A Simplified Treatment of Ramana's Exact Dual for Semidefinite Programming

Bruno F. Lourenço[*] Gábor Pataki[†]

May 26, 2022

Abstract

In semidefinite programming the dual may fail to attain its optimal value and there could be a duality gap, i.e., the primal and dual optimal values may differ. In a striking paper, Ramana [20] proposed a polynomial size extended dual that does not have these deficiencies and yields a number of fundamental results in complexity theory. In this work we walk the reader through a concise and self-contained derivation of Ramana's dual, relying mostly on elementary linear algebra.

Key words: semidefinite programming; duality; duality gap; facial reduction; Ramana's dual

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)

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

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 we write $\langle T, S\rangle := \operatorname{trace}(TS)$ to denote their inner product.

SDPs are an elegant generalization of linear programming, and they appear in a broad range of application areas. However, the duality theory of SDPs is much less satisfactory than that of linear

[*]Department of Statistical Inference and Mathematics, Institute of Statistical Mathematics, Japan. (bruno@ism.ac.jp)

[†]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, USA. (gabor@unc.edu)

programming. On the one hand, the optimal value of (P) is bounded from above by the optimal value of (D). On the other, (P) and (D) may not have optimal solutions, i.e., the supremum may not be a maximum, and the infimum may not be a minimum. Even worse, the optimal values of (P) and (D) may differ.

Example 1. *In the following classical pathological SDP*

$$\begin{array}{ll} \sup & 2x_1 \\ s.t. & x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \end{array} \tag{1.1}$$

the constraint is equivalent to $\left(\begin{smallmatrix} 1 & -x_1 \\ -x_1 & 0 \end{smallmatrix}\right) \succeq 0$*, so the only feasible solution is* $x_1 = 0$.

The dual, with a variable matrix $Y = (y_{ij})$*, is*

$$\begin{array}{lrcl} \inf & y_{11} & & \\ s.t. & y_{12} & = & 1 \\ & Y & \succeq & 0, \end{array} \tag{1.2}$$

wherein any $Y_\epsilon := \left(\begin{smallmatrix} \epsilon & 1 \\ 1 & 1/\epsilon \end{smallmatrix}\right)$ *with* $\epsilon > 0$ *is feasible. So we conclude that the infimum of* (1.2) *is* 0.

However, any Y *with* $y_{11} = 0$ *is not feasible in* (1.2)*, so its infimum is not attained.*

In a striking paper, Ramana [20] constructed a new dual problem that fixes most of the issues of the classical SDP dual. Ramana's dual has the following attractive traits:

(1) it does not assume anything about (P), other than it is feasible;

(2) it attains its optimal value, when that value is finite;

(3) its optimal value is the same as that of (P), so there is no duality gap;

(4) it yields important complexity implications. Among other things, it proves that deciding feasibility of SDPs in the Turing model is not NP-complete, unless NP = co-NP, which is an unlikely scenario according to most experts.

Ramana's dual sparked great excitement in the SDP community and inspired many followup papers. Ramana, Tunçel and Wolkowicz [22] connected it to the facial reduction algorithm of Borwein and Wolkowicz [3, 4]; Luo, Sturm, and Zhang [13] gave a different proof of its correctness; Ramana and Freund [21] showed that it has zero duality gap with its usual dual; and Klep and Schweighofer [8] constructed a dual with similar properties, which relies on machinery from real algebraic geometry. Ramana's work is often cited in surveys and books: see for example Drusvyatskiy and Wolkowicz [7], DeKlerk [5], Vandenberghe and Boyd [25], Nemirovski [14], and Laurent and Rendl [9]. It was used by DeKlerk, Roos and Terlaky [6] in self-dual embeddings. It is often mentioned in the discrete mathematics and theoretical computer science literature, see for example, Lovász [12] and O' Donnell [15].

Ramana's dual has been generalized in a number of directions: to conic linear programs over so called *nice cones* [16, Corollary 1], and even to arbitrary conic linear programs, which have a Ramana type dual [10, Theorem 2]. It has greatly inspired the authors to examine why the pathologies arise in the first place: see for example [11] and [17].

The known derivations of Ramana's dual rely on convex analysis, namely on the technique of facial reduction. Facial reduction originated in in the eighties [3, 4], then simplified variants were proposed by Waki and Muramatsu [26] and the second author of this note [16]. For a recent survey of facial reduction and its applications, we refer to [7]. On the other hand, the related dual of Klep and Schweighofer [8] employs algebraic geometry. These are two complementary approaches, tailored to readers trained either in convex analysis, or algebraic geometry.

In this work we give a short and elementary derivation of Ramana's dual that we hope will appeal to all audiences.

To set the stage, we define the operator $\mathcal{A}$ and its adjoint $\mathcal{A}^*$ as

$$\mathcal{A}x := \sum_{i=1}^{m} x_i A_i,\ \mathcal{A}^* Y = (\langle A_1, Y\rangle, \ldots, \langle A_m, Y\rangle)^\top,$$

where $x \in \mathbb{R}^m$ and Y is an $n \times n$ symmetric matrix.

In what follows, we assume that the primal (P) is feasible, and we denote by val() the optimal value of an optimization problem. We use the common convention that the optimal value of an "inf" problem is $+\infty$ exactly when it is infeasible. We denote by $\mathcal{S}^n$ the set of $n \times n$ symmetric matrices, and by $\mathcal{S}^n_+$ the set of symmetric psd matrices. For a matrix M (symmetric or not) $\mathcal{R}(M)$ stands for its range space.

Theorem 1. *Consider the optimization problem*

$$\left.\begin{array}{llll}
\inf & \langle B, U_{n+1} + V_{n+1}\rangle & & \\
s.t. & \mathcal{A}^*(U_{n+1} + V_{n+1}) = c & & \\
 & \mathcal{A}^*(U_i + V_i) = 0 & i = 1, \ldots, n & (1.3) \\
 & \langle B, U_i + V_i\rangle = 0 & i = 1, \ldots, n & (1.4) \\
 & V_i \in \tan(U_{i-1}) & i = 1, \ldots, n+1 & (1.5) \\
 & U_i \in \mathcal{S}^n_+ & i = 1, \ldots, n+1 & (1.6) \\
 & U_0 = V_0 = 0. & &
\end{array}\right\} \quad (\mathrm{D_{Ram}})$$

Here for $U \in \mathcal{S}^n_+$ the set $\tan(U)$ *is defined as*

$$\tan(U) \ = \ \{\, W + W^\top \,|\, W \in \mathbb{R}^{n\times n},\ \mathcal{R}(W) \subseteq \mathcal{R}(U)\,\}, \tag{1.7}$$

and called the tangent space of $\mathcal{S}^n_+$ *at* U.

Then

$$\mathrm{val}\,(\mathrm{P}) = \mathrm{val}\,(\mathrm{D_{Ram}}),$$

and $\mathrm{val}\,(\mathrm{D_{Ram}})$ *is attained when finite.* □

We call ($\mathrm{D_{Ram}}$) the *Ramana dual* of (P).

Ramana's dual at first may look mysterious. The reader may also object that, because of the tangent space constraint in (1.5), it is not even an SDP! We will take care of the latter issue soon, but we first explain the makeup of ($\mathrm{D_{Ram}}$). Variables U_0 and V_0 are included in it only for convenience, and variable V_1 is always zero since $U_0 = 0$.

It is straightforward that

$$\mathrm{val}\,(\mathrm{D_{Ram}}) \le \mathrm{val}\,(\mathrm{D}) \tag{1.8}$$

holds. Indeed, we can construct a feasible solution of ($\mathrm{D_{Ram}}$) from any feasible solution Y of (D): we can set $U_{n+1} := Y$ and the other U_i and all V_i to zero. This construction works, since the zero matrix is in the tangent space of $\mathcal{S}^n_+$ at any psd matrix: see the definition in (1.7).

However, in general $U_1, \ldots, U_n$ and the V_i will not be all zero. Loosely speaking, they ensure that $U_{n+1} + V_{n+1}$ lives in the "right" superset of $\mathcal{S}^n_+$ to ensure zero duality gap and attainment in (D_{Ram}). We illustrate this fact next.

Example 2. *(Example 1 continued) The Ramana dual of* (1.1) *has an attained* 0 *optimal value, as evidenced by the solution*

$$U_1 = V_1 = 0,\, U_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\, V_2 = 0,\; U_3 = 0,\; V_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{1.9}$$

Here $V_3 \in \tan(U_2)$ *because* V_3 *can be written as*

$$V_3 = \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{W} + \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{W^\top},$$

and $\mathcal{R}(W) \subseteq \mathcal{R}(U_2)$. *Note that* $U_3 + V_3$ *is not psd.*

In this example, as well as in later examples, we only list the U_i and V_i for $i \geq 1$ (since $U_0 = V_0 = 0$).

The next theorem, which uses ideas from Lemma 2.1 in [22], shows that (D_{Ram}) can be turned into a bona fide SDP:

Theorem 2. *Suppose* $U \in \mathcal{S}^n_+$. *Then*

$$\tan(U) = \left\{ W + W^\top \,\middle|\, W \in \mathbb{R}^{n \times n},\, \begin{pmatrix} U & W \\ W^\top & \beta I \end{pmatrix} \succeq 0 \text{ for some } \beta \in \mathbb{R} \right\}. \tag{1.10}$$

Therefore, (D_{Ram}) *can be expressed as an SDP with auxiliary variables* $W_i \in \mathbb{R}^{n \times n}$ *and* $\beta_i \in \mathbb{R}$ *for* $i = 1, \ldots, n+1$; *i.e., with polynomially many (in* n *and* m*) variables and constraints.*

We remark that $\tan(U)$ has a geometric meaning that is familiar from convex analysis; we will explain this in Section 4. However, to derive Ramana's dual, we only need the algebraic definition of $\tan(U)$ in (1.7).

Outline of the paper Our proofs use mostly basic linear algebra, but we rely on a classic, rudimentary strong duality result from conic linear programming, that we recap in Proposition 3.

We employ the notion of the relative interior of a convex set. However, we use it only for a very particular type of set. For $k \geq 0$ we define

$$\mathcal{S}^k_+ \oplus 0 := \left\{ \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{S}^n \,\middle|\, X \in \mathcal{S}^k_+ \right\}. \tag{1.11}$$

It is straightforward to show that the relative interior of $\mathcal{S}^k_+ \oplus 0$ is

$$\left\{ \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{S}^n \,\middle|\, X \text{ is } k \times k \text{ positive definite} \right\}, \tag{1.12}$$

and using this fact the reader will be able to follow all proofs.

In Section 2 we introduce the key ingredients of our proof: i) *rescaling* the operator $\mathcal{A}$ and the right hand side B; ii) a *maximum rank slack* in (P), in other words, a maximum rank psd matrix of the form $B - \mathcal{A}x$, where $x \in \mathbb{R}^m$; and iii) Proposition 3.

Section 3 has the main proofs, that we outline below:

- In Subsection 3.1, in Lemma 2 (with Lemma 1 as a preliminary result) we describe certificates for a maximum rank slack in (P). By "certificate" we mean a finite sequence of matrices that convince a third party that a maximum rank slack in (P) indeed has maximum rank.

- In Subsection 3.2 we present a semidefinite program ($\mathrm{D}_{\mathrm{strong}}$), which is a strong dual of (P). That is, ($\mathrm{D}_{\mathrm{strong}}$) is an "inf" problem which attains its optimal value, when finite, and

$$\mathrm{val}\,(\mathrm{P}) = \mathrm{val}\,(\mathrm{D}_{\mathrm{strong}}). \tag{1.13}$$

  The problem ($\mathrm{D}_{\mathrm{strong}}$) is essentially the same as (D), but in ($\mathrm{D}_{\mathrm{strong}}$) only a block of the variable matrix Y must be psd.

  At the same time, ($\mathrm{D}_{\mathrm{strong}}$) has a drawback: to write it down, we need to know a maximum rank slack in (P) explicitly. However, in general we do not know such a slack explicitly; we only know that one exists.

- In Subsection 3.3 we tie together the previously proved results and prove that ($\mathrm{D}_{\mathrm{strong}}$) is "mimicked" by Ramana's dual. We prove

$$\mathrm{val}\,(\mathrm{D}_{\mathrm{strong}}) = \mathrm{val}\,(\mathrm{D}_{\mathrm{Ram}}), \tag{1.14}$$

  and that ($\mathrm{D}_{\mathrm{Ram}}$) attains its optimal value, when it is finite.

  In particular, to prove the inequality $\geq$ in (1.14) we produce an optimal solution of ($\mathrm{D}_{\mathrm{Ram}}$) as follows: from an optimal solution of ($\mathrm{D}_{\mathrm{strong}}$) we produce U_{n+1} and V_{n+1}, and from the certificates for a maximum rank slack given in Subsection 3.1 we produce the other U_i and V_i.

  We then combine (1.13) and (1.14) and attainment in ($\mathrm{D}_{\mathrm{Ram}}$) and prove Theorem 1.

- In Subsection 3.4 we prove the SDP representation result Theorem 2.

In Section 4 we conclude: we present a larger example and explain the geometry of the set $\tan(U)$, which at first may look enigmatic.

2 Preliminaries

Principal submatrices and concatenation Suppose r and s are integers in $\{1, \ldots, n\}, r \leq s$, and $Y \in \mathcal{S}^n$. We then denote by $Y(r : s)$ the principal submatrix of Y indexed by rows and columns $r, r+1, \ldots, s$.

Further, we denote the concatenation of matrices A and B along the diagonal by $A \oplus B$, i.e.,

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

We naturally define the matrix $A \oplus B \oplus C$ as $(A \oplus B) \oplus C$.

2.1 Rescaling $\mathcal{A}$ and B, slacks, and maximum rank slacks

We will often rescale the operator $\mathcal{A}$ and the right hand side B to put our semidefinite programs into a more convenient form. The precise definition follows:

Definition 1. *We say that we* rescale *the operator $\mathcal{A}$ and the matrix B if we perform the operations*

$$\begin{array}{rcl} A_i & := & T^\top A_i T \text{ for } i = 1, \dots, m, \\ B & := & T^\top B T. \end{array} \tag{Rescale}$$

where T is a suitable invertible matrix.

Slacks and maximum rank slacks We first define slack matrices in (P), which generalize slack vectors in linear programming.

Definition 2. *We say that $S \succeq 0$ is a* slack matrix *or* slack *in* (P) *if $S = B - \mathcal{A}x$ for some $x \in \mathbb{R}^m$.*

Note that if the A_i and B are diagonal, then (P) is a linear program, and the diagonal of a slack matrix in (P) is just a slack vector in this linear program.

Since the rank of a slack matrix is a nonnegative integer, and it is at most n, the semidefinite program (P) has a slack of maximum rank. A maximum rank slack of (P) will be a key player in the rest of the paper: it measures "how centrally" the affine subspace $\{ B - \mathcal{A}x : x \in \mathbb{R}^m \}$ intersects the set of positive semidefinite matrices.

For convenience we make the following assumption.

Assumption 1. *There is a maximum rank slack in* (P) *of the form*

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

where $0 \le r \le n$, and, as usual, I_r stands for an identity matrix of order r.

For the rest of this paper we fix this Z and r.

We can ensure Assumption 1 by a suitable rescaling as follows. Suppose S is a maximum rank slack in (P), T is an invertible matrix of suitably scaled eigenvectors of S, and we perform the operations (Rescale). Afterwards in the new primal problem a maximum rank slack is $Z := T^\top S T$, which is in the required shape.

The maximum rank slack in (P) may not be unique. However, after we state Lemma 2, we will prove a slightly weaker statement: any maximum rank slack in (P) must be of the form $R \oplus 0$, where R is an order r symmetric positive definite matrix.

After the initial rescaling that created our maximum rank slack Z, we may rescale $\mathcal{A}$ and B several more times. In these subsequent rescalings the T transformation matrix will always be of the form

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

where M is an invertible, order $n - r$ matrix. These subsequent rescalings will keep Z in the same form, since for any such T matrix we have $T^\top Z T = Z$.

Example 3. *(Example 1 continued) As we discussed, in the SDP* (1) *the only feasible solution is $x_1 = 0$, hence the maximum rank slack is just the right hand side*

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.15}$$

That is, this SDP needs no rescaling, since Z already satisfies Assumption 1 (with $r = 1$).

2.2 Rescaling $\mathcal{A}$ and B keeps the optimal value of Ramana's dual the same

When we rescale $\mathcal{A}$ and $B,$ we keep the optimal value (and even the feasible set) of (P) the same. It is a bit less obvious that operations (Rescale) do not affect the optimal value Ramana's dual, so we prove that in Proposition 2.

First, in Proposition 1 we collect some useful properties of symmetric matrices.

Proposition 1. *The following statements hold:*

(1) Suppose $X, Y \in \mathcal{S}^n$ and T is invertible. Then

$$\langle X, Y \rangle \, = \, \langle T^\top X T, T^{-1} Y T^{-\top} \rangle.$$

(2) Suppose $U \in \mathcal{S}^n_+$ and T is invertible. Then

$$T^\top \tan(U) T \, = \, \tan(T^\top U T), \tag{2.16}$$

where $T^\top \tan(U) T$ is defined as $\{ T^\top V T \, | \, V \in \tan(U) \}$.

(3) Suppose $U \in \mathcal{S}^n_+$ is of the form $U = 0 \oplus I_s$, where $s \leq n$. Then $\tan(U)$ is the set of matrices in $\mathcal{S}^n$ of the form

$$\begin{pmatrix} \overbrace{0}^{n-s} & \overbrace{\times}^{s} \\ \times & \times \end{pmatrix}, \tag{2.17}$$

where the elements in the $\times$ blocks are arbitrary.

Proof Statement (1) follows directly from the properties of the trace:

$$\begin{array}{rcl} \langle X, Y \rangle & = & \mathrm{tr}(XY) \\ & = & \mathrm{tr}(X T T^{-1} Y T^{-\top} T^\top) \\ & = & \mathrm{tr}(T^\top X T T^{-1} Y T^{-\top}) \\ & = & \langle T^\top X T, T^{-1} Y T^{-\top} \rangle. \end{array}$$

To prove the inclusion $\subseteq$ in (2), suppose $\mathcal{R}(W) \subseteq \mathcal{R}(U)$, so $W + W^\top \in \tan(U)$. Then

$$\mathcal{R}(T^\top W T) \subseteq \mathcal{R}(T^\top U T),$$

hence $T^\top (W + W^\top) T \in \tan(T^\top U T),$ so the inclusion $\subseteq$ follows.

Next we apply the inclusion $\subseteq$ in (2.16), but now with $T^\top U T$ in place of U and T^{-1} in place of T, and we deduce

$$T^{-\top} \tan(T^\top U T) T^{-1} \subseteq \tan(U), \tag{2.18}$$

and left multiplying (2.18) by $T^\top$ and right multiplying by T yields the inclusion $\supseteq$ in (2).

The statement (3) follows from the definition of the tangent space in (1.7) and since for a matrix $W \in \mathbb{R}^{n \times n}$ we have $\mathcal{R}(W) \subseteq \mathcal{R}(U)$ exactly when the first $n - s$ rows of W are zero. $\square$

To build intuition, we first argue that rescaling keeps the optimal value of (D) the same. Indeed, suppose Y is feasible in (D) with objective value, say, $\alpha,$ and we apply the operations (Rescale) using an invertible matrix T. Then $T^{-1} Y T^{-\top}$ is feasible in (D) after rescaling, and has objective value $\alpha:$

this follows by item (1) in Proposition 1, with A_i or B in place of X. Since we can undo the rescaling with T by another rescaling (with T^{-1}), it follows that the optimal value of (D) is the same before and after rescaling.

A similar argument, given in Proposition 2, shows that the optimal value of ($\mathrm{D_{Ram}}$) stays the same after rescaling. The only difference is that we now also have to take care of the tangent space constraints (1.5).

Proposition 2. *The operations* (Rescale) *keep the optimal value of* ($\mathrm{D_{Ram}}$) *the same.*

Proof Suppose that $(U_j, V_j)_{j=0}^{n+1}$ is feasible in ($\mathrm{D_{Ram}}$) with objective value α before we performed (Rescale); we prove that

$$(U_j', V_j')_{j=0}^{n+1} := (T^{-1} U_j T^{-\top}, T^{-1} V_j T^{-\top})_{j=0}^{n+1}$$

is feasible afterwards, and has the same objective value.

Indeed, by (1) in Proposition 1 we have

$$\langle A_i, U_j + V_j \rangle \quad = \quad \langle T^\top A_i T, U_j' + V_j' \rangle$$

for all i and j. By the same logic we see that

$$\langle B, U_j + V_j \rangle \quad = \quad \langle T^\top B T, U_j' + V_j' \rangle$$

for all j. Thus, $(U_j', V_j')_{i=0}^{n+1}$ satisfies the equality constraints of ($\mathrm{D_{Ram}}$) after we executed (Rescale).

Also, for all j we have $V_j \in \tan(U_{j-1})$. Thus, by item (2) in Proposition 1 (with $T^{-\top}$ in place of T) we deduce

$$V_j' \in \tan(U_{j-1}')$$

for all j.

Summarizing, $(U_j', V_j')_{j=0}^{n+1}$ is feasible in ($\mathrm{D_{Ram}}$) with objective value α after we executed (Rescale), completing the proof. □

2.3 Strong duality under Slater's condition

We next state a classic strong duality result assuming the underlying space is $\mathcal{S}^n$, a special case most relevant for this work [1].

Suppose $K \subseteq \mathcal{S}^n$ is a closed convex cone [2] and let us denote its relative interior by $\operatorname{ri} K$, and its dual cone by K^*, i.e.,

$$K^* = \{ Y \in \mathcal{S}^n \mid \langle X, Y \rangle \geq 0 \, \forall \, X \in K \}.$$

Proposition 3. *Suppose $\mathcal{A}, B, K,$ and K^* are as previously defined, and*

$$B - \mathcal{A}x \in \operatorname{ri} K \text{ for some } x \in \mathbb{R}^m. \tag{2.19}$$

Then

$$\sup\{ c^\top x \mid B - \mathcal{A}x \in K \} = \inf\{ \langle B, Y \rangle \mid \mathcal{A}^* Y = c, Y \in K^* \} \tag{2.20}$$

and the optimal value of the "inf" problem is attained when it is finite. □

[1] Proposition 3 is usually stated in the space $\mathbb{R}^n$.

[2] That is, K is closed, convex, and $\lambda x \in K$ for all $x \in K$ and $\lambda \geq 0$.

When condition (2.19) holds, we say that the "sup" problem in (2.20) satisfies Slater's condition.

For better intuition, we next outline two important uses of Proposition 3. Each one corresponds to how large r, the rank of Z is.

First suppose $r = n$, and we choose $K = \mathcal{S}_+^n$. Then $K^* = \mathcal{S}_+^n$ holds as well, so the optimization problems in (2.20) are just the semidefinite programs (P) and (D). Further, (P) satisfies Slater's condition. Hence (D) has the same optimal value as (P), and attains this value when it is finite. In other words, the usual dual is just as good as Ramana's dual.

Second, suppose $r < n$ and we set $K = \mathcal{S}_+^r \oplus 0$. Since the dual cone of $\mathcal{S}_+^r$ is $\mathcal{S}_+^r$, itself, we have

$$K^* = \{Y \in \mathcal{S}^n \,|\, Y(1:r) \succeq 0\}, \tag{2.21}$$

i.e., in the dual cone only the upper left $r \times r$ block of matrices must be psd. Suppose in the constraint set of (P), namely in

$$B - \mathcal{A}x \in \mathcal{S}_+^n$$

we replace $\mathcal{S}_+^n$ by K and in the constraint set of (D) we replace $Y \in \mathcal{S}_+^n$ by $Y \in K^*$. Then we show in Lemma 3 that three interesting things happen. First, even though K is a smaller set than $\mathcal{S}_+^n$, the feasible set of (P) remains the same. Second, (P) satisfies Slater's condition. Third, by Proposition 3, the dual becomes the promised strong dual ($\mathrm{D_{strong}}$).

A bit surprisingly, short and self-contained proofs of Proposition 3 are rare in the literature. Fortunately, such a proof is given in Theorem 7 in the technical report by Luo, Sturm and Zhang [13]. However, proofs of more specific or more general statements are common. As to more specific ones, the result with "relative interior" replaced by "interior" appears in Section 3.2 of Renegar [23], in Section 2.4 of the textbook of Ben-Tal and Nemirovskii [1], and in Section 5.3 of Borwein and Lewis [2]. As to more general statements, Proposition 3 follows from Fenchel's duality theorem in Rockafellar's classic text [24, Theorem 31.4].

3 Proofs

Recall that whenever (P) is feasible, it has a maximum rank slack. Further, after a suitable rescaling, we fixed a maximum rank slack Z (with rank r) of the shape given in Assumption 1. This Z will play a key role in all proofs.

3.1 Certificates for the maximum rank slack

Lemma 1 below proves that all slacks of (P) have certain restrictions on their shape and rank.

Lemma 1. *Suppose $s \in \{r+1, \ldots, n\}$ is an integer. Then the following semidefinite system has a solution:*

$$\begin{array}{rcl} \mathcal{A}^*Y & = & 0 \\ \langle B, Y \rangle & = & 0 \\ Y & \in & \mathcal{S}^n \\ Y(1:s) & \in & \mathcal{S}_+^s \setminus \{0\}. \end{array} \tag{3.22}$$

□

The Y matrix given in Lemma 1 certifies that (P) has no slack whose order s leading principal submatrix is positive definite, and the rest is zero. Indeed, suppose $S = B - \mathcal{A}u$ is such a slack, then we have

$$0 = \langle B, Y\rangle - \langle \mathcal{A}^*Y, u\rangle = \langle B - \mathcal{A}u, Y\rangle = \langle S, Y\rangle. \tag{3.23}$$

However, $\langle S, Y\rangle > 0,$ which is a contradiction.

Example 4. *(Example 1 continued) As we previously discussed, in the SDP (1.1) the right hand side $Z = \left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)$ is the maximum rank slack, so $r = 1$. By Lemma 1 with $s = 2$ we produce*

$$Y = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.24}$$

to certify that (1.1) has no positive definite slack. In this tiny example, since Y in (3.24) certifies that (1.1) has no rank 2 slack, it also certifies that Z has maximum rank.

However, in larger examples we will need a finite sequence *of matrices to completely certify that Z has maximum rank. We will illustrate this fact in Example 6.*

Proof of Lemma 1 Let $K = \mathcal{S}^s_+ \oplus 0$. Consider the primal-dual pair of semidefinite programs

$$\sup_{x,t}\{\, t \mid B - \mathcal{A}x - t(I_s \oplus 0) \in K\} \text{ and } \inf_{Y}\{\langle B, Y\rangle \mid \mathcal{A}^*Y = 0, \langle (I_s \oplus 0), Y\rangle = 1,\, Y \in K^*\}, \tag{3.25}$$

and for brevity, define

$$S(x,t) := B - \mathcal{A}x - t(I_s \oplus 0) \text{ for } x \in \mathbb{R}^m \text{ and } t \in \mathbb{R}.$$

We first claim that the optimal values of the "sup" and "inf" problems in (3.25) are the same and the optimal value of the "inf" problem is attained when it is finite. For that, let $x \in \mathbb{R}^m$ be such that $Z = B - \mathcal{A}x$. Then the upper left $s \times s$ block of $S(x, -1)$ is positive definite, and the other elements of $S(x, -1)$ are zero. Thus $S(x, -1) \in \operatorname{ri} K,$ so the "sup" problem in (3.25) satisfies Slater's condition, hence our claim follows from Proposition 3.

Next we claim that the optimal value of both optimization problems in (3.25) is nonnegative. For that, again let $x \in \mathbb{R}^m$ be such that $Z = B - \mathcal{A}x$, then $S(x, 0) \in K$, so the optimal value of the "sup" problem is indeed nonnegative, and our claim follows.

Then we claim that the optimal value of both optimization problems in (3.25) is zero. To obtain a contradiction, suppose that $S(x,t) \in K$ for some $x \in \mathbb{R}^m$ and $t > 0$. Then the upper left $s \times s$ block of $B - \mathcal{A}x$ is positive definite, and the other elements are zero. Thus $B - \mathcal{A}x$ is a slack in (P) whose rank is larger than $r,$ which is the required contradiction.

Thus the "inf" problem in (3.25) has a feasible solution Y with objective value zero. By $Y \in K^*$ we get $Y(1:s) \succeq 0$ (see (2.21)) and by $\langle \big(I_s \oplus 0\big), Y\rangle = 1$ we get $Y(1:s) \neq 0$. Thus Y satisfies (3.22), as wanted. □

Lemma 1 gave a partial certificate for the maximum rank slack Z : the Y matrix in (3.24) certifies that Z has maximum rank among slacks with a fixed form. In contrast, Lemma 2 gives a complete certificate: it shows that Z has maximum rank among slacks of any form.

Lemma 2. *We can rescale $\mathcal{A}$ and B so that after the rescaling the following hold:*

(1) The Z matrix given in Assumption 1 is still a maximum rank slack in (P).

(2) There exist symmetric matrices $Y_1, \ldots, Y_k$ which are of the form

$$
Y_i := \begin{pmatrix} \overbrace{}^{n-\sum_{\ell=1}^{i} r_\ell} & \overbrace{}^{r_i} & \overbrace{\times}^{\sum_{\ell=1}^{i-1} r_\ell} \\ & I & \times \\ \times & \times & \times \end{pmatrix}, \qquad (3.26)
$$

and satisfy

$$
\begin{array}{rcl} \mathcal{A}^* Y_i & = & 0 \\ \langle B, Y_i \rangle & = & 0 \end{array} \qquad (3.27)
$$

for $i = 1, \ldots, k$. Here $k \geq 0$ and the r_i are positive integers such that $\sum_{i=1}^{k} r_i = n - r$.

□

Here, and in the sequel empty blocks in matrices contain all zeros, and $\times$ blocks may have arbitrary elements. (In some matrices we still explicitly indicate zero entries, if this helps readability.)

How do the Y_i in Lemma 2 certify that Z has maximum rank? To explain, let $S = B - \mathcal{A}u$ be an arbitrary slack in (P), where $u \in \mathbb{R}^m$. We present a simple argument using the Y_i to prove that the rank of S is at most r, in particular, that the last $n - r$ rows and columns of S are zero.

Using an argument similar to the one in (3.23) we first deduce $\langle S, Y_1 \rangle = 0$. Since $\langle S, Y_1 \rangle$ is the sum of the last r_1 diagonal elements of S, and these elements are all nonnegative, they must be all zero. Since $S \succeq 0$, we learn that the last r_1 rows and columns of S are zero.

We then repeat the above argument with Y_2 in place of Y_1. We have $\langle S, Y_2 \rangle = 0$. Since the last r_1 rows and columns of S are zero, $\langle S, Y_2 \rangle$ is the sum of the diagonal elements of S in rows numbered $n - r_1 - r_2 + 1, \ldots, n - r_1$. So these rows and colums are all zero.

Continuing, since the sum of all r_i is $n - r$, we deduce that the last $n - r$ rows and columns of S are zero, as required.

This argument also proves that Z is a maximum rank slack that is unique up to rescaling. Precisely, it proves that any rank r slack in (P) must look like $R \oplus 0$, where R is order r and positive definite.

Continuing Example 1, the Y matrix in (3.24) can serve as Y_1 for the SDP (1.1): here we can choose $k = 1$ so there is no need for other Y_i.

Proof of Lemma 2 For a nonnegative integer j we consider the following conditions:

(1) We have rescaled $\mathcal{A}$ and B so that Z is still a maximum rank slack in (P) after the rescaling.

(2) We have constructed Y_i for $i = 1, \ldots, j$ which are of the form required in (3.26) and satisfy the equations (3.27). Further, the r_i sizes of the identity blocks in the Y_i matrices are all positive.

We start with $j = 0$, then (1) is satisfied by assumption, and (2) is satisfied vacuously.

In a general step we assume that $j \geq 0$ and that conditions (1) and (2) hold with j. The argument after the statement of Lemma 2 implies that the identity blocks in the Y_i and the identity block in Z do not overlap, hence $\sum_{i=1}^{j} r_i \leq n - r$.

We define

$$s := n - \sum_{i=1}^{j} r_i,$$

hence $s \geq r$.

If $s = r$, we let $k = j$, and stop.

If $s > r$, then we will construct matrix Y_{j+1}, make sure that conditions (1) and (2) hold with $j+1$ in place of j, and increment j. Since all the r_i are positive, we can execute this step at most n times.

First we invoke Lemma 1 and produce $Y_{j+1} \in \mathcal{S}^n$ such that

$$\begin{array}{rcl} \mathcal{A}^* Y_{j+1} & = & 0 \\ \langle B, Y_{j+1} \rangle & = & 0 \\ Y_{j+1}(1:s) & \in & \mathcal{S}_+^s \setminus \{0\}. \end{array}$$

Let us recall the maximum rank slack Z from Assumption 1. Then using an argument just like in (3.23), we deduce $\langle Z, Y \rangle = 0$. Thus the first r rows and columns of $Y_{j+1}(1:s)$ are zero, meaning Y_{j+1} looks like

$$Y_{j+1} = \left(\begin{array}{c|c|c} \overbrace{}^{r} & \overbrace{\phantom{\bar{Y}\bar{Y}}}^{s-r} & \overbrace{\times}^{n-s} \\ \hline & \bar{Y} & \times \\ \hline \times & \times & \times \end{array} \right),$$

where $\bar{Y} \succeq 0$ and, as usual, the elements in the $\times$ blocks are arbitrary.

We next put Y_{j+1} into the required format, and rescale $\mathcal{A}$ and B to make sure that conditions (1) and (2) hold with $j+1$ in place of j. For that, let r_{j+1} be the rank of $\bar{Y}$, Q be an invertible matrix of eigenvectors of $\bar{Y}$ such that $Q^\top \bar{Y} Q = 0 \oplus I_{r_{j+1}}$ and

$$T := I_r \oplus Q \oplus I_{n-s}. \tag{3.28}$$

We next perform the operations

$$Y_i \quad := \quad T^\top Y_i T \quad \text{for } i = 1, \ldots, j+1. \tag{3.29}$$

These operations put Y_{j+1} into the form required in (3.26) , and keep $Y_1, \ldots, Y_j$ in the same form.

However, now the Y_i may not satisfy equations (3.27), so we correct that issue next. We perform the operations

$$\begin{array}{rcl} A_i & := & T^{-1} A_i T^{-\top} \quad \text{for } i = 1, \ldots, m, \\ B & := & T^{-1} B T^{-\top}, \end{array} \tag{3.30}$$

i.e., the (Rescale) operations with $T^{-\top}$ in place of T. After these operations the equations (3.27) hold (by part (1) of Proposition 1), so condition (2) holds with $j+1$.

Finally we show that condition (1) remains true. For that, we observe that after performing the operations in (3.30), the matrix $T^{-1} Z T^{-\top}$ is a slack of rank r in (P), hence it is a maximum rank slack. Also, by (3.28) we have

$$T^{-1} := I_r \oplus Q^{-1} \oplus I_{n-s}, \tag{3.31}$$

hence $T^{-1} Z T^{-\top} = Z$, so Z is a maximum rank slack in (P). Thus (1) holds, and the proof is complete. □

The proof of Lemma 2 gives a *facial reduction algorithm* to construct the Y_i. To explain this parlance, note that the set

$$F := \mathcal{S}_+^r \oplus 0$$

is a *face* of $\mathcal{S}_+^n$ [3] and the Y_i matrices *reduce* the set of feasible slacks of (P) to live in F. For simplicity, in our proofs we do not mention faces; there is no need, since F is perfectly captured by the maximum rank slack Z.

We note that the algorithm is theoretical, since to implement it, we must find the Y_i certificates in Lemma 2 in exact arithmetic; and for that, we must solve the pair of semidefinite programs in (3.25) in exact arithmetic. However, some heuristic implementations of facial reduction exist, see for example [19] and [27]. We further refer the reader to [4, 16, 26] for facial reduction algorithms for more general problems.

3.2 A strong dual, assuming we know a maximum rank slack

In this subsection we present our promised strong dual ($\mathrm{D}_{\mathrm{strong}}$).

Lemma 3. *Suppose that we rescaled $\mathcal{A}$ and B as stated in Lemma 2, and consider the optimization problem*

$$\begin{aligned} &\inf \ \langle B, Y\rangle \\ &s.t. \ \mathcal{A}^*Y = c \\ &\quad\ \ Y \in (\mathcal{S}_+^r \oplus 0)^*. \end{aligned} \tag{$\mathrm{D}_{\mathrm{strong}}$}$$

Then

$$\mathrm{val}\,(\mathrm{P}) = \mathrm{val}\,(\mathrm{D}_{\mathrm{strong}}),$$

and the optimal value of ($\mathrm{D}_{\mathrm{strong}}$) *is attained when it is finite.*

Proof By the argument after the statement of Lemma 2 we see that any slack in (P) is contained in $\mathcal{S}_+^r \oplus 0$. Thus (P) is equivalent to

$$\sup\{\, c^\top x \mid B - \mathcal{A}x \in \mathcal{S}_+^r \oplus 0 \,\}. \tag{3.32}$$

Again recall the maximum rank slack Z from Assumption 1. Since Z is in the relative interior of $\mathcal{S}_+^r \oplus 0$, we see that (3.32) satisfies Slater's condition. Thus by Proposition 3 the dual of (3.32) attains its optimal value, when it is finite, and this optimal value is the same as the optimal value of (3.32).

But the dual of (3.32) is just ($\mathrm{D}_{\mathrm{strong}}$), hence the proof is complete. □

We next illustrate Lemma 3.

Example 5. *(Example 1 continued) We repeat the SDP from Example 1 for convenience.*

$$\begin{aligned} \sup \quad & 2x_1 \\ s.t. \quad & x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned} \tag{3.33}$$

As we discussed, the only feasible solution is $x_1 = 0$, hence the optimal value of (3.33) is zero, and the right hand side in (3.33) is the maximum rank slack.

[3] This means two things: (i) it is a convex subset of $\mathcal{S}_+^n$ and (ii) if X, Y are in $\mathcal{S}_+^n$, and the open line segment $\{\lambda X + (1-\lambda)Y : \lambda \in (0,1)\}$ intersects F, then both X and Y must be in F.

Thus in the strong dual ($\mathrm{D}_{\mathrm{strong}}$) *of* (3.33) *only the upper left* 1×1 *block of* Y *must be psd. Hence* ($\mathrm{D}_{\mathrm{strong}}$) *is*

$$\begin{array}{rrcl} \inf & y_{11} & & \\ s.t. & y_{12} & = & 1 \\ & y_{11} & \geq & 0, \end{array} \tag{3.34}$$

which is just a linear program. The matrix

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

is an optimal solution of (3.34) *that attains the optimal value of* 0.

We note that the strong dual ($\mathrm{D}_{\mathrm{strong}}$) is essentially the same as the usual dual (D), however, it requires only a block of the variable matrix Y to be psd. Thus ($\mathrm{D}_{\mathrm{strong}}$) should even be easier to solve than (D)! So why not use it?

Here is the catch: to write down ($\mathrm{D}_{\mathrm{strong}}$) we would need to know a maximum rank slack in (P) explicitly. If we did, then by rescaling we could ensure that a maximum rank slack Z is in the shape required in Assumption 1, then write down ($\mathrm{D}_{\mathrm{strong}}$). Of course, in general we do not know a maximum rank slack explicitly, we only know that one exists.

However, in the next section we show that ($\mathrm{D}_{\mathrm{strong}}$) is "mimicked" by Ramana's dual, which has no need of a maximum rank slack; of course, Ramana's dual needs many more variables.

3.3 Proof of Theorem 1

In this subsection we complete the proof of Theorem 1. First, in Lemmas 4 and 5 we prove

$$\mathrm{val}\,(\mathrm{D}_{\mathrm{strong}}) = \mathrm{val}\,(\mathrm{D}_{\mathrm{Ram}}), \tag{3.35}$$

and if $\mathrm{val}\,(\mathrm{D}_{\mathrm{strong}})$ is finite, then ($\mathrm{D}_{\mathrm{Ram}}$) has a solution with that value.

Lemma 4.

$$\mathrm{val}\,(\mathrm{D}_{\mathrm{strong}}) \geq \mathrm{val}\,(\mathrm{D}_{\mathrm{Ram}}).$$

Further, when $\mathrm{val}\,(\mathrm{D}_{\mathrm{strong}})$ *is finite,* ($\mathrm{D}_{\mathrm{Ram}}$) *has a solution with that value.*

Proof If ($\mathrm{D}_{\mathrm{strong}}$) is infeasible, then there is nothing to prove, so let us assume it is feasible. Further, assume that we rescaled $\mathcal{A}$ and B as stated in Lemma 2. By Lemma 3 we have that ($\mathrm{D}_{\mathrm{strong}}$) has an optimal solution, and we choose $Y^* \in (\mathcal{S}^r_+ \oplus 0)^*$ to be an optimal solution.

Let $Y_1, \dots, Y_k$ be the matrices we constructed in Lemma 2. Recall that $k \leq n$. We will construct a feasible solution of ($\mathrm{D}_{\mathrm{Ram}}$) with value equal to $\mathrm{val}\,(\mathrm{D}_{\mathrm{strong}})$.

First we outline the idea. From (3.26) we see that each Y_i can be written as $Y_i = U_i + V_i$, where U_i is psd, and $V_i \in \tan(U_{i-1})$. Precisely, we may choose $U_i := 0 \oplus I_{r_1 + \cdots + r_i}$ for all i, then $V_i \in \tan(U_{i-1})$ follows from part (3) of Proposition 1.

Then we can decompose Y^* as

$$Y^* = U_{k+1} + V_{k+1}, \text{ with } U_{k+1} \in \mathcal{S}^n_+, \text{ and } V_{k+1} \in \tan(U_k).$$

In particular, we can choose $U_{k+1} := Y^*(1:r) \oplus 0$ and $V_{k+1} := Y^* - U_{k+1}$. Then $V_{k+1} \in \tan(U_k)$ follows from part (3) of Proposition 1 and $r_1 + \cdots + r_k = n - r$.

Thus, if we set $U_0 = V_0 = 0,$ we obtain a feasible solution to a variant of ($\mathrm{D}_{\mathrm{Ram}}$) in which n is replaced by k.

This plan is not quite perfect, since k may be strictly less than n. If this happens, we need to modify our plan, namely we need to add some zero U_i and V_i at the start to create a feasible solution of ($\mathrm{D}_{\mathrm{Ram}}$).

We now carry out this modified plan.

(1) The first few U_i and V_i are "padding": we set

$$U_0 = V_0 = \cdots = U_{n-k} = V_{n-k} = 0.$$

Then (1.3), (1.4), (1.5), and (1.6) hold in Ramana's dual for $i \leq n-k$.

(2) Then from $Y_1, \ldots Y_k$ we construct $U_{n-k+1}, V_{n-k+1}, \ldots, U_n, V_n$. We write

$$Y_i = \underbrace{\begin{pmatrix} \overbrace{\qquad}^{n-\sum_{\ell=1}^{i} r_\ell} & \overbrace{\qquad}^{r_i} & \overbrace{\qquad}^{\sum_{\ell=1}^{i-1} r_\ell} \\ & I & \\ & & I \end{pmatrix}}_{U_{n-k+i}} + \underbrace{\begin{pmatrix} \overbrace{\qquad}^{n-\sum_{\ell=1}^{i} r_\ell} & \overbrace{\qquad}^{r_i} & \overbrace{\qquad}^{\sum_{\ell=1}^{i-1} r_\ell} \\ & & \times \\ & & \times \\ \times & \times & \times \end{pmatrix}}_{V_{n-k+i}}$$

for $i = 1, \ldots, k$. In other words, we let U_{n-k+i} as above, then set $V_{n-k+i} := Y_i - U_{n-k+i}$. Then by Part (3) in Proposition 1 we have

$$V_{n-k+i} \in \tan(U_{n-k+i-1}) \text{ for } i = 1, \ldots, k,$$

so (1.3), (1.4), (1.5), and (1.6) hold in Ramana's dual for $i = n-k+1, \ldots, n$. (It is useful to note that $V_{n-k+1} = 0$.)

(3) We finally split Y^* to construct U_{n+1} and V_{n+1} : we write Y^* as

$$Y^* = \underbrace{\begin{pmatrix} Y^*(1:r) & 0 \\ 0 & 0 \end{pmatrix}}_{U_{n+1}} + \underbrace{\begin{pmatrix} \overbrace{0}^{r} & \overbrace{\times}^{n-r} \\ \times & \times \end{pmatrix}}_{V_{n+1}}.$$

That is, we let U_{n+1} be as above, then set $V_{n+1} := Y^* - U_{n+1}$. Then $\mathcal{A}^*(U_{n+1} + V_{n+1}) = c$ and $U_{n+1} \in \mathcal{S}^n_+$. Also, by Part (3) of Proposition 1 and $U_n = 0 \oplus I_{n-r}$ we have $V_{n+1} \in \tan(U_n)$.

In summary, $U_0, V_0, \ldots, U_{n+1}, V_{n+1}$ is a feasible solution of ($\mathrm{D}_{\mathrm{Ram}}$) with value equal to val($\mathrm{D}_{\mathrm{strong}}$), hence the proof is complete. □

We invite the reader to follow the recipe in the proof above, and construct the optimal solution of the Ramana dual of the problem (1.1). This solution was already given in Example 2, but it is fruitful to produce it from the following ingredients: the maximum rank slack Z which is just the right hand side in (1.1); and the $Y = Y_1$ matrix in in (3.24) that certifies that Z indeed has maximum rank.

We next prove the inequality $\leq$ in (1.14).

Lemma 5. *We have*

$$\text{val}\,(\text{D}_{\text{strong}}) \le \text{val}\,(\text{D}_{\text{Ram}}). \tag{3.36}$$

Proof If (D_{Ram}) is infeasible, then there is nothing to prove, so assume it is feasible, and let $U_0, V_0, \ldots, U_{n+1}, V_{n+1}$ be a feasible solution.

As before, let Z be our maximum rank slack in (P). Since we can write $Z = B - \mathcal{A}x$ for some $x \in \mathbb{R}^m$, (1.3) and (1.4) imply

$$\langle Z, U_i + V_i \rangle = 0 \text{ for } i = 1, \ldots, n. \tag{3.37}$$

Since $V_1 = 0$, from (3.37) we deduce $\langle Z, U_1 \rangle = 0$, hence the first r rows and columns of U_1 are zero. Let Q_1 be an invertible matrix of suitably scaled eigenvectors of the lower right order $n - r$ block of U_1,

$$T_1 = \begin{pmatrix} I_r & 0 \\ 0 & Q_1 \end{pmatrix},$$

and apply the operations

$$\begin{array}{rcl} U_i & := & T_1^\top U_i T_1 \\ V_i & := & T_1^\top V_i T_1 \end{array} \tag{3.38}$$

for all i. Then we also apply the (Rescale) operations with $T_1^{-\top}$ in place of T. Afterwards the U_i and V_i are still feasible in (D_{Ram}) and have the same objective value as they had before: this follows from the proof of Proposition 2.

This rescaling does not change the maximum rank slack Z in (P). By Proposition 2, it also does not change the optimal value of (D_{Ram}). Finally, it does not change the value of (D_{strong}): this follows by an argument similar to the one proving that rescaling keeps the optimal value of (D) the same (this argument was given just before Proposition 2).

After this rescaling U_1 looks like

$$U_1 = \begin{pmatrix} 0 & 0 \\ 0 & I_{r_1} \end{pmatrix}, \text{ where } 0 \le r_1 \le n - r. \tag{3.39}$$

Since $V_2 \in \tan(U_1)$, by part (3) in Proposition 1 we deduce that only the last r_1 rows and columns of V_2 can be nonzero, hence $\langle Z, V_2 \rangle = 0$. Again using (3.37) we deduce $\langle Z, U_2 \rangle = 0$.

Next we perform the operations (3.38) with a suitable invertible matrix T_2 in place of T_1 to ensure $U_2 = 0 \oplus I_{r_2}$ for some $0 \le r_2 \le n - r$ [4]. We also apply the (Rescale) operations with $T_2^{-\top}$ in place of T to keep the U_i and V_i feasible.

Continuing, we produce matrices $T_3, \ldots, T_n$, and apply the operations (3.38) with $T_3, \ldots, T_n$ in place of T_1. We also apply the (Rescale) operations with $T_3^{-\top}, \ldots, T_n^{-\top}$ in place of T. Afterwards, we have

$$U_n = \begin{pmatrix} 0 & 0 \\ 0 & I_{r_n} \end{pmatrix}, \text{ where } 0 \le r_n \le n - r. \tag{3.40}$$

Since $V_{n+1} \in \tan(U_n)$, again using part (3) in Proposition 1 we deduce that only the last r_n rows and columns of V_{n+1} can be nonzero. Also, $U_{n+1} \succeq 0$, so $U_{n+1} + V_{n+1}$ is a matrix whose upper left order r block is psd, in other words $U_{n+1} + V_{n+1}$ is feasible in (D_{strong}).

The proof is now complete.

[4] Afterwards U_1 may not look like in equation (3.39) anymore, but for our purposes this does not matter.

□

Note that in Lemma 5 we actually proved a stronger result, than what is strictly needed to prove (3.36). Namely, we proved that after rescaling $\mathcal{A}$ and B, and applying suitable similarity transformations to the U_i and V_i (to keep them feasible in ($\mathrm{D}_{\mathrm{Ram}}$)), the matrix $U_{n+1} + V_{n+1}$ is feasible in ($\mathrm{D}_{\mathrm{strong}}$).

We can now prove the main result of the paper.

Proof of Theorem 1: We have that

$$\mathrm{val}\,(\mathrm{P}) \;=\; \mathrm{val}\,(\mathrm{D}_{\mathrm{strong}}) \;=\; \mathrm{val}\,(\mathrm{D}_{\mathrm{Ram}}), \tag{3.41}$$

where the first equality comes from Lemma 3 and the second from Lemmas 4 and 5. We note that (3.41) holds both when the optimal value of (P) is finite, and when it is $+\infty$.

Further, if the optimal value of (P) is finite, then by Lemma 3 the SDP ($\mathrm{D}_{\mathrm{strong}}$) has a solution with the same value; and by Lemma 4 the problem ($\mathrm{D}_{\mathrm{Ram}}$) has a solution with that value. This completes the proof. □

In this work we assumed that (P) is feasible. On the other hand, when (P) is infeasible, Ramana's dual can provide a certificate to verify its infeasibility: for details we refer the reader to [20].

3.4 SDP representation: proof of Theorem 2

Let us fix $U \succeq 0$ and let $\tan'(U)$ be the set on the right hand side of equation (1.10). For $W \in \mathbb{R}^{n\times n}$ and $\beta \in \mathbb{R},$ define the matrix

$$M(W,\beta) \;:=\; \begin{pmatrix} U & W \\ W^\top & \beta I \end{pmatrix}. \tag{3.42}$$

To prove $\tan(U) \supseteq \tan'(U),$ suppose $W + W^\top \in \tan'(U)$ and fix β such that $M(W,\beta) \succeq 0.$

We want to show $\mathcal{R}(W) \subseteq \mathcal{R}(U),$ so to obtain a contradiction, assume this is not the case. Then the nullspace of U is not contained in the nullspace of $W^\top,$ so we can choose x such that $Ux = 0$ and $W^\top x \neq 0.$

Further, we pick y such that $2x^\top W y + \beta\|y\|^2 < 0.$ Letting $z^\top := (x^\top, y^\top),$ we deduce

$$z^\top M(W,\beta) z \;=\; 2x^\top W y + \beta\|y\|^2 < 0,$$

the desired contradiction.

To show $\tan(U) \subseteq \tan'(U)$ suppose $W + W^\top \in \tan(U),$ i.e., $\mathcal{R}(W) \subseteq \mathcal{R}(U).$ Hence $W = UH$ for some matrix $H.$ Define

$$R \;:=\; \begin{pmatrix} I & -H \\ 0 & I \end{pmatrix}.$$

Then by an elementary calculation,

$$R^\top M(W,\beta) R \;=\; \begin{pmatrix} U & 0 \\ 0 & \beta I - H^\top U H \end{pmatrix},$$

so if β is large enough, then $\beta I - H^\top U H \succeq 0.$ Hence $M(W,\beta) \succeq 0,$ so $W + W^\top \in \tan'(U),$ completing the proof. □

4 Conclusion: a larger example, and the tangent space

We first present a larger example to illustrate Ramana's dual and walk the reader through all steps in our previous discussions.

Example 6. *Consider the SDP*

$$\begin{array}{lll} \sup & x_2 + x_3 & \\ s.t. & x_1 \underbrace{\begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}}_{A_1} + x_2 \underbrace{\begin{pmatrix} 0 & & 1 & \\ & 1 & & \\ 1 & & 0 & \\ & & & 0 \end{pmatrix}}_{A_2} + x_3 \underbrace{\begin{pmatrix} 0 & & & \\ & 0 & & 1 \\ & & 1 & \\ & 1 & & 0 \end{pmatrix}}_{A_3} \preceq \underbrace{\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}}_{B}. \end{array} \tag{4.43}$$

We will proceed as follows. We first show that this SDP has a positive duality gap. Then we calculate the key players in the paper: the maximum rank slack of (4.43) in the form given in Assumption 1; an optimal solution of the strong dual ($\mathrm{D_{strong}}$); the Y_i certificates for the maximum rank slack (given in Lemma 2); and finally, an optimal solution to Ramana's dual.

(1) To calculate the primal optimal value we let x be feasible in (4.43), and

$$S := \begin{pmatrix} 1 - x_1 & & -x_2 & \\ & 1 - x_2 & & -x_3 \\ -x_2 & & -x_3 & \\ & -x_3 & & 0 \end{pmatrix} \tag{4.44}$$

the corresponding slack matrix. Since the lower right corner of S is 0, and $S \succeq 0$, the last row and column of S is zero. Thus $x_3 = 0$, so the $(3,3)$ element of S is 0. Hence the third row and column of S are also zero, so $x_2 = 0$.

We have learned that in any feasible solution

$$x_2 = x_3 = 0, \tag{4.45}$$

so the objective function is identically zero on the primal feasible set.

(2) To calculate the dual optimal value, suppose $Y \in \mathcal{S}^4_+$ is feasible in the dual. Since $\langle A_1, Y \rangle = y_{11} = 0$, and $Y \succeq 0$, the first row (and column) of Y is zero. Thus, we have $\langle A_2, Y \rangle = y_{22} = 1$, so the dual objective function is identically 1 on the feasible set.

For example,

$$Y = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 0 \end{pmatrix} \tag{4.46}$$

is an optimal solution in the dual.

(3) Next we compute the optimal value of ($\mathrm{D_{strong}}$). Because of (4.45), the maximum rank slack in the SDP (4.43) is just the right hand side. Thus, in the strong dual ($\mathrm{D_{strong}}$) only the upper left 2×2 block of Y must be positive semidefinite. It follows that

$$Y^* := \begin{pmatrix} 0 & 0 & 0.5 & 0 \\ 0 & 0 & & \\ 0.5 & & 1 & \\ 0 & & & 0 \end{pmatrix} \tag{4.47}$$

is optimal in (D_{strong}) *with value zero.*

(4) *The Y_i matrices from Lemma 2 below certify that the right hand side in* (4.43) *is the maximum rank slack:*

$$\begin{aligned} Y_1 &:= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & & \\ 0 & & 0 & \\ 0 & & & 1 \end{pmatrix}, \\ Y_2 &:= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & & -1 \\ 0 & & 2 & \\ 0 & -1 & & 0 \end{pmatrix}. \end{aligned} \tag{4.48}$$

Indeed, the Y_i are of the form required in (3.26) *and satisfy the equations* (3.27) [5].

(5) *We finally construct a solution of* (D_{Ram}) *with value* 0*, following the proof of Lemma 4. We first set U_i and V_i to 0 for $i = 0, 1, 2$.*

Then we split the Y_i and Y^* to define the other U_i and V_i in (D_{Ram})*:*

$$\begin{aligned} Y_1 &= \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & & \\ 0 & & 0 & \\ 0 & & & 1 \end{pmatrix}}_{U_3} + \underbrace{0}_{V_3}, \\ Y_2 &:= \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & & \\ 0 & & 2 & \\ 0 & & & 0 \end{pmatrix}}_{U_4} + \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & & -1 \\ 0 & & 0 & \\ 0 & -1 & & 0 \end{pmatrix}}_{V_4}, \\ Y^* &:= \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & & \\ 0 & & 1 & \\ 0 & & & 0 \end{pmatrix}}_{U_5} + \underbrace{\begin{pmatrix} 0 & 0 & 0.5 & 0 \\ 0 & 0 & & \\ 0.5 & & 0 & \\ 0 & & & 0 \end{pmatrix}}_{V_5}. \end{aligned} \tag{4.49}$$

Note that by part (3) *in Proposition 1 we have $V_4 \in \tan(U_3)$ and $V_5 \in \tan(U_4)$, thus $U_0, V_0, \ldots, U_5, V_5$ is indeed a solution of* (D_{Ram}) *with value* 0.

In the last remark we describe the geometry of the mysterious looking set $\tan(U)$ that plays a crucial role in Ramana's dual.

Suppose $U \in \mathcal{S}^n_+$. Recall that we call $\tan(U)$ the *tangent space of $\mathcal{S}^n_+$ at U* and we gave a purely algebraic definition in (1.7). An equivalent geometric expression is

$$\tan(U) = \left\{ V \in \mathcal{S}^n \,\middle|\, \frac{1}{\epsilon} \operatorname{dist}(U \pm \epsilon V, \mathcal{S}^n_+) \to 0 \text{ as } \epsilon \searrow 0 \right\}, \tag{4.50}$$

[5] The following argument may better explain the role of the Y_i matrices. If S is any slack, then $\langle S, Y_1 \rangle = \langle S, Y_2 \rangle = 0$. We invite the reader to check that these equations lead to the same argument that we gave in paragraph (1) that show the last two rows and columns of S are zero.

where $\operatorname{dist}(X, \mathcal{S}_+^n) = \inf\{\|X - Y\| \,|\, Y \in \mathcal{S}_+^n\}$ is the distance of matrix X from $\mathcal{S}_+^n$ [6]. For a detailed proof of the equivalence, see e.g. [18, Lemma 3].

For example, if

$$U = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then $V \in \tan(U)$ according to the algebraic definition (1.7). We also have $V \in \tan(U)$ according to the geometric definition (4.50), since we only need to change the upper left corner of $U \pm \epsilon V$ to ϵ^2 to make it psd.

We illustrate this on Figure 1. We let C be the set of psd matrices with trace 1, and we describe C with just two parameters, as

$$C = \left\{ \begin{pmatrix} x & y \\ y & 1-x \end{pmatrix} \mid 1 \geq x \geq 0,\, x(1-x) - y^2 \geq 0 \right\}.$$

Since we can rewrite the quadratic inequality as $(x - \frac{1}{2})^2 + y^2 \leq \frac{1}{4}$, the set C is a circle of radius $\frac{1}{2}$ centered at $(\frac{1}{2}, 0)$. Figure 1 shows C together with U and $U + \epsilon V$ for a small $\epsilon > 0$.

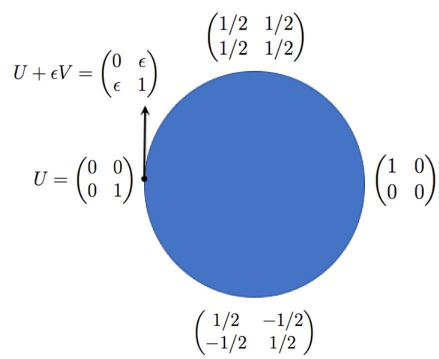

Figure 1: The matrix $U + \epsilon V$ is "almost" psd, but not quite

Nevertheless, it is interesting that one can completely derive Ramana's dual using only purely linear algebraic arguments about $\tan(U)$, rather than any geometric ones, and this is the route we followed.

Acknowledgements

Part of this paper was written during a visit to Chapel Hill by the first author and he would like to express his gratitude to Prof. Takashi Tsuchiya, for creating conditions that made this visit possible. The work of the first author was partially supported by the JSPS KAKENHI Grant Numbers JP15H02968 and JP19K20217.

The second author thanks Pravesh Kothari and Ryan O' Donnell for helpful discussions on SDP. The work of the second author was supported by the National Science Foundation, award DMS-1817272.

Both authors are grateful to Siyuan Chen, Alex Touzov, and Yuzixuan Zhu for their careful reading of the manuscript and their helpful comments.

[6] We can use any matrix norm, for example the spectral norm or the Frobenius norm.

Most importantly, we are very grateful to the anonymous referees whose comments and suggestions greatly improved the manuscript.

References

[1] Aharon Ben-Tal and Arkadii Nemirovskii. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. SIAM, Philadelphia, PA, 2001. 9

[2] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples, Second Edition.* CMS Books in Mathematics. Springer, 2005. 9

[3] Jonathan M. Borwein and Henry Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Aust. Math. Soc.*, 30:369–380, 1981. 2

[4] Jonathan M. Borwein and Henry Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. App.*, 83:495–530, 1981. 2, 13

[5] Etienne De Klerk. *Aspects of semidefinite programming: interior point algorithms and selected applications*, volume 65. Springer Science & Business Media, 2006. 2

[6] Etienne de Klerk, Tamás Terlaky, and Kees Roos. Self-dual embeddings. In *Handbook of Semidefinite Programming*, pages 111–138. Springer, 2000. 2

[7] Dmitriy Drusvyatskiy, Henry Wolkowicz, et al. The many faces of degeneracy in conic optimization. *Foundations and Trends® in Optimization*, 3(2):77–170, 2017. 2, 3

[8] 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. 2, 3

[9] Monique Laurent and Franz Rendl. Semidefinite programming and integer programming. *Handbooks in Operations Research and Management Science*, 12:393–514, 2005. 2

[10] Minghui Liu and Gábor Pataki. Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming. *Mathematical Programming*, 167(2):435–480, 2018. 2

[11] 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. 2

[12] László Lovász. *Graphs and geometry*, volume 65. American Mathematical Soc., 2019. 2

[13] 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. 2, 9

[14] Arkadi Nemirovski. Advances in convex optimization: conic programming. In *International Congress of Mathematicians*, volume 1, pages 413–444, 2007. 2

[15] Ryan O'Donnell. SOS is not obviously automatizable, even approximately. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. 2

[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. 2, 3, 13

[17] Gábor Pataki. Characterizing bad semidefinite programs: normal forms and short proofs. *SIAM Review*, 61(4):839–859, 2019. 2

[18] Gábor Pataki. Bad semidefinite programs: they all look the same. *SIAM J. Opt.*, 27(1):146–172, 2017. 20

[19] Frank Permenter and Pablo Parrilo. Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone. *Mathematical Programming*, pages 1–54, 2014. 13

[20] Motakuri V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Math. Program. Ser. B*, 77:129–162, 1997. 1, 2, 17

[21] Motakuri V. Ramana and Robert Freund. On the ELSD duality theory for SDP. Technical report, MIT, 1996. 2

[22] Motakuri V. Ramana, Levent Tunçel, and Henry Wolkowicz. Strong duality for semidefinite programming. *SIAM J. Opt.*, 7(3):641–662, 1997. 2, 4

[23] James Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* MPS-SIAM Series on Optimization. SIAM, Philadelphia, USA, 2001. 9

[24] Tyrrell R. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, USA, 1970. 9

[25] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996. 2

[26] Hayato Waki and Masakazu Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013. 3, 13

[27] Yuzixuan Zhu, Gábor Pataki, and Quoc Tran-Dinh. Sieve-SDP: a simple facial reduction algorithm to preprocess semidefinite programs. *Mathematical Programming Computation*, 11(3):503–586, 2019. 13