Possibilities and limitations of semidefinite approaches to polynomial optimization

Gennadiy Averkov

Brandenburg University of Technology

Seminar on Applied Combinatorics, Algebra, Topology & Statistics
KTH Royal Institute of Technology
Stockholm

Global optimization

Convex/linear optimization is the core of optimization theory:

- local methods give globally optimal solutions.
- local information is enough to certify optimality.

Finding a global optimum of a general non-linear problem is computationally challenging. Possible strategies:

- give it up and just find a locally optimal solution (with gradient decent, Newton etc.). That is a classical strategy.
- make case distinction and convexify (turn a non-convex problem into a convex one).

Finding a global optimum of a general non-linear problem is computationally challenging. Possible strategies:

- give it up and just find a locally optimal solution (with gradient decent, Newton etc.). That is a classical strategy.
- make case distinction and convexify (turn a non-convex problem into a convex one).

Good news:

- Real algebra can be used to develop general-purpose convexification strategies for polynomial optimization (as was pointed out by Lasserre).
- But we still need to think about different kinds of complexity issues

Algebra meets optimization: cultural differences

Algebraists	Optimizers
usually don't think about complexity	think a lot about complexity

Algebra meets optimization: cultural differences

Algebraists	Optimizers
usually don't think about complexity	think a lot about complexity
like deep and general math	ready to use any math

Algebra meets optimization: cultural differences

Algebraists	Optimizers
usually don't think about complexity	think a lot about complexity
like deep and general math	ready to use any math
don't tweak on algorithms	tweak on algorithms

Algebra meets optimization: cultural differences

Algebraists	Optimizers
usually don't think about complexity	think a lot about complexity
like deep and general math	ready to use any math
don't tweak on algorithms	tweak on algorithms
don't do numerical evaluation	do numerical evaluation

Algebra meets optimization: cultural differences

Algebraists	Optimizers
usually don't think about complexity	think a lot about complexity
like deep and general math	ready to use any math
don't tweak on algorithms	tweak on algorithms
don't do numerical evaluation	do numerical evaluation
speak their secret language	speak their secrete language

Our topic:

- using semidefinite programming
- in polynomial optimization
- and related challenges and limitations.

Conic and semidefinite programming

Conic programming

- *Conic programming*:

$$\inf\{c^\top x \,:\, Ax = b,\ x \in \underbrace{K}_{\text{closed convex cone}}\}.$$

- $K = \mathbb{R}^n_+ \rightsquigarrow$ *linear programming*
- $K = \mathcal{S}^k_+ := \{k \times k \text{ symmetric psd matrices over } \mathbb{R}\} \rightsquigarrow$ *semidefinite programming (SDP)*:

$$\inf\{c^\top \mathrm{vec}(X) \,:\, A\,\mathrm{vec}(X) = b,\ X \in \mathcal{S}^k_+\}$$

Semidefinite programming and LMIs

- Consider a $k \times k$ symmetric matrix
$$A(x) := \Big(a_{ij}(x)\Big)_{i,j=1,\ldots,k}$$
with entries $a_{ij}(x)$ being affine functions in $x \in \mathbb{R}^n$.
- The condition
$$A(x) \in \mathcal{S}_+^k$$
is called a *linear matrix inequality (LMI)* of size k on n real-valued variables $x \in \mathbb{R}^n$.
- The set
$$\{x \in \mathbb{R}^n \ : \ A(x) \in \mathcal{S}_+^k\}$$
is called a *spectrahedron*.
- *Semidefinite programming* is optimization of a linear function subject to finitely many LMIs.

Semidefinite programming, computational aspects

- SDP is efficiently solvable using interior-point methods under mild assumptions!
- **But:** If you can avoid LMIs of large size, you should really do that!
  - running time
  - numerical stability

Connections of SDP to other computational problems

1. Linear programming: linear constraints are LMIs of size 1
2. Solving a system of linear equations over $\mathbb{R}$ $\rightsquigarrow$ minimization of a convex quadratic function $\rightsquigarrow$ SDP:

$$\begin{aligned} Ax = b \qquad & \rightsquigarrow \qquad \min\left\{\|Ax - b\|_2^2 \,:\, x \in \mathbb{R}^n\right\} \\ & \rightsquigarrow \qquad \min\Big\{t \,:\, \underbrace{\|Ax - b\|_2 \le t}_{\substack{\text{can be modelled} \\ \text{as LMI}}}\Big\} \end{aligned}$$

3. Maximum eigenvalue of a symmetric matrix $\longrightarrow$ SDP with an LMI on one variable:

$$\min\Big\{\lambda \,:\, \underbrace{\lambda I - A \in \mathcal{S}_+^n}_{\text{LMI in } \lambda}\Big\}.$$

Applications

LMIs and SDP (frequently) allow to

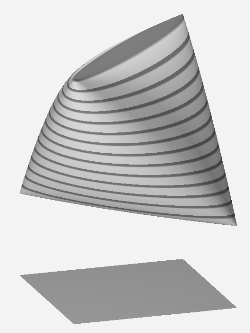
- convexify non-convex problems of algebraic nature
- solve optimization problems over semi-algebraic convex sets

Application areas

- Probability and statistics (moment problems)
- Coding theory
- Systems and control theory (Lyapunov stability)
- Combinatorial optimization (max cut and more)
- Discrete packing problems
- Global optimization (the case of polynomial optimization)
- ...

Example: eliptope in dimension three

$$\begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix} \in \mathcal{S}^3_+$$



Conic and semidefinite extended formulations

Conic extended formulation

- Assume we have fixed a convex cone K and we can solve conic problems with respect to K.
- We are given a convex set C and we want to solve linear problems over C:
$$\inf \{f(x) : x \in C\},$$
where $f : \mathbb{R}^n \to \mathbb{R}$.
- We say that C has a *K-lift*, if
$$C = \pi(K \cap H),$$
where H is an affine space and π a linear map.
- Then we can use conic programming with respect to the cone K:
$$\inf \{f(x) : x \in C\} = \inf\{\underbrace{f(\pi(y))}_{\text{linear}} : \underbrace{y \in K \cap H}_{\substack{\text{conic}\\ \text{constraint}}}\}.$$

Semidefinite extended formulation

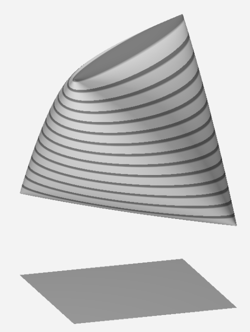
- We say that C has an *extended formulation* with m LMIs of size k if C is a linear image of a spectrahedron described by m LMIs of size k.
- Geometrically, this means that C has a K-lift for

$$K = (\mathcal{S}_+^k)^m = \underbrace{\mathcal{S}_+^k \times \ldots \times \mathcal{S}_+^k}_{m \text{ times}}.$$

- Optimizing linear functions over such C gets reduced to solving SDPs with m LMIs of size k.

Example: Square

$$\begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix} \in \mathcal{S}^3_+$$



The image under $\pi(x_1, x_2, x_3) = (x_1, x_2)$ of this spectrahedron is the square

$$[-1, 1]^2$$

Example: l_4-disc

Condition

$$x_1^4 + x_2^4 \le 1$$

can be lifted to a system of three conditions

$$\begin{array}{ccc} y_1^2 + y_2^2 \le 1, & x_1^2 \le y_1, & x_2^2 \le y_2. \\ \Updownarrow & \Updownarrow & \Updownarrow \\ \begin{pmatrix} 1 - y_1 & y_2 \\ y_2 & 1 + y_1 \end{pmatrix} \in \mathcal{S}_+^2 & \begin{pmatrix} y_1 & x_1 \\ x_1 & 1 \end{pmatrix} \in \mathcal{S}_+^2 & \begin{pmatrix} y_2 & x_2 \\ x_2 & 1 \end{pmatrix} \in \mathcal{S}_+^2 \end{array}$$

Semidefinite extension complexity

Definition

sxc(C), the *semidefinite extension complexity* of C, is the smallest k such that C has a semidefinite extended formulation with one LMI of size k.

Why **one** LMI?

$$A_1(x), \ldots, A_m(x) \in \mathcal{S}_+^k \quad \iff \quad \begin{pmatrix} A_1(x) & & \\ & \ddots & \\ & & A_m(x) \end{pmatrix} \in \mathcal{S}_+^{km}$$

- However, k and m have a different impact on running time in interior-point methods: k has a stronger influence. So, we suggest...

Semidefinite extension degree

Definition

sxdeg(C), the *semidefinite extension degree* of C, is the smallest k such that C has a semidefinite extended formulation with finitely many LMIs of size k.

Remember:

- size of the LMIs is more critical than their number. When the size is too large, interior-point methods get stuck!

Semidefinite extended formulations in polynomial optimization

Polynomial optimization (POP)

- Let $\mathbb{R}[x]$ be the ring of n-variate polynomials in variables $x_1, \ldots, x_n$ with coefficients in $\mathbb{R}$.
- *Constrained polynomial optimization (C-POP)*:

$$\inf \{ f(x) \, : \, x \in \mathbb{R}^n, \; g_1(x) \geq 0, \ldots, g_s(x) \geq 0 \} ,$$

  where $f, g_1, \ldots, g_s \in \mathbb{R}[x]$.
- *Unconstrained polynomial optimization (U-POP)*:

$$\inf \{ f(x) \, : \, x \in \mathbb{R}^n \} ,$$

  where $f \in \mathbb{R}[x]$.
- Both are hard problems in general, because nonnegativity of a polynomial is hard to decide!

Non-negativity vs. sum-of-squares property

- A polynomial $f \in \mathbb{R}[x]$ is called *sum of squares* (*SOS*) if

$$f = f_1^2 + \cdots + f_r^2$$

holds for finitely many polynomials $f_1, \ldots, f_r \in \mathbb{R}[x]$.

Convex cones in polynomial optimization

- Consider
$$\mathbb{R}[x]_d := \{f \in \mathbb{R}[x] \,:\, \deg f \leq d\},$$
where $\dim_{\mathbb{R}} \mathbb{R}[x] = \binom{n+d}{n}$.
- SOS cones:
$$\Sigma_{n,2d} := \{f \in \mathbb{R}[x]_{2d} \,:\, f \text{ is SOS}\} \subseteq \mathbb{R}[x]_{n,2d}.$$
- Cones of general non-negative polynomials:
$$P_{n,2d} := \{f \in \mathbb{R}[x]_{2d} \,:\, f \geq 0 \text{ on } \mathbb{R}^n\},$$
$$P_{n,2d}(X) := \{f \in \mathbb{R}[x]_{2d} \,:\, f \geq 0 \text{ on } X\} \qquad (X \subseteq \mathbb{R}^n).$$
- $\Sigma_{n,2d}$ is computationally much more tractable than $P_{n,2d}$ and $P_{n,2d}(X)$ in general, so frequently one uses SOS cones or some cones built up using SOS cones as substitutes for $P_{n,2d}$ and $P_{n,2d}(X)$.

SOS cones in U-POP

Let $f \in \mathbb{R}[x]_{n,2d}$.
Minimization as a search for the largest lower bound:

$$\inf\{f(x) : x \in \mathbb{R}^n\} = \sup\{\lambda \in \mathbb{R} : f - \lambda \in P_{n,2d}\}$$

This problem is too hard, but we can make it simpler:

$$\inf\{f(x) : x \in \mathbb{R}^n\} \geq \underbrace{\sup\{\lambda \in \mathbb{R} : f - \lambda \in \Sigma_{n,2d}\}}_{\text{so-called SOS relaxation}}$$

SOS cones in C-POP

Let $f, g_1, \dots, g_m \in \mathbb{R}[x]_{n,2d}$ and let

$$X := \{x \in \mathbb{R}^n \,:\, g_1(x) \geq 0, \dots, g_m(x) \geq 0\}.$$

Minimization as a search for the largest lower bound:

$$\inf\{f(x) \,:\, x \in X\} = \sup\{\lambda \in \mathbb{R} \,:\, f - \lambda \in P_{n,2d}(X)\}$$

This problem is too hard, but we can make it simpler:

$$\underbrace{\inf\{f(x) \,:\, x \in \mathbb{R}^n\} \geq \sup\{\lambda \in \mathbb{R} \,:\, f - \lambda \in \Sigma_{n,2d_0} + g_1\Sigma_{n,2d_1} + \cdots + g_m\Sigma_{n,2d_m}\}}_{\text{so-called hierarchy of SOS relaxations}}$$

Results in Real Algebra tell us that **qualitatively** this approach works (Positivstellensätze)

SOS relaxations and SDP

- **The key computational observation:** SOS relaxations can be formulated as SDPs.
- For example, the SOS relaxation for U-POP can be formulated as an SDP with 1 LMI of size $\binom{n+d}{n}$.

Semidefinite extended formulation of the SOS cone

- Consider
  - $v_{n,d}$, the vector of all monomials of degree $\leq d$ in n variables.
  - The linear bijection

    $$Y \in \mathcal{S}^k := \{k \times k \text{ symmetric matrices}\} \quad \mapsto \quad q_Y(u) := u^\top Y u$$

    between symmetric matrices and quadratic forms.
  - The linear map $\pi : \mathcal{S}^k \to \mathbb{R}[x]_{n,2d}$

    $$\pi(Y) := q_Y(v_{n,d})$$

    with

    $$k = \binom{n+d}{n}$$

- This gives a lifted representation of $\Sigma_{n,2d}$:

$$\begin{aligned}\pi(\mathcal{S}^k_+) &= \left\{q_Y(v_{n,d}) \;:\; Y \in \mathcal{S}^k_+\right\} \\ &= \Sigma_{n,2d}.\end{aligned}$$

Example: SDP formulation of $\Sigma_{1,4}$

$$f = f_0 + f_1 x + f_2 x + f_2 x^2 + f_3 x^3 + f_4 x^4 \in \mathbb{R}[x]_4$$

Identification:

$$f \in \mathbb{R}[x]_4 \longleftrightarrow \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} \in \mathbb{R}^5.$$

Consider the condition

$$\underbrace{\begin{pmatrix} 1 & x & x^2 \end{pmatrix}}_{v_{1,2}^\top} \underbrace{\begin{pmatrix} y_{00} & y_{01} & y_{02} \\ y_{01} & y_{11} & y_{12} \\ y_{02} & y_{12} & y_{22} \end{pmatrix}}_{Y} \underbrace{\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}}_{v_{1,2}} = f(x)$$

Example: SDP formulation of $\Sigma_{1,4}$

This gives a formulation of $\Sigma_{1,4}$.
Linear map $\pi : Y \mapsto f$:

$$\begin{aligned} y_{00} &= f_0 \\ 2y_{01} &= f_1 \\ y_{11} + 2y_{02} &= f_2 \\ 2y_{12} &= f_3 \\ y_{22} &= f_4 \end{aligned}$$

Linear matrix inequality:

$$\begin{pmatrix} y_{00} & y_{01} & y_{02} \\ y_{01} & y_{11} & y_{12} \\ y_{02} & y_{12} & y_{22} \end{pmatrix} \in \mathcal{S}_+^3$$

Example: SOS relaxation of U-POP with $n = 1, d = 2$

Maximize λ

for $\lambda, \; y_{00}, \; y_{01}, \; y_{02}, \; y_{11}, \; y_{12}, \; y_{22} \in \mathbb{R}$

subject to:

$$\begin{aligned} y_{00} + \lambda &= f_0 \\ 2y_{01} &= f_1 \\ y_{11} + 2y_{02} &= f_2 \\ 2y_{12} &= f_3 \\ y_{22} &= f_4 \end{aligned}$$

$$\begin{pmatrix} y_{00} & y_{01} & y_{02} \\ y_{01} & y_{11} & y_{12} \\ y_{02} & y_{12} & y_{22} \end{pmatrix} \in \mathcal{S}_+^3$$

Example: SOS relaxation of U-POP with $n = 1, d = 4$

If you want to have just one LMI without linear equations:

$$\begin{aligned} &\text{Maximize } f_0 - y_{00} \\ &\text{for } y_{00}, y_{11} \in \mathbb{R} \\ &\text{subject to:} \\ &\begin{pmatrix} y_{00} & \frac{1}{2} f_0 & \frac{1}{2}(f_2 - y_{11}) \\ \frac{1}{2} f_0 & y_{11} & \frac{1}{2} f_3 \\ \frac{1}{2}(f_2 - y_{11}) & \frac{1}{2} f_3 & f_4 \end{pmatrix} \in \mathcal{S}^3_+ \end{aligned}$$

This is an SDP with 1 LMI of size 3 on two variables.

New results

Combinatorial tool to bounding sxdeg(C) from below

Theorem A (A. SIAGA 2019)

Let $X \subseteq \mathbb{R}^n$ be a set with non-empty interior. Let $C \subseteq P_{n,2d}(X)$ be a closed convex cone such that there exist finite subsets S of X of arbitrarily large cardinality with the following property:

($*$) For every k-element subset T of S, some polynomial f in the cone C is equal to zero on T and is strictly positive on $S \setminus T$.

Then $\mathrm{sxdeg}(C) > k$.

Between $\Sigma_{n,2d}$ and $P_{n,2d}(X)$

More generally, if we want to approximate the non-negativity cone $P_{n,2d}$ from inside by cone that contains $\Sigma_{n,2d}$, we would have to pay a high price:

Corollary B

Let $X \subseteq \mathbb{R}^n$ be a set with non-empty interior and C be a closed convex cone satisfying $\Sigma_{n,2d} \subseteq C \subseteq P_{n,2d}(X)$. Then

$$\operatorname{sxdeg}(C) \geq \binom{n+d}{n}.$$

- This means, whenever we use a cone that contains $\Sigma_{n,2d}$ as a subset, we are forced to use LMIs of a large size.

All about $\Sigma_{n,2d}$

Corollary C

$\mathrm{sxdeg}(\Sigma_{n,2d}) = \mathrm{sxc}(\Sigma_{n,2d}) = \binom{n+d}{n}$.

- In other words, $\Sigma_{n,2d}$ has a semidefinite extended formulation with 1 LMI of size $\binom{n+d}{n}$ ...
- **but** has no semidefinite extended formulation with LMIs of a smaller size, no matter how many LMIs are used!
- **Conclusion**: If one wants to describe $\Sigma_{n,2d}$ exactly in the context of semidefinite programming, one is forced to use LMIs of large size.

- Note that $\mathcal{S}_+^k \simeq \Sigma_{k-1,2}$
- So, as a direct consequence of Theorem A we also obtain ...

Corollary D

$\mathrm{sxdeg}(\mathcal{S}_+^k) = k$

- Growth of the expressive power. Hierarchy:

$$\mathrm{SDR}(k) := \{S \subseteq \mathbb{R}^n \ : \ n \in \mathbb{N},\ \mathrm{sxdeg}(S) \leq k\}$$

satisfies

$$\underbrace{\mathrm{SDR}(1)}_{\text{polyhedra}} \subsetneq \underbrace{\mathrm{SDR}(2)}_{\text{SOC representable sets}} \subsetneq \mathrm{SDR}(3) \subsetneq \mathrm{SDR}(4) \subsetneq \cdots$$

Known cases of above results

- The presented formulas have been known only in a few cases (Fawzi 2018):
  - $\mathrm{sxdeg}(\mathcal{S}_+^3) = 3$ (Fawzi 2018)
  - $\mathrm{sxdeg}(\Sigma_{1,4}) = 3$ (Fawzi 2018)
- $\mathrm{sxdeg}(\Sigma_{1,4}) = 3$ is also mentioned by Ahmadi & Hall & Papachristodoulou & Saunderson & Zheng 2017.

Moment cones

We introduce the *moment cones*

$$\begin{aligned} M_{n,2d} &:= \overline{\text{cone}}(\{v_{n,2d}(x) \,:\, x \in \mathbb{R}^n\}), \\ M_{n,2d}(X) &:= \overline{\text{cone}}(\{v_{n,2d}(x) \,:\, x \in X\}) \qquad\qquad (X \subseteq \mathbb{R}^n). \end{aligned}$$

($\overline{\text{cone}}$ - topological closure of the convex conic hull)

$$\begin{array}{lcl} M_{n,2d} & \text{dual to} & P_{n,2d}, \\ M_{n,2d}(X) & \text{dual to} & P_{n,2d}(X). \end{array}$$

Moment cones occur in the context of probability.

Corollary for moment cones

Corollary E

For every $X \subseteq \mathbb{R}^n$ with non-empty interior,

$$\operatorname{sxdeg}(M_{n,2d}(X)) \geq \binom{n+d}{n}.$$

We know even more

Theorem (Claus Scheiderer 2018)

For every $X \subseteq \mathbb{R}^n$ with non-empty interior and $n, d \geq 2$, $(n, d) \neq (2, 2)$,

$$\operatorname{sxdeg}(M_{n,2d}(X)) = \infty.$$

Proof strategy

- **Tool 1 (convex optimization):** Results of Gouveia & Parrilo & Thomas 2013 relating
  - the existence of K-lifts for C to
  - the existence of K-factorizations of the slack matrix of C.
- **Tool 2 (combinatorics):** Ramsey theorem for hypergraphs.

Tool from convex optimization

Some Euclidean spaces

- $\mathbb{R}^n$:

$$\langle x \, , \, y \rangle := x^\top y.$$

- $\mathcal{S}^k$:

$$\left\langle (a_{ij}) \, , \, (b_{ij}) \right\rangle = \sum_{i,j} a_{ij} b_{ij}.$$

- $(\mathcal{S}^k)^m$:

$$\langle (A_1, \ldots, A_m) \, , \, (B_1, \ldots, B_m) \rangle = \langle A_1 \, , \, B_1 \rangle + \cdots + \langle A_m \, , \, B_m \rangle$$

Conic duality and the slack matrix

- If C is a convex cone in Euclidean space V, then

$$C^* = \{u \in V \,:\, \langle u\,,\, x\rangle \geq 0 \text{ for all } x \in C\}$$

  is the *dual cone* of C. (The dual cone captures all linear inequalities valid on C.)
- Note that $\mathcal{S}^k_+$ is self-dual: $(\mathcal{S}^k_+)^* = \mathcal{S}^k_+$.
- Consequently, $(\mathcal{S}^k_+)^m$ is self-dual too.
- We define the *slack matrix* of C as the $C \times C^*$ matrix

$$\big(\langle u\,,\, x\rangle\big)_{u\in C^*,\ x\in C}$$

  The slack matrix keeps the track of evaluations of all linear functionals non-negative on C at all points of C.

Conic factorization of the slack matrix

Slack-matrix theorem (Gouveia & Parrilo & Thomas 2013)

Let C be a closed convex cone. Then the following conditions are equivalent:

(i) C has a semidefinite extended formulation with m LMIs of size k.

(ii) The slack matrix is factorisable as

$$\Big(\langle u\,,\,x\rangle\Big)_{u\in C^*,\;x\in C}=\Big(\langle A_u\,,\,B_x\rangle\Big)_{u\in C^*,x\in C},$$

where $A_u, B_x \in (\mathcal{S}_+^k)^m$.

Combinatorial tool

Hypergraphs and colorings of hyperedges

- $G = (V, E)$ with
$$E \subseteq \binom{V}{t} := \{t\text{-element subsets of } V\}.$$
is a *t-uniform hypergraph* with set V of nodes and a set E of hyperedges.
- $t = 2 \rightsquigarrow$ a graph.
- $G = (V, E)$ with $E = \binom{V}{t}$ is called *complete*.
- We can color hyperedges.

Ramsey theorem

Theodore Motzkin: “Complete disorder is impossible”

Ramsey theorem

Theodore Motzkin: "Complete disorder is impossible"

Ramsey theorem for hyperpgraphs

For $t, n, c \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that every coloring of a complete t-uniform hypergraph on N nodes contains a monochromatic copy of a complete t-uniform hypergraph on n nodes.

Ramsey theorem

Theodore Motzkin: "Complete disorder is impossible"

Ramsey theorem for hyperpgraphs

For $t, n, c \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that every coloring of a complete t-uniform hypergraph on N nodes contains a monochromatic copy of a complete t-uniform hypergraph on n nodes.

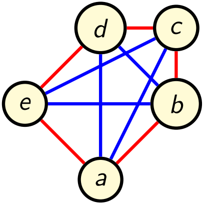
$$(t, n, c) \rightsquigarrow N$$

Definition

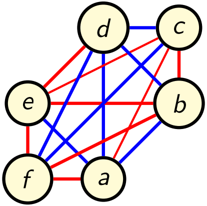
The least possible N in Ramsey theorem is called the *Ramsey number* and denote by $R_t(n; c)$.

Example: Ramsey Theorem

- Input: $t = 2$ (we are dealing with graphs), $n = 3$ (we are interested in monochromatic triangles), $c = 2$ (two colors).
- Output: $N = 6$. That is, if we color a complete graph on 6 nodes with two colors, there we will always have a monochromatic triangle.



no monochromatic triangles!



there is a monochromatic triangles!
(no matter how you 2-color)

Proof sketch

Proof ideas

- We prove
$$\mathrm{sxdeg}(\Sigma_{1,2d}) \geq d+1,$$
to illustrate the proof idea of Theorem A.
- Assuming that $\Sigma_{1,2d}$ has a semidefinite extended formulation with m LMIs of size d, we will arrive at a contradiction.
- We are going to use the slack-matrix theorem for
$$C = \Sigma_{1,2d}.$$

Considering $\Sigma^*_{1,2d}$

We can think of the elements of $\mathbb{R}[x]_{2d}$ as vectors from $\mathbb{R}^{2d+1}$. For example,

$$f(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + f_4 x^4 \in \mathbb{R}[x]_4 \longleftrightarrow \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} \in \mathbb{R}^5$$

With this in mind, we can introduce $\Sigma^*_{1,2d}$, the dual cone of $\Sigma_{n,2d}$.

Some elements of $\Sigma^*_{n,2d}$

Note that evaluation of $f \in \mathbb{R}[x]_{1,2d}$ at $x \in \mathbb{R}^n$ can be written as the scalar product For example,

$$f = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + f_4 x^4 = \left\langle \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix}, \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} \right\rangle$$

In particular, we see that $\begin{pmatrix} 1 \\ x^1 \\ \vdots \\ x^{2d} \end{pmatrix} \in \Sigma^*_{n,2d}$ for each $x \in \mathbb{R}^n$ because $f(x) \geq 0$ for $f \in \Sigma_{1,2d}$.

Sub-matrix of the slack matrix of $\Sigma_{n,2d}$

We thus conclude that the slack matrix of $\Sigma_{n,2d}$ contains the sub-matrix

$$\left(\ f(x)\ \right)_{f\in\Sigma_{1,2d},\ x\in\mathbb{R}}$$

That is the sub-matrix of the evaluations of SOS polynomials at points of $\mathbb{R}$.

Large finite sub-matrix of the slack matrix

We don't need the whole matrix but only a sufficiently large structured finite sub-matrix. Choose N distinct elements of $\mathbb{R}$, with N sufficiently large. For example, we can use the points $[N] = \{1, \ldots, N\}$. For each $T \in \binom{[N]}{d}$, we introduce the polynomial

$$f_T(x) = \left(\prod_{t \in T} (x - t) \right)^2 \in \Sigma_{1,2d}.$$

Zero pattern of the sub-matrix

We are going to use the sub-matrix

$$\left(\ f_T(s)\ \right)_{T\in\binom{[N]}{d},\ s\in[N]}$$

of the slack matrix.
Its zero pattern:

$$f_T(s)\begin{cases}=0 & s\in T,\\ >0 & s\notin T.\end{cases}$$

For example, for $N = 4$ and $d = 2$:

$$\begin{matrix}\{1,2\}\\ \{1,3\}\\ \{1,4\}\\ \{2,3\}\\ \{2,4\}\\ \{3,4\}\end{matrix}\qquad\begin{pmatrix}0 & 0 & + & +\\ 0 & + & 0 & +\\ 0 & + & + & 0\\ + & 0 & 0 & +\\ + & 0 & + & 0\\ + & + & 0 & 0\end{pmatrix}$$

Invoking slack-matrix theorem

By the slack-matrix theorem,

$$\left(\ f_T(s)\ \right)_{T \in \binom{[N]}{d},\ s \in [N]}$$

is factorizable with respect to the cone $(\mathcal{S}^d_+)^m$:

$$f_T(s) = \langle A_T\ ,\ B_s \rangle$$

holds for some

$$A_T = (A_{T,1}, \ldots, A_{T,m}) \in (\mathcal{S}^d_+)^m$$

and

$$B_s = (B_{s,1}, \ldots, B_{s,m}) \in (\mathcal{S}^d_+)^m.$$

Orthogonality of PSD matrices

Observation

For $F, G \in \mathcal{S}_+^k$:

$$\langle F\ ,\ G \rangle = 0 \qquad \Longleftrightarrow \qquad \mathrm{im}(F) \perp \mathrm{im}(G).$$

So, we do not need to know $A_{T,i}$ and $B_{s,i}$ exactly to check $\langle A_T\ ,\ B_s \rangle = 0$, it suffices to know the images.

Coloring sets T by m-tuples of dimensions

Consider

$$U_{T,i} := \sum_{t \in T} \mathrm{im}(B_{t,i}) \tag{1}$$

and

$$d_{T,i} := \dim(U_{T,i}).$$

Note that

$$\mathrm{im}(A_{T,i}) \perp U_{T,i}.$$

$\{0, \ldots, d\}^m$	our set of $(d+1)^m$ colors
$(d_{T,1}, \ldots, d_{T,m})$	the color of $T \in \binom{[N]}{d}$

Invoking Ramsey theorem

- By Ramsey theorem, if N is large enough, $\binom{W}{d}$ is monochromatic, for some $W \subseteq [N]$ with $|W| = d + 1$.
- $\Rightarrow$ All elements $T \in \binom{W}{d}$ have the same color $(d_1, \ldots, d_m) \in \{0, \ldots, d\}^m$.
- $\Rightarrow \dim(U_{T,i}) = d_i$ does not depend on $T \in \binom{W}{d}$
- By elementary linear algebra, $U_{T,i}$ does not depend of $T \in \binom{W}{d}$. This means:

Claim

For some vector spaces $U_1, \ldots, U_m$, one has $U_{T,i} = U_i$ for all $i \in [k]$ and $T \in \binom{W}{k}$.

Concluding the proof

- Since $|W| = d + 1$, we can choose an arbitrary decomposition $W = T \cup \{s\}$, where $T \in \binom{W}{d}$.
- In view of $0 = f_T(t) = \langle A_T, B_t \rangle$ for $t \in T$, we have $\langle A_{T,i}, B_{t,i} \rangle = 0$ for all $t \in T$ and $i \in [m]$.
- By the Observation, $\mathrm{im}(A_{T,i})$ is orthogonal to $\mathrm{im}(B_{t,i})$. Hence, $\mathrm{im}(A_{T,i})$ is orthogonal to $\sum_{t \in T} \mathrm{im}(B_{t,i}) = U_i$.
- By the choice of U_i, the linear space U_i contains all $\mathrm{im}(B_{w,i})$ with $w \in W$ as a subspace.
- Hence, $\mathrm{im}(A_{T,i})$ is orthogonal to $\mathrm{im}(B_{s,i})$.
- By the Observation, this means that $\langle A_{T,i}, B_{s,i} \rangle = 0$ holds for all $i \in [m]$.
- Thus, we have shown $\langle A_T, B_s \rangle = 0$. Since $s \notin T$, this contradicts $\langle A_T, B_s \rangle = f_T(s) > 0$.

Summary and outlook

Summary

Technology:

$$\underbrace{P_{n,2d}(X)}_{\text{intractable}} \xrightarrow{\text{'approximation'}} C \overset{\text{lifting}}{=\!=\!=} \pi\Big(\underbrace{H \cap K}_{\substack{\text{conic feasibility}\\ \text{set w.r.t. } K}} \Big) \xrightarrow{\text{solving a } K\text{-conic problem}}$$

Implementation of this template requires answering the questions:

- How to approximate? $\Longleftrightarrow$ How to choose C?
- What convex programming to use? Linear? Second-order cone? Semidefinite? Anything else? $\Longleftrightarrow$ How to choose K?
- How to lift? $\Longleftrightarrow$ What conic formulation of C to use?
- Which method to use for solving (large) K-conic problem?

Outlook

Lasserre hierarchies:

- C: an SOS cone or a truncated quadratic module
- $K = (\mathcal{S}_+^k)^m$ (semidefinite programming).

Alternatives:

- Other choices of C and K have also been suggested are being studied.
- More good choices of C and K?
- How to solve (large) conic and semidefinite problems?

Cones that are being studied/used:

- DSOS and SDSOS (Ahmadi, Majumdar 2019)
- SONC (De Wolff, Dressler, Iliman, Theobalad, 2016–)
- SAGE and relative entropy optimization (Chandrasekaran, Shah 2014)
- Sums of structured sparse SOS cones (A., Peters, Sager, in progress)

Enforcing small LMIs makes computations more tractable (numerical studies by A. & Peters & Sager 2019+). Optimization of a n-variate polynomials of degree $2d$ over a box. A snapshot of numerical evaluations:

	$n = 35, 2d = 4$	$n = 40, 2d = 4$
standard SOS	a few LMIs of size ≈ 700 ≈ 5.5 hours!	a few LMIs of size ≈ 800 SDP solver fails!
sparse SOS	lots of LMIs of size ≈ 40 15 seconds!	lots of LMIs of size ≈ 40 20 seconds!

Bounds of sparse SOS only slightly worse than the bounds of standard SOS.

Recent results:

- J. Saunderson (2020): A generalization of the obstruction from $\mathcal{S}_+^k$-lifts to general K-lifts.
- C. Scheiderer (2021): Every planar convex semi-algebraic set C satisfies $\mathrm{sxdeg}(C) \leq 2$.