(\mathcal{A}^*(\mathcal{S}^n_+))$ if and only if

$$\begin{array}{rl} \sup & \sum_{i=1}^m c_i x_i \\ s.t. & \sum_{i=1}^m x_i A_i \preceq 0 \end{array} \tag{4.36}$$

has optimal value zero, but its dual is infeasible. □

Thus, if (P_{SDH}) satisfies Assumption 1, then the characterizations of Theorems 1 and 2 apply.

More interestingly, Corollary 1 implies:

Corollary 2. *Suppose $\mathcal{A}$ and $\mathcal{A}^*$ are represented as in (4.35). Then $\mathcal{A}^*(\mathcal{S}^n_+)$ is*

(1) not closed if and only if the homogeneous system (P_{SDH}) has a reformulation of the form ($P_{SD,bad}$);

(2) closed if and only if the homogeneous system (P_{SDH}) has a reformulation of the form ($P_{SD,good}$).

We next illustrate Corollary 2 by continuing the previous examples. One the one hand, reformulating the map of Example 6 does not help either to verify nonclosedness of the image set, or to exhibit a vector in its frontier. Reformulating, however, does help a lot in Example 7.

Example 6 continued. *We can write the map in (4.32) as*

$$\mathcal{S}^2 \ni Y \to \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \bullet Y, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \bullet Y \right),$$

so the corresponding homogeneous semidefinite system is

$$x_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq 0,$$

whose standard bad reformulation is essentially the same:

$$x_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

(we just replaced the the right hand side by the maximum rank slack).

Example 7 continued. *The homogeneous semidefinite system corresponding to the map in (4.34) is*

$$x_1 \begin{pmatrix} 5 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 3 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq 0. \tag{4.37}$$

Its standard bad reformulation is [9]

$$x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.38}$$

Let $\mathcal{A}(x)$ be the left hand side in (4.37) and $\mathcal{A}'(x)$ the left hand side in (4.38). Then

$$\mathcal{A}'^*(Y) = (y_{11}, y_{11} + y_{22}, y_{22} + 2y_{13}),$$

and a calculation shows (for details, see Example 6 in [27])

$$\begin{array}{rcl} \text{closure}(\mathcal{A}'^*\mathcal{S}^3_+) & = & \{(\alpha, \beta, \gamma) \,:\, \beta \geq \alpha \geq 0\}, \\ \text{front}(\mathcal{A}'^*\mathcal{S}^3_+) & = & \{(0, \beta, \gamma) \,|\, \beta \geq 0,\, \beta \neq \gamma\}. \end{array} \tag{4.39}$$

The set $\mathcal{A}'^*(\mathcal{S}^3_+)$ is shown in Figure 5 in blue, and its frontier in red. Note that the blue diagonal segment on the red facet actually belongs to $\mathcal{A}'^*(\mathcal{S}^3_+)$.

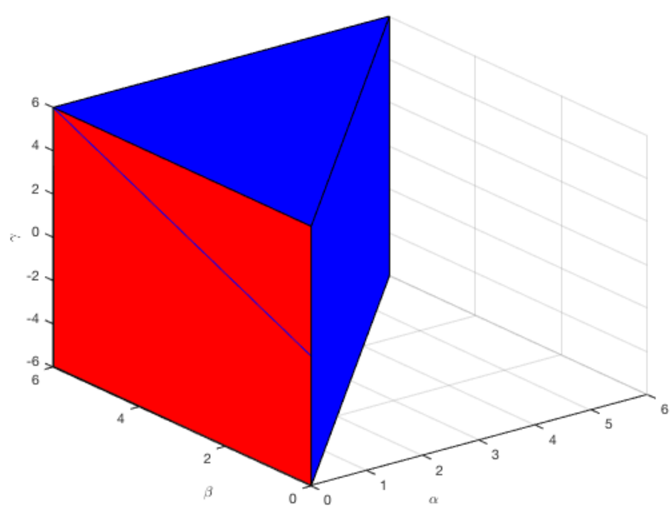

Figure 5: The set $\mathcal{A}'^*(\mathcal{S}^3_+)$ is in blue, and its frontier in red

[9]Let us denote the matrices in (4.37) by A_1, A_2, A_3. Then we get (4.38) by the operations $A_2 = A_2 - A_3; A_1 = A_1 - 2A_3; A_3 = A_3 - A_1 - A_2$ and replacing the right ahnd side by A_2.

The algebraic description of $\mathcal{A}'^*(\mathcal{S}^3_+)$ *is still not easy to find. However, its nonclosedness readily follows from Theorem 1, since we can choose* Z *as the right hand side in (4.38) and* V *as the coefficient matrix of* x_3.

We can also quickly exhibit an element in $\operatorname{front}(\mathcal{A}'^*(\mathcal{S}^3_+))$: *the optimal value of*

$$\sup \{ -x_3 \,|\, s.t.\, \mathcal{A}'(x) \preceq 0 \},$$

is 0, *but its dual is infeasible, hence*

$$(0,0,-1)^T \in \operatorname{front}(\mathcal{A}'^* \mathcal{S}^3_+).$$

Remark 3. To conclude this section, we connect Corollary 1 to results on asymptotes of convex sets, and to weak infeasibility in SDPs. Let us define the distance of sets S_1 and S_2 as

$$\operatorname{dist}(S_1, S_2) := \inf \{ \| x_1 - x_2 \| \;|\, x_1 \in S_1,\, x_2 \in S_2 \}.$$

Letting $H := \{ Y \,|\, \mathcal{A}^*(Y) = c \}$, a standard argument shows the following three statements are equivalent:

- $c \in \operatorname{front}(\mathcal{A}^*(\mathcal{S}^n_+))$;
- $H \cap \mathcal{S}^n_+ = \emptyset$, and $\operatorname{dist}(H, \mathcal{S}^n_+) = 0$;
- (SDD_c) is infeasible, and the alternative system

$$\begin{array}{rcl} \sum_{i=1}^m c_i x_i & = & 1 \\ \sum_{i=1}^m x_i A_i & \preceq & 0 \end{array} \tag{4.40}$$

  is also infeasible.

Two terminologies are used to express the equivalent statements above.

The first terminology says that H is an *(affine) asymptote* of $\mathcal{S}^n_+$. Asymptotes of convex sets were introduced in the classical paper [21]. For example,

$$H = \{ Y \in \mathcal{S}^2 \,:\, y_{11} = 0,\, y_{12} = 1 \}$$

is an asymptote of $\mathcal{S}^2_+$; to see why, we can intersect $\mathcal{S}^2_+$ with the hyperplane $y_{12} = 1$ and check that $y_{11} = 0$ is an asymptote of the resulting hyperbola: see the second part of Figure 1.

For more recent work on asymptotes, see [31], which shows that a convex set C has an asymptote if and only if there is a quadratic function that is convex and lower bounded on C, but does not attain its minimum.

The second terminology says that (SDD_c) is *weakly infeasible*. Note that when (SDP_c) has finite optimal value and the dual (SDD_c) is infeasible, it must be weakly infeasible. Indeed, suppose not; then the alternative system (4.40) has a feasible solution x, and adding a large multiple of x to a feasible solution of (SDP_c) proves the latter is unbounded, a contradiction.

For recent results on weak infeasibility in SDP and in conic linear programming, see [28] that proved that a weakly infeasible SDP over $\mathcal{S}^n_+$ has a weakly infeasible subsystem of dimension at most $n-1$. See Corollary 1 in [27] for a generalization to conic linear programs: that result uses a fundamental geometric parameter of the underlying cone, namely the length of the longest chain of faces.

5 Discussion and conclusion

We gave an elementary (almost purely linear algebraic) proof of a combinatorial characterization of pathological semidefinite systems. En route, we showed how to transform semidefinite systems into a canonical form to easily verify their pathological behavior. The canonical form also turned out to be useful for a related problem: it allows one to easily verify whether the linear image of $\mathcal{S}_+^n$ is closed.

We conclude with a discussion.

- Recalling our assumption that (P_{SD}) is feasible, we may ask: does studying its bad behavior help us understand *all* pathologies in SDPs?

  The answer is yes, at least to some extent. In particular, it helps us understand weak infeasibility, a well known pathology of infeasible SDPs: Remark 3 shows that all c that make (SDD_c) weakly infeasible are suitable objective functions associated with badly behaved *homogeneous* (hence feasible) systems.

  However, we cannot yet fully distinguish among bad objective functions; for example, we cannot tell which $c \in \mathbb{R}^m$ gives a finite positive duality gap, and which gives the more benign pathology of zero duality gap coupled with unattained dual optimal value.

- The interplay of semidefinite programming and algebraic geometry is a very active recent research area: see e.g. [8, 7, 32, 7, 48]. It would be interesting to see possible connections between our results and algebraic geometry.

- Note that the canonical systems ($P_{\text{SD,bad}}$) and ($P_{\text{SD,good}}$) are naturally split into two parts:

  – a "Slater part," namely the system $\sum_{i=1}^k x_i F_i \preceq I_r$, and
  – a "Redundant part," which corresponds to always zero variables $x_{k+1}, \ldots, x_m$.

  In ($P_{\text{SD,bad}}$) the "Redundant part" is responsible for the bad behavior.

  In ($P_{\text{SD,good}}$) the "Redundant part" is essentially linear: we can find the corresponding dual variable Y_{22} by solving a system of equations, then doing a linesearch.

- In more recent work, canonical forms of semidefinite – and more generally, of conic linear – systems turned out to be useful

  – to verify the infeasibility of an SDP (see [26]) and
  – to verify the infeasibility and weak infeasibility of conic linear programs: see [27].

- To construct the canonical systems, the bulk of the work is transforming the linear map

$$\mathbb{R}^m \ni x \to \mathcal{A}(x) = \sum_{i=1}^m x_i A_i.$$

  Indeed, operations (2)-(4) find an invertible linear map $M : \mathbb{R}^m \to \mathbb{R}^m$ so that $\mathcal{A}M$ is in an easier-to-handle form.

  Canonical forms of linear maps are ubiquitous in linear algebra: see, for example, the row echelon form, or the eigenvector decomposition of a matrix. This work (as well as [26] and [27]) shows that they are also useful in a somewhat unexpected area, the duality theory of conic linear programs.

Acknowledgement I am grateful to the referees and the Area Editor for their detailed and helpful feedback; to Cedric Josz, Dan Molzahn, and Hayato Waki for helpful discussions on SDP; and to Yuzixuan Zhu for her careful reading of the paper and her help with the figures.

References

[1] Alfred Auslender and Marc Teboulle. *Asymptotic cones and functions in optimization and variational inequalities.* Springer Science & Business Media, 2006. 5, 18

[2] George Phillip Barker and David Carlson. Cones of diagonally dominant matrices. *Pacific J. Math.*, 57:15–32, 1975. 18

[3] Alexander Barvinok. *A Course in Convexity.* Graduate Studies in Mathematics. AMS, 2002. 1, 2, 5, 18

[4] Heinz Bauschke and Jonathan M. Borwein. Conical open mapping theorems and regularity. In *Proceedings of the Centre for Mathematics and its Applications 36*, pages 1–10. Australian National University, 1999. 5, 18

[5] Aharon Ben-Tal and Arkadii Nemirovskii. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. SIAM, Philadelphia, PA, 2001. 1, 2

[6] Dimitri Bertsekas and Paul Tseng. Set intersection theorems and existence of optimal solutions. *Math. Program.*, 110:287–314, 2007. 5, 18

[7] Avinash Bhardwaj, Philipp Rostalski, and Raman Sanyal. Deciding polyhedrality of spectrahedra. *SIAM J. Opt.*, 25(3):1873–1884, 2015. 22

[8] Grigoriy Blekherman, Pablo Parrilo, and Rekha Thomas, editors. *Semidefinite Optimization and Convex Algebraic Geometry.* MOS/SIAM Series in Optimization. SIAM, 2012. 22

[9] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry.* SIAM, 2012. 10

[10] Frédéric J. Bonnans and Alexander Shapiro. *Perturbation analysis of optimization problems.* Springer Series in Operations Research. Springer-Verlag, 2000. 1, 2, 4

[11] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples, Second Edition.* CMS Books in Mathematics. Springer, 2005. 1, 2, 15, 19

[12] Jonathan M. Borwein and Warren B. Moors. Stability of closedness of convex cones under linear mappings. *J. Convex Anal.*, 16(3–4):699–705, 2009. 5, 18

[13] Jonathan M. Borwein and Warren B. Moors. Stability of closedness of convex cones under linear mappings ii. *Journal of Nonlinear Analysis and Optimization: Theory & Applications*, 1(1), 2010. 5, 18

[14] Jonathan M. Borwein and Henry Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. App.*, 83:495–530, 1981. 16

[15] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004. 1, 2

[16] Charles Delorme and Svatopluk Poljak. Combinatorial properties and the complexity of a max-cut approximation. *European Journal of Combinatorics*, 14(4):313–333, 1993. 9

[17] Richard J. Duffin. Infinite programs. In A.W. Tucker, editor, *Linear inequalities and Related Systems*, pages 157–170. Princeton University Press, 1956. 6

[18] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995. 9

[19] Didier Henrion and Milan Korda. Convex computation of the region of attraction of polynomial control systems. *IEEE Trans. Autom. Control*, 59(2):297–312, 2014. 5, 18

[20] Cédric Josz and Didier Henrion. Strong duality in Lasserre's hierarchy for polynomial optimization. *Optimization Letters*, 10(1):3–10, 2016. 10

[21] Victor Klee. Asymptotes and projections of convex sets. *Mathematica Scandinavica*, 8(2):356–362, 1961. 6, 21

[22] Igor Klep and Markus Schweighofer. An exact duality theory for semidefinite programming based on sums of squares. *Math. Oper. Res.*, 38(3):569–590, 2013. 4

[23] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001. 10

[24] Monique Laurent and Svatopluk Poljak. On the facial structure of the set of correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 17(3):530–547, 1996. 9

[25] Monique Laurent and Frank Vallentin. *Semidefinite Optimization.* Available from "`http://homepages.cwi.nl/~monique/master_SDP_2016.pdf`". 1, 2, 9

[26] Minghui Liu and Gábor Pataki. Exact duality in semidefinite programming based on elementary reformulations. *SIAM J. Opt.*, 25(3):1441–1454, 2015. 22

[27] Minghui Liu and Gábor Pataki. Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming. *Math. Program. Ser. A, to appear*, 2017. 19, 20, 21, 22

[28] Bruno Lourenco, Masakazu Muramatsu, and Takashi Tsuchiya. A structural geometrical analysis of weakly infeasible SDPs. *Journal of the Operations Research Society of Japan*, 59(3):241–257, 2015. 21

[29] László Lovász. Semidefinite programs and combinatorial optimization. In *Recent advances in algorithms and combinatorics*, pages 137–194. Springer, 2003. 9

[30] Zhi-Quan Luo, Jos Sturm, and Shuzhong Zhang. Duality results for conic convex programming. Technical Report Report 9719/A, Erasmus University Rotterdam, Econometric Institute, The Netherlands, 1997. 4

[31] Juan Enrique Martinez-Legaz, Dominikus Noll, and Wilfredo Sosa. Minimization of quadratic functions on convex sets without asymptotes. Technical report, 2018. 6, 21

[32] Jiawang Nie, Kristian Ranestad, and Bernd Sturmfels. The algebraic degree of semidefinite programming. *Mathematical Programming*, 122(2):379–405, 2010. 22

[33] Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96(2):293–320, 2003. 10

[34] Gábor Pataki. The geometry of semidefinite programming. In Romesh Saigal, Lieven Vandenberghe, and Henry Wolkowicz, editors, *Handbook of semidefinite programming*. Kluwer Academic Publishers, also available from www.unc.edu/˜pataki, 2000. 18

[35] Gábor Pataki. On the closedness of the linear image of a closed convex cone. *Math. Oper. Res.*, 32(2):395–412, 2007. 5, 6, 18, 19

[36] Gábor Pataki. Strong duality in conic linear programming: facial reduction and extended duals. In David Bailey, Heinz H. Bauschke, Frank Garvan, Michel Théra, Jon D. Vanderwerff, and Henry Wolkowicz, editors, *Proceedings of Jonfest: a conference in honour of the 60th birthday of Jon Borwein*. Springer, also available from http://arxiv.org/abs/1301.7717, 2013. 16

[37] Gábor Pataki. Bad semidefinite programs: they all look the same. *SIAM J. Opt.*, 27(1):146–172, 2017. 1, 2, 3, 4, 5, 6, 7, 12, 16, 17, 18, 19

[38] Frank Permenter and Pablo Parrilo. Partial facial reduction: simplified, equivalent sdps via approximations of the psd cone. *Mathematical Programming*, pages 1–54, 2014. 14, 16

[39] Svatopluk Poljak and Franz Rendl. Nonpolyhedral relaxations of graph-bisection problems. *SIAM Journal on Optimization*, 5(3):467–487, 1995. 9

[40] Motakuri V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Math. Program. Ser. B*, 77:129–162, 1997. 4, 8

[41] James Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* MPS-SIAM Series on Optimization. SIAM, Philadelphia, USA, 2001. 1, 2

[42] Tyrrel R. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, USA, 1970. 18

[43] Alexander Shapiro. Weighted minimum trace factor analysis. *Psychometrika*, 47(3):243–264, 1982. 9

[44] Michael J. Todd. Semidefinite optimization. *Acta Numer.*, 10:515–560, 2001. 1, 2, 8

[45] Levent Tunçel. *Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization.* Fields Institute Monographs, 2011. 1, 2, 4

[46] Levent Tunçel and Henry Wolkowicz. Strong duality and minimal representations for cone optimization. *Comput. Optim. Appl.*, 53:619–648, 2012. 4

[47] Lieven Vandenberghe and Steven Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996. 4

[48] Cynthia Vinzant. What is ... a spectrahedron? *Notices Amer. Math. Soc.*, 61(5):492–494, 2014. 22

[49] Hayato Waki. How to generate weakly infeasible semidefinite programs via Lasserre's relaxations for polynomial optimization. *Optim. Lett.*, 6(8):1883–1896, 2012. 10

[50] Hayato Waki and Masakazu Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013. 10, 16

[51] Yuzixuan Zhu, Gabor Pataki, and Quoc Tran-Dinh. Sieve-sdp: a simple facial reduction algorithm to preprocess semidefinite programs. *arXiv preprint arXiv:1710.08954*, 2017. 16