

On positive duality gaps in semidefinite programming

Gábor Pataki

June 27, 2018

Abstract

Semidefinite programs (SDPs) are optimization problems with linear constraints, and semidefinite matrix variables. SDPs are some of the most useful and versatile optimization problems to emerge since the 1990s, appearing in many contexts as combinatorial optimization, control theory and engineering. We study SDPs with positive duality gaps, i.e., different optimal values in the primal and dual problems. These SDPs are considered extremely pathological, they are often unsolvable, and they also serve as models of more general pathological convex programs.

We first fully characterize two variable SDPs with positive gaps: we transform them into a standard form which makes the positive gap easy to recognize. The transformation is very simple, as it mostly uses elementary row operations coming from Gaussian elimination. We next show that the two variable case sheds light on larger SDPs with positive gaps: we present SDPs in any dimension in which the positive gap is certified by the same structure as in the two variable case. We analyze an important parameter, the *singularity degree* of the duals of our SDPs and show that it is the largest that can result in a positive gap.

We complete the paper by generating a library of difficult SDPs with positive gaps (some of these SDPs have only two variables), and a computational study.

Key words: semidefinite programming; duality; positive duality gaps; facial reduction; singularity degree

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

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

1 Introduction

Consider the primal-dual pair of semidefinite programs (SDPs)

$$(P) \quad \begin{array}{ll} \sup & \sum_{i=1}^m c_i x_i \\ \text{s.t.} & \sum_{i=1}^m x_i A_i \preceq B \end{array} \quad \begin{array}{ll} \inf & B \bullet Y \\ \text{s.t.} & Y \succeq 0 \\ & A_i \bullet Y = c_i \quad (i = 1, \dots, m) \end{array} \quad (D)$$

where A_1, \dots, A_m , and B are $n \times n$ symmetric matrices and c_1, \dots, c_m are scalars. For symmetric matrices S and T we write $S \preceq T$ to say that $T - S$ is positive semidefinite (psd) and $T \bullet S := \text{trace}(TS)$ is their inner product.

SDPs are arguably some of the most useful and widespread optimization problems of the last few decades. They naturally generalize linear programs, and are applied in areas as diverse as combinatorial optimization, control theory, robotics, and machine learning.

While SDPs are useful, they are also often pathological (see [4, 3, 2]). In particular,

- they may not attain their optimal values;
- they may be infeasible, but have zero distance to the set of feasible instances. In that case we say they are *weakly infeasible*.
- the optimal values of (P) and (D) may differ. In that case we say there is a positive (duality) gap.

SDPs with positive duality gaps are often seen as the “most pathological/most interesting,” as they are in stark contrast with linear programs. They may also look innocent, but still defeat SDP solvers.

Example 1. *Let us consider the SDP*

$$\begin{aligned} \sup \quad & -x_2 \\ \text{s.t.} \quad & x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \succeq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned} \quad (P_{\text{small}})$$

Note that the $(1,3)$ element of A_2 is 1, but the $(3,3)$ element in all matrices is zero. Hence $x_2 = 0$ identically, so the optimal value of (P_{small}) is 0.

Let $Y = (y_{ij}) \succeq 0$ be the dual variable matrix. By the first dual constraint $y_{11} = 0$, so by psdness the first row and column of Y is zero, and the dual is equivalent to

$$\begin{aligned} \inf \quad & y_{22} \\ \text{s.t.} \quad & -y_{22} = -1, \end{aligned} \quad (1.1)$$

whose optimal solution is 1.

The Mosek commercial SDP solver reports that this SDP is “primal infeasible.”

Research in the last few years has greatly helped us to better understand pathological SDPs. In [17] we characterized pathological semidefinite *systems*, which yield an unattained dual value or positive gap for *some* $c \in \mathbb{R}^m$. However, [17] does not distinguish among “bad” objective functions: for example, it does not tell which c gives a positive gap, and which gives the more harmless pathology of zero gap and unattained dual value. The paper [23] showed how some positive gap SDPs can be found from a homogeneous primal-dual pair (with $c = 0$ and $B = 0$). See [12] for a study on weak infeasibility in SDP; and [10] and [9] for characterizations of infeasibility and weak infeasibility in SDPs and in conic LPs.

A related pathology of feasible conic LPs is *ill-posedness*, meaning zero distance to the set of infeasible instances. Ill posedness, and the distance to infeasibility of conic LPs was introduced in the seminal work [20]. For followup work, see e.g., [14, 7, 18].

This work is also motivated by [25], which develops a good understanding of the two trust region problem via studying the minimal, two variable case.

Despite intensive research on pathological SDPs, even basic questions about positive duality gaps are as yet unanswered. For example, we need $m \geq 2$ to have a positive gap, however, no such result has been published. No analysis of the $m = 2$ case is known.

Motivation and contributions One of our main motivations is the rather artificial look of Example 1: variable x_1 seems to exist just to ensure a zero block in the dual variable, and it is deceptively simple to verify the duality gap. Many other instances in the literature look quite similar. In this work we show that simple certificates of positive gaps exist in a large class of SDPs, not just in artificial looking examples.

We build on three main ideas. First, we analyze SDPs with just two variables, and transform them to a form that makes the positive gap (if any) easy to recognize. The transformation is very simple, as it mostly uses operations coming from Gaussian elimination.

Second, we show that the same structure that causes a positive gap in the two variable case appears in arbitrary higher dimensions as well, leading to a positive gap.

Third, we prove that our positive gap instances are, in a sense, best possible. We analyze an important parameter, the *singularity degree* of the dual SDPs, and show that they are the largest that permit a positive duality gap.

Our first result is

Theorem 1. *Suppose $m = 2$. Then $\text{val}(P) < \text{val}(D)$ iff (P) has a reformulation*

$$\begin{array}{l} \sup \quad c'_2 x_2 \\ \text{s.t.} \quad x_1 \left(\begin{array}{c|c|c|c} \Lambda & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \end{array} \right) + x_2 \left(\begin{array}{c|c|c|c} \times & \times & \times & M \\ \hline \times & \Sigma & & \\ \hline \times & & I_s & \\ \hline M^T & & & \end{array} \right) \preceq \left(\begin{array}{c|c|c|c} I_p & & & \\ \hline & I_{r-p} & & \\ \hline & & & \\ \hline & & & \end{array} \right), \end{array} \quad (P_{\text{ref}})$$

where Λ and Σ are diagonal, Λ is positive definite, $M \neq 0$, $c'_2 < 0$ and $s \geq 0$.¹

□

Here, and in the rest of the paper we denote by $\text{val}()$ the optimal value of an optimization problem. The partitioning of the matrices indicates their order, e.g., Λ has order p . Also, the blocks marked by \times symbols have arbitrary elements.

In Subsection 1.1 we precisely define “reformulations.” However, knowing only that a reformulated problem has a positive gap with its dual iff the original one does, we can already prove the “If” direction in Theorem 1. To build intuition, we give this proof below; we essentially reuse the argument from Example 1.

Proof of “If” in Theorem 1: Since $M \neq 0$, we have $x_2 = 0$ in any feasible solution of (P_{ref}) , so $\text{val}(P_{\text{ref}}) = 0$.

On the other hand, if Y is feasible in the dual of (P_{ref}) , then by the first dual constraint the upper left $p \times p$ block of Y is zero, and $Y \succeq 0$, so the first p rows and columns of Y are zero. So the dual is equivalent to

$$\begin{array}{l} \inf \quad \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \bullet Y' \\ \text{s.t.} \quad \begin{pmatrix} \Sigma & 0 \\ 0 & I_s \end{pmatrix} \bullet Y' = c'_2 \\ \quad \quad \quad Y' \succeq 0, \end{array} \quad (1.2)$$

¹Example 1 needs no reformulation and has $\Sigma = [-1]$ and $s = 0$.

whose optimal value is positive (since $c'_2 < 0$). □

The rest of the paper is organized as follows:

- In Section 2 we prove the “only if” direction of Theorem 1. We also prove Corollary 1 which shows that when $m = 2$, the “worst” pathology – positive gap coupled with unattained optimal value – does not happen. Here we also give Corollary 2 with a complete characterization of two variable semidefinite *systems* that admit a positive gap for some c .
- In Section 3 we present a class of SDPs in any dimension in which the same structure certifies the positive gap, as in the two variable case.
- In Section 4 we analyze an important parameter, the *singularity degree* of two related dual type SDPs. The first dual is (D) and the second is a homogeneous dual

$$\begin{aligned} A_i \bullet Y &= 0 \quad (i = 1, \dots, m) \\ B \bullet Y &= 0 \\ Y &\succeq 0. \end{aligned} \tag{HD}$$

We prove that the singularity degrees are $m - 1$ and m , respectively.

- In Section 5 we prove an auxiliary result, which we believe to be of independent interest: we show how to reformulate (P) - (D) so the maximum rank slack matrix in (P) and the maximum rank solution of (HD) both become easy to see.
- In Section 6 we show that our instances are, in a sense, best possible. We prove that the singularity degrees of (D) and of (HD) are always $\leq m$ and $\leq m + 1$, respectively, and when equality holds there is no duality gap.
- In Section 7 we generate a library of SDPs with positive gaps, many of them with only two variables. They are patterned after the infeasible and weakly infeasible SDPs described in [9]: their status can be verified by inspection in exact arithmetic, but are challenging for SDP solvers.

Reader’s guide Sections 2, 3 and 7 can be read with a minimal background in linear algebra and convex analysis, which we summarize in Subsection 1.1. Most of the proofs are elementary, and we illustrate our results with many examples.

Literature review A suitable *constraint qualification* (CQ) can prevent duality gaps: if (P) is *strictly feasible* ($B - \sum_i x_i A_i \succ 0$ for some x), then the gap is zero. Similarly, if (D) is strictly feasible (there is a feasible $Y \succ 0$) then the duality gap is zero. However, even if strict feasibility fails, the duality gap may still be zero.

Our understanding of pathological SDPs (and of other conic linear programs) is greatly helped by facial reduction algorithms. These algorithms regularize (P) or (D) by replacing the semidefiniteness constraint with membership in a face of the set of psd matrices. Facial reduction algorithms originated in [5]. Simplified versions appeared in [16, 24, 9].

Pathological SDPs are extremely difficult to solve by interior point methods. However, some recent implementations of facial reduction [19, 26] work on some pathological SDPs. We refer to [11] for an implementation of the Douglas-Rachford splitting algorithm to solve the weakly infeasible SDPs from [11].

Some geometry We remark that our positive gap SDPs give rise to sets with very interesting geometry. Define the map

$$S^n \times \mathbb{R} \ni (Y, y_0) \rightarrow f(Y, y_0) := (A_1 \bullet Y, \dots, A_m \bullet Y, B \bullet Y + y_0).$$

The closedness of the set $f(\mathcal{S}_+^n \times \mathbb{R}_+)$ plays an important role in SDP duality – see, e.g., Lemma 2 in [17]. For a set S let $\text{cl } S$ be its closure, and $\text{front } S := \text{cl } S \setminus S$ be its so called *frontier*. Using the A_i and B matrices of Example 1 a calculation (similar to the calculation in Example 6 in [9]) shows that

$$\begin{aligned} \text{cl } f(\mathcal{S}_+^n \times \mathbb{R}_+) &= \{(\alpha, \beta, \gamma) : \beta \geq \alpha \geq 0\}, \\ \text{front } f(\mathcal{S}_+^n \times \mathbb{R}_+) &= \{(0, \beta, \gamma) \mid -\beta > \gamma \geq 0\}. \end{aligned}$$

The set $f(\mathcal{S}_+^n \times \mathbb{R}_+)$ is shown on Figure 1 in blue, and its frontier in red.

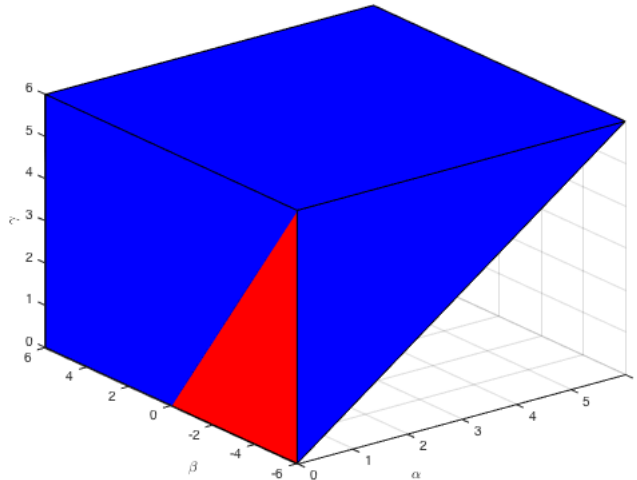


Figure 1: The set $f(\mathcal{S}_+^n \times \mathbb{R}_+)$ is in blue, and its frontier in red

1.1 Preliminaries

Reformulations The central definition of the paper is

Definition 1. We obtain an elementary reformulation, or reformulation of (P) - (D) by a sequence of the following operations:

- (1) Replace B by $B + \lambda A_j$, for some j and $\lambda \neq 0$.
- (2) Exchange (A_i, c_i) and (A_j, c_j) , where $i \neq j$.
- (3) Replace (A_i, c_i) by $\lambda(A_i, c_i) + \mu(A_j, c_j)$, where $\lambda \neq 0$.
- (4) Apply a rotation $T^T()T$ to all A_i and B , where T is an invertible matrix².

Note that operations (1)-(3) are indeed elementary row operations done on (D) . For example, operation (2) exchanges the constraints

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

Clearly, (P) and (D) attain their optimal values iff they do so after a reformulation, and reformulations also preserve duality gaps.

²We call the operation $T^T()T$ a rotation whenever T is invertible, even if it is not orthogonal.

Matrices We write \mathcal{S}^n , \mathcal{S}_+^n , and \mathcal{S}_{++}^n for the set of $n \times n$ symmetric, symmetric positive semidefinite (psd), and symmetric positive definite (pd) matrices, respectively.

For matrices A and B we let

$$A \oplus B := \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

We write $\mathcal{S}_+^r \oplus 0$ for the set of matrices whose upper left $r \times r$ block is psd, and the rest is zero. The dimension of the zero part will be clear from the context.

Cones, dual cones, and the Gordan-Stiemke theorem A set K is a *cone*, if $x \in K$, $\lambda \geq 0$ implies $\lambda x \in K$ and the dual cone of cone K is

$$K^* = \{y \mid \langle y, x \rangle \geq 0 \forall x \in K\}.$$

For brevity, we write

$$K^{*\setminus\perp} := K^* \setminus K^\perp. \tag{1.3}$$

Given an affine subspace H and a closed convex cone K such that $H \cap K \neq \emptyset$, the Gordan-Stiemke theorem states

$$H \cap \text{ri } K = \emptyset \Leftrightarrow H^\perp \cap K^{*\setminus\perp} \neq \emptyset, \tag{1.4}$$

where $\text{ri } K$ stands for the relative interior of K .

We call $Z \succeq 0$ a *slack matrix in (P)*, if $Z = B - \sum_{i=1}^m x_i A_i$ for some $x \in \mathbb{R}^m$. We make the following

Assumption 1. *Problem (P) is feasible, the A_i and B are linearly independent, B is of the form*

$$B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, \text{ where } 0 \leq r < n, \tag{1.5}$$

and it is the maximum rank slack in (P).

This assumption is easy to satisfy, at least theoretically. Suppose Z is a maximum rank slack in (P) and Q is a matrix of suitably scaled eigenvectors of Z . Replacing A_i by $Q^T A_i Q$ for all i and B by $T^T B T$ puts Z in the required form.

In all examples we will call the matrices on the left A_i and the right hand side B in the primal problem.

2 Theorem 1: proof of “Only if”

We now turn to the proof of the “Only if” direction in Theorem 1. The main idea is that (D) cannot be strictly feasible, otherwise the duality gap would be zero. We first make the lack of strict feasibility obvious by creating the constraint $(\Lambda \oplus 0) \bullet Y = 0$. This amounts to a step of a facial reduction algorithm [24, 16, 5]. Since in the two variable case we only need one step, we simply invoke the Gordan-Stiemke theorem once. Thus the proof is self-contained.

We need a basic lemma, whose proof is in Appendix A.

Lemma 1. *Let*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix},$$

where $A_{11} \in \mathcal{S}^{r_1}$, $A_{22} \in \mathcal{S}_+^{r_2}$.

Then there is an invertible matrix T such that

$$T^T A T = \left(\begin{array}{c|cc} \Sigma & 0 & W \\ \hline 0 & I_s & 0 \\ W^T & 0 & 0 \end{array} \right) \text{ and } T^T \begin{pmatrix} I_{r_1} & 0 \\ 0 & 0 \end{pmatrix} T = \begin{pmatrix} I_{r_1} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\Sigma \in \mathcal{S}^{r_1}$ is diagonal and $s \geq 0$.

□

Proof of "Only if" in Theorem 1 We reformulate (P) into (P_{ref}) in several steps, and we call the primal and dual problems (P) and (D) throughout the process. We call the constraint matrices on the left A'_1 and A'_2 throughout, starting with $A'_1 = A_1$ and $A'_2 = A_2$.

Case 1: (D) is feasible We break the proof into four parts: facial reduction step, transforming A'_1 , transforming A'_2 , and ensuring $c'_2 < 0$.

Facial reduction step Let

$$\begin{aligned} H &= \{ Y \mid A_i \bullet Y = c_i \forall i \} \\ &= \{ Y \mid A_i \bullet Y = 0 \forall i \} + Y_0, \end{aligned} \tag{2.6}$$

where $Y_0 \in H$ is arbitrary. Since (D) is not strictly feasible, by (1.4) there is $A'_1 \in H^\perp \cap (\mathcal{S}_+^n \setminus \{0\})$. Thus

$$\begin{aligned} A'_1 &= \lambda_1 A_1 + \lambda_2 A_2, \\ A'_1 \bullet Y_0 &= 0, \end{aligned}$$

hence

$$\lambda_1 c_1 + \lambda_2 c_2 = (\lambda_1 A_1 + \lambda_2 A_2) \bullet Y_0 = A'_1 \bullet Y_0 = 0,$$

so we can reformulate the feasible set of (D) as

$$\begin{aligned} A'_1 \bullet Y &= 0 \\ A'_2 \bullet Y &= c_2 \\ Y &\succeq 0 \end{aligned} \tag{2.7}$$

using only elementary row operations. These operations do not change B .

Transforming A'_1 Since $A'_1 \succeq 0$ and B is the maximum rank slack in (P) , the only nonzero components of A'_1 are in its upper left $r \times r$ block, otherwise $B - x_1 A'_1$ would be a slack with larger rank than r for $x_1 < 0$.

Let p be the rank of A'_1 , Q a matrix of length 1 eigenvectors of the upper left $r \times r$ block of A'_1 , set $T = Q \oplus I_{n-r}$, and apply the transformation $T^T()T$ to A'_1, A'_2 and B . After this A'_1 looks like

$$A'_1 = \left(\begin{array}{c|c|c|c} \Lambda & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \end{array} \right), \quad (2.8)$$

where $\Lambda \in \mathcal{S}_{++}^p$ is diagonal. From now on the double vertical and horizontal lines show the first r rows and columns in the relevant matrices. Also note that B is still in the same form (see Assumption 1).

Transforming A'_2 Let S be the lower $(n-r) \times (n-r)$ block of A'_2 . We claim that S cannot be indefinite, so suppose it is. Then the equation $S \bullet Y' = c'_2$ has a positive definite solution Y' . Then

$$Y := \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & Y' \end{array} \right). \quad (2.9)$$

is feasible in (D) with value 0, thus

$$0 \leq \text{val}(P) \leq \text{val}(D) \leq 0,$$

so the duality gap is zero, a contradiction.

We can now assume $S \succeq 0$ (if $S \preceq 0$, we multiply A'_2 and c'_2 by -1). Recall that Λ in (2.8) is $p \times p$, where $p \leq r$. Next we apply Lemma 1 with

$$\begin{aligned} A &:= \text{lower right } (n-p) \times (n-p) \text{ block of } A'_2, \\ r_1 &:= r-p, \\ r_2 &:= n-r. \end{aligned}$$

Let T be the invertible matrix supplied by Lemma 1, and apply the rotation $(I_p \oplus T)^T(I_p \oplus T)$ to A'_1, A'_2 and B . This operation keeps A'_1 as it was. It also keeps B as it was, since the rotation $T^T()T$ keeps $(I_{r-p} \oplus 0)$ the same. Now A'_2 looks like

$$A'_2 = \left(\begin{array}{c|c|c|c} \times & \times & \times & M \\ \hline \times & \Sigma & 0 & W \\ \hline \times & 0 & I_s & \\ \hline M^T & W^T & & \end{array} \right) \text{ for some } M \text{ and } W. \quad (2.10)$$

We claim that $W \neq 0$ or $M \neq 0$. Indeed if both were zero, then $B - x_2 A'_2$ would have larger rank than r for some $x_2 < 0$.

Thus $x_2 = 0$ in any feasible solution of (P) , hence $\text{val}(P) = 0$.

Next we claim $W = 0$, so suppose $W \neq 0$. Then we define

$$Y = \left(\begin{array}{c|c|c|c} & & & \\ \hline & \epsilon I & & * \\ \hline & & & \\ \hline & & * & \lambda I \end{array} \right),$$

where $\epsilon > 0$, we choose the “*” block so that $A'_2 \bullet Y = c'_2$, we choose λ to ensure $Y \succeq 0$, and the unspecified entries of Y are zero. Thus $B \bullet Y = (r-p)\epsilon$, so letting $\epsilon \rightarrow 0$ we deduce $\text{val}(D) = 0$, a contradiction.

Ensuring $c'_2 < 0$. We have $c'_2 \neq 0$, otherwise the duality gap would be zero. First, suppose $s > 0$ (s is the size of the identity block in A'_2) and to obtain a contradiction, assume $c'_2 > 0$. Then

$$Y := \begin{pmatrix} \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} & \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \\ \hline & \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \\ \hline & \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \\ \hline \end{pmatrix},$$

(where the unspecified entries are zero) is feasible in (D) with value 0, a contradiction. Next, suppose $s = 0$. If $c'_2 < 0$, then we are done; if $c'_2 > 0$, then we multiply A'_2 and c'_2 by -1 to ensure $c'_2 < 0$.

This completes the proof of Case 1.

Case 2: (D) is infeasible Since there is a positive duality gap, we have $\text{val}(P) < +\infty$.

Facial reduction step Consider the SDP

$$\begin{aligned} &\inf && -\lambda \\ &s.t. & A_i \bullet Y - \lambda c_i &= 0 \quad \forall i \\ & & Y &\succeq 0 \\ & & \lambda &\in \mathbb{R} \end{aligned} \tag{2.11}$$

and note that its optimal value is zero; if $\lambda > 0$ were feasible in it, then dividing Y by λ would give a feasible solution to (D) . We claim that

$$\nexists Y \succ 0 \text{ s.t. } (Y, \lambda) \text{ is feasible in (2.11) with some } \lambda, \tag{2.12}$$

so suppose that there is such a Y . We next construct an equivalent SDP in the standard dual form

$$\begin{aligned} &\inf && \bar{B} \bullet \bar{Y} \\ &s.t. & \bar{A}_i \bullet \bar{Y} &= 0 \quad \forall i \\ & & \bar{Y} &\in \mathcal{S}_+^{n+2}, \end{aligned} \tag{2.13}$$

where

$$\begin{aligned} \bar{A}_i &= A_i \oplus [-c_i] \oplus [c_i] \quad \forall i \\ \bar{B} &= [0] \oplus [-1] \oplus [1] \end{aligned}$$

(Note the free variable λ in (2.11) is split as $\lambda = \bar{y}_{n+1, n+1} - \bar{y}_{n+2, n+2}$). Since (2.13) is strictly feasible, and has optimal value 0, its dual is feasible, thus there is $\bar{x} \in \mathbb{R}^m$ s.t.

$$\begin{aligned} \sum_{i=1}^m \bar{x}_i A_i &\preceq 0 \\ \sum_{i=1}^m \bar{x}_i c_i &= 1. \end{aligned} \tag{2.14}$$

Adding $\lambda \bar{x}$ for a large $\lambda > 0$ to a feasible solution of (P) we deduce $\text{val}(P) = +\infty$, a contradiction.

The rest of the proof Up to now we proved (2.12), which means $\text{lin } H \cap \mathcal{S}_+^n = \emptyset$, where H is defined in (2.6). Since $(\text{lin } H)^\perp = H^\perp$, we next invoke the Gordan-Stiemke theorem (1.4) with $\text{lin } H$ in place of H and complete the proof as in Case 1. \square

The next corollary shows that the “worst” pathology (positive gap and unattained optimal values) does not happen when $m = 2$.

Corollary 1. *Suppose $m = 2$, (D) is feasible, and $\text{val}(P) < \text{val}(D)$. Then (P) attains its optimal value, and so does (D) if it is feasible.*

Proof Assume the conditions above hold and assume w.l.o.g. that we reformulated (P) into (P_{ref}) . Let (D_{ref}) be the dual of (P_{ref}) .

Clearly (P_{ref}) attains its optimal value, hence so does (P) .

If (D) is infeasible, then we are done, so assume it is feasible. Then so are (D_{ref}) and (1.2). In the latter Σ is diagonal, so (1.2) is just a linear program, which attains its optimal value. Hence so do (D_{ref}) and (D) . \square

In [17] we characterized when the semidefinite *system*

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

is *badly behaved*, meaning when there is $c \in \mathbb{R}^m$ such that (P) has a finite optimal value, but (D) has no solution with the same value. Hence it is natural to ask: when is (P_{SD}) “really” badly behaved, i.e., when is there $c \in \mathbb{R}^m$ such that the optimal values of (P) and (D) actually differ?

The following corollary of Theorem 1 answers this question when $m = 2$. It relies on *reformulating* (P_{SD}) , i.e., reformulating (P) with *some* $c \in \mathbb{R}^m$.

Corollary 2. *Suppose $m = 2$. Then there is (c_1, c_2) such that $\text{val}(P) < \text{val}(D)$ iff (P_{SD}) has a reformulation*

$$x_1 \left(\begin{array}{c|c|c|c} \Lambda & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \end{array} \right) + x_2 \left(\begin{array}{c|c|c|c} \times & \times & \times & M \\ \hline \times & \Sigma & & \\ \hline \times & & I_s & \\ \hline M^T & & & \end{array} \right) \preceq \left(\begin{array}{c|c|c|c} I_p & & & \\ \hline & I_{r-p} & & \\ \hline & & & \\ \hline & & & \end{array} \right),$$

where Λ and Σ are diagonal, Λ is positive definite, $M \neq 0$, and $s \geq 0$. \square

3 SDPs with similar structure in any dimension

We now show that the two variable case helps us understand positive gaps in larger SDPs: we present SDPs in any dimension in which the positive duality gap is certified by the same structure as in the two variable case.

First let us see a two variable example:

Example 2.

$$\begin{aligned} \sup \quad & -5x_2 \\ \text{s.t.} \quad & x_1 \left(\begin{array}{c|c|c|c} 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array} \right) + x_2 \left(\begin{array}{c|c|c|c} 0 & 0 & 0 & 2 \\ \hline 0 & -1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 2 & 0 & 0 & 0 \end{array} \right) \preceq \left(\begin{array}{c|c|c|c} 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array} \right). \end{aligned} \tag{3.15}$$

This SDP has zero value, while the value of its dual is 5.

The next example is a “larger version” of Example 2:

Example 3. Let us consider the SDP over \mathcal{S}_+^6

$$\begin{array}{l} \sup -5x_3 \\ \text{s.t. } x_1 \left(\begin{array}{c|c} 1 & \\ \hline & \end{array} \right) + x_2 \left(\begin{array}{c|c} 1 & 2 \\ \hline & 1 \end{array} \right) + x_3 \left(\begin{array}{c|c} & 2 \\ \hline -1 & 2 \\ & 1 \\ 2 & 2 \end{array} \right) \preceq \left(\begin{array}{c|c} 1 & \\ \hline & 1 \end{array} \right) \end{array}$$

and verify that there is a positive duality gap between it and its dual. The (6,6) element of all A_i and B is zero, but the (1,6) element of A_2 is nonzero, and the (2,6) element of A_3 is nonzero, so $x_2 = x_3 = 0$ in any feasible x .

On the other hand, suppose that $Y \succeq 0$ is feasible in the dual, then by $A_1 \bullet Y = 0$ the 1st row and column of Y is zero, and by $A_2 \bullet Y = 0$ the 2nd, and 4th rows and columns of Y are zero. So a feasible dual Y looks like

$$Y = \left(\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & \times & \times \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & \times & \times \\ 0 & 0 & \times & 0 & \times & \times \end{array} \right), \quad (3.16)$$

where the entries marked by \times may be nonzero. So the dual is equivalent to

$$\begin{array}{l} \inf \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \bullet Y' \\ \text{s.t.} \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \bullet Y' = -5, \\ \quad \quad \quad Y' \succeq 0, \end{array} \quad (3.17)$$

whose optimal value is 5.

To proceed, we need some notation.

Notation 1. For $Y \in \mathcal{S}^n$ and $\mathcal{I}, \mathcal{J} \subseteq \{1, \dots, n\}$ we define $Y(\mathcal{I}, \mathcal{J})$ as the submatrix of Y with rows in \mathcal{I} and columns in \mathcal{J} , and let

$$\begin{aligned} Y(\mathcal{I}) &:= Y(\mathcal{I}, \mathcal{I}), \\ Y(\mathcal{I}, :) &:= Y(\mathcal{I}, \{1, \dots, n\}). \end{aligned}$$

Given sets $\mathcal{I}_1, \dots, \mathcal{I}_k$ and $i \leq j$ we denote

$$\mathcal{I}_{i:j} := \mathcal{I}_i \cup \dots \cup \mathcal{I}_j.$$

We next state an algorithm, that, as we will prove, generates an SDP instance with the required properties.

Algorithm 1: Positive gap SDP

Let n , r , and m be integers with $\min\{r, n-r\} \geq m \geq 2$ and $g > 0$.

Let $c = -g \cdot e_m$, $B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

Partition $\{1, \dots, r\}$ into nonempty blocks $\mathcal{J}_1, \dots, \mathcal{J}_m$ and let $\mathcal{J}_{m+1} := \emptyset$.

Partition $\{r+1, \dots, n\}$ into nonempty blocks $\mathcal{I}_2, \dots, \mathcal{I}_{m+1}$ and let $\mathcal{I}_1 := \emptyset$.

Let $A_1(\mathcal{J}_1) = I$.

for $i = 2 : m$ **do**

Let $A_i(\mathcal{J}_i \cup \mathcal{I}_i) = I$.

Let $A_i(\mathcal{J}_{i-1}, \mathcal{I}_{m+1}) \neq 0$ and $A_i(\mathcal{I}_{i-1}, \mathcal{I}_{m+1}) \neq 0$.

end for

Let $A_m(\mathcal{J}_m) = -I$.

We first explain Algorithm 1, then analyze its output. First note that when $i = 2$ it skips setting $A_2(\mathcal{I}_1, \mathcal{I}_{m+1}) \neq 0$, since $\mathcal{I}_1 = \emptyset$. It also overwrites $A_m(\mathcal{J}_m)$ in the last step; we stated the algorithm in this manner to make it reasonably concise and clear.

Algorithm 1 can generate the following SDPs:

- The SDP of Example 2 with $n = 4$, $m = r = 2$, $\mathcal{J}_1 = \{1\}$, $\mathcal{J}_2 = \{2\}$, $\mathcal{I}_2 = \{3\}$, $\mathcal{I}_3 = \{4\}$.
- The SDP of Example 3 with $n = 6$, $m = r = 3$, $\mathcal{J}_1 = \{1\}$, $\mathcal{J}_2 = \{2\}$, $\mathcal{J}_3 = \{3\}$, $\mathcal{I}_2 = \{4\}$, $\mathcal{I}_3 = \{5\}$, $\mathcal{I}_4 = \{6\}$.

The algorithm sets some entries as nonzero in the “for” loop, and these are all ‘2’ in the above examples.

If in Algorithm 1 we set $\mathcal{I}_m := \emptyset$, then it still generates an SDP with a positive gap, as the proof of Theorem 2 (stated below) goes through with a minor change. This modified algorithm generates the SDP of Example 1 by choosing $m = 2$, $n = 3$, $\mathcal{J}_1 = \{1\}$, $\mathcal{J}_2 = \{2\}$, $\mathcal{I}_2 = \emptyset$, $\mathcal{I}_3 = \{3\}$.

Figure 2 shows the sparsity structure of the A_i and of the B in the instances generated by Algorithm 1 when $m = 2, 3$, and 4, and assuming that each index set is a singleton.

Theorem 2. *Suppose that (P) and (D) are output by Algorithm 1. Then*

$$0 = \text{val}(P) < \text{val}(D) = g.$$

In particular, (D) is equivalent to

$$\begin{aligned} \inf \quad & \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \bullet Y' \\ \text{s.t.} \quad & \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \bullet Y' = -g, \\ & Y' \succeq 0, \end{aligned} \tag{3.18}$$

which has optimal value g .

Proof Let x be feasible in (P). Since

$$A_i(\mathcal{J}_{i-1}, \mathcal{I}_{m+1}) \neq 0 \text{ and } A_i(\mathcal{I}_{i-1}, \mathcal{I}_{m+1}) \neq 0 \text{ for all } i \geq 2, \tag{3.19}$$

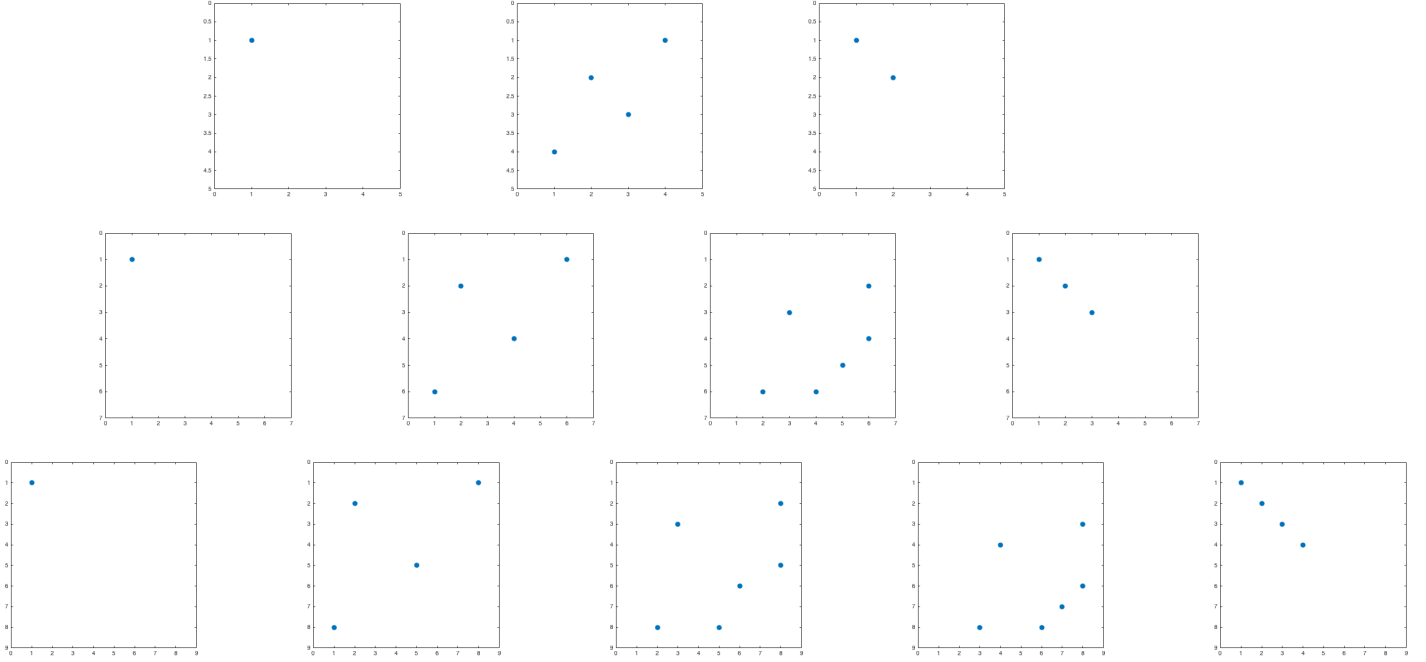


Figure 2: The sparsity structure of the instances with $m = 2$, $m = 3$ and $m = 4$

and the diagonal block corresponding to \mathcal{I}_{m+1} in all matrices is zero, we get $x_2 = \dots = x_m = 0$, and this proves $\text{val}(P) = 0$.

Next, let $Y \succeq 0$ and $1 \leq i \leq m - 1$. We claim that

$$A_1 \bullet Y = \dots = A_i \bullet Y = 0 \Rightarrow Y(\mathcal{J}_{1:i} \cup \mathcal{I}_{1:i}, :) = 0. \quad (3.20)$$

If $i = 1$ then (3.20) follows from $A_1(\mathcal{J}_1) \succ 0$.

Suppose $1 \leq i + 1 \leq m - 1$ and (3.20) is true for $1, \dots, i$. Since $A_{i+1}(\mathcal{J}_{i+1} \cup \mathcal{I}_{i+1}) \succ 0$ and all other nonzero entries of A_{i+1} are in the rows and columns indexed by $\mathcal{J}_{1:i} \cup \mathcal{I}_{1:i}$ (which are all zero in Y) the statement follows for $i + 1$ as well.

We next show A_m and Y that satisfies (3.20) for $i = 1, \dots, m - 1$. The \times blocks in Y may be nonzero, and the values of the elements in the \times blocks in A_m do not matter when calculating $A_m \bullet Y$.

$$A_m = \begin{pmatrix} \overbrace{\times}^{\mathcal{J}_{1:(m-1)}} & \overbrace{\times}^{\mathcal{J}_m} & \overbrace{\times}^{\mathcal{I}_{1:(m-1)}} & \overbrace{\times}^{\mathcal{I}_m} & \overbrace{\times}^{\mathcal{I}_{m+1}} \\ \times & -I & \times & 0 & 0 \\ \times & \times & \times & \times & \times \\ \times & 0 & \times & I & 0 \\ \times & 0 & \times & 0 & 0 \end{pmatrix}, Y = \begin{pmatrix} \overbrace{0}^{\mathcal{J}_{1:(m-1)}} & \overbrace{0}^{\mathcal{J}_m} & \overbrace{0}^{\mathcal{I}_{1:(m-1)}} & \overbrace{0}^{\mathcal{I}_m} & \overbrace{0}^{\mathcal{I}_{m+1}} \\ 0 & \times & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \times & 0 & \times & \times \\ 0 & \times & 0 & \times & \times \end{pmatrix}. \quad (3.21)$$

Thus we can set the nonzero minor of Y to make Y feasible in (D) . Hence (D) is equivalent to (3.18) and $\text{val}(D) = g$ follows. \square

4 The singularity degree of the duals of our positive gap SDPs

We now study our positive gap SDPs in more depth. We introduce an important parameter of SDPs, called the *singularity degree*, and show that the duals associated with our SDPs, namely (D) and (HD) , have singularity degree equal to $m - 1$ and m , respectively.

We first need some background.

We first need to define faces, facial reduction sequences, and the singularity degree. For more extensive treatments of these concepts, see e.g., [9].

Definition 2. *Given a closed convex cone K , a convex subset F of K is a face of K , if $x, y \in K$, $\frac{1}{2}(x + y) \in F$ implies $x, y \in F$.*

If H is an affine subspace with $H \cap K \neq \emptyset$, then the minimal cone of $H \cap K$ is the smallest face of K that contains $H \cap K$.

The minimal cone of (P) , (D) , and of (HD) is defined as the minimal cone of their feasible sets³.

We are mainly interested in the faces of \mathcal{S}_+^n , which have a simple and attractive description: they are

$$F = \left\{ T \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} T^T : X \in \mathcal{S}_+^r \right\}, \text{ with dual cone } F^* = \left\{ T^{-T} \begin{pmatrix} X & Z \\ Z^T & Y \end{pmatrix} T^{-1} : X \in \mathcal{S}_+^r \right\}, \quad (4.22)$$

where $0 \leq r \leq n$ and $T \in \mathbb{R}^{n \times n}$ is invertible (see, e.g., [15]).

A straightforward argument shows that if $H \subseteq \mathcal{S}^n$, then the minimal cone of $H \cap \mathcal{S}_+^n$ is the minimal face of \mathcal{S}_+^n that contains the maximum rank psd matrix of H .

The next definition is about constructing the minimal cone.

Definition 3. *Let $k \geq 1$ be an integer. The set of length k facial reduction sequences for K is*

$$\text{FR}_k(K) = \{ (y_1, \dots, y_k) : y_1 \in K^*, y_i \in (K \cap y_1^\perp \cap \dots \cap y_{i-1}^\perp)^* \text{ for } i = 2, \dots, k \}.$$

We say that such a sequence is strict, if $y_1 \in K^{\perp\perp}$ and $y_i \in (K \cap y_1^\perp \cap \dots \cap y_{i-1}^\perp)^{*\perp\perp}$ for all $i \geq 2$.*

In $\text{FR}_k(K)$ we drop the subscript when its value is clear from the context, or irrelevant.

Suppose H is an affine subspace with $H \cap K \neq \emptyset$. The singularity degree $d(H \cap K)$ of $H \cap K$ is the smallest k such that there is $(y_1, \dots, y_k) \in \text{FR}(K)$ with $y_i \in H^\perp$ for all i and

$$K \cap y_1^\perp \cdots \cap y_k^\perp = \text{the minimal cone of } H \cap K.$$

We say that such a (y_1, \dots, y_k) defines the minimal cone of $H \cap K$.

We can always construct a facial reduction sequence that defines the minimal cone by using a suitable facial reduction algorithm [5, 24, 16]; this can also be seen directly by iterating the Gordan-Stiemke theorem (1.4). Thus the singularity degree is well defined.

The singularity degree of SDPs was introduced in the seminal paper [21]. It was used to bound the distance of a symmetric matrix from $H \cap \mathcal{S}_+^n$, given the distances from H and from \mathcal{S}_+^n . More recently it was used in [6] to bound the rate of convergence of the alternating projection algorithm to such a set.

³The minimal cone of (P) is the minimal cone of $(\text{lin}\{A_1, \dots, A_m\} + B) \cap \mathcal{S}_+^n$.

Example 4. Let H be the linear subspace spanned by the matrices

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

and $K = \mathcal{S}_+^3$.

Then the minimal cone of $H \cap K$ is the set of nonnegative multiples of A_1 , and the sequence

$$Y_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, Y_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

defines this minimal cone. Since Y_1 is the maximum rank psd matrix in H^\perp , it is the best choice for the first member of such a facial reduction sequence, hence $d(H \cap \mathcal{S}_+^3) = 2$.

Example 5. (Example 2 continued) In this example the maximum rank feasible matrix in (D) and minimal cone of (D) are

$$Y = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } 0 \oplus \mathcal{S}_+^3, \text{ respectively,}$$

and the one element facial reduction sequence (A_1) defines this minimal cone.

Suppose next that Y is feasible in (HD) . Then by $B \bullet Y = 0$ the first two rows and columns of Y are zero, and by $A_2 \bullet Y = 0$ so is the third row and column. Thus the maximum rank matrix and the minimal cone of (HD) are

$$Y = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } 0 \oplus \mathcal{S}_+^1, \text{ respectively,}$$

and the sequence (B, A_2) defines the minimal cone.

Example 6. (Example 3 continued) Recall that any feasible matrix in the dual looks like in equation (3.16). Hence the minimal cone of (D) is

$$\{Y \succeq 0 \mid y_{ij} = y_{ji} = 0 \text{ for } i = 1, 2, 4 \text{ and all } j\},$$

and (A_1, A_2) defines this minimal cone. We can similarly argue that the minimal cone of (HD) is

$$0 \oplus \mathcal{S}_+^1,$$

and the sequence (B, A_2, A_3) defines this minimal cone.

Next, Lemma 2 explains the seemingly odd structure of the constraint matrices in the output of Algorithm 1: these matrices form facial reduction sequences.

Lemma 2. *Suppose that (P) and (D) are output by Algorithm 1. Then*

$$(A_1, \dots, A_{m-1}) \in \text{FR}_{m-1}(\mathcal{S}_+^n) \quad (4.23)$$

and

$$(B, A_2, \dots, A_m) \in \text{FR}_m(\mathcal{S}_+^n). \quad (4.24)$$

Proof Let $1 \leq i \leq m-2$, then (3.20) implies

$$F_i := \mathcal{S}_+^n \cap A_1^\perp \cap \dots \cap A_i^\perp = \{Y \succeq 0 \mid Y(\mathcal{J}_{1:i} \cup \mathcal{I}_{1:i}) = 0\}, \quad (4.25)$$

hence $A_{i+1} \in F_i^*$, as required.

The proof of (4.24) is analogous. □

It is interesting how we can change the facial reduction sequence in (4.23) to get the sequence in (4.24). First, we can replace A_1 by B , since the rangespace of B contains the rangespace of A_1 . Second, we can append A_m to the end of the sequence, even though A_m has a $-I$ diagonal block, since any matrix in $\mathcal{S}_+^n \cap B^\perp \cap A_2^\perp \cap \dots \cap A_{m-1}^\perp$ has zeros in its first r rows and columns. However, it is not hard to see that, in general, $(A_1, \dots, A_{m-1}, A_m)$ is *not* a facial reduction sequence.

The reader may wonder, why we connect positive gaps to the singularity degree of (D) and of (HD) , and not to the singularity degree of (P) . We could indeed do the latter, as we could exchange the roles of the primal and dual. However, we think that our treatment is more intuitive. To define the minimal cone of (D) we use a facial reduction sequence whose members are in $\text{lin}\{A_1, \dots, A_m\}$, and, as Lemma 2 and Theorem 3 show, these members are actually constraint matrices in (P) . Similarly, to define the minimal cone of (HD) we use a facial reduction sequence whose members are in $\text{lin}\{A_1, \dots, A_m, B\}$, and these members are also constraint matrices in (P) .

Theorem 3. *Suppose that (P) and (D) are output by Algorithm 1. Then*

$$d(D) = m - 1 \text{ and } d(HD) = m.$$

Proof Recalling equation (4.25), the proof of Theorem 2 implies that the minimal cone of (D) is contained in F_{m-1} . Since

$$Y = \begin{pmatrix} \overbrace{\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{matrix}}^{\mathcal{J}_{1:(m-1)}} & \overbrace{\begin{matrix} 0 \\ \lambda I \\ 0 \\ 0 \\ 0 \end{matrix}}^{\mathcal{J}_m} & \overbrace{\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{matrix}}^{\mathcal{I}_{1:(m-1)}} & \overbrace{\begin{matrix} 0 \\ 0 \\ 0 \\ I \\ 0 \end{matrix}}^{\mathcal{I}_m} & \overbrace{\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ I \end{matrix}}^{\mathcal{I}_{m+1}} \end{pmatrix}, \quad (4.26)$$

is feasible in (D) , where $\lambda > 0$ is suitably chosen, the minimal cone is exactly F_{m-1} . Taking into account (4.23), $d(D) \leq m - 1$ follows.

To prove $d(D) = m - 1$ we need to show that any facial reduction sequence in $\text{lin}\{A_1, \dots, A_{m-1}\}$ can reduce \mathcal{S}_+^n by at most as much as A_1, \dots, A_{m-1} . We prove this in the following claim.

Claim 1. Suppose that $(Y_1, \dots, Y_\ell) \in \text{FR}(\mathcal{S}_+^n)$ is strict, with $Y_i \in \text{lin}\{A_1, \dots, A_m\}$ for all i . Then

$$\ell < m$$

and

$$F_\ell = \mathcal{S}_+^n \cap Y_1^\perp \cap \dots \cap Y_\ell^\perp \quad (4.27)$$

for $i = 1, \dots, \ell$.

We now continue the proof of Theorem 3 and look at the singularity degree of (HD) . Suppose $Y \succeq 0$ is feasible in (HD) . Then $B \bullet Y = 0$ implies that the first r rows and columns of Y are zero; $A_2 \bullet Y$ implies that the rows and columns corresponding to \mathcal{I}_2 are zero, and so on. So Y looks like

$$Y = \begin{pmatrix} \overbrace{\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \oplus \end{matrix}}^{\mathcal{J}_{1:m}} & \overbrace{\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}}^{\mathcal{I}_{2:m}} & \overbrace{\begin{matrix} 0 \\ 0 \\ \oplus \end{matrix}}^{\mathcal{I}_{m+1}} \end{pmatrix},$$

where the \oplus symbol denotes a psd submatrix. Clearly there is a Y feasible in (HD) with $Y(\mathcal{I}_{m+1}) \succ 0$, thus (B, A_2, \dots, A_m) defines the minimal cone of (HD) , so $d(HD) \leq m$.

Equality follows from the claim below.

Claim 2. Suppose $Y_1, \dots, Y_\ell \in \text{FR}(\mathcal{S}_+^n)$ is strict, with $\ell \geq 2$ and $Y_i \in \text{lin}\{B, A_2, \dots, A_m\}$ for all i . Then

$$\mathcal{S}_+^n \cap Y_1^\perp \cap \dots \cap Y_\ell^\perp = \mathcal{S}_+^n \cap B^\perp \cap A_2^\perp \cap \dots \cap A_\ell^\perp \quad (4.28)$$

for $i = 2, \dots, \ell$.

□

The proof of Claim 2 is almost the same as the proof of Claim 1, hence we omit it. Given Claim 2 the proof of Theorem 3 is complete. □

5 The double reformulation

In this section we prove an auxiliary result, which we believe to be of independent interest (hence the separate section): we show how to reformulate (P) so its maximum rank slack and the maximum rank solution of (HD) both become easy to see.

First we define certain structured facial reduction sequences for \mathcal{S}_+^n which will be useful for this purpose. These sequences were originally introduced in [9].

Definition 4. We say that (M_1, \dots, M_k) is a regularized facial reduction sequence for \mathcal{S}_+^n if

$$M_i = \begin{pmatrix} \overbrace{\begin{matrix} \times & \times & \times \\ \times & I & 0 \\ \times & 0 & 0 \end{matrix}}^{r_1 + \dots + r_{i-1}} & \overbrace{\begin{matrix} \times \\ I \\ 0 \end{matrix}}^{r_i} & \overbrace{\begin{matrix} \times \\ 0 \\ 0 \end{matrix}}^{n - r_1 - \dots - r_i} \end{pmatrix}$$

for all i , where the r_i are nonnegative integers, and the \times symbols correspond to blocks with arbitrary elements.

We denote the set of such sequences by $\text{REGFR}_k(\mathcal{S}_+^n)$, dropping the subscript whenever possible.

Sometimes we will refer to such a sequence by the length r_i blocks

$$\begin{aligned}\mathcal{I}_1 &:= \{1, \dots, r_1\}, \\ \mathcal{I}_2 &:= \{r_1 + 1, \dots, r_1 + r_2\}, \\ &\vdots \\ \mathcal{I}_k &:= \left\{ \sum_{i=1}^{k-1} r_i + 1, \dots, \sum_{i=1}^k r_i \right\}.\end{aligned}$$

For instance, (A_1, A_2) in Example 4 is a regularized facial reduction sequence; so is (B, A_2) in Example 2; and so is (B, A_2, A_3) in Example 3. In fact, if an SDP is generated by Algorithm 1, then $(B, A_2, \dots, A_{m-1}) \in \text{REGFR}(\mathcal{S}_+^n)$.

The usefulness of such sequences comes from the following proposition.

Proposition 1. *Suppose $(M_1, \dots, M_k) \in \text{REGFR}(\mathcal{S}_+^n)$ with block sizes r_1, \dots, r_k and $Y \succeq 0$ satisfies*

$$M_1 \bullet Y = \dots = M_k \bullet Y = 0.$$

Then the first $r_1 + \dots + r_k$ rows and columns of Y are zero.

Proof Since $Y \succeq 0$ and $M_1 \bullet Y = 0$, the upper left $r_1 \times r_1$ block of Y is zero, hence the first r_1 rows and columns of Y are zero; by $M_2 \bullet Y = 0$ the next r_2 rows and columns of Y are zero, and so on. \square

Definition 5. *We say that*

$$\begin{aligned}\sup \sum_{i=1}^m c'_i x_i \\ \sum_{i=1}^m x_i A'_i \preceq B,\end{aligned} \tag{P'}$$

is a double reformulation of (P) if

- (1) B is the maximum rank slack in (P) and in (P') and
- (2) $(B, A'_1, \dots, A'_{d(HD)-1}) \in \text{REGFR}(\mathcal{S}_+^n)$ defines the minimal cone of (HD).

Theorem 4. *We have $d(HD) \leq m + 1$ and any SDP of the form (P) has a double reformulation. \square*

Before proving Theorem 4 we explain why it is interesting, and illustrate it. Recall that the minimal cone of (P) is the minimal cone of

$$(\text{lin}\{A_1, \dots, A_m\} + B) \cap \mathcal{S}_+^n.$$

To define this minimal cone, we use a facial reduction sequence whose members are orthogonal to all A_i and to B , hence its first member is feasible in (HD). To make the sequence short, it is of course preferable that this first member have maximum rank, and it is useful to have a certificate that it has maximum rank.

Now, to see what the maximum rank solution of (HD) is, we use the double reformulation of (P) as follows: if the block sizes in $(B, A'_1, \dots, A'_{d(HD)-1})$ are r_i for all i , then by Proposition 1 the first $\sum_i r_i$ rows and columns of Y must be zero.

Let us see an illustration:

Example 7. Consider the SDP

$$\begin{aligned} & \sup \quad 13x_1 - 5x_2 \\ & \text{s.t.} \quad x_1 \begin{pmatrix} 0 & 2 & 2 \\ & 1 & \\ 2 & & 0 \\ 2 & & & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & & & \\ & 0 & 2 & \\ & & 1 & \\ 2 & & & 0 \end{pmatrix} \succeq \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}, \end{aligned} \quad (5.29)$$

and let A_1 and A_2 denote the matrices on the left hand side, and B the right hand side. Then

- B is the maximum rank slack and
- $(B, A_1, A_2) \in \text{REGFR}(\mathcal{S}_+^4)$ with block sizes 1 defines the minimal cone of (HD) , which is the set of nonnegative multiples of

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

Thus $d(HD) \leq 3$ and we will see later that actually $d(HD) = 3$ holds.

Remark 5.1. We can add the terms

$$+x_3 \begin{pmatrix} 0 & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & 1 \\ & & 1 & 0 \end{pmatrix}$$

to the right hand side of (5.29), and it is still in a double reformulation form.

Proof of Theorem 4 For brevity, let

$$d := d(HD) \text{ and } L := \text{lin} \{B, A_1, \dots, A_m\}.$$

We first show

$$d \leq m + 1.$$

Indeed, the feasible set of (HD) is $L^\perp \cap \mathcal{S}_+^n$ hence Theorem 1 in [9] implies $d \leq \dim L = m + 1$, as required.

We next prove that

$$B \text{ is the maximum rank matrix in } L \cap \mathcal{S}_+^n. \quad (5.30)$$

To do so, assume that

$$B' = \sum_{j=1}^m \lambda_j A_j + \lambda B \succeq 0$$

has larger rank, where the λ_i and λ are suitable scalars. Then B' has a nonzero element outside its upper $r \times r$ block. Let $\epsilon > 0$ be such that $|\lambda|\epsilon < 1$, then

$$B'' := \frac{1}{1 + \lambda\epsilon} (B + \epsilon B')$$

is a slack in (P) with rank larger than r , a contradiction. Thus (5.30) follows.

Next we prove that there is A'_1, \dots, A'_{d-1} such that

$$(B, A'_1, \dots, A'_{d-1}) \in \text{FR}(\mathcal{S}_+^n) \text{ defines the minimal cone of } (HD), \text{ and} \quad (5.31)$$

$$A'_1, \dots, A'_m \in \text{lin}\{A_1, \dots, A_m\}. \quad (5.32)$$

Indeed, by definition, there is

$$(A'_0, A'_1, \dots, A'_{d-1}) \in \text{FR}(\mathcal{S}_+^n) \text{ which defines the minimal cone of } (HD), \text{ and}$$

$$A'_0, A'_1, \dots, A'_m \in L.$$

By equation (5.30) we can assume $A'_0 = B$, so (5.31) follows. Next, to ensure (5.32) let $i \geq 1$ and write $A'_i = \sum_{j=1}^k \lambda_j A_j + \lambda B$ for some λ_j and λ scalars. By definition, we have

$$A'_i \in (\mathcal{S}_+^n \cap B^\perp \cap A_1^\perp \cap \dots \cap A_{i-1}^\perp)^* \quad (5.33)$$

thus subtracting λB from A'_i maintains property (5.33). Doing this for $i = 1, 2, \dots, d-1$ we ensure (5.32).

Next we note that $(B, A'_1, \dots, A'_{d-1})$ are strict, hence by Theorem 1 in [9] they are linearly independent. Thus we can reformulate (P) using only operations (2) and (3) in Definition 1 to replace A_i by A'_i for $i = 1, \dots, m$, where A'_1, \dots, A'_{d-1} are as in (5.31) and A'_d, \dots, A'_m are suitable matrices.

Finally, by Lemma 2 in [9] there is an invertible matrix T of order n such that

$$(T^T B T, T^T A'_1 T, \dots, T^T A'_{d-1} T) \in \text{REGFR}(\mathcal{S}_+^n).$$

We apply this rotation and this completes the proof. □

6 The case of maximal singularity degree

We now look at the case when (D) or (HD) have maximal singularity degree (m or $m+1$, respectively) and prove that in these cases there is no duality gap.

The first result is fairly straightforward.

Proposition 2. *We have*

$$d(D) \leq m.$$

Furthermore, when $d(D) = m$ there is no duality gap.

Proof Let

$$\begin{aligned} H &= \{Y \in \mathcal{S}^n \mid A_i \bullet Y = c_i \forall i\} \\ &= \{Y \in \mathcal{S}^n \mid A_i \bullet Y = 0 \forall i\} + Y_0, \end{aligned}$$

where $Y_0 \in H$ is arbitrary. Then the feasible set of (D) is $H \cap \mathcal{S}_+^n$, so Theorem 1 in [9] implies

$$d(D) \leq \dim H^\perp \leq m.$$

Equality holds throughout exactly when $A_i \bullet Y_0 = 0$ for all i , i.e., when $c = 0$, in which case there is no gap.

□

The main result of this section follows.

Theorem 5. *Suppose $d(HD) = m + 1$. Then (D) is strictly feasible, and*

$$\text{val}(P) = \text{val}(D) = 0.$$

Proof Assume $d(HD) = m + 1$ and assume w.l.o.g. that we reformulated (P) into (P') as given in Theorem 4.

Assume that (B, A'_1, \dots, A'_m) is associated with index sets $\mathcal{I}_0, \mathcal{I}_1, \dots, \mathcal{I}_m$ (see Definition 3). For convenience set

$$A'_0 = B, \mathcal{I}_{m+1} = \{1, \dots, n\} \setminus \mathcal{I}_{1:m}.$$

For brevity, let us write \mathcal{I}_j for $\mathcal{I}_{j:(m+1)}$ for all j (we can do this without confusion, since $m + 1$ is the largest index).

We first prove

$$A'_i(\mathcal{I}_{i-1}, \mathcal{I}_{(i+1):}) \neq 0 \text{ for } i = 1, \dots, m. \quad (6.34)$$

Let $i \geq 1$ and suppose to the contrary that $A'_i(\mathcal{I}_{i-1}, \mathcal{I}_{(i+1):}) = 0$.

Then A'_{i-1} and A'_i look like

$$A'_{i-1} = \begin{pmatrix} \overbrace{\times}^{\mathcal{I}_{0:(i-2)}} & \overbrace{\times}^{\mathcal{I}_{i-1}} & \overbrace{\times}^{\mathcal{I}_i} & \overbrace{\times}^{\mathcal{I}_{(i+1):}} \\ \times & I & \times & \times \\ \times & & & \\ \times & & & \\ \times & & & \end{pmatrix}, \quad A'_i = \begin{pmatrix} \overbrace{\times}^{\mathcal{I}_{0:(i-2)}} & \overbrace{\times}^{\mathcal{I}_{i-1}} & \overbrace{\times}^{\mathcal{I}_i} & \overbrace{\times}^{\mathcal{I}_{(i+1):}} \\ \times & \times & \times & \times \\ \times & \times & \times & \\ \times & \times & I & \\ \times & & & \end{pmatrix},$$

where the unspecified entries are zero. Let $Y := \lambda A'_{i-1} + A'_i$ for some large $\lambda > 0$. Then by the Schur complement condition for positive definiteness we deduce

$$Y(\mathcal{I}_{(i-1):i}) \succ 0,$$

hence $A'_1, \dots, A'_{i-2}, Y, A'_{i+1}, \dots, A'_m$ is a shorter facial reduction sequence, which also defines the minimal cone of (HD) , a contradiction. We thus proved (6.34).

Next, let x be feasible in (P) ; we will prove

$$x = 0.$$

For that purpose, let

$$Z = B - \sum_{i=1}^m x_i A'_i.$$

Since $Z(\mathcal{I}_{m+1}) = 0$ and $A'_m(\mathcal{I}_{m-1}, \mathcal{I}_{m+1}) \neq 0$, we deduce $x_m = 0$.

Thus $Z(\mathcal{I}_{m:(m+1)}) = 0$. Since $A'_{m-1}(\mathcal{I}_{m-2}, \mathcal{I}_{m:(m+1)}) \neq 0$, we deduce $x_{m-1} = 0$ and so on. This proves $x = 0$.

The following algorithm constructs $Y \succ 0$ feasible for (D') , the dual of (P') . We call the algorithm a Staircase Algorithm, since it fills in the entries of Y in a staircase pattern.

Algorithm 2: Staircase Algorithm

Set $Y(\mathcal{I}_0) \succ 0$.
Set $Y(\mathcal{I}_1) \succ 0$.
for $i = 1 : m$ **do**
 (*) Set $Y(\mathcal{I}_{i-1}, \mathcal{I}_{i+1})$ to satisfy the i th equation in (D') .
 (**) Set $Y(\mathcal{I}_{i+1}) \succ 0$ to make $Y(\mathcal{I}_{0:i+1}) \succ 0$ while keeping
 all previously satisfied equations satisfied.
end for

Figure 6 illustrates how the Staircase Algorithm works when $m = 2$. It shows the order in which the entries of Y are defined, and the type of entries they are filled in with: first we set $Y(\mathcal{I}_0) \succ 0$; second, $Y(\mathcal{I}_1) \succ 0$; third, we set $Y(\mathcal{I}_0, \mathcal{I}_{2:3})$ to satisfy the constraint $A_1 \bullet Y = c'_1$; fourth, we set $Y(\mathcal{I}_2) \succ 0$ to ensure $Y(\mathcal{I}_{0:2}) \succ 0$, and so on.

$$Y = \begin{array}{|c|c|c|c|} \hline \textcircled{1, +} & & & \textcircled{3, c'_1} \\ \hline & \textcircled{2, +} & & \textcircled{5, c'_2} \\ \hline & & \textcircled{4, +} & \\ \hline & & & \textcircled{6, +} \\ \hline \end{array}$$

Figure 3: How the Staircase Algorithm works when $m = 2$.

We now argue that the Staircase Algorithm is correct. Clearly, steps (*) can be executed because of condition (6.34). Steps (**) can be executed with $i = 1, \dots, m-1$ because of the Schur complement condition for positive definiteness; since $A'_1(\mathcal{I}_{i+1}) = \dots = A'_i(\mathcal{I}_{i+1}) = 0$ it follows that the previously satisfied equations $1, \dots, i$ are not affected.

It remains to show that step (**) can be executed when $i = m$. Indeed, note that there is Y' feasible in (HD) such that

$$Y'(\mathcal{I}_{m+1}) \succ 0$$

and the other entries of Y' are zero. We add a large multiple of Y' to our current Y and this ensures $Y \succ 0$.

The proof is now complete. □

Example 8. (Example 7 continued) The blocks specified in equation (6.34) have entries equal to 2 in this example. Let us illustrate why they must be nonzero, assuming $d(HD) = 3$, so suppose we zero out the '2' elements in A_1 . Then

$$B + A_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & & & \\ & 0 & 2 & \\ & & 1 & \\ & 2 & & 0 \end{pmatrix},$$

so $(B + A_1, A_2)$ is a shorter facial reduction sequence which defines the same minimal cone. Similarly, if we zero out the '2' entries in A_2 then $(B, A_1 + A_2)$ becomes a shorter facial reduction sequence which defines the same minimal cone.

A possible Y constructed by the Staircase Algorithm is

$$Y = (y_{ij})_{i,j=1}^4 = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & -2 \\ 2 & 0 & 5 & 0 \\ 1 & -2 & 0 & 25 \end{pmatrix}$$

as follows:

- (1) set $y_{11} = 1$;
- (2) set $y_{22} = 1$;
- (3) set $y_{13} = y_{31} = 2, y_{14} = y_{41} = 1$;
- (4) set $y_{33} = 5$;
- (5) set $y_{24} = y_{42} = -2$;
- (6) set $y_{44} = 25$.

7 A computational study

Using Algorithm 1 we generated a library of SDPs with positive duality gaps, The problem generators are given three parameters: m , the number of variables; g , the duality gap; and bs , the blocksize. For our instances we chose $g = 10$.

Half of our instances are *clean*. These are straight outputs of Algorithm 1 and they are named as

`gap_clean_m_n_bs.mat`

where

- m is the number of variables;
- bs is the size of the \mathcal{J}_i and \mathcal{I}_i blocks in Algorithm 1;
- $n = 2 \cdot m \cdot bs$ is the order of the matrices (so the naming is a bit redundant, but we supply n anyway, for the sake of clarity).

The other half of our instances are *rotated*. These are obtained from the clean instances by applying a rotation $T^T()T$ to all A_i and B , where T is an invertible matrix with all integer entries, and condition number bounded by 150. They are named

`gap_rotated_m_n_bs.mat`.

The instances are stored in Sedumi format [22], in which the roles of c and B are interchanged. Each clean instance is given by

- A , which is a matrix with m rows and n^2 columns. The i th row of A contains matrix A_i of (P) stretched out as a vector;
- b , which is the c supplied by Algorithm 1, i.e., $b = -10 \cdot e_m$;
- c , which is the right hand side B of (P) , stretched out as a vector;
- $Y_{maxrank}$, the maximum rank feasible solution in (D) , given in equation (4.26);
- Y_{opt} , the optimal solution of (P) , obtained by padding an optimal solution of (3.18) with zeros.

Each rotated instance is given by

- A, b , and c , which describe the constraints, just like for the clean instances.
- $A_{clean}, b_{clean}, c_{clean}, Y_{maxrank_clean}, Y_{opt_clean}$, which describe the corresponding clean instance.
- Matrix T , which was used to obtain the rotated instance from the clean one.

All data is integral, so the user can verify in exact arithmetic that the duality gap is 10.

In our dataset m varies from 2 to 10 and bs from 1 to 5. Hence the smallest (m, n) pair is $(2, 4)$ and the largest is $(10, 100)$.

We tested two SDP solvers: the Mosek commercial solver [1] and the high precision SDP solver SDPA-GMP. We report two outputs. The first output is the status given by the solvers. The second output is the largest of the DIMACS errors, which measure the constraint violations, and the duality gap of the approximate solutions delivered by the solvers: see e.g., [13].

We report a subset of our computational results in Table 1. For the sake of brevity, we only report on instances with $m = 2, \dots, 10$ and with $bs = 1, 2$.

We can see that both Mosek and SDPA-GMP compute near optimal solutions as m gets larger, i.e., neither recognize the positive gap. Note that SDPA-GMP only reports “primal feasible” on many instances, however, it actually computes an approximate solution pair with largest DIMACS error on the order of 10^{-15} .

We remark that we tested the preprocessing method of [19] and Sieve-SDP [26] on the dual problems. Both these methods correctly preprocesses all “clean” instances, but cannot do anything with the “rotated” instances.

| Problem | Mosek | | SDPA-GMP | |
|--------------------|-------------------|------------|----------|------------|
| Name | Status | DIMACS | Status | DIMACS |
| gap_clean_2.4.1 | DUAL_ILLPOSED_CER | 9.0909e-01 | pdINF | 5.9040e-01 |
| gap_clean_2.8.2 | DUAL_ILLPOSED_CER | 1.0000e+00 | pdINF | 4.6449e-02 |
| gap_rotated_2.4.1 | NEAR_OPTIMAL | 1.6163e-03 | pdINF | 4.4210e-04 |
| gap_rotated_2.8.2 | OPTIMAL | 1.0862e-04 | pFEAS | 1.0307e-10 |
| gap_clean_3.6.1 | OPTIMAL | 3.7806e-07 | pdOPT | 3.6546e-18 |
| gap_clean_3.12.2 | OPTIMAL | 3.0678e-07 | pdOPT | 4.8461e-19 |
| gap_rotated_3.6.1 | OPTIMAL | 8.8476e-07 | pFEAS | 2.0780e-14 |
| gap_rotated_3.12.2 | OPTIMAL | 8.3928e-06 | pFEAS | 1.3092e-15 |

| | | | | |
|---------------------|--------------|------------|------------|------------|
| gap_clean_4.8_1 | OPTIMAL | 8.2543e-08 | pdOPT | 5.3023e-25 |
| gap_clean_4.16_2 | OPTIMAL | 6.2916e-08 | pFEAS | 4.7518e-11 |
| gap_rotated_4.8_1 | OPTIMAL | 9.1596e-06 | pFEAS_dINF | 1.2989e-14 |
| gap_rotated_4.16_2 | OPTIMAL | 4.3186e-05 | pFEAS | 5.5751e-16 |
| gap_clean_5.10_1 | OPTIMAL | 7.6266e-07 | pdOPT | 6.0187e-28 |
| gap_clean_5.20_2 | OPTIMAL | 8.9868e-08 | pdOPT | 3.5233e-29 |
| gap_rotated_5.10_1 | OPTIMAL | 9.6152e-06 | pFEAS | 1.9978e-16 |
| gap_rotated_5.20_2 | OPTIMAL | 1.7681e-05 | pFEAS | 1.9752e-16 |
| gap_clean_6.12_1 | OPTIMAL | 8.9917e-08 | pdOPT | 3.8321e-30 |
| gap_clean_6.24_2 | OPTIMAL | 1.7804e-07 | pdOPT | 1.2398e-30 |
| gap_rotated_6.12_1 | OPTIMAL | 1.1232e-05 | pFEAS | 3.5333e-16 |
| gap_rotated_6.24_2 | OPTIMAL | 6.6445e-06 | pFEAS | 8.2623e-16 |
| gap_clean_7.14_1 | OPTIMAL | 1.0910e-07 | pdOPT | 4.8748e-30 |
| gap_clean_7.28_2 | OPTIMAL | 1.3939e-07 | pdOPT | 1.5698e-30 |
| gap_rotated_7.14_1 | OPTIMAL | 1.0156e-05 | pFEAS | 1.5213e-16 |
| gap_rotated_7.28_2 | OPTIMAL | 1.0858e-05 | pFEAS | 8.8389e-17 |
| gap_clean_8.16_1 | OPTIMAL | 1.7786e-07 | pdOPT | 1.3095e-30 |
| gap_clean_8.32_2 | OPTIMAL | 1.5157e-07 | pdOPT | 2.6450e-30 |
| gap_rotated_8.16_1 | OPTIMAL | 1.4562e-05 | pFEAS | 7.3632e-16 |
| gap_rotated_8.32_2 | NEAR_OPTIMAL | 4.6842e-03 | pFEAS | 4.1471e-15 |
| gap_clean_9.18_1 | OPTIMAL | 1.9984e-07 | pdOPT | 1.5320e-30 |
| gap_clean_9.36_2 | OPTIMAL | 1.2970e-07 | pdOPT | 1.7916e-30 |
| gap_rotated_9.18_1 | OPTIMAL | 1.7398e-06 | pFEAS | 1.2526e-15 |
| gap_rotated_9.36_2 | OPTIMAL | 2.0730e-05 | pFEAS | 8.8892e-16 |
| gap_clean_10.20_1 | OPTIMAL | 8.6322e-07 | pdOPT | 1.8466e-30 |
| gap_clean_10.40_2 | OPTIMAL | 1.7111e-07 | pdOPT | 2.5063e-30 |
| gap_rotated_10.20_1 | OPTIMAL | 4.7172e-06 | pFEAS | 1.6902e-15 |
| gap_rotated_10.40_2 | NEAR_OPTIMAL | 1.3293e-03 | pFEAS | 1.0295e-15 |

Table 1: Computational results

8 Conclusion

We analyzed positive duality gaps in SDPs. First, we completely characterized positive gaps in the two variable case: we transformed two variable SDPs to a standard form, that makes the positive gap self-evident. Second, we showed that the two variable case helps us understand positive gaps in larger SDPs: the structure that causes a positive gap in the two variable case also appears in higher dimensions. We studied an intrinsic parameter, the singularity degree of the duals of our SDPs, and proved that these are the largest that permit a positive gap. Finally, we created a problem library of innocent looking, but very difficult SDPs, and showed that they are currently unsolvable by modern interior point methods.

It would be interesting to try the exact arithmetic SDP solver SPECTRA [8] on our library; or

perhaps a modification of the Douglas-Rachford algorithm [11] designed to detect weak infeasibility would work.

A Proofs of technical statements

A.1 Proof of Lemma 4

Let $Q_1 \in \mathbb{R}^{r_1 \times r_1}$ be a matrix of orthonormal eigenvectors of A_{11} , $Q_2 \in \mathbb{R}^{r_2 \times r_2}$ a matrix of suitably normalized eigenvectors of A_{22} , and $T_1 = Q_1 \oplus Q_2$.

Then

$$T_1^T M T_1 = \left(\begin{array}{c|cc} \Omega & V & W \\ \hline V^T & I_s & 0 \\ W^T & 0 & 0 \end{array} \right),$$

with $\Omega \in \mathcal{S}^{r_1}$, s is equal to the rank of A_{22} , and V and W are possibly nonzero.

Next, let

$$T_2 = \left(\begin{array}{c|cc} I_{r_1} & 0 & 0 \\ \hline -V^T & I_s & 0 \\ 0 & 0 & I_{r_2-s} \end{array} \right), \text{ then } T_2^T T_1^T M T_1 T_2 = \left(\begin{array}{c|cc} \Omega - VV^T & 0 & W \\ \hline 0 & I_s & 0 \\ W & 0 & 0 \end{array} \right).$$

Finally, let $Q_3 \in \mathbb{R}^{r_1 \times r_1}$ be a matrix of orthonormal eigenvectors of $\Omega - VV^T$, and $T_3 = Q_3 \oplus I_{r_2}$, then $T_3^T T_2^T T_1^T M T_1 T_2 T_3$ is in the required form.

Also note that applying the rotation $T_i^T()T_i$ to $I_{r_1} \oplus 0$ leaves this matrix the same for all i . Thus

$$T := T_1 T_2 T_3$$

will do. □

A.2 Proof of Claim 1

Let (Y_1, \dots, Y_ℓ) be as stated. First suppose $\ell = 1$ and write $Y_1 = \sum_{j=1}^m \lambda_{1j} A_j$ with some λ_{1j} reals. By definition $Y_1 \succeq 0$, and since $A_j(\mathcal{I}_{m+1}) = 0$ for all j , we deduce $Y_1(\mathcal{I}_{m+1}) = 0$.

Next note that

$$A_j(\mathcal{J}_{j-1} \cup \mathcal{I}_{j-1}, \mathcal{I}_{m+1}) \neq 0$$

for $j = 2, \dots, m$, so $\lambda_{12} = \dots = \lambda_{1m} = 0$. Since the sequence (Y_1) is strict, we deduce $\lambda_{11} > 0$, thus (4.27) follows.

Next, assume $\ell < m$ and that (4.27) is true for some ℓ , and consider a sequence $(Y_1, \dots, Y_{\ell+1})$. Then

$$\begin{aligned} F_\ell &= \mathcal{S}_+^n \cap Y_1^\perp \cap \dots \cap Y_\ell^\perp \\ &= \{ Y \succeq 0 \mid Y(\mathcal{J}_{1:\ell} \cup \mathcal{I}_{1:\ell}, \cdot) = 0 \}, \end{aligned} \tag{A.35}$$

where the first equality is from the induction hypothesis, and the second is from (4.25).

Define A'_1, \dots, A'_m and $Y'_{\ell+1}$ by deleting rows and columns corresponding to $\mathcal{I}_{1:\ell} \cup \mathcal{I}_{\ell+1}$ from A_1, \dots, A_m and $Y_{\ell+1}$. Then

$$A'_1 = \dots = A'_\ell = 0, \quad (\text{A.36})$$

$$A'_m \text{ is indefinite}, \quad (\text{A.37})$$

$$Y'_{\ell+1} \succeq 0, \quad (\text{A.38})$$

where the first two equations come from the definition of the A_i . In particular, (A.37) is easy to verify by looking at equation (3.21). The third follows since $Y_{\ell+1} \in (\mathcal{S}_+^n \cap Y_1^\perp \cap \dots \cap Y_\ell^\perp)^*$ and taking (A.35) into account.

Let us write $Y_{\ell+1} = \sum_{j=1}^m \lambda_{\ell+1,j} A_j$ with some $\lambda_{\ell+1,j}$ reals. By (A.36) we get

$$Y'_{\ell+1} = \sum_{j=\ell+1}^m \lambda_{\ell+1,j} A'_j. \quad (\text{A.39})$$

Since A'_m is indefinite, and $Y'_{\ell+1} \succeq 0$ we deduce $\ell + 1 < m$.

We now use a similar argument as before in the $\ell = 1$ case. We have $Y'_{\ell+1}(\mathcal{I}_{m+1}) = 0$ and

$$A'_j(\mathcal{J}_{j-1} \cup \mathcal{I}_{j-1}, \mathcal{I}_{m+1}) \neq 0$$

for $j = \ell + 2, \dots, m$, hence $\lambda_{\ell+1,\ell+2} = \dots = \lambda_{\ell+1,m} = 0$. Since the sequence is strict, we deduce $\lambda_{\ell+1,\ell+1} > 0$, so our statement follows for $\ell + 1$ as well. This completes the proof of Claim 1. \square

Acknowledgement I am grateful to Shu Lu and Minghui Liu for many helpful discussions and to Quoc Tran Dinh and Yuzixuan Zhu for help in the computational experiments. Special thanks are also due to Yuzixuan Zhu for a careful reading of the paper.

References

- [1] MOSEK ApS. The mosek optimization toolbox for matlab manual. version 7.0 (revision 126).”, 2014.
- [2] Aharon Ben-Tal and Arkadii Nemirovskii. *Lectures on modern convex optimization*. MPS/SIAM Series on Optimization. SIAM, Philadelphia, PA, 2001.
- [3] Frédéric J. Bonnans and Alexander Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, 2000.
- [4] Jonathan Borwein and Adrian S Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.
- [5] Jonathan M. Borwein and Henry Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. App.*, 83:495–530, 1981.
- [6] Dmitriy Drusvyatskiy, Guoyin Li, and Henry Wolkowicz. A note on alternating projections for ill-posed semidefinite feasibility problems. *Mathematical Programming*, 162(1-2):537–548, 2017.
- [7] Robert M Freund and Jorge R Vera. Some characterizations and properties of the distance to ill-posedness? and the condition measure of a conic linear system. *Mathematical Programming*, 86(2):225–260, 1999.

- [8] Didier Henrion, Simone Naldi, and Mohab Safey El Din. Spectra—a maple library for solving linear matrix inequalities in exact arithmetic. *Optimization Methods and Software*, pages 1–17, 2017.
- [9] Minghui Liu and Gábor Pataki. Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming. *Mathematical Programming*, pages 1–46, 2017.
- [10] Minghui Liu and Gábor Pataki. Exact duality in semidefinite programming based on elementary reformulations. *SIAM J. Opt.*, 25(3):1441–1454, 2015.
- [11] Yanli Liu, Ernest K Ryu, and Wotao Yin. A new use of Douglas-Rachford splitting and admm for identifying infeasible, unbounded, and pathological conic programs. *arXiv preprint arXiv:1706.02374*, 2017.
- [12] Bruno F Lourenço, 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, 2016.
- [13] Hans D Mittelmann. An independent benchmarking of sdp and socp solvers. *Mathematical Programming*, 95(2):407–430, 2003.
- [14] Manuel A Nunez. A characterization of ill-posed data instances for convex programming. *Mathematical programming*, 91(2):375–390, 2002.
- [15] Gábor Pataki. The geometry of semidefinite programming. In Romesh Saigal, Lieven Vandenbergh, and Henry Wolkowicz, editors, *Handbook of semidefinite programming*. Kluwer Academic Publishers, also available from www.unc.edu/~pataki, 2000.
- [16] 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.
- [17] Gábor Pataki. Bad semidefinite programs: they all look the same. *SIAM J. Opt.*, 27(1):146–172, 2017.
- [18] Javier Peña. Understanding the geometry of infeasible perturbations of a conic linear system. *SIAM Journal on Optimization*, 10(2):534–550, 2000.
- [19] Frank Permenter and Pablo Parrilo. Partial facial reduction: simplified, equivalent sdps via approximations of the psd cone. *Mathematical Programming*, pages 1–54, 2014.
- [20] James Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65(1-3):73–91, 1994.
- [21] Jos Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10:1228–1248, 2000.
- [22] Jos F Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.
- [23] Levent Tunçel and Henry Wolkowicz. Strong duality and minimal representations for cone optimization. *Comput. Optim. Appl.*, 53:619–648, 2012.
- [24] Hayato Waki and Masakazu Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013.
- [25] Boshi Yang and Samuel Burer. A two-variable approach to the two-trust-region subproblem. *SIAM Journal on Optimization*, 26(1):661–680, 2016.
- [26] 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.