Separation and relaxation for cones of quadratic forms*

Samuel Burer[†] Hongbo Dong[‡]

May 14, 2010
Revised: April 28, 2011

Abstract

Let $P \subseteq \Re_n$ be a pointed, polyhedral cone. In this paper, we study the cone $\mathcal{C} = \text{cone}\{xx^T : x \in P\}$ of quadratic forms. Understanding the structure of $\mathcal{C}$ is important for globally solving NP-hard quadratic programs over P. We establish key characteristics of $\mathcal{C}$ and construct a separation algorithm for $\mathcal{C}$ provided one can optimize with respect to a related cone over the boundary of P. This algorithm leads to a nonlinear representation of $\mathcal{C}$ and a class of tractable relaxations for $\mathcal{C}$, each of which improves a standard polyhedral-semidefinite relaxation of $\mathcal{C}$. The relaxation technique can further be applied recursively to obtain a hierarchy of relaxations, and for constant recursive depth, the hierarchy is tractable. We apply this theory to two important cases: P is the nonnegative orthant, in which case $\mathcal{C}$ is the cone of completely positive matrices; and P is the homogenized cone of the "box" $[0, 1]^n$. Through various results and examples, we demonstrate the strength of the theory for these cases. For example, we achieve for the first time a separation algorithm for 5×5 completely positive matrices.

Keywords: Quadratic form, Convex hull, Separation, Relaxation, Global optimization, Semidefinite programming

Mathematics Subject Classification: 90C20, 90C22, 90C26

1 Introduction

Let $P := \{x : Ax = 0, Bx \geq 0\} \subseteq \Re_n$ be a pointed, polyhedral cone such that $P \cap \{x : x_1 = 1\}$ is nonempty. We consider the following optimization problem:

$$\nu_* := \inf \{ \langle x, Hx \rangle : x \in P, \; x_1 = 1 \} . \tag{1}$$

*The research of both authors was supported in part by NSF Grant CCF-0545514.

[†]Department of Management Sciences, University of Iowa, Iowa City, IA, 52242-1994, USA. Email: samuel-burer@uiowa.edu

[‡]Corresponding author. Applied Mathematics and Computational Sciences, University of Iowa, Iowa City, IA, 52242-1419, USA. Email: hongbo-dong@uiowa.edu

Problem (1) models many NP-hard programs having a nonconvex quadratic objective and linear constraints. For example, $\inf\{\langle \tilde{x}, \tilde{H}\tilde{x}\rangle + 2\langle \tilde{c}, \tilde{x}\rangle : \tilde{A}\tilde{x} = \tilde{a},\ \tilde{B}\tilde{x} \geq \tilde{b}\}$ can be cast as (1) with

$$x = \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix}, \quad H = \begin{pmatrix} 0 & \tilde{c}^T \\ \tilde{c} & \tilde{H} \end{pmatrix}, \quad A = \begin{pmatrix} -\tilde{a} & \tilde{A} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ -\tilde{b} & \tilde{B} \end{pmatrix},$$

where the first row of B is included so that the resulting P is the homogenized feasible set. See the surveys [15, 16] for local and global methods for solving (1); BARON [26] provides general-purpose software. In this paper, we will examine two particular specifications of P in detail: (i) P equals the nonnegative orthant $\Re^+_n := \{x : x \geq 0\}$, i.e., A is empty and $B = I$; and (ii) P equals the homogenization of the box $\square_{n-1} := \{\tilde{x} : 0 \leq \tilde{x} \leq \tilde{e}\}$, e.g., A is empty and

$$B = \begin{pmatrix} 0 & \tilde{I} \\ \tilde{e} & -\tilde{I} \end{pmatrix},$$

where $\tilde{I}$ is the identity matrix of size $n-1$ and $\tilde{e}$ is the all-ones vector. Another specification of P, which is closely related to (i), is the homogenization of the standard simplex $\Delta_{n-1} := \{\tilde{x} \geq 0 : \tilde{e}^T\tilde{x} = 1\}$, e.g., $A = (-1, \tilde{e}^T)$ and $B = (0, \tilde{I})$.

For globally optimizing (1), a convex cone of interest is

$$\mathcal{C} := \mathcal{C}(P) := \text{cone}\{xx^T : x \in P\}, \tag{2}$$

where cone$(\cdot)$ indicates the convex conic hull, since this enables the convexification of (1) in the space $\mathcal{S}_n$ of symmetric matrices via the identity $\langle x, Hx\rangle = \langle H, xx^T\rangle$: $\nu_* = \inf\{\ \langle H, X\rangle : X \in \mathcal{C},\ X_{11} = 1\ \}$ (see Theorem 1 in Section 2). Sturm and Zhang [29] study $\mathcal{C}(P)$ for arbitrary sets P and prove several fundamental results, which we employ in this paper. While optimization over $\mathcal{C}$ may be tractable in specific cases, in general it is NP-hard. Nevertheless, even partial information about $\mathcal{C}$ can help to solve (1), for example, by providing a means to calculate tight lower bounds on ν_*. One standard, tractable relaxation of $\mathcal{C}$ is the following semidefinite relaxation [21, 27], where in addition to enforcing positive semidefiniteness, pairs of constraints defining P are multiplied and linearized to obtain valid constraints for $\mathcal{C}$:

$$\mathcal{D} := \mathcal{D}(P) := \left\{X \succeq 0 : AXA^T = 0,\ AXB^T = 0,\ BXB^T \geq 0\right\}. \tag{3}$$

$\mathcal{D}$ is a strong relaxation of $\mathcal{C}$ in the sense that it incorporates all constraints defining P to the fullest extent possible in the space $\mathcal{S}_n$, e.g., without introducing new variables.

Regarding $\mathcal{C}$, this paper investigates theoretical properties, algorithmic approaches, and improved convex relaxations. In Section 2, we establish some basic facts concerning $\mathcal{C}$, $\mathcal{D}$, and the dual cone $\mathcal{C}^*$. Then, in Section 3, we devise a separation procedure for $\mathcal{C}$ based on optimization over the dual cone of $\mathcal{C}_{\rm bd} := \mathcal{C}(\mathrm{bd}(P)) := \mathrm{cone}\{xx^T : x \in \mathrm{bd}(P)\}$. This separation procedure is especially attractive when $\mathcal{C}^*_{\rm bd}$ is tractable, e.g., has a representation with lower dimensional matrices. The separation algorithm then leads to a nonlinear formulation of $\mathcal{C}$ in terms of $\mathcal{C}_{\rm bd}$, which in turn motivates the construction of a new convex relaxation $\mathcal{C}(d)$ of $\mathcal{C}$, which depends on $\mathcal{C}_{\rm bd}$ and the choice of a "step direction" $d \in P$. We prove that the intersection of such $\mathcal{C}(d)$ over all $d \in P$ captures $\mathcal{C}$ exactly, i.e., $\mathcal{C} = \cap_{d \in P} \mathcal{C}(d)$. Further, one can relax $\mathcal{C}_{\rm bd}$ in $\mathcal{C}(d)$ to obtain a new tractable relaxation $\mathcal{D}(d)$ of $\mathcal{C}$, which is not weaker than $\mathcal{D}$. In total, we have $\mathcal{C} \subseteq \mathcal{C}(d) \subseteq \mathcal{D}(d) \subseteq \mathcal{D}$. Finally, we extend these ideas to construct a recursive hierarchy of convex relaxations for $\mathcal{C}$, which provides stronger and stronger approximations of $\mathcal{C}$. For a given choice of step directions and fixed recursive depth, each relaxation in the hierarchy is tractable.

In Sections 4 and 5, we tailor Section 3 to the two cases of P mentioned above: the nonnegative orthant $\Re^+_n$ and the homogenization $\mathrm{hom}(\square_{n-1})$ of the box. In the first case, $\mathcal{C}$ is the well-known cone of $n \times n$ completely positive matrices. The monograph [2] provides a recent survey; see also [18, 30, 31]. Recent relevant work includes [7, 12, 14, 19, 17]. In the second case, $\mathcal{C}$ is a fundamental convex cone for quadratic programs over bounded variables for which recent theoretical and computational advances include [1, 8, 9, 32, 33]. In each section, we review the literature on $\mathcal{C}$, focusing on known results in low dimensions. For example, in each of the two cases, respectively, it is known that $\mathcal{C} = \mathcal{D}$ if and only if $n \leq 4$ and $n \leq 3$. Combining this with the separation procedure of Section 3 and the observation that, in each case, $\mathrm{bd}(P)$ decomposes into several lower-dimensional versions of P, we establish separation procedures for the respective dimensions $n = 5$ and $n = 4$. In particular, this gives the first full separation procedure for 5×5 completely positive matrices, which extends the previous separation algorithms of [7, 12] for special classes of matrices in $\mathcal{D} \setminus \mathcal{C}$. Using a result of [1], we also achieve the first separation procedure for $\mathcal{C}(\mathrm{hom}(\square_4))$ when $n = 5$. Through several examples, we also demonstrate the strength of the relaxation hierarchies in these two cases.

We remark that the completely positive cone $\mathcal{C}(\Re^+_n)$ is actually relevant to any $P \subseteq \Re^+_n$ since then $\mathcal{C}(P) \subseteq \mathcal{C}(\Re^+_n)$. In other words, knowledge of $\mathcal{C}(\Re^+_n)$ can be applied directly to $\mathcal{C}(P)$. In fact, as long as P in any dimension contains n nonnegative variables, $\mathcal{C}(\Re^+_n)$ applies to those variables. For example, the separation procedure for $\mathcal{C}(\Re^+_5)$ can be applied to any subset of five nonnegative variables. In a similar manner, $\mathcal{C}(\mathrm{hom}(\square_{n-1}))$ applies to any P having $n - 1$ variables which, prior to homogenization, are in $[0, 1]^{n-1}$. Moreover, simple variable scalings allow the application of $\mathcal{C}(\mathrm{hom}(\square_{n-1}))$ to bounded variables generally.

In fact, the completely positive cone has even wider applicability and generality. By splitting free variables and adding slacks, (1) can be recast as $\inf\{\langle \hat{x}, \hat{H}\hat{x}\rangle : \hat{x} \in \hat{P}, \hat{x}_1 = 1\}$, where $\hat{P}$ is a polyhedron of the form $\{\hat{x} : \hat{A}\hat{x} = 0, \hat{x} \geq 0\}$ in some larger space $\Re_{\hat{n}}$. Then $\mathcal{C}(\hat{P}) \subseteq \mathcal{C}(\Re^+_{\hat{n}})$. Indeed, the results of Burer [6] imply $\mathcal{C}(\hat{P}) = \mathcal{C}(\Re^+_{\hat{n}}) \cap \{\hat{X} : \hat{A}\hat{X}\hat{A}^T = 0\}$. There may be, however, specific disadvantages to this general embedding of $\mathcal{C}(P)$ into $\mathcal{C}(\Re^+_{\hat{n}})$. For example, by transforming $\{\tilde{x} : 0 \leq \tilde{x} \leq \tilde{e}\}$ to $\{\binom{\tilde{x}}{\tilde{s}} \geq 0 : \tilde{x} + \tilde{s} = \tilde{e}\}$, $\mathcal{C}(\text{hom}(\square_{n-1}))$ may be embedded into $\mathcal{C}(\Re^+_{2n-1})$, but for $n = 4$, the separation procedure for $\mathcal{C}(\text{hom}(\square_3))$ of this paper is sacrificed since there is no known separation procedure for $\mathcal{C}(\Re^+_7)$. So we do not focus on such embeddings here.

1.1 Notation and terminology

Capital letters will indicate matrices; lower-case letters will indicate vectors; and Greek letters will indicate scalars. Let $\Re_n$ and $\mathcal{S}_n$ denote the Euclidian spaces of n-dimensional column vectors and $n \times n$ symmetric matrices, respectively. $\Re_n$ is endowed with the usual inner product $\langle x, y\rangle := x^T y$, and $\mathcal{S}_n$ is endowed with the trace inner product $\langle X, Y\rangle := \text{trace}(XY)$. Regarding matrix concatenation, a comma "," indicates horizontal concatenation, while a semicolon ";" indicates vertical concatentation.

We use $X \succeq 0$ to represent that X is positive semidefinite. We use e to represent a vector with all entries 1, and the vector e_j has j-th entry 1 and all other entries 0. Also, I denotes the identity matrix. Dimensions of these vectors and matrices will typically be clear from the context.

For a general, closed convex set S, we use $\text{int}(S)$ to denote the set of interior points of S and $\text{relint}(S)$ to denote the set of relative interior points. The boundary of S is written $\text{bd}(S)$. For the polyhedral cone P, we use $\text{relint}_>(P)$ to denote the set $\{x : Ax = 0, Bx > 0\}$.

2 Properties of the Cone of Quadratic Forms

$\mathcal{C} := \mathcal{C}(P)$ is closed by [29, Lemma 1] and pointed because the positive semidefinite cone contains it. The dual cone of $\mathcal{C}$ is defined as usual:

$$\mathcal{C}^* := \mathcal{C}(P)^* := \{Q : \langle Q, X\rangle \geq 0 \ \forall\ X \in \mathcal{C}\} = \left\{Q : x^T Q x \geq 0 \ \forall\ x \in P\right\}.$$

Generally, even testing whether a matrix is in $\mathcal{C}^*$ is co-NP-complete [23]. Some algorithms have been developed for this aim. For example, see [5] and reference therein.

Since $\mathcal{C}$ is closed and pointed, $\mathcal{C}^*$ has nonempty interior. Note that matrices in $\mathcal{C}^*$ correspond to the homogeneous quadratic functions, which are nonnegative over P. This property is sometimes called *copositivity over P*. $\mathcal{C}^*$ may also be interpreted as the convex cone of all valid inequalities for $\mathcal{C}$. These alternative viewpoints of $\mathcal{C}^*$ will be be used interchangeably throughout Section 3.

We first prove a claim mentioned in the Introduction.

Theorem 1. $\nu_* = \inf\{\langle Q, X\rangle : X \in \mathcal{C}, X_{11} = 1\}$.

Proof. Let ρ_* be the optimal value of the right-hand-side optimization problem. By standard techniques, the right-hand side is a relaxation of (1), and so $\nu_* \geq \rho_*$. If $\nu_* = -\infty$, then $\rho_* = -\infty$, and the result follows.

So assume ν_* is finite. Then $\langle x, Qx\rangle \geq 0$ for all $x \in P$ with $x_1 = 0$. Otherwise, any $x \in P$ with $x_1 = 0$ and $\langle x, Qx\rangle < 0$ would be a negative recession direction for (1).

Now, to prove $\nu_* \leq \rho_*$, let X be any feasible solution of the right-hand side. Because $X \in \mathcal{C}$, there exists a finite set $\{x_k\} \subset P$ such that $X = \sum_k x_k x_k^T$. We further break $\{x_k\}$ into two groups: those for which $[x_k]_1 > 0$ and those for which $[x_k]_1 = 0$. By scaling and the fact that $X_{11} = 1$, we may write

$$X = \sum_{k\,:\,[x_k]_1>0} \lambda_k \begin{pmatrix} 1 \\ \tilde{x}_k \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{x}_k \end{pmatrix}^T + \sum_{k\,:\,[x_k]_1=0} x_k x_k^T,$$

where $\lambda_k > 0$ and $\sum_{k\,:\,[x_k]_1>0} \lambda_k = 1$. So, by the definition of ν_* and the preceding paragraph,

$$\langle Q, X\rangle \geq \sum_{k\,:\,[x_k]_1>0} \lambda_k \nu_* + \sum_{k\,:\,[x_k]_1=0} 0 = \nu_*.$$

This implies $\rho_* \geq \nu_*$, as desired. □

We make three assumptions on P to simplify the presentation in the paper, the first two of which have been stated in the Introduction:

Assumption 1. *The set $P \cap \{x : x_1 = 1\}$ is nonempty.*

Assumption 2. *P is pointed, i.e., $\{x : Ax = 0, Bx = 0\} = \{0\}$*

Assumption 3. *The set* $\operatorname{relint}_>(P)$ *is nonempty, i.e., there exists $x^0 \in P$ such that $Bx^0 > 0$.*

Assumption 1 ensures feasibility of (1). Assumption 2 implies in particular that $\operatorname{rank}([A; B]) = n$. Together, Assumptions 2 and 3 imply that the slice $P \cap \{x : \langle e, Bx\rangle = 1\}$ is nonempty and bounded.

Note that Assumption 3 can in fact be made without loss of generality. If P does not satisfy Assumption 3, then we claim there exists a row b_i^T of B such that $\langle b_i, x\rangle = 0$ for all

$x \in P$, in which case $\langle b_i, x\rangle \geq 0$ can be moved into $Ax = 0$ and the process repeated until P has interior. The claim is true by the following proof of the contrapositive. Suppose B has m rows, and for each row b_j^T, pick $x^j \in P$ such that $\langle b_j, x^j\rangle > 0$. Then $x^0 := \frac{1}{m}\sum_{j=1}^m x^j$ satisfies $Bx^0 > 0$.

The above assumptions also imply that $\mathcal{C}$ and $\mathcal{D}$ have the same dimension.

Proposition 1. $\dim(\mathcal{C}) = \dim(\mathcal{D})$. *Furthermore, if $E \subset P$ contains the extreme rays of P, then* $\dim(\mathcal{C}(E)) = \dim(\mathcal{D})$.

Proof. We first consider the case when A is empty and claim that $\mathcal{C}$ is full-dimensional in $\mathcal{S}_n$. It suffices to find $n(n+1)/2$ linearly independent elements in $\mathcal{C}$. Let $\{r_j\}$ be the normalized extreme rays of P. By Assumptions 2 and 3, there are at least n which are linearly independent. We claim that $\{(r_j+r_k)(r_j+r_k)^T : j \leq k\}$ are the desired independent elements. Clearly, all are in $\mathcal{C}$. To see that they are also independent, we first note that, without loss of generality, we may pre- and post-multiply by a change-of-basis matrix to transform to the matrices $\{(e_j+e_k)(e_j+e_k)^T : j \leq k\}$, where e_j is the standard unit vector in $\Re_n$. It is then not difficult to verify that the matrices are linearly indepdendent. Since $\mathcal{C} \subseteq \mathcal{D}$ and $\mathcal{C}$ is full-dimensional, then $\mathcal{D}$ is full-dimensional also.

Now we consider the case when A is non-trivial. Let N be a matrix whose columns span $\{x : Ax = 0\}$, and define $R := \{y : BNy \geq 0\}$. Then $P = NR$, $\mathcal{C} = N\mathcal{C}(R)N^T$, and $\mathcal{D} = N\mathcal{D}(R)N^T$. Moreover, R is pointed and has interior in its own space due to Assumptions 2 and 3. Then, by the case when A is empty, $\dim(\mathcal{C}(R)) = \dim(\mathcal{D}(R))$, which implies the result.

The previous argument can be repeated to show $\dim(\mathcal{C}(E)) = \dim(\mathcal{D})$ if $E \subset \mathcal{P}$ contains the extreme rays of P. $\square$

For $X \in \mathcal{D}$, the conditions $X \succeq 0$ and $AXA^T = 0$ guarantee $AX = 0$, which clearly implies $AXB^T = 0$. Hence, $\mathcal{D}$ may also be written as

$$\mathcal{D} = \left\{X \succeq 0 : AX = 0,\ BXB^T \geq 0\right\}.$$

Additional properties of $\mathcal{D}$ are established in the following lemmas:

Lemma 1. *Let $X \in \mathcal{D}$. The following are equivalent: $BXB^T = 0$; $BX = 0$; $X = 0$.*

Proof. Assume $BXB^T = 0$, and let v be any column of XB^T. From $BXB^T = 0$ and $AXB^T = 0$, it clearly holds that $v \in \{x : Ax = 0, Bx = 0\}$, and so $v = 0$ by Assumption 2. Hence, $XB^T = 0$ or $BX = 0$. Now, since $AX = 0$ and $BX = 0$, the columns of X are in $\{x : Ax = 0, Bx = 0\}$, which implies $X = 0$. The reverse direction $X = 0 \Rightarrow BX = 0 \Rightarrow BXB^T = 0$ is obvious. $\square$

Lemma 2. *Let $X \in \mathcal{D}$, and define $x := XB^T e$. Then $x \in P$, and $X \neq 0$ implies $x \neq 0$. Furthermore, if $\langle B^T ee^T B, X\rangle = 1$, then $X - xx^T \succeq 0$ and $\langle B^T ee^T B, X - xx^T\rangle = 0$.*

Proof. Because $X \in \mathcal{D}$, one readily sees $Ax = AXB^T e = 0$ and $Bx = BXB^T e \geq 0$. So $x \in P$. If $X \neq 0$, then by Lemma 1, $0 \neq BXB^T \geq 0$. Hence

$$0 < \langle ee^T, BXB^T\rangle = \langle e, BXB^T e\rangle = \langle e, Bx\rangle,$$

which implies $x \neq 0$. Also, the equation

$$\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} = \begin{pmatrix} 1 & e^T BX \\ XB^T e & X \end{pmatrix} = \begin{pmatrix} e^T B \\ I \end{pmatrix} X \begin{pmatrix} B^T e & I \end{pmatrix}$$

and the Schur-complement theorem imply $X - xx^T \succeq 0$. Next, by using $\langle B^T ee^T B, X\rangle = 1$,

$$\begin{aligned} B^T ee^T B \left(X - xx^T\right) &= B^T ee^T BX - B^T ee^T Bxx^T \\ &= B^T ex^T - \langle e, Bx\rangle B^T ex^T \\ &= (1 - \langle e, Bx\rangle) B^T ex^T \\ &= 0. \end{aligned}$$

So $\langle B^T ee^T B, X - xx^T\rangle = 0$. □

3 Separation and Relaxation

In Section 3.1, we describe a separation algorithm for $\mathcal{C}$ under the assumption that one can optimize over a cone related to the boundary of P. Then in Section 3.2, we describe a nonlinear formulation of $\mathcal{C}$ by studying the dual of the optimization problem introduced in the separation algorithm. Finally, in Section 3.3, we discuss a new class of related convex relaxations of $\mathcal{C}$, which strengthen $\mathcal{D}$, and in Section 3.4, we extend this to a tractable hierarchy of relaxations, which provide even better approximations of $\mathcal{C}$.

3.1 A separation algorithm

In this subsection, we establish a separation algorithm for $\mathcal{C}$. The key idea is to consider matrices in $\mathcal{C}$ that are generated only from $x \in \text{bd}(P)$:

$$\mathcal{C}_{\text{bd}} := \mathcal{C}(\text{bd}(P)) \subseteq \mathcal{C}.$$

$\mathcal{C}_{\text{bd}}$ is closed by [29, Lemma 1] and pointed. $\mathcal{C}_{\text{bd}}^*$ denotes the dual cone of $\mathcal{C}_{\text{bd}}$, and $\text{int}(\mathcal{C}_{\text{bd}}^*) \neq \emptyset$. In fact,

$$\text{int}(\mathcal{C}_{\text{bd}}^*) \supseteq \text{int}(\mathcal{C}^*) = \{Q : \langle x, Qx \rangle > 0 \;\; \forall \;\; 0 \neq x \in P\}.$$

We may obtain alternative characterizations of $\mathcal{C}_{\text{bd}}$ and $\mathcal{C}_{\text{bd}}^*$ by breaking $\text{bd}(P)$ into pieces. Let b_i^T denote the i-th row of B, and let B_i denote the matrix gotten from B by deleting b_i^T. Define P_i to be the polyhedron resulting from P when the inequality $\langle b_i, x \rangle \geq 0$ is set to equality, i.e.,

$$P_i := \{x : Ax = 0, \; \langle b_i, x \rangle = 0, \; B_i x \geq 0\}.$$

Then $\text{bd}(P) = \cup_i P_i$. Defining $\mathcal{C}_i := \mathcal{C}(P_i)$ in analogy with (2), we see

$$\mathcal{C}_{\text{bd}} = \sum_i \mathcal{C}_i, \qquad \mathcal{C}_{\text{bd}}^* = \bigcap_i \mathcal{C}_i^*.$$

The inclusion $\mathcal{C}_{\text{bd}} \subseteq \mathcal{C}$ implies $\mathcal{C}_{\text{bd}}^* \supseteq \mathcal{C}^*$. The following important lemma provides conditions under which a matrix in $\mathcal{C}_{\text{bd}}^*$ is actually in $\mathcal{C}^*$.

Lemma 3. *Let $Q \in \mathcal{C}_{\text{bd}}^*$. If Q is not positive semidefinite on the linear subspace $\{x : Ax = 0, \langle e, Bx \rangle = 0\}$, then $Q \in \mathcal{C}^*$.*

Proof. To show $Q \in \mathcal{C}^*$, we must demonstrate that the quadratic function $q(x) := \langle x, Qx \rangle$ is nonnegative for all $x \in P$. By homogeneity, it suffices to show $q(x)$ is nonnegative over the nonempty, bounded slice $\hat{P} := P \cap \{x : \langle e, Bx \rangle = 1\}$.

We first claim that the minimum value of $q(x)$ over $\hat{P}$ is actually attained on $\text{bd}(\hat{P})$. To see this, let $x \in \text{relint}(\hat{P})$ be arbitrary, and let $d \in \{x : Ax = 0, \langle e, Bx \rangle = 0\}$ satisfy $\langle d, Qd \rangle < 0$, which exists by assumption. Also, ensure that d satisfies $\langle d, Qx \rangle \leq 0$; if not, simply replace d by $-d$. Then, for small $\epsilon > 0$, $x + \epsilon d$ is feasible, and

$$\langle x + \epsilon d, Q(x + \epsilon d) \rangle = \langle x, Qx \rangle + 2\epsilon \langle d, Qx \rangle + \epsilon^2 \langle d, Qd \rangle < \langle x, Qx \rangle.$$

Hence, x is not a global minimum, which proves the claim.

Now, since $Q \in \mathcal{C}_{\text{bd}}^*$, we know $q(x) \geq 0$ for all $x \in \text{bd}(\hat{P})$. By the preceding paragraph, this shows that $q(x)$ is nonnegative for all $x \in \hat{P}$, as desired. □

The separation procedure is stated as Algorithm 1. Since membership in $\mathcal{D}$ is necessary for membership in $\mathcal{C}$, Step 1 first separates $\bar{X}$ over $\mathcal{D}$. Step 2 then constitutes the main work of the algorithm. The idea of the optimization (4) is to minimize the inner product $\langle Q, \bar{X} \rangle$ over $Q \in \mathcal{C}_{\text{bd}}^*$ in order to separate based on the sign of κ. We first establish that κ is finite:

Lemma 4. *If $\bar{X} \in \mathcal{D}$, then the optimal value κ of (4) is finite.*

Algorithm 1 Separation over $\mathcal{C}$

Input: $\bar{X} \in \mathcal{S}_n$.

Output: YES if $\bar{X} \in \mathcal{C}$; otherwise, NO and $\bar{Q} \in \mathcal{C}^*$ separating $\bar{X}$ from $\mathcal{C}$.

1: If $\bar{X} \notin \mathcal{D}$, then return NO and $\bar{Q} \in \mathcal{D}^*$ separating $\bar{X}$ from $\mathcal{D}$.

2: For chosen $X^0 \in \text{relint}(\mathcal{C}_{\text{bd}})$, calculate the optimal value κ and an optimal solution $\bar{Q}$ of

$$\min_{Q \in \mathcal{S}_n} \left\{ \langle Q, \bar{X} \rangle : \begin{array}{c} Q \in \mathcal{C}^*_{\text{bd}}, \ \langle Q, X^0 \rangle \leq 1 \\ \langle \bar{X} B^T e, Q \bar{X} B^T e \rangle \geq 0 \end{array} \right\}. \tag{4}$$

3: If $\kappa \geq 0$, then return YES.

4: If $\kappa < 0$, then return NO and $\bar{Q}$.

Proof. We show that $\inf\{\langle Q, \bar{X} \rangle : Q \in \mathcal{C}^*_{\text{bd}}, \langle Q, X^0 \rangle \leq 1\}$ is finite or equivalently that the system

$$0 \neq \Delta Q \in \mathcal{C}^*_{\text{bd}}, \quad \langle \Delta Q, X^0 \rangle = 0, \quad \langle \Delta Q, \bar{X} \rangle < 0$$

is inconsistent. Suppose ΔQ satisfies the first two conditions. Since $X^0 \in \text{relint}(\mathcal{C}_{\text{bd}})$, it holds that $\langle \Delta Q, X \rangle = 0$ for all $X \in \text{span}(\mathcal{C}_{\text{bd}})$. By the second half of Proposition 1, we know $\text{span}(\mathcal{C}_{\text{bd}}) = \text{span}(\mathcal{D})$, and so $\bar{X} \in \text{span}(\mathcal{C}_{\text{bd}})$, which implies $\langle \Delta Q, \bar{X} \rangle = 0$. So the systm is inconsistent. □

It is worth noting that, while κ is finite, the feasible region of (4) is not bounded if A is non-trivial. For example, $A^T A$ is a direction of recession. However, by the discussion within Proposition 1, one can prove that all the unbounded directions are in $\text{span}(\mathcal{D})^\perp$, which equals $\{X \in \mathcal{S}_n : AX = 0\}^\perp$. Hence, those directions do not contribute to the objective $\langle Q, \bar{X} \rangle$ since $\bar{X} \in \mathcal{D}$. So, one could adjust (4) to have a bounded feasbile region without affecting the forthcoming proof of correctness of Algorithm 1 by simply constraining Q within the linear subspace $\text{span}(\mathcal{D})$. This equivalence with a bounded problem also shows that the minimum in (4) is actually attained, so that $\bar{Q}$ is well-defined.

Because $\mathcal{C}^*_{\text{bd}} \supseteq \mathcal{C}^*$, there is still a danger that Algorithm 1 reaches Step 4 and returns $\bar{Q} \notin \mathcal{C}^*$, rendering the algorithm incorrect. However, the following theorem proves this cannot happen; the critical result is Lemma 3.

Theorem 2. *Algorithm 1 correctly solves the separation problem for $\mathcal{C}$.*

Proof. If Algorithm 1 terminates in Step 1, then clearly $\bar{X} \notin \mathcal{C}$ and $\bar{Q} \in \mathcal{C}^*$ with $\langle \bar{Q}, \bar{X} \rangle < 0$. So assume the algorithm has reached Step 2. Then $\bar{X} \in \mathcal{D}$, and the optimal value κ in (4) is finite by Lemma 4. Define $\bar{x} := \bar{X} B^T e$; by Lemma 2, $\bar{x} \in P$.

Suppose $\kappa \geq 0$, causing the algorithm to terminate in Step 3. We claim every (suitably scaled) $Q \in \mathcal{C}^*$ is feasible for (4), which will establish $\bar{X} \in \mathcal{C}$. We already know $Q \in \mathcal{C}^*_{\text{bd}}$. We need to show $\langle \bar{x}, Q\bar{x}\rangle \geq 0$, which indeed holds since $\bar{x} \in P$.

Finally suppose $\kappa < 0$, causing termination in Step 4. In particular, $\bar{X} \neq 0$, and so by Lemma 1, $B\bar{X}B^T \neq 0$. Hence, without loss of generality, we scale $\bar{X}$ so that $\langle B^T ee^T B, \bar{X}\rangle = 1$. By Lemma 2, this implies also $\bar{X} - \bar{x}\bar{x}^T \succeq 0$ and $\langle B^T ee^T B, \bar{X} - xx^T\rangle = 0$.

We claim $\bar{Q} \in \mathcal{C}^*$, which will complete the proof. By Lemma 3, it suffices to show $\bar{Q}$ is not positive semidefinite on the linear subspace $\{x : Ax = 0, \langle e, Bx\rangle = 0\}$. To obtain a contradiction, suppose $\bar{Q}$ is positive semidefinite on this subspace. Then there exists $t \geq 0$ such that $\bar{Q} + t(A^T A + B^T ee^T B) \succeq 0$. We have

$$\begin{aligned}
0 &> \langle \bar{Q}, \bar{X}\rangle = \langle \bar{Q}, \bar{X} - \bar{x}\bar{x}^T\rangle + \langle \bar{x}, \bar{Q}\bar{x}\rangle \\
&= \langle \bar{Q} + t(A^T A + B^T ee^T B), \bar{X} - \bar{x}\bar{x}^T\rangle + \langle \bar{x}, \bar{Q}\bar{x}\rangle - t\langle A^T A + B^T ee^T B, \bar{X} - \bar{x}\bar{x}^T\rangle \\
&\geq \langle \bar{x}, \bar{Q}\bar{x}\rangle - t\langle A^T A + B^T ee^T B, \bar{X} - \bar{x}\bar{x}^T\rangle \\
&= \langle \bar{x}, \bar{Q}\bar{x}\rangle - t\langle B^T ee^T B, \bar{X} - \bar{x}\bar{x}^T\rangle \\
&= \langle \bar{x}, \bar{Q}\bar{x}\rangle,
\end{aligned}$$

which contradicts the feasibility of $\bar{Q}$ in (4). □

3.2 A nonlinear representation

We now examine the conic dual of the optimization problem (4), which appears within Algorithm 1. Defining $\bar{x} := \bar{X}B^T e$, the dual is

$$\begin{aligned}
\max \quad & \rho \\
\text{s.t.} \quad & \bar{X} = \rho X^0 + \lambda\, \bar{x}\bar{x}^T + Z \\
& \rho \leq 0,\ \lambda \geq 0,\ Z \in \mathcal{C}_{\text{bd}}
\end{aligned} \tag{5}$$

For non-trivial instances of (4) arising within Algorithm 1, i.e., when $0 \neq \bar{X} \in \mathcal{D}$, Lemma 2 implies $\bar{x} \neq 0$. Also, since $\text{int}(\mathcal{C}^*_{\text{bd}}) \supseteq \text{int}(\mathcal{C}^*) \neq \emptyset$ and $0 \neq \bar{x}\bar{x}^T \in \mathcal{C}$, there exists $Q \in \text{int}(\mathcal{C}^*)$ such that $\langle Q, \bar{x}\bar{x}^T\rangle > 0$ and $\langle Q, X^0\rangle < 1$. Therefore we know that (4) has non-empty interior. Hence, strong duality holds between (4) and (5), and the dual optimum in (5) is always attained.

This provides a nonlinear representation of $\mathcal{C}$:

Proposition 2. *$X \in \mathcal{C}$ if and only if $X \in \mathcal{D}$ and there exist $\lambda \geq 0$ and $Z \in \mathcal{C}_{\text{bd}}$ such that*

$$X = \lambda(XB^T e)(XB^T e)^T + Z \tag{6}$$

Proof. (⇒) Let $\bar{X}$ in (5) be the given X. Since Algorithm 1 terminates with $\kappa \geq 0$, by strong duality (5) has a feasible solution (ρ, λ, Z) with $\rho = 0$. This gives (6).

(⇐) Since $X \in \mathcal{D}$, $XB^Te \in P$ by Lemma 2, and so $(XB^Te)(XB^Te)^T \in \mathcal{C}$. Further, $Z \in \mathcal{C}_{bd} \subseteq \mathcal{C}$. Because X satisfies (6) with $\lambda \geq 0$, we see $X \in \mathcal{C}$. □

As a representation result, (6) is quite interesting since it demonstrates that the structure of $\mathcal{C}$ depends heavily on the structure of $\mathcal{C}_{\text{bd}}$. Algorithmically, however, the nonlinearity seems to preclude explicit computation except when X is fixed, which is the case within Algorithm 1. In the next subsection, we explore a similar, but different, representation of $\mathcal{C}$ that is amenable to computation. Still, the following proposition shows that, under a certain simple condition, the nonlinear representation reduces to a linear one:

Proposition 3. *Let $X \in \mathcal{C}$, and suppose BXB^T has a zero entry. Define $x := XB^Te \in P$ in accordance with Lemma 2. It holds that:*

(i) if $x \in \text{bd}(P)$, then $\lambda = 0$ in some representation (6) of X;

(ii) if $x \in \text{relint}_>(P)$, then $\lambda = 0$ in all representations (6) of X.

Proof. Write (6) as $X = \lambda xx^T + Z$. For (i), since $x \in \text{bd}(P)$, we have $xx^T \in \mathcal{C}_{\text{bd}}$. Hence, λxx^T can be subsumed into Z, yielding a new representation of X with $\lambda = 0$. For (ii), pre- and post-multiply $X = \lambda xx^T + Z$ by B and B^T, respectively, to yield

$$BXB^T = \lambda(Bx)(Bx)^T + BZB^T.$$

Because $Bx > 0$ and $BZB^T \geq 0$, the zero entry of BXB^T ensures $\lambda = 0$. □

Another interpretation of Proposition 3 is that $\mathcal{C}_{\text{bd}}$, which is an inner approximation of $\mathcal{C}$, is actually exact on certain faces of $\mathcal{C}$.

3.3 A class of convex relaxations

Motivated by the nonlinear representation (6) of $\mathcal{C}$ in the previous subsection, we now explore a new class of relaxations for $\mathcal{C}$, each of which is at least as strong as $\mathcal{D}$. The derivation of the relaxations are different than the derivation of (6), but they share similar structural elements.

Let a "step direction" $d \in P$ be fixed, and define

$$\mathcal{C}(d) := \mathcal{C}(P; d) := \left\{ X = \zeta dd^T + dz^T + zd^T + Z \ : \ \begin{array}{c} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \succeq 0 \\ z \in P, \ Z \in \mathcal{C}_{\text{bd}} \end{array} \right\}.$$

Proposition 4. *$\mathcal{C} \subseteq \mathcal{C}(d) \subseteq \mathcal{D}$ for all nonzero $d \in P$.*

Proof. Let $X \in \mathcal{C}(d)$. It is easy to verify that the equation

$$X = \begin{pmatrix} d & I \end{pmatrix} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \begin{pmatrix} d^T \\ I \end{pmatrix}$$

holds and implies $X \succeq 0$. Also $AX = 0$ and $BXB^T \geq 0$ hold because $d, z \in P$ and $Z \in \mathcal{C}_{\text{bd}} \subseteq \mathcal{C}$. So $\mathcal{C}(d) \subseteq \mathcal{D}$.

Next, we show that every extreme ray of $\mathcal{C}$ is in $\mathcal{C}(d)$, which will prove $\mathcal{C} \subseteq \mathcal{C}(d)$. Note every extreme ray of $\mathcal{C}$ has form $X := xx^T$ with $x \in P$ [29, Lemma 1]. Let $\alpha \geq 0$ be the smallest step-size such that $x - \alpha d \in \text{bd}(P)$. Note that α is well-defined because $-d \notin P$ due to Assumption 2. Then define

$$\begin{aligned} \zeta &:= \alpha^2 \\ z &:= \alpha(x - \alpha d) \in P \\ Z &:= (x - \alpha d)(x - \alpha d)^T \in \mathcal{C}_{\text{bd}} \end{aligned}$$

so that

$$\begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} = \begin{pmatrix} \alpha \\ x - \alpha d \end{pmatrix} \begin{pmatrix} \alpha \\ x - \alpha d \end{pmatrix}^T \succeq 0.$$

So $xx^T = \zeta dd^T + dz^T + zd^T + Z \in \mathcal{C}(d)$. □

The next proposition, which is analogous to Proposition 3, shows that if $d \in \text{relint}_>(P)$, then a zero entry in BXB^T forces $\zeta = 0$ and $z = 0$ in $\mathcal{C}(d)$.

Proposition 5. *Let $d \in \text{relint}_>(P)$, and let $X \in \mathcal{C}(d)$ be given with associated (ζ, z, Z). Suppose BXB^T has a zero entry. Then $\zeta = 0$ and $z = 0$.*

Proof. Pre- and post-multiplying $X = \zeta dd^T + dz^T + zd^T + Z$ by B and B^T, respectively, yields

$$BXB^T = \zeta(Bd)(Bd)^T + (Bd)(Bz)^T + (Bz)(Bd)^T + BZB^T,$$

Since $Bd > 0$, $Bz \geq 0$ and $BZB^T \geq 0$, the zero entry of BXB^T ensures $\zeta = 0$, which in turn forces $z = 0$ by positive semidefiniteness of $(\zeta, z^T; z, Z)$. □

We next show that, if we allow d to vary, the intersection of all $\mathcal{C}(d)$ captures $\mathcal{C}$ exactly.

Proposition 6. *$\mathcal{C} = \cap_{d \in P} \mathcal{C}(d)$.*

Proof. Since $\mathcal{C} \subseteq \mathcal{C}(d) \subseteq \mathcal{D}$, it suffices to show that, for any $X \in \mathcal{D} \setminus \mathcal{C}$, there exists $d \in P$ such that $X \notin \mathcal{C}(d)$. To do so, let $Q \in \mathcal{C}^*$ separate X with $\langle Q, X \rangle < 0$, and let d be a global minimizer of $\langle x, Qx \rangle$ over the nonempty, bounded slice $P \cap \{x : \langle e, Bx \rangle = 1\}$. We know $\delta := \langle d, Qd \rangle \geq 0$ since $Q \in \mathcal{C}^*$. In fact, we may assume without loss of generality that $\delta = 0$; otherwise, we may replace Q by the shifted $Q - \delta B^T ee^T B$. This shift ensures that the new Q remains in $\mathcal{C}^*$ because, for all $x \in P \cap \{x : \langle e, Bx \rangle = 1\}$,

$$\langle Q - \delta B^T ee^T B, xx^T \rangle = \langle Q, xx^T \rangle - \delta \geq 0.$$

Also, the new Q still satisfies $\langle Q, X \rangle < 0$ because $X \in \mathcal{D}$ implies

$$\langle Q - \delta B^T ee^T B, X \rangle \leq \langle Q, X \rangle < 0.$$

We claim $X \notin \mathcal{C}(d)$. Assuming otherwise, let (ζ, z, Z) be associated with X in $\mathcal{C}(d)$. Since $Z \in \mathcal{C}_{\text{bd}} \subseteq \mathcal{C}$, it holds that $\langle Q, Z \rangle \geq 0$. Hence,

$$0 > \langle Q, X \rangle = \langle Q, \zeta dd^T + dz^T + zd^T + Z \rangle = \zeta\delta + 2\langle Q, dz^T \rangle + \langle Q, Z \rangle \geq 2\langle Q, dz^T \rangle.$$

On the other hand, $d + \rho z \in P$ for all $\rho \geq 0$, and so

$$0 \leq \langle Q, (d + \rho z)(d + \rho z)^T \rangle = 2\rho\langle Q, dz^T \rangle + \rho^2 \langle Q, zz^T \rangle.$$

In particular, for $\rho > 0$, $\langle Q, dz^T \rangle \geq -\rho\langle Q, zz^T \rangle$. Taking $\rho \to 0$, we see $\langle Q, dz^T \rangle \geq 0$, a contradiction. □

Notice that $\mathcal{C}(d) = \mathcal{C}(\kappa d)$ for all $\kappa > 0$, and so, without loss of generality we may assume $\langle e, Bd \rangle = 1$. Also, in the context of the preceding proof, a similar argument as in Theorem 2 shows Q is not positive semidefinite on the linear subspace $\{x : Ax = 0, \langle e, Bx \rangle = 0\}$. This implies that Proposition 6 still holds even if d is taken only over $\text{bd}(P)$.

We now heuristically compare the equation $X = \zeta dd^T + dz^T + zd^T + Z$ for $X \in \mathcal{C}(d)$ with the nonlinear representation $X = \lambda(XB^T e)(XB^T e)^T + Z$ for $X \in \mathcal{C}_n$ as depicted in (6) of the previous subsection. The two representations share $Z \in \mathcal{C}_{\text{bd}}$, and for given $d \in P$, the rank-3, linear term $\zeta dd^T + dz^T + zd^T$ functions as an approximation of the rank-1, nonlinear term $\lambda(XB^T e)(XB^T e)^T$. By allowing d to vary over P, Proposition 6 shows that, in some sense, the rank-3, linear term is enough to enforce the rank-1, nonlinear term.

The two representations share an additional characteristic. Recall Proposition 3 showed that the nonlinear representation of $X \in \mathcal{C}$ simplifies to a linear one when BXB^T is known to have a zero component. In particular, the rank-1, nonlinear term $\lambda(XB^T e)(XB^T e)^T$

vanishes. For $\mathcal{C}(d)$ with $d \in \text{relint}_>(P)$, as shown in Proposition 5, the rank-3, linear term vanishes under the same condition.

A disadvantage of $\mathcal{C}(d)$ is that is generally intractable due to its dependence on $\mathcal{C}_{\text{bd}}$. However, using the equation $\mathcal{C}_{\text{bd}} = \sum_i \mathcal{C}_i$, we can relax $\mathcal{C}_i$ to $\mathcal{D}_i := \mathcal{D}(P_i)$ in $\mathcal{C}(d)$—where $\mathcal{D}(P_i)$ is defined in analogy with (3), respecting the fixed equality $\langle b_i, x \rangle = 0$—to obtain the following tractable relaxation:

$$\mathcal{D}(d) := \mathcal{D}(P; d) := \left\{ X = \zeta dd^T + dz^T + zd^T + Z \ : \ \begin{array}{c} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \succeq 0 \\ z \in P, \ Z \in \sum_i \mathcal{D}_i \end{array} \right\}.$$

Proposition 5 extends to $\mathcal{D}(d)$ by the same proof:

Proposition 7. *Let $d \in \text{relint}_>(P)$, and let $X \in \mathcal{D}(d)$ be given with associated (ζ, z, Z). Suppose BXB^T has a zero entry. Then $\zeta = 0$ and $z = 0$.*

In addition, $\mathcal{D}(d)$ is at least as strong as $\mathcal{D}$.

Theorem 3. *$\mathcal{D}(d) \subseteq \mathcal{D}$ for all nonzero $d \in P$.*

Proof. Let $X = \zeta dd^T + dz^T + zd^T + Z \in \mathcal{D}(d)$. We need to show that X satisfies all constraints of $\mathcal{D}$.

We first claim $X \succeq 0$. If $\zeta = 0$, then $z = 0$ and $X = Z \succeq 0$. If $\zeta > 0$, we have

$$\begin{aligned} X &= \zeta dd^T + dz^T + zd^T + \zeta^{-1} zz^T + Z - \zeta^{-1} zz^T \\ &= (\sqrt{\zeta} d + \sqrt{\zeta^{-1}} z)(\sqrt{\zeta} d + \sqrt{\zeta^{-1}} z)^T + Z - \zeta^{-1} zz^T, \end{aligned}$$

which expresses X as the sum of two semidefinite matrices, which implies $X \succeq 0$.

It remains to show that X satisfies all the linear constraints of $\mathcal{D}$. Since $P_i \subseteq P$ for all i, $\mathcal{D}_i \subseteq \mathcal{D}$ for all i. This implies Z satisfies all constraints. Since $d, z \in P$, we also see that dd^T and $dz^T + zd^T$ satisfy all constraints, which proves the result. □

Combining Proposition 4 and Theorem 3, we have the inclusions $\mathcal{C} \subseteq \mathcal{C}(d) \subseteq \mathcal{D}(d) \subseteq \mathcal{D}$. We were unable to prove or disprove whether $\cap_{d \in P} \mathcal{D}(d)$ equals $\mathcal{C}$.

3.4 A recursive hierarchy of convex relaxations

The ideas of the previous subsection can be extended to obtain a recursive hierarchy of convex relaxations for $\mathcal{C}$. In contrast to other hierarchies such as the one proposed by Parrilo [25] for completely positive matrices (see the discussion in Section 4.1), the number of relaxations in our hierarchy is finite. However, the hierarchy depends on a specific choice of vectors in P. In particular, a different choice of vectors yields a different finite hierarchy. In this sense, we actually propose an infinite family of finite hierarchies.

As it will turn out, $\mathcal{C}(d) := \mathcal{C}(P; d)$ and $\mathcal{D}(d) := \mathcal{D}(P; d)$ of the previous subsection will constitute the first level of the hierarchy for a specific choice $d \in P$. In order to define the higher levels of the hierarchy, we first need to introduce some definitions and notation.

Let m denote the number of inequalities in $Bx \geq 0$. For any $I \subseteq \{1, \ldots, m\}$, define

$$P_I := P \cap \{x : b_i^T x = 0 \ \forall \ i \in I\}$$

to be the face of P having all inequalities $\langle b_i, x\rangle \geq 0$, $i \in I$, set to equality. For example, $P_\emptyset = P$. Defining $J := \{1, \ldots, m\} \setminus I$, we write

$$P_I = \{x : A_I x = 0, B_J x \geq 0\},$$

where: A_I consists of the rows of A combined with the rows b_i^T of B for $i \in I$; and B_J consists of the remaining rows of B, i.e., b_j^T such that $j \in J$.

As mentioned above, the hierarchy will be defined in terms of a specific choice of vectors in P. Specifically, for all $I \subseteq \{1, \ldots, m\}$ with $|I| < m$, choose $v_I \in P_I$, and let $\mathcal{V}$ denote the collection of all v_I. If $P_I = \emptyset$, we consider the corresponding v_I to be non-existent in $\mathcal{V}$. The collection $\mathcal{V}$ will then serve as the basis for the hierarchy.

We next define a convex cone recursively in terms of P_I, $\mathcal{V}$, and any integer t satisfying $0 \leq t \leq |J|$. For notational convenience, let $P_{Ij} := P_{I \cup \{j\}}$. The convex cone is

$$\mathcal{C}^{(t)}(P_I; \mathcal{V}) := \left\{ X = \zeta v_I v_I^T + v_I z^T + z v_I^T + Z \ : \ \begin{array}{c} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \succeq 0 \\ z \in P_I, \ Z \in \sum_{j \in J} \mathcal{C}^{(t-1)}(P_{Ij}; \mathcal{V}) \end{array} \right\},$$

where the base-case of the recursion is defined as follows:

$$\mathcal{C}^{(0)}(P_\bullet; \mathcal{V}) := \mathcal{C}(P_\bullet),$$

i.e., the "zero-th" cone is simply the cone of completely positive matrices over the given argument. For example, consider $P_\emptyset = P$ with chosen $\mathcal{V}$, and define $P_j := P_{\{j\}}$. Then $t = 0$

yields $\mathcal{C}^{(0)}(P, \mathcal{V}) = \mathcal{C}(P)$, and $t = 1$ gives

$$\mathcal{C}^{(1)}(P; \mathcal{V}) := \left\{ X = \zeta v_\emptyset v_\emptyset^T + v_\emptyset z^T + z v_\emptyset^T + Z \ : \ \begin{array}{c} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \succeq 0 \\ z \in P, \ Z \in \sum_{j=1}^m \mathcal{C}(P_j) \end{array} \right\},$$

which is precisely $\mathcal{C}(P; d)$ of the previous subsection with $d = v_\emptyset$.

An important property of $\mathcal{C}^{(t)}(P_I; \mathcal{V})$ is the following inclusion relationship.

Lemma 5. *Given P_I, $\mathcal{V}$, and $0 \le t < |J|$, it holds that $\mathcal{C}^{(t)}(P_I; \mathcal{V}) \subseteq \mathcal{C}^{(t+1)}(P_I; \mathcal{V})$.*

Proof. Using a proof similar to that of Proposition 4, we can show $\mathcal{C}^{(0)}(P_\bullet; \mathcal{V}) \subseteq \mathcal{C}^{(1)}(P_\bullet; \mathcal{V})$ for any argument. So assuming $\mathcal{C}^{(t-1)}(P_\bullet; \mathcal{V}) \subseteq \mathcal{C}^{(t)}(P_\bullet; \mathcal{V})$, we proceed by induction. The only difference between the definitions of $\mathcal{C}^{(t)}(P_I; \mathcal{V})$ and $\mathcal{C}^{(t+1)}(P_I; \mathcal{V})$ is the respective constraints

$$Z \in \sum_{j \in J} \mathcal{C}^{(t-1)}(P_{Ij}; \mathcal{V}) \qquad \text{and} \qquad Z \in \sum_{j \in J} \mathcal{C}^{(t)}(P_{Ij}; \mathcal{V}).$$

By induction, $\mathcal{C}^{(t-1)}(P_{Ij}; \mathcal{V}) \subseteq \mathcal{C}^{(t)}(P_{Ij}; \mathcal{V})$, and so the second constraint is looser, which implies $\mathcal{C}^{(t)}(P_I; \mathcal{V}) \subseteq \mathcal{C}^{(t+1)}(P_I; \mathcal{V})$, as desired. □

This result establishes the following hierarchy of relaxations for $\mathcal{C}(P)$:

$$\mathcal{C}(P) = \mathcal{C}^{(0)}(P; \mathcal{V}) \subseteq \mathcal{C}^{(1)}(P; \mathcal{V}) \subseteq \cdots \subseteq \mathcal{C}^{(m)}(P; \mathcal{V}).$$

Via recursion, the relaxation $\mathcal{C}^{(t)}(P; \mathcal{V})$ is ultimately based on a number of convex cones of the generic form $\mathcal{C}(P_\bullet) = \mathcal{C}^{(0)}(P_\bullet; \mathcal{V})$. For this reason, in practice, one cannot expect to optimize efficiently over $\mathcal{C}^{(t)}(P; \mathcal{V})$ since $\mathcal{C}(P_\bullet)$ is generally intractable. However, if one further relaxes $\mathcal{C}(P_\bullet)$ to $\mathcal{D}(P_\bullet)$, then the resulting relaxation will be in fact tractable. Formally, in analogy with $\mathcal{C}^{(t)}(P; \mathcal{V})$, we define

$$\mathcal{D}^{(t)}(P_I; \mathcal{V}) := \left\{ X = \zeta v_I v_I^T + v_I z^T + z v_I^T + Z \ : \ \begin{array}{c} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \succeq 0 \\ z \in P_I, \ Z \in \sum_{j \in J} \mathcal{D}^{(t-1)}(P_{Ij}; \mathcal{V}) \end{array} \right\}$$

and

$$\mathcal{D}^{(0)}(P_\bullet; \mathcal{V}) := \mathcal{D}(P_\bullet),$$

For example, letting $v_j := v_{\{j\}}$ and $P_{jk} := P_{\{j,k\}}$ for notational convenience, we have

$$\mathcal{D}^{(2)}(P;\mathcal{V}) = \left\{ X = \zeta v_\emptyset v_\emptyset^T + v_\emptyset z^T + z v_\emptyset^T + Z \ : \ \begin{array}{c} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \succeq 0 \\ z \in P, \ Z \in \sum_{j=1}^m \mathcal{D}^{(1)}(P_j;\mathcal{V}) \end{array} \right\},$$

where

$$\mathcal{D}^{(1)}(P_j;\mathcal{V}) = \left\{ X = \zeta v_j v_j^T + v_j z^T + z v_j^T + Z \ : \ \begin{array}{c} \begin{pmatrix} \zeta & z^T \\ z & Z \end{pmatrix} \succeq 0 \\ z \in P_j, \ Z \in \sum_{k \neq j} \mathcal{D}(P_{jk}) \end{array} \right\}.$$

Just as Lemma 5 established an important inclusion property of $\mathcal{C}^{(t)}(P_I;\mathcal{V})$, a related—but different—inclusion holds for $\mathcal{D}^{(t)}(P_I;\mathcal{V})$.

Lemma 6. *Given P_I, $\mathcal{V}$, and $0 \le t < |J|$, it holds that $\mathcal{D}^{(t)}(P_I;\mathcal{V}) \supseteq \mathcal{D}^{(t+1)}(P_I;\mathcal{V})$.*

Proof. Using a proof similar to that of Theorem 3, we can show $\mathcal{D}^{(0)}(P_\bullet;\mathcal{V}) \supseteq \mathcal{D}^{(1)}(P_\bullet;\mathcal{V})$ for any $P_\bullet$. The proof now follows the proof of Lemma 5, except that all inclusions are reversed. □

This result establishes the following hierarchy of restrictions of $\mathcal{D}(P)$:

$$\mathcal{D}(P) = \mathcal{D}^{(0)}(P;\mathcal{V}) \supseteq \mathcal{D}^{(1)}(P;\mathcal{V}) \supseteq \cdots \supseteq \mathcal{D}^{(m)}(P;\mathcal{V}).$$

The next proposition brings together the above hierarchies for $\mathcal{C}(P)$ and $\mathcal{D}(P)$.

Proposition 8. *Given $\mathcal{V}$, it holds that*

$$\mathcal{C}(P) \subseteq \mathcal{C}^{(1)}(P;\mathcal{V}) \subseteq \cdots \subseteq \mathcal{C}^{(m)}(P;\mathcal{V}) \subseteq \mathcal{D}^{(m)}(P;\mathcal{V}) \subseteq \cdots \subseteq \mathcal{D}^{(1)}(P;\mathcal{V}) \subseteq \mathcal{D}(P).$$

Proof. We need only establish that $\mathcal{C}^{(m)}(P;\mathcal{V}) \subseteq \mathcal{D}^{(m)}(P;\mathcal{V})$ because the remaining inclusions follow from Lemmas 5 and 6.

In fact, we prove more generally that $\mathcal{C}^{(t)}(P_\bullet;\mathcal{V}) \subseteq \mathcal{D}^{(t)}(P_\bullet;\mathcal{V})$. This follows because the only true difference in the definitions of $\mathcal{C}^{(t)}(P_\bullet;\mathcal{V})$ and $\mathcal{D}^{(t)}(P_\bullet;\mathcal{V})$ occurs at the lowest level of recursion, where $\mathcal{C}(P_*)$ is relaxed to $\mathcal{D}(P_*)$. □

In the preceding proposition, the hierarchy is expressed fully via m levels, where m is the number of linear inequalities defining P. In some cases, one can prove strict inclusion for some levels. For example, for the completely positive case in Section 4, one can show that $\mathcal{C}^{(1)}(P, \mathcal{V})$ is a proper subset of $\mathcal{C}^{(2)}(P, \mathcal{V})$, if each $v_I \in \text{relint}_>(P_I)$ and $n \geq 5$. However, generally it may happen that some of the levels of the hierarchy may be identical, e.g., $\mathcal{C}^{(m-1)}(P; \mathcal{V}) = \mathcal{C}^{(m)}(P; \mathcal{V}) = \mathcal{D}^{(m)}(P; \mathcal{V}) = \mathcal{D}^{(m-1)}(P; \mathcal{V})$, making the full hierarchy unnecessary. For example, in Section 4, results therein imply that the levels of the hierarchy are identical for $t \geq m - 4$.

Finally, we claim that $D^{(t)}(P; \mathcal{V})$ is tractable when t is constant with respect to the size of P. Let p be the total number of linear constraints defining P (i.e., the number of linear equalities plus the number m of linear inequalities). We already know that $\mathcal{D}(P) = \mathcal{D}^{(0)}(P; \mathcal{V})$ requires a single positive-semidefinite variable of size $n \times n$, which is constrained by $O(p^2)$ linear constraints. In fact $\mathcal{D}(P_\bullet) = \mathcal{D}^{(0)}(P_\bullet; \mathcal{V})$ for any argument $P_\bullet$ has the same size since $P_\bullet$ also lives in $\Re^n$ and is defined by p linear constraints. Additionally, one can see from the recursion that $\mathcal{D}^{(t)}(P; \mathcal{V})$ is ultimately based on $O(m^t)$ cones of the form $\mathcal{D}(P_\bullet)$. So the description of $\mathcal{D}^{(t)}(P; \mathcal{V})$ requires roughly $O(m^t)$ semidefinite variables of size $n \times n$ and $O(m^t p^2)$ linear constraints. This shows that, if t is constant with respect to n and p, the level-t relaxation $\mathcal{D}^{(t)}(P; \mathcal{V})$ is tractable.

4 Completely Positive Matrices

In this section, we apply the results of Section 3 to the case when P equals the nonnegative orthant $\Re^+_n$.

4.1 Literature review

For $P = \Re^+_n$, $\mathcal{C}$ is the well-known cone of completely positive matrices, and $\mathcal{C}^*$ is the cone of copositive matrices. $\mathcal{C}$ is full-dimensional, e.g., $I + \frac{1}{n^2} ee^T$ is an interior point [11]; this is also implied by Proposition 1. The semidefinite relaxation $\mathcal{D}$ is the cone of so-called *doubly nonnegative* matrices $\{X \succeq 0 : X \geq 0\}$.

Optimizing a linear function over the completely positive cone with linear constraints is called copositive programming; see [13] for a recent survey. In [6], Burer proved that any nonconvex quadratic program having linear constraints, complementarity constraints, and binary and continuous variables may be modeled as a copositive program.

There are inner approximations for $\mathcal{C}^*$ in the literature. Parrilo [25] proposed a hierarchy of semidefinite inner approximations $\{\mathcal{K}^r\}$ of $\mathcal{C}^*$, in the sense that $\mathcal{D}^* = \mathcal{K}^0 \subseteq \mathcal{K}^1 \subseteq \cdots \subseteq \mathcal{C}^*$ and $\text{int}(\mathcal{C}^*) \subseteq \cup_r \mathcal{K}^r$. In [10], De Klerk and Pasechnik showed that there is another hierarchy of polyhedral cones, approximating $\mathcal{C}^*$ from inside, which satisfies a similar condition. Note the dual cones of these approximating cones can be seen as outer approximations of $\mathcal{C}$.

To emphasize the dimension n, we employ the notation $P^n := \Re_n^+$, $\mathcal{C}^n := \mathcal{C}(\Re_n^+)$ and $\mathcal{D}^n := \mathcal{D}(\Re_n^+)$. Maxfield and Minc [22] proved the following:

Theorem 4. *$\mathcal{C}^n = \mathcal{D}^n$ if and only if $n \leq 4$.*

For $n \geq 5$, relatively little is known about the structure of $\mathcal{C}^n$. Even when $n = 5$, characterizating $\mathcal{C}$ is very difficult. Berman and Xu [3] gives some partial (algebraic) rules to determine whether a given 5×5 matrix is completely positive or not. In the next subsection, we show that by applying the theory in Section 3, we are able to separate any 5×5 matrix from $\mathcal{C}^5$.

4.2 Application and examples

As mentioned in the Introduction, A empty and $B = I$ give rise to $P = \Re_n^+$. Assumptions 1–3 are straightforward to verify, so that all results in Section 3 apply.

To apply the separation algorithm, Algorithm 1, to $\mathcal{C}^n$, we note that

$$P_1^n := \left\{ \begin{pmatrix} 0 \\ \tilde{x} \end{pmatrix} : \tilde{x} \in P^{n-1} \right\}.$$

In other words, P_1^n is just a copy of P^{n-1} embedded in $\Re_n$. The same holds for P_i^n except the embedding sets the i-th coordinate to zero. Similarly,

$$C_1^n := \mathcal{C}(P_1^n) = \left\{ \begin{pmatrix} 0 & 0^T \\ 0 & \tilde{X} \end{pmatrix} : \tilde{X} \in \mathcal{C}^{n-1} \right\}$$

and C_i^n more generally are simply embeddings of $\mathcal{C}^{n-1}$. So Algorithm 1 amounts to the optimization problem (4) over n copies of the cone $(\mathcal{C}^{n-1})^*$. To complete the specification of Algorithm 1, we need a matrix $X^0 \in \operatorname{relint}(\sum_i C_i^n)$. One such choice, which we use in the examples below, is to define X_i^0 to be the appropriate embedding of $\tilde{I} + \frac{1}{(n-1)^2}\tilde{e}\tilde{e}^T \in \operatorname{int}(\mathcal{C}^{n-1})$ into C_i^n and then set $X^0 := \sum_i \frac{1}{n} X_i^0$.

Since $\mathcal{C}^{n-1} = \mathcal{D}^{n-1}$ is tractable for $n \leq 5$, we have the following corollary of Theorem 2:

Corollary 1. *Algorithm 1 correctly solves the separation problem for $\mathcal{C}(\Re_5^+)$.*

This is the first separation algorithm for $\mathcal{C}^5$ and successfully answers the open question as to whether or not $\mathcal{C}^5$ is tractable.

Example 1. *The paper [7] studies the set $\mathcal{D}^n \setminus \mathcal{C}^n$ of so-called* bad *matrices. In particular, the authors give a characterization of the extreme rays of $\mathcal{D}^5$, which are not in $\mathcal{C}^5$; they* *call these* extremely bad *matrices. The authors show how to separate* 5×5 *extremely bad matrices but are unable to separate nearby bad matrices.*

Consider the following 5×5 *extremely bad matrix* Z*, permutation matrix* P*, and doubly nonnegative matrix* $X := 0.93Z + 0.07PZP^T$*, which is close to* Z*:*

$$Z = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 6 & 2 & 0 & 0 \\ 0 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & 5 & 2 \\ 1 & 0 & 0 & 2 & 2 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad X = \begin{pmatrix} 1.00 & 0.93 & 0.07 & 0.07 & 0.93 \\ 0.93 & 5.65 & 1.86 & 0.14 & 0.07 \\ 0.07 & 1.86 & 1.07 & 0.93 & 0.14 \\ 0.07 & 0.14 & 0.93 & 5.07 & 1.86 \\ 0.93 & 0.07 & 0.14 & 1.86 & 2.21 \end{pmatrix}.$$

As detailed in [7], the copositive matrix

$$K = \begin{pmatrix} 9.00 & -4.50 & 10.50 & 4.50 & -7.50 \\ -4.50 & 2.25 & -5.25 & 2.25 & 3.75 \\ 10.50 & -5.25 & 12.25 & -5.25 & 8.75 \\ 4.50 & 2.25 & -5.25 & 2.25 & -3.75 \\ -7.50 & 3.75 & 8.75 & -3.75 & 6.25 \end{pmatrix}$$

separates Z*, but one can check that it does not separate* X*. By using Algorithm 1 in a Matlab implementation using SeDuMi [28], we are able to verify that* $X \notin \mathcal{C}^5$ *via separation with the copositive matrix*

$$Q \approx \begin{pmatrix} 0.380 & -0.190 & 0.409 & 0.161 & -0.263 \\ -0.190 & 0.095 & -0.205 & 0.080 & 0.131 \\ 0.409 & -0.205 & 0.441 & -0.173 & 0.283 \\ 0.161 & 0.080 & -0.173 & 0.068 & -0.111 \\ -0.263 & 0.131 & 0.283 & -0.111 & 0.182 \end{pmatrix}.$$

The SeDuMi solution time is approximately 0.1 seconds on a 2.40 GHz Linux machine.

We now turn our attention to the relaxations $\mathcal{C}(d)$ and $\mathcal{D}(d)$, for given $d \in \Re^+_n$, introduced in Section 3.3. Recall the inclusions $\mathcal{C} \subseteq \mathcal{C}(d) \subseteq \mathcal{D}(d) \subseteq \mathcal{D}$. Here, we do not include a superscript for the dimension n.

Example 2. *Let* $n = 5$*, in which case* $\mathcal{C}(d) = \mathcal{D}(d)$ *since* $\mathcal{C}_{\text{bd}}$ *is based on five copies of* $\mathcal{C}(\Re^+_4)$*. To illustrate the strength of* $\mathcal{D}(d)$ *for various* d*, we consider the optimization problem* $\mu(d) := \min\{\langle H, X\rangle : X \in \mathcal{D}(d), \langle I, X\rangle \leq 1\}$*, where* $H \in \mathcal{C}^* \setminus \mathcal{D}^*$ *is the well-known Horn* *matrix:*

$$H := \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}.$$

It is known that

$$\begin{aligned} 0 &= \min\{\langle H, X\rangle : X \in \mathcal{C}, \langle I, X\rangle \le 1\}, \\ -0.2361 \approx 2 - \sqrt{5} &= \min\{\langle H, X\rangle : X \in \mathcal{D}, \langle I, X\rangle \le 1\}. \end{aligned}$$

Hence, for any $d \ge 0$*, it must hold that* $-0.2361 \le \mu(d) \le 0$*, and the closeness to 0 gives an indirect indication of the strength of* $\mathcal{D}(d)$*. Table 1 summarizes our computational experiment.*

d	$\mu(d)$
$(1,0,0,0,0)^T$	-0.2361
$(1,1,0,0,0)^T$	0
$(1,0,1,0,0)^T$	-0.1249
$(1,1,1,0,0)^T$	-0.0787
$(1,1,1,1,0)^T$	0
$(1,1,1,1,1)^T$	0

Table 1: Strength of $\mathcal{D}_5(d)$ for various d

In our opinion, the preceding example and Proposition 7 suggest that $d = e$ is a reasonable default choice for $\mathcal{D}(d)$.

Example 3. *Again let* $n = 5$*, and let* Z *be an extremely bad matrix in* $\mathcal{D} \setminus \mathcal{C}$ *[7], which necessarily has a zero entry. We claim* $Z \notin \mathcal{D}(e)$*. Suppose for contradiction that* Z *is in* $\mathcal{D}(e)$*. Then Proposition 7 implies* $Z \in \sum_{i=1}^{5} D_i$*, where* $\mathcal{D}_i$ *is precisely all* 5×5 *completely positive matrices with row* i *and column* i *equal to 0. In other words,* $\mathcal{D}_i$ *is isomorphic to the* 4×4 *doubly nonnegative matrices, which are precisely the* 4×4 *completely positive matrices by Theorem 4. So* Z *is the sum of completely positive matrices and hence completely positive, which contradicts the assumption that* $Z \in \mathcal{D} \setminus \mathcal{C}$*.*

To further gauge the strength of $\mathcal{D}(e)$ *relative to* $\mathcal{C}$*, we calculate the distance of* Z *to* $\mathcal{D}(e)$ *and compare it to the distance of* Z *to* $\mathcal{C}$*. The distances are calculated by solving the* *following optimization problems:*

$$dist(Z, \mathcal{D}(e)) := \min\{\|X - Z\|_F : X \in \mathcal{D}(e)\},$$
$$dist(Z, \mathcal{C}) := \min\{\|X - Z\|_F : X \in \mathcal{C}\},$$

where $\|\cdot\|_F$ *is the Frobenius norm induced by the inner product* $\langle\cdot,\cdot\rangle$*. In practice,* $dist(Z,\mathcal{C})$ *is calculated by first solving over* $\mathcal{D}$ *and then repeatedly adding copositive cuts produced by Algorithm 1. We then calculate the percentage*

$$\frac{dist(Z, \mathcal{D}(e))}{dist(Z, \mathcal{C})} \times 100\%.$$

By definition, this percentage is between 0% and 100%, and the closer it is to 100%, the better $\mathcal{D}(e)$ *approximates* $\mathcal{C}$ *near* Z*.*

For 1000 randomly generated extremely bad Z *with* $\langle ee^T, Z\rangle = 1$*, we calculated the above percentage. The average of the 1000 percentages was 99.935%, and the standard deviation was 0.241%. This means that, on average,* $\mathcal{D}(e)$ *cuts away about 99.9% of the distance between the extremely bad matrix* Z *and* $\mathcal{C}$*. In addition the minimum percentage over all 1000* Z *was 97.958%. We feel this is convincing evidence that* $\mathcal{D}(e)$ *approximates* $\mathcal{C}$ *well and is certainly much stronger than* $\mathcal{D}$*.*

In the next example, we numerically examine the recursive hierarchy of tractable convex relaxations $\mathcal{D}^{(t)} \subseteq \cdots \subseteq \mathcal{D}^{(1)} \subseteq \mathcal{D}$ introduced in Section 3.4. The result shows that indeed $\mathcal{D}^{(t)}$ gets much stronger with increased depth t.

Example 4. *Let* $n = 7$*, and consider the following* 7×7 *exceptional and extremal copositive matrix (the so-called* Hoffman-Pereira matrix*) [18]:*

$$Q := \begin{pmatrix} 1 & -1 & 1 & 0 & 0 & 1 & -1 \\ -1 & 1 & -1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 & 1 \\ 1 & 0 & 0 & 1 & -1 & 1 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}$$

Also consider the optimization $\min\{\langle Q, X\rangle : \langle I, X\rangle \le 1, X \in \mathcal{C}\}$*, whose optimal value is 0 by definition. To compute lower bounds for this problem, we replace* $\mathcal{C}$ *by its level-1, level-2, and level-3 relaxations* $\mathcal{D}^{(1)}(\Re^7_+; \mathcal{V})$*,* $\mathcal{D}^{(2)}(\Re^7_+; \mathcal{V})$*, and* $\mathcal{D}^{(3)}(\Re^7_+; \mathcal{V})$*, respectively, where* $\mathcal{V}$

consists of vectors v_I *defined as follows:* $[v_I]_i = 0$ *for all* $i \in I$*, and* $[v_I]_j = 1$ *for all* $j \in J := \{1, \ldots, n\} \setminus I$*. Note that, when* $n = 7$*,* $\mathcal{D}^{(3)}(\Re^7_+; \mathcal{V})$ *is the deepest relaxation one needs to consider since it is based on* $\mathcal{D}(\Re^+_4) = \mathcal{C}(\Re^+_4)$*. We calculated the following four relaxation values:*

$$\begin{aligned}
-0.1099 &\approx \min \{\langle Q, X\rangle : \langle I, X\rangle \le 1, X \in \mathcal{D}\} \\
-0.0824 &\approx \min \left\{\langle Q, X\rangle : \langle I, X\rangle \le 1, X \in \mathcal{D}^{(1)}(\Re^7_+; \mathcal{V})\right\} \\
-0.0824 &\approx \min \left\{\langle Q, X\rangle : \langle I, X\rangle \le 1, X \in \mathcal{D}^{(2)}(\Re^7_+; \mathcal{V})\right\} \\
0.0000 &\approx \min \left\{\langle Q, X\rangle : \langle I, X\rangle \le 1, X \in \mathcal{D}^{(3)}(\Re^7_+; \mathcal{V})\right\}.
\end{aligned}$$

This experiment shows the potential power of the recursive relaxations. However, it should be noted that the relaxations grow exponentially in size with the recursive depth. For example, in our Yalmip/SeDuMi implementation, the first relaxation required 1.3 seconds, while the last one required 3.8.

4.3 The simplex

As mentioned in the Introduction, $\mathcal{C}(\Re^+_n)$ is closely related to $\mathcal{C}(\hom(\Delta_{n-1}))$. In the first case, A is empty and $B = I$, while in the second, $A = (-1, \tilde{e}^T)$ and $B = (0, \tilde{I})$. This case certainly satisfies Assumptions 1–3.

Anstreicher and Burer [1] showed that $\mathcal{C}(\hom(\Delta_{n-1})) = \mathcal{D}(\hom(\Delta_{n-1}))$ if and only if $n \le 5$. Note that this result does not follow directly from Theorem 4 since $\mathcal{C}(\hom(\Delta_4))$ is a subset of $\mathcal{C}(\Re^+_5)$, the completely positive matrices of size 5×5. Using Algorithm 1, we have for the first time a separation algorithm for $\mathcal{C}(\hom(\Delta_5))$, which is a subset of the 6×6 completely positive matrices:

Corollary 2. *Algorithm 1 correctly solves the separation problem for* $\mathcal{C}(\hom(\Delta_5))$*.*

To implement the separation algorithm, we note that, for $P = \hom(\Delta_{n-1})$, P_i is simply an embedding of $\hom(\Delta_{n-2})$. Also, a point X^0 can be generated using arguments found in the proof of Proposition 1.

5 The Box

In this section, we apply the results of Section 3 to the case when P equals the homogenization of the box $\Box_{n-1}$.

5.1 Literature review

For $P = \hom(\square_{n-1})$, $\mathcal{C}$ has been formally studied in the papers [1, 8]. In the latter paper, the slice

$$\{X \in \mathcal{C} : X_{11} = 1\} = \operatorname{conv}\left\{ \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix}^T : \tilde{x} \in \square_{n-1} \right\}$$

is denoted QPB_{n-1}. $\mathcal{C}$ is full-dimensional, e.g., [8] showed that

$$\lambda \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T + \sum_{j=1}^{n-1} \begin{pmatrix} 1 \\ \tilde{e}_j \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{e}_j \end{pmatrix}^T + \sum_{j=1}^{n-1} \begin{pmatrix} 1 \\ \frac{1}{2}\tilde{e}_j \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2}\tilde{e}_j \end{pmatrix}^T + \sum_{1 \le j < k \le n-1} \begin{pmatrix} 1 \\ \tilde{e}_j + \tilde{e}_k \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{e}_j + \tilde{e}_k \end{pmatrix}^T \right]$$

lies in the relative interior of QPB_{n-1}, where $\lambda := [n(n+1)/2]^{-1}$ and $\tilde{e}_j$ is the standard coordinate vector in $\Re_{n-1}$. Since QPB_{n-1} is a slice of $\mathcal{C}$, the above point is an interior point of $\mathcal{C}$. The semidefinite relaxation $\mathcal{D}$ has also been studied in several papers, and the slice $\{X \in \mathcal{D} : X_{11} = 1\}$ is often written PSD $\cap$ RLT since $\mathcal{D}$ combines both positive semidefiniteness and the linear inequalities arising from the reformulation-linearization technique of Sherali and Adams [27].

To emphasize the dimension, we use the notation $P^n := \hom(\square_{n-1})$, $\mathcal{C}^n := \mathcal{C}(\hom(\square_{n-1}))$ and $\mathcal{D}^n := \mathcal{D}(\hom(\square_{n-1}))$. Anstreicher and Burer [1] proved $QPB_{n-1} =$ PSD $\cap$ RLT if and only if $n \le 3$. In terms of the cones presented here, the result is as follows:

Theorem 5. *$\mathcal{C}^n = \mathcal{D}^n$ if and only if $n \le 3$.*

For $n > 3$, Burer and Letchford [8] showed that additional valid inequalities for $\mathcal{C}^n$ can be derived from valid inequalities of the *boolean quadric polytope* [24]

$$BQP_{n-1} := \operatorname{conv}\left\{ \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix}^T : \tilde{x} \in \{0,1\}^{n-1} \right\}.$$

Proposition 9. *Let $Q \in (BQP_{n-1})^*$ with $Q_{22} = \cdots = Q_{nn} = 0$. Then $Q \in (QPB_{n-1})^* = (\mathcal{C}^n)^*$.*

Such valid inequalities include for example the well-known triangle inequalities, which along with the RLT constraints, capture BQP_3 exactly. However, in [8], it is shown by example that the relaxation PSD $\cap$ RLT $\cap$ TRI which incorporates the triangle inequalities to tighten the slice $\{X \in \mathcal{D} : X_{11} = 1\}$ still does not characterize QPB_3 exactly. In terms of cones, $\mathcal{C}^4 \subsetneq \mathcal{D}^4 \cap$ TRI.

Using a different approach, [1] provides an exact, disjunctive formulation of $\mathcal{C}^4$. Otherwise, little is known about the structure of $\mathcal{C}^n$. In particular, it has not been known whether there is a separation procedure for $\mathcal{C}^4$ that is closely related to the cone $\mathcal{D}^4$ and whether separation over $\mathcal{C}^5$ is tractable.

5.2 More when $n = 4$

As mentioned in the previous subsection, the cone $\mathcal{C}^4 := \mathcal{C}(\text{hom}(\square_3))$ is properly contained in the cone $\mathcal{D}^4 \cap \text{TRI}$. Specifically, the set TRI incorporates the following four triangle inequalities:

$$\begin{aligned} X_{23} + X_{24} &\leq X_{12} + X_{34} \\ X_{23} + X_{34} &\leq X_{13} + X_{24} \\ X_{24} + X_{34} &\leq X_{14} + X_{23} \\ X_{12} + X_{13} + X_{14} &\leq X_{23} + X_{24} + X_{34} + 1. \end{aligned}$$

The following proposition provides a nonempty class of points that is guaranteed to be in this difference.

Proposition 10. *Suppose* $X \in \mathcal{D}^4 \cap \text{TRI}$ *satisfies the following conditions:* $\text{rank}(X) = 3$, $X_{22} > 0$, $X_{32} = X_{42} = 0$, $0 < X_{43} < \min\{X_{33}, X_{44}\}$, $X_{33} = X_{31}$, *and* $X_{44} = X_{41}$. *Then* $X \notin \mathcal{C}^4$.

Proof. We suppose $X \in \mathcal{C}^4$ and derive a contradiction. Write a minimal representation $X = \sum_{k=1}^K x^k (x^k)^T$, where $K \geq 3$ and $0 \neq x^k \in P^4$. In particular, by the characterization of P and that of all extreme rays of $\mathcal{C}$, $0 \neq x^k$ implies $x_1^k > 0$ for all k. Since $X_{22} > 0$, we may assume without loss of generality that $x_2^1 > 0$, which implies $x_3^1 = x_4^1 = 0$ since $X_{32} = X_{42} = 0$. Hence

$$X = \begin{pmatrix} x_1^1 \\ x_2^1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} x_1^1 \\ x_2^1 \\ 0 \\ 0 \end{pmatrix}^T + \sum_{k=2}^K x^k (x^k)^T.$$

Next, the conditions $X_{33} = X_{31}$ and $X_{44} = X_{41}$ imply $x_j^k \in \{0, x_1^k\}$ for all k and all $j \in \{3, 4\}$. Then, since $X_{43} > 0$, we may assume $x_3^2 = x_4^2 = x_1^2$, which in turn implies $x_2^2 = 0$ since $X_{32} = 0$. So

$$X = \begin{pmatrix} x_1^1 \\ x_2^1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} x_1^1 \\ x_2^1 \\ 0 \\ 0 \end{pmatrix}^T + \begin{pmatrix} x_1^2 \\ 0 \\ x_1^2 \\ x_1^2 \end{pmatrix} \begin{pmatrix} x_1^2 \\ 0 \\ x_1^2 \\ x_1^2 \end{pmatrix}^T + \sum_{k=3}^K x^k (x^k)^T.$$

Next, note that, for all $k \geq 3$, we have $x_3^k \neq x_1^k$ or $x_4^k \neq x_1^k$; otherwise, $x_2^k = 0$ and the representation for X is not minimal. So, for all $k \geq 3$, we have

$$x^k = \begin{pmatrix} x_1^k \\ 0 \\ x_1^k \\ 0 \end{pmatrix}, \quad x^k = \begin{pmatrix} x_1^k \\ 0 \\ 0 \\ x_1^k \end{pmatrix}, \quad \text{or} \quad x^k = \begin{pmatrix} x_1^k \\ x_2^k \\ 0 \\ 0 \end{pmatrix}.$$

Because $\text{rank}(X) = 3$, at most one of the first two possibilities may occur in the representation, and by minimality, if one possibility does occur, it occurs at most once. Suppose the first one occurs once; then $X_{43} = X_{44} = (x_1^2)^2$. If the second occurs once, then $X_{43} = X_{33} = (x_1^2)^2$. If neither occurs, then $X_{43} = X_{33} = X_{44} = (x_1^2)$. Whatever the case, this contradicts the condition $X_{43} < \min\{X_{33}, X_{44}\}$. □

One can solve a semidefinite program (e.g., with constraints such as $X_{22} \geq \varepsilon$ for fixed small $\varepsilon > 0$) to see that there exist $X \in \mathcal{D}^4 \cap \text{TRI}$ satisfying the conditions of the proposition. One concrete example is the following:

$$\begin{pmatrix} 1 & 1/3 & 1/3 & 1/3 \\ 1/3 & 1/4 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/15 \\ 1/3 & 0 & 1/15 & 1/3 \end{pmatrix}.$$

5.3 Application and examples

As mentioned in the Introduction, A empty and

$$B = \begin{pmatrix} 0 & \tilde{I} \\ \tilde{e} & -\tilde{I} \end{pmatrix}$$

give rise to $P = \text{hom}(\square_{n-1})$. Assumptions 1–3 are straightforward to verify, so that all results in Section 3 apply.

To apply the separation algorithm, Algorithm 1, to $\mathcal{C}^n$, we note that

$$P_{n-1}^n := \left\{ \begin{pmatrix} \tilde{x} \\ 0 \end{pmatrix} : \tilde{x} \in P^{n-1} \right\},$$
$$P_{2(n-1)}^n := \left\{ \begin{pmatrix} \tilde{x} \\ \tilde{x}_1 \end{pmatrix} : \tilde{x} \in P^{n-1} \right\}.$$

In other words, P_{n-1}^n and $P_{2(n-1)}^n$ are just copies of P^{n-1} embedded in $\Re_n$. The same holds for

P_i^n and P_{i+n-1}^n for $i = 1, \ldots, n-1$ generally except the embedding sets the i-th coordinate to zero and $\tilde{x}_1$, respectively. Similarly,

$$C_{n-1}^n := \mathcal{C}(P_{n-1}^n) = \left\{ \begin{pmatrix} \tilde{X} & 0 \\ 0^T & 0 \end{pmatrix} : \tilde{X} \in \mathcal{C}^{n-1} \right\},$$
$$C_{2(n-1)}^n := \mathcal{C}(P_{2(n-1)}^n) = \left\{ \begin{pmatrix} \tilde{X} & \tilde{X}_{\cdot 1} \\ \tilde{X}_{\cdot 1}^T & \tilde{X}_{11} \end{pmatrix} : \tilde{X} \in \mathcal{C}^{n-1} \right\},$$

and $\mathcal{C}_i^n$ and $\mathcal{C}_{n-1+i}^n$ are simply embeddings of $\mathcal{C}^{n-1}$. So Algorithm 1 amounts to the optimization problem (4) over $2(n-1)$ copies of the cone $(\mathcal{C}^{n-1})^*$. To complete the specification of Algorithm 1, we need a matrix $X^0 \in \operatorname{int}(\sum_{i=1}^{2(n-1)} \mathcal{C}_i^n)$. One such choice, which we use in the examples below, is to define X_i^0 to be the appropriate embedding of the $\square_{n-2}$ version of the interior point given in Section 5.1 and then set $X^0 := \sum_{i=1}^{2(n-1)} \frac{1}{2(n-1)} X_i^0$.

Since $\mathcal{C}^{n-1} = \mathcal{D}^{n-1}$ is tractable for $n \leq 3$, we have the following corollary of Theorem 2:

Corollary 3. *Algorithm 1 correctly solves the separation problem for* $\mathcal{C}^4 = \mathcal{C}(\text{hom}(\square_3))$.

This is the first separation algorithm for $\mathcal{C}^4$ that is closely related to the cone $\mathcal{D}^4$. Using the disjunctive formulation of $\mathcal{C}^4$ given in [1], we also obtain a separation procedure for $\mathcal{C}^5$.

Corollary 4. *Algorithm 1 correctly solves the separation problem for* $\mathcal{C}^5 = \mathcal{C}(\text{hom}(\square_4))$.

This answers the open question as to whether or not $\mathcal{C}^4$ and $\mathcal{C}^5$ are tractable.

Example 5. *Burer and Letchford [8] considered the following optimization:* $\min\{\langle \tilde{x}, \tilde{Q}\tilde{x} \rangle + 2\langle \tilde{c}, \tilde{x} \rangle : \tilde{x} \in \square_3\}$, *where*

$$\tilde{Q} = \begin{pmatrix} 2.25 & 3 & 3 \\ 3 & 0 & 0.5 \\ 3 & 0.5 & -1 \end{pmatrix}, \quad \tilde{c} = \begin{pmatrix} -1.5 \\ -0.5 \\ 0 \end{pmatrix}.$$

From basic principles, one can verify that the optimal value is -1. *However, after conversion to the form* $\{\langle Q, X \rangle : X \in \mathcal{C}^4, X_{11} = 1\}$ *and relaxation via* $\mathcal{C}^4 \subsetneq \mathcal{D}^4 \cap \text{TRI}$, *the relaxation value is approximately* -1.0929 *for a gap of 9.29%.*

In [12], Dong and Anstreicher reconsidered the same problem but from the point of view of its completely-positive formulation [6], which lies in the cone $\mathcal{C}(\Re_7^+)$. *By iteratively generating copositive cuts based on the zero structure of the solution to the relaxation over* $\mathcal{D}(\Re_7^+)$, *the gap to the optimal value can be closed to 0%. In a similar spirit, one could also separate* 5×5 *completely positive submatrices of the completely-positive formulation using Algorithm 1 in Section 4.*

Here, we focus on the box structure and generate cuts in $(\mathcal{C}^4)^*$ using Algorithm 1. Similar to [12], the gap is reduced to 0% after the addition of 20 cuts. In contrast to [12], however, our separation requires no special structure in the solution of the relaxation.

In Example 5, the relaxation over $\mathcal{D}^4 \cap \text{TRI}$ yields a gap of 9.29%, and so the optimal solution $\bar{X}$ of the relaxation is in the difference $\mathcal{D}^4 \cap \text{TRI} \setminus \mathcal{C}^4$. In fact, we found that the optimal $\bar{X}$ numerically satisfies the assumptions of Proposition 10, which is necessarily cut off by $\mathcal{D}^4(d)$ by the following corollary of Proposition 7.

Corollary 5. *Let $X \in \mathcal{D}^4 \cap \text{TRI}$ satisfy the assumptions of Proposition 10 so that $X \notin \mathcal{C}^4$, and let $d \in \text{relint}_>(P^4)$. Then $X \notin \mathcal{D}^4(d) := \mathcal{D}(P^4; d)$.*

Proof. We prove the contrapositive. Suppose $X \in \mathcal{D}^4(d)$. The condition $X_{42} = 0$ implies (via Proposition 7) that any X in $\mathcal{D}^4(d)$ can be written as the sum of six matrices (essentially) in $\mathcal{D}^3 = \mathcal{C}^3$ (by Theorem 5). This in turn implies $X \in \mathcal{C}^4$. □

Example 6. *We consider the same problem as in Example 5, where instead we solve over the relaxation $\mathcal{D}^4(d)$, where $d = (1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$. The above corollary guarantees that the gap of 9.29% must be improved. In fact, our Yalmip-CSDP [20, 4] implementation achieves a gap of 0%, i.e., it solves the quadratic problem over $\square_3$ exactly.*

Random numerical experiments suggest that encountering an X in the difference $\mathcal{D}^4 \cap \text{TRI} \setminus \mathcal{C}^4$ does not happen often. In other words, notwithstanding the preceding examples, $\mathcal{D}^4 \cap \text{TRI}$ seems to be a good approximation of $\mathcal{C}^4$. The following example provides some evidence for this viewpoint.

Example 7. *We created a simple procedure to generate randomly a $Z \in \mathcal{D}^4 \cap \text{TRI} \setminus \mathcal{C}^4$ with $Z_{11} = 1$ in accordance with Proposition 10. For 1000 such Z, we calculated the normalized distance*

$$\textit{normdist}(X, \mathcal{C}^4) := \frac{\min\{\|X - Z\|_F : X \in \mathcal{C}^4, X_{11} = 1\}}{\|Z\|_F}.$$

by first relaxing to $\mathcal{D}^4$ and then repeatedly adding $(\mathcal{C}^4)^*$ cuts produced by Algorithm 1. The average normalized distance was 0.0022 with standard deviation 0.0026. The maximum normalized distance was 0.0135.

6 Future Directions

There are many avenues to extend the current paper. It would be interesting to determine the relationship, if any, between our hierarchy for $\mathcal{C}(\Re^+_n)$ and Parrilo's hierarchy. The relationship between $\cap_{d \in P} \mathcal{D}(d)$ and $\mathcal{C}$ could also be explored. In particular, are these two sets equal? Given $X \in \mathcal{D}$, it would also be nice to compute $d \in P$ such that $X \notin \mathcal{D}(d)$ if such a d exists. This would allow one to generate d intelligently instead of using an *a priori* choice of d.

Another important direction for this research is to investigate the hierarchy of relaxations computationally for large n. This will undoubtedly be a challenge due to their large size. On the other hand, the relaxations are well structured, which may lead to opportunities for computational improvements. The results of this paper could also be applied to large n by focusing on small groups of nonnegative or bounded variables. For example, given n bounded variables in a quadratic program, one could focus separately on triples of variables and separate valid inequalities for $\mathcal{C}(\text{hom}(\square_3))$.

Acknowledgments

The authors wish to thank Kurt Anstreicher for many stimulating discussions on the topic of the paper, and to thank two anonymous referees for insightful and constructive comments which have significantly improved the quality of this paper.

References

[1] K. M. Anstreicher and S. Burer. Computable representations for convex hulls of low-dimensional quadratic forms. *Mathematical Programming (series B)*, 124:33–43, 2010.

[2] A. Berman and N. Shaked-Monderer. *Completely Positive Matrices*. World Scientific, 2003.

[3] A. Berman and C. Xu. 5×5 completely positive matrices. *Linear Algebra Appl.*, 393:55–71, 2004.

[4] B. Borchers. CSDP, a C library for semidefinite programming. *Optimization Methods and Software*, 11(1):613–623, 1999.

[5] S. Bundfuss and M. Dür. Algorithmic copositivity detection by simplicial partition. *Linear Algebra and its Applications*, 428:1511–1523, 2008.

[6] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120:479–495, 2009.

[7] S. Burer, K. M. Anstreicher, and M. Dür. The difference between 5×5 Doubly Nonnegative and Completely Positive Matrices. *Linear Algebra and its Applications*, 431(9):1539–1552, 2009.

[8] S. Burer and A. N. Letchford. On non-convex quadratic programming with box constraints. *SIAM Journal on Optimization*, 20(2):1073–1089, 2009.

[9] S. Burer and D. Vandenbussche. Globally solving box-constrained nonconvex quadratic programs with semidefinite-based finite branch-and-bound. *Comput. Optim. Appl.*, 43(2):181–195, 2009.

[10] E. de Klerk and D. V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM J. Optim.*, 12(4):875–892, 2002.

[11] P. J. C. Dickinson. An improved characterisation of the interior of the completely positive cone. *Electronic Journal of Linear Algebra*, 20:723–729, 2010.

[12] H. Dong and K. Anstreicher. Separating Doubly Nonnegative and Completely Positive Matrices. *Mathematical Programming*, 2010. Submitted; Available at `http://www.optimization-online.org/DB_HTML/2010/03/2562.html`.

[13] M. Dür. Copositive Programming - a Survey. In M. Deihl, F. Glineur, E. Jarlebring, and W. Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 3–20. Springer, Nov. 2010.

[14] M. Dür and G. Still. Interior points of the completely positive cone. *Electron. J. Linear Algebra*, 17:48–53, 2008.

[15] C. Floudas and V. Visweswaran. Quadratic optimization. In R. Horst and P. Pardalos, editors, *Handbook of global optimization*, pages 217–269. 1995.

[16] N. I. M. Gould and P. L. Toint. Numerical methods for large-scale non-convex quadratic programming. In *Trends in industrial and applied mathematics (Amritsar, 2001)*, volume 72 of *Appl. Optim.*, pages 149–179. Kluwer Acad. Publ., Dordrecht, 2002.

[17] J.-B. Hiriart-Urruty and A. Seeger. A variational approach to copositive matrices. *SIAM Rev.*, 52:593–629, November 2010.

[18] A. J. Hoffman and F. Pereira. On copositive matrices with -1, 0, 1 entries. *Journal of Combinatorial Theory (A)*, 14:302–309, 1973.

[19] C. R. Johnson and R. Reams. Spectral theory of copositive matrices. *Linear Algebra and its Applications*, 395:275–281, 2005.

[20] J. Löfberg. Yalmip: A toolbox for modeling and optimization in matalb. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[21] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[22] J. E. Maxfield and H. Minc. On the matrix equation $X'X = A$. *Proc. Edinburgh Math. Soc. (2)*, 13:125–129, 1962/1963.

[23] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Math. Programming*, 39(2):117–129, 1987.

[24] M. Padberg. The Boolean quadric polytope: some characteristics, facets and relatives. *Math. Programming*, 45(1, (Ser. B)):139–172, 1989.

[25] P. Parrilo. *Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.

[26] N. V. Sahinidis. BARON: a general purpose global optimization software package. *J. Glob. Optim.*, 8:201–205, 1996.

[27] H. D. Sherali and W. P. Adams. *A Reformulation-Linearization Technique (RLT) for Solving Discrete and Continuous Nonconvex Problems*. Kluwer, 1997.

[28] 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.

[29] J. F. Sturm and S. Zhang. On cones of nonnegative quadratic functions. *Math. Oper. Res.*, 28(2):246–267, 2003.

[30] H. Väliaho. Criteria for Copositive Matrices. *Linear Algebra and its applications*, 81:19–34, 1986.

[31] H. Väliaho. Almost Copositive Matrices. *Linear Algebra and its applications*, 116:121–134, 1989.

[32] D. Vandenbussche and G. Nemhauser. A branch-and-cut algorithm for nonconvex quadratic programs with box constraints. *Mathematical Programming*, 102(3):559–575, 2005.

[33] D. Vandenbussche and G. Nemhauser. A polyhedral study of nonconvex quadratic programs with box constraints. *Mathematical Programming*, 102(3):531–557, 2005.