Preprocessing and Regularization for Degenerate Semidefinite Programs

Yuen-Lam Cheung * Simon Schurr † Henry Wolkowicz ‡

February 8, 2013

University of Waterloo
Department of Combinatorics & Optimization
Waterloo, Ontario N2L 3G1, Canada

Research Report CORR 2011-02

Key words and phrases: semidefinite programming, preprocessing, degeneracy, strong duality, backwards stability.

AMS subject classifications: 90C46, 90C22, 90C25, 49K40, 65K10

Abstract

This paper presents a backward stable preprocessing technique for (nearly) ill-posed semidefinite programming, SDP, problems, i.e., programs for which the Slater constraint qualification, existence of strictly feasible points, (nearly) fails.

Current popular algorithms for semidefinite programming rely on *primal-dual interior-point, p-d i-p* methods. These algorithms require the Slater constraint qualification for both the primal and dual problems. This assumption guarantees the existence of Lagrange multipliers, well-posedness of the problem, and stability of algorithms. However, there are many instances of SDPs where the Slater constraint qualification fails or *nearly* fails. Our backward stable preprocessing technique is based on applying the Borwein-Wolkowicz facial reduction process to find a finite number, k, of *rank-revealing orthogonal rotations* of the problem. After an appropriate truncation, this results in a smaller, well-posed, *nearby* problem that satisfies the Robinson constraint qualification, and one that can be solved by standard SDP solvers. The case $k = 1$ is of particular interest and is characterized by strict complementarity of an auxiliary problem.

Contents

*Department of Combinatorics and Optimization, University of Waterloo, Ontario N2L 3G1, Canada Research supported by TATA Consultancy Services.

†Research supported by The Natural Sciences and Engineering Research Council of Canada. Email: spschurr@math.uwaterloo.ca

‡Research supported by The Natural Sciences and Engineering Research Council of Canada and by TATA Consultancy Services. Email: hwolkowicz@uwaterloo.ca

[0] URL for paper: orion.math.uwaterloo.ca/~hwolkowi/henry/reports/ABSTRACTS.html

List of Algorithms

List of Tables

List of Figures

1 Introduction

The aim of this paper is to develop a backward stable preprocessing technique to handle (nearly) ill-posed semidefinite programming, SDP, problems, i.e., programs for which the Slater constraint qualification (Slater CQ, or SCQ), the existence of strictly feasible points, (nearly) fails. The technique is based on applying the Borwein-Wolkowicz *facial reduction* process [11, 12] to find a finite number k of *rank-revealing orthogonal rotation* steps. Each step is based on solving an auxiliary problem (AP) where it and its dual satisfy the Slater CQ. After an appropriate truncation, this results in a smaller, well-posed, *nearby* problem for which the Robinson constraint qualification (RCQ) [52] holds; and one that can be solved by standard SDP solvers. In addition, the case $k = 1$ is of particular interest and is characterized by strict complementarity of the (AP).

In particular, we study SDPs of the following form

$$\text{(P)} \qquad v_P := \sup_y \{b^T y \ : \mathcal{A}^* y \preceq C\}, \tag{1.1}$$

where the optimal value v_P is finite, $b \in \mathbb{R}^m$, $C \in \mathbb{S}^n$, and $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is an onto linear transformation from the space $\mathbb{S}^n$ of $n \times n$ real symmetric matrices to $\mathbb{R}^m$. The adjoint of $\mathcal{A}$ is $\mathcal{A}^* y = \sum_{i=1}^m y_i A_i$, where $A_i \in \mathbb{S}^n, i = 1, \ldots, m$. The symbol $\preceq$ denotes the Löwner partial order induced by the cone $\mathbb{S}^n_+$ of positive semidefinite matrices, i.e., $\mathcal{A}^* y \preceq C$ if and only if $C - \mathcal{A}^* y \in \mathbb{S}^n_+$. (Note that the cone optimization problem (1.1) is commonly used as the dual problem in the SDP literature, though it is often the primal in the Linear Matrix Inequality (LMI) literature, e.g., [13].) If (P) is *strictly feasible*, then one can use standard solution techniques; if (P) is *strongly infeasible*, then one can set $v_P = -\infty$, e.g., [38, 43, 47, 62, 66]. If neither of these two feasibility conditions can be verified, then we apply our preprocessing technique that finds a rotation of the problem that is akin to *rank-revealing* matrix rotations. (See e.g., [58, 59] for equivalent matrix results.) This rotation finds an equivalent (nearly) block diagonal problem which allows for simple strong dualization by solving only the most significant block of (P) for which the Slater CQ holds. This is equivalent to restricting the original problem to a face of $\mathbb{S}^n_+$, i.e., the preprocessing can be considered as a *facial reduction* of (P) . Moreover, it provides a *backward stable* approach for solving (P) when it is feasible and the SCQ fails; and it solves a nearby problem when (P) is *weakly infeasible*.

The Lagrangian dual to (1.1) is

$$\text{(D)} \qquad v_D := \inf_X \left\{ \langle C, X \rangle : \mathcal{A}(X) = b, X \succeq 0 \right\}, \tag{1.2}$$

where $\langle C, X \rangle := \operatorname{trace} CX = \sum_{ij} C_{ij} X_{ij}$ denotes the trace inner product of the symmetric matrices C and X; and, $\mathcal{A}(X) = (\langle A_i, X \rangle) \in \mathbb{R}^m$. Weak duality $v_D \geq v_P$ follows easily. The usual constraint qualification (CQ) used for (P) is SCQ, i.e., strict feasibility $\mathcal{A}^*y \prec C$ (or $C - \mathcal{A}^*y \in \mathbb{S}^n_{++}$, the cone of positive definite matrices). If we assume the Slater CQ holds and the primal optimal value is finite, then strong duality holds, i.e., we have a zero duality gap and attainment of the dual optimal value. Strong duality results for (1.1) without any constraint qualification are given in [10, 11, 12, 72] and [48, 49], and more recently in [50, 65]. Related closure conditions appear in [44]; and, properties of problems where strong duality fails appear in [45].

General surveys on SDP are in e.g., [4, 63, 68, 74]. Further general results on SDP appear in the recent survey [31].

Many popular algorithms for (P) are based on Newton's method and a *primal-dual interior-point, p-d i-p,* approach, e.g., the codes (latest at the URLs in the citations) CSDP, SeDuMi, SDPT3, SDPA [9, 60, 67, 76]; see also the
SDP URL: www-user.tu-chemnitz.de/˜helmberg/sdp_software.html.
To find the search direction, these algorithms apply symmetrization in combination with block elimination to find the Newton search direction. The symmetrization and elimination steps both result in ill-conditioned linear systems, even for well conditioned SDP problems, e.g., [19, 73]. And, these methods are very susceptible to numerical difficulties and high iteration counts in the case when SCQ nearly fails, see e.g., [21, 22, 23, 24]. Our aim in this paper is to provide a stable regularization process based on orthogonal rotations for problems where strict feasibilty (nearly) fails. Related papers on regularization are e.g., [30, 39]; and papers on high accuracy solutions for algorithms SDPA-GMP,-QD,-DD are e.g., [77]. In addition, a popular approach uses a selfdual embedding e.g., [16, 17]. This approach results in SCQ holding by using homogenization and increasing the number of variables. In contrast, our approach reduces the size of the problem in a preprocessing step in order to guarantee SCQ.

1.1 Outline

We continue in Section 1.2 with preliminary notation and results for cone programming. In Section 2 we recall the history and outline the similarities and differences of what facial reduction means first for linear programming (LP), and then for ordinary convex programming (CP), and finally for SDP, which has elements from both LP and CP. Instances and applications where the SCQ fails are given in Section 2.3.1. Then, Section 3 presents the theoretical background and tools needed for the facial reduction algorithm for SDP. This includes results on strong duality in Section 3.1; and, various theorems of the alternative, with cones having both nonempty and empty interior, are given in Section 3.2. A stable auxiliary problem (3.5) for identifying the minimal face containing the feasible set is presented and studied in Section 3.3; see e.g., Theorem 3.11. In particular, we relate the question of transforming the unstable problem of finding the minimal face to the existence of a primal-dual optimal pair satisfying strict complementarity and to the number of steps in the facial reduction. See Remark 3.10 and Section 3.5. The resulting information from the auxiliary problem for problems where SCQ (nearly) fails is given in Theorem 3.15 and Propositions 3.16, 3.17. This information can be used to construct equivalent problems. In particular, a rank-revealing rotation is used in Section 3.4 to yield two equivalent problems that are useful in sensitivity analysis, see Theorem 3.20. In particular, this shows the backwards stability with respect to perturbations in the parameter β in the definition of the cone T_β for the problem. Truncating the (near) singular blocks to zero yields two smaller equivalent, regularized problems in Section 3.4.1.

The facial reduction is studied in Section 4. An outline of the facial reduction using a rank-revealing rotation process is given in Section 4.1. Backward stability results are presented in

Section 4.2.

Preliminary numerical tests, as well as a technique for generating instances with a finite duality gap useful for numerical tests, are given in Section 5. Concluding remarks appear in Section 6. (An index is included to help the reader, see page 49.)

1.2 Preliminary definitions

Let $(\mathcal{V}, \langle \cdot, \cdot \rangle_{\mathcal{V}})$ be a finite-dimensional inner product space, and K be a (closed) *convex cone* in $\mathcal{V}$, i.e., $\lambda K \subseteq K, \forall \lambda \geq 0$, and $K + K \subseteq K$. K is *pointed* if $K \cap (-K) = \{0\}$; K is *proper* if K is pointed and $\operatorname{int} K \neq \emptyset$; the *polar* or *dual cone* of K is $K^* := \{\phi : \langle \phi, k \rangle \geq 0, \forall k \in K\}$. We denote by $\preceq_K$ the partial order with respect to K. That is, $x_1 \preceq_K x_2$ means that $x_2 - x_1 \in K$. We also write $x_1 \prec_K x_2$ to mean that $x_2 - x_1 \in \operatorname{int} K$. In particular with $\mathcal{V} = \mathbb{S}^n$, $K = \mathbb{S}^n_+$ yields the partial order induced by the cone of positive semidefinite matrices in $\mathbb{S}^n$, i.e., the so-called Löwner partial order. We denote this simply with $X \preceq Y$ for $Y - X \in \mathbb{S}^n_+$. $\operatorname{cone}(S)$ denotes the convex cone generated by the set S. In particular, for any non-zero vector x, the *ray generated by* x is defined by $\operatorname{cone}(x)$. The ray generated by $s \in K$ is called an *extreme ray* if $0 \preceq_K u \preceq_K s$ implies that $u \in \operatorname{cone}(s)$. The subset $F \subseteq K$ is a *face of the cone* K, denoted $F \trianglelefteq K$, if

$$(s \in F, 0 \preceq_K u \preceq_K s) \implies (\operatorname{cone}(u) \subseteq F). \tag{1.3}$$

Equivalently, $F \trianglelefteq K$ if F is a cone and $\left(x, y \in K, \frac{1}{2}(x+y) \in F\right) \implies (\{x, y\} \subseteq F)$. If $F \trianglelefteq K$ but is not equal to K, we write $F \lhd K$. If $\{0\} \neq F \lhd K$, then F is a *proper face* of K. For $S \subseteq K$, we let face(S) denote the smallest face of K that contains S. A face $F \trianglelefteq K$ is an *exposed face* if it is the intersection of K with a hyperplane. The cone K is *facially exposed* if every face $F \trianglelefteq K$ is exposed. If $F \trianglelefteq K$, then the *conjugate face* is $F^c := K^* \cap \{F\}^\perp$. Note that the conjugate face F^c is *exposed* using any $s \in \operatorname{relint} F$ (where $\operatorname{relint} S$ denotes the *relative interior* of the set S), i.e., $F^c = K^* \cap \{s\}^\perp, \forall s \in \operatorname{relint} F$. In addition, note that $\mathbb{S}^n_+$ is self-dual (i.e., $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$) and is facially exposed.

For the general conic programming problem, the constraint linear transformation $\mathcal{A} : \mathcal{V} \to \mathcal{W}$ maps between two Euclidean spaces. The adjoint of $\mathcal{A}$ is denoted by $\mathcal{A}^* : \mathcal{W} \to \mathcal{V}$, and the Moore-Penrose generalized inverse of $\mathcal{A}$ is denoted by $\mathcal{A}^\dagger : \mathcal{W} \to \mathcal{V}$.

A linear conic program may take the form

$$(\mathrm{P}_{\mathrm{conic}}) \qquad v_P^{\mathrm{conic}} = \sup_y \{\langle b, y \rangle \ : \ C - \mathcal{A}^* y \succeq_K 0\}, \tag{1.4}$$

with $b \in \mathcal{W}$ and $C \in \mathcal{V}$. Its dual is given by

$$(\mathrm{D}_{\mathrm{conic}}) \qquad v_D^{\mathrm{conic}} = \inf_X \{\langle C, X \rangle \ : \mathcal{A}(X) = b, X \succeq_{K^*} 0\}. \tag{1.5}$$

Note that the Robinson constraint qualification (RCQ) is said to hold for the linear conic program $(\mathrm{P}_{\mathrm{conic}})$ if $0 \in \operatorname{int}(C - \mathcal{A}^*(\mathbb{R}^m) - \mathbb{S}^n_+)$; see [53]. As pointed out in [61], the Robinson CQ is equivalent to the Mangasarian-Fromovitz constraint qualification in the case of conventional nonlinear programming. Also, it is easy to see that the Slater CQ, strict feasibility, implies RCQ.

Denote the feasible solution and slack sets of (1.4) and (1.5) by $\mathcal{F}_P = \mathcal{F}_P^y = \{y : \ \mathcal{A}^* y \preceq_K C\}$, $\mathcal{F}_P^Z = \{Z : \ Z = C - \mathcal{A}^* y \succeq_K 0\}$, and $\mathcal{F}_D = \{X : \ \mathcal{A}(X) = b, \ X \succeq_{K^*} 0\}$, respectively. The *minimal face* of (1.4) is the intersection of all faces of K containing the feasible slack vectors:

$$f_P = f_P^Z := \operatorname{face}(C - \mathcal{A}^*(\mathcal{F}_P)) = \cap \{H \trianglelefteq K \ : \ C - \mathcal{A}^*(\mathcal{F}_P) \subseteq H\}.$$

Here, $\mathcal{A}^*(\mathcal{F}_P)$ is the linear image of the set $\mathcal{F}_P$ under $\mathcal{A}^*$.

We continue with the notation specifically for $\mathcal{V} = \mathbb{S}^n$, $K = \mathbb{S}^n_+$ and $\mathcal{W} = \mathbb{R}^m$. Then (1.4) (respectively, (1.5)) is the same as (1.1) (respectively, (1.2)). We let e_i denote the i-th unit vector, and $E_{ij} := \frac{1}{\sqrt{2}}(e_i e_j^T + e_j e_i^T)$ are the unit matrices in $\mathbb{S}^n$. for specific $A_i \in \mathbb{S}^n, i = 1, \ldots, m$. We let $\|\mathcal{A}\|_2$ denote the spectral norm of $\mathcal{A}$ and define the Frobenius norm (Hilbert-Schmidt norm) of $\mathcal{A}$ as $\|\mathcal{A}\|_F := \sqrt{\sum_{i=1}^m \|A_i\|_F^2}$.

Unless stated otherwise, all vector norms are assumed to be 2-norm, and all matrix norms in this paper are Frobenius norms. Then, e.g., [32, Chapter 5], for any $X \in \mathbb{S}^n$,

$$\|\mathcal{A}(X)\|_2 \leq \|\mathcal{A}\|_2 \|X\|_F \leq \|\mathcal{A}\|_F \|X\|_F. \tag{1.6}$$

We summarize our assumptions in the following.

Assumption 1.1. *$\mathcal{F}_P \neq \emptyset$; $\mathcal{A}$ is onto.*

2 Framework for Regularization/Preprocessing

The case of preprocessing for linear programming is well known. The situation for general convex programming is not. We now outline the preprocessing and facial reduction for the cases of: linear programming, (LP); ordinary convex programming, (CP); and SDP. We include details on motivation involving numerical stability and convergence for algorithms. In all three cases, the facial reduction can be regarded as a Robinson type regularization procedure.

2.1 The case of linear programming, LP

Preprocessing is essential for LP, in particular for the application of interior point methods. Suppose that the constraint in (1.4) is $\mathcal{A}^* y \preceq_K c$ with $K = \mathbb{R}^n_+$, the nonnegative orthant, i.e., it is equivalent to the elementwise inequality $A^T y \leq c, c \in \mathbb{R}^n$, with the (full row rank) matrix A being $m \times n$. Then (P_{conic}) and (D_{conic}) form the standard primal-dual LP pair. Preprocessing is an essential step in algorithms for solving LP, e.g., [20, 27, 35]. In particular, interior-point methods require strictly feasible points for both the primal and dual LPs. Under the assumption that $\mathcal{F}_P \neq \emptyset$, lack of strict feasibility for the primal is equivalent to the existence of an unbounded set of dual optimal solutions. This results in convergence problems, since current primal-dual interior point methods follow the *central path* and converge to the analytic center of the optimal set. From a standard Farkas' Lemma argument, we know that the Slater CQ, the existence of a strictly feasible point $A^T \hat{y} < c$, holds if and only if

$$\text{the system } \boxed{0 \neq d \geq 0, Ad = 0, c^T d = 0} \text{ is inconsistent.} \tag{2.1}$$

In fact, after a permutation of columns if needed, we can partition both A, c as

$$A = \begin{bmatrix} A^< & A^= \end{bmatrix}, \text{ with } A^= \text{ size } m \times t, \quad c = \begin{pmatrix} c^< \\ c^= \end{pmatrix},$$

so that we have

$$A^{<T} \hat{y} < c^<, \quad A^{=T} \hat{y} = c^=, \text{ for some } \hat{y} \in \mathbb{R}^m, \qquad \text{and } A^T y \leq c \implies A^{=T} y = c^=,$$

i.e. the constraints $A^{=T} y \leq c^{=}$ are the *implicit equality constraints*, with indices given in

$$\mathcal{P} := \{1, \ldots, n\}, \quad \mathcal{P}^{<} := \{1, \ldots, n-t\}, \quad \mathcal{P}^{=} := \{n-t+1, \ldots, n\}.$$

Moreover, the indices for $c^{=}$ (and columns of $A^{=}$) correspond to the indices in a *maximal positive* solution d in (2.1); and, the nonnegative linear dependence in (2.1) implies that there are redundant implicit equality constraints that we can discard, yielding the smaller $(A_R^{=})^T y = c_R^{=}$ with $A_R^{=}$ full column rank. Therefore, an equivalent problem to ($\mathrm{P}_{\mathrm{conic}}$) is

$$(\mathrm{P}_{reg}) \qquad v_P := \max\{b^T y \; : A^{<T} y \leq c^{<}, \; A_R^{=T} y = c_R^{=}\}. \tag{2.2}$$

And this LP satisfies the Robinson constraint qualification (RCQ); see Corollary 3.4, Item 2, below. In this case RCQ is equivalent to the Mangasarian-Fromovitz constraint qualification (MFCQ), i.e., there exists a feasible $\hat{y}$ which satisfies the inequality constraints strictly, $A^{<T} \hat{y} < c^{<}$, and the matrix $A^{=}$ for the equality constraints is full row rank, see e.g., [8, 40]. The MFCQ characterizes stability with respect to right-hand side perturbations and is equivalent to having a compact set of dual optimal solutions. Thus, recognizing and changing the implicit equality constraints to equality constraints and removing redundant equality constraints provides a simple *regularization of LP*.

Let f_P denote the minimal face of the LP. Then note that we can rewrite the constraint as

$$A^T y \preceq_{f_P} c, \quad \text{with } f_P := \{z \in \mathbb{R}^n_+ : z_i = 0, i \in \mathcal{P}^{=}\}.$$

Therefore, rewriting the constraint using the minimal face provides a regularization for LP. This is followed by discarding redundant equality constraints to obtain the MFCQ. This reduces the number of constraints and thus the dimension of the dual variables. Finally, the dimension of the problem can be further reduced by eliminating the equality constraints completely using the nullspace representation. However, this last step can result in loss of sparsity and is usually not done.

We can similarly use a theorem of the alternative to recognize failure of strict feasibility in the dual, i.e., the (in)consistency of the system $0 \neq A^T v \geq 0, b^T v = 0$. This corresponds to identifying which variables x_i are identically zero on the feasible set. The regularization then simply discards these variables along with the corresponding columns of A, c.

2.2 The case of ordinary convex programming, CP

We now move from LP to nonlinear convex programming. We consider the *ordinary convex program (CP)*

$$(\mathrm{CP}) \qquad v_{CP} := \sup\{b^T y \; : g(y) \leq 0\}, \tag{2.3}$$

where $g(y) = (g_i(y)) \in \mathbb{R}^n$, and $g_i : \mathbb{R}^m \to \mathbb{R}$ are convex functions, for all i. (Without loss of generality, we let the objective function $f(y) = b^T y$ be linear. This can always be achieved by replacing a concave objective function with a new variable $\sup t$, and adding a new constraint $-f(y) \leq -t$.) The quadratic programming case has been well studied, [28, 28, 42]. Some preprocessing results for the general CP case are known, e.g., [15]. However, preprocessing for general CP is not as well known as for LP. In fact, see [6], as for LP there is a set of *implicit equality constraints for CP*, i.e. we can partition the constraint index set $\mathcal{P} = \{1, \ldots, n\}$ into two sets

$$\mathcal{P}^{=} = \{i \in \mathcal{P} : y \text{ feasible } \implies g_i(y) = 0\}, \quad \mathcal{P}^{<} = \mathcal{P} \backslash \mathcal{P}^{=}. \tag{2.4}$$

Therefore, as above for LP, we can rewrite the constraints in CP using the minimal face f_P to get $g(y) \preceq_{f_P} 0$. However, this is not a true convex program since the new equality constraints are not affine. However, surprisingly the corresponding feasible set for the implicit equality constraints is convex, e.g., [6]. We include the result and a proof for completeness.

Lemma 2.1. *Let the convex program (CP) be given, and let $\mathcal{P}^=$ be defined as in (2.4). Then the set $\mathcal{F}^= := \{y : g_i(y) = 0, \forall i \in \mathcal{P}^=\}$ satisfies*

$$\mathcal{F}^= = \{y : g_i(y) \leq 0, \forall i \in \mathcal{P}^=\},$$

and thus is a convex set.

Proof. Let $g^=(y) = (g_i(y))_{i\in\mathcal{P}^=}$ and $g^<(y) = (g_i(y))_{i\in\mathcal{P}^<}$. By definition of $\mathcal{P}^<$, there exists a feasible $\hat{y} \in \mathcal{F}$ with $g^<(\hat{y}) < 0$; and, suppose that there exists $\bar{y}$ with $g^=(\bar{y}) \leq 0$, and $g_{i_0}(\bar{y}) < 0$, for some $i_0 \in \mathcal{P}^=$. Then for small $\alpha > 0$ the point $y_\alpha := \alpha\hat{y} + (1-\alpha)\bar{y} \in \mathcal{F}$ and $g_{i_0}(y_\alpha) < 0$. This contradicts the definition of $\mathcal{P}^=$. □

This means that we can regularize CP by replacing the implicit equality constraints as follows

$$(\mathrm{CP}_{reg}) \qquad v_{CP} := \sup\{b^T y \ : g^<(y) \leq 0, y \in \mathcal{F}^=\}. \tag{2.5}$$

The generalized Slater CQ holds for the *regularized convex program* (CP_{reg}). Let

$$\phi(\lambda) = \sup_{y\in\mathcal{F}^=} b^T y - \lambda^T g^<(y)$$

denote the *regularized dual functional for CP*. Then strong duality holds for CP with the *regularized dual program*, i.e.

$$\begin{array}{rcl} v_{CP} = v_{CPD} & := & \inf_{\lambda\geq 0} \phi(\lambda) \\ & = & \phi(\lambda^*), \end{array}$$

for some (dual optimal) $\lambda^* \geq 0$. The Karush-Kuhn-Tucker (KKT) optimality conditions applied to (2.5) imply that

$$\begin{array}{c} y^* \text{ is optimal for } \mathrm{CP}_{reg} \\ \text{if and only if} \end{array}$$
$$\begin{cases} y^* \in \mathcal{F} & \text{(primal feasibility)} \\ b - \nabla g^<(y^*)\lambda^* \in (\mathcal{F}^= - y^*)^*, \text{ for some } \lambda^* \geq 0 & \text{(dual feasibility)} \\ g^<(y^*)^T\lambda^* = 0 & \text{(complementary slackness)} \end{cases}$$

This differs from the standard KKT conditions in that we need the polar set

$$(\mathcal{F}^= - y^*)^* = \overline{\mathrm{cone}\ (\mathcal{F}^= - y^*)}^* = (D^=(y^*))^*, \tag{2.6}$$

where $D^=(y^*)$ denotes the *cone of directions of constancy* of the implicit equality constraints $\mathcal{P}^=$, e.g., [6]. Thus we need to be able to find this cone numerically, see, [71]. A backward stable algorithm for the cone of directions of constancy is presented in [37].

Note that a convex function f is faithfully convex if f is affine on a line segment only if it is affine on the whole line containing that segment; see [54]. Analytic convex functions are faithfully convex, as are strictly convex functions . For faithfully convex functions, the set $\mathcal{F}^=$ is an affine manifold, $\mathcal{F}^= = \{y : Vy = V\hat{y}\}$, where $\hat{y} \in \mathcal{F}$ is feasible, and the nullspace of the matrix V gives the intersection of the cones of directions of constancy $D^=$. Without loss of generality, let V be chosen full row rank. Then in this case we can rewrite the regularized problem as

$$(\mathrm{CP}_{reg}) \qquad v_{CP} := \sup\{b^T y \ : g^<(y) \le 0, Vy = V\hat{y}\}, \tag{2.7}$$

which is a convex program for which the MFCQ holds. Thus by identifying the implicit equalities and replacing them with the linear equalities that represent the cone of directions of constancy, we obtain the regularized convex program. If we let $g^R(y) = \begin{pmatrix} g^<(y) \\ Vy - V\hat{y} \end{pmatrix}$, then writing the constraint $g(y) \le 0$ using g^R and the minimal cone f_P as $g^R(y) \preceq_{f_P} 0$ results in the regularized CP for which MFCQ holds.

2.3 The case of semidefinite programming, SDP

Finally, we consider our case of interest, the SDP given in (1.1). In this case, the cone for the constraint partial order is $\mathbb{S}^n_+$, a *nonpolyhedral* cone. Thus we have elements of both LP and CP. Significant preprocessing is not done in current public domain SDP codes. Theoretical results are known, see e.g., [34] for results on redundant constraints using a probabilistic approach. However, [10], the notion of minimal face can be used to regularize SDP. Surprisingly, the above result for LP in (2.2) holds. A regularized problem for (P) for which strong duality holds has constraints of the form $\mathcal{A}^* y \preceq_{f_P} C$ without the need for an extra polar set as in (2.6) that is used in the CP case, i.e., changing the cone for the partial order regularizes the problem. However, as in the LP case where we had to discard redundant implicit equality constraints, extra work has to be done to ensure that the RCQ holds. The details for the facial reduction now follow in Section 3. An equivalent regularized problem is presented in Corollary 3.22, i.e., rather than a permutation of columns needed in the LP case, we perform a rotation of the problem constraint matrices, and then we get a similar division of the constraints as in (2.2); and, setting the implicit equality constraints to equality results in a regularized problem for which the RCQ holds.

2.3.1 Instances where the Slater CQ fails for SDP

Instances where SCQ fails for CP are given in [6]. It is known that the SCQ holds generically for SDP, e.g., [3]. However, there are surprisingly many SDPs that arise from relaxations of hard combinatorial problems where SCQ fails. In addition, there are many instances where the structure of the problems allows for exact facial reduction. This was shown for the quadratic assignment problem in [80] and for the graph partitioning problem in [75]. For these two instances, the barycenter of the feasible set is found explicitly and then used to project the problem onto the minimal face; thus we simultaneously regularize and simplify the problems. In general, the affine hull of the feasible solutions of the SDP are found and used to find Slater points. This is formalized and generalized in [64, 66]. In particular, SDP relaxations that arise from problems with matrix variables that have $0, 1$ constraints along with row and column constraints result in SDP relaxations where the Slater CQ fails.

Important applications occur in the facial reduction algorithm for sensor network localization and molecular conformation problems given in [36]. Cliques in the graph result in corresponding dimension reduction of the minimal face of the problem resulting in efficient and accurate solution techniques. Another instance is the SDP relaxation of the side chain positioning problem studied in [14]. Further Applications that exploit the failure of the Slater CQ for SDP relaxations appear in e.g., [1, 2, 5, 69].

3 Theory

We now present the theoretical tools that are needed for the facial reduction algorithm for SDP. This includes the well known results for strong duality, the theorems of the alternative to identify strict feasibility, and, in addition, a stable subproblem to apply the theorems of the alternative. Note that we use K to represent the cone $\mathbb{S}^n_+$ to emphasize that many of the results hold for more general closed convex cones.

3.1 Strong duality for cone optimization

We first summarize some results on *strong duality* for the conic convex program in the form (1.4). Strong duality for (1.4) means that there is a *zero duality gap*, $v_P^{\text{conic}} = v_D^{\text{conic}}$, and the dual optimal value v_D (1.5) is attained. However, it is easy to construct examples where strong duality fails, see e.g., [45, 49, 74] and Section 5, below.

It is well known that for a finite dimensional LP, strong duality fails only if the primal problem and/or its dual are infeasible. In fact, in LP both problems are feasible and both of the optimal values are attained (and equal) if, and only if, the optimal value of one of the problems is finite. In general (conic) convex optimization, the situation is more complicated, since the underlying cones in the primal and dual optimization problems need not be polyhedral. Consequently, even if a primal problem and its dual are feasible, a nonzero duality gap and/or non-attainment of the optimal values may ensue unless some *constraint qualification* holds; see e.g., [7, 55]. More specific examples for our cone situations appear in e.g., [38], [51, Section 3.2], and [63, Section 4].

Failure of strong duality is problematic, since many classes of p-d i-p algorithms require not only that a primal-dual pair of problems possess a zero duality gap, but also that the (generalized) Slater CQ holds for both primal and dual, i.e., that strict feasibility holds for both problems. In [10, 11, 12], an equivalent *strongly dualized primal problem* corresponding to (1.4), given by

$$\text{(SP)} \qquad v_{SP}^{\text{conic}} := \sup\{\langle b, y\rangle \ : \ \mathcal{A}^* y \preceq_{f_P} C\}, \tag{3.1}$$

where $f_P \unlhd K$ is the minimal face of K containing the feasible region of (1.4), is considered. The equivalence is in the sense that the feasible set is unchanged

$$\mathcal{A}^* y \preceq_K C \iff \mathcal{A}^* y \preceq_{f_P} C.$$

This means that for any face F we have

$$f_P \unlhd F \unlhd K \implies \{\mathcal{A}^* y \preceq_K C \iff \mathcal{A}^* y \preceq_F C\}.$$

The Lagrangian dual of (3.1) is given by

$$\text{(DSP)} \qquad v_{DSP}^{\text{conic}} := \inf\{\langle C, X\rangle \ : \ \mathcal{A}(X) = b, \ X \succeq_{f_P^*} 0\}. \tag{3.2}$$

We note that the linearity of the constraint means that an equality set of the type in (2.6) is not needed.

Theorem 3.1 ([10])**.** *Suppose that the optimal value* v_P^{conic} *in* (1.4) *is finite. Then strong duality holds for the pair* (3.1) *and* (3.2)*, or equivalently, for the pair* (1.4) *and* (3.2)*; i.e.,* $v_P^{\text{conic}} = v_{SP}^{\text{conic}} = v_{DSP}^{\text{conic}}$ *and the dual optimal value* v_{DSP}^{conic} *is attained.* ■

3.2 Theorems of the alternative

In this section, we state some theorems of the alternative for the Slater CQ of the conic convex program (1.4), which are essential to our reduction process. We first recall the notion of recession direction (for the dual (1.5)) and its relationship with the minimal face of the primal feasible region.

Definition 3.2. *The convex cone of* recession directions *for* (1.5) *is*

$$\mathcal{R}_{\mathrm{D}} := \{D \in \mathcal{V} : \mathcal{A}(D) = 0,\ \langle C, D\rangle = 0,\ D \succeq_{K^*} 0\}. \tag{3.3}$$

The cone $\mathcal{R}_{\mathrm{D}}$ consists of feasible directions for the homogeneous problem along which the dual objective function is constant.

Lemma 3.3. *Suppose that the feasible set* $\mathcal{F}_P \neq \emptyset$ *for* (1.4)*, and let* $0 \neq D \in \mathcal{R}_{\mathrm{D}}$*. Then the minimal face of* (1.4) *satisfies*

$$f_P \trianglelefteq K \cap \{D\}^\perp \triangleleft K.$$

Proof. We have

$$0 = \langle C, D\rangle - \langle \mathcal{F}_P, \mathcal{A}(D)\rangle = \langle C - \mathcal{A}^*(\mathcal{F}_P), D\rangle.$$

Hence $C - \mathcal{A}^*(\mathcal{F}_P) \subseteq \{D\}^\perp \cap K$, which is a face of K. It follows that $f_P \subseteq \{D\}^\perp \cap K$. The required result now follows from the fact that f_P is (by definition) a face of K, and D is nonzero. □

Lemma 3.3 indicates that if we are able to find an element $D \in \mathcal{R}_{\mathrm{D}}\backslash\{0\}$, then D gives us a smaller face of K that contains $\mathcal{F}_P^Z$. The following lemma shows that the existence of such a direction D is *equivalent* to the failure of the Slater CQ for a feasible program (1.4). The lemma specializes [12, Theorem 7.1] and forms the basis of our reduction process.

Lemma 3.4 ([12])**.** *Suppose that* $\operatorname{int} K \neq \emptyset$ *and* $\mathcal{F}_P \neq \emptyset$*. Then exactly one of the following two systems is consistent:*

1. $\mathcal{A}(D) = 0$, $\langle C, D\rangle = 0$, *and* $0 \neq D \succeq_{K^*} 0$ $\qquad$ ($\mathcal{R}_{\mathrm{D}}\backslash\{0\}$)
2. $\mathcal{A}^* y \prec_K C$ $\qquad$ *(Slater CQ)*

Proof. Suppose that D satisfies the system in Item 1. Then for all $y \in \mathcal{F}_P$, we have $\langle C - \mathcal{A}^*y, D\rangle = \langle C, D\rangle - \langle y, (\mathcal{A}(D))\rangle = 0$. Hence $\mathcal{F}_P^Z \subseteq K \cap \{D\}^\perp$. But $\{D\}^\perp \cap \operatorname{int} K = \emptyset$ as $0 \neq D \succeq_{K^*} 0$. This implies that the Slater CQ (as in Item 2) fails.

Conversely, suppose that the Slater CQ in Item 2 fails. We have $\operatorname{int} K \neq \emptyset$ and

$$0 \notin (\mathcal{A}^*(\mathbb{R}^m) - C) + \operatorname{int} K.$$

Therefore, we can find $D \neq 0$ to separate the open set $(\mathcal{A}^*(\mathbb{R}^m) - C) + \operatorname{int} K$ from 0. Hence we have

$$\langle D, Z\rangle \geq \langle D, C - \mathcal{A}^* y\rangle,$$

for all $Z \in K$ and $y \in \mathcal{W}$. This implies that $D \in K^*$ and $\langle D, C\rangle \leq \langle D, \mathcal{A}^* y\rangle$, for all $y \in \mathcal{W}$. This implies that $\langle \mathcal{A}(D), y\rangle = 0$ for all $y \in \mathcal{W}$; hence $\mathcal{A}(D) = 0$. To see that $\langle C, D\rangle = 0$, fix any $\hat{y} \in \mathcal{F}_P$. Then $0 \geq \langle D, C\rangle = \langle D, C - \mathcal{A}^*\hat{y}\rangle \geq 0$, so $\langle D, C\rangle = 0$. □

We have an equivalent characterization for the generalized Slater CQ for the dual problem. This can be used to extend our results to $(\mathrm{D}_{\mathrm{conic}})$.

Corollary 3.5. *Suppose that* $\operatorname{int} K^* \neq \emptyset$ *and* $\mathcal{F}_D \neq \emptyset$*. Then exactly one of the following two systems is consistent:*

1. $0 \neq \mathcal{A}^* v \succeq_K 0$*, and* $\langle b, v\rangle = 0$*.*
2. $\mathcal{A}(X) = b, X \succ_{K^*} 0$ *(generalized Slater CQ).*

Proof. Let $\mathcal{K}$ be a one-one linear transformation with range $\mathcal{R}(\mathcal{K}) = \mathcal{N}(\mathcal{A})$, and let $\hat{X}$ satisfy $\mathcal{A}(\hat{X}) = b$. Then, Item 2 is consistent if, and only if, there exists $\hat{u}$ such that $X = \hat{X} - \mathcal{K}\hat{u} \succ_{K^*} 0$. This is equivalent to $\mathcal{K}\hat{u} \prec_{K^*} \hat{X}$. Therefore, $\mathcal{K}, \hat{X}$ play the roles of $\mathcal{A}^*, C$, respectively, in Lemma 3.4. Therefore, an alternative system is $\mathcal{K}^*(Z) = 0, 0 \neq Z \succeq_K 0$, and $\langle \hat{X}, Z\rangle = 0$. Since $\mathcal{N}(\mathcal{K}^*) = \mathcal{R}(\mathcal{A}^*)$, this is equivalent to $0 \neq Z = \mathcal{A}^* v \succeq_K 0$, and $\langle \hat{X}, Z\rangle = 0$, or $0 \neq \mathcal{A}^* v \succeq_K 0$, and $\langle b, v\rangle = 0$. □

We can extend Lemma 3.4 to problems with additional equality constraints.

Corollary 3.6. *Consider the modification of the primal* (1.4) *obtained by adding equality constraints:*

$$(P_B) \qquad v_{P_B} := \sup\{\langle b, y\rangle \ : \mathcal{A}^* y \preceq_K C, \mathcal{B}y = f\}, \tag{3.4}$$

where $\mathcal{B} : \mathcal{W} \to \mathcal{W}'$ *is an onto linear transformation. Assume that* $\operatorname{int} K \neq \emptyset$ *and* (P_B) *is feasible. Let* $\bar{C} = C - \mathcal{A}^* \mathcal{B}^\dagger f$*. Then exactly one of the following two systems is consistent:*

1. $\mathcal{A}(D) + \mathcal{B}^* v = 0,\ \left\langle \bar{C}, D\right\rangle = 0,\ 0 \neq D \succeq_{K^*} 0.$
2. $\mathcal{A}^* y \prec_K C$*,* $\mathcal{B}y = f$*.*

Proof. Let $\bar{y} = \mathcal{B}^\dagger f$ be the particular solution (of minimum norm) of $\mathcal{B}y = f$. Since $\mathcal{B}$ is onto, we conclude that $\mathcal{B}y = f$ if, and only if, $y = \bar{y} + \mathcal{C}^* v$, for some v, where the range of the linear transformation $\mathcal{C}^*$ is equal to the nullspace of $\mathcal{B}$. We can now substitute for y and obtain the equivalent constraint $\mathcal{A}^*(\bar{y} + \mathcal{C}^* v) \preceq_K C$; equivalently we get $\mathcal{A}^* \mathcal{C}^* v \preceq_K C - \mathcal{A}^* \bar{y}$. Therefore, Item 2 holds at $y = \hat{y} = \bar{y} + \mathcal{C}^* \hat{v}$, for some $\hat{v}$, if, and only if, $\mathcal{A}^* \mathcal{C}^* \hat{v} \prec_K C - \mathcal{A}^* \bar{y}$. The result now follows immediately from Lemma 3.4 by equating the linear transformation $\mathcal{A}^* \mathcal{C}^*$ with $\mathcal{A}^*$ and the right-hand side $C - \mathcal{A}^* \bar{y}$ with C. Then the system in Item 1 in Lemma 3.4 becomes $\mathcal{C}(\mathcal{A}(D)) = 0, \langle (C - \mathcal{A}^* \bar{y}), D\rangle = 0$. The result follows since the nullspace of $\mathcal{C}$ is equal to the range of $\mathcal{B}^*$. □

We can also extend Lemma 3.4 to the important case where $\operatorname{int} K = \emptyset$. This occurs at each iteration of the facial reduction.

Corollary 3.7. *Suppose that* $\operatorname{int} K = \emptyset$*,* $\mathcal{F}_P \neq \emptyset$*, and* $C \in \operatorname{span}(K)$*. Then the linear manifold*

$$\mathbb{S}_y := \{y \in \mathcal{W} : C - \mathcal{A}^* y \in \operatorname{span}(K)\}$$

is a subspace. Moreover, let $\mathcal{P}$ *be a one-one linear transformation with*

$$\mathcal{R}(\mathcal{P}) = (\mathcal{A}^*)^\dagger \operatorname{span}(K).$$

Then exactly one of the following two systems is consistent:

1. *$\mathcal{P}^*\mathcal{A}(D) = 0$, $\langle C, D\rangle = 0$, $D \in \operatorname{span}(K)$, and $0 \neq D \succeq_{K^*} 0$.*

2. *$C - \mathcal{A}^* y \in \operatorname{relint} K$.*

Proof. Since $C \in \operatorname{span}(K) = K - K$, we get that $0 \in \mathbb{S}_y$, i.e., $\mathbb{S}_y$ is a subspace.

Let $\mathcal{T}$ denote an onto linear transformation acting on $\mathcal{V}$ such that the nullspace $\mathcal{N}(\mathcal{T}) = \operatorname{span}(K)^\perp$, and $\mathcal{T}^*$ is a partial isometry, i.e., $\mathcal{T}^* = \mathcal{T}^\dagger$. Therefore, $\mathcal{T}$ is one-to-one and is onto $\operatorname{span}(K)$. Then

$$\begin{array}{llll} \mathcal{A}^* y \preceq_K C & \iff & \mathcal{A}^* y \preceq_K C \text{ and } \mathcal{A}^* y \in \operatorname{span}(K), & \text{since } C \in K - K \\ & \iff & (\mathcal{A}^*\mathcal{P})w \preceq_K C,\ y = \mathcal{P}w, \text{ for some } w, & \text{by definition of } \mathcal{P} \\ & \iff & (\mathcal{T}\mathcal{A}^*\mathcal{P})w \preceq_{\mathcal{T}(K)} \mathcal{T}(C),\ y = \mathcal{T}w, \text{ for some } w, & \text{by definition of } \mathcal{T}, \end{array}$$

i.e., (1.1) is equivalent to

$$v_P := \sup\{\langle \mathcal{P}^* b, w\rangle : (\mathcal{T}\mathcal{A}^*\mathcal{P})w \preceq_{\mathcal{T}(K)} \mathcal{T}(C)\}.$$

The corresponding dual is

$$v_D := \inf\left\{\langle \mathcal{T}(C), D\rangle : \mathcal{P}^*\mathcal{A}\mathcal{T}^*(D) = \mathcal{P}^* b,\ D \succeq_{(\mathcal{T}(K))^*} 0\right\}.$$

By construction, $\operatorname{int} \mathcal{T}(K) \neq \emptyset$, so we may apply Lemma 3.4. We conclude that exactly one of the following two systems is consistent:

1. $\mathcal{P}^*\mathcal{A}\mathcal{T}^*(D) = 0$, $0 \neq D \succeq_{(\mathcal{T}(K))^*} 0$, and $\langle \mathcal{T}(C), D\rangle = 0$.

2. $(\mathcal{T}\mathcal{A}^*\mathcal{P})w \prec_{\mathcal{T}(K)} \mathcal{T}(D)$ (Slater CQ).

The required result follows, since we can now identify $\mathcal{T}^*(D)$ with $D \in \operatorname{span}(K)$, and $\mathcal{T}(C)$ with C. □

Remark 3.8. *Ideally, we would like to find $\hat{D} \in \operatorname{relint}\left(\mathcal{F}_P^Z\right)^c = \operatorname{relint}\left((C + \mathcal{R}(\mathcal{A}^*)) \cap K\right)^c$, since then we have found the minimal face $f_P = \{\hat{D}\}^\perp \cap K$. This is difficult to do numerically. Instead, Lemma 3.4 compromises and finds a point in a larger set $D \in \left(\mathcal{N}(\mathcal{A}) \cap \{C\}^\perp \cap K^*\right) \backslash \{0\}$. This allows for the reduction of $K \leftarrow K \cap \{D\}^\perp$. Repeating to find another D is difficult without the subspace reduction using $\mathcal{P}$ in Corollary 3.7. This emphasizes the importance of the minimal subspace form reduction as an aid to the minimal cone reduction, [65].*

A similar argument applies to the regularization of the dual as given in Corollary 3.5. Let $\mathcal{F}_D = (\hat{X} + \mathcal{N}(\mathcal{A})) \cap K^*$, where $\mathcal{A}(\hat{X}) = b$. We note that a compromise to finding $\hat{Z} \in \operatorname{relint}(\mathcal{F}_P^z)^c = \operatorname{relint}((\hat{X} + \mathcal{N}(\mathcal{A})) \cap K^*)^c$, $f_D = \{\hat{Z}\}^\perp \cap K^*$ is finding $Z \in (\mathcal{R}(\mathcal{A}^*) \cap \{\hat{X}\}^\perp \cap K) \backslash \{0\}$, where $0 = \langle Z, \hat{X}\rangle = \langle \mathcal{A}^* v, \hat{X}\rangle = \langle v, b\rangle$.

3.3 Stable auxiliary subproblem

From this section on we restrict the application of facial reduction to the SDP in (1.1). (Note that the notion of auxiliary problem as well as Theorems 3.11 and 3.15, below, apply to the more general conic convex program (1.4).) Each iteration of the facial reduction algorithm involves two steps. First, we apply Lemma 3.4 and find a point D in the relative interior of the recession cone $\mathcal{R}_{\mathrm{D}}$. Then, we project onto the span of the conjugate face $\{D\}^\perp \cap \mathbb{S}_+^n \supseteq f_P$. This yields a smaller dimensional equivalent problem. The first step to find D is well-suited for interior-point algorithms if we can formulate a suitable conic optimization problem. We now formulate and present the properties of a stable auxiliary problem for finding D. The following is well-known, e.g., [41, Theorems 10.4.1,10.4.7].

Theorem 3.9. *If the (generalized) Slater CQ holds for both primal problem* (1.1) *and dual problem* (1.2)*, then as the barrier parameter* $\mu \to 0^+$*, the primal-dual central path converges to a point* $(\hat{X}, \hat{y}, \hat{Z})$*, where* $\hat{Z} = C - \mathcal{A}^* \hat{y}$*, such that* $\hat{X}$ *is in the relative interior of the set of optimal solutions of* (1.2) *and* $(\hat{y}, \hat{Z})$ *is in the relative interior of the set of optimal solutions of* (1.1). ■

Remark 3.10. *Many polynomial time algorithms for SDP assume that the Newton search directions can be calculated accurately. However, difficulties can arise in calculating accurate search directions if the corresponding Jacobians become increasingly ill-conditioned. This is the case in most of the current implementations of interior point methods due to symmetrization and block elimination steps, see e.g., [19]. In addition, the ill-conditioning arises if the Jacobian of the optimality conditions is not full rank at the optimal solution, as is the case if strict complementarity fails for the SDP. This key question is discussed further in Section 3.5, below.*

According to Theorem 3.9, if we can formulate a pair of auxiliary primal-dual cone optimization problems, each with generalized Slater points such that the relative interior of $\mathcal{R}_{\mathrm{D}}$ coincides with the relative interior of the optimal solution set of one of our auxiliary problems, then we can design an interior-point algorithm for the auxiliary primal-dual pair, making sure that the iterates of our algorithm stay close to the central path (as they approach the optimal solution set) and generate our desired $X \in \operatorname{relint} \mathcal{R}_{\mathrm{D}}$.

This is precisely what we accomplish next. In the special case of $K = \mathbb{S}^n_+$, this corresponds to finding maximum rank feasible solutions for the underlying auxiliary SDPs, since the relative interiors of the faces are characterized by their maximal rank elements.

Define the linear transformation $\mathcal{A}_C : \mathbb{S}^n \to \mathbb{R}^{m+1}$ by

$$\mathcal{A}_C(D) = \begin{pmatrix} \mathcal{A}(D) \\ \langle C, D \rangle \end{pmatrix},$$

This presents a homogenized form of the constraint of (1.1) and combines the two constraints in Lemma 3.4, Item 1. Now consider the following conic optimization problem, which we shall henceforth refer to as the *auxiliary problem*.

$$(AP) \qquad \begin{array}{rcc} val_P^{aux} := & \min\limits_{\delta, D} & \delta \\ & \text{s.t.} & \|\mathcal{A}_C(D)\| \le \delta \\ & & \langle \frac{1}{\sqrt{n}} I, D \rangle = 1 \\ & & D \succeq 0. \end{array} \qquad (3.5)$$

This auxiliary problem is related to the study of the distances to infeasibility in e.g., [46]. The

Lagrangian dual of (3.5) is

$$\begin{aligned}
&\sup_{W\succeq 0,\begin{pmatrix}\beta\\ u\end{pmatrix}\succeq_{\mathcal{Q}}0} \inf_{\delta,D}\ \delta+\gamma\left(1-\left\langle D,\frac{1}{\sqrt{n}}I\right\rangle\right)-\langle W,D\rangle-\left\langle\begin{pmatrix}\beta\\ u\end{pmatrix},\begin{pmatrix}\delta\\ \mathcal{A}_C(D)\end{pmatrix}\right\rangle\\
&\qquad=\sup_{W\succeq 0,\begin{pmatrix}\beta\\ u\end{pmatrix}\succeq_{\mathcal{Q}}0} \inf_{\delta,D}\ \delta(1-\beta)-\left\langle D,\,\mathcal{A}_C^*u+\gamma\frac{1}{\sqrt{n}}I+W\right\rangle+\gamma,
\end{aligned}\tag{3.6}$$

where $\mathcal{Q}:=\left\{\begin{pmatrix}\beta\\ u\end{pmatrix}\in\mathbb{R}^{m+2}:\|u\|\leq\beta\right\}$ refers to the second order cone. Since the inner infimum of (3.6) is unconstrained, we get the following equivalent dual.

$$(DAP)\qquad \begin{array}{rcc} val_D^{aux}:= & \sup\limits_{\gamma,u,W} & \gamma\\ & \text{s.t.} & \mathcal{A}_C^*u+\gamma\frac{1}{\sqrt{n}}I+W=0\\ & & \|u\|\leq 1\\ & & W\succeq 0.\end{array}\tag{3.7}$$

A strictly feasible primal-dual point for (3.5) and (3.7) is given by

$$D=\frac{1}{\sqrt{n}}I,\ \delta>\left\|\mathcal{A}_C\left(\frac{1}{\sqrt{n}}I\right)\right\|,\quad\text{and}\quad \gamma=-1,\ u=0,\ W=\frac{1}{\sqrt{n}}I,\tag{3.8}$$

showing that the generalized Slater CQ holds for the pair (3.5)–(3.7).

Observe that the complexity of solving (3.5) is essentially that of solving the original dual (1.2). Recalling that if a path-following interior point method is applied to solve (3.5), one arrives at a point in the relative interior of the set of optimal solutions, a primal optimal solution (δ^*,D^*) obtained is such that D^* is of maximum rank.

3.3.1 Auxiliary problem information for minimal face of $\mathcal{F}_P^Z$

This section outlines some useful information that the auxiliary problem provides. Theoretically, in the case when the Slater CQ (nearly) fails for (1.1), the auxiliary problem provides a more refined description of the feasible region, as Theorem 3.11 shows. Computationally, the auxiliary problem gives a measure of how close the feasible region of (1.1) is to being a subset of a face of the cone of positive semidefinite matrices, as shown by: (i) the cosine-angle upper bound (near orthogonality) of the feasible set with the conjugate face given in Theorem 3.15; (ii) the cosine-angle lower bound (closeness) of the feasible set with a proper face of $\mathbb{S}_+^n$ in Proposition 3.16; and (iii) the near common block singularity bound for all the feasible slacks obtained after an appropriate orthogonal rotation, in Corollary 3.17.

We first illustrate the stability of the auxiliary problem and show how a primal-dual solution can be used to obtain useful information about the original pair of conic problems.

Theorem 3.11. *The primal-dual pair of problems* (3.5) *and* (3.7) *satisfy the generalized Slater CQ, both have optimal solutions, and their (nonnegative) optimal values are equal. Moreover, letting* (δ^*,D^*) *be an optimal solution of* (3.5)*, the following holds under the assumption that* $\mathcal{F}_P\neq\emptyset$*:*

1. *If* $\delta^* = 0$ *and* $D^* \succ 0$*, then the Slater CQ fails for* (1.1) *but the generalized Slater CQ holds for* (1.2)*. In fact, the primal minimal face and the only primal feasible (hence optimal) solution are*

$$f_P = \{0\}, \quad y^* = (\mathcal{A}^*)^\dagger(C).$$

2. *If* $\delta^* = 0$ *and* $D^* \not\succ 0$*, then the Slater CQ fails for* (1.1) *and the minimal face satisfies*

$$f_P \trianglelefteq \mathbb{S}^n_+ \cap \{D^*\}^\perp \triangleleft \mathbb{S}^n_+. \tag{3.9}$$

3. *If* $\delta^* > 0$*, then the Slater CQ holds for* (1.1)*.*

Proof. A strictly feasible pair for (3.5)–(3.7) is given in (3.8). Hence by strong duality both problems have equal optimal values and both values are attained.

1. Suppose that $\delta^* = 0$ and $D^* \succ 0$. It follows that $\mathcal{A}_C(D^*) = 0$ and $D^* \neq 0$. It follows from Lemma 3.3 that

$$f_P \trianglelefteq \mathbb{S}^n_+ \cap \{D^*\}^\perp = \{0\}.$$

   Hence all feasible points for (1.1) satisfy $C - \mathcal{A}^* y = 0$. Since $\mathcal{A}$ is onto, we conclude that the unique solution of this linear system is $y = (\mathcal{A}^*)^\dagger(C)$.

   Since $\mathcal{A}$ is onto, there exists $\bar{X}$ such that $\mathcal{A}(\bar{X}) = b$. Thus, for every $t \geq 0$, $\mathcal{A}(\bar{X} + tD^*) = b$, and for t large enough, $\bar{X} + tD^* \succ 0$. Therefore, the generalized Slater CQ holds for (1.2).

2. The result follows from Lemma 3.3.

3. If $\delta^* > 0$, then $\mathcal{R}_{\mathrm{D}} = \{0\}$, where $\mathcal{R}_{\mathrm{D}}$ was defined in (3.3). It follows from Lemma 3.4 that the Slater CQ holds for (1.1).

□

Remark 3.12. *Theorem 3.11 shows that if the primal problem* (1.1) *is feasible, then by definition of (AP) as in* (3.5), $\delta^* = 0$ *if, and only if,* $\mathcal{A}_C$ *has a right singular vector* D *such that* $D \succeq 0$ *and the corresponding singular value is zero, i.e., we could replace (AP) with* $\min\{\|\mathcal{A}_C(D)\| : \|D\| = 1, D \succeq 0\}$*. Therefore, we could solve (AP) using a basis for the nullspace of* $\mathcal{A}_C$*, e.g., using an onto linear function* $\mathcal{N}_{\mathcal{A}_C}$ *on* $\mathbb{S}^n$ *that satisfies* $\mathcal{R}(\mathcal{N}^*_{\mathcal{A}_C}) = \mathcal{N}(\mathcal{A}_C)$*, and an approach based on maximizing the smallest eigenvalue:*

$$\delta \approx \sup_y \left\{\lambda_{\min}(\mathcal{N}^*_{\mathcal{A}_C} y) : \operatorname{trace}(\mathcal{N}^*_{\mathcal{A}_C} y) = 1, \|y\| \leq 1\right\},$$

so, in the case when $\delta^* = 0$*, both (AP) and (DAP) can be seen as a max-min eigenvalue problem (subject to a bound and a linear constraint).*

Finding $0 \neq D \succeq 0$ *that solves* $\mathcal{A}_C(D) = 0$ *is also equivalent to the SDP*

$$\begin{array}{ll} \inf\limits_D & \|D\| \\ s.t. & \mathcal{A}_C(D) = 0, \ \langle I, D\rangle = \sqrt{n}, \ D \succeq 0, \end{array} \tag{3.10}$$

a program for which the Slater CQ generally fails. (See Item 2 of Theorem 3.11.) This suggests that the problem of finding the recession direction $0 \neq D \succeq 0$ *that certifies a failure for* (1.1) *to satisfy the Slater CQ may be a difficult problem.*

One may detect whether the Slater CQ fails for the dual (1.2) using the auxiliary problem (3.5) and its dual (3.7).

Proposition 3.13. *Assume that* (1.2) *is feasible, i.e., there exists* $\hat{X} \in \mathbb{S}^n_+$ *such that* $\mathcal{A}(\hat{X}) = b$. *Then we have that* X *is feasible for* (1.2) *if and only if*

$$X = \hat{X} + \mathcal{N}_{\mathcal{A}}^* y \succeq 0,$$

where $\mathcal{N}_{\mathcal{A}} : \mathbb{S}^n \to \mathbb{R}^{n(n+1)/2-m}$ *is an onto linear transformation such that* $\mathcal{R}(\mathcal{N}_{\mathcal{A}}^*) = \mathcal{N}(\mathcal{A})$. *Then the corresponding auxiliary problem*

$$\inf_{\delta,D} \ \delta \quad s.t. \quad \left\| \begin{pmatrix} \mathcal{N}_{\mathcal{A}}(D) \\ \langle \hat{X}, D \rangle \end{pmatrix} \right\| \leq \delta, \ \langle I, D \rangle = \sqrt{n}, \ D \succeq 0$$

either certifies that (1.2) *satisfies the Slater CQ, or that* 0 *is the only feasible slack of* (1.2), *or detects a smaller face of* $\mathbb{S}^n_+$ *containing* $\mathcal{F}_D$.

The results in Proposition 3.13 follows directly from the corresponding results for the primal problem (1.1). An alternative form of the auxiliary problem for (1.2) can be defined using the theorem of the alternative in Corollary 3.5.

Proposition 3.14. *Assume that* (1.2) *is feasible. The dual auxiliary problem*

$$\sup_{v,\lambda} \ \lambda \quad s.t. \quad (\mathcal{A}(I))^T v = 1, \ b^T v = 0, \ \mathcal{A}^* v \succeq \lambda I \tag{3.11}$$

determines if (1.2) *satisfies the Slater CQ. The dual of* (3.11) *is given by*

$$\inf_{\mu,\Omega} \ \mu_2 \quad s.t. \quad \langle I, \Omega \rangle = 1, \ \mathcal{A}(\Omega) - \mu_1 \mathcal{A}(I) - \mu_2 b = 0, \ \Omega \succeq 0, \tag{3.12}$$

and the following hold under the assumption that (1.2) *is feasible:*

(1) If (3.11) *is infeasible, then* (1.2) *must satisfy the Slater CQ.*

(2) If (3.11) *is feasible, then both* (3.11) *and* (3.12) *satisfy the Slater CQ. Moreover, the Slater CQ holds for* (1.2) *if and only if the optimal value of* (3.11) *is negative.*

(3) If (v^*, λ^*) *is an optimal solution of* (3.11) *with* $\lambda^* \geq 0$, *then* $\mathcal{F}_D \subseteq \mathbb{S}^n_+ \cap \{\mathcal{A}^* v^*\}^\perp \lhd \mathbb{S}^n_+$.

Since X *feasible for* (1.2) *implies that*

$$\langle \mathcal{A}^* v^*, X \rangle = (v^*)^T (\mathcal{A}(X)) = (v^*)^T b = 0,$$

we conclude that $\mathcal{F}_D \subseteq \mathbb{S}^n_+ \cap \{\mathcal{A}^* v^*\}^\perp \lhd \mathbb{S}^n_+$. *Therefore, if* (1.2) *fails the Slater CQ, then, by solving* (3.11), *we can obtain a proper face of* $\mathbb{S}^n_+$ *that contains the feasible region* $\mathcal{F}_D$ *of* (1.2).

Proof. The Lagrangian of (3.11) is given by

$$\begin{aligned} L(v, \lambda, \mu, \Omega) &= \lambda + \mu_1 (1 - (\mathcal{A}(I)^T v)) + \mu_2 (-b^T v) + \langle \Omega, \mathcal{A}^* v - \lambda I \rangle \\ &= \lambda (1 - \langle I, \Omega \rangle) + v^T (\mathcal{A}(\Omega) - \mu_1 \mathcal{A}(I) - \mu_2 b) + \mu_2. \end{aligned}$$

This yields the dual program (3.12).

If (3.11) is infeasible, then we must have $b \neq 0$ and $\mathcal{A}(I) = kb$ for some $k \in \mathbb{R}$. If $k > 0$, then $k^{-1}I$ is a Slater point for (1.2). If $k = 0$, then $\mathcal{A}(\hat{X} + \lambda I) = b$ and $\hat{X} + \lambda I \succ 0$ for any $\hat{X}$ satisfying $\mathcal{A}(\hat{X}) = b$ and sufficiently large $\lambda > 0$. If $k < 0$, then $\mathcal{A}(2\hat{X} + k^{-1}I) = b$ for $\hat{X} \succeq 0$ satisfying $\mathcal{A}(\hat{X}) = b$; and we have $2\hat{X} + k^{-1}I \succ 0$.

If (3.11) is feasible, i.e., if there exists $\hat{v}$ such that $(\mathcal{A}(I))^T v = 1$ and $b^T \hat{v} = 0$, then

$$(\hat{v}, \hat{\lambda}) = \left(\hat{v}, \hat{\lambda} = \lambda_{\min}(\mathcal{A}^*\hat{v}) - 1\right), \quad (\hat{\mu}, \hat{\Omega}) = \left(\begin{pmatrix} 1/n \\ 0 \end{pmatrix}, \frac{1}{n} I\right)$$

is strictly feasible for (3.11) and (3.12) respectively.

Let (v^*, λ^*) be an optimal solution of (3.12). If $\lambda^* \leq 0$, then for any $v \in \mathbb{R}^m$ with $\mathcal{A}^* y \succeq 0$ and $b^T v = 0$, v cannot be feasible for (3.11) so $\langle I, \mathcal{A}^* v \rangle \leq 0$. This implies that $\mathcal{A}^* v = 0$. By Corollary 3.5, the Slater CQ holds for (1.2). If $\lambda^* > 0$, then v^* certifies that the Slater CQ fails for (1.2), again by Corollary 3.5. □

The next result shows that δ^* from (AP) is a measure of how close the Slater CQ is to failing.

Theorem 3.15. *Let (δ^*, D^*) denote an optimal solution of the auxiliary problem (3.5). Then δ^* bounds how far the feasible primal slacks $Z = C - \mathcal{A}^* y \succeq 0$ are from orthogonality to D^*:*

$$0 \leq \sup_{0 \preceq Z = C - \mathcal{A}^* y \neq 0} \frac{\langle D^*, Z \rangle}{\|D^*\| \|Z\|} \leq \alpha(\mathcal{A}, C) := \begin{cases} \dfrac{\delta^*}{\sigma_{\min}(\mathcal{A})} & \text{if } C \in \mathcal{R}(\mathcal{A}^*), \\[2ex] \dfrac{\delta^*}{\sigma_{\min}(\mathcal{A}_C)} & \text{if } C \notin \mathcal{R}(\mathcal{A}^*). \end{cases} \tag{3.13}$$

Proof. Since $\langle \frac{1}{\sqrt{n}} I, D^* \rangle = 1$, we get

$$\|D^*\| \geq \frac{\left\langle \frac{1}{\sqrt{n}} I, D^* \right\rangle}{\|\frac{1}{\sqrt{n}} I\|} = \frac{1}{\frac{1}{\sqrt{n}} \|I\|} = 1.$$

If $C = \mathcal{A}^* y_C$ for some $y_C \in \mathbb{R}^m$, then for any $Z = C - \mathcal{A}^* y \succeq 0$,

$$\begin{aligned} \cos\theta_{D^*,Z} := \frac{\langle D^*, C - \mathcal{A}^* y \rangle}{\|D^*\| \|C - \mathcal{A}^* y\|} &\leq \frac{\langle \mathcal{A}(D^*), y_C - y \rangle}{\|\mathcal{A}^*(y_C - y)\|} \\ &\leq \frac{\|\mathcal{A}(D^*)\| \ \|y_C - y\|}{\sigma_{\min}(\mathcal{A}^*) \|y_C - y\|} \\ &\leq \frac{\delta^*}{\sigma_{\min}(\mathcal{A})}. \end{aligned}$$

If $C \notin \mathcal{R}(\mathcal{A}^*)$, then by Assumption 1.1, $\mathcal{A}_C$ is onto so $\langle D^*, C - \mathcal{A}^* y \rangle = \left\langle \mathcal{A}_C(D^*), \begin{pmatrix} -y \\ 1 \end{pmatrix} \right\rangle$ implies that $0 \preceq C - \mathcal{A}^* y \neq 0, \forall y \in \mathcal{F}_P$. Therefore the cosine of the angle $\theta_{D^*,Z}$ between D^* and

$Z = C - \mathcal{A}^* y \succeq 0$ is bounded by

$$\cos\theta_{D^*,Z} = \frac{\langle D^*, C - \mathcal{A}^* y\rangle}{\|D^*\|\|C-\mathcal{A}^*y\|} \leq \frac{\left\langle \mathcal{A}_C(D^*), \begin{pmatrix}-y\\1\end{pmatrix}\right\rangle}{\left\|\mathcal{A}_C^*\begin{pmatrix}-y\\1\end{pmatrix}\right\|}$$

$$\leq \frac{\|\mathcal{A}_C(D^*)\| \left\|\begin{pmatrix}-y\\1\end{pmatrix}\right\|}{\sigma_{\min}(\mathcal{A}_C)\left\|\begin{pmatrix}-y\\1\end{pmatrix}\right\|}$$

$$= \frac{\delta^*}{\sigma_{\min}(\mathcal{A}_C)}.$$

□

Theorem 3.15 provides a lower bound for the angle and distance between feasible slack vectors and the vector D^* on the boundary of $\mathbb{S}^n_+$. For our purposes, the theorem is only useful when $\alpha(\mathcal{A}, C)$ is small. Given that $\delta^* = \|\mathcal{A}_C(D^*)\|$, we see that the lower bound is independent of simple scaling of $\mathcal{A}_C$, though not necessarily independent of the conditioning of $\mathcal{A}_C$. Thus, δ^* provides qualitative information about both the conditioning of $\mathcal{A}_C$ and the distance to infeasibility.

We now strengthen the result in Theorem 3.15 by using more information from D^*. In applications we expect to choose the partitions of U and D^* to satisfy $\lambda_{\min}(D_+) >> \lambda_{\max}(D_\epsilon)$.

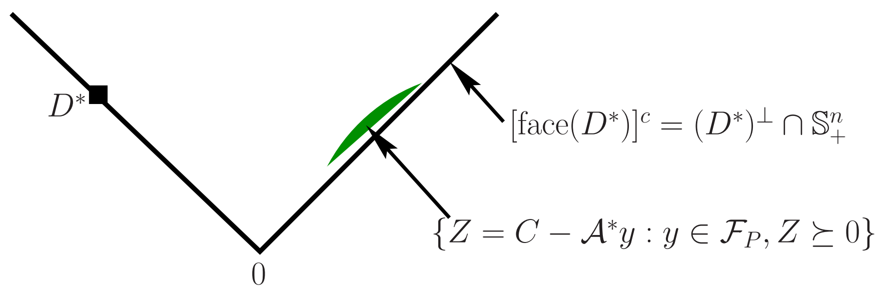

Figure 1: Minimal Face; $0 < \delta^* \ll 1$

Proposition 3.16. *Let (δ^*, D^*) denote an optimal solution of the auxiliary problem (3.5), and let*

$$D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P & Q \end{bmatrix}^T, \tag{3.14}$$

with $U = \begin{bmatrix} P & Q\end{bmatrix}$ orthogonal, and $D_+ \succ 0$.

Let $0 \neq Z := C - \mathcal{A}^* y \succeq 0$ and $Z_Q := QQ^T Z QQ^T$. Then Z_Q is the closest point in $\mathcal{R}(Q \cdot Q^T) \cap \mathbb{S}^n_+$ to Z; and, the cosine of the angle θ_{Z,Z_Q} between Z and the face $\mathcal{R}(Q \cdot Q^T) \cap \mathbb{S}^n_+$ satisfies

$$\cos\theta_{Z,Z_Q} := \frac{\langle Z, Z_Q\rangle}{\|Z\|\|Z_Q\|} = \frac{\|Q^T Z Q\|}{\|Z\|} \geq 1 - \alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}, \tag{3.15}$$

where $\alpha(\mathcal{A},C)$ is defined in (3.13). Thus the angle between any feasible slack and the face $\mathcal{R}(Q\cdot Q^T)\cap\mathbb{S}^n_+$ cannot be too large in the sense that

$$\inf_{0\neq Z=C-\mathcal{A}^*y\succeq 0}\cos\theta_{Z,Z_Q}\geq 1-\alpha(\mathcal{A},C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}.$$

Moreover, the normalized distance to the face is bounded as in

$$\|Z-Z_Q\|^2\leq 2\|Z\|^2\left[\alpha(\mathcal{A},C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}\right]. \tag{3.16}$$

Proof. Since $Z\succeq 0$, we have $Q^TZQ\in\operatorname{argmin}_{W\succeq 0}\|Z-QWQ^T\|$. This shows that $Z_Q:=QQ^TZQQ^T$ is the closest point in $\mathcal{R}(Q\cdot Q^T)\cap\mathbb{S}^n_+$ to Z. The expression for the angle in (3.15) follows using

$$\frac{\langle Z,Z_Q\rangle}{\|Z\|\|Z_Q\|}=\frac{\|Q^TZQ\|^2}{\|Z\|\|Q^TZQ\|}=\frac{\|Q^TZQ\|}{\|Z\|}. \tag{3.17}$$

From Theorem 3.15, we see that $0\neq Z=C-\mathcal{A}^*y\succeq 0$ implies that $\left\langle\frac{1}{\|Z\|}Z,D^*\right\rangle\leq\alpha(\mathcal{A},C)\|D^*\|$. Therefore, the optimal value of the following optimization problem provides a lower bound on the quantity in (3.17).

$$\begin{array}{rcll}\gamma_0 & := & \min\limits_Z & \|Q^TZQ\| \\ & & \text{s.t.} & \langle Z,D^*\rangle\leq\alpha(\mathcal{A},C)\|D^*\| \\ & & & \|Z\|^2=1,\quad Z\succeq 0.\end{array} \tag{3.18}$$

Since $\langle Z,D^*\rangle=\langle P^TZP,D_+\rangle+\langle Q^TZQ,D_\epsilon\rangle\geq\langle P^TZP,D_+\rangle$ whenever $Z\succeq 0$, we have

$$\begin{array}{rcll}\gamma_0\ \geq\ \gamma & := & \min\limits_Z & \|Q^TZQ\| \\ & & \text{s.t.} & \langle P^TZP,D_+\rangle\leq\alpha(\mathcal{A},C)\|D^*\| \\ & & & \|Z\|^2=1,\quad Z\succeq 0.\end{array} \tag{3.19}$$

It is possible to find the optimal value γ of (3.19). After the orthogonal rotation

$$Z=\begin{bmatrix}P & Q\end{bmatrix}\begin{bmatrix}S & V\\ V^T & W\end{bmatrix}\begin{bmatrix}P & Q\end{bmatrix}^T=PSP^T+PVQ^T+QV^TP^T+QWQ^T,$$

where $S\in\mathbb{S}^{n-\bar n}_+$, $W\in\mathbb{S}^{\bar n}_+$ and $V\in\mathbb{R}^{(n-\bar n)\times\bar n}$, (3.19) can be rewritten as

$$\begin{array}{rcll}\gamma & = & \min\limits_{S,V,W} & \|W\| \\ & & \text{s.t.} & \langle S,D_+\rangle\leq\alpha(\mathcal{A},C)\|D^*\| \\ & & & \|S\|^2+2\|V\|^2+\|W\|^2=1 \\ & & & \begin{bmatrix}S & V\\ V^T & W\end{bmatrix}\in\mathbb{S}^n_+.\end{array} \tag{3.20}$$

Since

$$\|V\|^2\leq\|S\|\|W\| \tag{3.21}$$

holds whenever $\begin{bmatrix} S & V \\ V^T & W \end{bmatrix} \succeq 0$, we have that $(\|S\| + \|W\|)^2 \geq \|S\|^2 + 2\|V\|^2 + \|W\|^2$. This yields

$$\begin{array}{rlcc} \gamma \geq & \bar{\gamma} := & \min_{S,V,W} & \|W\| \\ & & \text{s.t.} & \langle S, D_+\rangle \leq \alpha(\mathcal{A}, C)\|D^*\| \\ & & & \|S\| + \|W\| \geq 1 \\ & & & S \succeq 0,\ W \succeq 0. \end{array} \qquad \begin{array}{rcc} \bar{\gamma} \geq & \min\limits_{S} & 1 - \|S\| \\ & \text{s.t.} & \langle S, D_+\rangle \leq \alpha(\mathcal{A}, C)\|D^*\| \\ & & S \succeq 0 \end{array} \tag{3.22}$$

Since $\lambda_{\min}(D_+)\|S\| \leq \langle S, D_+\rangle \leq \alpha(\mathcal{A}, C)\|D^*\|$, we see that the objective value of the last optimization problem in (3.22) is bounded below by $1 - \alpha(\mathcal{A}, C)\|D^*\|/\lambda_{\min}(D_+)$. Now let u be a normalized eigenvector of D_+ corresponding to its smallest eigenvalue $\lambda_{\min}(D_+)$. Then $S^* = \frac{\alpha(\mathcal{A},C)\|D^*\|}{\lambda_{\min}(D_+)} uu^T$ solves the last optimization problem in (3.22), with corresponding optimal value $1 - \frac{\alpha(\mathcal{A},C)\|D^*\|}{\lambda_{\min}(D_+)}$.

Let $\beta := \min\left\{\frac{\alpha(\mathcal{A},C)\|D^*\|}{\lambda_{\min}(D_+)}, 1\right\}$. Then $\gamma \geq 1 - \beta$. Also,

$$\begin{bmatrix} S & V \\ V^T & W \end{bmatrix} := \begin{pmatrix} \sqrt{\beta} u \\ \sqrt{1-\beta} e_1 \end{pmatrix} \begin{pmatrix} \sqrt{\beta} u \\ \sqrt{1-\beta} e_1 \end{pmatrix}^T = \begin{bmatrix} \beta uu^T & \sqrt{\beta(1-\beta)} u e_1^T \\ \sqrt{\beta(1-\beta)} e_1 u^T & (1-\beta)\, e_1 e_1^T \end{bmatrix} \in \mathbb{S}^n_+.$$

Therefore (S, V, W) is feasible for (3.20), and attains an objective value $1 - \beta$. This shows that $\gamma = 1 - \beta$ and proves (3.15).

The last claim (3.16) follows immediately from

$$\begin{aligned} \|Z - Z_Q\|^2 &= \|Z\|^2 \left(1 - \frac{\|Q^T Z Q\|^2}{\|Z\|^2}\right) \\ &\leq \|Z\|^2 \left[1 - \left(1 - \alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}\right)^2\right] \\ &\leq 2\|Z\|^2 \alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}. \end{aligned}$$

□

These results are related to the extreme angles between vectors in a cone studied in [29, 33]. Moreover, it is related to the distances to infeasibility in e.g., [46], in which the distance to infeasibility is shown to provide backward and forward error bounds.

We now see that we can use the rotation $U = \begin{bmatrix} P & Q \end{bmatrix}$ obtained from the diagonalization of the optimal D^* in the auxiliary problem (3.5) to reveal *nearness to infeasibility*, as discussed in e.g., [46]. Or, in our approach, this reveals nearness to a facial decomposition. We use the following results to bound the size of certain blocks of a feasible slack Z.

Corollary 3.17. *Let* (δ^*, D^*) *denote an optimal solution of the auxiliary problem* (3.5)*, as in Theorem 3.15; and let*

$$D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P & Q \end{bmatrix}^T, \tag{3.23}$$

with $U = \begin{bmatrix} P & Q \end{bmatrix}$ *orthogonal, and* $D_+ \succ 0$*. Then for any feasible slack* $0 \neq Z = C - \mathcal{A}^* y \succeq 0$*, we have*

$$\operatorname{trace} P^T Z P \leq \alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}\|Z\|, \tag{3.24}$$

where $\alpha(\mathcal{A}, C)$ *is defined in* (3.13).

Proof. Since

$$
\begin{aligned}
\langle D^*, Z\rangle &= \left\langle \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix}, \begin{bmatrix} P^TZP & P^TZQ \\ Q^TZP & Q^TZQ \end{bmatrix} \right\rangle \\
&= \left\langle D_+ , P^TZP\right\rangle + \left\langle D_\epsilon , Q^TZQ\right\rangle \\
&\geq \left\langle D_+ , P^TZP\right\rangle \\
&\geq \lambda_{\min}(D_+) \operatorname{trace} P^TZP,
\end{aligned}
\tag{3.25}
$$

the claim follows from Theorem 3.15. □

Remark 3.18. *We now summarize the information available from a solution of the auxiliary problem, with optima $\delta^* \geq 0, D^* \not\succ 0$. We let $0 \neq Z = C - \mathcal{A}^*y \succeq 0$ denote a feasible slack. In particular, we emphasize the information obtained from the rotation U^TZU using the orthogonal U that block diagonalizes D^* and from the* closest *point $Z_Q = QQ^TZQQ^T$. We note that replacing all feasible Z with the* projected *Z_Q provides a nearby problem for the backwards stability argument. Alternatively, we can view the nearby problem by projecting the data $A_i \leftarrow QQ^TA_iQQ^T, \forall i$, $C \leftarrow QQ^TCQQ^T$.*

1. *From* (3.13) *in Theorem 3.15, we get a lower bound on the angle (upper bound on the cosine of the angle)*
$$\cos\theta_{D^*,Z} = \frac{\langle D^*, Z\rangle}{\|D^*\|\|Z\|} \leq \alpha(\mathcal{A}, C).$$
2. *In Proposition 3.16 with orthogonal $U = \begin{bmatrix} P & Q\end{bmatrix}$, we get upper bounds on the angle between a feasible slack and the face defined using $Q \cdot Q^T$ and on the normalized distance to the face.*
$$\cos\theta_{Z,Z_Q} := \frac{\langle Z, Z_Q\rangle}{\|Z\|\|Z_Q\|} = \frac{\|Q^TZQ\|}{\|Z\|} \geq 1 - \alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}.$$
$$\|Z - Z_Q\|^2 \leq 2\|Z\|^2 \left[\alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}\right].$$
3. *After the rotation using the orthogonal U, the* $(1, 1)$ *principal block is bounded as*
$$\operatorname{trace} P^TZP \leq \alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}\|Z\|.$$

3.4 Rank-revealing rotation and equivalent problems

We may use the results from Theorem 3.15 and Corollary 3.17 to get two *rotated* optimization problems equivalent to (1.1). The equivalent problems indicate that, in the case when δ^* is sufficiently small, it is possible to reduce the dimension of the problem and get a *nearby* problem that helps in the facial reduction. The two equivalent formulations can be used to illustrate backwards stability with respect to a perturbation of the cone $\mathbb{S}^n_+$.

First we need to find a suitable shift of C to allow a proper facial projection. This is used in Theorem 3.20, below.

Lemma 3.19. *Let $\delta^*, D^*, U = \begin{bmatrix} P & Q\end{bmatrix}, D_+, D_\epsilon$ be defined as in the hypothesis of Corollary 3.17. Let $(y_Q, W_Q) \in \mathbb{R}^m \times \mathbb{S}^{\bar{n}}$ be the best least squares solution to the equation $QWQ^T + \mathcal{A}^*y = C$, that is, (y_Q, W_Q) is the optimal solution of minimum norm to the linear least squares problem*

$$\min_{y,W} \frac{1}{2}\|C - (QWQ^T + \mathcal{A}^*y)\|^2. \tag{3.26}$$

Let $C_Q := QW_QQ^T$ *and* $C_{\rm res} := C - (C_Q + \mathcal{A}^*y_Q)$*. Then*

$$Q^T C_{\rm res} Q \;=\; 0, \quad \textit{and} \quad \mathcal{A}\left(C_{\rm res}\right) \;=\; 0. \tag{3.27}$$

Moreover, if $\delta^* = 0$*, then for any feasible solution* y *of* (1.1)*, we get*

$$C - \mathcal{A}^*y \in \mathcal{R}(Q \cdot Q^T), \tag{3.28}$$

and further $(y, Q^T(C - \mathcal{A}^*y)Q)$ *is an optimal solution of* (3.26)*, whose optimal value is zero.*

Proof. Let $\Omega(y, W) := \frac{1}{2}\|C - (QWQ^T + \mathcal{A}^*y)\|^2$. Since

$$\Omega(y,W) = \frac{1}{2}\|C\|^2 + \frac{1}{2}\|\mathcal{A}^*y\|^2 + \frac{1}{2}\|W\|^2 + \left\langle QWQ^T, \mathcal{A}^*y\right\rangle - \left\langle Q^T C\, Q, W\right\rangle - \left\langle \mathcal{A}(C), y\right\rangle,$$

we have (y_Q, W_Q) solves (3.26) if, and only if,

$$\begin{aligned} \nabla_y \Omega &= \mathcal{A}\left(QWQ^T - (C - \mathcal{A}^*y)\right) = 0, && (3.29)\\ \text{and} \quad \nabla_w \Omega &= W - \left[Q^T\left(C - \mathcal{A}^*y\right)Q\right] = 0. && (3.30) \end{aligned}$$

Then (3.27) follows immediately by substitution.

If $\delta^* = 0$, then $\langle D^*, A_i\rangle = 0$ for $i = 1, \ldots, m$ and $\langle D^*, C\rangle = 0$. Hence, for any $y \in \mathbb{R}^m$,

$$\langle D_+\,, P^T(C - \mathcal{A}^*y)P\rangle + \langle D_\epsilon\,, Q^T(C - \mathcal{A}^*y)Q\rangle = \langle D^*, C - \mathcal{A}^*y\rangle = 0.$$

If $C - \mathcal{A}^*y \succeq 0$, then we must have $P^T(C - \mathcal{A}^*y)P = 0$ (as $D_+ \succ 0$), and so $P^T(C - \mathcal{A}^*y)Q = 0$. Hence

$$\begin{aligned} C - \mathcal{A}^*y &= UU^T(C - \mathcal{A}^*y)UU^T \\ &= U\begin{bmatrix} P & Q\end{bmatrix}^T(C - \mathcal{A}^*y)\begin{bmatrix} P & Q\end{bmatrix}U^T\;, \\ &= QQ^T(C - \mathcal{A}^*y)QQ^T \end{aligned}$$

i.e., we conclude (3.28) holds.

The last statement now follows from substituting $W = Q^T(C - \mathcal{A}^*y)Q$ in (3.26). □

We can now use the rotation from Corollary 3.17 with a shift of C (to $C_{\rm res} + C_Q = C - \mathcal{A}^*y_Q$) to get two equivalent problems to (P). This emphasizes that when δ^* is *small*, then the auxiliary problem reveals a block structure with one principal block and three *small/negligible* blocks. If δ is small, then β in the following Theorem 3.20 is *small*. Then fixing $\beta = 0$ results in a nearby problem to (P) that illustrates backward stability of the facial reduction.

Theorem 3.20. *Let* $\delta^*, D^*, U = \begin{bmatrix} P & Q\end{bmatrix}, D_+, D_\epsilon$ *be defined as in the hypothesis of Corollary 3.17, and let* $y_Q, W_Q, C_Q, C_{\rm res}$ *be defined as in Lemma 3.19. Define the scalar*

$$\beta := \alpha(\mathcal{A}, C)\frac{\|D^*\|}{\lambda_{\min}(D_+)}, \tag{3.31}$$

and the convex cone $T_\beta \subseteq \mathbb{S}^n_+$ *partitioned appropriately as in* (3.23)*,*

$$T_\beta := \left\{ Z = \begin{bmatrix} A & B \\ B^T & C\end{bmatrix} \in \mathbb{S}^n_+ :\; \operatorname{trace} A \le \beta \operatorname{trace} Z \right\}. \tag{3.32}$$

Then we get the following two equivalent programs to (P) in (1.1)*:*

1. *using the rotation U and the cone T_β,*

$$v_P \quad = \quad \sup_y \left\{ b^T y : \begin{bmatrix} P^T Z P & P^T Z Q \\ Q^T Z P & Q^T Z Q \end{bmatrix} \succeq_{T_\beta} 0, Z = C - \mathcal{A}^* y \right\}; \tag{3.33}$$

2. *using (y_Q, W_Q),*

$$v_P \quad = \quad b^T y_Q + \sup_y \left\{ b^T y : \begin{bmatrix} P^T Z P & P^T Z Q \\ Q^T Z P & Q^T Z Q \end{bmatrix} \succeq_{T_\beta} 0, Z = C_{\text{res}} + C_Q - \mathcal{A}^* y \right\}. \tag{3.34}$$

Proof. From Corollary 3.17,

$$\mathcal{F}_P = \left\{ y : \begin{bmatrix} P^T Z P & P^T Z Q \\ Q^T Z P & Q^T Z Q \end{bmatrix} \succeq_{T_\beta} 0, Z = C - \mathcal{A}^* y \right\}. \tag{3.35}$$

hence the equivalence of (1.1) with (3.33) follows.

For (3.34), first note that for any $y \in \mathbb{R}^m$,

$$Z := C_{\text{res}} + C_Q - \mathcal{A}^* y = C - \mathcal{A}^*(y + y_Q),$$

so $Z \succeq 0$ if and only if $y + y_Q \in \mathcal{F}_P$, if and only if $Z \in T_\beta$. Hence

$$\mathcal{F}_P = y_Q + \left\{ y : \begin{bmatrix} P^T Z P & P^T Z Q \\ Q^T Z P & Q^T Z Q \end{bmatrix} \succeq_{T_\beta} 0, Z = C_{\text{res}} + Q W_Q Q^T - \mathcal{A}^* y \right\}, \tag{3.36}$$

and (3.34) follows. □

Remark 3.21. *As mentioned above, Theorem 3.20 illustrates the backwards stability of the facial reduction. It is difficult to state this precisely due to the shifts done and the changes to the constraints in the algorithm. For simplicity, we just discuss one iteration. The original problem (P) is equivalent to the problem in* (3.33)*. Therefore, a facial reduction step can be applied to the original problem or equivalently to* (3.33)*. We then perturb this problem in* (3.33) *by setting $\beta = 0$. The algorithm applied to this nearby problem with exact arithmetic will result in the same step.*

3.4.1 Reduction to two smaller problems

Following the results from Theorems 3.11 and 3.20, we focus on the case where $\delta^* = 0$ and $\mathcal{R}_{\text{D}} \cap \mathbb{S}^n_{++} = \emptyset$. In this case we get a proper face $Q\mathbb{S}^{\bar{n}}_+ Q^T \lhd \mathbb{S}^n_+$. We obtain two different equivalent formulations of the problem by restricting to this smaller face. In the first case, we stay in the same dimension for the domain variable y but decrease the constraint space and include equality constraints. In the second case, we eliminate the equality constraints and move to a smaller dimensional space for y. We first see that when we have found the minimal face, then we obtain an equivalent regularized problem as was done for LP in Section 2.1.

Corollary 3.22. *Suppose that the minimal face f_P of (P) is found using the orthogonal $U = \begin{bmatrix} P_{\text{fin}} & Q_{\text{fin}} \end{bmatrix}$, so that $f_P = Q_{\text{fin}} \mathbb{S}^r_+ Q_{\text{fin}}^T$, $0 < r < n$. Then an equivalent problem to (P) is*

$$\begin{array}{llcll} & v_P & = & \sup & b^T y \\ (P_{PQ,reg}) & & & s.t. & Q_{\text{fin}}^T (\mathcal{A}^* y) Q_{\text{fin}} \;\preceq\; Q_{\text{fin}}^T C Q_{\text{fin}} \\ & & & & \mathcal{A}_{\text{fin}}^* y \;=\; \mathcal{A}_{\text{fin}}^* y_{Q_{\text{fin}}}, \end{array} \tag{3.37}$$

where $(y_{Q_{\rm fin}}, W_{Q_{\rm fin}})$ *solves the least squares problem* $\min_{y,W} \|C - (\mathcal{A}^* y + Q_{\rm fin} W Q_{\rm fin}^T)\|$*, and* $\mathcal{A}^*_{\rm fin} : \mathbb{R}^m \to \mathbb{R}^t$ *is a full rank (onto) representation of the linear transformation*

$$y \mapsto \begin{bmatrix} P_{\rm fin}^T(\mathcal{A}^* y) P_{\rm fin} \\ Q_{\rm fin}^T(\mathcal{A}^* y) P_{\rm fin} \end{bmatrix}.$$

Moreover, $(P_{PQ,reg})$ *is regularized i.e., the RCQ holds.*

Proof. The result follows immediately from Theorem 3.20, since the definition of the minimal face implies that there exists a feasible $\hat{y}$ which satisfies the constraints in (3.37). The new equality constraint is constructed to be full rank and not change the feasible set. □

Alternatively, we now reduce (1.1) to an equivalent problem over a spectrahedron in a lower dimension using the spectral decomposition of D^*.

Proposition 3.23. *Let the notation and hypotheses in Theorem 3.20 hold with* $\delta^* = 0$ *and* $D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$*, where* $\begin{bmatrix} P & Q \end{bmatrix}$ *is orthogonal,* $Q \in \mathbb{R}^{n \times \bar{n}}$ *and* $D_+ \succ 0$*. Then*

$$v_P = \sup \left\{ b^T y : \begin{array}{l} Q^T(C - \mathcal{A}^* y) Q \succeq 0, \\ P^T(\mathcal{A}^* y) P = P^T(\mathcal{A}^* y_Q) P, \\ Q^T(\mathcal{A}^* y) P = Q^T(\mathcal{A}^* y_Q) P \end{array} \right\}. \tag{3.38}$$

Moreover:

1. *If* $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \{0\}$*, then for any* $y_1, y_2 \in \mathcal{F}_P$*,* $b^T y_1 = b^T y_2 = v_P$*.*
2. *If* $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) \neq \{0\}$*, and if, for some* $\bar{m} > 0$*,* $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ *is an injective linear map such that* $\mathcal{R}(\mathcal{A}^* \mathcal{P}) = \mathcal{R}(\mathcal{A}^*) \cap \mathcal{R}(Q \cdot Q^T)$*, then we have*

$$v_P = b^T y_Q \; + \; \sup_v \left\{ (\mathcal{P}^* b)^T v : W_Q - Q^T(\mathcal{A}^* \mathcal{P} v) Q \succeq 0 \right\}. \tag{3.39}$$

And, if v^* *is an optimal solution of* (3.39)*, then* $y^* = y_Q + \mathcal{P} v^*$ *is an optimal solution of* (1.1)*.*

Proof. Since $\delta^* = 0$, from Lemma 3.19 we have that $C = C_Q + \mathcal{A}^* y_Q, C_Q = Q W_Q Q^T$, for some $y_Q \in \mathbb{R}^m$ and $W_Q \in \mathbb{S}^{\bar{n}}$. Hence by (3.35),

$$\begin{aligned} \mathcal{F}_P &= \left\{ y \in \mathbb{R}^m : Q^T(C - \mathcal{A}^* y) Q \succeq 0, P^T(C - \mathcal{A}^* y) P = 0, Q^T(C - \mathcal{A}^* y) P = 0 \right\} \\ &= \left\{ y \in \mathbb{R}^m : Q^T(C - \mathcal{A}^* y) Q \succeq 0, P^T(\mathcal{A}^*(y - y_Q)) P = 0, Q^T(\mathcal{A}^*(y - y_Q)) P = 0 \right\}, \end{aligned} \tag{3.40}$$

and (3.38) follows.

1. Since $C - \mathcal{A}^* y \in \mathcal{R}(Q \cdot Q^T), \forall y \in \mathcal{F}_P$, we get $\mathcal{A}^*(y_2 - y_1) = (C - \mathcal{A}^* y_1) - (C - \mathcal{A}^* y_2) \in \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \{0\}$. Given that $\mathcal{A}$ is onto, we get $b = \mathcal{A}(\hat{X})$, for some $\hat{X} \in \mathbb{S}^n$, and

$$b^T(y_2 - y_1) = \left\langle \hat{X}, \mathcal{A}^*(y_2 - y_1) \right\rangle = 0.$$

2. From (3.40),

$$\begin{aligned}\mathcal{F}_P &= y_Q + \left\{y : W_Q - Q^T(\mathcal{A}^*y)Q \succeq 0, P^T(\mathcal{A}^*y)P = 0, Q^T(\mathcal{A}^*y)P = 0\right\}\\ &= y_Q + \left\{y : W_Q - Q^T(\mathcal{A}^*y)Q \succeq 0, \mathcal{A}^*y \in \mathcal{R}(Q\cdot Q^T)\right\}\\ &= y_Q + \left\{\mathcal{P}v : W_Q - Q^T(\mathcal{A}^*\mathcal{P}v)Q \succeq 0\right\},\end{aligned}$$

the last equality follows from the choice of $\mathcal{P}$. Therefore, (3.39) follows, and if v^* is an optimal solution of (3.39), then $y_Q + \mathcal{P}v^*$ is an optimal solution of (1.1).

□

Next we establish the existence of the operator $\mathcal{P}$ mentioned in Proposition 3.23.

Proposition 3.24. *For any $n\times n$ orthogonal matrix $U = \begin{bmatrix} P & Q\end{bmatrix}$ and any surjective linear operator $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ with $\bar m := \dim(\mathcal{R}(\mathcal{A}^*)\cap\mathcal{R}(Q\cdot Q^T)) > 0$, there exists a one-one linear transformation $\mathcal{P} : \mathbb{R}^{\bar m} \to \mathbb{R}^m$ that satisfies*

$$\mathcal{R}(\mathcal{A}^*\mathcal{P}) = \mathcal{R}(Q\cdot Q^T)\cap\mathcal{R}(\mathcal{A}^*), \tag{3.41}$$

$$\mathcal{R}(\mathcal{P}) = \mathcal{N}\left(P^T(\mathcal{A}^*\cdot)P\right)\cap\mathcal{N}\left(P^T(\mathcal{A}^*\cdot)Q\right). \tag{3.42}$$

Moreover, $\bar{\mathcal{A}} : \mathbb{S}^{\bar n} \to \mathbb{R}^{\bar m}$ is defined by

$$\bar{\mathcal{A}}^*(\cdot) := Q^T\big(\mathcal{A}^*\mathcal{P}(\cdot)\big)Q$$

is onto.

Proof. Recall that for any matrix $X \in \mathbb{S}^n$,

$$X = UU^TXUU^T = PP^TXPP^T + PP^TXQQ^T + QQ^TXPP^T + QQ^TXQQ^T.$$

Moreover, $P^TQ = 0$. Therefore, $X \in \mathcal{R}(Q\cdot Q^T)$ implies $P^TXP = 0$ and $P^TXQ = 0$. Conversely, $P^TXP = 0$ and $P^TXQ = 0$ implies $X = QQ^TXQQ^T$. Therefore $X \in \mathcal{R}(Q\cdot Q^T)$ if, and only if, $P^TXP = 0$ and $P^TXQ = 0$.

For any $y \in \mathbb{R}^m$, $\mathcal{A}^*y \in \mathcal{R}(Q\cdot Q^T)$ if, and only if,

$$\sum_{i=1}^m (P^TA_iP)y_i = 0 \quad\text{and}\quad \sum_{i=1}^m (P^TA_iQ)y_i = 0,$$

which holds if, and only if, $y \in \operatorname{span}\{\beta\}$, where $\beta := \{y_1,\dots,y_{\bar m}\}$ is a basis of the linear subspace

$$\left\{y : \sum_{i=1}^m (P^TA_iP)y_i = 0\right\} \cap \left\{y : \sum_{i=1}^m (P^TA_iQ)y_i = 0\right\} = \mathcal{N}\big(P^T(\mathcal{A}^*\cdot)P\big)\cap\mathcal{N}\big(P^T(\mathcal{A}^*\cdot)Q\big).$$

Now define $\mathcal{P} : \mathbb{R}^{\bar m} \to \mathbb{R}^m$ by

$$\mathcal{P}v = \sum_{i=1}^{\bar m} v_iy_i \quad \text{for } \lambda \in \mathbb{R}^{\bar m}.$$

Then, by definition of $\mathcal{P}$, we have

$$\mathcal{R}(\mathcal{A}^*\mathcal{P}) = \mathcal{R}(Q\cdot Q^T)\cap\mathcal{R}(\mathcal{A}^*) \quad\text{and}\quad \mathcal{R}(\mathcal{P}) = \mathcal{N}\left(P^T(\mathcal{A}^*\cdot)P\right)\cap\mathcal{N}\left(P^T(\mathcal{A}^*\cdot)Q\right).$$

The onto property of $\bar{\mathcal{A}}$ follows from (3.41) and the fact that both $\mathcal{P}, \mathcal{A}^*$ are one-one. Note that if $\bar{\mathcal{A}}^*v = 0$, noting that $\mathcal{A}^*\mathcal{P}v = QWQ^T$ for some $W \in \mathbb{S}^{\bar n}$ by (3.41), we have that $w = 0$ so $\mathcal{A}^*\mathcal{P}v = 0$. Since both $\mathcal{A}^*$ and $\mathcal{P}$ injective, we have that $v = 0$. □

3.5 LP, SDP and the role of strict complementarity

The (near) loss of the Slater CQ results in both theoretical and numerical difficulties, e.g., [46]. In addition, both theoretical and numerical difficulties arise from the loss of strict complementarity, [70]. The connection between strong duality, the Slater CQ, and strict complementarity is seen through the notion of complementarity partitions, [65]. We now see that this plays a key role in the stability and in determining the number of steps k for the facial reduction. In particular, we see that $k = 1$ is characterized by strict complementary slackness and therefore results in a stable formulation.

Definition 3.25. *The pair of faces $F_1 \unlhd K, F_2 \unlhd K^*$ form a* complementarity partition *of K, K^* if $F_1 \subseteq (F_2)^c$. (Equivalently, $F_2 \subseteq (F_1)^c$.) The partition is* proper *if both F_1 and F_2 are proper faces. The partition is* strict *if $(F_1)^c = F_2$ or $(F_2)^c = F_1$.*

We now see the importance of this notion for the facial reduction.

Theorem 3.26. *Let $\delta^* = 0, D^* \succeq 0$ be the optimum of (AP) with dual optimum (γ^*, u^*, W^*). Then the following are equivalent:*

1. *If $D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$ is a maximal rank element of $\mathcal{R}_{\mathrm{D}}$, where $\begin{bmatrix} P & Q \end{bmatrix}$ is orthogonal, $Q \in \mathbb{R}^{n \times \bar{n}}$ and $D_+ \succ 0$, then the reduced problem in (3.39) using D^* satisfies the Slater CQ; only one step of facial reduction is needed.*

2. *Strict complementarity holds for (AP); that is, the primal-dual optimal solution pair $(0, D^*), (0, u^*, W^*)$ for (3.5) and (3.7) satisfy $\operatorname{rank}(D^*) + \operatorname{rank}(W^*) = n$.*

3. *The faces of $\mathbb{S}^n_+$ defined by*

$$\begin{aligned} f^0_{aux,P} &:= \operatorname{face}\left(\{D \in \mathbb{S}^n : \mathcal{A}(D) = 0,\ \langle C, D\rangle = 0,\ D \succeq 0\}\right) \\ f^0_{aux,D} &:= \operatorname{face}\left(\left\{W \in \mathbb{S}^n : W = \mathcal{A}^*_C z \succeq 0,\ \text{for some } z \in \mathbb{R}^{\bar{m}+1}\right\}\right) \end{aligned}$$

form a strict complementarity partition of $\mathbb{S}^n_+$.

Proof. (1) $\iff$ (2): If (3.39) satisfies the Slater CQ, then there exists $\tilde{v} \in \mathbb{R}^{\bar{m}}$ such that $W_Q - \bar{\mathcal{A}}^*\tilde{v} \succ 0$. This implies that $\tilde{Z} := Q(W_Q - \bar{\mathcal{A}}^*\tilde{v})Q^T$ is of rank $\bar{n}$. Moreover,

$$0 \preceq \tilde{Z} = QW_QQ - \mathcal{A}^*\mathcal{P}\tilde{v} = C - \mathcal{A}^*(y_Q + \mathcal{P}\tilde{v}) = \mathcal{A}^*_C \begin{pmatrix} -(y_Q + \mathcal{P}\tilde{v}) \\ 1 \end{pmatrix}.$$

Hence, letting

$$\tilde{u} = \frac{\begin{pmatrix} y_Q + \mathcal{P}\tilde{v} \\ -1 \end{pmatrix}}{\left\| \begin{pmatrix} y_Q + \mathcal{P}\tilde{v} \\ -1 \end{pmatrix} \right\|} \quad \text{and} \quad \tilde{W} = \frac{1}{\left\| \begin{pmatrix} y_Q + \mathcal{P}\tilde{v} \\ -1 \end{pmatrix} \right\|} \tilde{Z},$$

we have that $(0, \tilde{u}, \tilde{W})$ is an optimal solution of (3.7). Since $\operatorname{rank}(D^*) + \operatorname{rank}(\tilde{W}) = (n - \bar{n}) + \bar{n} = n$, we get that strict complementarity holds.

Conversely, suppose that strict complementarity holds for (AP), and let D^* be a maximum rank optimal solution as described in the hypothesis of Item 1. Then there exists an optimal solution

$(0, u^*, W^*)$ for (3.7) such that $\text{rank}(W^*) = \bar{n}$. By complementary slackness, $0 = \langle D^*, W^* \rangle = \langle D_+, P^T W^* P \rangle$, so $W^* \in \mathcal{R}(Q \cdot Q^T)$ and $Q^T W^* Q \succ 0$. Let $u^* = \begin{pmatrix} \tilde{y} \\ -\tilde{\alpha} \end{pmatrix}$, so

$$W^* = \tilde{\alpha} C - \mathcal{A}^* \tilde{y} = \tilde{\alpha} C_Q - \mathcal{A}^*(\tilde{y} - \tilde{\alpha} y_Q).$$

Since $W^*, C_Q \in \mathcal{R}(Q \cdot Q^T)$ implies that $\mathcal{A}^*(\tilde{y} - \tilde{\alpha} y_Q) = \mathcal{A}^* \mathcal{P} \tilde{v}$ for some $\tilde{v} \in \mathbb{R}^{\bar{m}}$, we get

$$0 \prec Q^T W^* Q = \tilde{\alpha} \bar{C} - \bar{\mathcal{A}}^* \tilde{v}.$$

Without loss of generality, we may assume that $\tilde{\alpha} = \pm 1$ or 0. If $\tilde{\alpha} = 1$, then $\bar{C} - \bar{\mathcal{A}}^* \tilde{v} \succ 0$ is a Slater point for (3.39). Consider the remaining two cases. Since (1.1) is assumed to be feasible, the equivalent program (3.39) is also feasible so there exists $\hat{v}$ such that $\bar{C} - \bar{\mathcal{A}}^* \hat{v} \succeq 0$. If $\tilde{\alpha} = 0$, then $\bar{C} - \bar{\mathcal{A}}^*(\hat{v} + \tilde{v}) \succ 0$. If $\tilde{\alpha} = -1$, then $\bar{C} - \bar{\mathcal{A}}^*(2\hat{v} + \tilde{v}) \succ 0$. Hence (3.39) satisfies the Slater CQ.

(2) $\iff$ (3): Notice that $f^0_{aux,P}$ and $f^0_{aux,D}$ are the minimal faces of $\mathbb{S}^n_+$ containing the optimal slacks of (3.5) and (3.7) respectively, and that $f^0_{aux,P}, f^0_{aux,D}$ form a complementarity partition of $\mathbb{S}^n_+ = (\mathbb{S}^n_+)^*$. The complementarity partition is strict if and only if there exist primal-dual optimal slacks D^* and W^* such that $\text{rank}(D^*) + \text{rank}(W^*) = n$. Hence (2) and (3) are equivalent. □

In the special case where the Slater CQ fails and (1.1) is a linear program (and, more generally, the special case of optimizing over an arbitrary polyhedral cone, see e.g., [57, 56, 79, 78]), we see that one single iteration of facial reduction yields a reduced problem that satisfies the Slater CQ.

Corollary 3.27. *Assume that the optimal value of (AP) equals zero, with D^* being a maximum rank optimal solution of (AP). If $A_i = \text{Diag}(a_i)$ for some $a_i \in \mathbb{R}^n$, for $i = 1, \ldots, m$, and $C = \text{Diag}(c)$, for some $c \in \mathbb{R}^n$, then the reduced problem (3.39) satisfies the Slater CQ.*

Proof. In this diagonal case, the SDP is equivalent to an LP. The Goldman-Tucker Theorem [25] implies that there exists a required optimal primal-dual pair for (3.5) and (3.7) that satisfies strict complementarity, so Item 2 in Theorem 3.26 holds. By Theorem 3.26, the reduced problem (3.39) satisfies the Slater CQ. □

4 Facial Reduction

We now study facial reduction for (P) and its sensitivity analysis.

4.1 Two Types

We first outline two algorithms for facial reduction that find the minimal face f_P of (P) . Both are based on solving the auxiliary problem and applying Lemma 3.4. The first algorithm repeatedly finds a face F containing the minimal face and then projects the problem into $F - F$, thus reducing both the size of the constraints as well as the dimension of the variables till finally obtaining the Slater CQ. The second algorithm also repeatedly finds F; but then it identifies the implicit equality constraints till eventually obtaining MFCQ.

4.1.1 Dimension reduction and regularization for the Slater CQ

Suppose that Slater's CQ fails for our given input $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $C \in \mathbb{S}^n$, i.e., the minimal face $f_P \lhd F := \mathbb{S}^n_+$. Our procedure consists of a finite number of repetitions of the following two steps that begin with $k = n$.

1. We first identify $0 \neq D \in (f_P)^c$ using the auxiliary problem (3.5). This means that $f_P \unlhd F \leftarrow \left(\mathbb{S}^k_+ \cap \{D\}^\perp\right)$ and the interior of this new face F is empty.

2. We then project the problem (P) into span(F). Thus we reduce the dimension of the variables and size of the constraints of our problem; the new cone satisfies $\operatorname{int} F \neq \emptyset$. We set $k \leftarrow \dim(F)$.[1]

Therefore, in the case that $\operatorname{int} F = \emptyset$, we need to to obtain an equivalent problem to (P) in the subspace $\operatorname{span}(F) = F - F$. One essential step is finding a subspace intersection. We can apply the algorithm in e.g., [26, Thm 12.4.2]. In particular, by abuse of notation, let H_1, H_2 be matrices with orthonormal columns representing the orthonormal bases of the subspaces $\mathcal{H}_1, \mathcal{H}_2$, respectively. Then we need only find a singular value decomposition $H_1^T H_2 = U \Sigma V^T$ and find which singular vectors correspond to singular values $\Sigma_{ii}, i = 1, \ldots, r$, (close to) 1. Then both $H_1 U(:, 1 : r)$ and $H_2 V(:, 1 : r)$ provide matrices whose ranges yield the intersection. The cone $\mathbb{S}^n_+$ possesses a "self-replicating" structure. Therefore we choose an isometry $\mathcal{I}$ so that $\mathcal{I}(\mathbb{S}^n_+ \cap (F - F))$ is a smaller dimensional PSD cone $\mathbb{S}^r_+$.

Algorithm 4.1 outlines one iteration of facial reduction. The output returns an equivalent problem $(\bar{\mathcal{A}}, \bar{b}, \bar{C})$ on a smaller face of $\mathbb{S}^n_+$ that contains the set of feasible slacks $\mathcal{F}^Z_P$; and, we also obtain the linear transformation $\mathcal{P}$ and point y_Q, which are needed for recovering an optimal solution of the original problem (P) . (See Proposition 3.23.)

Two numerical aspects arising in Algorithm 4.1 need to be considered. The first issue concerns the determination of rank(D^*). In practice, the spectral decomposition of D^* would be of the form

$$D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix} \text{ with } D_\epsilon \approx 0, \quad \text{instead of} \quad D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}.$$

We need to decide which of the eigenvalues of D^* are small enough so that they can be safely rounded down to zero. This is important for the determination of Q, which gives the smaller face $\mathcal{R}(Q \cdot Q^T) \cap \mathbb{S}^n_+$ containing the feasible region $\mathcal{F}^Z_P$. The partitioning of D^* can be done by using similar techniques as in the determination of numerical rank. Assuming that $\lambda_1(D^*) \geq \lambda_2(D^*) \geq \cdots \geq \lambda_n(D^*) \geq 0$, the *numerical rank* rank(D^*, ε) of D^* with respect to a zero tolerance $\varepsilon > 0$ is defined via

$$\lambda_{\operatorname{rank}(D^*,\varepsilon)}(D^*) > \varepsilon \geq \lambda_{\operatorname{rank}(D^*,\varepsilon)+1}(D^*).$$

In implementing Algorithm 4.1, to determine the partitioning of D^*, we use the numerical rank with respect to $\frac{\varepsilon \|D^*\|}{\sqrt{n}}$ where $\varepsilon \in (0, 1)$ is fixed: take $r = \operatorname{rank}\left(D^*, \frac{\varepsilon\|D^*\|}{\sqrt{n}}\right)$,

$$D_+ = \operatorname{Diag}\left(\lambda_1(D^*), \ldots, \lambda_r(D^*)\right), \quad D_\epsilon = \operatorname{Diag}\left(\lambda_{r+1}(D^*), \ldots, \lambda_n(D^*)\right),$$

[1] Note that for numerical stability and well-posedness, it is essential that there exists Lagrange multipliers and that $\operatorname{int} F \neq \emptyset$. Regularization involves both finding a minimal face as well as a minimal subspace, see [65].

Algorithm 4.1: One iteration of facial reduction

Input($\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $C \in \mathbb{S}^n$);
Obtain an optimal solution (δ^*, D^*) of (AP)
if $\delta^* > 0$, **then**
  STOP; Slater CQ holds for $(\mathcal{A}, b, C)$.
else
  if $D^* \succ 0$, **then**
    STOP; generalized Slater CQ holds for $(\mathcal{A}, b, C)$ (see Theorem 3.11);
  else
    Obtain eigenvalue decomposition $D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$ as described in Proposition 3.23, with $Q \in \mathbb{R}^{n \times \bar{n}}$;
    if $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \{0\}$, **then**
      STOP; all feasible solutions of $\sup_y \{b^T y : C - \mathcal{A}^* y \succeq 0\}$ are optimal.
    else
      find $\bar{m}$, $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ satisfying the conditions in Proposition 3.23;
      solve (3.26) for (y_Q, W_Q);
      $\bar{C} \leftarrow W_Q$;
      $\bar{b} \leftarrow \mathcal{P}^* b$;
      $\bar{\mathcal{A}}^* \leftarrow Q^T(\mathcal{A}^* \mathcal{P}(\cdot))Q$;
      Output($\bar{\mathcal{A}} : \mathbb{S}^{\bar{n}} \to \mathbb{R}^{\bar{m}}$, $\bar{b} \in \mathbb{R}^{\bar{m}}$, $\bar{C} \in \mathbb{S}^{\bar{n}}$; $y_Q \in \mathbb{R}^m$, $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$);
    end if
  end if
end if

and partition $\begin{bmatrix} P & Q \end{bmatrix}$ accordingly. Then

$$\lambda_{\min}(D_+) > \frac{\varepsilon \|D^*\|}{\sqrt{n}} \geq \lambda_{\max}(D_\epsilon) \implies \|D_\epsilon\| \leq \varepsilon \|D^*\|.$$

Also,

$$\frac{\|D_\epsilon\|^2}{\|D_+\|^2} = \frac{\|D_\epsilon\|^2}{\|D^*\|^2 - \|D_\epsilon\|^2} \leq \frac{\varepsilon^2 \|D^*\|^2}{(1-\varepsilon^2)\|D^*\|^2} = \frac{1}{\varepsilon^{-2} - 1} \tag{4.1}$$

that is, D_ϵ is negligible comparing with D_+.

The second issue is the computation of intersection of subspaces, $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*)$ (and in particular, finding one-one map $\mathcal{P}$ such that $\mathcal{R}(\mathcal{A}^*\mathcal{P}) = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*)$). This can be done using the following result on subspace intersection.

Theorem 4.1 ([26], Section 12.4.3)**.** *Given $Q \in \mathbb{R}^{n \times \bar{n}}$ of full rank and onto linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, there exist $U_1^{\rm sp}, \ldots, U_{\min\{m, \bar{n}^2\}}^{\rm sp}, V_1^{\rm sp}, \ldots, V_{\min\{m, \bar{n}^2\}}^{\rm sp} \in \mathbb{S}^n$ such that*

$$\begin{array}{rcl} \sigma_1^{\rm sp} := \langle U_1^{\rm sp}, V_1^{\rm sp} \rangle & = & \max\left\{ \langle U, V \rangle : \|U\| = 1 = \|V\|,\ U \in \mathcal{R}(Q \cdot Q^T),\ V \in \mathcal{R}(\mathcal{A}^*) \right\}, \\ \sigma_k^{\rm sp} := \left\langle U_k^{\rm sp}, V_k^{\rm sp} \right\rangle & = & \max\left\{ \langle U, V \rangle : \|U\| = 1 = \|V\|,\ U \in \mathcal{R}(Q \cdot Q^T),\ V \in \mathcal{R}(\mathcal{A}^*), \right. \\ & & \left. \langle U, U_i^{\rm sp} \rangle = 0 = \langle V, V_i^{\rm sp} \rangle,\ \forall\, i = 1, \ldots, k-1 \right\}, \end{array} \tag{4.2}$$

for $k = 2, \ldots, \min\{m, \bar{n}^2\}$, *and* $1 \geq \sigma_1^{\rm sp} \geq \sigma_2^{\rm sp} \geq \cdots \geq \sigma_{\min\{m,\bar{n}^2\}}^{\rm sp} \geq 0$. *Suppose that*

$$\sigma_1^{\rm sp} = \cdots = \sigma_{\bar{m}}^{\rm sp} = 1 > \sigma_{\bar{m}+1}^{\rm sp} \geq \cdots \geq \sigma_{\min\{\bar{n},m\}}^{\rm sp}, \tag{4.3}$$

then

$$\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \operatorname{span}(U_1^{\rm sp}, \ldots, U_{\bar{m}}^{\rm sp}) = \operatorname{span}(V_1^{\rm sp}, \ldots, V_{\bar{m}}^{\rm sp}), \tag{4.4}$$

and $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ *defined by* $\mathcal{P}v = \sum_{i=1}^{\bar{m}} v_i y_i^{\rm sp}$ *for* $v \in \mathbb{R}^{\bar{m}}$, *where* $\mathcal{A}^* y_i^{\rm sp} = V_i^{\rm sp}$ *for* $i = 1, \ldots, \bar{m}$, *is one-one linear and satisfies* $\mathcal{R}(\mathcal{A}^*\mathcal{P}) = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*)$.

In practice, we do not get $\sigma_i^{\rm sp} = 1$ (for $i = 1, \ldots, \bar{m}$) exactly. For a fixed tolerance $\varepsilon^{\rm sp} \geq 0$, suppose that

$$1 \geq \sigma_1^{\rm sp} \geq \cdots \geq \sigma_{\bar{m}}^{\rm sp} \geq 1 - \varepsilon^{\rm sp} > \sigma_{\bar{m}+1}^{\rm sp} \geq \cdots \geq \sigma_{\min\{\bar{n},m\}}^{\rm sp} \geq 0. \tag{4.5}$$

Then we would take the approximation

$$\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) \approx \operatorname{span}(U_1^{\rm sp}, \ldots, U_{\bar{m}}^{\rm sp}) \approx \operatorname{span}(V_1^{\rm sp}, \ldots, V_{\bar{m}}^{\rm sp}). \tag{4.6}$$

Observe that with the chosen tolerance $\varepsilon^{\rm sp}$, we have that the cosines of the principal angles between $\mathcal{R}(Q \cdot Q^T)$ and $\operatorname{span}(V_1^{\rm sp}, \ldots, V_{\bar{m}}^{\rm sp})$ is no less than $1 - \varepsilon^{\rm sp}$; in particular, $\|U_k^{\rm sp} - V_k^{\rm sp}\|^2 \leq 2\varepsilon^{\rm sp}$ and $\|Q^T V_k^{\rm sp} Q\| \geq \sigma_k^{\rm sp} \geq 1 - \varepsilon^{\rm sp}$ for $k = 1, \ldots, \bar{m}$.

Remark 4.2. *Using* $V_1^{\rm sp}, \ldots, V_{\min\{m,\bar{n}^2\}}^{\rm sp}$ *from Theorem 4.1, we may replace* $A_1, \ldots, A_m$ *by* $V_1^{\rm sp}, \ldots, V_m^{\rm sp}$ *(which may require extending* $V_1^{\rm sp}, \ldots, V_{\min\{m,\bar{n}^2\}}^{\rm sp}$ *to a basis of* $\mathcal{R}(\mathcal{A}^*)$, *if* $m > \bar{n}^2$*).*

If the subspace intersection is exact (as in (4.3) *and* (4.4) *in Theorem 4.1), then* $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \operatorname{span}(A_1, \ldots, A_{\bar{m}})$ *would hold. If the intersection is inexact (as in* (4.5) *and* (4.6)*), then we may replace* $\mathcal{A}$ *by* $\breve{\mathcal{A}} : \mathbb{S}^n \to \mathbb{R}^m$, *defined by*

$$\breve{A}_i = \begin{cases} U_i^{\rm sp} & \text{if } i = 1, \ldots, \bar{m}, \\ V_i^{\rm sp} & \text{if } i = \bar{m}+1, \ldots, m, \end{cases}$$

which is a perturbation of $\mathcal{A}$ *with* $\|\mathcal{A}^* - \breve{\mathcal{A}}^*\|_F = \sqrt{\sum_{i=1}^{\bar{m}} \|U_i^{\rm sp} - V_i^{\rm sp}\|^2} \leq \sqrt{2\bar{m}\varepsilon^{\rm sp}}$. *Then* $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\breve{\mathcal{A}}^*) = \operatorname{span}(\breve{A}_1, \ldots, \breve{A}_{\bar{m}})$ *because* $\breve{A}_i \in \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\breve{\mathcal{A}}^*)$ *for* $i = 1, \ldots, \bar{m}$ *and*

$$\begin{aligned}
& \max_{U,V} \Big\{ \langle U, V \rangle : U \in \mathcal{R}(Q \cdot Q^T), \|U\| = 1, V \in \mathcal{R}(\breve{\mathcal{A}}^*), \|V\| = 1, \\
& \qquad\qquad \Big\langle U, U_j^{\rm sp} \Big\rangle = 0 = \Big\langle V, U_j^{\rm sp} \Big\rangle \ \forall\, j = 1, \ldots, \bar{m}, \Big\} \\
\leq\ & \max_{U,y} \left\{ \left\langle U, \sum_{i=1}^{\bar{m}} y_j U_j^{\rm sp} + \sum_{i=\bar{m}+1}^{m} y_j V_j^{\rm sp} \right\rangle : U \in \mathcal{R}(Q \cdot Q^T), \|U\| = 1, \|y\| = 1, \right. \\
& \qquad\qquad \left. \Big\langle U, U_j^{\rm sp} \Big\rangle = 0 \ \forall\, j = 1, \ldots, \bar{m}, \right\} \\
=\ & \max_{U,y} \left\{ \left\langle U, \sum_{i=\bar{m}+1}^{m} y_j V_j^{\rm sp} \right\rangle : U \in \mathcal{R}(Q \cdot Q^T), \|U\| = 1, \|y\| = 1, \Big\langle U, U_j^{\rm sp} \Big\rangle = 0 \ \forall\, j = 1, \ldots, \bar{m}, \right\} \\
=\ & \sigma_{\bar{m}+1}^{\rm sp} < 1 - \varepsilon^{\rm sp} < 1.
\end{aligned}$$

To increase the robustness of the computation of $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*)$ in deciding whether $\sigma_i^{\rm sp}$ is 1 or not, we may follow similar treatment in [18] where one decides which singular values are zero by checking the ratios between successive small singular values.

4.1.2 Implicit equality constraints and regularization for MFCQ

The second algorithm for facial reduction involves repeated use of two steps again.

1. We repeat step 1 in Section 4.1.1 and use (AP) to find the face F.
2. We then find the implicit equality constraints and ensure that they are linearly independent, see Corollary 3.22 and Proposition 3.23.

4.1.3 Preprocessing for the auxiliary problem

We can take advantage of the fact that eigenvalue-eigenvector calculations are efficient and accurate to obtain a more accurate optimal solution (δ^*, D^*) of (AP), i.e., to decide whether the linear system

$$\langle A_i, D\rangle = 0 \;\; \forall\, i = 1, \ldots, m+1 \quad (\text{where } A_{m+1} := C), \quad 0 \neq D \succeq 0 \tag{4.7}$$

has a solution, we can use Algorithm 4.2 as a preprocessor for Algorithm 4.1. More precisely,

Algorithm 4.2: Preprocessing for (AP)

Input($A_1, \ldots, A_m, A_{m+1} := C \in \mathbb{S}^n$);
Output(δ^*, $P \in \mathbb{R}^{n\times(n-\bar{n})}$, $D_+ \in \mathbb{S}^{n-\bar{n}}$ *satisfying* $D_+ \succ 0$*; (so* $D^* = PD_+P^T$*)*);
if *one of the* A_i *(*$i \in \{1, \ldots, m+1\}$*) is definite* **then**
    STOP; (4.7) does not have a solution.
else
    if *some of the* $A = \begin{bmatrix} U & \tilde{U} \end{bmatrix} \begin{bmatrix} \hat{D} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U^T \\ \tilde{U}^T \end{bmatrix} \in \{A_i : i = 1, \ldots, m+1\}$ *satisfies* $\hat{D} \succ 0$, **then**
        reduce the size using $A_i \leftarrow \tilde{U}^T A_i \tilde{U}, \forall i$;
    else
        if $\exists\, 0 \neq V \in \mathbb{R}^{n\times r}$ *such that* $A_i V = 0$ *for all* $i = 1, \ldots, m+1$, **then**
            % We get $\langle A_i, VV^T\rangle = 0 \;\forall\, i = 1, \ldots, m+1$;
            STOP; we get $\delta^* = 0, D^* = VV^T$ solves (AP);
        else
            U
        end if
        se an SDP solver to solve (AP).
    end if
end if

Algorithm 4.2 tries to find a solution D^* satisfying (4.7) without using an SDP solver. It attempts to find a vector v in the nullspace of all the A_i, and then sets $D^* = vv^T$. In addition, any semidefinite A_i allows a reduction to a smaller dimensional space.

4.2 Backward stability of one iteration of facial reduction

We now provide the details for one iteration of the main algorithm, see Theorem 4.9. Algorithm 4.1 involves many nontrivial subroutines, each of which would introduce some numerical errors. First we need to obtain an optimal solution (δ^*, D^*) of (AP); in practice we can only get an approximate optimal solution, as δ^* is never exactly zero, and we decide whether the true value of δ^* is zero when the computed value is only close to zero. Second we need to obtain the eigenvalue decomposition of D^*. There comes the issue of determining which of the nearly zero eigenvalues are indeed zero. (Since (AP) is not solved exactly, the approximate solution D^* would have eigenvalues that are positive but close to zero.) Finally, the subspace intersection $\mathcal{R}(Q\cdot Q^T)\cap\mathcal{R}(\mathcal{A}^*)$ (for finding $\bar{m}$ and $\mathcal{P}$) can only be computed approximately via a singular value decomposition, because in practice we would take singular vectors corresponding to singular values that are approximately (but not exactly) 1.

It is important that Algorithm 4.1 is robust against such numerical issues arising from the subroutines. We show that Algorithm 4.1 is backward stable (with respect to these three categories of numerical errors), i.e., for any given input $(\mathcal{A}, b, c)$, there exists $(\tilde{\mathcal{A}}, \tilde{b}, \tilde{C}) \approx (\mathcal{A}, b, C)$ such that the computed result of Algorithm 4.1 applied on $(\mathcal{A}, b, C)$ is equal to the exact result of the same algorithm applied on $(\tilde{\mathcal{A}}, \tilde{b}, \tilde{C})$ (when (AP) is solved exactly and the subspace intersection is determined exactly).

We first show that $\|C_{\text{res}}\|$ is relatively small, given a small $\alpha(\mathcal{A}, C)$.

Lemma 4.3. *Let y_Q, C_Q, C_{res} be defined as in Lemma 3.19. Then the norm of C_{res} is small in the sense that*

$$\|C_{\text{res}}\| \leq \sqrt{2}\left[\frac{\|D^*\|}{\lambda_{\min}(D_+)}\alpha(\mathcal{A}, C)\right]^{1/2}\left(\min_{Z=C-\mathcal{A}^*y\succeq 0}\|Z\|\right). \tag{4.8}$$

Proof. By optimality, for any $y \in \mathcal{F}_p$,

$$\|C_{\text{res}}\| \leq \min_W \|C - \mathcal{A}^*y - QWQ^T\| = \|Z - QQ^TZQQ^T\|,$$

where $Z := C - \mathcal{A}^*y$. Therefore (4.8) follows from Proposition 3.16. ☐

The following technical results shows the relationship between the quantity $\min_{\|y\|=1}\|\mathcal{A}^*y\|^2 - \|Q^T(\mathcal{A}^*y)Q\|^2$ and the cosine of the smallest principal angle between $\mathcal{R}(\mathcal{A}^*)$ and $\mathcal{R}(Q\cdot Q^T)$, defined in (4.2).

Lemma 4.4. *Let $Q \in \mathbb{R}^{n\times\bar{n}}$ satisfy $Q^TQ = I_{\bar{n}}$. Then*

$$\tau := \min_{\|y\|=1}\left\{\|\mathcal{A}^*y\|^2 - \|Q^T(\mathcal{A}^*y)Q\|^2\right\} \geq \left(1 - (\sigma_1^{\text{sp}})^2\right)\sigma_{\min}(\mathcal{A}^*)^2 \geq 0, \tag{4.9}$$

where σ_1^{sp} is defined in (4.2)*. Moreover,*

$$\tau = 0 \iff \sigma_1^{\text{sp}} = 1 \iff \mathcal{R}(Q\cdot Q^T)\cap\mathcal{R}(\mathcal{A}^*) \neq \{0\}. \tag{4.10}$$

Proof. By definition of σ_1^{sp},

$$\begin{aligned}
&\max_V\left\{\max_{\|U\|=1, U\in\mathcal{R}(Q\cdot Q^T)}\langle U, V\rangle : \|V\| = 1, V \in \mathcal{R}(\mathcal{A}^*)\right\}\\
\geq\ &\max_{\|U\|=1, U\in\mathcal{R}(Q\cdot Q^T)}\langle U, V_1^{\text{sp}}\rangle \quad\geq\quad \langle U_1^{\text{sp}}, V_1^{\text{sp}}\rangle \quad=\quad \sigma_1^{\text{sp}}\\
\geq\ &\max_V\left\{\max_{\|U\|=1, U\in\mathcal{R}(Q\cdot Q^T)}\langle U, V\rangle : \|V\| = 1, V \in \mathcal{R}(\mathcal{A}^*)\right\},
\end{aligned}$$

so equality holds througout, implying that

$$\begin{aligned}\sigma_1^{\mathrm{sp}} &= \max_V \left\{ \max_{\|U\|=1, U\in\mathcal{R}(Q\cdot Q^T)} \langle U, V\rangle : \|V\| = 1, V \in \mathcal{R}(\mathcal{A}^*) \right\} \\ &= \max_y \left\{ \max_{\|W\|=1} \langle QWQ^T, \mathcal{A}^* y\rangle : \|\mathcal{A}^* y\| = 1 \right\} \\ &= \max_y \left\{ \|Q^T(\mathcal{A}^* y)Q\| : \|\mathcal{A}^* y\| = 1 \right\}.\end{aligned}$$

Obviously, $\|\mathcal{A}^* y\| = 1$ implies that the orthogonal projection $QQ^T(\mathcal{A}^* y)QQ^T$ onto $\mathcal{R}(Q\cdot Q^T)$ is of norm no larger than one:

$$\|Q^T(\mathcal{A}^* y)Q\| = \|QQ^T(\mathcal{A}^* y)QQ^T\| \le \|\mathcal{A}^* y\| = 1. \tag{4.11}$$

Hence $\sigma_1^{\mathrm{sp}} \in [0,1]$. In addition, equality holds in (4.11) if and only if $\mathcal{A}^* y \in \mathcal{R}(Q\cdot Q^T)$, hence

$$\sigma_1^{\mathrm{sp}} = 1 \iff \mathcal{R}(\mathcal{A}^*) \cap \mathcal{R}(Q\cdot Q^T) \ne \{0\}. \tag{4.12}$$

Whenever $\|y\| = 1$, $\|\mathcal{A}^* y\| \ge \sigma_{\min}(\mathcal{A}^*)$. Hence

$$\begin{aligned}\tau &= \min_y \left\{ \|\mathcal{A}^* y\|^2 - \|Q^T(\mathcal{A}^* y)Q\|^2 : \|y\| = 1 \right\} \\ &= \sigma_{\min}(\mathcal{A}^*)^2 \min_y \left\{ \|\mathcal{A}^* y\|^2 - \|Q^T(\mathcal{A}^* y)Q\|^2 : \|y\| = \frac{1}{\sigma_{\min}(\mathcal{A}^*)} \right\} \\ &\ge \sigma_{\min}(\mathcal{A}^*)^2 \min_y \left\{ \|\mathcal{A}^* y\|^2 - \|Q^T(\mathcal{A}^* y)Q\|^2 : \|\mathcal{A}^* y\| \ge 1 \right\} \\ &= \sigma_{\min}(\mathcal{A}^*)^2 \min_y \left\{ \|\mathcal{A}^* y\|^2 - \|Q^T(\mathcal{A}^* y)Q\|^2 : \|\mathcal{A}^* y\| = 1 \right\} \\ &= \sigma_{\min}(\mathcal{A}^*)^2 \left(1 - \max_y \left\{ \|Q^T(\mathcal{A}^* y)Q\|^2 : \|\mathcal{A}^* y\| = 1 \right\} \right) \\ &= \sigma_{\min}(\mathcal{A}^*)^2 \left(1 - (\sigma_1^{\mathrm{sp}})^2 \right).\end{aligned}$$

This together with $\sigma_1^{\mathrm{sp}} \in [0,1]$ proves (4.9). If $\tau = 0$, then $\sigma_1^{\mathrm{sp}} = 1$ since $\sigma_{\min}(\mathcal{A}^*) > 0$. Then (4.12) implies that $\mathcal{R}(\mathcal{A}^*) \cap \mathcal{R}(Q\cdot Q^T) \ne \{0\}$. Conversely, if $\mathcal{R}(\mathcal{A}^*) \cap \mathcal{R}(Q\cdot Q^T) \ne \{0\}$, then there exists $\hat{y}$ such that $\|\hat{y}\| = 1$ and $\mathcal{A}^*\hat{y} \in \mathcal{R}(Q\cdot Q^T)$. This implies that

$$0 \le \tau \le \|\mathcal{A}^*\hat{y}\|^2 - \|Q^T(\mathcal{A}^*\hat{y})Q\|^2 = 0,$$

so $\tau = 0$. This together with (4.12) proves the second claim (4.10). □

Next we prove that two classes of matrices are positive semidefinite and show their eigenvalue bounds, which will be useful in the backward stability result.

Lemma 4.5. *Suppose $A_1, \ldots, A_m, D^* \in \mathbb{S}^n$. Then the matrix $\hat{M} \in \mathbb{S}^m$ defined by*

$$\hat{M}_{ij} = \langle A_i, D^*\rangle \langle A_j, D^*\rangle \quad (i,j = 1,\ldots,m)$$

is positive semidefinite. Moreover, the largest eigenvalue $\lambda_{\max}(\hat{M}) \le \sum_{i=1}^m \langle A_i, D^*\rangle^2$.

Proof. For any $y \in \mathbb{R}^m$,

$$y^T \hat{M} y = \sum_{i,j=1}^{m} \langle A_i, D^* \rangle \langle A_j, D^* \rangle y_i y_j = \left(\sum_{i=1}^{m} \langle A_i, D^* \rangle y_i \right)^2 .$$

Hence $\hat{M}$ is positive semidefinite. Moreover, by the Cauchy Schwarz inequality we have

$$y^T \hat{M} y = \left(\sum_{i=1}^{m} \langle A_i, D^* \rangle y_i \right)^2 \leq \left(\sum_{i=1}^{m} \langle A_i, D^* \rangle^2 \right) \|y\|_2^2.$$

Hence $\lambda_{\max}(\hat{M}) \leq \sum_{i=1}^{m} \langle A_i, D^* \rangle^2$. □

Lemma 4.6. *Suppose $A_1, \ldots, A_m \in \mathbb{S}^n$ and $Q \in \mathbb{R}^{n \times \bar{n}}$ has orthonomral columns. Then the matrix $M \in \mathbb{S}^m$ defined by*

$$M_{ij} = \langle A_i, A_j \rangle - \left\langle Q^T A_i Q, Q^T A_j Q \right\rangle, \quad i, j = 1, \ldots, m,$$

is positive semidefinite, with the smallest eigenvalue $\lambda_{\min}(M) \geq \tau$, where τ is defined in (4.9).

Proof. For any $y \in \mathbb{R}^m$, we have

$$y^T M y = \sum_{i,j=1}^{m} \langle y_i A_i, y_j A_j \rangle - \left\langle y_i Q^T A_i Q, y_j Q^T A_j Q \right\rangle = \|\mathcal{A}^* y\|^2 - \left\| Q^T (\mathcal{A}^* y) Q \right\|^2 \geq \tau \|y\|^2.$$

Hence $M \in \mathbb{S}^m_+$ and $\lambda_{\min}(M) \geq \tau$. □

The following lemma shows that when nonnegative δ^* is approximately zero and $D^* = P D_+ P^T + Q D_\epsilon Q^T \approx P D_+ P^T$ with $D_+ \succ 0$, under a mild assumption (4.15) it is possible to find a linear operator $\hat{\mathcal{A}}$ "near" $\mathcal{A}$ such that we can take the following approximation:

$$\delta^* \leftarrow 0, \quad D^* \leftarrow P D_+ P^T, \quad \mathcal{A}^* \leftarrow \hat{\mathcal{A}}^*,$$

and we maintain that $\hat{\mathcal{A}}(P D_+ P^T) = 0$ and $\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\hat{\mathcal{A}}^*)$.

Lemma 4.7. *Let $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m : X \mapsto (\langle A_i, X \rangle)$ be onto. Let $D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix} \in \mathbb{S}^n_+$, where $\begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, $D_+ \succ 0$ and $D_\epsilon \succeq 0$. Suppose that*

$$\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \operatorname{span}(A_1, \ldots, A_{\bar{m}}), \tag{4.13}$$

for some $\bar{m} \in \{1, \ldots, m\}$. Then

$$\min_{\|y\|=1,\, y \in \mathbb{R}^{m-\bar{m}}} \left\{ \left\| \sum_{i=1}^{m-\bar{m}} y_i A_{\bar{m}+i} \right\|^2 - \left\| \sum_{i=1}^{m-\bar{m}} y_i Q^T A_{\bar{m}+i} Q \right\|^2 \right\} > 0. \tag{4.14}$$

Assume that

$$\min_{\|y\|=1,\, y \in \mathbb{R}^{m-\bar{m}}} \left\{ \left\| \sum_{i=1}^{m-\bar{m}} y_i A_{\bar{m}+i} \right\|^2 - \left\| \sum_{i=1}^{m-\bar{m}} y_i Q^T A_{\bar{m}+i} Q \right\|^2 \right\} > \frac{2}{\|D_+\|^2} \left(\|\mathcal{A}(D^*)\|^2 + \|D_\epsilon\|^2 \sum_{i=\bar{m}+1}^{m} \|A_i\|^2 \right). \tag{4.15}$$

Define $\tilde{A}_i$ *to be the projection of* A_i *on* $\{PD_+P^T\}^\perp$*:*

$$\tilde{A}_i := A_i - \frac{\langle A_i, PD_+P^T\rangle}{\langle D_+, D_+\rangle} PD_+P^T, \quad \forall\, i = 1, \ldots, m. \tag{4.16}$$

Then

$$\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\tilde{\mathcal{A}}^*) = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*). \tag{4.17}$$

Proof. We first prove the strict inequality (4.14). First observe that since

$$\left\| \sum_{i=1}^{m-\bar{m}} y_i A_{\bar{m}+i} \right\|^2 - \left\| \sum_{i=1}^{m-\bar{m}} y_i Q^T A_{\bar{m}+i} Q \right\|^2 = \left\| \sum_{i=1}^{m-\bar{m}} y_i (A_{\bar{m}+i} - QQ^T A_{\bar{m}+i} QQ^T) \right\|^2 \geq 0,$$

the optimal value is always nonnegative. Let $\bar{y}$ solve the minimization problem in (4.14). If $\left\|\sum_{i=1}^{m-\bar{m}} \bar{y}_i A_{\bar{m}+i}\right\|^2 - \left\|\sum_{i=1}^{m-\bar{m}} \bar{y}_i Q^T A_{\bar{m}+i} Q\right\|^2 = 0$, then

$$0 \neq \sum_{i=1}^{m-\bar{m}} \bar{y}_i A_{\bar{m}+i} \in \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \operatorname{span}(A_1, \ldots, A_{\bar{m}}),$$

which is absurd since $A_1, \ldots, A_m$ are linearly independent.

Now we prove (4.17). Observe that for $j = 1, \ldots, \bar{m}$, $A_j \in \mathcal{R}(Q \cdot Q^T)$ so $\langle A_j, PD_+P^T\rangle = 0$, which implies that $\tilde{A}_j = A_j$. Moreover,

$$\operatorname{span}(A_1, \ldots, A_{\bar{m}}) \subseteq \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\tilde{\mathcal{A}}^*).$$

Conversely, suppose that $B := \tilde{\mathcal{A}}^* y \in \mathcal{R}(Q \cdot Q^T)$. Since $\tilde{A}_j = A_j \in \mathcal{R}(Q \cdot Q^T)$ for $j = 1, \ldots, \bar{m}$,

$$B = QQ^T BQQ^T \implies \sum_{j=\bar{m}+1}^{m} y_j (\tilde{A}_j - QQ^T \tilde{A}_j QQ^T) = 0$$

We show that $y_{\bar{m}+1} = \cdots = y_m = 0$. In fact, since $Q^T(PD_+P^T)Q = 0$, $\sum_{j=\bar{m}+1}^{m} y_j(\tilde{A}_j - QQ^T \tilde{A}_j QQ^T) = 0$ implies

$$\sum_{j=\bar{m}+1}^{m} y_j QQ^T A_j QQ^T = \sum_{j=\bar{m}+1}^{m} y_j A_j - \left(\sum_{j=\bar{m}+1}^{m} \frac{\langle A_j, PD_+P^T\rangle}{\langle D_+, D_+\rangle} y_j \right) PD_+P^T.$$

For $i = \bar{m}+1, \ldots, m$, taking inner product on both sides with A_i ,

$$\sum_{j=\bar{m}+1}^{m} \langle Q^T A_i Q, Q^T A_j Q\rangle\, y_j = \sum_{j=\bar{m}+1}^{m} \langle A_i, A_j\rangle\, y_j - \sum_{j=\bar{m}+1}^{m} \frac{\langle A_i, PD_+P^T\rangle \langle A_j, PD_+P^T\rangle}{\langle D_+, D_+\rangle} y_j,$$

which holds if, and only if,

$$(M - \tilde{M}) \begin{pmatrix} y_{\bar{m}+1} \\ \vdots \\ y_m \end{pmatrix} = 0, \tag{4.18}$$

where $M, \tilde{M} \in \mathbb{S}^{m-\bar{m}}$ are defined by

$$\begin{aligned} M_{(i-\bar{m}),(j-\bar{m})} &= \langle A_i, A_j \rangle - \left\langle Q^T A_i Q, Q^T A_j Q \right\rangle, \\ \tilde{M}_{(i-\bar{m}),(j-\bar{m})} &= \frac{\left\langle A_i, PD_+P^T \right\rangle \left\langle A_j, PD_+P^T \right\rangle}{\langle D_+, D_+ \rangle} \quad, \forall\, i,j = \bar{m}+1, \ldots, m. \end{aligned}$$

We show that (4.18) implies that $y_{\bar{m}+1} = \cdots = y_m = 0$ by proving that $M - \tilde{M}$ is indeed positive definite. By Lemmas 4.5 and 4.6,

$$\begin{aligned} \lambda_{\min}(M - \tilde{M}) &\geq \lambda_{\min}(M) - \lambda_{\max}(\tilde{M}) \\ &\geq \min_{\|y\|=1} \left\{ \left\| \sum_{i=1}^{m-\bar{m}} y_i A_{\bar{m}+i} \right\|^2 - \left\| \sum_{i=1}^{m-\bar{m}} y_i Q^T A_{\bar{m}+i} Q \right\|^2 \right\} - \frac{\sum_{i=\bar{m}+1}^{m} \left\langle A_i, PD_+P^T \right\rangle^2}{\langle D_+, D_+ \rangle}. \end{aligned}$$

To see that $\lambda_{\min}(M - \tilde{M}) > 0$, note that since $D^* = PD_+P^T + QD_\epsilon Q^T$, for all i,

$$\begin{aligned} \left| \left\langle A_i, PD_+P^T \right\rangle \right| &\leq \left| \langle A_i, D^* \rangle \right| + \left| \left\langle A_i, QD_\epsilon Q^T \right\rangle \right| \\ &\leq |\langle A_i, D^* \rangle| + \|A_i\| \|QD_\epsilon Q^T\| \\ &= |\langle A_i, D^* \rangle| + \|A_i\| \|D_\epsilon\| \\ &\leq \sqrt{2} \left(|\langle A_i, D^* \rangle|^2 + \|A_i\|^2 \|D_\epsilon\|^2 \right)^{1/2}. \end{aligned}$$

Hence

$$\sum_{i=\bar{m}+1}^{m} \left| \left\langle A_i, PD_+P^T \right\rangle \right|^2 \leq 2 \sum_{i=\bar{m}+1}^{m} \left(|\langle A_i, D^* \rangle|^2 + \|A_i\|^2 \|D_\epsilon\|^2 \right) \leq 2\|\mathcal{A}(D^*)\|^2 + 2\|D_\epsilon\|^2 \sum_{i=\bar{m}+1}^{m} \|A_i\|^2,$$

and that $\lambda_{\min}(M - \tilde{M}) > 0$ follows from the assumption (4.15). This implies that $y_{\bar{m}+1} = \cdots = y_m = 0$. Therefore $B = \sum_{i=1}^{\bar{m}} y_i \tilde{A}_i$, and by (4.13)

$$\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\tilde{\mathcal{A}}^*) = \operatorname{span}(A_1, \ldots, A_{\bar{m}}) = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*).$$

□

Remark 4.8. *We make a remark about the assumption* (4.15) *in Lemma 4.7. We argue that the right hand side expression*

$$\frac{2}{\|D_+\|^2} \left(\|\mathcal{A}(D^*)\|^2 + \|D_\epsilon\|^2 \sum_{i=\bar{m}+1}^{m} \|A_i\|^2 \right)$$

is close to zero (when $\delta^* \approx 0$ and when D_ϵ is chosen appropriately). Assume that the spectral decomposition of D^* is partitioned as described in Section 4.1.1. Then (since $\|D_\epsilon\| \leq \varepsilon \|D^*\|$)

$$\frac{2}{\|D_+\|^2} \|\mathcal{A}(D^*)\|^2 \leq \frac{2(\delta^*)^2}{\|D^*\|^2 - \|D_\epsilon\|^2} \leq \frac{2(\delta^*)^2}{\|D^*\|^2 - \varepsilon^2 \|D^*\|^2} \leq \frac{2n(\delta^*)^2}{1 - \varepsilon^2}$$

and

$$\frac{2\|D_\epsilon\|^2}{\|D_+\|^2} \sum_{i=\bar{m}+1}^{m} \|A_i\|^2 \leq \frac{2\varepsilon^2}{1 - \varepsilon^2} \sum_{i=\bar{m}+1}^{m} \|A_i\|^2.$$

Therefore as long as ε and δ^* are small enough (taking into account n and $\sum_{i=\bar{m}+1}^{m} \|A_i\|^2$), then the right hand side of (4.15) *would be close to zero.*

Here we provide the backward stability result for one step of the facial reduction algorithm. That is, we show that the smaller problem obtained from one step of facial reduction with $\delta^* \geq 0$ is equivalent to applying facial reduction exactly to an SDP instance "nearby" to the original SDP instance.

Theorem 4.9. *Suppose $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$ and $C \in \mathbb{S}^n$ are given so that* (1.1) *is feasible and Algorithm 4.1 returns (δ^*, D^*), with $0 \leq \delta^* \approx 0$ and spectral decomposition $D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$, and $(\bar{\mathcal{A}}, \bar{b}, \bar{C}, y_Q, \mathcal{P})$. In addition, assume that*

$$\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m : v \mapsto \begin{pmatrix} v \\ 0 \end{pmatrix}, \quad \textit{so } \mathcal{R}(\mathcal{A}^*\mathcal{P}) = \operatorname{span}(A_1, \ldots, A_{\bar{m}}).$$

Assume also that (4.15) *holds. For $i = 1, \ldots, m$, define $\tilde{A}_i \in \mathbb{S}^n$ as in* (4.16), *and $\tilde{\mathcal{A}}^* y := \sum_{i=1}^m y_i \tilde{A}_i$. Let $\tilde{C} = \tilde{\mathcal{A}}^* y_Q + Q\bar{C}Q^T$. Then $(\bar{\mathcal{A}}, \bar{b}, \bar{C})$ is the exact output of Algorithm 4.1 applied on $(\tilde{\mathcal{A}}, b, \tilde{C})$, that is, the following hold:*

(1) $\tilde{\mathcal{A}}_{\tilde{C}}(PD_+P^T) = \begin{pmatrix} \tilde{\mathcal{A}}(PD_+P^T) \\ \left\langle \tilde{C}, PD_+P^T \right\rangle \end{pmatrix} = 0,$

(2) $(y_Q, \bar{C})$ *solves*

$$\min_{y,Q} \frac{1}{2} \left\| \tilde{\mathcal{A}}^* y + QWQ^T - \tilde{C} \right\|^2. \tag{4.19}$$

(3) $\mathcal{R}(\tilde{\mathcal{A}}^*\mathcal{P}) = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\tilde{\mathcal{A}}^*)$.

Moreover, $(\tilde{\mathcal{A}}, b, \tilde{C})$ is close to $(\mathcal{A}, b, C)$ in the sense that

$$\sum_{i=1}^m \|A_i - \tilde{A}_i\|^2 \leq \frac{2}{\|D_+\|^2} \left((\delta^*)^2 + \|D_\epsilon\|^2 \sum_{i=1}^m \|A_i\|^2 \right), \tag{4.20}$$

$$\|C - \tilde{C}\| \leq \frac{\sqrt{2}}{\|D_+\|} \left((\delta^*)^2 + \|D_\epsilon\|^2 \sum_{i=1}^m \|A_i\|^2 \right)^{1/2} \|y_Q\| + \sqrt{2} \left[\frac{\|D^*\|}{\lambda_{\min}(D_+)} \alpha(\mathcal{A}, C) \right]^{1/2} \left(\min_{Z = C - \mathcal{A}^* y \succeq 0} \|Z\| \right), \tag{4.21}$$

where $\alpha(\mathcal{A}, c)$ is defined in (3.13).

Proof. First we show that $(\bar{\mathcal{A}}, \bar{b}, \bar{C})$ is the exact output of Algorithm 4.1 applied on $(\tilde{\mathcal{A}}, b, \tilde{C})$:

(1) For $i = 1, \ldots, m$, by definition of $\tilde{A}_i$ in (4.16), we have $\left\langle \tilde{A}_i, PD_+P^T \right\rangle = 0$. Hence $\tilde{\mathcal{A}}(PD_+P^T) = 0$. Also, $\left\langle \tilde{C}, PD_+P^T \right\rangle = y_Q^T(\tilde{\mathcal{A}}(PD_+P^T)) + \left\langle \bar{C}, Q^T(PD_+P^T)Q \right\rangle = 0$.

(2) By definition, $\tilde{C} - \tilde{\mathcal{A}}^* y_Q - Q\bar{C}Q^T = 0$, so $(y_Q, \bar{C})$ solves the least squares problem (4.19).

(3) Given (4.15), we have that

$$\mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\tilde{\mathcal{A}}^*) = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*) = \mathcal{R}(A_1, \ldots, A_{\bar{m}}) = \mathcal{R}(\tilde{A}_1, \ldots, \tilde{A}_{\bar{m}}) = \mathcal{R}(\tilde{\mathcal{A}}^*\mathcal{P}).$$

The results (4.20) and (4.21) follow easily:

$$\begin{array}{rcl}\sum_{i=1}^m \|A_i-\tilde{A}_i\|^2 = \sum_{i=1}^m \frac{\left|\langle A_i, PD_+P^T\rangle\right|^2}{\|D_+\|^2} & \leq & \sum_{i=1}^m \frac{2\left|\langle A_i, D^*\rangle\right|^2 + 2\|A_i\|^2\|D_\epsilon\|^2}{\|D_+\|^2} \\ & \leq & \frac{2}{\|D_+\|^2}\left((\delta^*)^2 + \|D_\epsilon\|^2\sum_{i=1}^m \|A_i\|^2\right),\end{array}$$

and

$$\begin{array}{rcl}\|C-\tilde{C}\| & \leq & \|\mathcal{A}^*y_Q - \tilde{\mathcal{A}}^*y_Q\| + \|C_{\mathrm{res}}\| \\ & \leq & \sum_{i=1}^m |(y_Q)_i|\|A_i-\tilde{A}_i\| + \|C_{\mathrm{res}}\| \\ & \leq & \|y_Q\|\left(\sum_{i=1}^m \|A_i-\tilde{A}_i\|^2\right)^{1/2} + \|C_{\mathrm{res}}\| \\ & \leq & \frac{\sqrt{2}}{\|D_+\|}\left((\delta^*)^2 + \|D_\epsilon\|^2\sum_{i=1}^m \|A_i\|^2\right)^{1/2} \|y_Q\| \\ & & \qquad + \sqrt{2}\left[\frac{\|D^*\|}{\lambda_{\min}(D_+)}\alpha(\mathcal{A},C)\right]^{1/2}\left(\min_{Z=C-\mathcal{A}^*y\succeq 0}\|Z\|\right),\end{array}$$

from (4.20) and (4.8). □

5 Test Problem Descriptions

5.1 Worst case instance

From Tunçel [66], we consider the following *worst case* problem instance in the sense that for $n \geq 3$, the facial reduction process in Algorithm 4.1 requires $n-1$ steps to obtain the minimal face. Let $b = e_2 \in \mathbb{R}^n$, $C = 0$, and $\mathcal{A} : \mathbb{S}^n_+ \to \mathbb{R}^n$ be defined by

$$A_1 = e_1e_1^T,\ A_2 = e_1e_2^T + e_2e_1^T,\ A_i = e_{i-1}e_{i-1}^T + e_1e_i^T + e_ie_1^T \text{ for } i = 3,\dots,n.$$

It is easy to see that

$$\mathcal{F}_P^Z = \left\{C - \mathcal{A}^*y \in \mathbb{S}^n_+ : y \in \mathbb{R}^n\right\} = \left\{\mu e_1e_1^T : \mu \geq 0\right\},$$

(so $\mathcal{F}_P^Z$ has empty interior) and

$$\sup\{b^Ty : C - \mathcal{A}^*y \succeq 0\} = \sup\{y_2 : -\mathcal{A}^*y = \mu e_1e_1^T, \mu \geq 0\} = 0,$$

which is attained by any feasible solution.

Now consider the auxiliary problem

$$\min \|\mathcal{A}_C(D)\| = \left[D_{11}^2 + 4D_{12}^2 + \sum_{i=3}^n (D_{i-1,i-1} + 2D_{1i})\right]^{1/2} \quad \text{s.t.} \quad \langle D, I\rangle = \sqrt{n},\ D \succeq 0.$$

An optimal solution is $D^* = \sqrt{n} e_n e_n^T$, which attains objective value zero. It is easy to see this is the only solution. More precisely, any solution D attaining objective value 0 must satisfy $D_{11} = 0$, and by the positive semidefiniteness constraint $D_{1,i} = 0$ for $i = 2, \ldots, n$ and so $D_{ii} = 0$ for $i = 2, \ldots, n-1$. So D_{nn} is the only nonzero entry and must equal $\sqrt{n}$ by the linear constraint $\langle D, I \rangle = \sqrt{n}$. Therefore, Q from Proposition 3.16 must have $n-1$ columns, implying that the reduced problem is in $\mathbb{S}^{n-1}$. Theoretically, each facial reduction step via the auxiliary problem can only reduce the dimension by one. Moreover, after each reduction step, we get the same SDP with n reduced by one. Hence it would take $n-1$ facial reduction steps before a reduced problem with strictly feasible solutions is found. This realizes the result in [12] on the upper bound of the number of facial reduction steps needed.

5.2 Generating instances with finite nonzero duality gaps

In this section we give a procedure for generating SDP instances with finite nonzero duality gaps. The algorithm is due to the results in [65, 70].

Algorithm 5.1: Generating SDP instance that has a finite nonzero duality gap

---

1 Input(*problem dimensions m, n; desired duality gap g*);

2 Output(*linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $C \in \mathbb{S}^n$ such that the corresponding primal dual pair* (1.1)-(1.2) *has a finite nonzero duality gap*);

1. Pick any positive integer r_1,r_3 that satisfy $r_1 + r_3 + 1 = n$, and any positive integer $p \le r_3$.

2. Choose $A_i \succeq 0$ for $i = 1, \ldots, p$ so that $\dim(\text{face}(\{A_i : i = 1, \ldots, p\})) = r_3$. Specifically, choose $A_1, \ldots, A_p$ so that

$$\text{face}(\{A_i : 1, \ldots, p\}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbb{S}^{r_3}_+ \end{bmatrix}. \tag{5.1}$$

3. Choose $A_{p+1}, \ldots, A_m$ of the form

$$A_i = \begin{bmatrix} 0 & 0 & (A_i)_{13} \\ 0 & (A_i)_{22} & * \\ (A_i)_{13}^T & * & * \end{bmatrix},$$

where an asterisk denotes a block having arbitrary elements, such that $(A_{p+1})_{13}, \ldots, (A_m)_{13}$ are linearly independent, and $(A_i)_{22} \succ 0$ for some $i \in \{p+1, \ldots, m\}$.

4. Pick

$$\bar{X} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sqrt{g} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{5.2}$$

5. Take $b = \mathcal{A}(\bar{X})$, $C = \bar{X}$.

Finite nonzero duality gaps and strict complementarity are closely tied together for cone optimization problems; using the concept of a *complementarity partition*, we can generate instances that fail to have strict complementarity; these in turn can be used to generate instances with finite nonzero duality gaps. See [65, 70].

Theorem 5.1. *Given any positive integers n, $m \le n(n+1)/2$ and any $g > 0$ as input for Algorithm 5.1, the following statements hold for the primal-dual pair* (1.1)-(1.2) *corresponding to the output data from Algorithm 5.1:*

1. *Both* (1.1) *and* (1.2) *are feasible.*
2. *All primal feasible points are optimal and $v_P = 0$.*
3. *All dual feasible point are optimal and $v_D = g > 0$.*

It follows that (1.1) *and* (1.2) *possess a finite positive duality gap.*

Proof. Consider the primal problem (1.1). (1.1) is feasible because $C := \bar{X}$ given in (5.2) is positive semidefinite. Note that by definition of $\mathcal{A}$ in Algorithm 5.1, for any $y \in \mathbb{R}^m$,

$$C - \sum_{i=1}^{p} y_i A_i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sqrt{g} & 0 \\ 0 & 0 & * \end{bmatrix} \text{ and } -\sum_{i=p+1}^{m} y_i A_i = \begin{bmatrix} 0 & 0 & * \\ 0 & * & * \\ * & * & * \end{bmatrix},$$

so if $y \in \mathbb{R}^m$ satisfies $Z := C - \mathcal{A}^* y \succeq 0$, then $\sum_{i=p+1}^{m} y_i A_i = 0$ must hold. This implies $\sum_{i=p+1}^{m} y_i (A_i)_{13} = 0$. Since $(A_{p+1})_{13}, \ldots, (A_m)_{13}$ are linearly independent, we must have $y_{p+1} = \cdots = y_m = 0$. Consequently, if y is feasible for (1.1), then

$$\mathcal{A}^* y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -Z_{33} \end{bmatrix}$$

for some $Z_{33} \succeq 0$. The corresponding objective value in (1.1) is given by

$$b^T y = \langle \bar{X}, \mathcal{A}^* y \rangle = 0.$$

This shows that the objective value of (1.1) is constant over the feasible region. Hence $v_P = 0$, and all primal feasible solutions are optimal.

Consider the dual problem (1.2). By the choice of b, $\bar{X} \succeq 0$ is a feasible solution, so (1.2) is feasible too. From (5.1), we have that $b_1 = \cdots = b_p = 0$. Let $X \succeq 0$ be feasible for (1.1). Then $\langle A_i, X \rangle = b_i = 0$ for $i = 1, \ldots, p$, implying that the (3,3) block of X must be zero by (5.1), so

$$X = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since $\alpha = (A_j)_{22} > 0$ for some $j \in \{p+1, \ldots, m\}$, we have that

$$\alpha X_{22} = \langle A_j, X \rangle = \langle A_j, \bar{X} \rangle = \alpha \sqrt{g},$$

so $X_{22} = \sqrt{g}$ and $\langle C, X \rangle = g$. Therefore the objective value of (1.2) is constant and equals $g > 0$ over the feasible region, and all feasible solutions are optimal. ∎

5.3 Numerical results

Table 1 shows a comparison of solving SDP instances *with* versus *without* facial reduction. Examples 1 through 9 are specially generated problems available online at the URL for this paper given in the footnote on page 1. In particular: Example 3 has a positive duality gap, $v_P = 0 < v_D = 1$; for Example 4, the dual is infeasible; in Example 5, the Slater CQ holds; Examples 9a,9b are instances of the worst case problems presented in Section 5.1. The remaining instances RandGen1-RandGen11 are generated randomly with most of them having a finite positive duality gap, as described in Section 5.2. These instances generically require only one iteration of facial reduction. The software package SeDuMi is used to solve the SDPs that arise.

Name	n	m	True primal optimal value	True dual optimal value	Primal optimal value with facial reduction	Primal optimal value without facial reduction
Example 1	3	2	0	0	0	-6.30238e-016
Example 2	3	2	0	1	0	+0.570395
Example 3	3	4	0	0	0	+6.91452e-005
Example 4	3	3	0	Infeas.	0	+Inf
Example 5	10	5	*	*	+5.02950e+02	+5.02950e+02
Example 6	6	8	1	1	+1	+1
Example 7	5	3	0	0	0	-2.76307e-012
Example 9a	20	20	0	Infeas.	0	Inf
Example 9b	100	100	0	Infeas.	0	Inf
RandGen1	10	5	0	1.4509	+1.5914e-015	+1.16729e-012
RandGen2	100	67	0	5.5288e+003	+1.1056e-010	NaN
RandGen4	200	140	0	2.6168e+004	+1.02803e-009	NaN
RandGen5	120	45	0	0.0381	-5.47393e-015	-1.63758e-015
RandGen6	320	140	0	2.5869e+005	+5.9077e-025	NaN
RandGen7	40	27	0	168.5226	-5.2203e-029	+5.64118e-011
RandGen8	60	40	0	4.1908	-2.03227e-029	NaN
RandGen9	60	40	0	61.0780	+5.61602e-015	-3.52291e-012
RandGen10	180	100	0	5.1461e+004	+2.47204e-010	NaN
RandGen11	255	150	0	4.6639e+004	+7.71685e-010	NaN

Table 1: Comparisons with/without facial reduction

One general observation is that, if the instance has primal-dual optimal solutions and has zero duality gap, SeDuMi is able to find the optimal solutions. However, if the instance has finite nonzero duality gaps, and if the instance is not too small, SeDuMi is unable to compute any solution, and returns NaN.

SeDuMi, based on self-dual embedding, embeds the input primal-dual pair into a larger SDP that satisfies the Slater CQ [16]. Theoretically, the lack of the Slater CQ in a given primal-dual pair is not an issue for SeDuMi. It is not known what exactly causes problem on SeDuMi when handling instances where a nonzero duality gap is present.

6 Conclusions and future work

In this paper we have presented a preprocessing technique for SDP problems where the Slater CQ (nearly) fails. This is based on solving a stable auxiliary problem that approximately identifies the minimal face for (P) . We have included a backward error analysis and some preliminary tests that successfully solve problems where the CQ fails and also problems that have a duality gap. The optimal value of our (AP) has significance as a measure of *nearness to infeasibility*.

Though our stable (AP) satisfied both the primal and dual generalized Slater CQ, high accuracy solutions were difficult to obtain for unstructured general problems. (AP) is equivalent to the underdetermined linear least squares problem

$$\min \|\mathcal{A}_C(D)\|_2^2 \quad \text{s.t.} \quad \langle I, D\rangle = \sqrt{n}, \quad D \succeq 0, \tag{6.1}$$

which is known to be difficult to solve. High accuracy solutions are essential in performing a proper facial reduction.

Extensions of some of our results can be made to general conic convex programming, in which case the partial orderings in (1.1) and (1.2) are induced by a proper closed convex cone K and the dual cone K^*, respectively.

References

[1] A. Alfakih, A. Khandani, and H. Wolkowicz. Solving Euclidean distance matrix completion problems via semidefinite programming. *Comput. Optim. Appl.*, 12(1-3):13–30, 1999. A tribute to Olvi Mangasarian. 10

[2] B. Alipanahi, N. Krislock, and A. Ghodsi. Manifold learning by semidefinite facial reduction. Technical Report Submitted to Machine Learning Journal, University of Waterloo, Waterloo, Ontario, 2010. 10

[3] F. Alizadeh, J-P.A. Haeberly, and M.L. Overton. Complementarity and nondegeneracy in semidefinite programming. *Math. Programming*, 77:111–128, 1997. 9

[4] A.F. Anjos and J.B. Lasserre, editors. *Handbook on Semidefinite, Conic and Polynomial Optimization*. International Series in Operations Research & Management Science. Springer-Verlag, 2011. 4

[5] M.F. Anjos and H. Wolkowicz. Strengthened semidefinite relaxations via a second lifting for the Max-Cut problem. *Discrete Appl. Math.*, 119(1-2):79–106, 2002. Foundations of heuristics in combinatorial optimization. 10

[6] A. Ben-Israel, A. Ben-Tal, and S. Zlobec. *Optimality in nonlinear programming: a feasible directions approach*. John Wiley & Sons Inc., New York, 1981. A Wiley-Interscience Publication. 7, 8, 9

[7] A. Ben-Israel, A. Charnes, and K. Kortanek. Duality and asymptotic solvability over cones. *Bulletin of American Mathematical Society*, 75(2):318–324, 1969. 10

[8] J.F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000. 7

[9] B. Borchers. CSDP, a C library for semidefinite programming. *Optim. Methods Softw.*, 11/12(1-4):613–623, 1999. projects.coin-or.org/Csdp. 4

[10] J.M. Borwein and H. Wolkowicz. Characterization of optimality for the abstract convex program with finite-dimensional range. *J. Austral. Math. Soc. Ser. A*, 30(4):390–411, 1980/81. 4, 9, 10, 11

[11] J.M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 30(3):369–380, 1980/81. 3, 4, 10

[12] J.M. Borwein and H. Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. Appl.*, 83(2):495–530, 1981. 3, 4, 10, 11, 40

[13] S. Boyd, V. Balakrishnan, E. Feron, and L. El Ghaoui. Control system analysis and synthesis via linear matrix inequalities. *Proc. ACC*, pages 2147–2154, 1993. 3

[14] F. Burkowski, Y-L. Cheung, and H. Wolkowicz. Efficient use of semidefinite programming for selection of rotamers in protein conformations. Technical Report 30 pages, submitted Dec. 2012, University of Waterloo, Waterloo, Ontario, 2012. 10

[15] R.J. Caron, A. Boneh, and S. Boneh. Redundancy. In *Advances in sensitivity analysis and parametric programming*, volume 6 of *Internat. Ser. Oper. Res. Management Sci.*, pages 13.1–13.41. Kluwer Acad. Publ., Boston, MA, 1997. 7

[16] E. de Klerk. *Interior point methods for semidefinite programming*. PhD thesis, Delft University, 1997. 4, 42

[17] E. de Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Applied Optimization Series. Kluwer Academic, Boston, MA, 2002. 4

[18] J. Demmel and B. Kågström. The generalized Schur decomposition of an arbitrary pencil $A-\lambda B$; robust software with error bounds and applications. part II: software and applications. *ACM Trans. Math. Softw.*, 19(2):175–201, June 1993. 31

[19] X.V. Doan, S. Kruk, and H. Wolkowicz. A robust algorithm for semidefinite programming. *Optim. Methods Softw.*, 27(4-5):667–693, 2012. 4, 14

[20] R. Fourer and D.M. Gay. Experience with a primal presolve algorithm. In *Large scale optimization (Gainesville, FL, 1993)*, pages 135–154. Kluwer Acad. Publ., Dordrecht, 1994. 6

[21] R.M. Freund. Complexity of an algorithm for finding an approximate solution of a semi-definite program with no regularity assumption. Technical Report OR 302-94, MIT, Cambridge, MA, 1994. 4

[22] R.M. Freund. Complexity of convex optimization using geometry-based measures and a reference point. *Math. Program.*, 99(2, Ser. A):197–221, 2004. 4

[23] R.M. Freund, F. Ordóñez, and K.C. Toh. Behavioral measures and their correlation with IPM iteration counts on semi-definite programming problems. USC-ISE working paper #2005-02, MIT, 2005. url: www-rcf.usc.edu/˜fordon/. 4

[24] R.M. Freund and J.R. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. Technical report, MIT, Cambridge, MA, 1997. 4

[25] A.J. Goldman and A.W. Tucker. Theory of linear programming. In *Linear inequalities and related systems*, pages 53–97. Princeton University Press, Princeton, N.J., 1956. Annals of Mathematics Studies, no. 38. 28

[26] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 3^{nd} edition, 1996. 29, 30

[27] J. Gondzio. Presolve analysis of linear programs prior to applying an interior point method. *INFORMS J. Comput.*, 9(1):73–91, 1997. 6

[28] N.I.M. Gould and Ph.L. Toint. Preprocessing for quadratic programming. *Math. Program.*, 100(1, Ser. B):95–132, 2004. 7

[29] D. Gourion and A. Seeger. Critical angles in polyhedral convex cones: numerical and statistical considerations. *Math. Programming*, 123(1):173–198, 2010. 21

[30] G. Gruber and F. Rendl. Computational experience with ill-posed problems in semidefinite programming. *Comput. Optim. Appl.*, 21(2):201–212, 2002. 4

[31] J-B. Hiriart-Urruty and J. Malick. A fresh variational-analysis look at the positive semide?nite matrices world. Technical report, University of Tolouse, Toulouse, France, 2010. 4

[32] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original. 6

[33] A. Iusem and A. Seeger. Searching for critical angles in a convex cone. *Math. Program.*, 120(1, Ser. B):3–25, 2009. 21

[34] S. Jibrin. *Redundancy in Semidefinite Programming*. PhD thesis, Carleton University, Ottawa, Ontario, Canada, 1997. 9

[35] M.H. Karwan, V. Lotfi, J. Telgen, and S. Zionts. *Redundancy in mathematical programming*. Springer-Verlag, New York, NY, 1983. 6

[36] N. Krislock and H. Wolkowicz. Explicit sensor network localization using semidefinite representations and facial reductions. *SIAM Journal on Optimization*, 20(5):2679–2708, 2010. 9

[37] M. Lamoureux and H. Wolkowicz. Numerical decomposition of a convex function. *J. Optim. Theory Appl.*, 47(1):51–64, 1985. 8

[38] Z-Q. Luo, J.F. Sturm, and S. Zhang. Conic convex programming and self-dual embedding. *Optim. Methods Softw.*, 14(3):169–218, 2000. 3, 10

[39] J. Malick, J. Povh, F. Rendl, and A. Wiegele. Regularization methods for semidefinite programming. *SIAM Journal on Optimization*, 20(1):336–356, 2009. 4

[40] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.*, 17:37–47, 1967. 7

[41] R.D.C. Monteiro and M.J. Todd. Path-following methods. In *Handbook of Semidefinite Programming*, pages 267–306. Kluwer Acad. Publ., Boston, MA, 2000. 14

[42] Cs Mszros and U. H. Suhl. Advanced preprocessing techniques for linear and quadratic programming. *OR Spectrum*, 25:575–595, 2003. 10.1007/s00291-003-0130-x. 7

[43] Y.E. Nesterov, M.J. Todd, and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Math. Program.*, 84(2, Ser. A):227–267, 1999. 3

[44] G. Pataki. On the closedness of the linear image of a closed convex cone. *Math. Oper. Res.*, 32(2):395–412, 2007. 4

[45] G. Pataki. Bad semidefinite programs: they all look the same. Technical report, Department of Operations Research, University of North Carolina, Chapel Hill, 2011. 4, 10

[46] J. Peña and J. Renegar. Computing approximate solutions for convex conic systems of constraints. *Math. Program.*, 87(3, Ser. A):351–383, 2000. 14, 21, 27

[47] I. Pólik and T. Terlaky. New stopping criteria for detecting infeasibility in conic optimization. *Optim. Lett.*, 3(2):187–198, 2009. 3

[48] M.V. Ramana. *An Algorithmic Analysis of Multiquadratic and Semidefinite Programming Problems*. PhD thesis, Johns Hopkins University, Baltimore, Md, 1993. 4

[49] M.V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Math. Programming*, 77(2):129–162, 1997. 4, 10

[50] M.V. Ramana, L. Tunçel, and H. Wolkowicz. Strong duality for semidefinite programming. *SIAM J. Optim.*, 7(3):641–662, 1997. 4

[51] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2001. 10

[52] S.M. Robinson. Stability theorems for systems of inequalities, part i: linear systems. *SIAM J. Numerical Analysis*, 12:754–769, 1975. 3

[53] S.M. Robinson. First order conditions for general nonlinear optimization. *SIAM J. Appl. Math.*, 30(4):597–607, 1976. 5

[54] R. Tyrrell Rockafellar. Some convex programs whose duals are linearly constrained. In *Nonlinear Programming (Proc. Sympos., Univ. of Wisconsin, Madison, Wis., 1970)*, pages 293–322. Academic Press, New York, 1970. 8

[55] R.T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks. 10

[56] A. Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming (Alicante, 1999)*, volume 57 of *Nonconvex Optim. Appl.*, pages 135–165. Kluwer Acad. Publ., Dordrecht, 2001. 28

[57] A. Shapiro and A. Nemirovskii. Duality of linear conic problems. Technical report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 2003. 28

[58] G.W. Stewart. Rank degeneracy. *SIAM J. Sci. Statist. Comput.*, 5(2):403–413, 1984. 3

[59] G.W. Stewart. Determining rank in the presence of error. In *Linear algebra for large scale and real-time applications (Leuven, 1992)*, volume 232 of *NATO Adv. Sci. Inst. Ser. E Appl. Sci.*, pages 275–291. Kluwer Acad. Publ., Dordrecht, 1993. 3

[60] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11/12(1-4):625–653, 1999. sedumi.ie.lehigh.edu. 4

[61] D. Sun. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Math. Oper. Res.*, 31(4):761–776, 2006. 5

[62] M. J. Todd and Y. Ye. Approximate Farkas lemmas and stopping rules for iterative infeasible-point algorithms for linear programming. *Math. Programming*, 81(1, Ser. A):1–21, 1998. 3

[63] M.J. Todd. Semidefinite programming. *Acta Numerica*, 10:515–560, 2001. 4, 10

[64] L. Tunçel. On the Slater condition for the SDP relaxations of nonconvex sets. *Oper. Res. Lett.*, 29(4):181–186, 2001. 9

[65] L. Tunçel and H. Wolkowicz. Strong duality and minimal representations for cone optimization. *coap*, 53(2):619–648, 2012. 4, 13, 27, 29, 40, 41

[66] Levent Tunçel. *Polyhedral and semidefinite programming methods in combinatorial optimization*, volume 27 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2010. 3, 9, 39

[67] R.H. Tütüncü, K.C. Toh, and M.J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Program.*, 95(2, Ser. B):189–217, 2003. www.math.nus.edu.sg/~mattohkc/sdpt3.html. 4

[68] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996. 4

[69] H. Waki, S. Kim, M. Kojima, and M. Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optim.*, 17(1):218–242, 2006. 10

[70] H. Wei and H. Wolkowicz. Generating and solving hard instances in semidefinite programming. *Math. Programming*, 125(1):31–45, 2010. 27, 40, 41

[71] H. Wolkowicz. Calculating the cone of directions of constancy. *J. Optim. Theory Appl.*, 25(3):451–457, 1978. 8

[72] H. Wolkowicz. Some applications of optimization in matrix theory. *Linear Algebra Appl.*, 40:101–118, 1981. 4

[73] H. Wolkowicz. Solving semidefinite programs using preconditioned conjugate gradients. *Optim. Methods Softw.*, 19(6):653–672, 2004. 4

[74] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of semidefinite programming*. International Series in Operations Research & Management Science, 27. Kluwer Academic Publishers, Boston, MA, 2000. Theory, algorithms, and applications. 4, 10

[75] H. Wolkowicz and Q. Zhao. Semidefinite programming relaxations for the graph partitioning problem. *Discrete Appl. Math.*, 96/97:461–479, 1999. Selected for the special Editors' Choice, Edition 1999. 9

[76] M. Yamashita, K. Fujisawa, and M. Kojima. Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0). *Optim. Methods Softw.*, 18(4):491–505, 2003. sdpa.indsys.chuo-u.ac.jp/sdpa/. 4

[77] M. Yamashita, K. Fujisawa, K. Nakata, M. Nakata, M. Fukuda, K. Kobayashi, and K. Goto. A high-performance software package for semidefinite programs: Sdpa7. Technical report, Dept. of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan, 2010. 4

[78] Constantin Zălinescu. On zero duality gap and the Farkas lemma for conic programming. *Math. Oper. Res.*, 33(4):991–1001, 2008. 28

[79] Constantin Zălinescu. On duality gap in linear conic problems. Technical report, University of Al. I. Cusa, Iasi, Romania, 2010. 28

[80] Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *J. Comb. Optim.*, 2(1):71–109, 1998. Semidefinite programming and interior-point approaches for combinatorial optimization problems (Fields Institute, Toronto, ON, 1996). 9

Index

A Replies from authors to referee report2

The replies from the authors follow below within the text/report provided by the referee. The replies are prefixed by REPLY:, and indented. These replies follow the questions/concerns raised in the file report2.pdf.

Most of the changes are limited to Section 4.

A.1 *Circular:* Theorem 4.4 and Proposition 4.7

As I indicated in the initial referee report, this manuscript contains interesting and original work but there were some weaknesses in the material concerning stability. Although the authors have revised the paper, it is disappointing to see no responses to several of the concerns clearly spelled out in the previous referee report. The authors have ignored the concerns about Theorem 4.4 and Proposition 4.7 previously raised. As I stated before, these results are circular and weak. In particular, the bound (4.5) in Theorem 4.4 can be rephrased as follows: there exists y feasible for the original problem such that

$$\|y - Pv\| \leq \frac{C_{res}}{C_{res} + \lambda_{\min}(\hat{Z})} \|\hat{y} - \mathcal{P}v\|$$

where $\hat{y}$ is feasible for the original problem and $\hat{Z} = C - \mathcal{A}^*\hat{y}$. This effectively bounds $\|y - Pv\|$ in terms of another quantity of the same kind. This type of tautology is obviously universally true and does not require any proof. I must be missing something and hence urge the authors to add some kind of discussion on the merit of Theorem 4.4. Perhaps there is something special about $\hat{y}$ and $\hat{Z}$? The same can be said about Proposition 4.7.

> **REPLY:** We have removed the section on sensitivity analysis. Therefore, these "circular" arguments are not a concern. Note that $\hat{y}$ was fixed and so the error bounds made sense for *large* v, i.e., $\|y - Pv\| \leq O(\|v\|)$ when $\|v\|$ is large.

A.2 Four additional ignored concerns

- First line in the proof of Lemma 4.2: since y is already fixed, I believe the inequality should read

$$\text{``}\|C_{\text{res}}\| = \min_W \|C - \mathcal{A}^*y - QWQ^T\| = \|Z - QQ^TZQQ^T\|\text{''}$$

> **REPLY:** The variable under the min has been changed. Please note that we have removed the old section 4.2 on sensitivity analysis, but have kept this Lemma. It is now Lemma 4.3. Also note that the first equality has been changed to an inequality since the y is arbitrary and not necessarily optimal.

- The matrix C_Q in Corollary 4.3 is not defined.

> **REPLY:** First, note that we no longer have this Corollary as it was part of the old Section 4.2. But, $C_Q = QW_QQ^T$ first appears in a comment following Lemma 3.19. We have added an appropriate definition in Lemma 3.19 and an index entry. We have changed the appearance throughout the paper.

- Last step in the proof of Theorem 4.4: I could not see why the inequality holds.

  REPLY: This Theorem was part of the old Section 4.2 and has been removed.

- The quantity κ in Corollary 4.5 is not defined.

  REPLY: This Corollary was part of the old Section 4.2 and has been removed.

A.3 New added subsection on backward stability

The authors have added a new subsection with a formal statement on backward stability (Theorem 4.9).

> **REPLY:** We would like to emphasize again that we included a modified backward stability statement in Section 3.4, i.e., backwards stability with respect to a perturbation in the cone.

This certainly addresses some of the main weaknesses mentioned above. However, I was unable to get the punch line in the last statement of the proof of Lemma 4.10: First, I do not see why it follows that $y_{\bar{m}+1} = \cdots = y_m = 0$.

> **REPLY:** Since we removed the section on sensitivity, Lemma 4.10 is now Lemma 4.7 in this new revision. We have added extra details to explain the conclusions of the Lemma, e.g., $y_{\bar{m}+1} = \cdots = y_m = 0$ is due to $\lambda_{\min}(M - \tilde{M}) > 0$ and equation (4.18).

Second, I do not see why this shows the desired inclusion.

> **REPLY:** This inclusion is now in Lemma 4.7, and the proof is on page 36.
>
> We started out with an arbitrary $B \in \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\tilde{\mathcal{A}}^*)$, i.e., $B = \sum_{j=1}^m y_j \tilde{A}_j \in \mathcal{R}(Q \cdot Q^T)$ for some y.
>
> $y_{\bar{m}+1} = \cdots = y_m = 0$ together with $\tilde{A}_j = A_j$ for $j = 1, \ldots, \bar{m}$ (from the second paragraph of the proof) implies that
>
> $$B = \sum_{j=1}^{m} y_j \tilde{A}_j = \sum_{j=1}^{\bar{m}} y_j \tilde{A}_j = \sum_{j=1}^{\bar{m}} y_j A_j \in \operatorname{span}\{A_1, \ldots, A_{\bar{m}}\},$$
>
> i.e., $B \in \mathcal{R}(Q \cdot Q^T) \cap \operatorname{span}\{A_1, \ldots, A_{\bar{m}}\} = \mathcal{R}(Q \cdot Q^T) \cap \mathcal{R}(\mathcal{A}^*)$, where the equality follows from the assumption (4.13).

In addition, Theorem 4.12 would be far stronger if the authors could argue that (4.15) typically holds. Does it?

> **REPLY:** We justify the assumption (4.15) in Remark 4.8.

A.4 Errors in cross references in the manuscript

- Proposition 4.6: “Corollary 3.20” does not exist. Should it be “Theorem 3.20”?

  REPLY: This cross reference was in the section that was removed.

  In addition, Z is not defined. Is it $Z := C_{res} + C_Q - \mathcal{A}^* y$?

  REPLY: This cross reference was in the section that was removed.

- Remark 4.11: ”Lemma 4.11” does not exist. Should it be ”Lemma 4.9” or ”Lemma 4.10”?

  REPLY: The cross reference was fixed using the latex label command.

- We found some additional errors, e.g. the word Proposition was missing in the reference to Prop 3.13 near the top of page 17.