Closing Duality Gaps of SDPs through Perturbation[*]

Takashi Tsuchiya[†]   Bruno F. Lourenço[‡]   Masakazu Muramatsu[§]

Takayuki Okuno[¶]

April 2023

Abstract

Let $(\mathbf{P}, \mathbf{D})$ be a primal-dual pair of SDPs with a nonzero finite duality gap. Under such circumstances, $\mathbf{P}$ and $\mathbf{D}$ are weakly feasible and if we perturb the problem data to recover strong feasibility, the (common) optimal value function v as a function of the perturbation is not well-defined at zero (unperturbed data) since there are "two different optimal values" $v(\mathbf{P})$ and $v(\mathbf{D})$, where $v(\mathbf{P})$ and $v(\mathbf{D})$ are the optimal values of $\mathbf{P}$ and $\mathbf{D}$ respectively. Thus, continuity of v is lost at zero though v is continuous elsewhere. Nevertheless, we show that a limiting version v_a of v is a well-defined monotone decreasing continuous bijective function connecting $v(\mathbf{P})$ and $v(\mathbf{D})$ with domain $[0, \pi/2]$ under the assumption that both $\mathbf{P}$ and $\mathbf{D}$ have singularity degree one. The domain $[0, \pi/2]$ corresponds to directions of perturbation defined in a certain manner. Thus, v_a "completely fills" the nonzero duality gap under a mild regularity condition. Our result is tight in that there exists an instance with singularity degree two for which v_a is not continuous.

Keywords: Semidefinite programs, nonzero duality gaps, perturbation, regularization, facial reduction

[†]National Graduate Institute for Policy Studies, 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677 Japan, e-mail: tsuchiya@grips.ac.jp

[‡]The Institute of Statistical Mathematics, Midori-cho 10-3, Tachikawa, 190-8562 Tokyo Japan, e-mail: bruno@ism.ac.jp

[§]The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585 Japan, e-mail: MasakazuMuramatsu@uec.ac.jp

[¶]Faculty of Science and Technology Department of Science and Technology, Seikei University, 3-3-1, Kita, Kichijoji, Musashino, Tokyo, 180-8633, Japan; Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027 Japan. email: takayuki-okuno@st.seikei.ac.jp

1 Introduction

Consider the standard form dual pair of semidefinite programs:

$$\min_{X} \ C \bullet X \ \text{ s.t. } A^i \bullet X = b_i, \ i = 1, \ldots, m, X \succeq 0 \tag{P}$$

$$\max_{y,S} \ b^T y \quad \text{s.t. } C - \sum_{i=1}^{m} A^i y_i = S, \ S \succeq 0, \tag{D}$$

where C, $A^i, i = 1, \ldots, m$, X, and S are real symmetric $n \times n$ matrices and $y \in \mathbb{R}^m$. We denote the optimal values of **P** and **D** by $v(\mathbf{P})$ and $v(\mathbf{D})$, respectively. We use analogous notation throughout this paper to represent optimal values. We assume that **P** and **D** are feasible but not necessarily strongly feasible, i.e., neither **P** nor **D** satisfy Slater's condition. Under this assumption, **P** and **D** may have a finite nonzero duality gap as shown in the following famous example adapted from Ramana [12]:

Example 1.1 ([12, Example 4]) *The problem* **D** *is*

$$\max \ y_1 \ \text{s.t.} \ \begin{pmatrix} 1 - y_1 & 0 & 0 \\ 0 & -y_2 & -y_1 \\ 0 & -y_1 & 0 \end{pmatrix} \succeq 0.$$

With that, we have

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A^1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad b_1 = 1.$$

We have $v(\mathbf{D}) = 0$ for this problem, because $y_1 = 0$ is the only possible value for the lower-right 2×2 submatrix to be positive semidefinite.

The associated primal **P** *is*

$$\min \ x_{11} \ \text{ s.t. } x_{11} + 2x_{23} = 1, \ x_{22} = 0, \ \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \succeq 0.$$

We have $v(\mathbf{P}) = 1$ for this problem, because $x_{23} = 0$ must hold for positive semidefiniteness of the lower-right 2×2 submatrix, which drives x_{11} to be 1.

In general, **P** and **D** might fail to have interior feasible solutions and this opens the possibility of nonzero duality gaps as in Example 1.1. On the other hand, most algorithms for SDP are constructed under the assumption that **P** and **D** have interior feasible solutions. This leads naturally to the question of how to solve SDPs with finite nonzero duality gaps with the usual algorithms. A simple approach to fix that is to perturb the problem data so that constraint qualifications are satisfied [11, 15, 17]. We can consider, for example, the following perturbed/regularized system $\mathbf{P}(\varepsilon, \eta)$ and $\mathbf{D}(\varepsilon, \eta)$ from [17].

Regularized Primal-Dual Standard Form SDP (RPD-SDP)

$$\mathbf{P}(\varepsilon, \eta): \quad \min_{X} (C + \varepsilon I) \bullet X \quad \text{s.t. } A^i \bullet X = b_i + \eta A^i \bullet I, \ i = 1, \ldots, m, \ X \succeq 0 \qquad (1.1)$$

and

$$\mathbf{D}(\varepsilon, \eta): \quad \max_{y, S} \sum_{i=1}^{m} (b_i + \eta A^i \bullet I) y_i \quad \text{s.t. } C - \sum_{i=1}^{m} A^i y_i + \varepsilon I = S, \quad S \succeq 0, \qquad (1.2)$$

where I denotes the $n \times n$ identity matrix. We call the pair (1.1) and (1.2) the *Regularized Primal-Dual Standard Form SDP or RPD-SDP for short*. $\mathbf{P}(\varepsilon, \eta)$ and $\mathbf{D}(\varepsilon, \eta)$ reduce to $\mathbf{P}$ and $\mathbf{D}$ when ε and η are set to zero. RPD-SDP is obtained by relaxing the semidefinite constraints $X \succeq 0$ of $\mathbf{P}$ and $S \succeq 0$ of $\mathbf{D}$ to $X \succeq -\eta I$ and $S \succeq -\varepsilon I$, respectively, and by redefining $X := X + \eta I$ and $S := S + \varepsilon I$.

Under the assumption that $\mathbf{P}$ and $\mathbf{D}$ are feasible, the perturbed problems admit interior feasible solutions for any $\varepsilon > 0$ and $\eta > 0$, so, in this sense, the lack of interior solutions of $\mathbf{P}$ and $\mathbf{D}$ is fixed. However, this is only useful if something can be said about how the optimal values of $\mathbf{D}(\epsilon, \eta)$ and $\mathbf{P}(\epsilon, \eta)$ relate to the optimal values of the original $\mathbf{P}$ and $\mathbf{D}$, so let us briefly examine this issue.

For $\varepsilon > 0$ and $\eta > 0$, Slater's condition is satisfied at both (1.1) and (1.2), so they have optimal solutions and a common optimal value, which we denote by $v(\varepsilon, \eta)$. In the sequel, $v(\cdot, \cdot)$ is referred to as *pd-regularized optimal value function*. We note that $v(\varepsilon, 0)$ and $v(0, \eta)$ are also well-defined for any $\varepsilon > 0$ and $\eta > 0$, since $\mathbf{P}(\varepsilon, 0)$ and $\mathbf{D}(0, \eta)$ have interior feasible solution for any $\varepsilon > 0$ and $\eta > 0$ and there is no duality gap in these cases according to the standard duality theory for convex programming.

But $v(0, 0)$ is different since it is not well-defined when there exists a finite nonzero duality gap, and ironically, *the value of $v(0, 0)$ is what we really wish to compute*. Thus, the perturbation/regularization approach might not be theoretically sound when there are nonzero duality gaps.

That said, we have recently analyzed the behavior of the pd-regularized optimal value function $v(\varepsilon, \eta)$ in the neighbourhood of $(\varepsilon, \eta) = (0, 0)$ and demonstrated that $v(\varepsilon, \eta)$ have a directional limit when approaching $(0, 0)$, see [17]. Let us define the directional limit

$$v_a(\theta) := \lim_{t \downarrow 0} v(t \cos \theta, t \sin \theta). \qquad (1.3)$$

The function v_a is referred to as *limiting pd-regularized optimal value function*. Then, the following theorem holds.

Theorem 1.2 ([17, Theorem 2]) *If $\mathbf{P}$ and $\mathbf{D}$ are feasible, the limiting pd-regularized optimal value function $v_a(\theta)$ has the following properties.*

1. $v_a(0) = v(\mathbf{P})$, $v_a(\pi/2) = v(\mathbf{D})$.

2. $v_a(\theta)$ is monotone decreasing in $[0, \pi/2]$ and is continuous on $(0, \pi/2)$.

In the case of Example 1.1, $v_a(\theta)$ is written as follows:

$$v_a(\theta) = \begin{cases} 1 - \tan\theta & \text{if } 0 \le \tan\theta \le \frac{1}{2}, \\ \frac{1}{4}\cot\theta & \text{if } 1/2 \le \tan\theta. \end{cases} \tag{1.4}$$

See [17, Example 1] for its derivation[1]. In this instance, it can be shown that v_a is a strictly monotone decreasing continuous function on the closed interval $[0, \pi/2]$. This implies that, for each value $w \in [v(\mathbf{D}), v(\mathbf{P})]$, there exists a unique angle $\theta \in [0, \pi/2]$ such that $v_a(\theta) = w$ holds. Thus, the limiting pd-regularized optimal value function v_a is a continuous bijection from $[0, \pi/2]$ to $[v(\mathbf{D}), v(\mathbf{P})]$. This shows that the gap is only superficial and hidden continuity behind the gap was brought to the surface. We may say that the nonzero duality gap is "filled completely" because we may reach any value in $[v(\mathbf{D}), v(\mathbf{P})]$ by appropriately selecting θ. It would be nice if this structure existed for any SDP with a finite nonzero duality gap.

Theorem 1.2, however, does not exclude the possibility that v_a is discontinuous on the boundary of domain at $\theta = 0$ and $\theta = \pi/2$. In particular, the continuity of $v_a(\theta)$ over the entire interval $[0, \pi/2]$ was left open in [17]. If v_a turns out to be continuous on the both end points $\theta = 0$ and $\theta = \pi/2$, then, v_a is a continuous function from $[0, \pi/2]$ to $[v(\mathbf{D}), v(\mathbf{P})]$, and as a consequence, by controlling the direction to which $(t\cos\theta, t\sin\theta)$ approaches $(0, 0)$ we may force $v_a(\theta)$ to assume any value in $[v(\mathbf{D}), v(\mathbf{P})]$.

Now we are ready to state our goal in this paper. We presents two results under the assumption of a finite nonzero duality gap, one is positive and the other negative. First, we will show that v_a is continuous at $\theta = 0$ and $\theta = \pi/2$ if the *singularity degree* of $\mathbf{D}$ and $\mathbf{P}$ are both one, and consequently, the limiting pd-regularized optimal value function v_a is a continuous bijective function from $[0, \pi/2]$ to $[v(\mathbf{D}), v(\mathbf{P})]$ in that case. Here, the singularity degree for a feasible SDP is defined as the minimum number of facial reduction steps necessary to regularize the problem (see Section 2). Then, second, we present an instance where v_a is discontinuous at $\theta = \pi/2$, thus showing that continuity cannot be expected to hold in general.

This paper is organized as follows. In Section 2, we discuss the notation and some prerequisite notions. In Section 3, we analyze the continuity of v_a under the assumption that the problems have singularity degree one. In Section 4 we present an instance to show that v_a may be discontinuous in general. Section 5 presents some concluding remarks.

2 Preliminaries

In this section, we introduce some terminology and mathematical definitions. The space of real $p \times p$ symmetric matrices will be denoted by $\mathcal{S}^p$. The space of $p \times q$ real matrices will be denoted by $\mathbb{R}^{p \times q}$. The cone of semidefinite matrices in $\mathcal{S}^p$ will be denoted by $\mathcal{S}^p_+$. The p-dimensional nonnegative orthant will be denoted by $\mathbb{R}^p_+$. In this paper, we are using the

---

[1]The formula obtained in [17] is slightly different than (1.4). But we obtain exactly the same formula by letting $\alpha = \cos\theta$ and $\beta = \sin\theta$.

convention that whenever the identity matrix I appears in a expression it has the "correct size" as to make the expression well-defined.

We recall that $\mathbf{P}$ is said to be *strongly feasible* if there exists a positive definite feasible solution X to $\mathbf{P}$. This is the same as saying that $\mathbf{P}$ satisfies *Slater's condition*. $\mathbf{P}$ is said to be *weakly feasible* if it is feasible but not strongly feasible. Similar definitions apply to $\mathbf{D}$.

In semidefinite programming it sometimes happens that the optimal value of $\mathbf{P}$ or $\mathbf{D}$ is finite, but no optimal solution exists[2]. In any case, provided that $\mathbf{P}$ is feasible, there exists a sequence of matrices $\{X^k\}$ such that the X^k are feasible to $\mathbf{P}$ and $C \bullet X^k$ converges to $v(\mathbf{P})$. We say that such a sequence is *an optimal sequence* (for $\mathbf{P}$). Analogously, an optimal sequence for $\mathbf{D}$ is a sequence $\{(y^k, S^k)\}$ for which $C - \sum_{i=1}^{m} A^i y_i^k = S^k$, $S^k \succeq 0$ and $b^T y^k \to v(\mathbf{D})$ holds.

Singularity Degree   Let $\mathcal{K} \subseteq \mathcal{E}$ be a closed convex cone contained in a finite-dimensional Euclidean space $\mathcal{E}$ equipped with an inner product $\langle \cdot, \cdot \rangle$. For a convex set $\mathcal{C} \subseteq \mathcal{E}$, let $\mathrm{ri}\,\mathcal{C}$ denote its relative interior. Also, let $\mathcal{C}^*$ and $\mathcal{C}^\perp$ denote the dual cone and orthogonal complement of $\mathcal{C}$ with respect to $\langle \cdot, \cdot \rangle$, respectively.

Let $\mathcal{V} \subseteq \mathcal{E}$ be an affine space such that $\mathcal{V} \cap \mathcal{K} \neq \emptyset$ and consider the following feasibility problem:

$$\text{find} \quad x \in \mathcal{V} \cap \mathcal{K}. \tag{Feas}$$

In this case, there exists an unique *minimal face* $\mathcal{F}$ of $\mathcal{K}$ with the property that $(\mathcal{V} \cap \mathcal{K}) \subseteq \mathcal{F}$ and $\mathcal{V} \cap (\mathrm{ri}\,\mathcal{F}) \neq \emptyset$. When (Feas) satisfies Slater's condition (i.e., $\mathcal{V} \cap (\mathrm{ri}\,\mathcal{K}) \neq \emptyset$), then the face $\mathcal{F}$ is $\mathcal{K}$ itself.

The process of finding the minimal face $\mathcal{F}$ is known as *facial reduction* and is one of the standard approaches for handling conic linear programs that fail to satisfy constraint qualifications, e.g., [2–4,7,9,16,19]. While the original problem might suffer from pathologies arising from a lack of constraint qualifications, identifying $\mathcal{F}$ makes it possible to construct a new equivalent problem where the absence of constraint qualifications is fixed.

The basic facial reduction algorithm as described, say, in [19], is based on the following observation. We have $\mathcal{V} \cap (\mathrm{ri}\,\mathcal{K}) = \emptyset$ (i.e., Slater's condition fails) if and only if $\mathcal{V}$ and $\mathrm{ri}\,\mathcal{K}$ can be *properly separated* by a hyperplane that does not contain $\mathrm{ri}\,\mathcal{K}$ entirely, see [13, Theorem 20.2]. Under the hypothesis that $\mathcal{V} \cap \mathcal{K} \neq \emptyset$, this implies the existence of $s \in (\mathcal{V}^\perp \cap \mathcal{K}^*) \setminus \mathcal{K}^\perp$, e.g., see [19, Lemma 3.2]. Then, letting

$$\mathcal{F} := \mathcal{K} \cap \{s\}^\perp,$$

$\mathcal{F}$ is a face of $\mathcal{K}$ strictly contained in $\mathcal{K}$ (since $s \notin \mathcal{K}^\perp$) with the property that $\mathcal{V} \cap \mathcal{K} \subseteq \mathcal{F}$. If it turns out that $\mathcal{V} \cap \mathrm{ri}\,\mathcal{F} \neq \emptyset$, then we are done. Otherwise, we can apply the same separation result again to $\mathcal{V}$ and $\mathcal{F}$ and repeat the process. Since the dimension of faces decrease at each step, this process must end in a finite number of steps, e.g., [19, Theorem 3.2]. This leads to a chain of ℓ faces of $\mathcal{K}$

$$\mathcal{F}_\ell \subsetneq \cdots \subsetneq \mathcal{F}_1 = \mathcal{K}$$

---

[2]Consider, for example, $\min x_{11}$ s.t. $\begin{pmatrix} x_{11} & 1 \\ 1 & x_{22} \end{pmatrix} \in \mathcal{S}_+^2$. The optimal value is 0 but not attained.

and $\ell - 1$ reducing directions $\{s_1, \ldots, s_{\ell-1}\}$ such that:

(i) for all $i \in \{1, \ldots, \ell - 1\}$, we have

$$s_i \in \mathcal{F}_i^* \cap \mathcal{V}^\perp \quad and \quad \mathcal{F}_{i+1} = \mathcal{F}_i \cap \{s_i\}^\perp. \tag{2.1}$$

(ii) $\mathcal{F}_\ell \cap \mathcal{V} = \mathcal{K} \cap \mathcal{V}$ and $\mathcal{V} \cap (\mathrm{ri}\, \mathcal{F}_\ell) \neq \emptyset$.

We note that ℓ must be bounded above by the dimension of $\mathcal{K}$, see also [7] for tighter bounds. Furthermore, item (ii) implies that $\mathcal{F}_\ell$ is the minimal face of $\mathcal{K}$ containing $\mathcal{K} \cap \mathcal{V}$.

There is freedom in the choice of the s_i in item (i), so difference choices of s_i's might lead to a smaller or larger ℓ. The *singularity degree* of (Feas) is the minimal number of *facial reduction steps* (i.e., the ℓ) in order to find the minimal face of (Feas) as in items (i) and (ii). The singularity degree was discussed extensively by Sturm in [16], although Sturm's definition of singularity degree is slightly different from the most current usage of the term, see [7, Footnote 3].

For the main result of this paper, we will only need a discussion of problems having singularity degree one, so we shall focus on that. (Feas) is said to have *singularity degree one* if it is feasible and there exists $s \in (\mathcal{V}^\perp \cap \mathcal{K}^*) \setminus \mathcal{K}^\perp$ such that

$$\mathcal{V} \cap (\mathrm{ri}\, (\mathcal{K} \cap \{s\}^\perp)) \neq \emptyset.$$

In the following, we discuss what this definition means for the problems $\mathbf{P}$ and $\mathbf{D}$. Letting $\mathcal{E} := \mathcal{S}^n$, $\mathcal{K} := \mathcal{S}_+^n$ and $\mathcal{V} := C + \mathcal{L}$, where $\mathcal{L}$ is the span of the A^i in $\mathbf{D}$, the problem (Feas) correspond to the feasible "slacks" associated to dual problem $\mathbf{D}$. Then, $\mathbf{D}$ has singularity degree one if and only if there exists a nonzero $X \in \mathcal{S}_+^n$, with $C \bullet X = 0$ and $A^i \bullet X = 0$ for all i in such a way that there exists $y \in \mathbb{R}^m$ satisfying

$$C - \sum_{i=1}^m A^i y_i \in \mathrm{ri}\, (\mathcal{K} \cap \{X\}^\perp).$$

Before we move further, we need to recall a classical characterization of the faces of $\mathcal{S}_+^n$. Namely, every face $\mathcal{F}$ of $\mathcal{S}_+^n$ is linearly isomorphic to a smaller positive semidefinite cone. Furthermore, there exists $r \leq n$ and a nonsingular matrix V such that

$$V \mathcal{F} V^T = \left\{ X \in \mathcal{S}^n \mid X = \begin{pmatrix} Y & 0 \\ 0 & 0 \end{pmatrix}, Y \in \mathcal{S}_+^r \right\}, \tag{2.2}$$

e.g., see [8] and also [1, Section 6].

Now we are ready to state the following proposition which plays a fundamental role in our analysis.

Proposition 2.1 *Suppose that $\mathbf{D}$ is feasible. Then the following statements hold.*

1. After appropriate rescaling[3], there exists $r \leq n$ such that defining $L(y) := C - \sum_{i=1}^{m} A^i y_i$, the following items hold:

   (a) $L(y) \in \mathcal{S}_+^n$ if and only if

   $$L(y) \in \left\{ X \in \mathcal{S}^n \mid X = \begin{pmatrix} Y & 0 \\ 0 & 0 \end{pmatrix}, Y \in \mathcal{S}_+^r \right\}.$$

   (b) (Slater's condition is satisfied for the reduced problem) There exists $y \in \mathbb{R}^m$ such that $L(y) = \begin{pmatrix} Y & 0 \\ 0 & 0 \end{pmatrix}$, $Y \in \mathrm{ri}\,(\mathcal{S}_+^r)$ (i.e., $Y \succ 0$).

   In particular, defining $L_{11} : \mathbb{R}^m \to \mathcal{S}^r, L_{12} : \mathbb{R}^m \to \mathbb{R}^{r \times (n-r)}, L_{22} : \mathbb{R}^m \to \mathcal{S}^{n-r}$ so that

   $$L(y) = \begin{pmatrix} L_{11}(y) & L_{12}(y) \\ L_{12}^T(y) & L_{22}(y) \end{pmatrix}, \qquad \forall y \in \mathbb{R}^m,$$

   D is equivalent to the following strongly feasible problem

   $$\mathbf{RD} \quad \max_y \ b^T y, \ \text{s.t.} \ L_{11}(y) \succeq 0, \ L_{12}(y) = 0, \ L_{22}(y) = 0$$

   and the optimal value of **RD** is equal to $v(\mathbf{D})$.

2. Under the setting of item 1, let

   $$\mathcal{T} := \left\{ tI = t \begin{pmatrix} I_{11} & 0 \\ 0 & I_{22} \end{pmatrix} \in \mathcal{S}^n \ \middle|\ t \in \mathbb{R} \right\},$$

   $$\mathcal{L} := \left\{ \begin{pmatrix} 0 & L_{12}(\hat{y}) \\ L_{12}(\hat{y})^T & L_{22}(\hat{y}) \end{pmatrix} \in \mathcal{S}^n \ \middle|\ \hat{y} \in \mathbb{R}^m \right\},$$

   where $I_{11} \in \mathcal{S}^r$ and $I_{22} \in \mathcal{S}^{n-r}$ are identity matrices. Note that $\mathcal{L}$ is a linear space because it contains the zero matrix (see **RD**). We take $\mathcal{T} + \mathcal{L} = \mathcal{T} \oplus \mathcal{L}$ as perturbation space, and consider the following SDP

   $$\mathbf{RD}(S) \quad \max_y \ b^T y, \ \text{s.t.} \ L_{11}(y) + s_{11} I_{11} \succeq 0, \ L_{12}(y) = S_{12}, \ L_{22}(y) + s_{11} I_{22} = S_{22}$$

   obtained by adding the perturbation

   $$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix} = s_{11} I + \begin{pmatrix} 0 & S_{12} \\ S_{12}^T & S_{22} - s_{11} I_{22} \end{pmatrix} \in \mathcal{T} \oplus \mathcal{L}$$

   to **RD**, where $S_{11} \in \mathcal{S}^r$, $S_{22} \in \mathcal{S}^{n-r}$ and $S_{12} \in \mathbb{R}^{r \times (n-r)}$, and $s_{11} \in \mathbb{R}$ is the $(1,1)$-element of S. We note that S is decomposed as the sum of $s_{11} I \in \mathcal{T}$ and $S - s_{11} I \in \mathcal{L}$, which implies that $L_{22}(\hat{y}) = S_{22} - s_{11} I_{22}$ holds for some $\hat{y} \in \mathbb{R}^m$.

   Let $w(S)$ be the optimal value function of the perturbed system[4] $\mathbf{RD}(S)$. Then, $w(0) = v(\mathbf{D})$ and $w(S)$ is continuous at $S = 0$.

---

[3]Rescaling **D** corresponds to selecting a nonsingular matrix V and replacing the C and the A^i with VCV^T and VA^iV^T. This transformation preserves the set of y that are feasible for **D** and also preserves the presence (or absence) of duality gaps. For the purposes of this paper, there is no difference between analyzing a problem or its rescalings.

[4]Compared with **RD**, the first constraint of $\mathbf{RD}(S)$ is perturbed by $S_{11} = s_{11} I_{11}$, the second is shifted by S_{12} from 0, and the third is shifted by $S_{22} - s_{11} I_{22}$ from 0.

3. *Under the setting of item 1, if the singularity degree of $\mathbf{D}$ is one, then there exists a nonzero $X \in \mathcal{S}^n_+$ of rank $n - r$ satisfying*

$$C \bullet X = 0, \quad A^i \bullet X = 0, \ i = 0, 1, \ldots, m, \quad X \succeq 0 \tag{2.3}$$

with the form

$$X = \begin{pmatrix} 0 & 0 \\ 0 & X_{22} \end{pmatrix}, \quad X_{22} \succ 0. \tag{2.4}$$

Proof. Items 1 and 3 are an amalgam of well-known results about $\mathcal{S}^n_+$ and minimal faces, so we only present a sketch of the proof. Let $\mathcal{V} := C + \mathcal{L}$, where $\mathcal{L}$ is the span of the A^i in $\mathbf{D}$. Let $\mathcal{F}$ the minimal face of $\mathcal{S}^n_+$ containing $\mathcal{V} \cap \mathcal{S}^n_+$. As discussed previously, such a face must be as in (2.2) and have the property that $(\mathcal{V} \cap \mathcal{S}^n_+) \subseteq \mathcal{F}$ and $\mathcal{V} \cap (\mathrm{ri}\,\mathcal{F}) \neq \emptyset$. Rescaling the C and A^i using V leads to the proof of item 1.

As for item 3, if the singularity degree is 1, then $\mathcal{F} = \mathcal{K} \cap \{X\}^\perp$, where $X \in \mathcal{S}^n_+ \cap \mathcal{V}^\perp$ is nonzero and satisfies $\mathcal{V} \cap (\mathrm{ri}\,(\mathcal{S}^n_+ \cap \{X\}^\perp))$. In view of (2.2), X has rank $n - r$ and satisfies (2.3) and (2.4).

It remains to prove item 2. Since $w(0)$ corresponds to the unperturbed problem, we have $w(0) = v(\mathbf{D})$. Then, since $\mathbf{RD}$ satisfies Slater's condition, continuity of $w(S)$ at $S = 0$ in item 2 is obtained as a consequence of Theorem 4.1.9 of [20]. For the sake of completeness, we describe a detailed proof in Appendix A. ∎

We note that even though Proposition 2.1 is presented for problems in dual format, a completely analogous discussion can be done for $\mathbf{P}$ if we let $\mathcal{K} := \mathcal{S}^n_+$, $\mathcal{V} := \{X \in \mathcal{S}^n \mid A^i \bullet X = b_i, i = 1, \ldots, m\}$ in (Feas).

3 Main Result and Proof

In this section, we show continuity of the limiting pd-regularized optimal value function v_a defined in (1.3) at $\theta = 0$ and $\theta = \pi/2$ under the assumption that the singularity degree of $\mathbf{P}$ and $\mathbf{D}$ is one. We also show bijectivity of v_a. More precisely, we prove the following.

Theorem 3.1 *Suppose that $\mathbf{P}$ and $\mathbf{D}$ are feasible. Then the following statements hold on the limiting pd-regularized optimal value function $v_a : [0, \pi/2] \to \mathbb{R}$.*

1. *If the singularity degree of $\mathbf{D}$ is one, then $v_a(\theta)$ is continuous at $\theta = \pi/2$.*

2. *If the singularity degree of $\mathbf{P}$ is one, then $v_a(\theta)$ is continuous at $\theta = 0$.*

3. *If the singularity degree of both $\mathbf{P}$ and $\mathbf{D}$ is one, then $v_a(\theta)$ is continuous at $\theta = 0$ and $\theta = \pi/2$. Furthermore, v_a is a monotonically decreasing continuous bijective function from $[0, \pi/2]$ to $[v(\mathbf{D}), v(\mathbf{P})]$.*

In [17], we introduced the function $\tilde{v} : \mathbb{R}_+ \cup \{\infty\} \to \mathbb{R}$ as

$$\tilde{v}(\beta) := \lim_{t \downarrow 0} v(t, t\beta) \ (0 \leq \beta < \infty), \ \tilde{v}(\infty) := \lim_{t \downarrow 0} v(0, t). \tag{3.1}$$

In what follows, we define the function $\bar{v} : \mathbb{R}_+ \cup \{\infty\} \to \mathbb{R}$ given by

$$\bar{v}(\alpha) := \lim_{t \downarrow 0} v(t\alpha, t) \ (0 \le \alpha < \infty), \quad \bar{v}(\infty) := \lim_{t \downarrow 0} v(t, 0). \tag{3.2}$$

The existence of the limit in (3.1) and (3.2) is shown in [17, Theorem 1]. We see, under the convention that $1/0 = \infty$ and $1/\infty = 0$, that

$$\bar{v}(\alpha) = \tilde{v}\left(\frac{1}{\alpha}\right), \quad \tilde{v}(\beta) = \bar{v}\left(\frac{1}{\beta}\right). \tag{3.3}$$

It also follows that

$$v_a(\theta) = \lim_{t \downarrow 0} v(t \cos\theta, t \sin\theta) = \tilde{v}(\tan\theta) = \bar{v}(\cot\theta). \tag{3.4}$$

We provide the following proposition on a few fundamental properties of $\tilde{v}$ and $\bar{v}$.

Proposition 3.2 *The following items hold.*

1. *$\tilde{v}(\beta)$ is monotone decreasing on $\mathbb{R}_+ \cup \{\infty\}$, convex on $\mathbb{R}_+$ and continuous on $\mathbb{R}_+ \backslash \{0\}$.*

2. *$\bar{v}(\alpha)$ is monotone increasing on $\mathbb{R}_+ \cup \{\infty\}$, concave on $\mathbb{R}_+$ and continuous on $\mathbb{R}_+ \backslash \{0\}$.*

Proof. Monotonicity and convexity of $\tilde{v}$ is given in [17, Theorem 4]. Given the convexity of $\tilde{v}$ on $\mathbb{R}_+$, continuity of $\tilde{v}(\alpha)$ on $\mathbb{R}_+ \backslash \{0\}$ follows from the well-known fact that a convex function is continuous over the relative interior of its domain, e.g., [14, Theorem 10.1]. Item 2 is the dual counterpart to item 1, and follows in a similar manner as outlined below. It was shown in [17, item 2 of Proposition 2] that for any fixed $t > 0$, $v(t\alpha, t)$ is a concave function in α. Furthermore, since the feasible region of $\mathbf{D}(t\alpha, t)$ gets larger as α increases (with respect to the y variable) while the objective function is unchanged, $v(t\alpha, t)$ is a monotone increasing function in α for any fixed $t > 0$. Thus, $v(t\alpha, t)$ is monotone increasing and concave in α for any fixed $t > 0$. Since $\bar{v}(\alpha) = \lim_{t \to 0} v(t\alpha, t)$, monotonicity and concavity of $\bar{v}$ is proved by taking the limit as $t \to 0$ in the similar manner as was done in the proof of monotonicity and convexity of $\tilde{v}$ in [17, Theorem 4]. ∎

Now we are ready to prove Theorem 3.1. Before proceeding to the proof, we comment on the linear independence of the $A^1, \ldots, A^m$. Theorem 3.1 itself does not require the assumption of linear independence of $A^1, \ldots, A^m$ to hold. Nevertheless, we assume linear independence of $A^1, \ldots, A^m$ in some parts of the proof. We remark that this is not an essential assumption and it avoids certain unnecessary complications. Indeed, even if $A^1, \ldots, A^m$ are not linearly independent, we can choose a subset $\{A^{i_1}, \ldots, A^{i_p}\}$, say, as a basis of the linear space spanned by $A^1, \ldots, A^m$. The pd-regularized optimal value function $v(\varepsilon, \eta)$ remains unchanged no matter whether $\{A^1, \ldots, A^m\}$ or $\{A^{i_1}, \ldots, A^{i_p}\}$ is used for representing the SDP under consideration.

3.1 Proof of item 1.

First we prove item 1 of Theorem 3.1. We start by recalling our main assumption for this proof.

We assume feasibility of $\mathbf{P}$ and $\mathbf{D}$, and assume that the singularity degree of $\mathbf{D}$ is one.

Furthermore, we assume that the problem data has been rescaled as in item 1 of Proposition 2.1. And, since the singularity degree is assumed to be one, we can further assume that we are in the setting of item 3 of Proposition 2.1. With that, by definition, $v(t\alpha, t)$ is the optimal value of the following problem,

$$\mathbf{D}(t\alpha, t) \quad \max \; b^T y - t(l(y) - c) \text{ s.t. } L(y) + t\alpha I \succeq 0, \tag{3.5}$$

where we define

$$l(y) := L(y) \bullet I, \; c := C \bullet I.$$

In what follows we will use the following convention, given an arbitrary $Z \in \mathcal{S}^n$, we will use Z_{11}, Z_{22} and Z_{12} to denote the blocks of Z according to the block division in Proposition 2.1, so that $Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22} \end{pmatrix}$ with $Z_{11} \in \mathcal{S}^r, Z_{22} \in \mathcal{S}^{n-r}, Z_{12} \in \mathbb{R}^{r \times (n-r)}$.

We start the proof with the following two preliminary lemmas.

Lemma 3.3 *The following statements hold:*

1. *$v(\mathbf{D}) \leq \bar{v}(\alpha) \leq v(\mathbf{P})$.*

2. *For $t > 0$, $v(t, 0)$ is finite and $\lim_{t \downarrow 0} v(t, 0) = \bar{v}(\infty) = v(\mathbf{P})$.*

3. *For $t > 0$, $v(0, t)$ is finite and $\lim_{t \downarrow 0} v(0, t) = \bar{v}(0) = v(\mathbf{D})$.*

Proof. Item 1 readily follows from Theorem 1.2, since $\bar{v}$ is the function v_a with a different parametrization, see (3.4). To prove item 2, we observe that $\mathbf{P}$ and $\mathbf{P}(t, 0)$ have the same feasible region. Since $\mathbf{P}$ is feasible, $\mathbf{P}(t, 0)$ is feasible and $v(t, 0) < +\infty$. Since $\mathbf{D}$ is feasible, $\mathbf{D}(t, 0)$ is also feasible. This means that $-\infty < v(t, 0)$. Thus, $v(t, 0)$ is finite. It follows from item 1 of Theorem 1.2 that $\lim_{t \downarrow 0} v(t, 0) = v(\mathbf{P})$. Item 3 is the dual counterpart of item 2 and follows analogously. ∎

Lemma 3.4 *There exists a constant $M > 0$ such that for every $t \geq 0$ the following implication holds:*

$$L_{22}(y) + tI_{22} \succeq 0 \quad \implies \quad tMI_{22} \succeq L_{22}(y) + tI_{22} \succeq 0. \tag{3.6}$$

Proof. Recall that we are under the setting of item 1 of Proposition 2.1 and $\mathbf{D}$ has singularity degree one. With X as in item 3 of Proposition 2.1, the rank of X_{22} is $n - r$. Since X_{22} is positive definite, there exists a constant $\kappa > 0$ such that

$$\|Y\| \leq \kappa X_{22} \bullet Y, \qquad \forall Y \in \mathcal{S}_+^{n-r}, \tag{3.7}$$

e.g., [5, Lemma 26].

Next, given some arbitrary $t > 0$, we define

$$\mathcal{C}_t := \{Y \in \mathcal{S}_+^{n-r} \mid Y = L_{22}(y) + tI_{22} \succeq 0, \ y \in \mathbb{R}^m\}.$$

By the assumptions on X we have $0 = X \bullet A^i = A_{22}^i \bullet X_{22}$ and $0 = C \bullet X = C_{22} \bullet X_{22}$. That is, $L_{22}(y) \bullet X_{22} = 0$ holds for every y. In view of (3.7), for $Y \in \mathcal{C}_t$, since Y belongs to $\mathcal{S}_+^{n-r}$ as well, we conclude that the following bound holds.

$$\|Y\| \leq \kappa X_{22} \bullet (L_{22}(y) + tI_{22}) = \kappa t(X_{22} \bullet I_{22}).$$

In particular, letting $M := \kappa(X_{22} \bullet I_{22})$, we conclude that the maximum eigenvalue of an arbitrary $Y \in \mathcal{C}_t$ satisfies $\lambda_{\max}(Y) \leq tM$ and M does not depend on t. In particular, $tMI_{22} \succeq L_{22}(y) + tI_{22}$ holds for all y such that $L_{22}(y) + tI_{22} \succeq 0$. ∎

In the following, we define

$$
\begin{aligned}
l_{12}^2(y) &:= L_{12}(y) \bullet L_{12}^T(y), \\
l_{11}(y) &:= L_{11}(y) \bullet I_{11}, \\
l_{22}(y) &:= L_{22}(y) \bullet I_{22}.
\end{aligned}
\tag{3.8}
$$

We will use M as a global constant satisfying (3.6) in Lemma 3.4. Let us consider the following problem.

$$\mathbf{RD1}(\alpha, t) \quad \max \ b^T y - \frac{l_{12}^2(y)}{M\alpha}, \quad \text{s.t.} \ \begin{pmatrix} L_{11}(y) + t\alpha I_{11} & L_{12}(y) \\ L_{12}^T(y) & L_{22}(y) + t\alpha I_{22} \end{pmatrix} \succeq 0.$$

Let $u_1(\alpha, t)$ be the optimal value function of $\mathbf{RD1}(\alpha, t)$. We show that $v(t\alpha, t)$ is majorized by the optimal value $u_1(\alpha, t)$ as follows.

Lemma 3.5 *For $t > 0$ and $\alpha > 0$, we have*

$$v(0, t) \leq v(t\alpha, t) \leq u_1(\alpha, t) + t^2\alpha n + tc. \tag{3.9}$$

Proof. We prove the first inequality. Recall that $v(0, t)$ is the optimal value of $\mathbf{D}(0, t)$ and $v(t\alpha, t)$ is the optimal value of $\mathbf{D}(t\alpha, t)$. Since the objective functions of the two problems are identical and the feasible region of $\mathbf{D}(t\alpha, t)$ contains the feasible region of $\mathbf{D}(0, t)$ (with respect to the y variable), we have $v(0, t) \leq v(t\alpha, t)$.

Now we prove the second inequality. Recalling the definitions in (3.8) and (3.5), $\mathbf{D}(t\alpha, t)$ can be written as follows.

$$\max \ b^T y + t(c - l_{11}(y) - l_{22}(y)), \ \text{s.t.} \ \begin{pmatrix} L_{11}(y) + t\alpha I_{11} & L_{12}(y) \\ L_{12}^T(y) & L_{22}(y) + t\alpha I_{22} \end{pmatrix} \succeq 0. \tag{3.10}$$

We note that $\mathbf{D}(t\alpha, t)$ and $\mathbf{RD1}(\alpha, t)$ share the same feasible region. Furthermore, $\mathbf{D}(t\alpha, t)$ satisfies Slater's condition so there exists a sequence of y^k corresponding to feasible solutions to $\mathbf{D}(t\alpha, t)$ such that the corresponding matrices are all positive definite and $(b^T y^k - t l_{11}(y^k) - t l_{22}(y^k) + tc)$ converges to the optimal value $\mathbf{D}(t\alpha, t)$. Thus, in order to establish

the second inequality in (3.9), it is enough to examine the y's that correspond to positive definite matrices.

So, let y be feasible solution to $\mathbf{D}(t\alpha, t)$ associated to a positive definite matrix. First, we find an upper bound on $-tl_{11}(y)$. Computing the Schur complement, we obtain

$$L_{11}(y) + t\alpha I_{11} - L_{12}(y)(L_{22}(y) + t\alpha I_{22})^{-1}L_{12}^T(y) \succeq 0. \tag{3.11}$$

Since $L_{22}(y) + t\alpha I_{22} \succeq 0$, we obtain $Mt\alpha I_{22} \succeq (L_{22}(y) + t\alpha I_{22})$ from (3.6). So each eigenvalue of $(L_{22}(y) + t\alpha I_{22})$ is less than or equal to $Mt\alpha$. This implies that $(L_{22}(y) + t\alpha I_{22})^{-1} \succeq \frac{I_{22}}{Mt\alpha}$. Therefore, we have

$$L_{12}(y)\left((L_{22} + t\alpha I_{22})^{-1} - \frac{I_{22}}{Mt\alpha}\right)L_{12}(y)^T \succeq 0,$$

which implies that

$$L_{12}(y)(L_{22} + t\alpha I_{22})^{-1}L_{12}(y)^T \succeq L_{12}(y)\frac{I_{22}}{Mt\alpha}L_{12}(y)^T.$$

Then, in view of (3.11), we obtain

$$L_{11}(y) + t\alpha I_{11} - \frac{1}{tM\alpha}L_{12}(y)L_{12}(y)^T \succeq 0.$$

Taking the inner-product with I_{11} and multiplying by t, we obtain

$$tl_{11}(y) + t^2\alpha r - \frac{1}{M\alpha}l_{12}^2(y) \geq 0. \tag{3.12}$$

Since $L_{22}(y) + t\alpha I_{22} \succeq 0$, we have $-tl_{22}(y) \leq t^2\alpha(n - r)$. From this inequality and (3.12), we see that the objective function of (3.10) is majorized by the objective function of **RD1** plus a constant $t^2\alpha n + tc$ as follows.

$$b^T y + t(c - l_{11}(y) - l_{22}(y)) \leq b^T y - \frac{l_{12}^2(y)}{M\alpha} + t^2\alpha n + tc.$$

Since the optimal value of (3.10) (or equivalently $\mathbf{D}(t\alpha, t)$) is $v(t\alpha, t)$ and the optimal value of $\mathbf{RD1}(\alpha, t)$ is $u_1(\alpha, t)$, the second inequality of (3.9) follows immediately from this inequality. ∎

Lemma 3.6 *Let $\alpha > 0$. Then, there exists a constant $\hat{t}_\alpha > 0$ depending on α such that, if $t \in (0, \hat{t}_\alpha)$ and $\{y^k\}$ is an optimal sequence of $\mathbf{RD1}(\alpha, t)$[5] then the following bound holds for sufficiently large k:*

$$l_{12}^2(y^k) \leq M(v(\mathbf{P}) - v(\mathbf{D}) + 2)\alpha = K\alpha, \tag{3.13}$$

where $K := M(v(\mathbf{P}) - v(\mathbf{D}) + 2)$ and M is a constant as in Lemma 3.4.

---

[5] $\mathbf{RD1}(\alpha, t)$ is not a linear SDP, but we can define optimal sequences analogously as in Section 2.

Proof. Recall that $\lim_{t\downarrow 0} v(t\alpha, 0) = v(\mathbf{P})$ and let $\hat{t}_\alpha > 0$ be small enough such that

$$v(t\alpha, 0) \leq v(\mathbf{P}) + \frac{1}{2} \tag{3.14}$$

holds for $0 < t < \hat{t}_\alpha$.

Let $t \in (0, \hat{t}_\alpha)$ and let $\{y^k\}$ be an optimal sequence to $\mathbf{RD1}(\alpha, t)$. Each y^k is also a feasible solution to $\mathbf{D}(t\alpha, 0)$, so the following inequality holds for every k

$$b^T y^k \leq v(t\alpha, 0). \tag{3.15}$$

Next, let $\{y_D^k\}$ be an optimal sequence to $\mathbf{D}$. Then $\lim_{k\to\infty} b^T y_D^k = v(\mathbf{D})$. Recalling that we are under the setting of Proposition 2.1, we have that $L_{12}(y_D^k) = 0$ holds which implies that $l_{12}^2(y_D^k) = 0$ holds as well, see (3.8). Furthermore, each y_D^k is feasible to $\mathbf{RD1}(\alpha, t)$ as well. In view for these facts, for sufficiently large k, we have $v(\mathbf{D}) - 1 \leq b^T y_D^k \leq v(\mathbf{D})$ which leads to

$$
\begin{aligned}
v(\mathbf{D}) - 1 &\leq b^T y_D^k \\
&= b^T y_D^k - \frac{1}{M\alpha} l_{12}^2(y_D^k) \\
&\leq v(\mathbf{RD1}(\alpha, t)) \\
&\leq b^T y^k - \frac{1}{M\alpha} l_{12}^2(y^k) + \frac{1}{2} \\
&\leq v(t\alpha, 0) - \frac{1}{M\alpha} l_{12}^2(y^k) + \frac{1}{2} \\
&\leq v(\mathbf{P}) + 1 - \frac{1}{M\alpha} l_{12}^2(y^k),
\end{aligned}
$$

where the first equality holds because $l_{12}^2(y_D^k) = 0$, the second inequality holds because y_D^k is feasible for $\mathbf{RD1}(\alpha, t)$, and the third inequality holds (for sufficiently large k) because $\{y^k\}$ is an optimal sequence. Moreover, the last two inequalities follow from (3.15) and (3.14), respectively. Overall, we have

$$\frac{1}{M\alpha} l_{12}^2(y^k) \leq v(\mathbf{P}) - v(\mathbf{D}) + 2,$$

for sufficiently large k. ∎

Now we consider

$$\mathbf{RD2}(\alpha, t) \quad \max \ b^T y, \quad \text{s.t.} \ \begin{pmatrix} L_{11}(y) + t\alpha I_{11} & L_{12}(y) \\ L_{12}^T(y) & L_{22}(y) + t\alpha I_{22} \end{pmatrix} \succeq 0, \ \ l_{12}^2(y) \leq K\alpha,$$

where K is as in Lemma 3.6. Let us denote by $u_2(\alpha, t)$ the optimal value of $\mathbf{RD2}(\alpha, t)$.

Lemma 3.7 *For $\alpha > 0$, the function defined by $\bar{u}(\alpha) := \lim_{t\downarrow 0} u_2(\alpha, t)$ is well-defined and finite.*

Proof. We note that $v(t\alpha, 0)$ is the optimal value of $\mathbf{D}(t\alpha, 0)$ and $\mathbf{D}(t\alpha, 0)$ is precisely the problem obtained by removing the constraint "$l_{12}^2(y) \leq K\alpha$" from $\mathbf{RD2}$. Therefore, we have

$$u_2(\alpha, t) \leq v(t\alpha, 0).$$

By item 2 of Lemma 3.3, for sufficiently small $t > 0$, $v(t\alpha, 0)$ is bounded from above by a finite value, which implies that the same is true for $u_2(\alpha, t)$. For fixed α, if we increase t, the feasible region of $\mathbf{RD2}(\alpha, t)$ enlarges (more precisely, it does not shrink). Therefore, $u_2(\alpha, t) \leq u_2(\alpha, \hat{t})$ holds if $0 < t \leq \hat{t}$, namely, $u_2(\alpha, t)$ is a monotone increasing function with respect to t. In addition, $u_2(\alpha, t)$ is bounded below by $v(\mathbf{D})$, since any feasible solution to $\mathbf{D}$ is feasible to $\mathbf{RD2}(\alpha, t)$ (again, we recall that we assume that Proposition 2.1 holds, so $L_{12}(y) = 0$ and $L_{22}(y) = 0$ if y is feasible to $\mathbf{D}$). Therefore, $\lim_{t\to 0} u_2(\alpha, t)$ exists and is finite, thus showing that $\bar{u}(\alpha)$ is a well-defined function assuming finite values. ∎

Now we are ready to finish the proof of item 1 of Theorem 3.1. For $\alpha > 0$ and $\hat{t}_\alpha > 0$ as in Lemma 3.6, an optimal sequence $\{y^k\}$ to $\mathbf{RD1}(\alpha, t)$ satisfies (3.13) for sufficiently large k. Furthermore, for each y^k, the objective value associated to $\mathbf{RD2}(\alpha, t)$ is greater or equal than the corresponding objective value associated to the problem $\mathbf{RD1}(\alpha, t)$. In view of these facts and of Lemma 3.6, we see that $u_1(\alpha, t) \leq u_2(\alpha, t)$. Combined with Lemma 3.5, we obtain

$$v(0, t) \leq v(t\alpha, t) \leq u_1(\alpha, t) + t^2 \alpha n + tc \leq u_2(\alpha, t) + t^2 \alpha n + tc. \qquad (3.16)$$

Recalling the definitions of $\bar{v}$ and $\bar{u}$ in (3.2) and Lemma 3.7, respectively, and invoking item 1 of Theorem 1.2, we let $t \to 0$ in (3.16) in order to obtain

$$v(\mathbf{D}) \leq \bar{v}(\alpha) \leq \bar{u}(\alpha).$$

Therefore, if we can show that

$$\lim_{\alpha\downarrow 0} \bar{u}(\alpha) = \lim_{\alpha\downarrow 0}\lim_{t\downarrow 0} u_2(\alpha, t) = v(\mathbf{D}), \qquad (3.17)$$

then $\lim_{\alpha\downarrow 0} \bar{v}(\alpha) = v(\mathbf{D})$ will hold as well, which, in view of (3.4) and Theorem 1.2 implies the continuity of v_a at $\theta = \pi/2$, which is what we wanted to show. We will prove (3.17) by contradiction.

Proof of (3.17)**:** First, we recall that, in view of item 1 of Proposition 2.1, every y that is feasible to $\mathbf{D}$ is feasible to $\mathbf{RD2}(\alpha, t)$, for every $\alpha > 0$, $t > 0$. Therefore

$$v(\mathbf{D}) \leq u_2(\alpha, t), \qquad \forall \alpha > 0, \forall t > 0,$$

which implies that

$$v(\mathbf{D}) \leq \bar{u}(\alpha), \qquad \forall \alpha > 0.$$

If (3.17) does not hold, there exists $\delta > 0$ and a positive sequence $\{\alpha^k\}$ such that $\lim_{k\to\infty} \alpha^k = 0$ for which

$$0 < 2\delta \leq |\bar{u}(\alpha^k) - v(\mathbf{D})| = \bar{u}(\alpha^k) - v(\mathbf{D})$$

14

holds for sufficiently large k. Next, since $\bar{u}(\alpha^k) = \lim_{t \downarrow 0} u_2(\alpha^k, t)$, for any given k, we can pick $t^k \in (0, 1/k]$ such that $|\bar{u}(\alpha^k) - u_2(\alpha^k, t^k)| \leq 1/k$.

Therefore, there exists a sequence $(\alpha^k, t^k) \to (0, 0)$ such that

$$u_2(\alpha^k, t^k) > v(\mathbf{D}) + \delta$$

holds for sufficiently large k. Then, for each k sufficiently large, there exists a feasible solution y^k to $\mathbf{RD2}(\alpha^k, t^k)$ satisfying

$$b^T y^k \geq v(\mathbf{D}) + \delta.$$

Next, we define $S_{11}^k := t^k \alpha^k I_{11}$, $S_{22}^k := L_{22}(y^k) + t^k \alpha^k I_{22}$, $S_{12}^k := L_{12}(y^k)$ and $S^k := \begin{pmatrix} S_{11}^k & S_{12}^k \\ (S_{12}^k)^T & S_{22}^k \end{pmatrix}$. Note that $S^k \in \mathcal{T} \oplus \mathcal{L}$ holds, where $\mathcal{T} \oplus \mathcal{L}$ was introduced as perturbation space in item 2 of Proposition 2.1. By definition, y^k is a feasible solution to the following problem.

$$\mathbf{RD3} \qquad \max_y \ b^T y, \quad \text{s.t.} \ \begin{pmatrix} L_{11}(y) + S_{11}^k & L_{12}(y) \\ L_{12}^T(y) & L_{22}(y) + t^k \alpha^k I_{22} \end{pmatrix} \succeq 0,$$
$$L_{22}(y) + t^k \alpha^k I_{22} = S_{22}^k, \quad L_{12}(y) = S_{12}^k, \quad l_{12}^2(y) \leq K \alpha^k.$$

Because y^k is also feasible to $\mathbf{RD2}(\alpha^k, t^k)$, we have $S_{22}^k = L_{22}(y^k) + t^k \alpha^k I_{22} \succeq 0$ and by Lemma 3.4, S_{22}^k goes to 0 as t^k and α^k goes to zero. Similarly, since $L_{12}(y^k) \bullet L_{12}(y^k) = l_{12}^2(y^k) \leq K \alpha^k$ holds, S_{12}^k also goes to zero as α^k goes to 0. By definition, S_{11}^k goes to zero as t^k and α^k goes to 0. So, overall S^k goes to 0 as $k \to \infty$.

Finally, recalling item 2 of Proposition 2.1, we note that y^k is also a feasible solution to $\mathbf{RD}(S^k)$ satisfying $b^T y^k \geq v(\mathbf{D}) + \delta$. Therefore, $w(S^k) \geq v(\mathbf{D}) + \delta$ and $S^k \to 0$ holds as k goes to infinity. But this contradicts the continuity of $w(S)$ at $S = 0$, that is, the conclusion of item 2 of Proposition 2.1. ∎

Thus, (3.17) must hold, and the proof of item 1 of Theorem 3.1 is complete.

3.2 Proof of item 2

Now, we proceed to prove item 2 of Theorem 3.1. Item 2 is the dual counterpart of item 1, and therefore one may argue that item 2 follows automatically from item 1 by primal-dual symmetry. But to be more precise, we proceed as follows. Our purpose is to show that $v_a(\theta)$ is continuous at $\theta = 0$, which is equivalent to continuity of $\tilde{v}(\beta)$ at $\beta = 0$. To this end, we rewrite $\mathbf{P}$ in the dual format and apply item 1.

Without loss of generality, we may assume the matrices $A_1, \ldots, A_m$ are linearly independent. Let $\bar{n} := n \times (n+1)/2 - m$. We rewrite $\mathbf{P}(\varepsilon, \eta)$ in the dual format by taking a basis $A_\perp^1, \ldots, A_\perp^{\bar{n}}$ of the space $\mathcal{X} := \{X \in \mathcal{S}^n \mid A^i \bullet X = 0, \ i = 1, \ldots, m\}$. Let X^* be a $n \times n$ matrix satisfying $A^i \bullet X_i^* = b_i$ for $i = 1, \ldots, m$. Representing $\mathcal{X}$ as $\{\widetilde{X} \in \mathcal{S}^n \mid \widetilde{X} = -\sum_{i=1}^{\bar{n}} A_\perp^i \tilde{y}_i\}$ and reversing the sign of the objective function to flip "max" and "min", we obtain the following representation of $\mathbf{P}(\varepsilon, \eta)$.

$$\max_{\tilde{y},\widetilde{X}} \sum_{i=1}^{\bar{n}} (\tilde{b}_i + \varepsilon A_\perp^i \bullet I)\tilde{y}_i - (C + \varepsilon I) \bullet (X^* + \eta I)$$

$$\text{s.t. } (X^* + \eta I) - \sum_{i=1}^{\bar{n}} A_\perp^i \tilde{y}_i = \widetilde{X}, \quad \widetilde{X} \succeq 0,$$

where $\tilde{b}_i = C^i \bullet A_\perp^i$ for $i = 1,\ldots,\bar{n}$. The optimal value of this problem is $-v(\varepsilon,\eta)$ if $(\varepsilon,\eta) \neq 0$, and $-v(\mathbf{P})$ if $(\varepsilon,\eta) = 0$.

Dropping the constant term $-(C + \varepsilon I) \bullet (X^* + \eta I)$, we obtain the following problem.

$$\mathbf{D}_{\mathrm{P}}(\eta,\varepsilon): \quad \max_{\tilde{y},\widetilde{X}} \sum_{i=1}^{\bar{n}} (\tilde{b}_i + \varepsilon A_\perp^i \bullet I)\tilde{y}_i$$

$$\text{s.t. } (X^* + \eta I) - \sum_{i=1}^{\bar{n}} A_\perp^i \tilde{y}_i = \widetilde{X}, \quad \widetilde{X} \succeq 0.$$

The dual problem of $\mathbf{D}_{\mathrm{P}}(\eta,\varepsilon)$ is

$$\mathbf{P}_{\mathrm{D}}(\eta,\varepsilon): \quad \min_{\widetilde{S}} (X^* + \varepsilon I) \bullet \widetilde{S}$$
$$\text{s.t. } A_\perp^i \bullet \widetilde{S} = \tilde{b}_i + \eta A_\perp^i \bullet I, \ i = 1,\ldots,\bar{n}, \ \widetilde{S} \succeq 0,$$

Note that $\mathbf{P}_{\mathrm{D}}(\eta,\varepsilon)$ coincides with $\mathbf{D}(\varepsilon,\eta)$ represented in standard form using $A_\perp^i$ ($i = 1,\ldots,\bar{n}$) and X^*, except that min and max are flipped by multiplying the objective function by -1 and the constant term $-(C + \varepsilon I) \bullet (X^* + \eta I)$ is dropped.

Let us denote by $v_1(\eta,\varepsilon)$ the common optimal value of $\mathbf{P}_{\mathrm{D}}(\eta,\varepsilon)$ and $\mathbf{D}_{\mathrm{P}}(\eta,\varepsilon)$ (if it exists). Since X part of $\mathbf{P}(\varepsilon,\eta)$ and $\widetilde{X}$ part of $\mathbf{D}_{\mathrm{P}}(\eta,\varepsilon)$ coincide, and since S part of $\mathbf{D}(\varepsilon,\eta)$ and $\widetilde{S}$ part of $\mathbf{P}_{\mathrm{D}}(\varepsilon,\eta)$ coincide, $v_1(\eta,\varepsilon)$ is well-defined if and only if $v(\varepsilon,\eta)$ is well-defined, and the two functions are connected by the relation

$$v_1(\eta,\varepsilon) = -v(\varepsilon,\eta) + (C + \varepsilon I) \bullet (X^* + \eta I) \tag{3.18}$$

for $(\varepsilon,\eta) \neq (0,0)$. In the case of $(\varepsilon,\eta) = (0,0)$, we have

$$v(\mathbf{D}_{\mathrm{P}}(0,0)) = -v(\mathbf{P}) + C \bullet X^*, \quad v(\mathbf{P}_{\mathrm{D}}(0,0)) = -v(\mathbf{D}) + C \bullet X^*. \tag{3.19}$$

Now, consider the primal-dual pair $\mathbf{D}_{\mathrm{P}}(0,0)$ and $\mathbf{P}_{\mathrm{D}}(0,0)$. It follows from (3.19) that $\mathbf{P}_{\mathrm{D}}(0,0)$ and $\mathbf{D}_{\mathrm{P}}(0,0)$ have a finite nonzero duality gap that is equal to the duality gap between $\mathbf{P}$ and $\mathbf{D}$. Furthermore, the singularity degree of $\mathbf{D}_{\mathrm{P}}(0,0)$ and $\mathbf{P}_{\mathrm{D}}(0,0)$ is one, because applying facial reduction to $\mathbf{D}_{\mathrm{P}}(0,0)$ and $\mathbf{P}_{\mathrm{D}}(0,0)$ corresponds to applying facial reduction to $\mathbf{P}$ and $\mathbf{D}$, respectively, which finishes in one step. Therefore, we can apply the result of item 1 to $\mathbf{D}_{\mathrm{P}}(0,0)$.

We define $\bar{v}_1(\beta) = \lim_{t \downarrow 0} v_1(t\beta, t)$. Then it follows from item 1 that $\bar{v}_1(\beta)$ is continuous at $\beta = 0$. Due to (3.18) and the definition of $\bar{v}_1$, we have $\bar{v}_1(\beta) = -\tilde{v}(\beta) + C \bullet X^*$. Consequently, $\tilde{v}(\beta)$ is continuous at $\beta = 0$, as we desired. ∎

3.3 Proof of item 3

We move on to a proof of item 3 of Theorem 3.1. The first half of item 3 follows immediately from items 1 and 2. In the following, we deal with the second half of item 3. Since

monotonicity and continuity of v_a on $[0, \pi/2]$ was already established, we focus on the bijectivity of v_a. To this end, we make use of Proposition 3.2, which ensures monotonicity, convexity/concavity, and continuity of $\tilde{v}$ and $\bar{v}$.

In view of (3.4), it is enough to argue that $\bar{v}$ is bijection from $[0, \infty) \cup \{+\infty\}$ to $[v(\mathbf{D}), v(\mathbf{P})]$ with the convention that $\bar{v}(+\infty) = \lim_{\alpha \to +\infty} \bar{v}(\alpha) = v(\mathbf{P})$ as in (3.2).

By contradiction, assume that $\bar{v}(\alpha)$ is (monotone increasing but) not strictly monotone increasing. Then, there exists an interval $J \subseteq \mathbb{R}_+$ where $\bar{v}$ is constant. Slightly abusing the notation, let us denote this constant by $\bar{v}(J)$. There are three cases to be considered:

(i) $\bar{v}(J) = \bar{v}(0) = v(\mathbf{D})$;

(ii) $v(\mathbf{D}) = \bar{v}(0) < \bar{v}(J) < \bar{v}(\infty) = v(\mathbf{P})$;

(iii) $\bar{v}(J) = \bar{v}(\infty) = v(\mathbf{P})$.

In case (i), we recall that $\bar{v}$ is monotone increasing, so $v(\mathbf{D})$ is the minimum value of $\bar{v}$ throughout $\mathbb{R}_+$. Since this minimum value is attained in an interval and $\bar{v}$ is concave, $\bar{v}$ must be constant throughout $\mathbb{R}_+$, see [14, Theorem 32.1] for the analogous fact about convex functions. That is, $\bar{v}(\alpha) = v(\mathbf{D}) < v(\mathbf{P})$ holds for all $\alpha > 0$. This contradicts the fact that $\lim_{\alpha \to \infty} \bar{v}(\alpha) = v(\mathbf{P})$. So case (i) cannot occur.

In case (ii), let x be a point in the relative interior of J, then x would be a local maximum of $\bar{v}$ that is not global. This contradicts the concavity of $\bar{v}$ and cannot occur.

Finally, in case (iii), since $\bar{v}$ is monotone increasing and $v(\mathbf{P})$ is the maximum value, J can be taken to be a closed half line of the form $[\alpha^*, +\infty)$, with $\alpha^* > 0$. Now we turn our attentions to $\tilde{v}$ and recall that $\bar{v}$ and $\tilde{v}$ are related as in (3.3). We have $\bar{v}(\alpha) = \tilde{v}(1/\alpha) = v(\mathbf{P})$ for all $\alpha \geq \alpha^*$. Recall that $\tilde{v}(0) = v(\mathbf{P})$, and thus $\tilde{v}(0) = \tilde{v}(1/\alpha) = v(\mathbf{P})$, for all $\alpha \geq \alpha^*$. Since $v(\mathbf{P})$ is the maximum value of the convex function $\tilde{v}$, $\tilde{v}$ must be constant throughout its domain, again by [14, Theorem 32.1]. Therefore, $\tilde{v}(\beta) = v(\mathbf{P})$ holds for all $\beta \in \mathbb{R}_+$. Therefore,

$$\lim_{\alpha \downarrow 0} \bar{v}(\alpha) = \lim_{\alpha \downarrow 0} \tilde{v}(1/\alpha) = v(\mathbf{P}) > v(\mathbf{D}) = \bar{v}(0),$$

which contradicts the continuity of $\bar{v}$ at $\alpha = 0$.

4 A Counter-Example

In the previous section, we established that the limiting pd-regularized optimal value function v_a is a monotonically decreasing continuous bijective function from $[0, \pi/2]$ to $[v(\mathbf{D}), v(\mathbf{P})]$ if $\mathbf{P}$ and $\mathbf{D}$ admit a finite nonzero duality gap and the singularity degree of $\mathbf{P}$ and $\mathbf{D}$ is one. In this section, we present a counter-example where $v_a([0, \pi/2]) \neq [v(\mathbf{D}), v(\mathbf{P})]$. The instance is obtained by modifying Example 1.1 and the singularity degree of this instance is two. We will show that continuity of $v_a(\theta)$ is violated at $\theta = \pi/2$.

Let us consider the following semidefinite program in dual format:

$$\max_{y \in \mathbb{R}^3} y_1 \text{ s.t.} \begin{pmatrix} 1 - y_1 & 0 & -y_3 & -y_3 \\ 0 & -y_2 & -y_1 & 0 \\ -y_3 & -y_1 & -y_3 & 0 \\ -y_3 & 0 & 0 & 0 \end{pmatrix} \succeq 0. \tag{4.1}$$

The $(4, 4)$-element of the matrix in (4.1) is zero, this forces y_3 to be 0 in order for $y \in \mathbb{R}^3$ to be feasible. This, however, implies that the $(3, 3)$-element is 0, which leads to $y_1 = 0$. Therefore, the optimal value of (4.1) is 0.

The corresponding primal $\mathbf{P}$ is

$$\min_{X \in \mathcal{S}^4} X_{11} \text{ s.t. } X_{11} + 2X_{23} = 1, \ X_{22} = 0, \ X_{33} + 2(X_{13} + X_{14}) = 0, \ X \succeq 0. \tag{4.2}$$

If X is feasible for (4.2), then we must have $X_{22} = 0$ and $X_{23} = 0$, which implies that $X_{11} = 1$. Therefore the optimal value of (4.2) is 1.

In conclusion, there is a duality gap between the primal dual pair of problems (4.1), (4.2). However, in contrast to Example 1.1, the singularity degree of (4.1) is two[6], so we are outside of the scope of Theorem 3.1.

Now we consider the problem $\mathbf{D}(\varepsilon, \eta)$ as in (1.2):

$$\max_{y \in \mathbb{R}^3} (1 + \eta)y_1 + \eta y_2 + \eta y_3 \text{ s.t.} \begin{pmatrix} 1 + \varepsilon - y_1 & 0 & -y_3 & -y_3 \\ 0 & \varepsilon - y_2 & -y_1 & 0 \\ -y_3 & -y_1 & -y_3 + \varepsilon & 0 \\ -y_3 & 0 & 0 & \varepsilon \end{pmatrix} \succeq 0.$$

Recalling the definition v_a in (3.4), we prove the following theorem, which states that the range of v_a is $\{0, 1\}$. Therefore, in this case, duality gap fails to be filled completely in the sense of Theorem 3.1.

Theorem 4.1 *The following statements hold for the problem (4.1):*

1. *$v_a(\theta) = v(\mathbf{P}) = 1$ for all $\theta \in [0, \pi/2)$,*

2. *$v_a(\pi/2) = v(\mathbf{D}) = 0$.*

Proof. In view of (3.4), it is enough to analyze $\bar{v}$ (defined in (3.2)). We will show that $\bar{v}(0) = 0$ and $\bar{v}(\alpha) = 1$ for all $\alpha > 0$. Since $v(\mathbf{D}) = \bar{v}(0) = 0$ and $v(\mathbf{P}) = \bar{v}(\infty) = 1$, this will imply that $\bar{v}$ is discontinuous at 0 and is continuous at $+\infty$. We analyze the problem $\mathbf{D}(t\alpha, t)$:

$$\max_{y \in \mathbb{R}^3} (1 + t)y_1 + ty_2 + ty_3 \text{ s.t.} \begin{pmatrix} 1 + t\alpha - y_1 & 0 & -y_3 & -y_3 \\ 0 & t\alpha - y_2 & -y_1 & 0 \\ -y_3 & -y_1 & -y_3 + t\alpha & 0 \\ -y_3 & 0 & 0 & t\alpha \end{pmatrix} \succeq 0. \tag{4.3}$$

---

[6]To see that, we observe that the first reducing direction s_1 in (2.1) must be a 4×4 matrix whose only nonzero and positive entry is its (4,4)-element. One can then confirm that the SDP obtained by performing one step of facial reduction with s_1 is essentially Example 1.1 which requires one more step of facial reduction to recover strong feasibility.

We already know that $0 \leq \bar{v}(\alpha) \leq 1$ for $0 \leq \alpha \leq \infty$. Let $\alpha > 0$ be fixed. In view of the definition of $\bar{v}$ in (3.2), in order to prove that $\bar{v}(\alpha) = 1$ holds, it is enough to show that for any $\delta > 0$, if $t > 0$ is sufficiently small, the problem (4.3) admits a feasible solution $y^t \in \mathbb{R}^3$ with objective value at least $1 - \delta$, i.e.,

$$(1+t)y_1^t + ty_2^t + ty_3^t \geq 1 - \delta. \tag{4.4}$$

This would imply $v(t\alpha, t) \geq 1 - \delta$ for sufficiently small t and thus $\bar{v}(\alpha) \geq 1 - \delta$ holds. Then, since δ is arbitrary, we obtain $\bar{v}(\alpha) = 1$.

In view of this, we focus our efforts on establishing the existence of feasible solutions as in (4.4). We will proceed by analyzing what conditions must a feasible solution to (4.3) satisfy. We note that since (4.3) is strongly feasible, there are feasible solutions corresponding to positive definite matrices such that the objective value is arbitrarily close to the optimal value. Therefore, it is enough to focus on analyzing the y's that are feasible to (4.3) and correspond to positive definite matrices.

So suppose that $y \in \mathbb{R}^3$ corresponds to a positive definite solution to (4.3). First, we take the Shur complement with respect to the $(4,4)$-element. Then, positive definiteness is equivalent to

$$\begin{pmatrix} 1 + t\alpha - y_1 - \frac{y_3^2}{t\alpha} & 0 & -y_3 \\ 0 & t\alpha - y_2 & -y_1 \\ -y_3 & -y_1 & -y_3 + t\alpha \end{pmatrix} \succ 0. \tag{4.5}$$

Next, we take the Shur complement of the left hand side of (4.5) with respect to the $(1,1)$-element. Then, positive definiteness is equivalent to the following conditions

$$1 + t\alpha - y_1 - \frac{y_3^2}{t\alpha} > 0, \quad \begin{pmatrix} t\alpha - y_2 & -y_1 \\ -y_1 & -y_3 + t\alpha - \frac{y_3^2}{1+t\alpha-y_1-\frac{y_3^2}{t\alpha}} \end{pmatrix} \succ 0.$$

This is equivalent to

$$\begin{aligned}
&1 + t\alpha - y_1 - \frac{y_3^2}{t\alpha} > 0, \\
&t\alpha - y_2 > 0, \quad (t\alpha - y_2)\left(-y_3 + t\alpha - \frac{y_3^2}{1+t\alpha-y_1-\frac{y_3^2}{t\alpha}}\right) - y_1^2 > 0.
\end{aligned} \tag{4.6}$$

Diving the third inequality by $\left(-y_3 + t\alpha - \frac{y_3^2}{1+t\alpha-y_1-\frac{y_3^2}{t\alpha}}\right)$ which is ensured to be positive in view of (4.6), we obtain that

$$\begin{aligned}
&t\alpha - y_2 > 0, \quad 1 + t\alpha - y_1 - \frac{y_3^2}{t\alpha} > 0, \\
&t\alpha - \frac{y_1^2}{-y_3+t\alpha-\frac{y_3^2}{1+t\alpha-y_1-\frac{y_3^2}{t\alpha}}} > y_2, \quad -y_3 + t\alpha - \frac{y_3^2}{1+t\alpha-y_1-\frac{y_3^2}{t\alpha}} > 0.
\end{aligned} \tag{4.7}$$

is equivalent to (4.6). Therefore, if $y \in \mathbb{R}^3$ satisfies (4.7), then y is a feasible solution to (4.3) corresponding to a positive definite matrix.

Let $y_3 := -\gamma\sqrt{\alpha t}$, and let $y_1 := 1 - \gamma^2 - \zeta$, where γ, ζ are arbitrary positive numbers. Then, we have

$$1 + t\alpha - y_1 - \frac{y_3^2}{t\alpha} = 1 + t\alpha - y_1 - \gamma^2 = t\alpha + \zeta > 0$$

and

$$-y_3 + t\alpha - \frac{y_3^2}{1 + t\alpha - y_1 - \frac{y_3^2}{t\alpha}} = \gamma\sqrt{\alpha t} + \alpha t - \frac{\gamma^2 \alpha t}{\alpha t + \zeta}.$$

Furthermore, let

$$y_2 := t\alpha - \frac{y_1^2}{-y_3 + t\alpha - \frac{y_3^2}{1 + t\alpha - y_1 - \frac{y_3^2}{t\alpha}}} - \xi = t\alpha - \frac{(1 - \gamma^2 - \zeta)^2}{\gamma\sqrt{\alpha t} + \alpha t - \frac{\gamma^2 \alpha t}{\alpha t + \zeta}} - \xi,$$

where $\xi > 0$ is an arbitrary positive number. Then $y := (y_1, y_2, y_3)$ defined as above constitutes a strict feasible solution to (4.3) when $t > 0$ is sufficiently small, since they satisfy (4.7) when $t > 0$ is sufficiently small. Furthermore, the objective value

$$\begin{aligned}(1 + t\alpha)y_1 + ty_2 + ty_3 &= (1 + t\alpha)(1 - \gamma^2 - \zeta) + t^2\alpha - \frac{t(1 - \gamma^2 - \zeta)^2}{\gamma\sqrt{\alpha t} + \alpha t - \frac{\gamma^2 \alpha t}{\alpha t + \zeta}} \\ &\quad -t\xi - \gamma t \sqrt{\alpha t},\end{aligned}$$

which can be arbitrarily close to $1 - \gamma^2 - \zeta$, if $t > 0$ is taken sufficiently close to zero. By taking $\gamma^2 + \zeta = \delta/2$, we obtain a strict feasible solution to (4.3) satisfying (4.4) for t sufficiently close to 0, as we desired. ∎

5 Concluding Discussion

In this paper, we analyzed the behavior of $v_a(\theta)$ which was introduced in [17] to bridge the primal and dual optimal value in the presence of a nonzero duality gap. We assumed that $\mathbf{P}$ and $\mathbf{D}$ are feasible and they have nonzero duality gap, and showed that, surprisingly, v_a is a monotone bijective function from $[0, \pi/2]$ to $[v(\mathbf{D}), v(\mathbf{P})]$ if the singularity degrees of $\mathbf{P}$ and $\mathbf{D}$ are both one, thus filling the duality gap completely. However, we also produced an example showing that when the singularity degree is higher, then v_a can be *discontinuous* at $\theta = 0$. The study of deeper relations between the discontinuity of v_a and the singularity degree of $\mathbf{P}$ and $\mathbf{D}$ is an interesting topic for further research. Another interesting direction of further research is to extend the results to the case where either $\mathbf{P}$ or $\mathbf{D}$ is weakly infeasible (or both!). We note that there are several papers focused on weakly infeasible problems and their underlying structure [6, 10, 18]. These works may provide clues on how to extend our results to the weakly infeasible case.

References

[1] G. P. Barker and D. Carlson. Cones of diagonally dominant matrices. *Pacific Journal of Mathematics*, 57(1):15–32, 1975.

[2] J. M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *Journal of the Australian Mathematical Society (Series A)*, 30(03):369–380, 1981.

[3] D. Drusvyatskiy and H. Wolkowicz. The many faces of degeneracy in conic optimization. Technical report, University of Washington, 2017.

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

[5] B. F. Lourenço. Amenable cones: error bounds without constraint qualifications. *Mathematical Programming*, 186:1–48, 2021.

[6] B. F. Lourenço, M. Muramatsu, and T. Tsuchiya. A structural geometrical analysis of weakly infeasible SDPs. *Journal of the Operations Research Society of Japan*, 59(3):241–257, 2016.

[7] B. F. Lourenço, M. Muramatsu, and T. Tsuchiya. Facial reduction and partial polyhedrality. *SIAM Journal on Optimization*, 28(3):2304–2326, 2018.

[8] G. Pataki. The geometry of semidefinite programming. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of semidefinite programming: theory, algorithms, and applications*. Kluwer Academic Publishers, online version at https://gaborpataki.web.unc.edu/wp-content/uploads/sites/14119/2018/07/chapter.pdf, 2000.

[9] G. Pataki. Strong duality in conic linear programming: Facial reduction and extended duals. In *Computational and Analytical Mathematics*, volume 50, pages 613–634. Springer New York, 2013.

[10] G. Pataki and A. Touzov. An echelon form of weakly infeasible semidefinite programs and bad projections of the psd cone. *Foundations of Computational Mathematics*, Nov 2022. doi:10.1007/s10208-022-09552-0.

[11] F. A. Potra and R. Sheng. A superlinearly convergent primal–dual infeasible–interior–point algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8:1007–1028, 1998.

[12] M. V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Mathematical Programming*, 77, 1995.

[13] R. T. Rockafellar. *Conjugate Duality and Optimization*. No. 16 in Conference Board of Math. Sciences Series. SIAM Publications, 1974.

[14] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1997.

[15] S. Sremac, H. Woerdeman, and H. Wolkowicz. Complete facial reduction in one step for spectrahedra. *arXiv e-prints:1710.07410*, 2017. arXiv:1710.07410.

[16] J. F. Sturm. Error bounds for linear matrix inequalities. *SIAM Journal on Optimization*, 10(4):1228–1248, Jan. 2000.

[17] T. Tsuchiya, B. F. Lourenço, M. Muramatsu, and T. Okuno. A limiting analysis on regularization of singular SDP and its implication to infeasible interior-point algorithms. *Mathematical Programming (published online, doi:10.1007/s10107-022-01891-8)*, Nov 2022.

[18] H. Waki. How to generate weakly infeasible semidefinite programs via Lasserre's relaxations for polynomial optimization. *Optimization Letters*, 6(8):1883–1896, Dec. 2012.

[19] H. Waki and M. Muramatsu. Facial reduction algorithms for conic optimization problems. *Journal of Optimization Theory and Applications*, 158(1):188–215, 2013.

[20] H. Wolkowicz, R. Saigal, and L. Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Springer, 2000.

A   Proof of item 2 of Proposition 2.1

In this appendix, we provide a proof of item 2 of Proposition 2.1. Without loss of generality, we may assume that $A^1, \ldots, A^m$ are linearly independent. Let $\mathcal{V} \coloneqq \{\sum_{i=1}^m A^i y_i \mid y \in \mathbb{R}^m\}$. Since **RD** is feasible, the subspace

$$\mathcal{W} \coloneqq \{W = \sum_{i=1}^m A^i y_i \mid y \in \mathbb{R}^m, \ W_{12} = 0, W_{22} = 0\}$$

of $\mathcal{V}$ is nonempty. Here we are using the convention that given $Z \in \mathcal{S}^n$, the matrices $Z_{11} \in \mathcal{S}^r, Z_{12} \in \mathbb{R}^{r \times (n-r)}, Z_{22} \in \mathcal{S}^{n-r}$ denote the blocks of Z according to the division given in Proposition 2.1. We also recall that, in view of item (1) of Proposition 2.1, there exists at least one y for which $L_{12}(y) = 0$ and $L_{22}(y) = 0$. This implies that $\mathcal{L}$ satisfies

$$\mathcal{L} = \left\{ \begin{pmatrix} 0 & \sum_{i=1}^m A_{12}^i y_i \\ \sum_{i=1}^m (A_{12}^i)^T y_i & \sum_{i=1}^m A_{22}^i y_i \end{pmatrix} \middle| y \in \mathbb{R}^m \right\}$$

and

$$\begin{pmatrix} 0 & C_{12} \\ C_{12}^T & C_{22} \end{pmatrix} \in \mathcal{L}. \tag{A.1}$$

In particular, $\mathcal{L}$ is a linear space. We define $\mathcal{Y}$ as the subspace of $\mathcal{S}^n$ whose $(1,1)$-block is zero, i.e., $Y \in \mathcal{S}^n$ belongs to $\mathcal{Y}$ if and only if $Y_{11} = 0$. Then, let $\pi : \mathcal{S}^n \to \mathcal{Y}$ denote the orthogonal projection onto $\mathcal{Y}$ so that the following holds for every $Y \in \mathcal{S}^n$:

$$\pi \left(\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^T & Y_{22} \end{pmatrix} \right) = \begin{pmatrix} 0 & Y_{12} \\ Y_{12}^T & Y_{22} \end{pmatrix}.$$

Recall that our perturbation space is $\mathcal{T} \oplus \mathcal{L}$. Given $S \in \mathcal{T} \oplus \mathcal{L}$, S is represented uniquely as the sum of $S_{\mathcal{T}} \in \mathcal{T}$ and $S_{\mathcal{L}} \in \mathcal{L}$. This decomposition is given as

$$S = s_{11}I + \begin{pmatrix} 0 & S_{12} \\ S_{12}^T & S_{22} - s_{11}I_{22} \end{pmatrix},$$

where $I = \begin{pmatrix} I_{11} & 0 \\ 0 & I_{22} \end{pmatrix}$. With that, we have

$$S_{\mathcal{T}} = s_{11}I \in \mathcal{T}, \ S_{\mathcal{L}} = \begin{pmatrix} 0 & S_{12} \\ S_{12}^T & S_{22} - s_{11}I_{22} \end{pmatrix} \in \mathcal{L}. \tag{A.2}$$

Now we are ready to proceed. For the proof of item 2 of Proposition 2.1, we take a suitable basis of $\mathcal{V}$ as described in the following proposition.

Proposition A.1 *There are matrices $B^1, \ldots, B^m \in \mathcal{S}^n$ with the following properties:*

1. *$\{B^1, \ldots, B^m\}$ forms a basis of $\mathcal{V}$.*

2. *$\{B^1, \ldots, B^k\}$ forms a basis of $\mathcal{W}$ $(k < m)$, in particular, $B^i_{12} = 0$ and $B^i_{22} = 0$ for all $i = 1, \ldots, k$.*

3. *Every element $V = \sum_{i=1}^m A^i y_i \in \mathcal{V}$ is written uniquely as*

$$V = \sum_{i=1}^k B^i z_i + \sum_{i=k+1}^m B^i z_i$$
$$= \begin{pmatrix} \sum_{i=1}^k B^i_{11} z_i & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \sum_{i=k+1}^m B^i_{11} z_i & \sum_{i=k+1}^m B^i_{12} z_i \\ \sum_{i=k+1}^m (B^i_{12})^T z_i & \sum_{i=k+1}^m B^i_{22} z_i \end{pmatrix}.$$

4. *There exists a nonsingular $m \times m$ matrix D such that $B^i = \sum_{j=1}^m d_{ij} A^j$ $(i = 1, \ldots, m)$ holds. If $y, z \in \mathbb{R}^m$ are such that $\sum_{i=1}^m A^i y_i = \sum_{j=1}^m B^j z_j$, then $y^T = z^T D$.*

Proof. To obtain the basis $B^1, \ldots, B^m$ satisfying items 1 and 2, we first construct a basis of $\mathcal{W}$ and then expand it by adding independent elements from $\mathcal{V}$ until the chosen elements forms a basis of $\mathcal{V}$. Item 3 is a direct consequence of items 1 and 2, and item 4 follows from a standard argument in linear algebra since $\{A^1, \ldots A^m\}$ and $\{B^1, \ldots, B^m\}$ are two bases of the same linear space $\mathcal{V}$. ∎

Lemma A.2 *With the B^i as in Proposition A.1, define $\tau : \mathbb{R}^{m-k} \to \mathcal{L}$, $\tau_{12} : \mathbb{R}^{m-k} \to \mathbb{R}^{r \times (n-r)}$ and $\tau_{22} : \mathbb{R}^{m-k} \to \mathcal{S}^{n-r}$ such that*

$$\tau(z_{k+1}, \ldots, z_m)$$
$$:= \begin{pmatrix} 0 & \tau_{12}(z_{k+1}, \ldots, z_m) \\ \tau_{12}(z_{k+1}, \ldots, z_m)^T & \tau_{22}(z_{k+1}, \ldots, z_m) \end{pmatrix}$$
$$:= \begin{pmatrix} 0 & \sum_{i=k+1}^m B^i_{12} z_i \\ \sum_{i=k+1}^m B^i_{12} z_i & \sum_{i=k+1}^m B^i_{22} z_i \end{pmatrix} = \pi \left(\sum_{i=k+1}^m B^i z_i \right).$$

Then, τ is a bijective linear map.

Proof. We first observe that τ is surjective. Recall that $\{B^1, \ldots, B^m\}$ is a basis of $\mathcal{V}$ and $\mathcal{L} = \pi(\mathcal{V})$. Pick any element X in $\mathcal{L}$. Since $\mathcal{L} = \pi(\mathcal{V})$, there exists $z \in \mathbb{R}^m$ such that $X = \pi(\sum_{i=1}^m B^i z_i)$. Then it follows that

$$X = \pi\left(\sum_{i=1}^m B^i z_i\right) = \pi\left(\sum_{i=k+1}^m B^i z_i\right) = \sum_{k+1}^m \pi(B^i) z_i = \tau(z^{k+1}, \ldots, z^m),$$

and this proves that τ is a surjection.

Next, we check that τ is injective by showing that its kernel is trivial. Suppose that $\tau(z_{k+1}, \ldots, z_m) = 0$. That is,

$$\begin{pmatrix} 0 & \sum_{i=k+1}^m B_{12}^i z_i \\ \sum_{i=k+1}^m B_{12}^i z_i & \sum_{i=k+1}^m B_{22}^i z_i \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

holds. Therefore, $\sum_{i=k+1}^m B^i z_i \in \mathcal{W}$. Since $B^1, \ldots, B^k$ form a basis for $\mathcal{W}$, there are $(z_1, \ldots, z_k)$ such that

$$\sum_{i=1}^k B^i z_i = \sum_{i=k+1}^m B^i z_i.$$

By the linear independence of the B^i, all the z_i must be 0, so τ is injective as claimed. Thus, τ is surjective and injective, and hence τ is bijective, and the proof is complete. ∎

Now, we are ready to prove item 2, i.e., the optimal value function $w(S)$ of the following problem:

RD(S) $\quad \min_y b^T y$ subject to $L_{11}(y) + s_{11} I_{11} \succeq 0$, $L_{12}(y) = S_{12}$, $L_{22}(y) + s_{11} I_{22} = S_{22}$,

where $S \in \mathcal{T} \oplus \mathcal{L}$, is continuous at $S = 0$. We note that $s_{11} I_{11} = S_{11}$ holds for any $S \in \mathcal{T} \oplus \mathcal{L}$.

To this end, we rewrite this problem in terms of $B^i, i = 1, \ldots, m$ and z, to obtain that

$$\min_z b^T D^T z \quad \text{s.t.} \quad \begin{aligned} & C_{11} + s_{11} I_{11} - \sum_{j=1}^m B_{11}^j z_j \succeq 0, \\ & C_{12} - \sum_{i=k+1}^m B_{12}^i z_i = S_{12}, \\ & C_{22} + s_{11} I_{22} - \sum_{i=k+1}^m B_{22}^i z_i = S_{22}. \end{aligned} \tag{A.3}$$

Now, observe that the two equality constraints in (A.3) can be written as

$$\tau_{12}(z_{k+1}, \ldots, z_m) = C_{12} - S_{12}, \quad \tau_{22}(z_{k+1}, \ldots, z_m) = C_{22} - (S_{22} - s_{11} I_{22}). \tag{A.4}$$

Since τ is a bijective mapping from $\mathbb{R}^{m-k}$ to $\mathcal{L}$ as was shown in Lemma A.2, its inverse mapping is well-defined. Let us denote by τ^{inv} the inverse mapping of τ. Slightly abusing the notation, we write $\tau^{\text{inv}}(\widetilde{L}_{12}, \widetilde{L}_{22})$ as a function of $\widetilde{L}_{12}$ and $\widetilde{L}_{22}$ implicitly assuming that $\begin{pmatrix} 0 & \widetilde{L}_{12} \\ \widetilde{L}_{12}^T & \widetilde{L}_{22} \end{pmatrix} \in \mathcal{L}$. Then, recalling that $\begin{pmatrix} 0 & C_{12} \\ C_{12}^T & C_{22} \end{pmatrix} \in \mathcal{L}$ and $\begin{pmatrix} 0 & S_{12} \\ S_{12}^T & S_{22} - s_{11} I \end{pmatrix} \in \mathcal{L}$ (see (A.1), (A.2)), we see that (A.4) implies that

$$(z_{k+1}, \ldots, z_m) = (\tau_{k+1}^{\text{inv}}(C_{12} - S_{12}, C_{22} - (S_{22} - s_{11} I_{22})), \ldots, \tau_m^{\text{inv}}(C_{12} - S_{12}, C_{22} - (S_{22} - s_{11} I_{22}))$$

must hold, that is, $(z_{k+1}, \ldots, z_m)$ is determined uniquely by the perturbation S in (A.3). Therefore, (A.3) is equivalent to the following semidefinite programs with respect to $(z_1, \ldots, z_k)$:

$$\min_{(z_1,\ldots,z_k)} \quad \sum_{i=1}^{k}(Db)_i z_i + \sum_{i=k+1}^{m}(Db)_i \tau_i^{\mathrm{inv}}(C_{12} - S_{12}, C_{22} - (S_{22} - s_{11}I_{22}))$$

$$\text{s.t.} \quad (C_{11} + s_{11}I_{11} - \sum_{j=k+1}^{m} B_{11}^j \tau_j^{\mathrm{inv}}(C_{12} - S_{12}, C_{22} - (S_{22} - s_{11}I_{22}))) \quad \text{(A.5)}$$

$$- \sum_{j=1}^{k} B_{11}^j z_j \succeq 0.$$

The optimal value function of (A.5) is is equal to $w(S)$.

Since τ^{inv} is a linear mapping and (A.1) and (A.2) holds, we have, for each $j = k + 1, \ldots, m$,

$$\tau_j^{\mathrm{inv}}(C_{12} - S_{12}, C_{22} - (S_{22} - s_{11}I_{22})) = \tau_j^{\mathrm{inv}}(C_{12}, C_{22}) - \tau_j^{\mathrm{inv}}(S_{12}, S_{22} - s_{11}I_{22}).$$

Therefore, we may further rewrite (A.5) as

$$\min_{(z_1,\ldots,z_k)} \sum_{i=1}^{k}(Db)_i z_i + \sum_{i=k+1}^{m}(Db)_i \tau_i^{\mathrm{inv}}(C_{12}, C_{22}) - \sum_{i=k+1}^{m}(Db)_i \tau_i^{\mathrm{inv}}(S_{12}, S_{22} - s_{11}I_{22})$$

$$\text{s.t.} \; (C_{11} + \sum_{j=k+1}^{m} B_{11}^j \tau_j^{\mathrm{inv}}(C_{12}, C_{22})) + s_{11}I_{11} \quad \text{(A.6)}$$

$$- \sum_{j=k+1}^{m} B_{11}^j \tau_j^{\mathrm{inv}}(S_{12}, S_{22} - s_{11}I_{22}) - \sum_{j=1}^{k} B_{11}^j z_j \succeq 0.$$

Overall, (A.6) is equivalent to $\mathbf{RD}(S)$ and the optimal value of (A.6) as a function of perturbation $S \in \mathcal{T} \oplus \mathcal{L}$ coincides with $w(S)$.

Since the second and third term of the objective function is a constant and a linear function of S, respectively, in order to establish continuity of $w(S)$ at $S = 0$, it is enough to show continuity of the optimal value function of the semidefinite program

$$\min_{(z_1,\ldots,z_k)} \sum_{i=1}^{k}(Db)_i z_i$$

$$\text{s.t.} \; (C_{11} + \sum_{j=k+1}^{m} B_{11}^j \tau_j^{\mathrm{inv}}(C_{12}, C_{22})) + s_{11}I_{11} \quad \text{(A.7)}$$

$$- \sum_{j=k+1}^{m} B_{11}^j \tau_j^{\mathrm{inv}}(S_{12}, S_{22} - s_{11}I_{22}) - \sum_{j=1}^{k} B_{11}^j z_j \succeq 0.$$

obtained by dropping the second and the third terms of the objective function in (A.6). The feasible region of (A.7) is the same as (A.6) and hence (A.7) is strongly feasible at

$S = 0$. The perturbation term

$$s_{11}I_{11} - \sum_{j=k+1}^{m} B_{11}^{j} \tau_{j}^{\text{inv}}(S_{12}, S_{22} - s_{11}I_{22})$$

exists only at "the constant part of the constraint" in (A.7) and this term vanishes at $S = 0$. Therefore, we can directly apply Theorem 4.1.9 of [20] to show that the optimal value function of (A.7) is continuous at $S = 0$ and so is the optimal value function of (A.6). This completes the proof.