Exploiting Algebraic Symmetry in Semidefinite Programs: Theory and Applications

Etienne de Klerk

Tilburg University, The Netherlands

SIAM Conference on Optimization, Boston, May 10th, 2008

What is the difference between method and device? A method is a device which you used twice.

George Pólya (1887 — 1985)

Part I: Theory

- Semidefinite programs (SDP's) with special structure;
- Matrix algebras and their decompositions;
- Matrix algebras arising from groups.

Standard form SDP

Primal problem

$$\min_{X \succeq 0} \text{trace}(A_0 X) \;\; \text{subject to} \;\; \text{trace}\,(A_k X) = b_k \;\; (k = 1, \ldots, m),$$

where the data matrices $A_i \in \mathbb{S}^{n \times n}$.

- $\mathbb{S}^{n \times n}$: symmetric $n \times n$ matrices;
- $X \succeq 0$: X symmetric positive semi-definite.

Sometimes we will add the additional constraint $X \geq 0$ (componentwise nonnegative).

SDP at a glance

- SDP may be solved in polynomial time with fixed precision in the real number model. (Ellipsoid algorithm of Nemirovski-Yudin);
- Interior point methods (IPM's) for LP were extended to SDP in the early 1990's (Nesterov-Nemirovski, Alizadeh, ...);
- Sparsity in LP data may be exploited effectively by IPM's, ...
- ... but this is not true for SDP.

Structured SDP instances

Three types of structure in the SDP data matrices $A_0, \ldots, A_m$ may be effectively exploited by IPM's:

- low rank of the A_i's; (Benson-Ye-Zhang; DSDP software)
- aggregate *chordal* sparsity pattern of the A_i's, including block diagonal structure; (Wolkowicz et al, Laurent; SDPA software)
- if the A_i's lie in a low dimensional matrix algebra (today's talk).

Matrix algebras

Definition

A set $\mathcal{A} \subseteq \mathbb{C}^{n\times n}$ is called a *matrix *-algebra* over $\mathbb{C}$ if, for all $X, Y \in \mathcal{A}$:

- $\alpha X + \beta Y \in \mathcal{A} \quad \forall \alpha, \beta \in \mathbb{C}$;
- $X^* \in \mathcal{A}$;
- $XY \in \mathcal{A}$.

Assumption

There is a 'low dimensional' matrix *-algebra $\mathcal{A}_{SDP} \supseteq \{A_0, \ldots, A_m\}$.

Link with SDP

Theorem

If the primal SDP problem and its dual problem meet the Slater condition, then there exists an optimal $X \in \mathcal{A}_{SDP}$.

Proof sketch:

- One may show that the central path of the primal SDP problem is contained in $\mathcal{A}_{SDP}$;
- This implies that there is an optimal primal point in $\mathcal{A}_{SDP}$.

cf.

Y. Kanno, M. Ohsaki, K. Murota and N. Katoh, Group symmetry in interior-point methods for semidefinite programming, *Optimization and Engineering*, 2(3): 293–320, 2001.

We may therefore restrict the primal problem to:

$$\min_{X \succeq 0} \{\text{trace}(A_0 X) \ : \ \text{trace}(A_k X) = b_k \quad (k = 1, \ldots, m), \ X \in \mathcal{A}_{SDP}\}.$$

Canonical decomposition of a matrix *-algebra $\mathcal{A}$

Theorem (Wedderburn (1907))

*Assume $\mathcal{A}$ is a matrix *-algebra over $\mathbb{C}$ that contains I. Then there is a unitary Q ($Q^*Q = I$) and some integer s such that*

$$Q^*\mathcal{A}Q = \begin{pmatrix} \mathcal{A}_1 & 0 & \cdots & 0 \\ 0 & \mathcal{A}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathcal{A}_s \end{pmatrix},$$

where each $\mathcal{A}_i \sim \mathbb{C}^{n_i \times n_i}$ for some integers n_i, and takes the form

$$\mathcal{A}_i = \left\{ \begin{pmatrix} A & 0 & \cdots & 0 \\ 0 & A & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & A \end{pmatrix} \;\middle|\; A \in \mathbb{C}^{n_i \times n_i} \right\} \quad (i = 1, \ldots, s).$$

Joseph Wedderburn (1882 – 1942)

"He was apparently a very shy man and much preferred looking at the blackboard to looking at the students. He had the galley proofs from his book 'Lectures on Matrices' pasted to cardboard for durability, and his 'lecturing' consisted of reading this out loud while simultaneously copying it onto the blackboard."

SDP reformulation

Assume we have a basis $B_1, \ldots, B_d$ of $\mathcal{A}_{SDP}$. Set $X = \sum_{i=1}^{d} x_i B_i$ to get

$$\min_{X \succeq 0} \{\text{trace}(A_0 X) \ : \ \text{trace}(A_k X) = b_k \quad (k = 1, \ldots, m), X \in \mathcal{A}_{SDP}\}$$

$$= \min_{x \in \mathbb{R}^d} \left\{ \sum_{i=1}^{d} x_i \text{trace}(A_0 B_i) \ : \ \sum_{i=1}^{d} x_i \text{trace}(A_k B_i) = b_k, \right.$$

$$\left. (k = 1, \ldots, m), \ \sum_{i=1}^{d} x_i B_i \succeq 0 \right\}.$$

- Replace the LMI by $\sum_{i=1}^{d} x_i Q^* B_i Q \succeq 0$ to get block-diagonal structure.
- Delete any identical copies of blocks in the block structure.

Remarks

- If a basis of $\mathcal{A}$ is known, the unitary matrix Q may be computed using only linear algebra.

Randomized algorithms:

W. Eberly and M. Giesbrecht, Efficient decomposition of separable algebras. *Journal of Symbolic Computation*, 37(1): 35–81, 2004.

K. Murota, Y. Kanno, M. Kojima and S. Kojima, A Numerical Algorithm for Block-Diagonal Decomposition of Matrix *-Algebras, Preprint 2007.

- Further reading:

Chapter X in: J.H.M. Wedderburn. *Lectures on Matrices*. AMS publishers, 1934.

K. Gatermann, P. A. Parrilo, Symmetry groups, semidefinite programs, and sums of squares. *J. Pure and Applied Algebra*, 192, 95–128, 2004.

F. Vallentin. Symmetry in semidefinite programs. arXiv:0706.4233

Matrix *-algebras from groups

How does the matrix *-algebra $\mathcal{A}_{SDP}$ arise in practice?

Definition

We define the *automorphism group* of a matrix $Z \in \mathbb{S}^{n\times n}$ as

$$\text{aut}(Z) = \{P \in \Pi_n \ : \ PZ = ZP\}$$

where Π_n is the set of $n \times n$ permutation matrices.

SDP symmetry assumption:

The multiplicative matrix group $\mathcal{G}_{SDP} := \bigcap_{i=0}^{m} \text{aut}(A_i)$ is non-trivial.

Note that $PA_i = A_iP$ for all $P \in \mathcal{G}_{SDP}$, $i = 0, \ldots, m$.

Thus we make take $\mathcal{A}_{SDP}$ as the commutant of $\mathcal{G}_{SDP}$:

$$\mathcal{A}_{SDP} = \left\{A \in \mathbb{C}^{n\times n} \ : \ PA = AP \text{ for all } P \in \mathcal{G}_{SDP}\right\}.$$

Matrix *-algebras from groups: properties

The commutant of a group of permutation matrices has a basis $B_1, \ldots, B_d$ with the following properties:

- The B_i's are 0-1 matrices;
- $\sum_{i=1}^{d} B_i = J$ (the all-ones matrix).

The B_i's correspond to the orbitals of the group.

Consequence

If $X = \sum_{i=1}^{d} x_i B_i$, then

$$X \succeq 0 \text{ and } X \geq 0 \iff \sum_{i=1}^{d} x_i B_i \succeq 0 \text{ and } x \geq 0.$$

Part II: Applications

- The Lovász ϑ-number for graphs;
- Error correcting binary codes;
- Kissing numbers;
- Crossing numbers of completely bipartite graphs;
- Quadratic assignment problems;

Lovász ϑ-function

A graph $G = (V, E)$ is given.

Lovász ϑ-function

$$\vartheta(G) := \max \operatorname{trace}(JX)$$

subject to

$$\begin{aligned} X_{ij} &= 0,\ \{i,j\} \in E\ (i \neq j) \\ \operatorname{trace}(X) &= 1 \\ X &\succeq 0, \end{aligned}$$

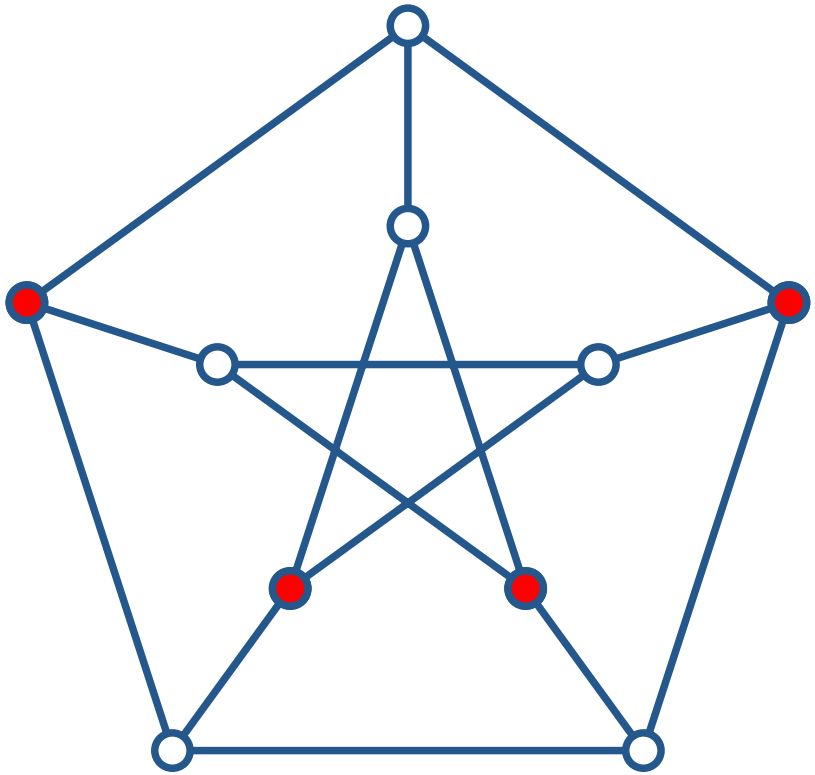
where e denotes the all-one vector.

Schrijver ϑ'-function

Add the additional constraint $X \geq 0$ to the ϑ problem.

Co-cliques

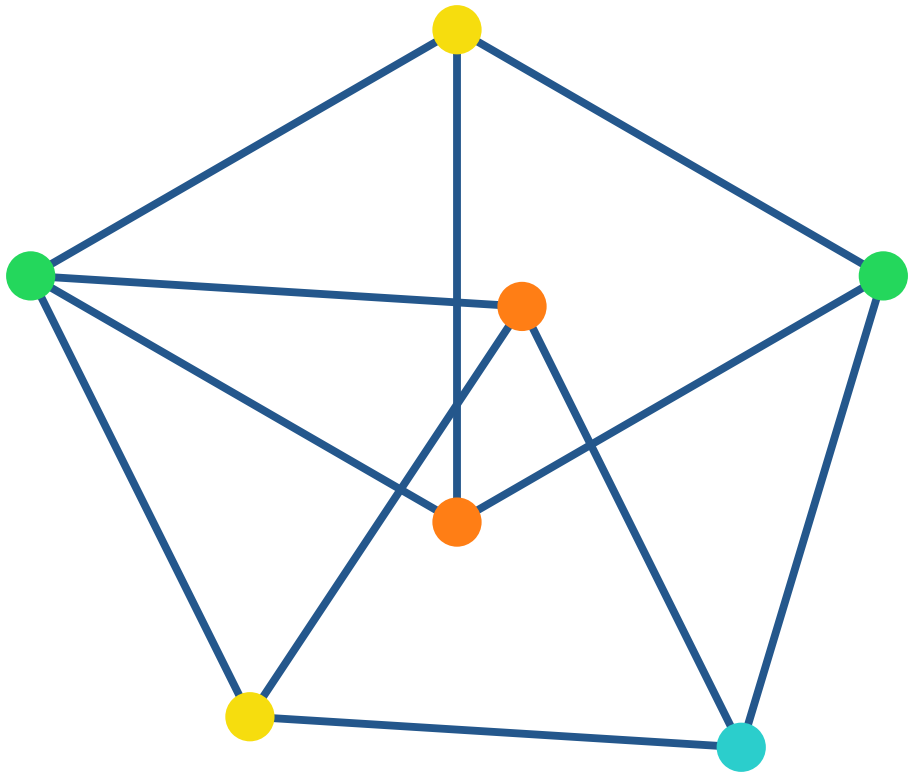
A *stable set* of $G = (V, E)$ is a subset $V' \subset V$ such that the *induced subgraph* on V' has no edges.



The stability number $\alpha(G)$ is the cardinality of the largest co-clique of G.

Vertex colourings

A (proper) vertex colouring is an assignment of colours to the vertices V of G such that endpoints of each edge are assigned different colours.



The smallest number of colours needed is called the chromatic number $\chi(G)$.

Lovász sandwich theorem

Theorem

$$\alpha(G) \leq \vartheta'(G) \leq \vartheta(G) \leq \chi(\bar{G}),$$

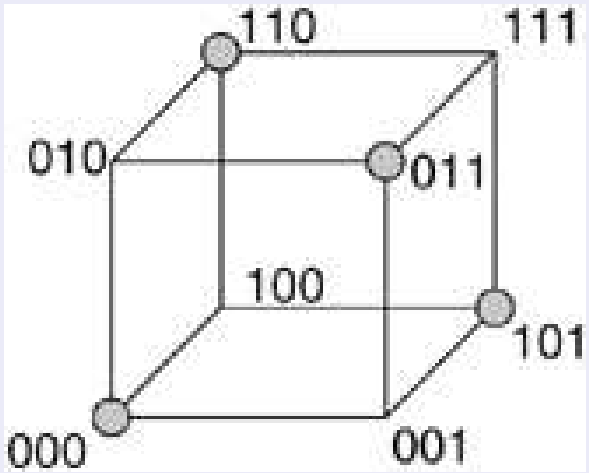
where $\bar{G}$ is the complementary graph of G.

- One may approximate $\alpha(G)$ or $\chi(G)$ by $\vartheta(G)$ (or $\vartheta'(G)$).
- If A is the adjacency matrix of G, then aut(A) is called the automorphism group of G.
- One may reduce the sizes of the SDP problems for graphs with large automorphism groups, like the Hamming graphs ...

The Hamming graph and binary codes

The Hamming graph $G(k, \delta)$ has vertices indexed by $\{0,1\}^k$ and vertices adjacent if they are at Hamming distance less than δ.

Hamming graph with $k = 3$ and $\delta = 2$.



Usual notation: $\alpha(G(k, \delta)) =: A(k, \delta)$. Thus $A(3, 2) = 4$ (see picture).

$A(k, \delta)$ is the maximum size of a binary code on k letters such that any two words are at a Hamming distance of at least δ.

ϑ'-of the Hamming graph

Equivalent formulation for ϑ':

$$\vartheta'(G) := \max_{X \succeq 0, X \geq 0} \{\text{trace}(JX) \mid \text{trace}(A+I)X = 1\},$$

where A is the adjacency matrix of the graph G. Thus $\mathcal{G}_{SDP} = \text{aut}(A)$.

- For the Hamming graph $|\text{aut}(A)| = 2^k k!$, and ....
- ... the commutant of $\text{aut}(A)$ is the commutative Bose-Mesner algebra of the Hamming scheme ...
- ... that has dimension $k+1$.
- Thus the SDP matrices may be reduced from the original size $n = 2^k$ to diagonal matrices of size $k+1$.

The resulting LP coincides with the LP bound of Delsarte.

A. Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Trans. Inform. Theory*, 25:425–429, 1979.

Improvements

A. Schrijver. New code upper bounds from the Terwilliger algebra. *IEEE Transactions on Information Theory*, 51:2859–2866, 2005.

In this paper, a stronger SDP bound for $A(k, \delta)$ is obtained as follows:

- a stronger SDP relaxation is constructed via 'lift-and-project' ...
- ... such that some symmetry is retained in the resulting SDP.
- $\mathcal{A}_{SDP}$ becomes the Terwilliger algebra of the Hamming scheme, a non-commutative algebra that contains the Bose-Mesner algebra of the Hamming scheme.
- The Terwilliger algebra has dimension $\binom{k+3}{3}$ and its canonical block-diagonalization is known.

Thus, improved upper bounds were computed for $A(19, 6)$, $A(23, 6)$, $A(25, 6)$, ...

Further improvements

Using other lift-and-project schemes, slightly better SDP bounds may be obtained.

M. Laurent. Strengthened semidefinite bounds for codes. *Mathematical Programming*, 109(2-3):239–261, 2007,

... and the approach may be extended to non-binary codes:

D. Gijswijt, A. Schrijver, H. Tanaka, New upper bounds for nonbinary codes based on the Terwilliger algebra and semidefinite programming, *Journal of Combinatorial Theory, Series A*, 113, 1719–1731, 2006.

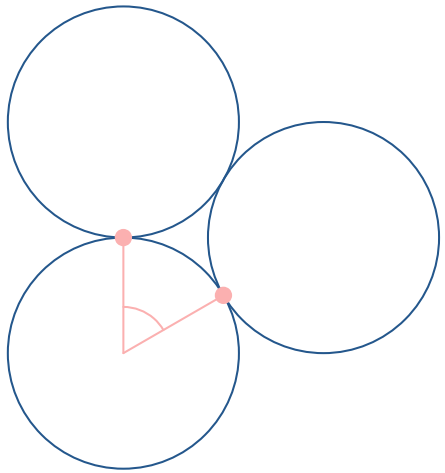
Kissing numbers

How many unit balls can touch a center ball in $\mathbb{R}^n$?

In $\mathbb{R}^3$ the answer is 12 (famous disagreement between Newton and Gregory).

Kissing numbers (ctd.)

- The kissing number problem may be formulated as a maximum stable set problem in an infinite graph ...
- with vertices the points on the unit ball, and two vertices adjacent if the angle between them is below a certain value.



Kissing numbers (ctd.)

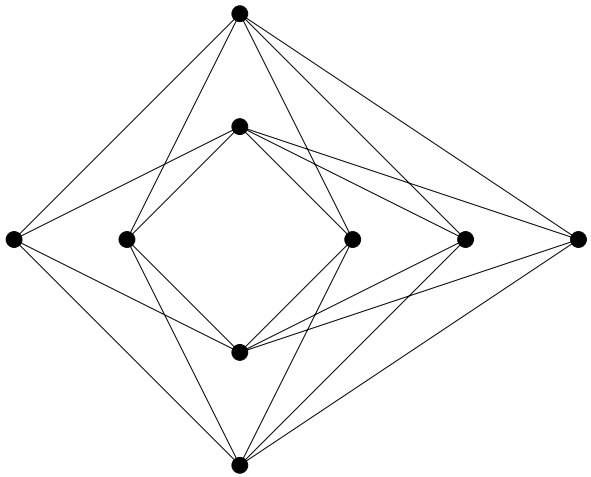
- Rough idea: generalize certain SDP relaxations of the maximum stable set problem to infinite graphs;
- Exploit the symmetry of the sphere (orthogonal group) to obtain a (finite) SDP relaxation.
- In this way, improved upper bounds on the kissing number were obtained in dimensions: $5, 6, 7, 9$ and 10.

C. Bachoc and F. Vallentin. New upper bounds for kissing numbers from semidefinite programming. *Journal of the AMS*, 21, 909-924, 2008.

Crossing numbers

The complete bipartite graph $K_{r,s}$ can be drawn in the plane with at most $Z(r,s)$ edges crossing, where

$$Z(r,s) = \left\lfloor \frac{r-1}{2} \right\rfloor \left\lfloor \frac{r}{2} \right\rfloor \left\lfloor \frac{s-1}{2} \right\rfloor \left\lfloor \frac{s}{2} \right\rfloor.$$



A drawing of $K_{4,5}$ with $Z(4,5) = 8$ crossings.

The smallest possible number of crossings is called the crossing number: $\mathrm{cr}(K_{r,s})$.

Crossing numbers (ctd.)

Conjecture (Zarankiewicz)

$$\mathrm{cr}(K_{r,s}) = Z(r,s).$$

(Open problem since Turán posed it in the 1940's).

- An SDP lower bound on $\mathrm{cr}(K_{r,s})$ was studied in:

E. de Klerk, J. Maharry, D.V. Pasechnik, B. Richter and G. Salazar. Improved bounds for the crossing numbers of $K_{m,n}$ and K_n. *SIAM J. Discr. Math.* 20:189–202, 2006,

E. de Klerk, D.V. Pasechnik and A. Schrijver. Reduction of symmetric semidefinite programs using the regular *-representation. *Mathematical Programming B*, 109(2-3):613-624, 2007,

- ... where the SDP matrix size is $n = (r-1)!$, ...
- ..., and $|\mathcal{G}_{SDP}| = 2r!$.

Crossing numbers (ctd.)

Using symmetry reduction of the SDP, it was shown that

$$0.859Z(r,s) \leq \mathrm{cr}(K_{r,s}) \leq Z(r,s)$$

if r or s is sufficiently large.

Quadratic assignment problem (QAP)

Definition (Trace formulation (Edwards 1977))

Given are symmetric $k \times k$ matrices A (distance matrix) and B (flow matrix).

$$\min_{X \in \Pi_k} \text{trace}(AXBX^T)$$

where Π_k is the set of $k \times k$ permutation matrices.

- QAP is NP-hard in the strong sense;
- Many applications, but very hard to solve in practice for $k \geq 30$.

SDP relaxation of QAP

An SDP relaxation of QAP was introduced in:

Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite Programming Relaxations for the Quadratic Assignment Problem. *Journal of Combinatorial Optimization*, **2**, 71–109, 1998.

- Matrix size $n = (k+1)^2$ — size reduction of the SDP essential if $k \geq 15$ ...
- The automorphism groups of A and B completely determine the SDP symmetry group $\mathcal{G}_{SDP}$;
- Thus symmetry reduction possible if aut(A) and/or aut(B) is large.

SDP relaxation of QAP: numerical results

Several instances in the QAPlib library have algebraic symmetry, e.g. the distance matrix is a Hamming distance matrix.

Some numerical results, after doing the SDP symmetry reduction:

instance	k	previous l.b.	SDP l.b.	best known u.b.	time(s)
esc64a	64	47	98	116	13
esc128a	128	2	54	64	140

E. de Klerk and R. Sotirov. Exploiting Group Symmetry in Semidefinite Programming Relaxations of the Quadratic Assignment Problem. Preprint, 2007.

Details in the talk by Renata Sotirov, session MS55, Monday 15:00.

TSP may be formulated as QAP with a circulant distance matrix — new SDP relaxation of TSP via symmetry reduction.

E. de Klerk, D.V. Pasechnik and R. Sotirov. On semidefinite programming relaxations of the traveling salesman problem. Preprint, 2007.

And, finally ...

- Symmetry reduction in SDP is the application of representation theory to reduce the size of specially structured SDP instances.
- The most notable applications are in computer assisted proofs (bounds on crossing numbers, kissing numbers, codes, ...)
- ... but also pre-processing of some SDP's arising in optimal design (truss design, QAP, ...)
- More applications in polynomial optimization, graph coloring, ...
- The device from 1979 has become a method!

The End

THANK YOU!