A *Short Course* on Semidefinite Programming

(in order of appearance)

Henry Wolkowicz, Etienne de Klerk, Levent Tunçel, Franz Rendl

hwolkowicz@uwaterloo.ca

Department of Combinatorics and Optimization

University of Waterloo

About these Notes:

Semidefinite Programming, SDP, refers to optimization problems where the vector variable is a symmetric matrix which is required to be positive semidefinite. Though SDPs (under various names) have been studied as far back as the 1940s, the interest has grown tremendously during the last ten years. This is partly due to the many diverse applications in e.g. engineering, combinatorial optimization, and statistics. Part of the interest is due to the great advances in efficient solutions for these types of problems.

These notes summarize the theory, algorithms, and applications for semidefinite programming. They were prepared for a minicourse given at the Workshop on **Large Scale Nonlinear and Semidefinite Programming**, held in memory of Jos Sturm, May 12, 2004, at the University of Waterloo.

Preface

The purpose of these notes is three fold: first, they provide a comprehensive treatment of the area of Semidefinite Programming, a new and exciting area in Optimization; second, the notes illustrate the strength of convex analysis in Optimization; third, they emphasize the interaction between theory and algorithms and solutions of practical problems.

1 Introduction and Motivation

1.1 Outline

- Basic Properties and Notation
- Examples
- Historical Notes

1.2 Basic Properties and Notation

Basic linear **Semidefinite Programming** looks just like Linear Programming

$$
(\mathbf{PSDP}) \qquad \begin{array}{rccc}
p^* = & \max & \operatorname{trace} CX & (\langle C, X\rangle) \\
& \text{s.t.} & \mathcal{A}(X) \cong \mathcal{A}X = b & \text{(linear)} \\
& & X \succeq 0, \;\; (X \in \mathcal{P}) & \text{(nonneg)}
\end{array}
$$

$C, X \in \mathcal{S}^n$
$\mathcal{S}^n$:= space of $n \times n$ symmetric matrices

Space of Symmetric Matrices

$A \in \mathcal{S}^n$:= space of $n \times n$ symmetric matrices

A is positive semidefinite (positive definite) $\quad (A \succeq 0 \; (A \succ 0))$
if $x^T A x \geq 0 (> 0), \; \forall x \neq 0.$

$\preceq$ denotes the Löwner partial order, [53]
$A \preceq B$ if $B - A \succeq 0$ (positive semidefinite)

TFAE:

1. $A \succeq 0 \quad (A \succ 0)$
2. the eigenvalues $\lambda(A) \geq 0 \quad (\lambda(A) > 0)$

Linear transformation $\mathcal{A}$; Adjoint:

$$\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m, \qquad \mathcal{A}^* : \mathbb{R}^m \to \mathcal{S}^n$$

$\mathcal{A}$ defined by associated set $\{A_i \in \mathcal{S}^n, i = 1, \ldots m\}$:

$$(\mathcal{A}X)_i = \operatorname{trace}(A_i X); \quad \mathcal{A}^* y = \sum_{i=1}^m y_i A_i,$$

where

$$\langle \mathcal{A}X, y \rangle = \langle X, \mathcal{A}^* y \rangle, \quad \forall X \in \mathcal{S}^n, \forall y \in \mathbb{R}^m$$

$\mathcal{P} = \mathcal{S}^n_+$ - cone of positive semidefinite matrices
replaces

$\mathbb{R}^n_+$ - nonnegative orthant

SDP or LMI

$$\operatorname{trace} CX = \langle C, X\rangle = \sum_{i=1}^n \sum_{j=1}^n C_{ij} X_{ij}$$

SDP is equivalent to (Linear Matrix Inequalities):

$$
\begin{array}{rrll}
& p^* = & \max & \sum_{i=1}^n \sum_{j=1}^n C_{ij} X_{ij} \\
(\mathbf{PSDP}) & & \text{s.t.} & \langle A_i, X\rangle = b_i, \qquad i = 1, \ldots m \\
& & & X \succeq 0, \quad (X \in \mathcal{P})
\end{array}
$$

1.2.1 Duality

payoff function; player Y to player X (Lagrangian)

$$L(X,y) := \operatorname{trace}(CX) + y^T(b - \mathcal{A}X) \ \ (= \langle C, X\rangle + \langle y, b - \mathcal{A}X\rangle)$$

Optimal (worst case) strategy for player X:

$$p^* = \max_{X \succeq 0} \min_{y} L(X,y)$$

For each fixed $X \succeq 0$: y free yields hidden constraint $b - \mathcal{A}X = 0$; recovers primal problem (PSDP).

Rewrite Lagrangian/payoff

$$
\begin{aligned}
L(X,y) &= \operatorname{trace}(CX) + y^T(b - \mathcal{A}X) \\
&= b^T y + \operatorname{trace}\,(C - \mathcal{A}^* y)\, X
\end{aligned}
$$

using adjoint operator, $\mathcal{A}^* y = \sum_i y_i A_i$

$$
\langle \mathcal{A}^* y, X \rangle = \langle y, \mathcal{A}X \rangle\,, \quad \forall X, y
$$

Then:

$$
p^* = \max_{X \succeq 0} \min_{y} L(X,y) \leq d^* := \min_{y} \max_{X \succeq 0} L(X,y)
$$

For dual: for each fixed y, that $X \succeq 0$ yields hidden constraint $g(y) := C - \mathcal{A}^* y \preceq 0$

Hidden Constraint

$$p^* = \max_{X \succeq 0} \min_y L(X,y) \leq d^* := \min_y \max_{X \succeq 0} L(X,y)$$

dual obtained from optimal strategy of competing player Y;
use *hidden constraint* $g(y) = C - \mathcal{A}^* y \preceq 0$

$$(\textbf{DSDP}) \qquad \begin{array}{rcc} d^* = & \min & b^T y \\ & \text{s.t.} & \mathcal{A}^* y \succeq C \end{array}$$

for the primal

$$(\textbf{PSDP}) \qquad \begin{array}{rcc} p^* = & \max & \operatorname{trace} CX \\ & \text{s.t.} & \mathcal{A}X = b \\ & & X \succeq 0 \end{array}$$

1.2.2 Weak Duality - Optimality

Proposition 1.2.2.1 *(Weak Duality) If X feas. in (PSDP), y feas. in (DSDP), $Z = \mathcal{A}^* y - C \succeq 0$ is slack variable, then*

$$\operatorname{trace} CX - b^T y = -\operatorname{trace} XZ \leq 0.$$

Proof. (Direct - using: $\operatorname{trace} ZX = \operatorname{trace} X^{1/2} X^{1/2} Z = \operatorname{trace} X^{1/2} Z^{1/2} Z^{1/2} X^{1/2} = \|X^{1/2} Z^{1/2}\|^2 \geq 0$;
Note: $XZ = 0$ iff $\operatorname{trace} XZ = 0$)

$$\begin{aligned} \operatorname{trace} CX - b^T y &= \operatorname{trace} (\mathcal{A}^* y - Z) X - b^T y \\ &= \operatorname{trace} y^T \mathcal{A} X - \operatorname{trace} ZX - b^T y \\ &= y^T (\mathcal{A} X - b) - \operatorname{trace} ZX = -\operatorname{trace} ZX, \end{aligned}$$

■

Characterization of optimality

for the
dual pair $X \succeq 0, y$ (and slack $Z \succeq 0$)

$$\begin{array}{rcl} \mathcal{A}^* y - Z = C & & \text{dual feasibility} \\ AX = b & & \text{primal feasibility} \\ ZX = 0 & & \text{complementary slackness} \end{array}$$

And

$$ZX = \mu I \qquad \text{perturbed}$$

Forms the basis for:
interior point methods (central path X_μ, y_μ, Z_μ);
(primal simplex method, dual simplex method)

Strong Duality

$$
\begin{aligned}
\text{primal value} &= \langle C, X \rangle \\
&= \langle \mathcal{A}^* y - Z, X \rangle && \text{dual feasibility} \\
&= \langle y, \mathcal{A} X \rangle - \langle Z, X \rangle && \text{adjoint} \\
&= \langle y, b \rangle - \langle Z, X \rangle && \text{primal feasibility} \\
&= \langle y, b \rangle && \text{complementary slackness} \\
&= \text{dual value}
\end{aligned}
$$

1.2.3 Preliminary Examples

Example 1.2.3.1 Minimizing the Maximum Eigenvalue
Arises in e.g. stability of differential equations

- *The mathematical problem:*
  - *given* $A(x) \in \mathcal{S}^n$ *depending linearly on vector* $x \in \mathbb{R}^m$
  - *Find* x *to minimize the maximum eigenvalue of* $A(x)$
- *SDP Model:*
  - *the largest eigenvalue* $\lambda_{\max}(A(x)) \leq \alpha$ <u>*iff*</u> $\lambda_{\max}(A(x) - \alpha I) \leq 0$ <u>*iff*</u> $A(x) - \alpha I \preceq 0.$
  - *The DSDP (in dual form) is:*

$$\max -\alpha \qquad \textit{s.t.} \quad A(x) - \alpha I \preceq 0.$$

Pseudoconvex Optimization

Example 1.2.3.2 Pseudoconvex (Nonlinear) Optimization

$$
\text{(PCP)} \qquad \begin{array}{rll} d^* = & \min & \dfrac{(c^T x)^2}{d^T x} \\ & \textit{s.t.} & Ax + b \geq 0, \end{array}
$$

(given $Ax + b \geq 0 \Rightarrow d^T x > 0$)
Then (using 2×2 determinant), (PCP) is equivalent to

$$
\begin{array}{rll} d^* = & \min & t \\ & \textit{s.t.} & \begin{bmatrix} \operatorname{Diag}(Ax+b) & 0 & 0 \\ 0 & t & c^T x \\ 0 & c^T x & d^T x \end{bmatrix} \succeq 0 \end{array}
$$

1.2.4 Further Properties, Definitions

- For matrices P, Q compatible for multiplication,

$$\operatorname{trace} PQ = \operatorname{trace} QP.$$

- Every symmetric matrix S has an orthogonal decomposition $S = Q\Lambda Q^T$, where Λ is diagonal with the eigenvalues of S on the diagonal, and Q has orthonormal columns consisting of eigenvectors of S.
- $X \succeq 0 \Rightarrow P^T X P \succeq 0$ (congruence)

• For S symmetric, TFAE:

1. $S \succeq 0 \quad (\succ 0)$
2. $v^T S v \geq 0, \forall v \in \mathbb{R}^n \quad (v^T S v > 0, \forall 0 \neq v \in \mathbb{R}^n)$
3. $\lambda(S) \geq 0,$ nonneg. eigs $\quad (> \; 0,$ pos. eigs$)$
4. $S = P^T P$ for some P (for some nonsingular P)
5. $S = (S^{\frac{1}{2}})^2$, for some $S^{\frac{1}{2}} \succeq 0 \quad (\succ 0)$

• SDP - Linear Matrix Inequality

$$g(y) = C - \mathcal{A}^* y \preceq 0$$

adjoint operator $\qquad \mathcal{A}^* y = \sum_i y_i A_i, \quad A_i = A_i^T$

• $g(y)$ is a <u>$\mathcal{P}$-convex constraint</u>, i.e.

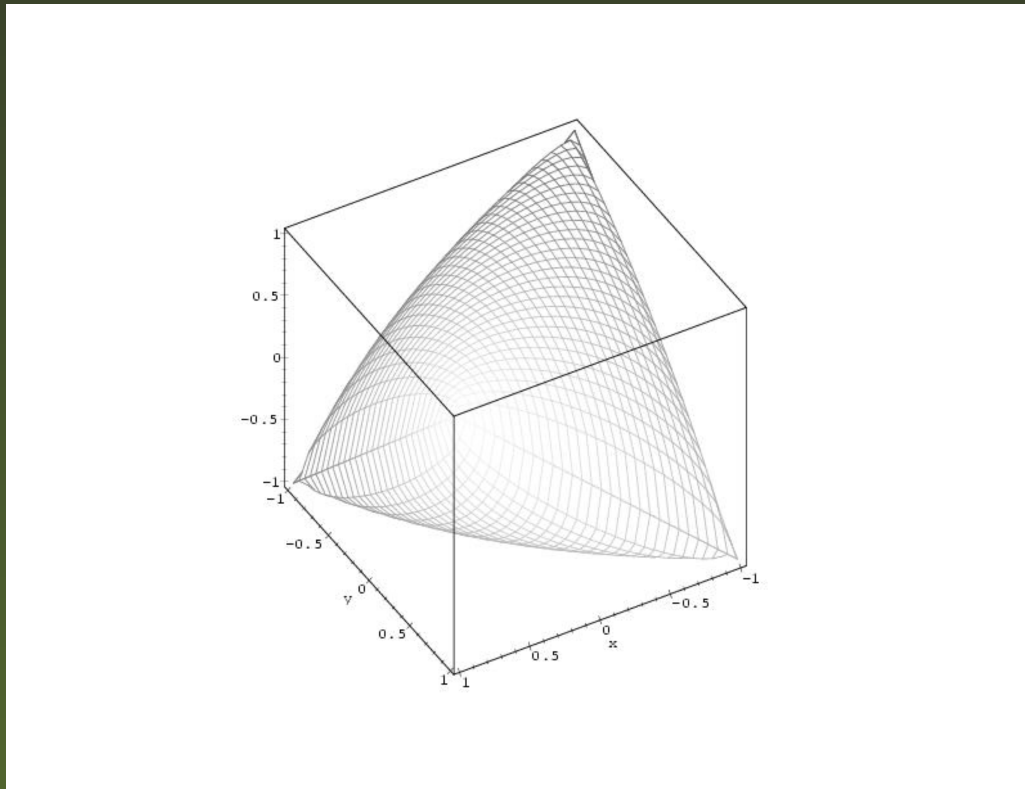
$$g(\lambda x + (1-\lambda) y) \preceq \lambda g(x) + (1-\lambda) g(y), \quad \forall 0 \leq \lambda \leq 1, \forall x, y \in \mathbb{R}^m$$

equivalently, for every $K \succeq 0$, the real valued function $f_K(x) := \operatorname{trace} K g(x)$ is a convex function, i.e.

$$f_K(\lambda x + (1-\lambda) y) \leq \lambda f_K(x) + (1-\lambda) f_K(y), \quad \forall 0 \leq \lambda \leq 1, \forall x, y \in \mathbb{R}^m$$

Geometry - Feasible Set

- *feasible set* $\{x \in \mathbb{R}^m : g(x) \preceq 0\}$ is convex set
- *optimum point* is on the boundary ($g(x)$ is singular)
- *boundary of feasible set* is *not* smooth (piecewise smooth, algebraic surface)



1.2.5 Historical Notes

- arguably most active area in optimization (see HANDBOOK OF SEMIDEFINITE PROGRAMMING: Theory, Algorithms, and Applications, 2000, [94], for comprehensive results, history, references, ... and books [15, 95, 96])
- Lyapunov over 100 years ago on stability analysis of differential equations
- Bohnenblust 1948 on geometry of the cone of SDPs, [10]
- Yakubovitch in the 1960's and Boyd and others on convex optimization in control in the 1980's, e.g. solving Ricatti Equations (called LMIs), e.g. [13],[92, 91]

More: Historical Notes

• matrix completion problems (another name for SDP) started early 1980's, continues to be a very active area of research, [19],[28], and e.g.: [47, 46, 45, 44, 48, 40, 32, 20, 41]. (More recently,it is being used to solve large scale SDPs.)

• combinatorial optimization applications 1980's: Lovász *theta function* [52]; the strong approximation results for the max-cut problem by Goemans-Williamson, e.g. [27], survey papers: [25, 26],[78].

• linear complementarity problems can be extended to problems over the cone of semidefinite matrices, e.g. [22, 39, 43, 42, 59].

More: Historical Notes

• Complexity, Distance to Ill-Posedness, and Condition Numbers
SDP is a convex program and it can be solved to any desired accuracy in polynomial time, see seminal work of Nesterov and Nemirovski e.g. [63, 64, 68, 66, 62, 65, 67]. Another measure of complexity is the distance to ill-posedness: e.g. work by Renegar [80, 84, 83, 82, 81].

• Cone Programming this is a generalization of SDP, also called *generalized linear programming*, in paper by Bellman and Fan 1963, [8]. Other books deal with problems over cones date back to 60s, e.g. [70],[34, 54, 38, 37, 76]. More recently, generalization of SDP to more general cones, e.g. Güler and Tuncel, [30, 90] and also Hauser [31]

More: Historical Notes

- Other Related Areas e.g.: Eigenvalue Functions, e.g. [14], [72, 73]; Financial Applications; Generalized Convexity, e.g. [85, 56]; Statistics; Nonlinear Programming; ...

1.3 Motivation/Examples/Applications

- Quadratic constrained quadratic programs
- Lovász theta function
- Statistics
- minimizing the L_2-operator norm of a matrix
- linear programming
- robust mathematical programming
- engineering, e.g. control theory
- Combinatorial Problems, e.g. the Max-Cut Problem

1.3.1 Quadratic Constrained Quadratic Programs

What is SEMIDEFINITE PROGRAMMING?
Why use it?

- Quadratic approximations are better than linear approximations. (For example, model $x \in \{0, 1\}$ using $x^2 - x = 0$.)
And, we can solve relaxations of quadratic approximations efficiently using semidefinite programming.

HOW DOES SDP arise from quadratic approximations?

Let $\quad q_i(y) = \frac{1}{2} y^T Q_i y + y^T b_i + c_i, \; y \in \mathbb{R}^n, \; i = 0, 1, \ldots, m$

$$(\mathbf{QQP}) \quad \begin{cases} q^* = & \min & q_0(y) \\ & \text{s.t.} & q_i(y) = 0 \\ & & i = 1, \ldots m \end{cases}$$

• Lagrangian: $\quad L(y, x) =$ (with x as Lagrange multipliers)

quadratic in y — linear in y — constant in y

$$\frac{1}{2} y^T (Q_0 - \sum_{i=1}^m x_i Q_i) y \; + \; y^T (b_0 - \sum_{i=1}^m x_i b_i) \; + \; (c_0 - \sum_{i=1}^m x_i c_i)$$

• Primal-Dual pair: $\quad q^* = \min_y \max_x L(y, x) \geq d^* = \max_x \min_y L(y, x)$

Homogenization

- homogenize (add y_0): $\quad y_0 y^T (b_0 - \sum_{i=1}^m x_i b_i), \;\; y_0^2 = 1.$

$$
\begin{aligned}
d^* &= \max_x \min_y \quad L(y,x) \\
&= \max_x \min_{y_0^2=1} \quad \tfrac{1}{2} y^T (Q_0 - \textstyle\sum_{i=1}^m x_i Q_i) y \quad (+t y_0^2) \\
&\qquad + y_0 y^T (b_0 - \textstyle\sum_{i=1}^m x_i b_i) \\
&\qquad + (c_0 - \textstyle\sum_{i=1}^m x_i c_i) \quad (-t)
\end{aligned}
$$

Hidden Constraint

with t as Lagrange mulitplier for $y_0^2 = 1$ constraint,
use *hidden semidefinite constraint* to yield SDP constraint

$$(\mathcal{A} : \mathbb{R}^{m+1} \to \mathcal{S}_{n+1}) \qquad B - \mathcal{A}\begin{pmatrix} t \\ x \end{pmatrix} \succeq 0.$$

$$B = \begin{pmatrix} 0 & b_0^T \\ b_0 & Q_0 \end{pmatrix}, \quad \mathcal{A}\begin{pmatrix} t \\ x \end{pmatrix} = \begin{bmatrix} -t & \sum_{i=1}^m x_i b_i^T \\ \sum_{i=1}^m x_i b_i & \sum_{i=1}^m x_i Q_i \end{bmatrix}$$

Dual Program

The dual program is equivalent to the SDP (with $c_0 = 0$)

$$(\mathbf{D}) \qquad \begin{array}{rll} d^* = & \sup & -\sum_{i=1}^m x_i c_i - t \\ & \text{s.t.} & \mathcal{A}\begin{pmatrix} t \\ x \end{pmatrix} \preceq B \\ & & x \in \mathbb{R}^m, t \in \mathbb{R} \end{array}$$

As in linear programming, the dual is obtained from the optimal strategy of the competing player:

$$(\mathbf{DD}) \qquad \begin{array}{rll} d^* = & \inf & \operatorname{trace} BU \\ & \text{s.t.} & \mathcal{A}^* U = \begin{pmatrix} -1 \\ -c \end{pmatrix} \\ & & U \succeq 0. \end{array}$$

1.3.2 Generalized Eigenvalue Problems for $X = X^T$

1. **The Generalized Eigenvalue Problem**
   Let M, A be two $n \times n$ symmetric matrices, $M \succ 0$.
   The set of eigenvalues of the *matrix pencil*, denoted $[M, A]$, is $\{\lambda \in \mathbb{R} : \lambda M - A \text{ is singular}\}$.

   $$\lambda_{\max}([M, X]) \leq t \quad \text{is equivalent to} \quad tM - X \succeq 0.$$

2. **Spectral Norm of Symmetric X**

   $$\{|\lambda_i(X)| \leq t,\ \forall i\} \text{ iff } \{tI - X \succeq 0,\ tI + X \succeq 0\}$$

3. **Sum of k largest eigenvalues of Symmetric X**
   Let $S_k(X)$ denote the sum of the largest k eigenvalues of X.

   $$S_k(X) \leq t \ \underline{\text{iff}} \ \ t - ks - \operatorname{trace} Z \succeq 0, Z \succeq 0, Z - X + sI \succeq 0.$$

1.3.3 SDP Application in Statistics

If $m_0, m_1, m_2, \ldots, m_{2n}$ are moments of some distribution, then

$$H(m_0, m_1, \ldots, m_{2n}) = \begin{bmatrix} m_0 & m_1 & \cdots & m_n \\ m_1 & \cdots & \cdots & m_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ m_n & m_{n+1} & \cdots & m_{2n} \end{bmatrix} \succeq 0$$

Find distribution with maximal variance and $l_i \leq m_i \leq u_i$:

$$\begin{array}{ll} \max & y \\ \text{s.t.} & \begin{bmatrix} m_2 - y & m_1 \\ m_1 & 1 \end{bmatrix} \succeq 0 \\ & l_i \leq m_i \leq u_i \, (i = 1, \ldots 2n) \end{array}$$

Further Application Areas

1.3.4 Minimizing the L_2-operator Norm of a Matrix

1.3.5 Linear Programming

1.3.6 Robust Mathematical Programming

1.3.7 Control Theory

1.3.8 Max-Cut Problem

(MC) is a combinatorial optimization problem on undirected graphs with weights on the edges.

Problem 1.3.8.1 *Find a partition of the set of vertices into two parts that maximizes the sum of the weights on the edges that have one end in each part of the partition.*

Quadratic Model of MC

given graph G with vertex set $\{1, \ldots, n\}$
weighted adjacency matrix $A(G) = (a_{ij})$, weight on edge ij
Laplacian matrix $L := \mathrm{Diag}\,(A(G) \cdot e) - A(G)$
Let $v \in \{\pm 1\}^n$ represent any cut in the graph

$$
\text{(MC)} \qquad \begin{array}{rll} \mu^* = & \max & \frac{1}{4} v^T L v \\ & \text{s.t.} & v_i^2 = 1, \quad i = 1, \ldots, n, \end{array}
$$

where μ^* denotes the optimal value of MC.

Some applications of MC

- Statistical physics: Finding the ground state of a spin glass according to the Ising model
- VLSI: Minimizing the number of vias in a two-sided circuit board
- Network design: Solving the separation problem in a cutting plane approach

Approaches to MC

The general MC problem is NP-hard, even though it is tractable for some classes of graphs, e.g. planar graphs. In fact, solving within relative error of 1 percent is NP-hard.
Techniques used for general MC include:

- Heuristics;
- Integer programming enumerative techniques (Branch-and-Bound);
- Approximation algorithms.

We seek *SDP relaxations* of MC (solvable in polynomial-time) that yield tight (upper) bounds on the optimal value of MC.

One Formulation of MC

Let $v \in \{\pm 1\}^n$, $n = |V|$, represent a cut in the graph G via $S = \{i : v_i = +1\}$ and $V \backslash S = \{i : v_i = -1\}$.
Then we can formulate MC as:

$$\begin{array}{rll} \mu^* = & \max & \sum\limits_{1 \le i < j \le n} w_{ij} \left(\frac{1 - v_i v_j}{2} \right) \\ & \text{s.t.} & v \in \{\pm 1\}^n. \end{array}$$

Equivalently,

$$\text{(MC1)} \qquad \begin{array}{rll} \mu^* = & \max & v^T Q\, v \\ & \text{s.t.} & v_i^2 = 1, \qquad i = 1, \ldots, n, \end{array}$$

where $Q = \frac{1}{4}(\mathrm{Diag}\,(A\,e) - A)$ (Laplacian L), and $A = (w_{ij})$ is the weighted adjacency matrix of G.

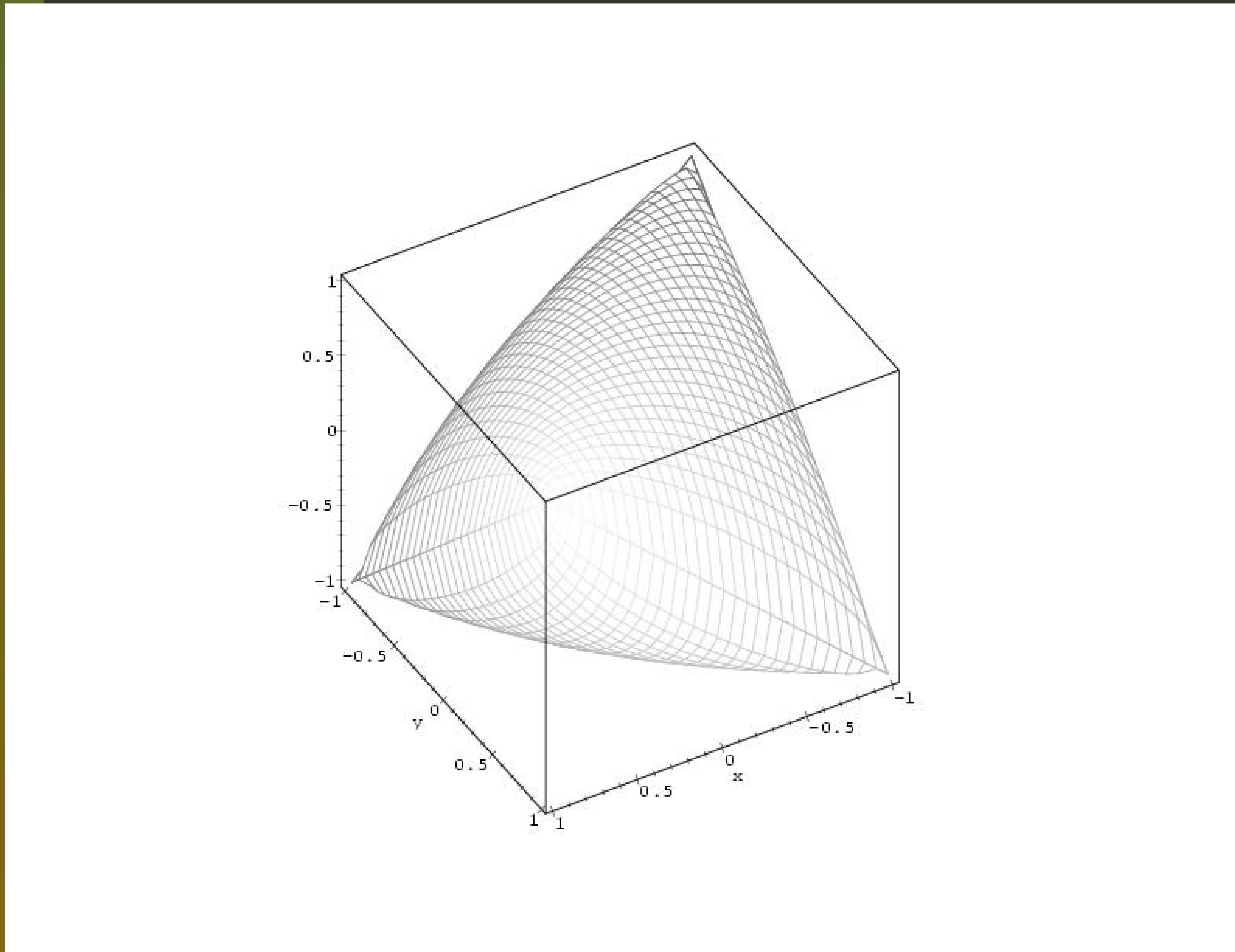
Equivalent Formulation of MC

With $Q := \frac{1}{4}L$, $X := vv^T, v \in \{\pm 1\}^n$,
then $v^T Qv = \operatorname{trace} QX$ and equivalent formulation is:

$$
\text{(MC1)} \qquad \begin{array}{rcl} \mu^* = & \max & \operatorname{trace} QX \\ & \text{s.t.} & \operatorname{diag}(X) = e \\ & & \operatorname{rank}(X) = 1 \\ & & X \succeq 0, X \in \mathcal{S}^n, \end{array}
$$

relax by deleting the *hard* constraint $\operatorname{rank}(X) = 1$ (get *elliptope*)

Elliptope for $n = 3$, [49]



GW Approximation Result

Goemans & Williamson proved that
if $w_{ij} \geq 0 \; \forall \, i, j$, then

$$\mu^* \geq \alpha \, \nu_1^*,$$

where $\alpha = \min_{0 \leq \theta \leq \pi} \frac{2}{\pi} \frac{\theta}{1 - \cos\theta} \approx 0.87856.$

Since $\frac{1}{\alpha} \approx 1.13823 \leq 1.14$, this implies

$$\mu^* \leq \nu_1^* \leq 1.14 \, \mu^*.$$

Other Approximations

Other such convex relaxations have been studied before.

The smallest convex set containing all the rank-one matrices X corresponding to cuts is the *cut polytope*:

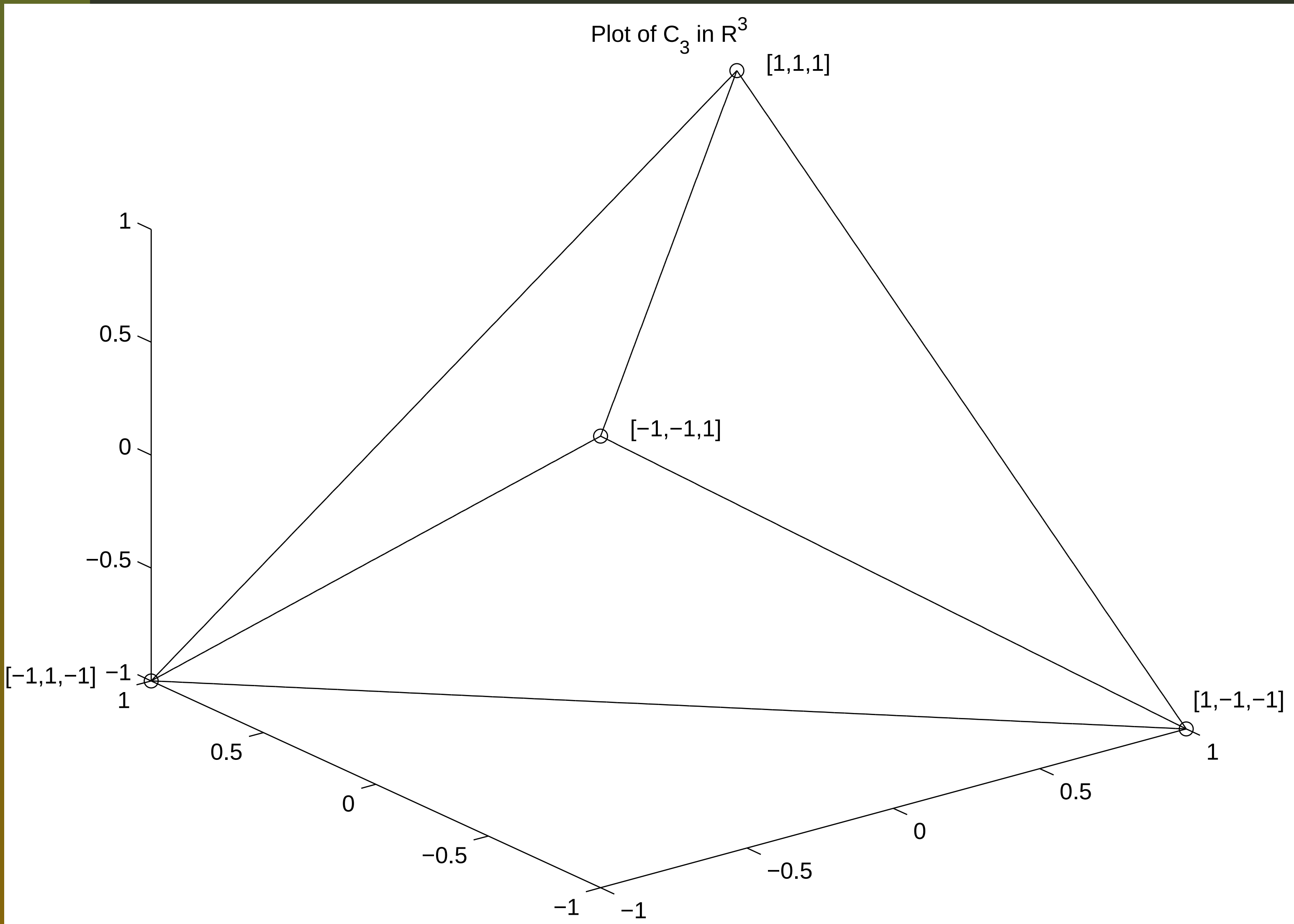
$$C_n := \mathrm{Conv}\{X : X = vv^T, v \in \{\pm 1\}^n\}.$$

In fact,

$$\begin{array}{rll} \mu^* = & \max & \operatorname{trace} QX \\ & \text{s.t.} & X \in C_n. \end{array}$$

However, it is not known how to optimize in polynomial-time over C_n.

Cut polytope for $n = 3$



Metric Polytope

Another convex relaxation is the *metric polytope* M_n, defined by

$$M_n := \{X \in \mathcal{S}^n : \operatorname{diag}(X) = e, \text{ and}$$
$$X_{ij} + X_{ik} + X_{jk} \geq -1, X_{ij} - X_{ik} - X_{jk} \geq -1,$$
$$- X_{ij} + X_{ik} - X_{jk} \geq -1, -X_{ij} - X_{ik} + X_{jk} \geq -1,$$
$$\forall\, 1 \leq i < j < k \leq n\}.$$

These inequalities model the fact that for any three mutually connected vertices in the graph, it is only possible to cut either zero or two of the edges joining them.

It is known that:

$C_n = M_n$ for $n \leq 4$, but $C_n \subsetneq M_n$ for $n \geq 5$.

2 Theory Of Cone Programming

2.1 Outline

- Convex cones and Partial Orders
- Convex Cone Programs
- Strong Duality
- Log Barrier and the Central Path

2.2 Convex Cones; Löwner Partial Order

Definition 2.2.0.1 *Let $\alpha \in \mathbb{R}$ and $S, T \subset \mathbb{R}^n$. Then*
$\alpha S = \{y : y = \alpha s,$ *for some* $s \in S\}$ *and*
$S + T = \{y : y = s + t,$ *for some* $s \in S, t \in T\}$

Definition 2.2.0.2 $\mathcal{K} \subset \mathbb{R}^n$ *is a* <u>*cone*</u> *if* $\alpha \mathcal{K} \subset K, \quad \forall \alpha > 0.$

Definition 2.2.0.3 *the cone* $\mathcal{K}$ *is a* <u>*convex cone*</u> *if* $\mathcal{K} + \mathcal{K} \subset \mathcal{K}$.

Definition 2.2.0.4 *A cone* $\mathcal{K}$ *is a* <u>*pointed cone*</u> *if*
$\mathcal{K} \cap (-\mathcal{K}) = \{0\}$.

Definition 2.2.0.5 *A cone* $\mathcal{K} \subset \mathbb{R}^n$ *is a* <u>*proper cone*</u> *if it is closed, pointed, and convex and has nonempty interior.*

Examples of Cones

Example 2.2.0.1 *open half line:* $\{x \in \mathbb{R} : x > 0\}$.

Example 2.2.0.2 *closed half line:* $\{x \in \mathbb{R} : x \geq 0\}$.

Example 2.2.0.3 *psd matrices, $\mathcal{P}$:* $\{X \in \mathcal{S}^n : X \succeq 0\}$.

Example 2.2.0.4 *Lorentz cone*
(ice-cream cone, second-order cone):
$L^m = \{x = (x_1, \ldots, x_m) \in \mathbb{R}^m : x_m \geq \sqrt{x_1^2 + \ldots + x_{m-1}^2}$.

Polar Cone

And, an important closed convex cone is:

Definition 2.2.0.6 *Polar (dual, conjugate) of set $\mathcal{S}$:*
$\mathcal{S}^+ := \{z : \langle x, z \rangle \geq 0, \forall x \in \mathcal{S}\}$.

Example 2.2.0.5 *nonnegative orthant, psd cone $\mathcal{P}$, and Lorentz cone L^m, are all self-polar, i.e. $\mathcal{K} = \mathcal{K}^+$.*

Further Examples

Definition 2.2.0.7 *Direct sum of two cones*

$\mathcal{K} \oplus \mathcal{L} := \{(k\ l) : k \in \mathcal{K}, l \in \mathcal{L}\}$.

Example 2.2.0.6 *Let L denote the half line in $\mathbb{R}$, then $\underbrace{L \oplus L \cdots \oplus L}_{n}$ is the n-dimension nonnegative orthant.*

Properties of Cones

Suppose $\mathcal{K}$, $\mathcal{K}_1$, $\mathcal{K}_2$ are proper cones, then:

- $\mathcal{K}^+$ is proper.
- $(\mathcal{K}^+)^+ = \mathcal{K}$.
- $(\mathcal{K}_1 \cap \mathcal{K}_2)^+ = \overline{\mathcal{K}_1^+ + \mathcal{K}_2^+}$.
- $(\mathcal{K}_1 + \mathcal{K}_2)^+ = \mathcal{K}_1^+ \cap \mathcal{K}_2^+$.
- $\mathcal{K}_1 \oplus \mathcal{K}_2$ is proper.
- $(\mathcal{K}_1 \oplus \mathcal{K}_2)^+ = \mathcal{K}_1^+ \oplus \mathcal{K}_2^+$.

Useful Lemma, e.g. to Prove Farkas' Lemma

Lemma 2.2.0.1

$$\mathcal{K} \textit{ is a closed convex cone} \quad \Longleftrightarrow \quad \mathcal{K} = \mathcal{K}^{++}.$$

Partial Orders

Definition 2.2.0.8 $x \succeq_{\mathcal{K}} y$ *(respectively* $x \succ_{\mathcal{K}} y$*) if* $x - y \in \mathcal{K}$ *(respectively* $x - y \in \operatorname{int} \mathcal{K}$*).*

Remark 2.2.0.1 *If* $\mathcal{K}$ *is a pointed, convex cone, then "*$\succeq_{\mathcal{K}}$*" is a* *(linear) partial order (reflexive, transitive, and antisymmetric):*

- $0 \in \mathcal{K} \Rightarrow x \succeq_{\mathcal{K}} x$ *(reflexive);*
- $\mathcal{K}$ *is convex* $\Rightarrow$ *if* $x \succeq_{\mathcal{K}} y$ *and* $y \succeq_{\mathcal{K}} z,$ *then* $x \succeq_{\mathcal{K}} z$ *(transitive);*
- $\mathcal{K}$ *is pointed* $\Rightarrow$ *if* $x \succeq_{\mathcal{K}} y$ *and* $y \succeq_{\mathcal{K}} x,$ *then* $x = y$ *(antisymmetric);*
- $\mathcal{K}$ *is convex cone* $\Rightarrow$
  if $a, b \geq 0, u \succeq_{\mathcal{K}} x$ *and* $v \succeq_{\mathcal{K}} y,$ *then* $au + bv \succeq_{\mathcal{K}} ax + by$
  (linear - homogeneous, additive);

2.3 Convex Cone Program

Let: $\mathcal{K}, \mathcal{L}$ be convex cones;
f real valued convex function, g is $\mathcal{K}$-convex, i.e.
$g(\alpha u + (1-\alpha)v) \preceq_{\mathcal{K}} \alpha g(u) + (1-\alpha)g(v), \forall 0 \leq \alpha \leq 1, \forall u, v$

$$
(CP) \qquad \begin{array}{rr} \mu^* = \min & f(x) \\ \text{s.t.} & g(x) \preceq_{\mathcal{K}} 0 \\ & x \succeq_{\mathcal{L}} 0 \end{array}
$$

with Lagrangian dual (weak duality)

$$
(DCP) \qquad \mu^* \geq \nu^* = \max_{y \in \mathcal{K}^+} \min_{x \in \mathcal{L}} \quad f(x) + \langle y, g(x) \rangle
$$

Linear Cone Optimization

Example 2.3.0.7 *If* $f(x) = \langle c, x\rangle, g(x) = b - \mathcal{A}x$*, then we have a linear cone programming problem, (LCP). The dual*

$$
(DLCP) \quad \begin{aligned}
\mu^* \geq \nu^* &= \max_{y \in \mathcal{K}^+} \min_{x \in \mathcal{L}} && \langle c, x\rangle + \langle y, b - \mathcal{A}x\rangle \\
&= \max_{y \in \mathcal{K}^+} \min_{x \in \mathcal{L}} && \langle c - \mathcal{A}^* y, x\rangle + \langle y, b\rangle \\
&= \max_{\substack{y \in \mathcal{K}^+ \\ c - \mathcal{A}^* y \in \mathcal{L}^+}} \min_{x \in \mathcal{L}} && \langle c - \mathcal{A}^* y, x\rangle + \langle y, b\rangle
\end{aligned}
$$

reduces to the elegant (LP or SDP type) form

$$
(DLCP) \quad \begin{array}{ccc}
\mu^* \geq \nu^* = \max & & \langle y, b\rangle \\
\text{s.t.} & & \mathcal{A}^* y \preceq_{\mathcal{L}^+} c \\
& & y \succeq_{\mathcal{K}^+} 0
\end{array}
$$

Convex Program

Example 2.3.0.8 *Any convex program:*

$$\begin{array}{rl} min & f(x) \\ \text{s.t.} & x \in C \end{array}$$

where $f(x)$ is convex function, C is convex set, can be transformed into Cone-LP:

1.
$$\min\{z : z \geq f(x), \quad x \in C\}$$
2. *constraints in (1) equivalent to $(z; x)$ in some convex set B.*
3. **Homogenize**; *add x_0, put B in the plane $x_0 = 1$, the convex cone is constructed by rays starting from 0 and having a point in B.*

2.4 Strong Duality

2.4.1 Outline

- Faces and minimal cones
- Optimality Conditions without constraint qualifcations (on web page only)

2.4.2 Facial Structure

Definition 2.4.2.1 *$\mathcal{F}$ is a* face *of a convex cone $\mathcal{K}$ (denoted $\mathcal{F} \lhd \mathcal{K}$) if $\mathcal{F} \subset \mathcal{K}$ and*

$$\frac{1}{2}(x_1 + x_2) \in \mathcal{F} \Rightarrow x_1, x_2 \in \mathcal{F}, \qquad \forall x_1, x_2 \in \mathcal{K}$$

Definition 2.4.2.2 *A face $\mathcal{F} \lhd \mathcal{K}$ is called* exposed *if*

$$\mathcal{F} = \mathcal{K} \cap \phi^{\perp}, \qquad \textit{for some } \phi \in \mathcal{K}^{+}$$

Definition 2.4.2.3 *A face $\mathcal{F} \lhd \mathcal{P}$ is called* projectionally exposed *if*

$$\mathcal{F} = Q\mathcal{P}Q^{t} \qquad \textit{for some matrix } Q$$

2.4.3 Weak Duality

The weak duality: $c^T x \geq b^T y$ holds in Cone-LP:

$$c^T x - b^T y = c^T x - (Ax)^T y = x^T (c - A^T y) = x^T s \geq 0$$

2.4.4 Extended Farkas' Lemma

Theorem 2.4.4.1 ***(Seperation Theorem)*** *If $K \subseteq \mathbb{R}^n$ is closed and convex, $b \in \mathbb{R}^n$, $b \notin K$, then $\exists a \in \mathbb{R}^n$, such that $\forall x \in K \,:\, a^T b < 0$ and $a^T x \geq 0$.*

Lemma 2.4.4.1 ***(Farkas' Lemma)*** *Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a closed, convex cone, $A \in \mathbb{R}^{m \times n}$, $A(\mathcal{K})$ closed, and $b \in \mathbb{R}^m$. Then:*

$$\{\exists x \succeq_{\mathcal{K}} 0 : Ax = b\} \iff \left\{A^T y \succeq_{\mathcal{K}^+} 0 \Rightarrow b^T y \geq 0\right\}.$$

Extended Farkas' Lemma

Proof.

Necessity:

Suppose $\exists x \in \mathcal{K} : Ax = b$. And suppose $A^T y \succeq_{\mathcal{K}^+} 0$. Then $b^T y = (Ax)^T y = x^T (A^T y) \geq 0$, since $x \succeq_K 0$.

Sufficiency:

Suppose that the RHS of (2.4.4.1) holds but the LHS fails, i.e. $b \notin A(\mathcal{K})$. Then separate, i.e. $\exists y \in \mathbb{R}_m$ such that $b^T y < 0$, and $\forall x \in \mathcal{K}$, $(Ax)^T y \geq 0$. Therefore, $A^T y \in K^+$ which implies $b^T y \geq 0$, contradiction. $\blacksquare$

A Related Papers

A.1 Outline

Following are summaries of several papers closely related to the above notes.

1. Alizadeh has notes that get updated each time he gives the course [1].
2. Nemirovski has course notes that he gave at Delft [9].
3. Todd, December, 2000 has a survey paper on SDP [89].

References

[1] F. ALIZADEH. Lecture notes on semidefinite programming. Technical report, Rutgers University, Piscataway, NJ, 2000. URL: http://karush.rutgers.edu/~alizadeh/.

[2] F. ALIZADEH, J-P.A. HAEBERLY, and M.L. OVERTON. A new primal-dual interior-point method for semidefinite programming. In J.G. Lewis, editor, *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra*, pages 113–117. SIAM, 1994.

[3] M.F. ANJOS and H. WOLKOWICZ. A strengthened SDP relaxation via a second lifting for the Max-Cut problem. *Discrete Appl. Math.*, 119:79–106, 2002.

[4] K.M. ANSTREICHER and N.W. BRIXIUS. Solving