Separating Doubly Nonnegative and Completely Positive Matrices

Hongbo Dong* and Kurt Anstreicher†

March 8, 2010

Abstract

The cone of Completely Positive (CP) matrices can be used to exactly formulate a variety of NP-Hard optimization problems. A tractable relaxation for CP matrices is provided by the cone of Doubly Nonnegative (DNN) matrices; that is, matrices that are both positive semidefinite and componentwise nonnegative. A natural problem in the optimization setting is then to separate a given DNN but non-CP matrix from the cone of CP matrices. We describe two different constructions for such a separation that apply to 5×5 matrices that are DNN but non-CP. We also describe a generalization that applies to larger DNN but non-CP matrices having block structure. Computational results illustrate the applicability of these separation procedures to generate improved bounds on difficult problems.

1 Introduction

Let $\mathcal{S}_n$ denote the set of $n \times n$ real symmetric matrices, $\mathcal{S}_n^+$ denote the cone of $n \times n$ real symmetric positive semidefinite matrices and $\mathcal{N}_n$ denote the cone of symmetric nonnegative $n\times n$ matrices. The cone of doubly nonnegative (DNN) matrices is then $\mathcal{D}_n = \mathcal{S}_n^+ \cap \mathcal{N}_n$. The cone of completely positive (CP) $n \times n$ matrices, denoted $\mathcal{C}_n$, consists of all matrices that can be written in the form AA^T where A is an $n \times k$ nonnegative matrix. The dual cone $\mathcal{C}_n^*$ is the cone of $n \times n$ copositive matrices, that is, matrices $X \in \mathcal{S}_n$ such that $y^T Xy \geq 0$ for every $y \in \Re_n^+$.

The CP cone is of interest in optimization due to the fact that a variety of NP-Hard problems can be represented as linear optimization problems over $\mathcal{C}_n$; see [Bur09] and references therein. On the other hand, the DNN cone provides a tractable relaxation $\mathcal{C}_n \subset \mathcal{D}_n$ since linear optimization over $\mathcal{D}_n$ can be performed using software for self-dual cones such as SeDuMi [Stu99]. A natural problem in the optimization setting is then to separate a given DNN but non-CP matrix from the CP cone. Since $\mathcal{D}_n = \mathcal{C}_n$ for $n \leq 4$ [BSM03], this separation problem is of interest for matrices with $n \geq 5$. The separation problem for $n = 5$

*Department of Applied Mathematics and Computational Sciences, University of Iowa, Iowa City, IA 52242. email:hondong@math.uiowa.edu

†Corresponding author. Department of Management Sciences, University of Iowa, Iowa City, IA 52242. email:kurt-anstreicher@uiowa.edu

was first considered in [BAD09]. Following the terminology of [BAD09], we say that an $n \times n$ matrix is *bad* if $X \in \mathcal{D}_n \setminus \mathcal{C}_n$, and *extremely bad* if X is a bad extreme ray of $\mathcal{D}_n$. In [BAD09] it is shown that if $X \in \mathcal{D}_5$ is extremely bad, then X can be constructively separated from $\mathcal{C}_5$ using a "cut" matrix $V \in \mathcal{C}_5^*$ that has $V \bullet X < 0$.

In this paper we describe new separation procedures that apply to matrices $X \in \mathcal{D}_n \setminus \mathcal{C}_n$. In Section 2, we generalize the separation procedure from [BAD09] to apply to a broader class of matrices $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ than the extremely bad matrices considered in [BAD09]. As in [BAD09], the cut matrix that is constructed is a transformation of the Horn matrix $H \in \mathcal{C}_5^* \setminus \mathcal{D}_5^*$. We also show that these "transformed Horn cuts" induce 10-dimensional faces of the 15-dimensional cone $\mathcal{C}_5$. In Section 3 we describe an even more general separation procedure that applies to any matrix $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ that is not componentwise strictly positive. The cut generation procedure in this case relies on the solution of a conic optimization problem. In Section 4 we describe separation procedures motivated by the procedure of Section 3 that apply to matrices in $\mathcal{D}_n \setminus \mathcal{C}_n$, $n > 5$ having block structure. In all cases, the goal is to find a copositive $V \in \mathcal{C}_n^*$ that separates X from $\mathcal{C}_n$ via $V \bullet X < 0$. In Section 5 we numerically apply the separation procedures developed in Sections 2-4 to selected test problems.

To date, most of the literature concerning the use of CP matrices in optimization has involved schemes for approximating the copositive cones $\mathcal{C}_n^*$. For example, [BdK02] and [dKP02] describe a hierarchy of cones $\mathcal{K}_n^r$, $r \geq 0$, with $\mathcal{K}_n^0 = \mathcal{D}_n^*$ and $\mathcal{K}_n^r \subset \mathcal{K}_n^{r+1} \subset \mathcal{C}_n^*$. A different approach to dynamically refining an approximation of $\mathcal{C}_n^*$ is taken in [BD09]. To our knowledge, the only paper other than [BAD09] that considers strengthening the DNN relaxation of a problem posed over CP matrices by generating cuts corresponding to copositve matrices is [BFL10]. However, the methodology described in [BFL10] is specific to one particular problem (max-clique), while our approach is independent of the underlying problem and assumes only that the goal is to obtain a CP solution.

Notation: For $X \in \mathcal{S}_n$, we sometimes write $X \succeq 0$ to mean $X \in \mathcal{S}_n^+$. We use e to denote a vector of appropriate dimension with each component equal to one, and $E = ee^T$. For a vector $v \geq 0$, $\sqrt{v}$ is the vector whose ith component is $\sqrt{v_i}$. For $n \times n$ matrices A and B, $A \circ B$ denotes the Hadamard product $(A \circ B)_{ij} = a_{ij} b_{ij}$, and $A \bullet B$ denotes the matrix inner product $A \bullet B = e^T(A \circ B)e$. If $a \in \Re^n$ and A is an $n \times n$ matrix, then $\mathrm{diag}(A)$ is the vector whose ith component is a_{ii}, while $\mathrm{Diag}(a)$ is the diagonal matrix with $\mathrm{diag}(\mathrm{Diag}(a)) = a$. For $X \in \mathcal{S}_n$, $G(X)$ denotes the undirected graph with vertex set $\{1, 2, \ldots, n\}$ and edge set $\{\{i \neq j\} : x_{ij} \neq 0\}$. We use $\mathrm{Co}\{\,\cdot\,\}$ to denote the convex hull of a set.

2 Separation based on the Horn matrix

In this section we describe a procedure for separating a bad 5×5 matrix from $\mathcal{C}_5$ that generalizes the separation procedure for extremely bad matrices in [BAD09]. Recall that from [BAD09], the class of extremely bad matrices (extreme rays of $\mathcal{D}_5$ that are not in $\mathcal{C}_5$) is

$$\mathcal{E}_5 := \{X \in \mathcal{D}_5 : \mathrm{rank}(X) = 3 \text{ and } G(X) \text{ is a 5-cycle}\}.$$

Note that when $G(X)$ is a 5-cycle, every vertex in $G(X)$ has degree equal to two. We will generalize the procedure of [BAD09] to apply to the case where $\mathrm{rank}(X) = 3$ and $G(X)$ has *at least one* vertex with degree two.

Our construction utilizes several results from [BX04] regarding matrices in $\mathcal{D}_5 \setminus \mathcal{C}_5$ of the form

$$X = \begin{pmatrix} X_{11} & \alpha_1 & \alpha_2 \\ \alpha_1^T & 1 & 0 \\ \alpha_2^T & 0 & 1 \end{pmatrix}, \tag{1}$$

where $X_{11} \in \mathcal{D}_3$. Following the notation of [BX04], for a matrix $X \in \mathcal{D}_5$ of the form (1), let C denote the Schur complement $C = X_{11} - \alpha_1\alpha_1^T - \alpha_2\alpha_2^T$. Let $\mu(C)$ denote the number of negative components above the diagonal in C, and if $\mu(C) > 0$ define

$$\lambda_4 = \min_{1 \le i < j \le 3} \left\{ \frac{x_{i4}x_{j4}}{-c_{ij}} \,\middle|\, c_{ij} < 0 \right\}, \quad \lambda_5 = \min_{1 \le i < j \le 3} \left\{ \frac{x_{i5}x_{j5}}{-c_{ij}} \,\middle|\, c_{ij} < 0 \right\}. \tag{2}$$

It is shown in [BX04, Theorem 2.5] that $X \in \mathcal{C}_5$ if $\mu(C) > 1$ and $\lambda_4 + \lambda_5 \ge 1$, and in [BX04, Theorem 3.1] that $X \in \mathcal{C}_5$ if $\mu(C) \ne 2$. Thus for a matrix $X \in \mathcal{D}_5$ of the form (1), $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ implies that $\mu(C) = 2$ and $\lambda_4 + \lambda_5 < 1$. Finally, when $\operatorname{rank}(X) = 3$, it is shown in [BX04, Theorem 4.2] that these conditions are also sufficient to have $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$. For convenience we repeat this result here.

Theorem 1. [BX04, Theorem 4.2] *Assume that $X \in \mathcal{D}_5$ has the form (1), with $\operatorname{rank}(X) = 3$. Then $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ if and only if $\mu(C) = 2$ and $\lambda_4 + \lambda_5 < 1$, where λ_4 and λ_5 are given by (2).*

As in [BAD09], the separation procedure developed here is based on a transformation of the well-known Horn matrix [HJ67],

$$H := \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix} \in \mathcal{C}_5^* \setminus \mathcal{D}_5^*.$$

Theorem 2. *Suppose that $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ has $\operatorname{rank}(X) = 3$, and $G(X)$ has a vertex of degree two. Then there exists a permutation matrix P and a diagonal matrix Λ with $\operatorname{diag}(\Lambda) > 0$ such that*

$$P\Lambda H \Lambda P^T \bullet X < 0.$$

Proof. To begin, consider a transformation of X of the form

$$\tilde{X} = \Sigma Q X Q^T \Sigma, \tag{3}$$

where Σ is a diagonal matrix with $\operatorname{diag}(\Sigma) > 0$ and Q is a permutation matrix. Then X satisfies the conditions of the theorem if and only if $\tilde{X}$ does. Moreover,

$$\begin{aligned} P\Lambda H\Lambda P^T \bullet \tilde{X} &= P\Lambda H\Lambda P^T \bullet \Sigma Q X Q^T \Sigma \\ &= P\Lambda H\Lambda P^T \bullet \Sigma(PP^T)QXQ^T(PP^T)\Sigma \\ &= Q^T P(P^T\Sigma P)\Lambda H\Lambda(P^T\Sigma P)P^T Q \bullet X \\ &= \tilde{P}\tilde{\Lambda}H\tilde{\Lambda}\tilde{P}^T \bullet X, \end{aligned}$$

where $\tilde{P} = Q^T P$ and $\tilde{\Lambda} = (P^T \Sigma P)\Lambda$. It follows that if we apply an initial transformation of the form (3) and show that the theorem holds for $\tilde{X}$, then it also holds for X. Below we will continue to refer to such a transformed matrix as X rather than $\tilde{X}$ to reduce notation.

Note that $\operatorname{diag}(X) > 0$, because otherwise $X \succeq 0$ implies that the only nonzero block in X is a 4×4 DNN matrix, and therefore $X \in \mathcal{C}_5$. Since $G(X)$ has a vertex of degree 2, after applying a suitable permutation we may assume that $x_{45} = x_{15} = 0$, $x_{25} > 0$, $x_{35} > 0$. The assumption that $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ then implies that $x_{14} > 0$, since otherwise it is easy to see that $G(X)$ would contain no 5-cycle, implying $X \in \mathcal{C}_5$ [BSM03]. Let C denote the Schur complement $C = X_{11} - \alpha_1 \alpha_1^T - \alpha_2 \alpha_2^T$. By Theorem 1, C has exactly two negative entries above the diagonal, so at least one of c_{13} and c_{23} must be negative. If necessary interchanging row/column 2 and 3, we can assume that $c_{13} < 0$. After applying a suitable diagonal scaling we may therefore assume that X has the form (1), with $\alpha_1 = (1, u, v)^T$ and $\alpha_2 = (0, 1, 1)^T$.

The fact that $\operatorname{rank}(X) = 3$ implies that $\operatorname{rank}(C) = 1$, and $X \succeq 0$ implies that $C \succeq 0$. Since $c_{13} < 0$, it follows that there are scalars $x > 0$, $y > 0$, $z > 0$ so that C has one of the following forms:

$$C = \begin{pmatrix} x \\ y \\ -z \end{pmatrix} \begin{pmatrix} x \\ y \\ -z \end{pmatrix}^T \quad \text{or} \quad C = \begin{pmatrix} x \\ -y \\ -z \end{pmatrix} \begin{pmatrix} x \\ -y \\ -z \end{pmatrix}^T .$$

We next show that in fact the second case is impossible under the assumptions of the theorem. Assume that

$$C = \begin{pmatrix} x \\ -y \\ -z \end{pmatrix} \begin{pmatrix} x \\ -y \\ -z \end{pmatrix}^T, \quad \text{so} \quad X_{11} = \begin{pmatrix} x^2+1 & u-xy & v-xz \\ u-xy & y^2+u^2+1 & yz+uv+1 \\ v-xz & yz+uv+1 & z^2+v^2+1 \end{pmatrix}. \tag{4}$$

By Theorem 1, $\lambda_4 + \lambda_5 < 1$, where λ_4 and λ_5 are defined as in (2). Obviously $\lambda_5 = 0$, and $\lambda_4 = \min\{\frac{u}{xy}, \frac{v}{xz}\}$. But $\lambda_4 = \frac{u}{xy} < 1 \Rightarrow -xy + u < 0 \Rightarrow x_{12} < 0$, which is impossible since $X \in \mathcal{D}_5$. Assuming that $\lambda_4 = \frac{v}{xy} < 1$ leads to a similar contradiction $x_{13} < 0$, and therefore (4) cannot occur. We may therefore conclude that

$$C = \begin{pmatrix} x \\ y \\ -z \end{pmatrix} \begin{pmatrix} x \\ y \\ -z \end{pmatrix}^T, \quad \text{so} \quad X_{11} = \begin{pmatrix} x^2+1 & xy+u & v-xz \\ xy+u & y^2+u^2+1 & uv-yz+1 \\ v-xz & uv-yz+1 & z^2+v^2+1 \end{pmatrix}. \tag{5}$$

Again $\lambda_5 = 0$, and now $\lambda_4 = \min\{\frac{v}{xz}, \frac{uv}{yz}\}$. Then $\lambda_4 = \frac{v}{xz} < 1 \Rightarrow x_{13} < 0$, which is impossible, so we must have $\lambda_4 = \frac{uv}{yz} < 1$.

Let T be the permuted Horn matrix

$$T := PHP^T = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{pmatrix}, \quad \text{where} \quad P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Computation using symbolic mathematical software yields

$$\operatorname{Det}(X \circ T) = -4x^2(yz - uv)^2 < 0.$$

Thus $X \circ T$ is nonsingular, and symbolic software obtains $(X \circ T)^{-1} = \frac{1}{2(uv-yz)} D_x M D_x$, where $D_x = \mathrm{Diag}((\frac{1}{x}, 1, 1, \frac{1}{x}, 1)^T)$ and

$$M = \begin{pmatrix} 2uv & z & y & vxy - uxz + 2uv & y+z \\ z & 0 & 1 & z + vx & 1 \\ y & 1 & 0 & y - ux & 1 \\ vxy - uxz + 2uv & z + vx & y - ux & 2(u+xy)(v-xz) & y+z+vx-ux \\ y+z & 1 & 1 & y+z+vx-ux & 2(1+uv-yz) \end{pmatrix}.$$

By using the inequalities $x_{13} = v - xz \geq 0$, $x_{23} = uv - yz + 1 \geq 0$ and an implied inequality $\lambda_4 < 1 \Rightarrow yz > uv \geq uxz \Rightarrow y > ux$, one can easily verify that $M \geq 0$ and therefore $(X \circ T)^{-1} \leq 0$, with at least one strictly negative component in each row.

Finally, define $w = -(X \circ T)^{-1} e > 0$, $V = \mathrm{Diag}(w) T \,\mathrm{Diag}(w) = P \Lambda H \Lambda P^T$ where $\Lambda = P^T \,\mathrm{Diag}(w) P$. Then

$$\begin{aligned} V \bullet X &= \mathrm{Diag}(w) T \,\mathrm{Diag}(w) \bullet X = (T \circ ww^T) \bullet X = ww^T \bullet (X \circ T) \\ &= w^T (X \circ T) w = e^T (X \circ T)^{-1} (X \circ T)(X \circ T)^{-1} e = e^T w < 0. \end{aligned}$$

$\square$

In Theorem 2, the condition that $G(X)$ has a vertex with degree 2 implies that $G(X)$ is a "book graph with CP pages." There exists an algebraic procedure to determine whether a matrix $X \in \mathcal{D}_n$ with such a graph is in $\mathcal{C}_n$; see [Bar98] or [BSM03, Theorem 2.18]. When X is extremely bad, it can be shown that the matrix $X \circ T$ in the proof of Theorem 2 is "almost copositive," which implies the facts that $X \circ T$ is nonsingular and $(X \circ T)^{-1} \leq 0$ [Väl89, Theorem 4.1]. However, we have been unable to show that $X \circ T$ is almost copositive under the more general assumptions of Theorem 2.

Let $V \in \mathcal{C}_5^*$, and consider a cut of the form $V \bullet X \geq 0$ that is valid for any $X \in \mathcal{C}_5$. From the standpoint of eliminating elements of $\mathcal{D}_5 \setminus \mathcal{C}_5$, it is desirable for the face

$$\mathcal{F}(\mathcal{C}_5, V) = \{X \in \mathcal{C}_5 : V \bullet X = 0\}$$

to have high dimension. We next show that the cut based on a transformed Horn matrix from Theorem 2 induces a 10-dimensional face of the 15-dimensional cone $\mathcal{C}_5$. (It is known [DS08] that $\mathcal{C}_n$ has an interior in the $n(n+1)/2$-dimensional space corresponding to the components of $X \in \mathcal{C}_n$ on or above the diagonal.)

Theorem 3. *Let $V = P \Lambda H \Lambda P^T$, where P is a permutation matrix and Λ is a diagonal matrix with* $\mathrm{diag}(\Lambda) > 0$. *Then* $\dim \mathcal{F}(\mathcal{C}_5, V) = 10$.

Proof. Without loss of generality we may take $P = \Lambda = I$, so $V = H$. The extreme rays of $\mathcal{F}(\mathcal{C}_5, H)$ are then matrices of the form $X = xx^T$, where $x \geq 0$ and $x^T H x = 0$, and any element of $\mathcal{F}(\mathcal{C}_5, H)$ can be written as a nonnegative combination of such extreme rays. To determine $\dim \mathcal{F}(\mathcal{C}_5, H)$ we must therefore determine the maximum number of such extreme rays that are linearly independent.

For any $X = xx^T$, $x \geq 0$ let $\mathcal{I}(x) = \{i : x(i) > 0\}$. We first consider which sets $\mathcal{I}(x)$ are possible for $X = xx^T \in \mathcal{F}(\mathcal{C}_5, H)$. It is easy to show that

$$\begin{aligned} x^T H X &= (x_1 - x_2 + x_3 + x_4 - x_5)^2 + 4x_2 x_4 + 4x_3(x_5 - x_4) \\ &= (x_1 - x_2 + x_3 - x_4 + x_5)^2 + 4x_2 x_5 + 4x_1(x_4 - x_5). \end{aligned}$$

The fact that $H \in \mathcal{C}_5^*$ then follows from the fact that for any $x \geq 0$, either $x_4 \geq x_5$ or $x_5 \geq x_4$. Moreover, $x > 0$ and $x^T Hx = 0$ would imply $x_4 = x_5 > 0$, and therefore $x_2 = 0$, a contradiction, so $|\mathcal{I}(x)| = 5$ is impossible. Similarly $|\mathcal{I}(x)| = 4$ implies that $x_4 = x_5 > 0$, and therefore $x_2 = 0$, so $(x_1 - x_2 + x_3 + x_4 - x_5)^2 = (x_1 + x_3)^2 > 0$, and $xx^T \notin \mathcal{F}(\mathcal{C}_5, H)$. Thus $|\mathcal{I}(x)| = 4$ is impossible. Finally $|\mathcal{I}(x)| = 1$ is impossible since $\operatorname{diag}(H) = e > 0$, so the only possibilities are $|\mathcal{I}(x)| = 2$ and $|\mathcal{I}(x)| = 3$.

Let H^+ and X^+ be the principal submatrices of H and $X = xx^T$, respectively, corresponding to the positive components of x. For $|\mathcal{I}(x)| = 2$, H^+ has the form

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and obviously $x^T Hx = 0$ is only possible in the first case. It follows that if $|\mathcal{I}(x)| = 2$, then $\mathcal{I}(x)$ is either $\{1,2\}$, $\{2,3\}$, $\{3,4\}$, $\{4,5\}$ or $\{5,1\}$, and in each case X^+ is a positive multiple of ee^T.

Next assume that $|\mathcal{I}(x)| = 3$. Clearly H^+ cannot have a row equal to $(1,1,1)$, and the 3×3 principal submatrices of H that do not contain such a row correspond to $\mathcal{I}(x)$ equal to $\{1,2,3\}$, $\{2,3,4\}$, $\{3,4,5\}$, $\{4,5,1\}$ and $\{5,1,2\}$. Consider $\mathcal{I}(x) = \{1,2,3\}$, so

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

Since $H^+ \succeq 0$, $x^T Hx = 0 \iff H^+ x^+ = 0$, where $x^+ = (x_1, x_2, x_3)^T$. It is then obvious that $x^+ = x_2(u, 1, 1-u)^T$ for some $0 < u < 1$. The vector t corresponding to the upper triangle of $x^+(x^+)^T$ then has the form

$$t = (x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33})^T = x_2^2(u^2, u, u(1-u), 1, (1-u), (1-u)^2)^T. \tag{6}$$

Any vector of the form (6) satisfies the equations

$$\begin{aligned} t_1 - 2t_2 + t_4 - t_6 &= 0 \\ t_1 - t_2 + t_3 &= 0 \\ t_2 - t_4 + t_5 &= 0. \end{aligned}$$

These equations are linearly independent, so there can be at most 3 linearly independent vectors of the form (6). However, the values $u = 0$ and $u = 1$ correspond to two of the matrices $X = xx^T$ with $|\mathcal{I}(x) = 2|$, so only one matrix with $0 < u < 1$ can be added while maintaining a linearly independent set. The same argument applies for all of the other possibilities having $|\mathcal{I}(x) = 3|$, so the maximum possible dimension for $\mathcal{F}(\mathcal{C}_5, H)$ is ten. Finally it can be verified numerically that the five $X = xx^T$ with distinct $|\mathcal{I}(x) = 2|$, and five with distinct $|\mathcal{I}(x) = 3|$, are in fact linearly independent. $\square$

As mentioned above, Theorem 2 can be viewed as a generalization of the separation result for extremely bad 5×5 matrices in [BAD09, Theorem 8]. To make this connection more explicit, we can use the fact [BAD09, Section 2] that $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ is extremely bad if and only if X can be written in the form

$$X = P\Lambda \begin{pmatrix} 1 & \beta_1 & 0 & 0 & 1 \\ \beta_1 & \beta_1^2+\beta_2^2+1 & \beta_2 & 0 & 0 \\ 0 & \beta_2 & 1 & 1 & 0 \\ 0 & 0 & 1 & \beta_2^2+1 & \beta_1\beta_2 \\ 1 & 0 & 0 & \beta_1\beta_2 & \beta_1^2+1 \end{pmatrix} \Lambda P^T,$$

where Λ is a positive diagonal matrix, P is a permutation matrix and $\beta_1, \beta_2 > 0$. With a suitable permutation of rows/columns and a slight modification in Λ, this characterization is equivalent to

$$X = P\Lambda \begin{pmatrix} \beta_1^2+1 & \beta_1\beta_2 & 0 & 1 & 0 \\ \beta_1\beta_2 & \beta_2^2+1 & 0 & 0 & 1 \\ 0 & 0 & \frac{\beta_1^2}{\beta_2^2}+1+\frac{1}{\beta_2^2} & \frac{\beta_1}{\beta_2} & 1 \\ 1 & 0 & \frac{\beta_1}{\beta_2} & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} \Lambda P^T,$$

corresponding to (5) with $u = 0$, $v = \frac{\beta_1}{\beta_2}$, $x = \beta_1$, $y = \beta_2$, $z = \frac{1}{\beta_2}$. Another interesting special case where Theorem 2 applies is as follows. Let H be the Horn matrix and consider the face of $\mathcal{D}_5$,

$$\mathcal{F}(\mathcal{D}_5, H) := \{X \in \mathcal{D}_5 : H \bullet X = 0\}.$$

In the optimization context an element of $\mathcal{F}(\mathcal{D}_5, H)$ could arise naturally via the following sequence. Suppose that an optimization problem posed over $\mathcal{D}_5$ has a solution X that satisfies the assumptions of Theorem 2 (for example, X is extremely bad). After adding a transformed Horn cut and re-solving, the new solution X' (after diagonal scaling and permutation) would likely be an extreme ray of $\mathcal{F}(\mathcal{D}_5, H)$. Burer (private communication) obtained the following characterization of the extreme rays of $\mathcal{F}(\mathcal{D}_5, H)$.

Theorem 4. *Let X be an extreme ray of $\mathcal{F}(\mathcal{D}_5, H)$. Then $\operatorname{rank}(X)$ equals 1 or 3. Further, if $\operatorname{rank}(X) = 3$, then $G(X)$ is either a 5-cycle, or a 5-cycle with a single additional chord.*

One can characterize the matrices in Theorem 4 using an argument similar to what is done in [BAD09, Section 2] to characterize extremely bad matrices. The result is that a matrix $X \in \mathcal{D}_5$ satisfies the conditions of Theorem 4 if and only if there exists a permutation matrix P, a positive diagonal matrix Λ and $\beta_1, \beta_2, \beta_3 > 0$, $\beta_2 \le \beta_1\beta_3$, such that

$$X = P\Lambda \begin{pmatrix} \beta_1^2+1 & \beta_1\beta_2 & 0 & 1 & 0 \\ \beta_1\beta_2 & \beta_2^2+1 & 1-\frac{\beta_2}{\beta_1\beta_3} & 0 & 1 \\ 0 & 1-\frac{\beta_2}{\beta_1\beta_3} & \frac{1}{\beta_3^2}+1+\frac{1}{\beta_1^2\beta_3^2} & \frac{1}{\beta_3} & 1 \\ 1 & 0 & \frac{1}{\beta_3} & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} \Lambda P^T,$$

correponding to (5) with $u = 0$, $v = \frac{1}{\beta_3}$, $x = \beta_1$, $y = \beta_2$, $z = \frac{1}{\beta_1\beta_3}$. Note that such an X is extremely bad if and only if $\beta_2 = \beta_1\beta_3$.

3 Separation based on conic programming

In this section we describe a separation procedure that applies to a broader class of matrices $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ than the procedure of the previous section. Let $X \in \mathcal{D}_5$, with at least one off-diagonal zero, and assume that $\mathrm{diag}(X) > 0$ since otherwise $X \in \mathcal{C}_5$ is immediate. After a permutation and diagonal scaling, X may be assumed to have the form (1). For such a matrix a useful characterization of $X \in \mathcal{C}_5$ is given by the following theorem from [BX04].

Theorem 5. [BX04, Theorem 2.1] *Let $X \in \mathcal{D}_5$ have the form* (1)*. Then $X \in \mathcal{C}_5$ if and only if there are matrices A_{11} and A_{22} such that $X_{11} = A_{11} + A_{22}$, and*

$$\begin{pmatrix} A_{ii} & \alpha_i \\ \alpha_i^T & 1 \end{pmatrix} \in \mathcal{D}_4, \quad i = 1, 2.$$

In [BX04], Theorem 5 is utilized only as a proof mechanism, but we now show that it has algorithmic consequences as well.

Theorem 6. *Assume that $X \in \mathcal{D}_5$ has the form* (1)*. Then $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ if and only if there is a matrix*

$$V = \begin{pmatrix} V_{11} & \beta_1 & \beta_2 \\ \beta_1^T & \gamma_1 & 0 \\ \beta_2^T & 0 & \gamma_2 \end{pmatrix} \quad \textit{such that} \quad \begin{pmatrix} V_{11} & \beta_i \\ \beta_i^T & \gamma_i \end{pmatrix} \in \mathcal{D}_4^*, \quad i = 1, 2,$$

and $V \bullet X < 0$.

Proof. Consider the "CP feasibility problem"

$$\begin{aligned} \text{(CPFP)} \quad \min \quad & 2\theta \\ \text{s.t.} \quad & \begin{pmatrix} A_{ii} & \alpha_i \\ \alpha_i^T & 1 \end{pmatrix} + \theta(I + E) \in \mathcal{D}_4, \quad i = 1, 2 \\ & A_{11} + A_{22} = X_{11} \\ & \theta \geq 0, \end{aligned}$$

where by assumption $\alpha_i \geq 0$, $i = 1, 2$ and $X_{11} \in \mathcal{D}_3$. By Theorem 5, $X \in \mathcal{C}_5$ if and only if the solution value in CPFP is zero. Using conic duality it is straightforward to verify that the dual of CPFP can be written

$$\begin{aligned} \text{(CPDP)} \quad \max \quad & -(V_{11} \bullet X_{11} + 2\alpha_1^T \beta_1 + 2\alpha_2^T \beta_2 + \gamma_1 + \gamma_2) \\ \text{s.t.} \quad & \begin{pmatrix} V_{11} & \beta_i \\ \beta_i^T & \gamma_i \end{pmatrix} \in \mathcal{D}_4^*, \quad i = 1, 2 \\ & (I + E) \bullet V_{11} + e^T \beta_1 + e^T \beta_2 + \gamma_1 + \gamma_2 \leq 1. \end{aligned}$$

Moreover CPFP and CPDP both have feasible interior solutions, so strong duality holds. The proof is completed by noting that the objective in CPDP is exactly $-V \bullet X$. □

Suppose that $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$, and V is a matrix that satisfies the conditions of Theorem 6. If $\tilde{X} \in \mathcal{C}_5$ is another matrix of the form (1), then Theorem 5 implies that $V \bullet \tilde{X} \geq 0$. However we cannot conclude that $V \in \mathcal{C}_5^*$ because $V \bullet \tilde{X} \geq 0$ only holds for $\tilde{X}$ of the form (1), in particular, $\tilde{x}_{45} = 0$. Fortunately, the matrix V can easily be "completed" to obtain a copositive matrix that still separates X from $\mathcal{C}_5$.

Theorem 7. *Suppose that $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ has the form* (1)*, and V satisfies the conditions of Theorem 6. Define*

$$V(s) = \begin{pmatrix} V_{11} & \beta_1 & \beta_2 \\ \beta_1^T & \gamma_1 & s \\ \beta_2^T & s & \gamma_2 \end{pmatrix}.$$

Then $V(s) \bullet X < 0$ for any s, and $V(s) \in \mathcal{C}_5^*$ for $s \geq \sqrt{\gamma_1\gamma_2}$.

Proof. The fact that $V(s) \bullet X = V \bullet X < 0$ is obvious from $x_{45} = 0$, and $V(s) \in \mathcal{C}_5^*$ for $s \geq \sqrt{\gamma_1\gamma_2}$ follows immediately from [HJR05, Theorem 1]. □

Theorems 6 and 7 provide a separation procedure that applies to any $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$ that is not componentwise strictly positive. In applying Theorem 6 to separate a given X, one can numerically minimize $X \bullet V$, where V satisfies the conditions in the theorem and is normalized via a condition such as $I \bullet V = 1$ or $(I + E) \bullet V = 1$. In [BX04, Theorem 6.1] it is claimed that if $X \in \mathcal{D}_5 \setminus \mathcal{C}_5$, $X > 0$, then there is a simple transformation of X that produces another matrix $\tilde{X} \in \mathcal{D}_5 \setminus \mathcal{C}_5$, $\tilde{X} \not> 0$, suggesting that a separation procedure for X could be based on applying the construction of Theorem 6 to $\tilde{X}$. However it is shown in [DA09] that in fact [BX04, Theorem 6.1] is false.

4 Separation for matrices with $n > 5$

Suppose now that $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ for $n > 5$. In order to separate X from $\mathcal{C}_n$, we could attempt to apply the procedure in Section 2 or Section 3 to candidate 5×5 principal submatrices of X. However, it is possible that all such submatrices are in $\mathcal{C}_5$ so that no cut based on a 5×5 principal submatrix can be found. In this section we consider extensions of the separation procedure developed in Section 3 to matrices $X \in \mathcal{D}_n \setminus \mathcal{C}_n$, $n > 5$ having block structure. In order to state the separation procedure in its most general form we will utilize the notion of a CP graph.

Definition 1. *Let G be an undirected graph on n vertices. Then G is called a CP graph if any matrix $X \in \mathcal{D}_n$ with $G(X) = G$ also has $X \in \mathcal{C}_n$.*

The main result on CP graphs is the following:

Proposition 1. [KB93] *An undirected graph on n vertices is a CP graph if and only if it contains no odd cycle of length 5 or greater.*

Our separation procedure is based on a simple observation for CP matrices of the form

$$X = \begin{pmatrix} X_{11} & X_{12} & X_{13} & \dots & X_{1k} \\ X_{12}^T & X_{22} & 0 & \dots & 0 \\ X_{13}^T & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ X_{1k}^T & 0 & \dots & 0 & X_{kk} \end{pmatrix}, \tag{7}$$

where $k \geq 3$, each X_{ii} is an $n_i \times n_i$ matrix, and $\sum_{i=1}^k n_i = n$.

Lemma 1. *Suppose that $X \in \mathcal{D}_n$ has the form* (7)*, $k \geq 3$, and let*

$$X^i = \begin{pmatrix} X_{11} & X_{1i} \\ X_{1i}^T & X_{ii} \end{pmatrix}, i = 2, \ldots, k.$$

Then $X \in \mathcal{C}_n$ if and only if there are matrices A_{ii}, $i = 2, \ldots, k$ such that $\sum_{i=2}^k A_{ii} = X_{11}$, and

$$\begin{pmatrix} A_{ii} & X_{1i} \\ X_{1i}^T & X_{ii} \end{pmatrix} \in \mathcal{C}_{n_1+n_i}, \quad i = 2, \ldots, k.$$

Moreover, if $G(X^i)$ is a CP graph for each $i = 2, \ldots, k$, then the above statement remains true with $\mathcal{C}_{n_1+n_i}$ replaced by $\mathcal{D}_{n_1+n_i}$.

Proof. Let $N_1 = 0$, $N_i = N_{i-1} + n_i$, $i = 2, \ldots, k$, and $\mathcal{I}_i = \{N_i + 1, \ldots, N_i + n_i\}$, $i = 1, \ldots, k$. Then $\mathcal{I}_i$ contains the indeces of the rows and columns of X corresponding to X_{ii}. From the structure of X, it is clear that $X \in \mathcal{C}_n$ if and only if there are nonnegative vectors a^{ij} such that

$$X = \sum_{i=2}^{k} \sum_{j=1}^{m_i} a^{ij} (a^{ij})^T,$$

where $a_l^{ij} = 0$ for $l \notin \mathcal{I}_1 \cup \mathcal{I}_i$. The lemma follows by deleting the rows and columns of $\sum_{j=1}^{m_i} a^{ij}(a^{ij})^T$ that are not in $\mathcal{I}_1 \cup \mathcal{I}_i$. That $\mathcal{C}_{n_1+n_i}$ can be replaced by $\mathcal{D}_{n_1+n_i}$ when $G(X^i)$ is a CP graph follows from Definition 1 and the fact that $G(A_{ii})$ must be a subgraph of $G(X_{11})$ for each i. □

Theorem 8. *Suppose that $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ has the form* (7)*, where $G(X^i)$ is a CP graph, $i = 2, \ldots, k$. Then there is a matrix*

$$V = \begin{pmatrix} V_{11} & V_{12} & V_{13} & \ldots & V_{1k} \\ V_{12}^T & V_{22} & 0 & \ldots & 0 \\ V_{13}^T & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ V_{1k}^T & 0 & \ldots & 0 & V_{kk} \end{pmatrix}$$

such that

$$\begin{pmatrix} V_{11} & V_{1i} \\ V_{1i}^T & V_{ii} \end{pmatrix} \in \mathcal{D}_{n_1+n_i}^*, \quad i = 2, \ldots, k,$$

and $V \bullet X < 0$. Moreover, the matrix

$$\tilde{V} = \begin{pmatrix} V_{11} & \ldots & V_{1k} \\ \vdots & \ddots & \vdots \\ V_{1k}^T & \ldots & V_{kk} \end{pmatrix},$$

where $V_{ij} = \sqrt{\mathrm{diag}(V_{ii})}\sqrt{\mathrm{diag}(V_{jj})}^T$, $2 \leq i \neq j \leq k$, has $\tilde{V} \in \mathcal{C}_n^*$ and $\tilde{V} \bullet X = V \bullet X < 0$.

Proof. The proof uses an argument very similar to that used to prove Theorems 6 and 7. □

When $n_1 = 2$, a matrix X satisfying the conditions of Theorem 8 has a book graph with k CP pages, and an algebraic procedure for testing if $X \in \mathcal{C}_n$ is known [Bar98]. There are several situations where we can immediately use Theorem 8 to separate a matrix of the form (7) with $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ from $\mathcal{C}_n$. One case, corresponding to $n_1 = 3$ and $n_i = 1$, $i = 2, \ldots, k$ can be viewed as a generalization of the separation procedure for a matrix of the form (1) in the previous section. Assuming that X has no zero diagonal components, after a symmetric permutation and diagonal rescaling, and setting $k \leftarrow k-1$, we may assume that such a matrix X has the form

$$X = \begin{pmatrix} X_{11} & \alpha_1 & \alpha_2 & \ldots & \alpha_k \\ \alpha_1^T & 1 & 0 & \ldots & 0 \\ \alpha_2^T & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \alpha_k^T & 0 & \ldots & 0 & 1 \end{pmatrix}. \tag{8}$$

Corollary 1. *Suppose that $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ has the form (8), where $X_{11} \in \mathcal{D}_3$. Then there is a matrix*

$$V = \begin{pmatrix} V_{11} & B \\ B^T & \mathrm{Diag}(\gamma) \end{pmatrix}$$

where $V_{11} \in \mathcal{D}_3^*$, $B = (\beta_1, \beta_2, \ldots, \beta_k)$ and $\gamma \geq 0$ such that

$$\begin{pmatrix} V_{11} & \beta_i \\ \beta_i^T & \gamma_i \end{pmatrix} \in \mathcal{D}_4^*, \quad i = 1, 2, \ldots, k$$

and $V \bullet X < 0$. Moreover, if $s = \sqrt{\gamma}$ then

$$V(s) = \begin{pmatrix} V_{11} & B \\ B^T & ss^T \end{pmatrix} \in \mathcal{C}_n^*.$$

Note that in the case where X has the structure (8), the vertices $4, \ldots, k+3$ form a stable set of size k in $G(X)$. A second case where the structure (7) can immediately be used to generate a cut separating a matrix $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ is when $n_i = 2$, $i = 1, \ldots, k$. In this case the subgraph of $G(X)$ on the vertices $3, \ldots, 2k$ is a matching, and the matrices

$$\begin{pmatrix} V_{11} & V_{1i} \\ V_{1i}^T & V_{ii} \end{pmatrix}$$

in Theorem 8 are again all in $\mathcal{D}_4^*$.

A second case where block structure can be used to generate cuts for a matrix $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ is when X has the form

$$X = \begin{pmatrix} I & X_{12} & X_{13} & \ldots & X_{1k} \\ X_{12}^T & I & X_{23} & \ldots & X_{2k} \\ X_{13}^T & X_{23}^T & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & X_{(k-1)k} \\ X_{1k}^T & 0 & \ldots & X_{(k-1)k}^T & I \end{pmatrix}, \tag{9}$$

where $k \geq 2$, each X_{ij} is an $n_i \times n_j$ matrix, and $\sum_{i=1}^k n_i = n$. The structure in (9) corresponds to a partitioning of the vertices $\{1, 2, \ldots, n\}$ into k stable sets in $G(X)$, of size $n_1, \ldots, n_k$ (note that $n_i = 1$ is allowed). In order to succinctly characterize when $X \in \mathcal{C}_n$ for such a matrix it is convenient to utilize a multi-index vector $p \in Z^k$, with components $1 \leq p_i \leq n_i$.

Lemma 2. *Suppose that $X \in \mathcal{D}_n$ has the structure (9). For each $p \in Z^k$ with $1 \le p_i \le n_i$, $i = 1 \ldots, n$, let X^p be the $k \times k$ matrix with components*

$$[X^p]_{ij} = \begin{cases} [X_{ij}]_{p_i p_j} & i \neq j, \\ a_i^p & i = j, \end{cases}$$

where $a^p \in \Re^k$. Then $X \in \mathcal{C}_n$ if and only if there are $a^p \in \Re^k$ so that $X^p \in \mathcal{C}_k$ for each p, and

$$\sum_{p: p_i = j} a_i^p = 1$$

for each $i = 1, \ldots, k$ and $1 \le j \le n_i$. If in addition $G(X^p)$ is a CP graph for each p, then the above statement holds with $\mathcal{D}_k$ in place of $\mathcal{C}_k$.

Proof. The proof is similar to that of Lemma 1, but uses the fact that if X has the form (9) and bb^T is a rank-one matrix in the CP-decomposition of X, then for each $i = 1, \ldots, k$, $b_j > 0$ for at most one $j \in \mathcal{I}(i)$. $\square$

Theorem 9. *Suppose that $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ has the form (9), where $G(X^p)$ is a CP graph for each $p \in Z^k$ with $1 \le p_i \le n_i$, $i = 1, \ldots, k$. Then there is a matrix*

$$V = \begin{pmatrix} \mathrm{Diag}(\gamma^1) & V_{12} & V_{13} & \ldots & V_{1k} \\ V_{12}^T & \mathrm{Diag}(\gamma^2) & V_{23} & \ldots & V_{2k} \\ V_{13}^T & V_{23}^T & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & V_{(k-1)k} \\ V_{1k}^T & 0 & \ldots & V_{(k-1)k}^T & \mathrm{Diag}(\gamma^k) \end{pmatrix},$$

with $V \bullet X < 0$, such that for every $p \in Z^k$ with $1 \le p_i \le n_i$, $i = 1, \ldots, k$, $V^p \in \mathcal{D}_k^*$, where

$$[V^p]_{ij} = \begin{cases} [V_{ij}]_{p_i p_j} & i \neq j, \\ \gamma_{p_i}^i & i = j. \end{cases}$$

Moreover, the matrix

$$\tilde{V} = \begin{pmatrix} V_{11} & \ldots & V_{1k} \\ \vdots & \ddots & \vdots \\ V_{1k}^T & \ldots & V_{kk} \end{pmatrix},$$

where $V_{ii} = \sqrt{\gamma^i} \sqrt{\gamma^i}^T$, has $\tilde{V} \in \mathcal{C}_n^*$ and $\tilde{V} \bullet X = V \bullet X < 0$.

Proof. The proof uses an argument very similar to that used to prove Theorems 6 and 7. $\square$

To close the section we mention two details regarding the applicability of Lemmas 1 and 2 to separate a given $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ from $\mathcal{C}_n$, via the cuts described in Theorems 8 and 9. First, in practice a given matrix X may have numerical entries that are small but not exactly zero. In such a case, Lemma 1 or 2 can be applied to a perturbed matrix $\tilde{X}$, where entries of X below a specified tolerance are set to zero in $\tilde{X}$. If a cut V separating $\tilde{X}$ from $\mathcal{C}_n$ is found and the zero tolerance is small, then $V \bullet X \approx V \bullet \tilde{X} < 0$, and V is very likely to also separate X from $\mathcal{C}_n$. Second, it is important to recognize that in practice Theorems 8 and

9 may provide a cut separating a given $X \in \mathcal{D}_n \setminus \mathcal{C}_n$ even when the sufficient conditions for generating such a cut are not satisfied. In particular, a cut of the form described in Theorem 8 may be found even when the condition that X^i is a CP graph for each i is not satisfied; similarly a cut of the form described in Theorem 9 may be found even when the condition that X^p is a CP graph for each p is not satisfied.

5 Applications

In this section we describe the results of applying the separation procedures developed in the paper to selected test problems. Consider an indefinite quadratic programming problem of the form

$$\begin{array}{lll} \text{(QP)} & \max & x^TQx + c^Tx \\ & \text{s.t.} & Ax = b \\ & & x \geq 0, \end{array}$$

where A is an $m \times n$ matrix. For the case of general Q, (QP) is an NP-Hard problem. Next define the matrices

$$Y = \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix}, \quad \tilde{Q} = \begin{pmatrix} 0 & c^T/2 \\ c/2 & Q \end{pmatrix}, \tag{10}$$

and let

$$QP(A,b) = \text{Co}\left\{ \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T : \; Ax = b, \; x \geq 0 \right\}. \tag{11}$$

Since the extreme points of $QP(A,b)$ correspond to feasible solutions of (QP), (QP) can be written as the linear optimization problem

$$\begin{array}{lll} \text{(QP)} & \max & \tilde{Q} \bullet Y \\ & \text{s.t.} & Y \in QP(A,b). \end{array}$$

The connection between (QP) and the topic of the paper is the following result showing that $QP(A,b)$ can be exactly represented using the CP cone.

Theorem 10. [Bur09] *Assume that $\{x : Ax = b, \; x \geq 0\}$ is bounded, and let $QP(A,b)$ be defined as in* (11). *Then $QP(A,b) = \{Y \in \mathcal{C}_{n+1} : a_i^Tx = b_i, a_i^TXa_i = b_i^2, i = 1, \ldots, m\}$.*

One well-studied case of (QP) is the box-constrained quadratic program

$$\begin{array}{lll} \text{(QPB)} & \max & x^TQx + c^Tx \\ & \text{s.t.} & 0 \leq x \leq e. \end{array}$$

In order to linearize (QPB) one can define Y and $\tilde{Q}$ as in (10) and write the objective as $\tilde{Q} \bullet Y$. There are then a number of different constraints that can be imposed on Y. For example, Y should satisfy with the well-known Reformulation-Linearization Technique (RLT) constraints

$$\{0, x_i + x_j - 1\} \leq x_{ij} \leq \{x_i, x_j\}, \quad 1 \leq i,j \leq n, \tag{12}$$

as well as the PSD condition $Y \succeq 0$. To apply Theorem 10 to (QPB), one can add slack variables s and write the constraints as $x + s = e$, $(x, s) \geq 0$. It is then natural to define an augmented matrix

$$Y^+ = \begin{pmatrix} 1 & x^T & s^T \\ x & X & Z \\ s & Z^T & S \end{pmatrix},$$

where S and Z relax ss^T and xs^T, respectively. The "squared constraints" from Theorem 10 then have the form $\mathrm{diag}(X + 2Z + S) = e$, and the CP representation of (QPB) can be written

$$\begin{array}{lll} (\mathrm{QPB})_{\mathrm{CP}} & \max & x^T Q x + c^T x \\ & \text{s.t.} & x + s = e, \quad \mathrm{Diag}(X + 2Z + S) = e, \\ & & Y^+ \in \mathcal{C}_{2n+1}. \end{array}$$

It can also be shown [AB10, Bur10] that replacing $\mathcal{C}_{2n+1}$ in $(\mathrm{QPB})_{\mathrm{CP}}$ with $\mathcal{D}_{2n+1}$ is equivalent to solving the relaxation of (QPB) that imposes the PSD condition $Y \succeq 0$ together with the RLT constraints (12). Moreover, this relaxation is tight for $n = 2$ but may not be for $n \geq 3$. In [BL09] it is shown that for (QPB) the off-diagonal components of Y can be constained to be in the Boolean Quadric Polytope (BQP). As a result, valid inequalities for the BQP can be imposed on the off-diagonal components of Y, an approach that was first suggested in [YF98]. For $n = 3$, the BQP is fully characterized by the RLT constraints and the well-known triangle (TRI) inequalities. However, it is shown in [BL09] that for $n \geq 3$, the PSD, RLT and TRI constraints on Y are still not sufficient to exactly represent (QPB). This is done by considering the (QPB) instance with $n = 3$ and

$$Q = \begin{pmatrix} -2.25 & -3.00 & -3.00 \\ -3.00 & 0.00 & -0.50 \\ -3.00 & -0.50 & 1.00 \end{pmatrix}, \qquad c = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}. \tag{13}$$

It is shown in [BL09] that the solution value for (QPB) with the data (13) is 1.0, but the maximum of $\tilde{Q} \bullet Y$ where $Y \succeq 0$ satisfies the RLT and TRI constraints is approximately 1.093.

For the data in (13), solving $\mathrm{QPB_{CP}}$, with $\mathcal{D}_7$ in place of $\mathcal{C}_7$ and the triangle inequalities added, results in the 7×7 matrix

$$Y^+ \approx \begin{pmatrix} 1.0000 & 0.1478 & 0.5681 & 0.5681 & 0.8522 & 0.4319 & 0.4319 \\ 0.1478 & 0.0901 & 0.0000 & 0.0000 & 0.0577 & 0.1478 & 0.1478 \\ 0.5681 & 0.0000 & 0.5681 & 0.2841 & 0.5681 & 0.0000 & 0.2840 \\ 0.5681 & 0.0000 & 0.2841 & 0.5681 & 0.5681 & 0.2840 & 0.0000 \\ 0.8522 & 0.0577 & 0.5681 & 0.5681 & 0.7944 & 0.2840 & 0.2840 \\ 0.4319 & 0.1478 & 0.0000 & 0.2840 & 0.2840 & 0.4319 & 0.1478 \\ 0.4319 & 0.1478 & 0.2840 & 0.0000 & 0.2840 & 0.1478 & 0.4319 \end{pmatrix}.$$

It is known [AB10] that Y^+ is CP if and only if the 6×6 matrix obtained by deleting its first row and column is CP. Further deleting the fifth row and column results in a 5×5 submatrix of Y^+ which is not CP. In fact, this 5×5 matrix meets all the conditions of Theorem 2, so a transformed Horn cut that separates Y^+ from $\mathcal{C}_7$ can be generated. Alternatively, a cut from

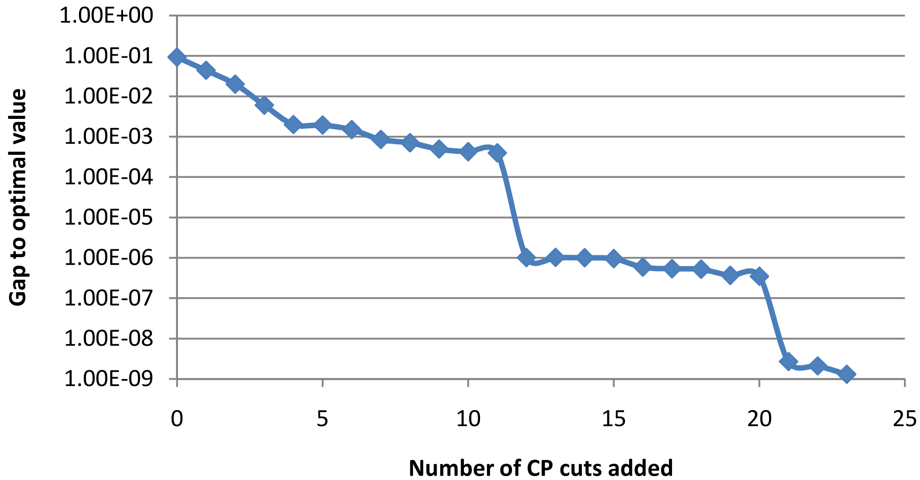

Figure 1: Gap to optimal value for Burer-Letchford QPB problem ($n = 3$)

Theorem 7 can be used. In either case imposing the new constraint and re-solving results in a new matrix Y^+ with lower objective value. The same 5×5 principal submatrix again fails to be CP, so another cut can be generated and the process repeated. In Figure 1 we show the effect of executing this process using the cuts from Theorem 7, normalized using $I \bullet V = 1$. After adding 21 cuts, the gap to the true solution value of the problem is reduced below 10^{-8}. (The rank condition in Theorem 2 fails after 3 cuts are added, but this has no effect on the cuts from Theorem 7.)

Before continuing, we remark that there are other known approaches that obtain an optimal value for the (QPB) instance (13) without branching. For example, in [AB10] it is shown that for $n = 3$, (QPB) can be exactly represented using DNN matrices via a triangulation of the 3-cube. This representation provides a tractable formulation that returns the exact optimal solution value for the original problem. A completely different approach is based on using the result of Theorem 10 and writing a dual for $(\mathrm{QPB})_{\mathrm{CP}}$ that involves the cone $\mathcal{C}^*_{n+1}$. It then turns out that using the cone $\mathcal{K}^1_7$ as an inner approximation of $\mathcal{C}^*_7$ also obtains the exact solution value for the Burer-Letchford instance (13). It should be noted, however, that both of these approaches involve "extended variable" formulations of the problem, whereas the procedure based on adding cuts operates in the original problem space.

Next we consider the problem of computing the maximum stable set in a graph. Let A be the adjacency matrix of a graph G on n vertices, and let α be the maximum size of a stable set in G. It is known [dKP02] that

$$\alpha^{-1} = \min\left\{(I + A) \bullet X : ee^T \bullet X = 1, X \in \mathcal{C}_n\right\}. \tag{14}$$

Relaxing $\mathcal{C}_n$ to $\mathcal{D}_n$ results in the Lovász-Schrijver bound

$$(\vartheta')^{-1} = \min\left\{(I + A) \bullet X : ee^T \bullet X = 1, X \in \mathcal{D}_n\right\}. \tag{15}$$

The bound ϑ' was first established (via a different derivation) by Schrijver as a strengthening of Lovász's ϑ number.

For our first example, with $n = 12$, let G_{12} be the complement of the graph corresponding to the vertices of a regular icosahedron [BdK02]. Then $\alpha = 3$ and $\vartheta' \approx 3.24$, a gap of approximately 8%. A notable feature of the stable set problem for G_{12} is that using the cone $\mathcal{K}^1_{12}$ to approximate the dual of (14) provides no improvement over $\mathcal{K}^0_{12}$, corresponding to the dual of (15) [BdK02]. For the solution matrix X from (15), the incidence matrix of $G(X)$ is

$$\begin{pmatrix} 0&1&1&1&1&1&0&0&0&0&0&0\\ 1&0&1&1&0&0&1&1&0&0&0&0\\ 1&1&0&0&1&0&1&0&1&0&0&0\\ 1&1&0&0&0&1&0&1&0&1&0&0\\ 1&0&1&0&0&1&0&0&1&0&1&0\\ 1&0&0&1&1&0&0&0&0&1&1&0\\ 0&1&1&0&0&0&0&1&1&0&0&1\\ 0&1&0&1&0&0&1&0&0&1&0&1\\ 0&0&1&0&1&0&1&0&0&0&1&1\\ 0&0&0&1&0&1&0&1&0&0&1&1\\ 0&0&0&0&1&1&0&0&1&1&0&1\\ 0&0&0&0&0&0&1&1&1&1&1&0 \end{pmatrix}. \tag{16}$$

It is then easy to see that there are many principal submatrices of the matrix X, up to size 9×9, that meet the conditions of Lemma 1. For example, consider the principal submatrix formed by omitting rows and columns $\{6, 8, 9\}$, partitioned into the sets $\mathcal{I}_1 = \{1, 2, 3, 10, 11, 12\}$, $\mathcal{I}_2 = \{4\}$, $\mathcal{I}_3 = \{5\}$, $\mathcal{I}_4 = \{7\}$. After a suitable permutation this 9×9 matrix has the form (7), with $n_1 = 6$, $n_2 = n_3 = n_4 = 1$, and it is easy to see that $G(X^i)$ is a CP graph for $i = 2, 3, 4$. However, it turns out that matrices A_{ii} satisfying the DNN feasibility constraints of Lemma 1 exist, demonstrating that this principal submatrix is in fact CP. This was the case for every principal submatrix satisfying the conditions of Lemma 1 that we examined.

We next consider applying Lemma 2. It turns out that the vertices of $G(X)$ can be partitioned into 4 disjoint stable sets of size 3; one such partition uses the sets $\{1, 7, 10\}$, $\{2, 6, 9\}$, $\{3, 8, 11\}$ and $\{4, 5, 12\}$. Then Lemma 2 applies, and $G(X^p)$ is a CP graph for each p since each X^p is a 4×4 matrix. The DNN feasibility system in Lemma 2 does not have a solution, so Theorem 9 can be used to generate a cut separating X from $\mathcal{C}_{12}$. Adding this one cut and re-solving, the gap to $1/\alpha = \frac{1}{3}$ is approximately 2×10^{-8}.

Before proceeding, it is worthwhile to note that for the stable set problem for G_{12}, Lemma 2 could actually be used to *reformulate* the problem (14) so that the exact solution value $\alpha = 1/3$ is obtained without adding any cuts. To see this, note that for problem (14) on a graph G with adjacency matrix A, it is obvious that $x_{ij} = 1 \implies a_{ij} = 0$, or in other words $G(X) \subset \overline{G}$, where $\overline{G}$ is the complement of G. For the graph G_{12}, the adjacency matrix for $\overline{G}$ is precisely that of $G(X)$ in (16). Applying Lemma 2, the condition $X \in \mathcal{C}_{12}$ in (14) can be replaced by an equivalent condition involving 81 matrices in $\mathcal{D}_4$, and the problem solved exactly. Compared to the procedure of first solving a relaxation over $\mathcal{D}_{12}$ and then adding a cut based on the solution $G(X)$, the reformulation has the advantage that only one problem is solved. However, prior knowledge of the underlying problem structure is essential to the reformulation but is unused by the procedure based on adding cuts.

We next consider several stable set problems from [PVZ07]. These problems, of sizes $n \in \{8, 11, 14, 17\}$, are specifically constructed to be difficult for SDP-based relaxations such as (15). We refer to the underlying graphs as G_8, G_{11}, G_{14} and G_{17}. First consider G_8, for

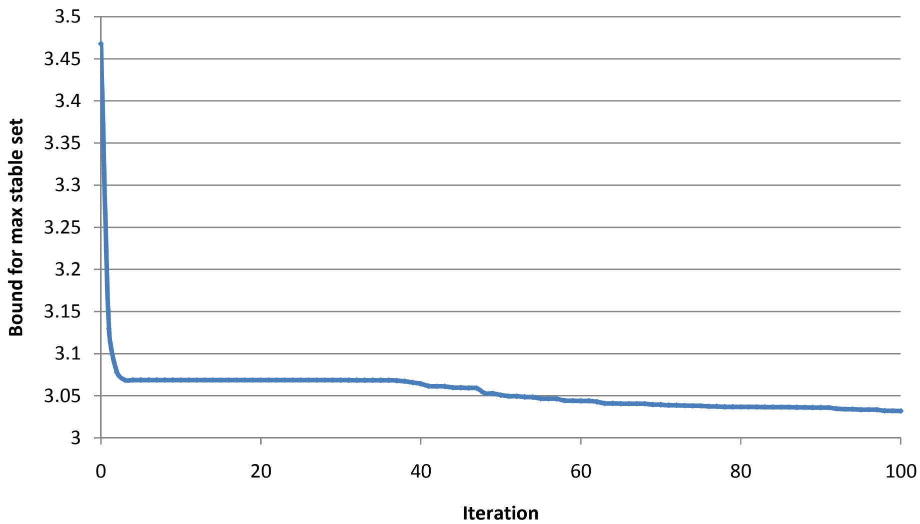

Figure 2: Bounds on max stable set for G_8

which $\alpha = 3$. The incidence matrix of $\overline{G}_8$ is

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \tag{17}$$

It is clear that the vertices of $\overline{G}_8$ can be partitioned into 4 stable sets of size 2. Since we could assume that $G(X) \in \overline{G}_8$ (see the discussion above), Lemma 2 could be used to reformulate the problem (14), as described above, so that the exact value of α is obtained. Instead, we apply the methodology of first solving the problem (15), and obtain a solution $X = X^0 \in \mathcal{D}_{12}$ with $G(X^0)$ as in (17). For each partitioning of $G(X^0)$ into 4 stable sets of size 2 we generate a cut from Theorem 9, and the augmented problem is then re-solved to obtain a new solution X^1. We then attempt to generate additional cuts based on $G(X^1)$, re-solve the problem to get a new solution X^2, and continue. On iteration i we generate one cut for each partitioning of $G(X^i)$ into 4 stable sets of size 2. The resulting bounds on the max stable set obtained for 100 iterations of this procedure are illustrated in Figure 2. The bound on $\alpha = 3$ drops to approximately 3.078 on iteration 2, and then decreases slowly on subsequent iterations. The number of cuts generated is 3 on iterations 0 and 1, 1 on iterations 2-30, and 3 or 4 per iteration thereafter. By comparison, in [BFL10] the gap to α for this problem is closed to zero, using copositivity cuts specialized for the max-clique problem posed on the complement graph $\overline{G}_8$.

For the graphs G_n, $n \in \{11, 14, 17\}$, it is not possible to use Lemma 7 or Lemma 9 to exactly reformulate the condition $X \in \mathcal{C}_n$ in terms of DNN matrices. For these problems we used the following computational approach. We first solved (15) to obtain the solution X and then found all possible structures consisting of 4 disjoint stable sets of size 2 in $G(X)$. (The

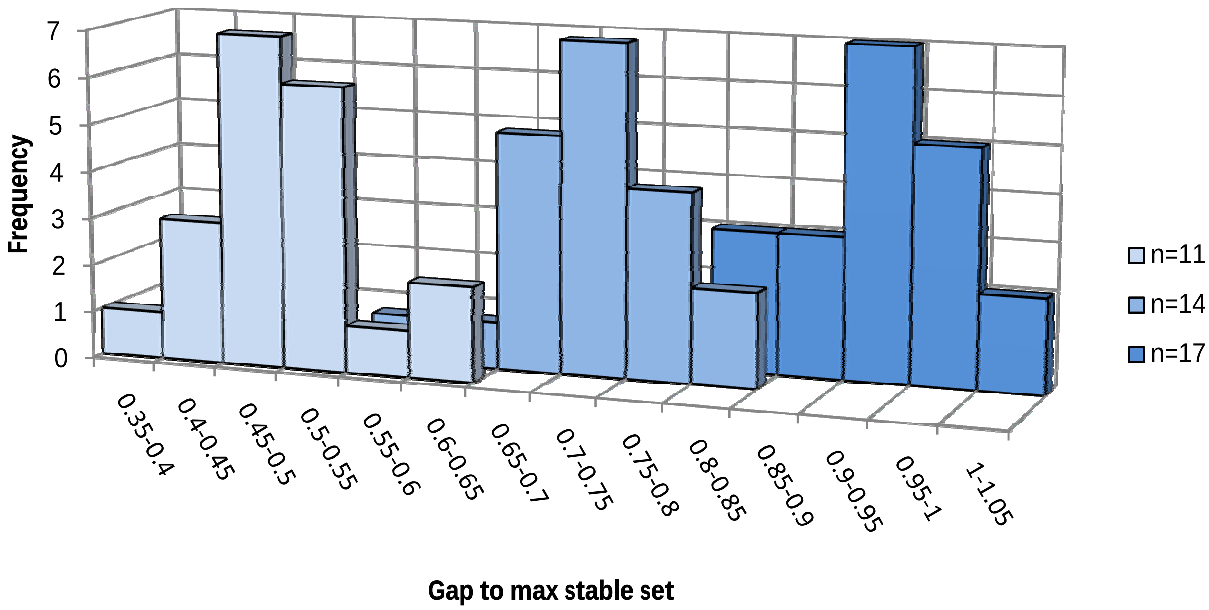

Figure 3: Gaps for stable set problems after addition of CP cuts

complement graphs $\overline{G}_i$ for these problems, like $\overline{G}_8$, contain no stable sets of size greater than 3.) We then randomly chose k of these structures to try to generate cuts based on Lemma 9 applied to the corresponding 8×8 principal submatrices of X. We used $k = 10, 15, 20$, respectively for the problems with $n = 11, 14, 17$. After adding all of the cuts found, we re-solved the problem to get a new bound on α. We performed the entire procedure 20 times for each problem, each time starting with the original solution X and randomly generating a new set of k structures from which to potentially generate cuts. The results are summarized in Figure 3 and Table 1. In Figure 3 we give the distribution of the gaps to α for the 20 bounds obtained for each problem. One clearly sees the increase in the average gap as n increases, and the variability in the gaps obtained for each problem. Note that for G_{17}, for which $\alpha = 6$ and $\vartheta' > 7$, the bound after adding CP cuts was below 7 on 18 of 20 runs. In Table 1 we give summary statistics for the results on each problem, including the min, mean and max bound obtained on the 20 runs, and the min, median and max number of cuts found on the k tries for each run. For comparison we also give the best bound ϑ^{cop} found in [BFL10]. From Table 1 it is clear that our results on these problems, while substantially improving on ϑ', are not as good as the best bounds obtained in [BFL10]. This is perhaps to be expected since:

1. Our methodology, unlike that of [BFL10], makes no explicit use of the underlying structure of the original problem (max stable set/max clique);

2. The results we report here are based on adding only 20 cuts to the original DNN solution $X^0 = X$, and could certainly be improved by adding more cuts to X^0 and/or adding cuts to subsequent solutions X^i, $i > 0$, as in the results reported for G_8 above.

Table 1: Results on stable set problems (20 runs for each problem)

Graph	α	ϑ'	ϑ^{cop}	Number of cuts: tries	min	median	max	Bound values: min	mean	max
G_{11}	4	4.694	4.280	10	8	9	10	4.352	4.500	4.634
G_{14}	5	5.916	5.485	15	6	10	15	5.595	5.722	5.825
G_{17}	6	7.134	6.657	20	6	10	15	6.828	6.930	7.034

6 Conclusion

It is known that completely positive matrices can be used to formulate a variety of NP-Hard optimization problems, such as the formulations for nonconvex (QP) in Theorem 10 and the maximum stable set problem in (14). Previous research based on such formulations has primarily focussed on obtaining improved approximations of the cones of copositive matrices associated with the dual problems. In this paper we suggest an alternative methodology based on better approximating the CP condition by adding copositive cuts to a DNN relaxation of the primal problem. Computational results on a number of small, difficult instances illustrate the promising potential of this approach.

References

[AB10] Kurt M. Anstreicher and Samuel Burer, *Computable representations for convex hulls of low-dimensional quadratic forms*, Math. Prog. B **to appear** (2010).

[BAD09] Samuel Burer, Kurt M. Anstreicher, and Mirjam Dür, *The difference between 5 × 5 doubly nonnegative and completely positive matrices*, Linear Algebra Appl. **431** (2009), 1539–1552.

[Bar98] Francesco Barioli, *Completely positive matrices with a book-graph*, Linear Algebra Appl. **277** (1998), no. 1-3, 11–31.

[BD09] Stefan Bundfuss and Mirjam Dür, *An adaptive linear approximation algorithm for copositive programs*, SIAM J. Optim. **20** (2009), no. 1, 30–53.

[BdK02] Immanuel M. Bomze and Etienne de Klerk, *Solving standard quadratic optimization problems via linear, semidefinite and copositive programming*, J. Global Optim. **24** (2002), 163–185.

[BFL10] Immanuel M. Bomze, Florian Frommlet, and Marco Locatelli, *Copositivity cuts for improving SDP bounds on the clique number*, Math. Prog. B **to appear** (2010).

[BL09] Samuel Burer and Adam N. Letchford, *On nonconvex quadratic programming with box constriants*, SIAM J. Optim. **20** (2009), no. 2, 1073–1089.

[BSM03] Abraham Berman and Naomi Shaked-Monderer, *Completely Positive Matrices*, World Scientific, 2003.

[Bur09] Samuel Burer, *On the copositive representation of binary and continuous nonconvex quadratic programs*, Math. Prog. **120** (2009), no. 2, 479–495.

[Bur10] ______, *Optimizing a polyhedral-semidefinite relaxation of completely positive programs*, Math. Prog. Comp. **to appear** (2010).

[BX04] Abraham Berman and Changqing Xu, 5×5 *Completely positive matrices*, Linear Algebra Appl. **393** (2004), 55–71.

[DA09] Hongbo Dong and Kurt Anstreicher, *On '5×5 Completely positive matrices'*, Working paper, Dept. of Management Sciences, University of Iowa (2009).

[dKP02] Etienne de Klerk and Dmitrii V. Pasechnik, *Approximation of the stability number of a graph via copositive programming*, SIAM J. Optim. **12** (2002), no. 4, 875–892.

[DS08] Mirjam Dür and Georg Still, *Interior points of the completely positive cone*, Electron. J. Linear Algebra **17** (2008), 48–53.

[HJ67] Marshall Hall Jr., *Combinatorial Theory*, Blaisdell Publishing Company, 1967.

[HJR05] Leslie Hogben, Charles R. Johnson, and Robert Reams, *The copositive completion problem*, Linear Algebra Appl. **408** (2005), 207–211.

[KB93] Natalia Kogan and Abraham Berman, *Characterization of completely positive graphs*, Discrete Math. **114** (1993), 297–304.

[PVZ07] Javier Peña, Juan Vera, and Luis F. Zuluaga, *Computing the stability number of a graph via linear and semidefinite programming*, SIAM J. Optim. **18** (2007), no. 1, 87–105.

[Stu99] Jos F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods and Software **11-12** (1999), 625–653.

[Väl89] Hannu Väliaho, *Almost copositive matrices*, Linear Algebra Appl. **116** (1989), 121–134.

[YF98] Yasutoshi Yajima and Tetsuya Fujie, *A polyhedral approach for nonconvex quadratic programming problems with box constraints*, J. Global Optim. **13** (1998), 151–170.